

STUDI  
E  
RICERCHE

Gianna Fusco

# TELLING FINDINGS

TRANSLATING ISLAMIC ARCHAEOLOGY  
THROUGH CORPORA

*LED*

—— Edizioni Universitarie di Lettere Economia Diritto ——





ISBN 978-88-7916-753-6

ISSN 1721-3096

Copyright © 2015

*LED Edizioni Universitarie di Lettere Economia Diritto*

Via Cervignano 4 - 20137 Milano

www.lededizioni.com - www.ledonline.it - E-mail: led@lededizioni.com

I diritti di riproduzione, memorizzazione e archiviazione elettronica, pubblicazione con qualsiasi mezzo analogico o digitale (comprese le copie fotostatiche, i supporti digitali e l'inserimento in banche dati) e i diritti di traduzione e di adattamento totale o parziale sono riservati per tutti i paesi.

---

Le fotocopie per uso personale del lettore possono essere effettuate nei limiti del 15% di ciascun volume/fascicolo di periodico dietro pagamento alla SIAE del compenso previsto dall'art. 68, commi 4 e 5, della legge 22 aprile 1941 n. 633.

Le riproduzioni effettuate per finalità di carattere professionale, economico o commerciale o comunque per uso diverso da quello personale possono essere effettuate a seguito di specifica autorizzazione rilasciata da: AIDRO, Corso di Porta Romana n. 108 - 20122 Milano  
E-mail segreteria@aidro.org <mailto:segreteria@aidro.org>  
sito web [www.aidro.org](http://www.aidro.org) <<http://www.aidro.org>>

---

*In copertina:*

Elaborazione grafica di Camilla Balsamo

*Stampa:* Digital Print Service

# CONTENTS

<i>Acknowledgments</i>	7
1. CORPORA AND TRANSLATION: AN INTRODUCTION	11
1.1 Corpus Linguistics: state of the art	11
1.2 Branching out	12
1.3 Corpus Linguistics and lexis	15
1.3.1 Definitions – 1.3.2 Types of corpora	
1.4 Corpora and translation	23
2. ARCHAEOLOGY IN TRANSLATION	35
2.1 Preamble	35
2.2 Language and archaeology	37
2.3 Translation, corpora and the Web	45
2.4 Can you see the foot under that bowl? A case study in Web-based documentation strategie	49
2.4.1 Compromising between synthesis and explicitation – 2.4.2 Opacify vs. opaque, or the opposition between process and result – 2.4.3 Competing descriptions	
3. A PARALLEL AND COMPARABLE CORPUS OF ISLAMIC ARCHAEOLOGY	59
3.1 The ArIIEL corpus	59
3.2 Corpus-based case studies on translation units	64
3.2.1 Scrittura / Grafia – 3.2.2 Ornato / Decorazione – 3.2.3 Campito / Campitura	
<i>Glossary</i>	85
<i>References</i>	111



# ACKNOWLEDGEMENTS

Stating that there are people without whom a certain book would simply not exist is almost a convention of that sub-genre of academic writing that is the acknowledgments section. Yet, indeed I owe the completion of this project to a number of friends and colleagues. First among them, Oriana Palusci, who is always a supportive and motivating presence in my professional life. She kindly invited me to take part in the organization of the conference *Translating East and West*, which was held at “l’Orientale” University of Naples in November 2012, and encouraged me to present at that conference. I started this project thanks to that opening and to the fruitful conversations I had there with several linguists. Through Oriana, I thank them all.

Camilla Balsamo is the person with whom I worked in closest contact for the entire duration of the writing process of this book. Camilla helped with data editing and processing, tested hypotheses, gave a crucial contribution to the final formatting of the graphic part of the book. In brief, she is the most helpful assistant a researcher could hope for and an intellectual sparring partner of unparalleled talent.

Maria Vittoria Fontana and Roberta Giunta are to me the idea itself of an archaeologist. For the past ten years they have been a friendly working connection, the gentle voices at the other end of concerned phone calls as a translator. With humour, kindness, and much understanding they introduced me to the wonders of Islamic archaeology.

The Department of Literary, Linguistic and Comparative Studies at “l’Orientale” is an invaluable source of inspiration for all of us working there. I owe much to all my colleagues who constantly provide an intellectually stimulating but laid-back atmosphere to my working days.

My endless gratitude goes to Donatella Izzo for reading every single thing I write, for sharing with me her wisdom and her immense knowledge, and for being supportive beyond any fairest expectation. And for the endless laughs, of course. Fiorenzo Iuliano is a role model of intellectual integrity, work ethic, and professional commitment. He is also among the closest friends I have in life, so I wish to thank him for being there, across distance and over the years.

I owe much fun to a network of amazing people I have met in professional milieus and to whom I am indebted for the daily joys of my life in Naples and away. Vincenzo Bavaro had me stay so many times at his and allowed me to cat-sit, which is one of my favourite summer activities. Serena Fusco and Carmen Gallo know the perfect timing for a cup of Neapolitan coffee, not to say about the wisdom that unmistakably accompanies it. Claire Fennell, Judy Moss and Leonardo Buonomo have treated me to countless nights of home-cooked food, couch and TV series, mixed with lively debates on politics and academia. Their heart-warming friendship is more appreciated than I am able to say in a couple of lines. Giuseppe Polise, the unexpected friend, has set a new standard of thoughtfulness that is best expressed through mischievous humour and good music, both of them delivered with amazing timing.

I have thought about writing this book for quite a long time. I eventually wrote it in the course of an extremely rewarding year of teaching, and completed it during an intense session of exams. I wish to thank here my students for having been in the past months the kind of learners any teacher dreams of. Some of them provided, however unwittingly, such a wealth of emotional support to deserve being mentioned by their name. Maria Chiara Aurilia, Antonio Di Donato, Stefano Morello, Anastasia Napolitano, Chiara Postorino: thanks for bringing a smile, coffee, and / or cookies to my office hours, all of them were extremely appreciated.

I would simply not be the person I am without my family, who has been teaching me love, care, and unconditional support for my entire life. My mother, Maria Grazia Marotta, is an inestimable source of daily good humour and intelligent conversation. My brother Domenico, my sister-in-law Delia, my cousin and best friend Michele Cinicola, my close and extended family, have always been there for me, indulging me when I seek solitude to



write, and making such a huge difference in my life that I cannot make it justice in words.

While I share the merits of this work, if any, with all the people I have mentioned here, any inaccuracies or mistakes, needless to say, are entirely my own responsibility.

I dedicate this book to the memory of my beloved father, Francesco Fusco, who taught me how to build bridges between where I am and where I want to be, and walk the distance with gusto.

G.F.



# 1.

## CORPORA AND TRANSLATION: AN INTRODUCTION

### 1.1. CORPUS LINGUISTICS: STATE OF THE ART

Corpus Linguistics has undoubtedly been one of the most influential areas of linguistic research over the past decades, a field of inquiry steadily growing and rapidly evolving, bearing witness, through major theoretical and methodological leaps forward, to its deep alliance with computer sciences and IT. As a field of study, in fact, Corpus Linguistics can be broadly defined today as being founded on the possibility itself of storing authentic language through the use of computers and analyzing it through specifically designed software (Freddi 2014) <sup>1</sup>.

The coupling of computer technology and authenticity of linguistic data in the early days of the discipline as we know it now constituted nothing less than a revolution in the approach to language, leading to the upheaval of long-held principles about language use and affecting not only linguistics as a disciplinary field, but also a larger arena of language learners, teachers, publishers, translators, in brief, anyone who had an interest in understanding how languages work and why. Its deep impact, extending well beyond the range of academic debate and resulting in the publication of truly revolutionary dictionaries and grammars, has

---

<sup>1</sup> A true pioneer in computer-based Corpus Linguistics, Jan Svartvik points out how corpora were all but unknown to linguists before IT: “While it is natural today to take ‘corpus linguistics’ to mean ‘electronic corpus linguistics’, we must not forget that there were language corpora BC, i.e. ‘before computers’. There were lexicographical corpora, especially that used for the *Oxford English Dictionary* which, by the time the first edition was completed in 1928, had accumulated over four million citation slips from two thousand volunteer readers” (2007, 12).

produced a paradigmatic shift in the way authoritativeness itself in language use is defined and acknowledged. A major landmark in this sense is certainly the publication of the first corpus-based dictionary of the English language, the *Collins COBUILD Dictionary*, produced by John Sinclair and his research team in 1987<sup>2</sup>. The impact of the discipline was equally eventful in the field of grammar studies, especially when the consistently pragmatic approach offered by corpora is contrasted with the highly abstract trend pioneered by Chomsky in the 1950s and 1960s. The landmark publication registering the impact of Corpus Linguistics in this sense is the *Longman Grammar of Spoken and Written English*, by Biber et al., in 1999. Such major outputs were made possible by significant advances in computer sciences and IT that eventually allowed linguists to gain analytic access to huge quantities of stored data, as compared to the smaller corpora they had available in the previous decades. Following Renouf (2007), it is thus possible to trace a significant shift from the era of relatively small corpora (i.e. from the 1960s on) to the era of the super corpora (i.e. from the 1980s on). Renouf appropriately identifies the advent of the Internet as another turning point, which again significantly broadened the research ground for linguists and gave rise to cyber-corpora and theories of the Web as corpus (from 1998 on)<sup>3</sup>.

## 1.2. BRANCHING OUT

As many successful research area, Corpus Linguistics soon seemed to branch out in a variety of interconnected subfields, each of them calling for highly specialized knowledge on the researcher's part, but contributing to the general aim of studying language from real instances of use. From Computational Linguistics to Corpus-based Translation Studies (CTS), from Natural Language Processing (NLP) to corpus-based contrastive lexicography, Corpus Linguistics is sometimes seen today as an umbrella discipline, which keeps together in a common vast scientific effort a

---

<sup>2</sup> COBUILD stands for "Collins Birmingham University International Language Database".

<sup>3</sup> The question of the Web as corpus is still an object of debate within corpus linguistics and will be discussed further in this book.

rich array of theories and methodologies for the understanding, studying and professional use of language. Since all corpus linguists use corpora with the purpose of basing their analyses “on real data – actual instances of speech or writing – rather than on data that are contrived or ‘made-up’”, Corpus Linguistics has been defined by some scholars in the field not as “a separate paradigm of linguistics but rather [as] a methodology” (Meyer 2004: xiii), a “methodological basis for pursuing linguistic research”, to use Leech’s words (1992: 105). Other prominent voices in the discipline, however, object to such a definition of Corpus Linguistics as a set of methods for the study of language, and advocate a major distinction within the field between what has come to be known as corpus-based and corpus-driven approaches.

The distinction was first introduced in these terms by Elena Tognini Bonelli (2001) and, although quite hotly debated within the discipline, is at the core of what is now known as the “neo-Firthian school” headed by John Sinclair, undoubtedly one of the major contributors to the rise to prominence of Corpus Linguistics within academic discourse, but also outside it, to wit, within the publishing industry as well as foreign language teaching and training. According to this view, corpus-based approaches can correctly be considered a methodology, as they use corpora to “expound, test or exemplify theories and descriptions that were formulated before large corpora became available to inform language study” (Tognini Bonelli 2001: 65). A corpus-driven linguist, on the other hand, sees corpora not simply as the source of data, but rather as the sole source of theory itself. The starting point for this methodological positioning is that of accepting the evidence (Tognini Bonelli 2001: 84) and then directly derive theory and descriptive statements from the observation of the type of information that only the corpus / corpora can yield, such as frequency distribution, recurring patterns, or the absence thereof. As theory and description of language strictly “reflect the evidence” (Sinclair 1991: 4), the corpus within this approach

is seen as more than a repository of examples to back pre-existing theories or a probabilistic extension to an already well-defined system ... The theory has no independent existence from the evidence and the general methodological path is clear: observation leads to hypothesis leads to generalisation leads to unification in theoretical statement (Tognini Bonelli 2001: 84-85).

In other words, corpus-driven linguistics investigates corpora in order to bring to light the theory of language that the corpus already embodies, and is for this reason also referred to as ‘corpus as theory’ and opposed to ‘corpus as method’<sup>4</sup>.

However, it has been noted that a sharp distinction between these two views of corpus work can result into the polarization of what is actually a quite blurred and often overlapping difference. As stated by McEnery and Hardie,

the corpus-based versus corpus-driven distinction implies a dichotomy where there is actually a sliding scale ... Within what would be dubbed corpus-based linguistics, we see an entire range of roles for the corpus, from providing (at most) a series of examples to illustrate a grammatical theory developed independently of corpus linguistics ... to being the source of most of the claims made... Similarly, studies by practitioners of so-called corpus-driven linguistics do not always rely solely on a corpus in the strict sense (151).

Along a similar line, and with specific regard to lexis, which as we will see, constitutes a core part of both the present study and Corpus Linguistics in general, Altenberg and Granger (2002) notice that the term ‘corpus-based’ can refer to

any work – theory-driven or data-driven – that makes use of a corpus for language description, but it is also used in a restricted sense to refer to studies which start from a model postulating a cross-linguistic difference or similarity on theoretical grounds and use a multilingual corpus to confirm, refute or enrich the theory. The latter approach, on the other hand, may start from an implicit or loosely formulated assumption but uses the corpus primarily to discover types and degrees of cross-linguistic correspondence and to arrive at theoretical statements. In practice, however, the distinction may be slight. The difference lies rather in the importance attached to the initial assumptions and the role that the data play in the analysis. (15)

The difference between the two approaches, in other words, cannot be presented, according to this view, as a radical opposition. Both of them will in fact make some use of a theory elaborated outside the corpora: in the case of what is called here ‘theory-

---

<sup>4</sup> For a quite exhaustive overview of the debate on corpus-based vs. corpus-driven approaches see McEnery, Hardie (2012), 149-152.

driven’, and that others call ‘corpus-based’ approach, a certain notion of a linguistic phenomenon is postulated (sometimes it would be better to say that it is actually inherited from traditional linguistics) and then tested by interrogating the corpus. The outcomes of this process may vary from a stronger validation of the theory, to its absolute rejection, to its refinement. Within a ‘data-driven’ (or, following Tognini Bonelli’s terminology, ‘corpus-driven’) methodology, on the other hand, the questions that we hope the corpus will help find the answer to reveal something about the underlying theory guiding the researcher. In other words, when interrogating the corpus, the researcher has already identified a significant aspect of language for investigation, and this constitutes implicit (i.e. not stated or motivated) or roughly sketched assumptions about the language(s) involved. In both cases, a theory of language, positioned along a scale going from unrefined intuition to highly sophisticated models of linguistic system, is formulated in a space located outside the corpus. Following this view of a difference in emphasis rather than in kind between the two approaches, the label ‘corpus-based’ will be used in its more extensive meaning throughout the present study, that is “as an umbrella term covering both types of corpus-informed studies” (Ibidem).

### 1.3. CORPUS LINGUISTICS AND LEXIS

One of the areas of language research where Corpus Linguistics has proven a game changer is, as I have just already hinted, lexicography, both within the scope of a single language and in the contrastive analysis of two or more languages. The emergence of Corpus Linguistics, in fact, has given a crucial contribution to the identification and theorization of the lexical level in language, with the concomitant shift in the understanding of several phenomena as connected with lexis rather than with syntax. We owe to Corpus Linguistics, in fact, the development of specific research tools for the analysis of collocation, concordances, and frequencies that have allowed linguists to find extensive evidence of the existence of phenomena that are “neither grammatical nor semantic” (Altenberg and Granger 2002: 4), but are rather connected to the position words usually occupy within sentences:

Corpus linguistics provides the methodology to take linguistics, and lexicology in particular, beyond the single word as the basic semantic unit. Rather than decontextualising words and describing their meanings in the isolation of a lexical entry, corpus linguistics breaks down the border between syntax and the lexicon by identifying semantic conglomerates in corpora, combining the parameters of recurrence, statistical significance and syntactic categorisation (Teubert 2002: 212).

As a consequence, lexis is no longer perceived as a linguistic phenomenon quite independent from grammar, but the two are actually understood as being interdependent (Sinclair 1991). This has given renewed impetus to the study of lexis also in a contrastive perspective, as the increasing use of corpora in fields such as translation studies and second language acquisition provided both the raw data for analysis and a horizon for future application.

All that has been said so far, especially with regard to issues of terminology and definition of the field, clearly hints at the complexity of Corpus Linguistics as a body of research and knowledge, and reveals the multifaceted character of the intellectual trajectories of its many subfields, which often intersect other bodies of scholarship in a truly interdisciplinary fashion. It may thus be useful, and perhaps even necessary, before moving on to a more detailed discussion of the corpora under examination here, to give some definitions of key terms that will help the reader navigate their use in the course of the present study. In the light of what has just been said as to the debates within the discipline, the first term that requires better definition is probably Corpus Linguistics itself.

### 1.3.1. *Definitions*

Corpus Linguistics is here intended as that methodology of linguistic investigation that aims at describing languages as they are used, on the basis of actual instances of written texts or spoken utterances gathered in corpora that are designed to be stored and analysed through the use of computers. Such investigation of corpora is understood as having theoretical implications for the understanding of specific aspects of language and the corpus is seen as a reliable source of authoritative usage.

As is all too obvious, the entire enterprise of corpus linguistics relies on the definition and compilation of corpora to be then



analysed. Defining what a corpus is, what its characteristics must be for it to count as a valid tool for linguistic analysis, however, is one of the issues still debated in the field. Definitions of corpus vary in fact from very rigid and narrow to ample and loosely delimited ones. A corpus can be defined very broadly as “any collection of texts (or partial texts) used for purposes of general linguistic analysis” (Meyer 2004: xii), or more specifically as “a computerized collection of authentic texts, amenable to automatic or semi-automatic processing or analysis” (Tognini Bonelli 2001: 55)<sup>5</sup>. Ideally, a corpus is not only authentic, but also representative (or balanced), that is large enough and inclusive enough and balanced enough with regard to the different instances of language it collects as to constitute a reliable statistic sample of the variety of language under scrutiny. Pointing at what is, for a field whose very existence is based on sampling, “more serious than academic point-scoring” (Leech 2007: 135), Geoffrey Leech underlines how “without representativeness, whatever is found to be true of a corpus, is simply true of that corpus – and cannot be extended to anything else” (Ibidem). The issue of representativeness is thus a particularly thorny one for scholars, who, for all their efforts in compiling carefully planned corpora, are aware that “the corpus is finite, but language is not” (McEnery, Hardie 2012: 17), so that representativeness remains, in the words of Leech’s himself and of others following him, the ‘holy grail’ of Corpus Linguistics, “something to strive for rather than something that can reasonably be attained” (Zanettin 2011: 15).

As can be easily inferred, the level of representativeness that can be achieved in a given corpus depends on several factors, among which also practicability and feasibility play an important role. The corpus, in other words, will be designed on the basis of both a certain set of ideas entertained by the research group about language, and the body of texts they can gain access to in order to digitalize them for inclusion into the analysis. A corpus may be, for example, unbalanced in terms of the percentage of written vs. spoken texts, with regard to the number of regional varieties featured, when it comes to the even representation of different social classes, in relation to gender or age, education, and so on. Such

---

<sup>5</sup> An overview of different definitions of corpora can be found in Freddi (2014: 9-10) and Milizia (2012: 21-22). See also Baker et al. (2006: 48-49).

differences may be the result of very practical issues, such as the type and number of texts available for inclusion in the corpus, but they also reflect the often implicit, under-theorized assumptions of the researcher(s) about language. These biases are to a certain extent unavoidable (Ahmad 2008: 61) and this is one of the reasons why studies in Corpus Linguistics should always include a detailed account of the way in which the corpus was designed.

### 1.3.2. *Types of corpora*

A first, basic distinction will then be operated between general (or reference) and specialized corpora. The former, which include some of the most important corpora of English, aim at being representative of a language as a whole thanks to the storage of huge quantities of real instances of language use. Among these, a benchmark for any other more recent corpus is constituted by the BNC (British National Corpus) which includes 100 million words of written and spoken English from the early 1970s to 1993.

Equally important and considerably bigger is the COCA (Corpus of Contemporary American English) which was started in 1990 and is still growing, counting 450 million words in 2012 and designed to be constantly updated with the inclusion of samples of both written and spoken language every year.

Both these corpora, as well as the huge BoE (Bank of English, comprising 500 million words), though aiming at achieving a very high standard of representativeness, privilege written over spoken texts, to the point that, in order to give due analytic and theoretical attention to spoken English, a separate corpus was designed, called CANCODE (Cambridge and Nottingham Corpus of Discourse) and including 5 million words of transcribed conversations. As the considerable difference in numbers clearly highlights, a major issue with spoken corpora and the main reason behind the underrepresentation of this variety of language in general corpora is the availability of raw data and the cost of their digitalization, in terms of both time and financial resources needed in order to tape, transcribe and digitalize them before they can be processed and analysed.

Among general corpora, a further distinction is to be made between static and dynamic corpora. Whereas the former desig-

nate corpora (such as the BNC) that try to give a faithful and detailed picture of a language in a certain time span and through a predetermined number of words (e.g. 100 million words), the latter refers to ever expanding corpora aiming at monitoring the evolution of the language over a longer time span. The above-mentioned COCA (developed and updated by Mark Davies at Brigham Young University, Utah) was the first dynamic corpus and is thus regarded as the research endeavour that opened up a new space of theoretical elaboration in the field.

The ever-expanding, open-ended corpus that seemed nothing less than a chimera, a utopian horizon, just a few decades ago is now regarded as one of the major achievements in Corpus Linguistics, firmly positioned at the top of the research agenda of prominent scholars who are devoted to elaborating ways of fully exploiting its enormous potentialities. Other challenges, however, lie ahead as technology rapidly evolves, resulting in an unprecedented impact on language practices and textual production. The very definition of text is in fact to be tackled in ways that account for the huge influence exerted on our lives and linguistic behaviour not only by the written or spoken text, but visual and multimedia ones as well. This concern is at the core of recent attempts to build multimodal corpora (Allwood 2008; Baldrý and Thibault 2006; Baldry 2007) that include “websites, films, or television programmes, i.e. texts that incorporate several types of data, such as images, sound and sometimes texts” (Murphy 2012: 42). Thus, the corpus will gather and organize, according to the research goals, a variety of semiotically encoded data, providing insights not only into language *stricto sensu*, but in communication as a multidimensional phenomenon involving sight, sound, and text. At the centre of this type of investigation is the construction of meaning “from a semiotic, social, linguistic and cultural point of view” (Ibidem), as the multimodal corpus “gives one an opportunity to capture not only written language (or a written transcription of spoken language) but provides an opportunity to include information of a contextual and cultural kind” (Allwood 2008: 212).

Multimodal corpora, however, also because of the considerable difficulties they present in terms of design and management, are still far from being available to a wide community of researcher, or to the general public, as happens with some of the

major general corpora. A typical characteristic of the latter, which contributes to making them searchable also by people who did not take part in their design and compilation, is constituted by the fact that they are usually annotated, that is each word in the corpus has been tagged or marked-up in order to carry with it some basic linguistic information (such as the part of speech and / or syntactic function); the tagging is what allows a series of computer-run linguistic queries by researchers, and interrogation of the corpus, as new tags are added to the existing ones, even from perspectives that had not been originally thought of by the compilers. According to the advocates of a 'corpus-driven' approach, however, this means that only annotated phenomena can be analysed, resulting in a huge influence of the corpus-designers' and subsequent taggers' pre-existing notions about language and its use on the research outcomes. For this very reason, a corpus-driven approach will privilege the analysis of non-annotated corpora, whose raw data can be investigated limiting the impact of the researchers' biases as to the function or meaning of a specific unit. Within this view, it is the corpus itself that will reveal such function and meaning of the unit under scrutiny by showing its behaviour in the context of actual instances of language use (Freddi 2014: 33).

Not all Corpus Linguistics is based, however, on the analysis of either annotated or non-annotated general super-corpora, and some very interesting insights come from sometimes smaller collections of digitalized texts, especially with regard to the immediate impact of research on very practical aspects of working with languages. These are specialized corpora that are not (and do not want to be) representative of a language in its entirety, but are especially designed to suit the needs of specific bodies of research and scientific endeavour. Among these, particularly important are the corpora designed to study Second Language Acquisition (SLA)<sup>6</sup>, and those bringing together texts across different lan-

---

<sup>6</sup> The most important one for the English language is certainly the ICLE (International Corpus of Learners of English) comprising 3,7 million words from learners of English with 21 different mother tongues, compiled in Louvain-le-Neuve and supervised by Sylviane Granger (Granger 2003). Also very significant, especially as it is used for the standardization of EFL tests and the description of bands of proficiency, is the CLC (Cambridge Learner Corpus). An interesting corpus designed to look at the affective / emotional side of language learning is the EFL - P.Æ.C.E (English as a

guages<sup>7</sup> for contrastive or translational research<sup>8</sup>.

A major distinction needs to be introduced here, namely that between parallel and comparable corpora, both of them being in turn further divided into subtypes according to the number of languages involved and / or the specific purpose guiding their design. Giving a definition of these corpora, however, can be in itself quite problematic since there is no general consensus in the Corpus Linguistics scientific community around the use of parallel vs. comparable, with the option “translation corpus” being propounded by some scholars (Aijmer and Altenberg, 1996; Granger, 1996) in addition to these<sup>9</sup>. Such prolonged debate is also due to the fact that “corpora involving more than one language are a relatively new phenomenon, with most research hailing from the 1990s” (McEnery and Xiao 2008: 19). Given the empirical nature of corpus work, it is then quite predictable that terminology will evolve and progressively move towards standardization as a growing body of scholarship in the field is published, which will account for the corpora that are being designed and the way they are put to use by researchers.

---

Foreign Language - Preferences. Aspettative / Expectations. Certainties. Emotions.), compiled by Liliana Landolfi at Università “l’Orientale” in Naples (Landolfi 2012).

<sup>7</sup> To say that a corpus brings together texts across different languages does not necessarily imply that the corpus will be bilingual or multilingual, since, as we will see, we can have monolingual corpora that still illuminate aspects of the mutual relation between languages or of translation as a process.

<sup>8</sup> Of great interest for scholars in different disciplines, from linguistics to law, is the EUROPARL (European Parliament Proceedings Parallel Corpus) which collects proceedings in the official languages of the European Community. Another parallel corpus (this time a French-German bilingual one) is the GeFrePaC (German French Parallel Corpus). It includes 30 million words and has been compiled by a research unit in Mannheim collecting official documents issued by the European Commission and translations of the European Parliament’s record of proceedings (Teubert 2002 : 203). A specific focus on the intersection between corpus linguistics and translation studies is offered by the TEC (Translational English Corpus), a 10 million words corpus compiled under the direction of Mona Baker and comprising translations into English from a variety of languages, both European and non-European. One of the first bilingual corpora is the ENPC (English-Norwegian Parallel Corpus), a bidirectional parallel corpus compiled in the 1990s and totalling 2,6 million words. Equally interesting is COMPARA, a “free, online parallel, bi-directional and extensible corpus of English and Portuguese literary texts” (Frankenberg-Garcia 2009: 51), totalling around 3 million words.

<sup>9</sup> For a review of the terminological controversy see McEnery and Xiao (2008: 19-20).

As has been convincingly argued by McEnery and Xiao, a crucial aspect to take into consideration when trying to settle this terminological issue is that of consistency in the criterion that is used in deciding whether corpora fall into one category or the other. In other words, if we decide that our definition of what a comparable corpus is relies on the sampling methodology (as opposed, for example, to the type of content of the corpus, or the number of languages involved, and so on), the same criterion is to be applied to the definition of what a parallel corpus is. (McEnery and Xiao 2008: 19). Based on the criterion of content, then, throughout the present study parallel and comparable corpora will be defined as follows:

- a parallel corpus is “a corpus that contains some source texts and their translations (McEnery and Xiao 2008: 20), or “texts that have been produced simultaneously in two or more languages” (Hunston 2002: 15), such as the EU and UN regulations;
- a comparable corpus is a corpus “built according to similar criteria” (Murphy 2012: 41) and “consisting of original, independent, monolingual texts, comparable across languages” (Ulrych 1999: 85) as they contain “components that are collected using the same sampling frame and similar balance and representativeness” (McEnery and Xiao 2008: 20).

In other words, parallel corpora gather texts originally written in one language alongside their translation(s) in one or more languages (Milizia 2012: 70), whereas comparable corpora include texts in different languages that have been selected to represent “the *same proportions* of the texts of the *same genres* in the *same domains* in a range of *different languages* in the *same sampling period*” (McEnery and Xiao 2008: 20, emphasis in the original) <sup>10</sup>.

Though the labels ‘parallel’ and ‘comparable’ are mostly used to refer to bilingual and multilingual corpora, research in transla-

---

<sup>10</sup> Some scholars (Hunston 2002, among others) maintain that the definition of comparable corpora extends to corpora sampling different varieties of the same language, such as the ICE (International Corpus of English). Others maintain that this view would virtually extend the definition of “comparable corpora” to include any collection of digitalized texts designed for linguistic analysis, including general corpora such as the BNC, as all corpora are ultimately created for comparative study of some kind, be it synchronic, diachronic, intralingual, or multilingual.

tion studies has also outlined a further sub-categorization (Murphy 2012: 41; Scarpa 2002: 146; D'Angelo 2008: 207) within these two major types to account for the existence of monolingual comparable corpora and of monolingual parallel corpora. The former include “two separate but comparable texts in the same language, one originally written in that language and the other a translation into it” (Ulrych 1999: 85), whereas the latter identifies corpora including several versions of a text “translated into the same language, generally by different translators, or different diachronic versions by the same translator” (Ibidem). Different parallel, comparable and hybrid corpora are being and will be created to help a variety of research trajectories and assist translators in many practical tasks. Quite big corpora, designed following the same guidelines for balance and representativeness that apply to general corpora, will be best suited for contrastive analyses of linguistic aspects such as lexis or grammar, while specialised parallel corpora (even when markedly small in the era of ever-growing cybercorpora) will be of great assistance for the extraction of specialized terminology, the study of translationese, the training of translators, not to speak of the advances in Machine Translation (MT) and the practical use by professional translators.

From the lively debate outlined above, it is clear how research at the intersection of translation studies and Corpus Linguistics is thriving and fruitful, as it not only allows for a wealth of contrastive analyses and translational theorization, but also fosters deeper insights into the interrelations of language(s) and culture(s).

#### 1.4. CORPORA AND TRANSLATION

A recurrent and quite consistent interest in translation as a complex phenomenon can be indentified in the work of some of the most prominent 20th-century linguists and scholars, from Firth to Malinowski to Catford<sup>11</sup>. Furthermore, with the emergence of corpus-based linguistic work that completely revolutionized the compilation of dictionaries and grammars, as well as the teaching

---

<sup>11</sup> An overview of the status of translation within linguistics during the 20<sup>th</sup> century can be found in Anderman and Rogers 2008.

of foreign languages, the extension of Corpus Linguistics to translation and bilingual / multilingual research seems a high predictable move. Yet, as noticed by Anderman, it is only in the 1980s that “linguists interested in corpus-based studies began to turn the potential role of corpora in the study of translated texts, initially literary texts such as novels” (2008: 13). Such interest emerged, quite understandably, outside the English speaking world, and more specifically, in Scandinavia, that is outside the academic environment that had been crucial to the rise and establishment of Corpus Linguistics and that would produce the first landmark publications directed at the general public and inspired by the new theories. It has been argued that in countries such as Norway, the interest toward then innovative corpus-based studies might have been fostered by the concurrent interest toward the approach to English by non-native speakers and students (Svartvik 2005). This is how Corpus Linguistics became involved with another emerging field, namely that of Translation Studies, through the compilation of a corpus comprising English novels translated into Swedish and a corresponding amount of literary texts originally written in Swedish, which allowed to observe phenomena such as translationese in Swedish novels translated from English (Anderman 2008: 13; Gellerstam 1986).

Since those early stages, the intersection of Translation Studies and Corpus Linguistics has produced a considerable amount of scholarship, with growing regularity and specialization at the beginning of the new century. Corpora are currently used as the theoretical and / or practical foundation for studies sweeping from a general investigation of translation, to specifically linguistic explorations of contrastive aspects of languages, from analyses of bilingual or multilingual literary archives to the training of new generations of translators<sup>12</sup>. These studies can be said to basically fall into two wide categories: theoretical vs. practical corpus-based translation studies. Within the framework of the theoretical approach, “corpora are used mainly to study the translation process by exploring how an idea in one language is conveyed in another language and by comparing the linguistic features and their frequencies in translated L2 texts and comparable L1 texts”

---

<sup>12</sup> See, among others, Haliday 2001, Kenni 2001, Riccardi 2002, Laviosa 2002, Zanettin et al, 2003,



(McEnery and Xiao 2008: 22). The practical approach, on the other hand, is oriented toward the impact of corpus-based research in translation in the professional field, that is to say, it looks at how corpora can help in the training of future translators, the editing of translations, the development of sophisticated applications for computer-assisted and machine translation systems (CAT and MT respectively). In both cases, the use of corpora in connection with the study of translation aims at accounting for

both language systems and language use, i.e. the task is not only to identify translation equivalents and 'systematic' correspondences between categories in different languages, but to specify to what extent and in what respect they express 'the same thing' and where similarities and differences should be located in a model of linguistic description (Altenberg and Granger 2002: 18).

As we have seen, the corpora that, being designed to allow observations across different languages, are ideally suited for this task are the parallel and comparable ones. Yet, designing a corpus or using an existing one for contrastive analysis or to enhance professional translations is all but an easy enterprise. First of all, a distinction has to be drawn between highly specialized and general corpora, as the quantity of data influence the kind of investigations that can be conducted. We have already seen how the compilation of a general corpus that aims at being statistically representative and satisfactorily balanced is an almost unattainable goal, despite decades of research and refinement of techniques and theories. The task becomes even more complex when more than one language is involved, as in order for the data to be comparable and used in contrastive analyses, the two corpora need to be designed on the basis of the same criteria and following exactly the same internal distribution of instances of language (written, spoken, formal, colloquial, class / gender / age-representative, etc.). Although some general corpora have been designed in order to mirror other existing collection of data<sup>13</sup> and are thus already used to draw comparisons between language systems and uses, scholars are aware that the correspondence between them can only be, at

---

<sup>13</sup> An example of this strategy is the Chinese National Corpus that has been designed using the same sampling criteria as the BNC and is currently used for contrastive studies between the two languages. (McEnery and Xiao 2008: 21)

best, asymmetrical, as similarity in design is not the same as perfect equivalence. The creation of a corpus, in fact, depends on too many variables to be as thoroughly controlled as to allow for faultless likeness between two collections of data gathered at different times and under distinct conditions. The perfect monolingual corpus, in any language, is still the object of a grandiose quest, yet the unyielding efforts of linguists have resulted into some landmark outputs that constitute unassailable certainties in today's scholarship. For this reason, it can be argued that, despite the prohibitive difficulties faced by whoever attempts the same endeavour in a bilingual or multilingual context, progressive success lies ahead, together with ever improving tools for research and increasing practical applicability of results.

The issue of correspondence could apparently be successfully addressed through the adoption of parallel corpora, that is, collections of texts in one language and their existing translations in one or more languages. However, compiling such a corpus outside the scope of a specific domain is virtually impossible, as translations exist only for specific texts and genres, which would be over-represented, at the expense of balance. Other possible issues raised by parallel corpora in linguistic analysis are constituted by the direction of the translation and the number of available translations in a given language for the same text (McEnery, Xiao 2008: 23-24). Many parallel corpora are in fact unidirectional, that is they include original texts in one language and their corresponding translations in the target language. This allows the observation of linguistic phenomena only in the direction of the translation – that is for example, we can use it to investigate how tense aspects of the verb are translated from L1 to L2 – but it does not give any insight into how the same language pattern is dealt with in the opposite direction. Moreover, several studies have demonstrated that “the special conditions of translation ... may be significantly different from ‘regular’ native-speaker production” (Alsina and De Cesaris 2002: 216) and that “translated language is translationese”. Since the “effect of source language on the translations is strong enough” to make the translation perceptibly different from a natural use of the target language, it is concluded that a “unidirectional parallel corpus is a poor basis for cross-linguistic contrast” (McEnery and Xiao 2008: 24). Significant mitigation of the impact of these issues would require the inclusion in the corpus of

a body of texts originally written in L2 and accompanied by their respective translations in L1. Such corpora are of course much more difficult to design, as they require a considerable amount of texts translated in both languages which might be not easy to locate, especially for certain language pairs, and yet “a well matched bidirectional parallel corpus can become the bridge that brings translation and contrastive studies together” (Ibidem).

The other question affecting the potential large use of parallel corpora in contrastive studies is connected, although quite indirectly, to the role of translators, who actually “create”, each of them individually, the texts in the corpus. A parallel corpus containing only one version of each text translated into the target language, in fact, will be deeply affected by the decisions made by the translators, what Malmkjar called “one individual’s introspection, albeit contextually and cotextually informed” (qtd. in McEnery and Xiao 2008: 23). Again, a practical solution to this problem can be envisaged in the inclusion in the corpus of as many translations of the same text as possible, so as to include the work of a number of different translators, but this obviously turns the compilation of a parallel corpus, especially one based on contemporary non-literary texts, into a really prohibitive task.

What for some authors is a vexing issue in corpus-based Translation Studies, that is, the crucial role played by the translators and their almost unquestionable authority over the texts they have produced, is for others a defining characteristic of the translated text which enhances its worth in linguistic and lexicographic studies, rather than undermining it. According to Wolfgang Teubert, “cross-linguistic lexicography in quest of meaning must pay close attention to the practice of translators. It is they who invent the translation equivalents for lexical expressions. For these translation equivalents are not discovered, they are *invented*” (2002: 191; emphasis added). Teubert underlines how the very possibility of mutual understanding across languages is generated and safeguarded by translators. The latter certainly play a creative role in producing what they – based on accumulated competence, professional training, and with the assistance of a variety of translation tools, from dictionaries to corpora – consider the L2 equivalent of a given text. Yet, such creativity should not be perceived as a hindrance to meaningful analysis by the linguist, not even within the context of bilingual contrastive lexicography, as the equivalence

posited by a translator is what allows communication and the very production of meaning in the target language. This view of translation as a deeply human activity clearly goes in opposite direction to another trend in corpus-based translation studies, that is, the attempt to improve MT applications to the point of achieving reliability in machine-produced translations. Some approaches to contrastive analysis do underline the possibilities opened up by corpora to produce specialized dictionaries and lists of equivalence in highly technical and specialized domains so as to ultimately make even MT and CAT effective and reliable. The lexicon and style of this type of texts are considered, in fact, as much less subject to the individual choices of authors and translators, and parallel corpora in this case are seen as a crucial resource both for specialized lexicography and especially for more practical uses in professional environments. It has been aptly argued, in fact, that “bilingual or multilingual corpora consisting of texts based either on translations produced by highly trained professionals or on comparable text production<sup>14</sup> thus play an essential role in ensuring that specialized dictionaries, glossaries and terminologies actually reflect the language used in the workplace” (Alsina and DeCesaris 2002: 215). Even in this case, though, where the express purpose is that of assisting professional translators with sophisticated bilingual tools, and training others in the same profession, the importance attached to the acts of interpretation by the translators whose work is included in the parallel corpus is paramount, as can be inferred from the passage quoted above. The concept of interpretation is here a crucial one. According to Teubert, in fact, translators

undertake to paraphrase a text in a different language so that the paraphrase will mean almost the same as the original text. In order to carry out their task, they have to understand the text. This means that they interpret the text. Text interpretation, however, is an action, not a process. Only human beings can do it. All computers can do is carry out processes. (191)<sup>15</sup>

---

<sup>14</sup> Comparable corpora in this context will be analysed below.

<sup>15</sup> Kraif says something very similar when he states: “Translation equivalence is above all a global property of the translation of a text. It is not a linguistic property, but a pragmatic one: the translation arrived at is a result of interpretative choices that are made in a specific situational context” (2002: 273).

Interpretation is then a distinctive mark of translation work, as the human beings involved in the process of translation, in order to locate the equivalent of a unit of meaning in the target language, need to have not just linguistic skills, but specific cultural competence allowing them to decode the message of the original text and encode it in a different language system. Translators in this sense are seen as authoritative sources of insights into two languages, their inner workings, their mutual stratified relationships (past translations, unidirectional or bidirectional cross-cultural knowledge and understanding, etc.), and thus as reliable allies in the enterprise of contrastive analysis and lexicography. As Altenberg and Granger put it, “when we use translation corpora<sup>16</sup>,... we can place more reliance on the translations found in the corpus. The corpus can be said to lend an element of empirical inter-subjectivity to the concept of equivalence, especially if the corpus represents a variety of translators” (Altenberg and Granger 2002: 17).

Corpus Linguistics, on the other hand, can be seen as a most precious ally for professional translators and translation trainees alike. Despite the antagonistic role the field is sometimes thought to play, especially with regard to those branches of research working toward the successful implementation of MT in specialized fields, Corpus Linguistics has unquestionably contributed to a new understanding of translation work and techniques as based on a more sophisticated competence than bilingual terminology and / or a generically holistic notion of culture. While preserving, in many cases, as we have seen, a kind of reverence for the mediating function of the translator, in fact, a major contribution of corpus theory to translation studies has been that of providing scientific descriptions of the fact that

the translation units, the text segments that are translated as a whole, are larger than the single word; they are phrases of two, three or many more words. The equivalents of these translation units do not have to be phrases of the same or a similar structure; a collocation can become a clause; a whole clause can be reduced to a single word; singulars become plurals and vice versa (Teubert 2002: 200).

---

<sup>16</sup> “Translation corpora” is the term these two authors use to refer to what others call parallel corpora.

The definition of such translation unit may vary according to the theoretical approach of the scholars who addressed it, yet the fact remains that corpora provided the empirical evidence to start theorizing a methodological approach to translation that fully accounts for the identification of meaning as being located at the intersection of lexicon with syntactic aspects of the language system. Parallel corpora are in this sense a precious resource for linguists, as they provide unique insights into language structures that professional translators have deemed to be equivalent. The theoretical take of Corpus Linguistics on this matters, on the other hand, can produce a paradigmatic shift in the elaboration of tools for translation assistance, such as dictionaries and software for CAT.

Kraif calls the bilingual pairs resulting from a parallel corpus “*textual segments that are translation equivalents*” (Kraif 2002: 273; emphasis in the original), and goes on to explain that “translation equivalence is not a linguistic property, but a pragmatic one: the translation arrived at is a result of interpretative choices that are made in a specific situational context” (ibidem). Tognini Bonelli and Manca call it “a specific function together with its formal realisations in L1” that the translator has to encode “into a chosen formal realisation [i.e. another set constituted by a function and its formal realisation] in the target language” (2004: 371).

Despite the different nuances in the definitions sketched above (which reflect the orientations of much current scholarship on the subject), what is clear is that by breaking down the border between syntax and lexicon (Teubert 2002: 212), Corpus Linguistics allows a complete rethinking of the concept of translation equivalence and calls for a comprehensive reorganization of translation and SLA tools, beginning with bilingual dictionaries that, though significant improvements have already been achieved, still lag behind the monumental advances that can be registered in the compilation of monolingual dictionaries, especially the ones addressing learners’ needs: “It is time to develop a new generation of dictionaries, dictionaries suitable for assisting translation not only into the translator’s native language but also into a foreign language, dictionaries that give their users the proper translation equivalent for each semantic unit they have to deal with” (Teubert 2002: 190). Teubert’s adamant stance as to the need for a radical upgrade of the tools for translation assistance highlights several important issues that are positioned at the intersection of contras-

tive Corpus Linguistics, translation studies, and translation practice. First of all, Teubert gets rid here of the quite entrenched belief that the highest standards in translation work can only be achieved when translating from what is the translator's L2 to his / her L1, the condition of native speaker being still seen as too precious a vantage point to be given up. Coherently with his view of the paramount value residing in translated texts as acts of creative realisation of equivalence between texts, Teubert seems to be positing here source and target text on exactly the same plan of linguistic prestige and the rightly assisted and trained translator as the ideal mediator in both directions. Secondly, he highlights how the focus of bilingual dictionaries has to finally shift from more or less decontextualized and isolated words to semantic units that are identified by "combining the parameters of recurrence, statistical significance and syntactic categorisation" in the available corpora (212). In order to achieve this goal, it is not enough for the bilingual lexicographer to validate translation candidates on the basis of evidence found in target language monolingual corpora. Although the equivalent thus identified is in many cases correct and acceptable, its drawback, according to Teubert, lies just in the fact that it does not reflect translation practices. While the latter are considered a potentially unreliable source of linguistic data as they might be heavily tainted with translationese, Teubert sees the resulting parallel corpora as precious tools since they carry within them the informed segmentation of the source text into the semantic conglomerates (and then translation units) identified by translators. Such segmentation is the crucial act of interpretation of the source text that allows the translator to establish a translation equivalence, that is, "a relationship between messages entrenched in two given contexts and backgrounds: the source and the target context", a global equivalence that clearly "does not imply equivalence at the level of linguistic units" (Kraif 2002: 273).

In order to be effectively turned into a tool for enhanced translation work, parallel corpora need to be aligned, that is, they need to be processed by software that identifies correspondences between the source text and the target texts and thus allows to examine the two sub-corpora concurrently. As units of meaning may vary in the way they are linguistically encoded in different languages, "bilingual alignment is not a negligible problem" (Kraif

2002: 275), as has been repeatedly noted:

While it is fairly straightforward to align entire texts, aligning paragraphs is only trivial when paragraph structure is preserved from source text to translation. Sentence alignment further complicates the issue, because translators can (and often do) join sentences together, split a sentence into two or more smaller sentences, delete entire sentences, reorder them or even add new sentences of their own, which were not present in the source text (Frankenberg-Garcia 2009: 61).

Existing alignment programs can only partially execute a satisfying alignment when the textual segment in the source text and in the translation are considerably asymmetric, so that the result often needs to be manually revised. Yet, it is alignment that turns parallel corpora into “repositories of translation units and their equivalents in the target language” so that they “can be processed and re-used in subsequent translations” (Teubert 2002: 193).

It is true that several strategies are devised or sought for in order to introduce some kind of counterweights to the choices of individual translators, especially in the form of comparable corpora to be analysed alongside the parallel one. Teubert himself, who, as we have seen, is a staunch advocate of the positive role played by translators, states that “ideally, parallel corpora should be viewed as complementary to comparable corpora” (qtd. in Alsina and DeCesaris 2002: 216), the latter being the repository of instances of use in the target language outside the space of mediation and interpretation constituted by the translating activity. The combination of a parallel corpus and a comparable one is unanimously considered by linguists as an ideal condition for any kind of contrastive and translational investigation (McEnery and Xiao 2008; Corpas Pastor and Seghiri 2009; Zanettin 2002; Teubert 2002; Altenberg and Granger 2002). Both linguistic theory and translation training and practice are in fact significantly enhanced when an aptly designed comparable corpus is compiled to complement the parallel one. Yet, even when the equivalences produced (or “invented”, to quote Teubert again) by translators are no longer the sole corpus-based evidence for linguistic correspondence, translation as a competence and as a more or less implicit activity emerges as a function of the linguist. A fair degree of reliance on the work of past translators, and on translation competence more generally, is showed for example by Tognini Bonelli in



her work on corpus-driven elicitation of units of meaning in comparable corpora:

the initial hypothesis positing one or more tentative matches between two or more *prima facie* units of meaning in SL [Source Language] and TL [Target Language] has to rely on the translator's intuition or past experience. Traditionally, standard reference works such as bilingual dictionaries attempted to provide this information. Recently we have witnessed the emergence of *translation corpora* (also referred to as *parallel corpora*)... I maintain that the use of a translation corpus at this stage, if available, will give us the benefit of such input in a more reliable manner and provide us with a range of possible translation pairs that have already been identified and used by translators, in other words verified by actual translation usage (2002: 81).

So actual translation usage is a benchmark of acceptable linguistic encoding and is the best starting point for the researcher / translator to begin checking the degree of correctness and naturalness of the chosen unit in the target language. It is interesting to notice how Tognini Bonelli, who is among the strongest supporters of what she herself has labelled "corpus-driven approach", eventually relies on what she consistently calls "*prima facie*" translations that can only emerge from the researcher / translator's background and initially have nothing to do with the data collected by the corpora under use. The interrogation of the corpus is then achieved only following a preparatory phase in which a decisive role is played by the bilingual competence of the subjects running the query, or of the translators who preceded them in analysing that portion of text and establishing a relationship of equivalence between this and another piece of text, real or hypothetical, in the target language. In other words, even a corpus-driven approach (which, as we have seen, maintains that the corpus itself is to be empirically used as the only source of theoretical understanding of the language) cannot in a multilingual context do without a pre-existing competence which, however, is again of paramount consequence in the final stage of the research process: "Any judgement about cross-linguistic equivalence (or similarity) must be based on the researcher's 'translation competence'. This is true at both ends of the analysis: initially, when items are selected for comparison, and finally, when the results of the comparison are evaluated," as in the end "the notion of equivalence is a matter of

judgement, reflecting either the researcher's or the translator's bilingual competence" (Altenberg and Granger 2002: 17, 18).

Despite the different emphasis on the role of the translator, scholars agree on the fact that Corpus Linguistics produced the final demise of any residual notion of equivalence based on the level of the word and also moved beyond the quite generic idea of context, toward an attempt at defining the basic unit of translation in truly linguistic terms.

## 2.

# ARCHAEOLOGY IN TRANSLATION

### 2.1. PREAMBLE

This chapter delineates the methodology and rationale behind the compilation of the corpora upon which the observations reported in the rest of this book are based. A few things need to be preliminarily pointed out, so as to clarify the attainable objectives and the scope of the present study, as well as to envision the trajectories for future developments of this research.

First of all, the corpora reflect an extremely empirical approach to corpus work, from compilation to use, as they are based on my first hand experience as a translator. For ten years, beginning in 2005, I have provided translations of academic texts from Italian into English for a group of scholars specializing in Islamic archaeology. The kind of translations I was called to execute varied greatly in terms of text genre (from catalogue entries to official reports, from academic essays to conference talks, from book reviews to paper abstracts), length (from a few lines to entire book chapters), and authorship (from affirmed scholars at the top of their university career to young doctoral candidates on the verge of publishing their first article). This gave me access to a wealth of linguistic data, in the form of specialized vocabulary, research domains (ranging from manuscripts iconography to architectural epigraphy, from pottery to metalworks), writing styles and academic discourse praxes. In this sense, these texts constitute, on the one hand, a very limited resource compared to the technological tools that are available today to corpus linguists. On the other hand, they are also an extremely homogeneous collection of data in a hyper-specialized field such as Islamic archaeology, something quite rare and for this reason of potential interest even from a theoretical point of view.

One point that needs to be clearly stated here is that only a very limited number of the texts that were the object of translation have been included in the first phase of this study that is presented here, and this is due to several reasons. To begin with, not all the translations that I executed were eventually published, as some of them were meant to be orally delivered at conferences or were intended for limited circulation in administrative contexts. This introduces a set of considerable differences among the texts at my disposal for the compilation of the corpus, due to the inherent dissimilarities in style of written and spoken language, the variation of register according to contexts, but also, and more crucially, because translations submitted for publication underwent the customary process of review and approval by editors and thus constitute, in the view of this writer, a much more reliable source of acceptable forms in the target language. Another safeguard of reliability, with specific regard to the specialized lexicon of the target scholarly community, is constituted by the fact that each translation was reviewed by the respective authors before submission to journals or book editors. This was a crucial passage that completed each work cycle, and can thus be regarded as a peculiar and essential aspect of the translation context in which the texts comprised in this corpus were produced. Being academic experts with an international scholarly profile in their own fields, the writers were in an ideal position to assess the correctness and / or degree of precision and acceptability of the choices that were made with regard to the highly specific vocabulary deployed in the description of findings.

Another criterion guiding the selection of data was the amount of human editing required by the texts before being processed by the chosen software. The present book is conceived, in fact, as a preliminary report on an ongoing research, a moment in which the outcomes of a much larger investigation are gathered for analysis and assessment before moving on to another phase of the project that will seek to broaden its scope and deepen its focus. This corpus does not claim to be satisfactorily balanced, nor ample enough to constitute conclusive evidence of the phenomena that will be described here, or to draw final theoretical conclusions as to the questions it raises. Rather, it is offered as a reflection on the early stages of the research aimed at refining its approach, while illuminating, from the vantage point of actual use in

professional environments, some current debates within Corpus Linguistics.

As I will illustrate in the course of the following pages, several aspects of said debates which span over the past few decades play a role in the constitution of these corpora and in the way they are used here and can be expanded in the future. Chief among these are the intersections between Corpus Linguistics and Translation Studies, the status of disposable and DIY corpora within the discipline and their theoretical soundness, the systematic and empirical use of the Internet for documentation and corpus creation, and the use of corpora for the extraction of specialized terminology and the compilation of glossaries. As to the latter point, the last part of this book presents a bilingual glossary of technical lexis in the domain of Islamic archaeology entirely compiled through corpus analysis.

As I hope to have made clear, this is still a work in progress, an attempt to move towards a better understanding of the overlapping methodological and theoretical insights Translation Studies shares with Corpus Linguistics. The wider scope of the research is the eventual compilation of a balanced and coherent corpus of Islamic archaeology to be used for the extraction of a much larger bilingual glossary alongside other units of meaning in the two languages involved, to be put to use in specialized translation as well as in translators' training. Such a glossary would also prove helpful to archaeologists who are not conversant in English, whereas the corpus from which it will be extracted would also benefit translators who specialize in this and in cognate fields and might use it as a source to build their own disposable corpora.

## 2.2. LANGUAGE AND ARCHAEOLOGY <sup>1</sup>

With globalization first looming at the horizon, then heavily impacting many aspects of human experience and knowledge, for quite a long time there has been a sort of generalized consensus

---

<sup>1</sup> This section and the following one expand and re-elaborate my contribution in the proceedings of the conference *Translating East and West* ("l'Orientale" University of Naples, November 2012). I take this chance to thank the conference organizer and volume editor Oriana Palusci for encouraging my work on this topic.

expressed from many quarters about the manifold risks of such a trend also in relation to translation, the circulation of information, the status of languages vis-à-vis more prestigious varieties, and ultimately cross-linguistic communication of all kinds. Yet, despite the understandable worries about the fact that English would have cannibalized most languages and erased them from international currency, taking their place in all sorts of contexts, there is ample evidence that internationalization and globalization have actually brought with them, at least so far, an increased demand for translations. Teubert points out how this is true at the present moment both in relation to the private contexts in which the expected spread of foreign languages (and English above all) did not happen, but also with regard to the public life of new political bodies such as the European Union. As to the former, he says that “while, in their professional lives, more and more people are learning to function in a bilingual or multilingual environment, it seems that, apart from a traditionally small polyglot elite, in their private lives they tend to cling to the language they grew up with” to the point that “globalisation of the media has opened up a new market for instant translation” (Teubert 2002: 189). When it comes to public life, “the necessity to translate agreements, contracts and all other documents that could have a legal impact (such as product descriptions and user instructions) into other languages” constitutes nothing less than a pressing urgency:

Whatever their source language (increasingly English, as we are all aware), these texts have to be localised, translated into the language(s) of the country whose jurisdiction is involved. As long as legal systems are not globalised, courts will accept documents as evidence only if they exist in the official language(s) of the country in which the court is situated (Teubert 2002: 189-190).

As can be seen, there is little doubt as to the role played by English in this increasingly globalised scenario, and it is equally evident how the other side of this phenomenon is what could be called a form of ‘compulsory internationalization’ through the use of English itself. A growing number of translations into and from English are required, in no time, from highly trained professionals. This tendency toward monolingual prestige pushes for a wider acceptance of English as the global interlingua in many arenas of international exchange, from economics to science to education,

to the point that being conversant in English is perceived as *the* essential asset to possess in order not to lag behind in the globalizing trend. Yet, it is quite interesting to notice, following Teubert in the above quoted passage, that a specialized professional domain in which the overcoming of geopolitical borders as we used to understand them has not resulted into the demise of local languages is just that of supranational laws and regulations. A specific case in point is represented by the European Parliament, where “less than 10% of [the] budget is spent on parliamentary work proper, while more than 90% is spent on interpretation and translation” (Teubert 2002: 190). As pointed out by Murphy, however, despite the EU obligation to produce all official regulations in a considerable number of languages, “the prevalent use of English by those working within the EU institutions is a clear sign that, regardless of legislation, practical concerns have the strongest influence on daily lives and working habits, and it is a fact that English serves as the *lingua franca* of the 21st century, particularly within Europe” (2012: 21). A good illustration of this sort of linguistic double standard is provided by the attitude of Finnish officials as reported by Koskinen, who underlines how they tend to operate within the European Commission using only English and overlooking altogether the subsequent translation into Finnish of official documents. The latter is disregarded not only because it requires a reasonable time to be produced, as opposed to the immediacy of working directly in English, but also because it is perceived as less reliable even from a juridical point of view than the ‘original’ version, which in the majority of cases was certainly not drafted in Finnish (Koskinen 2000).

Should the European Union insistence on the translation of legally binding documents in the local languages of member states be interpreted only as a function of its long-standing investment in multilingualism? (Grzega 2005) While there is no doubt that “multilingualism is seen as the strong point of the European Union, the emblem of its linguistic and cultural diversity” (Murphy 2012: 19), it is also true that this complex socio-political environment demanding “the production of translation of texts not to be read but to be there in case a need for them arises” (Teubert 2002: 190) brings to the fore the deep-seated cultural and psychic investment in mother tongues. Only the languages of long-held national identifications seem to bear within themselves the possi-

bility, the juridical capacity we might say, to express – and by expressing realise in the form of legally binding speech acts – the shift toward new forms of citizenship and community belonging.

The issue is considerably wider than the contingent problems raised by the still uneven knowledge of English as the dominant working language among political representatives, jurists, and the general population, and the need for accurate translations in the many languages of the European member states goes beyond the practicalities of effective communication between EU institutions and its citizenship (Gazzola 2006).

While English as a *lingua franca* and national languages actively compete for prestige and visibility in the arena of supranational law making processes – the former pushed to the fore by practical concerns, the latter driven to counterweight it with an awareness of the centrality of language to the sense of identity – it is commonly believed that academic discourse has practically, if not in principle, yielded to the predominance of English, which has become the working language of the greatest majority of scientific initiatives claiming any degree of internationality in scope. Scholars across a variety of disciplines are not only expected to know English to take part in conferences, but are also pushed to disseminate the results of their research in this language. This leads to what many consider (as we will see) the inevitable impoverishment of scientific production in languages other than English and the consequent limitation of the number of subjects who have actual access to knowledge. While internationalisation, in the form of fast circulation of research outcomes and researchers' mobility, is unanimously regarded as of utmost importance for advance in any field of knowledge, concerns are expressed by scholars within several academic subject formations as to the impact that the shift toward English monolingualism would produce in the disciplines themselves.

The frame of this debate also applies to archaeology, which as a field of study immersed in discourses about heritage, the past and its preservation, and narratives of belonging, cannot but be extremely interested in its terms and potential outcomes. According to several scholars working within the disciplinary boundaries of both archaeology and linguistics, archaeology itself is a narrative, even before excavations take place and findings are retrieved. Among these, Rosemary Joyce (2002), for example, has investi-



gated the intersections between the discourse produced by archaeology and narrative forms, concluding that

all archaeological discourse, regardless of its format and audience, is dialogic. The formation of marked genres – including site reports and more popular media, such as museum exhibits – are formalizations of specific dialogues, amenable to analysis as genres. Archaeology is a textual practice from the field through the lab and into all forms of dissemination (2002: 2).

Two points deserve closer attention here: first of all, the emphasis on the textual practices as the real essence of archaeology, which is particularly interesting since it shifts the perception of the field from the focus on tangible objects (findings) to the narratives through which they are interpreted and presented to the public. The second is the dialogic nature of this presentation, that is, the field produces bidirectional conversations with various interlocutors, deeply influencing the latter's understanding of the past, but also registering in a number of ways the very presence of other subjects in the exchange. This is of extreme consequence when the movement towards English, and the resistance it elicits, are considered alongside the role played by translators, who become cultural mediators between different 'narratives' of the same 'story'. The element of creativity stressed by Teubert thus becomes unmistakable. When archaeological discourse is understood as, among other things, a narrative form, the importance of national languages in the formation of the field emerges as an urgent issue in the age of the internationalization of academic research. The use of a common lingua franca among specialists, in fact, produces a narrative of knowledge that marks a shift from previous discourses in the same field, which were often rooted in the notion of archaeology as a nationalistic cultural project. In this sense, archaeological narrative genres, from excavation reports to museum exhibitions, contribute to a story both of the nation producing them and of its relation to its own past and that of other cultures. Even when, as is often the case with archaeological missions and the resulting scholarship, the focus is on geopolitical and cultural areas far removed from the one in which the researchers are professionally based, the narratives will also reflect views and needs located in the observing subject, besides dealing with the object of scrutiny. When this dialogic exchange becomes

monolingual, the inevitable outcome, according to some scholars, is the marginalization of the national culture due to the very fact that the scholarship produced in a country has to be circulated into a foreign language. This, in turn, would result into an impoverishment of national languages. As Estonian archaeologist Valter Lang puts it: “If we do not write about the results of our studies in the language we speak, its scientific terminology will inevitably degenerate, and before long, we will be simply unable to think scientifically in our mother tongue” (2000: 107-8)<sup>2</sup>. Although the deterioration of the scientific register of a language can occur in any field, Lang expresses his worried views with specific reference to research conducted by native speakers of the language adopted in the area that is simultaneously the object of archaeological study who publish almost exclusively in English, the latter being perceived as the dominant language of scientific communication, rather than as a mere *interlingua* or *lingua franca*. Other archaeologists have given voice to a less dramatic interpretation of this trend, seeing it as an opportunity for many scholars to practice their multilingualism and adjust each time to the specific publication venue, while simultaneously cultivating their mother tongue as the privileged medium of critical inquiry: “In the human sciences, the desperate following of so-called trends adopted from outside can sometimes only be done at a very high cost. On no account should we abandon our most significant tool for analysing reality – the rich and flexible symbol system which is our mother tongue” (Suhonen 2000: 121). Moreover, in response to Lang’s article, Bryan Boyd highlights how his position is predicated on the almost linear correspondence between the nationality of the researcher, their mother tongue, and their object of study as the expression of the same area, when not of the same nation. As he demonstrates through a humorous account of his own background<sup>3</sup>, Boyd effectively maintains that this case is actually

---

<sup>2</sup> Lang’s article is followed in the journal issue by two responses and is the leading contribution to a forum on the question of languages and archaeology.

<sup>3</sup> “I am myself Scottish but work in Wales, both countries’ languages having a dominant / subordinate relationship with English. My archaeological research normally takes place in Israel and the Palestinian Territories, which have political, historical and linguistic relationships not only with each other, but also with the languages of European, particularly British, colonialism. Finally, at the time of writing, I am working in Madagascar, the scientific communities of which are involved in de-

rather the exception than the rule in academia today, and in a field such as archaeology in particular.

Within this broad discourse of the intersections between archaeology and language, a specific case in point that, being notable for its complexity, can frame Boyd's argument in more general terms, is constituted exactly by the sub-field of Islamic archaeology. The latter in fact, identifies as its object of study a large array of cultural manifestations, from different epochs, produced in a wide territory extending from Europe to North Africa to the Near East, across many geopolitical borders and heterogeneous linguistic areas. Moreover, the scholarship on these varied cultural formations has been developed in the past centuries mainly by western scholars, as an effect of colonization, imperialism, and, later on, agreements between countries and research institutes leading to the excavation of eastern territories by western archaeological missions. The resulting studies were circulated in the languages bearing cultural and / or academic prestige at each given epoch, which basically means, until recently, French, German and English <sup>4</sup>.

This complex picture is of great relevance to delineate the academic, cultural, and political (in the broad sense) milieu in which the translator of this branch of scholarship operates. Though excellent scholarship in Islamic archaeology has been developed by researchers from different countries in their national languages, and knowledge of one of these is the very reason why certain individual translators are hired, what really enables the latter to carry out the task is their acquaintance also with the languages through which the discipline has been developed at international level. As cultural competence is indispensable in order to produce a translated text reflecting an adequate command of collocations and specialized terminology in the target language (Austermühl 2012: 71), translators can only proceed by acquiring said competence through access to relevant publications in the field and other domain related tools (such as monolingual or bilingual

---

bates surrounding the linguistic domination of the French colonial legacy. And, of course, I am writing in English for a Finnish journal!" (2000: 116)

<sup>4</sup> Excellent studies have been published in Italian by outstanding archaeologists who have led excavation missions in several countries for decades. Yet, their research has sometimes had limited circulation outside Italy also because of the barrier constituted by language.

glossaries). However informal and opportunistic the selection of these aids, the comprised texts constitute a corpus in the broadest (i.e. non scientific from the point of view of Corpus Linguistics) sense of the word, an often scarcely organized virtual or actual reference shelf translators consult to check the currency of a collocation or the degree of acceptability of a term. The number and type of language systems they are able to decipher in written form thus becomes an asset of the utmost importance. Trying to build up such a corpus to be used for present and future reference, they might in fact find themselves in the presence of texts written in languages they ignore, texts that are mediating discourses themselves, but which are the only widely circulating narratives in the field at the time when the results of excavations were published. This situation delineates a sort of paradox in which the languages that might be paramount in helping the translator gain the needed expertise in order to correctly translate highly technical texts are those through which the academic field has been constituted, rather than the language(s) spoken in the areas that were or are the cradle of the culture being translated <sup>5</sup>.

However, in the recent past, there has been a decisive shift from these once prestigious languages to the almost exclusive use of English for international conversation among Islamic archaeologists as well, as an effect of a more general pressure toward the adoption of English as the *lingua franca* of scientific communication. Since several specialists in this field started their academic career when other languages were the international standard for publications, not all of them are today conversant with English and thus the work of translators becomes paramount in contributing to the internationalization of the exchange. This is coherent with the trend observed in several studies and surveys (Cronin 2010; Beninatto and DePalma 2008; Van der Meer 2003) showing translation as a defining element and a central actor in globalization processes.

Because of this gradual movement toward English monolingualism, a generational gap can be noticed between senior scholars, who were mainly trained in their national tongue and other

---

<sup>5</sup> In other words, knowing Hindi or Arabic might be of no help to a translator if the publications in a field connected to the history and culture of India or the Arabic world were mainly written in other languages.

prestigious academic languages of the time, such as French, and their younger colleagues, who, being fully aware of its importance in our globalizing academic job market, often acquire excellent English alongside the core expertise of their elected field of research. This gap extends well beyond linguistic matters since, with regard to the type of observations archaeologists make on Islamic artifacts and manufacts, the generation trained directly in English tends to reproduce the discursive strategies and the interpretive paradigms elaborated in the Anglophone world, thus giving rise, as a matter of fact, to an at least partially different narrative for the same archaeological context than the one elaborated by their senior colleagues. The archaeologists of the latest generation are, in this sense, self-translated scholars who mediate between classes and textbooks in their national language and the exposure to a body of knowledge circulating internationally in English.

Within the context of the growing language expertise of these younger scholars, however, the mediation of professional translators can still be required for a number of reasons. Even when they master the technical jargon in their field, authors often lack, in fact, a sufficient degree of confidence in the use of the foreign language when it comes to the discursive and conceptual aspect of their writing; in other cases the services of a translator are required in order to re-publish abroad a text that is already circulating within a national context and whose rewriting by the author would be time consuming and still require a final revision by a linguistic expert. In conclusion, a translator who, thanks to his / her “context-adequate terminological and phraseological competence” (Austermühl 2012: 59) has become part of the professional expert community that is conversant in a specialized technical field (Király 2000) contributes to linguistically effective and time efficient writing and publishing strategies.

### 2.3. TRANSLATION, CORPORA AND THE WEB

Given the complex cultural framework constituted by the intersection of Islamic archaeology as a composite field of study and the narrative dimension of archaeology as a scientific discipline which I have sketched above, it is legitimate to ask what characteristics a translator should have to be able to successfully mediate

between and among such sophisticatedly layered contexts. Relevant questions here are the quantity and quality of expertise in Islamic culture that the translator should develop, the general acquaintance with archaeology that is expected of her, and the familiarity that might or might not be required with Arabic as the most prestigious language across the Islamic cultures due to the centrality of the Koran. In other words, in addition to being a cultural mediator, this highly specialized translator needs to get prepared to act as a mediating agent between specifically academic paradigms of knowledge.

The other crucial question is more directly connected to the object of this study and has to do with the ways in which Corpus Linguistics can have a positive spill-over effect on building the hyper-specialized expertise professional translators need in the age of compulsory academic internationalization. This impact is registered first and foremost in the growing popularity among teachers, learners, translators, and academics in general of corpora-based dictionaries designed to meet the needs of specific communities of speakers and writers. It is quite obvious, in fact, that, no matter how large a general corpus is and how sophisticated the resulting dictionaries are, “even the 100 million words BNC is ill-equipped to meet the needs of translators working with very specialised texts and confronted with specific terminology” (Zanettin 2002: 241), as such terminology would likely be obscure to untrained people even in one’s own mother tongue (Austermühl 2012: 71). However, although much has been written (as we have seen) about the boost that would come to this type of translation work from the availability of corpora combining parallel and comparable texts, the domain is in this case so specific that the lack (so far) of dedicated tools comes as no surprise.

Translators today often find a remedy (however partial or imperfect) to this need for glossaries of technical terms and their collocation by using the Internet to run searches of texts featuring the needed structures and thus build their own task-related corpora<sup>6</sup>. The practice finds its rationale, with reference to the case

---

<sup>6</sup> There is a growing literature on the issue of Internet-based corpora and their characteristics. A significant body of scholarship has been developed over the past few years on this topic in Spain (Corpas Pastor 2001 and 2004; Sánchez-Gijón 2003, among others). According to the aspect that is emphasised, these corpora can be

in point of Islamic archaeology, in the large availability of academic literature online, in the form of websites, web-based scientific journals and other documents (such as catalogues, teaching materials, PhD dissertations, etc.) that a growing number of Anglophone universities, museums, foundations, and research institutes provide access to.

The access granted by the Internet to virtually unlimited archives of texts in a number of languages has made the relevance of solid methodology in data retrieval (or documentation, as it is usually called) only too evident (Austermühl 2012; Lüdeling et al. 2007; Corpas Pastor and Seghiri 2009) and the acknowledgment of the virtuous uses of DIY and disposable corpora unavoidable despite the resistance of many scholars in the field of Corpus Linguistics who advocate a more theoretically sound approach to the definition, compilation and applied uses of corpora. This view is certainly at odds with the opportunistic use of often unbalanced, under-theorized, non representative, fluctuating collections of linguistic data in translation work, but there is no doubt that translators working within any specific field are in fact mediating not between two languages, but rather between two loosely bounded corpora comprising the instances of specialized language they have (had) access to. They develop in turn (and store in their computers) their own vast repertoire of words, grammar structures and discourse practices in the languages they use, which can be rightly considered highly specialized, though empirical and non-finite, corpora. The Internet, in other words, plays a crucial role today in providing an array of resources that was simply unthinkable of just a few years ago.

Yet, online sources are not just numerous, they are also spurious, uncontrolled, a corpus (or a growing body of corpora) which is never final and is not systematized. It has been argued that the Internet cannot be considered as a corpus due to its lack of representativeness and its virtually non-assessable size, both representativeness and assessable size being defining characteristics of corpora for many scholars in the field of Corpus Linguistics (McEnery & Wilson 1996). On the other hand, a corpus can also be intended in broader terms as “a collection of texts when con-

---

called *ad hoc* or *domain-specific* (Corpas Pastor 2001 and 2004), *disposable* and / or *DIY* (Zanettin 2002), *special purposes* (Fletcher 2004) or *electronic* (Varantola 1997).

sidered as an object of language or literary studies” (Kilgariff & Grefenstette 2003: 334). As for representativeness, it has been noticed that any corpus (just like the web) only represents itself (Ibid: 343). Moreover, highly specialized corpora do not need to be representative of the language as a whole and thus a very limited set of texts can be sufficient to develop effective language use on the part of the translators (Scott 2012: 8).

In the following section, I will illustrate some documentation techniques that make use of the most popular commercial search engine, i.e. Google, and a “Web as corpus” approach whose outcomes can eventually lead to the compilation of a comparable corpus of specialized texts (in this case within the field of Islamic archaeology) and thus turn into a “Web for corpus” strategy<sup>7</sup>. This work opens up a space for further cultural analysis of the connection between highly specialized fields of knowledge and the language(s) in which scientific research is carried out and circulated. Given the current prominence of English as the international language of research, an investigation in both a diachronic and synchronic perspective of the role played by language and translation in the definition and delimitation of specific academic fields would be certainly fruitful both theoretically and pragmatically. In other words, the IT-EN comparable corpora eventually produced by the mapping of scientific discourse I will just start outlining here would not only help translators localize into English the knowledge elaborated by Islamic archaeologists in other languages, but would also contribute to tracing the epistemological project of academic disciplines as evidenced in the corpus through which new scholarship is circulated within the field.

---

<sup>7</sup> The procedure I will present here is not much different from the one outlined by Austerlühl (2012), though I was not aware yet of his article on the subject when I started working on this topic in order to present at a conference in November 2012.



## 2.4. CAN YOU SEE THE FOOT UNDER THAT BOWL? A CASE STUDY IN WEB-BASED DOCUMENTATION STRATEGIES

One of the most common methods used by translators to check the appropriateness of specific words and their collocation is constituted by search engines (Google above all others) that allow looking for any word, expression, and structure within the immense “anarchical” collection of texts constituted by all accessible websites and online documents. Obviously, though a high number of occurrences suggests a higher frequency in the language as a whole of the fragment under scrutiny, mere statistics is all but adequate in order to determine whether or not a certain expression is the most appropriate available option in the target language. First of all, such web-based searches only allow estimating the frequency of a certain language item in the documents that can be retrieved by a search engine like Google, thus amounting to a rough approximation that cannot be assumed to proportionally represent the frequency of the word, neither in the natural language, nor in the sub-category constituted by the special lexicon the translator is working on. Indeed, arguably the majority of academic studies in a field as vast as that of archaeology are just unavailable in digital format and can only be accessed through libraries, often located in geographical areas the translator has no chance to visit. Moreover, given the extremely uneven quality of the sources and data accessible online (from academic publications to unedited and inaccurate websites produced by users who might even be very weak writers of English), translators need to exert extreme caution in using the results obtained from such searches. And yet, as we have seen, a good number of reliable, though often secondary, online sources allows today the bridging of the geographical gap between the translator and the special lexicon dispersed in so many otherwise inaccessible documents. In order to take advantage of the wealth of online information and minimize its risks, the first step to take is the filtering of results, an operation that can actually be assimilated to a narrowing down of a general query to a more coherent, and thus more reliable corpus. It is exactly through these processes of selection and filtering that the asymmetry between the Italian and the English lexicon of Islamic archaeology fully emerges and allows the appreciation of the epistemological shifts occurring in the academic discipline also

as an effect of the spread of English as international scientific language.

I will try to illustrate my points by way of three examples from some recent translations I have carried out for a bilingual Italian-English museum catalogue. These can be considered as case studies in how using Google as a “quick ‘n dirty corpus tool” (Robb 2003) in order to build a disposable corpus can actually be an empowering move on the translators’ part, while also demonstrating how their off-line cultural capital and professional networking can play a role in defining their personal corpora and, in turn, shaping the epistemological project of the academic fields in which they act. The analysis of specific problems raised by the translation of some expressions within a text such as the catalogue of a museum, which is simultaneously a linguistically sophisticated text and one with a significant circulation outside academia, allows in fact the observation of the way in which translation deploys its theoretical potential through the practice of translation itself, thus resulting into a theoretically self-producing practice. In terms of methodology, the following examples are based on a DIY corpus I built in 2010-2011 using Google and, as will be seen, combining general web searches with domain-specific ones. Since then, much has changed also in terms of online resources available to translators who wish to use the Internet, and more specifically Google, as a corpus tool. In 2010, in fact, Google released its own N-grams corpus, followed at the end of that same year by the corpus of American and British English based on the texts digitalized in Google Books. In 2011, linguist Mark Davies, based at Brigham Young University, released another corpus, based on the same collection of texts, but with additional search functions of crucial importance to linguistics and lexicography. Those tools were still in their early stages of public availability and were not known to me when I started collecting data from my own work as a translator for this study, but will certainly be included in the future stages of my research. Since the body of data available on and through the internet constantly changes, the type of documentation that I gathered in 2010-11 might well be no longer available online. The data I concomitantly downloaded and stored as a DIY, semi-disposable corpus, on the other hand, can be just as unreliable as a volatile collection of web searches, unless they are regularly integrated by new addition and constantly monitored to

decide which texts should be discarded as no longer representative. As a consequence, the examples I will discuss here are to be intended as an illustration of a bootstrap methodology for documentation and are not meant to be considered as representative of the kind of responses a search engine would give today to the same queries. However, as will be seen, they have also been tested against the comparable corpus compiled in the course of this study, called CompArIIEEL, whose characteristics are detailed in the following chapter.

My three examples are singled out from the translating process of a description of Islamic pottery in Italian to be put into English. First of all, let me point out that in all the cases I examine here each single word in Italian has an equivalent in English, the latter being also attested within the discipline. Thus, the asymmetry that will be observed between the corpora is produced in these cases by the fact that one language selects in its descriptions only some characteristics of the object under scrutiny, whereas the other accumulates more details. The asymmetry then is the result of the use of existing words within the discipline as a national field of study, rather than an inescapable issue due to the lack of specific words or to a radical difference in the ways in which the two languages express the same concepts and ideas. Let me start with my first example.

#### 2.4.1. *Compromising between synthesis and explicitation*

The Italian catalogue entry is titled: “Ceramica con rivestimento vetroso al piombo (ceramica invetriata)”. The descriptive and discursive modality adopted by the Italian author had no perfectly symmetrical correspondent in the English specialized lexicon, which to categorize the same class of findings uses the much more condensed label “glazed wares”. The English, in other words, produces as its default option the type of synthesis the Italian author puts into brackets (“ceramica invetriata”), thus constituting the text as immediately recognizable and understandable mainly to specialists, whereas the Italian heading tries somehow to explain the process of production of the artifacts, which can ideally be grasped also by non specialists.

The perfectly symmetrical English translation “Pottery with lead based glaze” is made up of forms widely used in the tech-

nolct pertaining to Islamic pottery, such as “lead based” and “glaze”, but if we try to verify the suitability of such a formula by looking it up on the internet, we find that the results obtained through a Google search are for the greatest majority in the domain of medicine and the definitions we are verifying are common mainly in publications relating to public health and the current ban on lead-based colours in the manufacturing of kitchen wares. Even the related “lead-glazed wares” is not as widely attested in the domain we are researching as to encourage an immediate decision in favour of this candidate. We only find such formula in three academic or museum based descriptions available online, which tells us that “pottery production”, or “characteristics of decorated pottery” are far too generic corpora. The additional and synthetic description given by the Italian text as if it were almost a case of intralinguistic translation (“ceramica invetriata”), shows how this language as well is capable of synthesis and of using a technical jargon. The English equivalent of this expression is “glazed pottery” or “glazed wares”, and they are widely adopted, as shown in the thousands of results given by Google, even when the search is limited to academic publications.

This preference for very synthetic formulation of the category can be a function both of a characteristic of the English language as a system, which tends towards synthesis, and of cultural prestige: synthesis, in other words, could somehow be predicated also upon the established outstanding tradition in academic scholarship that today accompanies scientific discourse in English with the mark of international excellence. The brief English description produces the impression of a sufficient clarity by removing information, in comparison with the Italian one, as if there were no need to further explain what is self-evident in its internationally understandable English formulation. Supposed transparency is thus constructed as a function of the dominant position occupied by the Anglophone academia, whereas the Italian source text reveals a desire on the author’s part to explain production processes and materials by accumulating details in the heading.

The other key concept to include in the label is that of the use of lead in the process of vitrification of the wares, so my widely attested form “glazed wares” was modified into “lead-glazed wares”. By empirically verifying the spread of such formula, I found more than 25,000 documents and websites, mostly

concerning ancient artifacts, which confirms this as the right form instead of the numerically more conspicuous “lead-based glaze” (55,000 documents). The Italian author’s attempt to distinguish between and keep together in the same text a more discursive description and a highly technical jargon (thus exemplifying the double function of the museum catalogue) finds a limit in the virtually non-existent circulation of the former in English. In order to avoid the risk of unnaturalness, I decided to use in English only the synthetic diction “Lead glazed wares”. However, this translation was a compromise between the author’s desire to retain in the classification of the wares a reference to the element used to vitrify them and what seemed to be in most contexts the default name of these artifacts in English, i.e. “glazed wares”. The comparable corpus CompArIIEl confirms the higher popularity of the shortest text in specialized discourse, as can be evinced by the number of occurrences of each label in the corpus.

Concordance					Concordance Plot	File View	Clusters/N-Gram	Collocate	Word Lis	Keyword Lis
Concordance Hits 41										
Hit	KWIC				File					
1		the medieval	lead glazed wares (		CompArIIEl.t					
2		as the	lead glazed wares		CompArIIEl.t					
3		main types –	lead glazed wares		CompArIIEl.t					
4		12th century.	Lead glazed wares		CompArIIEl.t					
5		of the	lead glazed wares,		CompArIIEl.t					
6		than the	lead glazed wares		CompArIIEl.t					
7		-sectioned pottery.	Lead glazed wares:		CompArIIEl.t					
8		with the	lead glazed wares,		CompArIIEl.t					
9		and Samanid	lead glazed wares,		CompArIIEl.t					
10		on monochrome	lead glazed wares		CompArIIEl.t					
11		case of	lead glazed wares)		CompArIIEl.t					
12		across the	lead glazed wares		CompArIIEl.t					
13		the typical	lead glazed wares		CompArIIEl.t					
14		Akhsiket’s	lead glazed wares.		CompArIIEl.t					
15		two green	lead glazed wares (		CompArIIEl.t					
16		of the	lead glazed wares		CompArIIEl.t					
17		al 1998, 242). The	lead glazed wares		CompArIIEl.t					
18		creating successful	lead glazed wares (		CompArIIEl.t					
19		on the	lead glazed wares,		CompArIIEl.t					
20		as the	lead glazed wares.		CompArIIEl.t					
21		, particularly the	lead glazed wares,		CompArIIEl.t					
22		on the	lead glazed wares		CompArIIEl.t					
23		that the	lead-glazed wares		CompArIIEl.t					

Figure 2.1. Lead glazed wares in CompArIIEl

“Lead glazed wares” occurs in fact 41 times in the corpus (Figure 2.1), whereas “glazed wares” occurs 136 times (Figure 2.2).

Concordance Hits 136			
Hit	KWIC	File	
1	percentage of glazed wares, although	CompArIIEl.t	
2	Valley. The glazed wares consist	CompArIIEl.t	
3	region, introducing glazed wares as	CompArIIEl.t	
4	applied. Alkali glazed wares have	CompArIIEl.t	
5	glass working. Glazed wares were	CompArIIEl.t	
6	for the glazed wares. Also,	CompArIIEl.t	
7	were in glazed wares. These	CompArIIEl.t	
8	Islamic period glazed wares became	CompArIIEl.t	
9	many varieties. Glazed wares form	CompArIIEl.t	
10	medieval lead glazed wares (Anarbaev	CompArIIEl.t	
11	Black- glazed wares: a.	CompArIIEl.t	
12	, when many glazed wares had	CompArIIEl.t	
13	wares. Alkali glazed wares Akhsiket'	CompArIIEl.t	
14	's alkali glazed wares fall	CompArIIEl.t	
15	the lead glazed wares at	CompArIIEl.t	
16	types - lead glazed wares which	CompArIIEl.t	
17	, and alkali glazed wares which	CompArIIEl.t	
18	century. Lead glazed wares can	CompArIIEl.t	
19	time. Alkali glazed wares have	CompArIIEl.t	
20	is the glazed wares found	CompArIIEl.t	
21	with the glazed wares and	CompArIIEl.t	
22	comparison with glazed wares - thus	CompArIIEl.t	
23	in the glazed wares. unglazed	CompArIIEl.t	

Figure 2.2. *Glazed wares in CompArIIEl*

Since the second cluster obviously includes the first – that is, all the 41 occurrences of “lead glazed wares” are also featured in the search “glazed wares” – we can conclude that the formula that does not mention the vitrifying agent is used in the corpus 85 times, twice as many times as the other candidate. A consultation with another archaeologist specializing in the same field finally brought to my knowledge that, since within the context of Islamic pottery the greatest majority of glazed wares are obtained through the use of lead, this detail is often taken for granted and omitted in catalogue descriptions, so that all “glazed wares” are “lead glazed wares” unless otherwise specified. Again, the CompArIIEl corpus reflects this insight by featuring only one instance of “lead-based” (Figure 2.3), and only to differentiate this process from the

other very common one in the context of Islamic archaeology, that is, alkali glazing.

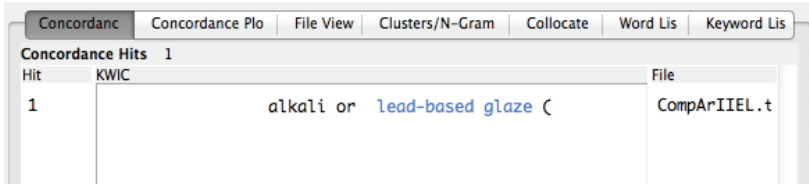


Figure 2.3 Lead-based glaze in CompArIIEl

#### 2.4.2. *Opacify vs. opaque, or the opposition between process and result*

Another interesting case was offered by the translation of “Ceramica con rivestimento vetroso opacizzante, monocromo”. The most immediate problem I was faced with was that of rendering “opacizzante”. Though “opacifying” is attested in the lexicon of pottery production, and I could thus consider it as a good candidate to translate the emphasis on the process rather than the result, again the few instances I found online were not from the specific field of academic discourse on archaeological findings. The adjective “opaque” was instead commonly used, with an emphasis on the outcome rather than on the process. I thus tested the possibility of using “opacified”, as in “opacified glaze”, which turned out to be fairly common (nearly 50,000 occurrences in Google) and was even attested in the “Dictionary of Ceramics”. A closer look at the micro-corpus rather than at the larger one, however, showed how the term is used consistently with reference to “tin glazing”, whereas my category was that of vitreous glazing. I had to finally choose “opaque”, again for the sake of clarity and naturalness in English.

My translation was then “Monochrome opaque glazed wares”, which was interestingly altered by the archaeologist who supervised the section on pottery into “Monochrome opaque *white* glazed wares”. I thus learned, and included in my personal corpus as a translator, that Anglophone Islamic archaeology considers the colour white a relevant feature in the categorization of

these wares and expects the description of this type of findings to be more detailed in terms of color, whereas the Italian categorization seems not to require any specification with regard to this feature. The asymmetry becomes significant when we have to categorize monochrome wares that are not white: they would be grouped together with the latter in Italian, whereas they would end up in a separate category in English.

#### 2.4.3. *Competing descriptions*

The final example I want to present here concerns the translation of “piede a disco, basso, ingobbiato”. In the description of the foot supporting a cup or a bowl, the Italian archaeologist pays attention to the shape of the abovementioned feature. The foot, in fact, can either be in the shape of a flat disc or in the shape of a tubular ring. This distinction emerges in the description made by the Italian author, but is not reflected in the English translation. This is because to an English speaking / writing scholar such distinction has no relevance in the characterization of the object. Reaching this conclusion was all but easy in the process of translating this text, but again the Internet and an appropriate use of the virtually ever-expanding corpora constituted by the texts and websites indexed by Google, proved very helpful in solving the issue of how to provide an appropriate description of this type of findings. Checking the currency of a literal translation of “a disco”, we find that the form “disc-shaped” has actually many occurrences (ca. 23,000), even in websites connected with the field of archaeology. Yet, the sources found in this case are not reliable since they, though written in English, are not the products of a mother tongue Anglophone context.

Since we are dealing here with an academic subject, it seems obvious to search academic corpora for evidence. If we restrict the search to .ac.uk websites (that is, the British academic domains), the occurrence drops to less than 5 instances. From my previous experience with translations in the same field, I was acquainted also with the Italian form “piede ad anello”, which led me to check the diffusion of its literal equivalent in English (which would be “ring-shaped foot”). Again, the search in scholarly British websites and publications gave no satisfactory outcomes



(2 results). On the other hand, there were numerous references to the foot of Islamic pottery vessels that did not mention the shape, neither as a disc, nor as a ring. So, I had to discuss how to appropriately describe this specific element with the editors of the publication, who were specialists in the field. It turned out that in the Anglophone world descriptions of the foot of cups / bowls are much less detailed than in Italian and basically amount to a statement about the presence or lack of such feature.

This is a clear example of the role played by language in shaping a field of knowledge. Before exactly the same type of artifact, the product of the same culture, and within the same scholarly discipline, the knowledge actually produced and circulated can vary according to the language used, so that we can say that looking at a single finding, archaeologists would tend to consider certain features rather than others as a function of the language they are using. The language shapes the narrative through which Islamic archaeology produces and reproduces itself as a discipline, and produces and circulates the knowledge about its own objects of study. In other words, what a translator does, especially at a time when scientific communication seems to be dominated by the use of a lingua franca, has potentially a huge impact on the continuous shifting boundaries marking academic disciplines.



### 3.

## A PARALLEL AND COMPARABLE CORPUS OF ISLAMIC ARCHAEOLOGY

### 3.1. THE ARIIEL CORPUS

This chapter presents the description and analysis of a complex corpus that is still in the making and is thus intended as a preliminary report on its current state of development and a reflection on the prospective uses of a future enhanced version of it.

The corpus has been called AriIEL from “Archaeology of Islam - Italian-English Lexicon”, and comprises a parallel corpus (ParAriIEL) and a comparable corpus (CompAriIEL) of texts on Islamic archaeology. The parallel corpus is constituted by two sub-corpora, one in Italian and the other in English, of academic texts originally written in Italian by different authors in a time span comprised between 2008 and 2012. In terms of genre, they range quite widely, from synthetic reports to catalogue entries to articles submitted to peer-reviewed journals.

At the time this book goes to press, the Italian corpus (ItAriIEL) counts 63,000 words and constitutes only a small fraction of a much wider collection of texts that are currently in the process of being selected and edited for future inclusion. The manual editing of texts is a customary phase in corpus compilation, as many texts feature characters that the available concordancers cannot recognise. Such operation, however, is of the utmost importance in the case of the body of texts under discussion as they are usually distinguished by the inclusion of images, captions, numbers, and, most crucially, quotations in different languages, chiefly Arabic. All these elements interfere with the processing of the text and thus need to be edited out so as to turn them into a set of linguistic data that the software can read and

analyse. The resulting documents then need to be converted and saved in a format that can be processed by the chosen concordancer <sup>1</sup> and another round of manual editing is needed in order to check that most (if not all) the critical elements have been removed. These operations are clearly time consuming and this is the main reason why, at the moment, the corpus is still considerably limited in terms of size. On the other hand, it is also very homogeneous in terms of topic and register, so it is ideally suited to analyses of bilingual parallel texts aiming at extracting specialized terminology and identify translation units.

The second sub-corpus is in English (EnArIIEl) and is constituted by the translation of each text that is included in the Italian one, all of them, as I have already anticipated in the previous chapter, carried out by the present writer and accepted for publication after peer reviewing and / or approval by editors. This was one of the guiding principles in the compilation of the corpora and another reason of its current limited size. Several translations, in fact, are still undergoing peer-reviewing and thus they were not considered final and acceptable candidates for inclusion at this stage. Others were never meant for publication, but rather for presentation at conferences. Again, at this initial stage of the research, the emphasis was placed on the stability, quality, and salience of the candidate texts for translational analysis, so the documents that had not yet been approved in some form by an external reader with linguistic expertise in the field of Islamic archaeology were ruled out. Also the English translations underwent considerable editing, as they obviously featured exactly the same amount of “interference” from other languages and media. It is in fact an interesting aspect of the translation context in which these texts were produced that the translator was constantly provided by authors with images, quotes in Arabic (which is not among the languages known to her), tables featuring statistics, etc., in brief rich apparatuses of information that apparently have nothing to do with the process of translating the parts in Italian into English, and yet was considered by all parties involved essential to carry out the task properly. Not only did pictures provide crucial insights into the characteristics of objects and places described in

---

<sup>1</sup> The data in this study were analysed using AntConc, the freeware software elaborated by Laurence Anthony and available for free download from the web.

the text, but even quotations from the Koran in Arabic were useful to understand the discursive practices of Islamic archaeology as an academic field. The EnArIIEl currently counts 55,000 words, a figure that seems to counter the widespread notion that translated texts are longer than original ones <sup>2</sup>.

The parallel corpus has then been aligned, that is, it has been processed by a software that segmented it into smaller units in Italian and aligned them to the corresponding units in English <sup>3</sup>. The alignment took place at the level of the sentence and a further round of editing was required as translation might result into a change of the original make up of a passage. Sentences might be grouped together or split into smaller ones, according to the syntax of the target language.

As has been repeatedly noted in discussions about translation and corpora, no matter how useful, coherent, and even rare a parallel corpus is, the ideal context for research in this field is given by the complementary presence of both parallel and comparable corpora. The compilation of a comparable corpus of academic texts in the field of Islamic archaeology was surely suggested by theoretical concerns, but was also the natural consequence of the documentation work carried out in the years during which the texts in ItArIIEl were translated. As a result of the use of the Web As Corpus to assist translation (a practice I have illustrated in Chapter 2), dozens of academic texts dealing with archaeological excavations and findings were searched for, read, evaluated, and archived in order to constitute a virtual corpus, that is, “a reliable resource ... at a minimal cost, based on texts mined from the Internet, to satisfy the translator’s documentation needs” (Corpas Pastor and Seghiri 2009: 78). <sup>4</sup> As the names by which these corpora are usually designated (virtual, disposable, ephemeral) indicate, these tools are normally discarded when a new translation task takes precedence. In this case, however, the same

---

<sup>2</sup> According to several studies this might be due to the phenomenon of explicitation, that is, “the process of rendering information which is only implicit in the source text explicit in the target text” (Frankenber 2009: 48). For a review of the literature on explicitation as a translation universal see Frankenberg 2009: 50-51.

<sup>3</sup> The software used for this project is LF Aligner, version 3.12, available for free download from the web.

<sup>4</sup> Documentation techniques using the web are illustrated, among others, by Corpas Pastor and Seghiri 2009.

person worked in close contact with a quite stable group of scholars carrying out translations for several years, and thus ephemeral corpora gradually evolved into less volatile entities. Eventually, these texts were edited to become part of the comparable corpus called CompArIIEl. For the purposes of this study, only ten texts were selected for inclusion, seeking a certain degree of balance with the parallel corpus in terms of genres and topic. All documents are academic, so the register is the same, written by native speakers of English, published in some Anglophone context, and publicly available on the Internet. The comparable corpus thus compiled counts 190,000 words, so it is roughly three times larger than the Italian and the English ones. In the compilation of such a corpus, an *ad hoc* tool for pragmatic uses, “quality takes priority over quantity” (Corpas Pastor and Seghiri 2009: 80) and homogeneity was deemed a much more relevant aspect than size.

None of the texts was annotated, in line with an approach to corpus work that aims rather at deriving information directly from the interrogation of corpora than at verifying existing theories about language through them. The approach is also markedly empirical, and reflects a staunch belief in translation as practice and in the role of the translator as mediator between cultures. I thus align my research through and about corpora in translation studies to a vision of translation that “not only highlights the human dimension of the profession of translation but also favors a humanistic perspective within the discipline of Translation Studies” (Austermühl 2012: 57).

The resulting complex corpus ArIIEl (Table 3.1) is then a bilingual, monodirectional, non-annotated corpus, which is made up of a comparable monolingual corpus in English and a parallel bilingual (Italian-English) one. The decisions that were made in compiling it were often opportunistic and mostly conditioned by the available translated texts, a “compromise between design criteria and practical constraints” (Zanettin 2002: 22), as is naturally the case with this kind of corpora<sup>5</sup>. Despite the theoretical limitations placed by such constraints on the present corpus, it has to be noted that the difficulties in gathering enough texts of the right type, covering a satisfactory range of topics within a highly specialized field, along with their translations is what makes parallel

---

<sup>5</sup> For a discussion of such decision-making processes see Frankenberg 2009.

corpora for specific purposes very rare. At the same time, this is also what makes them particularly interesting both from a pragmatic point of view and from a more theoretical stance, as they provide a repertoire of translation equivalents that cannot be found even in specialized bilingual dictionaries, provided that the latter are available for the desired language pair.

*Table 3.1. The ArIIEL Corpus*

<b>ArIIEL: Archaeology of Islam – Italian-English Lexicon</b>			
ParArIIEL		CompArIIEL	
<i>ItArIIEL</i>	<i>EnArIIEL</i>		
63,000	55,000		
118,000		190,000	

With specific reference to Archaeology, and to Islamic archaeology in particular, dictionaries and glossaries are available in English, but no bilingual tool for the pair Italian-English is available to this day, despite the fact that Italian archaeologists have been internationally recognised for decades as leading figures in the field. Their peculiar position as teachers and writers who, by effect of globalization and of deep-rooted power asymmetry within the academic world, disseminate their knowledge in a language that is not their mother tongue makes their cultural production particularly relevant to anyone who is interested in both scientific discourse and the language(s) in which it is encoded. Do specialised translations, be they corpus-based or not, play a role in the shifts that can be observed in scientific discourse when the paradigm of knowledge encoded in a certain language (in this case English) is used as the standard of “naturalness” for translated texts?

Projecting the present work into the future, one of the steps that lie ahead for the researcher is the compilation of a monolingual comparable corpus of Islamic archaeology in a diachronic perspective. The latter would allow in fact the observation of changes in the discursive practices of Italian archaeologists over the decades and eventually trace the influence exerted on them by English academic discourse.

### 3.2. CORPUS-BASED CASE STUDIES ON TRANSLATION UNITS

In what remains of this chapter I will illustrate through some examples the type of insights into translation practices and issues offered by domain-based parallel and comparable corpora. As will become evident through the discussion of the data, technical jargon and scientific language are only apparently easier to transfer from one language system to another. Different degrees of asymmetry between Italian and English turn the identification and translation of specific units of meaning into a complex act of interpretation that requires not just linguistic competence, but widely cultural and domain-specific ones as well. From understanding the influence exerted on terminology by the dominant theoretical paradigms within the field, to getting around the problem of translating a word for which there is no lexical equivalent in the target language, the case studies discussed below shed some light on the ways in which corpus-based tools can assist translators and reveal much about translation as a process. The corpus being monodirectional, only translations from Italian into English will be discussed.

#### 3.2.1. *Scrittura / Grafia*

As we have seen in the previous chapters, one of the main uses of parallel corpora, and the reason why they are simultaneously rare and coveted by researchers and translators alike, is the possibility they lend to look at aligned segments of texts in two languages and observe regularities, idiosyncrasies, and asymmetries, from lexis to grammar, in a contrastive perspective. In this section I will discuss some examples from the ParArIIEl corpus, trying to illustrate the translation issues raised by the specific domain of Islamic archaeology and what the translation practice here sampled can reveal about the asymmetry between Italian and English, not only in terms of available language patterns, but also in terms of discursive practices within academic fields.

I will start by looking at what is apparently a very common word, with straightforward meaning and translation from Italian into English, that is, *scrittura*. What Tognini Bonelli (2004) would refer to as the *prima facie* translation candidate for *scrittura* is



*writing*, which in a contrastive perspective, is already revealing of a well-known morphosyntactic phenomenon such as the formation of nouns from verbs in English. Yet, when we look at the way in which units of meaning containing this word are translated into English in the corpus (Table 3.2), we see that the rendering of *scrittura* is neither straightforward nor univocal.

*Table 3.2. Scrittura in ParArIIEl*

la <b>scrittura</b> , dono divino, alla quale si affidano	reverent respect for <b>writing</b> – God’s gift, bearing sometimes clear messages,
Nell’accezione che vede la <b>scrittura</b> dipinta in bruno	the version with the <b>inscription</b> painted in brown
La presenza del nome di Dio su una coppa con <b>scrittura</b> incompleta e imprecisa	The presence of the name of God on a bowl featuring an incomplete and inaccurate <b>inscription</b>
ornato pseudoepigrafico con <b>lettere in cufico</b> apicato	pseudo-epigraphic decoration in <b>apicated Kufic lettering</b>
La <b>scrittura</b> è ugualmente incisa con tratto sottile e poco profondo ed è eseguita in cufico	The <b>inscription</b> is also engraved in a thin, shallow trait and is executed in Kufic
<b>Scrittura</b> incisa, in cufico con apici triangolari, bilobati e qualche elemento vegetale	The <b>writing</b> is engraved, in Kufic, with triangular and bilobated apexes, and some vegetal elements
<b>Scrittura</b> eseguita in leggero rilievo, in cufico con ampi apici triangolari	The <b>writing</b> is carved in low relief, in Kufic with large triangular apexes
secondo la medesima tecnica adottata per la <b>scrittura</b>	following the same technique used for <b>the writing</b> .
Lo stile di <b>scrittura</b> delle steli d’Egitto	<b>The writing style</b> of the Egyptian stelae
<b>La scrittura incisa</b> si ritrova sui due documenti più antichi	<b>The engraved writing</b> is found on the two most ancient documents

è eseguita in <b>scrittura corsiva</b> , priva di punti diacritici	is executed in <b>cursive writing</b> , without diacritical marks
<b>uno stile di scrittura</b> identico al precedente e ugualmente in lingua persiana	executed in a <b>writing style</b> that is identical to the one we find in the large central band and similarly in Persian.
dell'oggetto eseguito in <b>scrittura</b> nasta'liq	in nasta'liq <b>writing</b>
introduce le iscrizioni di augurio eseguite in <b>scrittura</b> cufica.	introduce the well-wishing inscriptions executed in <b>Kufic</b> <b>script</b> .
due nuove iscrizioni a nome del celebre regnante zanghide, entrambe eseguite in <b>scrittura</b> <b>corsiva</b>	two other inscriptions bearing the name of the famous Zangid ruler would have been identified, <b>both</b> <b>written in cursive</b>
alcune di queste sono eseguite in <b>scrittura cufica</b>	and some of them are in <b>Kufic</b> <b>script</b>

First of all, the context is that of using writing as a form of decoration of objects or buildings, a peculiar feature of the Islamic world and its arts. Here *scrittura* has four different translations into English: *writing*, *inscription*, *lettering*, and *script*. If we look at the fragments containing each of the English candidates, we might notice a certain degree of regularity in the choices that were made <sup>6</sup>:

- la **scrittura**, dono divino, alla quale si affidano / *reverent respect for **writing** – God's gift, bearing sometimes clear messages*
- **Scrittura** incisa, in cufico con apici triangolari, bilobati e qualche elemento vegetale / *The **writing** is engraved, in Kufic, with triangular and bilobated apexes, and some vegetal elements*
- **Scrittura** eseguita in leggero rilievo, in cufico con ampi apici triangolari / *The **writing** is carved in low relief, in Kufic with large triangular apexes*

<sup>6</sup> The segments featured in each pair are obviously asymmetric, only partially overlapping, as a consequence of the different syntactic behaviour of English and Italian. Being instances of authentic language, they have not been edited as they reflect the translation context under discussion.

- secondo la medesima tecnica adottata per la **scrittura** / *following the same technique used for **the writing***
- La **scrittura incisa** si ritrova sui due documenti più antichi / *The **en-graved writing** is found on the two most ancient documents*
- è eseguita in **scrittura corsiva**, priva di punti diacritici / *is executed in **cursive writing**, without diacritical marks*
- **uno stile di scrittura** identico al precedente e ugualmente in lingua persiana / *executed in **a writing style** that is identical to the one we find in the large central band and similarly in Persian*
- dell'oggetto eseguito in **scrittura nasta'liq** / *in nasta'liq **writing***

*Writing* tends in fact to emphasise the act of producing a written text, as well as the techniques used and the forms that the texts take as a consequence. The use of *inscription*, instead, tends to evoke content alongside the physical presence of letters and words used as a decoration:

- Nell'accezione che vede la **scrittura** dipinta in bruno / *the version with the **inscription** painted in brown*
- La presenza del nome di Dio su una coppa con **scrittura** incompleta e imprecisa / *The presence of the name of God on a bowl featuring an incomplete and inaccurate inscription*
- La **scrittura** è ugualmente incisa con tratto sottile e poco profondo ed è eseguita in cufico / *The **inscription** is also engraved in a thin, shallow trait and is executed in Kufic*

When there is a reference to the type of letters that were used, which in many cases in the corpus is kufic, an alternative to both *writing* and *inscription* is *script*.

- introduce le iscrizioni di augurio eseguite in **scrittura cufica** / *introduce the well-wishing inscriptions executed in **Kufic script***
- alcune di queste sono eseguite in **scrittura cufica** / *and some of them are in **Kufic script***

*Lettering* is only used once, and this reflects the fact that, although the word is attested in the domain-based literature, it is definitely rarer than the other three candidates.

- ornato pseudoepigrafico con **lettere in cufico** apicato / *pseudo-epigraphic decoration in **apicated Kufic lettering***

All the terms attested in the parallel corpus as translations of *scrittura* can also be found in the comparable corpus, which confirms

that they are indeed possible equivalents of the term found in the source text.

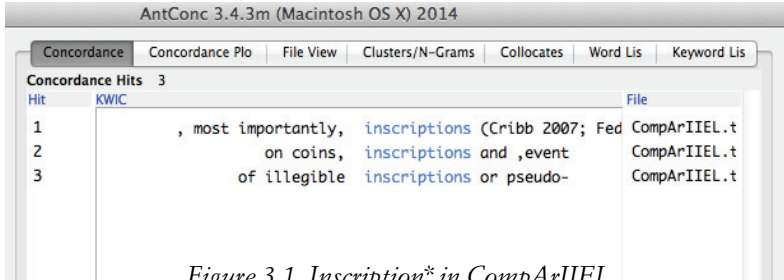


Figure 3.1. *Inscription\* in CompArIIEl*

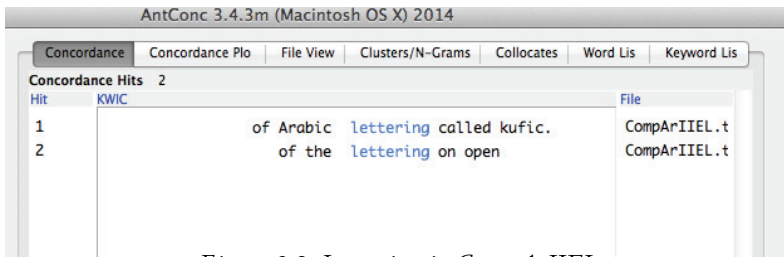


Figure 3.2. *Lettering in CompArIIEl*

Looking at the last two fragments in Table 3.2 allows us also to see how the translation sometimes seeks to embed in the target language some morphosyntactic features of the source language, whereas in other cases it privileges a more straightforward rendering. The tension between these often opposite objectives can determine a higher or lower degree of naturalness of the translation:

- *eseguite in corsivo* / *written in cursive*
- *eseguite in scrittura cufica* / *in kufic script*

In the first case, the Italian *eseguite*, which reveals a preference of the language (and / or the writer) for a verb expressing a generic action here applied to writing is merged with the English word for *scrittura* (n.), that is, *writing* (n.) and the concept of “executed in cursive writing” is synthetically rendered by embedding the action

in the chosen translation unit as *written in cursive*. In the second case, the verb *eseguito* is dropped entirely and the choice of the translation equivalent of *eseguito in scrittura cufica* simply becomes *are in kufic*, that is a status as opposed to an action.

Besides being one of the translation candidates for *scrittura*, script is attested in the parallel corpus also as one of the translations for another word belonging to the semantic field of writing, that is, *grafia* (Table 3.3).

*Table 3.3. Grafia in ParArIIEl*

in una <b>grafia</b> minuta ma decifrabile.	in a very tiny, yet decipherable <b>script</b> .
sia la <b>grafia</b> ornamentale cufica sia quella corsiva	executed inscriptions both in ornamental Kufic and in cursive (naskhi) script
infine una <b>grafia</b> intermedia risultata peculiare della ceramica	as well as in an intermediate type of <b>script</b> which
Il valore religioso che l'Islam attribuisce alla <b>grafia</b>	The religious value attributed by Islam to <b>writing</b>
in una elegante <b>grafia</b> cufica, l'iscrizione che recita:	painted inscription in an <b>elegant Kufic lettering</b> saying:
nonostante la <b>grafia</b> sia stata alterata dal ceramista / decoratore	Despite the potter / decorator's extremely skillful alteration of <b>writing</b> , it is often possible to decipher
sia la <b>grafia</b> sottesa da una linea continua la quale,	and the style of the <b>script</b> , resting on a continuous line,
sia per la <b>grafia</b> , nella quale si potrebbe riconoscere,	and with reference to the <b>writing</b> , which a slightly stretched reading might interpret as
L'intento di 'animare' la <b>grafia</b> si esprime sulle ceramiche	the intention of 'animating' the <b>inscription</b> ,
lo stile del volatile ad ali spiegate nella <b>grafia</b> dell'iscrizione	to reproduce the style of the bird with spreading wings in <b>the inscription</b>

alcuni particolari della <b>grafia</b> e il doppio tratto intorno al bordo	some details of the <b>script</b> and the double line running along the rim
in una <b>grafia</b> cufica con richiami ornitomorfi, v. Gouchani 1986, nn.	in Kufic <b>script</b> with ornithomorphic traits: see Gouchani 1986, nos.
per una <b>grafia</b> più corsiva v. Gouchani 1986, nn.	for a cursive <b>script</b> , see Gouchani 1986, nos.
è dipinta in un'elegante <b>grafia</b> cufica, l'iscrizione che recita	we find a painted inscription in an elegant <b>Kufic script</b> saying:
con una caratteristica <b>grafia</b> cufica composta di lettere profilate in nero e campite di verde o rosso	with a peculiar <b>Kufic script</b> composed of letters outlined in black and filled in green or red

The translation units for *grafia* are *script*, *inscription*, *writing*, and *lettering*, that is, a cluster of words we have already seen as possible translations of *scrittura*, hinting at the semantic overlapping of the two terms both in Italian and in English. The slippery distinction between content and form that we have already noticed in the pairs *scrittura-inscription* / *writing* seems to be confirmed by the translation equivalents of *grafia*. Although it occasionally hints at the possibility of reading / deciphering a text, script tends to be used quite consistently to refer to the material form of the writing, its very presence on an object. When the emphasis is on the action from which the script is produced, we find again writing. Particularly interesting is the fact that both when translating *scrittura* and when translating *grafia*, a reference to religion elicits *writing* as the chosen candidate:

- la **scrittura**, dono divino, alla quale si affidano / *reverent respect for **writing** – God's gift, bearing sometimes clear messages*
- Il valore religioso che l'Islam attribuisce alla **grafia** / *The religious value attributed by Islam to **writing***

There is religious reverence in the Islamic culture towards the possibility of encoding meaning into language through the alphabet, and thus, in this context, writing cannot be reduced to its material form. *Lettering* is once again the least attested form, confirming the awareness on the translator's part of its low frequency

in English. In this case, the choice of *kufic lettering* over the more widely attested option *kufic script* seems to be motivated by the close by presence of the assonant *inscription*:

- dipinta in un'elegante grafia cufica, l'iscrizione che recita / *a painted inscription in an elegant kufic lettering saying*

### 3.2.2. *Ornato / Decorazione*

The asymmetry between the technical terminology of two languages can take different forms, one of these being the presence of several (partial) synonyms for the same concept in one language vs. the lack of an equal number of alternatives in the other. The analysis of the data yielded by the ParArIIEL and CompArIIEL corpora reveals such a case to be represented by the cluster *ornato / decorazione*, which is quite consistently translated into English as *decoration* or some other word derived from the same root. (Table 3.4 and Table 3.5).

*Table 3.4. Ornato in ParArIIEL*

moltiplicare all'infinito varianti di ogni tipo di <b>ornato</b> e di combinarle insieme per ottenere una rappresentazione immaginifica	to endlessly multiply versions of any <b>decorative pattern</b> and mix them together to produce highly imaginative representations
ricorre all'esterno lo stesso tipo di <b>ornato</b> con tratti e cerchi	on whose outside we find the same <b>decorative pattern</b> of dashes and circles
Un <b>ornato</b> molto simile	A <b>decoration</b> very similar to the one
Decorazione con <b>ornato</b> vegetale	Decoration with vegetable <b>motif</b>
MO110 che, pur essendo dipinta in bruno e rosso, denuncia, per la disposizione <b>dell'ornato</b> ,	MO110 which, though being painted in red and brown, in terms of <b>ornaments</b> positioning shows a
<b>Ornato</b> esclusivamente geometrico è invece quello che decora la coppa n.	An exclusively geometrical <b>motif</b> decorates bowl no.

Questo <b>ornato</b> , che si vuole far derivare dall'incipit delle sure del Corano	This <b>motif</b> , usually considered as derived from the incipit of the Quran's sure
L' <b>ornato</b> , alquanto ridotto, si dispone su tutta la superficie	The <b>decoration</b> , rather diminutive, is applied so as to occupy the entire surface
un <b>ornato</b> con effetto marmorizzato sarebbe stato dipinto in bianco sotto una vetrina incolore ancora oggi brillante.	a <b>marbled decoration</b> in white would have been painted under a colourless glaze which is still shiny today.
la sua forma alquanto sottile e l' <b>ornato</b> pseudoepigrafico con lettere in cufico apicato	its rather slender form and the pseudo-epigraphic <b>decoration</b> in apicated Kufic lettering
Un esempio di <b>ornato</b> ottenuto con l'incisione e il traforo è custodito al MNAO	A example of <b>decoration</b> executed using the engraving and the openwork techniques is preserved at the MNAO
Infine l' <b>ornato</b> della coppa turchese n.	Finally, the <b>decoration</b> of the turquoise bowl no.
per dare maggior risalto ai dettagli dell' <b>ornato</b> ,	in order to give greater emphasis to the details of the <b>decoration</b> ,
per meglio far risaltare l' <b>ornato</b> che è spesso anche vistoso ed elaborato	to better bring out the <b>decoration</b> that is often bold and elaborate
è inciso un medaglione circolare apicato con <b>ornato</b> vegetale	a small circular medallion is carved and apexed with vegetal <b>ornaments</b> .
L'elemento superiore, leggermente piriforme, è invece anepigrafico ed è <b>ornato</b> con una sequenza di otto lobi leggermente concavi	The upper element, which is slightly pear-shaped, is instead non-epigraphic and <b>decorated</b> with a sequence of eight slightly concave lobes
Sulla parete sono presenti tre cartigli rettangolari con i lati brevi concavi <b>ornati</b> ognuno con un tralcio sinusoidale con semipalmette	On the wall there are three rectangular cartouches with concave short sides, each of them <b>decorated</b> with a sinusoidal spray with half-palmettes



tra le lettere su un <b>campo ornato</b> con linee continue a zig-zag:	among the letters against a <b>background adorned</b> with continuous zig-zag lines:
---	--

*Table 3.5. Decorazione in ParArIIEl*

generalmente <b>senza decorazione</b>	generally <b>undecorated</b>
Nella <b>decorazione</b> della ceramica	the <b>decoration</b> of pottery
la <b>decorazione</b> ottenuta mediante sigilli	the <b>decoration</b> obtained by applying seals
risiede nella loro <b>decorazione</b>	lies in their <b>decoration</b>
oggetti con <b>decorazione</b> a stampo	items with moulded <b>decoration</b>
ceramiche con <b>decorazione</b> dipinta	pottery with painted <b>decoration</b>
dare risalto ai colori della <b>decorazione</b>	to give great emphasis to the colours of the <b>decoration</b>
dare un maggior risalto alla <b>decorazione</b> dipinta	better emphasise the painted <b>decoration</b>
nella <b>decorazione</b> dipinta	the painted <b>decoration</b>
Viene così definita quella produzione di oggetti d'uso comune e di <b>decorazione</b> architettonica	This is the name given to that production of utilitarian wares and architectural <b>decorations</b>
<b>decorazione</b> di queste ceramiche,	the <b>decoration</b> of these vessels
<b>Decorazione</b> epigrafica	Epigraphic <b>decoration</b>
<b>Decorazione</b> con reminiscenze epigrafiche	<b>Decoration</b> with epigraphic reminiscences
<b>Decorazione</b> con ornato vegetale	<b>Decoration</b> with vegetable motif
<b>decorazione</b> geometrica	<b>geometrical</b> decoration
Due larghi cartigli contrapposti contenenti grafemi nei quali si individua solo la lettera k	Finally, two wide opposing cartouches containing graphemes – among which the only

costituiscono infine la <b>decorazione</b> della coppa n.	decipherable one is the letter k – <b>decorate</b> bowl no.
La <b>decorazione</b> a settori con reminiscenze epigrafiche	The <b>decoration</b> divided into panels and with epigraphic reminiscences

Let us begin with *ornato*, which in Italian can be both a noun (e.g. *ornato vegetale*) and the past participle of the verb *ornare* (e.g. *l'elemento superiore...è ornato*). Such distinction, which in Italian is not marked by any change in the word spelling or pronunciation, is reflected in the English translation equivalents *decoration* and *decorated*. The list of identified candidates for the translation of *ornato* is completed by *ornaments*, *decorative pattern*, and *motif*.

The translator repeatedly selects some derivation of *decorat*<sup>7</sup> trying to consistently represent in the target language the distinction between noun and verb that is found in the source text. The resulting English words have quite transparent equivalents in Italian, all of them being derivations of *decor*<sup>\*</sup> and covering all the required grammar classes (*decorare* / *decorato* / *decorativo* / *decorazione*). Likewise, a cluster of words exists in English which have the same etymological origin as *ornato* (*ornate* / *ornament* / *ornamental*), yet only twice in this corpus does the translator opt for one of these, a fact that invites closer analysis of the possible reasons behind such a choice. The two translation units are the following<sup>7</sup>:

- un medaglione circolare apicato con **ornato** vegetale / a small circular medallion is carved and apexed with vegetal **ornaments**
- MO110 che, pur essendo dipinta in bruno e rosso, denuncia, per la posizione **dell'ornato** / MO110 which, though being painted in red and brown, in terms of **ornaments** shows

In the first case, we are in the presence of a sentence marked by a concentration of domain specific jargon (*apicato*, *ornato vegetale*) coupled with a certain 'quaintness' signalled by the use of *medaglione*. The latter, in fact, is a much less common, we could even say literary, variant of *piccolo medaglione*, and its use

<sup>7</sup> Throughout this section, the examples provided often expand the translation units properly defined, in order to help the reader follow the argument by providing longer chunks of text.

contributes to the general tone of the sentence by conferring it a markedly aesthetic quality. The English translation equivalent can only be *small medallion*, as the language system does not allow for any variant on the name as the Italian one does. The choice of *ornament* might then be explained as an attempt to compensate for the lack of an adequate equivalent for the stylistically motivated *medaglione*.

If we look at the second case, we might notice a similar translational context, as the verb in the Italian text is certainly not a common term for the description of the characteristics of an object. Again, the English version seems to compensate for some linguistic sophistication that is lost in the rendering of *denuncia* with *shows*, while recuperating it through the use of *ornaments* instead of *decoration*. As the word is not to be found in our comparable corpus, one way to confirm that this is actually a good candidate is, as we have seen in the previous chapter, a Web search of relevant sites through key terms such as the combination of ‘ornaments Islamic pottery wares’. The search through these keywords aims at eliciting results from specialized, hopefully reliable websites thanks to the presence of multi word units (such as ‘Islamic pottery’) that are vastly used in the domain specific literature. Among the first results yielded by this search there is the website of the Encyclopaedia Iranica, an international scholarly project based at Columbia University. In this absolutely reliable source the word *ornament* is attested in the chapter devoted to ‘Ceramics’, making it a very good candidate for the translation. The final approval given by the editors of the publication confirms that *ornament* and its derivations are acceptable lexical options in the context of the argument developed by the author in Italian. Yet, the consistent predilection for one of its synonyms finds its motivation when we turn to the comparable corpus to validate all the translation equivalents against native speakers’ usage in the same field. As shown in Figure 3.3, and as I have anticipated above, the CompArIIEL corpus has not a single occurrence for *ornament* / *ornaments* and the only word with this root that is present is the adjective *ornamental*.

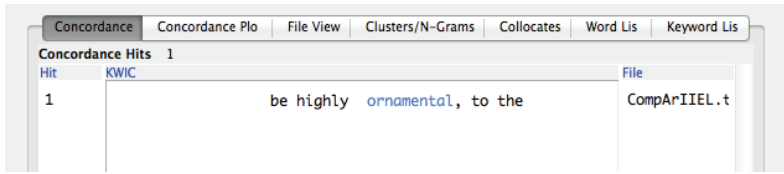


Figure 3.3. *Ornament\** in CompArIIEl

This might suggest that the corpus is unbalanced in the sense that not enough texts were included in its compilation to represent a satisfactory repertoire of lexis in the domain of decorated pottery. One way of verifying this is to search for the other translational candidates for *ornato*, that is, above all *decoration* and *decorated*. Figures no. 3.4 and no. 3.5 clearly show how the field of decorated wares is amply represented in the corpus, as *decoration* is present 73 times, and *decorated* 11 times, in both cases with collocates such as ‘unglazed wares’, ‘pottery’, ‘glazed ceramic’. We can thus quite safely conclude that, although *ornament* is one of the words that can be used to describe Islamic wares, it is a much less common option compared to decoration.

In two instances *ornato* (noun) is translated as *decorative pattern*, a choice that can be explained by looking at the immediate context in the Italian source text. By paying attention to the unit of meaning rather than the word, we see that in both cases *ornato* collocates with *tipo di*, that is the reference is not to the fact that the object is decorated, but to the typology of the decoration and its repetition to produce a pattern. In order to provide an adequate, that is functionally complete (Tognini Bonelli 2002) English version of *ornato* in these contexts it is necessary to translate the entire semantic unit *tipo di ornato*, as it clearly cannot be further segmented without losing part of its meaning.

Finally, when the source text refers to the specific characteristics featured in the decoration of an item, the translation equivalent of *ornato* is *motif*, which is also attested as collocating with *decoration* and providing details to what would otherwise be a very generic description (*decorazione con ornato vegetale* / *decoration with vegetal motif*).

Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Concordance Hits 73						
Hit	KWIC					File
1	glazes, and	decoration	methods such			CompArIIEl.t
2	. and texture,	decoration	colour and			CompArIIEl.t
3	glazed ceramic	decoration	came into			CompArIIEl.t
4	body fabrics,	decoration	methods, and			CompArIIEl.t
5	blue painted	decoration	created by			CompArIIEl.t
6	'splashes', sparse	decoration	with little			CompArIIEl.t
7	for the	decoration	of earthenwares,			CompArIIEl.t
8	a popular	decoration	method for			CompArIIEl.t
9	surface for	decoration.	Slip paints			CompArIIEl.t
10	a monochrome	decoration.	In the			CompArIIEl.t
11	widely used	decoration	technique on			CompArIIEl.t
12	, when the	decoration	style reached			CompArIIEl.t
13	protects the	decoration	from weathering			CompArIIEl.t
14	specialised ware.	Decoration	was applied			CompArIIEl.t
15	a visual	decoration.	Typically decora			CompArIIEl.t
16	forms and	decoration	patterns.	Figur	CompArIIEl.t	
17	. Raw materials,	decoration	methods and			CompArIIEl.t
18	and visible	decoration	methods, as			CompArIIEl.t
19	variability in	decoration	methods, and			CompArIIEl.t
20	types and	decoration	motifs. From			CompArIIEl.t
21	, styles and	decoration	techniques, and			CompArIIEl.t
22	on the	decoration	methods. As			CompArIIEl.t
23	type and	decoration	style for			CompArIIEl.t

*Figure 3.4. Decoration in CompArIIEl*

Concordance				Concordance Plot	File View	Clusters/N-Grams	Collocates	Word Lis	Keyword Lis
Concordance Hits 11									
Hit	KWIC	File							
1	) , with slip-	decorated	wares and	CompArIIEl.t					
2	render the	decorated	surface of	CompArIIEl.t					
3	vessels are	decorated	with incised,	CompArIIEl.t					
4	0.	Decorated	unglazed wares,	CompArIIEl.t					
5	3.22. Examples of	decorated	fineware sherds.	CompArIIEl.t					
6	are occasionally	decorated	with moulded	CompArIIEl.t					
7	of the	decorated	unglazed wares	CompArIIEl.t					
8	glazed and	decorated	pottery, are	CompArIIEl.t					
9	white slip,	decorated	with black	CompArIIEl.t					
10	or cup,	decorated	on the	CompArIIEl.t					
11	or brown	decorated	sherds were	CompArIIEl.t					

*Figure 3.5. Decorated in CompArIIEl*

This analysis allows us to conclude that the Italian word *ornato* is a polysemous term whose different meanings are revealed by its collocates. It is the elicitation of a segment of language containing both *ornato* and the significant collocate as the semantic unit of translation that enables its encoding in a corresponding equivalent in English.

Table 3.5 above shows the repeated occurrence of *decorazione* in Italian, which is straightforwardly translated as *decoration*, a quite transparent equivalence. While *ornato* seems to be fragmented into its synonyms when it is transferred in English, *decorazione* seems only to have one possible equivalent, which further explains why other terms might be needed to cover the many meanings of *ornato*. Moreover, it is interesting to notice how the absence of decoration (which could be rendered as *without decoration* in English), is actually expressed with *undecorated*: this highlights a feature of the English language, which has a predilection for morphological synthesis whenever this is allowed by the context. In this case, the prefix *un-* performs the same semantic function as the Italian preposition *senza*, merging a two-word term (Chodkiewicz et al. 2002) into a one-word translation equivalent in the target language.

### 3.2.3. *Campito* / *Campitura*

An extreme case of lexical asymmetry highlighted by the analysis of the parallel corpus is constituted by the cluster of words that are used to refer to and describe a very specific type of decoration, that is, what in Italian is called *campitura* and in English can be defined as “the painting of the background”<sup>8</sup>. (Table 3.6)

Table 3.6. *Campitura* / *Campito* in ParArIIEl

su un fondo <b>campito</b> con ornati vegetali	against a background <b>filled</b> with vegetal decorations
su un fondo <b>campito</b> con ornati vegetali	some vegetal elements used as <b>background filler</b>

<sup>8</sup> This is the translation provided by the online *Garzanti* bilingual Italian-English dictionary, one of the few to even list the word *campitura*, which is identified as belonging to the technolact of the fine arts ([garzantilinguistica.it](http://www.garzantilinguistica.it))

fascia circolare continua <b>campita</b> con una sequenza di triangoli	<b>background featuring</b> a sequence of triangles
un fondo <b>campito</b> da brevi segmenti obliqui	on a <b>background filled</b> with short slant segments
una larga fascia variamente <b>campita</b>	employs wide band, variously <b>filled in</b>
quattro spazi contenenti ciascuno una larga palmetta <b>campita</b> di spirali	four panels, each of them containing a large palmette <b>filled in</b> with whorls.
spazi triangolari <b>campiti</b> da un reticolo puntinato in rosso e in bruno	triangular spaces <b>filled in</b> with a grid of dots in red and brown
gli elementi di <b>campitura</b> come i cerchi con punti e tratti paralleli	the elements <b>filling the background</b> such as the dotted circles and the parallel dashes
i motivi della decorazione e per la caratteristica <b>campitura</b> a grossi punti si possono attribuire alla stessa officina	the motifs of the decoration are painted and the peculiar dotted <b>filling</b> allow us
fondo <b>campito</b> da brevi segmenti obliqui	against a background <b>decorated</b> with short slant lines
il blu anche scuro per le <b>campiture</b>	the blue (also in a dark shade) for the <b>filling</b>
conserva la <b>campitura</b> a grossi punti del corpo e i segni del nastro intorno al collo	emphasised with large dotted stamps and, finally, a wide band, <b>variously filled in</b> , marks the base of the neck
provviste di code scorpioniche, su un fondo <b>campito</b> con ornati vegetali.	scorpion's tails, against a background <b>filled with</b> vegetal decorations
una caratteristica grafia cufica composta di lettere profilate in nero e <b>campite</b> di verde o rosso	with a peculiar Kufic script composed of letters outlined in black and <b>filled in</b> green or red
sia gli elementi di <b>campitura</b> come i cerchi con punti e tratti paralleli	and the elements <b>filling</b> the background such as the dotted circles and the parallel dashes
utilizza larghe fasce variamente <b>campite</b>	for example, employs wide band, variously <b>filled in</b>

The translator is faced here with a quite challenging task as the Italian lexicon features both a verb (*campire*, *campito*) and a noun (*campitura*), whereas the English lexicon seems to have no single word to express the same very specific concept. The available linguistic options in the two languages differ not only with regard to the number of words needed to express the idea (i.e. the unit of meaning or unit of translation), but also in terms of the domains they refer to. *Campire* / *campitura* are words that by no means would be used in non-specialized contexts, and have a very specific semantic value that is unambiguously evoked whenever they are used. The definition of *campire* provided by the Treccani Encyclopedia helps clarify the point:

Campire: In pittura, fare il campo, dare cioè risalto al fondo, o a una zona delimitata, segnata in precedenza da un contorno, stendendo il colore in modo uniforme; nel disegno può farsi anche mediante tratteggio (Treccani.it)

Translating this definition into English is, again, far from easy, as it uses another expression (*fare il campo*) that is drawn from the same cluster of Italian semantic units that seem to be missing in English and that would be obscure even to native speakers of Italian who are not conversant in the jargon of the fine arts. The following explication, however, allows us to understand that *campire* means to evenly paint the background, or a delimited area that has been outlined in advance, so as to bring it out.

As can be seen, *campire* and its morphological derivations are highly technical words untainted by polysemy. As such, they belong to a category of words for which translation is usually believed to be straightforward and unproblematic. Yet, the only translation available in bilingual dictionaries is concocted from common words that have meanings utterly unrelated to the world of fine arts and can be variously combined with other terms in totally different contexts (whereas *campitura* is rigidly domain-specific). Furthermore, even when the entire segment “the painting of the background” is treated as a multiword term (Chodkiewicz et al. 2002) and used as a single unit of meaning it can still refer not only to the domain of art, but quite generically to any act of painting of any sort of background. The use of *campito* / *campitura*, instead, not only says that we are describing a painted back-



ground, but, as is clarified by the encyclopedia entry quoted above, also reveals how that background is treated as a field (*campo*), that is, it is delimited, its contour has been delineated before the colour or other decorative element is applied. Moreover, such decoration is evenly distributed and its function is that of bringing out the background. This wealth of information is completely lost in the English version, as nothing in “painting of the background” reveals anything about the technique (the even distribution of colour or decorative motif) and the function (to emphasize the importance of the background making it a central element of the composition).

A certain attention to the possibility of recuperating part of this loss of meaning can be noticed in the consistent use (we might say the insistence on the use) of *filled* / *filling* in English as a way to hint at the even and complete covering of the background by the decorative motif:

- su un fondo **campito** con ornati vegetali / *against a background **filled** with vegetal decorations*
- una larga fascia variamente **campita** / *employs wide band, variously **filled in***
- spazi triangolari **campiti** da un reticolo puntinato in rosso e in bruno / *triangular spaces **filled in** with a grid of dots in red and brown*
- gli elementi di **campitura** come i cerchi con punti e tratti paralleli / *the elements **filling the background** such as the dotted circles and the parallel dashes*
- il blu anche scuro per le **campiture** / *the blue (also in a dark shade) for the **filling***

The form *filling* is not attested in the comparable corpus, and *filled*, though attested, never refers to the painting of the background (Figure 3.6). This can be seen as evidence that these “invented translation equivalents” (Teubert 2002: 191) are an attempt on the part of the translator to find a remedy for the asymmetry between the specialized lexicons of fine arts in Italian and in English. The provided translations try to bring into the target language as much meaning as possible from the semantic unit in the source language, thus making a compromise between naturalness and scientific accuracy.

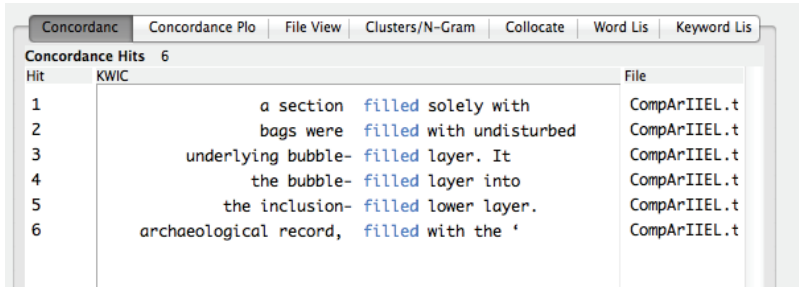


Figure 3.6. Filled in CompArIIEl

That the semantic field covered by the English word *background* does not fully overlap with the one covered by the Italian term *campo* (from which *campire* / *campito* / *campitura* are derived) is further demonstrated by the fact that another word is widely used by Italian archaeologists with reference to a background that is not a *campo*, and that word is *fondo* (Table 3.7).

Table 3.7. Fondo in ParArIIEl

un motivo zoomorfo su un <b>fondo</b> di tralci	a zoomorphic motif against a vine-scrolls <b>background</b>
dipinte soprattutto in blu su <b>fondo</b> bianco	mostly painted in blue on a white <b>background</b>
sul <b>fondo</b> bianco opaco	on an opaque white <b>background</b>
si stagliano su un vistoso <b>fondo</b> giallo / mostarda	stand out against a vivid mustard yellow <b>background</b>
Ceramiche dipinte in bruno e in bruno e rosso su <b>fondo</b> chiaro.	Pottery painted in brown or brown and red on a pale <b>background</b> .
dal tratto preciso e netto ai quali la <b>superficie bianca</b> , e volutamente sgombra, dà particolare risalto	characterized by a neat and precise stroke which stands out against the white <b>background</b>
Decorazine pseudoepigrafica (bruno su <b>fondo</b> bianco)	Pseudo-epigraphic decoration (brown on a white <b>background</b> )
ceramica policroma su <b>fondo</b> chiaro	polychrome pottery on a pale <b>background</b>

As in the case of *decorazione* / *decoration* illustrated above, here too we have just one consistent translation for the Italian word, whose English correspondent however has a more generic meaning and covers a range of semantic areas, from the description of what in a decorative motif is not in the foreground, to a contoured and filled-in area in a decoration (*campo*), to abstract concepts such as circumstances and historical context. All these uses can be evinced by looking at the concordance of *background* in the comparable corpus (Figure 3.7).

Concordance Hits 23			
Hit	KWIC	File	
1	used for background research. Severc	CompArIIEl.t	
2	a firm background of use	CompArIIEl.t	
3	and archaeological background to the	CompArIIEl.t	
4	, first, the background to the	CompArIIEl.t	
5	future research. Background to the	CompArIIEl.t	
6	, provides a background of political	CompArIIEl.t	
7	plain white background make the	CompArIIEl.t	
8	be either background elements, or,	CompArIIEl.t	
9	theme. Other background patterns includ	CompArIIEl.t	
10	purple-tinged 'background' colours of	CompArIIEl.t	
11	transparent lilac background. The green	CompArIIEl.t	
12	lilac-coloured 'background' glaze. BSE	CompArIIEl.t	
13	. The 'lilac' background colour is	CompArIIEl.t	
14	white engobe background. The pattern	CompArIIEl.t	
15	century black- background motifs in	CompArIIEl.t	
16	as a background layer on	CompArIIEl.t	
17	a white background. The technique	CompArIIEl.t	
18	a light background. The background	CompArIIEl.t	
19	background. The background is a	CompArIIEl.t	
20	a lilac background glaze. This	CompArIIEl.t	
21	better geological background for context,	CompArIIEl.t	
22	level of background variation, which,	CompArIIEl.t	
23	a white background to the	CompArIIEl.t	

Figure 3.7. *Background in CompArIIEl*

The case studies presented so far illustrate how the combination of parallel and comparable corpora in domain-based linguistic work constitutes a valuable tool for both theoretical investigation

of contrastive aspects of languages and practical use in professional settings. In the next stage of this project, the three corpora will be enlarged with the addition of texts that are currently in the process of being edited and assessed for inclusion and that will provide linguistic data in areas of Islamic archaeology that are relatively underrepresented in the current setup. While there is no doubt, in fact, that pottery and inscriptions are strongly represented in the ParArIIEL corpus at the moment, other areas of archaeological studies will be included, such as a rich body of texts on metalwork and burial sites. Likewise, the CompArIIEL corpus will be expanded with a conspicuous number of texts featuring these new terminological areas and further refinements will be brought with regard to its balance and representativeness of the specialized lexis of Islamic archaeology.

# GLOSSARIO / GLOSSARY

<b>A</b>
<b>abradere:</b> abrade
<b>addorsato:</b> back to back
<b>a disco:</b> ( <i>di piede o supporto</i> ) disc-shaped
<b>adornare:</b> adorn
<b>ageminata:</b> inlaid
<b>agente opacizzante:</b> opaque (NB: <i>in inglese non vi è mai riferimento al processo di opacizzazione, ma solo all'effetto. Es.: l'utilizzo dell'agente opacizzante ha permesso di dare un maggior risalto alla decorazione dipinta: opacity allowed to bring out the painted decoration.</i> )
<b>à la barbotine:</b> à la barbotine ( <i>tecnica di decorazione consistente nel soffiare argilla liquida con una cannuccia su una ceramica al fine di creare un motivo decorativo.</i> )
<b>alonatura:</b> halos
<b>altorilievo:</b> high-relief
<b>analisi:</b> analysis --- <b>dei materiali:</b> materials analysis --- <b>stratigrafica:</b> stratigrafic analysis --- <b>strutturale:</b> structural analysis
<b>ansa:</b> handle
<b>ansula:</b> ansula; small handle
<b>antropomorfo:</b> anthropomorphic
<b>apicato:</b> apicated; apexed

<b>apici:</b> apexes --- <b>triangolari:</b> triangular apexes
<b>apicatura:</b> apexes; <i>alcune apicature assumono la forma di elementi vegetali: some apexes take the shape of vegetal elements</i>
<b>approccio multidisciplinare:</b> multidisciplinary approach
<b>arabesco:</b> arabesque
<b>arco:</b> arch --- <b>pentalobato:</b> five-lobed arch; <i>arco / archetti a tutto sesto incrociati: crossed round arches</i>
<b>architetonico:</b> architectonic, architectural
<b>archivi digitalizzati:</b> digitalized archives
<b>aree cimiteriali:</b> cemeteries; burial sites
<b>argilla:</b> clay --- <b>bianca:</b> white clay --- <b>cruda:</b> raw clay --- <b>liquida:</b> liquid clay, à la barbotine --- <b>naturale:</b> natural clay
<b>arte fittile:</b> pottery
<b>arti figurative:</b> figurative arts
<b>artigiano:</b> ( <i>della ceramica</i> ) potter; ceramist
<b>asse:</b> axis
<b>aste:</b> ( <i>delle lettere</i> ) stems
<b>attestato:</b> attested

<b>B</b>
<b>banda:</b> band --- <b>circolare:</b> circular band; <i>banda circolare continua: continuous circular band</i> --- <b>diametricale:</b> diametral band <b>bande concentriche:</b> concentric bands
<b>base:</b> ( <i>di lampada</i> ) base --- <b>a cupola:</b> dome-shaped base --- <b>a campana:</b> bell-shaped base --- <b>a saliera:</b> salt cellar base --- <b>a semi-cupola:</b> semi-dome shaped base
<b>base piatta:</b> flat base; flat bottom
<b>base strombata:</b> splayed base
<b>bassa carinatura:</b> low carinated sides
<b>bassorilievo:</b> bas-relief
<b>bilobati:</b> bilobated; bilobed
<b>blocco:</b> block --- <b>epigrafico:</b> epigraphic blocks
<b>bolli:</b> stamps
<b>borace:</b> borax
<b>bordo :</b> rim --- <b>estroflesso:</b> everted rim --- <b>festonato:</b> festooned rim --- <b>in rilievo:</b> embossed rim --- <b>lobato:</b> lobed rim --- <b>rientrante:</b> recessed rim --- <b>roverso:</b> reversed rim --- <b>verticale:</b> straight rim --- <b>breve bordo rialzato:</b> short raised rim
<b>brocche:</b> jugs, water strainers
<b>bronzo:</b> bronze --- <b>fuso:</b> cast bronze
<b>brucia-incenso:</b> incense-burner
<b>bruno:</b> ( <i>colore</i> ) brown

**bugnato:** ashlar

**paramento ---:** ashlar face



<b>C</b>
<b>calamo:</b> quill
<b>calcare:</b> limestone
<b>campito:</b> filled (NB: <i>l'inglese non ha un termine tecnico equivalente all'italiano 'campito' o 'campitura' e fa riferimento in questo contesto a un più generico 'background'. Es.: fondo campito con ornati vegetali: background filled with vegetal decorations; lettere profilate in nero e campite di verde: letters outlined in black and filled in green; fascia campita con una sequenza di triangoli: band featuring a sequence of triangles.</i> )
<b>campitura:</b> background filling; <i>gli elementi di campitura: the elements filling the background; la caratteristica campitura a grossi punti: the characteristic dotted filling; il blu scuro per le campiture: the dark blue for the filling; campitura con brevi segmenti obliqui: hatching.</i>
<b>campo epigrafico:</b> epigraphic field
<b>candelieri:</b> torch-stands
<b>cantonale:</b> quoin
<b>caolino:</b> kaolin
<b>capitello:</b> capital; <i>capitello a bulbo: bulb shaped capital</i>
<b>capomastro:</b> mastermason
<b>caratteri cufici:</b> Kufic script; kufic writing; kufic lettering (NB: <i>alcuni ritengono che Kufic vada sempre scritto con la maiuscola, seguendo l'ipotesi che derivi dall'antica città di Kufa, ma il consenso non è unanime né sull'etimo né sull'ortografia; la dizione 'lettering', benché attestata in letteratura, è piuttosto rara.</i> )
<b>carenatura:</b> flaring; <i>carenatura bassa: low fairing</i>
<b>cartiglio:</b> cartouche --- <b>composito:</b> composite cartouche --- <b>epigrafico rovesciato:</b> reversed epigraphic cartouche
<b>catalogazione museale:</b> cataloguing
<b>ceramica:</b> ceramics; pottery; vessels; wares --- <b>ceramica a colature:</b> splashed wares --- <b>ceramica a lustro metallico:</b> lustre wares

- **ceramica a pelle d'uovo**: eggshell wares
- **ceramica comune**: unglazed pottery / wares
- **ceramica con rivestimento vetroso al piombo**: (*ceramica invetriata*) (lead) glazed wares / pottery  
(NB: *se non altrimenti specificato, nell'ambito dell'archeologia islamica in contesto anglofono (e in modo crescente anche in Italia) l'invetriatura si intende implicitamente al piombo, poiché questo è il procedimento riscontrato in modo preponderante nei ritrovamenti a disposizione degli studiosi. Quando l'invetriatura è eseguita con altri materiali diversi dal piombo, l'elemento caratterizzante viene invece specificato. Es.: alkali glazed.*)
- **ceramica con rivestimento vetroso opacizzante monocromo**: (monochrome) opaque white-glazed wares  
(NB: *poiché i reperti di questo tipo sono solitamente monocromi, l'uso in inglese ritiene superflua questa osservazione e predilige invece l'indicazione del colore.*)
- **ceramica d'acqua**: water jugs
- **ceramica d'uso comune**: unglazed pottery
- **ceramica da cucina**: kitchen wares
- **ceramica di pasta artificiale**: stonepaste; fritware; siliceous stonepaste  
(NB: *la pasta artificiale in inglese è sempre 'stonepaste'; se nella composizione è presente il vetro, è diffusa anche la dizione 'fritware'.*)
- **ceramica di pasta artificiale con decorazione monocroma**: stonepaste with monochrome decoration; turquoise stonepaste; turquoise wares  
(NB: *questo tipo di ceramiche è caratterizzato dal colore azzurro, per cui sono spesso designate sinteticamente in inglese come 'turquoise wares / stonepaste'.*)
- **ceramica dipinta**: pottery with painted decoration
- **ceramica dipinta e incisa sotto vetrina al piombo**: pottery with underglaze painted and incised decoration
- **ceramica dipinta su ingobbio sotto vetrina trasparente**: slip-painted pottery / wares; slipware; lead glazed slipware  
(NB: *la dizione 'slip-painted' è largamente la più comune anche in contesti italofoeni, dal momento che questo tipo di decorazione è stato scoperto in ambito anglofono e la sua denominazione si è poi diffusa inalterata anche ad altri paesi.*)
- **ceramica dipinta su rivestimento vetroso opacizzante**: pottery with opaque glaze and painted decoration; pottery painted on an opaque white glaze
- **ceramica di tipo bianco crema importata dalla Cina**: white porcelain imported from China; Chinese white porcelain
- **ceramica in pasta silicea con colature in blu sotto rivestimento**

<p><b>alcalino:</b> stonepaste with splashed decoration under alkali glaze (NB: <i>le decorazioni a colatura, 'splashed', si sottintendono essere blu in inglese</i>)</p> <p>--- <b>ceramica in pasta silicea con decorazione policroma sotto rivestimento alcalino:</b> stonepaste with painted decoration under alkali glaze</p> <p>--- <b>ceramica in pasta silicea con rivestimento alcalino:</b> alkali glazed stonepaste; (<i>in presenza di vetro nella composizione</i>) fritware</p> <p>--- <b>ceramica invetriata:</b> glazed pottery</p> <p>--- <b>ceramica non invetriata:</b> unglazed pottery</p> <p>--- <b>ceramica non rivestita:</b> unglazed pottery</p> <p>--- <b>ceramica policroma:</b> pottery with polichrome decoration</p> <p>--- <b>ceramica rivestita:</b> glazed pottery</p> <p>--- <b>ceramica rivestita con vetrina al piombo:</b> (lead) glazed pottery</p> <p>--- <b>ceramiche dipinte in bruno / nero, verde, giallo su ingobbio bianco:</b> pottery with brown / black, green, yellow decoration on a white slip; pottery with brown / black, green, yellow decoration under a transparent glaze</p> <p>--- <b>ceramiche dipinte su ingobbio colorato:</b> pottery with painted decoration on a coloured slip</p>
<b>ceramista:</b> ceramist; potter
<b>ceramista decoratore:</b> potter painter
<b>cittadella:</b> citadel
<b>colature radiali:</b> radial splashes
<b>collezione:</b> collection
<b>collo a rocchetto:</b> ( <i>di coppa</i> ) waisted neck
<b>collocazione:</b> ( <i>di scavo o reperto</i> ) location
<b>colonna:</b> column
<b>colonnina:</b> small column
<b>coniazione:</b> mintage, minting
<b>consistenza:</b> ( <i>dell'argilla</i> ) stiffness
<b>contenitori:</b> ( <i>vasellame atto a contenere</i> ) containers, wares
<b>contraffatto:</b> counterfeit
<p><b>coperchio:</b> lid; cover</p> <p>--- <b>a medaglione:</b> medallion shaped lid</p>

<b>---</b> a <b>cupola</b> : dome-shaped cover / lid
<b>coppa</b> : bowl <b>---</b> <b>caremata</b> : carinated bowl <b>---</b> <b>con corpo sub-sferico</b> : bowl with bulbous sides <b>---</b> <b>da vino</b> : wine bowls <b>---</b> <b>con pareti oblique</b> : bowl with flaring sides <b>---</b> <b>con profilo ricurvo</b> : bowl with rounded sides
<b>copricapo</b> : headgear
<b>corda</b> : rope; <i>cordicella da sospensione</i> : suspension rope
<b>cornice</b> : frame <b>---</b> <b>continua</b> : circular continuous frame <b>---</b> <b>in rilievo</b> : frame in relief <b>---</b> <b>cornice incisa</b> : incised frame; engraved frame; carved frame (NB: la scelta tra 'incised', 'engraved' e 'carved' è determinata dalla relativa profondità dell'incisione, con progressiva maggiore penetrazione da inciso [incised] a intagliato [engraved] a scolpito [carved]. L'uso rivela spesso sovrapposizioni semantiche soprattutto tra 'incised' e 'engraved'.) <b>---</b> <b>concentriche</b> : concentric frames
<b>corpo figulino</b> : earthenware
<b>corpo subglobulare</b> : boubus body
<b>corsivo ceramico</b> : ceramic cursive
<b>cottura</b> : firing
<b>cresta a forma di tralcio</b> : vine-shaped crest
<b>crogiuolo</b> : crucible
<b>cromia</b> : palette
<b>cufico</b> : kufic (NB: alcuni ritengono che Kufic vada sempre scritto con la maiuscola, seguendo l'ipotesi che derivi dall'antica città di Kufa, ma il consenso non è unanime né sull'etimo né sull'ortografia; la dizione 'lettering', benché attestata in letteratura, è piuttosto rara.) <b>---</b> <b>annodato</b> : knotted kufic / cufic <b>---</b> <b>apicato</b> : apicated kufic / cufic; apexed kufic / cufic <b>---</b> <b>fiorito</b> : flowered kufic / cufic
<b>cuoriforme</b> : heart-shaped

<b>D</b>
<b>datare:</b> dating
<b>datazione:</b> dating
<b>dati documentari:</b> documentary data
<b>dati paleografici:</b> paleographic data
<b>decifrazione:</b> deciphering
<b>decorazione:</b> decoration --- <b>a stampo:</b> cast decoration --- <b>a bassorilievo:</b> bas-relief decoration --- <b>a settori:</b> decoration divided into panels --- <b>a traforo:</b> openwork --- <b>ageminata:</b> inlaid decoration --- <b>architettónica:</b> architectural decoration --- <b>bucherellata:</b> pitted decoration --- <b>dipinta:</b> painted decoration --- <b>epigrafica:</b> epigraphic decoration; inscription --- <b>geometrica:</b> geometric decoration --- <b>graffita:</b> sgraffito; <i>ceramica con decorazione graffita: sgraffito wares</i> --- <b>intagliata:</b> engraved decoration; carved decoration --- <b>intagliata a risparmio:</b> champlevé (NB: ' <i>champlevé</i> ' si utilizza solo in riferimento a materiali morbidi, come le paste ceramiche; per i materiali duri, come la pietra, la dizione corretta è ' <i>carved</i> '.) --- <b>pseudo-epigrafica:</b> pseudo-epigraphic decoration --- <b>radiale:</b> radial decoration
<b>di grande diffusione:</b> ubiquitous
<b>diaframma:</b> partition wall; <i>diaframma a tre archi: three-arch partition wall</i>
<b>dipinto:</b> painted
<b>disposto a vortice:</b> radially-positioned
<b>documento epigrafico:</b> epigraphic document

<b>E</b>
<b>edificio:</b> building
<b>egira:</b> hegira
<b>elemento:</b> element --- <b>decorativo:</b> decorative element --- <b>di forma globulare:</b> globular-shaped element --- <b>elemento in aggetto:</b> projecting element
<b>epigrafia:</b> epigraphy --- <b>monumentale:</b> monumental epigraphy
<b>epigrafico:</b> epigraphic
<b>epitaffio:</b> epitaph
<b>esemplare:</b> sample; specimen; item; <i>esemplari con rivestimenti vetrosi:</i> items of the glazed type

<b>F</b>
<b>facciata:</b> facade
<b>fascia:</b> band --- <b>epigrafica:</b> epigraphic band --- <b>puntinata:</b> dotted stripe --- <b>superiore:</b> upper band --- <b>inferiore:</b> lower band
<b>fascetta:</b> small band
<b>fattura:</b> craftsmanship
<b>feldspato:</b> feldspar
<b>fiasca del pellegrino:</b> pilgrim's flask
<b>figure ritorte:</b> winding figures
<b>filtri di brocche:</b> jug strainers
<b>fiore quadripetalo:</b> four-lobed flower
<b>fiorone:</b> big flower --- <b>dai petali circolari:</b> dots-petalled flower --- <b>stilizzato:</b> stylized flower
<b>foglia:</b> leaf --- <b>acquatica:</b> water weeds --- <b>trilobata:</b> trilobed leaf --- <b>'a giglio':</b> 'lily-shaped' leaf
<b>fonte:</b> source
<b>forma:</b> shape --- <b>'a coda':</b> ( <i>del versatoio</i> ) tail-shaped pouring lip --- <b>'ad ali':</b> ( <i>dell'ornato</i> ) wing-shaped decoration
<b>forno:</b> kiln
<b>frammentario:</b> fragmented
<b>frammento:</b> fragment; --- <b>frammenti ceramici:</b> ceramic fragments; potsherds
<b>fusto:</b> ( <i>di lampada</i> ) shaft

<b>G</b>
<b>galloni:</b> chevrons
<b>globetto:</b> globe
<b>grafemi:</b> graphemes
<b>grafia:</b> script; writing; lettering; handwriting --- <b>animata:</b> animated script --- <b>cufica:</b> kufic script; ( <i>raro</i> ) kufic lettering --- <b>minuta:</b> tiny script --- <b>ornamentale:</b> decorative script --- <b>fiorita:</b> floriate script
<b>guttiforme:</b> drop-shaped; <i>coda guttiforme:</i> drop-shaped tail



<b>I</b>
<b>impasto artificiale:</b> stonepaste; <i>(in presenza di vetro nella composizione) frit</i>
<b>in rilievo:</b> in relief
<b>incisione:</b> carving; carving technique; engraving;
<b>inciso:</b> incised; engraved; carved (NB: la scelta tra 'incised', 'engraved' e 'carved' è determinata dalla relativa profondità dell'incisione, con progressiva maggiore penetrazione da inciso [incised] a intagliato [engraved] a scolpito [carved]. L'uso rivela spesso sovrapposizioni semantiche soprattutto tra 'incised' e 'engraved'.)
<b>ingobbato:</b> slip-painted; engobed
<b>ingobbio:</b> slip
<b>intagliato:</b> engraved
<b>intaglio a risparmio (champlevé):</b> champlevé carving (NB: 'champlevé' si utilizza solo in riferimento a materiali morbidi, come le paste ceramiche; per i materiali duri, come la pietra, la dizione corretta è 'carved'. La dizione francese è comunque molto comune anche in contesto italiano)
<b>integro:</b> <i>(di reperto)</i> undamaged
<b>inventario:</b> inventory
<b>invetriato / a:</b> glazed
<b>invetriatura:</b> glaze; glazing --- <b>alcalina:</b> alkali glaze; <i>(più raro)</i> alkaline glaze
<b>invetriatura opaca:</b> opaque glaze; opaque glazing
<b>iscrizioni:</b> inscriptions
<b>iscrizioni monumentali:</b> monumental inscriptions

<b>L</b>
<b>labbro svasato:</b> flared lip
<b>lastra:</b> slab
<b>lega:</b> alloy
<b>leggibile:</b> legible
<b>linea di base:</b> baseline
<b>lobi:</b> lobes --- <b>concavi:</b> concave lobes --- <b>guttiformi:</b> drop-shaped lobes
<b>lucerna:</b> oil lamp --- <b>da sospensione:</b> hanging oil lamp --- <b>da tavolo:</b> standing oil lamps
<b>lustro:</b> lustre (BE); luster (AE) --- <b>lustro metallico:</b> (ceramica) lusterware; <i>produzione del lustro in Iran: Iranian lusterware</i> --- <b>lustro policromo:</b> polychrome lustre

<b>M</b>
<b>manufatto:</b> work; artefact (BE); artifact (AE)
<b>materiale ceramico:</b> pottery --- <b>privo di rivestimento:</b> unglazed wares
<b>matrice:</b> mould; matrix
<b>mattone:</b> brick --- <b>cotto:</b> baked brick
<b>medaglione:</b> medallion --- <b>perforato:</b> openwork medallion
<b>metallo:</b> ( <i>materia</i> ) metal; ( <i>oggetto in metallo</i> ) metalwork
<b>minareti:</b> minarets
<b>modanatura tortile:</b> spiral moulding
<b>momenti costruttivi:</b> construction phases
<b>moneta:</b> coin; coinage; <i>tipi monetari:</i> coinages
<b>monoansato:</b> one-handled
<b>monocromo:</b> monochrome
<b>monumentale:</b> monumental
<b>mortaio:</b> mortar
<b>motivo:</b> motif --- <b>a pennacchio:</b> panache motif --- <b>geometrico:</b> geometric motif --- <b>a corda:</b> rope motif --- <b>a macchie:</b> mottled motif; ( <i>di ceramica</i> ) splashed --- <b>a nodo:</b> knotted motif --- <b>a palmette:</b> palmette motif --- <b>ad arco:</b> arch-shaped motif --- <b>epigrafico:</b> epigraphic motif --- <b>floreale:</b> floral motif --- <b>in rilievo:</b> motif in relief
<b>murature:</b> stonework; masonry
<b>muro:</b> wall --- <b>di cinta:</b> bailey

<b>N</b>
<b>nastro:</b> ribbon --- <b>puntinato:</b> dotted ribbons
<b>nicchia:</b> niche; ( <i>in mattoni</i> ) brick niche
<b>nodo:</b> knot --- <b>quadrilobato:</b> quatrefoil knot
<b>numismatica:</b> numismatic

<b>O</b>
<b>opacizzante:</b> opaque
<b>opere murarie:</b> masonry; stonework
<b>orlo:</b> rim --- <b>arrotondato:</b> rounded rim --- <b>dentellato:</b> notched rim; scalloped rim; indented rim --- <b>liscio:</b> smooth rim --- <b>piatto ed estroflesso:</b> flat everted rim --- <b>rigonfio:</b> bulging rim --- <b>verticale:</b> vertical rim
<b>ornato:</b> ( <i>sostantivo</i> ) ornament; decoration; ( <i>verbo</i> ) decorated --- <b>con tratti e cerchi:</b> decorative motif of dashes and circles --- <b>geometrico:</b> geometric decoration --- <b>traforato:</b> openwork decoration <b>tipo di ---:</b> decorative motif
<b>ossidi:</b> oxide --- <b>di metallo:</b> metallic oxides --- <b>di piombo:</b> lead oxide

<b>P</b>
<b>palmetta:</b> palmetta --- <b>alata:</b> winged palmette motif --- <b>poilobata:</b> polylobed palmette <b>semi-palmette:</b> half palmettes
<b>paramento bugnato:</b> ashlar face
<b>parete:</b> wall; ( <i>di ceramica</i> ) side; --- <b>obliqua:</b> ( <i>di ceramica</i> ) flaring side --- <b>ricurva:</b> curving sides; <i>parete rovesciata esterna, leggermente ricurva: slightly curved and reversed external wall</i> --- <b>verticale:</b> vertical wall
<b>pasta artificiale:</b> stonepaste; frit; (NB: <i>in inglese la pasta artificiale si indica in modo generico come 'stonepaste'; se nella composizione è presente il vetro, è diffusa anche la dizione frit</i> )
<b>patina:</b> patina
<b>petalo:</b> petal --- <b>lanceolato:</b> lanceolate petal
<b>piattello:</b> ( <i>di lampada</i> ) small plate
<b>piede:</b> foot --- <b>a disco:</b> disc foot --- <b>ad anello:</b> ring foot --- <b>strombato:</b> splayed foot
<b>piriforme:</b> pear-shaped
<b>pittura su ingobbio:</b> slip painting
<b>portalampada:</b> lamp-stands
<b>privo di data:</b> undated
<b>privo di decorazione:</b> undecorated
<b>produzione fittile:</b> pottery
<b>profilo arrotondato:</b> ( <i>di spalla</i> ) round shoulder; ( <i>di muro</i> ) round wall --- <b>profilo svasato:</b> flaring walls
<b>programma epigrafico:</b> epigraphic program

<b>prospetto:</b> (di edificio) facade
<b>punti diacritici:</b> diacritical dots
<b>puntinate:</b> dotted

**Q**

**quota:** level; *quota di imposta della volta:* springing level of the vault



<b>R</b>
<b>rame:</b> copper --- <b>stagnato:</b> tinned copper
<b>reperto:</b> finding; find
<b>repertorio:</b> repertoire; repertory --- <b>decorativo:</b> decorative repertoire / repertory --- <b>floreale:</b> repertoire / repertory of floral motifs --- <b>geometrico:</b> repertoire / repertory of geometric motifs --- <b>ornamentale:</b> ornamental repertory
<b>restauro:</b> restoration
<b>resti:</b> remains
<b>reticolo:</b> grid --- <b>puntinato:</b> dotted grid
<b>retticurvilineo:</b> curvirectilinear
<b>ricostruzione:</b> reconstruction
<b>rinvenimento:</b> finding; find ( <i>usati di solito nelle forme plurali: findings, finds</i> )
<b>rivestimento:</b> ( <i>di parete</i> ) wall facing; ( <i>di ceramica</i> ) glaze --- <b>alkalino:</b> alkali glaze; glazing --- <b>alcalino colorato:</b> coloured alkali-glaze slip --- <b>alcalino trasparente:</b> transparent alkali glaze --- <b>vetroso al piombo:</b> lead glaze
<b>rosetta:</b> small rose
<b>rovine:</b> ruins

<b>S</b>
<b>sagoma:</b> outline
<b>sagomato:</b> outlined
<b>sala:</b> room --- <b>a cupola:</b> domed room
<b>sali alcalini:</b> alkali salts
<b>scanalature:</b> grooves
<b>scansione:</b> subdivision; <i>scansione ritmica delle superfici: even subdivision of the surfaces</i>
<b>scavato:</b> excavated
<b>scavo:</b> excavation
<b>schema decorativo:</b> pattern
<b>scodella:</b> bowl
<b>scolpito in rilievo:</b> carved in relief
<b>scrittura:</b> writing; script; inscription --- <b>corsiva:</b> cursive --- <b>incisa:</b> engraved inscription; incised inscription --- <b>scolpita:</b> carved writing; <i>scrittura scolpita in leggero rilievo: inscription carved in low relief</i> <b>stile di ---:</b> writing style; lettering
<b>semi-palmette:</b> half palmettes
<b>settore:</b> ( <i>con riferimento alle aree in cui è suddivisa e organizzata una decorazione</i> ) panel --- <b>radiali:</b> radial panels --- <b>triangolari:</b> triangular panels
<b>sfondo:</b> background
<b>sigillo:</b> seal
<b>sito:</b> site; <i>sito archeologico: archaeological site</i>
<b>smerlato:</b> scalloped
<b>smussato:</b> chamfered
<b>spalla:</b> (di coppa) shoulder; <i>coppa con breve spalla leggermente</i>

<i>arrotondata</i> : bowl with bulbous sides; bowl with short, slightly rounded shoulder
<b>spazio</b> : space
<b>spiraliforme</b> : spiralling; <i>decorazione spiraliforme</i> : spiralling decoration
<b>squadrato</b> : squared
<b>stagno</b> : tin
<b>stampato</b> : printed
<b>stato di conservazione</b> : state of preservation
<b>stela</b> : stela; <i>pl.</i> stelae, steles --- <b>funerarie</b> : funerary stelae / steles
<b>stilizzato</b> : stylized
<b>stratigrafico</b> : stratigraphic
<b>superficie</b> : surface
<b>svasato</b> : flared

<b>T</b>
<b>terme:</b> thermal baths
<b>terminazione:</b> ending --- <b>trilobata:</b> trilobated ending
<b>testimonianza:</b> testimony; witness; <i>testimonianze archeologiche:</i> <i>archaeological testimonies / witness</i>
<b>testo benaugurale:</b> well wishing text; text expressing good wishes
<b>testo frammentario:</b> fragmented text
<b>tonalità:</b> shades
<b>tono:</b> ( <i>di colore</i> ) palette
<b>tornio:</b> wheel; <i>ceramiche eseguite al tornio:</i> <i>wheel thrown pottery</i>
<b>traccia:</b> trace
<b>traforo:</b> openwork
<b>tralcio:</b> vine; vine-scroll; spray; <i>decorazione con tralcio pendente:</i> <i>decoration with hanging vine-branch; tralcio sinusoidale con</i> <i>semipalmette:</i> <i>sinusoidal spray with half-palmettes</i> --- <b>fogliati:</b> leafed vines --- <b>spiraliforme:</b> vine-scroll; vine scrollwork
<b>transenna:</b> transennas
<b>tratteggio:</b> dashes
<b>treccia a due capi:</b> two-strand braided motif
<b>trilobati:</b> trilobated; trilobed

<b>V</b>
<b>vano:</b> hall --- <b>di accesso:</b> entrance hall
<b>vasellame:</b> ware; vessels; pottery --- <b>da cucina:</b> kitchenware
<b>verde di rame:</b> copper oxide; copper green
<b>versatoio:</b> pouring lip; spout; <i>versatoio orizzontale lungo e stretto:</i> <i>long and narrow horizontal pouring lip</i>
<b>vetrificazione:</b> vitrification; glazing
<b>vetrina:</b> glaze --- <b>al piombo:</b> lead glaze --- <b>alcalina:</b> alkaline glaze; alkali glaze --- <b>opacizzante:</b> opaque glaze

<b>Z</b>
<b>zecca:</b> mint
<b>zircone:</b> zircon
<b>zoomorfo:</b> zoomorphic

## REFERENCES

- Ahmad, Khurshid. "Being in Text and Text in Being: Notes on Representative Texts." Anderman and Rogers 60-94.
- Aijmer, Karin, and Bengt Altenberg, eds. *Advances in Corpus Linguistics. Papers from the 23rd International Conference on English Language Research on Computerized Corpora (ICAME 23)*. Göteborg 22-26 May 2002. Amsterdam: Rodopi, 2004. Print.
- Aijmer, Karin, Bengt Altenberg, and Mats Johansson, eds. *Languages in Contrast: Papers from a Symposium on Text-Based Cross-Linguistic Studies*. Lund, 4-5 March 1994. Lund: Lund UP, 1996. Print.
- Allwood, Jens. "Multimodal Corpora." *Corpus Linguistics. An International Handbook*. Ed. Anke Lüdeling and Merja Kytö. Berlin, New York: Walter de Gruyter, 2008. 207-25. Print.
- Alsina, Victòria, and Janet DeCesaris. "Bilingual Lexicography, Overlapping Polysemy, and Corpus Use." Altenberg and Granger 215-29.
- Altenberg, Bengt, and Sylviane Granger, eds. *Lexis in Contrast. Corpus-Based Approaches*. Amsterdam, Philadelphia: Benjamins, 2002. Print.
- . "Recent Trends in Cross-Linguistic Lexical Studies." Altenberg and Granger 3-48.
- Anderman, Gunilla, and Margaret Rogers, eds. *Incorporating Corpora. The Linguist and the Translator*. Clevedon, Buffalo, Toronto: Multilingual Matters, 2008. Print.
- . "The Linguist and the Translator." Anderman and Rogers 5-17.
- Assis Rosa, Alexandra. "Descriptive Translation Studies (DTS)." Gambler and van Doorslaer 94-104.
- Austermühl, Frank. "Using Concept Mapping and the Web as Corpus to Develop Terminological Competence among Translators and Interpreters." *Translation Spaces* 1 (2012): 54-79. Print.

- Baker, Mona. "Corpora in Translation Studies – An Overview and Some Suggestions for Future Research." *Target* 7 (1995): 223-43. Print.
- Baker, Paul, Andrew Hardie, and Tony McEnery, eds. *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh UP, 2006. Print.
- Baldry, Anthony. "The Role of Multimodal Concordancers in Multimodal Corpus Linguistics." *New Directions in The Analysis of Multimodal Discourse*. Ed. Terry Royce and Wendy Bowcher. Mahwah, NJ: Erlbaum, 2007. 173-93. Print.
- Baldry, Anthony, and Paul J. Thibault. "Multimodal Corpus Linguistics." *System and Corpus: Exploring Connections*. Ed. Geoff Thompson and Susan Hunston. London: Equinox, 2006. 164-83. Print.
- Beeby, Allison, Patricia Rodríguez Inés, and Pilar Sánchez-Gijón, eds. *Corpus Use and Translating. Corpus Use for Learning to Translate and Learning Corpus Use to Translate*. Amsterdam, Philadelphia: Benjamins, 2009. Print
- Beninato, Renato S., and Donald A. DePalma. "Ranking of Top 25 Translation Companies." *Commons Sense Advisory* 29 May 2008. Web. 20 May 2015.  
[http://www.commonsenseadvisory.com/Research/All\\_Users/080528\\_QT\\_2008\\_top\\_25\\_lsps/tabid/1492/Default.aspx](http://www.commonsenseadvisory.com/Research/All_Users/080528_QT_2008_top_25_lsps/tabid/1492/Default.aspx)
- Bowker, Lynne. "Using Specialised Native-Language Corpora as a Translation Resource: A Pilot Study." *Meta: journal des traducteurs/Meta: Translators' Journal* 43.4 (1998): 631-51. Print.
- Bowker, Lynne, and Des Fisher. "Computer-aided Translation." *Gambler and van Doorslaer* 60-5.
- Bowker, Lynne, and Jennifer Pearson. *Working with Specialized Language. A Practical Guide to Using Corpora*. London, New York: Routledge, 2002. Print.
- Boyd, Brian. "Some Comments Inspired by Valter Lang's 'Archaeology and Language.'" *Fennoscandia Archaeologica* XVII (2000): 115-7. Print.
- Cabré, M. Teresa. "Terminology and Translation." *Gambler and van Doorslaer* 356-65.
- Charteris-Black, Jonathan. "Why 'An Angel Rides in the Whirlwind and Directs the Storm': A Corpus-Based Comparative Study of Metaphor in British and American Political Discourse." *Aijmer and Altenberg* 133-50.
- Chodkiewicz, Christine, Didier Bourigault, and John Humbley. "Making a Workable Glossary out of a Specialised Corpus: Term Extraction and Expert Knowledge." *Altenberg and Granger* 249-67.



- Claridge, Claudia. "Constructing a Corpus from the Web: Message Boards." Hundt, Nesselhauf and Biewer 87-108.
- Corpas Pastor, Gloria. "Compilacion de un corpus *ad hoc* para la enseñanza de la traducción inversa especializada." *Trans: Revista de Traductología* 5 (2001): 155–184. Print.
- . 2004. "Localización de recursos y compilación de corpus via Internet: Aplicaciones para la didáctica de la traducción médica especializada." *Manual de documentación y terminología para la traducción especializada*, Gonzalo García, Ed. Consuelo and Valentin García Yebra. Madrid: Arco/Libros. 223–257. Print.
- Corpas Pastor, Gloria, and Miriam Seghiri. "Virtual Corpora as Documentation Resources: Translating Travel Insurance Documents (English-Spanish)". Beeby, Rodríguez Inés and Sánchez-Gijón 75-107.
- Cronin, Michael. "Globalization and Translation." Gambier and van Doorslaer 131-40.
- D'Angelo, Larissa. "Creating a Corpus for the Analysis of Identity Traits in English Specialized Discourse." Torsello, Ackerley and Castello 199-210.
- Diemer, Stefan. "Corpus Linguistics with Google?" *Proceedings of ISLE 2 Boston*. 2011. Web. 27 March 2014.  
<[www.bu.edu/isle/files/2012/01/Stefan-Diemer-Corpus-Linguistics-with-Google.pdf](http://www.bu.edu/isle/files/2012/01/Stefan-Diemer-Corpus-Linguistics-with-Google.pdf)>
- Dyvik, Helge. "Translations as Semantic Mirrors: From Parallel Corpus to Wordnet." Aijmer and Altenberg 311-25.
- Facchinetti, Roberta, ed. *Corpus Linguistics 25 Years On*. Amsterdam, New York: Rodopi, 2007. Print.
- Fletcher, William H. "Facilitating the Compilation and Dissemination of Ad-Hoc Web Corpora." *The Fifth International Conference on Teaching and Language Corpora*. 2004. Web. 25 January 2015.  
<[http://www.kwicfinder.com/Facilitating\\_Compilation\\_and\\_Dissemination\\_of\\_Ad-Hoc\\_Web\\_Corpora.pdf](http://www.kwicfinder.com/Facilitating_Compilation_and_Dissemination_of_Ad-Hoc_Web_Corpora.pdf)> 2004.
- Folaron, Deborah. "Networking and Volunteer Translators." Gambler and van Doorslaer 231-4.
- Frankenberg-Garcia, Ana. "Are Translations Longer than Source Texts? A Corpus-Based Study of Explicitation." Beeby, Rodríguez Inés and Sánchez-Gijón 47-58.
- . "Compiling and Using a Parallel Corpus for Research in Translation." *International Journal of Translation* 21.1-2 (2009): 57-71. Print.

- Gambier, Yves, and Luc van Doorslaer, eds. *Handbook of Translation Studies*. Amsterdam, Philadelphia: Benjamins, 2010. Print.
- Gatto, Maristella. "The 'Body' and the 'Web': The Web as Corpus Ten Years On." *ICAME Journal* 35 (2011): 35-58. Print.
- Gazzola, Michele. "Managing Multilingualism in the European Union: Language Policy Evaluation for the European Parliament." *Language Policy* 5.4 (2006): 395-419. Print.
- Gellerstam, Martin. "Translationese in Swedish Novels Translated from English." Wollin and Lindquist 88-95.
- Granger, Sylviane. "From CA to CIA and back: An Integrated Approach to Computerised Bilingual and Learner Corpora." Aijmer, Altenberg and Johansson 38-51.
- . "The International Corpus of Learner English: A New Resource for Foreign Language Learning and Teaching and Second Language Acquisition Research." *TESOL Quarterly* 37.3 (2003): 538-46. Print.
- Grzega, Joachim. "Reflections on Concepts of English for Europe. British English, American English, Euro-English, Global English." *Journal for Euro-Linguistics* 2 (2005): 44-64. Print.
- Halliday, M.A.K., Wolfgang Teubert, Colin Yallop, and Anna Cermakova, *Lexicology and Corpus Linguistics. An Introduction*. London, New York: Continuum, 2004. Print.
- Halliday, M.A.K. "Towards a theory of good translation." *Exploring Translation and Multilingual Text Production: Beyond Content*. Ed. Erich Steiner and Colin Yallop. Berlin and New York: Mouton de Gruyter 2001. 13-18. Print.
- Hundt, Marianne, Nadja Nesselhauf, and Carolin Biewer, eds. *Corpus Linguistics and the Web*. Amsterdam, New York: Rodopi, 2007. Print.
- Hunston, Susan. *Corpora in Applied Linguistics*. Cambridge: Cambridge UP, 2002. Print.
- Joyce, Rosemary A., Robert W. Preucel, Jeanne Lopiparo, Carolyn Guyer, and Michael Joyce. *The Languages of Archaeology. Dialogue, Narrative, and Writing*. Oxford: Blackwell, 2002. Print.
- Kenny, Dorothy. "Corpora in translation studies." *Routledge Encyclopedia of Translation Studies* Ed. Mona Baker (assisted by K. Malmkjar). London: Routledge 2001. 50-53. Print.
- Kilgarriff, Adam, and Gregory Grefenstette. "Introduction to the Special Issue on the Web as Corpus." *Computational Linguistics* 29.3 (2003): 333-47. Print.

- Kiraly, Don. *A Social Constructivist Approach to Translator Education: Empowerment from Theory to Practice*. Manchester: St. Jerome. 2000. Print.
- Koskinen, Kaisa. "Institutional Illusions: Translating in the EU Commission." *The Translator* 6 (2000): 49-66. Print.
- Kraif, Olivier. "Translation Alignment and Lexical Correspondences. A Methodological Reflection." Altenberg and Granger 271-89.
- Landolfi, Liliana. *P.Æ.C.E. An Italian-English Corpus Based on EFL Students*. Napoli: Liguori, 2012. Print.
- Lang, Valter. "Archaeology and Language." *Fennoscandia Archaeologica* XVII (2000): 103-10. Print.
- Laviosa, Sara. "Corpora." Gambier and van Doorslaer 80-6.
- . *Corpus-Based Translation Studies. Theory, Findings, Applications*. Amsterdam, New York: Rodopi, 2002. Print.
- Leech, Geoffrey. "Corpora and Theories of Linguistic Performance." *Directions in Corpus Linguistics: Proceedings of the Nobel Symposium 82. Stockholm, 4–8 August 1991*. Ed. Jan Svartvik. Berlin: Mouton de Gruyter, 1992. 105-22. Print.
- . "New Resources, or Just Better Old Ones? The Holy Grail of Representativeness." Hundt, Nesselhauf and Biewer 133-49.
- Lüdeling, Anke, and Merja Kytö, eds. *Corpus Linguistics. An International Handbook*. Berlin, New York: Walter de Gruyter, 2008. Print.
- Lüdeling, Anke, Stefan Evert, and Marco Baroni. "Using Web Data for Linguistic Purposes." Hundt, Nesselhauf and Biewer 7-24.
- Malmkjar, Kirsten. "Norms and Nature in Translation Studies." Anderman and Rogers 49-59.
- Mauranen, Anna. "Universal Tendencies in Translation." Anderman and Rogers 32-48.
- McEnery, Tony, and Andrew Hardie. *Corpus Linguistics. Method, Theory and Practice*. Cambridge: Cambridge UP, 2012. Print.
- McEnery, Tony, and Andrew Wilson. *Corpus Linguistics*. Edinburgh: Edinburgh UP, 1996. Print.
- Meyer, Charles F. *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge UP, 2004. Print.
- McEnery, Tony, and Richard Xiao, "Parallel and Comparable Corpora: What is Happening?" Anderman and Rogers 18-31.
- Meyer, Charles F. *English Corpus Linguistics. An Introduction*. Cambridge: Cambridge UP, 2004. Print.

- Milizia, Denise. *Phraseology in Political Discourse. A Corpus Linguistics Approach in the Classroom*. Milano: LED, 2014. Print.
- Murphy, Amanda C. *Editing Specialised Texts in English. A Corpus-Assisted Analysis*. 2nd ed. Milano: LED, 2012. Print.
- Pym, Anthony. "Redefining Translation Competence in an Electronic Age. In Defence of a Minimalist Approach." *Meta: journal des traducteurs/Meta: Translators' Journal* 48.4 (2003): 481-97. Web. 19 July 2014. <[http:// id.erudit.org/iderudit/008533ar](http://id.erudit.org/iderudit/008533ar)>
- Rathje, William L., Michael Shanks, and Christopher Witmore, eds. *Archaeology in the Making. Conversations through a Discipline*. London, New York: Routledge, 2013. Print.
- Renouf, Antoinette, "Corpus Development 25 Years On: From Super-Corpus to Cyber-Corpus." Facchinetti 27-49.
- Renouf, Antoinette, Andrew Kehoe, and David Mezquiriz. "The Accidental Corpus: Some Issues in Extracting Linguistic Information from the Web." Aijmer and Altenberg 403-19.
- Resnik, Philip, and Noah A. Smith. "The Web as a Parallel Corpus." *Computational Linguistics* 29.3 (2003): 349-80. Print.
- Riccardi, Alessandra, ed. *Translation Studies. Perspectives on an Emerging Discipline*. Cambridge: Cambridge UP, 2002. Print.
- Robb, Thomas. "Google as a Quick 'n Dirty Corpus Tool." *The Electronic Journal for English as a Second Language* 7:2 (2003). Web. 14 January 2014. <[http:// www.teslej.org/wordpress/issues/volume7/ej26/ej26int/](http://www.teslej.org/wordpress/issues/volume7/ej26/ej26int/)>
- Salkie, Raphael. "Two Types of Translation Equivalence." Altenberg and Granger 51-71.
- Sánchez-Gijón, Pilar. "Developing Documentation Skills to Build Do-It-Yourself Corpora in the Specialised Translation Course." Beeby, Rodríguez Inés and Sánchez-Gijón 109-27.
- . "És la web pública la nova biblioteca del traductor?" *Tradumàtica: Traducció i tecnologies de la informació i la comunicació* 2 (2003): 1-7. Web. 10 April 2015. <http://www.raco.cat/index.php/Tradumatica/article/view/29147/28982>
- Scarpa, Federica. "Closer and Closer Apart? Specialized Translation in a Cognitive Perspective." Riccardi 133-49.
- Scott, Juliette. "Towards Professional Uptake of DIY Electronic Corpora in Legal Genres." *Salford Working Papers in Translation and Interpreting* 1 (2012): 1-22. Print.

- Sinclair, John. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press, 1991. Print.
- Snell-Hornby, Mary. "The Turns of Translation Studies." Gambler and van Doorslaer 366-70.
- Suhonen, Mervi. "Comment on Valter Lang's Article Archaeology and Language." *Fennoscandia Archaeologica* XVII (2000): 118-21. Print.
- Svartvik, Jan. "Corpus Linguistics 25+ Years On." Facchinetti 11-25.
- . "Edited Extract from 'A Life in Linguistics.'" *The European English Messenger* 14.1 (2005): 34-44. Print.
- Teubert, Wolfgang. "The Role of Parallel Corpora in Translation and Multilingual Lexicography." Altenberg and Granger 189-214.
- Tognini-Bonelli, Elena. *Corpus Linguistics at Work*. Amsterdam, Philadelphia: Benjamins, 2001. Print.
- . "Functionally Complete Units of Meaning across English and Italian: Towards a Corpus-Driven Approach." Altenberg and Granger 73-95.
- Tognini-Bonelli, Elena, and Elena Manca. "Welcoming Children, Pets and Guests: Towards Functional Equivalence in the Languages of 'Agriturismo' and 'Farmhouse Holidays.'" Aijmer and Altenberg 371-85.
- Torsello, Carol Taylor, Katherine Ackerley, and Erik Castello, eds. *Corpora for University Language Teachers*. Bern: Peter Lang, 2008. Print.
- Ulrych, Margherita. *Focus on the Translator in a Multidisciplinary Perspective*. Padova: Unipress, 1999. Print.
- van der Meer, Jaap. "At Last Translation Automation Becomes a Reality: An Anthology of the Translation Market." *EAMT/CLAW Conference*. Web. 9 May 2014. <<http://www.mt-archive.info/CLT-2003-VanDerMeer.pdf>>
- Viana, Vander, Sonia Zyngier, and Geoff Barnbrook, eds. *Perspectives on Corpus Linguistics*. Amsterdam, Philadelphia: Benjamins, 2011. Print.
- Varantola, Krista. "Translators, dictionaries and text corpora." *I corpora nella didattica della Traduzione*. Ed. Silvia Bernardini and Federico Zanettin. Bologna: CLUEB, 1997. 117-133. Print.
- . "Disposable Corpora as Intelligent Tools in Translation." *Cadernos de Tradução IX – Tradução e Corpora* 1.9 (2002): 171-89. Print.

- Wittenburg, Peter. "Preprocessing Multimodal Corpora." Lüdeling and Kytö 664-85.
- Wollin, Lars, and Hans Lindquist, eds. *Translation Studies in Scandinavia. Proceedings from the Scandinavian Symposium on Translation Theory (SSOTT) II, Lund 14-15 June, 1985*. Lund: CWK Gleerup, 1986. Print.
- Zanettin, Federico. "DIY Corpora: The WWW and the Translator." *Training the Language Services Provider for the New Millennium*. Ed. Belinda Maia, Jonathan Haller, and Margherita Urlrych. Porto: Faculdade de Letras, Universidade do Porto 2002. 239-48. Print.
- \_\_\_\_\_. "Translation and Corpus Design." *SYNAPS – A Journal of Professional Communication* 26 (2011): 14-23. Print.
- Zanettin, Federico, Silvia Bernadini, and Dominic Stewart, eds. *Corpora in Translator Education*. Manchester: St. Jerome, 2003. Print.
- Zimmer, Ben. "Bigger, Better Google Ngrams: Brace Yourself for the Power of Grammar." *The Atlantic*. 18 Oct. 2012. Web. 18 July 2014.
- <<http://www.theatlantic.com/technology/archive/2012/10/bigger-better-google-ngrams-braceyourself-for-the-power-of-grammar/263487/>>



*Volumi dal catalogo LED:*

- M.C. Rizzardi - M. Barsi • *Metodi in classe per insegnare la lingua straniera* • con DVD  
I. Teorie Applicazioni Materiali - II. Lezioni dal vivo • 2 voll.
- Studi sul portfolio europeo delle lingue digitale. Contributi alla giornata di studio. Università degli Studi di Milano, 28 gennaio 2005* • A cura di E. Landone • e-book
- R. Salvi • *Global English. A Global Debate*
- R. Salvi - J. Bowker - E. Incelli, A. Pontesilli • *English in Context. Exploration in a Grammar of Discourse*
- M.G. Guido • *Mediating Cultures. A Cognitive Approach to English Discourse for the Social Science*
- L. Lombardo - L. Haarman - J. Morley - C. Taylor • *Massed Medias: Linguistic Tools for Interpreting Media Discourse*
- G. Mansfield • *Changing Channels. Media Language in (Inter)action*
- A.C. Derouaux • *Guidebook to Translating from Italian into English*
- G. Garzone - F. Miglioli • *From Culture and Institutions to Language*
- G. Barbieri - L. Codeluppi • *From Business and Economics to Language*
- L. Clegg - S. Garton • *Reading for Research in Economics and Business*
- L. Clegg • *From Farm to Fork. English for Food Sciences*
- G. Alessi - S. Garton • *Reading for Research in Psychology*
- A. Partington - C. Taylor • *Persuasion in Politics. A Textbook. II Edition*
- C.M. de B. Clark • *Views in the News. A Textbook*
- A. Murphy • *Editing Specialised Texts in English. A Corpus-assisted Analysis. II Edition*
- S. Campagna - C. Boggio • *Multimodal Business and Economics*
- D. Milizia • *Phraseology in Political Discourse. A Corpus Linguistics Approach in the Classroom*
- C. Cucchi • *Non-native English within the European Parliament*
- S. Anselmi • *On Self-translation. An Exploration in Self-translators' Teloi and Strategies*
- G. Garzone • *Le traduzioni come fuzzy set. Percorsi teorici e applicativi*
- G. Fusco • *Telling Findings. Translating Islamic Archaeology through Corpora*

---

*Lingue Culture Mediazioni / Languages Cultures Mediation* - <http://www.ledonline.it/LCM-Journal>

*Linguae & - Rivista di Lingue e Culture Moderne* - <http://www.ledonline.it/linguae>

Il catalogo aggiornato di LED - Edizioni Universitarie di Lettere Economia Diritto è consultabile all'indirizzo web <http://www.lededizioni.com>, dove si possono trovare anche informazioni dettagliate sui volumi sopra citati: di tutti si può consultare il sommario, di alcuni vengono date un certo numero di pagine in lettura, di altri è disponibile il testo integrale. Tutti i volumi possono essere ordinati on line.