

The Eurasian Latin Archive

From Modelling to Corpus Use

Emmanuela Carbé

Università Ca' Foscari Venezia
emmanuela.carbe@unive.it

Abstract

The *Eurasian Latin Archive (ELA)* is a digital platform designed to collect and analyse Latin and multilingual texts relating to East Asia from the medieval and early modern periods. Initially developed within the *DAS-MeMo* project (2018-2020) and subsequently expanded through the *SERICA* framework, *ELA* offers textual data and computational tools for corpus-based research. This article presents *ELA* as a research environment, outlining its data model, technical architecture, and NLP workflow. Through selected examples, it shows how users can query, filter and compare documents. Beyond a functional overview, the article reflects on some methodological challenges in modelling multilingual Latin corpora and sketches possible directions for future extensions of the project.

Keywords: Data Modelling; NLP; Digital Platform; Multilingual texts; Eurasian Latin.

Eurasian Latin Archive (ELA) è una piattaforma digitale progettata per raccogliere e analizzare testi latini e multilingui relativi all'Asia orientale dal periodo medievale alla prima età moderna. Inizialmente sviluppata nell'ambito del progetto DAS-MeMo (2018-2020) e successivamente ampliata attraverso il framework SERICA, ELA mette a disposizione dati testuali e strumenti computazionali per la ricerca basata su corpora. Questo contributo presenta ELA come ambiente di ricerca, delineandone il modello dei dati, l'architettura tecnica e il flusso di lavoro di elaborazione del linguaggio naturale (NLP). Attraverso alcuni esempi selezionati, mostra come gli utenti possano interrogare, filtrare e confrontare i documenti. Oltre a una panoramica funzionale, l'articolo riflette su alcune sfide metodologiche nella modellazione di corpora in lingua latina e multilingui abbozzando possibili direzioni per gli sviluppi futuri del progetto.

Parole chiave: Modellazione dei dati; NLP; Piattaforma digitale; Testi multilingue; Latino eurasiatico.

1. Introduction

The *Eurasian Latin Archive (ELA)* is a digital platform devoted to Latin and multilingual texts concerning East Asia from the medieval and early modern periods. Directed by Francesco Stella, the project is rooted in the tradition of the Centre for Comparative Studies “I Deug-Su” and

focuses on texts that document the circulation of Latin across Asia, especially in missionary contexts, and that negotiate cultural difference through a learned yet increasingly hybrid language, marked by both diachronic and synchronic variation ([12], [13]).

The theoretical premises of *ELA* were first outlined by Stella in 2017 at the conferences *Medialatinitas* in Vienna and *Silk Roads Again II: Eurasian Studies in the Digital Age* in Seoul [11] and were subsequently developed within the two-year project *DAS-MeMo*¹. In this start-up phase (2018-2020) the team defined the conceptual model, established a workflow for digitisation and TEI encoding, implemented an initial prototype of the Digital Library on a Java EE and Elastic-Search architecture, and produced a first version of the NLP suite *ELA Tools* ([3], [4])². A public prototype of the platform was released in January 2020 and presented at the first *Global Latin. Exploratory Workshop* in Siena, where it was tested by an international group of researchers ([12], [2]).

A second phase of development took place in 2022 within the *SERICA* project [4]. In this phase the user interface of the *ELA* platform was refined, downloadable formats and public APIs were introduced, the statistical interface was extended so that users can work more easily with selections of documents as *subcorpora*. In addition, the architecture was configured so that TEI-encoded documents can point to facsimiles stored in external repositories via references in the TEI documents.

ELA addresses a specific gap in the current infrastructure for Latin studies. Large-scale digital libraries of Latin texts tend to privilege canonical and predominantly European corpora, while missionary letters, reports, grammars, and chronicles concerning East Asia ([7], [8]) are scattered across archives, unevenly digitised, and often linguistically hybrid, presenting encoding, NLP pipelines and philological issues that have not yet been fully addressed. *ELA* proposes a focused, research-driven corpus in which different forms of Latin systematically interact within a multilingual and multicultural substrate, and in which questions about cultural translation and global circulation can be pursued through both close reading and quantitative analysis.

In the light of these eight years of *Eurasian Latin Archive*, it seems useful to reconsider the project both as a model and as a working environment: first by retracing its passage from conceptual design to technical implementation, and then by examining how the platform is used in practice, with a view to inviting discussion within the wider community of scholars working on digital Latin corpora and Eurasian studies.

2. Conceptual and Formal Structure

ELA starts from a modelling problem: how can models and metamodels of text encoding be brought together to represent the semantic field of complex documents and, at the same time, obtain reliable linguistic data for corpus analysis? The corpus consists of medieval and early modern Latin texts that are neither monolingual nor confined to a European horizon, and whose

¹ The website of the *DAS-MeMo* project is available at <https://dasmemo.unisi.it/en/> (11/25).

² The *ELA* Platform is available at ela.unisi.it. The *ELA Tools* suite, co-developed by Francesco Garosi, is published on GitHub (<https://github.com/EurasianLatinArchive>). A fixed 2025 version of *ELA Tools*, prepared by Giulio Quaresima, is now available at https://github.com/DFCLAM/ELA_TOOLS/tree/master/TOOLS (see Modelling linguistic data: *ELA Tools*).

Latin is often far from the classical norm, marked by missionary usage, orthographic variation and interference from other languages. The documents identified are frequently multilingual, mixing Latin with vernacular Italian, Portuguese, Spanish, French or Chinese on the same page and sometimes within a single sentence. Their transmission is equally heterogeneous, ranging from printed series and composite volumes to single manuscripts now preserved in institutional archives, and extending to modern collections edited according to very different criteria.

As a first step, this modelling question was grounded in an empirical corpus through a systematic census of sources: more than 300 items were identified and described using *OpenRefine* as a data-management environment. Each record combined a Dublin Core metadata set with links to external authority files and catalogues (*VIAF*, *WorldCat*, *Wikidata*) and to digital libraries where scans or transcriptions were already available. At the same time, the record traced the internal history of each item within *ELA* (e.g., who digitised it and when, which OCR software was used, who corrected the transcription, which licence applies). In this way the census evolved from a simple list of sources into a first instantiation of the data model and later became the backbone of the TEI headers used to encode the documents.

Conceptually, the model was shaped in dialogue with *ALIM* (*Archivio della Latinità Italiana del Medioevo*)³, with its architecture and its use of TEI P5 for the document's editions. *ELA* inherits from *ALIM* the idea of a modular, standards-based Digital Library, but adapts it to a different set of problems. Instead of focusing on a relatively homogeneous tradition (Latin texts produced in medieval Italy), it had to deal with a corpus defined less by geography than by contact zones, where Latin describes and translates East Asia in a variety of communicative situations. This required a stronger emphasis on multilingual markup, on authority control for places and persons, and on the explicit recording of language mixtures.

On the formal side, these requirements converged in a customised TEI P5 schema. The project adopted a compact subset focused on the basic hierarchy of `<text>`, `<body>`, `<div>`, and `<p>`. The critical apparatus has been deliberately left aside. The priority was to include elements and attributes needed to register linguistic and semantic information, rather than variant readings and editorial layers. Particular attention was given to the `<foreign>` element and to the systematic use of the `@xml:lang` attribute, which together allow code-switching to be marked at word level and make multilingual passages visible to NLP tools. A pair of short extracts from Michele Ruggieri's Latin compilation of Confucian doctrine illustrates this approach:

```
<head>
<foreign xml:lang="zh">大学</foreign> Liber Primus. Humana
Institutio
</head>

<p>Regia idest <placeName key="Khanbaliq"
ref="https://www.geonames.org/1816670/">Pacchin</placeName>
et <placeName key="Nanking"
ref="https://www.geonames.org/1799962/">Nanchin</placeName>
```

³ The Digital Library is available at <https://en.alim.unisi.it/> (11/25).

quae mille stadiis ambitu est, tantum secunda populi
habitatio est.</p>

<p>In Poematibus est: aves dulciter canentes aureique
coloris considunt in vertice montium, qui vocantur Chieu et
Ju. Quo loco <persName key="Confucius"
ref="http://www.viaf.org/89664672">Confucius</persName> ait,
loca tuta aves sibi delegunt, an decet homines esse avibus
inferiores?</p>⁴

Here the Chinese title of the Confucian text is encoded as foreign material in the heading, with its language attribute, while the toponyms are linked to GeoNames and Confucius is associated with a VIAF identifier. The same strategy is applied to other language insertions, so that the linguistic profile of each document can be reconstructed both qualitatively and through automatic analysis. Rather than normalising away such variation, the model preserves it as evidence. In addition to multilingual markup, abbreviations are encoded using the TEI <choice> mechanism, pairing <abbr> and <expan>. For instance:

```
<choice>
  <abbr>Illmae</abbr>
  <expan>Illustrissimae</expan>
</choice>
```

This representation preserves the original siglum while also exposing an expanded reading that can, where appropriate, be used as a more regularised string in tokenisation and lemmatisation workflows. The aim is not to impose a single normalisation strategy, but to keep open both the documentary form and a form that may be more tractable for downstream analysis.

3. Modelling linguistic data: *ELA* Tools

The TEI model is also the starting point of a chain of computational representations. The aim is to let the same textual object circulate across different layers: a textual layer (the encoded transcription), a linguistic layer (tokens, lemmas, part-of-speech), and a semantic layer (named entities referring to external authorities). To explore this question, in 2019 Francesco Garosi co-developed an experimental toolkit, the *ELA* Tools, adapted to the current infrastructure in 2025 by Giulio Quaresima.

The toolkit takes as input the TEI files and processes them through CLTK and NLTK [1], generating a JSON file in which each token is associated with its lemma, part-of-speech tag, and basic distributional information. In this pipeline the TEI markup is not a mere pretext: language tags and <foreign> elements guide the selection of tokens that are sent to the “Latin lemmatiser”, while <choice> structures, with <abbr>/<expan> pairs, provide expanded forms that can be used to obtain more stable lemmas in highly abbreviated texts. On top of this enriched representation, the tool computes standard corpus-linguistic measures (frequency lists,

⁴ Michaelis Rogerius, *Humanae institutionis ratio* [6].

TTR, concordances, collocations and n-grams). In the *ELA* platform these JSON files act as a bridge between the TEI layer and the interactive interface [Fig. 1].

Info ×

Info Stats Words Latin Words Latin Stop Words Lemmas Collocations N-Grams Persons Places Maps

Show 10 entries Search:

Number of words	797
Number of Latin words	797
Number of Latin words without stop words	596
Numbers of Latin lemmas	441
Number of Latin types	540
Min lengths of Latin types	1
Max lengths of Latin types	15
Average lengths of Latin types	6.9463
Latin type-token ratio (TTR)	0.6775

Showing 1 to 9 of 9 entries Previous 1 Next

Fig. 1 – Statistical panel generated by *ELA* Tools for a letter by Andreas de Perusio [14]: 372-377

The results of the *ELA* Tool thus enable the visualisation of statistical panels on the platform and exploratory views that allow users to move from individual passages to patterns emerging across the entire corpus, including, where possible, a visualisation of the georeferencing of the places mentioned in the document [Fig. 2].

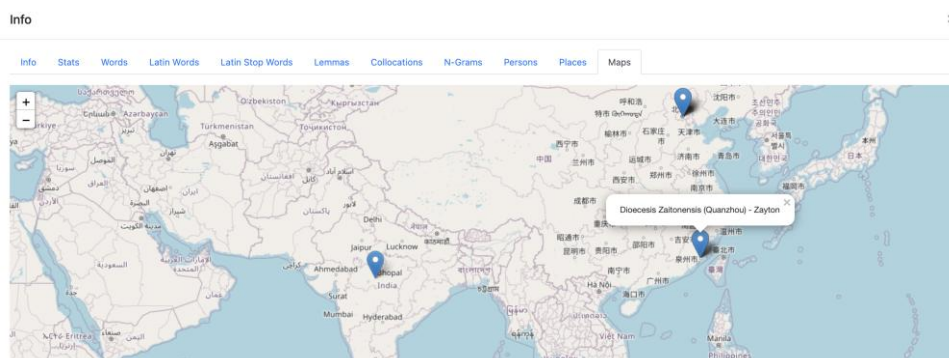


Fig. 2 – “Maps” view for an *Epistola* of Andreas de Perusio [14]: 372-377, displaying the georeferenced distribution of the places mentioned in the text.

The same infrastructure is used for what we call re-tagging experiments. Starting from the named entities already identified and encoded in TEI (person and place names linked to *VIAF*, *Wikidata*, *Pleiades* or *GeoNames*), the tool builds lookup tables that can be applied to new texts or to revised versions of the same text. In this way, occurrences of known entities are tagged automatically in the TEI files. This semi-automatic propagation of tags is intentionally conservative: it accelerates the markup of recurrent names and toponyms, but it still relies on a

controlled vocabulary and on manual verification, especially in ambiguous cases such as homonyms, overlapping designations or shifts between places.

At the present stage, *ELA Tools* still focuses on the Latin layer of the corpus. Non-Latin segments marked in TEI with <foreign> and the relevant @xml:lang values are preserved and indexed, but they are treated as opaque strings in the NLP pipeline. This is a conscious limitation of the current prototype, which relies on general-purpose Latin models in CLTK, trained primarily on more standard varieties of Latin. Their performance is acceptable for basic exploration but less reliable when confronted with missionary usage, heavy orthographic variation, or interference from other languages. Future work will therefore include the creation of domain-specific training sets and the systematic evaluation of alternative lemmatisers, with the aim of increasing the precision and recall of lemmatisation on post-classical Latin ([10], [9]).

In this sense, the toolkit is not presented as a finished solution but as a laboratory in which the adequacy of the TEI model can be tested against concrete NLP tasks. Misalignments between lemmas and orthographic variants, discrepancies between automatically recognised entities and those encoded in TEI, or tensions between language tags and actual usage become diagnostic signals. They feed back into the ongoing process of modelling Eurasian Latin and into the next steps, where the main challenge will be to bring the non-Latin component of the corpus into the same analytical frame.

Finally, *ELA Tools* is accompanied by a small set of Jupyter notebooks. Among these, GlobalStats⁵ aggregates the statistics produced by *ELA Tools* to compute corpus-level measures. GlobalStats can be run on the whole *ELA* collection or on *subcorpora* organised as flat directories, and produces reports, including token and lemma counts, TTR, frequency lists, and distributions of persons and places. These outputs can then be reused for more specific experiments and visualisations.

4. Exploring the corpus: search, *subcorpora*, and APIs

From the user’s point of view, *ELA* can be understood as a search environment that exposes different levels of description of the same objects: bibliographical metadata, encoded texts, linguistic annotation, and named-entity information. The interface offers several entry points: “document list”, “browsing views” and “full-text search”, which can be combined in complex queries. In the second development phase, supported by the *SERICA* Project, this search layer was extended with tools for building *subcorpora* and for running comparative statistics directly in the web interface.

The most immediate access to the corpus is provided by the Document list [Fig. 3]. Here users can see, in tabular form, the list of the documents available with basic metadata (Author, Title, Date). From this list it is possible to visualize the text and the associated information and statistical panels, moving quickly between descriptive metadata, transcription, and quantitative views.

⁵ See https://github.com/DFCLAM/ELA_TOOLS/blob/master/TOOLS/nbk_globalstats.ipynb.

Author	Title	Date	Share	View	Info
Andreas de Perusio O.F.M.	Epistola	1326			
Augustinus Sancti Paschali O.F.M.	Epistola ad illum. Bernardinum Della Chiesa, 16 Dec. 1684	1684			
Augustinus Sancti Paschali O.F.M.	Responsum ad intimationem D. Maigrot 30 nov. 1689	1689			
Augustinus Sancti Paschali O.F.M.	Responsum ad Procuratorem Generalem, 11 Dec. 1689	1689			
Augustinus Sancti Paschali O.F.M.	Responsa ad D. Carolum Maigrot, 15 et 22 Mart. 1689	1689			

Fig. 3 – The “Document list”

The Browse section offers some exploratory paths: users can navigate the corpus by author, places or persons mentioned in the text, date of the document, document type, collections, following the internal organisation of the archive. Browsing a given author, it is possible to switch to the document view and to the statistical panels [Fig. 4].

Browse by

Author

- Andreas de Perusio O.F.M.
- Augustinus Sancti Paschali O.F.M.
- Benedictus Polonus O.F.M.
- Cardim, Antonius Franciscus S.J.
- Colladus, Didacus O.P.
- Criminalis, Antonius S.J.
- Franciscus Montefuscus S.J.
- Gonçalves de Câmara, Ludovicus S.J.
- Gulielmus de Rubruquis O.F.M.
- Ibañez, Bonaventura O.F.M.
- Ignotus
- Intorcetta, Prosperus S.J.
- Iohannes de Marignollis O.F.M.

Author	Title	Date	Share	View	Info
Ibañez, Bonaventura O.F.M.	Epistola ad Cardinales S.C.P.F., mense Iulio 1672	1672			
Ibañez, Bonaventura O.F.M.	Epistola ad Episcopos Turkinenses, 5 febr. 1674	1674			
Ibañez, Bonaventura O.F.M.	Epistola ad Episcopos Turkinenses, 29 Ian. 1675	1675			
Ibañez, Bonaventura O.F.M.	Epistola ad Cardinales S.C.P.F., 5 nov. 1672	1672			
Ibañez, Bonaventura O.F.M.	Brevis Relatio	1668			

Showing 1 to 5 of 5 entries

Previous **1** Next

Fig. 4 – Browsing by authors

A third entry point is the search interface, which exposes a search engine working on both word forms and lemmas. Queries can use Lucene syntax and combine lexical conditions with metadata filters. Users can perform simple word or lemma searches, exact phrase queries by means of quotation marks, wildcard searches, as well as fuzzy and proximity searches [Fig. 5].

The screenshot shows a search interface with a sidebar on the left and a results table on the right. The sidebar contains a search box with the text "frater missionis"~10, a Lemmas section, and a Filters section with dropdown menus for "Has", "Select tag", and "Select value". Below the filters are "Search" and "Full corpus" buttons. The results table has columns for Author, Title, Date, Compare, Share, View, and Info. The first entry is by Augustinus Sancti Paschali O.F.M. with the title "Responsum ad intimationem D. Maigrot 30 nov. 1689" and the date 1689. The table shows "Showing 1 to 1 of 1 entries" and navigation buttons for "Previous", "1", and "Next".

Fig. 5 – Searching “frater” and “missionis” within the maximum distance of ten words

In the current version of *ELA*, developed thanks to the *SERICA* project, these search options are integrated with tools for working on *subcorpora*. In practical terms, the interface allows researchers to assemble “baskets” of texts to compare them. As an illustration, Fig. 6 compares two seventeenth-century documents: *Brevis Relatio de Numero, et Qualitate Christianorum apud Sinas*, written by Martino Martini (1655), and the *Brevis Relatio* by Bonaventura Ibañez (1668; [15]: 75-94).

The first document is a Jesuit *memorandum* that surveys the geography of the mission, counts Christians by province, and presents the China mission as a largely Jesuit enterprise marked by steady numerical growth. It includes a catalogue of names and numbers of converts. Other orders are mentioned only once: *in Provinciâ Fokien sive Chincheo (quam solam ex quindecim Sinici Imperij frequentârunt alij Religiosi à nostris Ordines)*. By contrast, the Franciscan Bonaventura Ibañez casts Simone Martini not as a neutral chronicler of the China mission but as an active agent of exclusion. He reports concrete episodes, for example when Martini ordered his servants not to allow Dominican and Franciscan fathers to enter his house anymore (*P. Martinus statim præcepit servis suis ne deinceps sinerent P. Dominicanos nec Franciscanos in suam domum ingredi sed dicerent eis se non esse domi*).

The screenshot shows two side-by-side panels, Union A and Union B, each with a search bar and a table of statistics. Union A has 6029 words, 4881 Latin words without stop words, and 6029 Latin lemmas. Union B has 4807 words, 3709 Latin words without stop words, and 4807 Latin lemmas. Both panels show "Showing 1 to 3 of 3 entries" and have "Download all data" and "Download current table" buttons.

Fig. 6 – Comparing Relations with the A | B basket tool

Martini’s text contains 6.215 tokens and has a type-token ratio of 0.487; the average sentence length is 7.50 words. It mentions 53 different persons and 21 different place names. Ibañez’s text is shorter: it contains 4.807 tokens, with a type-token ratio of 0.3686, an average sentence length of 7.33 words, and references to 21 persons and 26 place names.

In what follows, these two documents are analysed with the “Compare” function, assigning Martini’s text to “Basket A” and Ibañez’s to “Basket B”. Once two or more texts have been included in the baskets, the interface aggregates the JSON outputs produced by *ELA Tools* and computes corpus-level measures on the fly. In practice, this makes it possible to move from the close reading of individual letters to systematic comparisons between *subcorpora*.

In this example, by inspecting the “A × B comparison” we can identify the intersection of named entities: both texts mention Franciscus Xaverius and Franciscus Brancati, and they share a common set of place names. Besides China and Rome, the overlap includes Fokien, Xantung, Chekiang, Nanking, and several others [Fig. 7].

Comparison info ×

Stats Words Latin Words Latin Stop Words Lemmas Persons Places

Name	↕	Frequency ↕
China		9
Fokien		4
Roma		3
Chantong		2
Chekiang		2
Lusitania		2
Philippinas		1
Manila		1
Hispania		1
Nanking		1

Showing 1 to 10 of 10 entries Previous 1 Next

Fig. 7 – “A x B comparison” of place entities in the two *Relationes*

If we look at lemmas in Martini that do not appear in Ibañez (“A – B”), a small but coherent group emerges: *fidei*, *ecclesiae*, *virtus* and *doctor* (referring here to Chinese literati and degree-holders). This cluster fits Martini’s tendency to describe the China mission in explicitly ecclesial terms and in connection with high-status groups. The lemmas in Ibañez’s text that are absent from Martini’s *relatio* (“B – A”) suggest a different lexical profile. Items such as *minister*, *neophytus*, *ecclesiae* highlight pastoral roles and the administration of the sacraments, while *iter*, *vicus* and *metropolis* point to repeated journeys and specific local settings. Verbs like *respondeo* and *redeo* underline the back-and-forth structure of the narrative, with letters, replies, and returns. Overall, Ibañez’s account is organised less around provinces and numbers, and more around ministers, villages, and concrete exchanges.

This small contrastive example already illustrates a form of scalable reading: by moving back and forth between basket-level statistics and close reading, *ELA* makes it possible to turn individual documents into entry points for broader, corpus-based analyses of missionary discourse. Much more could be done with larger *subcorpora* and more complex queries, but even this minimal experiment shows how corpus work can help to reshape the way these texts are read.

Alongside the graphical interface, *ELA* provides a public web API, documented through an online Postman collection. The API offers machine-readable access to core resources exposed on the platform. It was designed from the outset so that external projects (for instance within the *SERICA* framework) can query it directly, in order to make interoperability and data exchange as straightforward as possible. This makes it easier to integrate *ELA* into reproducible workflows and to connect it with other infrastructures.

5. Conclusions

Like most digital projects, *ELA* quickly disproved the illusion that a platform is finished once it goes online. The *SERICA* phase made visible what had already been implicit in *DAS-MeMo*: maintenance is not a purely technical afterthought but a continuation of the same intellectual work that produced the model in the first place. Each correction in the schema, each revision in the metadata, each adjustment in the interface reopens the questions that shaped the project from the start: what should be represented, and how can that representation remain valid over time? Maintenance has involved iterative redesigns, updates to documentation, and small but meaningful corrections.

This continuous work has also had a didactic dimension. *ELA* has been used in teaching contexts to introduce students to the idea that encoding and data structures are forms of interpretation.

At the same time, much remains to be done. The corpus needs to grow through new encodings; the NLP pipeline and the lemmatiser require further training and evaluation; and the multilingual nature of the material still calls for more systematic solutions. The life of the project, like that of any digital object, depends on its capacity to remain open: open to revision, to reuse, and, inevitably, to failings. A good model, like a good theory, is porous; its value lies not in its completeness but in its ability to generate new questions.

6. Acknowledgements

Eurasian Latin Archive is a project led by Francesco Stella at the Centre for Comparative Studies I Deug-Su of the University of Siena. Its initial development took place within the *DAS-MeMo* project (dasmemo.unisi.it), co-funded by Regione Toscana and carried out at the University of Siena (Departments of Philology and Literary Criticism, and of Information Engineering and Mathematics) in partnership with QuestIT and Pacini Editore; the core team included, among others, Leonardo Rigutini, Achille Globo, and Monica Bianchini. The author worked as post-doc fellow in close collaboration with QuestIT associate Nicola Giannelli (2018-2020), with contributions from the research fellows Irene Volpi (2019), Costanza Marzucchi (2020), and Silvia Bettella (2021). A second phase of implementation was realised in 2022 within the *SERICA* project (P.I. Chiara Ombretta Tommasi), with support from QuestIT (Nicola Giannelli, Alessandro Gronchi, Chiara Pugliese). Since 2024, the development of the platform has benefited from the work of Paola Mocella and Giulio Quaresima. Finally, the author wishes to thank Francesco Garosi, who co-developed the ELA-Tools suite.

References

- [1] Burns, Patrick J. 2019. “Building a Text Analysis Pipeline for Classical Languages”. In *Digital Classical Philology: Ancient Greek and Latin in the Digital Revolution*, edited by Monica Berti, 159-176. De Gruyter. <https://doi.org/10.1515/9783110599572-010>
- [2] Carbé, Emmanuela. 2020. “Eurasian Latin Archive”. *Itineraria* 19: 247-251.
- [3] Carbé, Emmanuela, and Nicola Giannelli. 2020. “ELA: fasi del progetto, bilanci e prospettive”. In *Atti del IX Convegno Annuale AIUCD. La svolta inevitabile: sfide e prospettive per l'Informatica Umanistica*, edited by Cristina Marras, Marco Passarotti, Greta Franzini, Eleonora Litta, 61-66. <https://doi.org/10.6092/unibo/amsacta/6316>
- [4] Carbé, Emmanuela, and Nicola Giannelli. 2021. “First steps in building the *Eurasian Latin Archive*”. *Qualitative and Quantitative Methods in Libraries* 9 (4): 285-300. <https://www.qqml-journal.net/index.php/qqml/article/view/573/603>
- [5] Carbé, Emmanuela, Andrea Balbo, Chiara O. Tommasi, Francesco Stella, e Mario G.C.A. Cimino. 2022. “Idee, persone, *realia*: un ambiente digitale per la Via della Seta». In *Culture digitali. Intersezioni: filosofia, arti, media. Proceedings AIUCD 2022*, edited by Fabio Ciraci, Giulia Miglietta, and Carola Gatto, 68-70. <https://doi.org/10.6092/unibo/amsacta/6848>
- [6] Ferrero, Michele. 2019. *Il primo Confucio latino: trascrizione, traduzione e commento di un manoscritto inedito di Michele Ruggieri SJ (1543-1607)*. LAS.
- [7] Fischer, Susanna. 2024. “Latin orientalism: Travel and pilgrimage literature”. In *Latin Literatures of Medieval and Early Modern Times in Europe and Beyond. A millennium heritage*, edited by Francesco Stella, co-edited by Lucia Doležalová and Danuta Shanzer, 296-307. John Benjamins. <https://doi.org/10.1075/chlel.34.18fis>
- [8] Golvers, Noël. 2024. “Central and East Asia”. In *Latin Literatures of Medieval and Early Modern Times in Europe and Beyond. A millennium heritage*, edited by Francesco Stella, co-edited by Lucia Doležalová and Danuta Shanzer, 308-323. John Benjamins. <https://doi.org/10.1075/chlel.34.19gol>
- [9] Mambrini Francesco, and Marco C. Passarotti. 2023. “The *LiLa* Lemma Bank: A Knowledge Base of Latin Canonical Forms”. *Journal of Open Humanities Data* 9 (28): 1-5. <https://doi.org/10.5334/johd.145>
- [10] Korkiakangas, Timo. 2020. “Theoretical and pragmatic considerations on the lemmatization of non-standard Early Medieval Latin charters”. *Current Approaches in Latin Lemmatization*, edited by Marco Passarotti. 58 (1): 67-94. <https://doi.org/10.4454/ssl.v58i1.277>
- [11] Stella, Francesco. 2017. “From Medieval Latin to Global Latin: The Eurasian Digital Archive”. In *Silk Roads Again II: Eurasian Studies in the Digital Age*, 245-269. Hanyang University, Asia-Pacific Research Center.
- [12] Stella, Francesco. 2020. “Global Latin”. *Itineraria* 19: 3-22.
- [13] Stella, Francesco. 2024. “A “postcolonial” approach to medieval Latin literature?”. In *Latin Literatures of Medieval and Early Modern Times in Europe and Beyond. A millennium*

heritage, edited by Francesco Stella, co-edited by Lucia Doležalová and Danuta Shanzer, 335-344. John Benjamins. <https://doi.org/10.1075/chlel.34.21ste>

- [14] Van den Wyngaert, Anastase. 1929. *Sinica Franciscana, I: Itinera et relationes fratrum Minorum saeculi XIII et XIV*. Quaracchi.
- [15] Van den Wyngaert, Anastase. 1936. *Sinica Franciscana, III: Relationes et epistolas fratrum minorum saeculi XVII*. Quaracchi.