

## Detecting Eastern and Western Names in the Latin Corpus of the *SERICA* Project – With Special Regard to the *Confucius Sinarum Philosophus* (1687) as a Case Study

Andrea Balbo, Elisa Della Calce

**Abstract:** This paper aims to describe the role of ICT within the *SERICA* (*Sino-European Religious Intersections in Central Asia. Interactive Texts and Intelligent Networks*) Project, especially by focusing on the corpus of Latin texts we are progressively building. Particular attention will be paid to the annotation of Named Entities (NEs) through *Recogito*<sup>1</sup> from a very peculiar 17<sup>th</sup> century Latin text, entitled *Confucius Sinarum Philosophus* (*CSP*) and edited by the Jesuits Prospero Intorcetta, Christian Herdrich, François de Rougemont, and Philippe Couplet in 1687. Despite including the translation of three Confucian texts (*Daxue*, *Zhongyong*, *Lunyu*), the *CSP* contains various references to Graeco-Hellenistic and Roman literature, and this comes as unsurprising since the Jesuit *Ratio Studiorum* (1599) was indebted to pagan classical authors. Yet the reception of ancient Latin literature can be further investigated by resorting to digital technologies. The annotation and the extraction of NEs allow in fact to take into account an extensive amount of data and to establish a first mapping concerning the impact of classical antiquity on the *CSP*, so as to detect which authors were mentioned more often and to reflect on their pattern of distribution within the work.

### 1. The *SERICA* Project<sup>2</sup>

The *SERICA* (*Sino-European Religious Intersections in Central Asia. Interactive Texts and Intelligent Networks*) Project has contributed to bringing to scholars' attention Eurasian studies or Silk Road(s) studies that are counted among the liveliest and most attractive subjects in global history. This specific research field has just started receiving in Italy the attention it deserves, although the wide-ranging perspective and the complexity involved in such a complicated network, which spans at least two millennia and more than ten thousand kilometres, have not always been taken into due account. These studies have often uncovered the role of travelling merchants and missionaries, who from Late Antiquity onwards started tracking the earliest cultural path which joined East Asia and Europe in a structural and permanent interchange, which reached its peak in the 16<sup>th</sup> and 18<sup>th</sup> centuries, with the production of a huge number of texts and the creation of multilingual libraries, especially in China. These works aimed at circulating to the Far East elements of classical European education, and at the same time communicated to Europe the first widespread information on geography, customs and cultural

---

1 <https://recogito.pelagios.org/> (last access 20.03.2026).

2 This paper has been adapted from a presentation given at the International Workshop *Nomina Omina. Detecting and Preserving Ancient Greek and Latin Proper Names in the Age of Artificial Intelligence* (Leipzig University, June 27–29 2024). Paragraphs 1–2 are by Andrea Balbo and paragraphs 3–4 are by Elisa Della Calce. Our thanks go in particular to Monica Berti, for organizing this Workshop and for her precious advice (especially on the 'semantic annotation'), to Philip Barras and Simone Mollea for their thorough reading and useful suggestions. It goes without saying that only the authors are responsible for any remaining mistakes.

bases of the civilizations of East Asia. At the Western end of the Silk Road(s), in Anatolia and in the Levant, the entangled interactions of the Western and Eastern Christianities and the Muslim world in the Middle Ages have also received sustained – though uneven – scholarly attention. New approaches are now underlining the dialogue and controversy within what can be regarded as a pervasive Mediterranean ‘culture of disputation’. A reassessment of these texts and other sources bearing witness to Christian-Muslim interactions (for example ethnographic works) appears badly needed, so as to shed light on their intended audience, wider circulation and later reception.

This multifaceted process of communication and cultural exchange has been the subject of many historical studies, always on the basis of a small selection of documents and almost never from the point of view of the linguistic intersection between the civilizations involved, nor from the perspective of Eastern contacts with Europe and the West. Our scope is precisely to bring together an interdisciplinary team of experts and at the same time to promote further national or international collaborations underlining the connections with other similar projects. In this regard, while from a strictly administrative standpoint the team is mainly composed by scholars belonging to the Universities of Pisa and Turin, a constellation of partner universities or institutions is also involved and cooperations are already established with Siena, Venice, Naples ‘L’Orientale’, Bologna-Ravenna, Seoul, and Beijing. The research team includes historians, philologists, and computer scientists.<sup>3</sup> While the role of philologists and specialists of the various Oriental languages is indispensable, as they guarantee a precise and rigorous interpretation and contextualization of all the scrutinized texts, the presence of information scientists and the creation of collaborative virtual environments is a further added value of this project, which promotes the convergence of technologies and applications between different domains. In this sense, the present project has brought together and unified research in the philological, linguistic and historical fields, which, despite having reached very high levels of refinement, still remain distinct and separate and too often do not create a dialogue between them.

By selecting some of the most significant case studies within a vast period (400 BC–1700 AD), we have been investigating how books and manuscripts reached Europe and were sometimes further disseminated through translations. Such text circulation attests not only to diplomatic-commercial relations, but also to an osmotic interaction between different agencies, faiths, social systems and cultural universes that moved along these routes. Considering “material things as entry points into history”,<sup>4</sup> the project’s digital platform and virtual mapping of the historical-geographical paths either mentioned in these books or actually covered by their circulation, seek to highlight, in its broadest sense, the uninterrupted contact between Asia and Europe. Choosing a long historical period and a wide geographical area permits the adoption of a trans-disciplinary agenda, which promotes a rethinking of traditional methodologies and has the aim of achieving a non-Eurocentric reading of the past: both these aspects allow us to reflect on and answer crucial questions for contemporary society.

Consequently, the novelty in our approach can be summarized as a twofold one: it presents and makes available new documents, often unpublished or scarcely known, for most of them have never been translated into Italian or any modern language. Secondly, this has been made possible also thanks to the pivotal role played by the ICT, which permits the evaluation and interpretation of the data by combining them into a united, transversal and dynamic corpus, capable of exploiting the potential of information technologies. New diachronic perspectives of comparative studies have thus been opened; in particular, the strict cooperation with information scientists permits the development of a digital library and a new corpus of texts and documents from medieval and early modern times concerning the Middle and the Far East. By proposing a fresh new approach to this kind of literature, we wish to place the study of East-West trajectory into the wider scientific discussion about Eurasian cultural exchanges, in particular considering how Eastern sources dealt with and were received in the West. Fur-

---

3 See <https://serica.unipi.it/il-team> (last access 29.08.2025).

4 Ulrich et al. (2015), XI.

thermore, it is difficult to find platforms that combine inter-linguistic and philological studies with historical reconstructions and a clear visual rendering.

One of the objectives of the *SERICA* Project is the development of a digital library on the website,<sup>5</sup> which aims to include texts regarding the Central Asian routes between China and Europe. Many of these texts were written in Latin by Jesuit missionaries who travelled to China between the 16<sup>th</sup> and the 18<sup>th</sup> centuries and drew up several works to spread in Europe Chinese history and culture, along with Confucian thought.

In this contribution, I intend to describe the digital side of the project, especially by focusing on the results achieved and its future perspectives (§2), while Elisa Della Calce aims to show a semantic annotation experiment (§3–§4) – which was started during an Erasmus Training stay at Leipzig University in 2024 (22.04–03.05.24) – by using *Recogito* on the *Confucius Sinarum Philosophus* (henceforth *CSP*).

The *CSP* is one of the crucial and longest texts that we planned to include in *SERICA*'s database: it was printed in Paris in 1687 and edited by the Jesuits Prospero Intorcetta, Christian Herdtrich, François de Rougemont, and Philippe Couplet, with the aim of spreading Confucianism in Europe. The Jesuits resorted to Latin as a medium to make Confucianism more understandable to Western cultivated readers. The *CSP* contains a prefatory epistle, dedicated to Louis XIV; an introduction (*Proëmiialis Declaratio*); a life of Confucius; the Latin translation of three Confucian Books, namely *Daxue* (The Great Learning), *Zhongyong* (The Doctrine of the Mean), *Lunyu* (The Analects), and some chronological tables related to Chinese monarchy.<sup>6</sup>

## 2. The Role of ICT within the *SERICA* Project

Thanks to collaboration with private companies, which strengthens the links between the university and the local area, once fully realized, the website will feature a highly usable and effective user interface, and will be structured according to different phases of design prototyping (from Wireframing to Graphic Design), including the definition of general categories of users (Proto-Person); the construction of Use cases and Users stories; the description of navigation paths (User journeys); and the analysis of the effectiveness of the proposal through interviews and user testing. Once completed, it will feature a digital library equipped with innovative semantic web and artificial intelligence tools and a diachronic array of interactive maps that make use of geolocalisation tools. Data have been collected into an interactive database, with visual mapping of the texts and objects documenting mutual exchanges, mainly those from East to West. The final result will present a navigable interactive and 'talking' diachronic map of the various routes, where images and historical details will contribute to reconstruct a context as precise as possible of the knowledge that a traveller might have had.

This digital library is being realized using suites that include functions such as uploading texts and metadata, semantic enrichment, and data management. Such a knowledge base will represent a concrete model (and the first one in Italy) to display interactive data of historical relationship between different Eastern countries and Europe (for a similar project see the CHCD at Boston University). In addition, the book and other textual materials have become an integral part of the platform. Particular attention is being paid to the processes of linguistic analysis (through Natural Language Processing techniques), so as to normalize the texts and extract the greatest possible amount of structured information to be used as input for the advanced queries and analyses of artificial intelligence. A context of particular interest in the research area of the project is multilingualism, i.e. having texts written in different

5 <https://serica.unipi.it/> (last access 29.08.2025).

6 The text of *CSP* is currently available in a PDF format on <https://archive.org/> (last access 29.08.2025). Text quotations from this work have been adapted to modern typographic criteria (e.g. & → *et*; <j> → <i>).

languages – among which ancient Greek, Latin, Chinese, Arabic – which must be normalized and adapted to a common knowledge model in order to be able to create machine learning models with the greatest possible quantity of homogeneous data so as to guarantee greater accuracy. Another important aspect is the temporal diversity of the origin of the sources, which means that NLP analysis might be particularly difficult, since normally the available tools only consider the current versions of a language and not its older or classical versions. In particular, its hybrid nature (digital library-interactive map-database) makes it a pilot model in the Italian research panorama of Asian language studies, which can be connected to an ever-growing network of information, both national and international, cooperating with projects currently already active or in progress. As already suggested, the creation of an information framework structured through semantic and thematic paths and by means of a virtual guide based on web ontologies and deep learning is central to the success of the project.

The final goal is to create a virtual place that can be used not only by experts for academic purposes, but also by a wider audience for educational or divulgation purposes. Although the project is carried out in a scientific manner by experts from different academic sectors, one can also envisage the potential didactic impact, involving many students in the study and the use of these digital tools. Interdisciplinary teaching also increases student learning and encourages an intertwined approach at school and at the academy. In particular, we are proud to mention the fact that the project permitted a temporary enrolling of at least twenty young researchers, some of whom still undergraduates, or at the early stage of their professional academic career.

Uploading documents on *SERICA*'s website enables users to carry out targeted lexical searches (see fig. 1) by accessing texts that previously were digitized and spread only in a non-searchable format. This process involves several steps. To begin with, the document, depending on how it looks or has been digitized, is transcribed or processed by an OCR software. Afterwards, the document is subjected to linguistic-typographical correction and normalization in order to remove any mistakes. At this stage, as far as Latin texts are concerned, the abbreviations are replaced by their corresponding words (e.g. & > *et*), the grapheme <j> is transcribed as <i>, and all accents on vowels are removed (e.g. *Proëmialis Declaratio*, p. IX: *eò etiam magis, quòd placuit > eo etiam magis, quod placuit*) so as to facilitate an exact lexical search on the database. Once this step is completed, the text is transferred into XML format, tagged according to the TEI standard, and uploaded on the site.

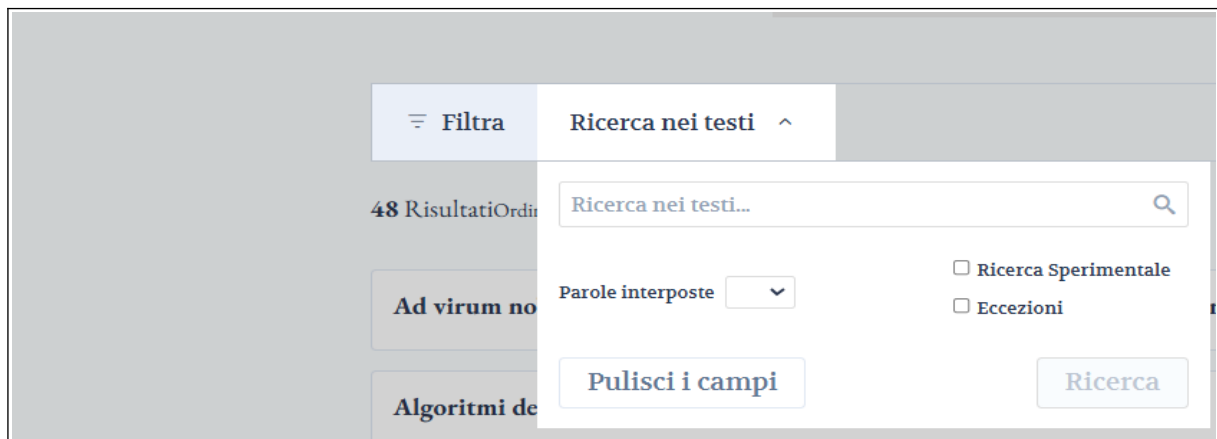


Fig. 1: The search mask of *SERICA*'s website.

### 3. CSP and Annotation through *Recogito*

This section focuses on the relationship between my literary interests in the *CSP* and the semantic Annotation process by means of *Recogito*, “an open-source, free and online semantic annotation tool developed by the Pelagios Network”.<sup>7</sup>

My starting point was to question what impact this annotation and extraction process may have on my research goals, which particularly concern the influence of Latin authors in the *CSP*. Despite being composed in 17<sup>th</sup> century, this work includes several echoes of and quotations from pagan Latin literature, especially from Cicero. The Jesuit education was based on the *Ratio Studiorum* (1599), which put the emphasis on readings suitable to Catholic morality in the Counter-Reformation era. In this regard, a substantial number of Graeco-Roman classics was recommended to be read and learned, among which Cicero was a model for language and style.<sup>8</sup>

In light of this, the annotation of Named Entities (NEs)<sup>9</sup> or, more precisely, of proper personal names (henceforth PPNs) enabled me to consider a larger amount of data and create a mapping of the influence of Western ancient classics in the *CSP*, in order to investigate *which* authors and characters were most quoted from pagan Classical Literature. It follows that I privileged the extraction and analysis of PPNs over those of place names, yet limiting myself to some general considerations on applying *Recogito*'s gazetteers in a text as complex as the *CSP*. Since the annotation has not yet been concluded, the results I intend to share are only related to the first sections of the *CSP* (the *EP* = *Epistula praefatoria* and the *PD* = *Proëmialis Declaratio*).<sup>10</sup>

First of all, since automatic annotation via Herodotus Latin NER in *Recogito* proved unsuccessful, I proceeded with a manual annotation of the Latin text,<sup>11</sup> also in the belief that, due to the particular status of the *CSP*'s structure and content, a ‘close reading approach’ can contribute to a first text analysis.<sup>12</sup>

---

7 Del Rio Riande / Vitale (2020), 2. In addition to this, cf. also Simon et al. (2017) and Gregory (2021).

8 Cf. e.g. on this Balbo (2020), Balbo (2022), 112–117, Della Calce/Mollea (2023), 46–48, and Della Calce (2024), 279–282 (with bibliography).

9 For a general overview of the linguistic annotation, cf. at least Montemagni (2023), 162–176, Berti (2019), who in particular dealt with the annotation of ancient Greek, and Berti (2023), 316–318. Cf. also Yousef et al. (2023), who instead focused on “a processing pipeline to transfer NE annotations from a text in modern languages to parallel texts in classical or low-resourced languages” (175). On the annotation as one of the Pelagios Network’s “core Activities”, cf. <https://pelagios.org/activities/annotation/> (last access 01.07.2025).

10 This paper refers to my use of *Recogito* mostly during the period of my training stay in Leipzig and of the *Nomina Omina* Workshop. The data and the number of annotations extracted, however, have been revised and updated in view of these proceedings and analyzed only with respect to the PPNs.

11 As far as Latin and NER are concerned, cf. at least e.g. Beersmans et al. (2023), 1–2, also with reference to Ehrmann et al. (2021) as well as to earlier bibliography, such as Erdmann et al. (2016) and (2019): “for modern high-resource languages, generic NER off-the-shelf solutions, focusing mainly on identifying locations, organizations and people, can produce highly accurate annotations. For historical languages, even prolific ones like Latin, the task remains a challenge, in part due to a lack of annotated corpora and tools (Ehrmann et al., 2021)”. For a short overview on NER algorithms provided by *Recogito*, cf. del Rio Riande/Vitale (2020), 5 and Berti (2019), 1, n. 3: “*Recogito* (<https://recogito.pelagios.org/> [last access 20.03.2026]) provides automatic NER tagging for historical data using Stanford CoreNLP (English, French, German and Spanish), experimental Latin NER with the Herodotus Latin NER plugin [...] and experimental Hebrew NER with the Kima NER plugin (<https://geo-kima.org> [last access 01.07.2025])”.

12 Cf. on this Montemagni (2023), 176–177, on the wake of Moretti (2005)'s dichotomy “Close Reading – Distant Reading”: “l’annotazione linguistica può potenziare le funzionalità di ricerca di tipo Close Reading permettendo le ricerche per lemma, invece che per forma, oppure ricerche mediante schemi che combinano variamente informazione (morfo)sintattica e lessicale. Gli stessi testi diventano esplorabili mediante tecniche di Distant Reading, che permettono l’estrazione di generalizzazioni a partire dai testi”.

The *CSP*, in fact, is a very articulated work which was exclusively composed in Latin. The Jesuits transliterated all Chinese names into Latin characters, but their initial plan was to juxtapose the Chinese and the Latin. Yet the Chinese text was in the end excluded, as N. Dew (2009) and others have pointed out:

“Originally, the plan had been to include the original Chinese texts with the translations, but this aim had to be abandoned because of the practical difficulties involved in printing the Chinese – although in some passages a remnant of this intention survived, in the form of superscript numerals that would have guided the reader from each Latin word to the corresponding Chinese character”.<sup>13</sup>

What is more, the Jesuits’ transliteration of Chinese characters normally diverges from modern transcription: for instance, the grapheme <ç> does not always appear transliterated in the same way, as is the case with *Heu çie* > Houji<sup>14</sup> and *ço xi* > Zuoshi.<sup>15</sup> Therefore, since I have no specific sinological background, I have classified the extracted data so as to juxtapose the Jesuit transcription of Chinese names and that which is transliterated into modern Chinese, according to T. Meynard’s edition of the *CSP*, which includes the English translation of the text from the *EP* to the first Confucian Book (*The Great Learning*).

For all names extracted I have also provided the link to repositories such as *Wikidata*, in order to give a unique identifier for each name and then distinguish it from other similar names that refer to different entities.<sup>16</sup> As far as the Western names are concerned, I preferred reporting both the original transcription and the corresponding lemma (e.g. *Ciceronem* → *Cicero*).

### 3.1. What Criteria for the Annotation?

Since the *CSP* is a very peculiar text at a thematic and structural level, it is extremely important to establish some strict criteria for this annotation. To begin with, I annotated only the PPNs which were explicitly cited. I therefore took into account historical or legendary names, while I did not consider names related to divinities, honorific attributes or titles, dynasties, religious or philosophical congregations and ethnonyms. In line with this, I excluded antonomasias, such as *Philosophus* to indicate *Confucius*, and personifications.

Despite focusing on the PPNs and, particularly, on the Graeco-Hellenistic and Roman ones, I limit myself, in passing, to some remarks on the annotation of place names through *Recogito* and its available gazetteers, which I quote below:

- **HistoGIS** – A GIS repository for historical temporalized spatial data by the Austrian Centre for Digital Humanities
- **Pleiades** – *Pleiades Gazetteer of the Ancient World*
- **CHGIS** – *China Historical GIS*
- **DPP Places** – Places from the Digitizing Patterns of Power project

---

13 Dew (2009), 210. Cf. also Della Calce (2024), 287–288.

14 <https://www.wikidata.org/wiki/Q1207613> (last access 01.07.2025). According to transcription by Meynard (2011), 189.

15 <https://www.wikidata.org/wiki/Q230192> (last access 01.07.2025). According to transcription by Meynard (2011), 102.

16 This is all the more important the more the name Thomas, who corresponds to Han Lin (*PD*, p. CIX) according to Meynard (2011), 227, can be confused with the name of the saint and Doctor of the Church, Thomas Aquinas (*PD*, p. LXIX and XCIII).

- **DARE** – *Digital Atlas of the Roman Empire*
- **MoEML** – *Map of Early Modern London*
- **HGIS de las Indias** – *Historical-Geographic Information System for Spanish America (1701-1808)*
- **GeoNames** – A subset of *GeoNames* populated places, countries and first-level administrative divisions
- **Kima** – *Kima Historical Gazetteer – place names in the Hebrew script.*

Regardless of the fact that I did not select the gazetteers which were unsuitable to the context of composition of the *CSP* (e.g. *MoEML* and *HGIS de las Indias*), they do not always convey the exact geographical information which the *CSP* provides at historical and geo-political levels. For example, in the *EP* there are 11 occurrences of place names, that is, *Gallia/Galliae* (6 occurrences), *Sina* (2 occurrences), *Africa*, *Asia*, *Europa* (1 occurrence respectively). However, would it be preferable to resort to the gazetteer *Pleiades*, thereby preserving the Latin uses and tradition, or to other gazetteers such as *Geonames*, which enable us to select only the modern ‘France’? By the same token, the gazetteer *Pleiades* does not annotate place names as *Asia* in the *EP*, as it designates only a geographical limited portion of this continent in the ancient world. It is then preferable to use other gazetteers, such as *Geonames*, for place names relating to modern regions. With regard to China, progress has been achieved in representing the printed Western Maps of China to 1735, as is the case with the *Regnum Chinae* edited in 2022 by Marco Caboara. And this could be used as an important resource to refine any geolocalization activity, by integrating, for instance, further notes and references which could be useful in describing how better to interpret place names in their historical context.

Also regarding the analysis of the PPNs, other interesting incongruities emerge:

- the name attributed to the Buddha is variously expressed (*Xe*, *Xe Kia*, *Xaca*, *Foe*, according to Jesuit transcription),<sup>17</sup>
- when Jesuits allude to *Doctores Sinenses* or those who have a double name, according to Western and Eastern traditions (especially Jesuit missionaries, *Li Mateo* corresponds to *Matthaeus Riccius*), it is better to explicit both names to disambiguate (e.g. Xu Guangqi = Paul Siu),<sup>18</sup>
- some mythical characters are inconsistently transliterated, as is the case with *Aedipus* (*PD*, p. LIX) and *Oedipus* (*PD*, p. XVIII),
- *Recogito* does not allow any continuous annotation when the original phrasing seems to be interrupted (*PD*, p. CIX: *Matthaei scilicet Riccii*; *PD*, p. CX, *Iulius vero Aleni*).

Nonetheless, *Recogito* enabled me to download the annotations as spreadsheet data in CSV format. This means that I could further analyze the data tables, especially those relating to Graeco-Hellenistic and Roman names, in order to provide a more precise overview of their use and frequency.

---

17 *PD*, pp. XXVII–XXVIII: *ex hac [scil. Mo-ye] natus est illi filius, Xe primum, sive Xe Kia dictus (quo etiam nomine tota Bonziorum colluvies ac superstitione significatur; Japonii tamen corrupto vocabulo Sinico Xaca fecere) deinde, cum trigessimum attigit aetatis annum, Foe nominatus.* Furthermore, the reference to the so-called brothers Cheng (*PD*, pp. XXXIV–XXXVI, along with Meynard 2011, 127, n. 58 and 130, n. 9) could not be separated so as to distinguish two different figures and *Wikidata* resource was adapted accordingly (s.v. Cheng brothers, <https://www.wikidata.org/wiki/Q48880876> [last access 01.07.2025]).

18 <https://www.wikidata.org/wiki/Q420427> (last access 01.07.2025). Cf. *infra* n. 20.

5 PPNs have been extracted from the *EP*, namely 3 occurrences of Confucius and 1 of Louis XIV and Philippe Couplet respectively. This implies that Couplet was interested in highlighting only the names that were crucial both to emphasize the main character of the work (*Confucius*) and to accomplish the rhetorical aims related to his dedication by mentioning himself as editor and Louis XIV as recipient.

By contrast, the data extracted from the *PD* are definitely more, amounting to 516 annotations of PPNs. 76 occurrences are relating to *Confucius*<sup>19</sup> and 326 occurrences to Eastern/Middle Eastern names, not only transliterated Chinese names, but also names from Buddhist and Hindu as well as Assyrian and Jewish tradition.<sup>20</sup> As I said before, I am not going to dwell on this latter category, given my different research purposes. In this sense, Graeco-Hellenistic and Roman PPNs amount to 33 occurrences, as is evident from the following table (Tab. 1), which summarizes the remaining data (114 occurrences).

Category	PPNs (Occurrences)	PPNs (Characters)
1. Pagan Graeco-Hellenistic and Roman culture	33	16
2. Christian culture and tradition from Antiquity to Middle Ages	25	9
3. Western names related to Modern Age (from the Second Half of 15 <sup>th</sup> century to 17 <sup>th</sup> century approximately) <sup>21</sup>	56	15

**Tab. 1: Categories of Western PPNs.**

The first category reflects the highest number in terms of referents, thus contributing further to confirming the Jesuits' knowledge of several pagan characters and their interest in Western ancient literature. While showing the highest number of occurrences, the third category has fewer referents than the first, since Matteo Ricci is the most mentioned character (37 occurrences), which is unsurprising when one considers Ricci's crucial role in applying the Jesuit method of 'accommodation'.<sup>22</sup> Accordingly, the proportion between Jesuit and non-Jesuit characters shows the prevalence of the former over the latter in terms of occurrences (50 vs. 6) and referents (10 vs. 5).

---

19 These occurrences consist of the Latin name *Confucius*, the name *Cum-çu* (corresponding to Kongzi: cf. Meynard [2011], 175) and the reference to the Jesuit missionary Giulio Aleni (1582–1649, <https://www.wikidata.org/wiki/Q2707504> [last access 01.07.2025]) as *Occidentis advena Confucius (Si lai cum çu)*. Also the occurrence *Confucius* in the capture at *PD*, p. XLII has been included.

20 The name *Confucius* (see *supra*, n. 19) as well as Matteo Ricci's Chinese name (*Li Mateo*) have been excluded. Those who are referred to *Doctores Sinenses* or Chinese people converted to Christianity (Ignatius, Leo, Luke, Mathias, Michael, Paul, Peter, Philip, Thomas, cf. *PD*, p. CV and CIX, along with Meynard [2011], 222–223 and 227 for the corresponding Chinese names) have been included. By the same token, the name *Cham Colaüs*, in the capture at *PD*, p. XLII, has been included. He corresponds to Colaüs Zhang, *scil.* Zhang Juzheng (Meynard [2011], 141), who lived between 1525 and 1582, <https://www.wikidata.org/wiki/Q197234> (last access 01.07.2025).

21 In the case of discontinuous syntagms (*Matthaei scilicet Riccii; Iulius vero Aleni*) two occurrences have been included in the total amount.

22 Cf. Mungello (1985), 44–73 and Catto (2014).

In the second category, pride of place goes to Lactantius (7 occurrences),<sup>23</sup> followed by Jerome (6 occurrences) and Augustine of Hippo (4 occurrences),<sup>24</sup> whereas Cicero (5 occurrences) is the most mentioned author in the category of Graeco-Hellenistic and Roman names. Then, in order of importance, there are 3 occurrences of Oedipus, Orpheus, Trismegistus, Plato respectively, 2 occurrences of Aristotle, Epictetus, Pythagoras, Seneca, Socrates respectively, and 1 occurrence of Aeneas, Hesiod, Homer, Plutarch, Solon, Varro respectively. This statistical survey allows us to draw some further remarks, starting from the central role played by Cicero. He is, in fact, the only figure of pagan antiquity who is called by different names (2 occurrences of *Cicero* and 3 occurrences of *Tullius*), thereby showing how the Jesuit authors were familiar with him. Not unsurprisingly the *Ratio Studiorum* (1599) attributed a crucial role to Ciceronian works as far as rhetoric, style and philosophy are concerned.

Furthermore, it is also worth considering that *Oedipus*, *Orpheus* and *Trismegistus* recur more than others. On the one hand, *Orpheus* and *Trismegistus* are commonly connected with the notion of *prisca theologia*. In this regard, D. E. Mungello, referring to *China illustrata* (1667), composed by another Jesuit, Athanasius Kircher, has pointed out:

“Kircher saw certain vestiges of Christianity in China. [...] For Kircher, the strongest religious influences on China were the Egyptian and Greek pagan religions. Kircher’s position is understandable only in terms of his groundings in the Hermetic tradition of Christian apologetics which argued that certain pagan texts contained vestiges of Christianity. The primary texts of this tradition were ascribed to Hermes Trismegistus, Orpheus and Pythagoras and the tradition is referred to variously as Hermetism or *prisca theologia* (Ancient Theology)”.<sup>25</sup>

On the other hand, *Oedipus*, who solved the famous Sphinx’s riddle, is associated with two Chinese rulers at *PD*, p. XVIII and LIX, who are therefore regarded as *Aedipi*.

*Lineolis et quidem paucis tota res constat: Nos eas proxime hic depingemus, unaque declarabimus, quoties et quomodo variatae figuras novas, et quasi nova rerum significata conficiant. Annis mille et octingentis Monarchia steterat, cum tandem Oedipus apparuit, Regulus, inquam, Ven vum: hic lineolis octo octies inter se mutuo commutatis conatus est, octo rerum principium mutuas vicissitudines exponere.*

“Everything in the world is classified according to a set of drawings of short lines. We shall draw these lines below and also demonstrate how they form every day new figures with almost new significations. The monarchy had stood for one thousand eight hundred years when at last a kind of Oedipus appeared: the minor king Wen Wang came to explain the permutations of the eight principles by combining and recombining the eight trigrams (groups of three lines) eight times.”

*PD*, p. XVIII; transl. Meynard (2011), 101.

---

23 Cf. von Collani (1990), 41–42: “Ph. Couplet cites several times Aurelius Augustine (354–430), Jerome (347–420), Thomas Aquinas (1225–1274) and finally relatively often the early Latin Father Lactantius (ca. 250/260–after 317). Although not precisely stated, it is easily seen that Ph. Couplet takes the quotations from the *opus maximum* by Lactantius *De divinis institutionibus libri VII*. Lactantius regards the true religion being supported by the revelations of paganism, for instance by the prophecies of the sibyls and Hermes Trismegistos, who predicted the future true Redeemer”.

24 Also St. Paul and Thomas Aquinas (2 occurrences respectively) as well as Basil of Caesarea, John the Evangelist, John Chrysostom (1 occurrence respectively) are included in this category.

25 Mungello (1985), 136–137.

Ye kim, sive eum qui de mutationibus inscribitur; quippe cuius author est idem qui gentis Sini-  
cae Fundator Fo hi. Recte tu quidem: At, amabo te, quid tandem libri fuit, cuius Authorem Fo hi  
praedicat? Figurae aenigmatae quatuor et sexaginta, sive lineolae 384 partim continuas par-  
tim interruptas et praeterea nihil. Bene habet. At si aenigmatae, ergo perobscurae; si tam ob-  
scurae, ergo aedipo fuit opus qui lucem afferret. [...] Aedipi fuerunt magnus ille Princeps et  
quasi conditor Cheu Familiae tertiae Imperialis Ven vam dictus, nec non eiusdem filius Cheu  
cum. Hi solverunt aenigma, et figuras interpretati sunt.

“It is entitled the *Yijing*, or the *Book of Changes*, written by Fuxi, the Founder of the Chinese  
race. You may well ask: ‘But, do tell, what is in this book whose author you claim to be Fuxi? It  
is nothing more than sixty-four mysterious figures, or 384 lines, some being continuous, others  
broken’. That is all very well. But these figures are mysterious to the point of being utterly ob-  
scure, so obscure that it would take another Oedipus to elucidate them. [...] Wen Wang, this  
great Prince who can almost be considered the founder of the Third Dynasty Zhou, and his son  
Zhou Gong, were the true Oedipuses, in that they solved the riddle and interpreted the figures.”  
*PD*, p. LIX; transl. Meynard (2011), 163–164.

It is also interesting to observe that *Sinicus noster Epictetus* corresponds to *Confucius* at *PD*, pp. XIII–  
XIV.<sup>26</sup> In this case, a Greek name conveys a Chinese one, so as to emphasize analogies between two  
philosophical traditions which actually diverged from a chronological and geographical viewpoint.  
Furthermore, *Confucius* as *Sinicus Epictetus* is compared to Socrates, Plato, Seneca, and Plutarch, who  
are cited in the plural form (*et vero in Europa illa, ubi iam Socrates, et Platones, ubi Senecae,  
Plutarchi prope viluerunt, an speremus fieri posse ut plausum referat Sinicus noster Epictetus?*). This  
can be regarded not only as a simple literary embellishment, but also as a rhetorical strategy to empha-  
size the relevance of their spread across Europe.<sup>27</sup> By the same token, the famous Athenian legislator  
Solon is compared to the ancient legendary Chinese rulers Yao and Shun. For this reason he is men-  
tioned in the plural form: *illi [...] veri Solones gentis Sinicae* (*PD*, p. XV).

#### 4. Annotating Personal Names in the *EP* and *PD*. Some Final Re- marks

Although only a part of the *CSP* has been analyzed, the annotation and extraction of PPNs may be a  
useful instrument for more detailed analyses in Jesuit Latin texts. As for Graeco-Hellenistic and Ro-  
man personal names, to which particular attention has been paid, it is true that they are a minority  
compared to Eastern names. However, it is equally true that, by combining digital tools with tradi-  
tional methods of analysis, it was possible to reflect on these names in a de-contextualized form and,  
at the same time, to analyze them further in specific passages, by creating interesting intersections of  
linguistic, literary, and thematic viewpoints. In other words, close and distant reading appeared com-  
plementary both to carry out a manual annotation of names and to analyze an amount of data from a  
comparative perspective in terms of referents and occurrences.<sup>28</sup>

26 On *Sinicus Epictetus* cf. Tommasi (2020), 80.

27 Consequently, for consistency with this passage, I also included the occurrence *Platones* (*PD*, p. LVIII: *Platones subinde aliquos audire te credas, aliosve Philosophos haudquaquam male sentientes de Deo*).

28 Cf. *supra*, §3.

In this sense, two main patterns of citation of these Graeco-Hellenistic and Roman personal names emerged as basically functional to:

1. compare Western and Eastern intellectual figures (Solon, Orpheus, Oedipus), at times by resorting to a ‘catalogue’, as is the case with *PD*, p. LXXVII, which relates to “the early knowledge and worship of God”<sup>29</sup>: *apud Graecos Socrates, Pythagoras, Plato, Epictetus; et apud Latinos, Varro, Tullius, Seneca aliique Philosophi de Deo multa recte senserunt, atque scripserunt.*
2. introduce a source, especially by reporting quotations or other authors’ thought, as is the case with Cicero at *PD*, p. LXXXIX, where the phrase *quot enim, teste Tullio, hominum linguae, tot nomina Deorum* derived from Cicero’s *De natura deorum* (1,84)<sup>30</sup> or, to give another example, with Lactantius.<sup>31</sup>

To conclude, among some future prospects, we would especially like to focus on collecting more data, by extending the analysis to the entire *CSP*, in order to understand fully how these digital techniques can contribute to building a bridge between Eastern and Western cultures in the same way as Latin used to be considered a vehicular language by Jesuit Fathers.

---

29 Meynard (2011), 183.

30 Cf. Della Calce (2024), 290–292.

31 Cf. e.g. *PD*, p. LVIII: *et quemadmodum Lactantius dicebat a nullo Ciceronem, quam ab ipso Cicerone vehementius posse refutari, ita nec hos novatores a nullis certius, quam a seipsis refutari posse; PD*, p. LXXXIII: *et de illis quidem aliarum gentium superstitionibus, deque ipsarum prisca religione constat ex primis litterarum cuiusque gentis monumentis ex Orpheo, inquam, et primis Poëtarum Hesiodo, Homero, et c. qui et ipsi (uti idem Lactantius ait) multo ante natum Philosophiae nomen fuerunt, et habiti sunt sapientes, et tamen tam inepta de Deo Deorumque generationibus figmenta et fabulas protulerunt.*

## References

- Balbo (2020): A. Balbo, *Classics, Latin and Greek Authors in the Proemialis Declaratio of Confucius Sinarum Philosophus (1687)*, *Itineraria* 19 (2020), 153–172.
- Balbo (2022): A. Balbo, *The Epistula praefatoria of the Confucius Sinarum Philosophus: A Rhetorical Analysis in Search of Cicero and Seneca*, in: A. Balbo / J. Ahn / K. Kim (eds.), *Eastern and Western Civilizations. Searching for a Respublica Romanosinica*, Berlin / Boston 2022, 111–130.
- Beersmans et al. (2023): M. Beersmans / E. de Graaf / T. Van de Cruys / M. Fantoli, *Training and Evaluation of Named Entity Recognition Models for Classical Latin*, in: A. Anderson / S. Gordin / B. Li / Y. Liu / M. C. Passarotti (eds.), *Proceedings of the Ancient Language Processing Workshop, Shoumen 2023*, 1–12.
- Berti (2019): M. Berti, *Named Entity Annotation for Ancient Greek with INCEpTION*, in: K. Simov / M. Eskevich (eds.), *CLARIN Annual Conference Proceedings, Leipzig 2019*, 1–4.
- Berti (2023): M. Berti, *L'antichità greco-romana e le tecnologie digitali*, in: F. Ciotti (ed.), *Digital Humanities. Metodi, strumenti, saperi*, Roma 2023, 312–324.
- Caboara (2022): M. Caboara, *Regnum Chinae: The Printed Western Maps of China to 1735*, Leiden, 2022.
- Catto (2014): M. Catto, *Il mito gesuitico della Cina*, in: F. D'Arelli, Matteo Ricci. *L'altro e diverso mondo della Cina*, Milano 2014, 9–27.
- Della Calce (2024): E. Della Calce, *Cicerone nel Confucius Sinarum Philosophus (1687). Discussione di alcuni casi emblematici tra filosofia, religione e morale*, *Ciceroniana online* 8/1 (2024), 277–308.
- Della Calce / Mollea (2023): E. Della Calce / S. Mollea, *Seneca in Jesuit Thought: the Case of Antonio Possevino and His Secondary Sources (Muret, Ortino, Perera, Erasmus, and Others)*, *Latinitas Series Nova* 11 (2023), 45–69.
- Del Rio Riande / Vitale (2020): G. del Rio Riande / V. Vitale, *Recogito-in-a-Box: From Annotation to Digital Edition*, *Modern Languages Open* (2020), 1–13.
- Dew (2009): N. Dew, *Orientalism in Louis XIV's France*, Oxford / New York 2009.
- Ehrmann et al. (2021): M. Ehrmann / A. Hamdi / E. Linhares Pontes / M. Romanello / A. Doucet, *Named Entity Recognition and Classification on Historical Documents: A Survey*, arXiv:2109.11406 [cs.CL] – *ACM Computing Surveys* 56/2 (2023), 1–47.
- Erdmann et al. (2016): A. Erdmann / C. Brown / B. Joseph / M. Janse / P. Ajaka / M. Elsner / M.-C. de Marneffe, *Challenges and Solutions for Latin Named Entity Recognition*, in: E. Hinrichs / M. Hinrichs / T. Trippel (eds.), *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*, Osaka 2016, 85–93.
- Erdmann et al. (2019): A. Erdmann / D. J. Wrisley / B. Allen / C. Brown / S. Cohen-Bodénès / M. Elsner / Y. Feng / B. Joseph / B. Joyeux-Prunel / M.-C. de Marneffe, *Practical, Efficient, and Customizable Active Learning for Named Entity Recognition in the Digital Humanities*, in: J. Burstein / C. Doran / T. Solorio (eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis (MN.) 2019, 2223–2234.
- Gregory (2021): I. Gregory, Barker, Elton, Leif Isaksen, Rebecca Kahn, Rainer Simon, Valeria Vitale, and the Pelagios Network, Project Creators. *Recogito: Semantic Annotation without the Pointy Brackets. Other*, *Renaissance and Reformation* 44/3 (2021), 243–249.

- Meynard (2011): T. Meynard S. J. (ed.), *Confucius Sinarum Philosophus* (1687). The First Translation of the Confucian Classics, Roma 2011.
- Montemagni (2023): S. Montemagni, *Trattamento automatico del linguaggio e Digital Humanities: metodi e strumenti, sfide*, in: F. Ciotti (ed.), *Digital Humanities. Metodi, strumenti, saperi*, Roma 2023, 160–177.
- Moretti (2005): F. Moretti, *La letteratura vista da lontano*, Torino 2005.
- Mungello (1985): D. Mungello, *Curious Land: Jesuit Accommodation and the Origins of Sinology*, Stuttgart 1985.
- Simon et al. (2017): R. Simon / E. Barker / L. Isaksen / P. de Soto Cañamares, *Linked Data Annotation Without the Pointy Brackets: Introducing Recogito 2*, *Journal of Map & Geography Libraries* 13/1 (2017), 111–132.
- Tommasi (2020): C. O. Tommasi, *Epictetus in the Forbidden City. Accommodation and Resilience in Matteo Ricci’s “Twenty-five Paragraphs”*, *Itineraria* 19 (2020), 73–104.
- Ulrich et al. (2015): L. T. Ulrich / I. Gaskell / S. J. Schechner / S. A. Carter, with Photographs by Samantha S. B. van Gerbig, *Tangible Things: Making History through Objects*, New York 2015.
- von Collani (1990): C. von Collani, *Philippe Couplet’s Missionary Attitude Towards the Chinese in Confucius Sinarum Philosophus*, in: J. Heyndrickx, (ed.) *Philippe Couplet, S. J. (1623–1693). The Man Who Brought China to Europe*, Nettetal 1990, 37–54.
- Yousef et al. (2023): T. Yousef / C. Palladino / G. Heyer / S. Jänicke, *Named Entity Annotation Projection Applied to Classical Languages*, in: S. Degaetano-Ortlieb / A. Kazantseva / N. Reiter / S. Szpakowicz (eds.), *Proceedings of the 7<sup>th</sup> Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, Dubrovnik 2023, 175–182.

### Figure and Table References

Fig. 1: The search mask of *SERICA*’s website.

Tab. 1: Categories of Western PPNs.

### Author Contact Information<sup>32</sup>

Prof. Andrea Balbo  
Università di Torino  
E-mail: [andrea.balbo@unito.it](mailto:andrea.balbo@unito.it)

Dr. Elisa Della Calce  
Università di Torino  
E-mail: [elisa.dellacalce@unito.it](mailto:elisa.dellacalce@unito.it)

---

<sup>32</sup> The rights pertaining to content, text, graphics, and images, unless otherwise noted, are reserved by the authors. This contribution is licensed under CC BY-SA 4.0.