PHILOLOGIA HISPALENSIS

ESTUDIOS LITERARIOS

PHILOLOGIA HISPALENSIS

AÑO 2025
VOL. XXXIX/2

ESTUDIOS LITERARIOS

FACULTAD DE FILOLOGÍA

UNIVERSIDAD DE SEVILLA

ÍNDICE

Reseñas de libros

TAMING THE BEAST: TRANSCRIBING HERNANDO COLÓN'S
LIBRO DE LOS EPÍTOMES

DOMANDO A LA BESTIA: TRANSCRIPCIÓN DEL LIBRO DE LOS EPÍTOMES
DE HERNANDO COLÓN

Matthew Driscoll
Department of Nordic Studies and Linguistics, University of Copenhagen
mjd@hum.ku.dk
0000-0001-9662-1271

Alessandro Gnasso
Department of Nordic Studies and Linguistics, University of Copenhagen
agnasso@libero.it
0000-0001-5964-2989

Abstract

In 2019, manuscript AM 377 fol. of the Arnamagnæan Collection in Copenhagen was identified as the fair copy of the *Libro de los epítomes*, one of the main catalogues of the 16th-century library of Hernando Colón. This discovery prompted the development of a research project centered on the *Libro* as a key source for the understanding of the functioning of Colón's library, possibly the largest in the world at the time.

This article introduces the project and describes the Copenhagen copy of the *Libro de los epítomes*, its nature, the circumstances of its arrival and preservation in Denmark and its role in Colón's library. It further reports on the TEI transcription of the manuscript developed as part of the project, focusing on a few noteworthy methodological and technical aspects of the enterprise and on the insights gained in the course of the project.

Keywords: *Libro de los epítomes*, manuscript transcription, TEI encoding, Hernando Colón, Biblioteca Capitular Colombina.

Resumen

En 2019, el manuscrito AM 377 fol. de la Colección Arnamagnæana de Copenhague fue identificado como la copia fiel del *Libro de los epítomes*, uno de los principales catálogos de la biblioteca del siglo XVI de Hernando Colón. Este hallazgo impulsó el desarrollo de un proyecto de investigación centrado en el *Libro* como fuente clave para la comprensión del funcionamiento de la biblioteca de Colón, posiblemente la mayor del mundo en aquella época.

Este artículo presenta el proyecto y describe el ejemplar de Copenhague del *Libro de los epítomes*, su naturaleza, las circunstancias de su llegada y conservación en Dinamarca y su función en la biblioteca de Colón. Además, informa sobre la transcripción TEI del manuscrito realizada en el marco del proyecto, centrándose en algunos aspectos metodológicos y técnicos dignos de mención y en los conocimientos adquiridos en el transcurso del proyecto.

Palabras clave: Libro de los Epítomes, transcripción de manuscritos, codificación TEI, Hernando Colón, Biblioteca Capitular Colombina.

1. *El libro de los epítomes*: an introduction

The Arnamagnæan Institute, part of the Department of Nordic Studies and Linguistics at the University of Copenhagen, houses an important collection of manuscripts and archivalia collected by the Icelandic scholar Árni Magnússon and bequeathed by him to the university upon his death in 1730. The collection comprises mostly early Scandinavian manuscripts, the majority from Árni's native Iceland (2300 items), although there are also significant Danish (250), Norwegian (282) and Swedish (32) holdings. There are also around 100 items of continental provenance, including some 22 Spanish manuscripts. In 2019, one of these, with the shelfmark AM 377 fol., was identified as the hitherto presumed lost fair copy of the so-called *Libro de los epítomes*, one of a number of bibliographic tools from the library of Hernando Colón (1488-1539), second son of the famous navigator Cristóbal Colón, known in English as Christopher Columbus.[1]

In the first decades of the 16th century, Hernando Colón set out to build a universal library, one which would contain "todos los libros y de todas las lenguas y facultades que se podran por la christiandad y en fuera della hallar"[2] (Felipe, 1539/1941: 241). By the time of his death in 1539, Colón's library comprised over 15000 volumes,

---

[1] This article builds upon the research undertaken by the *Book of Books* project, based at the University of Copenhagen and supported by the Carlsberg Foundation and by a private donor. An in-depth study of AM 377 fol. is in preparation by Matilde Malaspina (forthcoming), to whom we express our gratitude for many constructive discussions about the manuscript over the course of the project and for sharing her provisional findings with us. An online facsimile of the manuscript is available at https://handrit.is/manuscript/view/en/AM02-0377.

[2] "All the books, in all languages and disciplines, that can be found within Christendom and without". Unless otherwise noted, all translations are by the authors.

the bulk of them printed books, possibly the largest private collection in the world at the time.[3]

To manage it all he designed a revolutionary cataloguing system (essentially anticipating many aspects of modern library science). In his eyes, the four principal inventories of his library were lists of "Autores", "Ciencias", i. e. subjects, "Materias", themes or keywords, and "Epítomes", summaries of the contents of each of the books in the library. Modern scholars make ample use of two further inventories from the library, the "Abecedarium B", an index of authors, titles and incipits in alphabetical order, and the "Registrum B", an accessions list containing information on where and when each book was published, where and when it was bought and how much Colón paid for it (both in local currencies and in Spanish ducats). All of these inventories and lists were carefully cross-referenced. All also survive in the Biblioteca Capitular Colombina in Seville, apart from the *Libro de los epítomes*, which was long believed to be lost.[4]

This "rediscovery" of the fair copy of the *Libro* led, in 2020, to the establishment of the research project "The Book of Books: Hernando Colón's *Libro de los epítomes*", which received funding from the Carlsberg Foundation and a private donor.[5]

The manuscript AM 377 fol. has thus been part of the Arnamagnæan Collection since the end of the 17th century, but exactly how and when it entered Árni's collection is not entirely clear.[6] What is clear is that its connection with Hernando Colón's library was unknown at the time of its acquisition by Árni, as it appears to have been for some time before that.

The description of the manuscript in the catalogue of the Arnamagnæan collection prepared by Kristan Kålund reads simply: "Udførligt (ikke nordisk) forfatter-lexicon, på Latin"[7] (Kålund, 1888: 296). An equally brief description of the manuscript in the first catalogue of the collection, prepared by Jón Ólafsson frá Grunnavík, Árni Magnússon's amanuensis, shortly after Árni's death in 1730, reads: "De variis authoribus, eorumque scriptis judicia, signatus est liber E. S."[8] (manuscript Copenhagen, Arnamagnæan Collection, AM 384 fol.).

---

[3] The best over-all study of Colón's library is still Marín Martínez (1970). See also Marín Martínez (1993) and Ruiz Asencio (2008). In English, see McDonald (2004, 2005), Wilson-Lee (2018) and Pérez Fernández and Wilson-Lee (2021).

[4] For an introduction to the functioning and cross-referencing of these inventories, see Malaspina (2024).

[5] For more information on the "Book of Books" project, see the project's website: https://bookof-books.ku.dk/

[6] Árni Magnússon acquired many of his manuscripts –about forty, including at least eight of the Spanish manuscripts– from the collection of the aristocrat and bibliophile Jens Rosenkrantz (1640-1695), but there is no evidence that the *Libro* was one of those. On manuscripts in the Arnamagnæan Collection acquired by Árni from Rosenkranz, see Overgaard (1996).

[7] "Comprehensive dictionary of (non-Nordic) authors, in Latin".

[8] "On various authors, and judgements on their writings, the book is marked E. S.".

Most of the Spanish manuscripts in Árni's collection, at least 18 of them, including this one, had been part of the collection of Cornelius Pedersen Lerche (1615-1681), who was Danish envoy to Spain in the mid-17th century, though, as was said, whether Árni Magnússon acquired the manuscript directly from him or via another Danish collector is not known. In the catalogue of Lerche's library prepared when his books were auctioned off following his death, the entry on the manuscript reads "Judicia XVIII. De variis Authoribus, eorumq[ue] scriptis, MS." (*Catalogus Selectissimorum Librorum*, 1682: 164), and is thus essentially identical to Jón Ólafsson's description, which can hardly be coincidental.

At the same time as the manuscript was identified as the lost *Libro*, N. Kıvılcım Yavuz (2019) discovered that it had been part of the library of Gaspar de Guzmán, the Count-Duke of Olivares (1587-1645), before being acquired by Lerche and brought to Denmark. This is demonstrated by an inventory of Guzmán's library originally compiled in 1627, although only surviving in later copies. Here the entry on the manuscript reads: "Bibliotheca virorum illustrium et opera quae scripserunt, magnus codex 2330 aucthorum [*sic*] fol. E.5."[9] (manuscript Madrid, Real Biblioteca del Palacio Real de Madrid, II/1781, f. 126r). This suggests that even in Guzmán's time, the manuscript's connection to Colón's library was unknown –or at least not openly acknowledged. Guzmán's original signature, "E.5.", later misread by Jón Ólafsson and Kålund as "E.S.", is still prominently visible on the recto of what is now, and clearly was already at the time, the manuscript's first leaf.

AM 377 fol. is indeed a "magnus codex" –and with a thickness of nearly 15 cm, the thickest book in the Arnamagnæan Collection– comprising in its present state 974 leaves, with an average size of 215 mm × 308 mm. Originally it must have been significantly larger, given that it is, and has been at least since Guzmán's time, acephalous, lacking an unknown number of folios at the beginning, possibly around 70, so that it begins in the middle of epitome 140. It ends, as Guzmán's catalogue indicates, with epitome 2330, which is incomplete.

Within the volume there are a further three lacunae; these have been filled with blank leaves in modern paper, after the manuscript came to Denmark. The lacunae are:

1. ff. 297r-302v (12 leaves), previously containing 273 epitomes 1131-1404 (1130 breaks off toward the end; only the last few lines of epitome 1405 remain). Such a large number of epitomes would not have been able to fit on just 12 leaves; given the average of two epitomes per leaf, the number of missing leaves is more likely to have been in the vicinity of 135.
2. ff. 311r-318v (16 leaves), containing 15 epitomes 1416-1431 (epitomes 1415 and 1432 are both complete).

---

[9] "A library of distinguished men and the works they have written, a large book of 2330 authors. Folio E.5.".

3. ff. 605r-618v (28 leaves), containing 23 epitomes 1560-1583 (the end of 1559 and the beginning of 1584 are missing).

The final leaves of the original book block show water damage, with material losses to parts of each page, of increasing severity, from f. 957 to f. 973 (the last leaf of the original book block; ff. 974-980 are later flyleaves). The written area is affected from f. 966 onwards, causing partial losses in the text of epitomes 2311-2330. In addition, some worm holes extend from the last leaf into the book block, with traces visible up to f. 887.

There is, for some reason, no epitome 1551 in the *Libro*, and there are two epitomes numbered 822. There are thus 1871 complete epitomes plus six incomplete ones, of which only the beginnings or the endings remain.[10]

Parts of a draft copy of the *Libro* survive at the Biblioteca Capitular Colombina in Seville (Sevilla, Biblioteca Capitular Colombina, 10-1-17). The surviving portions are in rather poor shape and difficult to read in places, but nevertheless remain a valuable source of information. They contain 819 complete epitomes and ca. 10 incomplete ones. Of these, about 527 epitomes survive in both the draft and fair copies, although they are not necessarily legible in both.

Following Colón's death, Juan Pérez, Colón's principal librarian, recorded how the library was organised in a remarkable document titled "Memoria de las obras y libros de don Hernando Colón", which survives in a copy in the Real Biblioteca in Madrid.[11] In the "Memoria", the *Libro de los epítomes* is described as "un libro grande de cuadernos de pliego de pliego [*sic*] cosidos en pergamino blanco, espcritos [*sic*] de mano, de buena letra"[12] (Pérez, ca. 1539-1544/1970: 52). Pérez also states that there also were "dos libros viejos de cuarto de pliego, escritos de mano, de muy mala letra, encuadernadas [*sic*] en pergamino y atados entrambos juntos, los cuales son de epítomes de libros de la librería"[13] (Pérez, ca. 1539-1544/1970: 51). It seems obvious that the manuscripts here described are, respectively, the fair copy of the *Libro*, now AM 377 fol. in the Arnamagnæan Collection,[14] and the draft copy, still partially preserved, as mentioned, at the Biblioteca Capitular Colombina in Seville.

---

[10] We are grateful to Matilde Malaspina for this data, which will be examined in greater detail in Malaspina (forthcoming).

[11] Madrid, Real Biblioteca del Palacio Real de Madrid, II/652 (9). The reference edition, by Tomás Marín Martínez, is Pérez (ca. 1539-1544/1970). English translations are available in McDonald (2004: 269-277) and Pérez Fernández and Wilson-Lee (2021: 201-226).

[12] "A big book with sheets in folio, bound in white vellum, written by hand in good handwriting".

[13] "Two old books in quarto, written by hand in very bad handwriting, bound in parchment and tied together, which contain epitomes of the books of the library".

[14] AM 377 fol. is currently in a late 18th-century cardboard binding, but traces of one or more previous parchment bindings can be seen in the spine.

The *Libro de los epítomes*, as its name indicates, was meant to contain short summaries of the books in the library. As Pérez explains in the "Memoria":

> El gran provecho de este libro de los epítomes está muy notorio porque [por] él en breve se podrá saber lo sustançial que el libro trata difusamente, y si uno no puede tener muchos libros para lleellos, a lo menos t[e]rná uno que le dará una muestra de lo que se trata en muchos, y de allí resulta que si le paresçiere bien el libro y la materia de que trata, comprallo ha, y si no, dexallo ha, y no se engañará en comprarlo, porque hay muchos libros de grandes y hinchados títulos y después no tractan lo que prometen y esto hazen los impresores por engañar. [15] (Pérez, ca. 1539-1544/1970: 51)

We discovered quickly in our work with the *Libro* that Colón makes a clear distinction between a "work" and its physical manifestation in a given book. This means, among other things, that the epitomes generally do not indicate which editions of the works being summarised were used. To find that out, one has to consult the other inventories, in particular the "Abecedarium B" and the "Registrum B". It also means that in many cases there are multiple epitomes for a single book. A typical example is the case of a work and one or more commentaries on it, which are dealt with in separate epitomes, even when these commentaries have never appeared separately from the work they refer to. As a result, although there are 1878 separate epitomes preserved in the fair copy of the *Libro*, these represent only c. 1450 bibliographical items (books). In extreme cases, there can be as many as 12 separate epitomes of works contained in a single volume. This is for example the case with epitomes 1939-1950, referring to the medical miscellany *Articella medicinae*, containing works by Johannitius (Abū Zayd Ḥunayn ibn Isḥāq al-'Ibādī) (Venice, 1507), or with epitomes 1458-1469, referring to Johannes Annius's *Antiquitatum variarum volumina XVII* (Paris, 1515). These represent a minority, however, with around 1220 epitomes dealing with single bibliographical items in their entirety. In a very few cases, the opposite is true, where an epitome deals with multiple bibliographical items. The works of Plato summarised in epitome 1444, for example, were accessed from a number of unrelated editions.

The text is written in various hands. Matilde Malaspina (forthcoming) identifies nine principal hands, responsible for transcribing the text, and six secondary hands, responsible for interventions such as corrections and annotations; among these latter is the hand of Hernando Colón.

---

[15] "The great benefit of this book of epitomes is very obvious because it allows you to know briefly that which the book treats more diffusely, and if a person cannot afford to have many books to read, at least he will have one that will give him a sense of the contents of many books, and this allows him, if he finds the book and the subject it treats fitting, to buy it, and if he does not, then he will leave it, and will not be cheated in buying it, since there are many books of grand and swollen titles which then do not deal with what they promise, and publishers do this in order to cheat readers".

Juan Pérez explains in his "Memoria" how work on the *Libro* proceeded (or, perhaps, how it was ideally supposed to proceed). Colón employed qualified scholars, whom he called *sumistas*, to read and summarise the contents of the books in the library; the summaries would then be edited and corrected by Colón himself and subsequently transcribed into the *Libro*. Pérez explains that

> y es verdad que aunque he dicho arriba que los epitomes que estan trasladados en limpio se han de corregir, pero no todos, porque alguno de ellos estan ya corregidos y son los que tienen escrito en el margin 'vidi' que quiere dezir el tal epitome esta visto y corregido y aun otros hay a la poste los cuales corregimos yo y el maestro y no tienen señal pero puedese ver en que hallaran en ellos algunas partes borradas y emendadas en la margen.[16] (Pérez, ca. 1539-1544/1970: 53)

There are many cases, particularly on the first 30 leaves of the manuscript, of such corrections, both interlinear and marginal. The word "vidi" appears next to most entries on these early leaves, but stops suddenly on f. 34v. The majority of the entries do not appear to have been revised. Many do contain errors and omissions, however.

The epitomes themselves differ greatly in length. Dividing the 1871 epitomes which survive in full by 954 written leaves in the manuscript, we get an average of close to two epitomes per leaf, one per page. The actual lengths vary widely, however. The longest epitome in terms of number of pages is 1454, dealing with the *Disputata in quatuor libros Magistri Sententiarum* by the 14th-century Spanish Franciscan Guillen Rubio, which occupies 26 leaves (ff. 387v-413r), 52 pages, while many of the shortest epitomes are a mere three or four lines, scarcely stating more than the name of the author and subject of the book.

2. The transcription

One of the main deliverables of the "Book of Books" project is a full transcription of the text, roughly 400,000 words of 16th-century Latin.[17] The transcription is encoded in XML in accordance with the Text Encoding Initiative's (TEI) encoding guidelines (P5). TEI encoding is a well-established standard, widely used in the humanities

---

[16] "And it is true that although I have said that the epitomes that are transcribed in clean must be corrected, this does not apply to all of them, since some are already corrected. These epitomes have the word 'vidi' written in the margin, which means that an epitome has been checked and corrected. Finally, there are others that were corrected by the *maestro* and by myself, and they are not marked, though they can be distinguished because some parts of them have been erased and corrected in the margin".

[17] A single epitome, number 508, is in Spanish rather than Latin. Its text is thought to have been composed by Hernando Colón himself.

both for long-term preservation of data and easy interoperability. The resulting XML files can include multiple layers of markup and metadata.

The initial release of the transcription will contain the entire text of AM 377 fol., and encode the basic structures of both the text (for example, its division into epitomes) and the manuscript (for example, its division into pages and lines). All additions, deletions, corrections and marginal annotations will be included, with each intervening hand explicitly identified. The transcription will include deleted text which is still readable; all intermediate stages of correction and textual revision will be represented as accurately as possible. This initial release is intended to serve as the basis for a critical edition of the *Libro*, and as a preliminary stage for further study of the text. Successive releases will refine the encoding, revise inaccuracies and add metadata as provided by (or needed for) ongoing research, such as explicit identification of all authors named in the text.

TEI is a mature and extensively tested standard, able to represent in its encoding most, if not all, of the features of the source text (and, in our case, of the source manuscript) which a transcriber may be expected to encounter and to desire to express in the transcription. Difficulties in encoding may arise, but are often limited to unusual features of the source material. This holds true for AM 377 fol., which is relatively conventional in most of its aspects, but contains some heavily corrected and rewritten passages: the most significant difficulties in an otherwise relatively straightforward encoding all relate to these passages.

In this context, we believe that a general report on the transcription work would for the most part be rather unremarkable, substantially repeating the same experiences of many other similar projects. Instead, the remainder of this article will address three specific aspects of the work which proved relatively unusual (without being unique or overly specific to this project) and which we believe might be of interest to other researchers.

2.1. Multiple planned outcomes for the transcription

The transcription was initially conceived as the basis for a printed edition and translation, to be published by Oxford University Press, but further outcomes and reuses for the transcription emerged as the project matured. Early in the development it was decided to include a plaintext version of the transcription in the *Book of Books* database we have developed, hosted by the Consortium of European Research Libraires (CERL), which will eventually contain information not just on the books epitomised in the *Libro* but on all the books in Colón's library –a step towards a virtual reconstruction of this monumental library. Over time, further re-uses were envisaged, such as a second plaintext version (with different conversion criteria) optimised for natural language processing (NLP) and a corpus of transcriptions suitable for usage as ground truth in the development of a handwritten text recognition (HTR) model.

We decided that the transcription should aim to contain all data needed for all (re)uses in a single TEI document. XSLT transformations would then extract from this document the data required for each intended use and format it accordingly. Each of the additional outcomes brought its own requirements and desiderata, which may not align with each other. In addition, these were to be balanced against the practical necessity of keeping the encoding (relatively) simple to manipulate via XSLT, which effectively ruled out introducing conflicting hierarchies.[18] In particular, it is worth noting that HTR ground truth would benefit from radically different normalisation and encoding criteria than the edition.[19] Such drastic changes could not be accommodated, but minor adjustments in the encoding (for example, explicit and more detailed encoding of punctuation and blank spaces) were introduced where possible. The edition remained the main outcome, and most of the encoding and normalisation criteria reflect this; any "additional" desiderata from other projected outcomes were considered secondary and kept to the bare minimum.

TEI encoding (and XML encoding in general) is theoretically human-readable as well as machine-readable. In practice, it may contain large amounts of encoding markup interspersed within the text, as well as multiple textual layers (e. g. notes, deletions, corrections or rewritings). As these increase in size or frequency, the text (or, eventually, the multiple, co-occurring texts) may become rather difficult to read, even to scholars familiar with TEI: beyond a certain threshold, a TEI document is effectively better treated as computer code than as human-readable marked-up text. To facilitate reading, proofreading and circulation of transcriptions of individual epitomes both within and outside the project, we thus devised a series of scripts to export the text to Office Open XML format (the format employed by Microsoft Word).[20] This effectively became an additional, informal outcome.

---

[18] XML structures can directly represent only one hierarchy. This causes difficulties when one intends to represent two competing hierarchies at the same time, for example the division of a work into pages as well as its division into chapters. Workarounds are possible, but add complexity to the encoding.

[19] A transcription optimised for HTR should employ what the TEI Guidelines call "embedded transcription" encoding, i. e. a transcription structured around the representation of (usually text-bearing) areas of a digital facsimile image, rather than around the (reconstructed) flow of text. In addition, it would benefit from diplomatic or semi-diplomatic transcription criteria.

[20] This is done through a short chain of custom scripts, first converting from XML to markdown using XSLT, and then converting the resulting markdown to Office Open XML (Word) using Pandoc. Since TEI can encode significantly more information than Office Open XML can easily and intuitively visualise, some information must be discarded in the process. Deciding what information should be retained in the output format, and how it should be visualised, was in fact the most complex aspect of development. Office Open XML was preferred to XHTML in order to allow readers to more easily add (and send back) comments or corrections. The TEI consortium provides its own conversion scripts from TEI to Office Open XML, but they did not cover all our needs (especially in regard to visualising corrections).

2.2. Transcription size, collaboration and use of XInclude

The size of the manuscript influenced a number of basic decisions and preliminary considerations. Even before the formal start of the project, it was decided not to crowdsource a preliminary transcription, but to have, in principle, a single transcriber responsible for the entire work. Over the course of the project, nevertheless, a handful of additional researchers contributed transcriptions of a number of epitomes, as well as corrections to the initial transcriptions; these contributions are acknowledged within the transcription through use of the "responsibility" (@resp) attribute. In order to facilitate coordination of multiple transcribers potentially active at the same time, the text of the work-in-progress transcription (i. e., the contents of /TEI[1]/text[1]) was split into a number of arbitrarily-defined sections, each stored in a separate TEI document. This allows multiple contributors to be active at the same time without any synchronisation issue, as long as each of them works on a separate document (i. e., file) at a time. File exchange between contributors, as well as versioning, is handled through a Git repository. As a significant side benefit, working on relatively small "section" files, rather than on an extremely large single file, greatly improves the responsiveness of the XML editor, especially in connection with validation.

The TEI document containing the complete transcription is generated dynamically from a "main" file which retrieves the relevant sections of the various "source" (section) files via XInclude instructions.[21] Compared to other mechanisms (e. g. XSLT), XInclude has the advantage of real-time native support in the XML editor used by the project, oXygen XML Editor. This allows the most recent modifications to a source file (including unsaved ones) to be constantly mirrored by the main file, without having to save the source file or regenerate the main file.

It should be noted that native support of XInclude in oXygen is limited to including a single element and its descendants per instruction; it is not possible to include "all descendants of a given element" without also including the element itself.[22] This causes a small difficulty in our case, as epitomes should be encoded as <div type="epitome"> element children of <body>:

---

[21] The TEI Guidelines discuss XInclude in section 17.9, but only in the context of stand-off annotation. Our usage employs the same framework for a different purpose. The Guidelines use the term "external file" for what we call the "main file", this is simply a matter of nomenclature, reflecting the different usage perspectives.

[22] This is due to the fact that oXygen only natively supports the XPointer element() scheme. Most editors either share the same limitation or do not support XInclude at all.

```
<body>
   <div xml:id="e0140" type="epitome" n="140">[...]</div>
   <div xml:id="e0141" type="epitome" n="141">[...]</div>
   [...]
</body>²³
```

Importing several hundred <div> elements is impractical, both because of their number and because in the course of transcription the transcriber might insert additional nodes between two consecutive <div> epitome elements, for example a page beginning (<pb/>), a note (<note>) or even an XML comment. They would then have to insert a matching XInclude instruction in the main file, which would defeat the purpose of limiting concurrent modifications of the main file. Importing multiple <body> elements (one from each "source" file) and their descendants is possible, but would produce a XML document which is not valid TEI, as TEI only allows a single <body> element in <text>.

As a workaround, in each "section" document we wrapped all children of <body> in an arbitrary <div type="xinclude_wrapper"> element, which is the element actually included in the "main" TEI document. Thus, the source file for epitomes 140-508 (e0140-e0508.xml) reads:

```
<body>
   <div xml:id="ee0140-0508" type="xinclude_wrapper">
      <pb xml:id="f001r"/>
      <div xml:id="e0140" type="epitome" n="140">[...]</div>
      [...]
      <div xml:id="e0508" type="epitome" n="508">[...]</div>
   </div>
</body>
```

The main file includes these wrappers as:

```
<body>
   <xi:include href="e0140-e0508.xml" xpointer="element(ee0140-0508)"
   parse="xml"/>
   <xi:include href="e0509-e0801.xml" xpointer="element(ee0509-0801)"
   parse="xml"/>
   [...]
</body>
```

---

²³ This and all following examples are simplified pseudo-code. "[...]" indicates the omission of any number of nodes.

After the XInclude instructions are processed, the resulting "complete" TEI document retains the spurious "wrapper" <div> elements (together with all their content, for example the initial <pb/>):

```
<body>
   <div xml:id="ee0140-0508" type="xinclude_wrapper">
      <pb xml:id="f001r"/>
      <div xml:id="e0140" type="epitome" n="140">[...]</div>
      [...]
      <div xml:id="e0508" type="epitome" n="508">[...]</div>
   </div>
   <div xml:id="ee0509-0801" type="xinclude_wrapper">
      <div xml:id="e0509" type="epitome" n="509">[...]</div>
      [...]
      <div xml:id="e0801" type="epitome" n="801">[...]</div>
   </div>
   [...]
</body>
```

Since these <div type="xinclude_wrapper"> elements are utterly arbitrary divisions which do not correspond to anything present in the source manuscript or its text, their usage is a pure workaround, even if they are valid against the default tei_all schema. For this reason, they are only retained for day-to-day work; when the "complete" transcription is to be saved as a single self-contained file (as opposed to a dynamically generated XML document), an XSLT transformation removes the wrapper <div> elements, moving their content directly under <body>. Thus, these wrapper elements will not appear in the versions of record of the transcription, which will be released as a single XML file, to simplify long-term archiving and reference.

There is a further caveat in connection to our usage of XInclude: the mechanism is not aware of any cross-reference or pointer, and does nothing to ensure their coherence. For example, in each source file, each epitome's <div> references the person(s) responsible for its transcription, encoding and eventual revision by referencing, through its @resp attribute, a <respStmt> (responsibility statement) element present in the source file. When these @resp attributes are included (together with their <div> parent) in the main file, there is no guarantee that this latter file will contain a matching <respStmt>. Using XInclude to import the contents of <respStmt> elements from the source files runs into the same type of problem encountered with <div> children of <body>. In our case we found it simpler to periodically check the consistency of all cross-references and pointers (not just @resp) in the main file at once via a dedicated XSLT script.

2.3. Encoding corrections and modifications

Some sections of the manuscript present a large number of corrections. These range from self-correction by the copyist (who often evidently had difficulty deciphering the exemplar), to interventions by a corrector (who was presumably checking the text against the exemplar), to complete rewriting of entire passages. Since the very beginning of the project, the transcription has aimed to document all stages of corrections and modifications to the text, by transcribing all deleted text which is still readable and reporting the presence of unreadable deletions (using the <gap> element).

While this is relatively straightforward to encode for cases involving single interventions on the text, it can become challenging in the presence of multiple, consecutive corrections or modifications affecting the same passage. In such cases, documenting all stages of modification of the text requires one not just to decipher each individual intervention, but also to reconstruct their overall sequence; unsurprisingly, the complexity of both tasks increases significantly as the number of successive interventions grows. In addition, there are inevitable structural difficulties in representing (extensive) modifications in XML, as this means representing multiple textual stages at the same time, each of which may have a different set of textual hierarchies. The end result is that the hierarchies can overlap, or conflict with non-textual ones (such as the division in pages and lines). This is relatively common, and TEI provides some workarounds for such cases, for example the <delSpan> element.

As an example, on f. 14v, both the end of epitome 190 and the beginning of epitome 191 have been deleted in a single act, i.e. the deletion spans across the hierarchical separation between two <div> elements. On paper, this appears to be the ideal case for <delSpan>, and could have been encoded as follows (for simplicity, the relations between the various <del> and <add> elements in the passage are left undefined):

```
<div xml:id="e0190">
   [...]
   <ab>
      [...]
      <lb break="no"/>iore reperitur.
         <add><!-- large addition from the right margin -->[...]</add>
         <del>Adiectus est preterea</del>
         <delSpan spanTo="#s1end">libellu<del>s</del><add>m</add>
         qu<del>i</del><add>e<ex>m</ex></add>dam cuius auc
      <lb break="no"/>tor ignoratur [...]
      [...]
   </ab>
</div>
```

```
<div xml:id="e0191">
   <lb/><head>191</head>
   <ab>
      <lb/>Joannes Stofflerinus vir natione germanus [...]
      <anchor xml:id="s1end"/>
      <pb xml:id="f015r"/>
      [...]
   </ab>
</div>
```

In practice, the sequence of rewriting still gives rise to implicit conflicts between hierarchies. The deleted text has been replaced in part by an additional epitome (in the left margin, which must be inserted between 190 and 191), and in part by rewriting the incipit of epitome 191 in the upper margin of the facing page.[24] While <delSpan> as an empty element does not itself cross the hierarchical separation between the <div> elements, its effective meaning does, and it would be extremely confusing (to say the least) to transcribe the added epitome between the <delSpan> and its matching <anchor>. To use <delSpan>, then, it should be fragmented in two elements (one for ep. 190, the other for ep. 191), aggregated through @prev and @next attributes. The replacement text (from the right margin) and the replacement epitome (from the left and lower margin) could then be transcribed (as <add> elements) between them. Finally, a <substJoin> element would aggregate the <delSpan> elements and the <add> elements into a single textual substitution.

In practice, this removes many of the advantages provided by <delSpan> over a series of fragmented <del> elements aggregated through @prev, @next and <substJoin> as needed. In fact, in general we found that whenever complex features are encountered, using fragmented elements is often simpler (albeit much more verbose) than using spans (or, more accurately, than inevitably using a mixture of span and non-span elements). Verbosity is of course a significant downside, as the resulting code may effectively stop being human-readable due to this. Such coding is significantly simpler to process via XSLT, however, and (perhaps counter-intuitively) appears to be somewhat simpler to revise and maintain (a human can eventually still decode it, given enough time).

---

[24] The new epitome is added in the left and lower margin (incipit "Libellum quedam…") without an explicit epitome number, but with a "vidi" note. We arbitrarily labelled it "epitome 190 bis" for convenience. The addition in the right margin (incipit "quequidem membra…") was inserted at a previous stage (its final, now-deleted line replaced "Adiectus est preterea" in the text) and is not connected to the deletion currently examined except for the final line (which was deleted at the same time as the large deletion in the main writing area). The large deletion thus begins from "libellum quemdam" (note the changed case, matching the changes in the verb from "adiectus est" to "adiecit").

For reference, the actual encoding of the passage where the large deletion starts (lines 41-42 on f. 14v) was ultimately encoded along the lines of that shown here (simplified pseudo-code, indenting spaces and line breaks before the elements with @xml:id s5, s6, s7, s8, s13 and s14 were added purely for readability and should be ignored):

```
<lb n="41" break="no"/>iore reperitur
<add xml:id="s1" seq="4" change="#c4b">
    quequide<ex>m</ex><!-- addition from the right margin -->[...]
</add>
<del xml:id="s2" seq="4" change="#c4b">
    adiectus est preterea
</del>
<del xml:id="s3" next="#s11" seq="5" change="#c5">
    libellu
        <del xml:id="s5" seq="4" change="#c4b"><seg change="#s1">s</seg></del>
        <add xml:id="s6" seq="4" change="#c4b"><ex>m</ex></add>
    qu
        <del xml:id="s7" seq="4" change="#c4b"><seg change="#c1">i</seg></del>
        <add xml:id="s8" seq="4" change="#c4b">e<ex>m</ex></add>
    dam cuius auc
</del>
[...]
<lb n="42" break="no"/>
<del xml:id="s11" prev="#s3" next="#s15" seq="5" change="#c5">
    tor ignoratur algorism<ex>us</ex> vulgo nuncupat
        <del xml:id="s13" seq="4" change="#c4b">
            <seg change="#c1"><ex>us</ex></seg>
        </del>
        <add xml:id="s14" seq="4" change="#c4b">
            <ex>um</ex>
        </add>
    in quo ęedem q<ex>uae</ex> dic
</del>
```

The verbosity of the passage is immediately evident, as is the consequent difficulty in reading what would otherwise be rather straightforward code.[25] The fragmented <del> (and <add>) elements are aggregated by via @next and @prev; the various interventions are in turn aggregated via <substJoin> elements (not shown in the example). The @seq attribute indicates the relative sequence of interventions in each <substJoin> group. The @change element is used systematically to explicitly indicate

---

[25] On a computer, automatic colour-coding of the XML code facilitates reading to some extent.

at which stage of textual revision each fragment of text was added or deleted; a lack of evidence suitable for synchronizing stages across epitomes means that most of them can only be defined in the context of a given epitome. In the case of text originally written as part of the initial transcription (not inserted by a later addition), the <seg> element is used as a container to explicitly express a @change attribute.

The case of epitome 190 is exceptional. An investigation of the epitome by Matilde Malaspina has identified between five and nine stages of revision to a text which was originally fewer than 20 lines long. The precise number is uncertain as some revisions can be assumed to have taken place at the same time, even if no evidence remains for this; in any case the consequences of such uncertainty on the overall encoding are limited. Other epitomes have also been subjected to successive rewritings, such as epitomes 167 (ff. 7v-9r) and 233 (ff. 25v-27r), but their modifications are not as complex to encode.

Our work was complicated by the co-occurrence of three factors: an exceptionally complex sequence of rewriting, a commitment to document every stage of rewriting and (least importantly) the desire to be able to satisfy as many of the transcription outcomes mentioned above, which required the avoidance of excessively convoluted encoding. In practice, the latter was abandoned; the Office Open XML conversion script is still unable to process epitome 190 correctly (an Office Open XML version of the epitome was instead manually curated).

It might be questioned whether the decision to document every stage of rewriting was justified, especially since many of them will not be relevant for an edition, for example a trivial self-correction by the copyist. Even if they are, inclusion in a printed edition of more than a couple of revision stages would be technically very demanding. In the case of epitome 190, anyway, a full transcription of all stages was in any case necessary in the context of the *Book of Books* project. Epitome 190 and the other reworked epitomes were studied in detail by Matilde Malaspina (see Malaspina, forthcoming) in order to investigate the characteristics the epitomes were expected to have, through an examination of the apparent deficiencies in the text(s) of these epitomes addressed by later revisions. A full transcription was necessary for the analysis; once this was produced, it seemed opportune to curate and encode it in TEI format, however complex the process was.[26] A similar argument applies to most of the other partially rewritten epitomes: while the original and intermediate texts might not be considered relevant enough to appear in a printed edition, a study of their evolution ultimately

---

[26] It should be mentioned that the encoding of epitomes 190 and 191 makes exceptional use of some markup not otherwise used (or only very rarely used) in the rest of the transcription. This includes the use of @change (which was originally added to assist with XSLT transformations) and explicit numbering (@n) of line beginnings. This additional information can be easily removed (e.g. via XSLT) for the sake of uniformity in encoding, although it is perhaps doubtful whether uniformity is more relevant than practicality in such an exceptional case.

allowed a significant improvement in our understanding of how the *Libro* was composed, copied and revised.

On the other hand, the question remains valid for the large numbers of frequent corrections, most of which appear to address copyist misreadings. The corrections are often made by overwriting the original text, often only affecting single letters across multiple words (e. g. grammatical case markers in a noun and in the adjective relating to it). This requires recourse to <substJoin>, with a consequent increase in markup verbosity even for relatively simple interventions.

3. In lieu of a conclusion: some remarks on the readability of manuscript transcriptions

TEI is often considered to be highly human-readable. While generally true, this is actually dependant on the ratio between text and encoding markup. In most TEI documents, or in fact in most TEI use case scenarios, this ratio is arguably not so skewed as to significantly hinder readability. In other words, readability issues tend to arise in edge cases; epitome 190 certainly is one. In our transcription, noteworthy difficulties affect fewer than a dozen passages out of 1891 written pages, and serious problems only arise in a handful of them.

Nevertheless, such difficulties do exist, and are (relatively) more frequent in some types of TEI documents, such as manuscript transcriptions. In the specific case of epitome 190, some of the difficulties are due to the nature of the information represented in the document, as it encodes a diachronic sequence of textual revisions, rather than "a text" in the common meaning of the term. At the same time, the XML encoding, while verbose and intrusive, is not particularly complex (at least in terms of element nesting). The problem therefore is probably mostly one of visualisation, rather than structure.

Author contributions

Matthew Driscoll wrote part 1 of the paper, Alessandro Gnasso wrote parts 2 and 3 of the paper. Both authors revised all parts of the article and contributed equally to the work.

References

Catalogus selectissimorum librorum omnium facultatum & linguarum, nobilissimi & illustrissimi d[omini] b[eatae] m[emoriae] Cornelii Lerke ... (1682). Literis Conradi Hartvigi Neuhofii.

Kålund, K. (1888). *Katalog over den Arnamagnæanske handskriftsamling, 1. Hæfte*. Gyldendalske Boghandel.

Malaspina, M. (2024). Aesopian Texts and Aesopian Editions in the Library of Hernando Colón. *The Library*, *25*(4), 389–429. https://doi.org/10.1093/library/fpae037

Malaspina, M. (forthcoming). *Copenhagen, Arnamagnæan Institute, AM 377 fol.: Some preliminary observations on the structure and content of Hernando Colón's Libro de los epitomes* [Manuscript in preparation].

Marín Martínez, T. (1970). *«Memoria de las obras y libros de Hernando Colón» del bachiller Juan Pérez*. [Cátedra de Paleografía y Diplomatica].

Marín Martínez, T. (1993). Hernando Colón y la Biblioteca Colombina. In T. Marín Martínez, J. M. Ruiz Asencio, & K. Wagner (Eds.), *Catálogo concordado de la biblioteca de Hernando Colón: Tomo I: 1-400* (pp. 19-352). Fundación MAPFRE América; Cabildo de la Catedral de Sevilla.

McDonald, M. P. (2004). *The print collection of Ferdinand Columbus (1488-1539): A Renaissance collector in Seville*. The British Museum Press.

McDonald, M. P. (2005). *Ferdinand Columbus: Renaissance collector (1488-1539)*. The British Museum Press.

Overgaard, M. (1996). Manuscripta Rosenkrantziana. *Bibliotheca Arnamagnæana*, *40* (*Opuscula*, 10), 262-285.

Pérez Fernández, J. M., & Wilson-Lee, E. (2021). *Hernando Colón's New World of Books: Toward a Cartography of Knowledge*. Yale University Press. https://doi.org/10.2307/j.ctv1b0fw2s

Ruiz Asencio, J. M. (2008). *La Biblioteca de Hernando Colón: Una aventura bibliográfica en el siglo XVI: Lección inaugural del Curso Académico 2008-2009*. Universidad de Valladolid. http://uvadoc.uva.es/handle/10324/4542

Wilson-Lee, E. (2018). *The catalogue of shipwrecked books: Young Columbus and the quest for a universal library*. William Collins.

Yavuz, N. K. (2019, April, 15). *Hernando Colón's Book of Books: AM 377 fol.* https://manuscript.ku.dk/motm/hernando-colons-book-of-books

Primary Sources

Copenhagen, Arnamagnæan Collection, AM 377 fol. Digital facsimile https://handrit.is/manuscript/view/en/AM02-0377

Copenhagen, Arnamagnæan Collection, AM 384 fol. *Catalogus librorum Arnæ Magnæi*.

Felipe, M. (1941). Declaraciones de Marcos Felipe al testamento de don Hernando Colón (11 de noviembre de 1539). In J. Hernández Díaz, & A. Muro Orejón (Eds.). *El testamento de Hernando Colón y otros documentos por su biografía* (pp. 226-246). Instituto Hispano Cubano de Historia de América. (Original work written 1539).

Madrid, Real Biblioteca del Palacio Real de Madrid, II/1781. *Bibliotheca selecta del Conde-duque de Sanlúcar, gran chanciller, de materias hebreas, griegas, arábigas, latinas, castellanas, francesas, tudescas, italianas, lemosinas, portuguesas, etc.* Digital facsimile https://rbdigital.realbiblioteca.es/s/realbiblioteca/item/12690

Madrid, Real Biblioteca del Palacio Real de Madrid, II/652 (9). *Memoria de la orden que llevava Colón en su librería.* Digital facsimile https://rbdigital.realbiblioteca.es/s/realbiblioteca/item/2509

Pérez, J. (1970). Memoria de las obras y libros de Hernando Colón. In T. Marín Martínez (Ed.), *Memoria de las obras y libros de Hernando Colón» del bachiller Juan Pérez.* [s.n.] (Original work written ca. 1539-1544)

Sevilla, Biblioteca Capitular Colombina, 10-1-17.

UNIVERSIDAD DE SEVILLA

ueus

**FACULTAD DE FILOLOGÍA
UNIVERSIDAD DE SEVILLA**