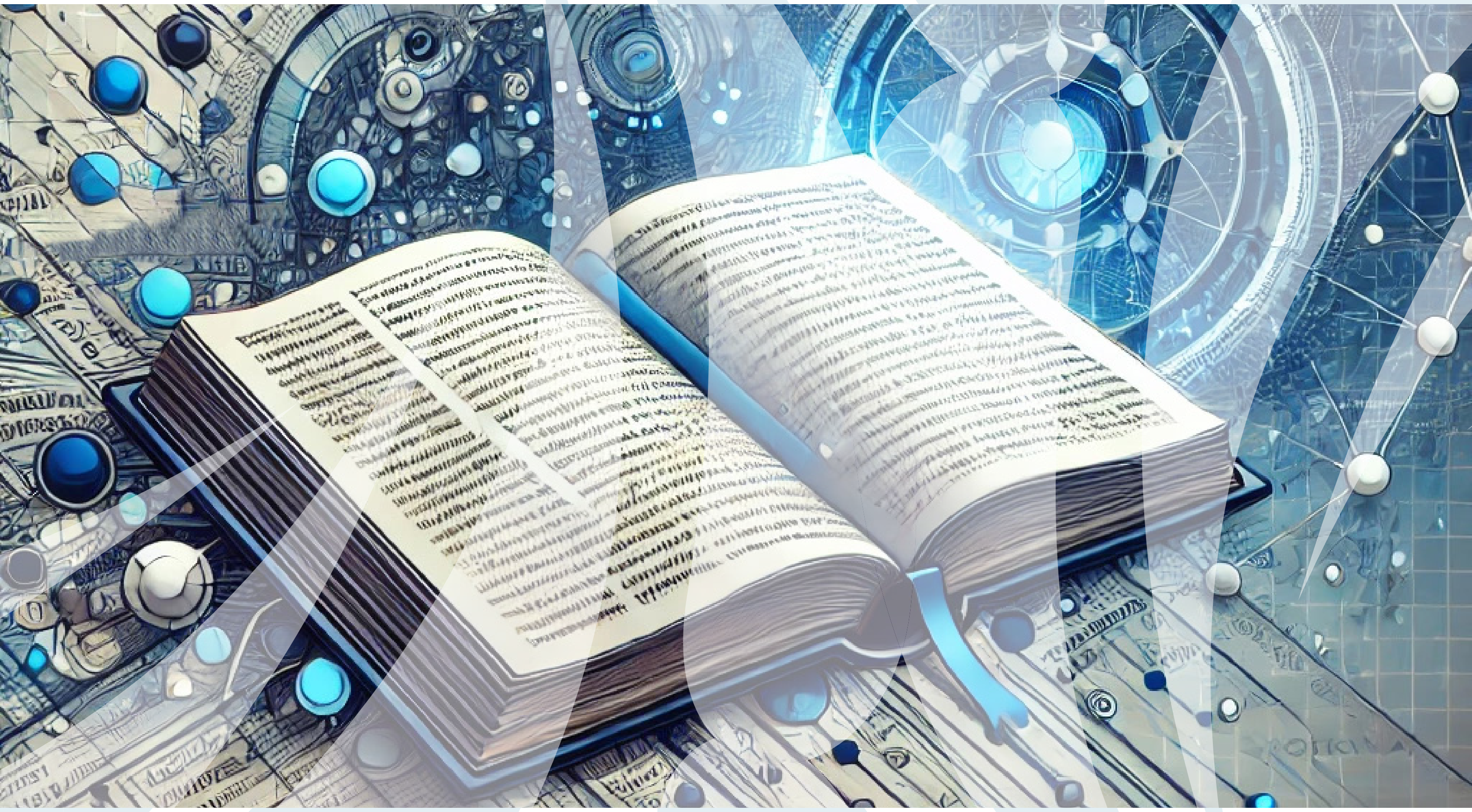
Kristina Š. Despot
Ana Ostroški Anić
Ivana Brač (Eds.)

Lexicography and Semantics



Proceedings of the
XXI EURALEX International Congress
8–12 October 2024
Cavtat, Croatia

Nakladnik: Institut za hrvatski jezik
Za nakladnika: Željko Jozić

Knjiga: Lexicography and Semantics,
Proceedings of the
XXI EURALEX International Congress

Urednice: Kristina Štrkalj Despot
Ana Ostroški Anić
Ivana Brač

Tehničko uređenje i računalni slog: Elena Vrbanić
Oblikovanje naslovnice: Elena Vrbanić
Tisak:

© This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License, 2024.

ISBN 978-953-7967-77-2

Kristina Š. Despot, Ana Ostroški Anić, Ivana Brač (Eds.)

Lexicography and Semantics

Proceedings of the
XXI EURALEX International Congress
8–12 October 2024
Cavtat, Croatia

TABLE OF CONTENTS

FOREWORD......11

PART I: CONFERENCE OVERVIEW

Acknowledgements......15
  Main Sponsors......15
  Sponsors......16
  Programme Committee......17
  Local Organising Committee......17
  Scientific Committee......17
Overview of Keynotes and Workshops......19
Keynotes......19
Keynotes of the Workshop Figurative Language and Large Language Models......19
Workshops......19

PART II: PROCEEDINGS

Chapter I. Lexicography and Semantics......21

Tony Veale
You Talk Funny! Someday Me Talk Funny Too!
– On Learning to See the Humorous Side of Familiar Words......22

Valeria Caruso, Lucia di Pace
Words for Choosing Food in *ALMA – Multimedia Atlas of Bio/Cultural Food*......35

Janet DeCesaris, Mercè Lorente Casafont
Old Words, New Terms – Semantic Broadening and
Narrowing in the Vocabulary of the Circular Economy......49

Maucha Gamonal, Adriana Pagano, Tiago Torrent, Ely Matos, Arthur Lorenzi
Automated Semantic Frame Annotation –
An Exploratory Study in the Health Domain......61

Ellert Thor Johannsson, Thordis Ulfarsdottir
The Role of Semantic Fields in Contemporary Icelandic Dictionaries......75

Robert Krovetz
Morpho-Semantics and Dictionary Entries......86

Haniva Yunita Leo
Do Indonesian Speakers Feel ‘pain’? NSM and Corpus-Based
Approach to the Cross-Linguistic Concept of ‘pain’ in Bahasa Indonesia......101

Michael Nguyen, Peter Juel Henrichsen
The Locum Hyphen – A Formal Approach to the Lexicalization
of Multiword Expressions With Rich Internal Semantics......113

Pär Nilsson
Figurative, Transferred or Extended use? The Use of Semantic Labels in the
First Edition and the Revised Version of the Swedish Academy Dictionary............................122

Sanni Nimb, Ida Flörke, Sussi Olsen, Bolette S. Pedersen, Nathalie C. H. Sørensen
COR.SEM, a New Formal Semantic Lexicon for Danish..131

Andrej Perdih, Kozma Ahačič, Nataša Jakop, Nina Ledinek, Špela Petric Žižić
Semantic Information on the Franček Educational Language Portal for Slovenian...............144

Emma Sköldberg, Shafqat Mumtaz Virk, Pauline Sander,
Simon Hengchen and Dominik Schlechtweg
Revealing Semantic Variation in Swedish Using Computational Models of
Semantic Proximity – Results From Lexicographical Experiments...158

Petra Storjohann
Synonyms in Contrast – A Dynamic and Descriptive
Resource for New Semantic Equivalents...172

Lars Trap-Jensen, Henrik Lorentzen
Back to Basics – Meaning Description for Human Users and for Computers.........................179

Ene Vainik, Geda Paulsen, Heete Sahkai, Jelena Kallas,
Arvi Tavast, Arvi Tavast, Kristina Koppel
From a Dictionary to a Constructicon – Putting the Basics on the Map.................................196

Chapter II. Lexicography and Language Technologies..204

Igor Boguslavsky
Lexical Resources for a Semantic Parser of Russian
– Argument Structure of Ordinal Adjectives...205

Nataliia Cheilytko, Ruprecht von Waldenfels
Word Embeddings for Detecting Lexical Semantic Change in Ukrainian...............................216

Lian Chen, Wenjun Sun, Flora Badin
Innovation in Phraseomatics – DiCoP Project and DiCoP-Text Corpus
for the Enrichment of Language Models and Automatic Translation.......................................227

Enikő Héja, Kata Gábor, László Simon, Veronika Lipp
Graph-based Detection of Hungarian Adjectival Meaning
Structures via Monolingual Static Embeddings...235

Geraint Paul Rees, Isabel Gibert
A Textbook or ChatGPT – Which Helps Novice
Programmers Most with Unknown Terms?...248

Bálint Sass
The "Dependency Tree Fragments" Model for Querying a Constructicon..............................257

Dragana Špica, Benedikt Perak
Enhancing Japanese Lexical Networks Using Large Language Models – Extracting Synonyms and Antonyms with GPT-4o....265

Chapter III. Dictionary Writing Systems and Lexicographic Tools....285

Michaela Denisová, Gilles-Maurice de Schryver, Pavel Rychlý
The Automatic Determination of Translation Equivalents in Lexicography: What Works and What Doesn't?....286

Stephanie Evert, Christine Ganslmayer, Christian Rink
Multi-Level Analysis as a Systematic Approach to Evaluating the Quality of AI-Generated Dictionary Entries....298

Carolina Flinz, Daniel Henkel, Valeria Zotti, Sabrina Ballestracci
A Multilingual Parallel Corpus for the Lexical Information System LBC – Recent Progress and Future Perspectives....316

Peter Juel Henrichsen
Make Each Morph Count – A New Approach to Computational Lexicography for Text Processing....329

František Kovařík, Vojtěch Kovář, Marek Blahuš
On Rapid Annotation of Czech Headwords – Analysing the First Tasks of Czech Dictionary Express....336

Irene Renau, Rogelio Nazar, Daniel Mora Melanchthon
Towards the Automatic Generation of a Pattern-Based Dictionary of Spanish Verbs....345

Chapter IV. Reports on Lexicographical and Lexicological Projects....361

Juris Baldunčiks, Silga Sviķe
Pages of Latvian Historical Slang Dictionary: *dzeršana* ('drinking')....362

María Auxiliadora Barrios Rodríguez
Diretes, a Spanish Monolingual Dictionary Based on Lexical-Semantic Relations....369

Luke Omoyemi Akinremi, María José Domínguez Vázquez
The Erasmus Mundus Joint Masters in Lexicography – EMJM-EMLex – New Developments and Goals....383

Dwayne Ellul
Maltese Lexicography – A Historical Context and the Current State....389

Ivana Filipović Petrović, Kristina Kocijan
Creating the Dataset of Croatian Verbal Idioms – Automatic Identification in a Corpus and Lexicographic Implementation....405

Louise Holmer, Ann Lillieström, Emma Sköldberg, Jonatan Uppström
Time to Say Goodbye Revisited – On the Exclusion of Headwords from the Swedish Academy Glossary (SAOL)........419

Kathryn Hudson
The Benefits of Bio(lexicography) – A Topical Approach to Lexicographic Practice........429

Boris Kern
Considering Word Formation in Compiling Dictionaries........438

Anas Fahad Khan, Ana Salgado, Isuri Anuradha, Rute Costa, Chamila Liyanage, John P. McCrae, Atul K. Ojha, Priya Rani and Francesca Frontini
Cultural HeritAge and Multilingual Understanding through lexiCal Archives (CHAMUÇA) – Portuguese Borrowings in Contemporary Asian Languages........449

Veronika Kolářová, Jiří Mírovský
Looking for Sense in Nonsense – Valency of Negative Forms of Nouns and Adjectives in the NomVallex Lexicon........459

Kusujiro Miyoshi
John Pickering's Reference Materials for His *Vocabulary* (1816) – Transcending the Bounds of Dictionaries........471

Pär Nilsson
Report on the Revision of the Swedish Academy Dictionary – and the Search for "Old Neologisms"........481

Sanni Nimb, Nathalie C. H. Sørensen, Jonas Jensen
Making Danish Thesaurus Data Available to Researchers – The WebDDB project........497

Christian-Emil Smith Ore, Oddrun Grønvik
The Spoken Word as Represented in Norsk Ordbok........504

Petya Osenova, Kiril Simov
All About Words! An Integrated Dictionaries Portal for Bulgarian........520

Vanja Štefanec, Krešimir Šojat, Matea Filko
CroDeriv – Search and Visualization Interface........529

Chapter V. Bi- and Multilingual Lexicography........535

Vladimír Benko, Zuzana Kříhová, Boris Lehečka, Darina Vystrčilová
Persian to Czech Dictionary – A Traditional Dictionary in the Era of AI?........536

Cormac Breathnach, Pádraig Ó Mianáin
Making a Molehill out of a Mountain: Technical and Editorial Considerations in Producing the Concise English-Irish Dictionary (2020)........551

Elina Chadjipapa, Zoe Gavriilidou
Helix – A Bilingual Illustrated Dictionary for Greek Heritage Learners........566

Rajna Dragićević, Yury Makarov, Daria Ryzhova, Yulia Shapich, Ekaterina Yakushkina
A New Serbian-Russian Dictionary.....576

Mariusz Piotr Kamiński
The Contribution of Bilingualized Entries and Vocabulary Knowledge to the Learner's Success in Sentence Completion – The Case of *jump* Verbs.....584

Evelina Kirsakmene
False Friends in General Bilingual Dictionaries (English and French into Latvian.....591

Daria Lazić
Lexicographic Treatment of Vocabulary Related to Age – The Example of Croatian and Danish.....600

Irina Lobzhanidze, Rusudan Gersamia, Nino Tsulaia
Compiling a Bilingual Megrelian-English Online Dictionary – Preserving Endangered Kartvelian Languages.....614

Chenlu Yu
A German-Chinese e-Dictionary of Manufacturing Technology in Automotive Industry – Entry Design.....627

Chapter VI. Specialized Lexicography, Terminology, and Terminography.....635

Andrea Abel, Natascia Ralli
Gender in Electronic Dictionaries and Terminology Databases – State of the Art and Future Directions.....636

Ieda Maria Alves, Beatriz Curti-Contessoto, Ana Maria Ribeiro de Jesus
Challenges of Creating a Medical Dictionary for a Low Literacy Audience in Brazil – Focusing on Politically Marked Terms Related to the COVID-19 Pandemic.....653

Lynne Bowker
Eponyms, EDI and Terminology Planning in the Biological Sciences.....662

Theresa Kruse, Ulrich Heid, Barbara Schmidt-Thieme
Mathematics Students as Lexicographers – Learning Domain Concepts and Their Relations by Designing Dictionary Articles and Concept Maps.....673

David Lindemann
Teaching Terminology through Wikibase and Wikidata.....680

Bruno Nahod
Can We Substitute Field Experts with Customized Large Language Model in Processing Specialized Languages? – A Case Study.....686

Chapter VII. Dictionary (in) Use..699

Margit Langemets, Lydia Risberg, Kristel Algvere
To Dream or Not to Dream About 'Correct' Meanings?
– Insights into the User Experience Survey..700

Tinatin Margalitadze, Katalin P. Márkus
Cross-Border Collaboration in Teaching Dictionary Skills..720

Barbora Štěpánková, Lucie Poláková, Jana Šindlerová, Michal Novák
What Can Dictionaries Tell Us About Pragmatic Markers
– Building the Lexicon of Epistemic and Evidential Markers in Czech..728

Chapter VIII. Historical and Dialect Lexicography..742

David Lindemann, Mikel Alonso
Linking Historical Corpus Data and Annotations Using Wikibase..743

Ivana Lovrić Jović, Martina Kramarić
The Dubrovnik Idiom Through Time – Crafting a Diachronic Dictionary..749

Magdalena Majdak
Defining Meanings in Historical Dictionaries – The Case of the
Electronic Dictionary of the 17th- and 18th-Century Polish..763

Martina Waclawičová
Dialect Dictionary and Lexicalization of Dialect Phenomena..776

Leonardo Zilio, Besim Kabashi
Using Neural Machine Translation for Normalising Historical Documents..783

Carolina Flinz, Daniel Henkel, Valeria Zotti, and
Sabrina Ballestracci

A MULTILINGUAL PARALLEL CORPUS FOR THE LEXICAL INFORMATION SYSTEM LBC

Recent Progress and Future Perspectives

Abstract The LBC-Platform (https://www.lessicobeniculturali.net) is a comprehensive lexical information system that aims to integrate various types of corpora and resources: dictionaries, concordances, monolingual Language for Special Purposes (LSP) corpora in different languages and LSP parallel corpora. Designed for users interested in cultural heritage, the platform provides free access to resources and tools such as NoSketch Engine, facilitating an in-depth understanding of the language of Fine Arts — encompassing painting, architecture, and sculpture — and highlighting its multidisciplinary nature and diverse discourse types. Numerous studies have demonstrated the platform's utility for various purposes, including lexicography, translation, and didactics. This paper focuses on the preparation of the multilingual Vasari parallel corpus, part of the LBC Project's four planned subcorpora (Vasari corpus, Literature corpus, Museum Web-Site Corpus, Travel Guide Corpus). Special attention is given to the Italian-English, Italian-French, and Italian-German components of the corpus, based on G. Vasari's *Le vite de' più eccellenti architetti, pittori, et scultori italiani* (1568) and its translations. In addition to methodological choices including alignment strategies and potential technical solutions, a small case study on the words *disegno* ('drawing') and *disegnare* ('to draw') will serve to illustrate the current status of the project and its relevance for LBC dictionaries, as well as future perspectives.

Keywords cultural heritage; parallel corpora; Language for Specific Purposes; artistic lexicon; translation

1. Introduction[1]

Corpora of different kinds (LSP monolingual comparable corpora, LSP parallel corpora) form the core of the LBC-Platform,[2] a lexical information system (Klosa, 2016) created for different types of users interested in cultural heritage (Billero, Farina, & Nicolás Martínez, 2020) by the "Multilingual Cultural Heritage Lexicon" Research Unit of the University of Florence, in collaboration with scholars from Italian and foreign universities (Bologna, Paris 8, Pisa, Milan etc.). Resources (corpora, concordances)

---

[1] The four authors have written the paper jointly. Carolina Flinz and Sabrina Ballestracci are responsible for §1 and §2.3.3, Daniel Henkel for §2.2 and §2.3.1, Valeria Zotti for §2.1 and §2.3.2. Sections §3 and §4 were written jointly. In particular Carolina Flinz and Sabrina Ballestracci are responsible for the Italian-German part, Daniel Henkel for Italian-English and Valeria Zotti for Italian-French. We would also like to thank the trainees from the University of Florence (Sara Biagini, Carlotta Buricchi, Maria Chiara Susini) for the Italian-German alignment, as well as Valeria Zotti's trainee from the University of Bologna, Elisa Squadrito, for the Italian-English alignment and part of the Italian-French one.

[2] Cfr. https://www.lessicobeniculturali.net

This paper is part of the publication: Despot, K. Š., Ostroški Anić, A., & Brač, I. (Eds.). (2024). *Lexicography and Semantics. Proceedings of the XXI EURALEX International Congress.* Institute for the Croatian Language.

and tools (NoSketch Engine) are freely accessible for the user (Billero, 2020), providing a comprehensive vision of the language of Fine Arts (painting, architecture, and sculpture), characterised by multidisciplinarity and interplay between discourse-types (Zotti, 2022).

The use and application of the comparable corpora present on the platform for different aims (lexicography, translation, didactics etc.) have been already demonstrated in different studies (most recently, see Ballestracci, Buffagni, & Flinz, 2024; Farina & Sini 2024; Natali, 2024).

A preliminary study by Luporini (2023), focusing on De Vere's English translation of Vasari's masterpiece, which had not been aligned yet at the time of the study (see 2.1), shows how information provided by general dictionaries can be enriched also by information from specialized parallel corpora. Analysing the word *disegno* ('drawing') and its collocations in Italian, together with a small number of translated segments[3] into English, Luporini reveals the inherent polysemy of this lexeme, described as a "cline" from abstract (skill) to concrete (graphic representation), underlining the need for a parallel database to resolve translation difficulties.

The focus of this paper is on the preparation of the multilingual Vasari parallel corpus, one of the four subcorpora (Vasari corpus, Literature corpus, Museum Web-Site-Corpus, Travel Guide Corpus) of the LBC-Project which are in planning. In particular, attention will be given to the Italian-English, Italian-French and Italian-German components of the corpus based on a series of artists' biographies published in 1568, Giorgio Vasari's *Le vite de più eccellenti architetti, pittori, et scultori italiani* ("Lives of the Most Excellent Painters, Sculptors, and Architects," abridged in *Le Vite*) and its translations.

The discussion of methodological issues will encompass alignment (see 2.), possible technical solutions, the status of the project at present by giving a small case study (see 3.), its relevance for the LBC-dictionaries and future perspectives.

2. The Vasari Parallel Corpora: Methodology and First Results

2.1 The Source Text: *Le Vite* by Vasari

The decision to build a corpus on Vasari's *Le Vite* is due to its historic importance as the seminal text in Art History. In *Le Vite*, Vasari, a renowned Renaissance Italian painter and architect, describes the lives and works of Italian artists from the early 14th century (e.g., Cimabue, Giotto) to the second half of the 16th century (e.g., Leonardo da Vinci, Raphael and Michelangelo), for which he is recognized as the originator of Italian art criticism (Biffi, 2017).

Vasari published two editions of his work, the Torrentiniana in 1550, followed by the amplified Giuntina in 1568. In the 1568 edition, there are a total of 155 lives dedicated to various artists (about 834,000 words). Throughout a work as vast as *Le Vite*, which touches on many problems of a technical nature, such as artistic techniques, materials and tools, Vasari tried to conserve the technical vocabulary used by the artists who had preceded him (Le Mollé, 1988, p. 9). However, he did not have complete command of the Italian vocabulary, which was still evolving, and was sometimes obliged to 'invent' his own vocabulary to denote concepts that had either fallen into disuse, because the tools, techniques and recipes had been abandoned (Le Mollé, 1988, p. 7) or were entirely new and not yet stabilised in use. Vasari's text is thus considered at the origin of the artistic lexicon.

---

[3] Before sentence-level alignment had been completed, it was impossible to automatically extract source/target segment pairs, hence the use of a small sample of source/target pairs taken manually from both texts.

His text has been translated into several languages over the centuries, thus contributing to the development of a pan-European lexicon of artistic terminology (Zotti & Henkel, 2024, p. 13). These translations are the means by which the Italian artistic proto-lexicon has been adapted to other languages. Translations are therefore a field in which specialists from different linguistic and cultural backgrounds explore the networks of meanings in Vasari's text to forge new vocabulary and concepts (Dubus & Fiorato, 2017, p. 32). The analysis of translations of *Le Vite*, mostly based on the more complete Giuntina edition, reveals different lexical choices over almost five centuries in different languages.

The translations gave rise to a veritable 'artistic terminology' which has subsequently continued to evolve. Over the centuries, this specificity has fostered an abundance of cross-linguistic synonyms, parasynonyms or hyponyms in translation which reflect the complexity inherent in describing an artistic heritage linked to a unique historical, social and geographical context. Whereas the Italian source text, however, has remained unchanged since the 16th century, the ongoing creation of a multilingual database of target texts belonging to different centuries is necessarily diachronic in essence. Most translations of *Le Vite* made in the 17th and 18th centuries focus only on the biographies of the artists known best in the culture of the target language. For example, one of the first translations in France in the 17th century, Daret (1651), concerns the life of the painter Raphael. However, given that abridged translations are much more difficult to align, requiring many more manual interventions to deal with untranslated portions, preference has been given to complete translations for the purpose of creating a parallel database.

2.2 Alignment Methodology

The constitution of the parallel corpus follows a common alignment methodology for all languages. Source and target texts are pre-aligned at sentence level using LF Aligner and manually proofread by student interns following a common protocol under the supervision of the project coordinators, with further quality control carried out using Okapi Checkmate to detect suspicious disparities in length between source and target segments, a potential sign of misalignment. After automatic (pre)alignment, the first step in the proofreading process focuses on empty cells, as these necessarily indicate non-correspondence between the source and target texts. Most often, correspondences can be found between sentences in the source text and sentences or sentence fragments in the target text. However, when additions or omissions are found, these are marked respectively as #NoSource# or #NoTarget#. Subsequently, as the risk of misalignment is greater for groups of sentences than for single sentences, the second step in the proofreading process focuses on sentences which have been grouped together into a single segment. Although "segments" generally correspond to "sentences", whenever sentences in the source text correspond to more than one sentence in the target text, LF Aligner groups them together automatically. When this occurs in the other direction, i.e., two or more sentences in the source correspond to a single sentence in the target text, a source segment containing several sentences would be incompatible with other translations in other languages or from other periods, as other translators could hardly be expected to have made exactly the same choice to combine more than one sentence in exactly the same places. In addition to correcting misaligned segments, the proofreading protocol thus involves re-establishing the original sentence divisions in the source text so that it can serve as a common denominator for other translations from different periods or into different languages, with modifications in sentence segmentation (groupings, partial sentences) restricted solely to the target text.

Following alignment and proofreading, both source and target texts are tagged for Part of Speech (POS) and lemma in TreeTagger and converted to several formats: UTF-8 text (.txt), spreadsheets (.ods and .xlsx) and Translation Memory eXchange (.tmx, a specific XML standard for encoding bilingual or multilingual documents) to ensure compatibility with the greatest possible number of software tools.

2.3 Current Status and First Results

In the following three sections we briefly outline the current status of the alignment work for the Italian-English corpus, the Italian-French corpus and the Italian-German corpus.

2.3.1 Italian-English

Although *Le Vite* has been translated into English at different periods (Aglionby, 1685; Foster, 1850 *inter alia*), De Vere's 1912–1915 English translation in 10 volumes was selected for inclusion in the LBC parallel database as it is the only English translation currently available in electronic format and is also the closest to a full translation, with only one section omitted *Introduzzione Di Messer Giorgio Vasari Pittore Aretino Alle Tre Arti Del Disegno* (33,693 words, 4.4% of the total).

The English translation by De Vere was fully aligned with the Italian source text in 2022–23. De Vere's translation is largely literal, with very few additions or omissions, and therefore relatively easy to align. The main corrections involved sentence divisions, as very long sentences in Italian were often divided into shorter sentences in English, although occasionally English sentences overlapped with two sentences in Italian.

In the following passage, for example, while the last segment pair contains single sentences in both languages, the first source sentence in Italian (*E prima andarono in mala parte … la memoria dell'une e dell'altre.*) corresponds in the English translation to a shorter sentence and the first half of a longer sentence (with tildes "~~~" marking the merger), the second half of which corresponds to the second sentence in Italian (Table 1).

Table 1: Example of alignment for Italian-English from the *Proemio delle Vite* (Vasari, 1568; De Vere, 1912–1915)

Italian (Vasari 1568)	**English (Vasari 1912–1915, translated by De Vere)**
(1) E prima andarono in mala parte la pittura e la scoltura, come arti che più per diletto che per altro servivano; e l'altra, cioè l'architettura, come necessaria e utile alla salute del corpo, andò continuando, ma non già nella sua perfezzione e bontà; e se non fusse stato che le sculture e le pitture rappresentavano inanzi agl'occhi di chi nasceva di mano in mano, coloro che n'erano stati onorati per dar loro perpetua vita, se ne sarebbe tosto spento la memoria dell'une e dell'altre.	And the first to fall into decay were painting and sculpture, as being arts that served more for pleasure than for use, while the other—namely, architecture—as being necessary and useful for bodily weal, continued to exist, but no longer in its perfection and excellence. ~~~ And if it had not been that the sculptures and pictures presented, to the eyes of those who were born from day to day, those who had been thereby honoured to the end that they might have eternal life, there would soon have been lost the memory of both;
(2) Là dove alcune ne conservarono per l'imagine e per l'inscrizioni poste nell'architetture private e nelle publiche, cioè negli anfiteatri, ne' teatri, nelle terme, negli acquedotti, ne' tempii, negli obelisci, ne' colossi, nelle piramidi, negli archi, nelle conserve, e negli erarii, e, finalmente, nelle sepulture medesime; delle quali furono distrutte una gran parte da gente barbara et efferata, che altro non avevano d'uomo che l'effigie e 'l nome.	whereas some of them survived in the images and in the inscriptions placed in private houses, as well as in public buildings, namely, in the amphitheatres, the theatres, the baths, the aqueducts, the temples, the obelisks, the colossi, the pyramids, the arches, the reservoirs, the public treasuries, and finally, in the very tombs, whereof a great part was destroyed by a barbarous and savage race who had nothing in them of man but the shape and the name.
(3) Questi fra gli altri furono i Visigoti, i quali avendo creato Alarico loro re, assalirono l'Italia e Roma, e la saccheggiorno due volte e senza rispetto di cosa alcuna.	These, among others, were the Visigoths, who, having created Alaric their King, assailed Italy and Rome and sacked the city twice without respect for anything whatsoever.

In each case, the necessary changes in sentence division have been restricted to the target text only, so that every segment pair corresponds to exactly one complete sentence in the original Italian source text.

2.3.2 Italian-French

For the alignment of the Italian-French database a first test (Zotti, 2017) was carried out manually for four French translations of Vasari's *Vita di Michelagnolo Buonarroti*, the longest biographical account and the most often studied of the different biographies. Given Michelangelo's well-known versatility, the analysis of this section proved a valuable source of lexical and artistic knowledge in the fields of painting, sculpture and architecture. This test showed the value of building a diachronic parallel corpus for qualitative analysis of artistic terminology as it has evolved over time (Zotti & Henkel, 2024).

Since 2021, the preparation of a parallel corpus including two different French translations (Weiss, 1903; Leclanché & Jeanron, 1839–1842) has been under way, but is proceeding very slowly as the alignment process has encountered problems due to the nature of the translations: in one case the translation is abridged, while the other translation is rather free in terms of sentence structure.

The first French translation to be selected for alignment was that of Charles Weiss (1903), judged by the French specialists of the team to be the best in terms of quality (Farina & Billero, 2020). This translation, which was not available in digital format, was first converted into electronic format from the OCR-generated file available on the Internet Archive and then corrected manually by the LBC-Vasari team on Wikisource, where it is now available in digital format (Weiss, 1903). Despite its superior stylistic quality compared with other French translations, Charles Weiss' translation is incomplete: one biography is missing from the first part, five from the second and thirty-nine from the third (approx. 30% of the full text). For this reason, the decision was made to first supplement the biographies missing in Weiss (1903) with those taken from the complete translation by Leclanché & Jeanron (1839–1842), and then to complete the second alignment, once the proofreading of Leclanché & Jeanron (1839–1842) was finished.

The alignment process showed that Weiss' text contains many passages best described as "summaries" which proved exceedingly difficult to align as clear correspondences could not be established with specific sentences in the source text (cf. Table 2). The following passage shows several problems commonly encountered in this translation:

Table 2: Example of alignment for Italian-French from the *Vita di Giotto* (Vasari, 1568; Weiss, 1903)

Italian (Vasari, 1568)	French (Vasari, 1903, translated by Weiss)
(1) Questa opera dagl'artefici pittori era molto lodata, e particolarmente da Michelagnolo Buonarroti, il quale affermava, come si disse altra volta, la proprietà di questa istoria dipinta non potere essere più simile al vero di quello ch'ell'era.	Cette œuvre qui était très admirée par tous les peintres, particulièrement par Michel-Ange Buonarroti,
(2) Questa tavoletta, dico, essendo venuta in considerazione, da che si diede fuora la prima volta il libro di queste Vite, è stata poi levata via da chi che sia: che forse per amor dell'arte e per pietà, parendogli che fusse poco stimata, si è fatto, come disse il nostro poeta, spietato.	a été depuis enlevée par je ne sais qui, estimant peut-être qu'on ne l'appréciait pas assez[50].
(3) E veramente fu in que' tempi un miracolo, che Giotto avesse tanta vaghezza nel dipignere, considerando massimamente che egli imparò l'arte, in un certo modo, senza maestro.	#NoTarget#
(4) Dopo queste cose mise mano l'anno 1334 a dì 9 di luglio al campanile di S. Maria del Fiore; il fondamento del quale fu, essendo stato cavato venti braccia a dentro, una platea di pietre forti, in quella parte donde si era cavata acqua e ghiaia; sopra la quale platea, fatto poi un buon getto che venne alto dodici braccia dal primo fondamento, fece fare il rimanente, cioè l'altre otto braccia di muro a mano.	Le 9 juillet de l'an 1334, Giotto mit la main au campanile de Santa Maria del Fiore. ~~~ Après avoir fouillé le sol à la profondeur de vingt brasses, il établit comme fondation une couche de pierres dures, dans la partie qu'on avait asséchée, et y posa un massif de béton haut de douze brasses qu'il termina ensuite par huit brasses en pierres de taille.

In the first two segment pairs (1, 2), the corresponding segment in the target text, which is considerably shorter than the source sentence, is not a complete sentence syntactically and is incomplete semantically with respect to the information contained in the source sentence. The third source segment (3) has no counterpart whatsoever in the target text, while in the fourth segment pair (4), the source sentence corresponds to two sentences in the target text.

Such gaps and disparities occur continuously throughout Weiss' French text. Although Weiss' translation is more or less literal at the beginning of each biography, the omissions and reductions become more and more frequent as it progresses, resulting in serious discontinuities in the automatic alignment, which relies on segment length as one of the main criteria to identify corresponding segment pairs.

Because of these problems, the second translation by Leclanché & Jeanron (1839–1842), which is the first complete French translation of Vasari's *Le Vite*, is currently undergoing proofreading and alignment. This translation is analytical, based on the tradition of earlier studies of the work, since the two translators took into account the translations that already existed in England and Germany, with many notes alongside the French version of Vasari's original text. The French translation, *stricto sensu*, mainly by Léopold Leclanché, a lawyer and politician, is in fact enriched by a vast array of commentaries by Jeanron, a painter, draughtsman and lithographer. These comments, placed at the end of each Artist's biography, are marked with a paragraph separator symbol which will help to detach them from the text itself, thus facilitating alignment of the full version. Leclanché & Jeanron's translation, which corresponds almost perfectly to the Giuntina edition of 1568, even though Leclanché often rephrases to explain certain artistic concepts for which no specific terms existed in Vasari's time, laid the foundations for future translations of Vasari's text (Zotti, 2017), all of which are based on this edition, which they improve especially with respect to terminological accuracy (Weiss, 1903; and later Chastel, 1981).

2.3.3 Italian-German

For German, Kupper's 2008 edition was selected. It is based on the first German translation of *Le Vite* that has come down to us, edited by Schorn and Förster in 1832–1849, to which Kupper made some minor changes, such as normalising spelling according to new conventions. Unlike many earlier editions, this is an almost complete translation, despite certain omissions as revealed in the alignment process.

The first attempt at alignment highlighted some shortcomings of the text in German translation. In particular, it came to light that a few chapters of the initial part of the Italian text had not been translated, namely: the dedication *Agli artefici del disegno*, the *Proemio di tutta l'opera* and the *Introduzzione di messer Giorgio Vasari pittore aretino alle tre arti del disegno, cioè architettura pittura e scoltura e prima dell'architettura*. As these parts were at the beginning, a new alignment phase was necessary, conducted by removing the parts not included in the German text from the Italian text and then inserting them again after the new automatic (pre)alignment.

The analysis of the alignment revealed a small number of #NoSource# cells (additions), but a larger number of #NoTarget# cells (omissions). The former are principally cells that correspond to quotations in the source text, which in the target text are also translated into German. The latter are mainly anecdotes about the lives of artists, which are omitted or summarised, i.e., condensed into fewer sentences. Moreover, where there are complex sentences in Italian, the German translation tends to divide them into several sentences of greater simplicity:

Table 3: Example of alignment for Italian-German of the *Vita di* Cimabue (Vasari, 1568; Kupper, 2008)

Italian (Vasari, 1568)	**German (Vasari, 2008, translated by Kupper)**
(1) Avendo poi preso a fare per i monaci di Vall'Ombrosa nella Badia di Santa Trinita di Fiorenza una gran tavola, mostrò in quell'opera, usandovi gran diligenza per rispondere alla fama che già era conceputa di lui, migliore invenzione, e bel modo nell'attitudini d'una Nostra Donna, che fece col Figliuolo in braccio e con molti Angeli intorno che l'adoravano in campo d'oro; la qual tavola finita, fu posta, da que' monaci in sull'altar maggiore di detta chiesa, donde essendo poi levata, per dar quel luogo alla tavola che v'è oggi di Alesso Baldovinetti, fu posta in una cappella minor della navata sinistra di detta chiesa.	Darauf übernahm er für die Mönche von Vallombrosa in der Abtei von Santa Trinita in Florenz eine große Tafel. Er wendete großen Fleiß auf dies Werk, um dem Ruf zu genügen, den er sich schon erworben hatte, und zeigte darin noch viel bessere Erfindung und schöne Stellungen. ~~~ Es war eine Mutter Gottes, mit dem Kind auf dem Arm und vielen Engeln umher, welche sie anbeten, auf Goldgrund. ~~~ Dieses Bild stellten jene Mönche über dem Hauptaltar ihrer Kirche auf, von wo es später weggenommen und in eine kleinere Kapelle des linken Seitenschiffs der Kirche gebracht ward, um dem Gemälde von Alessio Baldovinetti Platz zu machen, welches noch heute auf jenem Hauptaltar steht.

In the future, the corpus will be expanded to include the translation by the *Kunsthistorisches Institut* (KHI) in Florence (2004–2015), a translation that stands out from all previous ones because of its focus on Vasari's style and specialised vocabulary.

3. Case Study: *disegno/disegnare* and Their Translation Counterparts

As indicated above, Vasari's parallel database will be integrated in the LBC-Platform whose one of the main objectives is facilitating the translation, in different foreign languages, of texts belonging to diverse discourse types in the domain of Fine Arts. The parallel corpus will thus certify new translation equivalents of the lemmas (words and collocations) which will be included in the future LBC-dictionary so as to provide quality lexicographical information (Farina & Billero, 2020), given that the multilingual lexicographic and lexical resources in print and online are not always adequate or helpful in identifying possible translation counterparts. In our small case study, we will concentrate on the lexemes *disegno* ('drawing', noun)[4] and *disegnare* ('to draw', verb) in order to give a glimpse of this future innovative lexicographical application. These two lexemes have different variants of meaning, enriched over time, entering in the specialised lexicons of various disciplines or areas of experience (Ballestracci, 2023, p. 22). In Vasari's masterpiece they have a complex and multifaceted semantic potential, characterised by multiple meanings. The frequency is high: the noun/

---

[4] These lexemes have been studied previously by Luporini (2023) in one English translation and by Balestracci (2023) in the *Vita di Raffaello* and its German translation (Kupper, 2008 and KHI), using the LBC database while alignment of source/target texts was still in progress.

lemma *disegno* has 1433 occurrences and the verb/lemma *disegnare* 530 occurrences, among the most frequent nouns and verbs in the entire work (*disegno* is the 7th most common noun, following *cosa* ('thing'), *opera* ('work'), *figura* ('figure'), *mano* ('hand'), *tempo* ('time') and *anno* ('year'), while *disegnare* is the 30th most common verb).

Interesting patterns are revealed by examining their counterparts in the aligned corpora. In the English translation by De Vere (1912–1915), *disegno* corresponds in order of descending frequency to *design* (142), *drawing* (53), *draughtsmanship* (11), *draw* (8), *draughtsman* (3), *likeness* (1), *proof* (1), while *disegnare* is rendered with *to draw* (49), *to design* (16), *drawing* (10), *outline* (4), *draughtsman* (2), *draughtsmanship* (2), *to copy* (1), *to intend* (1), *to propose* (1), *to take* (1), *to trace* (1). These findings confirm Luporini's (2023) observations concerning the three most common interpretations of *disegno*: *design*, *drawing* and *draughtsmanship*, and offer two new (albeit much less common) alternative translations: *proof* and *likeness*. The same three most common counterparts of the noun *disegno* are present among the most common translations for the verb *disegnare*, as well as a fourth technical meaning corresponding to *outline*, *trace* and a fifth, non-technical interpretation rendered by *intend* or *propose*. It is interesting to note, with reference to Luporini's "cline" from abstract to concrete, that the verb *disegnare* is most often interpreted in the most concrete sense of *draw/drawing*, while the median interpretation *design* is more common for the noun *disegno*. For both, the most abstract interpretation corresponding to *draughtsmanship* is the least common of the three main meanings. The parallel database thus provides a comprehensive overview of the semantic profile of these words as revealed in their alternative translation counterparts.

In French, so far only the chapters known as the *Teoriche* (especially *Introduzzione Di Messer Giorgio Vasari Pittore Aretino Alle Tre Arti Del Disegno*) have been semi-automatically aligned with Weiss (1903) for a total of about 50.000 words. In these chapters, which are missing from most other translations but are particularly interesting from a terminological point of view, the most common translation counterparts for *disegno* are: *dessin* (81), *dessiner* (3), *composer* (1), *composition* (1), *contour* (1), *dessein* (1), *modèle* (1), *propriété* (1). The verb *disegnare* most commonly corresponds to: *dessiner* (22), *dessin* (4), *chercher* (1), *marquer* (1), *reproduire* (1), *tracé* (1), *traiter* (1). The polysemy of *disegno* and *disegnare* is thus apparent in their translation counterparts.

In German, only a quarter of the Italian source text has been aligned so far (24,112 words). The lemma *disegno* (330 occurrences, of which 289 are of *disegno* and 50 of the plural form *disegni*) is translated mostly by *Zeichnung* ('drawing') (131) and *Zeichnungen* ('drawings') (32) often adding the corresponding word in Italian in square brackets [*disegno*], in those cases where the translated term has a more precise meaning than the German (see also Ballestracci, 2023, pp. 22–23). Other translation counterparts with substantives include *Zeichenkunst* (24) ('drawing art', mostly followed by the Italian *disegno* in square brackets in those cases); *Plan* ('plan'), when it is used with verbs like *entwerfen* ('to draft') or *zeichnen* ('to draw'); *Angabe/Angaben* (5) ('indication/indications'); *Blatt/Blätter* (3) ('sheet/sheets'); *Entwurf/Entwürfe* (2) ('draft/drafts'); *Modell* (1) ('model'); *Arbeit* (1) ('work') and *Werk* (1) ('work'). In some cases the lemma is translated with verbs like *zeichnen* ('to draw') or substantive verbs like *Zeichnen* ('drawing'). In a few cases the lemma is not translated at all. Interesting cases, in which the lexeme has other meanings, are the translations of *disegno* with *nach richtigem Verhältnis* ('according to the correct ratio'), with *was hast du vor?* ('what are you planning to do?') and with *so gut* ('so good'). The verb *disegnare* (65) is translated mainly by the verb *zeichnen* (46) ('to draw') or with *ausführen* (3) ('to carry out'), *entwerfen* (2) ('to draft'), *malen* (1) ('to paint'), *darstellen* (1) ('to represent'). In some cases there the verb is translated with the nouns *Zeichnung* (3) ('drawing'), *Blatt* (1) ('sheet') with the phraseme *von seiner Hand* (1) ('by his Hand') or not translated at all. This quick overview of some of the most used translation counterparts in German highlights several potential meanings (more general, more specific, phraseological) of the lexemes in the source language.

4. Conclusion

Parallel corpora are not only important for the linguistic analysis of translation but also from a contrastive point of view: work on parallel corpora makes it possible to conduct quantitative and qualitative analyses on large quantities of data, focusing on Vasari's style and specialised vocabulary, comparing lexical and syntactic aspects between original texts and translations.

The polysemy of lexemes like *disegno* can be revealed in translation, delineating interlingual correspondences that are useful for specialist translation in the field of art. Careful qualitative analysis elucidates semantic peculiarities that are central for lexicographic purposes. In the specific case study presented herein, analysis of the translations shows that, over time, ambiguities in Vasari's use of 16th-c. Italian are replaced over time by greater terminological precision in other languages.

Indeed, the study of *disegno* and *disegnare* and their counterparts in English, French and German reveals that in Vasari, these lexemes were used with a broad range of meanings including, on the one hand, technical meanings close to terminology and, on the other hand, descriptive meanings belonging to the general language and also phraseological uses.

In English, De Vere tried to render these different meanings as precisely as possible, hence a greater lexical variety which can also be seen as a sign or consequence of the development of artistic terminology between 16th-c. Italian and 20th-c. English. Also, the French and German translators chose corresponding terms that have now become stabilized, according to their purpose and intended public.

Future work in the LBC project will involve the alignment of other translations. In French the alignment of Leclanché & Jeanron (1839–1842) will be completed, followed by the alignment of translation by a renowned French art historian, Chastel (1981). In German, for example, the alignment of the translation of KHI is planned, which contains even more counterparts for *disegno* highlighting its semantic complexity (Ballestracci, 2023, p. 30). The alignment of *Le Vite* with its translations in other languages of the LBC Project like Spanish and Russian is planned as well.

No comprehensive inquiry into translation in the field of the fine arts and art history has yet been undertaken, due to the technical difficulties inherent in creating diachronic parallel corpora, but the work carried out by the LBC team is an attempt to promote new research in this field which will have a clear impact also in the field of cultural heritage.

References

Aglionby, W. (1685). *Painting lllustrated in Three Dialogues. Includes the first English translation of Giorgio Vasari's writings on the lives of eminent artists.* John Gain.

Ballestracci, S. (2023). Da un segno a tanti segni. L'emergere della polisemia del termine disegno nelle traduzioni tedesche delle Vite di Vasari. In V. Zotti, & M. Turci (a cura di.), *Nuove strategie per la traduzione del lessico artistico* (pp. 21–38). Firenze University Press.

Ballestracci, S., Buffagni, C. & Flinz (Eds.), (2024). Das deutsche LBC-Korpus: Zusammenstellung und Anwendung. In S. Ballestracci, C. Buffagni, & C. Flinz, *Das deutsche LBC-Korpus.* Firenze University Press. https://corpora.lessicobeniculturali.net/de/deutsche-LBC-korpus-zusammenstellung-anwendung

Biffi, M. (2017). *L'italiano dell'arte, in L'Italiano. Conoscere una lingua formidabile.* Accademia della Crusca.

Billero, R., Farina, A., & Nicolás Martínez, M. C. (2020) (Eds.). *I Corpora LBC. Informatica Umanistica per il Lessico dei Beni Culturali.* Firenze University Press.

Billero, R. (2020). Implementazione di software per la gestione dei corpora LBC. In R. Billero, A. Farina, & M. C. Nicolás Martínez (Eds.), *I Corpora LBC. Informatica Umanistica per il Lessico dei Beni Culturali* (pp. 19–32). Firenze University Press.

Daret, P. (1651). *Abregé de la vie de Raphaël Sansio d'Urbin, tres-excellent peintre & architecte; ou il est traitté de ses oeuvres, des stampes qui en ont esté gravées, tant par Marc-Antoine Bolognois qu'autres excellent [« sic »] graveurs.* Trad. en françois par P. Daret, graveur.

Dubus, P., & Fiorato, C. L. (2017) (Eds.). *La réception des Vite de Giorgio Vasari dans l'Europe des XVIe-XVIIIe siècle.* Librairie Droz.

Farina, A., & Billero, R. (2020). Corpus in "Natural" Language Versus "Translation" Language: LBC Corpora, A Tool for Bilingual Lexicographic Writing. In: D. F. Iezzi, D. Mayaffre, & M. Misuraca (Eds.), *Text Analytics. JADT 2018. Studies in Classification, Data Analysis, and Knowledge Organization.* Springer, Cham. https://doi.org/10.1007/978-3-030-52680-1_14

Farina, A., & Sini, L. (2024). Le Corpus LBC Français. In A. Farina, *Corpus LBC Français.* Firenze University Press. https://corpora.lessicobeniculturali.net/fr/corpus-LBC-francais

Klosa, A. (2016). Der lexikographische Prozess im Projekt elexiko. In V. Hildenbrandt, & A. Klosa (Eds.), *Lexikographische Prozesse bei Internetwörterbüchern* (pp. 29–38). OPAL 1/2016 Institut für Deutsche Sprache.

Le Mollé, R. (1988). *Georges Vasari et le vocabulaire de la critique d'art dans les* Vite. Editions littéraires et linguistique de l'université de Grenoble.

LBC: *Lessico dei Beni Culturali.* Retrieved July 30, 2024, from https://www.lessicobeniculturali.net

LF Aligner. Retrieved July 30, 2024, from https://sourceforge.net/projects/aligner/

Luporini, A. (2023). Exploring the Lexis of Art Through a Specialized Corpus: A Bilingual Italian-English Perspective. *International Journal of English Linguistics*, *13*(7), 28–38.

Natali, I. (2024). The English Corpus for the LBC project. In I. Natali, *The LBC English Corpus.* Firenze University Press. https://corpora.lessicobeniculturali.net/en/english-corpus-LBC-project

Okapi Checkmate. Retrieved July 30, 2024, from https://okapiframework.org/wiki/index.php?title=CheckMate

Vasari, G. (1550). *Le Vite de' più eccellenti Pittori, Scultori et Architettori.* Edizione Torrentiniana.

Vasari, G. (1568). *Le Vite de' più eccellenti Pittori, Scultori et Architettori.* Edizione Giuntina.

Vasari, G. (1832–1849). *Leben der ausgezeichnetsten Maler, Bildhauer und Baumeister. Von Cimabue bis zum Jahre 1567*, a cura di Ludwig Schorn e Ernst Förster. Cotta.

Vasari, G. (1839–1842). *Vies des peintres, sculpteurs et architectes.* Trad. et commentaires par Léopold Leclanché et Philippe-Auguste Jeanron. Tessier.

Vasari, G. (1850). *Lives of the most eminent painters, sculptors, and architects: translated from the Italian of Giorgio Vasari.* With notes and illus., chiefly selected from various commentators by Mrs. Jonathan Foster. Robarts – University of Toronto.

Vasari, G. (1903). *Les vies des plus excellents peintres, sculpteurs, et architectes de Giorgio Vasari*, traduction de Charles Weiss. Dorbon-Aine. Available on wikisource :<https://fr.wikisource.org/wiki/Les_vies_des_plus_excellents_peintres,_sculpteurs,_et_architectes>

Vasari, G. (1912–1915). *Lives of the Most Eminent Painters, Sculptors and Architects* (10 volumes; London: Macmillan and Co. and the Medici Society, 1912–1915), trans. by Gaston du C. De Vere.

Vasari, G. (1981). *La vie des meilleurs peintres, sculpteurs et architectes*, traduction d'André Chastel. Berger-Levrault.

Vasari, G. (2004–2015). *Deutsche Ausgabe der 'Vite' Giorgio Vasaris*, a cura di A. Nova con M. Burioni, K. Burzer, S. Feser, H. Gründler, & F. Jonietz. Wagenbach.

Vasari, G. (2008). *Leben der ausgezeichnetsten Maler, Bildhauer und Baumeister von Cimabue bis zum Jahre 1567*, a cura di D. Kupper. Directmedia Publishing.

Zotti, V. (2017). L'integrazione di corpora paralleli di traduzione alla descrizione lessicografica della lingua dell'arte: l'esempio delle traduzioni francesi delle Vite di Vasari. In V. Zotti, & A. Pano Alaman (Eds.), *Informatica umanistica. Risorse e strumenti per lo studio del lessico dei beni culturali* (pp. 105–134). Firenze University Press.

Zotti, V. (2022). La langue des Beaux-Arts: dimension interculturelle et enjeux terminologiques des corpus comparables. In C. Grimaldi et al. (Eds.), *Terminologia e interculturalità. Problematiche e prospettive* (pp. 229–245). I libri di Emil.

Zotti, V., & Henkel, D. (2024). The diachronic evolution of artistic terminology in translation. Building a parallel corpus of Giorgio Vasari's Le Vite. *CORPUS*, *25*, 1–20.

Contact information

Carolina Flinz
Università degli Studi di Milano
carolina.flinz@unimi.it

Daniel Henkel
Université Paris 8
daniel.henkel@univ-paris8.fr

Valeria Zotti
Alma Mater Studiorum – di Bologna
valeria.zotti@unibo.it

Sabrina Ballestracci
Università degli Studi di Firenze
sabrina.ballestracci@unifi.it