

LESSICO MULTILINGUE DEI BENI CULTURALI

# Informatica umanistica

Risorse e strumenti per lo studio  
del lessico dei beni culturali

a cura di  
VALERIA ZOTTI  
ANA PANO ALAMÁN



STRUMENTI  
PER LA DIDATTICA E LA RICERCA

– 193 –

## LESSICO MULTILINGUE DEI BENI CULTURALI

### *Comitato Scientifico*

Annick Farina, Direttrice (*Università di Firenze*)  
Christina Samson, Direttrice (*Università di Firenze*)  
Sabrina Ballestracci (*Università di Firenze*)  
Marco Biffi (*Università di Firenze e Accademia della Crusca*)  
Elena Carpi (*Università di Pisa*)  
Dave Coniam (*University of Hong Kong*)  
Christina Dechamps (*Universidade Nova de Lisboa*)  
Isabella Gagliardi (*Università di Firenze*)  
Marcello Garzaniti (*Università di Firenze*)  
Paul Geyer (*Universität Bonn*)  
Donata Levi (*Università di Udine*)  
Valentina Pedone (*Università di Firenze*)  
Federica Rossi (*Kunsthistorisches Institut di Firenze*)  
Geoffrey Williams (*Université de Bretagne Sud*)

### *Comitato scientifico dell'opera*

Silvia Cacchiani (*Università di Modena e Reggio Emilia*), Elena Carpi (*Università di Pisa*), Francesca Chessa (*Università di Cagliari*), Isabella Chiari (*Università di Roma "La Sapienza"*), Dave Coniam (*University of Hong Kong*), Cosimo De Giovanni (*Università di Cagliari*), Marcello Garzaniti (*Università di Firenze*), Nicole Maroger (*Università di Firenze*), Marie-France Merger (*Università di Pisa*), Carlota Nicolás (*Università di Firenze*), Sara Radighieri (*Università di Modena e Reggio Emilia*), Rachele Raus (*Università di Torino*), Lorella Sini (*Università di Pisa*),  
Geoffrey Williams (*Université de Bretagne Sud*)

### *Titoli pubblicati*

Raus R., Cappelli G., Flinz C. (édité par), *Le guide touristique: lieu de rencontre entre lexique et images du patrimoine culturel. Vol. II*  
Zotti V., Pano Alamán A. (a cura di), *Informatica umanistica. Risorse e strumenti per lo studio del lessico dei beni culturali*

# **Informatica umanistica**

Risorse e strumenti per lo studio del lessico  
dei beni culturali

*a cura di*

Valeria Zotti e Ana Pano Alamán

Firenze University Press

2017

Informatica umanistica : risorse e strumenti per lo studio del lessico dei beni culturali / a cura di Valeria Zotti, Ana Pano Alamán. – Firenze : Firenze University Press, 2017. (Strumenti per la didattica e la ricerca ; 193)

<http://digital.casalini.it/9788864535463>

ISBN 978-88-6453-545-6 (print)  
ISBN 978-88-6453-546-3 (online)

Progetto grafico di Alberto Pizarro Fernández, Pagina Maestra snc  
Immagine di copertina: © Dwnld777 | Dreamstime

*Certificazione scientifica delle Opere*

Tutti i volumi pubblicati sono soggetti ad un processo di referaggio esterno di cui sono responsabili il Consiglio editoriale della FUP e i Consigli scientifici delle singole collane. Le opere pubblicate nel catalogo della FUP sono valutate e approvate dal Consiglio editoriale della casa editrice. Per una descrizione più analitica del processo di referaggio si rimanda ai documenti ufficiali pubblicati sul catalogo on-line della casa editrice ([www.fupress.com](http://www.fupress.com)).

*Consiglio editoriale Firenze University Press*

A. Dolfi (Presidente), M. Boddi, A. Bucelli, R. Casalbuoni, M. Garzaniti, M.C. Grisolia, P. Guarnieri, R. Lanfredini, A. Lenzi, P. Lo Nostro, G. Mari, A. Mariani, P.M. Mariano, S. Marinai, R. Minuti, P. Nanni, G. Nigro, A. Perulli, M.C. Torricelli.

La presente opera è rilasciata nei termini della licenza Creative Commons Attribution 4.0 International (CC BY 4.0: <https://creativecommons.org/licenses/by/4.0/legalcode>).

This book is printed on acid-free paper

CC 2017 Firenze University Press  
Università degli Studi di Firenze  
Firenze University Press  
via Cittadella, 7, 50144 Firenze, Italy  
[www.fupress.com](http://www.fupress.com)  
*Printed in Italy*

INTRODUZIONE <i>Valeria Zotti, Ana Pano Alamán</i>	7
GUIDEBOOKS OF FLORENCE FOR A SPECIALISED LEXICAL DATABASE. A CORPUS-DRIVEN LINGUISTIC ANALYSIS <i>Christina Samson</i>	17
ON THE LANGUAGE OF FLORENCE ART MUSEUM WEBSITES: THE ITALIAN TEXTS OF THE «VIRTUAL TOUR» <i>Giuliana Diani</i>	33
TOURISME CULTUREL SUR INTERNET. LES NOMS PROPRES DES ÉDITIONS ORIGINALES DE RABELAIS <i>Denis Maurel, Nathalie Friburger, Iris Eshkol-Taravella</i>	47
VALORIZZARE GLI SCRITTI DI LEONARDO DA VINCI PER MEZZO DELLE NUOVE TECNOLOGIE: L'ARCHIVIO DIGITALE E-LEO <i>Monica Taddei</i>	67
LE VOYAGE EN FRANCE DU PRINCE DE MACHIAVEL. L'OUTIL HYPERMACHIAVEL ET SES EFFETS DE SENS <i>Jean-Claude Zancarini, Séverine Gedzelman</i>	83
L'INTEGRAZIONE DI CORPORA PARALLELI DI TRADUZIONE ALLA DESCRIZIONE LESSICOGRAFICA DELLA LINGUA DELL'ARTE: L'ESEMPIO DELLE TRADUZIONI FRANCESI DELLE VITE DI VASARI <i>Valeria Zotti</i>	105

6 Informática umanistica

WIKIPEDIA: POSIBILIDADES Y LÍMITES PARA LA EXTRACCIÓN DE TERMINOLOGÍA MULTILINGÜE SOBRE EL ARTE <i>Ana Pano Alamán</i>	135
L'INFORMAZIONE DIGITALE E IL WEB SEMANTICO. IL CASO DELLE SCHOLARLY DIGITAL EDITIONS <i>Francesca Tomasi</i>	157
NOTE SUGLI AUTORI	175

V. Zotti  
A. Pano Alamán

## Introduzione

L'informatica umanistica è un campo di studio in continua evoluzione<sup>1</sup>. Del termine esistono molteplici definizioni che dipendono dal posizionamento di chi lo circoscrive in una determinata area o dal modo in cui si intende la stessa disciplina, più orientata verso le tecnologie e le metodologie di indagine, oppure rivolta 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). In senso stretto, alcuni studiosi la definiscono come uno spazio di creazione di strumenti informatici e di risorse digitali a disposizione dei ricercatori (Rieger 2010). E mentre per Svensson (2010) è fondamentalmente un'intersezione tra le scienze umanistiche e le tecnologie dell'informazione, Carter l'associa all'evoluzione degli studi umanistici tradizionali, «making use of the tools of the day in order to study and explore the limits of the human condition» (2013: XI).

D'altra parte, il concetto stesso di 'informatica umanistica' o *humanities computing*, in inglese, sembra essere limitato o non più adeguato per designare questa disciplina in costante trasformazione<sup>2</sup>. Se con questo sintagma

<sup>1</sup> Per una panoramica completa si veda Schreibman, Siemens, Unsworth (2004). Il rapido sviluppo e il crescente interesse verso la disciplina da parte dei ricercatori si manifestano attraverso la creazione di molteplici associazioni nazionali ed internazionali di settore, ad esempio: The Association for Computers and the Humanities, <<http://www.ach.org>>; The Alliance of Digital Humanities, <<http://www.adho.org>>; in Italia, l'Associazione per l'Informatica Umanistica e la Cultura Digitale, <<http://www.umanisticadigitale.it/>>; e, rispettivamente in ambito ispanofono e francofono, le associazioni Humanidades Digitales Hispánicas, <<http://www.humanidadesdigitales.org>> e Humanistica. Association francophone des humanités numériques, <<http://www.humanisti.ca/>>. Vanno segnalate anche le numerose riviste dedicate alla disciplina; a questo proposito vedasi la sezione ad essa dedicata nel sito dell'European Association for Digital Humanities: <<http://eadh.org/publications/all>>.

<sup>2</sup> Cf. la riflessione di Spence (2014: 39) in merito alla questione terminologica, al carattere collaborativo delle *digital humanities* (p. 40) e allo stretto rapporto tra sviluppo tecnologico ed evoluzione dell'agenda scientifica al loro interno (p. 38).



si faceva riferimento all'insieme di ricerche che sancivano l'unione della ricerca umanistica e i computer – siamo nei primi anni dell'applicazione di software all'analisi stilometrica di testi –, oggi si privilegia il concetto di *digital humanities*, *humanités numériques* in francese o *humanidades digitales* in spagnolo, per riferirsi a una realtà più vasta che riguarda non solo le metodologie di costituzione 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.

Indipendentemente dal termine utilizzato e senza entrare nel dibattito sulla corrispondenza o meno tra i concetti di 'digital humanities' e 'informatica umanistica', è evidente che oggi l'informatica e il paradigma digitale arricchiscono e potenziano la ricerca e l'insegnamento nelle scienze umanistiche. Tutti gli studiosi concordano sul fatto che «it is through the digital humanities that culture can be understood differently as a result of the digitation process» (Carter 2013: XI). In altre parole, l'incontro tra informatica e scienze umane permette di avvicinarsi alla cultura in modo inedito e di aprire alle nuove generazioni di ricercatori, docenti e studenti la possibilità di gestire facilmente una molteplicità di strumenti messi a loro disposizione per lo studio e per la diffusione dei risultati del loro lavoro.

Per quanto riguarda la ricerca, non si può negare che il computer sia un grande alleato degli umanisti e che abbia portato con sé una rivoluzione nelle loro pratiche e metodologie quotidiane. Alcuni affermano che lo sviluppo di programmi informatici sempre più sofisticati, l'espansione e la trasformazione di Internet, così come la digitalizzazione generalizzata di testi scritti e orali, hanno avuto e stanno avendo un impatto simile a quello della stampa nel Rinascimento, nella misura in cui inaugurano una nuova era scientifica per le scienze umane e sociali (Brossaud, Reber 2007: 17). Già negli anni Novanta, Marcos Marín affermava che il computer stava cambiando «la naturaleza y el valor de la comunicación en dimensiones más profundas que la imprenta o el tubo de rayos catódico. [...] Los estudiosos del ser humano como ser individual y, por la comunicación, social, los humanistas, no han permanecido al margen de esa innovación» (1994: 7). Per gli studiosi delle scienze umane, infatti, sono oggi disponibili numerosi linguaggi di programmazione ed applicazioni che consentono loro di condurre ricerche che fino a pochi anni fa non erano possibili o che richiedevano competenze informatiche avanzate. L'analisi di un ampio numero di testi nei campi della linguistica, la letteratura, la filosofia, la storia o l'arte dipendeva dalle conoscenze e dalle capacità dei ricercatori che dovevano essere competenti non solo nelle proprie aree di formazione, ma anche, ad esempio, in programmazione di software o in linguaggi di codifica sia per poter creare strumenti di indagine che per usarli. Attualmente, invece, gli umani-

sti non devono necessariamente saper scrivere ‘codice’ o ideare applicazioni per sé stessi o ad uso di altri ricercatori. Hanno a disposizione strumenti di analisi e risorse digitali pronti per l’uso o anche facilmente adattabili alle proprie necessità, che permettono di svolgere ricerche in larga misura diversificate, approfondite e precise. Al giorno d’oggi, gli strumenti di comunicazione, di collaborazione e di visualizzazione di grandi quantità di dati facilitano anche la connessione tra esperti della stessa area o di aree affini e lo sviluppo di ricerche contraddistinte dall’interdisciplinarietà, dall’incrocio tra metodologie qualitative e quantitative e dallo scambio di conoscenze e di pratiche a livello globale (Marcos Marín 1994; Spence 2015). L’impatto del digitale sul lavoro degli umanisti riguarda ancora altri aspetti. Ad esempio, Davidson (2008) ha osservato che, come risultato delle possibilità collaborative e comunicative messe a disposizione dalle reti sociali virtuali, da strumenti di lavoro condiviso o da sistemi aperti di codifica di testi che possono essere implementati da chiunque, i ricercatori in scienze umane e sociali hanno oggi l’opportunità di decentrare l’autorità, mettendo l’informatica umanistica al centro di una rivoluzione accademica (Brossaud, Reber 2007: 18).

In relazione con l’insegnamento, invece, gli strumenti e le risorse esistenti, nonché quelli creati da e per gli esperti delle aree umanistiche, promuovono l’adozione di metodologie di apprendimento innovative rivolte alle nuove generazioni e non solo. Oggi comunichiamo e lavoriamo in modo diverso, siamo a nostro agio con gran parte delle piattaforme esistenti sul web e accediamo a una grande quantità di dati attraverso smartphone, computer e altri dispositivi, con diverse finalità. In questo contesto, il docente dispone di una variegata gamma di risorse utili non solo a promuovere un apprendimento basato sull’esperienza diretta, attraverso la creazione, l’implementazione di progetti applicativi e la ricerca di dati in modo efficace ed efficiente, ma anche a incoraggiare, nelle nuove generazioni di studiosi delle scienze umane, lo sviluppo di un atteggiamento critico di fronte alla complessità che il paradigma digitale porta con sé. Di fatto, nelle facoltà e nelle scuole umanistiche sono ormai numerosi i corsi di studi dedicati a formare ‘umanisti digitali’, il che ha portato diversi studiosi a ripensare la formazione dei futuri linguisti, storici o esperti in letteratura, tra gli altri, fornendo strumenti diversificati, indirizzati alla ricerca e alla didattica in questo campo (Tomasi 2008; Numerico *et al.* 2010).

Questo volume affronta dunque una realtà estremamente vasta e multiforme, della quale è possibile fornire soltanto una parziale panoramica. Per questo motivo, la presenta da una prospettiva concreta, specializzata e circoscritta allo studio del lessico dell’arte e dei beni culturali. Le ricerche e le riflessioni contenute in questo libro si inseriscono infatti all’interno del progetto di ricerca *Lessico multilingue dei Beni Culturali* (LBC), nato nel 2013 per iniziativa dell’Unità di ricerca LBC del Dipartimento di Lingue, Letterature e Studi interculturali dell’Università

di Firenze e condotto in collaborazione con diverse Università italiane e straniere, il cui obiettivo principale è la realizzazione di un dizionario plurilingue in formato elettronico del lessico dei beni culturali (Garzanti, Farina 2013; Farina 2016). I contributi del volume esplorano diverse metodologie di analisi, condotte per lo più dal punto di vista della linguistica applicata, della traduttologia e della lessicografia, e presentano risorse, strumenti e piattaforme disponibili in rete, dedicati allo studio del lessico del patrimonio culturale e del discorso sull'arte, nonché alla traduzione in più lingue di testi relativi ai beni artistici. Inoltre, le riflessioni contenute nei diversi contributi raccolti si presentano in una prospettiva volutamente plurilingue, che è propria del progetto LBC e che si manifesta sia nelle diverse lingue dei saggi presentati (inglese, italiano, francese e spagnolo) sia nella scelta di analizzare il lessico dell'arte in lingue e culture differenti.

Il volume si apre con due testi corrispondenti a quella che, secondo Schnapp e Presner (2009), è la prima fase dell'informatica umanistica<sup>3</sup>, una fase più quantitativa, incentrata sulla raccolta di dati in database di grandi dimensioni e su tipologie di analisi condotte secondo i metodi della linguistica dei corpora. I contributi di Christina Samson e di Giuliana Diani, attraverso l'approccio della *corpus-driven linguistics* e dell'analisi del discorso, prendono infatti in esame dei corpora testuali di guide turistiche on-line in lingua inglese e di siti web museali fiorentini in lingua italiana, rispettivamente, allo scopo di analizzare quantitativamente e qualitativamente il lessico dei beni culturali fiorentini e di avviare una riflessione sulla traduzione di alcuni termini in diverse lingue nell'ambito della divulgazione del patrimonio culturale italiano.

Alla prima fase dell'informatica umanistica viene anche ricondotta l'era dei primi linguaggi del web e della diffusione di linguaggi di marcatura (*markup language*), concepiti per l'annotazione o codifica formale di testi previamente digitalizzati. Su questa scia, il saggio di Denis Maurel, Nathalie Friburger e Iris Eshkol-Taravella presenta il progetto *Renom*, una proposta di annotazione, mediante il linguaggio di marcatura XML basato sullo standard della TEI (*Text Encoding Initiative*), per etichettare nomi propri di persona o di luogo all'interno di testi del Rinascimento francese. L'obiettivo del progetto è creare un portale web che permetta di associare la navigazione in opere letterarie francesi con le visite turistiche nella regione Centre, terra di Pierre de Ronsard e di François Rabelais, così come di Gargantua e Pantagruel, in modo da promuovere il turismo culturale di quest'area.

La seconda fase della disciplina è, per Schnapp e Presner, «qualitative, interpretative, experimental, emotive, [and] generative in character» (2009: 2). L'informatica umanistica sembra ora andare oltre l'analisi te-

<sup>3</sup> Sulle fasi storiche della disciplina, sul suo statuto epistemologico e sul progetto culturale che essa inaugura, vedasi Numerico e Vespignani (2003: 9-16).

stuale e l'*encoding* o codifica di testi per volgere verso nuovi paradigmi disciplinari e metodologie ibride di ricerca che producono, ad esempio, software e piattaforme di lavoro atte a interagire con diverse fonti di conoscenza e diversi tipi di dati, ovvero, i testi che risultano da un processo di digitalizzazione precedente e i testi 'nati' in ambiente digitale. Due contributi di questo volume si inseriscono pienamente in questa fase, in quanto descrivono strumenti che sono stati concepiti e realizzati sin dall'origine per analizzare dati disponibili in formato elettronico. Il primo è l'archivio digitale *e-Leo*, di cui si occupa il saggio di Monica Taddei. Si tratta di uno strumento avanzato per la fruizione e lo studio della collezione completa in formato digitale delle opere di Leonardo da Vinci, posseduta dalla Biblioteca leonardiana di Vinci. Il secondo è il software *HyperMachiavel*, descritto nel volume dai suoi ideatori e sviluppatori, Jean-Claude Zancarini e Séverine Gedzelman. Nel contesto dell'edizione critica digitale, questo programma rende possibile consultare e visualizzare online dei corpora di testi allineati e paralleli, formati per esempio da un testo e dalle sue traduzioni o da diverse edizioni dello stesso testo. *HyperMachiavel* ne favorisce anche l'esplorazione lessicale e concettuale attraverso diverse funzionalità e dispositivi intuitivi, tra cui l'annotazione semi-automatica dei testi, pensata per rispondere a rilevanti problematiche di studio nel campo della traduzione.

Questo stesso software, sviluppato originariamente per l'analisi delle traduzioni francesi de *Il Principe* di Machiavelli, è stato adottato e adattato dal gruppo di ricerca LBC alla stregua di un'altra tendenza che caratterizza l'informatica, ovvero, la riusabilità. Oggi, molti ricercatori nelle scienze umane possono sviluppare progetti usando strumenti ed applicazioni creati da esperti di area informatica o del loro stesso settore scientifico senza che sia necessario modificarli, oppure apportando piccole modifiche sulla base delle proprie necessità di ricerca. In questo volume, Valeria Zotti illustra l'applicazione delle funzioni presenti in *HyperMachiavel* all'analisi del corpus parallelo delle traduzioni francesi de *Le Vite* di Giorgio Vasari (*HyperVasari*), con l'obiettivo di dimostrare che l'esplorazione avanzata di un corpus digitalizzato e annotato lessicalmente e semanticamente, resa possibile da questo strumento, permette di integrare e precisare le informazioni fornite dalle principali risorse lessicografiche e terminologiche esistenti per la traduzione della lingua dell'arte.

Come è noto, lo sviluppo delle *digital humanities* va di pari passo con lo sviluppo delle tecnologie e, in particolare, del web, motivo per cui, secondo gli studiosi, questa seconda fase va anche associata a un ulteriore momento di espansione del web, il web 2.0 (O'Reilly 2005). Se nel cosiddetto web 1.0, l'informazione veniva creata e distribuita da pochi esperti e consumata da utenti che semplicemente cliccavano su un link per accedere a dati multimediali e ipertestuali accessibili in rete, nel web 2.0 o web sociale, qualsiasi persona ha la possibilità di creare, descrivere e distribuire

contenuti digitali, nonché di interagire con dati creati da altri utenti, modificarli e collaborare al loro sviluppo con altri utenti ancora, anche nel momento stesso in cui essi vengono creati (Carter 2013: 12). D'altra parte, le reti sociali virtuali come Facebook o Google+, i blog e i microblog come Twitter, gli spazi di condivisione di dati, quali YouTube o Instagram, per nominare solo quelli più diffusi, vengono usati sempre di più per comunicare all'interno delle discipline umanistiche e per diffondere le proprie ricerche attraverso testi, ipertesti, immagini e video spesso etichettati grazie a degli identificatori o *tag*, in un contesto di ricerca più ampio che richiede nuovi linguaggi.

Gli studiosi utilizzano quotidianamente strumenti di video broadcasting, di conferenze audio o di audiocasting (Skype, Oovoo, Spreaker, tra molti altri), accedono ad ambienti di lavoro in collaborazione o di condivisione di documenti, e fanno spesso ricorso ad applicazioni di *screen sharing* (Join.me, ScreenLeap), con i quali si è in grado di raggiungere qualsiasi utente nel mondo dal proprio computer, tablet o smartphone. Questi canali offrono ai ricercatori modi complessi per comunicare, collaborare, condividere i contenuti del loro lavoro e diffonderli. Altre piattaforme, invece, sono utili spazi di consultazione di grandi quantità di dati, i quali possono essere implementati dagli studiosi in scienze umanistiche. Proprio in relazione con le prospettive aperte da queste risorse, Ana Pano Alamán propone una riflessione critica sulle opportunità e sui limiti dell'enciclopedia collaborativa Wikipedia per l'estrazione automatica di dati e per la successiva creazione di banche dati terminologiche sull'arte e sul patrimonio culturale. Nell'ambito del progetto LBC e partendo dall'analisi qualitativa-comparativa di alcune voci dell'enciclopedia relative ai termini della pittura, l'autrice esplora le possibilità che la nota piattaforma offre al traduttore specializzato nel campo dell'arte e del patrimonio culturale, nonché l'eventuale contributo del progetto LBC all'arricchimento dell'enciclopedia online.

In relazione con questa fase dello sviluppo del web, è d'obbligo menzionare il concetto di 'web semantico', la cui definizione, secondo il *Nuovo Soggettario Thesaurus* della Biblioteca Nazionale Centrale di Firenze, è

Implementazione del World Wide Web come fonte d'informazione e di conoscenza, attribuendo ad *agenti software la capacità di analizzare il significato dei documenti in esso presenti* e dunque di selezionarli o confrontarli in modo semanticamente rilevante o di inferirne conseguenze che non siano già esplicitate (BGC) [corsivo nostro]<sup>4</sup>.

<sup>4</sup> Disponibile all'indirizzo: <<http://thes.bncf.firenze.sbn.it/termine.php?id=48388&menuR=2&menuS=2>> (02/2017).

In questo caso, la ricerca in ambito umanistico si confronta con altri concetti complessi quali web 3.0, web di dati, Linked data, tag, metadati, semantica e sistemi esperti. L'informatica umanistica si trova così di fronte a un nuovo paradigma che permette di descrivere i dati non solo dal punto di vista formale, ma anche dal punto di vista del loro significato, grazie a metadati e sistemi di lettura degli stessi molto più sofisticati. Di fatto, in questo momento ci si avvia, secondo Carter (2013: XI), verso una terza fase delle *digital humanities*, quella degli strumenti e degli ambienti già creati che stanno adesso evolvendo o migliorando. Alcuni ricercatori e sviluppatori web lanciano oggi applicazioni che pertengono al web 2.0, ma che fanno progredire la ricerca verso il cosiddetto web socio-semantico (Brossaud, Reber 2007: 20). Questo è caratterizzato principalmente da interazioni sociali che permettono di creare rappresentazioni dei dati esplicite e semanticamente ricche di conoscenza. Bisogna ricordare che il web si intende qui come un sistema di 'intelligenza collettiva', che è capace di fornire informazioni attraverso il contributo delle persone e che integra ed utilizza tecnologie e metodologie proprie del web semantico, del software sociale e del web 2.0.

È in questa fase di transizione che si colloca il contributo di Francesca Tomasi. La studiosa applica i presupposti del web semantico allo studio della collezione delle *Lettere* di Vespasiano da Bisticci. Il saggio tratta questo importante cambiamento in corso, fornendo una ricca panoramica sull'evoluzione dei linguaggi del web, sulle nuove forme di rappresentazione dei testi incentrate sul dato e non più sul documento, e sulla descrizione dell'informazione con sistemi Linked Open Data (LOD). Il testo offre dunque un'ampia riflessione sulle metodologie di edizione digitale che si avviano man mano verso il *knowledge site*, ambiente semantico di accesso alla conoscenza, formato dai dati di un testo, dalle stringhe di dati interpretati e descritti con metadati, e dalla relazione di questi dati con quelli che sono disponibili sul web. In questo modo, assistiamo a dei cambiamenti che aprono prospettive inedite nel campo dell'informatica umanistica per quanto riguarda il lavoro sui testi: infatti, gli iperdocumenti non sono più duplicazioni dei documenti cartacei e non possono essere ricondotti semplicemente al prodotto della digitalizzazione di uno scritto; essi si liberano dalle forme tradizionali di lettura e possono essere modificati e arricchiti all'infinito.

In definitiva, l'informatica umanistica, e le *digital humanities*, si presentano come una rottura epistemologica, come un profondo cambiamento nelle strutture delle discipline coinvolte (Brossaud, Reber 2007: 47). Non siamo confrontati soltanto con un nuovo tipo di dati, sommersi dall'accesso ad un numero di risorse documentarie immense o messi di fronte a nuove possibilità di collaborazione tra studiosi. Come si evince dalle riflessioni raccolte in questo volume, il vasto campo dell'informatica umanistica tocca il cuore teorico di alcune discipline, la loro organizzazione e le loro implica-

zioni sociali e politiche (Brossaud, Reber 2007: 24; Spence 2014). Di fronte a questa trasformazione in corso, l'umanista digitale dovrebbe compiere un doppio scarto: «riscoprire le proprie radici e aprirsi al rinnovamento. [...] [e] riconoscere che il sapere umanistico non può più crescere e diffondersi senza gli strumenti di comunicazione, rappresentazione e organizzazione delle informazioni» (Numerico, Vespignani 2003: 8-9). Questo volume intende essere un ulteriore passo in questo senso, per mostrare come le risorse e gli strumenti offerti dall'informatica umanistica portano con sé un deciso rinnovamento nell'ambito dello studio e della divulgazione della lingua dell'arte. Questo vuole essere però un passo in avanti il più possibile dinamico, poiché sappiamo che gli strumenti digitali messi a disposizione dei ricercatori continueranno ad evolversi e a perfezionarsi negli anni a venire aprendo nuove prospettive di studio.

### Riferimenti bibliografici

- Brossaud C., Reber B. (eds.) 2007, *Humanités numériques 1. Nouvelles technologies cognitives et épistémologie*, Lavoisier, Paris.
- Carter B. W. 2013, *Digital Humanities. Current Perspective, Practices, and Research*, Emerald, Bingley.
- Davidson C. N. 2008, *Humanities 2.0: promise, perils, predictions*, «Publications of the Modern Language Association of America (PMLA)», CXXIII (3): 707-717.
- 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.publifarum.farum.it/ezone\\_articles.php?art\\_id=335](http://www.publifarum.farum.it/ezone_articles.php?art_id=335)> (01/2017).
- 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. (coord.), *Strategie e programmazione della conservazione e trasmissibilità del patrimonio culturale*, Edizioni Fidei Signa, Roma: 500-509.
- Marcos Marín F. 1994, *Informática y Humanidades*, Gredos, Madrid.
- Numerico T., Fiormonte D., Tomasi F. 2010, *L'umanista digitale*, il Mulino, Bologna.
- Numerico T., Vespignani A. 2003, *Informatica per le scienze umanistiche*, il Mulino, Bologna.
- O'Reilly T. 2005, *What is Web 2.0.*, <<http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>> (01/2017).
- Rieger O. 2010, *Framing Digital Humanities: The role of new media in humanities scholarship*, «First Monday», XV (10).
- Schnapp J., Presner P. 2009, *Digital humanities manifesto 2.0*. <[http://www.humanitiesblast.com/manifesto/Manifesto\\_V2.pdf](http://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf)> (01/2017).
- Schreibman S., Siemens R., Unsworth J. 2004, *A Companion to Digital Humanities*, Blackwell, Oxford, <<http://www.digitalhumanities.org/companion/>> (01/2017).

- Spence P. 2014, *Centros y fronteras: el panorama internacional de las humanidades digitales*, *Humanidades digitales*, «Janus», Anexo 1: 37-61, <[http://ruc.udc.es/dspace/bitstream/handle/2183/13576/HD\\_art\\_3.pdf?sequence=1](http://ruc.udc.es/dspace/bitstream/handle/2183/13576/HD_art_3.pdf?sequence=1)> (01/2017).
- Svensson P. 2010, *Landscape of Digital Humanities*, «Digital Humanities Quarterly», IV (1), <<http://digitalhumanities.org/dhq/vol/4/1/000080/000080.html>> (01/2017).
- Tomasi F. 2008, *Metodologie informatiche e discipline umanistiche*, Carocci, Roma.





C. Samson

## Guidebooks of Florence for a specialised lexical database. A corpus-driven linguistic analysis

*Abstract:* For long guidebooks have been considered a resource for history of tourism studies. They have been included in genre analysis by mainly focusing on their textual, visual content, and spatial descriptions while being seen as a support to the dissemination of culture online. However, few studies have analysed common and proper nouns in guidebook corpora on which specialised online dictionaries are based. The purpose of this study is, therefore, to analyse the lexicon of Florentine heritage in order to bolster translators' and students' knowledge of the linguistic cultural aspects of Florentine heritage. By adopting a *corpus-driven linguistic* approach, common and proper nouns with their *clusters/n-grams* are quantitatively analysed in a corpus of online guidebooks of Florence. The emerging data are then qualitatively interpreted through discourse analysis to highlight how the repeated use of *clusters/n-grams* form a network and a variation of meaning within the corpus.

*Keywords:* corpus linguistics, heritage, Florence, nouns, phraseology.

*Riassunto:* Le guide turistiche sono state a lungo considerate una fonte per lo studio della storia del turismo. Sono state incluse nell'analisi di genere con particolare attenzione al loro contenuto testuale, visuale, alla loro descrizione spaziale e sono anche state considerate un ausilio alla diffusione della cultura online. Tuttavia, pochi studi si sono incentrati sull'analisi dei nomi comuni e propri nei corpora di guide turistiche a partire dai quali vengono costituiti database di dizionari specialistici online. Lo scopo dello studio è, perciò, di analizzare il lessico dei beni culturali fiorentini per ampliare le conoscenze di traduttori e studenti sugli aspetti linguistico-culturali del patrimonio culturale fiorentino. Attraverso l'approccio della *corpus-driven linguistics*, i nomi comuni e propri con i relativi *cluster/n-gram* vengono analizzati quantitativamente in un corpus di guide di Firenze online. I dati sono successivamente interpretati qualitativamente mediante un'analisi del discorso

per rivelare come l'uso ripetuto dei *cluster/n-gram* formi un network ed una variazione di significato all'interno del corpus.

*Parole chiave:* linguistica dei corpora, patrimonio, Firenze, nomi, fraseologia.

## 1. Introduction

For long, guidebooks have contributed to construe generic histories of tourism (Bruner 2004), or to investigate people's narratives about their travel and tourism experiences (Beck 2006). These texts have also been included in genre studies (Denti 2012) by analysing their textual and visual content (Bhattacharyya 1997), their descriptions of space and/or identity of heritage sites (Samson 2011), or the way they have popularised museums and art on the Internet (Samson 2012). In contrast, there is a paucity of studies on the use of common and proper nouns to describe heritage in guidebooks forming corpora as a database for online dictionaries.

Heritage includes a large range of goods. Its definition changes over time and space depending on the variety of dimensions (symbolic, cultural, national identity-oriented, social and suchlike) included in the concept (Chastel 1986). Benhamou (2011) argues that heritage can be seen as a social construction whose boundaries are unstable and blurred with a two-fold source of extension: historical additions and an enlargement of the concept towards other items, such as gardens, industrial buildings, and so on. Consequently, heritage is not only about tangible material artifacts and/or intangible forms of the past, but it is also about the meanings placed upon them, the representations created for them (Smith 2006).

Studies on the compilation of multilingual dictionaries focusing on heritage, and deriving from comparable and/or parallel databases, have hardly been developed (Teubert 2007). Most research has addressed the automatic compilation of lists of words and the development of automatic extractors of terms without considering the potential of a corpus as a source of information to give account of the use of lexical items (Alonso *et al.* 2012) in construing extended or multi-word units of meaning.

By extended units of meaning Sinclair (1996) refers to a core word (node) that incorporates other words in the co-text that appear to be co-selected with it and form a regular pattern. These are multi-word units, i.e., they are defined by the strict correlation existing between a node and its context. They involve both lexical and grammatical realizations and only when they have reached their pragmatic function can they be seen as 'functionally complete' (Tognini-Bonelli 1996).

This paper, as part of a wider research project – *Il Lessico dei Beni Culturali di Firenze* including the creation of comparative databases in seven languages (Farina 2015) – analyses how Florentine heritage is described in a corpus of online heritage guidebooks of Florence (OHGFLO) by adopting a

corpus-driven linguistic (CDL) approach. This is integrated with discourse analysis, given that the aim of discourse analysis is to identify the conventional meanings and values expressed in a corpus of texts (Groom 2010).

A *corpus* can be defined as a computerised collection of authentic texts, amenable to automatic or semiautomatic processing or analysis. The texts are selected according to explicit criteria (content/genre/register, etc.) with a specific purpose in mind, in order to capture the regularities of a language, a language variety or a sub-language (Tognini-Bonelli 2001). A CDL investigation starts by automatically extracting lexical items from the entire corpus (OHGFLO). The research is carried out on whole texts and not on text samples. Working with samples (for e.g. on the first 2,000 words of each text) carries the risk of missing important items that are characteristic of the text type under scrutiny and tend to occur outside the text sections covered in the samples (Sinclair 1991). By adopting CDL, instead, the corpus tells us what the facts are, as the narratives talk for themselves (Tognini-Bonelli 2001). This means that the relative most frequent keywords and their recurring clusters in OHGFLO emerge directly from the corpus itself, without being adjusted to fit pre-existing categories of the analyst (Table 2), with computer software applied to the corpus (Sinclair 1992).

Thus, the purpose of this study is to shed light on the lexicon of Florentine cultural heritage, given the paucity of studies referring to the lexis used to describe Florence which is a renowned example of Italian 'art city'. The study, furthermore, aims to provide translators and students with linguistic and cultural information emerging from the key common and proper nouns and the use of their clusters/n-grams to construe networks and meaning across OHGFLO. Clusters/n-grams refer to the identification of the commonest collocations providing more context than what may be attained by a single-word analysis.

The remainder of the paper is organized as follows. Section 2 defines the difference between common and proper nouns; section 3 discusses the function of clusters and phraseology. Section 4 describes the corpus and explains the methodology used to generate the data whereas the findings are analysed in section 5. Final conclusions are drawn in section 6.

## 2. Common and proper nouns

Common nouns are nouns that are generalised to a class of referents. Halliday (2004: 326) claims that:

[...] they name all the classes of phenomena that the language admits as things, and hence as participants in processes of any kind. There is a long tradition of characterising such phenomena as a list of very general categories, e.g. persons, other living beings, objects (concrete or abstract) collectives, institutions. These relate to a cline of potential agency, that is the likelihood of functioning as Actor/Agent in the clause (2004: 326).

Searle (1958) argues that common nouns denote, name, or point out a certain object or class of objects. Common nouns convey or imply some qualities or facts concerning them. In other words, all such nouns have a meaning, or are connotative. By contrast, proper nouns do not specify any characteristics, they convey no meaning, they are non-connotative, since «they function not as descriptions, but as pegs on which to hang descriptions» (Searle 1958: 172). In other words, they are affixed to one object not to convey any fact about it, but to enable you to speak about it. Marmaridou (1989: 355-356) argues that proper nouns may be attributed to more than one referent, yet, in discourse the encoder refers to a specific referent, situated in a given time and space. In order to understand which referent the encoder is referring to, the decoder must possess a competence of the name system as well as the chunks of encyclopaedic knowledge associated with a name to establish a link between proper noun and referent. Only when the decoder retrieves associated information from his/her knowledge, the 'virtual' referent is actualised, and the proper noun becomes a 'rigid designator'.

Thus, although proper nouns constitute a class of linguistic items sharing features with both nouns and deictics, they differ in various respects. Both proper nouns and deictics lack lexical meaning and have a referential function; but, while the interpretation of deictics depends on the situational context, the interpretation of proper nouns depends on the linguistic context and the encyclopaedic knowledge. In interpreting the proper noun, the decoder first has to recognise whether its use is referential or figurative, relying on the linguistic context; then, s/he will activate encyclopaedic knowledge or recur to her/his lexical competence, if the item is lexicalized. Moreover, proper nouns refer to a 'fixed' referent, while deictics to a referent that can vary according to the situational context (Pierini 2008).

Given such features, proper nouns are usually excluded from monolingual dictionaries as the only way to describe them is by detailing their referents, which is the main concern of encyclopaedias. Nevertheless, Farina (2015) asserts that they can be defined in a dictionary since, as common nouns, proper nouns have hyperonyms or related synonyms and associations that are used to define them. These associations or clusters are considered, in this study, knowledge of the world, as they represent the memory of what a culture has associated with them and they contribute to construing the meaning of heritage proper nouns belonging to a particular culture.

### **3. Clusters and recurrent phraseology**

Clusters, also defined by Biber *et al.* (2004) as lexical bundles, are words which are found repeatedly together in each others' company in a sequence forming phrases (Scott 2010). Clusters are based on the assumption that words are not to be seen as elements in isolation that can be slotted

into syntactic frameworks, but as forming larger units or, as Sinclair (1996) terms them, as extended units of meaning.

Since the meaning of words lies in their use and use cannot exist in isolation, use can only be recognised and analysed contextually and functionally, as Firth (1957) argues. Consequently, language is to be seen as the vector of continuous repetitions in the social process (Firth 1957), that is, people linguistically act systematically; their lexical patterns entail patterns of meaning and every distinct sense of a word is associated with a distinction in form. In other words, form and meaning are inseparable (Sinclair 2004). Williams (2012) views clusters as statistically based chains of collocations which form recurrent phraseology. Recurrent clusters, thus, build networks in the form of phraseology which are linked through the process of collocation. The idea that collocations «cluster» forming interwoven meaning networks comes from Phillips (1985). Phillips's aim was the study of metastructure within texts and the notion of 'aboutness'. Following this, Williams (2000) hypothesises that the patterns of co-occurrence forming the collocational networks will be unique to any sublanguage and serve to define the frames of reference within that sublanguage. Williams and Millon (2010) state that collocational networks not only demonstrate thematic patterns, but they also show the most significant lexical units which out of the analysis of monolingual corpora form the main cognitive nodes of a specific corpus. Therefore, the studies developed show that chains of collocations also constitute a powerful tool for headword selection.

Moreover, cluster networks enable the analyst/translator not only to look at the immediate environment of a search word, but also to link it outwards to the wider meaning context. This enables to isolate lexical units in the Sinclairian sense (Williams 2010) and foreground the connotations which give sense to common and proper nouns in a particular culture (Samson 2016a). Contextual meaning is therefore vital as, on the one hand, simple surface equivalence can hide important connotative differences between or among common and proper nouns. On the other hand, different situational contexts, specialised languages or specific genres develop clusters which are unique to that environment (Williams 2001) which, in this case, is a corpus of online guidebooks of Florence.

#### **4. Corpus and methodology**

The corpus was compiled by downloading online heritage guidebooks of Florence in English to form the OHGFLO, which describe the main monuments, museums, places, public figures and artists of Florence. All the webpages describing Florentine heritage sites were saved as text files (txt), in order to be processed by WordSmith Tools 5.0 (WST) (Scott 2010), a piece of commercial computer software.

Although the corpus is small, it presents the advantage of being homogeneous, as regards time of publication (2013-2014), genre, and it is in line with Coniam (2004), Ghadessy *et al.* (2001) and Sinclair (2001), who note that a small corpus, properly constructed, can be viewed as a body of relevant and reliable evidence. The corpus includes 130 files, comprising a total of about 74,000 words. The online heritage guidebooks of Florence are listed in Table 1.

Table 1 – Online heritage guidebooks of Florence – OHGFLO (06/2017).

<<http://www.museumsinflorence.com/index.html>>  
 <<http://www.viator.com/Florence/d519-ttd>>  
 <<http://www.lonelyplanet.com/italy/florence>>

The methodology adopted in this study is a mixed one. It starts with a corpus-driven analysis (Tognini-Bonelli 2001) whose main feature is commitment to the data it starts from and from which it tries to derive observational and theoretical findings, while not losing contact with the corpus (Römer 2011). In this method, the text is processed directly without having previously annotated it in its parts (Sinclair 2000), that is, without the addition of information/elements with the purpose of providing linguistic/grammatical/structural information, such as part of speech, semantics, pragmatics, prosody, interaction and many others. In this way, the centrality of the texts forming a corpus is pivotal, as findings are directly derived from the corpus and not filtered through existing concepts that are supposed to take place. Consequently, a more objective extraction of lexis can occur by calling upon statistically significant co-occurrence patterns in a text which are then qualitatively analysed. Römer (2011) claims that, even if intuition is not banished from corpus-driven analysis, evidence from the corpus should never be ignored, since it can lead to new theoretical insights of language, to observations that previously were not possible to make. Consequently, the method leads to see how language really works, its actual facts, how it is used in communicative situations, as in the case of online heritage guidebooks of Florence.

*WST 5.0* through its *WordList* function automatically generates word lists based on input text files. The main outputs of *WordList* are a statistical frequency list about the words in the OHGFLO corpus. The function provides additional information, such as the number of texts and the percentage of occurrence in texts in the word list. Thus, a Word list for the OHGFLO corpus was generated and compared with a reference corpus DATA<sup>1</sup> (approx. 700,000 words) to obtain a Keyword List which is a re-

<sup>1</sup> DATA is a reference corpus including files of entire texts collected by the author during her participation to national interuniversity research groups sponsored by the Ministry of Education, University and Research of Italy. DATA is formed by files of published written economics lectures, industrial products, surgery products, EU and non-EU museum descriptions, collections, exhibitions, narrative guidebooks.

finement to the word list production (Coniam 2004). *WST 5.0* calculates the keywords by comparing the frequency of each word in the smaller list of the two wordlists with the frequency of the same word in the reference wordlist. All words which appeared in the smaller list were considered to be unusually frequent, in comparison with what one would have expected on the basis of the reference corpus (Scott 2010).

A further step was to discover the clusters emerging from the *WST 5.0 Concordancer* function which generates statistical counts per 1,000 key words in the corpus. Since frequency was not seen as self explanatory, but as data that needs to be explained, a qualitative analysis of the organisation of language above the sentence or above the clause (Stubbs 1983) proceeded by analysing the key clusters which constitute multi-word units of meaning in OHGFLO.

## 5. Analysis and findings

The first feature analysed are the lexical items emerging from the corpus. Table 2 presents the first twenty five most frequent words of OHGFLO which were determined by taking the cut-off score of 1,000 or above as the log-likelihood statistic.

Table 2 – OHGFLO – Most frequent words.

N	Word	Freq.	%
1	THE	8,689	11.73
2	OF	3,928	5.30
3	AND	2,408	3.25
4	IN	1,908	2.58
5	#	1,743	2.35
6	TO	1,503	2.03
7	A	1,158	1.56
8	BY	1,116	1.51
9	WAS	826	1.12
10	IS	644	0.87
11	WITH	599	0.81
12	ON	523	0.71
13	FOR	441	0.60
14	FROM	432	0.58
15	AS	400	0.54
16	IT	387	0.52
17	<b>CENTURY</b>	<b>373</b>	<b>0.50</b>
18	THAT	368	0.50
19	WHICH	340	0.46
20	HIS	323	0.44
21	ARE	292	0.39
22	<b>MEDICI</b>	<b>292</b>	<b>0.39</b>
23	THIS	250	0.34
24	WERE	241	0.33
25	ITS	240	0.32



The snapshot provided in Table 2 indicates that the most frequent common noun «century» refers to a period of time whereas the most frequent proper noun *Medici*, refers to the most influential family in Renaissance Florence. As mentioned, a refinement to the wordlist production can be obtained through *WST 5.0's KeyWord* which allows the analysis to be extended by comparing the high-frequency words in the corpus with a reference corpus (DATA). The results among the first twenty-three most frequent key words in OHGFLO are listed in Table 3 (function words excluded).

Table 3 – OHGFLO Key word list.

N	Key word	Freq.
7	<b>CENTURY</b>	804
9	WAS	1,759
10	<b>MEDICI</b>	515
15	MUSEUM	423
16	WE	133
17	PALAZZO	345
18	FLORENCE	535
19	BUILDING	346
20	CHAPEL	325
21	WORKS	460
22	MICHELANGELO	288
23	RENAISSANCE	279

As can be seen, the common noun 'century' and the proper noun *Medici* are confirmed as the relative most frequent key words in OHGFLO. To provide more context to the single word analysis, a cluster/n-gram analysis was undertaken by identifying the commonest collocations of 'century' and by carrying out a Concordancer statistical count per 1,000 words, as shown in Table 4:

Table 4 – Concordance for key common noun 'century'.

<p>me rare examples from the 16th <b>century</b>. The rooms, quipped remains after the seventeenth <b>century</b> alterations. Along w isters, built in the 16th-17th <b>century</b> by Ammannati and his of civic memories. In the 19th <b>century</b> the church received a er the period between the 18th <b>century</b> and the present-day.</p>
--

The repeated collocations of the key word 'century' in OHGFLO, as shown in Table 4, highlight its semantic preferences which are linked to two or more words within a short space of each other. These can be

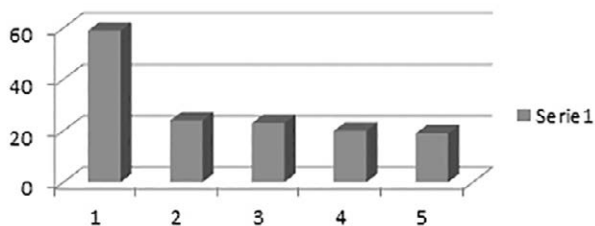
considered crucial aid for translators and students who are unfamiliar with, or need to acquire further knowledge of Florence and its history, as the repeated collocations provide ‘century’s’ connotations and contextual meaning. In Table 4, the collocations mainly refer to the interior space of museums or churches, the historical changes which Florentine built heritage underwent, the collections of most famous Renaissance artists, or the development of specific areas of the city over a period of time. The concordances, thus, are very useful as they point at the nodes’ typicality or uniqueness (Scott 2010) which characterise OHGFLO.

To examine the semantic preferences of the key common noun ‘century’ in more depth, a 4-word cluster/n-gram search of the concordances in Table 2 was undertaken. The top five clusters emerging are listed in Table 5 and depicted in Figure 1:

Table 5 – Most frequent 4-word clusters/n-grams for key common noun ‘century’.

N	Cluster	Freq.
1	THE END OF THE	59
2	THE BEGINNING OF THE	24
3	SECOND HALF OF THE	23
4	IN THE 16TH CENTURY	20
5	OF THE 19TH CENTURY	19

Figure 1 – Most frequent 4-word clusters/n-grams for key common noun ‘century’.



For space reasons, the analysis will only focus on the cluster/n-gram THE END OF THE and its discourse functions. These were identified by looking at its proximity to a consistent series of collocates that beyond the limit of the cluster share its semantic preferences which not always include the key proper noun ‘century’ as, for example, when referring to space. Moreover, the recurrent use of the same cluster across the corpus (59 times) indicates that it generates recurring cluster networks across OHGFLO. The qualitative interpretation of THE END OF THE shows that it repeatedly co-occurs with referential expressions that identify an entity or single out some particular attribute of that entity as being especially important (Biber, Barbieri 2007).

## 5.1 Spatial reference

In examples (1) and (2) below the cluster/n-gram acquires the meaning of spatial reference by collocating cataphorically with nouns horizontally defining a building's internal space ('aisle', 'corridor') within the use of frames of reference (FoR) (Levinson 2003). These are coordinate systems used to compute and specify the location of objects with respect to other objects in every utterance occurring in a particular spatio-temporal situation (Levinson 2003). The orientational features of language are expressed through deixis which, in this case, includes the use of expressions directing the browser's gaze through the interior space («At the end of the right aisle», «at the end of the corridor», «at the end of the room») of a museum, church or cloister. In addition, the descriptions portray also the content of the spaces ('arks', 'altarpiece') by specifying their attributes («two-tone in glazed terracotta») while evaluating them («fine») for someone not sharing the same visual experience (Smyth 2008). This hints at the describer acting as mediator between reality and a verbal icon of it (Merlini Barbaresi 2009), that is, the verbal subjective description of objects in space, which stimulates their visual and contextual representation in the addressee's mind. The spatial references are indeed integrated by fictive motion verbs («leading», «gives access») in examples (1) and (3) that express no explicit motion or state change, but include a mental simulation (Matlock 2004). Such simulations have a crucial role in construing representations of a destination<sup>2</sup> in guidebooks, since they not only stimulate interest in potential tourists to actually visit the places described, but they are also essential to translators and students needing appropriate background knowledge of Florence.

(1) At **the end of the** *right aisle* there is a doorway *leading* to an oratory *containing two Arks*. Museum of Hebraic Art and Culture txt

(2) At **the end of the** corridor there is a *Medici chapel* with a *fine two-tone altarpiece in glazed terracotta by Andrea della Robbia*. Santa Croce txt

(3) The small cloister *gives access* to the Refectory and at **the end of the** room it *leads* to the Large Cloister. Santa Maria Novella txt

## 5.2 Temporal reference of events

Furthermore, the analysis of the cluster/n-gram THE END OF THE appears to be crucial for translators and students when investigating other usages of the cluster/n-gram. As the data shows, the cluster/n-gram has also a time reference function characterising the historical narration of the works

<sup>2</sup> For the construal of heritage representations in guidebooks see Samson (2016b).

of art which are typified by the use of past tense («were painted», «were invited»)<sup>3</sup>. The aim of identifying and underlining the importance attributed to the content of the church is explicitly underscored by the use of evaluative adjectives («important») and by referring to artists' names (Andrea della Robbia, Domenico Ghirlandaio, Filippino Lippi), or to the name of those (*Umiliati*) granted the honour of becoming members of the Florentine public administration, as pointed out in examples (2), (4) and (5).

(4) Towards **the end of the 15th century** two *important* frescoes were painted for the church by *Domenico Ghirlandaio* and *Filippino Lippi*. Santa Maria Novella txt

(5) By **the end of the 13th century** the *Umiliati* were invited to become members of the public administration. Cenacolo di Ognissanti txt

### 5.3 Temporal value appearance

The recurring collocation of THE END OF THE with a specific time period («fifteenth century»), proper names of churches and streets (*Santa Croce*, *Via Larga*), adjectives («maximum») and the proper noun (*Medici*) also points at the semantic preference for a connotation of power. This can be seen in examples (6) and (7).

(6) At **the end of the** fifteenth century Santa Croce church reached its *maximum* extension and *importance*.

(7) At **the end of the** *fifteenth century*, Via Larga was seen pre-eminently as the street of the *Medici*. Palazzo Medici Riccardi txt

A Concordancer statistical count per 1,000 words was also carried out for the relative most frequent key proper noun *Medici*, and the collocations emerging are shown in Table 6:

Table 6 – Concordance for key proper noun *Medici*.

<p>ceiling among the most admired in the <b>Medici</b> residence. The walls of Man, commissioned by the <b>Medici</b>, which represented the built by Michelozzo for Cosimo de <b>Medici</b>, where a considerable the buildings also contain the <b>Medici</b> chapels with their crypt furniture and textiles from the <b>Medici</b> collections and those</p>
--

In Table 6 the semantic preferences of *Medici* allow translators and students to understand the importance of this historical family by the key word's repeated references to the buildings, tombs, and churches built un-

<sup>3</sup> For the use of verb tenses in guidebooks see Samson (2016b).

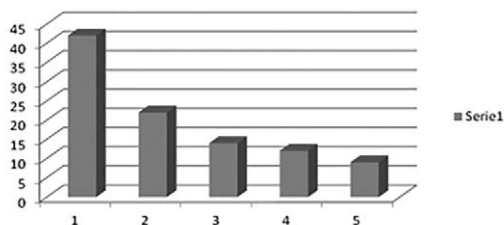
der the Medici, where they were buried, the type of their art collections. However, as already mentioned, an in-depth examination of the network of meanings construed across OHGFLO, may be achieved by a cluster search of the key proper noun *Medici*. The results are listed in Table 7.

Table 7 – Cluster – Key proper noun *Medici*.

N	Cluster	Freq.
1	OF THE MEDICI FAMILY	42
2	COSIMO I DE MEDICI	22
3	MEMBERS OF THE MEDICI	14
4	OF THE MEDICI DYNASTY	12
5	OF COSIMO I DE MEDICI	9

Figure 2 depicts the considerable variations in terms of frequency of the first five most frequent 4-word clusters/n-grams for proper noun *Medici* which create a cluster network in OHGFLO:

Figure 2 – Most frequent 4-word clusters/n-grams for key proper noun *Medici*.



#### 5.4 Buildings as symbols of power

As the analysis foregrounds, the relative most frequent cluster OF THE MEDICI FAMILY is repeatedly used to underline the importance and power of the Medici family in Renaissance Florence. This is expressed not by factual descriptions of battles, or any other specific activity the various members of the Medici family were engaged in, but through the descriptions of how any space related to them was divided, or used within churches («crypt», «grandiose octagonal chapel») and/or buildings (*Pitti*). Also the evaluation of spatial dimension («immense dome», «fourteen magnificent rooms») and the precious building materials used («polychrome marble» and «pietre dure») to decorate internal spaces acquire connotations of power and value, as shown in examples (8), (9) and (10):

- (8) The church buildings also contain the Medici chapels with their crypt, in which lie the remains of 50 members **of the Medici family**. Underneath

the church are buried both Cosimo the Elder and Donatello. Basilica of San Lorenzo txt

(9) The Chapel of the Princes was begun in the early 17th century to become the mausoleum **of the Medici family** grand-dukes. This grandiose octagonal chapel, with its immense dome, is entirely faced with polychrome marble and pietre dure. Medici Chapel txt

(10) The Royal Apartments consist of fourteen magnificent rooms which were the home **of the Medici family** and, from 1865, of the king of Italy. Pitti txt

### 5.5 Art as a symbol of power

The cluster OF THE MEDICI FAMILY is also recurrently used to underline in the guidebooks' narration that art was another medium used by the Medici family to convey their power to the public. This is shown in example (11) wherein a statue symbolising astuteness winning over brute force was chosen to be the symbol of Florence:

(11) Michelangelo sculpted this statue between 1502 and 1504; this was the most commonly portrayed Biblical character in the Renaissance, because he symbolises astuteness winning over brute force. The statue became the symbol of the city right from the time **of the Medici family**. Accademia Gallery txt

Also in example (12) allegorical paintings are used to represent the Medici family's power and possessions:

(12) The allegorical paintings on the ceiling and the walls narrate the triumphal Return of Grand Duke Cosimo I to Florence, illustrate the possessions **of the Medici family** and the Stories of the Conquest of Pisa and Siena. Salone dei 500 txt

All the above examples are instances of the proper noun *Medici* that is defined by its collocates and clusters while construing expanded units of meaning across OHGFLO. The repeated clusters/n-grams can thus be seen as a valid support to translators and students needing to acquire, or to widen their knowledge of Florence's heritage, since, as mentioned, the clusters/n-grams of their common and proper nouns in OHGFLO represent the memory of the city's Renaissance culture.

### 6. Conclusion

This study has highlighted the most frequent key common and proper nouns characterising the lexicon of cultural heritage in OHGFLO. Through

a Corpus-Driven Linguistic approach the use of the two most frequent key common and proper nouns, 'century' and *Medici*, have been analysed and their repeated collocations have led to the identification of their most frequent cluster networks across OHGFLO.

Furthermore, the study has attempted to show how the recurring clusters/n-grams can help translators and students acquire and/or bolster their knowledge about Florence's heritage. For instance, the semantic preferences of 'century's' cluster/n-gram THE END OF THE are useful in that they underline spatial reference, temporal reference of events, and temporal value appearance with the purpose of singling out, through descriptions and narration, particular attributes of Florentine tangible and intangible heritage. More specifically, the cluster emerges as being multifunctional, in that it is used to highlight particular features characterising interior spaces, the value of the artworks and artists belonging to a specific period of time as well as to identify the importance of historical events and the value attributed to spaces and places of Florence.

By contrast, the semantic preferences of *Medici's* most frequent cluster OF THE MEDICI FAMILY convey the connotation of power to the proper noun. Its repeated use has the function of creating extended units of meaning that provide translators and students an awareness of the power and wealth of the Medici family through spatial descriptions, the richness characterising the building materials implemented and the artistic representations. The cluster networks across OHGFLO, thus, shed light on how the proper noun *Medici* as any common noun implies qualities of, and coincides with power.

The findings, therefore, suggest that the use of recurring key common and proper nouns and their cluster/n-gram networks are a crucial means for translators and students when consulting a cultural heritage dictionary. Clusters allow to discover how key nouns can vary their connotations in different situational contexts and how, more importantly, not only common but also proper nouns contribute to construe meaning in an online dictionary of Florentine cultural heritage.

## References

- Alonso A., Blancafort H., de Groc C., Million C., Williams G. (eds.) 2012, *MET-RICC: Harnessing Comparable Corpora for Multilingual Lexicon Development*, 15th EURALEX International Congress, Oslo.
- Beck U. 2006, *The Cosmopolitan Vision*, Polity Press, Cambridge.
- Benhamou F. 2011, *Heritage*, in Towse R. (ed.), *A Handbook of Cultural Economics*, Elgar, London: 255-262.
- Bhattacharyya D.P. 1997, *Mediating India: An analysis of a guidebook*, «Annals of Tourism Research», XXIV (2): 371-389.
- Biber D., Conrad S., Cortes V. 2004, *If you look at... Lexical bundles in university lectures and textbooks*, «Applied Linguistics», 25: 371-405.

- Biber D., Barbieri F. 2007, *Lexical bundles in university spoken and written registers*, «English for Specific Purposes», 26: 263-286.
- Bruner E. 2004, *Culture on Tour. Ethnographies of Travel*, Chicago University Press, Chicago.
- Chastel A. 1986, *La notion de patrimoine*, in Nora P. (éd.), *Les Lieux de Mémoire*, Gallimard, Paris: 405-450.
- Coniam D. 2004, *Concordancing oneself: Constructing individual textual profiles*, «International Journal of Corpus Linguistics», IX (2): 271-298.
- Denti O. 2012, *The Island of Sardinia from travel books to travel guides*, in Fodde L., Van Den Abbeele G. (eds.), «Textus», XXV (1): 37-50.
- Farina A. 2015, *Guideline proposal for the description and translation of proper nouns in a multilingual cultural heritage dictionary of Florence*, in Karpova O. M., Kartashkova F. (eds.), *Life Beyond Dictionaries*, Cambridge Scholars Publishing, Newcastle upon Tyne: 122-132.
- Firth J. R. 1957, *Papers in Linguistics, 1934-1951*, Oxford University Press, Oxford.
- Ghadessy M., Henry A., Roseberry R. L. (eds.) 2001, *Small Corpus Studies and ELT: Theory and Practice*, John Benjamins, Amsterdam.
- Groom N. 2010, *Closed-class keywords and corpus-driven discourse analysis*, in Bondi M., Scott M. (eds.), *Keyness in Texts*, John Benjamins, Amsterdam: 59-78.
- Halliday M.A.K. 2004, *An Introduction to Functional Grammar*, Hodder Education, London.
- Levinson S. 2003, *Space in Language and Cognition. Explorations in Cognitive Diversity*, Cambridge University Press, Cambridge.
- Marmaridou A. S. 1989, *Proper names in communication*, «Journal of Linguistics», XXV (2): 355-372.
- Matlock T. 2004, *Fictive motion as cognitive simulation*, «Memory and Cognition», XXXII (8): 1389-1400.
- Merlini Barbaresi L. 2009, *The speaker's imprint in descriptive discourse*, in Tucker P., Radighieri S. (eds.), *Point of View: Description and Evaluation across Discourses*, Officina, Roma: 15-36.
- Phillips M. 1985, *Aspects of Text Structure: An Investigation of the Lexical Organization of Text*, North Holland, Amsterdam.
- Pierini P. 2008, *Opening a Pandora's box: Proper names in English phraseology*, «Linguistics Online», XXXVI (4): 43-52.
- Römer U. 2011, *Corpus research applications in second language teaching*, «Annual Review of Applied Linguistics», 31: 205-225.
- Samson C. 2011, *Ex-sacred territories on the Internet. Examples of space, identity and discourse interconnectedness in museum websites*, «Rassegna Italiana di Linguistica Applicata», I (1-2): 245-226.
- Samson C. 2012, *From cultural islands to popular sites. Semantic sequences typifying museum descriptions on the Web*, in Bongo G., Caliendo G. (eds.), *The Language of Popularisation: Theoretical and Descriptive Models*, Peter Lang, Bern: 139-161.
- Samson C. 2016a, *Moving between words: Keywords and phraseological networks in (English) guidebooks of Florence*, *Vestnik* 1(26), Ivanovo State University, 47-52.
- Samson C. 2016b (forthcoming), *Construing built heritage representations. A corpus-driven analysis of Florence guidebooks*, in Farina A., Samson C. (eds.), *Le*



- Passé dans le Présent: La Langue du Patrimoine / Past in Present: The Language of Heritage*, Florence University Press, Firenze.
- Scott M. 2010a, *Wordsmith Tools, 5.0.*, Oxford University Press, Oxford.
- Scott M. 2010b, *What can corpus software do*, in O'Keeffe, A., McCarthy, M. (eds.), *The Routledge Handbook of Corpus Linguistics*, Routledge, London: 136-151.
- Searle J. R. 1958, *Proper names*, «Mind», 67: 166-173.
- Sinclair J. M. 1991, *Corpus, Concordance, Collocation*, Oxford University Press, Oxford.
- Sinclair J. M. 1992, *The Automatic Analysis of Corpora*, Mouton de Gruyter, Berlin-New York: 379-397.
- Sinclair J. M. 1996, *The search for units of meaning*, «Textus», IX (1): 75-106.
- Sinclair J. M. 2000, *Current issues in corpus linguistics*, in Rossini Favretti R. (ed.), *Linguistica e Informatica. Corpora, Multimedialità e Percorsi di Apprendimento*, Bulzoni, Roma: 29-38.
- Sinclair J. M. 2001, *Preface*, in Ghadessy M., Henry A., Roseberry R. L. (eds.), *Small Corpus Studies and ELT: Theory and Practice*, John Benjamins, Amsterdam: VII-XV.
- Sinclair J. M. 2004, *Trust the Text. Language, Corpus, and Discourse*, Routledge, London.
- Smith L. 2006, *Uses of Heritage*, Routledge, Abingdon.
- Smyth F. 2008, *Constructing place, directing practice? Using travel guidebooks, Working paper. Sociology Centre for Narrative and Auto/Biographical Studies*, University of Edinburgh, Edinburgh.
- Stubbs M. 1983, *Discourse Analysis: The Sociolinguistic Analysis of Natural Language*, Oxford, Basil Blackwell.
- Teubert W. (ed.) 2007, *Text Corpora and Multilingual Lexicography*, John Benjamins, Amsterdam.
- Tognini-Bonelli E. 1996, *Corpus Theory and Practice*, Pescia, Tuscan Word Centre.
- Tognini-Bonelli E. 2001, *Corpus Linguistics at Work*, John Benjamins, Amsterdam.
- Williams G., Millon C. 2010, *Going organic: Building an experimental bottom-up dictionary of verbs in science*, in Dykstra A., Schoonheim T. (eds.), *Proceedings of the XIV EURALEX International Congress*, Fryske Akademy, Leeuwarden: 1251-1257.
- Williams G. 2000, *Collocational networks as the realization of a specialised textual environment*, DGFS, Philipps Universität, Marburg.
- Williams G. 2001, *Mediating between lexis and texts: collocational networks in specialised corpora*, «ASP», XXXI (33): 1-12.
- Williams G. 2012, *Bringing data and dictionary together: Real science in real dictionaries*, in Bolton A., Thomas Rowley-Jolivet S. E. (eds.), *Corpus-informed Research and Learning in ESP: Issues and Applications*, John Benjamins, Amsterdam: 219-240.

G. Diani

## On the language of Florence art museum websites: the Italian texts of the «virtual tour»

*Abstract:* This paper investigates the language of the Italian explanatory texts accompanying the artworks displayed on a virtual tour offered by Florence art museum websites. As shown by applied linguistics research conducted over the past twenty years, great attention has been paid to the way museums communicate with their users through a range of genres such as brochures, catalogues, exhibition press releases and announcements, wall captions. Using a corpus and adopting a discourse perspective, this study aims at investigating the main linguistic features which are used to disseminate the Florentine cultural heritage through the texts of the virtual tour in museum websites that still remain an under-explored area of linguistic inquiry. Some implications can be drawn from this study for research on the lexis of Italian and other languages referring to Italian cultural heritage.

*Keywords:* Florence art museum websites, virtual tour, on-line texts, linguistic analysis.

*Riassunto:* Il presente contributo intende avviare una riflessione sull'aspetto linguistico dei testi redatti in italiano che accompagnano il visitatore nella visita virtuale dei musei fiorentini in una prospettiva di indagine che affronta la diffusione sul web di informazioni legate al patrimonio culturale. Come testimoniano le ricerche condotte negli ultimi vent'anni in ambito linguistico, la comunicazione museale è stata al centro di un interesse crescente e il contributo più evidente si è avuto nel campo degli studi sul linguaggio di specifici generi promozionali quali brochure, catalogo, comunicato stampa della mostra/collezione del museo, didascalia a parete di un'opera. Servendosi degli strumenti di analisi che derivano tanto dalla linguistica dei corpora quanto dall'analisi del discorso, questo studio si propone di tracciare un profilo linguistico dei testi in italiano che descrivono il patrimonio

del museo attraverso una visita virtuale a cui poca attenzione è stata posta finora. L'analisi quantitativa e qualitativa offre spunti utili per ricerche sul lessico delle diverse lingue in relazione alla lingua italiana nell'ambito della divulgazione del patrimonio culturale italiano.

*Parole chiave:* siti museali fiorentini, visita virtuale, testi on-line, analisi linguistica.

## I. Introduction

Museums represent «cultural agents, trying to realize their basically educational aims in a rapidly changing cultural market» (Bondi 2009: 111). Communication plays a pivotal role in the way museums present their cultural identity (Drotner, Schrøder 2013). Catalogues, brochures, and a whole range of museum texts (Ravelli 2006) contribute to the dissemination of cultural knowledge. But also websites, as Bondi (2009: 113) rightly points out, offer «interesting material for an analysis of how museums communicate with their users, in the face of important technological and discursive changes brought about by the World Wide Web».

With the advent of digital media technologies, communication tools have increased their degree of multimodality, defined by Kress and van Leeuwen (2001: 20) as «the use of several semiotic modes in the design of a semiotic product or event». This is particularly the case with websites and their webpages that, as research has revealed (e.g. Shepherd, Watters 2004; Askehave, Ellerup Nielsen 2005; Garzone 2007; Adami 2015), are characterized by a multimodal and multisemiotic content echoing Ravelli's (2006) process of intersemiosis, i.e. the combination across different sign systems which produces discourse complexity. Complexity is also influenced by the presence of hyperlinks which allow website users to move around the pages, thus determining a «non-linear reading path» (Lemke 2005), and interrupting a traditional print text reading process (Samson 2014). Visual perception is particularly central in art appreciation and «provides the obvious justification for the role played by the visual mode in museum web texts» (Bondi 2009: 113). Virtual museums offer «an alternative to visiting physical showrooms with a huge access to works of art, retrospectives on the life of an artist, new acquisitions, current and past exhibitions, quizzes, databases, and so on» (Bernier 2002).

Substantial work has been devoted to issues of web design by marketing and museum experts alike (McLean 1997; Kotler, Kotler 1998; Cataldo, Paraventi 2007, to name but a few), as well as by practical networks (e.g. The MINERVA project, Minerva Ec 2003). Little linguistic research, however, has been paid to museum communication. Most of the existing research has looked at promotional texts such as exhibition press announcement, exhibition presentation, brochure, catalogue, walltext (Ravelli 1996, 2006;

Purser 2000; Hofinger, Ventola 2004; Atkins *et al.* 2008; Lazzeretti, Bondi 2012; Pierroux, Ludvigsen 2013; Maci 2015; Lazzeretti 2016). In this paper I focus on the museum website that still remains an under-explored genre (Bondi 2009; Pierroux, Skjulstad 2011; Samson 2011, 2014).

The purpose of a museum website is to promote the museum itself. It can easily be seen to belong to the colony of promotional genres (Bhatia 2004: 59). As pointed out by Bhatia (2004: 133), «a positive description and evaluation of the product, service or idea being promoted» are typical of promotional genres. The concepts of description and evaluation are therefore crucial for the scope of this paper, based on a linguistic study of the Italian texts included in the «virtual tour» (*Visita il museo*) section of the home page of Florence art museum websites. The aim of this paper is to investigate the main linguistic features which are used to disseminate the Florentine cultural heritage through on-line texts from a corpus and a discourse perspective. The present article begins by outlining the design of the corpus, as well as the methods applied in the analysis (section 2). In section 3 the results of the analysis are presented. The findings are summarized and conclusions are drawn in section 4.

## 2. Corpus and methods

### 2.1 Corpus

The study is based on the analysis of the Italian explanatory texts accompanying the artworks displayed on a «virtual tour» offered by Florence art museum websites. For the compilation of the corpus, the home page of the *Polo Museale Fiorentino* (<<http://www.polomuseale.firenze.it>>) was consulted. It includes the websites of Florence art museums. From there, all the museums available were taken, for a total of twenty-one, as listed in Table 1.

*Table 1 – List of the museums accounted in compiling the corpus.*

---

Galleria degli Uffizi
Galleria dell'Accademia
Galleria Palatina
Museo Nazionale del Bargello
Palazzo Pitti
Pitti Galleria d'Arte Moderna
Galleria del Costume
Museo degli Argenti
Museo delle Porcellane
Giardino di Boboli
Musei delle Cappelle Medicee
Museo di San Marco

Museo delle Carrozze  
 Chiesa e Museo di  
 Orsanmichele  
 Corridoio Vasariano  
 Museo di Palazzo Davanzati  
 Cenacolo di Ognissanti  
 Cenacolo di Andrea del Santo  
 Cenacolo di Fuligno  
 Cenacolo di Sant'Apollonia  
 Chiostro dello Scalzo

---

Each museum website forming the corpus offers a section dedicated to a «virtual tour» of the museum providing a room-by-room tour of the whole museum. The browser/visitor can navigate from room to room by clicking map locations or by following arrow links that connect the rooms and floors of the museum. Each room is provided with explanatory texts accompanying the artworks preserved inside, as shown in Figure 1 below. The corpus comprises all the texts available, for a total of 24,599 words. Since I focused on the linguistic features of the texts under examination, images and any other graphical elements were excluded, as well as any texts organised by lists of links or scattered around the webpages.

Figure 1 – A room virtual tour.

Galleria dell'Accademia  
 Visite | Informazioni | Contatti | Visita | Museo | Spazio Eventi e mostre

LUNEDÌ e MERCOLEDÌ  
 9.00 - 19.00

Tribuna del David

PIANO TERRA

Tribuna del David

La Galleria dell'Accademia ospita dal 1873 il David di Michelangelo, una delle sculture più famose al mondo e tappa imprescindibile per chiunque visiti la città di Firenze.

La statua, realizzata in marmo bianco e alta oltre 4 metri, fu scolpita da Michelangelo tra il 1501 e il 1504 per essere collocata presso la Cattedrale di Firenze, ma l'innovazione bellica del regno francese, con il governo francese David rifugiato in esilio a nell'esperto parere di un cune dell'armata, fece prendere agli amministratori della città una diversa collocazione, individuata nella Piazza della Signoria, davanti a Palazzo Vecchio, dove è rimasta per oltre 250 anni.

Esistono degli ideali della Firenze repubblicana nei primi anni del XVI secolo, la grande figura del David è dunque progressivamente assorbita di perfezione formale, caratterizzata da una bellezza fuori del tempo resa laboriosa e palpabile dalla tensione della muscolatura, dalla partecipazione emotiva rivelata dalla sguardo angosciato, dalla presa in tensione che prelude al movimento.

La scultura è stata in questi mesi oggetto di una pulitura e di un restauro a carattere aperto, così da consentire la visione dell'opera da parte dei visitatori del Museo anche durante i lavori di restauro, che si è conclusi nel maggio 2004.

## 2.2 Methods

The methodology adopted for this study combines a corpus and a discourse perspective. The corpus-based approach seems to be particularly useful in analyzing specific lexico-syntactic choices characterizing the texts under examination, although, when looking at evaluative expressions, a discourse-based approach becomes essential. Thus, corpus evidence was used

bearing in mind Römer's (2008: 126) claim that evaluative lexis cannot be identified with quantitative methods alone and her choice to move «from an automatic computer-based to a mainly manual but in part computer-assisted [...] type of analysis». This echoes Hunston's (2011: 4) statement that «evaluative language is more suited to text-based than to corpus-based enquiry».

The analysis was based on a preliminary process of keyword identification. The keyword list is automatically generated by the Keywords programme which is part of the WordSmith Tools suite of corpus analysis software (Scott 2008, version 5.0). I started with an overview of the keyword list obtained comparing the corpus with a reference corpus of general Italian: the CORIS/CODIS (*CORpus di Italiano Scritto/Corpus Dinamico di Italiano Scritto*, Rossini Favretti 2000). From such a keyword list I selected the Salient Grammatical Words (SGWs, in Gledhill's 2000 terms), featuring among the highest scoring keywords in the list within the first 50 positions. For each keyword, I extracted a random sample of concordance lines. The concordances were then submitted for analysis, in order to investigate the typical lexico-syntactic phraseological arrangements they are involved in (i.e. typical collocates and grammar patterns).

Two arguments may be put forward in favour of SGWs as the most suitable starting point to identify lexico-syntactic phenomena. First, the statistical algorithm that operates as the basis for the Keywords programme tends as a norm to remove grammar words from the output keyword list. This is an obvious consequence of the fact that grammar words are the most frequent word-forms found in any corpus. Accordingly, the very appearance of a 'closed-class' keyword (Groom 2010: 59) in a keyword list testifies to its statistical significance as regards the research corpus. The second argument is that, arguably, SGWs ensure a better data coverage as opposed to, for example, lexical keywords. Since, as just mentioned, grammar words are the most commonly found in any given corpus, it may be argued that an analysis starting from a selection of such high frequency word-forms will account for a proportion of the overall corpus data that is significantly larger than the one accounted for by a comparable or even larger selection of lexical keywords (Sinclair 1999). More so, if considering that SGWs are not to be examined as isolated items but together with their lexical and syntactic environments as attested by concordance lines. In other words, the results obtained by an analysis based on grammatical keywords may be expected to provide a potentially more accurate and more representative picture of the overall lexico-syntactic phenomena occurring in a specialized corpus as a whole.

### 3. Results and discussion

The list of 4,951 keywords as automatically generated by the WordSmith Tools software has been reduced to the top 50 word-forms. Starting

from a general overview of the list, the first 50 keywords were divided into three groups according to general criteria as follows. A first group includes grammar words that are the most frequent word-forms in the corpus, occupying the top twenty places in the list: definite articles (*la, il, le, gli*), ranking respectively at the 3<sup>rd</sup>, the 4<sup>th</sup>, the 11<sup>th</sup>, and the 19<sup>th</sup> positions in the list; indefinite articles *un, una* rank 8<sup>th</sup> and 17<sup>th</sup> respectively; prepositions (*di, del, della, dei, nel, nella, per, con*), whose ranking in the list is 1<sup>st</sup>, 2<sup>nd</sup>, 5<sup>th</sup>, 9<sup>th</sup>, 12<sup>th</sup>, 20<sup>th</sup>, 10<sup>th</sup>, 13<sup>th</sup>. A second group includes keywords that are fairly clearly related to the physical objects housed by the art museum (*opere, dipinti, collezione, pittura, scultura*) or to the museum's space (*sala, piano*). A third group includes keywords that are related to the language of evaluation. It comprises evaluative adjectives like *grande* (meaning 'remarkable'), *importanti, famose*.

As mentioned in 2.2, I adopt the SGW approach to keywords as a starting point to investigate the corpus through concordance analysis, especially in view of the phenomena of lexical co-occurrence. Following Gledhill (1996), I decided to focus on the prepositions *di/del/della/dei* as they are symptomatic of longer stretches of regular phraseology. The results of an investigation of a random sample of 100 concordance lines for each preposition attest the repeated occurrence of the grammar pattern 'n + prep + n' (es. *parte di un antico convento, ultimo piano del grande edificio, capolavoro della pittura italiana*). Within this pattern, attention was paid to repeated word-clusters (Scott 2001) of the prepositions under examination, that is, words that are found repeatedly in their company. Tables 2-5 below include all the most frequent clusters of the prepositions under scrutiny, together with their respective frequencies (within the first five positions).

Table 2 – Top 5 clusters of *di* in the corpus.

<i>cluster</i>	<i>nr. of occurrences</i>
opere di	32
dipinti di	17
di michelangelo	16
opera di	15
parte di	12

Table 3 – Top 5 clusters of *del* in the corpus.

<i>cluster</i>	<i>nr. of occurrences</i>
metà del	30
del secolo	24
fine del	16
piano del	11
parte del	8

Table 4 – Top 5 clusters of della in the corpus.

<i>cluster</i>	<i>nr. of occurrences</i>
della pittura	13
della galleria	12
della scultura	12
della collezione	11
interno della	5

Table 5 – Top 5 clusters of dei in the corpus.

<i>cluster</i>	<i>nr. of occurrences</i>
uno dei	18
dei Medici	11
dei pittori	7
dei più	6
dei capolavori	4

The most frequent clusters of the prepositions examined may be divided into three groups. A first group includes words related to the cultural heritage of the museum and the historical importance of its artists, paintings and sculptures (*opere, dipinti, pittura, scultura, collezione, pittori, capolavori, Medici, Michelangelo, metà del secolo, fine del secolo*), as shown in the following examples:

- (1) Nei vasti locali del convento sono collocate importanti **opere** della prima **metà del** Cinquecento.
- (2) La sala ospita un **capolavoro della pittura** italiana del XIV secolo.
- (3) Nella sala sono esposti **dipinti**, databili tra la prima **metà del** secolo XII e gli inizi del secolo XIV.

A second group includes words, having a deictic function (*parte di, parte del, interno della*) that highlight one of the main functions of the museum's «virtual tour», based on the cognitive process of perception in space. The browser/visitor is informed about sections of the museum as regions of space including the museum as a whole as well as directed to that section by a virtual eye, which simultaneously allows «reading-as-such», in this case the written text within the webpage, and the «navigating mode», entailing a shift from one descriptive museum webpage to another (Samson 2011). This is exemplified by the following excerpts:

- (4) Dalle finestre si gode una splendida vista sull'Arno e le colline circostanti e si può osservare **parte del Corridoio vasariano**.
- (5) Entrando **all'interno della sala** il visitatore si trova di fronte tre opere fondamentali per ricostruire la formazione e l'attività giovanile di Leonardo da Vinci.



A deictic function is also revealed in the cluster **primo/secondo piano del Palazzo/Galleria** pointing to fixed specific areas of the building and guiding the browser through the museum's space, as shown in the following examples:

(6) **Disceso al primo piano della Galleria**, il visitatore attraversa alcuni ambienti ancora da restaurare per giungere infine nella splendida sala che si affaccia, attraverso grandi finestre, da un lato sull'Arno e dall'altro sul piazzale degli Uffizi.

(7) **Al secondo piano del Museo**, in cima alla costruzione, si può osservare un notevole, unico panorama di Firenze.

A third group include words that are exploited in the expression of evaluation (*uno dei, dei più*) when found in the grammar pattern 'number + prep + NPs/superlative', as illustrated in the following examples:

(8) Di fronte all'ingresso della sala è collocato **uno dei dipinti più famosi** della Galleria [...].

(9) A Gentile da Fabriano, considerato **uno dei maggiori pittori italiani**, appartengono due delle opere più famose della sala.

(10) [...] È **uno dei più famosi musei** del mondo per le sue straordinarie collezioni di dipinti e di statue antiche.

As the examples show, the explanatory texts are characterized by evaluative language stressing the importance and uniqueness of the museum. The browser/visitor is mainly provided with some details of the collections and/or artworks in the museum which are described with highly evaluative language. The grammar pattern 'number + prep + NPs/superlative' underscores the high value of the artworks preserved in the museum through the use of superlative adjectives. This evaluative pattern contributes to create a valuable voice for the museum, showing the degree of expertise of the museum in the field of Arts.

Similarly, the highly evaluative function of the explanatory texts examined is exemplified by the recurring grammar patterns 'prep + noun + evaluative adjective/ noun + verb', as shown in examples (11) and (12):

(11) Tra le opere **più famose del Mantegna** sono [...].

(12) Tra le sculture di **maggior importanza** esposte nella sala si segnalano [...].

Or by «evaluative verb + (adj) + noun» as in (13) to (15):

(13) Nelle sale del primo piano si può **ammirare un insieme di grandiosi politici tardo gotici** [...].

(14) Un posto a parte **merita il bel busto con testa di Antinoo** [...].

(15) Nel vasto ambiente **spiccano le tre tavole cuspidate** di Duccio di Boninsegna, Cimabue e Giotto [...].

Of the more relevant aspects of the findings, it is the dominance of an evaluation of importance that is the most striking (albeit predictable) in the texts examined. Importance is attached to the museum, to the art object or to its features as a way of attracting the browser/visitor's attention to it. These 'publicity' texts, although serving an informative function, reflect the promotional nature of the museum website through a significant use of evaluative language, contributing to the positive image of the museum itself. The quality of the museum and its artworks are qualified as *grande*, *importanti*, *famose* that were found scoring at the highest rankings in the keyword list within the first 50 positions: more specifically, they rank 25<sup>th</sup>, 38<sup>th</sup> and 45<sup>th</sup>, as opposed to their corresponding singular and plural forms (*grandi*, *importante*, *famosa*, *famoso*, *famosi*) that rank outside the 50 keyword list that was adopted as a basis for the present study.

They score respectively as the 309<sup>th</sup> (*famosi*), 623<sup>rd</sup> (*grandi*), 670<sup>th</sup> (*importante*), 1,118<sup>th</sup> (*famosa*), 1,154<sup>th</sup> (*famoso*) in the complete keyword list. Below are 10 lines each of *grande*, *importanti* and *famose* that illustrate the use of these evaluative adjectives conveying quality and importance:

### **grande**

di codici miniati. Nelle sue opere, di	grande eleganza, il sentimento
artista, detto dello Scalzo, fondata nel 1376. Il	grande pittore fiorentino
L'allestimento di questa sala è di	grande efficacia didattica,
la sala espone un'opera di	grande interesse, il Polittico
esistenziali che si riflette in opere di	grande suggestione anche per i
e piani di appoggio, raggiungendo effetti di	grande raffinatezza attraverso la
è opera rimasta incompiuta, ma di	grande complessità concettuale
nel gennaio del 1506, che suscitò una	grande ammirazione negli
i più rilevanti e prestigiosi del	grande collezionismo italiano
figurativo personalissimo e vivace, che ebbe	grande impatto sui pittori coevi

### **importanti**

Nei vasti locali del convento sono collocati	importanti dipinti della prima metà del
no. Nel complesso vasariano sono ospitate altre	importanti raccolte: la Collezione Contini
sala del Caravaggio La sala ospita tre	importanti dipinti del Caravaggio. Sulla parete
Accanto a quelle di Leonardo troviamo esposte	importanti pale d'altare di tre maestri attivi

le statue disposte nel corridoio vi sono	importanti sculture antiche come Amore e
dell'episodio. Sono esposti anche alcuni suoi	importanti ritratti, sia di tono aulico, come il
Gotico internazionale Sono qui raccolte alcune	importanti opere eseguite da Lorenzo
diventato Museo Nazionale, alcune delle più	importanti sculture del Rinascimento, tra le
ntemporanei. Nelle vetrine sono visibili alcune	importanti terracotte, tra cui si segnala il bozze
e alle maioliche italiane, la collezione annovera	importanti esemplari di ceramica spagnola di
<b>famose</b>	
anticaglie, citate da Plinio, delle più	famose: il Laconte, l'Ercole e il
italiani, appartengono due delle opere più	famose della sala: Quattro Santi dal
copertura originaria. È una delle più	famose della Galleria perchè ospita alcuni tra
opere di Sandro Botticelli, le più	famose sono La Primavera e La Nascita di
il Correggio. Tra le opere più	famose del Mantegna sono la piccola Madonna delle
famose	di Michelangelo, una delle sculture più famose al mondo e tappa imprescindibile per
dei Lanzi, numerose riduzioni di sue opere	famose quali il Mercurio volante, in più
detto Il Riccio la raccolta contiene opere	famose come il Satiro con Vaso e conchiglia,
due angeli, uno delle opere più	famose della Galleria. Nella sala sono esposti
cicala (V secolo). Le opere più	famose e rappresentative dell'oreficeria longob

What these lines show is that *grande* is nearly always in the pattern 'di grande + noun' and appears as the most typical attributive adjective of nouns conveying quality and value of the artworks preserved in the museum: *eleganza, efficacia, suggestione, raffinatezza, ammirazione, impatto*. Similarly, *importanti* and *famose* appear as a relevant collocates of nouns referring to the collections of the museum (*dipinti, raccolte, sculture, ritratti, opere*).

An in-depth reading of all the texts featured by the corpus revealed the presence of other, less frequent, evaluative adjectives that emphasize the quality of the museum, contributing to the construction of its positive image. In order to highlight it, the artworks are qualified as *straordinaria/o/eli, notevole/i, preziosa/o/ile, splendorosi/i, maggiori/i, bellissima/o/a*, as shown in the following examples:

- (16) La sala accoglie una **preziosa** selezione di capolavori senesi del Trecento. I due *straordinari* dipinti che si fronteggiano provengono dal duomo di Siena.
- (17) Nel primo ambiente è esposto un nucleo di epigrafi romane di **notevole** importanza storica.
- (18) Assai più ricca e formata da **splendidi** esemplari, la vetrina degli Avori [...].
- (19) La pittura veneta della seconda metà del Cinquecento è qui rappresentata da alcuni dei suoi **maggiori** protagonisti.
- (20) Al secolo XV risalgono invece la **bellissima** Scacchiera, che reca sul verso il gioco del 'tric-trac' [...].

#### 4. Concluding remarks

The results confirm that the promotional intent is very strong in the Italian explanatory texts accompanying the «virtual tour» offered by Florence art museum websites pursued through a combination of description and evaluation. Because of that intent, the language of evaluation is almost inevitable. Indeed, we may expect the vast majority of the texts accompanying the «virtual tour» of the museum to include linguistic elements that do not simply identify physical objects, but rather signal attitudinal reactions to the objects.

As emerged from the analysis, the texts under scrutiny are characterized by an overwhelming use of evaluative adjectives expressing global praise for the museum. An evaluation of quality and importance is attached to the museum, its featured artists and their artworks as a way of attracting online audience and increasing the number of potential visitors. Typical examples of evaluative language are adjectives as *grande/i*, *importante/i*, *famoso/a/i/e*, *straordinario/ole/i*, *notevole/i*, *prezioso/a/i/e*, *splendido/i*, *maggiore/i*, *bellissimo/a*, nouns as *eleganza*, *raffinatezza*, *ammirazione*, and verbs as *ammirare*, *spiccare*, *meritare*.

The analysis also revealed recurrent lexical and grammar patterns. For instance, the nouns *opere*, *dipinti*, *collezione*, *pittura*, *scultura*, *capolavori*, appearing as the most frequent clusters of the prepositions *di/del/della/dei*, occur in descriptive paragraphs of the texts examined, introducing cultural and historical details. The nouns *parte* and *piano* – occupying the top five places in the cluster list – have a deictic function in the descriptive parts of the museum's «virtual tour», guiding the browser/visitor through its space. The grammar patterns like 'number + prep + NPs/superlative' and 'prep + noun + evaluative adjective/noun + verb', on the other hand, highlight the high value of the artworks preserved in the museum through the use of superlative or evaluative adjectives.

The findings shed light on how museums and their websites try to make their 'products' appealing (Dann 1996) with the aim of representing

the expertise of the museum and, thus, contributing to the construction of its identity as a cultural institution. Because of the small corpus used in this study, there are of course limitations to the generalisability of the results. It would be useful to compare similar websites as representative of Italian cultural heritage so as to shed more light on the trends emerged from the present study as well as on other tendencies within and beyond those found above. In addition, a comparative analysis with other languages may be worthwhile.

Some implications can be drawn from this study for researchers interested in the study of lexis of Italian and other languages referring to Italian cultural heritage (specialist translators, lexicographers). Cultural heritage is considered as one of the foundations of cultural identity of a community, and language plays a vital role in the expression and transmission of heritage. As UNESCO has stated (2003: 1), «all intangible cultural domains [...] depend on language for their day-to-day practice and inter-generational transmission». Thus studying the lexis of different languages used to describe cultural heritage may be the starting point for the elaboration of tools (dictionaries, reference materials on technical translations) that are able to convey such knowledge.

## References

- Adami E. 2015, *What's in a Click? A Social Semiotic Framework for the Multimodal Analysis of Website Interactivity*, «Visual Communication», XIV (2): 133-153.
- Askehave I., Ellerup Nielsen A. 2005, *What are the Characteristics of Digital Genres? Genre Theory from a Multimodal Perspective*, in *Proceedings 38th Hawaii International Conference on System Sciences*, «HICSS», 4: 98-108.
- Atkins L. J., Velez L., Goudy D., Dunbar K. N. 2008, *The Unintended Effects of Interactive Objects and Labels in the science Museum*, «Science Education», XCIII (1): 161-184.
- Bernier R. 2002, *Uses of Virtual Museums: The French Viewpoint*, *Museums and the Web*, <<http://www.museumsandtheweb.com/mw2002/papers/bernier/bernier.html>> (01/2016).
- Bhatia V. K. 2004, *Worlds of Written Discourse*, Continuum, London.
- Bondi M. 2009, *Perspective and Position in Museum Websites*, in Radighieri S., Tucker P. (eds.), *Point of View: Description and Evaluation across Discourses*, Officina Edizioni, Roma: 113-127.
- Cataldo L., Paraventi M. 2007, *Il Museo Oggi: Linee Guida per una Museologia Contemporanea*, Hoepli, Milano.
- Dann G. 1996, *The Language of Tourism. A Sociolinguistic Perspective*, CBA, Wallingford.
- Drotner K., Schröder, K. C. 2013, *Museum Communication and Social Media. The Connected Museum*, Routledge, London.
- Garzone G. 2007, *Genres, Multimodality and the World Wide Web: Theoretical Issues*, in Garzone G., Poncini G., Catenaccio P. (eds.), *Multimodality in Corpo-*

- rate *Communication: Web Genres and Discursive Identity*, FrancoAngeli, Milano: 5-30.
- Gledhill C. 1996, *Science as a Collocation. Phraseology in Cancer Research Articles*, in Botley S., Glass J., McEnery T., Wilson A. (eds.), *Proceedings of Teaching and Language Corpora 1996*, «UCREL Technical Papers», 9: 108-126.
- Gledhill C. 2000, *Collocations in Science Writing*, Gunter Narr, Tübingen.
- Groom N. 2010, *Closed-class Keywords and Corpus-driven Discourse Analysis*, in Bondi M., Scott M. (eds.), *Keyness in Texts*, John Benjamins, Amsterdam: 59-78.
- Hofinger A., Ventola E. 2004, *Multimodality in Operation: Language and Picture in a Museum*, in Ventola E. (ed.), *Perspectives on Multimodality*, John Benjamins, Amsterdam: 193-209.
- Hunston S. 2011, *Corpus Approaches to Evaluation*, Routledge, London.
- Kotler N., Kotler P. 1998, *Museum Strategy and Marketing: Designing Missions, Building Audiences, Generating Revenue and Resources*, Jossey-Bass, San Francisco, CA.
- Kress G. R., van Leeuwen T. 2001, *Multimodal Discourse*, Arnold, London.
- Lazzeretti C., Bondi M. 2012, *A Hypnotic Viewing Experience. Promotional Features in the Language of Exhibition Press Announcements*, «Pragmatics», XXII (4): 567-589.
- Lazzeretti C. 2016, *The Language of Museum Communication. A Diachronic Perspective*, Palgrave Macmillan, Basingstoke.
- Lemke J. L. 2005, *Multimedia Genres and Traversals*, «Folia Linguistica», XXXIX (1-2): 45-56.
- Maci M. S. 2015, *For Your Eyes Only: How Museum Walltexts Communicate East and West. The Case of the Peggy Guggenheim Foundation*, «Lingue Culture Mediazioni/Languages Cultures Mediation», II (1): 135-155.
- McLean F. 1997, *Marketing the Museum*, Routledge, London.
- Minerva Europe 2003, *Handbook for Quality in Cultural Web Sites*, <<http://www.minervaeurope.org/publications/qualitycommentary/qualitycommentary-050314final.pdf>> (01/2016).
- Pierroux P., Ludvigsen S. 2013, *Communication Interrupted: Textual Practices and Digital Interactives in Art Museums*, in Schröder K., Drotner K. (eds.), *The Connected Museum: Social Media and Museum Communication*, Routledge, London: 153-176.
- Pierroux P., Skjulstad S. 2011, *Composing a Public Image Online: Art Museums and Narratives of Architecture in Web Mediation*, «Computers and Composition», XXVIII (3): 205-214.
- Purser E. 2000, *Telling Stories: Text Analysis in a Museum*, in Ventola E. (ed.), *Discourse and Community: Doing Functional Linguistics*, Gunter Narr, Tübingen: 169-198.
- Ravelli L. 1996, *Making Language Accessible: Successful Text Writing for Museum Visitors*, «Linguistics and Education», 8: 367-387.
- Ravelli L. 2006, *Museum Texts: Communication Frameworks*, Routledge, London.
- Römer U. 2008, *Identification Impossible? A Corpus Approach to Realisations of Evaluative Meaning in Academic Writing*, «Functions of Language», XV (1): 115-130.

- Rossini Favretti R. 2000, *Progettazione e Costruzione di un Corpus di Italiano Scritto: CORIS/CODIS*, in Rossini Favretti R. (ed.), *Linguistica e Informatica. Multi-medialità, Corpora e Percorsi di Apprendimento*, Bulzoni, Roma: 39-56.
- Samson C. 2011, *Ex-sacred Territories on the Internet. Examples of Space, Identity and Discourse Interconnectedness in Museum Websites*, «Rassegna Italiana di Linguistica Applicata», 1-2: 245-262.
- Samson C. 2014, *From Cultural Islands to Popular Sites. Semantic Sequences Typifying Museum Descriptions on the Web*, in Bongo G., Caliendo G. (eds.), *The Language of Popularization / Die Sprache der Popularisierung*, Peter Lang, Bern: 139-161.
- Scott M. 2001, *Comparing Corpora and Identifying Key Words, Collocations and Frequency Distributions through the WordSmith Tools Suite of Computer Software*, in Ghadessy M., Henry A., Roseberry R. (eds.), *Small Corpus Studies and ELT*, John Benjamins, Amsterdam: 47-67.
- Scott M. 2008, *WordSmith Tools (Version 5.0)*, Oxford University Press, Oxford.
- Shepherd M., Watters C. 2004, *Identifying Web Genre: Hitting A Moving Target*, in *Proceedings of the WWW2004 Conference. Workshop on Measuring Web Search Effectiveness: The User Perspective*, ACM, New York.
- Sinclair J. 1999, *A Way with Common Words*, in Hasselgard H., Oksefjell S. (eds.), *Out of Corpora: Studies in Honour of Stig Johansson*, Rodopi, Amsterdam: 157-179.
- UNESCO 2003, *Linguistic Diversity in Selected UNESCO's Normative Texts*, <<http://www.unesco.org/new/en/culture/themes/endangered-languages/linguistic-diversity-in-unesco-normative-texts/>> (10/2016).

D. Maurel  
N. Friburger  
I. Eshkol-Taravella

## Tourisme culturel sur Internet. Les noms propres des éditions originales de Rabelais

*Résumé:* Pour exploiter le contexte culturel des œuvres présentées dans un cadre touristique de développement régional, l'idée est venue de créer un site permettant d'associer la navigation dans les œuvres via les noms propres et le déplacement touristique en Région Centre, terre de Ronsard et de Rabelais, mais aussi de Gargantua et de Pantagruel. Le projet Renom vise à permettre une meilleure exploitation de ce patrimoine en fournissant à l'internaute des outils pour naviguer à partir des noms de personne ou de lieu, et en l'invitant à compléter sa consultation par une visite de la Touraine. Cet article décrit l'ajout d'une recherche de noms propres à une chaîne de transcription de textes de la Renaissance. La recherche de noms propres dans des textes de la Renaissance présente deux difficultés qui la complexifient: une grande variété d'écriture, l'orthographe n'étant pas encore fixée, et la présence d'un grand nombre de balises XML de formatage TEI, car les textes gardent trace de la présentation de l'édition originale. L'objectif est d'ajouter à ce balisage TEI un balisage supplémentaire des noms propres avec leurs extensions à gauche ou parfois à droite. Nos programmes ont apporté une aide précieuse aux experts annotateurs et sont aujourd'hui intégrés à la chaîne de transcription.

*Mots-clés:* tourisme culturel, noms propres, textes de la Renaissance, cascades de graphe, CasSys.

*Abstract:* Allo scopo di sviluppare il turismo culturale e di valorizzare il patrimonio della regione Centre-Val de Loire, terra di Ronsard e di Rabelais, ma anche di Gargantua e Pantagruel, è nato il progetto Renom, che ha realizzato un sito-web attraverso il quale viene data la possibilità all'internauta di navigare in opere letterarie regionali, per esempio nei testi originali di Rabelais e nelle mappe di viaggio di Gargantua, a partire da nomi di persona o di luogo, con l'invito di completare la consultazione con una visita



reale della Touraine. Questo contributo descrive nello specifico l'aggiunta di una ricerca dei nomi propri (Named Entity Recognition) ad un processo di trascrizione di testi del Rinascimento. La ricerca dei nomi propri nei testi di quest'epoca presenta due difficoltà che la rendono particolarmente complessa: una grande varietà di grafie, dato che l'ortografia non era ancora stabilizzata, e la presenza di un numero notevole di etichette XML-TEI, poiché i testi conservano traccia della presentazione dell'edizione originale. È stata aggiunta a questa etichettatura TEI una etichettatura supplementare di nomi propri con le loro estensioni a sinistra o, talvolta, a destra. Un prezioso aiuto è stato fornito agli annotatori esperti, implementando il programma gratuito CasSys per processare i testi con cascate di grafi Unitex e costruendo quattro dizionari e otto cascate specifiche, oggi integrate al processo di trascrizione.

*Parole chiave:* turismo culturale, nomi propri, testi del Rinascimento, cascate di grafi, CasSys.

## I. Introduction

Dans le cadre de ce qu'on appelle aujourd'hui les *Humanités numériques*, les bibliothèques et les centres de recherche sur les documents anciens souhaitent mettre à disposition du public des textes numérisés et, souvent, enrichis. C'est le cas par exemple du site Gallica<sup>1</sup> de la Bibliothèque Nationale de France ou du site des Bibliothèques Virtuelles Humanistes (BVH)<sup>2</sup> du Centre d'Études Supérieures de la Renaissance (CESR) de l'Université François-Rabelais de Tours. Pour réaliser le site des BVH, le CESR a conçu une 'chaîne de transcription' pour réaliser la transcription des ouvrages originaux de la Renaissance en sa possession. Cette chaîne comprend la transcription elle-même, sa relecture, son enrichissement par des balises respectant la *Text Encoding Initiative*<sup>3</sup> (TEI) afin d'annoter le format des ouvrages originaux dans le texte lui-même, ainsi que la correction (toujours précisée par des balises TEI) de certaines séquences présentes dans le document<sup>4</sup>.

Pour exploiter le contexte culturel des œuvres présentées dans un cadre touristique de développement régional, l'idée est venue de créer un site permettant d'associer la navigation dans les œuvres via les noms propres et le déplacement touristique en Région Centre, terre de Ronsard et de Rabelais, mais aussi de Gargantua et de Pantagruel. Le projet Renom<sup>5</sup> vise à permettre une meilleure exploitation de ce patrimoine en fournissant à l'internaute des

<sup>1</sup> <<http://gallica.bnf.fr/>> (06/2017).

<sup>2</sup> <<http://www.bvh.univ-tours.fr/>> (06/2017).

<sup>3</sup> <<http://www.tei-c.org/index.xml>> (06/2017).

<sup>4</sup> Entre autres l'article aujourd'hui écrit *l'* qui est collé au mot dans les corpus étudiés. Un exemple est donné dans le Tableau 1.

<sup>5</sup> <[http://tln.li.univ-tours.fr/Tln\\_Renom.html](http://tln.li.univ-tours.fr/Tln_Renom.html)> (06/2017).

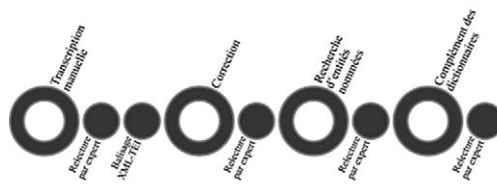
outils pour naviguer à partir des noms de personne ou de lieu, et en l’invitant à compléter sa consultation par une visite de la Touraine (par exemple, le musée de La Devinière, s’il s’agit de Rabelais, mais aussi les châteaux qu’il mentionne). Il sera même possible de ‘localiser’ virtuellement des lieux tels que Thélème (qui n’a jamais existé) en fonction des données fournies par le texte. Le premier site Renom (voir la Figure 1), publié le 6 février 2014, permet le lien entre le texte rabelaisien et les cartes géographiques; le site à venir ajoutera des pointeurs vers les lieux touristiques aux alentours, en lien avec le personnage (château de Chinon) ou l’auteur (la Devinière). Concrètement, les programmes développés dans le cadre du projet Renom effectuent une recherche des noms propres et des informations principales les concernant (qui sont ces personnes ou personnages? où sont situés ces lieux?). Cette recherche est aujourd’hui intégrée à la chaîne de transcription du CESR (voir la Figure 2) et permet aussi l’enrichissement des dictionnaires du Centre.

Figure 1 – Une page du site Renom.



Après une présentation des données et des outils (section 2), viendra la méthodologie suivie (section 3) et, pour finir, l’évaluation des résultats obtenus (section 4).

Figure 2 – La chaîne de transcription du CESR.



## 2. Données et outils

### 2.1 Le corpus

Les œuvres qui composent le corpus du travail sont:

- les *Discours fantastiques*, de Justin Tonnelier (édition de 1566);
- le *Courtisan*, de Baldassare de Castiglione (édition de 1538);
- le *Voyage de Tours* (édition de 1560) et *l'Élégie sur les troubles d'Amboise* (édition de 1563) de Pierre de Ronsard;
- *Gargantua* (édition de 1542), *Pantagruel* (édition de 1542), le *Tiers Livre* (éditions de 1546 et de 1552), le *Quart Livre* (éditions de 1548 et de 1552) et *Brève déclaration* (édition de 1552), de François Rabelais.

La chaîne de transcription du CESR produit un corpus très particulier par sa fidélité complète à la mise en page de l'édition originale, transcrite en suivant la norme XML-TEI. Le texte ne comporte pas d'accents, ou très exceptionnellement. La majuscule n'est pas toujours présente. Pour économiser de la place, les éditeurs de l'époque n'hésitaient pas à couper les mots, à les orthographier différemment suivant leur position, à les abrégéer ou à les finir après une parenthèse à la fin de la ligne suivante. De plus, la transcription mémorise les modifications de taille de certaines lettres ainsi que les bas de page, numéro de page, etc. Pour faciliter la lecture des ouvrages, un mot coupé en bas de page peut être réécrit en haut de la page suivante dans la marge. Tout cela est présent dans le corpus transcrit. De plus, la deuxième étape de la chaîne ajoute quelques corrections orthographiques, principalement une apostrophe pour décoller l'article du mot qu'il détermine ou l'écriture développée d'une forme abrégée. Le Tableau 1 présente un exemple pour chacune de ces situations<sup>6</sup>.

Les annotations des noms propres doivent venir s'ajouter au texte balisé actuel. Pour trouver les noms propres, il faut donc un outil permettant à la fois de préserver les balises actuelles et de les prendre en compte si nécessaire dans l'annotation. Nous avons donc choisi de ne pas utiliser d'outils XML classiques, mais des cascades de graphes dont les premiers graphes traitent le balisage XML-TEI du format d'impression.

Tableau 1 – Exemples de balises TEI présentes dans le corpus.

Taille de	<hi rend="larger">E</hi>N ceste mesme heure Gargan
caractères, puis	<lb rend="hyphen"/>tua [...]
coupure de mot	<lb/> fut adverty [...] comment Picrocho-
sans et avec tiret	<lb rend="hyphen"/>le estoit rempare a la Rocheclermaud

<sup>6</sup> C'est nous qui ajoutons les coupures de texte pour rendre les exemples plus lisibles.

Complément en fin de ligne suivante <sup>a</sup>	<pre>&lt;l&gt;Clers, basauchiens mangeurs du popu- &lt;lb rend="hyphen"/&gt;&lt;hi rend="bottom"&gt;(laire.&lt;/hi&gt;&lt;/l&gt; &lt;l&gt;Officiaulx, scribes, &amp;amp; pharisiens&lt;/l&gt;</pre>
Bas de page	<pre>&lt;lb/&gt;je luy cede la mestayrie de la Pomar- &lt;fw place="bot-center" type="sig"&gt;M ij&lt;/fw&gt; &lt;pb n="180" xml:id="B360446201_B343_2_0180"/&gt; &lt;fw place="top-left" type="pageNum"&gt;[90v]&lt;/fw&gt; &lt;lb rend="hyphen"/&gt;diere, a perpetuite pour luy &amp;amp; les siens &lt;lb/&gt;possedable en franc alloy.</pre>
Correction <sup>b</sup>	<pre>&lt;lb/&gt;[...] saint &lt;lb/&gt;Thomas &lt;choice&gt;&lt;orig&gt;Langloys&lt;/orig&gt;&lt;reg&gt;L'angloys&lt;/ reg&gt;&lt;/choice&gt; voulut bien pour &lt;lb/&gt;yceulx mourir</pre>

<sup>a</sup> La parenthèse apparaît, dans le texte original, en fin de ligne, mais le transcripateur l'a déplacée au début: Clers, basauchiens mangeurs du popu- Officiaulx, scribes, &amp; pharisiens (laire).

<sup>b</sup> Dans le texte original, on avait seulement la forme *Langloys*: [...] saint Thomas Langloys voulut bien pour yceulx mourir.

Prenons l'exemple suivant où *Panarge* est corrigé par le transcripateur en *Panurge*<sup>7</sup>:

```
<choice>
  <sic>Panarge</sic>
  <corr>Panurge</corr>
</choice>
neust faict esvanouyr
```

Le balisage des noms propres ne devait concerner, ni *Panarge*, ni *Panurge*, mais l'ensemble de la correction :

```
<persName>
  <choice>
    <sic>Panarge</sic>
    <corr>Panurge</corr>
  </choice>
</persName>
neust faict esvanouyr
```

## 2.2 L'outil

Un état de l'art concernant les techniques de reconnaissance des noms propres pourra être trouvé dans Nadeau, Sekine (2009); l'idée principale que nous utilisons est ce que MacDonald (1996) appelle «internal and ex-

<sup>7</sup> C'est nous qui allons à la ligne ici et dans les autres exemples.

ternal evidence», c'est-à-dire le contexte local interne ou externe au nom propre lui-même (un prénom, une profession, etc.). Pour traiter l'ensemble du corpus qui nous a été proposé, nous avons utilisé et amélioré le système libre CasSys (Friburger, Maurel 2004) qui permet de construire et d'utiliser des cascades de transducteurs (Abney 1991), ou plutôt de graphes, du fait de son intégration à la plateforme Unitex (Paumier 2003).

Dans une cascade de graphes (Figure 3), le premier graphe modifie par insertion ou remplacement le texte d'origine, le second modifie le texte résultant, etc. Les exemples de la section 4 illustreront le fonctionnement de CasSys. D'autres exemples sont disponibles dans Maurel *et al.* (2011).

Figure 3 – Déroulement d'une cascade.



### 3. Méthodologie

#### 3.1 La gestion des multiples formes d'écriture

Pour gérer l'existence de multiples formes d'écriture des mots ou des noms propres, nous avons construit des dictionnaires (voir section 3.3.1). Pour les mots courants, c'est la 'forme vedette' qui relie les différentes formes entre elles. Pour les noms propres, c'est un identifiant. Lorsque le nom propre reconnu ne figure pas dans les dictionnaires, le système en construit un. Par exemple, la consultation d'un dictionnaire permet de savoir que le *Col Agnel* et le *col de laignel* ont le même identifiant, *loc\_agnel-col*, et désignent donc un même lieu. Parfois, des personnages d'ouvrages différents portent le même nom sans désigner la même personne et ils reçoivent alors des experts des identifiants différents. Après le traitement d'un ouvrage, les noms propres non présents dans le dictionnaire et leurs identifiants construits sont soumis aux experts du CESR pour validation et intégrés aux dictionnaires pour le traitement des autres ouvrages.

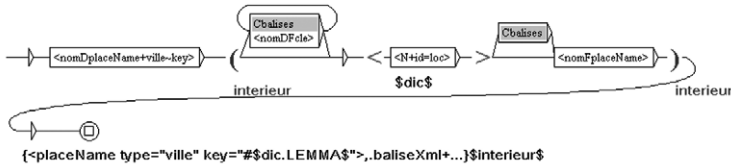
Dans l'exemple ci-dessous, le graphe de la Figure 4 trouve, entre deux balises *placeName*, un nom de lieu (boîte  $N+id=loc$ ) dans le dictionnaire, lui associe un nom de variable ( $\$dic\$$ ) et ajoute l'identifiant, qui occupe dans le dictionnaire la position de lemme ( $\$dic.LEMMA\$$ ), comme valeur de l'attribut *key*:

```
mestaiers de <placeName>Seuille</placeName> & de <placeName>Synays</placeName>.
```

Devient:

mestaiers de <placeName key=>#loc\_seuilly>Seuille</placeName> & de  
 <placeName key=>#loc\_cinai>Synays</placeName>.

Figure 4 – Un graphe de récupération de l’identifiant dans le dictionnaire.



### 3.2 La typologie adoptée

La typologie adoptée est celle de la TEI: les noms propres reconnus sont répartis en quatre types, les lieux géographiques (*geogName*), les lieux administratifs (*placeName*), les organisations (*orgName*) et, enfin, les personnes ou personnages (*persName*). Lorsqu’une adaptation au corpus est nécessaire, elle porte sur les attributs des balises.

#### 3.2.1 Les lieux géographiques et administratifs

Les lieux géographiques sont de deux types, d’une part, les géonymes: ‘montagnes’, ‘plaines’, ‘plateaux’, ‘grottes’ (*geo*) et, d’autre part, les hydronymes: ‘océan’, ‘mer’, ‘rivière’, ‘lac’, ‘étang’ (*hydro*). Les précisions géographiques sont intégrées à la balise et elles-mêmes balisées (*geogFeat*):

```
<geogName type="geo" key=>#loc_montsinai>
  <geogFeat>mont</geogFeat>
  Sinai
</geogName>
<geogName type=>hydro" key=>#loc_loire">
  <geogFeat>rivière</geogFeat>
  de Loyre
</geogName>
```

Les lieux administratifs peuvent éventuellement être typés (‘ville’, ‘pays’, ‘batiment ou domaine’):

```
<placeName type=>ville" key=>#loc_seuilly">Seuille</placeName>
<placeName type=>pays" key=>#loc_france">France</placeName>
<placeName type=>domaine" key=>#loc_lapommardiere">
  mestayrie de la Pomardiere
</placeName>
```

Les deux types de lieux peuvent être imbriqués l'un dans l'autre:

```
<placeName type=»batiment»>
  Palais de
  <placeName type=»ville» key=»#loc_poitiers»>Poitiers</placeName>
</placeName>
<placeName key=»#loc_guevede»>
  gue de
  <geogName type=»hydro» key=»#loc_vede»>Vede</geogName>
</placeName>
<geogName key=»#loc_ilescanaries»>
  isles de
  <placeName key=»#loc_canaries»>Canarre</placeName>
</geogName>
```

### 3.2.2 Les organisations

Les organisations sont partagées en trois types ('peuple', 'domaine' ou 'communaute') et n'ont pas d'identifiant associé:

```
<orgName type=»domaine»>
  Royaume de
  <placeName type=»pays» key=»#loc_france»>France</placeName>
</orgName>
```

Elles peuvent éventuellement être imbriquées:

```
<orgName type =»domaine»>Royaume des
  <orgName type = »peuple»>Dipsodes</orgName>
</orgName>
```

Lorsqu'il est difficile de distinguer entre un lieu et une organisation, un double balisage est possible:

```
<placeName type=»batiment» key=»#loch_coingnaufondabbaye»>
  <orgName type=»communaute»>
    abbaye de
    <placeName type=»ville» key=»#loch_coingnaufond»>
      Coingnaufond
    </placeName>
  </orgName>
</placeName>
```

### 3.2.3 Les personnes

Le balisage le plus simple consiste à reconnaître les personnes et à leur associer leur identifiant (*key*):

```
<persName key=#pers_aristote>>Aristote</persName>
```

Ce balisage peut être complété en interne par un balisage des prénoms (*forename*), des noms (*surname*) et des particules (*nameLink*):

```
<persName key=#pers_francoisconnan>>
  <forename>François</forename>
  <nameLink>de</nameLink>
  <surname>Connan</surname>
</persName>
```

Enfin, le balisage doit être étendu par des titres ou des civilités (*role-Name*) qui sont typés: titres nobiliaires, militaires ou religieux, fonctions et civilités honorifiques. Lorsqu'un titre comporte un nom de lieu ou une organisation, celui-ci est aussi balisé. Enfin, les précisions familiales (*gen-Name*) et les surnoms (*addName*) entrent aussi dans le balisage:

```
<persName key=#pers_huguesthierrysalel>>
  <forename>Hugues</forename>
  <forename>Thierry</forename>
  <surname>Salel</surname>
  <genName>l'ainé</genName>,
  <roleName type=#noble>>
    seigneur de
    <placeName type=#ville key=#loc_seuilly>>Seuille</placeName>
  </roleName>
</persName>
```

Comme pour les organisations, un double balisage est parfois possible:

```
<placeName key=#loc_saintmesmesdechinon>>
  <persName key=#pers_saintmesmesdechinon>>
    <roleName type=#religion>>saint</roleName>
    Mesmes de
    <placeName type=#ville>>Chinon</placeName>
  </persName>
</placeName>
```



### 3.3 Les ressources créées pour le projet

#### 3.3.1 Les dictionnaires

Les cascades réalisées consultent quatre dictionnaires, avant le passage des graphes:

- trois dictionnaires de noms propres, gérés par le CESR, respectivement pour les personnes, les lieux et les organisations; ces dictionnaires sont augmentés à chaque analyse par les nouveaux noms propres reconnus, après validation par les experts. Ils contiennent pour chaque entrée un nom, un identifiant, un type, éventuellement des traits ('ville', 'pays', 'batiment') et des variantes;

*ancenis,loc\_ancenis.N+id=loc:ms*  
*ancenys,loc\_ancenis.N+id=loc:ms*

- un dictionnaire créé par nous et contenant des mots déclencheurs de la reconnaissance (titres, charges, rôles, famille, prénoms) sous de multiples formes d'écriture possibles à la Renaissance, la graphie des mots n'étant pas fixée.

*capitaine,N+Militaire:ms*  
*capiteine,capitaine.N+Militaire:ms*  
*cappitaine,capitaine.N+Militaire:ms*

Pour créer ce dictionnaire de mots déclencheurs, nous nous sommes servis des dictionnaires contemporains que nous possédions et des corpus annotés en noms propres pour y repérer des formes manquantes ou des variantes orthographiques. Le Tableau 2 présente le nombre d'entrées à la livraison des cascades. Les dictionnaires sont constitués des noms propres présents dans les ouvrages et sont augmentés au fur et à mesure du traitement des documents.

Tableau 2 – Taille des dictionnaires (en nombre d'entrées).

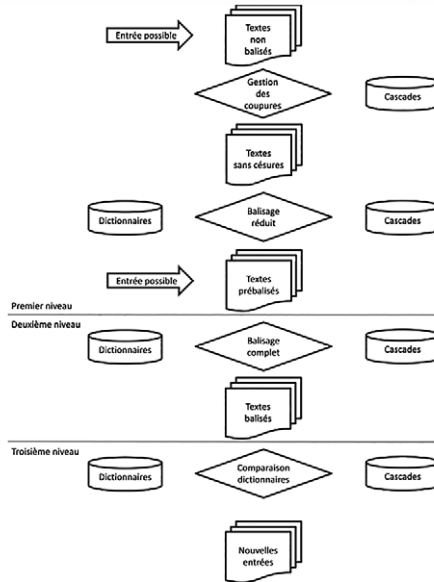
Personnes	1 149
Lieux	990
Organisations	57
Déclencheurs	2 608

#### 3.3.2 Le système de cascades créé

Étant donné la diversité de l'état des textes à analyser, nous avons adopté, pour optimiser la maintenance de l'ensemble, un système à trois niveaux. Le premier niveau analyse un texte où les noms propres ne sont

pas déjà balisés et le transforme en un texte conforme aux textes déjà balisés par le CESR. Le deuxième niveau augmente le balisage et le rend conforme à la typologie. Le troisième niveau, plus technique, génère une liste d'entrées à ajouter éventuellement dans les dictionnaires de noms propres (ajouts ensuite supervisés par les experts du CESR). La Figure 5 présente les trois niveaux de cette architecture.

Figure 5 – Les trois niveaux de l'architecture du projet Renom.



### 3.4 Les trois premières étapes

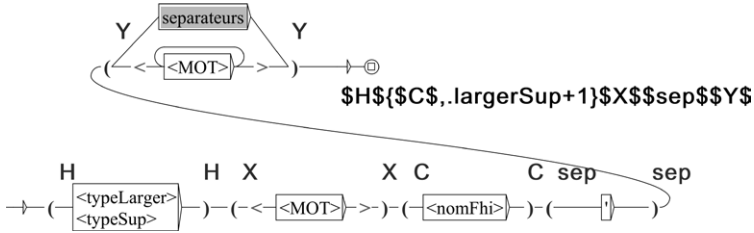
#### 3.4.1 La gestion des coupures

Comme cela a été dit à la section 3.1, les textes du corpus respectent la mise en page de l'édition originale et comportent un grand nombre de coupures de mot, ce qui rend l'utilisation des dictionnaires impossible. La première cascade d'analyse reconnaît tout d'abord les balises XML (comme expliqué à la section 1), puis reconstruit les mots coupés par des balises de format, de saut de ligne ou de bas de page. Par exemple, le graphe de la Figure 6 reconstruit les mots séparés par une balise de format (*hi*), lorsque la séparation porte sur la première lettre<sup>8</sup>. On englobe la balise ouvrante dans une expression polylexi-

<sup>8</sup> Dans le système Unitex, lorsqu'on est, comme ici, en mode morphologique, l'expression <MOT> désigne en fait une lettre.

cale *largerSup* avec un trait indiquant le nombre de lettres pour reconstruire le texte au final; la balise fermante est placée juste après cette dernière, avant le mot reconstitué. Cela permet de traiter l'exemple suivant du Tableau 1:

Figure 6 – Le graphe de gestion de la balise 'hi'.



*<hi rend=>larger>>E</hi>N ceste mesme heure*

Qui devient tout d'abord:

```
{<hi rend=>larger>>, baliseXml+nomDhi+typeLarger}
{
  </hi>, baliseXml+nomFhi}
, .largerSup+1}
EN ceste mesme heure
```

Puis:

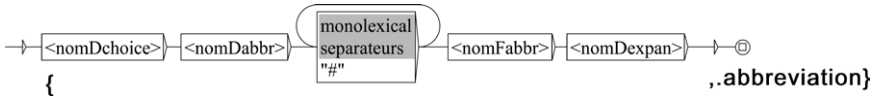
```
<csc>
  <form><hi rend=>larger>><\form>
  <code>baliseXml<\code>
  <code>nomDhi<\code>
  <code>typeLarger<\code>
<\csc>
<csc>
  <form>
    <csc>
      <form></hi><\form>
      <code>baliseXml<\code>
      <code>nomFhi<\code>
    <\csc>
  <\form>
  <code>largerSup<\code>
  <code>1<\code>
<\csc>EN ceste mesme heure
```

La cascade de synthèse qui suit reconstruit le texte en ajoutant un attribut *value* dans les balises TEI de coupure pour mémoriser le nombre de lettres à décaler lors de la reconstruction:

```
<hi rend=»larger» value=»1»></hi>EN ceste mesme heure
```

Les graphes de bas de page et de correction sont légèrement plus complexes, puisqu’il s’agit non seulement de reconstruire (pour les bas de page) ou de choisir le bon mot (pour les corrections), mais aussi de cacher du texte pour la suite de l’analyse. Notons qu’une correction peut se trouver coupée elle aussi par un saut de ligne ou même un bas de page. Le graphe de la Figure 7, utilisé pour l’expansion des abréviations, ne laisse visible que la forme corrigée, la forme d’origine étant cachée.

Figure 7 – Un graphe pour gérer les abréviations.



Ce graphe agit donc par exemple comme suit:

```
<choice>
  <abbr>PAN.</abbr>
  <expan>PANURGE</expan>
</choice>
```

Devient:

```
{<choice>,.baliseXml+nomDchoice}
  {<abbr>,.baliseXml+nomDabbr}PAN.{</abbr>,.baliseXml+nomFabbr}
  {<expan>,.baliseXml+nomDexpan}
  PANURGE
  {</expan>,.baliseXml+nomFexpan}
{</choice>,.baliseXml+nomFchoice}
```

Et:

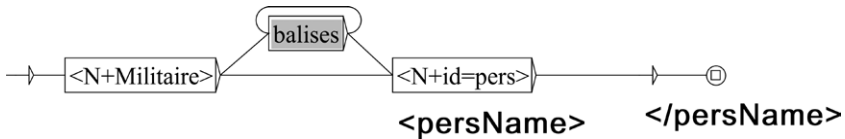
```
{
  {<choice>,.baliseXml+nomDchoice}
  {<abbr>,.baliseXml+nomDabbr}PAN.{</abbr>,.baliseXml+nomFabbr}
  {<expan>,.baliseXml+nomDexpan}
,.abbreviation}
PANURGE
{</expan>,.baliseXml+nomFexpan}
{</choice>,.baliseXml+nomFchoice}
```

### 3.4.2 Le balisage réduit

La deuxième cascade d'analyse commence par l'utilisation des dictionnaires pour reconnaître et baliser les noms propres déjà connus, comme *Gargantua* dans l'exemple ci-dessus. Cependant, les graphes de balisage utilisent aussi le contexte, car plusieurs entrées des dictionnaires sur lieu, personne ou organisation sont ambiguës. Le graphe de la Figure 8 balise un *persName* dans le contexte d'un déclencheur de type 'Militaire':

```
<lb/> cens chevaux legiers soubz la conduite
<lb/> du capitaine Engoulevent, pour descou
<lb rend=>hyphen/>vrir le pays, &amp; scavoir si embuche aulcune
<lb/> cens chevaux legiers soubz la conduite
<lb/> du capitaine <persName>Engoulevent</persName>, pour descou
<lb rend=>hyphen/>vrir le pays, &amp; scavoir si embuche aulcune
```

Figure 8 – Un graphe contextuel pour étiqueter les mots des dictionnaires.



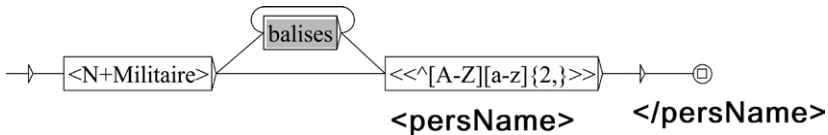
Lorsque le contexte ne permet pas de désambiguïser, un double balisage est inséré, comme expliqué dans la section 4.2.3. Après le balisage des mots du dictionnaire, des graphes contextuels insèrent des balises lorsqu'un mot commençant par une majuscule suit le contexte. Ces graphes utilisent des contextes similaires aux précédents. Le graphe de la Figure 9 reprend le même contexte que celui de la Figure 8 pour l'appliquer aux mots commençant par une majuscule et non déjà balisés, comme:

```
<lb/> du chevalereux capitaine Moses
```

Qui devient:

```
<lb/> du chevalereux capitaine <persName>Moses</persName>
```

Figure 9 – Un graphe contextuel pour étiqueter les mots commençant par une majuscule.



La cascade de synthèse complète la reconnaissance en reconstruisant le texte d'origine, augmenté de balises *persName*, *geogName*, *placeName* et

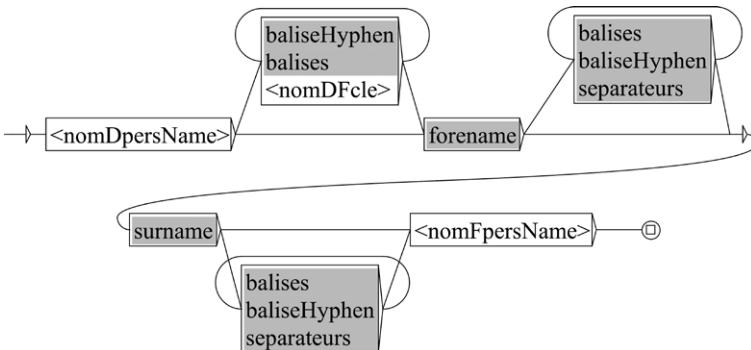
*orgName*. Ces dernières comprennent des attributs uniquement si ceux-ci proviennent directement de la consultation du dictionnaire.

### 3.4.3 Le balisage complet

La troisième cascade d'analyse opère sur le texte prébalisé avec les cinq balises *name*<sup>9</sup>, *persName*, *geogName*, *placeName* et *orgName*. Comme cela a été expliqué, ce texte est le résultat de la deuxième cascade ou provient d'un étiquetage manuel effectué avant le projet par le CESR. Dans un premier temps, les graphes préparent l'analyse du texte en travaillant sur les balises XML-TEI. À chaque fois que cela est possible, ces graphes sont bien sûr mutualisés et sont appelés par plusieurs cascades. Ensuite, la consultation des dictionnaires complète éventuellement les attributs manquants dans les balises (identifiant ou type) et, parfois, les réordonne. Puis cette cascade gère, d'une part, le balisage interne et les imbrications, et, d'autre part, l'extension du balisage. Enfin, elle crée les identifiants des noms propres reconnus non présents dans les dictionnaires.

Le balisage interne concerne deux types de balises, les *persName*, avec en particulier l'indication du prénom et du nom, et les *geogName*, avec l'introduction de la balise *geogFeat*. Pour la reconnaissance des prénoms, nous disposons d'une liste dans le dictionnaire de la Renaissance. Il est possible qu'une personne porte plusieurs prénoms (*hugues thierry salet*<sup>10</sup>) ou un nom polylexical (*Jan Trivolse Guallo*) et que certains noms comportent une particule à baliser elle aussi (*Ulrich Thierry du Gallet*). Le graphe de la Figure 10 balise un prénom et un nom (dans les sous-graphes *foreName* et *surName*), toujours en prévoyant des balises optionnelles:

Figure 10 – Le graphe prénoms-noms.



<sup>9</sup> Ces balises placées manuellement par le CESR sont aussitôt transformées en balise *persName*.

<sup>10</sup> Ce nom est en minuscule dans le texte.

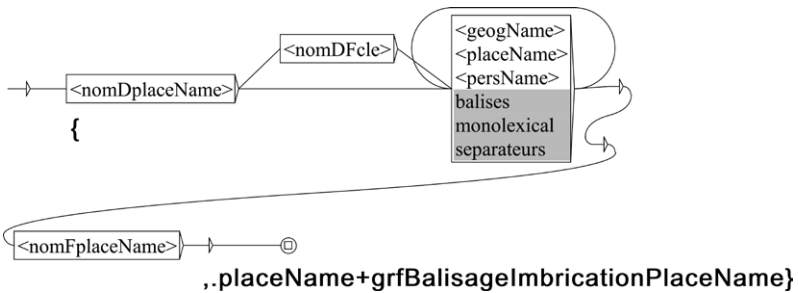
Pour ce qui est des *geogName*, l'insertion de la balise *geogFeat* est plus simple, puisque nous avons une liste des expansions possibles et que celles-ci se situent toujours à gauche du nom propre. Une fois le balisage interne a été effectué, les noms propres sont eux aussi considérés comme un seul bloc polylexical, afin de pouvoir gérer les imbrications.

Dans une imbrication, chaque élément a un identifiant. Le système d'identifiants peut donc se révéler assez complexe à mettre en place. Par exemple, on souhaite construire trois identifiants pour le château du gué de Vede:

```
<placeName key=#loc_chasteauduguedevede> dic=#non>>
  chateau du
  <placeName key=#loc_guedevede> dic=#non>>
    Gue de
    <geogName key=#loc_vede>>
      Vede
    </geogName>
  </placeName>
</placeName>
```

Le graphe des imbrications est itératif, afin de reconnaître plusieurs niveaux d'imbrications. Il s'agit en fait d'un graphe qui appelle quatre sous-graphes, chacun consacré à un des quatre types possibles. La Figure 11 présente ce graphe itératif et le sous-graphe qui gère une insertion dans un *placeName*:

Figure 11 – Le graphe des insertions et le sous-graphe pour un *placeName*.

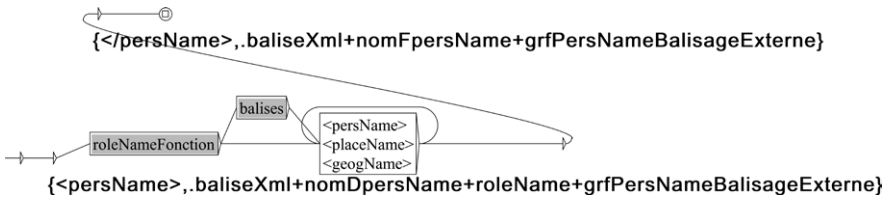


Comme cela a été annoncé à la section 3.2.3, l'extension du balisage correspond à des titres nobiliaires, militaires ou religieux, à des fonctions et civilités honorifiques, ainsi qu'à des précisions familiales et des surnoms. Les noms propres reconnus étant à ce moment considérés comme des expressions polylexicales, il s'agit d'ajouter des balises à gauche et à droite, une balise englobante *persName* et des balises *roleName*, *genName* ou *addName*. Le graphe de la Figure 12 étend le balisage à gauche d'un *persName*, comme dans l'exemple qui suit:

```

tu as
<lb>ton precepteur
<persName key=#pers_epistemon» dic=»non»>
    Epistemon
<persName>
dont
<choice>
    <orig>lun<orig>
    <reg>l'un<reg>
<choice>
par
<lb>vives & amp; vocables instructions
tu as
<lb>ton
<persName key=#pers_epistemon» dic=»non»>
    <roleName type=»fonction»>
        precepteur
    <roleName>
        Epistemon
<persName>
dont
<choice>
    <orig>lun<orig>
    <reg>l'un<reg>
<choice>
par
<lb>vives & amp; vocables instructions
    
```

Figure 12 – Graphe d'extension du balisage à gauche d'un persName.



#### 4. Évaluation

L'évaluation de notre système a été réalisée sur les deux ouvrages de Pierre de Ronsard présents dans notre corpus, le *Voyage de Tours* et l'*Élégie sur les troubles d'Amboise*, documents qui n'ont pas été utilisés pour construire les cascades. Rappelons que l'essentiel de notre corpus était



composé d'ouvrages de François Rabelais. Pour cette évaluation, nous avons utilisé les mesures classiques du rappel et de la précision (Figure 13).

Figure 13 – Rappel et précision.

$$\text{Rappel} = \frac{\text{nombre de noms propres correctement détectés}}{\text{nombre de noms propres réels}} = 80,9\%$$

$$\text{Précision} = \frac{\text{nombre de noms propres correctement détectés}}{\text{nombre de noms propres détectés}} = 77,9\%$$

Donnons quelques exemples d'erreurs:

- Nom propre non reconnu. *Jesus* apparaît sans contexte ci-dessous et n'était pas présent dans le dictionnaire:

`<l>Du saint nom, ou Jesus en la croix attaché,</l>`

- Erreur de reconnaissance. *Sinon* est un nom de personne dans notre dictionnaire de la Renaissance; pour éviter un trop grand nombre d'erreurs, les mots ambigus avec un mot du vocabulaire commun ne sont reconnus qu'avec une majuscule initiale. Mais ici, il apparaît en début de ligne, avec une majuscule, car il s'agit de vers (balise *l*):

`<l>Sa puissance est cruelle, & n'a point d'autre  
</b/><space unit="mm" quantity="4">  
<choice><orig>ieu</orig><reg>jeu</reg></choice>,</l>  
<l><persName key="persm_sinon">Sinon</persName>  
de rebrusler nos coeurs à petit feu,</l>`

- Erreur de type. *Antioche* et *Sydon* sont étiquetés *persName* au lieu de *placeName*; l'erreur provient de la présence de *Antioche* uniquement comme *persName* dans le dictionnaire (erreur corrigée):

`<l>Gagner la  
<placeName type="pays" key="loc_palestine">Palestine</placeName>  
, & toute l'Idumee,</l>  
<l>Tyr,  
<persName key="pers_sydon" dic="non">Sydon</persName>,  
<persName key="persf_antioche">Antioche</persName>  
& la ville nommée</l>`

- Erreur de frontière: l'extension du balisage au mot 'vallee' n'a pas été réalisée, à cause d'une mauvaise description de la correction (erreur corrigée):

```
<l rend=»positif»>Quel passe-temps prends-tu d'habiter la
  <lb/><space unit=»mm» quantity=»6»/>
  <choice><orig>uallee</orig><reg>vallee</reg></choice></l>
<l>De
  <placeName type=»batiment» key=»#loc_bourgueil»>Bourgueil</placeName>,
  où [...]</l>
```

## 5. Conclusion

Nous avons présenté l'ajout d'un module à la chaîne de transcription des textes de la Renaissance utilisée au CESR, module chargé de la recherche des noms propres dans des textes de la Renaissance, balisés dans un format XML-TEI conservant la présentation exacte des éditions originales. La particularité de ce corpus nous a conduits à adopter une méthodologie adaptée. Pour cela nous avons conçu des dictionnaires de noms propres et de mots déclencheurs contenant des multiples formes d'écriture et des cascades de graphes Unitex permettant de traiter, d'une part, des textes sans aucun repérage manuel de noms propres et, d'autre part, des textes où les noms propres sont déjà partiellement marqués. Après chaque traitement d'un ouvrage, les dictionnaires de noms propres sont augmentés après supervision d'un expert. Les experts du CESR ont considéré nos résultats comme tout-à-fait satisfaisants et permettant un précieux gain de temps dans l'annotation. Notre système est donc aujourd'hui opérationnel chez eux et sert à l'enrichissement des textes qu'ils transcrivent. Le site internet créé permet aux touristes visitant un château ou un hameau cité dans les œuvres de Rabelais de retrouver dans le texte original les mentions qu'en fait l'auteur. Inversement, les cartes générées sur ce site à partir des toponymes du texte les invitent à se déplacer en Touraine sur les traces de Gargantua.

Ce travail a été financé par le programme de recherche d'intérêt régional de la Région Centre. Les auteurs remercient les collègues du CESR, particulièrement Sandrine Breuil, Marie-Luce Demonet, Jorge Fins et Marie Olivron.

## Références

- Abney S. 1991, *Parsing By Chunks*, in Berwick R. C. et al. (ed.), *Principle-Based Parsing: Computation and Psycholinguistics*, Kluwer Academic Publishers, Dordrecht: 257-278.
- Friburger N., Maurel D. 2004, *Finite-state transducer cascade to extract named entities in texts*, «Theoretical Computer Science», 313: 94-104.

- MacDonald D. 1996, *Internal and external evidence in the identification and semantic categorisation of Proper Names*, in Boguraev B., Pustejovsky J. (eds.), *Corpus Processing for Lexical Acquisition*, The MIT Press, Cambridge, MA: 21-39.
- Maurel D., Friburger N., Antoine J.-Y., Eshkol-Taravella I., Nouvel D. 2011, *Cascades de transducteurs autour de la reconnaissance des entités nommées*, «Traitement automatique des langues», LII (1): 69-96.
- Nadeau N., Sekine S. 2009, *A survey of named entity recognition and classification*, in Satoshi S., Ranchhod E. (eds.), *Named Entities: Recognition, classification and use*, John Benjamins, Amsterdam: 3-28.
- Paumier S. 2003, *De la Reconnaissance de Formes Linguistiques à l'Analyse Syntaxique*, Thèse de Doctorat en Informatique, Université de Marne-la-Vallée.

M. Taddei

## Valorizzare gli scritti di Leonardo da Vinci per mezzo delle nuove tecnologie: l'archivio digitale *e-Leo*

*Abstract:* In today's information environment, which is further oriented towards the digital paradigm, the Library of Leonardo da Vinci has created e-Leo, a digital archive whose main objective is to enhance and preserve the complete collection of works possessed by the library. At the same time, the goal of the project is to offer to the international public an advanced tool for the study of Leonardo's works. This paper presents the most innovative features of this digital library: its different archive consultation functions, which allow the user to view the texts integrally, within a complex combination of text and drawing, as well as to improve the consultation of a corpus which lacks in many cases a sequential order; the study also presents and discusses e-Leo's search functions within the Index of drawings, the Leonardo's Glossary, and the texts; these features represent the most innovative aspect of the project, since they have been implemented using an advanced classification system and according international standards (Iconclass) and through the most advanced Text Mining tools currently available.

*Keywords:* Leonardo da Vinci, digital archives, Leonardian Library of Vinci, history of technics, history of science.

*Riassunto:* Nel contesto informativo attuale, sempre più orientato verso il digitale, è nato e-Leo, un archivio digitale realizzato dalla Biblioteca leonardiana di Vinci con l'obiettivo di valorizzare e di salvaguardare la collezione completa delle opere di Leonardo da Vinci da essa posseduta e, al tempo stesso, di offrire ad un pubblico internazionale uno strumento avanzato di studio delle opere di questo artista. Questo contributo presenta le caratteristiche più innovative di questa biblioteca digitale: le diverse funzioni di consultazione dell'archivio, che permettono di visualizzare il testo leonar-

diano nell'integrità e complessità di testo e disegno, nonché di ottimizzare la consultazione di un corpus in molti casi del tutto privo di ordine sequenziale; le funzioni di ricerca, nell'Indice dei disegni, nel Glossario leonardiano e nel testo, che rappresentano l'aspetto più inedito del progetto, in quanto realizzate servendosi di un sistema di classificazione all'avanguardia e rispondente a standard internazionali (Iconclass). Queste funzioni verranno prossimamente implementate attraverso gli strumenti più avanzati di Text Mining attualmente disponibili.

*Parole chiave:* Leonardo da Vinci, archivi digitali, Biblioteca Leonardiana, storia della tecnica, storia della scienza.

## I. Introduzione

*e-Leo. Archivio digitale di storia della tecnica e della scienza*<sup>1</sup> è stato realizzato e sviluppato dalla Biblioteca leonardiana di Vinci a partire dal 2005, inizialmente con la collaborazione scientifica del Dipartimento di Meccanica e Tecnologia dell'Università di Firenze ed il Centro di linguistica storica e teorica: italiano, lingue europee, lingue orientali (C.L.I.E.O) dell'Università di Firenze, e con l'apporto della società di consulenza e ricerche informatiche Synthema di Pisa.

Ha anche ricevuto un cofinanziamento della Commissione Europea entro il programma di azioni *La tecnica come bene culturale / Technics as Cultural Heritage* nel 2005. Successivamente ha potuto svilupparsi grazie ad altre risorse messe a disposizione o reperite dal Comune di Vinci, dal C.L.I.E.O. e dalla Regione Toscana.

Il progetto si è concretizzato con la messa in linea della prima versione nel marzo 2007 all'indirizzo <<http://www.leonardodigitale.com>>, dove è tuttora consultabile ad accesso libero. Il progetto *e-Leo* prende le mosse dalla volontà di valorizzare (anche salvaguardandola ed escludendola quindi il più possibile da movimentazioni) la 'collezione speciale' della Biblioteca leonardiana di Vinci e, contemporaneamente, di offrire uno strumento più avanzato di studio delle opere di Leonardo da Vinci. Si è poi ritenuto che il trattamento del lascito leonardiano, secondo le modalità che vedremo,

<sup>1</sup> Nanni R., Cascini G., Taddei M. (idea e progetto di) 2007, *e-Leo. Archivio digitale di storia della tecnica e della scienza*, Vinci, Biblioteca Leonardiana, <<http://www.leonardodigitale.com>>. Figura fondamentale per l'ideazione, realizzazione e sviluppo del progetto è stato senza dubbio Romano Nanni, direttore della Biblioteca leonardiana fino al 2014. Sotto la sua direzione un gruppo multidisciplinare composto da informatici, tecnici, grafici, bibliotecari, studiosi (ingegneri, fisici, storici dell'architettura, della lingua, ecc.) ha lavorato alla costituzione di un vero e proprio portale. Per i crediti completi si rimanda alla sezione *Crediti* a cui si accede dalla home dell'archivio. Per una bibliografia essenziale su *e-Leo*, oltre a quanto si citerà nelle note successive relativamente alle singole sezioni dell'archivio, si segnala: Nanni (2011), Biffi (2011) e Taddei (2005).

rappresentasse una base di partenza per un più ampio e generalizzato trattamento, secondo criteri affini, di testi manoscritti di argomento tecnico-scientifico del tardo Medioevo e del Rinascimento.

L'esperienza è dunque radicata nella tipologia molto particolare della collezione della Biblioteca leonardiana<sup>2</sup>, che possiede il corpus di tutte le opere di Leonardo da Vinci pubblicato a partire dalla prima edizione del *Trattato della pittura* del 1651, dalle edizioni più lontane nel tempo alle più recenti pubblicazioni in facsimile, tra cui quelle dell'Edizione nazionale dei manoscritti e dei disegni di Leonardo da Vinci<sup>3</sup>. Un corpus ingente come numero di fogli: 22 codici manoscritti (includendo il *Codice Leicester*, ora sfasciolato), 3 ampie raccolte di fogli rilegati in volumi (*Codice Atlantico*, il corpus di disegni della Royal Library di Windsor, il *Codice Arundel*), numerosi fogli sparsi in collezioni pubbliche e private. Un corpus frammentato come collocazione, prezioso e quindi difficilmente accessibile sia per utenti occasionali sia per ricercatori e studiosi, cosa che fra l'altro dà un valore aggiunto ad una raccolta che lo raggruppa e rende accessibile. Ma anche un corpus estremamente complesso per le sue caratteristiche intrinseche, vale a dire: per il rapporto tra testo ed immagine che lo caratterizza graficamente, per la frammentarietà dei disegni e dei progetti, raramente compiuti e sistematizzati, per la difficoltà di lettura dei testi perché essi stessi frammentari, oltre che redatti specularmente. La differenza con la produzione di disegni da parte di altri artisti italiani del Cinquecento è netta, sia come quantità sia come natura. Leonardo lavora su libri di appunti e taccuini di disegni secondo sequenze non lineari, le sue pagine abbondano di note e disegni sparsi in apparenza senza rapporto fra di loro<sup>4</sup>.

*e-Léo* nasce quindi in questo particolare contesto in un momento, tra il 2005 ed il 2006, in cui la discussione sulle biblioteche digitali attraversava il mondo della biblioteconomia italiana e dei beni culturali più in generale<sup>5</sup>. Nell'ambito della molteplicità di modelli possibili ed esperiti in Italia e

<sup>2</sup> Per ulteriori informazioni sulla biblioteca e sulle sue collezioni si veda: <<http://www.bibliotecaleonardiana.it/bbl/bb-leo/bb-leo-home.shtml>>.

<sup>3</sup> L'edizione fu prima a cura della Reale Commissione Vinciana (1923-1952), poi della Commissione Vinciana (1973-1980), infine Sotto l'Alto Patronato del Presidente della Repubblica Italiana (1986-).

<sup>4</sup> Per una riflessione sull'eredità manoscritta di Leonardo si rimanda, tra gli altri, a Bambach (2009).

<sup>5</sup> Si pensi per esempio al *Gruppo di studio sulle biblioteche digitali* costituitosi in seno all'Associazione Italiana Biblioteche, il quale ha operato dal 2003 al 2011 (<<http://www.aib.it/aib/cg/gbdig.htm3>>); alle linee guida riguardanti alcune tipologie di materiali pubblicate tra il 2004 ed il 2006 dall'ICCU-Istituto Centrale per il Catalogo unico delle Biblioteche italiane e per le informazioni bibliografiche, al *Gruppo di studio sugli standard e le applicazioni di metadati nei beni culturali* costituito dall'ICCU al fine di coordinare a livello nazionale le implementazioni di metadati nei progetti di digitalizzazione nei diversi settori dei beni culturali e di raccordare le iniziative italiane con quelle europee ed internazionali, alle *Raccomandazioni sulla digitalizzazione e l'accessibilità online del materiale culturale e sulla conservazione digitale* (2006/585/CE)

all'estero, *eLeo* – come vedremo – è stato fin dall'inizio non un mero *repository* di documenti digitalizzati finalizzato alla preservazione e fruizione del patrimonio della biblioteca, ma una realizzazione, per così dire, più avanzata per l'analisi dei testi ed il loro studio, nello specifico le opere di Leonardo da Vinci, ma anche testi tardo medievali e rinascimentali di argomento tecnico-scientifico.

Nello schema (Fig. 1) sono sintetizzate le funzioni di consultazione dell'archivio.

1. Il fulcro su cui ruotano i vari strumenti è il sistema di consultazione (sezione 'Sfoggia') che consente di sfogliare i codici o la sequenza dei disegni nelle varie raccolte e leggerne i testi<sup>6</sup>.
2. Il testo è ricercabile in maniera, per così dire tradizionale, parola per parola nella sezione 'Ricerca nel testo'<sup>7</sup>. A breve potrà anche essere interrogato ad un livello semanticamente più raffinato tramite sistemi di analisi linguistico-statistica (lo abbiamo chiamato livello 2): i risultati dell'interrogazione potranno cioè essere analizzati e classificati in maniera da creare gruppi affini (*cluster*) di testi e identificare i possibili vari campi d'impiego del termine di partenza.
3. Il testo è ricercabile nella sua componente grafica per singolo disegno nella sezione 'Indice dei disegni'<sup>8</sup>. Viene quindi fornita un'ulteriore chiave di accesso al contenuto, anche laddove il disegno non è accompagnato da note o si trova in contesti testuali d'altro argomento.
4. Il 'Glossario' fornisce un ausilio linguistico alla consultazione<sup>9</sup>, facilitando l'accesso ad un testo spesso complesso, caratterizzato dalla presenza di 'lessici' settoriali. Per come è strutturato, si pone comunque anche come «strumento di analisi utile alla storia della linguistica italiana» (Biffi 2011: 105)<sup>10</sup>.

della Commissione [Europea] del 24 agosto 2006, ribadite e rinnovate da quelle del 2011 (cfr. *Gazzetta Ufficiale dell'Unione Europea* del 29/10/2011 L 283/39), in cui per l'Italia veniva indicato come obiettivo da raggiungere entro il 2015 il numero di ben 3.705.000 documenti da far confluire nelle banche dati di Europeana.

<sup>6</sup> Alla data attuale (marzo 2016) nella sezione 'Sfoggia' sono consultabili 12.594 carte. Le opere presenti ed i relativi riferimenti bibliografici sono elencati alla pagina 'Descrizione del progetto', a cui si accede dalla home <<http://www.leonardodigitale.com>>.

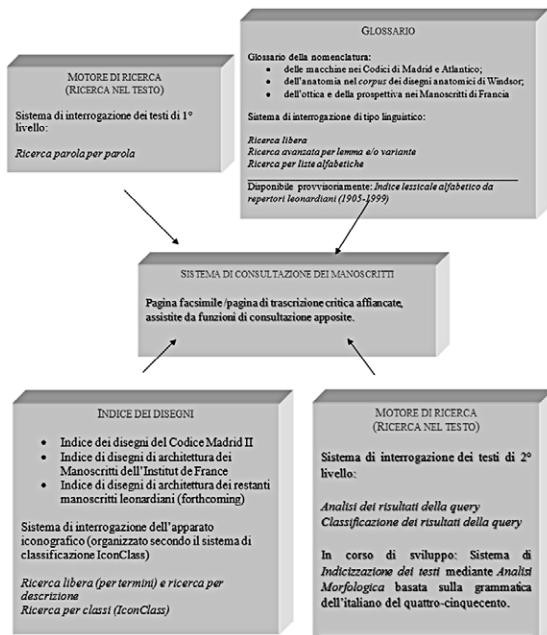
<sup>7</sup> Alla data attuale (marzo 2016) sono ricercabili 8.950 pagine di testo. Le 28 opere selezionabili per la ricerca testuale sono quelle elencate nel menù di ricerca nella sezione 'Ricerca nel testo'.

<sup>8</sup> Alla data attuale (marzo 2016) nella sezione 'Indice dei disegni' sono stati classificati e indicizzati nel sistema 746 disegni.

<sup>9</sup> Alla data attuale (marzo 2016) i lemmi analizzati sono 882.

<sup>10</sup> Per una riflessione sull'importanza e funzione del 'Glossario', si veda: Manni, Biffi (2008) e Manni, Biffi (2011): Introduzione e Criteri di allestimento del Glossario (le voci del glossario sono disponibili anche in *e-Leo* nella sezione 'Glossario' > 'Nomenclatura delle macchine nei Codici di Madrid e Atlantico'); Quaglino (2013): Introduzione e Criteri di allestimento del Glossario (le voci del glossario sono disponibili anche in *e-Leo* nella sezione 'Glossario' > 'Nomenclatura dell'ottica e della prospettiva nei manoscritti dell'Institut de France'); e Manni (2015).

Figura 1 – Schema di sintesi delle funzioni di consultazione di e-Leo.



L'interfaccia di *e-Leo* è per ora in italiano ed inglese, i contenuti delle varie sezioni sono invece per lo più solo in italiano, ad eccezione di alcune edizioni in lingue europee del *Trattato della pittura* e, da poco, della traduzione in inglese dei *Manoscritti dell'Institut de France* a cura di John Venerella<sup>11</sup>. Scendendo nel dettaglio dell'architettura e delle funzioni dell'archivio, vediamo come *e-Leo* è strutturato nelle seguenti sezioni: 'Sfoggia i manoscritti' | 'Indice dei disegni' | 'Glossario' | 'Ricerca nel testo'. In 'Sfoggia i manoscritti' il testo leonardiano è presentato nella integrità e complessità di testo e disegno, affiancati e consultabili in simultanea. Il trattamento dei testi ad oggi ha interessato i facsimili di tutti i codici di Leonardo (il *Codice Atlantico* fra l'altro è presente in due edizioni), parte della raccolta di Windsor, il *Libro di pittura*, alcune edizioni del *Trattato della pittura* in italiano, inglese, francese, spagnolo e tedesco ed infine un piccolo nucleo di opere di Francesco di Giorgio Martini.

La digitalizzazione ha previsto diverse fasi di lavorazione ed un sistema di controllo che ne ha tracciato lo stato e la qualità. Il materiale è stato digitalizzato ad una risoluzione non inferiore ai 300 dpi (in genere 400 dpi)

<sup>11</sup> La traduzione è stata edita a stampa in 12 volumi tra il 1999 ed il 2007: Leonardo da Vinci 1999-2007, *The Manuscripts of Leonardo da Vinci in the Institut de France. Manuscript A-M*, traduzione e annotazioni di J. Venerella, Ente Raccolta Vinciana, Milano. Messa a disposizione dal traduttore, è stata adattata allo strumento informatico consentendone la ricerca testuale in *e-Leo*.



in formato TIFF non compresso per tutte le dimensioni dei materiali con sistema di ripresa con fotocamere digitali professionali Nikon D800E<sup>12</sup>.

Per ogni carta o pagina sono state prodotte tre immagini di diverso formato come di seguito specificato:

- Formato non compresso di conservazione per l'archivio dei master fuori linea: TIFF 6.0 non compresso, con risoluzione di 300 dpi (non interpolati), RGB.
- Formato di consultazione compresso ad alta definizione: compresso in formato JPEG standard a 300 dpi, destinato alla consultazione in rete locale o alla produzione di eventuali supporti multimediali ad alta risoluzione.
- Formato di consultazione compresso a bassa definizione: compresso in formato JPEG standard a 300 dpi di qualità inferiore, destinato alla consultazione sul web.

Prima della messa in linea le immagini sono state post prodotte su piattaforma Apple con applicativi Adobe per interventi migliorativi su luce/contrasto/nitidezza. Durante tutte le fasi di digitalizzazione e di elaborazione dei documenti sono previsti controlli sistematici e a campione per garantire la perfetta corrispondenza qualitativa con il prototipo realizzato a inizio del lavoro e con le specifiche di progetto. Tutte le immagini sono state memorizzate secondo uno schema di *file system* concordato con noi, in modo da identificare univocamente ogni immagine all'interno di una struttura e facilitare le successive fasi di indicizzazione e catalogazione. Per ognuna delle opere sono stati prodotti i metadati relativi ai file di immagine secondo lo standard MAG versione 2.0.1 dell'Istituto per il Catalogo unico delle Biblioteche italiane e per le informazioni bibliografiche<sup>13</sup>.

Le trascrizioni critiche a stampa dei manoscritti sono state acquisite tramite OCR; i file di testo sono poi stati revisionati a vista uno ad uno, i primi anche pesantemente in quanto la messa a punto di una routine di autocorrezione è stata realizzata per fasi successive mano a mano che procedevamo con l'esperienza. Alcuni dei testi in linea sono invece stati direttamente trascritti a mano perché il riconoscimento ottico avrebbe dato risultati pesimi e quindi improponibili<sup>14</sup>.

<sup>12</sup> La campagna di digitalizzazione è stata realizzata da Gronchi Fotoarte di Pisa tra il 2006 ed il 2011.

<sup>13</sup> Si veda: <[http://www.iccu.sbn.it/opencms/opencms/it/main/standard/metadati/pagina\\_267.html?sessionId=E7723EEF88598A93F9856E7C44BA803B](http://www.iccu.sbn.it/opencms/opencms/it/main/standard/metadati/pagina_267.html?sessionId=E7723EEF88598A93F9856E7C44BA803B)> (04/2016).

<sup>14</sup> Le acquisizioni tramite OCR sono state eseguite da Gronchi Fotoarte di Pisa (2005) solo per il *Codice Atlantico* nell'edizione Firenze Giunti, per lo più da Alessio Veracini (dal 2002 ad oggi); il lavoro di revisione è stato effettuato da quest'ultimo in collaborazione con il gruppo di lavoro della Biblioteca Leonardiana, in particolare Monica Taddei e Simona Minichelli. La trascrizione manuale di alcune delle opere è stata eseguita, tra il 2007 ed il 2011, da: Cecil Glydel per il *Codice Leicester* (Cogliati, Milano, 1909); Simona Minichelli per il *Tractat von der Mahlerey*

## 2. Sezione 'Sfoggia i manoscritti'

La sezione comprende tasti normalmente utilizzati e diffusi in sistemi per la consultazione di testi in formato digitale, quali i tasti avanti, indietro, posizionamento all'inizio o alla fine del documento, spostamento alla pagina prescelta, lo zoom (possibile fino al 200%), la rotazione di 90° in 90°; il *thumbnail* consente uno scorrimento veloce attraverso il manoscritto di cinque o cinquanta pagine avanti e indietro o dall'inizio alla fine e viceversa.

La maschera di consultazione è stata però anche adattata alle specificità dei testi che ci siamo trovati a trattare, per cui sono stati sviluppati una serie di strumenti ad hoc:

- ogni pagina del testo leonardiano è mostrata a fronte della corrispondente pagina di trascrizione critica;
- la pagina di trascrizione è generalmente corredata del relativo apparato critico di note ed indicazioni topografiche al fine di facilitare l'orientamento nella pagina;
- la pagina di trascrizione è redatta in sequenza lineare, secondo il modello prevalente adottato nelle edizioni facsimilari e che abbiamo rivisto o adattato solo in alcuni casi; laddove non lo era, nello specifico nei *Codici di Madrid*, la sequenza è stata da noi ricostruita e corredata di indicazioni topografiche di orientamento nel testo, in quanto l'edizione critica curata da Ladislao Reti<sup>15</sup>, l'unica ad oggi disponibile, ne era priva e riproduceva invece mimeticamente sulla carta di trascrizione la struttura non lineare del testo leonardiano (si veda per esempio la c. 87v al *Codice di Madrid I* nella Figura 2, da confrontare con la stessa carta in linea in e-Leo).

Ogni segmento unitario e autonomo di testo, di disegno o di testo-disegno all'interno della singola pagina del manoscritto è stato reso sensibile in modo che, selezionato, richiami il corrispondente punto della trascrizione digitalizzata a fronte. La funzione, lunga e laboriosa da realizzare, consente di ottimizzare, sul piano anche didattico, la consultazione di un corpus in molti casi del tutto privo di un qualsiasi tipo di ordine sequenziale. Ne risulta una lettura assai agevole che contemporaneamente mantiene un rapporto di verifica a vista con la spesso inusuale composizione originaria della pagina leonardiana. La funzione è attivabile dal tasto 'Link testi' (ad esclusione che per il *Codice Arundel*, che ne è privo), per cui da una qualsiasi parte dell'immagine posso visualizzare la corrispondente porzione di trascrizione. È poi attiva, ad eccezione che per il *Codice Atlantico*, il *Codice Arundel* e i *Codici di Madrid*, anche la funzione inversa di 'Linkback', la qua-

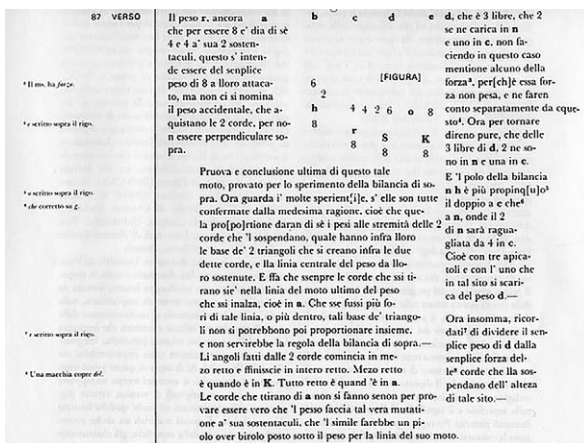
(Cristoph Weigel, Nürnberg, 1747), il *Traité de la peinture* (Jacques Langlois, Paris, 1651) e di una parte del *Trattato della pittura* (Giacomo Langlois, Parigi, 1651).

<sup>15</sup> Leonardo da Vinci 1974, *I Codici di Madrid [nella] Biblioteca nazionale di Madrid*, trascrizioni a cura di L. Reti, Giunti Barbèra, Firenze.

le consente di raggiungere dalla trascrizione la porzione corrispondente del manoscritto semplicemente cliccando su 'Evidenzia':

- il sistema offre l'opzione di riflessione (tasto 'Rifletti'), oltre che quella più usuale di rotazione a sinistra e destra (tasto 'Ruota'), la quale consente di riflettere la scrittura speculare di Leonardo, al fine di agevolare una lettura o verifica diretta del testo originario;
- sono disponibili funzioni di ridimensionamento dello spazio di lavoro atte ad agevolare lo studio dei disegni e del *ductus*: le immagini possono essere trascinate, allargate e ridotte, il testo di trascrizione può essere chiuso, spostato, riposizionato, allungato e sovrapposto all'immagine anche in trasparenza;
- sono disponibili opzioni di 'Reference' quali note e, in alcuni casi, tavole delle corrispondenze, utili per la consultazione di quei materiali che presentano sistemi di numerazione e riferimento diversi, come il corpus dei disegni anatomici di *Windsor* e le edizioni del *Codice Atlantico*.

Figura 2 – Pagina di trascrizione della c. 87v al Codice di Madrid I nell'edizione Giunti Barbèra (1974).



### 3. Funzioni di ricerca

Le funzioni di ricerca rappresentano probabilmente l'aspetto più inedito del progetto, sotto il profilo sia scientifico sia tecnico, anche rispetto ad altri sistemi di consultazione di testi della tradizione rinascimentale disponibili in linea. Sicuramente per noi è stata ed è la parte più impegnativa dal punto di vista scientifico, che ha richiesto e richiede la sinergia tra numerose competenze non presenti all'interno del nostro istituto. Le funzioni di ricerca comprendono: ricerca nell'Indice dei disegni, nel Glossario leonardiano e nel testo.

### 3.1 Indice dei disegni

L'Indice dei disegni in linea ad oggi comprende quello relativo al *Codice di Madrid II*, realizzato da Romano Nanni e Davide Russo, e quello dei disegni di architettura nei *Manoscritti di Francia*, a cura di Emanuela Ferretti e Davide Turrini. Già realizzati e da mettere in linea a breve sono l'Indice dei disegni del *Codice di Madrid I* di Romano Nanni e Davide Russo, e quello dei disegni di architettura di Emanuela Ferretti e Davide Turrini relativo a tutto il restante corpus manoscritto leonardiano.

Lo strumento è ovviamente finalizzato all'indagine iconografica, a dare cioè un accesso al corpus consentendo una ricerca a tutto campo nel manoscritto, anche laddove il disegno non è accompagnato o commentato da un testo, e quindi non recuperabile partendo da una normale ricerca testuale. La base ideale della ricerca iconografica è poter accedere e disporre di una documentazione organica e articolata, possibilmente ampia, ordinata in modo sistematico per soggetti (e sottolineo in modo sistematico), meglio se ricercabile in maniera automatica.

Prendendo le mosse da questa considerazione di partenza, la questione da affrontare nella fase progettuale da parte del gruppo di lavoro era quindi decidere come organizzare e rendere disponibile questa documentazione iconografica, e scegliere quale linguaggio usare per indicizzarla. La nostra preferenza è andata al sistema di classificazione internazionale Iconclass, abbreviazione di ICONographic CLASSification System, sulla base del quale sono stati classificati e resi interrogabili i disegni secondo un sistema che ordina gerarchicamente per classi e sottoclassi le definizioni di oggetti, persone, eventi e idee astratte<sup>16</sup>.

La scelta di Iconclass è venuta dopo una serie di valutazioni: 1. perché temevamo che il sistema fortemente gerarchizzato potesse crearci qualche problema sia in fase di indicizzazione, sia in fase di realizzazione dell'interfaccia di consultazione; 2. perché le esperienze maggiori di utilizzo di questo sistema erano nel campo dell'arte figurativa, mentre la nostra documentazione da trattare afferiva ai campi più svariati, scienza e tecnica incluse. Iconclass presentava, d'altro canto, notevoli vantaggi: 1. era un sistema internazionale e quindi capace sulla carta di favorire possibilità di circolazione di dati e interoperabilità; 2. era multilingue: sviluppato in versione inglese, nel 2005 era già stato tradotto nelle principali lingue eu-

<sup>16</sup> Sorvolerò sulla storia e contesto di sviluppo di Iconclass. Per ulteriori informazioni rimando al sito <<http://www.iconclass.nl>>. La traduzione italiana del sistema di classificazione è pubblicata nel 2000 dall'Istituto Centrale per il Catalogo e la Documentazione: Waal 2002. Un utile ragguglio sul tema è rappresentato dalla traduzione in italiano del volume di Roelof van Straten del 1985, *Inleiding in de iconografie*: Straten 2009. Per inquadrare brevemente il dibattito in Italia sull'utilizzo di questo sistema di classificazione si rimanda a: Lattanzi, Colalucci (1992) e Plances (1998). La raccolta di risorse sul tema tratta da AIB-WEB all'indirizzo <<http://www.aib.it/?s=iconclass>>.

ropee; la traduzione in italiano risale al 2000 a cura dell'Istituto Centrale per il Catalogo e la Documentazione (cfr. nota 15); 3. era già stato sviluppato in precedenza in un software e, quindi, adatto ad un utilizzo in ambiente elettronico; 4. era potenzialmente in grado di poter funzionare bene come sistema di organizzazione documentale a tutto campo, essendo una struttura 'vuota' di codici alfanumerici, o notazioni, che rinviano a temi e soggetti iconografici afferenti a tutti i campi del sapere: religione, natura, essere umano, uomo in generale, società, idee e concetti astratti, storia, Bibbia, letteratura, mitologia classica e storia antica; 5. e, infine, il panorama italiano delle biblioteche e dei beni culturali nel 2004-2005 non offriva strumenti di catalogazione semantica altrettanto sistematici ed esaustivi pensati per i materiali iconografici.

Per quanto riguarda la banca dati di interrogazione, consapevoli della necessità di semplificare l'accesso all'indice da parte degli utenti non esperti, abbiamo tecnicamente creato un semplice ed intuitivo accesso per termini, oltre a offrire la possibilità di ricercarlo tramite il sistema Iconclass. Nella 'Ricerca per termini' infatti i disegni possono essere semplicemente ricercati per parole chiave in modalità libera, per un descrittore o un qualsiasi termine presente nella descrizione della sottoclasse o del singolo disegno, senza che sia necessaria alcuna conoscenza del sistema di classificazione sotteso. Nella 'Ricerca per classi' invece i disegni possono essere ricercati per classi strutturate secondo il sistema di classificazione Iconclass, in una gerarchia ad albero, dinamicamente dalle classi superiori si scende alle relative sottoclassi e poi giù fino al singolo disegno. Chi conosce il sistema di notazioni di Iconclass ha anche l'opportunità di interrogarlo partendo da queste.

### 3.2 Glossario leonardiano

Altra funzione di ricerca è offerta da uno strumento di interrogazione di tipo linguistico, il Glossario leonardiano. Il progetto e la direzione generale sono di Paola Manni e Marco Biffi, nell'ambito di una collaborazione tra Biblioteca leonardiana e C.L.I.E.O. dell'Università di Firenze. Questa sezione comprende per ora tre glossari, quello dei sostantivi della meccanica applicata relativi ai *Codici di Madrid* e *Codice Atlantico* a cura di Paola Manni e Marco Biffi, quello dei termini dell'ottica e della prospettiva nei *Manoscritti dell'Institut de France* a cura di Margherita Quaglino e una parte (circa la metà dei lemmi analizzati) dei termini anatomici tratti dal corpus dei disegni leonardiani di *Windsor* a cura di Maria Rosaria D'Anzi. Di quest'ultimo è in corso una revisione complessiva ed un ampliamento da parte di Rosa Piro.

I Glossari presenti in *e-Leo*, oltre che la collaborazione del C.L.I.E.O., hanno visto di volta in volta la collaborazione di gruppi di ricerca di varie università, che hanno coinvolto, oltre a linguisti, anche ingegneri, fisici

ottici, storici dell'anatomia, ecc., a seconda dell'ambito di indagine. Tutti i glossari, quelli già online e quelli a venire, hanno e avranno la stessa impostazione e rispondono a criteri espressamente approntati per questo progetto dal gruppo di lavoro del C.L.I.E.O.

### 3.2.1 Glossari

Tecnicamente il sistema consente di effettuare ricerche libere, ricerche avanzate (per lemma e/o variante) e per liste alfabetiche. Il sistema restituisce la definizione, le varianti formali in cui il termine si presenta nei manoscritti leonardiani, l'eventuale presenza di una glossa d'autore e i lemmi correlati. Ogni lemma è collegato alla variante o alle varianti, da cui si risale a tutte le occorrenze del termine nel testo leonardiano. Ogni lemma è approfondito da una scheda lessicografica che contiene varie informazioni: lemma esponente, varianti grafiche, citazioni tratte dal ms (manoscritto), lista di frequenza e *index locorum*, lemmi correlati, eventuale trattamento del termine in precedenti edizioni di testi leonardiani, corrispondenze volgari e latine, indicazione di eventuale prima attestazione.

### 3.2.2 Indice lessicale alfabetico

A fianco del Glossario in corso di sviluppo, è presente anche questo strumento realizzato tramite assemblaggio di repertori leonardiani posseduti dalla Biblioteca e pubblicati tra il 1905 ed il 1999. È uno strumento che riteniamo debba essere sempre di più residuale, testimonianza storica di una stagione di studi linguistici sostanzialmente superati, con limiti quantitativi e scientifici. Pur restando a disposizione come sussidio di consultazione, questo indice ha un carattere compilativo e temporaneo ed il suo utilizzo come strumento di ricerca sarà sostanzialmente soppiantato dai Glossari che mano a mano andremo a rendere consultabili. Consente una ricerca libera e per liste ed il risultato della ricerca è una lista di forme con relative descrizioni.

## 3.3 Ricerca nel testo

### 3.3.1 Ricerca di livello I

*e-Leo* offre una ricerca, per così dire, di livello 1, ossia una ricerca tradizionale basata su combinazioni booleane. Consente di ricercare per parola chiave o stringa all'interno di tutta la banca dati di *e-Leo*. La ricerca è raffinabile e può essere limitata ad uno o più manoscritti e/o testi (si veda il menu sulla sinistra che consente la selezione). Il risultato della ricerca può essere salvato foglio per foglio o completamente (con 'Seleziona tutti i fogli') e poi visualizzato nell'interfaccia di consultazione. Al momento, per

la tecnologia software utilizzata (Flash di Adobe), la parola ricercata non viene evidenziata<sup>17</sup>.

### 3.3.2 Ricerca avanzata

Una ricerca più avanzata è quella a cui si potrà accedere, dopo aver effettuato una ricerca nel testo, da 'Analizza' e 'Classifica' (Figg. 3-5). Questa parte del sistema, sviluppata dalla ditta Synthema, non è però ancora funzionante in maniera completa. È in corso di sviluppo la sua versione finale (si veda quanto accennato alla nota 16).

Figura 3 – Analisi linguistica.

**SYNTHEMA** LABORATORIO DI SEMANTICO TECNOLOGIE COGNITIVE **Analisi linguistica** federico.neri@synthema.it

Indicizzazione dei testi mediante

**Analisi Morfologica-Sintattica**

(base(n)+base(adj) ⇒ "linea centrale", "rota maggiore", "rota grande" ... )  
 (n+prep+base(n) ⇒ "ciento della rota", "diametro della rota" ... )

**Analisi Statistica** (n > v<sub>1</sub>, v<sub>1</sub> = f(SW-MW, nTot, ...))

- Soglia di Frequenza (log(α(v<sub>1</sub>/n) \* β(nTot/nDoc) \* f(POS)\*...) > v<sub>2</sub>)
- Soglia di Significatività

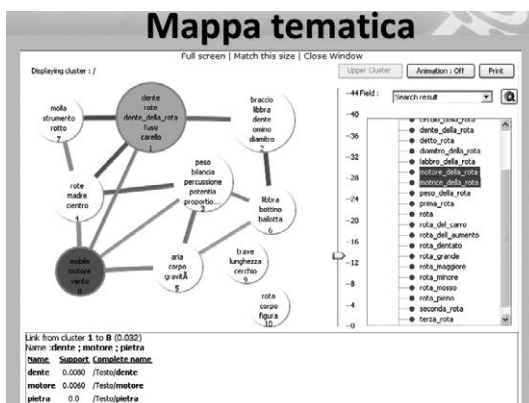
➔ **Le informazioni più rilevanti vengono identificate in automatico**

<sup>17</sup> È allo studio un programma di aggiornamento di *e-Leo* che vedrà la sostituzione della tecnologia Flash con lo standard HTML5 e che quindi consentirà di ovviare a questo evidente limite, nonché, fra le altre cose, di superare la struttura a frames, che sebbene supportata da tutti i browser attuali non è compatibile al 100% con l'indicizzazione dei motori di ricerca, di risolvere il problema della compatibilità con ogni apparato multimediale presente in commercio (sia esso un PC con qualsiasi sistema operativo, o un tablet o uno smartphone, senza escludere smart TV e altre apparecchiature ibride e consolle di gioco), di mettere a disposizione sistemi di *Text Mining* più avanzati. La nuova versione di *e-Leo* andrà in produzione presumibilmente entro il 2018 secondo un piano di sviluppo pluriennale che traghetterà *e-Leo* ad un vero e proprio *e-Leo 2.0* articolato intorno essenzialmente a tre obiettivi: affrontare e risolvere problemi di aggiornamento e sviluppo della tecnologia software impiegata, nonché della grafica; completare la messa in linea del sistema di *Text Mining* assistito da un motore di ricerca calibrato sull'italiano del Quattro-Cinquecento, come sopra detto; offrire ulteriori servizi per l'utente specialistico, anche integrandolo in uno spazio informativo unico con altri strumenti e banche dati specialistiche in un'ottica che guardi sempre di più verso l'interoperabilità sintattica.

Figura 4 – Classificazione.



Figura 5 – Mappa tematica.



Gli strumenti di analisi e di classificazione aggiungeranno al sistema funzionalità attualmente disponibili negli strumenti più avanzati di *Text Mining*: i testi analizzati saranno infatti indicizzati con descrittori lessicali dedotti automaticamente mediante analisi morfologica e statistica del testo. La capacità di lemmatizzare i termini di un testo sulla base di sequenze morfologiche tipiche della lingua offrirà la possibilità di identificare per un dato documento anche quelle combinazioni di termini che risultano nell'insieme rilevanti per la concettualizzazione del documento stesso. Ad oggi, mancando però una grammatica del linguaggio leonardiano che consenta di compiere correttamente tale operazione, questa funzionalità risulta incompleta. Il lavoro di definizione delle regole di flessione della lingua usata da Leonardo è però in corso per conto della Biblioteca Leonardiana e sarà messa in linea entro il prossimo anno.



Tuttavia, anche allo stato attuale, senza una grammatica dell'italiano del Quattro-Cinquecento, sarà possibile confrontare i documenti (le trascrizioni) risultato di una ricerca ed effettuare, sulla base delle co-occorrenze di termini, una classificazione automatica di questi documenti, la quale consente di raggruppare i testi in sottoclassi omogenee per contenuti, tali quindi da consentire una navigazione più rapida nel caso di grandi quantità di documenti. Non secondaria anche la possibilità di trovare nessi non scontati fra documenti diversi che possano essere messi in relazione concettualmente, pur non contenendo l'esatta parola chiave con cui si è lanciata la ricerca. Quest'ultima funzionalità si è dimostrata estremamente utile in diverse ricostruzioni di macchine e meccanismi realizzate intorno al 2004 dal Dipartimento di Meccanica e Tecnologie Industriali dell'Università degli Studi di Firenze per il Museo leonardiano di Vinci<sup>18</sup>.

#### 4. Conclusioni

Per concludere mi riallaccio a quanto detto inizialmente, ossia che il nostro obiettivo nel progettare *e-Leo* era stato di: 1. valorizzare il patrimonio della Biblioteca leonardiana in un contesto informativo sempre più orientato verso il digitale come quello attuale; 2. supportare la consultazione e l'accesso al contenuto – grafico e testuale – dell'opera manoscritta leonardiana con risorse scientifiche più o meno avanzate che potessero meglio rispondere alle necessità di ricerca degli utenti. La nostra biblioteca è un centro di ricerca e documentazione che promuove lo studio e la ricerca in un ambito estremamente ricco di impulsi di tutti i generi e di tutti i livelli, quale quello degli studi intorno a Leonardo da Vinci. Si trova a 'gestire', nonché valorizzare, un patrimonio speciale e molto raro, che si rivolge sostanzialmente ad utenti specialistici e quindi ad un bacino geografico di riferimento tutt'altro che locale, anche internazionale. È ubicata in un piccolo comune periferico a 40 km da Firenze sulle colline ai piedi del Montalbano, con una viabilità non eccezionale ed un'accessibilità fisica quindi limitata, elementi che vanno ad aggiungersi ad una difficoltà di impatto che la biblioteca speciale e di ricerca può avere, e ha, di per sé<sup>19</sup>. Ecco che la nostra risposta è stata quella di sfruttare le potenzialità che la 'biblioteca digitale' ci ha offerto, trasformando la nostra biblioteca da 'luogo fisico' anche in uno 'spazio virtuale'.

<sup>18</sup> Una riflessione, anche metodologica, su questo tema è fornita in Cascini, Nanni, Russo (2004: 419-435) e Russo (2013: 183-217). I risultati di questo nuovo approccio all'analisi tecnica di macchine, utilizzato in occasione dell'ampliamento e rinnovamento della collezione di modelli del Museo leonardiano (<<http://www.museoleonardiano.it>>), sono illustrati in: Neuwahl, Russo (2013: 226-319) e Nanni (2013: 226-319).

<sup>19</sup> Si rimanda a Manfron, Papi (2015) e ai riferimenti bibliografici in esso contenuti.

La risposta è stata *e-Leo*: uno strumento di accesso per utenti remoti di tutto il mondo<sup>20</sup> che possono accedere a distanza, senza restrizioni, alla collezione della biblioteca e agli strumenti di ricerca avanzata che essa offre relativamente al corpus leonardiano, alle fonti leonardiane e a quant'altro l'archivio si proporrà di ospitare; uno strumento di conoscenza dell'opera manoscritta di Leonardo da Vinci tramite la valorizzazione e la facilitazione dell'accesso ai testi leonardiani; infine, uno strumento che, partendo dai testi e dal loro studio rilancia numerosi campi di ricerca nell'ambito degli studi leonardiani (da quelli di linguistica a quelli di storia della meccanica, di ottica, di architettura ecc., oltre a quelli dell'informatica umanistica).

## Bibliografia

- Bambach C. C. 2009, *Una eredità difficile: i disegni e manoscritti di Leonardo tra mito e documento*, XLII Lettura Vinciana, 14 aprile 2007, Vinci, Giunti, Firenze.
- Biffi M. 2011, *E-Leo. Archivio digitale di storia della tecnica e della scienza* (<http://www.leonardodigitale.com>), «Studi italiani», XLV (1): 201-207.
- Cascini G., Nanni R., Russo D. 2004, *A contribution to history of technology: analyzing Leonardo's textile machines and his inventive process by TRIZ methods*, in Cascini G. (a cura di), *TRIZ Future Conference*, Firenze University Press, Firenze: 419-435.
- Lattanzi M., Colalucci F. (a cura di) 1992, *Vocabolario di controllo ricavato dalle parole chiave del sistema di classificazione iconografica Iconclass. Nomi propri storici, letterari e geografici. Locuzioni iconografiche*, ICCD, Roma, <<http://www.iccd.beniculturali.it/getFile.php?id=184>> (03/2016).
- Manfron A., Papi F. 2015, *Primi spunti di riflessione su impatto e valore delle biblioteche storiche e di ricerca: una sfida e un'opportunità*, «AIB Studi», LV (3): 359-372.
- Manni P. 2015, *Sulla lingua tecnico-scientifica di Leonardo. Bilancio di un decennio fecondo*, «Studi di Memofonte», 15: 44-52, <[http://www.memofonte.it/home/files/pdf/XV\\_2015\\_MANNI.pdf](http://www.memofonte.it/home/files/pdf/XV_2015_MANNI.pdf)> (03/2016).
- Manni P., Biffi M. 2008, *Saggio di un Glossario leonardiano*, «Studi di lessicografia italiana», 25: 45-63.
- Manni P., Biffi M. 2011a, *Introduzione*, in Manni P., Biffi M. (a cura di), *Glossario Leonardiano. Nomenclatura delle macchine nei Codici di Madrid e Atlantico*, Olschki, Firenze: XIX-XXVII.
- Manni P., Biffi M. 2001b, *Criteri di allestimento del Glossario*, in Manni P., Biffi M. (a cura di), *Glossario Leonardiano. Nomenclatura delle macchine nei Codici di Madrid e Atlantico*, Olschki, Firenze: XXIX-XXXV.

<sup>20</sup> Nel 2015 *e-Leo* ha contato: 27.347 visitatori diversi, 59.541 visite, 2.773.816 accessi, 1.194.395 pagine consultate. Le nazioni di provenienza degli accessi sono state le seguenti: 26,48% Italia, 19,07% Stati Uniti d'America, 9,38% Germania, 7,98% Francia, 7,48% Federazione Russa, 6,93% Spagna, 5,61% Turchia, 4,58% Gran Bretagna, 2,82% Repubblica Ceca, 0,94% Cina, il restante 8,73% altri paesi.

- Nanni R. 2011, *E-Leo: Online Digital Archive for studying the history of Renaissance science and technology*, in Cappellini V. (ed.), *EVA 2011, Conference, workshops, meetings, training & exhibition*, Pitagora, Bologna: 88-93.
- Plances E. 1998, *Soggettario iconografico. Introduzione*, ICCD, Roma, <<http://www.iccd.beniculturali.it/getFile.php?id=185>> (03/2016).
- Quaglino M. 2013a, *Introduzione*, in Quaglino M. (a cura di), *Glossario Leonardiano. Nomenclatura dell'ottica e della prospettiva nei Codici di Francia*, Olschki, Firenze: IX-XVII.
- Quaglino M. 2013b, *Criteri di allestimento del Glossario*, in Quaglino M. (a cura di), *Glossario Leonardiano. Nomenclatura dell'ottica e della prospettiva nei Codici di Francia*, Olschki, Firenze: XVII-XXI.
- Russo D. 2013, *Nuove ipotesi sulle macchine da cantiere brunelleschiane*, in Nanni R. (a cura di), *Leonardo e le arti meccaniche*, Skira, Milano: 183-217.
- Straten R. van 2009, *Introduzione all'iconografia*, edizione italiana a cura di R. Cassanelli, Jaca Book, Milano.
- Taddei M. 2005, *E-leo: archivio digitale di storia della tecnica e della scienza: the Leonardo Library's reference and research tool for disseminating and enhancing knowledge of Leonardo's work. State of the art and future prospects*, 24 novembre 2015, Académie de l'Architecture, Parigi, intervento nel seminario *Dessins, modèles, architecture, histoire de l'art: l'étude historique et la restitution d'hypothèses à l'âge du numérique*, École pratique des Hautes Études e Laboratoire MAP (UMR 3495 CNRS/MCC), Paris.
- Waal H. van de 2002, *Iconclass*, ed. italiana a cura di M. Lattanzi, S. Ciofetta, E. Plances, ICCD, Roma.

J.-C. Zancarini  
S. Gedzelman

## Le voyage en France du *Prince* de Machiavel. L'outil *HyperMachiavel* et ses effets de sens

*Résumé:* Cette contribution présente *HyperMachiavel*, un outil que nous avons développé dans le but de comparer le texte du *Prince* de Machiavel avec ses différentes traductions françaises. Les fonctions offertes par cet outil, à savoir la vue synoptique, l'affichage des index des formes dans chaque langue et de leur fréquence, l'exploration des équivalences par le biais de la mise en évidence graphique, permettent de réfléchir à la fois à la complexité du texte d'origine et aux sens multiples des mots utilisés par l'auteur. *HyperMachiavel* donne aussi la possibilité d'analyser l'acte de traduction de chaque traducteur en montrant les effets produits selon les choix qui sont faits vis-à-vis du lexique du texte original. Deux exemples sont donnés ici: le premier porte sur les langues politiques de départ et d'arrivée, notamment sur le mot *stato* et sur ses traductions; le second sur la façon dont un des traducteurs du *Prince*, Amelot de la Houssaye, traduit le style et le lexique de Machiavel.

*Mots-clés:* Machiavel, langue politique, histoire des traductions, corpus parallèle, TEI.

*Riassunto:* Questo contributo presenta *HyperMachiavel*, uno strumento che è stato realizzato allo scopo di confrontare il testo del *Principe* di Machiavelli con le sue diverse traduzioni francesi. Le funzioni offerte da *HyperMachiavel*, ovvero, la vista sinottica, la visualizzazione degli indici delle forme in ogni lingua e della loro frequenza, l'esplorazione delle equivalenze attraverso l'evidenziazione grafica, permettono di riflettere al tempo stesso sulla complessità del testo originale e sui molteplici sensi delle parole utilizzate dall'autore. *HyperMachiavel* offre anche la possibilità di analizzare l'atto del tradurre di ogni traduttore mostrando gli effetti prodotti a seconda delle

scelte effectuées respecto al lessico del testo originale. Due esempi vengono proposti: il primo riguarda le lingue politiche di partenza e di arrivo, precisamente la parola 'stato' e le sue traduzioni; il secondo il modo in cui uno dei traduttori del *Principe*, Amelot de la Houssaye, traduce lo stile e il lessico di Machiavelli.

*Parole chiave:* Machiavelli, linguaggio politico, storia delle traduzioni, corpus parallelo, TEI.

## I. Introduction

L'outil *HyperMachiavel* permet de comparer un texte (ici le *Prince*) avec ses différentes traductions, de façon à pouvoir réfléchir d'une part à la complexité même du texte d'origine et aux sens multiples des mots utilisés par l'auteur et, d'autre part, à ce qui fait la spécificité d'un traduire en acte et aux effets qu'il produit. Le choix du *Prince* pour la construction de l'outil n'est évidemment pas le fruit du hasard mais découle de l'intimité que nous avons avec ce texte, traduit et commenté par nos soins (Machiavel 2000: 640): nous avons donc décidé de comparer l'*editio princeps* (Blado 1532) avec (pour commencer) les trois traductions françaises du XVI<sup>e</sup> siècle: Jacques de Vintimille (1546), Gaspard d'Auvergne et Jacques Cappel (1553), Jacques Gohory (1571). Dans un second temps, nous avons ajouté la principale traduction du XVII<sup>e</sup> siècle, celle de Amelot de la Houssaye (1683-1694) dans laquelle on a lu le *Prince* jusqu'à la fin du XVIII<sup>e</sup> siècle. La mise au point de cet outil n'a pu se faire que sur la base d'une traduction 'à plusieurs interprètes' connaissant chacun des 'langues différentes', ainsi qu'il en alla souvent, du XII<sup>e</sup> au XVI<sup>e</sup> siècle dans les processus de traduction de l'arabe vers le latin (D'Alverny 1989)<sup>1</sup>: dans le cas de la 'paire' qui a conçu et réalisé l'HM, il y avait aussi deux langues différentes qui communiquaient; d'un côté, la pratique-théorie de la traduction et de la réflexion politique, de l'autre, la programmation informatique. Après avoir présenté les caractéristiques de l'outil HM nous donnerons deux exemples de la réflexion que permet son usage, le premier portant sur les langues politiques de départ et d'arrivée et le second sur la façon dont un des traducteurs du *Prince* (Amelot de la Houssaye) traduit le 'style' et le lexique de Machiavel.

## 2. Présentation du logiciel *HyperMachiavel* (HM)

Le développement du logiciel HM s'inscrit dans un moment où la philologie bascule vers un monde plus systémique et plus opératoire. L'édi-

<sup>1</sup> Elle donne des exemples qui vont du XII<sup>e</sup> siècle (en Espagne) au XVI<sup>e</sup> siècle (à Venise).

tion numérique est en pleine expansion, les ressources sont disponibles pour une plus grande communauté au-delà de celle des chercheurs; l'usage des textes devient donc multiple et, de ce fait, il est nécessaire de prendre en compte qu'ils peuvent évoluer. Quel que soit leur usage, ces ressources sont au centre et pour que leur manipulation et leur exploitation ne soient pas entravées, il faut d'emblée les penser dans un modèle et un format suffisamment ouverts et interopérables: nous détaillerons donc les différentes modalités d'importations vers le logiciel, qui se fondent principalement sur les propositions d'encodage du consortium TEI (*Text Encoding Initiative*).

Nous avons intégré trois dimensions nécessaires à la manipulation et à l'interrogation d'un corpus parallèle:

- la première dimension concerne l'appareillage de la lecture synoptique comme elle est aussi proposée en édition numérique, pour voir différentes facettes et transcriptions d'un même texte;
- la deuxième intègre quelques outils classiques utilisés en linguistique de corpus, à savoir des index et des concordanciers monolingues ou bilingues. En lien avec les index, nous proposons également des fonctions autour des dictionnaires, plus couramment dédiées au travail des lexicographes;
- la troisième aborde les fonctions dédiées à l'alignement de lexique et à plusieurs idées de visualisation pour aider l'utilisateur à explorer ces équivalences traductionnelles, qu'elles soient faites ou restent à faire.

## 2.1 L'appareillage de la lecture synoptique

La lecture des textes en parallèle dans HM est centrale, chaque fonction nous ramène à cette vue synoptique permettant le choix dans la disposition des textes, la re-contextualisation du sens des mots étudiés. Pour pouvoir proposer la fonction d'alignement des textes, cette vue paraît fondamentale, cependant les outils d'alignement ne mettent généralement en scène qu'une paire de textes (parfois plusieurs paires), la paire étant composée d'un texte 'cible' et d'un texte 'source'<sup>2</sup>. Pour notre objectif d'étude de traductions multiples, aller au-delà du 'bitexte' est un apport réel; ces vues synoptiques témoignent également d'une nouvelle pratique dans les éditions numériques en ligne, qui confrontent parfois différentes éditions d'un même texte, comme il est possible de le faire avec l'application web *Versioning Machine*<sup>3</sup>, ou différentes facettes d'un seul manuscrit avec ses

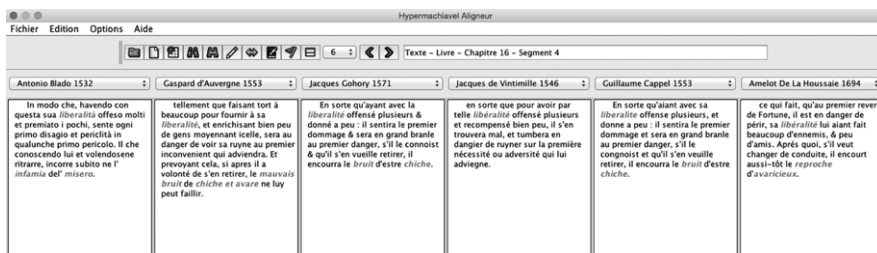
<sup>2</sup> Dans le contexte de l'alignement automatique, Harris (1998) introduit une nouvelle notion avec le terme 'bitexte' et montre ainsi que chaque texte n'est plus tout à fait une unité à part entière, conception très différente de l'usage que nous souhaitons faire de nos textes parallèles.

<sup>3</sup> <<http://v-machine.org/>> (06/2017).

transcriptions multiples (normalisée, diplomatique, etc.), comme c'est le cas pour la *Queste del Saint Graal* publiée dans le portail web TXM<sup>4</sup>.

L'outil HM orchestre donc cette vue (Fig. 1) avec le nombre de textes dont l'utilisateur a besoin pour étudier le style des traducteurs, aligner une nouvelle traduction ou bien mettre en relation des équivalences traductionnelles au niveau des occurrences entre un et plusieurs textes. Au niveau de l'encodage pour représenter les corpus parallèles, HM ne gère plus qu'un modèle de données en XML.

Figure 1 – Vue synoptique présentant les cinq traductions françaises en parallèle de la version princeps italienne du Principe de Machiavel.



Le modèle TMX<sup>5</sup>, propre à l'utilisation dans les outils d'alignement, opère un aplatissement de la structure d'origine et présente un fichier corpus pour chaque paire de langue (it-fr), obligeant à la manipulation d'autant de paires de texte qu'il existe de combinaisons source-cible. Il entre donc en conflit avec l'idée de donner une autre vie à ces textes, telle que l'édition numérique, et introduit une grande complexité pour la gestion du parallélisme entre plusieurs sources textuelles ayant la même langue. Notre choix s'est donc porté sur la TEI qui, elle, peut considérer des équivalences à différents niveaux du texte (*div*, *p*, *s*, *span*, etc.). HM en réduit cependant les possibilités en exigeant une parallélisation stricte des divisions et des sous-segments philologiques, en plus du système de liaison par les balises *linkGrp*<sup>6</sup> et les attributs *id* de chaque élément. Un exemple est donné dans l'Extrait 1 ci-dessous :

```
<div n="16" type="Chapter" xml:id="i1_Ch16">
  <head type="Title" xml:id="i1_Ch16-Seg0"/>
  <seg n="1" type="Segment" xml:id="i1_Ch16-Seg1">Cominciandomi
adunche a le prime soprascritte qualità, dico come sarebbe bene esser'
tenuto liberale.</seg>
</div>
```

<sup>4</sup> <<http://txm.bfm-corpus.org/?command=documentation&path=/GRAAL>> (06/2017).

<sup>5</sup> <[https://en.wikipedia.org/wiki/Translation\\_Memory\\_eXchange](https://en.wikipedia.org/wiki/Translation_Memory_eXchange)> (06/2017).

<sup>6</sup> <<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-linkGrp.html>> (06/2017).

```

<div n="16" type="Chapter" xml:id="f3_Ch16">
  <head type="Title" xml:id="f3_Ch16-Seg0"/>
  <seg n="1" type="Segment" xml:id="f3_Ch16-Seg1">Pour déclarier donc
  au menu le bien et le mal de toutes les meurs et qualités des hommes, que
  cy dessus ont esté touchées, je diz, en commençant à la première, qu'il
  seroit bien bon d'estre estimé liberale,</seg>
</div>
<linkGrp type="alignment" domains="#i1_Ch16 #f3_Ch16">
  <link target="#i1_Ch16-Seg0 #f3_Ch16-Seg0" />
  <link target="#i1_Ch16-Seg1 #f3_Ch16-Seg1" />
</linkGrp>

```

Extrait 1 du fichier corpus présentant tous les textes avec une structure identique, et le système pour l'alignement.

En ce qui concerne les équivalences en deçà d'un paragraphe, au niveau d'un mot ou d'un petit groupe de mots, nous avons fait le choix d'élaborer un modèle XML *stand-off*, propre à HM et qui pourrait devenir une proposition d'extension de la TEI. L'Extrait 3 montre ainsi un exemple d'équivalences, dont chacune fait référence aux annotations dans le corpus de l'Extrait 2. Ce modèle permet l'adjonction d'autres éléments liés au travail d'annotation et de mise en liaison; il peut, par exemple, coder le statut des équivalences établies (incertain/certain) ou l'indication de dates et d'auteurs ou bien encore indiquer en commentaire la nature de la traduction en langue naturelle (périphrase, etc.) ou avec un système de codes, si possible déjà existant et élaboré dans le cadre d'une communauté de spécialistes:

```

<div n="16" type="Chapter" xml:id="i1_Ch16">
  <head type="Title" xml:id="i1_Ch16-Seg0"/>
  <seg n="1" type="Segment" xml:id="i1_Ch16-Seg1">Cominciandomi
  adunche a le prime soprascritte <seg ana="Hyperprince_Dictionary_
  it.xml#C_234" xml:id="Annot_1683">qualità</seg>, dico come sarebbe
  bene esser' tenuto liberale.</seg>
</div>
<div n="16" type="Chapter" xml:id="f3_Ch16">
  <head type="Title" xml:id="f3_Ch16-Seg0"/>
  <seg n="1" type="Segment" xml:id="f3_Ch16-Seg1">Pour déclarier donc
  au menu le bien et le mal de toutes les meurs et <seg ana="Hyperprince_
  Dictionary_frm.xml#C_213" xml:id="Annot_9537">qualités</seg> des
  hommes, que cy dessus ont esté touchées, je diz, en commençant à la pre-
  mière, qu'il seroit bien bon d'estre estimé liberale,</seg>
</div>

```



Extrait 2 du fichier corpus présentant le même extrait, avec les annotations *qualità* et ‘qualités’, respectivement dans la version de texte de A. Blado, 1532 (i1) et de J. De Vintimille, 1546 (f3).

```
<equivalence loc="Ch16-Seg1" id="I_500">
  <word id="Annot_6909" refText="f1">
    <text>qualitez</text>
    <comment />
  </word>
  <word id="Annot_4312" refText="f5">
    <text>qualitez</text>
    <comment />
  </word>
  <word id="Annot_9537" refText="f3">
    <text>qualités</text>
    <comment />
  </word>
  <word id="Annot_1683" refText="i1">
    <text>qualità</text>
    <comment />
  </word>
  <word id="Annot_12120" refText="f6">
    <text>qualitez</text>
    <comment />
  </word>
  <word id="Annot_17613" refText="f7">
    <text>qualitez</text>
    <comment>ajouté automatiquement en juin 2014</comment>
  </word>
  <comment />
</equivalence>
```

Extrait 3 du fichier des équivalences, avec la correspondance entre l’occurrence du texte italien (i1) *qualità* et celles des traductions (f1, f3, f5, f6, f7).

À propos du niveau de l’alignement, HM permet de re-segmenter manuellement l’œuvre plus finement pour que les fonctions de recherche et d’alignement semi-automatique soient optimisées et que la lecture synoptique et l’interprétation soient facilitées. Cependant une difficulté peut alors être rencontrée lors de la mise en équivalence d’occurrences, qui se situe à cheval sur deux segments philologiques différents. Lors de l’ajout de la traduction du XVII<sup>e</sup> siècle (Amelot de La Houssaye) dans le corpus *HyperPrince*, nous avons en effet constaté un certain nombre d’inversion du texte d’origine. Une fonction de fusion des segments est à ce jour proposée mais *in fine* l’outil devra implémenter un mécanisme d’annotation d’équivalences un peu plus souple.

## 2.2 La manipulation du lexique

Lors de la construction d'un corpus dans HM, chaque texte est tokenisé, ce qui permet un affichage des index de formes dans chaque langue. Cette synthèse s'apparente à celle du vocabulaire proposé en textométrie. La disposition alignée de ce vocabulaire, avec pour chaque forme, les textes côte à côte indiquant la fréquence de cette forme chez chacun, peut faire état de l'usage plus ou moins important d'une forme par rapport aux autres. Dans un contexte de corpus constitué uniquement de textes antérieurs au XVII<sup>e</sup> siècle, cette lecture comparative entre les textes semble une voie intéressante pour l'étude du contexte stylistique de l'auteur, et de la grande variabilité orthographique de manière chronologique (exemple de 'toutefois', Fig. 2 et Fig. 3).

Figure 2 – Les différentes formes des 'toutefois', réparties chez chaque auteur français.

Forme lexicale	Nb Gaspard d'Avvergne 1553	Nb Jacques Gohory 1571	Nb Jacques de Vintimille 1546	Nb Guillaume Cappel 1553	Nb Amelot De La Houssaie 1694
tours	1	0	2	0	0
tous	76	67	115	64	63
tousiours	68	69	0	70	0
tousjour	0	0	1	0	0
tousjours	0	0	37	0	0
tout	97	73	94	84	78
toute	36	42	49	45	36
touteffois	0	0	0	0	0
touteffoys	0	0	25	0	0
toutefois	2	16	0	2	1
toutfois	0	0	1	0	0
toutes	50	58	71	58	60
toutesfois	24	6	0	22	0
toutesfoys	2	0	0	0	0
touts	0	0	0	1	0
toutfois	0	0	0	0	0
toutteffois	0	0	10	0	0
toutz	0	0	1	0	0

Figure 3 – Une concordance KWIC appelée depuis la vue vocabulaire, après sélection des diverses formes de 'toutefois'.

The screenshot displays the HyperMachavel software interface. At the top, there are tabs for 'Equivalences', 'Recherche bilingue', 'Résultats', and 'Vocabulaire'. Below these is a search bar with 'Recherche rapide' and 'Voir contexte' buttons. A table shows the frequency of various forms of 'toutefois' for five authors: Gaspard d'Avvergne 1553, Jacques Gohory 1571, Jacques de Vintimille 1546, Guillaume Cappel 1553, and Amelot De La Houssaie 1694. Below the table, there are several text excerpts from different authors, each with a highlighted instance of 'toutefois' or a related form. At the bottom, a search results window shows 52 results for the query 'toutesfois', listing the ID, location, text, and context for each occurrence.

À partir des index de formes, un autre travail peut être mis en route dans l'optique d'une élaboration de dictionnaires monolingues ou même bilingues et par la suite d'une lemmatisation partielle ou complète du corpus. Pour des corpus de textes anciens dont les ressources dictionnairiques n'existaient pas encore (*Vocabolario della Crusca* en 1612 et *Dictionnaire de l'Académie française* en 1694), il semble plus intéressant de confier ces opérations à l'utilisateur. Il est néanmoins possible au moment de la première importation des textes d'utiliser *TreeTagger*<sup>7</sup> (avec pour l'instant les langues suivantes: it, fr, en) à défaut d'autres ressources disponibles.

HM scénarise l'édition de ces formes canoniques, normalisées, telles des entrées de dictionnaire, au moyen d'une arborescence présentant deux niveaux de profondeur: 'Lemme', 'Forme', et un troisième niveau, ajouté pour signaler le lien direct avec les textes: 'Occurrence'. Ces informations autour du lemme sont ajoutées à la fois aux fichiers de *tokenization* avec l'attribut *idLemma* et dans les fichiers dictionnaires dont le modèle proposé est TEI (voir les extraits 4 et 5). Le travail du dictionnaire avec fusion et rattachement de formes à un autre lemme est particulièrement facilité par ce type d'affichage (déplacement d'un nœud dans l'arbre par glisser-déposer):

```
<Text>
  <Token debIndex="0" endIndex="13" id="0" idLemma="C_3305"
indexed="True" pos="" type="WORD">Cominciandomi</Token>
  <Token debIndex="13" endIndex="14" id="1" indexed="True"
type="SPACE"></Token>
  <Token debIndex="14" endIndex="21" id="2" idLemma="C_3306"
indexed="True" pos="" type="WORD">adunche</Token>
  <Token debIndex="21" endIndex="22" id="3" indexed="True"
type="SPACE"></Token>
  <Token debIndex="22" endIndex="23" id="4" idLemma="C_19"
indexed="True" pos="" type="WORD">a</Token>
  <Token debIndex="23" endIndex="24" id="5" indexed="True"
type="SPACE"></Token>
  <Token debIndex="24" endIndex="26" id="6" idLemma="C_12"
indexed="True" pos="" type="WORD">le</Token>
  <Token debIndex="26" endIndex="27" id="7" indexed="True"
type="SPACE"></Token>
  <Token debIndex="27" endIndex="32" id="8" idLemma="C_3307"
indexed="True" pos="" type="WORD">prime</Token>
  <Token debIndex="32" endIndex="33" id="9" indexed="True"
type="SPACE"></Token>
```

<sup>7</sup> <<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>> (06/2017).

```

<Token debIndex="33" endIndex="45" id="10" idLemma="C_3308"
indexed="True" pos="" type="WORD">soprascritte</Token>
  <Token debIndex="45" endIndex="46" id="11" indexed="True"
type="SPACE"> </Token>
  <Token debIndex="46" endIndex="53" id="12" idLemma="C_234"
indexed="True" pos="" type="WORD">qualità</Token>
  <Token debIndex="53" endIndex="54" id="13" indexed="True"
type="PUNCT">,</Token>
  <Token debIndex="54" endIndex="55" id="14" indexed="True"
type="SPACE"> </Token>
  <Token debIndex="55" endIndex="59" id="15" idLemma="C_1450"
indexed="True" pos="" type="WORD">dico</Token>

```

Extrait 4: Extrait du fichier de tokenization correspondant à la deuxième division du Chapitre 16, après le titre.

```

<entry xml:lang="it" xml:id="C_234">
  <form>
    <orth type="lemma">qualità</orth>
    <orth>qualitati</orth>
    <orth>qualitati di tempi</orth>
    <orth>qualità de tempi</orth>
    <orth>qualità di tempo</orth>
  </form>
  <gramGrp><pos /></gramGrp>
  <sense><note /></sense>
</entry>

```

Extrait 5: Extrait du fichier du dictionnaire italien avec la déclaration d'un pseudo-lemme *qualità*.

## 2.3 Les équivalences

L'originalité de l'outil se situe également au niveau de l'acquisition des équivalences et de leur exploration. Les méthodes et techniques d'alignement automatique, dont un excellent état de l'art a été produit par J. Véronis en 2000, ont parfois des interfaces intégrant une recherche bilingue mais elles livrent le plus souvent des index bilingues dénués de retour au texte. Nous nous situons plutôt dans un paradigme où les opérations sont faites manuellement avec la possibilité de conserver la relation au texte pour permettre un contrôle et une exploration de l'alignement lexical. Les algorithmes de recherche dans HM tirent tout de même parti d'une mémoire des équivalences, qui ont été validées au fur et à mesure. Le tableau des résultats de cette recherche bilingue montre l'absence et/ou la co-pré-

sence du motif recherché en source et en cible et affiche une colonne par texte (exemple dans la Figure 4):

Figure 4 – Tableau des résultats d’une recherche bilingue «spetie» (it.) et ‘espece’ (fr.).

The screenshot shows a window titled 'Recherche bilingue' with tabs for 'Vocabulaire', 'Equivalences', and 'Résultats'. The 'Résultats' tab is active, displaying 16 results for the query 'spetie <=> espece'. The results are organized in a table with columns for 'Localisation', 'Source', and 'Cible'. Each row represents a text source and its corresponding target word.

Localisation	Source	Cible
Texte - Livre ...	spetie	especies
Texte - Livre ...		espece
Texte - Livre ...		espece
Texte - Livre ...		espece
Texte - Livre ...		espece
Texte - Livre ...		espece
Texte - Livre ...		espece
Texte - Livre ...		espece
Texte - Livre ...		espece
Texte - Livre ...		espece
Texte - Livre ...		espece
Texte - Livre ...		espece
Texte - Livre ...		espece
Texte - Livre ...		espece
Texte - Livre ...		espece
Texte - Livre ...		espece

La validation d’une ou plusieurs équivalences selon les lignes permet d’aboutir à un tableau semblable (Fig. 5) mais enrichi à mesure que l’utilisateur le complète avec les résultats d’autres recherches bilingues ou avec l’éditeur d’équivalence (Fig. 6):

Figure 5 – Quelques équivalences traductionnelles au niveau des occurrences du lemme «ordine» (it).

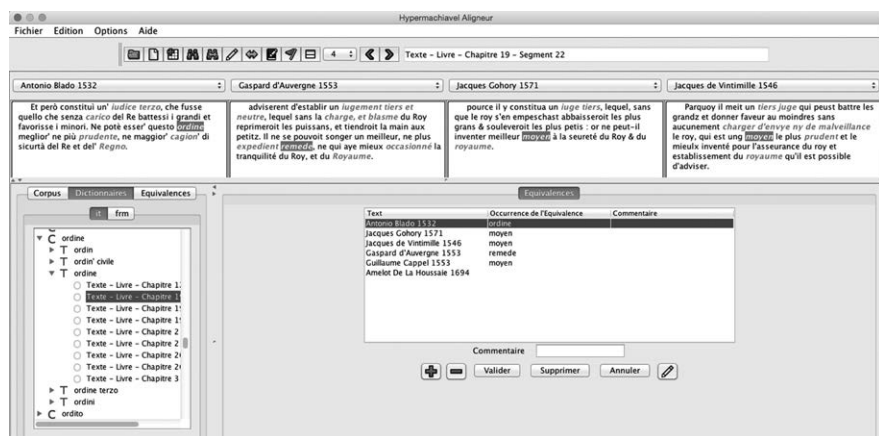
The screenshot shows a window titled 'Hypermachavi Aligneur' with a menu bar (Fichier, Edition, Options, Aide) and a toolbar. The main window displays search results for the query 'null' (likely a typo for 'ordine'). The results are organized in a table with columns for 'Localisation', 'Cible', and 'Source'. Each row represents a text source and its corresponding target word.

Localisation	Cible	Source
Texte - Livre - Chapitre 1	ordre	ordine
Texte - Livre - Chapitre 1	ordre	ordine
Texte - Livre - Chapitre 1	terce institution	ordine
Texte - Livre - Chapitre 1	ordres	ordine
Texte - Livre - Chapitre 1	ordres	ordine
Texte - Livre - Chapitre 1	ordres	ordine
Texte - Livre - Chapitre 1	ordres	ordine
Texte - Livre - Chapitre 1	ordres	ordine
Texte - Livre - Chapitre 1	ordres	ordine
Texte - Livre - Chapitre 1	ordres	ordine
Texte - Livre - Chapitre 1	ordres	ordine
Texte - Livre - Chapitre 1	ordres	ordine
Texte - Livre - Chapitre 1	ordres	ordine
Texte - Livre - Chapitre 1	ordres	ordine
Texte - Livre - Chapitre 1	ordres	ordine
Texte - Livre - Chapitre 1	ordres	ordine
Texte - Livre - Chapitre 1	ordres	ordine

Pour l’exploration des équivalences validées, plusieurs affichages ont été développés, dont une perspective sous forme de graphes (avec la librairie *Prefuse*)<sup>8</sup>: nous en donnons des exemples ci-après.

<sup>8</sup> Information Visualisation Toolkit, <<http://www.prefuse.org>> (06/2017).

Figure 6 – Cas d'un groupe d'équivalences traductionnelles pour une des occurrences d'«ordine» à compléter manuellement avec l'éditeur d'équivalence.



### 3. HM et les effets de sens qu'il produit

#### 3.1 Stato et ses traductions: polysémie, 'brouillage', 'tension du sens'

Le graphe de niveau 1 de l'outil HM (Fig. 7) donne l'ensemble des traductions utilisées pour un mot (ici: *stato*) par l'ensemble des traducteurs. Il génère également un tableau du nombre des occurrences des mots utilisés pour traduire, ce qui permet de voir les traductions les plus fréquentes et celles qui sont exceptionnelles. Un graphe de niveau 2 (Fig. 8) donne, outre les traductions du mot envisagé, les autres mots italiens qui ont été traduits par les mots français qui ont servi à traduire *stato*. Nous envisagerons tour à tour le graphe niveau 1 de *stato*, puis le graphe niveau 2, avant de revenir sur les traductions françaises principales de *stato*: 'état', 'seigneurie', 'pays'.

On constate une grande polysémie du terme, qui se marque par l'emploi de plus de quarante termes français. Il y a cependant trois traductions majoritaires de *stato*: 'état', 'seigneurie' et 'pays'.

L'aller-retour italien-français-italien met en évidence un lien de sens entre la majorité des mots du vocabulaire des institutions politiques. Pour éviter les biais qui pourraient avoir été provoqués par des mots français rarement, voire exceptionnellement, utilisés pour traduire *stato*, nous allons revenir uniquement sur les trois mots qui sont le plus fréquemment utilisés par les traducteurs pour rendre *stato* ('état', 'seigneurie', 'pays') et voir quels autres mots italiens ils servent à traduire (Fig. 9-10-11).

Figure 7 – Graphe «stato» niveau 1.

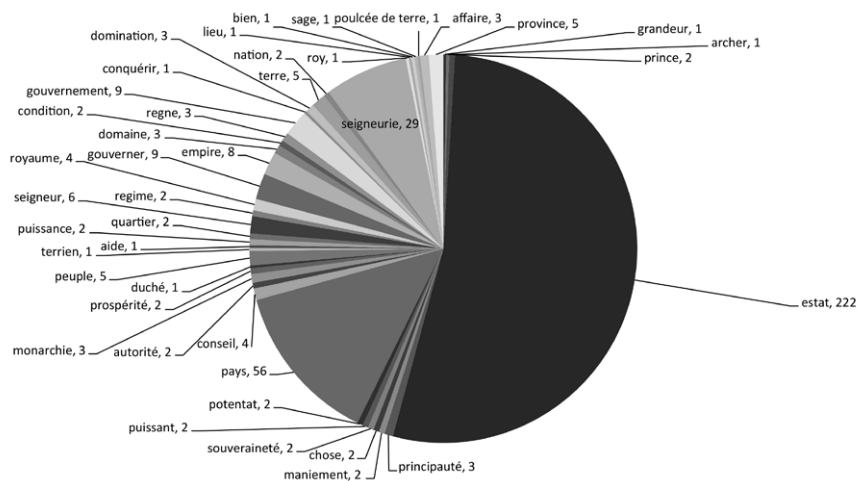


Figure 8 – Graphe «stato» niveau 2.

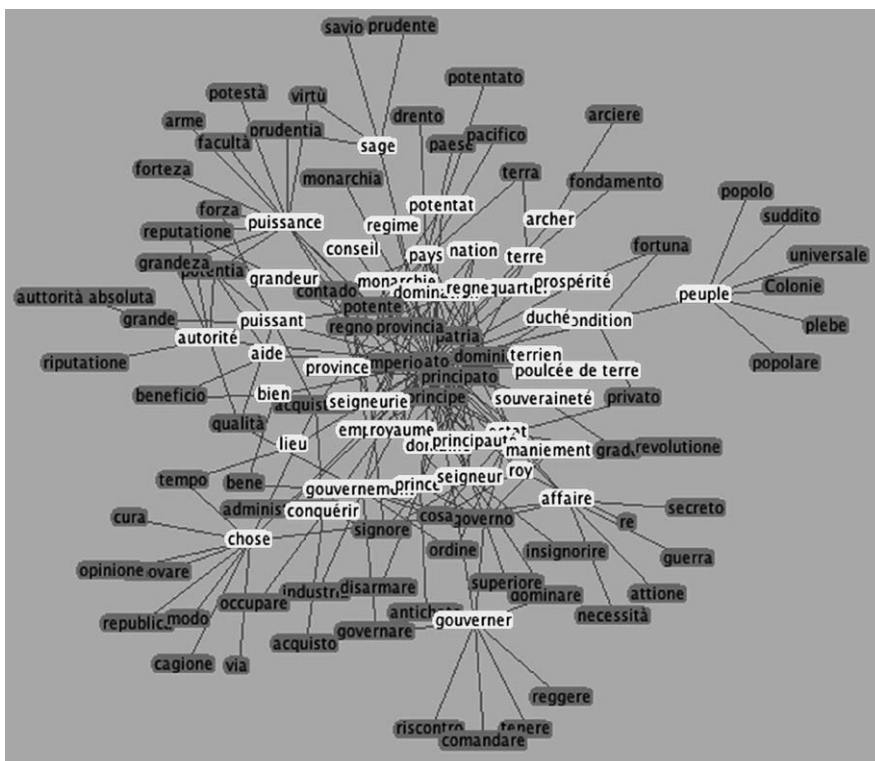


Figure 9 – Graphe 'estat' niveau 1. 'Estat' sert à traduire «stato», «mais aussi imperio», «dominio», «principato», «governo», «principe».

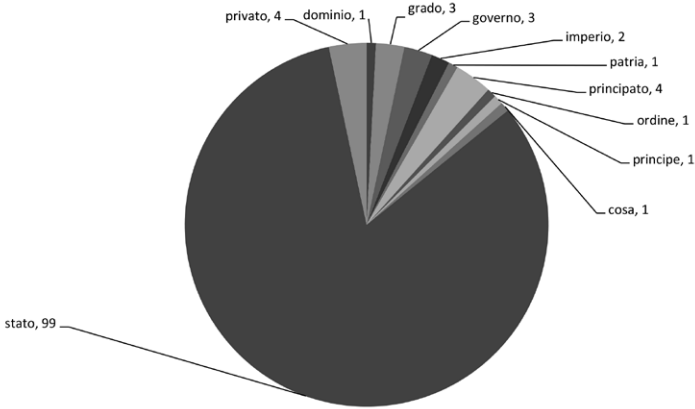


Figure 10 – Graphe 'pays' niveau 1. 'Pays' sert à traduire surtout «stato et provincia mais aussi paese», «dominio», «regno», «patria».

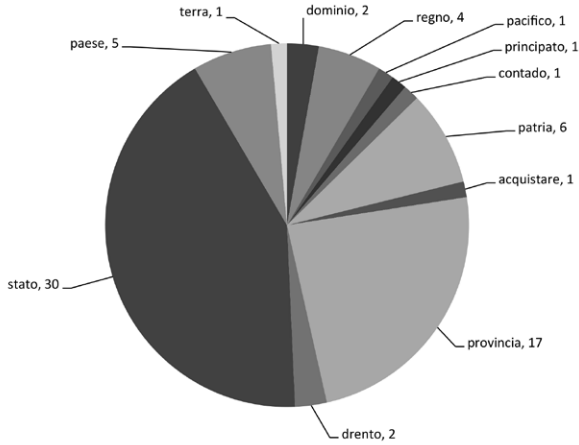
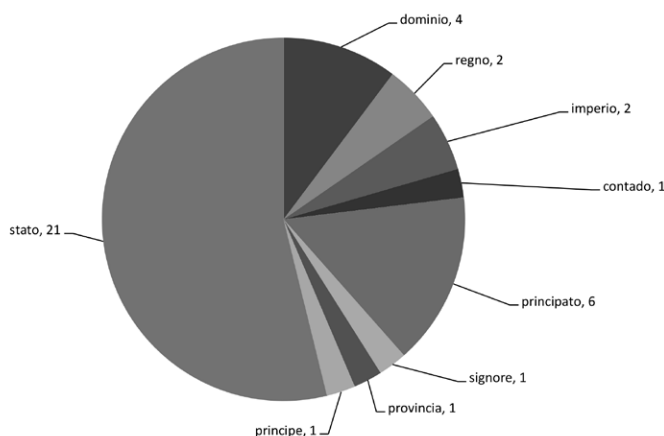




Figure 11 – Graphe ‘seigneurie’ niveau 1. ‘Seigneurie’ traduit *stato* mais aussi «*principato*», «*dominio*», «*imperio*», «*regno*».



On peut continuer cette vérification en revenant à l’italien comme point de départ et en examinant les traductions françaises de trois des mots (*dominio*, *imperio*, *principato*) qui ont été traduits par les trois mots français les plus utilisés pour traduire *stato*. *Dominio* est traduit principalement par ‘seigneurie’, ‘domination’, ‘règne’, ‘pays’, ‘domaine’. *Imperio* l’est très majoritairement par ‘empire’ ou ‘empereur’, mais aussi par ‘seigneurie’, ‘puissance’, ‘gouvernement’, ‘estat’. *Principato* est traduit généralement par ‘principauté’ ou par ‘prince’, mais également par ‘monarchie’, ‘seigneurie’, ‘estat’, ‘empereur’, ‘empire’, ‘souveraineté’, ‘seigneur’, ‘royaume’.

L’aller-retour italien>français>italien nous permet donc d’énoncer plusieurs thèses:

1. Sur le sens même de *stato* chez Machiavel: les mots qui entourent *stato* en italien et en français, ces quasi-synonymes, désignent trois sens principaux: territoire (pays, province; *paese*, *provincia*, *patria*, *contado*); forme de gouvernement (principauté, empire, monarchie, royaume, gouvernement; *governo*, *regno*, *principato*, *imperio*); sens politico-juridique de domination et commandement (puissance, domination, seigneurie, souveraineté, *dominio*, *imperio*);
2. Sur la langue du *Prince* et sur les effets des traductions du vocabulaire de la politique. Les ‘choses nouvelles’ que Machiavel énonce et dont il recherche les fonctionnements sont complexes; leurs ‘territoires sémantiques’ se recoupent et se chevauchent. Il y a là une tension constante du sens, une polysémie des mots utilisés pour dire la politique. Cela s’explique par la façon dont Machiavel dessine en creux, précisément en utilisant les mêmes termes dans des sens parfois différents, les objets nouveaux ou les formes de l’agir politique qu’il voulait comprendre et

dont il voulait faire comprendre le fonctionnement à ses lecteurs. Cela implique de lire *Le Prince* en admettant que cohabitent d'emblée différentes significations des termes, qui ne doivent donc pas être considérés au sens strict comme des concepts, faute de quoi on serait amené à parler d'incohérences ou de contradictions ou à vouloir à toute force reconstruire une cohérence au prix de torsions;

3. Sur la langue politique en Italie et en France (au moins au XVI<sup>e</sup> et au XVII<sup>e</sup> siècles): il y a un 'brouillage' d'origine dans l'analyse des fonctionnements de la politique; ce brouillage est renforcé par les effets de la traduction, une série de termes tendant à devenir 'équivalents' et à désigner globalement le lieu et les formes du pouvoir. Cette thèse nous paraît devoir s'appliquer non seulement à Machiavel mais aussi à la langue politique française qui le traduit.

### **3.2 L'analyse du traduire en acte: la traduction d'Amelot de la Houssaye**

HM permet aussi d'analyser l'acte de traduction de chaque traducteur en montrant quels choix sont faits vis-à-vis du lexique du texte original, de quelles façons sont traitées l'écriture et la structure du texte de départ (ordre de l'argumentation, structure des phrases, ajouts ou suppressions). Pour l'analyse du lexique nous utilisons la fonction de mise en évidence graphique des équivalences traductionnelles; pour l'analyse de l'écriture, nous utilisons la fonction 'vue synoptique' de notre outil. Notre méthode d'investigation n'est en aucun cas normative: il ne s'agit pas de dire qui traduit bien ou mal et encore moins d'explicitier ce que devrait être une 'traduction parfaite'. Nous tendons à analyser ce qui est fait par un traducteur quand il traduit un texte, donc à décrire le plus finement possible les actes de traduction et à rendre compte de ces choix. Dans une contribution récente (Zancarini 2015), nous avons mené une analyse comparative entre les cinq traducteurs du XVI<sup>e</sup> et du XVII<sup>e</sup> siècle. Ici nous nous concentrerons sur le cas d'Amelot de la Houssaye, le traducteur du *Prince* de la fin du XVII<sup>e</sup> siècle (1683-1694), dont la traduction est lue et fait référence jusqu'à la fin du XVIII<sup>e</sup> siècle (Barthas 2014).

Dans la préface de son édition, Amelot de la Houssaye définit la façon dont il conçoit sa traduction:

Elle est si fidèle, que ie pourois me vanter, qu'il seroit assez difficile d'en faire une, qui le fût davantage; & si claire que je ne crois pas, qu'il s'y trouve rien, qu'il faille lire plus d'une fois, pour l'entendre, quoiqu'il y ait dans l'original quelques endroits, qui ne sont pas tout-à-fait intelligibles.

Ce qui doit attirer notre attention c'est la dernière remarque: «quoiqu'il y ait dans l'original quelques endroits, qui ne sont pas tout-à-fait intel-

ligibles». Amelot entend rendre le texte plus intelligible qu'il ne l'est et pour ce faire il modifiera si cela lui semble nécessaire «le fil du discours». Cette dernière expression est utilisée dans une note marginale du chapitre xv. Amelot décide de ne pas mettre dans le corps du texte les réflexions de Machiavel sur la différence de sens entre *avaro* et *misero* et il justifie ce choix en note:

Machiavel use du mot, *Misero*, qui est un mot florentin, parceque, dit il, *avaro* en notre langue signifie aussi un homme, qui s'enrichit de rapines, au lieu que nous appelons *Misero*, celui qui épargne trop le sien. Parentese qui rompt le fil du discours, & que pour cela j'ai trouvé mieux de métre à la marge.

Cette décision de ne pas «rompre le fil du discours» est au fondement de sa façon de traduire; la volonté de rendre le texte «intelligible» l'amène ainsi très souvent à simplifier le texte ou à modifier l'ordre et le rythme de l'argumentation; cela a aussi des effets sur ses choix lexicaux, comme nous le verrons dans un second temps.

### 3.2.1 Simplification du texte

VII 37 (Blado) Il che se li fusse riuscito, che gli riusciva l'anno medesimo che Alessandro morì, s' acquistava tante forze et tanta reputatione che per se stesso si sarebbe retto, **senza dependere da la Fortuna o forza d'altri, ma solo da la potentia et virtù sua.**

(Amelot) Et si cela eût réüssi, comme il fut arrivé sans doute l'Année même qu'Aléxandre mourut, il devenoit si puissant & si acrédité, qu'il eût pû se soutenir lui même, **sans dépendre nullement d'autrui.**

XI 1 (Blado) Restaci solamente al' presente a' ragionare de' Principati Ecclesiastici, **circa quali tutte le difficoltà sono avanti che si posseghino, perche s'acquistano o per Virtù o per Fortuna, et senza l'una et l'altra si mantengono...**

(Amelot) Il ne me reste plus à parler, que des Principautés Eclésiastiques, **qui sont difficiles à aquerir, mais faciles a conserver...**

On sent la volonté d'aller «à l'essentiel», de laisser de côté ce qu'on pourrait considérer comme des répétitions ou des explications superflues. Cette volonté de réduction du texte est visible massivement dans la longueur de chaque segment (en vue synoptique) et du texte pris globalement: le *Prince* d'Amelot est légèrement plus court que *Il Principe* de Blado alors que tous les autres traducteurs sont nettement plus longs (et que statistiquement une traduction en français est toujours plus longue que le texte italien original).

### 3.2.2 Modification de l'ordre et du rythme de l'argumentation

VII 47-49 (Blado) Per tanto il Duca innanzi ad ogni cosa doveva crear' Papa uno Spagnuolo, et, non potendo, doveva consentire che fusse Roano et non San' Pietro ad vincula. Et chi crede che nè personaggi grandi i beneficij nuovi faccino dimenticare l'ingiurie vecchie s'inganna. **Errò adunque il Duca in questa elettione**, et fù cagion' de l'ultima rovina sua.

(Amelot) Ainsi **la prudence vouloit**, qu'il essayât premièrement de faire élire un Espagnol, &, ne le pouvant pas, qu'il acceptât le Cardinal de Rouen, & non Saint-Pierre-aux-Liens, **qui fut cause de sa ruine**. Tant se trompent ceux, qui croient, que les bienfaits nouveaux font oublier aux Grans les anciennes offenses.

On note la reconstruction de l'argumentation de Machiavel: le passage ne finit plus avec l'énoncé de la cause de la ruine de César Borgia (il a fait une erreur lors de l'élection du pape!), mais sur une maxime qui s'énonce comme une morale de l'histoire; la «prudence» qu'Amelot met désormais au début du passage, et dont Machiavel ne fait pas mention, aurait consisté au fond à connaître cette maxime. On voit comment la reconstruction d'Amelot fait disparaître l'historicisation radicale de l'analyse machiavéenne.

### 3.2.3 Effets sur les choix lexicaux

La première remarque sur les usages lexicaux d'Amelot porte sur le grand nombre des cas où Amelot ne traduit pas un mot italien: c'est le taux de non-traduction le plus élevé de tous les traducteurs considérés; cela apparaîtra 'à l'œil nu', dans la représentation graphique des trois exemples que nous allons prendre, *ordine*, *virtù* et *stato* (Fig. 12-13-14). La deuxième remarque c'est que – à l'exception notable de *stato* dont nous rendrons compte – Amelot tend à ne pas respecter la polysémie des mots utilisés par Machiavel. Un mot italien est traduit par plusieurs mots français, selon le sens qui paraît convenir à Amelot dans tel ou tel contexte pour que le texte soit 'intelligible' par un lecteur français.

*Ordine* est un terme important du lexique machiavélien; il a des sens multiples qui vont de l'ordre dans lequel on met en troupe aux ordres qu'on lui donne et aux formes d'organisations que l'on met en place (les traducteurs modernes traduisent souvent par 'institutions', Machiavel 2000: 573-578). Chez Machiavel, il n'y a donc pas solution de continuité entre ces différents usages du terme *ordine* et on sent toujours, sous son emploi, l'idée récurrente d'une mise en place matérielle, d'une façon de se placer ou de placer les hommes et les choses afin d'agir. Amelot comprend fort bien les sens divers d'*ordine*: les quatre mots qu'il utilise le plus souvent ('loix', 'usages', 'ordres' et 'coutumes') dessinent la complexité du

terme. Il refuse la polysémie du terme machiavélien et tente de trouver le mot français le plus approprié au contexte et chaque fois qu'il ne trouve pas, il ne traduit pas.

Figure 12 – Graphe «Ordine» Amelot.

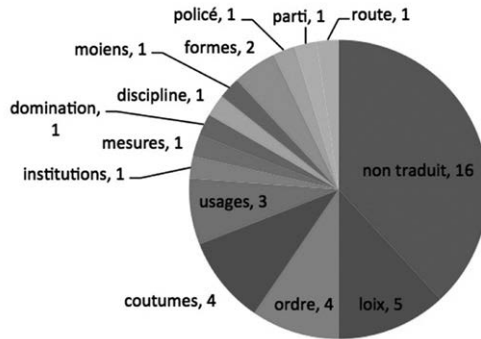
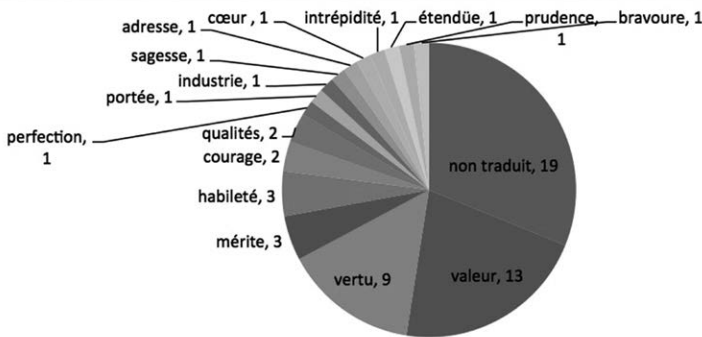
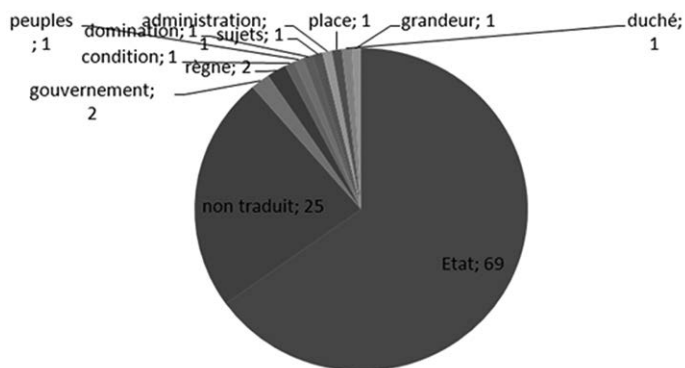


Figure 13 – Graphe «Virtù» Amelot.



*Virtù* est un des grands mots machiavéliens, souvent utilisé en paire avec *fortuna*. Le chapitre xxv du *Prince* tout entier vise à élucider les rapports de la *virtù* comme capacité d'agir en acteur politique et de la *fortuna*. Les traducteurs du XVI<sup>e</sup> siècle, même Gaspard d'Auvergne qui est pourtant celui qui tend le plus vers l'éclatement du vocabulaire, utilisent massivement 'vertu' pour traduire *virtù*. Ce n'est plus le cas dans la traduction d'Amelot: quand il traduit le mot, il utilise 'valeur' et une série d'autres qualités ('bravoure', 'industrie', 'courage', 'habileté', etc.) et n'emploie 'vertu' que lorsque *virtù* semble désigner une qualité morale. Pour Amelot, le lien avec le sens large de la *virtus* latine (qui existe dans la *virtù* machiavélienne et que sentent encore les traducteurs du XVI<sup>e</sup> siècle) s'est brisé (il faudra attendre la période révolutionnaire pour que 'vertu' retrouve ce sens large).

Figure 14 – Graphe «*Stato*» Amelot.

Quant à la traduction de *stato*, sur le sens duquel il est inutile de revenir puisque nous l'avons élucidé plus haut, c'est le seul cas de terme machiavélien pour lequel Amelot, contrairement à son habitude lexicale, ne traduit pas par de nombreux mots différents suivant les contextes. Il tend en effet à faire de *stato* un concept, l'État avec une majuscule; il traduit d'ailleurs aussi par 'État' d'autres mots du lexique machiavélien (*principato*, *regno*, *dominio*). Quand *stato* lui semble ne pas avoir cette valeur conceptuelle, Amelot ne traduit pas. À partir de la traduction d'Amelot (et pratiquement jusqu'à nos jours) Machiavel devient un 'théoricien de l'État' et le *Prince* un traité politique sur l'État.

#### 4. Conclusion

Certains grands textes, traduits et retraduits, constituent une part non négligeable du patrimoine intellectuel de l'Europe. L'histoire des traductions et la reconstruction du voyage des textes sont des outils pour comprendre les modalités de la constitution de ce patrimoine textuel commun. Catalogues des traductions, reconstitution des biographies des traducteurs et des éditeurs, mise en évidence des réseaux internationaux des libraires, imprimeurs et éditeurs ou bien encore analyse des pièces liminaires des traductions sont des méthodes éprouvées et efficaces pour effectuer ce travail nécessaire. L'outil HM ne prétend pas s'y substituer mais permet d'enrichir l'approche des questions de traduction. Il va au-delà des intentions explicites des traducteurs quand ils décident de traduire un texte, il n'entend pas donner les recettes de la bonne voire de la parfaite traduction mais il permet de décrire avec précision les actes réellement effectués par les traducteurs, de mettre en évidence ce que fait concrètement le traducteur vis-à-vis du lexique ou bien de la façon d'écrire de l'auteur qu'il traduit. Ce faisant, il permet également de s'interroger sur l'état des deux langues, celle de l'original et celle vers la-

quelle on traduit. Nos interrogations de départ portaient, avec le choix du *Prince*, sur la langue politique mais l'outil HM est utilisable pour tous les textes sur lesquels se sont accumulées les traductions. Notre outil est libre et gratuit: nous le concevons comme un 'petit cadeau' à l'attention de ceux et celles qui estiment que l'histoire des traductions et des façons de traduire est une tâche de notre présent.

## Bibliographie

- Barlow M., Bowker L. 2004, *Bilingual concordancers and translation memories: A comparative evaluation*, in Yuste E. (ed.), *Coling 2004 Second International Workshop on Language Resources for Translation Work, Research and Training*, Coling, Genève: 52-61.
- Barthas J. 2014, *Amelot de la Houssaye, Abraham-Nicolas*, in Sasso G., Inglese G. (dir.), *Enciclopedia machiavelliana*, Treccani, Roma, vol. I: 49-51.
- D'Alverny M.-T. 1989, *Les traductions à deux interprètes: d'arabe en langue vernaculaire et de la langue vernaculaire en latin*, in G. Contamine (ed.), *Traduction et traducteurs au Moyen Âge: Actes du colloque international du Centre National de la Recherche Scientifique*, CNRS, Paris: 193-206.
- Guillot C., Lavrentiev A., Rainsford T., Marchello-Nizia C., Heiden S. 2014, *La «philologie numérique»: tentative de définition d'un nouvel objet éditorial*, in Buchi É., Chauveau J.-P., Pierrel J.-M. (dir.), *Actes du XXVII<sup>e</sup> Congrès international de linguistique et de philologie romanes*, 3 vol., Société de linguistique romane/ÉLiPhi, Strasbourg, non paginé.
- Fleury S., Zimina M. 2008, *Utilisations de mkAlign pour la traduction philologique*, in *Actes JADT 2008. Journées Internationales d'Analyse statistique des Données Textuelles*, Lyon.
- Harris B. 1998, *Are you bi-textual?*, «Language Technology», 7: 41-45.
- Lay M.-H., Pincemin B. 2010, *Pour une exploration humaniste des textes: AnaLog*, in *Proceedings of JADT'2010*, Rome: 1045-1056.
- Machiavel N. 2000, *De Principatibus. Le Prince*, texte italien de G. Inglese, traduction et appareil critique de J.-C. Zancarini et J.-L. Fournel, PUF, coll. «Fondements de la politique», Paris.
- Melamed I. D. 1998, *Empirical methods for exploiting parallel texts*, in *Dissertations available from ProQuest, Paper AAI9829948*, <<http://repository.upenn.edu/dissertations/AAI9829948>> (06/2017).
- Pincemin B., Heiden S., Lay M.-H., Leblanc J.-M., Viprey J.-M. 2010, *Fonctionnalités textométriques: proposition de typologie selon un point de vue utilisateur*, in *Proceedings of JADT'2010*, Rome: 341-353.
- Schreibman S., Hanlon A., Daugherty S., Ross T. 2007, *The Versioning Machine v3.1: A Tool for Displaying and Comparing Different Versions of Literary Texts*, in *Digital Humanities 2007, Conference abstracts*, Urbana, IL: 195-196.
- TEI Consortium 2008, *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 1.1.0. <<http://www.tei-c.org/Guidelines/P5>> (06/2017).

- Véronis J. 2000, *From the Rosetta stone to the information society. A survey of parallel text processing*, in Véronis J. (ed.), *Parallel text processing: alignment and use of translation corpora*, Kluwer Academic Publishers, Dordrecht: 1-24.
- Zancarini J.-C. 2015, «*Et Favellar francese non gli spiace*». *Sulle traduzioni francesi del Principe, XVI-XVII secolo*, in Anselmi G. M., Caporali R., Galli G. (eds.), *Machiavelli Cinquecento, Mezzo millennio del Principe*, Mimesis, Milano-Udine: 73-90.





## L'integrazione di corpora paralleli di traduzione alla descrizione lessicografica della lingua dell'arte: l'esempio delle traduzioni francesi delle *Vite* di Vasari

*Abstract:* Questo contributo illustra in quale misura i dati estrapolati dal corpus parallelo di traduzione del progetto LBC integrano le informazioni contenute nelle principali risorse lessicografiche e terminologiche per quanto concerne la lingua dell'arte. Dopo aver presentato le caratteristiche del progetto e, più dettagliatamente, il corpus italo-francese, l'analisi verte su due termini dell'ambito della pittura e della scultura, 'tondo' e 'figura', attestati nel sotto-corpus della *Vita di Michelangelo* di Giorgio Vasari. L'obiettivo è dimostrare che la consultazione di un corpus parallelo costruito *ad hoc* offre un patrimonio linguistico estremamente ricco che non può che essere posseduto da numerosi traduttori specializzati e specialisti della lingua dell'arte, e che né un traduttore generalista né un dizionario o un'enciclopedia può detenere. Corpora e dizionari, interrogati in maniera appropriata, sono dunque complementari, poiché permettono al traduttore di formarsi un'idea completa o, in ogni caso, molto ampia delle molteplici opzioni di traduzione, a partire da quelle più aderenti al testo di partenza a quelle più creative o aperte alla diversità di realizzazione dei fenomeni discorsivi.

*Parole chiave:* lessico, arte, corpus, traduzione, dizionario.

*Résumé:* Dans cette contribution nous investiguons dans quelle mesure le corpus parallèle du projet LBC complète les informations fournies par les principales ressources lexicographiques et terminologiques en ce qui concerne la traduction de la langue de l'art. Après avoir illustré les principales caractéristiques de ce projet et présenté, plus en profondeur, le corpus parallèle italien-français, nous nous arrêtons sur l'analyse de deux termes appartenant aux domaines de la peinture et de la sculpture, *tondo* et *figura*, qui sont attestés dans le sous-corpus de la *Vie de Michel-Ange* de Giorgio Vasari. Le but est de montrer que la consultation d'un corpus parallèle ciblé offre la richesse d'un patrimoine linguistique qui n'est possédée que par

plusieurs traducteurs spécialisés et spécialistes de la langue de l'art et que ni un traducteur généraliste ni un dictionnaire ou une encyclopédie ne peut détenir. Corpus et dictionnaires, interrogés de manière appropriée, sont ainsi complémentaires car ils permettent au traducteur de se former une idée complète ou, de toute façon, très large des multiples options de traduction, à partir de celles qui sont plus adhérentes au texte de départ à celles plus créatives ou ouvertes à la diversité de réalisation des phénomènes discursifs.

*Mots-clés*: lexique, art, corpus, traduction, dictionnaire.

## I. Introduzione

Nel corso degli ultimi trent'anni l'integrazione di strumenti tecnologici nel campo delle scienze umane ha permesso di coniugare la conoscenza linguistico-culturale (il sapere tradizionale) e la competenza tecnico-professionale (tecnologica e applicativa o *know-how*). L'avvento dell'informatica umanistica (*humanités numériques* in francese, *digital Humanities* in inglese) ha significato una grande rottura epistemologica che ha provocato un profondo cambiamento delle strutture delle discipline coinvolte, comprese quelle della linguistica (Brossaud, Reber 2007: 47). In particolare, la comparsa di corpus testuali in formato elettronico che, secondo Schnapp e Presner (2009), ha caratterizzato soprattutto la prima fase delle *digital humanities*, più quantitativa, incentrata maggiormente sulla raccolta di dati in database di grandi dimensioni, ha consentito ai ricercatori di lavorare su dati prima inaccessibili che hanno aperto nuove prospettive di studio (cfr. tra tanti Habert, Nazarenko, Salem 1997; Granger *et al.* 2003; Williams 2005).

Anche la ricerca in traduzione ha tratto vantaggio dagli orientamenti più recenti nel campo della linguistica dei corpora (Aston 1999; Baker 1999; Beeby *et al.* 2009; Bernardini, Castagnoli 2008; Zanettin *et al.* 2003), in particolare attraverso la creazione ed esplorazione di corpora testuali paralleli. Per corpora paralleli si intendono corpus di traduzione allineati, formati da testi in lingua originale (SL, *source language*) e dalle loro traduzioni (TL, *target language*). Come osservato da Chiari (2007: 53), «i corpora paralleli sono di utilità eccezionale per la predisposizione di strumenti di traduzione e per la didattica delle lingue straniere», in quanto permettono di consultare i termini e le loro traduzioni in contesto e di stabilire una serie di corrispondenze tra le porzioni di testo tradotte frase per frase. Grazie a questi tipi di corpora, i ricercatori lavorano dunque oggi non sulla «traduzione» in generale ma su «fatti traduttivi attestati», hanno cioè la possibilità di osservare un oggetto di studio esterno e delimitato, composto da più traduzioni pubblicate a cui viene riconosciuta una legittimità, data dalla ricezione (da parte del pubblico) o dall'emissione (le istituzioni o case editrici che le diffondono).

L'attenzione è quindi posta oggi sulla comparazione di elementi linguistici tra due (o più) lingue messe a confronto e, in particolare sui fenomeni

discorsivi (Biber *et al.* 2008), cioè sulle diverse modalità di tradurre una parola o un'associazione in funzione del contesto o dell'ambiente lessicale circostante, in ottica sincronica, o delle epoche, in ottica diacronica. Si tratta quindi di un approccio descrittivo, analitico e pragmatico dei fenomeni traduttivi fondato sull'osservazione aperta del discorso nella diversità delle sue realizzazioni (Kübler, Bordet, Pecman 2011: 582) a partire da corpus autentici e convalidati da professionisti.

In questo contributo ci soffermeremo sul corpus parallelo del progetto di ricerca *Lessico dei Beni Culturali* (LBC), progetto nato nel 2013 per iniziativa dell'Unità di ricerca LBC del Dipartimento di Lingue, Letterature e Studi interculturali dell'Università di Firenze e condotto in collaborazione con diverse Università italiane e straniere, il cui obiettivo principale è la realizzazione di un dizionario plurilingue in formato elettronico del lessico dei beni culturali (cfr. Farina 2016)<sup>1</sup>. Cercheremo di mostrare in quale misura questo corpus parallelo completa le informazioni fornite dalle risorse lessicografiche e terminologiche disponibili per quanto concerne la traduzione della lingua dell'arte. Dopo aver illustrato le principali caratteristiche del corpus parallelo del progetto LBC e presentato, nello specifico, il corpus parallelo italiano-francese, forniremo due esempi ('tondo' e 'figura') di termini appartenenti all'ambito della pittura e della scultura attestati nel sotto-corpus della *Vita di Michelagnolo Buonarroti*. Mostreremo in seguito in che modo il dizionario del progetto LBC potrà colmare alcune lacune importanti relative alla descrizione dei termini e delle espressioni della lingua dell'arte e alla loro traduzione in diverse lingue.

## 2. Il corpus parallelo del progetto LBC

Le *Vite de' più eccellenti pittori, scultori e architettori* di Giorgio Vasari, note nel mondo intero per la descrizione del patrimonio artistico fiorentino, sono state scelte come testo di riferimento nell'ambito del progetto LBC. Si è ritenuto che quest'opera possieda le caratteristiche adeguate per lo studio della lingua dell'arte, in quanto «primo e più completo tentativo di decodificare la vita artistica fiorentina e italiana del periodo rinascimentale» (Garzaniti, Farina 2013: 84). Inoltre, l'indiscutibile statuto del suo autore<sup>2</sup>, Giorgio Vasari, riconosciuto come massimo esperto d'arte del Rinascimento italiano, è garanzia della pertinenza del corpus e della sua capacità di rappresentare questa disciplina.

Di quest'opera sono state effettuate numerose traduzioni in diverse lingue straniere e in diverse epoche. Questo fattore non secondario rende

<sup>1</sup> Per informazioni dettagliate sul progetto LBC si rimanda al sito web del progetto: <<http://www.lessicobeniculturali.net/contenuti/il-progetto/818>> (10/2016).

<sup>2</sup> Secondo Pearson (1998: 60) lo statuto dell'autore è uno dei criteri da valutare per inserire determinati testi all'interno di un corpus.

possibile condurre, all'interno del corpus di riferimento, sia analisi in ottica contrastiva tra diverse coppie di lingue, con l'italiano che funge sempre da lingua pivot, che valutazioni in prospettiva diacronica sui diversi modi di tradurre i termini artistici nel corso del tempo. Vedremo infatti in seguito, attraverso alcuni esempi, che, con l'evolversi della disciplina, alcuni termini nascono e si radicano nella terminologia artistica solo in determinate epoche per poi eventualmente essere soppiantati da altri (es. 'tondo'). Inoltre, l'esame delle traduzioni selezionate, effettuate nella maggior parte dei casi da specialisti, traduttori professionisti o studiosi dell'arte, rivela che alcune di esse possono essere prese a modello per la traduzione dei termini dell'arte, poiché forniscono soluzioni più adeguate rispetto a quelle presenti nelle principali risorse lessicografiche consultate abitualmente dai traduttori.

Precisiamo che il corpus parallelo Vasari è tuttavia rappresentativo di un aspetto soltanto della lingua dell'arte, la lingua parlata dal Vasari nel '500, in cui convivono «le due anime delle lingue delle arti, umanistica e tecnica», in quanto Vasari «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» (*Enciclopedia Treccani*<sup>3</sup>). Le diverse traduzioni delle *Vite* nel corso dei secoli attestano dunque l'evoluzione della lingua dell'arte in uso esclusivamente nel Rinascimento italiano. Per conciliare la classica descrizione del patrimonio con la sua lettura contemporanea, nell'ambito del progetto LBC, il corpus Vasari sarà messo a confronto e integrato con la descrizione della lingua dell'arte fornita da un altro corpus parallelo contemporaneo, il corpus plurilingue delle guide turistiche di Firenze, su cui non ci soffermeremo in questo contributo.

## 2.1 HyperMachiavel

Per la costituzione del corpus parallelo Vasari, l'équipe LBC ha beneficiato del lavoro svolto dal laboratorio di ricerca *Triangle* dell'École Normale Supérieure di Lione, che ha realizzato, sotto la direzione scientifica del professor Jean-Claude Zancarini e la guida tecnica dell'ingegnere informatico Séverine Gedzelman, il software *HyperMachiavel*, uno strumento di allineamento e confronto di traduzioni, nonché di annotazione lessicale<sup>4</sup>. Nel contesto dell'edizione critica elettronica e dei corpus digitali, questo software, concepito originariamente per l'analisi de *Il Principe* di Machiavelli, propone varie funzionalità per la costituzione di un corpus di testi paralleli (es. di un dato testo e delle sue traduzioni, o di diverse edizioni di uno stesso testo) e permette l'esplorazione lessicale e concettuale attraverso diverse funzioni

<sup>3</sup> Voce 'arte e critica d'arte': <[http://www.treccani.it/enciclopedia/arte-e-critica-d-arte-lingua-dell\\_%28Enciclopedia-dell%27Italiano%29/](http://www.treccani.it/enciclopedia/arte-e-critica-d-arte-lingua-dell_%28Enciclopedia-dell%27Italiano%29/)> (10/2016).

<sup>4</sup> Cfr. il contributo di Zancarini e Gedzelman in questo volume.

di ricerca e di annotazione semi-automatica, nel tentativo di rispondere in questo modo alle esigenze degli studiosi di traduzione e di filologia testuale.

## 2.2 HyperVasari

Abbiamo chiamato *HyperVasari* l'adeguamento e la customizzazione del software *HyperMachiavel* alle necessità del progetto LBC. *HyperVasari* rende possibile confrontare automaticamente le due edizioni italiane delle *Vite* di Giorgio Vasari e le loro traduzioni nelle diverse lingue del progetto (Fig. 1).

Figura 1 – I contesti del termine 'tondo' e le loro traduzioni francesi in *HyperVasari*.

The screenshot shows the HyperMachiavel software interface. At the top, there are tabs for different editions: 1550\_Torrentiniana, 1568\_Giuntina, 1839\_Leclanché, 1981\_Chastel, 2002\_Luciani, and 2007\_Grasset. The main window displays parallel text in Italian and French. Below the text, there is a search results table.

Localisation	Source [1550_	Source [1568_	Cible [1839_Le	Cible [1981_Ch	Cible [2002_Lu	Cible [2007_Gr	A relier
Texte - Oeuvre - ...	tondezza	tondezza					<input type="checkbox"/>
Texte - Oeuvre - ...	tonde	tonde					<input type="checkbox"/>
Texte - Oeuvre - ...	tonda	tonda					<input type="checkbox"/>
Texte - Oeuvre - ...	tondo	tondo					<input type="checkbox"/>
Texte - Oeuvre - ...	tonde	tonde		tondi			<input type="checkbox"/>
Texte - Oeuvre - ...	tonde	tonde					<input type="checkbox"/>
Texte - Oeuvre - ...	tonde	tonde					<input type="checkbox"/>
Texte - Oeuvre - ...	tondo	tondo		tondo			<input type="checkbox"/>
Texte - Oeuvre - ...			rotonde				<input type="checkbox"/>

I corpus paralleli di traduzione di Vasari nelle diverse lingue del progetto LBC sono attualmente in corso di realizzazione. Illusteremo dettagliatamente nel prossimo paragrafo il corpus delle traduzioni francesi, ad oggi il corpus parallelo allo stadio più avanzato di raccolta e analisi dei dati.

## 3. Il corpus italiano-francese

Il corpus italiano-francese è composto dalle due edizioni originali delle *Vite* di Vasari in lingua italiana, la prima, l'edizione Torrentiniana del 1550, e la seconda, l'edizione aumentata del 1568, nota come edizione Giuntina, e da cinque traduzioni in lingua francese pubblicate in periodi differenti, di cui solo quattro sono state inserite attualmente nel software *HyperVasari* (Leclanché 1839, Chastel 1981, Luciani 2002, Gerard Powell 2007). Di seguito presentiamo brevemente le peculiarità di ognuna delle traduzioni attualmente presenti nel corpus italo-francese.

### 3.1 Leclanché

La prima traduzione, datata 1839, è stata effettuata da Léopold Leclanché, traduttore di Giorgio Vasari e di Benvenuto Cellini, ed è commentata in alcune parti da Philippe-Auguste Jeanron. Nel proemio alla traduzione, gli stessi Leclanché et Jeanron (1839: 1) spiegano di essersi dedicati a tale impresa per rendere fruibile al pubblico francese un testo così celebre:

Le livre du Vasari, à force d'avoir été recommandé à l'attention par l'éloge ou le blâme, à force d'avoir été compilé, cité, abrégé, délayé, traité enfin de mille manières, est devenu célèbre en France, sans jamais y avoir été traduit. Nous venons combler cette lacune et en cela nous croyons faire une œuvre utile.

Questa traduzione, che presenta una quasi perfetta corrispondenza con l'edizione Giuntina, ha posto le basi delle future traduzioni vasariane. I due traduttori dovettero tenere conto delle traduzioni già esistenti in Inghilterra e Germania, dove «le texte original du Vasari et sa version ne marchent guère qu'escortés d'une masse compacte de commentaires, d'interprétations ou d'additions» (Leclanché, Jeanron 1839: 19-20), e cercarono di riprodurre lo stesso lavoro per l'edizione francese. Si tratta di una traduzione commentata e annotata di cui Leclanché e Jeanron (1839: 26) affermano: «Voilà dans quel sens et pourquoi nous nous sommes permis d'ajouter aux vies des artistes les plus célèbres nos annotations ou commentaires». La loro traduzione si presenta dunque come una traduzione analitica, che tiene conto sia della tradizione di studi precedenti dell'opera che degli errori e delle lacune commesse da Vasari. Risulta molto esaustiva in alcune parti e tuttavia approssimativa in altre, come l'analisi di alcuni esempi rivela (cfr. paragrafi 4 e 5 di questo contributo).

Nel 1903 viene pubblicata dall'editore Foulard la traduzione rivista dell'edizione di Leclanché, per mano di Charles Weiss, che verrà criticata aspramente in seguito. Si tratta della traduzione attualmente assente in *HyperVasari*.

### 3.2 Chastel

La seconda traduzione presente nel corpus parallelo italo-francese è stata realizzata nel 1981 da André Chastel, celebre storico dell'arte in Francia,

l'une des figures les plus marquantes de l'histoire de l'art dans la seconde moitié du XX<sup>e</sup> siècle, tant par le rayonnement international de ses très nombreuses publications, [...] que par l'emprise de son activité féconde sur les développements de la discipline en France (Brunon, en ligne).

Chastel (1964: 288) ha osservato riguardo alle traduzioni francesi esistenti:

En France, [...] la traduction de Leclanché (1839-1842), est déplorable; celle de Charles Weiss en deux volumes (1903, réédition 1913) est partielle, arbitrairement tronquée, semée de fautes grossières et dépourvue de tout commentaire utile. Le champ est donc libre pour un travail d'ensemble.

Chastel è quindi l'autore di un lavoro monumentale volto a dare una giusta interpretazione e revisione dell'opera nel tentativo di ridarle vita nella lingua e tradizione francese. La sua edizione si presenta in effetti più completa e fedele allo spirito del Vasari. Nella *Revue de l'art* (1990: 6), si legge a proposito della traduzione da lui diretta:

Grâce à lui Vasari ira rejoindre les grands classiques de la littérature universelle dans la Bibliothèque de la Pléiade plaçant ainsi l'histoire de l'art au niveau suprême qu'il avait souhaité toute sa vie.

La traduzione di Chastel è in effetti una traduzione dotta, redatta da un fine conoscitore dell'arte e della cultura italiane, come è dimostrato dall'utilizzo di numerosi termini dell'arte che veicolano concetti importanti di questa disciplina, termini quali 'tondo', 'ignudi', 'terribilità', mentre in altre edizioni questi termini vengono appiattiti e tradotti in francese in modo approssimativo (vedremo in seguito nel dettaglio l'esempio del termine 'tondo'). La traduzione di Chastel è apparsa parzialmente nel 1966 per i soli pittori toscani presso l'editore Hermann, mentre tra il 1981 e il 1989 ne è stata pubblicata la traduzione integrale presso l'editore Berger-Levrault<sup>5</sup>. Anche questa traduzione presenta una quasi perfetta corrispondenza con l'edizione Giuntina del 1568.

### 3.3 Luciani

Altre due edizioni sono apparse nel corso del XX secolo, molto diverse l'una dall'altra e ognuna con un'impronta propria. La terza traduzione presente nel corpus è stata curata da Gérard Luciani, professore emerito all'Università Stendhal di Grenoble specialista del XVIII secolo italiano, è stata editata da Gallimard nel 2002 ed è l'unica traduzione basata integralmente sull'edizione Torrentiniana di Vasari del 1550. Si tratta di un'edizione bilingue con testo a fronte, notevolmente ridotta e rimaneggiata rispetto alle edizioni italiane, che presenta poche corrispondenze con le edizioni italiane originali. Vi sono state selezionate le vite più importanti e, all'interno di queste, sono stati omessi numerosi passaggi poiché sono state prese in considerazione soltanto le opere più significative di ogni artista (ad es. è assente tutta la descrizione della Cupola di San Pietro in Vaticano progettata da

<sup>5</sup> La versione integrale è stata rieditata nel 2005 da Actes Sud in due volumi. Si tratta dell'edizione utilizzata per la costituzione di questo corpus.



Michelangelo). Più che di una compiuta traduzione dell'opera del Vasari, si tratta dunque di una rielaborazione a scopo divulgativo.

### 3.4 Gerard Powell

La quarta traduzione è apparentemente un'esaustiva riedizione delle traduzioni di Leclanché (1839) e di Weiss (1903), effettuata per l'editore Grasset nel 2007 da Véronique Gerard Powell, docente di storia dell'arte presso l'Università di Parigi IV Sorbona. Questa edizione è molto approfondita e, nonostante non costituisca una traduzione integrale come quella di Chastel, viene molto studiata oggi. Da un'analisi accurata della *Vita* di Michelangelo, si evince che alcune parti relative all'architettura sono assenti (anche qui tutta la sezione dedicata alla cupola di San Pietro è stata omessa), mentre la corrispondenza con l'edizione italiana è quasi perfetta per le parti riguardanti la pittura. Powell ha quindi riorganizzato in modo più immediato e comprensibile soltanto una selezione delle *Vite*, con un approccio moderno e in una lingua attuale.

## 4. Il sotto-corpus della *Vita di Michelagnolo Buonarroti*

Ad oggi, le *Vite* che sono state oggetto di uno spoglio completo per la costituzione del corpus parallelo di traduzione francese sono: la *Vita di Cimabue pittore fiorentino* (ed. Giuntina 1568: 83-87), la *Vita di Giotto pittore, scultore et architetto fiorentino* (ed. Giuntina 1568: 119-134) ed infine la *Vita di Michelagnolo Buonarroti fiorentino pittore scultore, architetto* (ed. Giuntina 1568: 715-796)<sup>6</sup>. Per testare il corpus ed effettuare le prime analisi di un campione di termini della lingua dell'arte, si è scelto di partire da quest'ultima, la *Vita di Michelagnolo Buonarroti*, a causa di diverse ragioni. Si tratta della vita più lunga, con la trattazione più ampia e tra le più studiate al mondo. Inoltre, data la nota produttività di Michelangelo, è rappresentativa di tre diversi ambiti artistici in cui egli è stato maestro: pittura, scultura e architettura. Questo sotto-corpus, nonostante sia di piccole dimensioni nella prospettiva della ricerca in linguistica dei corpora (è composto complessivamente da circa 120.000 parole), presenta materiale linguistico rilevante per l'analisi della lingua dell'arte. Di seguito, ci soffermeremo su alcuni esempi relativi agli ambiti della scultura e della pittura, in particolare sui termini 'tondo' e 'figura'.

## 5. Analisi di 'tondo' nella *Vita di Michelagnolo Buonarroti*

Il termine risulta già diffuso nel XVI secolo ed è infatti attestato nella *Vita di Michelagnolo Buonarroti* dove viene usato come sostantivo per due

<sup>6</sup> Lo spoglio della *Vita di Sandro Botticello pittor fiorentino* è attualmente in corso.

volte. Come osservato da Grassi-Pepe (1994: 999), Vasari ha distinto opportunamente tra 'tondo di marmo' e 'tondo di pittura'. Precisando l'ambito di applicazione ('marmo', per scultura, e 'pittura'), Vasari è riuscito in questo modo ad evitare ambiguità interpretative e ad essere estremamente preciso:

Et ancora in questo tempo abbozzò e non finì due **tondi di marmo**, uno a Taddeo Taddei, oggi in casa sua, et a Bartolomeo Pitti ne cominciò uno altro [...]. (Vasari 1568: 21)

Venne volontà ad Agnolo Doni, cittadino fiorentino amico suo... d'aver alcuna cosa di Michelagnolo; per che gli cominciò un **tondo di pittura**, dentrovi una Nostra Donna, la quale inginocchiata con amendua le gambe, ha in sulle braccia un putto e porgelo a Giuseppo, che lo riceve [...] (Vasari 1568: 22)

Il termine 'tondo' designa due referenti distinti nella lingua dell'arte italiana: un quadro di forma circolare, nel sotto-ambito della pittura, e un basso-rilievo di forma rotonda, in quello della scultura. Si tratta dunque di un caso di omonimia in terminografia (Cabr  1993: 218) e, per questo, sar  necessario condurre la loro analisi distintamente.

## 5.1 La consultazione del corpus parallelo

L'analisi della prima citazione, all'interno del corpus parallelo delle traduzioni francesi delle *Vite*, rivela scelte traduttive differenti.

Tabella 1 – Traduzioni di 'tondo di marmo' nel corpus parallelo Michelangelo.

Vasari	1550	Et ancora in questo tempo abbozzò e non finì due <b>tondi di marmo</b> , uno a Taddeo Taddei, oggi in casa sua, et a Bartolomeo Pitti ne cominciò uno altro,
Vasari	1568	Et ancora in questo tempo abbozzò e non finì due <b>tondi di marmo</b> , uno a Taddeo Taddei, oggi in casa sua, et a Bartolomeo Pitti ne cominciò un altro,
Leclanché	1839	Dans le m�me temps, il �baucha deux <b>bas-reliefs circulaires en marbre</b> ; l'un, pour Taddeo Taddei, chez lequel on le voit aujourd'hui; et l'autre, destin� � Bartolommeo Pitti,
Chastel	1981	il �baucha sans les terminer deux <b>tondi de marbre</b> , dont un pour Taddeo Taddei, qui se trouve toujours chez lui et l'autre pour Bartolomeo Pitti,
Luciani	2002	X
Gerard Powell	2007	A cette �poque, il �baucha, sans les terminer, deux <b>m�dailleurs de marbre</b> : l'un pour Taddeo Taddei, actuellement dans la maison de celui-ci, l'autre pour Bartolommeo Pitti.

In un caso ‘tondi di marmo’ al plurale viene tradotto con *bas-reliefs circulaires* (Leclanché 1839); in un altro caso è stato impiegato il prestito italiano *tondi* (Chastel 1981); nella traduzione di Luciani (2002), tutta la citazione è stata omessa; infine, nella traduzione più recente rivista da Gerard Powell (2007), ‘tondi di marmo’ è stato tradotto con *médailles*.

Per quanto riguarda la seconda citazione in cui è presente l’unica occorrenza di ‘tondo di pittura’ di tutto il sotto-corpus Michelangelo, che fa riferimento precisamente al *Tondo Doni*, sintagma che non si era ancora affermato in quell’epoca, troviamo i traducenti *tableau circulaire* e *tableau rond*, rispettivamente nelle traduzioni di Leclanché (1839) e di Gerard Powell (2007), e ancora una volta il prestito ‘tondo’ nelle traduzioni di Chastel (1981) e Luciani (2002). Segnaliamo che nell’edizione di Luciani (2002: 223) è presente una nota esplicativa in cui il traduttore precisa che, in questa occorrenza, il termine ‘tondo’ impiegato nella traduzione francese ha il significato di quadro e non di tondo di scultura come in altri passaggi precedenti delle *Vite*: «ici tondo a le sens de tableau, et non plus de médaille comme on l’a vu ailleurs».

Tabella 2 – Traduzioni di ‘tondo di pittura’ nel corpus parallelo Michelangelo.

Vasari	1550	per che gli cominciò un <b>tondo di pittura</b> , ch’è dentrovi una Nostra Donna, la quale, inginocchiata con amendua le gambe, alza in su le braccia un putto e porgelo a Giuseppe, che lo riceve;
Vasari	1568	per che gli cominciò un <b>tondo di pittura</b> , dentrovi una Nostra Donna, la quale, inginocchiata con amendua le gambe, ha in sulle braccia un putto e porgelo a Giuseppe che lo riceve;
Leclanché	1839	Michel-Ange exécute ensuite, pour Agnolo Doni, citoyen de Florence, riche amateur de ses amis, un <b>tableau circulaire</b> , où il peignit la Vierge à genoux, tenant son fils entre ses bras et le présentant à saint Joseph.
Chastel	1981	et celui-ci entreprit la peinture d’un <b>tondo</b> avec la Madone: les deux jambes repliées, elle soulève de ses bras un petit enfant et le tend à Joseph qui le reçoit.
Luciani	2002	Il commença pour lui un <b>tondo</b> peint avec une Madone qui, les deux genoux à terre, a dans les bras un petit enfant et le tend à Joseph qui le reçoit.
Gerard Powell	2007	Michel-Ange commença pour lui un <b>tableau rond</b> dans lequel la Vierge, posée sur les deux genoux, élève l’Enfant Jésus et le tend à Joseph qui le prend.

Riteniamo utile raffrontare questi passaggi con uno della *Vita* di Botticelli<sup>7</sup>, l’altro artista rinascimentale celebre per i suoi tondi di pittura, in cui è presente una occorrenza di ‘tondo’.

<sup>7</sup> Come indicato nella nota 6, questo sotto-corpus non è stato ancora oggetto di uno spoglio e di un’analisi completa.

Tabella 3. Traduzioni di 'tondo di pittura' nel corpus parallelo Botticelli.

Vasari	1550	Per la città in diverse case fece <b>tondi</b> di sua mano, e femmine ignude assai,
Vasari	1568	Per la città in diverse case fece <b>tondi</b> di sua mano, e femmine ignude assai,
Leclanché	1839	et, dans diverses maisons de Florence, des <b>médailleurs</b> et quantité de femmes nues.
Chastel	1981	Chez différents particuliers de la ville, il peignit de sa main de nombreux <b>tondi</b> et des nus féminin;
Luciani	2002	À travers la ville, dans diverses maisons, il fit de sa main de nombreux <b>tondi</b> et des nus de femmes.
Powell	2007	Dans plusieurs maisons de la ville, il fit des <b>tableaux ronds</b> et de nus féminins:

Vasari si riferisce qui indubbiamente non a 'tondi' di scultura, ma a dipinti («il peignit de sa main de nombreux *tondi*», Chastel 1981). Notiamo che Chastel (1981) e Luciani (2002) mantengono in tutta la traduzione delle *Vite* il traduce *tondo* per l'ambito della pittura, mentre Leclanché (1839), che aveva utilizzato *tableau circulaire* nel passaggio relativo al *Tondo Doni* di Michelangelo, utilizza *médailлон* per tradurre i tondi dipinti da Botticelli. Il corpus parallelo fornisce in definitiva ben cinque traduzioni diversi di 'tondo': *bas-relief circulaire*, *tondo*, *médailлон*, *tableau circulaire*, *tableau rond*, in contesti a partire dai quali è possibile evincere l'ambito di applicazione appropriato (scultura o pittura).

## 5.2 La descrizione lessicografica del termine italiano 'tondo'

### 5.2.1 I dizionari bilingui

La consultazione dei principali dizionari bilingui disponibili sul mercato fornisce informazioni generiche e imprecise e, inoltre, spesso contrastanti se viene consultato più di un dizionario alla volta, come si evince dagli articoli riportati sotto:

Ferrante-Cassiani 1991: **tondo** [...] s.m. [...] 3 (*pitt. e scult.*) médailлон || **scultura a tutto tondo**, sculpture en ronde bosse ; **a mezzo tondo**, en demi-bosse f. [...]

Larousse Francese 2006: **tondo** [...] II s.m. [...] 4 (*Pitt, Scult*) médailлон, tondo [...]

Garzanti Francese 2007: **tondo** n.m. [...] 2 (*pitt., scultura*) **médailлон**

Hachette-Paravia 2007: **tondo** [...] II m. [...] 3 (ART.) médailлон; **scultura a tutto** ~ *ronde-bosse* [...]

Boch 2014: **tondo** s.m. [...] 2 (*pitt.*) **tondo** (*tableau de forme circulaire*) [...]

Secondo i dizionari *Ferrante-Cassiani* (1991) e *Garzanti Francese* (2007), il traducete *médaille* sarebbe appropriato sia nell'ambito della pittura che in quello della scultura, come desunto dall'analisi del corpus che riserva questo traducete ai 'tondi di marmo' nel sotto-corpus Michelangelo (Tabella 1, traduzione a cura di Gerard Powell 2007) e ai 'tondi di pittura' nel sotto-corpus Botticelli (Tabella 3, traduzione di Leclanché 1839). Anche il dizionario *Hachette-Paravia* (2007) propone questo unico traducete con una indicazione settoriale generica (ART. per Arte). Lo stesso traducete *médaille* è presente anche nel *Larousse Francese* (2006), ma accompagnato anche dal traducete *tondo* e senza l'aggiunta di una precisazione semantica che permetta di far capire all'utente se questi due traduceti siano intercambiabili e utilizzabili sia nell'ambito della pittura che della scultura. Il dizionario *Boch* (2014) propone invece come traducete soltanto il prestito *tondo*, ed esclusivamente per l'ambito della pittura (indicato dalla marca settoriale *pitt.*), aggiungendo una precisazione semantica («tableau de forme circulaire») o variante di traduzione, che avevamo trovato in forma simile nella traduzione di Leclanché (1839) (Tabella 2). La consultazione di questo dizionario bilingue non fornisce invece alcun aiuto per la traduzione di 'tondi di marmo'. Infine è degna di nota, nel *Ferrante-Cassiani* (1991) e nell'*Hachette-Paravia* (2007), l'inserzione della traduzione di alcuni sintagmi tecnici del dominio dell'arte, «scultura a tutto tondo, ronde-bosse», «a mezzo tondo, en demi-bosse», generalmente poco presenti in queste risorse, come osservato da Farina (2015: 50):

Les dictionnaires bilingues existants laissent une place plus ou moins grande aux vocabulaires spécialisés, mais ne leur consacrent de toute façon que peu de place, puisqu'ils se basent tous sur les nomenclatures de monolingues à nomenclature restreinte. Pour les mots considérés comme techniques, il est très rare de trouver des articles complexes dans les dictionnaires généraux français-italien.

Dato che gli esempi riportati sopra mostrano chiaramente l'incompletezza e la scarsa affidabilità dei dizionari bilingui, per quanto concerne nello specifico la traduzione della lingua dell'arte, si rivela opportuno consultare risorse monolingui italiane generali e specializzate. Queste dovrebbero innanzitutto permettere ai traduttori di chiarire ed approfondire il concetto espresso dal termine artistico 'tondo' in italiano, prima che la corrispondenza semantica e concettuale di questo termine con i suoi supposti traduceti in lingua francese venga verificata all'interno di risorse francofone.

### 5.2.2 I dizionari monolingui

In tutte le risorse monolingui consultate è presente una definizione logica iperspecifica (Pruvost 2006: 172) in cui il primo termine iperonimico è

sempre il termine 'dipinto', seguito dal tratto semantico distintivo «di forma rotonda» o «di forma circolare», ad indicare che nella lingua generale il termine 'tondo' rientra indubbiamente nel campo della pittura. Fanno eccezione il *Garzanti Italiano* (on line), dove viene indicato come termine iperonimo anche il 'bassorilievo', e lo *Zingarelli* (2013), in cui associato a 'tondo' è presente anche l'iperonimo 'decorazione'. La circolarità è il tratto semantico distintivo comune a tutti i referenti di tondo:

*Sabatini Coletti*: **tondo s.m. 1** Cerchio, circonferenza; oggetto di forma circolare: *un t. di cartone, di legno*; in partic. **dipinto** su tela o su tavola di forma circolare o qualsiasi elemento decorativo con le stesse caratteristiche: *i t. di Michelangelo; uno splendido t. in ceramica* || fig.in t., a t., tutt'intorno.

*Garzanti Italiano*: **tondo 3.** nelle arti figurative, **dipinto** o bassorilievo di forma rotonda: *un tondo di terracotta; i toni di Botticelli, di Michelangelo.*

*Zingarelli 2013*: **tondo s.m 2 Dipinto**, decorazione avente forma circolare: *un t. di terracotta; il t. Doni di Michelangelo.*

*Enciclopedia Treccani*: **tondo c. Dipinto** su tela, su tavola o anche affrescato, di forma circolare anziché quadrangolare: *il tondo Doni di Michelangelo; quell'espressione tenue di sofferenza e di stanchezza, che forma l'umano incanto delle Vergini ne' toni fiorentini del tempo di Cosimo* (D'Annunzio); anche, qualsiasi elemento decorativo di forma circolare, usato nell'ornamentazione, cromatica o plastica, di pareti, volte, soffitti, mobili, ecc.

La consultazione di risorse specializzate dell'ambito artistico fornisce informazioni aggiuntive:

Masciotta (1969), *Dizionario di termini artistici*: **tondo** (da *ritondo*, rotondo, e questi dal latino *rotundus*, derivato di *rota*, ruota). **Pittura** o scultura a bassorilievo di forma rotonda. Fu usato specialmente dagli artisti italiani del Rinascimento. Gli si dà il nome italiano anche nelle altre lingue.

La *Nuova enciclopedia dell'arte* Garzanti (2000): **tondo Pittura** o bassorilievo di forma rotonda; ebbe grande fortuna in epoca rinascimentale.

Ferrari (2002), *Dizionario di arte e di architettura*: **tondo Opera pittorica** o scultorea a rilievo di forma rotonda.

Il concetto di 'dipinto' (pittura o opera pittorica) è ancora presente. È inoltre qui menzionato sistematicamente anche l'altro referente del termine, 'basso-rilievo', relativo all'ambito della scultura, che era assente nella maggior parte dei dizionari di lingua generali. Nelle risorse specializzate non figura

invece nessuna menzione del possibile uso del termine 'tondo' ad indicare un «elemento decorativo» nell'ambito delle arti decorative. La consultazione del dizionario di Masciotta (1969) si rivela interessante in ottica traduttiva, perché fornisce alcune indicazioni sulla traduzione nelle principali lingue europee (francese, inglese e tedesco) dei termini presenti nella nomenclatura, come precisato nella prefazione. Secondo Masciotta, il termine italiano 'tondo' viene preso in prestito da queste lingue («gli si dà il nome italiano anche nelle altre lingue») per designare una «pittura o scultura a bassorilievo di forma rotonda». Questo dato induce a verificare l'esattezza di questa prima informazione traduttiva e ad esaminare più approfonditamente il prestito *tondo* in francese, che, ricordiamo, risulta attestato nel corpus parallelo sia nella traduzione di Chastel del 1981 per la pittura e la scultura (Tabella 1, 2 e 3) che nella traduzione di Luciani del 2002 per la pittura (Tabella 2 e 3).

### 5.3 Il traduttore francese *tondo*

La ricerca del prestito *tondo* nelle principali risorse lessicografiche francesi contemporanee (*Petit Robert* 2016, *TLFi*) e in diverse risorse terminologiche e/o specializzate del passato e del presente (Lacombe 1753, Boutard 1826, Adeline 1884, Baudry 2002), nonché nell'*Encyclopédie* (Diderot et D'Alembert dir. 1751-1772), non fornisce alcun risultato. Il prestito italiano *tondo* sembra non essere stato lessicalizzato nella lingua francese. Si potrebbe dunque dubitare del dato fornito da Masciotta (1969). Pertanto si è ritenuto necessario condurre ricerche più approfondite sul prestito *tondo* in altre risorse francesi specializzate con l'obiettivo di sciogliere questo dubbio.

Il *Dictionnaire de la Peinture Larousse* (2003)<sup>8</sup> attribuisce a *tondo* esclusivamente il significato di quadro di forma circolare:

*Dictionnaire de la Peinture Larousse* (2003): **tondo** Tableau de forme circulaire (abréviation de l'italien *rotondo*, 'rond'), particulièrement en vogue en Italie à la Renaissance et notamment à Florence, où, dès l'époque du Gothique international (Maître du Jugement de Pâris du Bargello), il apparaît sous la forme du *desco da parto* (plateau d'accouchée) et devient un format adopté pour représenter la *Vierge et l'Enfant* ou la *Sainte Famille* (Michel-Ange, *Tondo Doni*, Offices).

Nell'*Encyclopaedia Universalis*, Jean Rudel, pittore, scrittore e professore emerito di Storia dell'Arte all'Università di Parigi-Sorbona, dedica a questo prestito un lungo articolo<sup>9</sup> in cui si trova conferma del fatto che il termine *tondo* designa innanzitutto un quadro di forma rotonda («tableau de forme circulaire») diffuso nel XV secolo in Italia. Rudel aggiunge un'informazione nuova che precisa chiaramente le caratteristiche del referente: il *tondo* è gene-

<sup>8</sup> Consultato su <<http://www.larousse.fr/archives/peinture>> (05/2016).

<sup>9</sup> Consultato su <<http://www.universalis.fr/encyclopedie/tondo/>> (05/2016).

ralmente costituito da un pannello di legno circondato da una cornice di dimensioni abbastanza grandi che evoca le ghirlande all'antica che circondano la stessa forma circolare in scultura, motivo per cui lo stesso termine ha come possibile referente anche *encadrement décoratif*, come Rudel spiega, dalla seconda metà del XVI secolo, *tondo* viene impiegato per indicare le cornici ornamentali di forma rotonda che decorano i soffitti dell'epoca. Qualsiasi riferimento all'ambito della scultura in quanto tale è tuttavia assente.

La consultazione delle principali risorse terminologiche francofone si rivela analogamente poco soddisfacente per quanto concerne questo aspetto. *Termium Plus*, la banca dati linguistica e terminologica del governo del Canada, principalmente bilingue francese e inglese, si limita a segnalare l'esistenza del termine ricercato in francese, nel settore dell'Archeologia. Questa informazione non conferma all'utente se *tondo* (fr) possa essere utilizzato anche in pittura e scultura. Tuttavia, l'articolo relativo alla lingua inglese, sempre riferito all'ambito dell'Archeologia, si rivela interessante, perché presenta il traduttore inglese *circular medallion* seguito da un'informazione semantica: «a circular tablet bearing carved or embossed figures» (una tavoletta circolare che racchiude figure incise o in rilievo). Si tratta dell'unico riferimento trovato relativo alla possibilità che *tondo* in francese possa riferirsi anche ad una figura in rilievo o basso-rilievo.

Il *Grand dictionnaire terminologique* (GDT) online<sup>10</sup> fornisce una definizione di *tondo* (fr) che conferma le osservazioni fatte finora: il campo concettuale del termine *tondo* in francese non corrisponde a quello del termine 'tondo' in italiano, da cui è stato preso in prestito. Secondo il GDT, il nome maschile *tondo* è il termine privilegiato per designare un quadro di forma rotonda (definizione: «tableau de forme ronde»), all'interno dell'ambito della pittura e del disegno artistici<sup>11</sup>. Neanche qui l'ambito della scultura appare. Quest'ultima risorsa, così come le altre risorse lessicografiche specializzate, si soffermano esclusivamente sul referente 'quadro', indipendentemente dal tipo di soggetto rappresentato, quindi all'ambito della pittura. Nessuna presenta in definitiva il referente 'basso-rilievo' relativo all'ambito della scultura per il francese (con l'eccezione di *Termium Plus* nella definizione del termine in lingua inglese), diversamente dalle risorse specializzate italiane illustrate sopra.

Tutti questi dati inducono quindi a concludere che, contrariamente a quanto attestato da Masciotta, *tondo* può essere usato in francese in ambito specializzato solo per designare un quadro di forma rotonda nell'ambito della pittura e che quindi esista un altro termine francese per designare il basso-rilievo circolare tipico del Rinascimento. Tornando alle traduzioni rilevate nel corpus, 'tondo di pittura' tradotto con il prestito francese *tondo*

<sup>10</sup> Consultato su <<http://www.larousse.fr/archives/peinture>> (05/2016).

<sup>11</sup> Il GDT presenta inoltre un'osservazione linguistica utile nell'ottica della traduzione relativa alla forma plurale di *tondo*, che segue la formazione del plurale della lingua italiana attraverso l'uso del morfema *-i* (*dei tondi*).



da Chastel (1981) e da Luciani (2002) (Tabella 2) può essere dunque considerata una scelta adeguata. Chastel sembra invece aver commesso un errore utilizzando lo stesso prestito nella traduzione di *tondi de marbre* (Tabella 1). Ricordiamo che Leclanché (1839) aveva proposto *bas-reliefs circulaires* e Gerard Powell (2007) che, come abbiamo visto, ha ripreso recentemente questa traduzione attualizzandola, ha modificato la proposta di Leclanché impiegando il traduceur *médaille*. Sofferiamoci dunque su quest'ultimo termine per verificare se questa proposta traduca adeguatamente 'tondo' nel campo della scultura, dato che il prestito *tondo* francese è stato ormai definitivamente escluso per questo ambito.

#### 5.4 Il traduceur francese *médaille*

Il termine è presente nelle principali risorse lessicografiche monolingui generali francesi:

*Petit Robert*: **médaille** n.m. 1 Portrait ou sujet sculpté, peint, dessiné ou gravé dans un cadre circulaire ou ovale. Lambris orné de médailles. 2 Petit **bas-relief circulaire** représentant une effigie. [...]

*TLFi*: **médaille** ARCHIT. **Bas-relief circulaire ou ovale** représentant une effigie. Synon. médaille. Un médaille d'empereur romain. *Cet Alcazar a pourtant été dessiné par Alonzo Berruguete; les trophées, les bas-reliefs, les médailles de sa façade sont fouillés par un ciseau fier, hardi, patient* (GAUTIER, *Tra les montes*, 1843, p. 222). [...]

*Encyclopaedia Universalis*: **médaille** : nom masculin singulier [...]

- cadre rond ou ovale entourant un portrait miniature [...]
- **bas-relief** ou élément décoratif de forme ronde ou ovale [...]

*Dictionnaire de l'Académie française*, 9<sup>e</sup> édition en cours (1986-): **médaille** n. m. XVI<sup>e</sup> siècle. Emprunté de l'italien *medaglione* 'grande médaille'. 1. Médaille qui surpasse en poids et en volume les médailles ordinaires. *Un médaille d'or, d'argent, de bronze*. 2. **Bas-relief de forme circulaire ou ovale**, représentant un personnage illustre ou une action mémorable. *Médaille de marbre, d'ivoire. Une façade ornée de médailles*. Par anal. Tableau de même forme. *Les portraits en médaille* ou, simplement, *les médailles de David d'Angers*. Se dit aussi parfois d'un court portrait littéraire d'un personnage célèbre. Par ext. Tout élément décoratif, tout ornement de forme circulaire ou ovale. *Un fauteuil à médaille*, dont le dossier est ovale. [...]

Adeline J. (1884), *Lexique des termes d'art*: **médaille** Portraits ou sujets peints, dessinés, gravés ou **sculptés dans un entourage circulaire ou de forme elliptique**.- Se dit aussi de motifs de décoration architecturale inscrits ou placés à l'intérieur de cartouches de forme circulaire ou elliptique.

Baudry M.-T. (2002), *La Sculpture: méthode et vocabulaire: médaillon*, n.m. **Relief de forme circulaire ou elliptique.**

Secondo queste fonti il termine *médaille* designa diversi referenti. Per le accezioni relative al campo dell'arte, si riferisce innanzitutto ad un bassorilievo circolare o ovale nell'ambito della scultura. Può essere inoltre utilizzato per indicare un ritratto o soggetto realizzato all'interno di una cornice rotonda o ovale, indipendentemente dalla tecnica artistica impiegata (pittura, scultura, disegno, incisione). Per estensione, può designare qualsiasi elemento decorativo di forma circolare o ovale. Per finire, per alcune risorse consultate, *médaille* può rinviare alla cornice rotonda o ellittica che circonda il ritratto o, per analogia, a un quadro circolare o ovale. Il termine *médaille* si configura dunque come polisemico in quanto nel corso del tempo ha acquisito numerosi significati, tra cui anche quello di quadro di forma rotonda nell'ambito della pittura, mentre il prestito *tondo*, rispetto al termine 'tondo' italiano, ha mantenuto nella lingua francese soltanto un significato più specifico relativo al solo ambito della pittura.

### 5.5. La scheda 'tondo' nel dizionario plurilingue *Lessico Beni Culturali*

Sia nelle risorse lessicografiche generali e specializzate che nel corpus parallelo, il termine francese *tondo* rinvia principalmente al referente «quadro circolare» o «cornice decorativa dei soffitti», e mai strettamente ad un basso-rilievo e quindi all'ambito della scultura. La ricerca nelle stesse risorse ha rivelato che il termine francese *médaille* si riferisce invece soprattutto al basso-rilievo, anche se può designare per estensione altri referenti. Attraverso questa analisi, si è evinto che il campo concettuale del termine 'tondo' in italiano non corrisponde né a quello del termine francese *tondo*, da cui è stato preso in prestito, né a quello di *médaille*. I due traduttori francesi *tondo* e *médaille* sono dunque equivalenti parziali. Diversamente da quanto avviene attualmente nei dizionari bilingui, nella scheda lessicografica del dizionario plurilingue LBC essi verranno adeguatamente distinti, il primo come termine dell'ambito della pittura di uso specialistico, e il secondo principalmente come termine divulgativo dell'ambito della scultura e in uso anche, per analogia, nell'ambito della pittura (vedremo un esempio di scheda lessicografica LBC nel prossimo paragrafo).

Infatti, la presenza del termine *médaille* in tutte le risorse francesi generali consultate ha suggerito che questo termine è di uso corrente e può essere impiegato anche da utenti non specialisti. La presenza invece del termine *tondo* in francese soltanto in risorse specializzate dell'ambito della pittura ha permesso di evincere che il prestito *tondo* è di uso specialistico. Abbiamo dunque avuto conferma di quanto osservato succintamente da Garzaniti e Farina (2013: 507) quando affermano che 'tondo' in italiano e

*tondo* in francese presentano nell'ambito dell'arte un problema di sinonimia e connotazione:

in francese la parola italiana 'tondo', usata per riferirsi a un quadro tondo viene tradotta *médailлон* nei testi divulgativi (descrizioni di opere d'arte in guide turistiche), mentre gli specialisti di storia dell'arte usano talvolta l'italianismo 'tondo'<sup>12</sup>.

## 6. Analisi di 'figura' nella Vita di Michelagnolo Buonarroti

Se la scelta di analizzare il termine 'tondo' è stata dettata dall'eterogeneità dei traduttori riscontrati nelle diverse traduzioni francesi delle *Vite*, il criterio di selezione del termine 'figura' è stato invece l'elevato numero di occorrenze di questa forma all'interno del sotto-corpus Michelangelo. Inserendo nella maschera di ricerca di *HyperVasari* l'espressione regolare 'figur.'<sup>13</sup>, si ottengono 68 risultati (Fig. 2) che includono tutte le forme flesse di 'figura', tra cui, oltre alla forma plurale, 'figure', anche la forma diminutiva al plurale 'figurine'.

La consultazione del corpus parallelo delle traduzioni francesi rivela che, nella maggior parte dei casi, queste forme flesse sono state tradotte con l'equivalente francese *figure*: questo traduttore appare 8 volte nella traduzione di Chastel (1981), 2 volte nelle traduzioni di Leclanché (1839) e di Luciani (2002) e 7 volte in quella a cura di Gerard Powell (2007).

È possibile selezionare una traduzione in particolare per visionare le opzioni adottate da ogni traduttore e confrontarne le scelte lessicali, sintattiche o stilistiche, anche in ottica diacronica. La presenza limitata del traduttore *figure* nella traduzione di Luciani è dovuta al fatto che si tratta di una traduzione non integrale del testo, mentre nel caso di Leclanché si riscontra che quest'ultimo utilizza parole diverse per tradurre il termine 'figura' a seconda del contesto. Delle quattro traduzioni esaminate, la traduzione di Leclanché mostra infatti in generale la maggiore varietà lessicale. Come si evince dall'analisi effettuata automaticamente in *HyperVasari*, egli impiega sei traduttori diversi (*figure, statue, colosse, Bacchus, David, Apollon*) e in un caso sostituisce il sostantivo con il pronome neutro *en* (Fig. 3).

<sup>12</sup> A nostro avviso questa affermazione dovrebbe essere attenuata, dato che abbiamo riscontrato in una delle guide turistiche consultate, quindi in un testo divulgativo, il *Guide Officiel de la Galerie des Offices* (Fossi 2014b), un'attestazione del termine 'tondo', riferito al celebre quadro *Tondo Doni*, tradotta in francese con il prestito *tondo*: It. «Il celebre **tondo**, con le figure come proiettate in uno specchio convesso [...]» (p. 65); Fr. «Ce célèbre **tondo**, dont les personnages semblent comme projetés dans un miroir convexe [...]» (p. 63).

<sup>13</sup> Nel linguaggio informatico, il punto indica la ricerca di qualsiasi carattere che segue una determinata forma. Questo tipo di ricerca è necessaria perché in *HyperVasari* l'opzione per lemmatizzare con il programma *TreeTagger* al momento non è funzionante.

Figura 2 – Risultati della ricerca delle forme flesse di 'figura' in HyperVasari.

Hypermachavel Aligneur

Fichier Edition Options Aide

Texte - Oeuvre - Book - Segment 169

1550\_Torrenstiana | 1568\_Giuntina | 1839\_Leclanché | 1981\_Chastel | 2002\_Luciani | 2007\_Grasset

tiene la croce, il quale è una *figura* miracolosissima, che fu posto nella Minerva, allato alla cappella maggiore, per messer Antonio Metelli.

*figura* mirabilissima, che fu posto nella Minerva, allato alla cappella maggiore, per messer Antonio Metelli.

près du chœur, la belle *statue* du Christ à la croix, que lui avait commandée Messer Antonio Metelli.

Christ nu tenant la croix, *figure* tout à fait admirable qui fut mise près du chœur à la Minerva par messire Antonio Metelli.

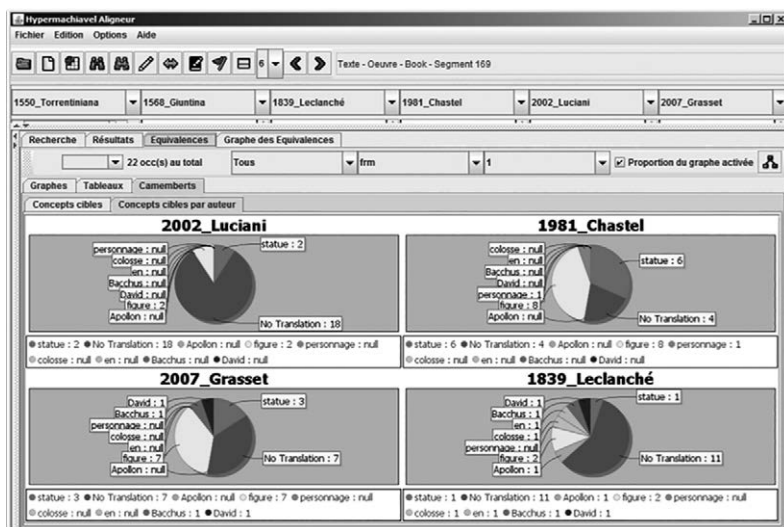
*figure* admirable qui fut placée dans l'église de la Minerva, à côté de la grande chapelle, pour Messire Antonio Metelli.

Recherche Résultats Equivalences

68 résultats à la requête figur.

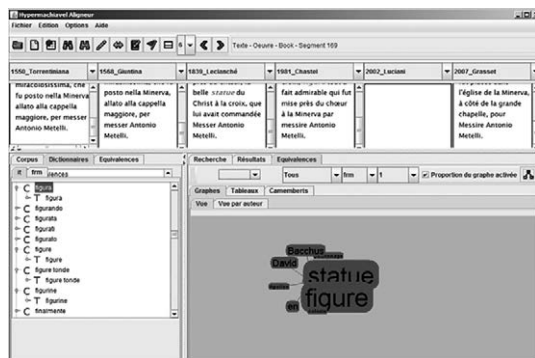
ID	Localisation	Texte	Contexte Gauche	Pivot	Contexte Droit
Annot_74	Texte - Oeuvre - Book	1550_Torrenstiana	modello tutto pieno di	figura	et adorno di cose
Annot_75	Texte - Oeuvre - Book	1550_Torrenstiana	i brevi tempo dur	figura	di marmo fin le
Annot_76	Texte - Oeuvre - Book	1550_Torrenstiana	lavorava di terra cotta	figura	fonde che da Bertoldo
Annot_77	Texte - Oeuvre - Book	1550_Torrenstiana	volontà, ne cavasse u.	figura	
Annot_78	Texte - Oeuvre - Book	1550_Torrenstiana	il quale è una	figura	miracolosissima, ch
Annot_79	Texte - Oeuvre - Book	1550_Torrenstiana	gi far fare una	figura	
Annot_80	Texte - Oeuvre - Book	1550_Torrenstiana	esaminando potersi	figura	di quel sasso cavare
Annot_81	Texte - Oeuvre - Book	1550_Torrenstiana	legname fortissimo.	figura	col canapi sospeser...
Annot_82	Texte - Oeuvre - Book	1550_Torrenstiana	gli pareva di quella	figura	rispose il Francia che
Annot_83	Texte - Oeuvre - Book	1550_Torrenstiana	già mai. Nella qual	figura	questa conoscenza s
Annot_84	Texte - Oeuvre - Book	1550_Torrenstiana	ne abbozzò una parte	figura	i Prigioni in varie
Annot_85	Texte - Oeuvre - Book	1568_Giuntina	medaglio, gioie, cam	figura	di bronzo, pitture, dis
Annot_86	Texte - Oeuvre - Book	1568_Giuntina	sopra tre o quattro	figura	bozzate di sua mano
Annot_87	Texte - Oeuvre - Book	1568_Giuntina	prima delle quali era	figura	per lo Studio, perciò

Figura 3 – I diversi traducenti delle forme flesse di 'figura' in HyperVasari.



L'attivazione delle proporzioni del grafico in base alla frequenza di apparizione dei traducenti (Fig. 4) mostra chiaramente che i traducenti *figure* e *statue* risultano complessivamente essere i più frequenti. Vengono infatti utilizzati rispettivamente 24 e 15 volte nelle quattro traduzioni:

Figura 4 – I traducenti più frequenti delle forme flesse di ‘figura’ in HyperVasari.



È interessante il dato che il termine italiano ‘figura’ sia stato tradotto con l’equivalente francese *statue* in diversi casi: viene attestata 1 occorrenza nella traduzione di Leclanché (1839), ben 6 in quella di Chastel (1981), 2 in Luciani (2002) e 3 nella traduzione di Gerard Powell (2007). Questi primi dati, estrapolati dal corpus parallelo, costituiscono dunque il punto di partenza per un’analisi approfondita del concetto di ‘figura’ nell’ambito della scultura.

Ancora una volta la consultazione delle risorse lessicali monolingui italiane è necessaria per chiarire il significato di questo termine. Le risorse italiane consultate (Zingarelli 2013, Sabatini-Coletti, *Enciclopedia Treccani*, Grassi-Pepe 1994) presentano una definizione simile del termine ‘figura’ che, per necessità di sintesi, riassumiamo nel modo seguente: nelle arti figurative, si tratta della «rappresentazione dipinta, disegnata o scolpita di un essere vivente (uomo, donna o animale)». Verifichiamo se le definizioni dei due traducenti principali scelti dai traduttori francesi (*figure* e *statue*) presentano gli stessi tratti semantici distintivi e il medesimo ambito di applicazione del termine italiano e possono quindi essere considerati traducenti pertinenti.

## 6.1 Il traducente *figure*

La ricerca del termine francese *figure* nelle risorse relative a questa lingua fornisce i seguenti risultati. La consultazione dei monolingui si rivela soddisfacente: tutte le definizioni di *figure* corrispondono perfettamente al concetto italiano espresso dal termine *figura* in quanto si ritrovano i tratti «rappresentazione dipinta, disegnata o scolpita di un essere vivente (uomo, donna o animale)».

*Petit Robert*: **figure** n. f. **Représentation** d’un personnage > **effigie, portrait, statue.**

*TLFi*: **figure** **Représentation** au naturel, **par le dessin, la peinture, la sculpture** (ou le moulage), d’un être pris en entier.

Adeline J. (1884), *Lexique des termes d'art*: **figure** Se dit d'une **représentation** d'homme, de femme ou d'animal dessinée, **peinte ou sculptée**. Orner un paysage de *figures*, c'est y ajouter des personnages et des animaux ; [...] Une figure de *grandeur nature* est une représentation d'homme ou de femme de grandeur naturelle ; une *demi-figure* est celle qui représente un personnage jusqu'à la ceinture seulement, et une figure *demi-nature* est une représentation à moitié de la taille ordinaire. Enfin une *figure d'académie* désigne la dimension ordinaire des figures peintes et modelées dans les écoles des beaux-arts et dont la proportion est un peu au-dessous de la demi-nature.

Baudry M-T. (2002), *La Sculpture: méthode et vocabulaire*: **figure Représentation en ronde-bosse, en haut relief ou en bas-relief** du corps humain ou animal.

Soltanto una risorsa specializzata, l'opera di Baudry (2002), interamente dedicata alle tecniche e alla terminologia dell'ambito della scultura, riporta informazioni apparentemente nuove sul concetto di 'figura': il termine francese *figure* può essere impiegato indistintamente per designare la rappresentazione a tutto tondo, in alto rilievo o in basso rilievo, umano o animale, preso per intero («représentation en ronde-bosse, en haut relief ou en bas-relief d'un corps, humain ou animal, pris en entier»).

Queste precisazioni riguardo alla tecnica di scultura impiegata per rappresentare i soggetti non sono presenti nelle risorse italiane, generali e specializzate, relativamente al termine *figure*, anche se si può considerare che l'informazione data esplicitamente da Baudry sia implicita nel participio passato «scolpita» (riferito a «rappresentazione») trovato nelle risorse lessicografiche italiane.

Conducendo ricerche supplementari alla voce 'scultura' (per «figura scolpita») nel Grassi-Pepe (1994), scopriamo che originariamente il termine rinvia esclusivamente alle figure realizzate attraverso la tecnica dell'intaglio o dell'incisione e che, dalla seconda metà del XVI secolo, il termine 'scultura' ha subito un ampliamento semantico grazie agli sviluppi tecnici realizzati in questo ambito, iniziando a designare le figure realizzate con qualsiasi tipo di materia o tecnica:

tutti i corpi composti e formati d'ogni sorte di materia, pur che in loro sia cavo, o rilievo, o di basso, o di mezzo, o di tondo rilievo circondato [...] basta che sia inteso Scultura tutte le figure che nelle materie si formano [...].

A partire da questi dati, si desume che il concetto di 'figura scolpita' rinvia implicitamente alla «rappresentazione a tutto tondo, in alto rilievo o in basso rilievo di un corpo, umano o animale». L'utilizzo del web come corpus di attestazioni ne dà ulteriore conferma: la ricerca di 'figura scolpi-

ta' su Google.it produce risultati che confermano questa conclusione: sono 'figure scolpite' da Michelangelo sia la *Centauromachie* (1492), celebre esempio di basso rilievo, che il *David* (1501-1504), nota rappresentazione a tutto tondo. Il traduttore francese *figure* che, come abbiamo visto, si applica a qualsiasi tipo di scultura, risulta dunque essere adeguato.

## 6.2 Il traduttore *statue*

Il secondo traduttore riscontrato con una frequenza abbastanza elevata nel corpus parallelo di traduzione francese è *statue* che, nelle risorse lessicografiche francesi, presenta tratti semantici distintivi diversi rispetto a *figure*:

*Petit Robert*: **statue** Ouvrage de sculpture **en ronde-bosse** représentant en entier un être vivant [réel ou imaginaire]

*TLFi*: **statue** subst. fém. **A.** Ouvrage **de plein relief**, sculpté ou moulé, représentant en entier un être animé (p. oppos. à *tête*, *buste*), et par allégorie, une qualité abstraite, un élément ou un effet de la nature. Synon. *figure*. [...]

Adeline J. (1884), *Lexique des termes d'art*: **statue** Ouvrage de sculpture **en ronde bosse** représentant la figure humaine. Une statue en marbre, en bronze, en pierre, en bois. *Une statue colossale*.

Baudry M-T. (2002), *La Sculpture: méthode et vocabulaire*: **statue** représentation **en ronde-bosse**.

*Termium Plus*: **statue** Ouvrage de sculpture **en ronde bosse** représentant en entier un être vivant (humain ou animal).

Il principale tratto distintivo che differisce è il tratto «en ronde bosse», detto anche «de plein relief»<sup>14</sup>. Il termine francese *statue* designa esclusivamente una rappresentazione a tutto tondo, motivo per cui questo termine deve essere preso in una accezione più ristretta, come viene sottolineato ancora da Baudry (2002) che precisa: «ce serait donc à tort que l'on qualifierait de statues les gisants en demi ou haut-relief». *Figure* e *statue* si configurano dunque come due termini che appartengono allo stesso campo concettuale e che sono legati tra loro da una relazione d'inclusione: *figure* è l'iperonimo di *statue*, e quest'ultimo il suo iponimo.

<sup>14</sup> «On appelle figure de plein-relief, ou de ronde-bosse, celle qui est isolée et terminée en toutes les vues» (Lacombe, J. 1753, *Dictionnaire portatif des beaux-arts*).

Una risorsa di supporto alla traduzione dovrebbe fornire indicazioni semantiche distintive relative ai possibili traduttori, mostrando le relazioni di inclusione esistenti. Questo tipo di informazioni, assenti nei principali dizionari bilingui sul mercato che, alla voce *figura*, presentano il solo traduttore univoco *figure*, per l'ambito dell'arte in generale (*Boch*) e per designare sia un'immagine dipinta che scolpita (*Garzanti Francese*), sono invece presenti nelle schede lessicografiche bilingui francese-italiano del dizionario plurilingue LBC (Fig. 5)<sup>15</sup>. Esse guidano il traduttore nell'uso corretto dei diversi traduttori possibili, come nel caso di *statue* che, come dimostrato, è un traduttore assolutamente appropriato se impiegato per riferirsi a 'figure scolpite' (*en ronde bosse*).

Figura 5 – Esempio di scheda lessicografica bilingue italiano-francese LBC del termine 'figura'.

Lessico plurilingue dei Beni Culturali fiorentini				
				
ricerca		fonti	il gruppo	collaborazioni
Lingua fonte italiano		Lingua target français	Ricerca: figura	
lessico	enciclopedia	etimologia	corpus parallelo	concordanze
<sup>2</sup> <i>figura</i> (n.f.) [f'iɡura]  SCULTURA ELEMENTO	Rappresentazione scolpita, a tutto tondo, in altorilievo o in basso rilievo, di un uomo, una donna o un animale.  (SN) <i>figura serpentinata, mezza figura, figura allegorica, figura di prua</i> ; (LOC) <i>figura a tutto tondo</i> .  I. LEOTTA			traduzioni ← citazioni analisi associazioni
<i>représentation en ronde-bosse, en haut relief ou en bas relief</i>		■ ■ <sup>1</sup> <i>figure</i> (n.f.) [fiɡy:R] SCULP. (ELEM)		
<i>représentation en ronde-bosse</i>		■ ■ <sup>2</sup> <i>statue</i> (n.f.) [sta ty] SCULP. (ELEM)		

Tornando al corpus parallelo, riportiamo soltanto due diverse attestazioni di 'figura' presenti nel sotto-corpus Michelangelo e le sue traduzioni in cui si nota che due traduttori francesi, Leclanché (Tabella 4) e Gerard Powell (Tabella 5), utilizzano come traduttore il termine francese *statue*.

<sup>15</sup> Menzioniamo altre caratteristiche delle schede lessicografiche LBC che possono essere di grande utilità per il traduttore: indicazione semantica distintiva, definizione, informazione etimologica, raccomandazioni d'uso, problemi traduttivi (fiche bilingue 'figura'), fotografie o illustrazioni (fiches monolingue *figure* et *statue*).



Tabella 4 – Traduzioni di 'figura' nel corpus parallelo Michelangelo.

Vasari	1550	Mandò in questo tempo Pietro Urbano pistolese, suo creato, a Roma a mettere in opera un Cristo ignudo che tiene la croce, il quale è una <b>figura</b> miracolosissima, che fu posto nella Minerva, allato alla cappella maggiore, per messer Antonio Metelli.
Vasari	1568	Mandò in quello tempo Pietro Urbano pistolese, suo creato, a Roma a mettere in un opera un Cristo ignudo che tiene la croce, il quale è una <b>figura</b> mirabilissima, che fu posto nella Minerva, allato alla cappella maggiore, per messer Antonio Metelli.
Leclanché	1839	A cette époque, Michel-Ange envoya à Rome Pietro Urbano de Pistoia, son élève, pour placer dans l'église de la Minerva, près du chœur, la belle <b>statue</b> du Christ à la croix, que lui avait commandée Messer Antonio Metelli.
Chastel	1981	A son élève Pietro Urbano de Pistoia il expédia à Rome pour le mettre en place un <i>Christ nu tenant la croix</i> , <b>figure</b> tout à fait admirable qui fut mise près du chœur à la Minerva par messire Antonio Metelli.
Luciani	2002	X
Gerard Powell	2007	A cette époque, il envoya Pietro Urbano de Pistoia, son élève, à Rome, pour mettre en œuvre un Christ nu tenant la croix, <b>figure</b> admirable qui fut placée dans l'église de la Minerva, à côté de la grande chapelle, pour Messire Antonio Metelli.

Tabella 5 – Altre traduzioni di 'figura' nel corpus parallelo Michelangelo.

Vasari	1550	Per che Giuliano da San Gallo et Antonio suo fratello fecero un castello di legname fortissimo, e quella <b>figura</b> coi canapi sospesero a quello, acciò che scotendosi non si troncasse, anzi venisse crollandosi sempre;
Vasari	1568	Per che Giuliano da S. Gallo et Antonio suo fratello fecero un castello di legname fortissimo, e quella <b>figura</b> con i canapi sospesero a quello, acciò che scotendosi non si troncasse, anzi venisse crollandosi sempre;
Leclanché	1839	Giuliano et Antonio San-Gallo, à l'aide d'une machine fort ingénieuse, transportèrent sans danger, en 1504, ce <b>colosse</b> sur la place de' Signori.
Chastel	1981	Giuliano de Sangallo et son frère Antonio fabriquèrent une cage de bois très solide et le <b>personnage</b> y fut suspendu par des cordes de chanvre pour éviter qu'il ne soit brisé par les secousses ou même ne s'effondre;
Luciani	2002	Aussi Giuliano da Sangallo et Antonio son frère construisirent un châssis en bois très robuste, et y suspendirent cette <b>figure</b> avec des câbles de chanvre, afin qu'elle ne se brise pas dan les secousses, mais qu'au contraire elle puisse constamment remuer.
Gerard Powell	2007	Giuliano da San Gallo et Antonio, son frère, firent un énorme chariot de bois et y suspendirent la <b>statue</b> avec des cordes, de manière qu'elle ne fût pas fracassée par les heurts, mais qu'elle se balançât continuellement;

Entrambe le opere, il *Christ nu tenant la croix*, referente della prima citazione, e il *David* referente del secondo passaggio, sono rappresentazioni a tutto tondo, a dimostrazione del fatto che le scelte dei due traduttori francesi sono pertinenti e ragionate e che, per questo, meritano di essere segnalate in una risorsa per la traduzione del lessico artistico.

## 6. Conclusioni

In questo contributo abbiamo cercato di mostrare che la consultazione di un corpus parallelo di traduzione consente di prendere coscienza delle difficoltà di traduzione di un dato termine, sulla base dell'osservazione delle strategie adottate da diversi traduttori e, talvolta, anche degli errori o incoerenze da loro commessi. Non è stata tuttavia nostra intenzione portare giudizi sulla correttezza o meno di determinate traduzioni, bensì abbiamo voluto mostrare in che misura l'esplorazione di un corpus parallelo permetta di analizzare le possibili opzioni traduttive in un'ottica descrittiva e documentaria.

Un corpus diacronico, composto da traduzioni realizzate in diverse epoche, come quello costruito per il progetto LBC, ha inoltre il merito di illustrare il contesto storico-culturale e l'evoluzione di un termine, dei suoi referenti e dei suoi traduttori. Lo abbiamo visto con il primo esempio analizzato dettagliatamente, il termine italiano 'tondo', in uso nel lessico artistico italiano del Rinascimento nell'ambito della pittura e della scultura per indicare sia un quadro di forma rotonda che una figura in rilievo o basso-rilievo, che è stato preso in prestito in seguito dalla lingua francese. Dal confronto delle quattro traduzioni francesi della *Vita* di Michelangelo, si è osservato che diversi traduttori (*bas-relief circulaire*, *tondo*, *médaille*, *tableau circulaire*, *tableau rond*) sono stati impiegati nel corso del tempo per esprimere queste due accezioni o concetti<sup>16</sup>. Laddove le traduzioni non mostravano sempre una coerenza interna, come nel caso di Leclanché (1839), che traduce il 'tondo di pittura' talvolta con *tableau circulaire* (Tabella 2, sotto-corpus Michelangelo) talvolta con *médaille* (Tabella 3, sotto-corpus Botticelli), o come nel caso del traduttore *tondo* utilizzato da Chastel (1981) per tradurre sia i 'tondi di pittura' che i 'tondi di scultura' del testo vasariano, la consultazione di diverse risorse lessicografiche e terminologiche ha permesso di precisare il campo di applicazione esatto di ogni traduttore. Il *GDT* per esempio si è rivelato utile per chiarire che il campo concettuale del termine *tondo* in francese non corrisponde esattamente a quello italiano, in quanto in lingua francese è il termine privilegiato per designare esclusivamente un quadro di forma rotonda nell'ambito della pittura e del disegno artistico, e per inferire, di

<sup>16</sup> Entrambe le denominazioni sono valide, a seconda del punto di vista adottato, lessicografico o terminologico.

conseguenza, che un altro traducevole sarebbe stato adeguato per tradurre il 'tondo di scultura'. Dinanzi all'assenza di informazioni chiare nei dizionari bilingui – abbiamo visto nello specifico che i dizionari bilingui consultati forniscono informazioni contraddittorie che non possono che indurre in confusione l'utente, in quanto i possibili traducevoli non vengono distinti chiaramente per ogni ambito (pittura e scultura), ma vengono o presentati come sinonimi interscambiabili<sup>17</sup>, oppure uno stesso traducevole viene proposto per entrambi gli ambiti, o ancora un'unica marca settoriale generica come ART. è associata ad un solo traducevole –, la consultazione del corpus parallelo si è rivelata utile per reperire gli altri possibili traducevoli di 'tondo di scultura' all'interno delle traduzioni pubblicate convalidate da studiosi dell'arte professionisti. È il caso della proposta attestata nella traduzione più recente a cura di Gerard Powell (2007), punto di partenza per una verifica, ancora una volta lessicografica, terminologica ed enciclopedica, della adeguatezza del traducevole proposto per l'ambito della scultura, *médaille*, che si è rivelato valido e pertinente, nonché attuale.

Il lavoro del traduttore si configura dunque oggi idealmente come un andirivieni costante tra consultazione di un corpus parallelo affidabile e verifica in fonti specializzate. Corpus e dizionari, interrogati in modo appropriato, sono così complementari e permettono al traduttore di formarsi un'idea completa o comunque molto ampia delle molteplici possibilità di traduzione, dalle più aderenti al testo di partenza alle più creative o aperte alla diversità di realizzazione dei fenomeni discorsivi.

È quanto è stato illustrato anche dal secondo esempio analizzato in questo contributo: le traduzioni esaminate del termine 'figura' nel sotto-corpus della *Vita* di Michelangelo mostrano una grande varietà di scelte traduttive, sia all'interno di una stessa traduzione (particolarmente evidente in Leclanché 1839), che nelle diverse traduzioni realizzate in varie epoche. L'esplorazione del corpus ha permesso di trovare numerose occorrenze di un traducevole, il termine *statue*, che, pur non presentando esattamente gli stessi tratti semantici distintivi del suo traducevole letterale *figure*, risulta assolutamente appropriato nei contesti in cui il referente è solo una rappresentazione a tutto tondo, e anzi aggiunge alla traduzione maggiore precisione evitando ambiguità interpretative. Una scelta traduttiva di questo tipo, basata sulla conoscenza della relazione d'inclusione che esiste tra il termine iperonimo *figure*, che designa una rappresentazione sia a tutto tondo, sia in alto rilievo, che in basso rilievo, e il suo iponimo *statue*, non può che essere effettuata da uno specialista dell'arte che conosce la terminologia del settore e che è consapevole di non rischiare di commettere un'inesattezza concettuale scostandosi leggermente dal testo di partenza. Il corpus offre quindi la ricchezza

<sup>17</sup> Per esempio, il *Larousse Francese* 2006 propone i traducevoli *médaille* e *tondo* indifferente per la pittura e la scultura.

di un patrimonio linguistico posseduto da traduttori specialisti dell'ambito artistico che né un traduttore cosiddetto generalista né una risorsa consultata abitualmente da un traduttore, come un dizionario bilingue, possiede.

Il dizionario plurilingue del progetto LBC, attualmente in corso di realizzazione, è stato concepito con l'obiettivo di integrare questa ricchezza, guidando il traduttore nella scelta dei traducanti in uso nel linguaggio artistico sulla base di un'analisi accurata delle preferenze espresse da traduttori francesi riconosciuti, nonché delle informazioni fornite da risorse lessicografiche e terminologiche affidabili. Il dizionario LBC ha pertanto la vocazione di imporsi come uno strumento per eccellenza di supporto al lavoro del traduttore del linguaggio artistico, spesso abbandonato a sé stesso a causa della mancanza, deplorata anche da guide e accompagnatori turistici, sul mercato attuale di strumenti di riferimento per questo ambito. In questo contributo ci siamo soffermati sulla sinergia possibile tra corpora paralleli di traduzione e risorse lessicografiche che descrivono la lingua dell'arte, tutti selezionati per la loro affidabilità, qualità che è alla base del progetto stesso del dizionario LBC, sviluppato sin dall'origine esclusivamente su supporto elettronico per offrire all'utente target informazioni precise, accurate e varie, in più lingue, sulle strutture semantiche e concettuali del lessico artistico e sulla sua traduzione.

## Bibliografia

- Aston G. 1999, *Corpus use and learning to translate*, «Textus», 12: 289-313.
- Baker M. 1999, *The role of corpora in investigating the linguistic behavior of translators*, «International Journal of Corpus Linguistics», 4: 281-298.
- Beeby A., Rodríguez I., Sánchez-Gijón P. 2009, *Corpus Use and Translating*, John Benjamins, Amsterdam.
- Bernardini S., Castagnoli S. 2008, *Corpora for translator education and translation practice*, in Rodrigo Y. (ed.), *Topics in Language Resources for Translation and Localisation*, John Benjamins, Amsterdam: 39-55.
- Biber D., Connor U., Upton T. 2008, *Discourse on the move*, John Benjamins, Amsterdam.
- Brossaud C., Reber B. (eds.) 2007, *Humanités numériques 1. Nouvelles technologies cognitives et épistémologie*, Lavoisier, Paris.
- Brunon H. (2012), *Qui était André Chastel?*, *Le centre André Chastel*, <<http://www.centrechastel.paris-sorbonne.fr/page/qui-etait-andre-chastel>> (10/2016).
- Cabré M.T. 1993, *La terminología: Teoría, metodología y aplicaciones*, Editorial Antártida/Empúries, Barcelona.
- Chastel A. 1964, *Histoire de la Renaissance*, «Sciences historiques et philologiques», XCVII (1): 287-290, <[http://www.persee.fr/doc/ephe\\_0000-0001\\_1964\\_num\\_1\\_1\\_4846](http://www.persee.fr/doc/ephe_0000-0001_1964_num_1_1_4846)> (10/2016).
- Chiari I. 2007, *Introduzione alla linguistica computazionale*, Laterza, Roma-Bari.
- 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 tem-*

- pera et guazzo, in Carpi E. (ed.), *Prospettive multilingue e interdisciplinari nel discorso specialistico*, Pisa University Press, Pisa: 33-57.
- 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: 1-10, <[http://publifarum.farum.it/ezone\\_articles.php?art\\_id=335](http://publifarum.farum.it/ezone_articles.php?art_id=335)> (10/2016).
- 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. (a cura di), *Strategie e Programmazione della Conservazione e Trasmissibilità del Patrimonio Culturale*, Fidei Signa, Roma: 500-509.
- Granger S., Lerot J., Petch-Tyson S. 2003, *Corpus-based approaches to Contrastive Linguistics and Translation Studies*, Rodopi, Amsterdam-New York.
- Habert B., Nazarenko A., Salem A. 1997, *Les linguistiques de corpus*, Armand Colin, Paris.
- Kübler N., Bordet G., Pecman M. 2011, *La linguistique de corpus entretient-elle d'étroites relations avec la traduction pragmatique ?*, in Van Campenhoudt M., Lino T., Costa R. (eds.), *Passeurs de mots, passeurs d'espoir. Lexicologie, terminologie et traduction face au défi de la diversité*, Éditions des archives contemporaines/Agence Universitaire de la Francophonie, Paris: 579-591.
- L'Homme M.-C. 2003, *Les combinaisons lexicales spécialisées (CLS). Description lexicographique et intégration aux banques de terminologie*, in Grossmann F., Tutin A. (éds.), *Les Collocations: analyse et traitement*, De Werelt, Amsterdam: 89-103.
- Pearson J. 1998, *Terms in Context*, John Benjamins, Amsterdam.
- Pruvost J. 2006, *Les dictionnaires français : Outils d'une langue et d'une culture*, Ophrys, Paris.
- Revue de l'Art 1990, *André Chastel*, n. 89: 5-7, <[http://www.persee.fr/doc/rvart\\_0035-1326\\_1990\\_num\\_89\\_1\\_347845](http://www.persee.fr/doc/rvart_0035-1326_1990_num_89_1_347845)> (10/2016).
- Schnapp J., Presner P. 2009, *Digital humanities manifesto 2.0.*, <[http://www.humanitiesblast.com/manifesto/Manifesto\\_V2.pdf](http://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf)> (10/2016).
- Williams G. (ed.) 2005, *La linguistique de corpus*, Presses Universitaires de Rennes, Rennes.
- Zanettin F., Bernardini S., Stewart D. (eds.) 2003, *Corpora in Translator Education*, St-Jerome, Manchester.

### Corpus parallelo

- Vasari G. 1550, *Le vite de' piu eccellenti architetti, pittori, et scultori italiani, da Cimabue insino a' tempi nostri*, Lorenzo Torrentino, Firenze (ed. Torrentiniana).
- Vasari G. 1568, *Le vite de' piü ecelenti pittori, scultori ed architettori*, Giunti, Firenze (ed. Giuntina).
- Vasari G. 1839-42, *Vies des peintres, sculpteurs et architectes*, Tessier, Paris (traduzione di Ph.-A. Jeanron e L. Leclanché).
- Vasari G. 1966, *Les peintres toscans*, Hermann, Paris (traduzione di A. Chastel).
- Vasari G. 1981, *La vie des meilleurs peintres, sculpteurs et architectes*, Berger-Levrault, Paris [1995, Actes Sud 2005] (traduzione di A. Chastel).

Vasari G. 2002, *Vies d'artistes*, Gallimard, Paris (traduzione di G. Luciani, edizione bilingue francese-italiano).

Vasari G. 2007, *Vies des artistes*, Grasset, Paris (a cura di V. Gerard Powell).

#### Guide turistiche

Fossi G. 2014a, *Galleria degli Uffizi. Guida ufficiale. Tutte le opere*, Giunti, Firenze.

Fossi G. 2014b, *Galerie des Offices. Guide officiel. Toutes les œuvres*, Giunti, Firenze (traduzione di H. Aste, D. Chevalier e S. Lomnitz).

#### Dizionari ed enciclopedie

Adeline J. 1884, *Lexique des termes d'art*, Quantin, Paris.

Baudry M.-T. 2002, *La Sculpture: méthode et vocabulaire*, Ed. du Patrimoine, Paris.

Boch 2014, *Dizionario francese-italiano/italiano-francese di Raoul Boch*, sesta edizione a cura di C. Salvioni Boch, Zanichelli-Le Robert, Bologna.

Boutard M. 1826, *Dictionnaire des arts du dessin, la peinture, la sculpture, la gravure et l'architecture*, Ch. Gosselin libraire, Paris.

Dictionnaire de l'Académie 1986, *Dictionnaire de l'Académie française*, nona edizione in corso, 1986-, <<http://atilf.atilf.fr/academie9.htm>>.

Dictionnaire de la Peinture Larousse 2003, <<http://www.larousse.fr/archives/peinture>>.

Diderot D., D'Alembert J.-B. (dir.) 1751-1772, *Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers, par une Société de Gens de Lettres*, Briasson, Paris, <<https://encyclopedia.uchicago.edu/>>.

*Enciclopedia Treccani*, <<http://www.treccani.it>>.

*Encyclopaedia Universalis*, <<http://www.universalis.fr>>.

Ferrante V., Cassiani E. (ed.) 1991, *Dizionario Moderno italiano-francese*, Sei, Torino.

Ferrari S. 2002, *Dizionario di arte e architettura*, Mondadori, Milano.

*Grande Dizionario Italiano Garzanti*, <<http://www.garzantilinguistica.it>>.

*Grande Dizionario Francese Garzanti* 2007, Milano.

*Grand dictionnaire terminologique (GDT)*, <<http://gdt.oqlf.gouv.qc.ca/index.aspx>>.

Grassi L., Pepe M. 1994, *Dizionario dei termini artistici*, Tea, Torino.

Hachette-Paravia 2007, *DIF Hachette-Paravia. Dizionario francese-italiano, italiano-francese*, Paravia, Torino.

*La Nuova Enciclopedia dell'arte Garzanti* 2000, Garzanti, Milano.

Lacombe J. 1753, *Dictionnaire Portatif Des Beaux-Arts Ou Abrégé de Ce Qui Concerne L'Architecture, Le Sculpture, Le Peinture, Le Gravure, Le Poésie & La Musique: Avec ... Talent de Personnes Qui Se Sont Distinguées*, Jean-Th. Herissant et Les Frères Estienne, Paris.

Larousse Francese 2006, *Il Larousse Francese. Français-italien, italiano-francese*, a cura di Edigeo, Rizzoli-Larousse, Milano-Paris.

Masciotta M. 1969, *Dizionario di termini artistici*, Le Monnier, Firenze.

Petit Robert 2013, *Dictionnaire alphabétique et analogique de la langue française*, Nouvelle édition du *Petit Robert* de Paul Robert, texte remanié et amplifié sous la direction de Josette Rey-Debove et Alain Rey, éd. Le Robert, Paris.

*Sabatini Coletti Dizionario della lingua italiana*, <<http://dizionari.corriere.it>>.  
*Termium Plus*, <<http://www.termiumplus.gc.ca>>.  
*Trésor de la Langue Française informatisé (TLFi)*, <<http://atilf.atilf.fr/>>.  
Zingarelli N. 2013, *Vocabolario della lingua italiana*, Zanichelli, Bologna.

## Wikipedia: posibilidades y límites para la extracción de terminología multilingüe sobre el arte

*Resumen:* El objetivo de este artículo es reflexionar sobre las oportunidades y límites que presentan las herramientas 2.0 y los recursos disponibles en línea, como Wikipedia, para la elaboración de bases de datos y recursos lexicográficos relacionados con el patrimonio artístico y cultural. Dado que buena parte de la información contenida en Wikipedia se considera útil para los traductores especializados, cabe interrogarse sobre el papel que esta puede desempeñar en la constitución y mejora de bases de datos multilingües sobre léxico especializado, pero también, de qué manera proyectos como *Lessico dei Beni Culturali* pueden integrarse y enriquecer este recurso colaborativo. Para ilustrar estas cuestiones se presentan los resultados de un análisis cualitativo y comparativo de dos voces relacionadas con las técnicas y las materias de la pintura publicadas en las versiones italiana y española de la enciclopedia.

*Palabras clave:* wikipedia, lenguaje del arte, lexicografía, terminología.

*Abstract:* L'obiettivo di questo articolo è quello di riflettere sulle opportunità e sui limiti che presentano gli strumenti e le risorse 2.0 disponibili in rete, come ad esempio Wikipedia, per lo sviluppo di database e risorse lessicografiche e terminologiche legate al patrimonio artistico e culturale. Dal momento che gran parte delle informazioni contenute in Wikipedia è utilizzata oggi da traduttori specializzati, è possibile interrogarsi sul ruolo che essa possa svolgere nella creazione e nel miglioramento di database multilingue su lessico specializzato di progetti destinati a questo tipo di utenti, quali il progetto *Lessico dei Beni Culturali*; ma anche su come questo stesso progetto possa integrare o arricchire l'enciclopedia collaborativa. Al fine di illustrare questa riflessione, si presentano i risultati di un'analisi qualitativa e comparativa di due articoli relativi alle tecniche e materiali della pittura (tempera, guazzo), pubblicati nelle versioni italiana e spagnola dell'enciclopedia.

Parole chiave: wikipedia, linguaggio dell'arte, lexicografia, terminologia.



## I. Introducción

Señala Águila que «la cuestión de los medios tecnológicos ha originado toda una nueva concepción de la filología que posee como imagen precisa la noción de docuverso» (2006: 2), esto es, en palabras de Millán, una serie de textos interconectados «entre los que se puede mover el investigador y al que también contribuirá su propia investigación, creada en un entorno digital, y puesta en conexión con su comunidad» (1999: 160). En este contexto, la actividad del lexicógrafo se transforma puesto que, como apunta Alvar Ezquerro, la lexicografía tiene hoy como objetivo principal la recopilación documental para la elaboración de bases de datos con información estructurada y amplia, que puede ayudar en la posible redacción de un diccionario. La lexicografía se enfrenta así a la constitución de macrobases, dotadas de los programas y herramientas adecuados «para que los destinatarios puedan efectuar múltiples consultas y gestionar la ingente cantidad de datos puesta a su alcance» (2003: 226).

Como es sabido, las bases de datos deben diseñarse de acuerdo a una concepción previa de la finalidad y contenido de las mismas, sabiendo que su estructuración en campos definidos determina tanto la uniformidad como la sistematización de los artículos, facilitando la detección de deficiencias. Tales bases pueden ser de creación nueva, aunque es difícil prescindir de la tradición lexicográfica, o pueden crearse a partir de la recuperación de datos ya existentes. Por ejemplo, los datos lingüísticos relativos al patrimonio artístico y cultural pueden extraerse de corpus elaborados a partir de guías de viaje o de textos clave de la historia del arte, tal y como está planteado en el Proyecto *Lessico multilingue dei Beni Culturali* (LBC)<sup>1</sup>, cuyo objetivo principal es investigar el léxico del arte y del patrimonio cultural en distintas lenguas y en relación a la lengua italiana, en el contexto de los bienes culturales de la ciudad de Florencia, modelo por excelencia de ciudad de arte.

Desde este enfoque y considerando el peso que tiene hoy en día la información sobre el léxico del patrimonio cultural y artístico disponible en línea a través de sitios web, blogs, foros especializados, glosarios y enciclopedias y diccionarios colaborativos, cabe interrogarse sobre el papel que pueden tener estos recursos en la elaboración de bases de datos lexicográficos y terminográficos multilingües. En este trabajo nos centramos en Wikipedia, la enciclopedia multilingüe accesible en línea. Nuestro propósito

<sup>1</sup> El proyecto, coordinado por la Università di Firenze, prevé la elaboración de un diccionario en siete lenguas (alemán, chino, español, francés, inglés, portugués, ruso), en el que se incluirán distintos lemas con sus significados en relación a la lengua italiana. Los lemas proceden de bases de datos que incluyen palabras clave provenientes de corpus de textos relacionados con el arte como la edición italiana y las traducciones de *Le Vite de' più eccellenti architetti, pittori et scultori italiani* de Giorgio Vasari (1550, 1568), véase: <<http://www.lessicobeniculturali.net/contenuti/il-progetto/818>> (10/2016) y el artículo de Zotti en este volumen.

es indagar las posibilidades y límites de este recurso para la extracción de datos, principalmente términos específicos, relacionados con el léxico del arte. En primer lugar, abordamos las principales características de Wikipedia e ilustramos algunas de sus aplicaciones en el ámbito de la creación de diccionarios y ontologías; en segundo lugar, exploramos este recurso desde un punto de vista lexicológico y lingüístico-discursivo<sup>2</sup>, mediante el análisis de los artículos relativos a los términos italianos de la pintura, *tempera* y *guazzo*, y de los artículos correspondientes a estos términos en la versión española de la enciclopedia, siguiendo los planteamientos de Carpi (2016) y, en parte, Farina (2015).

## 2. Plataformas y herramientas 2.0

Aunque el concepto de Web 2.0 ha conocido distintas definiciones, todos coinciden en señalar que este se relaciona con la centralidad de plataformas basadas en la conexión entre individuos y comunidades, así como en la generación de contenidos por parte de personas que no son necesariamente competentes en programación. El concepto presupone la convergencia de medios, la producción en colaboración de contenidos y el desarrollo de las llamadas «inteligencias colectivas» (Lévy 1994). O'Reilly (2005) señala que la Web 2.0 permite crear dicha «inteligencia» a partir del intercambio de conocimientos, gracias a herramientas que facilitan la creación y edición de textos, la inclusión de imágenes en distintos espacios y la actualización de las propias redes de contactos.

De acuerdo con Ribes (2007), estas actividades pueden dividirse en dos grandes grupos: la producción de contenidos, por un lado, y la optimización de recursos, por otro. Probablemente, los ejemplos más claros de producción y edición colaborativa son la enciclopedia en línea Wikipedia y el proyecto lexicográfico asociado, Wiktionary (en inglés), plataformas que se rigen por modelos de distribución y propiedad intelectual libre (gnu Free Documentation License y Creative Commons<sup>3</sup>) y que emplean *software* abierto para el diseño y el mantenimiento del sistema, apoyándose en MediaWiki, una aplicación basada en PHP y MySQL que se implantó a inicios del 2000 y que ha sido mejorada en los últimos años. Ambos proyectos se basan en que un artículo sobre un determinado tema o una palabra y su definición pueden ser creados por cualquier usuario de la web, publicarse en cada una de esas plataformas y ser editados, esto es, modificados por cualquier otro usuario en el mismo espacio. Se trata de una práctica que aplica para la creación y mejora de contenidos la máxima de Raymond (1999) so-

<sup>2</sup> Coincidimos con Chiari (2012) en que proyectos de lexicografía colaborativa como Wiktionary o Wikipedia, por ejemplo, no se han analizado aún adecuadamente desde estos enfoques.

<sup>3</sup> <<http://www.gnu.org/copyleft/fdl.html>> (10/2016).

bre el *software* libre, o la llamada *Ley de Linus*, según la cual: «A la vista de suficientes ojos, todos los errores resultan evidentes»<sup>4</sup>.

A pesar de que en estos espacios la producción textual en colaboración representa un tipo de discurso en el que aparentemente no se observan diferencias entre las distintas aportaciones, los usuarios, cuando editan, suelen interactuar y negociar los contenidos que se publican dejando constancia de dicha interacción en los llamados *historiales*, en los que aparecen los comentarios y actualizaciones de lo que ha hecho cada uno<sup>5</sup>. Por otra parte, un grupo de editores, votados por el gran número de ediciones que realizan, revisan las enmiendas y propuestas de inclusión de nuevos artículos. Estos usuarios están capacitados para modificar los artículos, conservarlos total o parcialmente y controlar su veracidad en función de las políticas que guían el proyecto, a saber, la redacción de los artículos desde un punto de vista neutral y la inclusión en ellos de referencias fiables a las fuentes de la información publicada (Colomer, Ramos 2012). Huvila (2010) señala que quienes más contribuyen a la edición de artículos en Wikipedia suelen ser investigadores (*investigators*), que se apoyan en fuentes científicas de carácter divulgativo; académicos (*scholars*), que en cierto modo reutilizan los contenidos de sus propias investigaciones; y, por último, los editores (*editors*), mencionados previamente, que se ocupan sobre todo de los aspectos de gestión, corrección y traducción de los artículos incluidos en la enciclopedia en línea. Los datos recogidos por Guiote (2011), por medio de entrevistas, señalan que los libros y revistas especializados son las fuentes más utilizadas para documentar los contenidos; les siguen las webs oficiales, las enciclopedias tradicionales y la experiencia personal.

En cuanto a su uso por parte de traductores, Alonso (2014) destaca que estos utilizan Wikipedia con distintos propósitos durante el proceso de traducción: documentales, terminológicos, para identificar fuentes sobre un determinado tema, visualizar las imágenes asociadas a un término y negociar las soluciones de traducción con terceros (i.e. clientes). Asimismo, los traductores entrevistados destacan la utilidad de Wikipedia para comprender mejor el texto de origen o alguno de sus elementos, gracias a la visualización de imágenes contenidas en los artículos. Estos valoran positivamente también su fácil acceso, el hecho de estar siempre disponible, el alto grado de actualización de muchos de los artículos, su extensa cobertura temática y su carácter multilingüe. Sin embargo, aluden a la necesidad de contrastar la solución que extraen con otras fuentes consideradas más

<sup>4</sup> No obstante, Feller *et al.* (2005) advierten de que en este tipo de proyectos no siempre es seguro que el código se revise de manera exhaustiva ni que la revisión por pares sea más eficaz que otros procedimientos de detección de errores.

<sup>5</sup> La Fundación Wikimedia mantiene el historial público de todas las acciones realizadas en sus proyectos. En general, es posible acceder a él a través de las estadísticas de los mismos: <<http://stats.wikimedia.org>> (09/2016).

fiables, pues son muy conscientes de los límites relativos a la fiabilidad de los artículos; además, señalan como aspectos críticos la extensión a veces excesiva de los artículos, la calidad de los mismos en determinados campos y el hecho de que sean, en muchas ocasiones, traducciones no declaradas de otros artículos.

### 3. Wikipedia: aplicaciones para diccionarios y ontologías

Wikipedia se presenta como una enciclopedia alfabética. En cuanto a su estructura, cada página web corresponde a un artículo que contiene, en la parte superior del mismo, un apartado de 'Discusión' sobre su contenido, un espacio al que se accede para 'Editar' el artículo y un apartado 'Historial', donde es posible consultar los cambios realizados.

Los artículos, clasificados mediante distintas 'categorías', se estructuran en títulos y subtítulos, párrafos generalmente breves, un índice que se genera automáticamente una vez superadas las tres secciones o epígrafes, una serie de referencias bibliográficas, fuentes recomendadas como lecturas adicionales y un conjunto variable de enlaces externos, lo que refuerza su dimensión hipertextual. La mayor parte de los artículos combina texto, imágenes (fotografías, reproducción de pinturas, mapas, grabados, dibujos, etc.) y enlaces a muy distintos documentos (texto, audio, video, sitios web), que los vinculan generalmente a contenidos de la web externos a la enciclopedia.

Wikipedia ha sido analizada desde diferentes enfoques y con diversos objetivos (cfr. Mesgari *et al.* 2015), entre los que destacan la extracción de información léxica y conceptual para la elaboración de diccionarios y ontologías (Milne *et al.* 2006; Zesch 2010; Müller, Gurevych 2009; Erdmann *et al.* 2008, Oliver, Climent 2012; entre otros) y el análisis de distintos elementos de las páginas de Wikipedia. Por ejemplo, Spandana *et al.* (2014) utilizan las 'categorías', entendidas como dominios de especialidad, con el objeto de aplicarlas a proyectos de carácter lexicográfico existentes, como WordNet, para mejorarlos. La comparación llevada a cabo por estos autores entre las categorías de Wikipedia y los dominios de WordNet para inglés, francés, alemán, italiano, portugués y español muestra que al menos el 85% de los dominios de WordNet presentan una correspondencia con las categorías de Wikipedia para cada uno de los idiomas considerados. No obstante, cabe señalar que el sistema de categorías que utiliza Wikipedia, diseñado para encontrar y navegar por grupos de artículos relacionados, no se utiliza adecuadamente como mecanismo de acceso a otros contenidos, debido quizá a una falta de sensibilidad o de formación de los editores a la hora de asignar categorías a los artículos (Zazo *et al.* 2015).

Por su parte, Tyers y Pienaar (2008) plantean la elaboración de diccionarios bilingües mediante los llamados *interwikis* o enlaces interlingüísticos – que permiten pasar del artículo escrito en una lengua a un artículo equi-

valente en otra –, a través de la recopilación de lista de palabras en cuatro lenguas de origen y el uso de los enlaces asociados, con el objeto de encontrar su equivalente en cuatro lenguas de llegada. Vivaldi y Rodríguez (2011) proponen en cambio un método para la extracción de información terminológica en español e inglés relativa a la Medicina, la Economía y la Química, trabajando con candidatos a términos (TC) de esos dominios extraídos de Wikipedia. En los dos trabajos mencionados, los resultados confirman que esta es una fuente de información útil para la obtención de términos relacionados con un dominio determinado, aunque en el último estudio se señala la necesidad de confrontar los candidatos a términos con corpus de lengua de especialidad y listas de términos extraídas de los mismos.

Otros proyectos señalables son el de Erdmann *et al.* (2008), quienes utilizan los enlaces entre artículos, el sistema de redirección de páginas de Wikipedia y los enlaces textuales en el interior de los artículos, con el fin de crear un diccionario bilingüe (inglés-japonés) de términos especializados; y el de Goldman y Scherrer (2012), dedicado a extraer correspondencias léxicas bilingües de entidades nombradas de cara a la traducción. Después de descargar los ficheros de archivo de Wikipedia y de filtrar las páginas de redirección y desambiguación en dos direcciones (inglés↔francés), estos investigadores obtuvieron un 57% de términos equivalentes entre las dos versiones. Pero, como en el caso precedente, para evaluar la adecuación de las listas de correspondencias obtenidas, los autores sugieren compararlas con corpus textuales externos para cada idioma.

#### **4. Presencia y tratamiento de *tempera* y *guazzo* en Wikipedia**

Tras esta rápida panorámica de las características y aplicaciones de Wikipedia como recurso útil para traductores especializados y como fuente disponible para la extracción automática de datos de cara a la elaboración o mejora de recursos lexicográficos y terminológicos existentes, llevamos a cabo un análisis cualitativo de algunos contenidos de este recurso con el objeto de profundizar en las oportunidades y límites indicados en estudios previos. Realizamos una revisión crítica y comparada de los artículos en italiano de los términos *tempera* y *guazzo* y de sus correspondientes en español. Sin pretensión de exhaustividad y siendo conscientes de la necesidad de ampliar el análisis a otros términos, para los fines explorativos de este trabajo se han seleccionado estas voces propias de la pintura siguiendo los planteamientos de Farina (2015) y Carpi (2016), cuyos análisis surgen de la reflexión de los equipos de investigación de español y francés del proyecto LBC, al constatar que: «[...] dichos términos presentan características polisémicas y un alto grado de variación semántica que se manifiesta a lo largo de los siglos» (Carpi 2016: 113-114). Ambos plantean, pues, cuestiones que pueden enriquecer el análisis de los artículos presentes en la enciclopedia en línea sobre todo en una perspectiva contrastiva.

Las pautas seguidas por estas autoras para el análisis de los términos *tempera* y *guazzo* y para obtener posibles equivalencias de traducción en español y francés, han sido, principalmente, la consulta de recursos lexicográficos monolingües y bilingües, desde las perspectivas diacrónica y sincrónica, así como la exploración del uso de estos términos en textos especializados en la pintura. En concreto, para el español, Carpi ha consultado también los diccionarios incluidos en los *Tesoros del Patrimonio Cultural de España* (TPCE), la base de datos Corpes XXI y la biblioteca digital Google Books. Asimismo, para verificar la validez de las propuestas de traducción, utiliza las páginas web de los *Musei Vaticani*, del *Polo Museale Fiorentino* y del Museo del Prado. Reproducimos aquí una tabla que recoge los principales resultados del análisis de esta investigadora (2016), con las acepciones y posibles equivalentes de traducción de *tempera* y *guazzo* en español:

Tabla 1 – *Tempera y guazzo, acepciones y posibles traductores en español* (Carpi 2016: 122).

IT	acepción	aglutinante	ES
tempera	Solución colorante, técnica, cuadro pintado con esta técnica	agua y todo tipo de aglutinante (hasta el siglo XVII)	temple
tempera	Solución colorante, técnica, cuadro pintado con esta técnica	agua y aglutinante de huevo (animal)	temple
tempera	Solución colorante, técnica, cuadro pintado con esta técnica	agua y aglutinante vegetal	aguada/ gouache
tempera	Solución colorante de tipo industrial en tubos o en tarros	caseína (aglutinante animal)	témpera
guazzo	Solución colorante, técnica, cuadro pintado con esta técnica	agua y aglutinante vegetal	aguada
gouache	Solución colorante, técnica, cuadro pintado con esta técnica	agua y aglutinante vegetal (mod.)	aguada/ gouache

Estos datos nos servirán de base para el presente análisis. Nos centramos, en concreto, en el título del artículo, el breve texto de introducción o de definición del término, el índice de contenidos y las referencias y enlaces externos de los artículos en italiano y en español, relativos a estos dos términos de la pintura. Se busca comparar así las distintas secciones de los artículos en cuestión para ver en qué medida esta información puede considerarse equivalente a la que se obtiene mediante la consulta de diccionarios y de textos especializados, con el fin de dar respuesta a la pregunta: ¿puede la información proporcionada por Wikipedia integrarse en bases de datos sobre léxico especializado?

#### 4.1 *Tempera* / ‘Temple’

Si se busca el término *tempera* (IT) en la versión italiana de Wikipedia, el usuario accede a una página de desambiguación que remite, entre otras opciones, a *pittura a tempera*. A partir de este artículo y siguiendo el enlace interlingüístico ‘Español’ del menú de la izquierda, es posible acceder al artículo ‘Pintura al temple’ (ES)<sup>6</sup>. Enseguida se constata que existe una relación directa entre estos dos artículos y, en lo que se refiere al título, que existe una correspondencia exacta entre los dos sintagmas *pittura a tempera* y ‘pintura al temple’, que establece la equivalencia entre ambos términos, indicando, además, el dominio de especialización al que pertenecen *tempera* y ‘temple’, esto es, la pintura. Como todos los artículos de Wikipedia, estos presentan un primer párrafo que proporciona una breve definición o descripción del término. En este caso, los artículos de las dos versiones difieren en lo que respecta a la extensión, pero presentan un contenido similar que se centra en la pintura al temple como técnica y sobre todo en los tipos de aglutinante que se utilizan para mezclar los pigmentos en el agua. Como es posible observar en la tabla comparativa (Tabla 2) que sigue, gran parte de estos datos se enlazan con otras entradas en Wikipedia (subrayado en el texto), lo que permite al usuario acceder a otros datos, por ejemplo, sobre los aglutinantes utilizados, y en el caso del artículo en español, sobre los estilos artísticos en los que se utiliza esta técnica.

Comparando ambos artículos con los datos recogidos en la Tabla 1 (Carpi 2016: 122), vemos que los dos párrafos contienen dos de las tres acepciones señaladas, esto es, que se trata de una ‘solución colorante’ y de una ‘técnica’, mientras que se deja de lado la tercera acepción: ‘cuadro pintado con esta técnica’, aunque el sintagma ‘pintura al temple’ parece contener ya dicho significado si se atiende a la polisemia que presenta la palabra ‘pintura’<sup>7</sup>. Asimismo, se indican los tipos de aglutinante, aunque esto se hace de forma un tanto desordenada, lo que puede desorientar al lector.

En la Tabla 1, ‘temple’ se relaciona con los tipos de aglutinante: «agua, y todo tipo de aglutinante hasta el siglo XVII» y «agua y aglutinante de huevo (animal)». En el primer párrafo relativo al término en español vemos que se ofrecen, por un lado, datos sobre los tipos de aglutinante animal (e implícitamente, vegetal); por otro lado, un dato histórico sobre el uso de esta técnica en la pintura medieval y renacentista, lo que podría relacionarse con la indicación temporal de la segunda acepción de ‘temple’: «hasta el siglo XVII». En el artículo en italiano no se ofrece en cambio información histó-

<sup>6</sup> <[https://it.wikipedia.org/wiki/Pittura\\_a\\_tempera](https://it.wikipedia.org/wiki/Pittura_a_tempera)> (09/2016).

<[https://es.wikipedia.org/wiki/Pintura\\_al\\_temple](https://es.wikipedia.org/wiki/Pintura_al_temple)> (09/2016). Los términos subrayados en los párrafos corresponden a los enlaces presentes en la introducción de cada artículo.

<sup>7</sup> Definición «Pintura: 1. f. Arte de pintar; 2. f. Tabla, lámina o lienzo en que está pintado algo.; 3. f. Obra pintada; 5. f. Acción de pintar. Procedieron a la pintura de la pared» (*Diccionario de la Real Academia Española*, 2014, 23ª ed.).

rica, pero el texto es mucho más detallado acerca de los tipos de aglutinante, que incluyen explícitamente los de tipo animal y vegetal. Sin embargo, estos datos se mezclan en un texto en el que curiosamente no se menciona el término *agglutinante*, por lo que se pierde e incluso se confunde la acción de estos materiales. Por otro lado, en el texto italiano se ofrece información especializada relativa a la preparación del soporte cuando se utiliza esta técnica, un aspecto que no recoge la introducción del artículo en español.

Tabla 2 – Primer párrafo de los artículos ‘*Pittura a tempera*’ y ‘*Pittura al temple*’ en Wikipedia.

<u>Pittura a tempera</u>	<u>Pittura al temple</u>
<p>La <u>pittura a tempera</u> (dal latino Temperare – mescolare) è una tecnica che si avvale dell’uso di un colore preparato mescolando <u>pigmenti</u> in polvere con un legante formato da un’<u>emulsione</u> in fase acquosa (parti oleose in minoranza che ‘nuotano’ sospese in forma di piccole gocce nell’acqua).            Le parti solubili in acqua dell’emulsione che costituiscono il <u>colore a tempera</u> sono <u>uovo</u>, <u>caseina</u>, colle animali, gomme vegetali, amidi; quelle insolubili sono olii (<u>olio di lino</u>, di noce, di papavero, anche in forma pre-polimerizzata), resine, cere, lacche. Per quanto riguarda l’uovo viene utilizzato intero, oppure solo il tuorlo o solo il bianco.            I supporti utilizzati, analogamente alla tecnica ad olio e ad altre tecniche, vengono preparati con una <u>imprimitura</u> che impegna entrambi i lati del supporto per evitare deformazioni causate da un differente grado di <u>umidità</u> tra le due superfici.</p>	<p>La <u>pittura al temple</u>, también conocida como <u>témpera</u>, es una técnica de pintura en la que el <u>disolvente</u> del <u>pigmento</u> es el agua y el <u>aglutinante</u> (también denominado temple o engrosador) algún tipo de grasa animal, <u>glicerina</u>, huevo, <u>caseína</u>, otras materias orgánicas o goma. Históricamente, la <u>pittura al temple</u> es característica de la <u>Edad Media</u> europea. Puede considerarse característico de los estilos <u>románico</u> y <u>gótico</u> en el <u>occidente europeo</u>, y de los <u>iconos bizantinos</u> y <u>ortodoxos</u>, en <u>Europa Oriental</u>.</p>

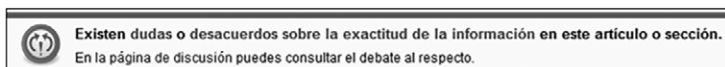
Un elemento más problemático si cabe, en el caso de querer establecer una equivalencia entre estos dos artículos para la extracción de datos similares, es el de la mención en el texto en italiano del sintagma *colore a tempera*, que parece presentarse como sinónimo de *pittura a tempera*, pero que en este caso remite no a ‘técnica’, segunda acepción recogida en la Tabla 1, sino a ‘solución colorante’, esto es, la primera acepción. La inclusión de este sintagma en la misma definición de *pittura a tempera* puede confundir al usuario que busca un significado unívoco para uno u otro término.

De forma similar, la indicación de ‘témpera’ como sinónimo de ‘pittura al temple’ en la primera línea del texto en español puede confundir a quien consulta el artículo. El término ‘témpera’ se recoge en la Tabla 1 como equivalente de *tempera* (IT), es decir, «solución colorante de tipo industrial en



tubos o en tarros», cuyo aglutinante principal es la caseína; esta cuarta acepción del término y el significado específico asociado a la solución industrial se pierden al presentarse ambas palabras como sinónimas. En este sentido, es interesante consultar la página de discusión<sup>8</sup> del artículo en cuestión. De hecho, en la parte superior del mismo aparece una advertencia:

Figura 1 – Advertencia sobre posibles desacuerdos en torno a «Pintura al temple».



Por ejemplo, en uno de los comentarios se destaca que:

[...] al poner tempera [sic] en el buscador me redirecciona aquí, puede que esté mal informada pero hasta donde yo sé, la tempera y la pintura al temple no son lo mismo, ¿o es la ténpera un tipo de pintura al temple? [...] – 15:32 4 jun 2011.

Este comentario ilustra bien la confusión que puede plantear la consideración de estos dos términos como sinónimos cuando se busca una definición lo más precisa posible en los artículos dedicados a la pintura y a sus técnicas y materiales. En el comentario que sigue se señalan los puntos críticos del artículo, como pueden ser las nomenclaturas y definiciones. Cabe notar en este caso la alusión a la necesidad de consultar o contrastar los datos con tratados y estudios especializados:

La entrada que se refiere a la pintura al temple reclama mi atención porque resulta muy desordenada al tiempo que entra en incorrecciones de base y fundamento, tanto de nomenclaturas como de definiciones y fórmulas de elaboración y sobre todo, maneras de aplicación, seguramente producto de un desconocimiento práctico y de una escasa consulta de tratados y otros estudios elaborados a propósito [...] 19:08 28 dic 2014.

Si analizamos, por otra parte, los índices que contienen los dos artículos de Wikipedia objeto del presente análisis – ambos superan los tres

<sup>8</sup> En la parte superior de la página, modificada por última vez el 29 de noviembre de 2015, se lee: «Esta página se encuentra en el área de interés del Wikiproyecto Pintura, un esfuerzo colaborativo cuyo fin es mejorar y organizar todos los contenidos relacionados con ese tema. Puedes visitar la página de discusión del proyecto para colaborar y hacer preguntas o sugerencias». Se trata de un proyecto que contiene exclusivamente artículos sobre pintores, corrientes pictóricas y técnicas y que busca ofrecer información especializada sobre estos temas fomentando, en primer lugar, la colaboración entre expertos. Para demostrar la utilidad del método colaborativo de mejora de los artículos, se afirma: «2009: En un intenso trabajo colectivo se mejoraron 23 artículos de los más visitados y de los esenciales de pintura. Estos 23 artículos mejorados reciben ahora 990.942 visitas mensuales. Antes de la mejora recibían 551.734 visitas al mes».

epígrafes o secciones –, notamos una diferencia en cuanto al número de ítems, el orden de presentación de los mismos y su contenido, lo que puede explicar en parte las diferencias señaladas en el párrafo de inicio:

Tabla 3 – Índice de contenidos de los artículos ‘Pittura a tempera’ y ‘Pittura al temple’ en Wikipedia.

<b>Pittura a tempera</b>	<b>Pittura al temple</b>
1 Definizione dei diversi tipi di tempera	1 Historia
2 Cenni storici	2 Técnica
3 Bibliografia	2.1 Temple de huevo
4 Altri progetti	2.2 Témpera grassa
5 Collegamenti esterni	2.3 Pigmentos
	2.4 Aplicación
	2.5 Soporte
	3 Artistas
	4 Galería
	5 Véase también
	6 Referencias
	7 Bibliografía
	8 Enlaces externos

De entrada, observamos que en el artículo en italiano, las cuestiones históricas se tratan en una sección 2. Cenni storici y se amplían a través de enlaces en 4. Altri progetti. En el español, el índice parece más elaborado en la medida en que se dedica un espacio específico a las técnicas de pigmentación, desarrollando así los pocos datos ofrecidos en la introducción, así como al soporte utilizado (2.5 Soporte); se incluyen, además, dos secciones: 3. Artistas y 4. Galería, que no aparecen en el artículo correspondiente italiano. Este tipo de información, que enlaza distintas partes del mismo documento y de otros textos en Wikipedia, permite profundizar en el tema al ampliar el alcance semántico del término y el conocimiento enciclopédico asociado a él. No obstante, y esto dificulta también la extracción de datos en una perspectiva bilingüe, la desigualdad de contenidos entre uno y otro artículo obliga al usuario que busca la equivalencia más adecuada para *tempera*, navegar por numerosas páginas para llegar después a los mismos resultados.

Por ejemplo, en base a la búsqueda lexicográfica llevada a cabo por Carpi (2016), la tercera acepción de *tempera* / ‘temple’ es «solución colorante, técnica», cuyo aglutinante es «agua y aglutinante vegetal». Su equivalente en español sería ‘aguada’ o ‘gouache’, un término que se menciona en el artículo ‘Pittura al temple’ pero solo en la sección ‘Véase también’, en la parte inferior de la página en forma de enlace a otro artículo: ‘Gouache’ (ES)<sup>9</sup>, cuando podría aparecer en la descripción inicial.

En este otro artículo se lee el siguiente párrafo introductorio:

<sup>9</sup> <<https://es.wikipedia.org/wiki/Gouache>> (09/2016).

El *gouache* o aguada (del galicismo *gouache* que puede derivar del italianismo *guazzo*), es color diluido en agua sola, o con diversos ingredientes, como goma, miel, etc. También es la pintura realizada con esta técnica [...].

El texto proporciona información de carácter etimológico y pone en relación el francés *gouache* y el italiano *guazzo*, un dato que puede ser útil para el traductor; además, indica que se trata de una solución colorante que utiliza el agua o el aglutinante vegetal como la goma, un dato importante que no se incluía de forma explícita en ‘Pintura al temple’. Destaca aquí también la mención a una de las acepciones de *guazzo* y *gouache*, esto es, «cuadro pintado con esta técnica» (v. Tabla 1), que no se indicaba tampoco en el primer párrafo de ‘Pittura a tempera’ y ‘Pintura al temple’, siendo este, sin embargo, uno de los significados de *tempera*.

En cuanto a las referencias contenidas en estos dos artículos, se observa que en el italiano *pittura a tempera*, la Bibliografía remite curiosamente a dos textos especializados en alemán y a uno en inglés, lo que lleva a pensar en una posible traducción del artículo a partir de sus correspondientes alemán e inglés en la enciclopedia en línea.

Tabla 4 – Referencias y enlaces en los artículos ‘Pittura a tempera’ y ‘Pintura al temple’ en Wikipedia.

---

– Max Doerner – Malmaterial und seine Verwendung im Bilde, achzente Auflage, F. Enke Verlag, Stuttgart 1994.	– DDAA (1984). <i>Historia Universal de Arte: Volumen 8</i> . Madrid, SARPE.
– Markus Stegmann, René Zey – Lexikon der Modernen Kunst, Techniken und Stile, Verlag Gruner Jahr AG & Co., 2001.	<a href="#">Wikimedia Commons</a> alberga contenido multimedia sobre <b><u>Pintura al temple</u></b> .
– Rutherford J. Gettens, George L. Stout – Painting Materials, A short Encyclopedia, Dover Publications, Inc., New York 1966.	<a href="#">¿Qué es la pintura al temple?</a>

---

[Wikimedia Commons](#) contiene immagini o altri file su *pittura a tempera*  
*Pittura a tempera*, in *Tesaurus del Nuovo soggettario*, BNCF, marzo 2013.

---

Sin embargo, tras consultar los enlaces interlingüísticos IT-DE e IT-EN y las referencias contenidas en ‘Tempera’ (DE) y ‘Tempera’ (EN), se constata que solo una de ellas coincide con las del artículo en italiano, Doerner (1994). Este dato nos lleva a reflexionar sobre el peso que pueden tener estas referencias en la elaboración del artículo, de qué manera se han utilizado para redactar y/o contrastar los datos que se ofrecen y por qué se han seleccionado estas y no otras fuentes, por ejemplo, en italiano.

En cuanto al artículo en español, la referencia a la *Historia Universal del Arte* (1984) lleva a pensar en cambio que se trata de un artículo redactado

originalmente en español y que su autor o autores han consultado un texto de carácter enciclopédico, más apropiado para esta labor.

Es innegable que uno de los datos más útiles para quien consulta estos artículos es el conjunto de enlaces que aparecen bajo las referencias al proyecto Wikimedia Commons<sup>10</sup>. Mediante estos enlaces, el usuario puede acceder a un gran número de imágenes relativas, en este caso, a *tempera* y ‘temple’, que, como se apuntaba más arriba, facilitan la comprensión de un término específico, utilizado por ejemplo en un texto especializado en pintura que se esté traduciendo, y ello gracias a la visualización de dichas imágenes, que pueden aparecer también en el artículo en la sección ‘Galería’.

Por último, cabe destacar en el artículo de Wikipedia en italiano el enlace al *Tesaurus del Nuovo Sogettario* (2013) de la Universidad de Florencia<sup>11</sup>, en el que se distinguen claramente los términos *pittura a tempera* y *tempere*, lo que también permite al usuario verificar las acepciones del término y resolver las posibles dudas surgidas en la lectura del párrafo introductorio. En el artículo en español, el enlace ‘¿Qué es la pintura al temple?’ reenvía a un vídeo en YouTube en el que se explica detalladamente la técnica. Como se observa, el carácter multimodal de la enciclopedia en línea facilita la comprensión del término y los elementos que lo definen, a pesar también de que la aportación de más datos de los necesarios puede desorientar al usuario.

#### 4.2 Guazzo y ‘aguada’ o ‘gouache’

En lo que respecta a *guazzo* cabe señalar aquí también la existencia de un enlace directo entre los artículos ‘Guazzo’ (IT) y ‘Gouache’ (ES). Atendiendo a los equivalentes en español recogidos en la Tabla 1 para *guazzo*, es posible afirmar que los interwikis o enlaces interlingüísticos permiten identificar fácilmente los equivalentes de traducción, al menos entre este par de lenguas; se trata de datos que pueden extraerse automáticamente para poder elaborar recursos lexicográficos bilingües, aunque, como apuntaban los estudios mencionados, cabe adoptar ciertas precauciones.

Los textos de introducción de ambos artículos aportan información etimológica, estableciendo una correspondencia entre los términos en italiano y español, y entre estos y el francés *gouache*, correspondencia que documentan Farina (2015) y Carpi (2016).

<sup>10</sup> También conocido como WikiCommons o Commons es un proyecto de la Fundación Wikimedia que se creó el 7 de septiembre de 2004; el objetivo es constituir un gran archivo de imágenes y otros formatos multimedia para otros proyectos, como Wikipedia. Sus archivos deben haber sido liberados antes bajo licencias libres.

<sup>11</sup> <<http://thes.bncf.firenze.sbn.it/termine.php?id=27682>> (10/2016).

Tabla 5 – Primer párrafo de los artículos ‘Guazzo’ y ‘Gouache’ en Wikipedia.

Guazzo	Gouache
<p>Il guazzo, noto anche nella forma <u>france-se gouache</u>, è un tipo di <u>colore a tempera</u> reso più pesante e opaco con l'aggiunta di un <u>pigmento bianco</u> (per esempio <u>biacca</u> o gesso) mescolato con la <u>gomma arabica</u> (un tempo era preferita la <u>gomma adragante</u>). Il risultato è appunto un colore più coprente e più luminoso rispetto al normale colore a tempera.</p> <p>Il termine può anche indicare sia la tecnica di pittura che i dipinti eseguiti con questo tipo di colore.</p>	<p>El <i>gouache</i> o aguada (del <u>galicismo</u> <i>gouache</i> que puede derivar del <u>italianismo</u> <i>guazzo</i>), es color diluido en agua sola, o con diversos ingredientes, como goma, miel, etc. También es la pintura realizada con esta técnica. Es diferente de la <u>acuarela</u>, pues los colores son opacos.</p>

En el artículo en italiano, el término *guazzo* se relaciona explícitamente con *tempera* a través del sintagma «un tipo di colore a tempera», de forma que se conecta con el significado de este término y con parte de la información contenida en el artículo relativo al mismo, que hemos explorado previamente. El texto contiene, además, un enlace a otro artículo<sup>12</sup>, más breve, en el que se da una definición de *colore a tempera* y se amplía la información sobre esta técnica, su disponibilidad en el mercado e incluso su aplicación sobre ciertos soportes desde un punto de vista histórico.

Por otra parte, el artículo en español, que ya recoge el galicismo en el título, menciona en su introducción el término equivalente ‘aguada’<sup>13</sup>, que define como «color diluido en agua sola, o con diversos ingredientes [...]». De forma diferente estos artículos remiten a un tipo de color que se obtiene mezclando distintos «ingredientes», como se indica en el artículo, aunque sería más adecuado utilizar ‘aglutinantes’; sin embargo, solamente el artículo en español indica que el principal elemento para poder hablar de ‘gouache’ o ‘aguada’ es el agua. Curiosamente, el texto en italiano no menciona este elemento y habla de un «colore più coprente e più luminoso rispetto al colore a tempera», un dato que puede resultar enigmático para el usuario, que está obligado a consultar el artículo ‘Colore a tempera’ mediante el enlace, si quiere obtener datos más precisos y fiables.

En este caso, ambos textos indican que se trata no solo de una técnica sino también de la pintura realizada con la misma, e incluyen datos útiles sobre los «ingredientes» utilizados en su preparación, entre otros, la goma; nótese, además, cómo en italiano se especifica el tipo de goma: *arabica* o *adragante*, y se señala la opacidad como rasgo principal de este *colore a tem-*

<sup>12</sup> <[https://it.wikipedia.org/wiki/Colore\\_a\\_tempera](https://it.wikipedia.org/wiki/Colore_a_tempera)> (10/2016).

<sup>13</sup> En Wikipedia en español existe el artículo ‘Aguada’, pero solo se llega a él directamente a través del motor de búsqueda de la plataforma y no a través de interwikis: «La aguada es una técnica pictórica que consiste en mezclar en distintas cantidades de [sic] agua o alcohol con diversas tintas» <<https://es.wikipedia.org/wiki/Aguada>> (06/2017).

*pera*. En el artículo español se señala correctamente la diferencia entre esta solución colorante y la ‘acuarela’, debida precisamente a la opacidad de los colores, un elemento importante para distinguir la aguada de otro tipo de técnica en la que se utilice el agua.

En lo que se refiere al índice, el artículo en español presenta, aparentemente, un mayor número de contenidos y algunos datos que no recoge el texto en italiano, como la información sobre las superficies a las que se aplica la aguada o gouache, o incluso sobre el tipo de pinturas y el equipo necesario para realizar obras con esta técnica.

Tabla 6 – Índice de contenidos de los artículos ‘Guazzo’ y ‘Gouache’ en Wikipedia.

1 Etimología	1 Historia
2 Storia	2 Características
3 Tecnica	3 Superficies
4 Note	3.1 Papel
5 Voci correlate	3.2 Tableros
6 Altri progetti	4 Pinturas y equipo
7 Collegamenti esterni	4.1 Gouache
	4.2 Colores de póster y en polvo
	4.3 Pinceles y caballetes
	5 Actualidad
	6 Referencias
	7 Bibliografía
	8 Enlaces externos

El artículo en italiano presenta una estructura con un número menor de ítems y en un par de casos estos son diferentes respecto al artículo de la enciclopedia en español, por lo que no es posible considerar los índices como textos equivalentes. En ambos artículos se dedica espacio a la historia de esta técnica y a las cuestiones relativas a los aglutinantes, aspecto que el primero recoge en el más específico ‘3. Tecnica’ y el segundo, en el apartado menos intuitivo ‘2. Características’. Mientras que en el texto en italiano se ofrece información sobre la etimología de *guazzo*, un dato que se incluye brevemente en la introducción del artículo en español, en este se informa acerca de las superficies y equipos utilizados con esta técnica, aspectos que no se señalan en el artículo ‘Guazzo’.

Los datos recogidos en estos dos artículos permiten entender y situar ambos términos dentro del ámbito de la pintura; en ellos se ofrece información especializada sobre la etimología del término, la historia de la técnica, los materiales que se emplean para poder trabajar con ella, las superficies y los equipos. Si bien parte de esta información parte de esta información remite a otros artículos y a otros términos – de forma que se obliga al usuario a consultar otras fuentes –, se trata de datos que pueden ser reutilizados para enriquecer la información léxica sobre estos términos en recursos lexicográficos y terminológicos monolingües. Resulta más difícil extraer esta

información para implementar diccionarios y glosarios bilingües sobre la pintura, debido a las numerosas discordancias observadas entre las definiciones y las estructuras de ambos artículos.

En este sentido cabe señalar también las diferencias notables que presentan en los apartados ‘Collegamenti esterni’ (IT) y ‘Bibliografía’ y ‘Enlaces externos’ (ES).

Tabla 7 – Referencias y enlaces en los artículos ‘Guazzo’ y ‘Gouache’ en Wikipedia.

<p>– Definizione su <i>Il dizionario della lingua italiana De Mauro</i>, Pearson Paravia Bruno Mondadori, 1999-2008. URL consultato il 18 maggio 2009, ma ora non più attivo.</p>	<p>– Hayes, Colin. (1980). Guía completa de pintura y dibujo, técnicas y materiales. España: Hermann Blume Ediciones. ISBN: 84-7214-275-2</p>
<p>– Bruno Pierozzi, <i>Manuale della tempera, inforestauro</i>. URL consultato il 23 marzo 2012.</p>	<p>0.1. <b>Enlaces externos</b>  <a href="#">Wikimedia Commons</a> alberga contenido multimedia sobre <b>Gouache</b>.          Colección de Gouaches digitalizados en la <a href="#">Biblioteca Digital Hispánica</a> de la Biblioteca Nacional de España</p>

Como se observa en la Tabla 7, la sección relativa a enlaces externos en el artículo en italiano remite en realidad a una fuente lexicográfica: *Il Dizionario della lingua italiana* de Tullio De Mauro, y a un texto especializado, *Manuale della tempera*, a los que se puede acceder mediante sendos enlaces, de ahí que se hayan insertado en la sección de enlaces, que es fija en todos los artículos de Wikipedia. En el artículo en español se proporcionan quizá de forma más correcta, por un lado, una fuente bibliográfica completa que corresponde a un texto especializado: una guía sobre pintura y dibujo; por otro lado, una serie de enlaces externos. El artículo no menciona, sin embargo, fuentes lexicográficas específicas, por lo que no es posible identificar claramente la fuente de la definición proporcionada en la introducción del mismo. Sí se incluyen enlaces a Wikimedia Commons y a la Biblioteca Digital Hispánica de la Biblioteca Nacional de España, que remiten a una serie de contenidos multimedia, principalmente: imágenes de obras realizadas con la técnica del *gouache* o aguada, y que permiten al usuario comprobar de qué tipo de técnica se trata. Como ya se apuntaba en relación con *tempera* y ‘temple’, estos son recursos útiles para completar la definición y descripción proporcionadas en el artículo, aunque parece necesario confrontar este tipo de información con otras fuentes, léxicográficas o enciclopédicas, y con manuales y textos especializados que puedan cubrir distintas épocas.

Por último, si se consideran las páginas de discusión de estos dos artículos, notamos que la página del artículo en italiano ha sido eliminada el 5 de enero de 2008 por la aparente poca utilidad de los comentarios: «cancello discussioni prive di info». En el español, algunos mensajes permiten entender los motivos de las correcciones aportadas: «Entre las correcciones que he aportado quiero indicar las dos siguientes: la palabra *gouache* se escribe

siempre en cursiva (consúltese *gouache* en el DRAE)», mientras que la mayoría señala la necesidad de precisar el alcance semántico del término relacionándolo con otros términos del mismo ámbito, a la vez que se critica la tendencia, en algunos casos, a la confusión por similitud:

En este artículo se dice que guache [sic] equivale a tempera y hay otro artículo sobre esta (que no hace referencia a guache).

**Pero sucede que la t mpera (o pintura al temple) es una t cnica pict rica muy diferente de la aguada (gouache o guazzo), esto se nota a primera vista cuando se mira una t mpera o cuando se mira una aguada/gouache (no espa olizo de un traslado el galicismo porque mundialmente, incluso en los pa ses hispan fonos, se usa preferentemente entre los artistas el galicismo *gouache*). En una enciclopedia inteligente se requiere el m ximo an lisis posible (a n cuando se busque dial cticamente la s ntesis), fusionar (fundir) por meras similitudes es confundir.** 02:24 20 feb 2011 [ nfasis nuestra].

## 5. A modo de conclusi n

Las herramientas 2.0 como Wikipedia presentan no pocas oportunidades para el estudio y divulgaci n del l xico especializado, sobre todo en una  ptica de optimizaci n de recursos como puede ser, por ejemplo, la extracci n autom tica de informaci n multiling e para elaborar bases de datos terminol gicas. El inter s del investigador en este tipo de herramienta reside en su creciente uso por parte de redactores y traductores, que lo utilizan por la riqueza y diversidad de datos que ponen a disposici n de cualquier persona y que resultan de la colaboraci n entre usuarios, algunos de ellos expertos. No en vano, como afirma Chiari (2012: 105):

[...] garantendo un arricchimento della risorsa piuttosto rapido, [la collaborativit ] ha spinto soprattutto chi si occupa di sviluppo di risorse linguistiche ad avvicinarsi ai progetti di tipo Wiki per estrarre automaticamente informazione linguistica da aggregare ad uso computazionale o da integrare e arricchire con risorse linguistiche *expert-based*.

La exploraci n llevada a cabo en este trabajo, aun limitada al an lisis de la presencia y tratamiento de dos t rminos, permite confirmar, en la direcci n de estudios previos, que es posible extraer equivalentes terminol gicos entre los pares de lenguas italiano-espa ol a trav s de los *interwikis* o enlaces interling isticos de la enciclopedia; adem s, se corrobora que los numerosos datos de car cter hist rico, los relativos a las t cnicas y soportes, am n de las im genes y las referencias contenidas en los art culos, pueden ser  tiles para aclarar el significado de voces especializadas asociadas a la pintura y encontrar sus posibles equivalentes de traducci n. Y es que, como se ala Carpi:



Los textos que tratan temas relativos a las bellas artes, por ser el resultado de una traducción intersemiótica – transposición en palabras de un texto icónico – y por mantener una estricta relación con objetos del mundo extralingüístico, pueden plantear relevantes problemas de traducción (2016: 114).

Por este motivo, la extracción de datos multimedia y la inclusión de este tipo de contenidos – en particular los de dominio público – en diccionarios especializados, como el que propone el proyecto *Lessico dei Beni Culturali* pueden resultar de gran utilidad. También es posible, como plantean González y Gamallo (2010), utilizar Wikipedia como fuente multilingüe de corpus comparables, adoptando ciertas categorías temáticas (e.g. Pintura) como elementos restrictivos. Sin embargo, en relación con la extracción automática de datos de Wikipedia para integrar bases de datos existentes o para elaborar diccionarios y glosarios fiables, es necesario seguir investigando. Como apuntan algunos estudios citados, para elaborar diccionarios utilizando las categorías y enlaces de Wikipedia es esencial contrastar los resultados obtenidos con herramientas y recursos como Wordnet, y con corpus de textos especializados, como pueden ser los textos paralelos de traducciones de tratados de arte o, en el caso que nos ocupa, manuales y guías de pintura. Cabe por tanto adoptar cierta prudencia hacia la enciclopedia en línea, en particular en relación con las definiciones y explicaciones contenidas en la parte introductoria de los artículos publicados. Parece, pues, necesario interrogarse sobre los criterios de selección y descripción de los artículos en base a fuentes correctas, fiables y completas. Como se ha visto, es difícil controlar la veracidad de los datos y existe una amplia desigualdad en el tratamiento y presentación de los mismos. También es fundamental entender el papel que juega la traducción en la elaboración de algunos artículos. Se ha comprobado cómo el artículo ‘Pittura a tempera’ se apoyaba en distintas referencias escritas en inglés y en alemán, lo que lleva a suponer algún tipo de adaptación o traducción a partir de esas referencias para la redacción del mismo. Además, la traducción en Wikipedia está basada en el trabajo de voluntarios que probablemente no cuentan con un método traductor compartido.

En todo caso, es posible afirmar que la enciclopedia en línea puede contribuir a dar visibilidad a proyectos sobre léxico especializado. El progresivo refinamiento de sus controles, la atención de los medios académicos y científicos y la tendencia que revelan sus estadísticas de uso parecen dirigir esta herramienta hacia un estatus de referencia obligada. Algunos estudios (Lally, Dunford 2007; Szajewski 2013) han demostrado que la inclusión en determinados artículos de Wikipedia de enlaces hipertextuales a las colecciones digitales universitarias garantiza una mayor visibilidad de las mismas y la consulta frecuente de sus documentos. Desde esta perspectiva, los proyectos sobre léxico especializado plurilingüe como LBC pueden aprovechar el impacto de estas plataformas para aumentar su visibilidad en la

Red, erigiéndose además en puntos de referencia para el conocimiento del léxico, en este caso, artístico. Incluido en una sección de referencias y de enlaces externos, el diccionario y los recursos que van surgiendo de este proyecto (i.e. corpus paralelo de traducciones de las *Vidas* de Vasari) pueden contribuir a mejorar los contenidos de los artículos de Wikipedia, esto es, las voces relativas a campos del saber menos representados, puesto que se apoyan en metodologías específicas de investigación lexicográfica y traductológica.

## Referencias

- Águila Escobar G. 2006, *Las nuevas tecnologías al servicio de la lexicografía*, in Villayandre Llamazares M. (ed.), *Actas del XXXV Simposio internacional de la Sociedad Española de Lingüística*, Universidad de León, León: 1-23.
- Alonso E. 2014, *Interacciones sociales y tecnológicas en el entorno profesional de la traducción*, «Tonos Digital. Revista de Estudios Filológicos», 17: 1-29, <<http://www.tonosdigital.es/ojs/index.php/tonos/article/view/1124/692>> (08/2016).
- Alonso E. 2015, *Una aproximación a Wikipedia como polisistema cultural*, «Convergencia. Revista de Ciencias Sociales», XXII (68): 125-149, <<http://www.redalyc.org/pdf/105/10536227005.pdf>> (09/2016).
- Alvar Ezquerro M. 2003, *La estación lexicográfica*, in Martín Zorraquino, M. A. Aliaga J. L. (eds.), *La lexicografía hispánica ante el siglo XXI*, Institución Fernando el Católico, Zaragoza: 225-231.
- 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.
- Chiari I. 2012, *Il dato empirico in lessicografia: dizionari tradizionali e collaborativi a confronto*, «Bollettino di italianistica», 2: 94-125.
- Colomer J. C., Ramos J. 2012, *Wikipedia en la encrucijada*, «Revista de Estudios Culturales», 1: 115-118.
- Erdmann M., Nakayama K., Hara T., Nishio S. 2008, *An approach for extracting bilingual terminology from Wikipedia*, in Haritsa J. R., Kotagiri R., Pudi V. (eds.), *Database Systems for Advanced Applications*, DASFAA, New Delhi: 380-392.
- Farina A. 2015, *Un parcours du combattant: la recherche de traductions de la langue d'art dans les ressources lexicographiques et terminologiques. L'exemple de tempera e guazzo*, in Carpi E. (ed.), *Prospettive plurilingui e interdisciplinari nel discorso specialistico*, Pisa University Press, Pisa: 33-57.
- Feller J., Fitzgerald B., Hissam S. A., Lakhani K. R. 2005, *Perspectives on free and open source software*, The MIT Press, Cambridge, MA.
- Goldman J.-P., Scherrer Y. 2012, *Création automatique de dictionnaires bilingues d'entités nommées grâce à Wikipédia*, «Nouveaux cahiers de linguistique française», 30: 213-227, <[http://clf.unige.ch/files/8114/4102/7432/11\\_Goldman\\_Scherrer\\_Nclf30\\_213-227.pdf](http://clf.unige.ch/files/8114/4102/7432/11_Goldman_Scherrer_Nclf30_213-227.pdf)> (10/2016).
- Guiote A. 2011, *Entre los bastidores del conocimiento 2.0: el perfil, las reflexiones y las rutinas de acción de los constructores del contenido en Wikipedia*, «Zer», 16: 13-36, <<http://www.ehu.es/zer/hemeroteca/pdfs/zer30-01-guiote.pdf>> (09/2016).

- Huvila I. 2010, *Where does the information come from? Information source use patterns in Wikipedia*, «Information Research», XV (3), <<http://www.informationr.net/ir/15-3/paper433.html>> (09/2016).
- Lally A., Dunford C. 2007, *Using Wikipedia to extend digital collections*, «D-Lib Magazine», 13 (5/6), <<http://www.dlib.org/dlib/may07/lally/05lally.html>> (09/2016).
- Lévy P. 1994, *L'intelligence collective. Pour une anthropologie du cyberspace*, La Découverte, Paris.
- Mesgari M., Okoli Ch., Mehdi M., Nielsen F. A., Lanamäki A. 2015, *The sum of all human knowledge: A systematic review of scholarly research on the content of Wikipedia*, «Journal of the Association for Information Science and Technology», LXVI (2): 219-245.
- Millán J. A. 1999, *Estaciones filológicas*, in Blecua J. M., Clavería G., Sánchez C., Torruella J. (eds.), *Filología e informática. Nuevas tecnologías en los estudios filológicos*, Universitat Autònoma de Barcelona, Barcelona: 143-164.
- Milne D., Medelyan O., Witten I. H. 2006, *Mining domain-specific thesauri from wikipedia: A case study*, in *Proceedings IEEE/WIC/ACM International Conference on Web Intelligence*, IEEE, Washington, DC: 442-448.
- Müller C., Gurevych I. 2009, *Using Wikipedia and Wiktionary in domain-specific information retrieval*, in Peters C., Deselaers Th., Ferro N., Gonzalo J., Jones G. J. F., Kurimo M., Mandl Th., Peñas A., Petras V. (eds.), *Evaluating Systems for Multilingual and Multimodal Information Access*, Springer, Berlin: 219-226.
- Oliver A., Climent S. 2012, *Using Wikipedia to develop language resources: WordNet 3.0 in Catalan and Spanish*, «Digithum», 14: 15-24, <<https://digithum.uoc.edu/articles/abstract/10.7238/d.v0i14.1474/>> (06/2017).
- O'Reilly T. 2005, *What is Web 2.0? Design patterns and business models for the next generation of software*, O'Reilly Network, <<http://oreilly.com/web2/archive/what-is-web-20.html>> (09/2016).
- Raymond E. S. 1997, *The Cathedral and the Bazaar*, O'Reilly, Sebastopol, CA.
- Ribes X. 2007, *La Web 2.0. El valor de los metadatos y de la inteligencia colectiva*, «Telos», 73: 36-43, <<https://telos.fundaciontelefonica.com/telos/articulooperpectiva.asp?idarticulo=2&rev=73.htm>> (09/2016).
- Spandana G., Strapparava C., Nastase V. 2014, *Mapping WordNet Domains, WordNet Topics and Wikipedia Categories to Generate Multilingual Domain Specific Resources*, in *Proceedings 9th International Conference on Language Resources and Evaluation*, Reykjavik: 1117-1121.
- Szajewski M. 2013, *Using Wikipedia to enhance the visibility of digitized archival assets*, «D-Lib Magazine», XIX (3-4), <<http://www.dlib.org/dlib/march13/szajewski/03szajewski.html>> (10/2016).
- Tyers M. F., Pieanaar J. A. 2008, *Extracting Bilingual Word Pairs from Wikipedia*, in *Proceedings 6th International Conference on Language Resources and Evaluation*, SALT MIL, Marrakesh: 19-22.
- Vivaldi J., Rodríguez H. 2011, *Extracting terminology from Wikipedia*, «Procesamiento del Lenguaje Natural», 47: 65-73, <[http://rua.ua.es/dspace/bitstream/10045/18514/1/PLN\\_47\\_07.pdf](http://rua.ua.es/dspace/bitstream/10045/18514/1/PLN_47_07.pdf)> (09/2016).

- Zazo A., Alonso J. L., Figuerola C. G. 2015, *Edición de contenidos en un entorno colaborativo: el caso de la Wikipedia en español*, «Scire: Representación y organización del conocimiento», XXI (2): 57-67.
- Zesch T. 2010, *What's the difference? Comparing expert-built and collaboratively-built lexical semantic resources*, in Calzolari N., Baroni P., Monachini M., Soria C. (eds.), *Proceedings 2nd European Language Resources and Technologies*, ACL Press: 91-92.



F. Tomasi

## L'informazione digitale e il Web semantico. Il caso delle *scholarly digital editions*

*Riassunto:* Il rinnovamento dei linguaggi del Web ha portato con sé un rinnovamento delle metodologie della rappresentazione delle risorse informative, che va dall'attenzione al documento all'attenzione verso il dato, ogni elemento identificato in grado di veicolare informazione in un testo. Le forme della rappresentazione del testo digitale si confrontano dunque con nuovi modi di intendere l'informazione quali Linked Open Data (LOD), che adotta un approccio 'data centrico' capace di sottrarre ambiguità alle sequenze di stringhe di caratteri in un testo, non interpretabili con sistemi di rappresentazione formale. Il presente lavoro esplora i linguaggi del Web semantico e il rapporto di questi nuovi linguaggi con la filologia e descrive il processo di costruzione dei LOD applicato alla collezione delle *Lettere* di Vespasiano da Bisticci. L'obiettivo è illustrare come, dall'edizione digitale (documenti HTML) è possibile passare al *knowledge site*, un ambiente semantico di accesso alla conoscenza formato dai dati di un testo, dalle stringhe interpretate al suo interno e dal rapporto di questi con altri dati sul Web.

*Parole chiave:* rappresentazione dati, web semantico, linked open data, filologia, lettere, Vespasiano da Bisticci.

*Abstract:* The renewal of Web languages has brought with it a renewal of the methodologies available to represent information resources, which puts the attention from the document itself to the data, that is, each item situated in a specific position of the text, which conveys information. The ways and methods to represent digital texts face, thus, new systems capable of understanding information, such as Linked Open Data (LOD), which adopts a 'data centric approach' able to remove ambiguities within a specific sequence of character strings that are not interpretable with formal representation systems. This paper explores these Semantic Web languages and their relation with philology; it describes the construction process of LOD applied to the

collection of letters written by Vespasiano da Bisticci. The aim of the study is to highlight how digital edition (HTML documents) moves towards the 'knowledge site', a semantic environment that allows the access to knowledge through the data contained in a given text, the interpreted strings within it and the relationship of these and other data on the Web.

*Keywords:* data representation, Semantic Web, Linked Open Data, philology, letters, Vespasiano da Bisticci.

## 1. Introduzione

L'evoluzione del Web ha determinato un cambiamento importante nei sistemi di rappresentazione digitale delle risorse informative. Il rinnovamento dei linguaggi del Web ha portato con sé un rinnovamento delle metodologie della rappresentazione. Dall'attenzione al documento si è passati all'attenzione al dato, ad ogni elemento atomico, identificato e qualificato, in grado di veicolare l'informazione. Lo scopo finale è quello di acquisire conoscenza dagli oggetti digitali. Le *humanities* non sono rimaste immuni da questo cambiamento. E le forme della rappresentazione del testo digitale si sono dovute confrontare con una *suite* di nuovi linguaggi, che sono anche nuovi modi di intendere l'informazione. Non è più il documento che trasmette l'informazione, ma è invece il dato la risorsa da cui acquisire il sapere. Questo ha significato fare uno sforzo per tradurre i documenti in dati. In particolare il fenomeno dei *Linked Open Data* (abbreviato in LOD) ha determinato nuovi sistemi per avviare tale trasformazione nel contesto del *Web of data*. Il caso delle edizioni digitali di testi letterari è un caso di rinnovamento nel settore delle *humanities* e in particolare della filologia e dell'ecdotica. L'approccio cosiddetto 'data centrico' vuole sottrarre l'ambiguità e l'indeterminatezza delle sequenze di stringhe di caratteri non interpretate con un sistema di rappresentazione formale in grado di consentire alla macchina di comprendere il dato.

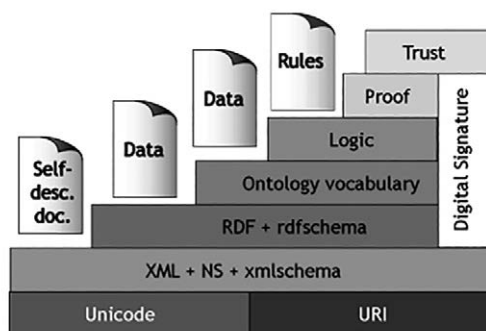
Dopo una disamina dunque dei linguaggi del Web semantico, e del rapporto di questi nuovi linguaggi con la filologia, viene esemplificato il processo di costruzione dei LOD applicato ad un caso reale: la collezione delle *Lettere* di Vespasiano da Bisticci. Dall'edizione digitale, sequenza di documenti HTML, si passa al *knowledge site*, cioè ad un ambiente semantico di accesso alla conoscenza veicolata tanto dai dati, come stringhe interpretate, quanto dal loro rapporto con altri dati sul Web, in un contesto di relazioni multiple e consistenti.

## 2. Dal documento al dato

Da quando nel 2001 Berners-Lee presentò il progetto del Web semantico (o Web 3.0) molta strada è stata fatta. Applicazioni, progettisti ma anche

utenti hanno saputo ricollocare il ruolo dell'informazione digitale alla luce dell'arricchimento concettuale che il Web semantico ha determinato. Se la definizione di nuovi standard, ma anche il recupero di linguaggi consolidati, sono stati un elemento importante nell'elaborazione del *Semantic Web Stack*<sup>1</sup> (Fig. 1), cioè nella sua architettura per *layer*, una consistente novità è stata la ridefinizione stessa del concetto di informazione. Il Web 3.0 ha valorizzato quei linguaggi di rappresentazione capaci di tradurre il processo, noto anche, nei sistemi di *knowledge management*, come piramide della conoscenza, che dal dato (il fatto) deve portare alla saggezza (la conoscenza applicata), passando per l'informazione (il dato portatore di significato) e la conoscenza (l'informazione organizzata).

Figura 1 – *Semantic Web Stack*. W3C, *Semantic Web - XML2000* by Tim Berners-Lee, slide 10, <<https://www.w3.org/2000/Talks/1206-xml2k-tbl/slide10-0.html>>.



Ecco che al primissimo livello dello *stack* si colloca lo standard internazionale per la codifica dei caratteri (*Unicode*) e il sistema di identificazione univoca dei dati (*URI* – *Uniform Resource Identifier*); al secondo livello stanno i linguaggi formali per la descrizione dei dati, in particolare la sintassi XML, ovvero il meta-linguaggio, il sistema dello spazio dei nomi (*Namespace*), per identificare univocamente l'uso di un certo vocabolario descritto poi negli Schemi (*xmlschema*), per associare appunto alla sintassi un vocabolario di rappresentazione del dato. Il terzo livello è quello del formato di descrizione della risorsa, ovvero RDF, che permette di descrivere il dato come una tripla, o anche un'asserito, costituito da soggetto-predicato-oggetto assieme allo schema (*rdfschema*), che fornisce le specifiche del vocabolario di descrizione per arricchire semanticamente il dato. Si sale ancora all'ontologia: la concettualizzazione del dominio attraverso classi e proprietà o anche la rappresentazione della conoscen-

<sup>1</sup> Detto anche *Semantic Web Layer Cake*. Cfr. W3C Semantic Web Activity, <<https://www.w3.org/2001/sw/>> (10/2016).



za (*Ontology vocabulary*). Per finire con i tre livelli: lo strato logico (*logic*) necessario a supportare le inferenze, lo strato delle dimostrazioni (*proof*), per associare le regole di inferenza agli assiomi e quindi lo strato della fiducia (*trust*), per garantire l'autenticità degli asserti e stabilire la loro provenienza. Questo sforzo di definizione di linguaggi e livelli impone, dal punto di vista della rappresentazione digitale, di considerare non più il documento come oggetto portatore di informazione, ma il dato inteso come ogni risorsa identificata e qualificata.

Figura 2 – L'evoluzione del Web. Radar Networks & Nova Spivack, 2007, <<http://novaspivack.typepad.com/RadarNetworksTowardsAWebOS.jpg>>.

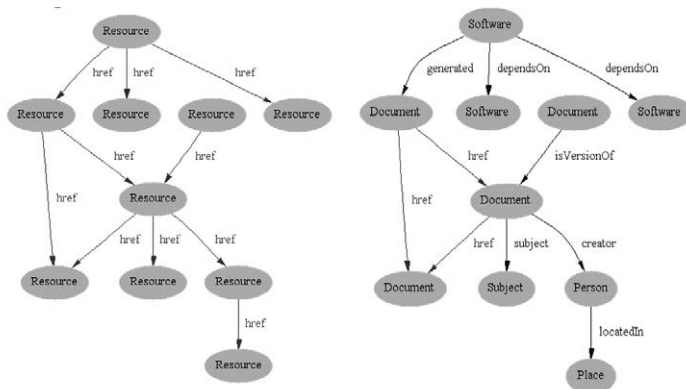


Detto altrimenti, da un approccio tipicamente documento-centrico del Web 1.0, passando per una fase ibrida del Web 2.0<sup>2</sup>, ciò che distingue il Web 3.0 è la scelta di un approccio data-centrico al concetto stesso di informazione<sup>3</sup> (Fig. 2). Il Web 1.0 ha individuato nel documento HTML il focus della rappresentazione digitale e nel link lo strumento per creare collegamenti fra documenti. Il Web 2.0 ha introdotto l'importante concetto della democratizzazione del sapere, concentrando l'attenzione sull'idea di un Web come ambiente collaborativo e partecipato, in cui il *crowdsourcing*, o anche la generazione di contenuti da parte di un utente non più passivo fruitore ma attivo produttore, è la chiave per la valorizzazione del sapere. Il Web 3.0 ha spostato l'enfasi sul dato, la risorsa qualificata, e sulle relazioni tipizzate, ovvero semantiche (Fig. 3).

<sup>2</sup> Per una disamina del concetto di Web 2.0 si vedano le pagine del suo 'ideatore', Tim O'Reilly, *What is Web 2.0.*, <<http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>> (10/2016).

<sup>3</sup> Il Semantic Web è associato al termine Web 3.0. Secondo il W3C, «the Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries» (*W3C Semantic Web Activity*, World Wide Web Consortium (W3C), November 7, 2011, <<https://www.w3.org/2001/sw/>> 10/2016). Il termine è stato coniato da Tim Berners-Lee, «for a web of data that can be processed by machines» (Berners-Lee, Hendler, Lassila 2001).

Figura 3 – Fra Web 1.0 (risorse, ovvero documenti, fra di loro collegate) e Web 3.0 (ogni risorsa è identificata in un dato collegato ad altri dati attraverso relazioni semanticamente esplicite).



Il focus si è allora spostato dal «documento» al «dato», inteso come un qualunque oggetto portatore di informazione, sia esso appunto una risorsa (testo, immagine, audio e video identificato come contenuto) o un 'frammento', l'elemento atomico (una persona, un luogo, una data, un evento ecc.).

Storicamente la contrapposizione documento/dato ha riguardato anche componenti tecnologiche della rappresentazione informatica. Il documento, classico esempio di file non strutturato, è da sempre stato l'oggetto su cui i linguaggi di markup hanno esercitato sforzi finalizzati alla modellazione, con l'obiettivo di produrre oggetti semi-strutturati.

Da un documento, inteso come sequenza di stringhe di caratteri non altrimenti interpretate, i linguaggi di markup, attraverso una notazione formale, aggiungono struttura, per un primo livello di formalizzazione del contenuto del documento stesso (Fig. 4).

La complessità della rappresentazione informatica che i linguaggi di markup *embedded* – in cui il markup è interno al documento stesso – portano con sé ha arricchito il dibattito intorno ai sistemi formali per la descrizione del documento. Il fine è quello di superare la nozione di 'testo' come organizzazione rigidamente gerarchica dei suoi elementi costitutivi, che i linguaggi di markup a base SGML (come HTML) o le sue derivazioni (come XML) impongono<sup>4</sup>. I linguaggi di markup *standoff* si muovono in questa prospettiva<sup>5</sup>, esternalizzando il markup in file a parte, e limitando nel documento una serie di puntatori a file esterni. Il dato è invece fondamentale legato al concetto di database e quindi di informazione strutturata.

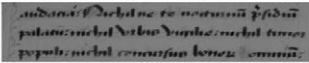
<sup>4</sup> Sui linguaggi di markup si vedano almeno: DeRose, Durand, Mylonas, Renear 1997; Sperberg-McQueen, Huitfeldt, Renear (2000).

<sup>5</sup> Per una disamina dei linguaggi di markup *standoff*, si veda Schmidt (2010).

La rappresentazione astratta dello schema logico del database (le strutture dati definite da tabella, grafo, albero od oggetti) supera le difficoltà dell'annotazione incorporata nel documento, agevolando la definizione di un modello esterno che si astrae dal documento per focalizzarsi sui suoi metadati descrittivi.

Figura 4 – Trascrizione, attraverso un linguaggio di markup, di una fonte primaria per segnalare alcuni fenomeni della scrittura. In alto l'immagine di una porzione di manoscritto. Sotto l'immagine, a sinistra, una marcatura formale e, a destra, la legenda per interpretare la marcatura. Nella parte bassa della figura, due metodi per visualizzare lo stesso documento marcato: a sinistra, una visualizzazione diplomatica, rispettosa della segnaletica del manoscritto e che replica i fenomeni della fonte primaria; a destra, una trascrizione critica che risolve gli elementi del manoscritto in forma leggibile per l'utente finale.

Immagine: Cicero, *De Catilinari oratione IV*, Baltimore, Johns Hopkins University, Milton S. Eisenhower Library, MSB 03, f. 134r.



**Markup XML/TEI (P4) della porzione di manoscritto in immagine**

```

audacia <orig reg="1">esegno paragrafematico</orig>
chi id="01"><del>Nihil te nocentium praevidui</del>
te nocentium<del> expan="m">atratto sopra:</del></del>
p<abbr expan="rae">esopra punto:</abbr>
sidi<abbr expan="m">atratto sopra:</abbr>
<orig reg="2">pe</orig>
<orig reg="3">svirgola medievale:</orig>
nisi<del> corra="s">con</del>
<orig reg="4">svirgola medievale:</orig>
nisi<del> corra="s">con</del>
populi<orig reg="5">svirgola medievale:</orig>
nisi<del> corra="s">con</del>
bono<del> abbe expan="m">atratto sopra:</del></del>
<del rend="rasura"></del>
omni<del> abbe expan="m">atratto sopra:</del></del>
<orig reg="6">svirgola medievale:</orig>

```

LEGENDA	
<orig>	testo come si legge nel manoscritto;
reg	di la versione normalizzata
<del>	segnala la presenza di un errore o di un luogo su cui l'editore intende intervenire;
corra	riporta la versione corretta o modificata
<abbr>	indica un'abbreviazione (espressa con riferimento all'entità carattere in grado di riprodurre il segno abbreviativo);
expan	scioglie l'abbreviazione
<del>	indica una cancellazione;
rend	specifica le modalità della cancellazione
<del>	segnala una qualche forma di enfasi;
id	si associa all'entità un identificativo univoco

Dal file XML possono essere prodotti due diversi output. Questo è possibile associando allo stesso documento un diverso foglio di stile. Nel caso della **trascrizione diplomatica** si rispettano i segni di scrittura, rendendoli così come compaiono nel manoscritto, tentando anche di riprodurre i fenomeni ricostruiti (segni speciali); nel caso della **trascrizione critica** il testo viene sostituito con i valori degli attributi (<orig> per l'elemento <del>, <expan> per l'elemento <abbr>, <reg> per l'elemento <orig>). Con l'ausilio di elementi tipografici (id e color) i fenomeni (abbreviazioni, correzioni e cancellazioni) possono essere resi in layout.

<p><b>Esempio di trascrizione diplomatica</b></p> <pre> audacia: Nihil te nocentium praevidui palatii nichil urbis vigiliis: nichil timor populi: nichil concursuum bonoꝝ omnium: </pre>	<p><b>Esempio di trascrizione critica</b></p> <pre> audacia: Nihil te nocentium praesidium Palatii, nihil urbis vigiliis, nihil timor populi, nihil concursus bonoꝝ omnium, </pre>
--	--

Data questa rapida disamina è evidente che il Web semantico ha comportato uno slittamento tanto sul piano delle scelte tecnologiche quanto su quello della teoria della testualità digitale. L'esempio più evidente sono i Linked Data<sup>6</sup>. Il Semantic Web usa linguaggi di markup consolidati, primo fra tutti XML, eletto dal consorzio W3 come il linguaggio per l'interscambio fra documenti<sup>7</sup>. Si tratta quindi di un linguaggio dichiarativo, *embedded*, che lega il testo alla sua descrizione. Ma il Semantic Web adotta anche il RDF<sup>8</sup>, il 'modello per triple' (soggetto, predicato, oggetto) attraverso cui descrivere ogni possibile contenuto. Qualunque cosa dotata di URI, ovvero, di un indirizzo univoco per la sua riconoscibilità di rete, può essere

<sup>6</sup> Si veda <<http://linkeddata.org/>> per tutorial, guide, *tools*, *datasets* e risorse. La primissima risorsa ad aver trattato del tema dei cosiddetti LOD è a cura dello stesso fondatore del WWW, Tim Berners-Lee, nel 2006, <<https://www.w3.org/DesignIssues/LinkedData.html>> (10/2016).

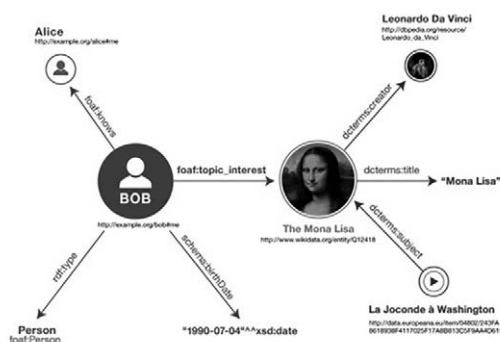
<sup>7</sup> L'ultima raccomandazione XML del W3C si può leggere all'indirizzo <<https://www.w3.org/TR/2006/REC-xml-20060816/>> (10/2016).

<sup>8</sup> La suite di *RDF recommendations* si può leggere in <[https://www.w3.org/standards/techs/rdf#w3c\\_all](https://www.w3.org/standards/techs/rdf#w3c_all)> (10/2016).

variamente collegata a qualsiasi altra cosa altrettanto dotata di URI o descrivibile attraverso un 'letterale', cioè una stringa.

Tipicamente, fatta eccezione per RDFa, che prevede l'annidamento delle 'triple' nel testo del documento, analogamente al funzionamento del markup *embedded*<sup>9</sup>, la descrizione RDF risiede esternamente al documento, formando un nuovo oggetto che si presenta come una sequenza di asserti. I Linked Data, che possiamo definire come una porzione di Web semantico o anche come una sua realizzazione<sup>10</sup>, sono questo: dati in formato aperto, identificati univocamente ed espressi in RDF (Fig. 5).

Figura 5 – Esempio di visualizzazione di una tripla RDF attraverso un grafo. I dati, identificati univocamente attraverso URI, sono posti fra di loro in relazione attraverso collegamenti semantici espliciti (<<https://www.w3.org/TR/rdf11-primer/>>, 10/2016).



Le relazioni fra URI della tripla RDF fanno riferimento al modello cui la collezione si ispira e che è tipicamente l'ontologia. Da un punto di vista teorico, l'ontologia, al di là delle sue implementazioni, rappresenta la concettualizzazione astratta di un dominio. Essa quindi è il modello concettuale su cui si basa la descrizione dei dati. Ecco che anche un documento diventa allora una rappresentazione dei suoi contenuti attraverso un processo di astrazione delle sue proprietà descrittive per popolare un *dataset* di asserti. In questo modo, dal documento si passa al dato e al suo modello astratto di concettualizzazione (cfr. *infra* par. 4).

### 3. Il ruolo della filologia

La filologia non è immune dal processo di manipolazione di dati. Definire gli elementi significativi del testo (dalle aggiunte alle abbreviazioni,

<sup>9</sup> Per RDFa Core 1.1 (terza edizione) si veda <<https://www.w3.org/TR/rdfa-syntax/>> (10/2016).

<sup>10</sup> Non è un caso se la W3C Semantic Web Activity sia stata sostituita nel dicembre 2013 dalla W3C Data Activity, <<https://www.w3.org/2013/data/>> (10/2016).

dalle note prosopografiche a quelle storiche, dagli interventi di mano d'autore all'apparato delle varianti) vuol dire individuare dati per arricchirli con altri dati e stabilire relazioni fra dati nel documento e altri dati provenienti da altri documenti.

In generale potremmo dire che l'edizione digitale si può descrivere come un processo. Tale processo è un complesso percorso che, da una fase di progettazione iniziale, attraversa una serie di scelte tecnologiche, per approdare alla disseminazione. Ogni fase di questo processo segue un modello ideale ed astratto che identifica le caratteristiche dell'edizione. Il principio della modellazione dell'edizione digitale<sup>11</sup> trova, in particolare nel momento di gestione del testo, una delle componenti costitutive forti per ogni successivo intervento computazionale. Il modello del testo è dunque un'esigenza necessaria e determinante. Modellare significa ragionare sia sulle componenti logiche interne del testo, che sulla rete delle relazioni interne, che ancora sulle componenti extra-testuali o trans-testuali che ne arricchiscono la funzione (ai livelli di intertesto, paratesto, metatesto, ipertesto e architesto). Questi sono i dati che andranno formalizzati attraverso gli adeguati strumenti concettuali per estrarre nuova conoscenza.

Ipotizzando un'edizione diplomatica o interpretativa a testimone unico, la concettualizzazione si risolve, a livello zero, nelle scelte adottate in fase di trascrizione. In particolare, nel caso di testimoni manoscritti non saranno irrilevanti le decisioni prese per il processo di normalizzazione: la segnalazione di caratteristiche della scrittura, l'interpunzione, la modernizzazione delle grafie, le abbreviazioni, le correzioni sono solo alcune delle operazioni su cui si esercita lo spirito critico dell'editore. In un contesto digitale ogni intervento dovrà essere espresso secondo principi formali, ovvero utilizzando linguaggi computabili. Solo ciò che viene formalizzato potrà essere trattato in modo automatico. Tipicamente il processo di formalizzazione del testo trova nel markup XML/TEI<sup>12</sup> la sua più naturale realizzazione. L'annotazione *embedded*, tramite una sintassi condivisa come XML e un vocabolario altrettanto standardizzato nella comunità TEI, traduce le ipotesi interpretative dell'ermeneuta in un sistema formale e quindi computabile (Fig. 4).

È innegabile che all'analisi del modello del testo si affianchi l'analisi delle relazioni interne fra le componenti testuali. Naturalmente i dati presenti nel testo sono fra loro in necessaria relazione. E solo la definizione delle relazioni dà significato ai dati. Esempi possono essere le connessioni che nascono dalle risposte ai quesiti: chi, dove, quando e cosa – i.e. una data e

<sup>11</sup> I cui requisiti possono essere ad esempio nelle linee guida MLA (Modern Language Association).

<sup>12</sup> Si veda il sito ufficiale della TEI (*Text Encoding Initiative*) all'indirizzo <<http://tei-c.org>> (10/2016).

un luogo, un luogo e una persona, una persona e un evento. Le relazioni, nativamente latenti, danno quindi consistenza al significato dei dati rispetto al contesto, e cioè il testo in cui sono inseriti. La rete delle relazioni può essere formalizzata attraverso asserti, tipicamente RDF, che permettono ai dati di trasformarsi in contenuto computabile.

Un terzo livello di modellazione è rappresentato invece dalle connessioni che il testo intrattiene con altri testi o con il commento. Ogni forma di correlazione fra il testo e tutto ciò che contribuisce a dare consistenza semantica al testo stabilisce il dialogo fra dati. Sarà necessario definire quelle opportune relazioni che trasformano un silos di dati in un *network*. Questo è il principio che governa l'evoluzione del Web semantico, ovvero i Linked Open Data:

Linked Data is about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods. More specifically, Wikipedia defines Linked Data as 'a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF'<sup>13</sup>.

Non è sufficiente però adottare URI, RDF e dati aperti. Al fine di garantire interoperabilità, interscambio, dialogo e reale efficacia semantica dovrà essere prevista non solo la creazione di ontologie ad hoc per il dominio, ma soprattutto il riuso di modelli concettuali già esistenti. Sarà allora utile esemplificare quanto detto su un caso reale: l'edizione digitale delle *Lettere di Vespasiano da Bisticci*<sup>14</sup>.

#### 4. Il caso Vespasiano da Bisticci e le sue lettere

Vespasiano da Bisticci, o anche «*princeps omnium librariorum Florentinorum*»<sup>15</sup>, è stato un copista e libraio fiorentino vissuto nel corso del XV secolo. Le lettere, inviate e ricevute, non sono un vero e proprio epistolario, ma costituiscono una raccolta risultato dell'aggregazione da parte di moderni editori. Esse coprono la gamma degli interessi del libraio e aprono una finestra sulla tipologia delle sue occupazioni, fornendo un quadro nitido, nonostante l'esiguità del *corpus* (costituito, ad oggi, da sole quarantaquattro lettere, delle quali ventidue inviate e ventidue ricevute,

<sup>13</sup> <<http://linkeddata.org>> (10/2016).

<sup>14</sup> URI: <<http://vespasianodabisticciletters.unibo.it/>>; DOI: 10.6092/unibo/vespasianodabisticciletters.

<sup>15</sup> Così lo appella uno dei più prolifici copisti della scuola di Vespasiano, Gherardo del Ciria-go, cfr. de la Mare 1985.

che coprono il periodo 1446-1498<sup>16</sup>), di alcune questioni fondative dell'umanesimo: le biblioteche degli umanisti e dei mecenati delle principali corti europee, le modalità della circolazione dei codici manoscritti, i rapporti politici, le traduzioni e i volgarizzamenti.

Le missive testimoniano relazioni commerciali, rapporti di amicizia, scambi di vedute politiche, narrazioni di fatti storici, discussioni di diatribe teologiche. E la varietà dello status dei corrispondenti rispecchia la molteplicità dei contenuti trattati: amici, maestri, 'dipendenti', signori e principi, semplici committenti.

Le lettere di Vespasiano sono raccolte in istituti di conservazione diversi e hanno quindi una diversa origine materiale: manoscritti miscelanei e di dedica, documenti d'archivio di varia natura, testi a stampa<sup>17</sup>. Lo scopo dell'edizione digitale è di creare una collezione di risorse diverse che dal testo delle lettere consentano all'utente finale di accedere ad elementi di apparato, che documentino tanto le relazioni interne quanto la rete dei collegamenti esterni, citando dalla *homepage* del progetto (Fig. 6).

Le lettere, inviate e ricevute e fino ad oggi rintracciate, sono navigabili attraverso un sistema di faccette tematiche (corrispondente, data, luogo, segnatura). L'edizione è accompagnata da strumenti filologici di orientamento e guida (indice delle parole sottoforma di *authorities*, tavola sinottica, nota filologica, descrizione dei testimoni). Le lettere sono

<sup>16</sup> Vari sono gli editori delle lettere. L'ultimo è il Cagni, che ha realizzato un'edizione pressoché integrale (quarantadue lettere): Cagni (1969).

<sup>17</sup> Il libraio decise di far trascrivere in codice unico solo un piccolo corpus, corrispondente ad un parziale raggruppamento delle lettere ricevute, realizzato forse per compiacere Alessandro di Pierfilippo Pandolfini e i membri dell'Accademia Fiorentina cui il codice era dedicato. Si tratta del ms Plut. 90 sup 30. Ma è un caso isolato. Il resto del carteggio rinvenuto fino ad oggi proviene da fonti diverse. Alcune lettere sono confluite in un codice cartaceo miscelaneo: Firenze, Biblioteca Nazionale, Magl. VIII, 1390 (ff. 102, 102v, 103v, 41v, 47v, 48v) che raccoglie, per la maggior parte, note e lettere di Donato Acciaiuoli (sia lettere dell'umanista che lettere scritte per Vespasiano da Bisticci). Un documento è esso stesso un manoscritto di sole 10 carte (Firenze, Biblioteca Nazionale, Ms. Naz. II.XI.34, ff. 1-10v). Altre tre lettere sono carte di codici della Vaticana (Roma, Biblioteca Apostolica Vaticana, Vat. lat. 326, ff. 26r-v; Reg. lat. 1612, ff. 157v-158, ff. 158-158v). Una lettera si trova in un codice dell'Universitaria (Bologna, Biblioteca Universitaria, Italiano 1452, ff. 347-349). Per quanto riguarda i documenti d'archivio ci interessano quattro fondi: tre dell'Archivio di Stato e uno della Biblioteca Comunale di Forlì: Firenze, Archivio di Stato, *Mediceo avanti il Principato* (1373-1557; 165 pezzi; inventario a stampa) filza XVII n. 165, filza XVII n. 176, filza XII n. 413, filza XXVIII n. 701, filza XXV n. 377. Riproduzione digitale integrale in <<http://www.archiviodistato.firenze.it/Map/>>. Firenze, Archivio di Stato, *Carte Strozziiane* (1336-1905; 3497 pezzi; inventario) serie I, filza 253 n. 2; serie III, vol. 145 ff. 86-87v; serie I, vol. 137 ff. 290rv; serie I, vol. 137 ff. 288-289v. Firenze, Archivio di Stato, *Notarile Antecosimiano* (1237-1589; 21489 pezzi; inventario) T. 498 allegato primo e secondo all'atto del 6 settembre 1479 e allegato all'atto del 14 febbraio 1479 (il notaio è Ser Leonardo da Colle). Forlì, Biblioteca Comunale, *Autografi Piancastelli*, n. 85 e n. 102. Quattro documenti sono stati trascritti direttamente da due testi a stampa. Si tratta di carte provenienti dal fondo aragonese (cfr. F. Trinchera, *Codice Aragonese*, Napoli, 1866-1874, III voll.), distrutto dai tedeschi nel 1943.

accompagnate da informazioni contestuali (la raccolta delle lettere, i corrispondenti, i manoscritti realizzati dalla scuola, i copisti di Vespasiano, le biblioteche prodotte) necessarie ad inserire i documenti in una prospettiva metatestuale in evoluzione e aggiornamento costante, anche in futura dimensione collaborativa.

Figura 6 – Vespasiano da Bisticci, *Lettere*. Homepage <<http://vespasianodabisticci-letters.unibo.it>>.

**Vespasiano da Bisticci, Lettere**  
A digital edition

Homepage Per corrispondente Per luogo Per data Per segnatura

**Authorities**  
Ricerca guidata attraverso elenchi.  
L'edizione è stata indicizzata al fine di estrarre persone, codici, lessica, citazioni.  
Authorities »

**Tavola sinottica**  
Tabella relazionale.  
Una tabella che raccoglie, segnatura, mittente, destinatario, data cronica, data copia, fonte.  
Tavola sinottica »

**Nota filologica**  
I principi della trascrizione.  
Identificazione e descrizione di segni della scrittura, segni dell'eccezione, segni del supporto, segni dell'editore.  
Nota filologica »

**I testimoni**  
Le testimonie della raccolta.  
La descrizione degli esemplari, codici miscelanei e di dedica, documenti d'archivio, testi a stampa.  
I testimoni »

Al fine di sfruttare le tecnologie del Web semantico, con lo scopo di arricchire la base di conoscenza, il progetto ha richiesto un'adeguata modellazione degli scopi del lavoro editoriale. L'obiettivo è stato di creare un *Web environment* che desse consistenza al processo tripartito sopra descritto: la marcatura del documento, la costruzione delle relazioni fra dati, la definizione delle relazioni a livello di network.

In prima battuta il markup XML/TEI ha avuto l'obiettivo primario di identificare le persone menzionate nei documenti, i codici manoscritti greci e latini menzionati, richiesti o proposti da/a Vespasiano dai suoi committenti, il lessico tecnico della copia e della tradizione manoscritta. Ogni elemento marcato (persone/codici/lessico) popola un indice che associa la stringa alla corrispondente *authority* e istituisce le prime relazioni (Figg. 7, 8 e 9).



Figura 7 – Authority delle persone: Codice identificativo, Entry, Controllo di autorità, Forme parallele del nome, Forma attestata nel documento, Link diretto alla Lettera.

ID	Entry/Person	Authority control	Accepted parallel forms	Attested forms	Letter
P1	Acciaiuoli, Agnolo (d. post 1467)		Acciaiuoli, Anglo	Agnolo	8
P2	Acciaiuoli, Donato (1429-1478)	VIAF: 56292841 LCN: n81133865 SIRE: IT\CCIAISBU\157380	See authorities	Donato	1. (2). (3). (4). (5). 6. (20). 21. 22. 24
P3	Acciaiuoli, Jacopo		Acciaiuoli, Jacopo	Jacobo, Jacopo Acciaiuoli	6. 8. 22. 24
P4	Acciaiuoli, Piero		Acciaiuoli, Piero	Petro, Piero, Piero Acciaiuoli	6. 8. 22. 23. 24-41
P5	Alberto			Alberto scrittore	22
P6	Albizi degli, Luca (1459-1502)			Lucha deg'Albizi	22

La marcatura di ogni singola lettera consente, attraverso l'attributo @ref, di associare ogni fenomeno ad un file RDF esterno (per l'esattezza ad una porzione specifica del file RDF che corrisponde ad ogni fenomeno rilevato), nella forma: ns:metadata (namespace TEI:elemento TEI) + rdf files#XML:ID (file RDF e porzione del file corrispondente all'entità descritta) + string (forma attestata nel documento) + ns:metadata (chiusura dell'elemento). Ecco qualche esempio di estrazione della marcatura:

### Persname

```
<tei:persname ref="people.rdf#PS">Piero Strozi</tei:persname>
<tei:persname ref="people.rdf#FT">Pipo</tei:persname>
<tei:persname ref="people.rdf#VdB">Vespasiano di Filippo</tei:persname>
<tei:persname ref="people.rdf#PdM">Piero di Chosimo de' Medici
</tei:persname>
```

### Manuscripts

```
<tei:bibl ref="manuscripts.rdf#P_SN">
  <tei:author>Prinio</tei:author>
</tei:bibl>
<tei:bibl ref="manuscripts.rdf#L_D_III">
  <tei:title>Decha</tei:title>
</tei:bibl>
<tei:bibl ref="manuscripts.rdf#L_D_IV_E">
  <tei:title>abbreviations</tei:title>
</tei:bibl>
```

### Lexicon

```
<tei:term type="binding" ref="lexicon.rdf#leg">legaranno</tei:term>
<tei:term type="illumination" ref="lexicon.rdf#min">min</tei:term>
```

Figura 8 – Authority dei codici manoscritti: Codice identificativo, Entry, Autore, Titolo, Tipo di commissione (richiesto/realizzato/menzionato/copiato/proposto), Committente, Segnatura (quando il codice è stato identificato), Link diretto alla Lettera.

Codici

In ordine di menzione (cronologico). Accedi a: [ordine alfabetico per autore](#)

ms <sup>o</sup>	Entry:Codex	Author	Title	Commission	Person	Shelfmark	Letter
C1	Lactantium super Statium	Lattanzio [Placido]	[Commento di] su [Trebaiside di] Stazio	Realizzato per	Filippo Podocataro	Perduto	2
C2	Lycyrgl et Numme vitas (conversas)	[Plutarco, trad. Filelfo]	Vite di Licurgo e Numa [in traduzione]	Richiesto a	Filippo Podocataro	Forse codice utilizzato per esemplare medico (Campana, Una lettera)	3
C3	Tertulliani et Athanasij et Gregorij Nazanzeni	Tertulliano ed Atanasio e Gregorio Nazianzeno		Realizzato per	William Gray	Assente dal catalogo Baillou di Oxford (Mynors)	5
C4	De Vitis Plutarchi	Plutarco	Vite	In attesa di conferma per la realizzazione da	William Gray	Assente dal catalogo Baillou di Oxford (Mynors)	5

Figura 9 – Authority del lessico: Codice identificativo, Entry, Area semantica (formato/scrittura/supporto/legatura/costo/miniatura/aspetti filologici), Codice a cui è riferito il lemma, Link diretto alla Lettera.

Lessico

In ordine di menzione (cronologico). Accedi a: [ordine alfabetico](#)

ms <sup>o</sup>	Entry:Term	Area	Related to codex	Letter
L1	ad fragmenta (scribere)	format	--	2
L2	ad volumina (scribere)	format	--	2
L3	littere/litteris	writing	--	2
L4	scriptores/scriptoris	writing	--	2
L5	latus quinquaginta lineas, versus vero singuli elementa septuaginta	format	--	2
L6	grossi sex (pretium unius voluminis)	cost	--	2
L7	formam	writing	--	2
L8	optimum sane et emendatum	philology	[Commento di] Lattanzio Placido su Stazio	2
L9	forma scripture fusa	writing	[Commento di] Lattanzio Placido su Stazio	2

Le relazioni interne, espresse come asserti RDF, sono collocate in un file a parte, dove ogni persona è collegata al manoscritto – richiesto, copiato, realizzato, proposto o menzionato – e ogni manoscritto al lessico utilizzato per descriverlo, secondo l'area semantica di pertinenza. Nello specifico, ogni fenomeno marcato, espresso attraverso il valore dell'attributo @ref, intrattiene una relazione (predicato espresso in modo formale) con un oggetto, che sarà un altro valore di @ref, e che avrà quindi il suo URI, o un letterale, cioè una stringa in vocabolario naturale (Fig. 10).

La tripla inizia a definire anche un primo livello di relazioni con altre risorse Web o altri LOD datasets. Nel caso, ad esempio, delle persone la naturale associazione è, da un lato, con altri *authority files* che forniscono

la forma controllata dei nomi o che consentono di esercitare il controllo di autorità (es. Viaf, LCCN, SBN, ISNI ecc.), dall'altro, con risorse Web in qualche modo pertinenti (come il dataset di Wikipedia, Dbpedia, o altre pagine Web pertinenti). Questo è un esempio di descrizione che utilizza il principio delle relazioni fra gli elementi interni al testo e altri oggetti (notazione non formale):

```
<persName xml:id="LdM">Lorenzo de' Medici</persName>
xml:id=LdM
Access_key: Medici, Lorenzo de'
Dbpedia URI: http://it.dbpedia.org/page/Lorenzo_de%27_Medici
VIAF permalink: http://viaf.org/viaf/54169908
Father-of: Piero de' Medici (URI)
Born-in: Firenze (URI)
Died-in: Firenze (URI)
Went-in: Roma (URI) – function: marriage
Trip-to: Venezia (URI) – function: military mission
Born-when: 1449 (xsd:integer)
Died-when: 1492 (xsd:integer)
Iconography: http://www.google.com/search?hl=en&q=lorenzo+de%27+medici
&um=1&ie=UTF-8&tbm=isch&source=og&sa=N&tab=wi&ei=_1vcUOeAJqeI4A
SikoCYBQ&biw=1146&bih=709&sei=AVzcUNnjIOXV4gTVloHQCC
Biography: http://it.wikipedia.org/wiki/Lorenzo_de'_Medici
Author-of: http://it.wikisource.org/wiki/
Canti_carnascaleschi_%28Lorenzo_de%27_Medici%29
Attested-in: http://it.wikiquote.org/wiki/Maria
```

Ultime riflessioni in particolare sono dunque necessarie su classi e predicati scelti per introdurre in conclusione il tema delle ontologie e la loro funzione di concettualizzazione semantica di dominio. Se l'organizzazione concettuale attraverso l'identificazione delle classi, che nel nostro caso possono equivalere alle persone, ai codici e al lessico, ci permettono di istituire un primo livello di semantizzazione, è con i predicati che si stabiliscono le relazioni fra classi e, nel caso di una base di conoscenza, fra i valori delle classi, cioè le istanze.

Costruire una nuova ontologia significa infatti concettualizzare il dominio di riferimento attraverso un'astrazione delle sue caratteristiche, nella forma di classi e predicati. Ma quando si vuole descrivere un dominio, tipicamente, il primo passo è verificare se altre esistenti ontologie possono essere sufficientemente espressive a descrivere il dominio stesso. Il riuso di modelli ontologici è infatti un'azione preventiva, necessaria a garantire massimo interscambio fra i dati basati sullo stesso modello semantico. Laddove però le ontologie esistenti non siano sufficienti o non siano adeguate, allora sarà opportuno creare nuove soluzioni espressive. Creare una

nuova ontologia potrà anche prevedere, contestualmente alla definizione del nuovo processo astrattivo, il recupero di classi o predicati già esistenti per arricchire il modello. Varrà un esempio, vale a dire l'entry «Vespasiano da Bisticci» su Dbpedia, ovvero il progetto di estrazione di stringhe di informazione da Wikipedia per rappresentarle in formato strutturato (Fig. 11). Nell'ontologia di Dbpedia si può notare sia la creazione di proprietà ad hoc (dbprop-it:proprietà), sia il riuso di proprietà provenienti da altre ontologie (dcterms e foaf, ad esempio). Il valore della proprietà può essere un letterale o un URL/URI.

Figura 10 – Esprimere le relazioni in RDF: (soggetto = URI), predicato (nome della relazione), oggetto (URI o letterale).

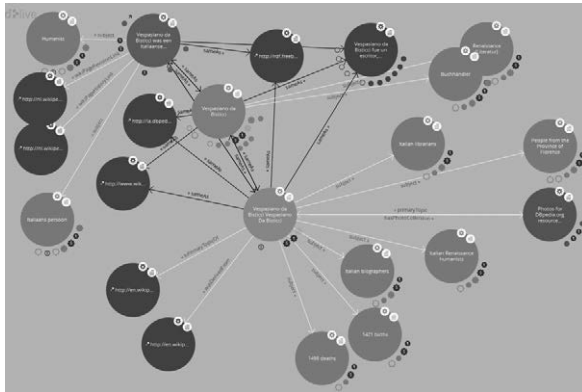
SUBJECT	PREDICATES	OBJECT
people.rdf#Pdm URI: <a href="http://vespasianodabisticci.letters/people/Pdm">http://vespasianodabisticci.letters/people/Pdm</a>	has_normalized_form	Medici, Piero de' Dbpedia: <a href="http://eu.dbpedia.org/page/Piero_de_Medici">http://eu.dbpedia.org/page/Piero_de_Medici</a> VIAF: <a href="http://viaf.org/viaf/25406033">http://viaf.org/viaf/25406033</a>
	has_variant_forms	Piero, Piero di Cosimo de' Medici, Principe di Firenze
	is_owner_of	manuscripts.rdf#P_SN manuscripts.rdf#L_D_111 manuscripts.rdf#L_D_TV_E
SUBJECT	PREDICATES	OBJECT
manuscripts.rdf#P_SN URI: <a href="http://vespasianodabisticci.letters/manuscripts/P_SN">http://vespasianodabisticci.letters/manuscripts/P_SN</a>	has_normalized_form	Pinco, Storia naturale
	is_requested_by	people.rdf#Pdm
	is_owned_by	people.rdf#Pdm
	is_illuminated_by	people.rdf#FT
SUBJECT	PREDICATES	OBJECT
lexicon.rdf#min URI: <a href="http://vespasianodabisticci.letters/lexicon/min">http://vespasianodabisticci.letters/lexicon/min</a>	has_normalized_form	miniare, miniatura, miniato
	is_referred_to	manuscripts.rdf#L_D_TV_E

Figura 11 – Entry «Vespasiano da Bisticci» su Dbpedia.it. L'entry su Dbpedia.org: <[http://dbpedia.org/page/Vespasiano\\_da\\_Bisticci](http://dbpedia.org/page/Vespasiano_da_Bisticci)> (10/2016).

Property	Value
dbpedia-owl:birthPlace	■ <a href="#">dbpedia-it:Rignano_sull'Arno</a>
dbpedia-owl:birthYear	■ 1421-01-01 00:00:00 (xsd:date)
dbpedia-owl:deathYear	■ 1498-01-01 00:00:00 (xsd:date)
dbpedia-owl:wikiPageExternalLink	■ <a href="http://www.treccani.it/enciclopedia/vespasiano_res-5b1c59c7-8bb8-11dc-8e9d-0016357ee551_(Enciclopedia-Italiana)/">http://www.treccani.it/enciclopedia/vespasiano_res-5b1c59c7-8bb8-11dc-8e9d-0016357ee551_(Enciclopedia-Italiana)/</a> ■ <a href="http://www.italica.ai.it/inascimento/parole_chiave/schede/vespasianoobisti.htm">http://www.italica.ai.it/inascimento/parole_chiave/schede/vespasianoobisti.htm</a> ■ <a href="http://firenzeccarlostia.blogspot.com/2008/02/bisticci-comune-di-riignano-sullarno.html">http://firenzeccarlostia.blogspot.com/2008/02/bisticci-comune-di-riignano-sullarno.html</a> ■ <a href="http://bivio.signum.sns.it/bwWorkTOC.php?authorSign=VespasianodaBisticci&amp;titleSign=LeVite">http://bivio.signum.sns.it/bwWorkTOC.php?authorSign=VespasianodaBisticci&amp;titleSign=LeVite</a>
dbpedia-owl:wikiPageID	■ 1788315 (xsd:integer)
dbprop-it:annomorte	■ 1498 (xsd:integer)
dbprop-it:annonascita	■ 1421 (xsd:integer)
dbprop-it:attività	■ scrittore ■ umanista
dbprop-it:attivitàaltre	■ e libralo
dbprop-it:epoca	■ 1400 (xsd:integer)
dbprop-it:giornomesemorte	■ -07-27
dbprop-it:luogomorte	■ Antella
dbprop-it:luogomortelink	■ Antella
dbprop-it:luogonascita	■ Bisticci
dbprop-it:luogonascitalink	■ <a href="#">Rignano sull'Arno</a>
dbprop-it:nazionalità	■ italiano
dbprop-it:nome	■ Vespasiano da Bisticci
dbprop-it:preattività	■ fu uno
dbprop-it: sesso	■ M

I diversi tipi di relazione che ogni entità in un dominio può intrattenere coprono diversi aspetti della rete di connessioni. Un esempio di visualizzazione attraverso LodLive<sup>18</sup> (Fig. 12) può far capire la tipologia di network che da ogni entry (in questo caso la voce «Vespasiano da Bisticci») è possibile istituire. Con LodLive è possibile ‘tradurre’ un dataset RDF in un sistema di *browsing friendly*. Le entries sono le entità raffigurate in un cerchio e le frecce etichettate sono le proprietà che permettono di mettere in una relazione, esplicitamente dichiarata, le entries:

Figura 12 – La entry «Vespasiano da Bisticci» visualizzata attraverso LodLive (alcune relazioni). <[http://lodlive.it/?http://dbpedia.org/resource/Vespasiano\\_da\\_Bisticci](http://lodlive.it/?http://dbpedia.org/resource/Vespasiano_da_Bisticci)> (10/2016).



## 5. Conclusioni

La prospettiva del Web semantico offre quindi innegabili vantaggi al mondo della filologia. Una più ampia e massiccia adesione ai modelli semantici garantirebbe una maggiore efficacia nell'*information retrieval*; inoltre, un aumento consistente di testi digitali basati su queste tecnologie arricchirebbe la base di conoscenza offrendo un'esperienza ancora più significativa all'utente finale. Attraverso i LOD, ogni concetto è identificato univocamente e le relazioni che ogni entità intrattiene vengono esplicitate, in modo tale che la conoscenza da parte del ricercatore finale risulta aumentata attraverso il corpus di dati correlati che ogni concetto intrattiene. Ogni edizione non è dunque più una monade isolata, ma una base di conoscenza semantica perché arricchita attraverso tutte le informazioni che il Web è capace di fornire su ogni singolo dato che dall'edizione è possibile

<sup>18</sup> «LodLive project provides a demonstration of the use of Linked Data standards (RDF, SPARQL) to browse RDF resources. The application aims to spread linked data principles using a simple and friendly interface with reusable techniques», <<http://en.lodlive.it/>> (10/2016).

estrarre. Il progetto *knowledge graph* di Google<sup>19</sup> ci permette una riflessione finale:

The Knowledge Graph is a knowledge base used by Google to enhance its search engine's search results with semantic-search information gathered from a wide variety of sources. Knowledge Graph display was added to Google's search engine in 2012, starting in the United States, having been announced on May 16, 2012.

It provides structured and detailed information about the topic in addition to a list of links to other sites.

The goal is that users would be able to use this information to resolve their query without having to navigate to other sites and assemble the information themselves.

## Bibliografia

- Berners-Lee T. 2006, *Linked Data*. Last modified June 18, 2009, <<https://www.w3.org/DesignIssues/LinkedData.html>> (10/2016).
- Berners-Lee T., Hendler J., Lassila O. 2001, *The Semantic Web*, «Scientific American Magazine», May 17, <<https://www.scientificamerican.com/article/the-semantic-web/>> (10/2016).
- Cagni G. M. 1969, *Vespasiano da Bisticci e il suo epistolario*, Edizioni di Storia e Letteratura, Roma: 117-185.
- Ciula A., Spence P., Vieira J. M. 2008, *Expressing Complex Associations in Medieval Historical Documents: The Henry III Fine Rolls Project*, «Literary and Linguistic Computing», 23(3), 311-325 <doi:10.1093/lc/fqn018>.
- Daquino M., Tomasi F. 2015, *Ontological approaches to information description and extraction in the cultural heritage domain*, in Tomasi F., Rosselli Del Turco R., Tammaro A. M. (eds.), *Humanities and Their Methods in the Digital Ecosystem*, ACM, New York: 1-7.
- de la Mare A. C. 1985, *New research on humanistic scribes in Florence*, in Garzelli A. (a cura di), *Miniatura fiorentina del Rinascimento 1440-1525*, Giunta Regionale Toscana-La Nuova Italia, Firenze, App. III, 17 e 18: 567.
- DeRose S. J., Durand D. G., Mylonas E., Renear A. H. 1997, *What Is Text, Really?*, «ACM SIGDOC Asterisk: The Journal of Computer Documentation», XXI (3): 1-24 <doi:10.1145/264842.264843> (10/2016).
- Eide Ø. 2015, *Ontologies, Data Modeling, and TEI*, «Journal of the Text Encoding Initiative», 8: 1-22, <<http://jtei.revues.org/1191>> (11/2016).
- Modern Language Association 2011, *Guidelines for Editors of Scholarly Editions*. Last revised 29 June, <<https://www.mla.org/Resources/Research/Surveys-Reports-and-Other-Documents/Publishing-and-Scholarship/Re>

<sup>19</sup> Una descrizione del *knowledge graph* di Google è disponibile in <<http://www.google.com/intl/es419/insideseach/features/search/knowledge.html>> (10/2016).

- ports-from-the-MLA-Committee-on-Scholarly-Editions/Guidelines-for-Editors-of-Scholarly-Editions> (11/2016).
- O'Reilly T. 2005, *What is Web 2.0*, <<http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>> (11/2016).
- Schmidt D. 2010, *The inadequacy of embedded markup for cultural heritage texts*, «Literary and Linguistic Computing», XXV (3): 337-356.
- Singhal A. 2012, *Introducing the knowledge graph: things, not strings*. Google blog, <<https://googleblog.blogspot.it/2012/05/introducing-knowledge-graph-things-not.html>> (11/2016).
- Sperberg-McQueen C. M., Huitfeldt C., Renear A. 2000, *Meaning and Interpretation of Markup*, «Markup Languages: Theory & Practice», II (3): 215-234.
- TEI Consortium 2015, *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 2.8.0. Last updated April 6 2015, <<http://www.tei-c.org/Vault/P5/2.8.0/doc/tei-p5-doc/en/html/>> (11/2016).
- Tomasi F. 2013, *L'edizione digitale e la rappresentazione della conoscenza. Un esempio: Vespasiano da Bisticci e le sue lettere*, «Ecdotica», 9: 264-286.
- Tummarello G., Morbidoni Ch., Pierazzo E. 2005, *Toward Textual Encoding Based on RDF*, in *Proceedings 9th ICC International Conference on Electronic Publishing*, Peeters, Leuven: 57-63, <<http://elpub.scix.net/data/works/att/206el-pub2005.content.pdf>> (11/2016).

**Giuliana Diani** è professoressa associata di Lingua e Traduzione inglese presso l'Università di Modena e Reggio Emilia. Ha ottenuto un *Master in Language studies* presso l'Università di Lancaster e il dottorato di ricerca in Linguistica inglese presso l'Università di Pisa. Ha lavorato su vari aspetti dell'analisi del discorso e dell'inglese accademico, con particolare riferimento alla variazione linguistica nei diversi generi accademici, discipline e culture, attraverso l'analisi di corpora specializzati. Le sue ultime ricerche si concentrano sull'analisi linguistica negli ambiti della diffusione della conoscenza nel settore turistico e giuridico. Tra gli ultimi libri da lei pubblicati vi sono: con Paul Thompson, *English for Academic Purposes: Approaches and Implications* (Cambridge Scholars, 2015); *Reviewing Academic Research in the Disciplines: Insights into the Book Review Article in English* (Officina Edizioni, 2012); e con Ke Hyland, *Academic Evaluation: Review Genres in University Settings* (Palgrave Macmillan, 2009).

**Iris Eshkoll-Taravella** è ricercatrice abilitata presso l'Università di Orléans, membro del laboratorio LLL (Laboratoire Ligérien de Linguistique – UMR 7270), lavora nell'ambito della linguistica dei corpora e del trattamento automatico delle lingue. Ha dedicato la maggior parte delle sue ricerche all'annotazione di corpora, che si tratti di sintassi (etichettatura morfosintattica), semantica e/o pragmatica (*named entities*, indicizzazione automatica di persone, riformulazioni, sentimenti ecc.). Le applicazioni delle sue ricerche spaziano da interviste trascritte a titoli di carte geografiche, testi procedurali e articoli di «Le Monde».

**Nathalie Friburger** è ricercatrice abilitata in Informatica e Linguistica, membro del LI (Laboratoire d'informatique – EA6300) dell'Università François-Rabelais di Tours, si occupa del Trattamento Automatico delle



Lingue e delle cascate di trasduttori a stati finiti, di *Text data mining* e, in particolare, della ricerca di *Named entities* in testi di diversa provenienza. Quest'ultima ricerca ha dato luogo a diverse partecipazioni alle campagne di valutazione *Ester* 2008, e in seguito, con ottimi risultati, a *Etape* nel 2012 (sistema CasSys), progetto su testi orali trascritti manualmente e automaticamente. Ha codiretto nel 2013 il numero tematico n. 54 (2) *Entités nommées* della rivista TAL (*Traitement automatique des langues*).

**Séverine Gedzelman**, ingegnere informatico, è membro del laboratorio Triangle (CNRS, ENS Lione). Ha sviluppato il programma *HyperMachiavel* dal 2008 per rispondere alle esigenze di analisi e di annotazione di corpora allineati. Ha partecipato a varie reti di diffusione della ricerca (<https://ahnenslyon.hypotheses.org/>) e a progetti di *Digital Humanities*, tra cui il corpus online *HyperPrince* sulle prime traduzioni francesi del *Principe* di Machiavelli (<http://hyperprince.ens-lyon.fr>) e *Siprojuris*, una base prosopografica sugli universitari di giurisprudenza (<http://siprojuris.symogih.org/>). I suoi interessi si concentrano sui metodi qualitativi e quantitativi di ricerca e sulla loro estensione negli strumenti digitali; ha inoltre partecipato alla riflessione sul loro impatto nella ricerca in scienze umane e sociali attraverso la rete di APN (*Ateliers Pratiques Numériques*).

**Denis Maurel** ha conseguito un dottorato di ricerca in Informatica presso l'Università di Parigi VII nel 1989, l'abilitazione a dirigere ricerche (HDR) nel 1996, l'abilitazione ad insegnare (*Qualification*) in Informatica e in Linguistica, ed è professore titolare presso l'Università François-Rabelais di Tours dal 2000. È responsabile del gruppo di ricerca BdTln (Banche dati e Trattamento delle lingue naturali) del LI (Laboratoire d'informatique – EA6300). È inoltre membro associato del LLL (Laboratoire ligérien de linguistique - UMR 7270) e del comitato direttivo delle conferenze CIAA (*International Conference on Implementation and Application of Automata*), nonché responsabile delle recensioni della rivista TAL (*Traitement automatique des langues*). Le sue ricerche vertono sul trattamento automatico delle lingue, principalmente mediante automi e trasduttori a stati finiti: i lessici, il trattamento automatico dei nomi propri (progetto *Prolex*) e, meno di recente, le metodologie di comunicazione con persone disabili.

**Ana Pano Alamán** è ricercatrice in Lingua e Traduzione spagnola presso il Dipartimento di Lingue, Letterature e Culture Moderne dell'Università di Bologna. Si occupa da anni del rapporto tra lingua, letteratura e nuove tecnologie. Le sue principali aree di ricerca sono l'analisi linguistica e pragmatica del discorso digitale, le applicazioni delle tecnologie (*corpus linguistics* e strumenti 2.0) alla ricerca in lingua e linguistica e la storia della traduzione. Su questi temi ha pubblicato numerosi articoli su riviste nazionali e internazionali e la monografia *Dialogar en la Red. La lengua española en chats*,

*emails, foros y blogs* (Peter Lang, 2008); è co-autrice dei volumi *El discurso político en Twitter* (Anthropos, 2013), *El español coloquial en las redes sociales* (Arco, 2013); *Los avatares del Quijote en Europa* (Cátedra, 2010) e *Tradurre un classico della scienza. Traduzioni dell'Origine delle specie in Francia, Italia e Spagna* (BUP, 2015). È membro dell'associazione internazionale *Humanidades Digitales Hispánicas*.

**Christina Samson** è docente di Lingua inglese presso la Scuola di Studi Umanistici e della Formazione e presso la Scuola di Economia e Management dell'Università degli Studi di Firenze. È titolare di un dottorato di ricerca in Linguistica Applicata. La sua ricerca è principalmente incentrata sull'analisi di corpora specialistici di generi diversi, integrata con l'analisi del discorso. La sua attenzione si è particolarmente rivolta al lessico e al discorso dei beni culturali e alla descrizione verbale del punto di vista, dello spazio, della valutazione e del discorso promozionale nei siti museali, nelle guide online, come pure nella rendicontazione annuale aziendale online. È co-fondatrice dell'Unità di Ricerca *Lessico dei Beni Culturali* del Dipartimento di Lingue, Lettere e Studi Interculturali (Università di Firenze) e co-coordina la creazione del portale *Dizionario dei Beni Culturali Fiorentini* rivolto a traduttori e studenti universitari.

**Monica Taddei** è bibliotecaria presso la Biblioteca leonardiana di Vinci. Centro di ricerca e documentazione per gli studi leonardiani. Nell'ambito della sua attività cura tra l'altro la redazione della *Bibliografia internazionale leonardiana* che viene pubblicata a stampa con cadenza generalmente biennale sulla rivista *Raccolta Vinciana* dall'Ente Raccolta Vinciana di Milano, oltre che online nel catalogo della biblioteca. È responsabile delle due collane editoriali del Centro 'Lectura Vinciane' e 'Biblioteca leonardiana. Studi e documenti'. I suoi studi vertono su questioni di bibliografia, biblioteconomia, in particolare sui temi della conservazione, valorizzazione, comunicazione, fruizione e promozione delle biblioteche speciali e dei loro strumenti di accesso/servizi all'utenza in relazione al patrimonio bibliografico e documentario della biblioteca di appartenenza. A partire dal 2005 circa, anche a seguito dell'ideazione e progettazione di *e-Leo. Archivio digitale di storia della tecnica e della scienza* di cui è co-autrice insieme a Romano Nanni e Gaetano Cascini, il suo interesse per il tema della valorizzazione delle biblioteche speciali si è focalizzato sullo sviluppo e realizzazione di funzionalità di base ed avanzate, anche semanticamente raffinate, per l'accesso, analisi e studio di depositi documentari non propriamente definibili di 'pubblica lettura', quali quello del *corpus* leonardiano dell'istituto vinciano.

**Francesca Tomasi** è ricercatrice dal 2006 presso il Dipartimento di Filologia Classica e Italianistica dell'Università di Bologna. Si occupa di Informatica umanistica (*Digital humanities*) con un'attenzione speciale ai metodi della

progettazione computazionale di dati umanistici nel solco dei sistemi di organizzazione della conoscenza, in particolare, nel settore dell'edizione digitale di testi e documenti nei sistemi di biblioteche ed archivi digitali. È *visiting researcher* al DDH (*Department of Digital Humanities*) del King's College di Londra dal 2009. Dal 2014 è Presidente della BDU (Biblioteca di Discipline Umanistiche) dell'Università di Bologna. Ha partecipato a numerosi convegni di *digital humanities* sia come relatore che come membro del comitato scientifico e ha pubblicato numerosi contributi in materia su rivista e in volume. È autrice di cinque monografie (l'ultima con D. Fioromonte e T. Numerico, *The Digital Humanist. A Critical Inquiry*. Punctum books, 2015) e curatrice di numerosi volumi, nonché responsabile dell'edizione digitale delle *Lettere* di Vespasiano da Bisticci.

**Jean-Claude Zancarini**, professore emerito di Italianistica, è membro del laboratorio Triangle (CNRS, *École Normale Supérieure* de Lyon). La sua ricerca verte sui legami tra storia, lingua e politica nell'Italia del XVI secolo, in particolare sul pensiero politico fiorentino all'epoca delle guerre d'Italia. Conduce in parallelo studi filologici (cura e traduzione di testi di Savonarola, Guicciardini e Machiavelli) e lavori di esegesi (con J.-L. Fournel ha pubblicato *La politique de l'expérience. Savonarole, Guicciardini et le républicanisme florentin*, Edizioni dell'Orso, 2002 e *La grammaire de la République. Langage de la politique chez Francesco Guicciardini (1483-1540)*, Droz, 2009). Insieme a Séverine Gedzelman, ha ideato e realizzato lo strumento *Hyper-Machiavel*, presentato in questo volume.

**Valeria Zotti** è ricercatrice in Lingua e Traduzione francese presso l'Università di Bologna dal 2006 e abilitata al ruolo di professore di seconda fascia dal 2013. Ha conseguito un Master in lessicografia, terminografia e trattamento automatico dei corpora (Università di Lille) ed è titolare di un dottorato di ricerca in Francesistica (Università di Bari e Università di Cergy-Pontoise). Ha inoltre effettuato diversi stage di ricerca in Canada (Università Laval à Québec). I suoi campi di ricerca sono la lessicologia, la lessicografia francese, monolingue, bilingue e differenziale, e il trattamento automatico della lingua. Attualmente si occupa di traduzione letteraria e nuove tecnologie (progetto QU.IT.) e della terminologia del linguaggio artistico e gastronomico (progetto LBC). Oltre a numerosi articoli pubblicati in Italia e all'estero, è autrice della monografia *Dictionnaire bilingue et francophonie. Le français québécois* (Schena, 2007), della banca dati di traduzione QU.IT (Università di Bologna, CRR-MM, 2013) ed è coautrice di varie curatele tra cui si segnalano: *Traduire des français: des mots et des mondes*, *Études de Linguistique Appliquée* (Didier, 2011, n. 164) e *La variation des français. Dictionnaires, bases de données, corpus* (Honoré Champion, 2014).

STRUMENTI  
PER LA DIDATTICA E LA RICERCA

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, *Lomini 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 Bertirrotti, *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 Piovani, *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 concorsua-*

- li 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 Burresti (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 Masseti, *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 empoiese*
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 Scampoli, *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), *MU-SINT. 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 Glimpse at Work. Educational Perspectives*
  130. Raffaele Donvito, *L'innovazione nei servizi: i percorsi di innovazione nel re-tailing 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 Welllitteratur. 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, *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 P.O.M.A. Museo*
157. Fortunato Sorrentino, Maria Chiara Pettenati, *Orizzonti di Conoscenza. Strumenti digitali, metodi e prospettive per l'uomo del terzo millennio*
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 Urbani, 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'archeosismologia 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'Ipertensione ARTERIOSA*
175. Enzo Manzato, 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 esperien-*



- ze 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*



