

Towards a Linguistic Linked Open Data Resource for Italian Cultural Heritage: The *Lessico dei Beni Culturali* Corpus

Riccardo Billero

Free University of Bozen

riccardo.billero@unibz.it

Abstract

We present an ongoing effort to bridge the *Lessico dei Beni Culturali* (LBC), a multilingual lexicographic project covering Italian cultural heritage terminology, with the Linguistic Linked Open Data (LLOD) ecosystem. The LBC corpus spans five centuries of art-historical writing, from fifteenth- and sixteenth-century treatises by Alberti, Leonardo, and Vasari to nineteenth-century works by Stendhal and Burckhardt and contemporary tourist guides to Florence, with source texts in several European languages alongside their translations. The resource has already undergone automatic linguistic annotation and term extraction, but lacks structured lexical representation in any standard LLOD formalism. We describe the current state of the resource, identify the main challenges for its publication as Linked Data — including the modelling of culturally-bound terms (*realia*), historical proper nouns, and multilingual source texts of different registers — and outline a roadmap towards its representation in OntoLex-Lemon (McCrae et al., 2017) and its alignment with existing LLOD resources such as the Getty Vocabularies (Getty Research Institute, 2024a) and Wikidata (Vrandečić and Krötzsch, 2014). By sharing this work with the LLOD community, we expect input on best practices for historical-artistic and cultural heritage lexicons that will raise interoperability between resources from different sources, generating new information and increasing the value of existing data.

Keywords: Linguistic Linked Open Data, cultural heritage, lexicography, multilingual corpus, digital humanities, corpus annotation

1. Introduction

The Linguistic Linked Open Data (LLOD) cloud has grown substantially over the past decade, yet certain specialised domains remain conspicuously underrepresented. Among these is the language of cultural heritage: the rich, historically layered vocabulary used to describe artworks, monuments, artists, and artistic practices. This gap is especially striking in the case of Italian, whose cultural heritage terminology forms the backbone of art history discourse worldwide yet lacks a structured, interoperable lexical representation.

The *Lessico dei Beni Culturali* (LBC; “Lexicon of Cultural Heritage”) project directly addresses this gap from the perspective of multilingual lexicography and specialised translation (Garzaniti and Farina, 2013; Billero and Nicolás Martínez, 2017; Billero, 2020a; Billero et al., 2020; Ballestracci, 2023; Flinz et al., 2024). The project has assembled a corpus of parallel and comparable texts spanning five centuries of art-historical writing, in Italian and seven other languages (Chinese, English, French, German, Portuguese, Russian, Spanish).

What makes the LBC corpus particularly distinctive is its multilingual nature and the broad diachronic range and variety of genres of its source texts: it ranges from foundational Italian treatises — Alberti’s *I dieci libri dell’architettura* (Alberti, 1550) and *Della pittura* (Alberti, 1436), Leonardo da Vinci’s *Trattato della pittura* (da Vinci,

1651), and Vasari’s *Le Vite de’ più eccellenti pittori, scultori e architettori* (Vasari, 1568) — to works originally composed in other European languages, such as Stendhal’s *Rome, Naples et Florence* (Stendhal, 1817) and Jacob Burckhardt’s *Geschichte der Renaissance in Italien* (Burckhardt, 1867), through to contemporary tourist guides to Florence (Cicarelli Roming et al., 2016).

2. Related Work

Within the LLOD ecosystem, significant efforts in the cultural heritage domain have focused on museum collections and archival metadata rather than on the language used to describe them. The Europeana Data Model provides a framework for aggregating cultural heritage objects across European institutions (Europeana Foundation, 2017), while CIDOC-CRM (Doerr, 2003) offers a rich ontology for cultural heritage events and objects; neither, however, is designed to represent the multilingual lexical variation that characterises art-historical discourse across centuries.

At the lexicographic level, OntoLex-Lemon (McCrae et al., 2017) has become the de facto standard for representing lexical resources as Linked Data, with applications ranging from general-purpose resources such as English WordNet (McCrae et al., 2014, 2020) to domain-specific terminological databases. Its application to art-historical and cultural heritage lexica specifically, however, remains relatively underexplored.

The LBC project addresses precisely this gap: its diachronic, multilingual corpus and existing lexicographic infrastructure make it a strong candidate for OntoLex-Lemon encoding, though substantial work remains before a full LLOD publication can be achieved.

3. The LBC Resource

3.1. Corpus composition

The LBC corpus spans roughly five centuries of art-historical writing, forming a continuum from Renaissance treatises to contemporary tourist publications, with source texts in different European languages.

Renaissance and early modern treatises. The earliest layer of the corpus consists of Italian treatises that established the vocabulary of Western art history: Leon Battista Alberti's *I dieci libri dell'architettura* (Alberti, 1550) and *Della pittura* (Alberti, 1436), Leonardo da Vinci's *Trattato della pittura* (da Vinci, 1651) (compiled posthumously), and Benvenuto Cellini's *Vita* (Cellini, 1728) (written c. 1558–1563). These texts are available in Italian with translations in the project's target languages.

Vasari's *Vite*. Vasari's *Le Vite de' più eccellenti pittori, scultori e architettori* (Vasari, 1568) is the foundational text of Western art historiography. Written in sixteenth-century Tuscan Italian, it provides technical descriptions and related terminology of paintings, sculptures, and architectural works alongside biographical accounts of Renaissance artists. The LBC project has assembled the Italian original together with translations in other project languages, forming a set of *parallel corpora* of exceptional depth.

Nineteenth-century critical writing. Texts such as Stendhal's *Rome, Naples et Florence* (Stendhal, 1817) and Burckhardt's *Geschichte der Renaissance in Italien* (Burckhardt, 1867) introduce a non-Italian scholarly perspective on Florentine cultural heritage and expand the source language range of the corpus beyond Italian.

Contemporary tourist guides. Tourist guides to Florence (Ciccarelli Roming et al., 2016) provide an accessible, descriptive register sharing the same terminological core as the historical sources while operating under very different communicative constraints. Together with the earlier texts, they form a *comparable corpus* that enables diachronic and register-based analysis.

3.2. Corpus access and current processing

Corpus query interface. Of the full LBC corpus, six sub-corpora have so far been published as individual open-access volumes by Firenze University Press (Lanini, 2024; Carpi and Pano Alamán, 2024; Ballestracci et al., 2024; Natali, 2024; Farina, 2024; Rossi and Zhukova, 2024), all searchable via a local installation of NoSketchEngine (Kilgarriff et al., 2014) hosted at <http://corpora.lessicobeniculturali.net/> (Billero, 2020b).

Users can submit simple keyword queries or formulate complex searches using the Corpus Query Language (CQL), which allows fine-grained retrieval based on word forms, lemmas, and part-of-speech tags. CQL queries can be combined with metadata filters to restrict searches to specific source texts, time periods, or language pairs, making it possible to trace the diachronic evolution of terms across the corpus.

Expert concordance collection. A curated collection of KWIC (Key Word in Context) concordances, selected and annotated by domain experts, is accessible via the lexicon interface at <http://lexicon.lessicobeniculturali.net/>. These concordances are linked to specific lemmas and are currently available for four of the seven target languages: French, German, Russian, and Spanish, providing high-quality usage examples of domain-specific terms across languages and registers.

Linguistic annotation and term extraction. The corpus texts have already undergone automatic linguistic annotation, including tokenisation, part-of-speech tagging, and lemmatisation, carried out with TreeTagger (Schmid, 1994, 1995) and related tools for the languages supported. Domain-specific terms have subsequently been extracted and are linked to the lemma entries in the lexicon interface. These preprocessing steps represent a significant investment and provide a solid foundation for subsequent LLOD modelling, though the accuracy of automatic annotation for historical and specialised texts poses known challenges that will require targeted post-correction before LLOD encoding.

Prototype dictionary. A prototype dictionary, organised as bilingual entries from Italian to each target language, is under development and is intended for open-access publication by Firenze University Press (Farina et al., 2024).

3.3. Lexical content and scope

The LBC targets three main categories of lexical items, each posing distinct challenges for LLOD representation:

- **Common nouns of art:** technical terms describing techniques, materials, styles, and genres (*affresco*, *chiaroscuro*, *pietra serena*, *predella*). Many of these terms originate in Italian and have been borrowed or calqued into other European languages, making Italian the primary source for terminological standardisation in art history.
- **Proper nouns specific to Florence and its history:** artists, patrons, political figures, place names, street names, monuments, and titles of artworks.
- **Florentine *realia*:** culturally bound items including foods, festivals, and practices that are specific to the Florentine cultural context and resist direct translation.

This tripartite structure, together with the diachronic variation of the language across five centuries of texts, maps naturally onto the distinctions drawn in LLOD modelling between lexical entries, named entities, and culture-specific concepts. Beyond these lexical categories, two further challenges arise from the corpus itself: the multilingual origin of its source texts (translation directionality) and its five-century diachronic span (register variation), both discussed in Section 4.

4. Challenges for LLOD Modelling

4.1. Historical and multilingual proper nouns

The LBC corpus contains a large number of historically attested proper nouns recorded under different orthographic forms and in different scripts across the seven languages. *Michelangelo Buonarroti* appears as *Michel-Ange* in French, *Miguel Ángel* in Spanish, and under Cyrillic transliterations in Russian.

Wikidata (Vrandecic and Kröttsch, 2014) already provides multilingual labels for most major artists and monuments and is available as LLOD. Linking LBC entries to Wikidata Q-items via `owl:sameAs` or `skos:exactMatch` would provide a ready-made multilingual backbone. However, less prominent figures — minor painters, local patrons, Florentine guild officials — may be absent from Wikidata or poorly described, requiring new entity creation or the use of more specialised resources such as the Getty Union List of Artist Names (ULAN) (Getty Research Institute,

2024c). A similar challenge arises for historical place names: Florentine toponyms appear under different forms across languages and centuries — from *Firenze* to *Florence*, *Florenz*, and *Fiorenza* — and their alignment to the Getty Thesaurus of Geographic Names (TGN) (Getty Research Institute, 2024b) will require careful historical disambiguation.

4.2. Modelling *realia*

Realia — culture-specific items that lack direct equivalents in other languages — are one of the central concerns of the LBC project (Farina, 2014) and one of the most theoretically interesting challenges for LLOD modelling. A term such as *calcio storico* (a historical Florentine ball game) or *canto* (a civic district of Florence) does not simply translate; it *migrates* into target languages with varying degrees of adaptation, borrowing, paraphrase, or loss.

Standard LLOD formalisms such as OntoLex-Lemon (McCrae et al., 2017) are well suited to representing translation equivalents when equivalence holds, but are less expressive for capturing *degrees of cultural approximation*. One possible approach is to model *realia* as `ontolex:LexicalConcept` instances with asymmetric `skos:closeMatch` or `skos:relatedMatch` relations to their approximate translations, supplementing these with `skos:note` annotations that capture the nature of the mismatch.

The same framework handles specialised art vocabulary more straightforwardly: *affresco* would be encoded as an `ontolex:LexicalEntry` linked to a `ontolex:LexicalConcept`, with translation equivalents (*fresco* in Spanish, *fresque* *fresque* in French, *Fresko*) in German) in their respective lexicons all pointing to the same concept node, itself aligned via `skos:exactMatch` to Getty AAT 300177433 (*frescoes*).

The following sketches illustrate the proposed encoding for *affresco*:

```
:affresco_it a ontolex:LexicalEntry ;
  ontolex:language "it" ;
  ontolex:sense :affresco_sense .
:affresco_sense
  ontolex:reference :fresco_concept .
:fresco_concept a ontolex:LexicalConcept ;
  skos:exactMatch aat:300177433 .
:fresco_en a ontolex:LexicalEntry ;
  ontolex:language "en" ;
  ontolex:sense :fresco_en_sense .
:fresco_en_sense
  ontolex:reference :fresco_concept .
```

4.3. Multilingual source texts and translation directionality

Unlike most parallel corpora used in LLOD projects, the LBC corpus includes texts originally composed in several languages, not only Italian. This complicates standard assumptions about translation directionality: rather than a single source language with multiple target translations, the corpus presents a web of source-target relationships that varies by text. For instance, Burckhardt’s German text on the Italian Renaissance has been translated into Italian, making Italian both a source and a target language depending on the text pair considered. The LIME metadata module (Fiorelli et al., 2015) for OntoLex-Lemon (McCrae et al., 2017) would need to be applied carefully to represent these relationships without forcing Italian into the role of sole pivot language. Concretely, each source text would be associated with a distinct `lime:Lexicon` instance carrying its own `lime:language` and `lime:linguisticCatalog` properties. Translation relations between lexicons would be expressed at the `ontolex:LexicalSense` level using `vartrans:translatableAs` from the OntoLex Vartrans module, allowing Burckhardt’s German lexicon to relate directly to its Italian and other translations without Italian acting as intermediary.

4.4. Diachronic register variation

The juxtaposition of fifteenth-century treatises with twenty-first-century tourist guides raises a fundamental issue for lexical representation: the same term may carry different denotations, connotations, or pragmatic values across registers and centuries. *Maniera*, for instance, is a key Vasarian aesthetic concept but functions as a generic stylistic descriptor in contemporary tourist texts.

OntoLex-Lemon’s `ontolex:usage` and `lexinfo:register` (Cimiano et al., 2011) properties can capture register distinctions, while separate `lime:Lexicon` instances could encode the different sub-corpora as distinct lexical perspectives on the same domain ontology.

4.5. Alignment with existing LLOD resources

The cultural heritage domain is one of the few areas where substantial, professionally curated LLOD-compatible resources already exist:

- **Getty Art & Architecture Thesaurus (AAT)** (Getty Research Institute, 2024a): hierarchical vocabulary of artistic concepts and techniques, available as Linked Data at <https://vocab.getty.edu/aat/>.

- **Getty Union List of Artist Names (ULAN)** (Getty Research Institute, 2024c): authority file for artists and architects with multilingual labels, at <https://vocab.getty.edu/ulan/>.
- **Getty Thesaurus of Geographic Names (TGN)** (Getty Research Institute, 2024b): geographic thesaurus covering historical and contemporary place names, at <https://vocab.getty.edu/tgn/>.
- **Wikidata** (Vrandečić and Krötzsch, 2014): encyclopaedic knowledge graph with strong multilingual coverage of artists, artworks, and monuments, queryable via SPARQL at <https://query.wikidata.org/>.

Aligning LBC entries with these resources would immediately embed the lexicon into a rich semantic context. For example, the LBC entry for *pietra serena* — the grey sandstone characteristic of Florentine Renaissance architecture — could be linked to the corresponding Getty AAT concept, while *Brunelleschi* could be linked via `skos:exactMatch` to Getty ULAN 500018169 and Wikidata Q174330, simultaneously enriching all three knowledge graphs with the Italian-language perspective of the LBC. A key challenge is that the Getty vocabularies are English-centric and do not always provide the nuanced Italian-language perspective and multilingual dimension that the LBC is designed to offer, and many minor historical figures in the corpus are absent from all existing authority files.

5. Towards an LLOD Roadmap for LBC

We propose the following incremental steps towards the publication of the LBC as an LLOD resource, starting from the processing already completed.

Step 1 — Quality assessment and post-correction of existing annotation. The corpus has already been tokenised, POS-tagged, and lemmatised using TreeTagger (Schmid, 1994, 1995) and related tools, and domain terms have been extracted and linked to lemma entries in the lexicon interface. TreeTagger offers robust models for several of the languages covered by the LBC, but coverage is uneven across all languages of the project, and for some target languages dedicated tools are required. A further source of degradation is the diachronic range of the corpus: TreeTagger models are trained on contemporary corpora, and their performance degrades on historical texts where orthographic and morphological variation is

substantial, as well as on domain-specific nominal compounds. Quality assessment will therefore require targeted manual post-correction by specialists with both technical and linguistic competence in the individual languages. Error typologies identified during this phase will inform post-correction heuristics applicable to similar multilingual historical corpora.

For the historical Italian sub-corpus in particular, the main difficulties stem from pre-standardisation graphemic and orthographic variability; archaic or obsolete morphological forms not always recognised by models trained on contemporary Italian; specialist humanistic vocabulary not found in standard dictionaries; and interference arising from modern editorial conventions. The most frequent errors relate to POS disambiguation in the presence of archaic spellings, incorrect lemmatisation of historical verb forms, and incorrect recognition of proper names and Latinisms. Phase 1 therefore does not involve fully automated annotation, but rather a hybrid workflow: post-correction will be supported by specific guidelines for historical Italian, developed in collaboration with Italian linguists, with the aim of achieving a level of reliability consistent with the objectives of semantic querying and LLOD modelling.

Step 2 — OntoLex-Lemon encoding of extracted terms. Encode the existing extracted terms in OntoLex-Lemon (McCrae et al., 2017), with separate `lime:Lexicon` instances for each language and for each diachronic sub-corpus. The expert-curated concordances already available in the lexicon interface for French, German, Russian, and Spanish can serve as high-quality usage examples within the `ontolex:LexicalSense` entries.

Step 3 — Entity linking and LLOD alignment. Link named entities (artists, places, artworks) to Getty ULAN (Getty Research Institute, 2024c), Getty TGN (Getty Research Institute, 2024b), and Wikidata (Vrandečić and Krötzsch, 2014) via `owl:sameAs` and `skos:exactMatch`. Link conceptual terms (art techniques, styles, genres) to the Getty AAT (Getty Research Institute, 2024a).

Step 4 — Modelling of the specialised artistic lexicon, *realia*, and translation asymmetries. Develop a lightweight extension to the OntoLex-Lemon model to represent degrees of translational equivalence for both specialised artistic vocabulary and culture-specific *realia*, building on the SKOS mapping vocabulary (Miles and Bechhofer, 2009) and drawing on literature on the lexicographic treatment of culture-specific items (Aixelà, 1996).

Step 5 — Publication and interlinking. Publish the resulting lexicon as RDF under an open licence, with a SPARQL endpoint, and submit it for inclusion in the LLOD cloud diagram (Chiarcos et al., 2012).

Steps 1–2 are the immediate next objectives; Steps 3–5 represent medium-term goals for which community input is actively sought.

6. Conclusion

We have presented the LBC project as a candidate for LLOD publication, described its corpus and current state of processing, and outlined the principal challenges its conversion to Linked Data poses. The cultural heritage domain is underrepresented in the LLOD cloud despite the existence of high-quality complementary resources such as the Getty Vocabularies; the LBC has the potential to fill a significant gap, particularly with respect to Italian-centric multilingual coverage and the lexicography of historical texts.

We therefore invite members of the LDL community to share experience with OntoLex-Lemon modelling for specialised or historical lexica, strategies for representing culture-specific concepts and translational asymmetries in RDF, and alignment methodologies for parallel corpora and knowledge graphs.

We hope this paper will serve as a starting point for a productive discussion at LDL-2026, contribute to the growing intersection of digital humanities and the Semantic Web, and offer a replicable model for other historical multilingual cultural heritage lexica currently outside the Linked Data cloud.

7. Limitations

The work presented here is at an early stage. No LLOD implementation has yet been produced, and the modelling proposals in Section 5 remain programmatic. The accuracy of TreeTagger on historical Italian is known to be lower than on contemporary language: lemmatisation errors are particularly frequent for sixteenth-century verb forms and domain-specific nominal compounds, and will require targeted manual post-correction before high-quality LLOD encoding can be achieved. The proposed treatment of *realia* via SKOS mapping properties is theoretically motivated but has not yet been validated against the full range of cases in the corpus. Finally, the alignment with Getty and Wikidata has been illustrated with selected examples only; systematic alignment will require semi-automatic methods and expert validation.

8. Acknowledgements

The *Lessico dei Beni Culturali* project is hosted at the Department of Education, Languages, Inter-cultures, Literatures and Psychology (FORLILPSI) of the University of Florence and is published by Firenze University Press.

9. Bibliographical References

- Javier Franco Aixelà. 1996. Culture-specific items in translation. In Román Álvarez and M. Carmen África Vidal, editors, *Translation, Power, Subversion*, pages 52–78. Multilingual Matters, Clevedon.
- Leon Battista Alberti. 1436. *Della pittura*. Three books. Italian version of *De pictura*.
- Leon Battista Alberti. 1550. *I dieci libri dell'architettura*. Lorenzo Torrentino, Florence. Italian translation of *De re aedificatoria* (1452).
- Sabrina Ballestracci. 2023. Da un segno a tanti segni. L'emergere della polisemia del termine “disegno” nelle traduzioni tedesche delle “Vite” di Vasari. In Valeria Zotti and Monica Turci, editors, *Nuove strategie per la traduzione del lessico artistico: da Giorgio Vasari a un corpus plurilingue dei beni culturali*, pages 21–38. Firenze University Press, Firenze.
- Riccardo Billero. 2020a. Cultural heritage lexicon: A case study. In Ana Pano Alamán and Valeria Zotti, editors, *The Language of Art and Cultural Heritage: A Plurilingual and Digital Perspective*, pages 86–103. Cambridge Scholars Publishing.
- Riccardo Billero. 2020b. [Implementazione di software per la gestione dei corpora LBC](#). In Riccardo Billero, Annick Farina, and María Carlota Nicolás Martínez, editors, *I Corpora LBC. Informatica Umanistica per il Lessico dei Beni Culturali*, pages 19–32. Firenze University Press, Firenze.
- Riccardo Billero, Annick Farina, and María Carlota Nicolás Martínez. 2020. [I Corpora LBC. Informatica Umanistica per il Lessico dei Beni Culturali](#). Firenze University Press, Firenze.
- Riccardo Billero and María Carlota Nicolás Martínez. 2017. Nuove risorse per la ricerca del lessico del patrimonio culturale: corpora multilingue LBC. *CHIMERA: Romance Corpora and Linguistic Studies*, 4(2).
- Jacob Burckhardt. 1867. *Geschichte der Renaissance in Italien*. Ebner & Seubert, Stuttgart.
- Benvenuto Cellini. 1728. *Vita*. Naples. Written c. 1558–1563; first printed edition.
- Christian Chiarcos, Sebastian Nordhoff, and Sebastian Hellmann. 2012. Linked data in linguistics: Representing and connecting language data and language metadata. In *Proceedings of the 1st Workshop on Linked Data in Linguistics (LDL-2012)*, Frankfurt am Main, Germany. CEUR Workshop Proceedings.
- C. Ciccarelli Roming, T. Jepson, and T. Fisher. 2016. *Florenz. Perfekte Tage in der Toskana – Metropole*. Taschen, München.
- Philipp Cimiano, Paul Buitelaar, John McCrae, and Michael Sintek. 2011. LexInfo: A declarative model for the lexicon-ontology interface. *Journal of Web Semantics*, 9(1):29–51.
- Leonardo da Vinci. 1651. *Trattato della pittura*. Compiled posthumously; first printed edition, Paris: Giacomo Langlois.
- Martin Doerr. 2003. The CIDOC conceptual reference module: An ontological approach to semantic interoperability of metadata. *AI Magazine*, 24(3):75–92.
- Europeana Foundation. 2017. European data model documentation. <https://pro.europeana.eu/page/edm-documentation>. Accessed: February 2026.
- Annick Farina. 2014. Descrivere e tradurre il patrimonio gastronomico italiano: le proposte del lessico plurilingue dei Beni Culturali. In Francesca Chessa, Cosimo De Giovanni, and Maria Teresa Zanola, editors, *La terminologia dell'agroalimentare*, pages 55–66. Franco Angeli, Milano.
- Annick Farina, Riccardo Billero, and María Carlota Nicolás Martínez. 2024. [Presentation of the LBC database](#). In I. Natali, editor, *The LBC English Corpus*. Firenze University Press, Firenze. E-ISBN: 979-12-215-0307-4. <https://corpora.lessicobeniculturali.net/en/>.
- Manuel Fiorelli, Armando Stellato, John P. McCrae, Philipp Cimiano, and Maria Teresa Pazienza. 2015. LIME: The metadata module for OntoLex-Lemon. In *Proceedings of the 12th European Semantic Web Conference (ESWC 2015)*, pages 321–336, Portorož, Slovenia. Springer.
- Carolina Flinz, Valeria Zotti, D. Henkel, and Sabrina Ballestracci. 2024. A multilingual parallel corpus for the lexical information system LBC:

- Recent progress and future perspectives. In *Lexicography and Semantics*. Cambridge Scholars Publishing.
- Marcello Garzaniti and Annick Farina. 2013. Un portale per la comunicazione e la divulgazione del patrimonio culturale: progettare un lessico multilingue dei beni culturali on-line. In A. Filipovic and W. Troiano, editors, *Strategie e Programmazione della Conservazione e Trasmissibilità del Patrimonio Culturale*, pages 500–509. Edizioni scientifiche Fidei Signa, Roma.
- Getty Research Institute. 2024a. Art & architecture thesaurus online. <https://vocab.getty.edu/aat/>. Accessed: February 2026.
- Getty Research Institute. 2024b. Thesaurus of geographic names online. <https://vocab.getty.edu/tgn/>. Accessed: February 2026.
- Getty Research Institute. 2024c. Union list of artist names online. <https://vocab.getty.edu/ulan/>. Accessed: February 2026.
- Adam Kilgarriff, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, and Vít Suchomel. 2014. *The Sketch Engine: Ten years on*. In *Lexicography*, volume 1, pages 7–36.
- John P. McCrae, Julia Bosque-Gil, Jorge Gracia, Paul Buitelaar, and Philipp Cimiano. 2017. The OntoLex-Lemon model: Development and applications. In *Proceedings of eLex 2017*, pages 587–597, Leiden, The Netherlands.
- John P. McCrae, Christiane Fellbaum, and Philipp Cimiano. 2014. Publishing and linking WordNet using Lemon and RDF. In *Proceedings of the 3rd Workshop on Linked Data in Linguistics (LDL-2014)*, Reykjavik, Iceland. European Language Resources Association (ELRA).
- John P. McCrae, Alexandre Rademaker, Ewa Rudnicka, and Francis Bond. 2020. *English WordNet 2020: Improving and extending a WordNet for English using an open-source methodology*. In *Proceedings of the LREC 2020 Workshop on Multimodal Wordnets (MMW2020)*, pages 14–19, Marseille, France. European Language Resources Association (ELRA).
- Alistair Miles and Sean Bechhofer. 2009. SKOS simple knowledge organization system reference. W3C Recommendation, <https://www.w3.org/TR/skos-reference/>.
- Helmut Schmid. 1994. Probabilistic part-of-speech tagging using decision trees. In *Proceedings of the International Conference on New Methods in Language Processing*, pages 44–49, Manchester, UK.
- Helmut Schmid. 1995. Improvements in part-of-speech tagging with an application to German. In *Proceedings of the ACL SIGDAT-Workshop*, Dublin, Ireland.
- Stendhal. 1817. *Rome, Naples et Florence*. De-launay, Paris.
- Giorgio Vasari. 1568. *Le Vite de' più eccellenti pittori, scultori e architettori*. Giunti, Florence. Second enlarged edition; first edition 1550.
- Denny Vrandečić and Markus Krötzsch. 2014. *Wikidata: A free collaborative knowledge base*. *Communications of the ACM*, 57(10):78–85.

10. Language Resource References

- Sabrina Ballestracci, Claudia Buffagni, and Carolina Flinz. 2024. *Das deutsche LBC-korpus*. E-ISBN: 979-12-215-0311-1. <https://corpora.lessicobeniculturali.net/de/>.
- Elena Carpi and Ana Pano Alamán. 2024. *Corpus LBC español*. E-ISBN: 978-88-5518-035-1. <https://corpora.lessicobeniculturali.net/es/>.
- Annick Farina. 2024. *Corpus LBC français*. E-ISBN: 979-12-215-0309-8. <https://corpora.lessicobeniculturali.net/fr/>.
- Ludovica Lanini. 2024. *Corpus LBC italiano*. E-ISBN: 979-12-215-0305-0. <https://corpora.lessicobeniculturali.net/it/>.
- Ilaria Natali. 2024. *The LBC English corpus*. E-ISBN: 979-12-215-0307-4. <https://corpora.lessicobeniculturali.net/en/>.
- Valentina Rossi and Natalia Zhukova. 2024. *Russkij korpus «Leksika kul'turnogo nasledija»*. E-ISBN: 979-12-215-0313-5. <https://corpora.lessicobeniculturali.net/ru/>.