

# The Language of Art and Cultural Heritage



# The Language of Art and Cultural Heritage:

*A Plurilingual and  
Digital Perspective*

Edited by

Ana Pano Alamán and Valeria Zotti

Cambridge  
Scholars  
Publishing



The Language of Art and Cultural Heritage:  
A Plurilingual and Digital Perspective

Edited by Ana Pano Alamán and Valeria Zotti

This book first published 2020

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data  
A catalogue record for this book is available from the British Library

Copyright © 2020 by Ana Pano Alamán, Valeria Zotti and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-5275-4567-9

ISBN (13): 978-1-5275-4567-0

The papers of this volume were submitted to double blind peer review.

*Scientific Committee*

Claudia Buffagni (Foreigners University of Siena)  
Elena Carpi (University of Pisa)  
Annick Farina (University of Florence)  
Sabrina Fusari (University of Bologna)  
Ludovica Lanini (University of Roma La Sapienza)  
Antonella Luporini (University of Bologna)  
Marie-France Merger (University of Pisa)  
Ilaria Natali (University of Florence)  
Ana Pano Alamán (University of Bologna)  
Monica Perotto (University of Bologna)  
Valentina Rossi (University of Florence)  
Christina Muriel Samson (University of Florence)  
Rachele Raus (University of Turin)  
Lorella Sini (University of Pisa)  
Valeria Zotti (University of Bologna)

*Editorial board*

Umberto Faraone  
Camilla Lucarelli  
Ana Pano Alamán  
Valeria Zotti

Centro di Studi Linguistico-Culturali (CeSLiC)  
Dipartimento di Lingue, Letterature e Culture Moderne  
University of Bologna



# TABLE OF CONTENTS

Introduction .....	1
<i>Ana Pano Alamán and Valeria Zotti</i>	
<b>Section 1. Corpus Linguistics</b>	
Chapter One.....	12
On Translating Art and Heritage Discourse from Italian into English: From a Learner Corpus to a Specialized Corpus	
<i>Monica Turci and Gaia Aragrande</i>	
Chapter Two.....	39
Between Philology, Lexicography and Art History: The ItalArt project	
<i>Alessandro Aresti</i>	
Chapter Three .....	59
Reactividad emocional ante la obra de arte en la lengua extranjera. Observación de producciones lingüísticas multimodales en un grupo de estudiantes itálofonos	
<i>Irene Buttazzi and María Cecilia Ainciburu</i>	
Chapter Four.....	86
Cultural Heritage Lexicon: A Case Study	
<i>Riccardo Billero</i>	
Chapter Five.....	104
Il termine russo <i>frjag</i> e le sue radici nelle relazioni culturali e artistiche fra la Russia e l'Italia	
<i>Marcello Garzaniti</i>	
Chapter Six.....	120
Lexical Association as an Indicator of Conceptual Convergence and Divergence in English, French and Italian	
<i>Daniel Henkel</i>	

## Section 2. e-Lexicography

Chapter Seven.....	140
Defining Architecture: Considerations in the Field of Architectural Corpora <i>Monika Bogdanowska</i>	
Chapter Eight.....	157
Typologie des dictionnaires numériques de français illustrée par des mots du vocabulaire patrimonial <i>Nathalie Gasiglia</i>	
Chapter Nine.....	182
Art and Cultural Heritage in the Slovenian Tourism Corpus and E-dictionary <i>Vesna Mikolič</i>	
Chapter Ten .....	190
Architecture in the 1701 <i>Dictionnaire Universel</i> : Encoding and Analysing Architectural Terminology with Digital Humanities Methodologies <i>Geoffrey Williams</i>	

## Section 3. Web genres

Chapter Eleven .....	210
An approach to Information Diffusion of Spanish World Heritage in the Italian Wikipedia <i>Lola García-Santiago and María Dolores Olvera-Lobo</i>	
Chapter Twelve .....	234
Le Patrimoine sarde en français et en ligne: le discours touristique, du papier au numérique <i>Lorenzo Devilla</i>	
Chapter Thirteen.....	248
Digital Spaces of Collaboration in Aesthetic Counter Narratives: <i>Hamedullah. The Road Home and The Mirror Project</i> <i>Alessandra Rizzo</i>	
Contributors.....	279



# INTRODUCTION

ANA PANO ALAMÁN AND VALERIA ZOTTI

Communicating art has become a crucial and a challenging mission, since artistic and cultural heritage, as well as tourism, represent key socio-economic and strategic development sectors worldwide (Chabra *et al.* 2003; Martelloni 2004; Loulanski 2006). Exhibitions, museums, tangible and intangible heritage assets, cultural events, and artistic objects, among others, feed into a market where individuals interested in heritage conservation and dissemination, foundations, associations, academia, along with public and private enterprises, play a pivotal role. In order to stimulate this economic and social potential, art and cultural heritage need to be brought into the focus of attention of the wide public through effective communicative and promotional strategies (Colbert *et al.* 1994; Chiapparini 2009, 2011).

Today, these strategies are made up of a wide variety of discursive and semiotic techniques that are becoming increasingly sophisticated, since they concur in both traditional text typologies, such as brochures, catalogues and magazines, and innovative online digital media. Hence, professionals in the fields of communication and marketing (e.g. journalists, translators, media planners, and community managers who work in the art and cultural sectors) and social actors of the tourism, culture and creative industries (such as museum directors, exhibition curators, or tourist guides) face a series of challenges that call for the design of new methods and tools improving art and cultural heritage research, preservation and dissemination. In this evolving context, it appears evident that Information and Communication Technologies are a key factor in the innovation and the improvement of such strategies and techniques (EU DigiCULT 2002; Sylaiou and Papaioannou 2018; Wiener 2018).

On the other hand, the discourse on art and cultural heritage is becoming increasingly complex and multilingual these days. It moves across the boundaries that divide sciences and technology from social sciences and humanities, for instance, incorporating in the text types aforementioned, terms that are frequently used in History or in Literature,

and terms adopted in Mathematics or Chemistry discourses. Indeed, one of the most eloquent cases of this complexity is the terminology of architecture, where engineering terms and art history expressions appear to intertwine to a certain extent, according to different text genres. Moreover, since art and cultural heritage contents are simultaneously disseminated worldwide through globalised media and cultural exchanges (Gardner-Chloros 2004), the linguistic and discursive strategies adopted for the creation of such contents need to be necessarily multilingual (see Minerva Europe, 2005).

In order to explore both the digital and the plurilingual character of the language of art and cultural heritage, as well as to analyse the challenges that the cultural and tourism sectors face in a globalised and interconnected world, this book provides an overview of current research projects that deliver and critically examine a series of digital linguistic tools and lexical resources that may contribute to the design and the implementation of new strategies within art and cultural heritage communication. This is why the present volume mainly adopts a digital and plurilingual perspective.

These projects deal with the above-mentioned challenges, drawing on disciplines such as linguistics, translation studies, lexicography, terminology, and specialised communication, and implementing a digital humanities approach to the study of art and cultural heritage discourses (Zotti and Pano Alamán 2017). Because of the inherent complexity of the issues explored, and according to the different areas of research considered, the contributions included in this volume are organised into three interrelated sections: corpus linguistics, e-lexicography, and web genres.

Some challenging issues addressed in *Section 1* (Chapters One-Six), devoted to Corpus Linguistics, are the design and the implementation of learner corpora for the study of art and architectural heritage within original and translated specialised texts; the constitution of corpora that are specifically designed for the improvement of lexical resources in the art and the cultural sectors; the adoption of interlinguistic and intercultural methods for the analysis of artistic words from a diachronic viewpoint; the design of a specialised lexicon of arts; and the treatment and exploitation of specialised corpora extracted from the Web.

In *Chapter One*, **Monica Turci** and **Gaia Aragrande** report on a pilot study focusing on the translation from Italian into English of texts on “heritage tourism”, leading towards the creation of a learner translation corpus. This study of the non-professional translation of art and heritage texts shows how teaching Language for Specific Purposes and Specialised

Translation may benefit from the application of Corpus Linguistics methods. This project is the result of a collaboration between the University of Bologna and a group of public and private organisations, which are active in the field of tourism and heritage conservation and promotion, such as the Italian Ministry of Culture and Tourism of the Emilia-Romagna region.

In *Chapter Two*, **Alessandro Aresti** presents a Marie Curie project entitled *ItalArt. The Language of the Arts and Architecture in Italy between the Middle Ages and the Renaissance*, which aims at developing the *Opera del Vocabolario Italiano* (OVI). The OVI is an online dictionary of historical-artistic terms in Italian that may contribute to understanding the formation and development of an Italian lexicon of the arts and architecture from the Middle Ages to the Renaissance. The core of the project is the digital edition of two significant texts that belong to the art historiography tradition, *Documenti per la storia dell'arte senese* (1854-1856) and *La scrittura di artisti italiani (secoli XIV-XVII)* (1876), both published by the art historian Gaetano Milanesi (1813-1895). Aresti presents and examines this corpus of approximately 1200 documents for digitisation and lemmatization purposes, in order to implement the OVI dictionary.

Written by **Irene Buttazzi** and **Maria Cecilia Ainciburu**, *Chapter Three* discusses whether an artwork, such as a painting, and the words used for its description, have an influence on the linguistic production of foreign language learners, while describing their aesthetic experience in front of the same artwork. Adopting a Cognitive Linguistics approach, the authors explore a corpus of written texts, which have been produced by a group of Italian-speaking learners of Spanish as a foreign language, after being exposed to three paintings by the Colombian artist Botero, and which convey different emotions to the viewer. The authors carry out a qualitative analysis of the learners' descriptions of their "emotional" (positive, negative, or neutral) response to the paintings, which have been produced both in Italian and in Spanish. The analysis conducted by Buttazzi and Ainciburu reveals that learners process the artistic *stimuli* differently, depending greatly on the use of their L1 or L2. The results also suggest that language plays a key role in our cognitive and aesthetic responses to art.

In *Chapter Four*, **Riccardo Billero** illustrates the project *Cultural Heritage Lexicon*, carried out by the Research Unit *Lessico dei Beni Culturali* (LBC), which aims at creating digital resources and tools, such as specialised corpora and an e-dictionary of art, that may prove useful for art historians, specialised translators, tour operators, and exhibition

curators. Billero describes the process managed by the research unit to implement six comparable corpora on the artistic and cultural heritage of Florence and the Tuscany region, in English, French, Spanish, Russian, Italian and German. As the author explains, all these corpora, which are accessible from the *Lessico dei Beni Culturali* website<sup>1</sup>, have been specifically designed for the extraction of lemmas related to the arts in different languages, leading to the creation of a multilingual expert e-dictionary that will be available online.

In *Chapter Five*, **Marcello Garzaniti** examines a case study of particular relevance for art history and for the analysis of the artistic lexis in Russian, by exploring the Russian *Lessico dei Beni Culturali* corpus. Through an extensive enquiry of the frequencies and concordances automatically obtained using Sketch Engine, the author scrutinizes the term *frjag*, which designates the Italian architects and engineers who worked in Russia in the late 15th century, such as those who participated in the building of the Kremlin. Adopting a diachronic and an intercultural approach, Garzaniti highlights the importance of using corpus-based methods and digital tools for the lexicographical treatment of complex artistic terms.

**Daniel Henkel** shows, in *Chapter Six*, how comparable corpora extracted from Wikipedia, using *BootCaT*—a versatile tool both for corpus building and for finding terminology and translations—offer interesting insights into the ways in which the English, French and Italian linguistic and cultural communities describe music, painting, and sculpture. From his analysis of these three corpora in the three languages, Henkel provides a list of highly specific and high-frequency keywords using log likelihood and odds ratio to measure specificity. Despite many similarities, some keywords in one language do not have translational equivalents for the corresponding domain in the other two languages. As the author argues, some of these disparities can be explained by the lesser specificity of polysemous words, but others can be seen as the result of an underlying divergence in the ways different language communities understand and formulate artistic and cultural concepts.

*Section 2* (Chapters Seven-Ten) is focused on e-Lexicography or digital lexicography, and, thus, on the impact that the digital paradigm has had on dictionaries in the last decades. The contributions gathered within this section examine the achievements of the so-called “new lexicography” in the fields of arts and cultural heritage. Special attention is devoted to a selection of lexical and terminological resources already available through digital devices, as well as their implementation for various uses. This

---

<sup>1</sup> <http://www.lessicobeniculturali.net>

section focuses on digital methods and tools applied to the analysis and creation of online dictionaries, which are specific to architecture and tourism heritage. Among the issues explored in this section, we shall mention the transformation of digital lexical resources; the role of open and collaborative lexical resources that challenge traditional dictionaries as authoritative lexical reference tools; the relation of search engines *vis à vis* traditional dictionaries; and lexicography specialised in art and cultural heritage.

In *Chapter Seven*, **Monika Bogdanowska** gives a thorough overview of popular and academic online dictionaries as well as glossaries of architectural terminology in different languages. The author discusses their advantages and drawbacks on the basis of a series of criteria established by an international research project carried out at the Cracow University of Technology, which aimed at compiling a multilingual dictionary in the fields of conservation and restoration of architectural heritage. The project gave birth to the *Interdisciplinary Multilingual Dictionary*, a terminological online database with thousands of entries in Polish, English, German, French, and Italian. This resource constitutes a first important step in compiling the *New Dictionary of Architecture*. Assuming that architecture is a vast research area with a great variety of sub-domains, territory specificities and construction typologies across the world, and through different periods of history, Bogdanowska argues in favour of a dictionary that allows users to access cultural data as well. From this perspective, a dictionary is not seen as a mere collection of terms and their equivalents, but rather as a platform for intercultural exchange.

In *Chapter Eight*, **Nathalie Gasiglia** critically discusses the transposition of printed monolingual dictionaries of French to their digital version. This exhaustive study has been conducted by Gasiglia and Pierre Corbin, within the *Master in Lexicography, Terminography and Natural Language Processing* at the University of Lille. Their main objective is to monitor the market evolution of such dictionaries, by taking into consideration the needs of the target users and the scope of the dictionaries themselves, as well as their format and the coding of the data included within them. The author examines general language dictionaries and dictionaries for learners, both available in print and in digital editions, providing an accurate analysis of the nomenclature and of the different kinds of microstructural information included in the dictionaries for the entries *art*, *balustre* and *gouache*. Gasiglia's study proves that digital dictionaries do not offer a more extensive overview or a broader semantic and syntactic description of these nouns, compared to print dictionaries. The author, thus, concludes that an in-depth analysis of the market

segments identified for the printed dictionaries should be necessary before implementing digital dictionaries addressed to the same market segments.

*Chapter Nine*, by **Vesna Mikolič**, addresses a number of conceptual, methodological and technical issues related to the design and creation of a corpus and an e-dictionary focused on artistic, architectural, and landscape vocabulary of Slovenian tourism heritage. The author outlines the phases of a project whose objective was the creation of a specialized corpus of tourism texts in Slovenian, that included translated texts from Slovenian into Italian, English and German, which had been previously designed and disseminated by the key institutions involved in promoting tourism in the country. The *TURistični Korpus* (TURK), which is available online, comprises more than 30 million words and gathers texts that belong to different subtypes of tourism such as nature and natural heritage, from sports to health resorts, and from rural tourism to adventure tourism. Other types, like promotional texts for cultural events or business conventions, are related to urban spaces. Mikolič shows how the criteria adopted for gathering, classifying, and annotating these tourism texts, and for the implementation of the open access TURK corpus, have proved to be advantageous for the design of a collaborative e-dictionary of Slovenian tourism (TURS). The aim of these two projects is to define those areas that may be strategic for the Slovenian tourism sector, and create digital resources useful for investigating tourism discourse.

Assuming that dictionaries provide fascinating windows on the past, **Geoffrey Williams** explores, in *Chapter Ten*, the 1701 Basnage de Beauval edition of Furetière's *Dictionnaire Universel*, which is a major, although little exploited, lexical source. Whilst Furetière set out to create a universal dictionary that would include a series of terms related to arts, crafts, and sciences, de Beauval turned it into an encyclopaedia that included many technical and historical data. Hence, it has become an unvaluable resource for research projects, such as the BasNum project, whose objective is the automatic extraction of terminology, particularly of architecture terms, from this historical document. The methodologies and technologies employed in the BasNum project are thoroughly discussed. As Williams claims, in order to extract those terms, the source needs to be digitised and explored using Computer Assisted Qualitative Data Analysis Software (CAQDAS). The author then goes on to describe in detail the preliminary analysis of the dictionary using CAQDAS and the application of the XML-TEI mark-up language for annotating the 1701 and later editions of the dictionary.

The last Section of the volume, *Section 3* (Chapters Eleven-Thirteen), deals with some textual, hypertextual, multimodal, and interactive aspects

of the so-called web genres, and with the impact of the technological affordances characterising them in promotional discourses of art and cultural heritage. The topics included in this section concern the representation of art and heritage on web 2.0 platforms like Wikipedia; the adaptation of printed tourism texts to digital environments; and the characteristics of audiovisual and multilingual forms of art, like online documentaries and digital storytelling, and the challenges that they pose to translation.

In regard to the first topic, **Lola García-Santiago** and **María Dolores Olvera-Lobo** present, in *Chapter Eleven*, the results of a quantitative and qualitative analysis of a corpus of Wikipedia entries on world cultural heritage sites located in Spain and recognised by UNESCO. The analysis provides a detailed assessment of both the Spanish and the Italian versions of the online encyclopaedia. The authors identify a number of Spanish and Italian entries, whose features bear both similarities and differences, including named entities, historical data, editing policies, and the presence of interwikis, as well as the employment of user orientation and reference mechanisms in the entries. Their work demonstrates that Wikipedia is a fundamental web resource for the dissemination of cultural and artistic heritage, but that it is necessary to develop coherent and interrelated entries in its different versions.

In *Chapter Twelve*, **Lorenzo Devilla** provides an interesting panorama on the way in which the traditional discourse of guidebooks changes when the same texts are transferred to the web. In particular, Devilla scrutinizes a corpus of print tourist guidebooks to Sardinia (Italy) in French, and their “corresponding” website versions in the same language. Following the methods of discourse analysis and digital discourse analysis, the author concludes that there are significant linguistic and discursive variances between the online and the print versions of the corpus, principally due to the fact that on web platforms users are able to interact with both the authors of the online guidebooks, with other travellers, and with other internet users.

The last contribution, in *Chapter Thirteen*, written by **Alessandra Rizzo**, draws on recent research on narrative theory in translating dissent, and on issues of (re)narration in Translation Studies. By adopting a Systemic Functional Linguistic approach, Rizzo examines in detail a corpus of works belonging to the creative and cultural industries from the perspectives of multimodality, audiovisual, and documentary translation. As case studies, the author explores two counter narratives, the documentary *Hamedullah. The Road Home* and *The Mirror Project*, which focus on the ideological and structural diversities between misleading and

oversimplified news reporting. As Rizzo demonstrates, the narrative modalities employed in these visual arts and digital audiovisual products enable and stimulate collaborative translational practices that include online fansubbing and fandubbing projects. These translation phenomena result also from a broad diffusion of migratory artistic projects on the web.

As all papers just presented show, this volume intends to document how the resources and tools offered by humanistic computing bring with them a decisive renewal in the study of art and cultural heritage communication. This is why the view adopted is heterogenous. Indeed, the digital tools made available for researchers continue to evolve, opening up new study perspectives. The book constitutes in fact a further step in the advance of research in the Digital Humanities methods applied to the study of art and heritage language, adopted within the *Multilanguage Cultural Heritage Lexis* or *Lessico multilingue dei Beni Culturali* (LBC) research project, which was established in 2013 at the University of Florence.

The project assumes that an increasing number and a wide variety of texts on Italian cultural heritage are available today, both online and in print. Provided in different languages, these works attempt to satisfy an international public and its increasing demand for information on Italian cultural heritage. However, there are currently no specific tools, such as dictionaries or reference materials on technical translations, or methods to help train specialist writers, translators or other experts involved in the promotion and dissemination of cultural and tourism heritage. The *LBC* project aims at creating an e-dictionary, currently a work in progress, in French, English, Spanish, German, Portuguese, Russian, and Chinese, which will incorporate a series of words (lemmas) emerging from text corpora related to art, including common nouns and proper nouns referring to Renaissance art and its history. The project is implementing other digital resources, such as a Giorgio Vasari's encyclopaedia, an etymological dictionary, and translation parallel corpora. Although, as mentioned before, this book is mainly addressed to specialist writers and translators, as well as to scholars in the fields of art history and cultural heritage communication and promotion, this collection of studies will also be of interest to language and culture experts, IT engineers specialized in Natural Language Processing (NLP), lexicographers, terminologists, and foreign language teachers, involved in projects on digital and cultural heritage like LBC, not to mention public and private stakeholders working in cultural and artistic heritage management.



## Bibliography

- Blake Wiener, James. "Information and Communication Technologies in Cultural Heritage & Tourism." *Ancient History Encyclopaedia*, 13 June 2018. Accessed October 13, 2019.  
<https://www.ancient.eu/article/1242/information--communication-technologies-in-cultura>
- Chabra, Depak, Robert Healy, and Erin Sills. "Staged Authenticity and Heritage Tourism." *Annals of Tourism Research* 30, 3 (2003): 702–719.
- Chiapparini, Alessandra. "Communication and Culture. Why and How Communication Should Become a Support Instrument in the Heritage Valorization Process?" In *Building Materials and Building Technology to Preserve Built Heritage*, edited by Luc Schueremans, 15–33. München: WTA Publications, 2009.
- . "Communication and Culture. Why and How Communication Should Become a Support Instrument in Heritage Conservation Process." In *Building Materials and Building Technology to Preserve Built Heritage*, edited by Rostislav Drochytka and Stepan Bohus, 146–156. München: WTA Publications, 2011.
- Colbert, François, Johanne Brunet, Philippe Ravanas, Mariachiara Restuccia, Yannik St. James, and J. Dennis Rich. *Le marketing des arts et de la culture*. Québec: Gaetan Morin Chenelière Education, 2014, 4th edition.
- Council of Europe. "Framework Convention on the Value of Cultural Heritage for Society." *Council of Europe Treaty Series*, no. 199, 27th October, 2005.
- . *Heritage and Beyond*. Strasbourg: Council of Europe Publishing, 2009.
- EU DigiCULT Report. *Technological landscapes for tomorrow's cultural economy*. Luxembourg: Official Publications of the European Communities, 2002. Accessed October 15, 2019.  
[https://www.digicult.info/pages/report2002/dc\\_fullreport\\_230602\\_scren.pdf](https://www.digicult.info/pages/report2002/dc_fullreport_230602_scren.pdf)
- Gardner-Chloros, Penelope. "Multilingualism and the arts: Introduction." *International Journal of Bilingualism* 18, no. 2 (2014): 95–98.
- Loulanski, Tolina. "Cultural Heritage in Socio-Economic Development: Local and Global Perspectives." *Environments* 34, no. 2 (2006): 51–69.
- Martelloni, Rossella. *Sistemi Culturali. L'interazione dinamica dei beni culturali con il turismo, il mercato, il territorio, le imprese*. Roma: Peraldo Editore, 2004.

Minerva Europe. *Quality Principles for Cultural Websites: A Handbook*. 2005. Accessed October 13, 2019.

<http://www.minervaeurope.org/publications/qualitycommentary/qualitycommentary050314final.pdf>

Sylaiou, Stella, and Georgios Papaioannou. "The Use of ICT in the promotion of arts and cultural heritage education in museums." *International Conference on Strategic Innovative Marketing and Tourism (ICSIMAT)*, 2018.

Zotti, Valeria, and Ana Pano Alamán (eds). *Informatica umanistica. Risorse e strumenti per lo studio del lessico dei beni culturali*. Firenze: Firenze University Press, 2017.

**SECTION 1.**  
**CORPUS LINGUISTICS**

CHAPTER ONE

ON TRANSLATING ART AND  
HERITAGE DISCOURSE IN TOURISM  
FROM ITALIAN INTO ENGLISH:  
A PILOT STUDY

MONICA TURCI AND GAIA ARAGRANDE

**1. Pilot study: aims, text types, and participants**

This article reports on a pilot study aimed at providing information on the translation of heritage tourism texts from Italian into English conducted by Italian-L1 university students with an advanced knowledge of general English. It is the result of a collaboration between the University of Bologna and public and private organisations operating in the field of tourism. Also, it has been created and coordinated by three members of the teaching staff working for the course of the language of tourism as specialised language in the International Master Degree in “Language, Society and Communication”: the convenor of the course, a teaching assistant who is also a scholar researching in translation and corpus linguistics and a native speaker language teacher.

The organisations involved that have provided texts are the *Segretariato Regionale del Ministero dei beni e delle attività culturali e del turismo per l'Emilia Romagna* (henceforth Segretariato) and I.TA.CÀ Festival of Sustainable Tourism (henceforth I.TA.CÀ). This pilot study was presented to the class as one of three exam options students could choose from. For this reason, in this paragraph we will refer to the corpus of texts commissioned as “translation option”. In order to map students’ profiles and attitudes during the translation, students had to answer an online questionnaire created via Google Forms aimed at obtaining data for this sample of participants that, we believe, will prove significant for future developments of our study.

Though these organisations and the texts they commissioned have much in common, there are also important differences that need to be underlined. Hence, the following two sections (1.1. and 1.2) include more detailed information concerning these aspects. As concerns common aspects, the texts commissioned by the the Segretariato and I.TA.CÀ are available in electronic format and address a wide audience of tourists and visitors, in order to promote forms of art and heritage tourism. Finally, section 1.3 provides a linguistic profile of the participants of this pilot study.

### 1.1 *Segretariato*

The Segretariato is a public administrative office responsible for coordinating and implementing national policies from the Italian Ministry of Culture and Tourism (MiBACT) at regional level. Following a recent re-organisation of the Italian public administration, the mission of the Segretariato has been re-fashioned; where previously this organisation was in charge of protecting the cultural art and heritage of the region, now it is responsible also for promoting it.

The texts from the Segretariato corpus constitute precisely an example of this activity; they are part of a project named “Cantiere Estense” focusing on works of art and architecture connected to the ancient House of the Estense family (1471-1598) in what used to be the Duchies of Modena, Reggio Emilia and Ferrara. The description of the project, the way this frames texts and the texts themselves provide an interesting, if not uncommon, case in Italy of a clash between two ways of communicating art and heritage. In particular, the *tourER* (<http://www.tourer.it>) is a multimodal and interactive map that functions as a search engine for the locations and information connected to the Estense heritage. Far from providing solely the channel and medium to access our texts, the *tourER* discursively frames them within a consumer-driven heritage experience (Apostolakis 2003: 796) that combines a user-friendly and intuitive technology with short, simple and engaging messages that directly address a wide audience of potential visitors inviting them to play an active role in the tourist experience, to construct their own personal journeys, to discover new routes off-the-beaten track and even to contribute with their comments and photographs.

In contrast to the *tourER*'s engaging communicative style and strategies, the texts on the Estense project commissioned for translation are monomodal, devoid of linguistic resources that address readers. These texts are mainly descriptive, employing highly specialised lexis on art and

architecture; this suggests that they are targeted towards a small elite of art experts or highly educated arts-core visitors often associated with the *Grand Tour* (Buczowska and Banaszkiwicz 2015: 53) rather than contemporary tourists. The contrast between the *tourER* website and the texts commissioned for translation, in particular in relation to the potential audience –a wide audience of contemporary experiential visitors on the one hand, and a restricted elite on the other– is bound to complicate the translation task and has certainly proved problematic for the students taking part in our pilot project, as we shall see. Suffice for the moment to say that 76% of students found that the Segretariato texts were the most complicated ones.

## 1.2. I.TA.CÀ

I.TA.CÀ started in Bologna in 2008 as a group of small local non-governmental organisations that wanted to find new ways to promote their territories through sustainable tourism. Though it has operated mostly at local and national level, it has recently gained international recognition after it was been awarded a prize for excellence and innovation in tourism by the World Tourist Organization.

I.TA.CÀ's mission and the texts it has commissioned provide more explicit and consistent connections with some of the most salient aspects of contemporary heritage tourism than the Segretariato ones, in particular as concerns forms of tourism and audience's engagement. The I.TA.CÀ manifesto lists amongst its several activities "guided tours aimed at enhancing the cultural and historical heritage of the area" (<https://www.festivalitaca.net/it-a-ca-migrants-and-travelers-festival-of-responsible-tourism>). Moreover, this provides references to heritage tourism through invitations to participants to get involved. The long list of the activities organised by I.TA.CÀ, that are described in detail in the texts commissioned for translation, go from hiking to museum visits, from yoga classes to experiential meals reinforcing the connection between I.TA.CÀ and heritage tourism in that they provide an example of what Francesconi has effectively and succinctly described as the "multifarious and multifaceted" sides of heritage tourism that encompass a wide category of attractions ranging from "natural or cultural forms, including objects, places and practices" (2016: 153).

In conclusion, the I.TA.CÀ's texts fulfil the same descriptive function of the Segretariato ones, at the same time they differ in that their topics and communicative functions are more consistent with aims and forms of

contemporary heritage tourism addressing a wide audience of experiential tourists, whose interest encompasses numerous attractions and activities.

### **1.3. Participants**

This pilot study originated in the classroom involving the students enrolled in the course of “English for Specific Purposes” during the 2018/2019 academic year. In this course, students can choose between 4 exam options, among these the “translation option” that essentially provided the material for this investigation was chosen by the 36.6% of the students enrolled.

The remaining options focused on different, and yet connected, writing skills: critical and close reading skills if students chose to write an essay analysing an aspect and/or a text connected to tourism, and creative writing skills if they chose to write a short brochure or guided tour relating a tourist experience, place or event that has never been discussed in promotional and informative material written in English. Overall these options were aimed at having students reflect on the multivariied aspects and functions of the language of tourism. The translation option in particular focused attention on art and heritage tourism.

Students were asked to spontaneously form working groups from a minimum of 2 to a maximum of 5 and were encouraged to work as a team, rather than translating individually a part of the texts assigned. A total of 16 groups were formed. The “translation option” thus can be defined as a “take-home” exam in that no limitations had been set by the convenor and the tutor of the course as far as work organisation and resources were concerned. Indeed, each group was independent in their elaboration of the translation task, although the tutor of the course was available for a limited number of meetings in order to provide students technical support and advice on resources and translation-related issues.

Students had two deadlines: after a month from the assignment they had to hand in a draft of their translations (referred to later as Tourist Learner-Translation Corpus), which they could discuss with the convenor and the tutor of the course. They then had a fortnight to make changes and corrections and hand in the final version of the translation for marking.

It was mandatory for students to fulfil some requirements additional to those necessary to be enrolled in the Master Course; namely students had to have either a valid certificate showing that they had a knowledge of English at the level C1 in all skills according the Common European Framework of Reference for Languages (CEFR), or a mark >26/30 in

“English Language 3” from their BA Degree<sup>1</sup>. Previous translation experience was strongly recommended but not compulsory. Regarding this last aspect, the questionnaire mentioned above showed that 83% of students had previous experience in translation, either from Italian into English or from English into Italian. Table 1-1 below provides a summary of the “translation option” that is the core of our study.

Translation option	
Exam type	Take-home
Time available	6 weeks
Number of groups	16 (2 to 5 people per group)
Workload per person	Approx. 2000 words
Requirements (compulsory)	C1 (CEFR) or mark >26/30 in last BA English exam
Desiderata (optional)	Previous experience in translation

**Table 1-1.** “Translation option” exam: information and characteristics.

As to course content and preparation in relation to the translation option, connections were made between translating skills and writing skills specifically in the field of tourist discourse. It was pointed out to students that in order to produce a good translation, background knowledge of the style and function of tourist discourse in English was of paramount importance. On the translation front, this recommendation was meant to orient students to privilege, whenever possible, instrumental TT-oriented translation strategies rather than documentary ST-preserving methods (Nord 2005).

As concerns properties, techniques and registers of the language of tourism, students were asked to read Dann’s seminal study (1996). Though this is undoubtedly one of the most complete and accessible texts for students that are new to this specialised language, as well as providing the theoretical framework for several scholars researching translation for the language of tourism (see, for example, Gandin 2013: 326), we felt this study did not provide useful background information for students who were engaging with texts about heritage tourism. As Dann makes clear in the introduction, this book deals mainly with mass tourism that in many respects is at the opposite end of heritage tourism (see Apostolakis 2003: 795).

As far as translation theories and strategies are concerned, two *ad hoc* workshops were organised by the tutor of the course. As previous

---

<sup>1</sup> This is equivalent to a range of marks that goes from a 2:1 to a 1st in the English system.



experience in translation was not required, the aim of these workshops was to unpack the meaning of TT-oriented translation, as well as to provide practical examples and illustration of useful online resources. In particular, these workshops focused on the meaning and implications of covert translation (House 2006) and the related concepts of “cultural filter” and Culture Specific Items (CSIs) in order to foreground strategies and present examples of adaptation and localisation.

As another limitations of Dann’s text is the fact that it tends to privilege the conative function of the language of tourism and examples of promotional material, the workshops provided a particular thorough illustration of functions in translation (Reiss 2001), showing that tourist texts use a medley of informative, expressive and operative functions. Regarding online translation tools, students were provided with a list of online dictionaries, tips for effective search engine usage, as well as an introduction to the effectiveness of comparable and parallel corpora for translation. While doing this it became clear that, although all the students are digital native, the majority of them –with the sole exception of those who had attended a BA Degree at the Scuola Interpreti e Traduttori (SSLMIT), today Department of Translation and Interpreting (DIT), at the University of Bologna– were not familiar with online corpora; hence workshops ended with a hand-on session on the use of Sketch Engine.

## **2. TLTC: Corpus and methodology**

This paper employs Corpus Linguistics (CL) as a methodology to analyse students’ translations of a number of tourist texts, in particular it resorts to the sub-discipline of Corpus-based Translation Studies, which “can be defined as the use of corpus linguistic technologies to inform and elucidate the translation process” (Kruger *et al.* 2011). Back in 1995, Baker suggested a treble typology of translation corpora: “comparable”, “parallel”, and “multilingual” corpora (1995: 230-235). This typology refers to corpora that are specifically built and envisaged for research on translated language and translation practices, but they can also be useful in other related areas of research, such as Second Language Acquisition, Lexicography, Terminology, and Translator Education (Laviosa 2002).

In line with Baker’s (1995) typology of translation corpora, we can define ours as a specialised parallel corpus of learner translations. As sections 1.1 and 1.2 above explained, the texts students were asked to translate are mainly informative. As concerns the type of audience, given that the original environment of these texts is a publicly available website and that the texts are aimed at tourists, we have envisaged an audience of

non-experts. Hence, despite the highly specialised nature of the Segretariato texts, we asked our students to translate picturing a lay-reader as final user of the text, privileging simplification and normalisation whenever possible.

A further methodological aspect that is central to this case study is the role of Translation Studies in the field of Language for Specific Purposes (LSP). Indeed, the intersection of contrastive Translation Studies, CL and LSP can prove to be fruitful on many levels: research, but most importantly, teaching. Following Gandin (2016), we firmly believe that the use of translation (both as a process/activity and as a product) should be rediscovered in the teaching of the LSP of tourism as a L2. Moreover, this case study shows how the use of CL tools can be beneficial in both areas of research and teaching. As Gandin (2016: 80) puts it, CL:

can be effectively employed in contexts of L2 learners of English at university level in order to:

- a) explore and learn the different linguistic, stylistic and pragmatic properties of the language of tourism by means of authentic linguistic materials;
- b) recognise and understand the most/least successful strategies characterising the translation of tourist texts;
- c) discover corpus linguistics methodologies, notions, design criteria, technical tools and procedures for the collection, analysis and interpretation of complex linguistic phenomena, particularly when applied to the study of translation universals.

As such, the interplay of CL and Translation Studies can produce synergies from the classroom to the research group, using the first to feed the second and, in turn, the second to improve the first, creating a virtuous circle of knowledge which is by no means just theoretical but deeply rooted in practice.

The Tourist Learner-Translation Corpus (henceforth TLTC) consists of two sub-corpora, which are to be considered parallel corpora. Indeed, TLTC is composed by a source texts corpus (ST-TLTC) and a target texts corpus (TT-TLTC). The TLTC is mono-directional, in that the direction of STs translation is from Italian (Source Language - SL) to English (Target Language - TL). Intertext editor (Vondříčka 2014) was used to align STs and TTs and uploaded the output spreadsheets on Sketch Engine in order to be analysed.

Regarding the way we exploited this corpus in our analysis, following McEnery and Wilson (2011), we claim that corpus-driven and corpus-based approaches do not exclude each other, but rather are complementary (McEnery *et al.* 2006). Indeed, the role of the corpus within this study is

both of providing the starting point for the analysis (corpus-driven) and of substantiating and/or reject hypotheses made a priori by the analyst (corpus-based). The insights, which this corpus is able to provide within ESP and non-professional translation settings, are obtained mainly through the exploration of keywords, concordance lines and collocation patterns, which the following section (3) reports about. In a contrastive perspective, keywords can provide a good starting point for analysis alongside the exploration of normalised frequencies and concordance lines. Finally, with a corpus like TLTC, concordance lines are quite manageable in numbers, thus we felt encouraged to qualitatively analyse them, starting from an exploration of keywords for both sub-corpora considering them a statistically-relevant starting point for more detailed and fine-grained corpus-based analyses (McEnery and Wilson 2011; McEnery and Hardie 2012).

The TLTC was automatically POS-tagged using the Italian TreeTagger part of speech by Baroni (2004), which is a built-in feature of Sketch Engine. The latter is the platform we used to extract wordlists and keywords, as we will explain later. Table 1-2 sums up the main information about the TLTC.

<b>TLTC Parallel corpus</b>	<b>ST-ITA</b>	<b>TT-ENG</b>
<b>TOKENS</b>	45,285	48,275
<b>TYPES</b>	7,790	6,500
<b>TT-R</b>	17.2%	13.5%
<b>SENTENCES</b>	2,157	2,330

**Table 1-2.** TLTC information.

As Table 1-2 shows, TLTC is a very small corpus, this is because it was manually compiled and, ideally, it will serve as a teaching tool in ESP classes. Despite being smaller in terms of tokens, the ST corpus presents a higher lexical variety (TT-R Type/Token Ratio), which, following Baker (1993), Toury (1995) and Blum-Kulka (2001), can be ascribed to the existence of the so-called translation universals or laws, which are “[...] features which typically occur in translated texts rather than original utterances and which are not the result of interference from specific linguistic systems” (Baker 1993: 243).

Although Baker’s definition of “universal features of translation” seems to exclude “interference”, other scholars (Toury 1995; Mauranen 2004) proved that a lower TT lexical density can be the result of the tendencies of simplification, explicitation and interference. Being this a small corpus, we were able to close-read all the translations and confirm that SL interference is surely present in our data set. Though this is not

enough to call for a universal feature of translation, it certainly explains a higher type-token ratio in our ST corpus.

The case study presented here employs a mixed quantitative and qualitative methodological approach, exploiting CL to extract frequencies and keywords, then comparing our results in the two sub-corpora and qualitatively analysing collocational patterns and gaps. Indeed, both sub-corpora were analysed by means of Sketch Engine featured Wordlists, KeyWords, WordSketch and Concordance functions (using as reference corpora *IT-TenTen2016* and *EN-TenTen2015*). Taking the lead from the quantitative data provided by the Keywords lists we extracted, we moved to a more fine-grained linguistic analysis, which mainly focused on the translation of terminology and syntax, using Sketch Engine's concordancer.

### 3. Results

Our first step into the exploration of the TLTC was extracting frequencies (Wordlists comprehending lexical items only) and Keyword lists, in particular we extracted one- and multi-term keywords from both the ST corpus and the TT corpus. Since keywords in a corpus highlight the “aboutness” of the corpus (Baker and McEnery 2015), our main aim with keywords extraction was both to unveil peculiar uses of the TL in learners' translations and to identify LSP items. Tables 1-3 and 1-4 contain 14 keywords and key multi-terms chosen among the top 100 items in our keyword extraction. The keywords lists were cleaned up as to eliminate names of places and people, because these are not items that require translation and are usually transferred to the TT without being altered. In this way tourists are able to find a correspondence between names of places/historical figures in the text and in the site they are visiting. Finally, our selection was mainly guided by the items that were particularly key in the ST corpus, starting from the assumption that students may have found the translation of those terms more difficult.

ONE-TERM KEYWORDS	KEY-NESS	ST-TLTC FREQ (p.10,000 w)		ONE TERM KEY-WORDS	KEY-NESS	TT-TLTC FREQ (p.10,000 w)
Mastio	177.490	2.87	1	Ducal	621.870	10.77
Merli	161.130	5.29	2	Chiesa	473.990	6
Rocca	124.600	12.36	3	Quadrangular	437.740	5.17

Fortilizio	113.900	2.21	4	Palazzo	396.000	10.56
Levatoio	113.260	2.21	5	Façade	348.400	10.56
Convento	116.610	5.52	6	Loggia	276.820	3.52
Torrione	110.940	3.97	7	Castello	273.590	4.14
Impaginare	103.820	2.21	8	Palazzo	266.070	10.56
Pietrame	102.290	1.76	9	Portico	257.640	4.14
Arenaria	98.380	1.1 3.09	10	Fortress	209.070	13.87
Transetto	95.430	2.42	11	Transept	188.510	2.48
Duca	93.730	28.9	12	Ducal	177.620	10.77
Escursionistica	87.790	1.76	13	Torrione	164.990	2.07
Scalone	86.980	2.21	14	Romanes- que	160.850	2.69

**Table 1-3.** Selected keywords from TLTC Parallel Corpus.

<b>MULTI-TERM ST KEY WORDS</b>	<b>KEY NESS</b>	<b>ST-TLTC FREQ (p.10,000 w)</b>		<b>MULTI-TERM TT KEY WORDS</b>	<b>KEY NESS</b>	<b>TT-TLTC FREQ (p.10,000 w)</b>
architetto ducale	287.600	5.29	1	ducal architect	256.650	2.9
corpi di fabbrica	249.480	6.85	2	environmental hiking guide	183.610	2.27
pianta quadrangolare	213.430	2.87	3	hiking guide	179.130	2.27
coda di rondine	167.940	2.87	4	southern side	171.070	2.48
cortile interno	154.160	4.15	5	internal courtyard	163.390	2.07
fronte meridionale	147.770	1.98	6	bell tower	162.220	2.48
corpo di fabbrica	144.410	6.85	7	ducal palace	128.820	3.72
campagna edilizia	136.470	2.21	8	quadrangular plan	110.560	1.45
ambientale escursionistica	127.380	2.21	9	quadrangular tower	110.560	1.24
perimetro difensivo	117.590	5.52	10	inner courtyard	99.700	1.24

addizione erculea	114.730	5.52	11	corner tower	92.300	1.45
fronte settentrionale	113.840	5.52	12	northern side	91.680	1.45
rilievo collinare	112.440	1.32	13	open gallery	90.410	1.65
dispositivo militare	112.010	1.32	14	ancient church	83.910	1.24

**Table 1-4.** Selected key multi-terms from TLTC Parallel Corpus.

Keywords in a corpus give hints about which words/compounds are key in that particular corpus if compared to larger general corpora (Baker and McEnery 2015). Accordingly, here keyword extraction tool was used to isolate lexical items, which orient research either to specific tourism-related SL-terminology or to TL unusual terms and mistranslations.

From these lists and a first exploration of some concordance lines, we were able to identify four main areas in which, we supposed, students may have encountered some difficulties. The four areas are the following:

1. Untranslatable items: Culture Specific Items (CSIs) and specialized or domain-related language;
2. Mistranslations in nominal groups: calques and/or lexis;
3. High vs. Low frequency translational choices;
4. Punctuation and syntax.

In what follows, we will discuss in detail our results in each of the above-mentioned areas.

### 3.1. Untranslatable items

The first area, “untranslatable items”, certainly was the most rich in terms of example availability in our corpus. Indeed, the nature of the texts students were asked to translate is not only extremely culture-specific, but often entails also domain-specific terminology, whose translation is notoriously problematic. Before going any further, we should clarify what we mean by “untranslatable items”. This labelling may lead to believe that there are some lexical items that are simply untranslatable; however, here, we refer to what the students, according to our investigation, found particularly difficult to translate, to what they conceived as “untranslatable”. Thanks to a close and qualitative investigation of parallel concordance lines, guided by a previous keyword list exploration, we were