

Corpus in “Natural” Language Versus “Translation” Language: LBC Corpora, A Tool for Bilingual Lexicographic Writing



Annick Farina and Riccardo Billero

Abstract The aim of this paper is to describe the work done to exploit the LBC database for the purpose of translation analysis as a resource to edit the bilingual lexical sections of our dictionaries of cultural heritage (in nine languages). This database, made up of nine corresponding corpora, contains texts whose subject is cultural heritage, ranging from technical texts on art history to books on art appreciation, such as tour guides, and travel books highlighting Italian art and culture. We will illustrate the different questions with the SketchEngine LBC French corpus, made up at the moment of 3,000,000 words. Our particular interest here is in research that not only orients lexical choices for translators but that also precedes the selection of bilingual quotations (from our Italian/French parallel corpus) and that we rely on for editing an optional element of the file called “translation notes.” We will rely on this as much for works on “universals of translation” already described by Baker (Corpus linguistics and translation studies. Implications and applications. In Baker M et al (eds) *Text and technology*. Benjamins, Amsterdam/Philadelphia, pp 233–250 (1993)) as for studies aimed at improving translation quality assessment (TQA). We will show how a targeted consultation of different corpora and subcorpora that the database allows us to distinguish (“natural language” vs “translation,” “technical texts” vs “popularization texts” or “literary texts”) can help us identify approximations or translation errors, so as to build quality comparative lexicographical information.

Riccardo Billero wrote parts 1 and 2 of this article: Parts 3, 4, and 5 were written by Annick Farina. The IT support for the project (the database and the statistical results) was produced by Riccardo Billero.

A. Farina (✉) · R. Billero (✉)
Università Degli Studi di Firenze, Florence, Italy
e-mail: annick.farina@unifi.it

R. Billero
e-mail: riccardo.billero@gmail.com

1 Introduction

The *Lessico dei Beni Culturali* project and the research group by the same name have existed since 2013 within the Educational Department of Languages, Interculture and Psychology (FORLILPSI) of the University of Florence, with the following objectives:

- to create a reference platform for the use and study of the heritage lexicon;
- to create text corpora for all the languages of the project;
- to provide tools for consulting these corpora to enable and promote research in different fields such as linguistics and the Humanities in general;
- to offer access to corpora to all types of users, by offering a dialogue with all those who want to contribute in a constructive way to enriching the platform;
- to write multilingual dictionaries of art and heritage words using corpora as a resource.

This work reflects the wide variety of cultural and scientific interests of the members of our research unit (specialists in linguistics, lexicography, computational linguistics, translation, literatures from the different countries affected by the project, history of art, etc.) all of which converge on the study of the heritage lexicon in a broad spectrum perspective that favors opening up to the world and its diversity.

The ultimate goal of the *Lessico dei Beni Culturali* project is to create monolingual dictionaries of nine different languages according for the specific use of describing (and translating descriptions) the Tuscan heritage mainly in texts for popular utilization (tourist guides, museum sites, etc.). To do this, we have created textual databases, which we are completing as a work in progress,¹ which will be used for various tasks, from creating lists of entries to drafting monolingual lexicographic/terminological files and translation files linking the entries of the different languages to each other (for the description of these databases see Billero et al. [3]).

It is the use of these textual databases for the drafting of bilingual translation files that we will illustrate here,² based on the analysis of different queries on NoSketchEngine (mainly statistics and context) from our French LBC corpus, currently composed of approximately three million words. In particular, we will compare data from several comparable subcorpora of texts “in natural language” and texts “in translation.” We will also propose a first comparison of results from a subset of the Italian corpus with a subset containing the French translations of the same texts, which constitute fragmentary material for the moment because we are still working on the insertion of the texts with the aim of creating parallel bases of translation from Italian into all the languages of the project.

¹Consultable at <https://corpora.lessicobeniculturali.net/>.

²Apropos of using our databases to create monolingual dictionaries, see the article by Nicolás and Lanini in this volume. We are currently building the list of entries of the different languages following the model they have defined for Italian. The bilingual link between these different entries will only be possible once we have established our parallel databases.

We will show how a targeted consultation of the various corpora and subcorpora that the database allows us to distinguish (“natural language” Italian vs “natural language” French, French “in translation” vs “natural language” French, “specialized texts” French vs “popularization” French vs “literary” French) can help us identify approximations or translation errors, and hence provide highly accurate comparative lexicographic information.

2 The LBC Corpora

One of the main resources of the LBC Research Unit is the LBC corpora. Begun in 2016, they have been designed from the start as a tool to be used not only by the group for writing dictionaries but that should also be accessible for other types of uses and users, as a work tool for translators or as a documentary source for specialists in art history or literature.

For this reason, in designing the corpus and corpus interrogation model for extensive use by specialists, we have defined three basic needs:

1. to furnish useful linguistic material to create and serve as a document base for writing multilingual entries in LBC dictionaries;
2. to furnish linguistic material that can serve as a basis for linguistic, literary, or more broadly cultural studies;
3. to furnish useful textual material for spreading knowledge of the Tuscan heritage to the general public.

In creating the LBC corpora, we have therefore chosen texts that have as their main subject Italian artistic heritage and, more in particular, that convey a perspective “from without” on Florence and Tuscany, and descriptions that are together objective and personal. We have since started to broaden the type of texts, integrating works on Italy.

Some of the texts that appear in our corpora are:

- Biographies, for example, Benvenuto Cellini’s autobiography (*La vita* and its translations in the different languages of the project) or Dumas’ *Une année à Florence*;
- Art texts or having art as their subject, for example, Vasari (*Le Vite* and its translations) or Alarcón’s *Vidas de los pintores y estatuarios*;
- Novels, for example, Forster’s *A Room with a View* or Fernandez’ *Dans la main de l’Ange*;
- Essays, for example, Feuillet’s *L’art italien* or Burckhardt’s *Geschichte der Renaissance*;
- Tourist guides such as the *Routard Guide* or *GeoGuide*;
- Specialized dictionaries, such as Viollet Le Duc’s *Dictionnaire de l’architecture*.

These texts originate from different eras, from the Renaissance to the present day. For each text, both its dates of composition and its year of publication, of

translation when applicable, and publication of the translation are recorded. For us, the most important dates are those of its composition, as regards “natural” language texts, and that of the translation, as regards translation texts. What interests us is the representative character of these texts in relation to their use of language and not for the type of discourse pertaining to heritage. However, we also leave the possibility for users of the corpora to search with criteria other than ours.

The different texts were collected and underwent a work process illustrated in detail in Billero [2]: these texts, accompanied by bibliographic type metadata, have been integrated into the different corpora published online through NoSketchEngine.

Because of the use we want to make of corpora for our project, we needed the proper software for processing and analyzing data from our corpus. Software belonging to this category can be of two types: what can only be installed on a computer and what can be accessed from an online service (and so available on the internet).

A research on the most recent techniques allowed us to identify three of the main software types that interested us—and which therefore also allowed us to make our corpus accessible to persons outside the group: CWB (IMS Open Corpus Work-Bench), INL BlackLab, and SketchEngine. While the first two are available free of charge, SketchEngine must normally be purchased, though it is free for members of European research institutions. There is also a free version, NoSketchEngine, which lacks some of the commercial version functions, which are mainly useful for lexicographic studies such as thesauruses and combinatorial word sketch maps.

All of the above-mentioned software has a number of basic functions that are fundamental for querying, allowing both simple searches and searches using regular expressions or the “Corpus Query Language” (CQL acronym), a query language defined according to graphic or lemmatized forms or grammatical categories, created at the base for CWB, to then be integrated into the other software mentioned. They also allow the use of lemmatization tools such as TreeTagger, Freeling, RFTagger ... These three software programs also have a programming interface app for software-to-software communication (API) that allows them to perform query operations or introduce data through additional software created specifically for the purposes of such research.

CWB and SketchEngine by now enjoy excellent documentation and decent recognition in the university community. This is not the case with BlackLab, which has shortcomings in this regard.

SketchEngine offers some combinatorial word functions that are very useful for creating lists of specific terms of the various arts and that can be used to relate words that refer to the same art objects, to delimit areas of knowledge or references, for example, to monuments. They can also be used to differentiate objects from people or to disambiguate proper names that refer to these two categories. However, these possibilities are not available in the free NoSketchEngine version.

In the end, we chose SketchEngine as the corpus management software in our research activities, while we use NoSketchEngine to online consultation of our corpus.

3 Comparison Between “Natural Language” and “Translation” Corpus: A Perspective Halfway Between Descriptive and Prescriptive Translatology

We rely on analyses which do not consider the language of translation as a “third code” [6], in the belief that translated texts find their place perfectly within a single textual base of the same language, alongside texts “in natural language.” However, on the model of proposals for using a translation corpus for a didactic purpose, for teaching both languages and translation, it seems necessary to offer the possibility of consulting the database in separate subcorpora that group texts of both types and define criteria for evaluating translated texts to be integrated into the database. To do this, we have created separate subcorpora of texts translated into all the languages of the project. These corpora are useful to us as translation memory tools for working on the bilingual part of our lexicographical files in a perspective that is more prescriptive than descriptive.

As will be shown in comparing results from our French LBC database “in natural language” and “in translation” with a corpus of almost 100,000 words currently not integrated into our database (and so not available online) consisting of translations of “popular” works into French (in particular tourist guides of Tuscany and museum brochures), some of the texts that interest us present features similar to “translationese,” which could only distort queries of the base that aim at attesting forms or structures typical of French as it is written and spoken by most native speakers without interference from another language. We intend to define acceptability thresholds for texts “in translation” so as to introduce into our published corpus only texts that we will have recognized as similar to “natural language” texts, while limiting their proportion in our corpus so that the latter category remains preponderant.

3.1 Descriptive and Prescriptive Information in LBC Dictionaries: Universals and Deviations

Following Baker [1], we start from the principle that there are translation universals which will serve as a framework for illustrating the different types of queries carried out within our subcorpora and for comparing the results obtained. It is on these universals that we depend to provide the descriptive part of the detailed comparative lexicographic information present in the bilingual part of our dictionaries.³ This information corresponds first to the observation of parallel corpora, which certify the translation of the lemmas (words or collocations) described by the dictionary,

³For a more precise description of this comparative linguistic information inside our dictionaries, as we defined it in a first proposal for a drafting protocol which we intend to review during the year 2020, we refer to [4].

by appearing in bilingual quotations within the bilingual part of the article. We will analyze in particular:

- the simplification (mainly, as regards our corpus, the choice of hyperonyms to translate certain more specific terms) which we will depend on in the drafting phase of our dictionaries to provide any added semantic information that accompanies the suggested translation: distinctive features specific to the lemma, which are not rendered by the translation, will be indicated, with or without parentheses, after the translation (e.g., *tavola* translated as *peinture (sur bois)-painting (on wood)*—and *tavoletta* translated as *(petite) peinture (sur bois)*—(small) painting (on wood))
- leveling (non-respect of register, for example, the choice of technical rather than everyday terms, and vice versa). All entries have a usage mark indication. In the case of a translation that involves a shift of register, this change will be noted in the “translation note” part or will appear in the part reserved for distinctive semantic indicators, where several translations of the same lemma are possible, with or without loss of register. This is the case, for example, of the (unmarked) Italian word *tondo* compared to (unmarked) *médailion* and the Italianism *tondo* (technical term used mainly in French in art historians).

Baker also analyzes the explicitation which is particularly frequent in the texts which interest us because it is almost systematically employed when using an Italianism, in particular for *realia*, which are given a special treatment in our dictionaries (cf. [4, 5]). It would be possible to systematically search for this type of data in our corpus by extracting all the occurrences of “type of” or “sort of” or the elements indicated in parentheses, but we have deliberately left out this category, which is too strongly linked to the subject described by our texts and to stylistic choices shared between the authors of “natural language” texts and translators in the context of our database, and therefore would not allow us to illustrate by comparing the two types of linguistic constraint resources related to translation operations.⁴ We have also left aside “normalization” or “conservatism” which is ill-suited to our material, little conducive to variation or to lexical and stylistic exploration.

Unlike Baker ([1]: 243), who defines translation universals as “features which typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems,” we have adopted a fairly prescriptive, or rather didactic perspective, taking into consideration the phenomena of interference (influence of the source language on the target language) frequent in translation operations between two close languages like Italian and French, and in texts whose quality is far from homogeneous. Interference is, in our view, the source not only of many cases of simplification and leveling deviations found in our comparisons but of other manifestations which can be likened to losses resulting from the translation operation, or even to errors or translation inaccuracies. The translation quality assessment (TQA) model and, in particular, the different types of quality

⁴The abundant use of Italianisms is a dominant feature of the tourist guides analyzed, comparable to a desire by their authors to give these texts a «hint of Italianness» (Farina [5]: 61).

measures which can guide the translator toward an improvement in fluidity and precision can serve as a reference for this (cf. “Multidimensional Quality Metrics” [11]). These analyses mainly orient us toward the choosing a position that may seem to run counter to utilizing a descriptive corpus like Baker’s. In fact, it presents itself as an accessory allowing the user of our dictionaries to make choices, on the basis of a descriptive exploitation of the resources consulted, as we have already described it, and the indication of statistical data resulting from frequency analyses like the ones we will present below. The editor of the lexicographical files will be able to decide, if necessary and when our analyses of these data urge him or her to identify errors or deviations that could be reduced, not to suggest a form that appears in the base as a translation (indicating in the part of the article providing statistical information on the translations found) or to write the “translation note” part, optional in our bilingual articles, to advise users on their choices by explaining why certain forms may be preferred to others.

4 Natural Language Versus Translated Language: Observation of the Corpus

The difference in frequency of words or collocations present in comparable corpora containing French texts in “natural language” and texts which come from a French translation can allow us to identify forms chosen under the influence of the source language.

4.1 *Zero Frequency in Natural Language Texts*

We compared the list of words present in the LBC subcorpus of popular texts written in French containing 270,000 words with a non-integrated corpus in the database for the moment of the same text types but in translation (93,000 words) by performing a list of words found exclusively in the “in translation” subcorpus.

– errors

Most of the forms encountered can be considered errors, such as the absence of circumflex accents (*cloître*), the influence of Italian spelling on French spelling (*baroque*), excessive “Frenchization” in spelling (*Caliari*) or by using a French translation whose usage would call for the Italian form (*Sainte-Réparate* is used in French to designate the person or the cathedral of Nice but not the Santa Reparata church of Florence, for which the French form is not attested anywhere in the LBC database) or vice versa (*Giove* is never used in Italian in our corpus, but is always translated as *Jupiter*), the use of words that have nothing to do with the description of Florentine heritage, probably because they correspond to a meaning of the source word that

applies to other contexts (e.g., *coursive* in a description of Florentine monuments,⁵ or *panonceau* to refer to the panels of the Gates of Paradise). This kind of error does not give rise to the production of targeted information inside dictionaries except, in the case of a high frequency of the error (as with *panonceau*, which is found in several sources and with a total of 28 occurrences but not for *coursive*, which is attested in just one text) to justify the insertion of an Italian entry and its French translation even in cases where the frequency of the form in the corpus would not justify it. The presence, even in a small proportion of errors of this type in a text, is enough for us to dismiss the work from the published corpus.

– leveling

We can distinguish forms that correspond to a “pragmatic” or stylistic difference between French and Italian which do not interest lexicographically, such as the use of *mentionnons* in several translation texts, which is not found in any of the texts from the complete database, or certain simple past forms (*décora*, *succéda*) which are not used in popular texts in “natural” French. These are forms that correspond to different standards relating to the text types in the corpus: a more in-depth analysis would probably show little or no attested use of “nous” in French tourist guides, and the infrequent use of forms in the simple past in comparison to the compound past or present, etc.

What interests us much more in this comparison is identifying forms that, while being perfectly “correct” in French, can be considered out of context with respect to the usages attested in the same type of context in natural language. The difference in the use of a word not attested can have the effect of an “anachronism” (difference in frequency of use in synchrony). This is the case, for example, of the *grand-ducal* adjective and the past participle *paraphé*, whose Italian equivalents are more frequent in the present-day language than are their literal French translations. The difference in register can also apply in the case of a difference in “technicity.” The adjective *autographe* present in several popularization sources in translation is absent from texts of the same type in our natural language corpus, but we find some occurrences of it in more specialized texts of the general corpus. The difference in register will give rise to a differentiated marking between the lemma in the source language and its attested translation.

⁵In the sentences: “Vers le début du XVIème siècle on pensa décorer le tambour externe de la coupole par une *coursive* en marbre qui en fasse tout le tour.”; “Écoutant l’avis autorisé de Michel-Ange qui définit cette *coursive* une cage à grillons, les autorités firent cesser les travaux.”; and “Sous les petites arcades de la *coursive* du corps du palais se trouvent de nombreux emblèmes liés à la vie politique de la commune de Florence” in Riccardo NESTI, *Florence - Histoire, art, folklore, tous les chefs-d’œuvre* (Florence, ATS Italia Editrice, nd [2008]), the word *coursive* probably suggested itself as a translation for *ballatoio*, which refers to a gallery or a balcony in reference to buildings. This word is equivalent to *ballatoio* in French only in the context of describing a ship.

4.2 *Frequency Difference in Source Texts Compared to Target Texts*

To illustrate phenomena of simplification, we have queried two subcorpora of our LBC database made up of 51 lives from Giorgio Vasari’s *Le vite de ‘più eccellenti pittori, scultori e architettori* (1568) and their French translations into (Leclanché-Weiss translation, 1900).

Unable to base ourselves on statistics from parallel translation databases (for the description of these databases, see [12], we focused on French words that had a high frequency, by comparing this frequency with that of the closest common word in Italian (same meaning, same distinctive features). This allowed us to note frequency differences that will lead us to a more in-depth study to define analog networks in both languages which enable us to suggest translation links that help to avoid a loss of precision (Figs. 1 and 2).

Tableau, for example, has a frequency of 2232 per million words in our French subcorpus, while *quadro* has a frequency of 793 per million words in the Italian subcorpus containing the same texts in the original language. A large number of *quadro* hyponyms are in effect translated by *tableau* into French. If this loss is probably compensated by the addition of distinctive features that accompany the word, we think that the translator can only gain in precision if we propose other forms to him or her to render the meanings of these various hyponyms.

5 Conclusion

The comparison of results concerning the frequency of forms within the LBC corpus allowed us to illustrate the use of different subcorpora to orient both descriptive and normative information that we wish to provide in the bilingual part of our LBC dictionaries.

We will consider, even if it remains to seen [...], that an over- or under-representation of a given linguistic phenomenon can correspond to a violation of the constraint of usage [...] and that a good translation must tend towards a blending between the original language and the translated language. [9]

The application of methods aimed at verifying the quality of translations and the creation of tools based on critical analyses of existing translations, by comparing

tableau VasariFR

Query **tableau** 376 (2,232.10 per million)

Page of 2 [Next](#) | [Last](#)

FR/it_TEC_... de San Gilio. Il fit également à détrempe le **tableau** du maître-autel et les fresques d'une chapelle

FR/it_TEC_... . Sur la voûte sont quatre Patriarches, et le **tableau** d'autel représente une Trinité, avec saint

FR/it_TEC_... qu'ils sont; aussi voit-on toujours dans ses **tableaux** des fleuves, des ponts, des rochers, des herbes

FR/it_TEC_... sur des moulages de statues antiques et sur des **tableaux** sur toile qu'il fit venir de divers endroits,

FR/it_TEC_... pas plus de dix-sept ans quand il peignit le **tableau** du maître-autel de Santa Sofia, à Padoue, qui

FR/it_TEC_... à cette chapelle, il peignit encore un **tableau** qui fut placé à l'autel de saint Luc, dans l'

FR/it_TEC_... , où il inscrivit son nom. Il fit, à Vérone, un **tableau** pour l'autel de San Cristofano et de Saint'

FR/it_TEC_... , appartenant aux Frères olivétains, il fit le **tableau** du maître-autel, qui est très beau, et

FR/it_TEC_... son séjour à Vérone, il produisit plusieurs **tableaux** qu'il envoya en différents endroits. L'abbé de

FR/it_TEC_... d'anges chantant d'une grâce admirable. Ce **tableau** fut regardé comme une chose rare et l'est encore

FR/it_TEC_... château de Mantoue, pour la chapelle, un petit **tableau** , sur lequel sont des figures pas très grandes

FR/it_TEC_... extrayant des pierres pour divers travaux. Ce **tableau** est exécuté avec tant de patience et de finesse

FR/it_TEC_... . L'une de ses dernières productions fut un **tableau** à détrempe, à Santa Maria della Vittoria,

FR/it_TEC_... l'armée vénitienne contre les Français. Sur ce **tableau** , peint en détrempe et placé au maître-autel,

FR/it_TEC_... au naturel, au point qu'il paraît vivant. Ce **tableau** , qui plut et plaît encore à quiconque le voit,

FR/it_TEC_... , à San Miniato fra le Torri de Florence, un **tableau** représentant l'Assomption de la Vierge avec

FR/it_TEC_... le maître-autel de Santa Lucia de' Bardi, un **tableau** représentant la Vierge tenant son Fils, entre

FR/it_TEC_... , ayant envoyé au roi Alphonse Ier de Naples un **tableau** rempli de figures et peint à l'huile par Jean,

FR/it_TEC_... appelé un jour à Naples, il entendit parler du **tableau** à l'huile de Jean de Bruges que possédait le roi

FR/it_TEC_... plaisirs. Il peignit alors un grand nombre de **tableaux** à l'huile, d'après la manière qu'il avait

FR/it_TEC_... à la fin une grande renommée, on lui commanda un **tableau** pour l'église San Cassano ; il ne ménagea ni son

FR/it_TEC_... d'Antonello. Celui-ci, après avoir terminé le **tableau** de San Cassano, fit un grand nombre de tableaux

FR/it_TEC_... le tableau de San Cassano, fit un grand nombre de **tableaux** et de portraits, pour plusieurs gentilshommes

FR/it_TEC_... beaucoup le coloris, ils firent un **tableau** à l'huile pour le cardinal de Portugal, et qui

FR/it_TEC_... , dans la chapelle des Pucci, Antonio peignit le **tableau** de l'autel, qui est une œuvre rare et excellente

FR/it_TEC_... les traits de Gino, fils de Lodovico Capponi. Ce **tableau** est l'œuvre la plus estimée qu'Antonio ait

FR/it_TEC_... cet instrument. Antonio Pucci donna pour ce **tableau** , qui fut terminé l'an 1475, trois cents écus, en

FR/it_TEC_... Médicis, il a peint pour Laurent l'Ancien trois **tableaux** d'Hercule, chacun de cinq brasses : dans le

FR/it_TEC_... à la Vierge le peuple d'Arezzo ; sur ce **tableau** , il représenta la place de la ville, la maison de

FR/it_TEC_... . Dans la même ville d'Arezzo, il fit un **tableau** représentant l'archange Raphaël , qui est

FR/it_TEC_... de succès. Il peignit aussi quantité de **tableaux** qu'il envoya hors de Florence. Il y en a un, à

FR/it_TEC_... . À San Marco de Florence, il fit, au-dessus du **tableau** du Pardon, un demi-cintre plein de figures très

FR/it_TEC_... à toutes les dimensions soit enfin parce que les **tableaux** sur toile peuvent être transportés où l'on veut

FR/it_TEC_... de la Croix, Gentile ajouta sept ou huit **tableaux** où il représenta le miracle de la Croix qui est

FR/it_TEC_... partie de cette école ou confrérie. Le dernier **tableau** , rempli de belles inventions, représente la

Fig. 1 Search for the word “tableau” in Vasari’s *Vies* subcorpus translated into French

them, in particular, with productions which do not go through the mediation of another language, should yield an optimization of the natural character of the translated texts and of their precision, an essential aim for the disseminating quality information.

The screenshot shows a search interface for the word "quadro" in the VasariIT corpus. The search bar contains "quadro" and the VasariIT logo. Below the search bar, it indicates "Search in VasariIT" and "Query quadro 172 (793.64 per million)". The results are displayed as a list of text snippets, each starting with a truncated identifier "IT/it_TEC_...". The snippets describe various artworks and scenes, such as paintings in the Badia di Fiesoli, the Sala de' Dugento, and the Chiesa Vecchia in Zoccolo, mentioning artists like S. Francesco, S. Domenico, and S. Apollonia.

Fig. 2 Search for the word “quadri” in the Italian subcorpus of Vasari’s *Vite*

References

1. Baker M (1993) Corpus linguistics and translation studies. Implications and applications. In: Baker M et al (eds) *Text and technology*. Benjamins, Amsterdam/Philadelphia, pp 233–250
2. Billero R (2020) Cultural heritage lexicon: a case study. In: Alamán AP, Zotti V (eds) *The language of art and culture heritage: a plurilingual and digital perspective*. Cambridge Scholars Publishing, pp 86–103. ISBN: 978-1-5275-4567-0
3. Billero R, Nicolas Martine MC (2017) Nuove risorse per la ricerca del lessico del patrimonio culturale: corpora multilingue LBC. *CHIMERA Romance Corpora and Linguistic Studies* 4(2):203–216. ISSN : 2386-2629
4. Farina A (2016) Le portail lexicographique du Lessico plurilingue dei Beni Culturali, outil pour le professionnel, instrument de divulgation du savoir patrimonial et atelier didactique, *PUBLIF@RUM*, vol 24. https://publiforum.farum.it/ezine_articles.php?id=335
5. Farina A (2014) Descrivere e tradurre il patrimonio gastronomico italiano: le proposte del Lessico plurilingue dei Beni Culturali. In Chessa F, De Giovanni C (eds) *La terminologia dell’agroalimentare*. Franco Angeli, Milano, pp 55–66
6. Frawley W (1984) Prolegomenon to a theory of translation. In: Frawley W (ed) *Translation: literary, linguistic and philosophical perspectives*. University of Delaware Press, Newark, pp 159–175

7. Johansson S, Hofland K (1994) Towards an English-Norwegian parallel corpus. In Fries U et al (eds) *Creating and using English language corpora*. Rodopi, Amsterdam, pp 25–37
8. Lanini L, Nicolas Martinez MC (2018) Verso un dizionario *corpus-based* del lessico dei beni culturali: procedure di estrazione del lemmario. In: JADT'18 Proceedings of the 14th international conference on statistical analysis of textual data, pp 108–116, ISBN 978-88-3293-137-2, UniversItalia
9. Looock R, Mariaule M, Oster C (2013) Traductologie de corpus et qualité : étude de cas. Tralogy - Session 5 - Assessing Quality in MT/Mesure de la qualité en TA. <https://odel.irevues.inist.fr/tralogy/index.php?id=188>
10. Looock R (2016) *La Traductologie de corpus*. Presses Universitaires Septentrion, Villeneuve-d'Ascq
11. Uszkoreit H, Burchardt A, Lommel A (2013) A new model of translation quality assessment tralogy—Session 5—Assessing Quality in MT/Mesure de la qualité en TA. <https://odel.irevues.inist.fr/tralogy/index.php?id=319>
12. Zotti V (2017) L'integrazione di corpora paralleli di traduzione alla descrizione lessicografica della lingua dell'arte : l'esempio delle traduzioni francesi delle Vite di Vasari. In Zotti V, Pano A (eds) *Informatica Umanistica. Risorse e strumenti per lo studio del lessico dei beni culturali*. Firenze University Press.