Exploring the Lexis of Art Through a Specialized Corpus: A Bilingual Italian-English Perspective

Antonella Luporini[1]

[1] Department of Modern Languages, Literatures and Cultures, University of Bologna, Bologna, Italy

Correspondence: Antonella Luporini, University of Bologna, Department of Modern Languages, Literatures and Cultures, via Cartoleria, 5, 40124 Bologna, Italy. Tel: +39-051-20 97 240. E-mail: antonella.luporini@unibo.it

Abstract

This study presents an application of a specialized corpus, including texts specifically related to art and cultural heritage, to the analysis of artistic vocabulary in a bilingual (Italian-English) perspective, focusing on the Italian lemmas *opera*, *figura* and *disegno* and their English translation equivalents. The starting point is the Italian corpus that is being developed under the research project *Lessico multilingue dei beni culturali* ('Multilingual art and cultural heritage vocabulary', LBC), available online in open access through NoSketchEngine. First, a lemmatized nounlist ordered by frequency of occurrence is extracted from the corpus, leading to the selection of the above-mentioned focus words, in view of both their frequency and status as technical terms within the domain of art (though exhibiting different levels of technicality). These are further investigated by extracting collocates and KWIC concordances, leading to the identification of several specialized collocations and domain-specific senses. The analysis subsequently moves from corpus to dictionary, exploring the extent to which the patterns emerging from corpus investigation are accounted for in the entries for *opera*, *figura* and *disegno* in four Italian-English bilingual dictionaries. From this viewpoint, the study also aims to show how specialized corpus data can be used for the extraction of collocations, terms, and context-specific word senses, which may in turn be used both to enrich the information provided by currently available general dictionaries, and to work towards the creation of a large-scale specialized bilingual dictionary, which is non-existent to date.

Keywords: art lexis, bilingual lexicography, collocations, polysemy, specialized corpus, Italian-English translation

1. Introduction

The specialized language of art is going through a process of visibilization, thanks to the growing number of public initiatives and online resources fostering the preservation of cultural heritage and promoting sustainable tourism, such as the UNESCO World Heritage List, the European Union's Heritage Days, or the events organized by the National Trusts for England, Wales and Northern Ireland, Scotland and Italy, to name a few. Domain-specific vocabulary is consequently being used in an increasing variety of genres and contexts – promotional/informative web contents, webinars and museum virtual tours, alongside more traditional resources such as tourist guidebooks and brochures – all, by their very nature, addressing a multilingual audience.

At the same time, specialized bilingual lexicographic resources that could support, e.g., translators, web content creators and visitor bureaux in the multilingual description of art and cultural heritage are still scarce (Flinz, 2023). With reference to the Italian-English language pair, while several monolingual dictionaries, glossaries and thesauri of art terms are available (Note 1), a bilingual resource with comparable levels of comprehensiveness and specialization, to date, does not exist (Note 2).

Specialized corpora of texts from the domain of art and cultural heritage would also represent a key resource for translation and lexicographic purposes. Indeed, the fact that corpus data can reliably support different decision-making processes involved in the creation and updating of dictionaries is a given of modern lexicography: among these, defining lemma-lists, choosing relevant examples and retrieving complex information (lexical-semantic, pragmatic, collocational/colligational, connotational etc.) to be included in dictionary entries (see, e.g., Berti & Pinnavaia, 2012; Faaß, 2018; Rundell, 2018). However, until a few years ago, there were no freely accessible corpora of this kind (Flinz, 2023).

With a view to filling this gap, back in 2016 the research unit *Lessico multilingue dei beni culturali* ('Multilingual art and cultural heritage vocabulary', henceforth LBC) at the University of Florence (Italy) launched a project for the creation of a large databank of texts related to the domain of art and cultural heritage, belonging to different text-types (technical, literary, informative), and in different languages – Italian, English, French, German, Russian and Spanish (see Farina & Nicolás Martínez, 2020). The project was subsequently extended to other universities, among which Bologna, Milano Statale, Paris 8 and Pisa. At its current stage of development, it has led to the creation of six specialized comparable corpora (one for each of the project languages), including at least 1 million words, which are freely accessible online through NoSketch Engine, a free version of the corpus management system Sketch Engine (Note 3). This resource remains one of a kind to date, both in terms of coverage and accessibility. The ultimate goal of the project, currently work-in-progress, is the creation and open-access publication of a corpus-based, specialized multilingual dictionary, made up of single monolingual dictionaries with dynamic connections among the entries in the various languages, and of several parallel corpora, where source texts in the project languages will be aligned with their available translations (see Zotti, 2017; Farina & Billero, 2020; Flinz, 2023).

Against the above background, we will now move on to describe an application of the currently available LBC corpora to the analysis of the lexis of art in an Italian-as-source/English-as-target perspective. This contribution qualifies as a corpus-driven study (Tognini-Bonelli, 2001), in that the choice of items to be investigated and the analysis/identification of patterns are fundamentally driven by corpus criteria, as explained in the 'Method' section below. Focus is on the Italian lemmas *opera*, *figura* and *disegno*, and related translation equivalents in English. By comparing corpus evidence with the information provided by several bilingual dictionary entries, this study ultimately aims to offer new perspectives on the lexicographical choices made with reference to these words, from the specific perspective of the translation of art vocabulary.

2. Method

Since it was decided, at the study design stage, to focus on Italian as the source language, the Italian LBC corpus (Note 4) provides the starting point for this study. This corpus currently consists of approximately 1,186,000 word tokens, mostly from two of the text-type categories envisaged by the project: technical and literary (64% and 28.3%, respectively). A key source of texts for this corpus is Giorgio Vasari's *Le vite*: a collection of biographies of the most eminent Italian painters, sculptors and architects, written in the 16th century. *Le Vite* is a foundational text for the description of Italian Renaissance art and for art history more generally (Note 5); due to its comprehensiveness and complexity, it may be treated as a small specialized corpus in its own right (Luporini, 2023).

The analysis took place through the steps listed below.

2.1 Nounlist

The first step of research involved retrieving a wordlist from the Italian LBC corpus, with the aim of identifying lexical candidates to be examined. A wordlist is a list of all the words included in the corpus, ordered by raw frequency of occurrence; wordlists can also be lemmatized and annotated with part-of-speech information (see Baker, Hardie, & McEnery, 2006, p. 169).

Nouns are particularly important in the creation of technical vocabularies, as indicated by the etymology of the word 'nomenclature' itself: "(from the Latin *nomen calare*) first appeared in French and English at the beginning of the 16th century, with the meaning of 'glossary' or '*list of names*'" (Rey, 1995, p. 11; emphasis added). Taking this into consideration, a wordlist tailored to the purposes of this study was generated, including only the nouns that appear in the corpus, in a lemmatized form: we shall refer to this as a 'nounlist'. This was done by choosing 'find tags matching regular expression (regex)' in the advanced wordlist options, filling in the regex slot with the tag NOUN.* (Note 6), and, finally, selecting 'display results as lemmas'. This procedure, which was devised as a 'shortcut' to the specialized words included in the corpus, also has the advantage of automatically eliminating from the list all the grammatical words that, due to their frequency, typically rank very high independently of the text-type, but only create 'noise' when analysis is oriented towards content words.

From the final output, to be illustrated in Section 3, the three items *opera*, *figura* and *disegno* were selected as foci for this study, based on their frequency and status as specialized vocabulary items.

2.2 Collocations and KWIC Concordances

The second step of the analysis involved examining the behavior of the focus words emerging from the previous step through collocations and KWIC concordances. Collocates are words that tend to co-occur with a node word in a significant way in a corpus (see Baker, Hardie, & McEnery, 2006, p. 36 ff.). NoSketch Engine – just like its

fully functional counterpart, Sketch Engine – provides various options for collocate identification. For this study, the collocational window (the span within which the system will look for collocates) was set to -3/+3 – from three words to the left to three words to the right of the node word; this is narrower than the 'traditional' -5/+5 window, but is supposed to yield more precise results in terms of strength of association (cf. Bartsch & Evert, 2014, p. 57). Candidate collocates were retrieved in lemma form and ordered by LogDice score, a statistical measure of the strength of collocational association (Note 7). One problem with this step was that the lists of collocates thus generated were skewed by an overabundance of grammatical words, and even punctuation marks, which occupied the top positions without being particularly revealing (Note 8). For this reason, it was decided to work on a significant portion of the collocate lists, taking into account all the candidates with a LogDice value higher than 7.0. Collocations worth further examination within the context and objectives of this study were then singled out. Concretely, this was done by 'moving back and forth' between the collocate lists and the related KWIC concordances, showing the collocates in their original, extended (sentence-length) co-text (see Baker, Hardie, & McEnery, 2006, p. 42 ff.).

2.3 Looking at Dictionary Entries Through the Lens of the Corpus

The final step of the analysis involved moving from corpus to dictionary, with a view to assessing if, and to what extent, the specific collocations and word senses resulting from 2.2 above are included in the entries for our focus words. To this end, a set of four bilingual Italian-English resources was identified. This included two renowned free online dictionaries, the Collins Italian-English Dictionary online and the Cambridge Italian-English Dictionary online (Note 9), and the digital editions of two authoritative dictionaries, both requiring subscription: the Oxford Italian-English Dictionary and *Il Ragazzini* Zanichelli – undoubtedly a best-seller in Italy, featuring "the richest and most up-to-date range of aids for Italian students of English" (Iamartino, 2019, p. 149) (Note 10). This set is necessarily limited, due to the qualitative nature of this study, but it still aims at being representative of different publishing/lexicographic traditions (UK- and Italy-based companies) and also of different user needs. In fact – in the absence of a specialized bilingual dictionary focusing exclusively on the language of art, and with the growing demand for multilingual texts in this sector (cf. the Introduction) – it was hypothesized that free online dictionaries like Collins and Cambridge are more likely to be consulted by non-professional translators than Oxford or Zanichelli.

3. Results and Discussion

3.1 Nounlist

The first step that will be illustrated is the choice of the focus words for this study. Taking into account the upper part of the LBC Italian corpus nounlist, showing the most frequent noun lemmas in the corpus (Table 1), the three interconnected items *opera*, *figura* and *disegno* were selected, being interpreted as items belonging to art vocabulary.

Table 1. LBC Italian corpus nounlist: top ten noun lemmas by frequency of occurrence

#	Noun (lemma)	Frequency	#	Noun (lemma)	Frequency
1	cosa	4,925	6	anno	2,063
2	**opera**	4,081	7	modo	1,939
3	tempo	2,356	8	parte	1,762
4	mano	2,324	9	**disegno**	1,647
5	**figura**	2,100	10	città	1,635

With reference to their being interconnected, *opera* may be taken as the superordinate, with *figura* and *disegno* as co-hyponyms (i.e., *figura* and *disegno* as 'types of' opera). However, between *opera* and *disegno*, on the one hand, and *figura*, on the other, a lexical relation of meronymy could also be found, depending on the context (i.e., *figura* as 'part of' an *opera* or *disegno*). At the same time, it must be noted that our focus words also instantiate different levels of *technicality*, understood as "the degree to which a term is specialized and exclusively used by experts in a domain" (Hätty et al., 2020, p. 2883; see also Dima, 2012, pp. 95–96). *Disegno* – even though it is commonly used in general language, and its meaning has been metaphorically extended to different domains – still enjoys full status as an element of artistic terminology, as also highlighted by its etymology (Note 11). In fact, *disegno* is also the only lexical unit from this set having a dedicated entry in an Italian-language art encyclopedia that was consulted as a reference (*La nuova enciclopedia dell'arte Garzanti*, 1986). *Figura* is also etymologically linked to the domain of arts and crafts, broadly speaking (Note 12), but the connection is arguably more opaque in this case, probably also as a consequence of a broader process of polysemization.

Similar considerations can be made with reference to *opera*, the item with the lowest level of technicality in this set (Note 13). By the same token, *mano* and *modo*, which had also been originally identified as potentially relevant items, were eventually put aside for a future study, after looking at their occurrences in KWIC concordances, and noticing that the more technical senses 'layer of color' and 'artistic manner' are not predominant in the corpus. All in all, the fact that technical or semi-technical terms rank so high may also be taken as a positive sign concerning the quality of this specialized corpus.

3.2 Collocations and KWIC Concordances

3.2.1 Opera

Within the collocate list for *opera* generated with the criteria described in Section 2.2 above (Note 14), the lemmas with a LogDice value higher than 7.0 correspond roughly to the first 120. From this set, two candidate collocates worth further investigation were singled out: *bocca* (no. 62) and *fresco* (no. 103). The related concordance lines show that *opera*, on the one hand, and *bocca* and *fresco*, on the other, occur together within the expressions *bocca d'opera* (35 hits) and *opera in/a fresco* (14 hits).

Bocca d'opera is a lexicalized expression, commonly found as a single word in present-day Italian, which refers to the part of the stage in a theatre that is framed by the proscenium arch. Authoritative Italian dictionaries such as *Zingarelli*, *Sabatini-Coletti* and *Devoto-Oli* explain it by establishing a correspondence with *boccascena* (usually translated as 'proscenium'), even though, according to the *Manuale della lingua teatrale* by Papiol (1909), the two terms actually refer to different parts of the front stage; there is also a difference between *proscenio* and *boccascena* (Note 15). The multiword expression *bocca d'opera* in the LBC corpus is found within a single text: Giordani's *Intorno al Gran Teatro del Comune e ad altri minori in Bologna* (1855).

As for *opera in fresco* and *opera a fresco* (lit. 'work in fresco'), corpus data point to them as virtually exclusive to *Le Vite*, where Vasari uses them in place of *affresco* ('fresco'). In fact, *affresco* as a single word in Italian appears for the first time in the 19th century, from the bigram (*dipingere*) *a fresco*, of which *in fresco* is a less frequent variant. Indeed, a search for *affresco* and *affreschi* in the Italian LBC corpus yields only results from later texts, such as Gualandi's *Tre giorni in Bologna o Guida per la città e i suoi contorni* (1865), or Baraldi's *Alla scoperta dei segreti perduti di Bologna* (2016).

3.2.2 Figura

As was the case with *opera*, in the collocate list for *figura* (Note 16), the candidates with a LogDice value higher than 7.0 amounted to roughly 120. Six were shortlisted in this case: *storia* (no. 3); *tondo* (no. 12); *mezzo* (no. 18); *rilievo* (no. 23), *fresco* (no. 27) and *terra* (no. 69). By looking at the related concordance lines, several specialized expressions were identified, all occurring in the corpus as part of Vasari's *Le Vite*: these are listed below, together with explanations and English translation equivalents.

Storia di/in figure (31 hits), referring to the painted or sculpted representation of an event made up of various parts. En. 'scene with figures'.

Figura tonda (35 hits), denoting well-rounded sculpted or carved figures, in full relief. As Vasari himself explains in the *Introduction to the three arts of design* that opens *Le Vite*, "[s]uch figures [those showing due proportion, grace, design and perfection] we call figures 'in the round', provided that all the parts appear finished, just as one sees them in a man, when walking round him" (Maclehose & Brown, 1907, p. 147). En. 'figure (carved/cast) in the round', 'round figure', or 'figure in full relief'.

Mezza figura (25 hits), referring to the representation (painting, drawing, or sculpture) of a man or woman from head to waist. En. 'half-length figure'.

Figura di (*basso/mezzo/gran/tondo*) *rilievo* (23 hits), denoting sculpted figures, landscapes, or architectural/ornamental elements that emerge from the background in a more or less neat way. En. 'figures in (low/half/strong/full) relief'.

Figura in/a fresco (26 hits), referring to 'frescoed' figures. En. 'figure in fresco'.

Figura di terra (15 hits), i.e., sculpted figures made with *terra* in the sense of clay. En. 'figure in clay'.

3.2.3 Disegno

The candidates with LogDice higher than 7.0 in this case amounted to roughly 110 (Note 17). From this subset, the following were selected for further investigation: *modello* (no. 4), *invenzione* (no. 5), *grazia* (no. 8), *cartone* (no. 11), *ordine* (no. 15) and *giudizio* (no. 31), all occurring within nominal group complexes of the kind '*disegno* and x': *disegno e modello* (or *modello e disegno*, 41 hits); *disegno e invenzione* (or *invenzione e disegno*,

33 hits); *disegno e grazia* (or *grazia e disegno*, 28 hits); *disegno e cartone* (or *cartone e disegno*, 21 hits); *disegno e ordine* (or *ordine e disegno*, 27 hits) and, finally, *disegno e giudizio* (or *giudizio e disegno*, 13 hits). The related concordance lines were, once again, almost exclusively extracted from *Le Vite*.

As can be noted, these expressions cannot be said to realize specialized, terminologically relevant collocations (as was the case with *opera* and *figura*); their relevance for the present study rather lies in the fact that they emphasize the inherent polysemy of the word *disegno* in Vasari's masterpiece, already shown by previous contrastive translation studies on other language pairs (Carpi & Pano Alamán, 2019, on Italian and Spanish; Ballestracci, 2023, on Italian and German).

In fact, the analysis of the above-mentioned expressions in context by means of concordance lines points to three main senses of the word *disegno*, which emerge neatly from its association (in some cases, contrast) with the collocates being considered, and which can be put on a cline, from the more concrete to the more abstract: (i) *disegno* as 'graphic representation' resulting from the act of drawing; (ii) *disegno* in the sense of 'preparatory draft', but also 'project', 'plan'; (iii) by extension, *disegno* as 'skill', 'talent', also 'inventive ability'. This breakdown obviously also impacts on the choice of an appropriate translation equivalent. Through the following concordance lines, we illustrate the difference between senses (i) and (iii) in relation to the collocation *invenzione e disegno*, by also comparing the Italian source text with De Vere's *Lives of the Most Eminent Painters, Sculptors and Architects* (1912–1915), which remains to date the only extant complete English translation of *Le Vite*. In (1), *disegno* basically refers to the (positively evaluated) final result, and is translated as 'drawing' by De Vere. In (2) and (3), *disegno* stands for 'skill', 'ability' and is translated by De Vere as 'draughtsmanship' and 'design', respectively.

(1) Sonovi oltre ciò molte grottesche e altre cose lavorate di chiaroscuro simili al marmo e fatte stranamente con **invenzione e disegno** bellissimo (*Vita di Filippo Lippi pittore fiorentino*).

Besides this, it contains many grotesques and other things wrought in chiaroscuro to resemble marble, executed in strange fashion with **invention and** most beautiful **drawing** (De Vere, 1913, vol. 4, p. 8).

(2) Aveva Cosimo, sentendosi debole **d'invenzione e di disegno**, cercato di occultare il suo deffetto (*Vita di Cosimo Rosselli*).

Cosimo, feeling himself weak in **invention and draughtsmanship**, had sought to conceal his shortcomings (De Vere, 1912, vol. 3, p. 189).

(3) ella sarebbe certo bellissima tra tutte le cose d'Andrea: al quale, se la natura avesse dato gentilezza nel colorire come ella gli diede **invenzione e disegno**, egli sarebbe veramente stato tenuto maraviglioso (*Vita di Andrea Dal Castagno e Domenico Veneziano*).

it would certainly be the most beautiful of all the works of Andrea. And if Nature had given grace of colouring to this craftsman, even as she gave him **invention and design**, he would have been held truly marvellous (De Vere, 1912, vol. 3, p. 100).

At the same time, *disegno* as 'draft', 'project' or 'plan' – to be technically distinguished from a 'model' or a 'cartoon' – may also be rendered through 'design' in English, as shown by the following concordance lines related to *disegno e modello* and *disegni e cartoni*:

(4) egli, secondo ch'io truovo, fece il **disegno e modello** del palazzo de' Governatori della città d'Ancona (*Vita di Margaritone*).

he, according to what I find, made the **design and model** of the Palazzo de' Governatori in the city of Ancona (De Vere, 1912, vol. 1, p. 66).

(5) oltre a molti **disegni e modelli** che vi fece di abitazioni private e publiche (*Vita di Michelozzo Michelozzi*).

besides many **designs and models** that he made for private dwellings and public buildings (De Vere, 1912, vol. 2, p. 261).

(6) e con quanta diligenza esso Puntormo conducesse a perfezione e ponesse ottimamente in pittura i **disegni e 'cartoni** di Michelagnolo (*Vita di Jacopo da Pontormo*)

with what diligence Pontormo carried to completion and executed excellently well the **designs and cartoons** of Michelagnolo (De Vere, 1914, vol. 7, p. 172).

All these senses of *disegno* are likely to be found also in contemporary texts describing works of art or promoting cultural heritage, and are, therefore, relevant also from the perspective of bilingual lexicography and translation. Other 'nuances' of meaning of *disegno* – and potential translation equivalents – might be unveiled by

future studies comparing *Le Vite* with different English translations through a wider and more systematic approach. This will be made possible by the implementation of parallel corpora, which is on the LBC project agenda (see Section 1).

3.3 From Corpus to Dictionary

The last stage of the analysis involved checking the entries for *opera*, *figura* and *disegno* in the Italian-English section of the bilingual dictionaries listed in Section 2.3, in order to look for the specialized collocations and multiple senses of the focus words that emerge from the corpus. Table 2 summarizes the main findings.

Table 2. Bilingual dictionary entries for *opera*, *figura*, *disegno*: presence (✓) or absence (✗) of corpus findings

Search item	Collins Ita-En Dictionary online	Cambridge Ita-En Dictionary online	Oxford Ita-En Dictionary	*Il Ragazzini* Zanichelli
Bocca d'opera	✗ [Only *boccascena* as separate entry > 'proscenium']	✗	✗ [Only *boccascena* as separate entry > 'proscenium']	✓ [*Boccadopera* present as separate entry > 'proscenium']
Opera in/a fresco	✗	✗	✗	✗
Storia di/in figure	✗	✗	✗	✗
Figura tonda	✗	✓ > 'round figure' [*figura rotonda*, not indicated as 'Art']	✓ > 'round figure' [*figura tonda*, not indicated as 'Art']	✗
Mezza figura	✓ > 'half-length portrait'	✗	✗	✓ > 'half-length portrait'
Figura di (*basso/mezzo/gran/tondo*) *rilievo*	✗	✗	✗	✗
Figura in/a fresco	✗	✗	✗	✗
Figura di terra	✗	✗	✗	✗
Disegno as 'graphic representation' – sense (i)	✓ > 'drawing'; (*schizzo*) 'sketch'	✓ > 'drawing'; 'sketch'; 'design'	✓ > 'drawing'; 'picture'	✓ > 'drawing'; (*schizzo*) 'sketch'
Disegno as 'draft', 'project', 'plan' – sense (ii)	✓ > 'outline'; 'plan'; 'project' [all indicated as 'figurative']	✓ > 'sketch'; 'outline'; 'design'	✓ > 'design'; 'scheme'	✓ > 'outline'; 'sketch'; 'design'; 'plan'
Disegno as 'skill', 'talent', 'inventive ability' – sense (iii)	✗	✗	✗	✓ > 'draftsmanship' 'draughtsmanship';

As can be seen from Table 2, none of the consulted dictionaries includes all the information emerging from our corpus investigation. This is perhaps not surprising, considering that these are general dictionaries. However, given the paucity of specialized reference resources currently available for the translation of art lexis – and the consequent need for expert and non-expert translators to rely on general resources – a few additions based on corpus evidence would be desirable. These concern especially the entries for *figura* and *disegno*. The findings related to *bocca d'opera* and *opera in fresco*, which are archaic forms only weakly linked to *opera* in present-day language, are probably not as relevant for general dictionaries as they could be for the creation of a specialized comprehensive dictionary.

Regarding *figura*, it would be worth adding the specialized collocations/terms *storia di/in figure* ('scene with figures') and *figura di terra* ('figure in clay') that are absent from all the examined entries, especially considering that the words *storia* and *terra* may be misleading for Italian speakers who are not versed in art terminology and may interpret and translate them as 'story' and 'earth'/'soil', respectively. *Mezza figura* could also be incorporated in the Cambridge and Oxford dictionaries, while the superordinate 'half-length figure' could be included along with 'half-length portrait' in Collins and *Il Ragazzini*, thereby more explicitly including carved realizations, in line with corpus evidence. As for *figura tonda*, the translation equivalents 'figure in the round' and 'figure in full relief' could also be added and explicitly signalled as belonging to the field of art. In the case of *figura di rilievo*, since the bigram *in rilievo*, translated as 'in relief', can be found under *rilievo* in all the

dictionaries under scrutiny, it would probably be unnecessary to also mention it in the entry for *figura* (although it must be noted that Cambridge only mentions 'in relief' with specific reference to the globe or a map). However, the entries for *rilievo* could also accommodate the collocations *di basso/mezzo/gran/tondo rilievo*, with the respective translation equivalents 'in low'/'half'/'strong'/'full' relief. Currently, only Collins and Ragazzini have both 'high-relief' and 'bas-relief'; Cambridge only mentions 'bas-relief' and Oxford does not give any specification. As for *figura in/a fresco*, the observations made above with reference to *opera in/a fresco* remain valid.

Concerning *disegno*, while all the dictionaries account for senses (i) and (ii) as emerging from the corpus analysis presented above, sense (iii) can only be found under this lemma in *Il Ragazzini*, translated as 'draughtsmanship' (AmE 'draftsmanship'). Cambridge, under the label "*arte di disegnare*", provides translations that appear to apply more to the subject taught in schools than to a type of art ('drawing', 'design', 'graphic arts') – therefore, these were not included in Table 2. A more comprehensive account of the complex polysemy of this word would be provided by the other dictionaries if they acknowledged this sense as well. Additional corpus-assisted studies are needed to provide evidence that may shed light on the contextual difference between apparently interchangeable equivalents like 'design' and 'sketch', on the one hand, and 'design' and 'cartoon', on the other. For instance, Cambridge, differently from the other dictionaries, includes 'design' among the possible equivalents for our sense (i); Cambridge and *Il Ragazzini* mention 'sketch' among those for sense (ii); *Il Ragazzini* translates the collocation *disegno preparatorio* – under *disegno* as 'graphic representation', our sense (i) – as 'preparatory drawing' or 'cartoon', but this does not seem to fit in well with corpus evidence showing the existence of the collocation *disegni e cartoni* in Vasari (see Section 3.2.3). These remain open questions to be investigated with the indispensable aid of parallel corpora that comprise more recent or contemporary source texts and translations.

4. Conclusions and Outlook

This study has presented a possible application of the specialized corpora that are currently being developed as part of the LBC project to the analysis of the lexis of art for bilingual lexicographic purposes. More specifically, it has shown how specialized corpus data can be used for the extraction of collocations, terms, and context-specific word senses, which may be used both to enrich the information provided by currently available general dictionaries and to work towards the creation of a truly specialized lexicographic resource. More extensive studies on Italian-English pairs are definitely needed in this perspective; indeed, the parallel corpora foreseen by the LBC project will enable more systematic contrastive analysis that can shed light on both the source and the target language. A possible limitation of this study has to do with the corpus data themselves, which mostly come from ancient texts, especially *Le Vite*. While most terminology introduced by Vasari in his foundational work remains valid at present (and this is the reason why efforts were made by the LBC team to include the full text of both editions in the Italian corpus), more recent and contemporary texts are needed to make the corpus itself – and the investigations based on or driven by it – more balanced. From this perspective, the Italian LBC corpus should (and, as a 'monitor' corpus, will) be updated so as to include more texts, encompassing a wider range of historical periods covered. In the final analysis, this shows how the results of corpus-assisted studies can provide feedback also from the viewpoint of corpus 'maintenance'.

Acknowledgments

References

Adamska-Sałaciak, A. (2020). Bilingual Learners' Dictionaries in the Lexicographic Landscape. In M. Szczyrbak & A. Tereszkiewicz (Eds.), *Languages in Contact and Contrast. A festschrift for Professor Elżbieta Mańczak-Wohlfeld on the Occasion of Her 70th Birthday* (pp. 37–47). Krakow, Jagiellonian University Press.

Baker, P., Hardie, A., & McEnery, T. (2006). *A Glossary of Corpus Linguistics*. Edinburgh, Edinburgh University Press. https://doi.org/10.1515/9780748626908

Ballestracci, S. (2023). Da un segno a tanti segni. L'emergere della polisemia del termine *disegno* nelle traduzioni tedesche delle *Vite* di Vasari. In V. Zotti & M. Turci (Eds.), *Nuove strategie per la traduzione del lessico artistico. Da Giorgio Vasari a un corpus plurilingue dei beni culturali* (pp. 21–38). Firenze, Firenze University Press. https://doi.org/10.36253/979-12-215-0061-5.04

Berti, B., & Pinnavaia, L. (2012). Towards a bilingual Italian-English dictionary of collocations. In R. Facchinetti (Ed.), *English Dictionaries as Cultural Mines* (pp. 201–221). Newcastle upon Tyne, Cambridge Scholar.

Bartsch, S., & Evert, S. (2014). Towards a Firthian notion of collocation. *OPAL – Online publizierte Arbeiten zur Linguistik*, *2*, 48–61.

Carpi, E., & Pano Alamán, A. (2019). Diseño, Desarrollo y aplicaciones del Corpus LBC sobre el léxico del arte y el patrimonio cultural. *Revista de Lexicografía*, *25*, 149–160. https://doi.org/10.17979/rlex.2019.25.0.5992

De Vere, G. (1912–1915). *Lives of the Most Eminent Painters, Sculptors and Architects by Giorgio Vasari* (Vol. 1–10). London, Macmillan and Co. Ld. & The Medici Society.

Dima, G. (2012). A terminological approach to dictionary entries. A case study. *PROCEDIA – Social and Behavioral Sciences*, *63*, 93–98. https://doi.org/10.1016/j.sbspro.2012.10.016

Faaß, G. (2018). Lexicography and corpus linguistics. In P. A. Fuertes-Olivera (Ed.), *The Routledge Handbook of Lexicography* (pp. 123–137). Abingdon/New York, Routledge. https://doi.org/10.4324/9781315104942-9

Farina, A., & Billero, R. (2020). Corpus in "natural" language versus "translation" language: LBC corpora, a tool for bilingual lexicographic writing. In D. F. Iezzi, D. Mayaffre & M. Misuraca (Eds.), *Text Analytics: Advances and Challenges. JADT 2018. Studies in Classification, Data Analysis, and Knowledge Organization* (pp. 167–178). Cham, Springer. https://doi.org/10.1007/978-3-030-52680-1_14

Farina, A., & Nicolás Martínez, M. C. (2020). La banca dati LBC. In R. Billero, A. Farina & M. C. Nicolás Martínez (Eds.), *I corpora LBC. Informatica umanistica per il lessico dei beni culturali* (pp. 7–17). Firenze, Firenze University Press. https://doi.org/10.36253/978-88-5518-253-9

Flinz, C. (2023). The multifunctional LBC-Corpora: different aims depending on the user. *Lexicographica*, *39*, 191–208. https://doi.org/10.1515/lex-2023-0010

Hätty, A., Schlechtweg, D., Dorna, M., & Schulteim Walde, S. (2020). Predicting degrees of technicality in automatic terminology extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 2883–2889). Kerrville, Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.258

Iamartino, G. (2019). Recent bilingual English-Italian lexicography: Insights into usage. *Dictionaries: Journal of the Dictionary Society of North America*, *40*(2), 133–154. https://doi.org/10.1353/dic.2019.0017

Luporini, A. (2023). 'Metaphors of art' in the English translations of Giorgio Vasari's *Le Vite*: A software-assisted enquiry. In V. Zotti & M. Turci (Eds.), *Nuove strategie per la traduzione del lessico artistico. Da Giorgio Vasari a un corpus plurilingue dei beni culturali* (pp. 65–92). Firenze, Firenze University Press. https://doi.org/10.36253/979-12-215-0061-5.06

Maclehose, L. S., & Brown, G. B. (Trans. & Ed.). (1907). *Vasari on Technique, Being the Introduction to the Three Arts of Design, Architecture, Sculpture and Painting, by Giorgio Vasari*. London, J.M. Dent & Co.

Papiol. (1909). *Manuale della lingua teatrale*. Milan, Società Editoriale Milanese.

Rey, A. (1995). *Essays on Terminology, Translated and Edited by J. C. Sager*. Amsterdam/Philadelphia, Benjamins. https://doi.org/10.1075/btl.9

Rundell, M. (2018). Searching for extended units of meaning – and what to do when you find them. *Lexicography*, *5*, 5–21. https://doi.org/10.1007/s40607-018-0042-1

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam/Philadelphia, Benjamins. https://doi.org/10.1075/scl.6

Zotti, V. (2017). L'integrazione di corpora paralleli di traduzione alla descrizione lessicografica della lingua dell'arte: l'esempio delle traduzioni francesi delle *Vite* di Vasari. In V. Zotti & A. Pano Alamán (Eds.), *Informatica umanistica. Risorse e strumenti per lo studio del lessico dei beni culturali* (pp. 105–134). Firenze, Firenze University Press. https://doi.org/10.36253/978-88-6453-546-3

Notes

Note 1. E.g., for English, the *Oxford Dictionary of Art*, the *Oxford Concise Dictionary of Art Terms*, or the Getty Research Institute *Art and Architecture Thesaurus Online*

(https://www.getty.edu/research/tools/vocabularies/aat/index.html); for Italian, the *Dizionario Enciclopedico dell'Arte* Mondadori, the *Dizionario Arte* Jaca Book, or the *Dizionario dell'Arte* Baldini Castoldi Dalai, which is, in fact, translated from the *Oxford Dictionary of Art*. With reference to English, an up-to-date list of monolingual resources is maintained by the Berkeley Library of the University of California (https://guides.lib.berkeley.edu/c.php?g=478634&p=3273764).

Note 2. Some bilingual technical dictionaries (e.g., the *McGraw-Hill* published by Zanichelli, or the *Nuovo Marolli* and the *Dizionario tecnico dell'edilizia e dell'architettura* published by Hoepli) also include terms related to the artistic domain, especially to the sub-field of architecture, but are not specialized in this sense. Margherita Palli's *Dizionario teatrale* (Quodlibet, 2021) provides translation equivalents in multiple languages of terms related to the sub-field of theatre (architecture and performance).

Note 3. The LBC corpora can be accessed from https://corpora.lessicobeniculturali.net/. For a detailed overview of corpus construction criteria, see Farina and Billero (2020); Flinz (2023).

Note 4. https://corpora.lessicobeniculturali.net/it/

Note 5. There are two editions of Giorgio Vasari's text: the first was published in 1550 as *Le Vite de' più Eccellenti Architetti, Pittori et Scultori Italiani*; the second, revised and extended edition was published in 1568, with the title *Le vite dei più eccellenti Pittori, Scultori, e Architettori*. The two editions are also referred to as 'Torrentiniana' and 'Giuntina', respectively, after the names of the typographers that first printed and circulated them. Both editions are included in the Italian LBC corpus and make up most of its texts. Descriptive statistics concerning corpus composition can be found at the URL provided above.

Note 6. The appropriate tags to be used in regular expressions depend on the tagset used for the corpus, which can be found under 'corpus information' in Sketch Engine and NoSketch Engine.

Note 7. LogDice is the preferred statistic measure for computing collocations on SketchEngine, for several reasons: see https://www.sketchengine.eu/my_keywords/logdice/.

Note 8. The LogDice measure usually helps to avoid this issue, but it was not as effective as expected in this case, possibly because of the still limited size of the corpus we are considering (LogDice performs well on very big corpora), or because of underlying inaccuracies in lemmatization and part-of-speech tagging of the corpus itself.

Note 9. The two dictionaries can be consulted at the following URLs: https://www.collinsdictionary.com/dictionary/italian-english; https://dictionary.cambridge.org/it/dizionario/. The Cambridge Ita-En Dictionary is in fact based on the Cambridge corpus and on other two dictionaries, Global and Password by Kernerman Publishing, which are best defined as *semi-bilingual* learners' dictionaries (more details on the distinction, which is not deemed relevant for the purposes of this study, can be found in Adamska-Sałaciak, 2020, p. 43).

Note 10. For both the Oxford Dictionary and *Il Ragazzini*, digital editions, access was granted by the University of Bologna.

Note 11. The deverbal noun *disegno* comes from *disegnare*, literally 'to represent something by drawing lines on a surface' (*L'etimologico. Vocabolario della lingua italiana*, Le Monnier).

Note 12. According to the etymological dictionary, Italian *figura* is a borrowing from Latin *figūra* – 'image', 'effigy', 'statue', 'appearance' – in turn derived from *fingĕre*, literally 'to mould', 'to craft' (*L'etimologico. Vocabolario della lingua italiana*, Le Monnier).

Note 13. Cf. the entry for *opera* in the *Vocabolario Treccani*, which mentions *opera d'arte* (En. 'artwork') as only one of the possible, more concrete meanings of the word, together with several others, belonging to different fields: e.g., *opera letteraria* (En. 'literary work'), *opera in musica* (En. 'opera'), but also *opera dell'ingegno* (En. 'intellectual work'), *opera viva* (En. 'quickwork' – in shipbuilding, the bottom part of a ship). Indeed, It. *opera* comes from Latin *ŏpĕra*, plural form of *opus*, *opĕris*, which simply means 'product of work' (see https://www.treccani.it/vocabolario/opera/).

Note 14. The collocate list for *opera* can be found here: https://shorturl.at/gqzQT.

Note 15. Indeed, the *Vocabolario Treccani* defines *boccadopera* without establishing links with *boccascena* (see: https://www.treccani.it/vocabolario/boccadopera/). See also the entry for *boccascena*, mentioning its relationship with *proscenio*: https://www.treccani.it/vocabolario/boccascena/.

Note 16. The collocate list for *figura* can be found here: https://shorturl.at/isDEG.

Note 17. The collocate list for *disegno* can be found here: https://shorturl.at/vLWY3.