

II. EDITING TEXTS WITH A COMPLEX TRANSMISSION

CRITICAL DIGITAL EDITIONS OF CHRISTIAN APOCRYPHAL LITERATURE IN LATIN AND GREEK: TRANSCRIPTION AND COLLATION OF THE ACTS OF *BARNABAS*

Caroline Macé, Maïeul Rouquette, Violeta Seretan,
Frédéric Amsler, Patrick Andrist, Cecilia Antonelli

Université de Lausanne

Abstract

The project *Editer Numériquement la Littérature Apocryphe Chrétienne* (ENLAC), which we are presenting here, aims at producing critical digital editions of a few Latin or Greek Christian apocryphal writings. The project will also provide a suite of interconnected tools, existing or new, for computer-assisted critical editing of works preserved in complex textual traditions. The tools will be open-source and the editions produced will be open-access.

In this article we are explaining the first steps of the process, i.e. the preparation of transcriptions

of each manuscript and their collation by CollateX in the form of a variant-graph. In the next steps the variant-graph will be transferred in Stemmaweb, which will be used to produce the critical editions. Taking the example of the Greek *Acts of Barnabas* (fifth century), preserved in twelve manuscripts, we are illustrating the principles that we followed for our diplomatic transcriptions in XML-TEI. Some peculiarities of mediaeval Greek manuscripts and how we are dealing with them, are presented, such as word splitting, punctuation, abbreviations, and corrections.

1. Aim and methodology

The project *Editer Numériquement la Littérature Apocryphe Chrétienne* (ENLAC) at the Swiss Institute of Biblical Sciences of the University of Lausanne, led by Frédéric Amsler and financed by the Swiss National Science Foundation, aims to produce critical digital editions of three dif-

KEYWORDS: Apocrypha / Greek Manuscripts /
Digital Critical Edition / Variant-Graph / Diplomatic Transcription

ferent works in Latin and Greek, each presenting specific philological challenges.¹ These editions are intended as fully critical and fully digital (or “born digital”). The project will also produce a suite of interconnected tools, existing or novel, for computer-assisted critical editing of works preserved in complex textual traditions.

In line with the digital focus of the project, the editions will be published in open access, both in a static (PDF) and a dynamic (website) form. However, they will also be published in a traditional way, as volumes of the Brepols series *Corpus Christianorum Series Apocryphorum*. This double format of publication, printed and digital, is necessary because established series act as guarantors of scholarly quality and as repositories of the editions.² The series in which the editions are to be published sets up specific rules for the critical editions of apocryphal texts, and our project will abide by these rules.

To comply with the scholarly standards of the field, the project aims to produce critical editions recognised for their high level of quality and accuracy. A critically edited text is a hypothetical reconstruction of the most ancient textual state of a work, based on the comparison and classification of the witnesses that preserve it. This is not the place to justify the necessity of such a critical approach to texts, and in what follows we will take it for granted.³ The specificity of our project is that it seeks to take full advantage of the possibilities offered by digital tools and methods in order not only to display the edition and witnesses, but also, more importantly, to produce the edition itself. Unlike in other edition projects, the digital technology serves for the creation, not only the visualisation, of the end product.

The methodology of the project relies, partly, on software modules developed in previous research projects. It is notably the case of Collatex and Stemmaweb,⁴ the first produced in the COST Action *Interedition*

1 An additional sub-project is devoted to the production of a multilingual alignment of the different versions of the *Pseudo-Clementine Romance*, on the basis of the already edited Greek, Syriac and Latin texts. This sub-project is aimed at reconstructing a hypothetical “basis text” (Grundschrift) of the *Pseudo-Clementine Romance*.

2 About the role of established series for the production of critical editions see Macé 2016.

3 See, for example, Trovato 2014.

4 The modules are described on the following websites: collatex.net (see Dekker *et al.* 2015) and stemmaweb.net (see Andrews-Macé 2013).

led by the Huygens Institut (2008-2012), and the second in the project *The Tree of Texts* carried out at KU Leuven (2010-2012). Both software modules are based on the idea that textual variation is best represented as a graph.⁵ Our project is expected to lead to the improvement and fine-tuning of these existing tools, as well as of the underlying graphbased model of textual variation (see Rouquette-Seretan, *forthc.*).

The process of critical digital editing consists of the following steps: (1) creation of TEI-XML files containing a scholarly diplomatic transcription of each witness, (2) collation of the TEI-XML files in CollateX, which produces a variant graph, (3) text-critical analysis based on the variant graph in Stemmaweb, where the whole edition data is stored as a Neo4j graph database.⁶ Stemmaweb will be used to analyse the differences between the manuscripts, to produce a *stemma codicum*, and to create the edition with its apparatuses.

In the present article, we will first present the works that will be edited in our project, focusing on the *Acts of Barnabas* as a case study (§ 2), then we will detail the first two steps of the process, i.e., the transcription and the collation of the manuscripts (§§ 3-5). Preliminary conclusions and perspectives for the future steps are outlined in § 6.

2. Works

The following works will be edited in the framework of the project:

- *Acts of Barnabas* (CANT 285, BHG 225, BHL 983-985)
- *Lives of the Prophets* (BHG 1585-1590)
- *Dialogue of Timothy and Aquila* (CPG 7794).

All three works belong to the early period of Christian history and were composed in Greek or, as in the case of the *Lives of the Prophets*, were preserved primarily in Greek. The *Acts of Barnabas* were translated into Latin in the early Middle Ages, and later into Old Slavonic. The

5 See the definition of a variant graph by Tara L. Andrews in the online *Parvum Lexicon Stemmatologicum* (wiki.helsinki.fi/display/stemmatology/Variant+graph) (released in November 2015, last accessed December, 2018).

6 At the end of the project, the data will also be stored on a free platform, possibly the nationally-centralised platform that is being built for the preservation of data in the humanities (knora.org) (last accessed December, 2018).

Dialogue of Timothy and Aquila is known in an Old Slavonic translation, and the *Lives of the Prophets* exists in many languages (in Latin and nearly all languages of eastern Christianity).

These works are preserved in a variable number of manuscripts: from about 12 for the *Acts of Barnabas* in Greek and for the *Dialogue of Timothy and Aquila*, up to more than 100 for the *Lives of the Prophets* and the *Acts of Barnabas* in Latin. They belong to different genres or are at the crossing of different genres: hagiography, biography, polemical dialogue. Because they were not considered canonical or authorial, they were subjected to deliberate interventions and changes in the course of their transmission, a situation which led to the existence of several recensions.⁷

Although one editor will be mainly responsible for each work and the editing sub-projects will be independent from one another, the project as a whole is carried out collaboratively. For instance, important issues are discussed in common; transcriptions and annotations are revised by other members of the team; and most importantly, the specifications of the editing pipeline are elicited and reviewed by all the sub-projects together.

For the purposes of the present article, we will focus our discussion on the *Acts of Barnabas*. These were written in Greek between 431 and 488 by members of the Church of Salamis in Cyprus.⁸ They narrate the journeys of Barnabas together with the apostle Paul, then the separation of Paul and Barnabas, and the evangelising mission of Barnabas and John-Mark, his companion and the alleged narrator of the *Acts*, in the island of Cyprus. At the end of the story, Barnabas is martyred in Salamis, his body is burnt, and his ashes are buried in a crypt by John-Mark, who then sails off to Egypt.

The *Acts* were translated into Latin and Old Slavonic.⁹ In Latin, the text was quite popular and is known through several recensions (BHL 983-986) and more than one hundred manuscripts. In the framework of this project, we will provide critical editions of the Latin texts as well, based on a complete examination of this large manuscript tradition.

The Greek text has been edited several times, the most recent edition being that of Maximilien Bonnet (1903). This edition is based on six Greek manuscripts and on editions of two Latin recensions (BHL 985

7 The three works have already been edited, albeit unsatisfactorily.

8 An English translation can be found in Snyder 2016, pp. 327-336; a French translation in Norelli 2001, pp. 629-642. See also Rouquette 2017.

9 Snyder 2016, p. 318 and n. 6.

and BHL 983). Bonnet was also aware of the existence of two more manuscripts, but did not use them. Four additional Greek witnesses were identified by Jacques Noret in 1994,¹⁰ making it necessary to produce a new edition. The twelve Greek witnesses are listed below in alphabetical order of their sigla.¹¹

- B Hagion Oros (Athos), Monê Batopediou, 84; ff. 202r-208r; saec. IX-X.
- C Cambridge University Library, Add. 4489; f. 11r-v; saec. VIII ex. (palimpsest fragment).
- F Hagion Oros (Athos), Monê Philotheou, 8 (Lambros 1771); ff. 62v-65v; saec. XI.
- K Hagion Oros (Athos), Monê Koutloumousiou, 2 (Lambros 3071); ff. 202r-208r; saec. XIII.
- L Hagion Oros (Athos), Monê Megistês Lauras, Γ 87 (Eustratiades 327); f. 220v-224v; saec. XI.
- M Messina, Biblioteca Regionale Universitaria, S. Salv. 29; ff. 81-83; AD 1307-1308.
- N Messina, Biblioteca Regionale Universitaria, S. Salv. 26; ff. 59-63v; saec. XIII.
- P Paris, BNF, grec 1470; ff. 55v-57v and 66-67; AD 890.
- Q Paris, BNF, grec 1219; ff. 31v-37v; saec. XI.
- R Paris, BNF, grec 1179A; f. 2; saec. XI (fragmentary).
- U Città del Vaticano, BAV, Vat. gr. 821; ff. 99v-103; saec. XII.
- V Città del Vaticano, BAV, Vat. gr. 1667; ff. 105-110; saec. X.

3. Transcription principles

The first step towards producing a critical digital edition is the full manual transcription of each available witness, which will be the input of the automated collation software CollateX.¹²

10 Noret 1994, p. 160 n. 5.

11 For more information about the manuscripts, see the Pinakes database (pinakes.irht.cnrs.fr) (last accessed December, 2018).

12 In non-digital editorial projects, the collation is performed manually, skipping the step of transcription. The substitution of the first traditional step (manual collation) with two different steps (manual transcription + automatic collation) was first developed by Peter Robinson, who created Collate, the ancestor of CollateX (see Robinson 1989). For a

Although computer tools assisting scholars in their task of transcribing manuscripts do exist (for example, T-PEN or Transkribus),¹³ none of them is suitable for our purpose. T-PEN, initially considered for transcription assistance in our project, adopts a line-by-line processing approach that makes it inappropriate for our project due to the time-expensive manual adjustment of line borders. Moreover, T-PEN is designed for facsimile transcriptions that encode palaeographic and codicological features of the manuscript, whereas for our project we need diplomatic transcriptions, to be used for collation.¹⁴ Transkribus, one of the most recent and best-known transcription tools, offers an automated transcription feature based on machine learning. However, the training data acquisition is a real bottleneck in our case. Transkribus needs an average of 30 to 40 pages of the same manuscript in order to learn to transcribe by comparing images and previous manual transcriptions. In our project, we work with many different manuscripts and relatively short texts. Therefore, we could not benefit from the automatic transcription feature of Transkribus, and continued our manual transcription using text editors with XML support, as explained below.

To make the process of transcription faster and safer, instead of starting from scratch, we use a model, in this case Bonnet's edition (1903). If there were no pre-existing edition, we would first transcribe one of the most complete and legible manuscripts and use that transcription as a model. For each new witness, we adapt this model. This method is a compromise between the transcription from scratch and the traditional method of comparing the manuscripts to a base text. This way, we take advantage of the two methods, producing transcriptions faithful to the manuscripts and checking our reading of the manuscripts against a pre-existing model. Another advantage is the possibility of inserting recurrent XML tags directly into the model.

Each witness is transcribed using TEI-XML schema.¹⁵ The information we encode falls into five main categories:

discussion of the respective advantages and shortcomings of both methods see Macé-Andrews 2015, as well as Macé-Gippert *forthc.*, and the section on collation by Tara L. Andrews in Roelli *et al.* *forthc.* About automated collation in general, see Nury 2018.

13 See t-pen.org and transkribus.eu (last accessed December, 2018).

14 On the differences between facsimile and diplomatic transcriptions, see for example the Menota handbook v. 2.0 (menota.org – last accessed December, 2018).

15 <http://www.tei-c.org/guidelines/> (last accessed December, 2018).

1. reference of the manuscript and its abbreviated identification (siglum);
2. main physical divisions of the text, namely folios (<pb />), columns (<cb />) and lines (<lb />);
3. titles, paragraphs and textual units (<head>, <p> and <seg>);
4. cases of complex textual situations (corrections, scribal notes, etc.) or uncertainties in our interpretation of the writing;
5. a few editorial annotations directly related to the specific manuscript.

The text itself is transcribed as it appears in the manuscript, respecting all its textual features. However the palaeographical peculiarities are not encoded, because they are not relevant for our purpose.

The XML-enabled text editor finally selected for our project is Atom (see fig. 1). It was chosen because it is free, easy to learn, and customizable for specific projects. Moreover, it has an integrated interface to the Git versioning system, which we use for the collaborative and iterative encoding and revision of transcriptions.

As a matter of fact, in order to ensure a high scholarly quality for the transcriptions, each transcription is checked by another member of the team and the changes are validated by the main editor. We record the different stages of each transcription using the Git versioning system. This allows us to not only to keep a history of file versions (and therefore of the transcription process), but also to have a convenient reviewing process, the corrections being easily checked and merged into the main file version if validated by the editor. Another advantage of using Git is that it provides a backup system, as the history can be stored in a distant server. In our project, we use the server of the free software support association Framasoft.¹⁶

At the transcription stage, we do not perform any textual analysis, apart from adding a very few editorial notes, and we do not yet tag any sources or proper names etc. The annotation is postponed until the editorial process proper, which will take place in Stemmaweb. During the editorial process, we may need to correct possible mistakes that might surface after the comparison of the witnesses. Therefore, the transcriptions will be redisplayed after correction in Stemmaweb and the final form will be posted on a dedicated website.

¹⁶ framagit.org (last accessed December, 2018). About Git, see for example Chacon-Straub 2014.

The TEI guidelines foresee many possibilities of encoding, out of which we selected a subset of tags and attributes useful for our project, and which could be adapted for other projects. This subset is summarised and explained in our *Manuscript Transcription Tutorial* that will be made available on the project website. Our TEI subset is formally described as a document-type-definition file (DTD), which declares the tags and the attributes that can be used, together with the relationships between tags and the allowed values for attributes. The DTD allows us to automatically check the compliance of the transcriptions to our XML encoding conventions, whereas the official TEI DTD is used for checking the adherence to the more general TEI guidelines.

The TEI-XML encoding is illustrated through a series of specific situations in the next section.

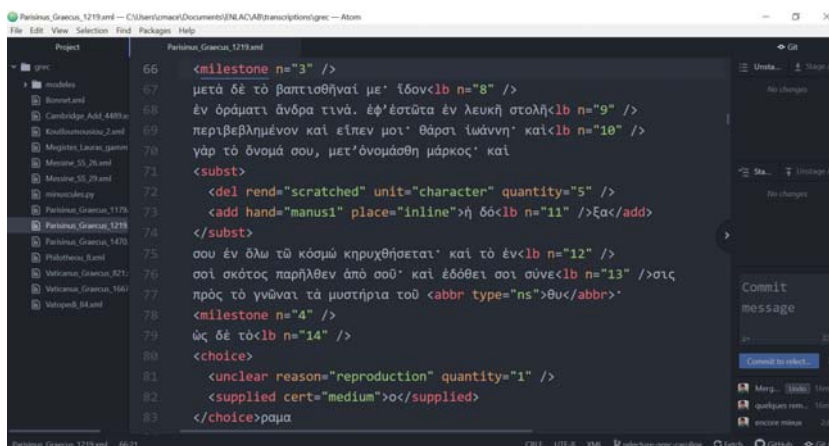


Fig. 1. Transcription of the text of Q f. 32r ll. 7-14, using Atom.

4. Transcription peculiarities

The transcriptions are made on the basis of reproductions provided by the libraries in which the manuscripts are kept. These reproductions are of varying degrees of quality, from excellent colour images showing every detail of the manuscript, to very poor reproductions of old microfilms, where letters are sometimes invisible or blurred. This is especially the case of manuscripts kept on Mount Athos, for which it is impossible to obtain better-quality images. Trivial as this observation may seem, it has an undeniable impact on the quality of our transcriptions, and some parts of

the manuscripts had to be indicated as illegible in the transcriptions due to the bad quality of the reproductions. If possible, difficult passages are checked *in situ* on the actual manuscripts.

It should be noted moreover that we do not intend to publish the images of the manuscripts as part of the edition. This task does not belong, in our view, to the editorial process, but should remain within the purview of other scholarly institutions, such as libraries.

As mentioned above, our purpose in transcribing the manuscripts is twofold: to provide textually (as opposed to palaeographically) reliable transcriptions of each witness (at the end of the editorial process), and to collate the transcriptions using CollateX, so that the variant graph can be used in Stemmaweb for a text-critical analysis. This double aim explains the decisions we take when performing the transcriptions.

4.1. Word splitting

Word splitting is the basis of any comparison. In Greek minuscule manuscripts, the division of words is not so much indicated by spaces (which are often difficult to discern, if visible at all), as by the presence of a spiritus on the vowel at the beginning of a word (of course if that word begins with a vowel) and of an accent on one of the last three syllables of a word. In many manuscripts, spiritus and accents are used in a way that shows that the rules for the use of these diacritics were not actively known any longer, and that copyists may have had some difficulties in deciding upon word divisions.

We introduce spaces to divide words, keeping obviously erroneous word divisions, such as μετὰ σὲ αὐτοῦ for μετὰ σεαυτοῦ – preposition with the reflexive pronoun of the second person in the genitive (see fig. 2a-b).¹⁷ In a case such as κατῆντήσαμεν (see fig. 3), we keep it in one word, despite the presence of two accents, because κατῆν- alone and -τήσαμεν alone do not exist as words. Many prepositions do not bear any accent, but we write them nevertheless as separate words. Conversely, several preverbs (prefixed prepositions at the beginning of a verb) receive an accent, but they cannot be considered independent words, because they are not followed by a substantive, for example συνάκολουθῆσαι (see fig. 4).

¹⁷ The same erroneous word splitting is found in two manuscripts (M and N), probably related to one another, thus showing that apparently very trivial mistakes can sometimes be stemmatically relevant.

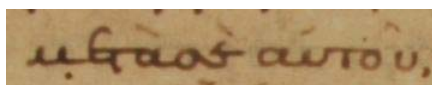


Fig. 2a. N f. 60ra l. 14.

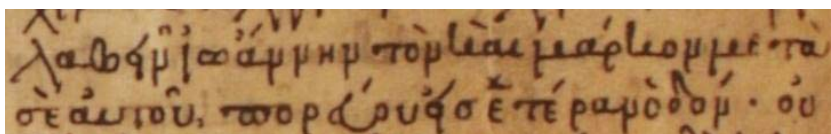


Fig. 2b. M f. 81vb ll. 16-17: μετά at the end of l. 16 and σε αὐτοῦ at the beginning of l. 17.

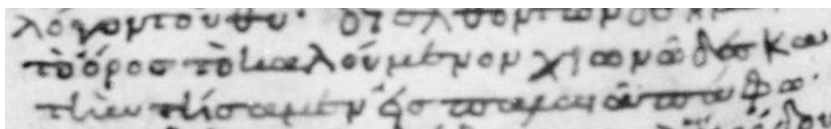


Fig. 3. K f. 205v ll. 21-22: κα- at the end of l. 21 and τηντήσαμεν at the beginning of l. 22.



Fig. 4. Q f. 32v l. 24.

To avoid having vowels in hiatus, the last vowel of some prepositions and adverbs can be elided, with or without a sign of elision (apostrophe), for example παρ' ἐμοῦ instead of παρὰ ἐμοῦ (see fig. 5). We introduce a space between the two words (παρ' ἐμοῦ) to facilitate the collation, although there should be no space there in the final output.¹⁸

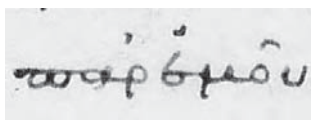


Fig. 5. Q f. 34r l. 22.

¹⁸ In many modern editions, a space is arbitrarily introduced after the apostrophe, but this modern practice does not make sense: see Noret 2007.

An additional problem concerning accentuation is that it is sometimes difficult on reproductions (even colour ones) to distinguish the type of accent or spiritus (copyists are sometimes very careless about it). When in doubt, we follow the standard accentuation. Corrections pertaining to accentuation are not coded as such: we transcribe the text as it is after correction, because it is in most cases impossible to distinguish the hands (see fig. 6).

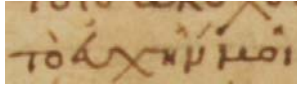


Fig. 6. N f. 60 r l. 10. It seems that τὸ εὐχὴν μοι was corrected to τὸ εὐχὴν μοι. We transcribed this string of text as τὸ εὐχὴν μοι.

4.2. Punctuation

Mediaeval copyists used a system of punctuation with dots, which, in theory, if we imagine letters written between two lines (up- and down-strokes notwithstanding), could be written on the lower line (lower dot), between the two lines (middle dot) or on the upper line (upper dot). This theoretical system is normally reduced to two positions of the dots, positions which are often difficult to distinguish with any certainty. In addition, two (:) or three dots (:.) are used to indicate very strong punctuation. The comma is not often found, and is probably a later development of the lower dot. The interrogation mark (;) is rarely used, and normally only if the interrogative clause is not introduced explicitly by an interrogative pronoun or adverb.¹⁹ See fig. 7 for an example in manuscript Q, where one semi-colon is found, as well as commas here and there. Moreover, the accentuation of the oxytone words is conditioned by the punctuation, as strong punctuation would normally require the oxytone word before it to be accented as an acute, not as a grave.

19 See Giannouli 2014.

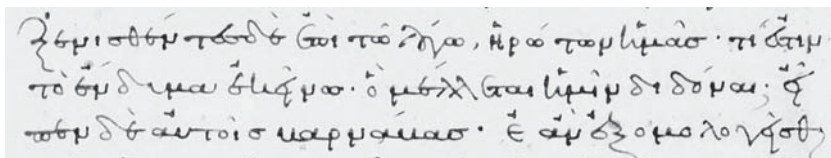


Fig. 7. Q f. 34r ll. 10-12: see the comma in l. 10, and a semi-colon after an interrogative clause introduced by an interrogative pronoun (τί) on l. 11; note that the relative pronoun ὁ is preceded by a middle dot, something that would be impossible in a standardised Greek punctuation.

The mediaeval system of punctuation found in manuscripts is adapted in modern editions and transformed into a system with four signs: comma for weak punctuation, middle dot for a stronger division, final (lower) dot at the end of a sentence, semicolon for any interrogative clause. In modern editions, the guiding principle for punctuation is the syntactical and semantic division of the text, whereas in mediaeval manuscripts the punctuation was meant to help in reading the text aloud.²⁰ If an adjustment of the mediaeval system of punctuation is therefore necessary in a critical edition (also because that system is variable from one manuscript to another and often inconsistent even within one and the same manuscript), in the transcriptions we tend to respect the punctuation of the manuscripts, with two qualifications: we have to use standard Unicode punctuation signs (reduced to the system of modern editions), and the difference of position of the dots in the manuscripts is often indiscernible, thus forcing us to use only the lower and the upper dots.

4.3. Abbreviations

We silently resolve all abbreviations, except the so-called *nomina sacra*, which are encoded as such in TEI-XML (<abbr type="ns">) and automatically expanded in a pre-processing step before the collation (see § 5.1). The *nomina sacra* are abbreviations by contraction of a limited number of common words, such as God, man, heaven, etc., and of proper names, such as Jesus, Christ, etc. (see figg. 8a-b). They are relatively standardised in Greek manuscripts and stable over a long period of time, some of them being already present in uncial manuscripts and in papyri.²¹ They can be accented or not, depending on the manuscripts, and we reproduce in our transcription the presence or absence of the accentuation.

20 See Parkes 1998; see also several contributions in Fasseur-Rochelois 2016.

21 See Traube 1907 and Paap 1959.

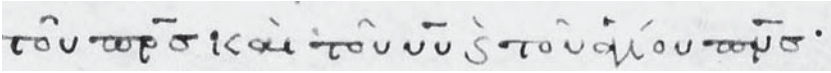


Fig. 8a. Q f. 34r l. 20: the *nomina sacra* are not accented.

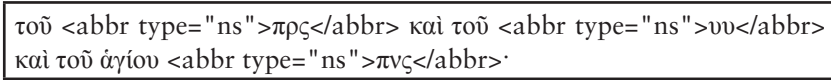


Fig. 8b. Transcription of the text in fig. 8a with XML-tags for the *nomina sacra*. πατρ stands for πατήρ ('father' in genitive), υἱ for υἱός ('son' in genitive), and πνε for πνεύματος ('spirit' in genitive).

Other forms of abbreviation in mediaeval Greek manuscripts include, on the one hand, the use of tachygraphic signs, which are numerous and vary in the course of time, but are normally not ambiguous (see figg. 9a-b), and, on the other hand, abbreviations by suspension, where the ending of the word is not expressed. Normally the decision about which ending to supply is made obvious by the context. When there is a doubt, in either type of abbreviations, we may decide to indicate with brackets that the presence of a letter is uncertain, as in the case of the abbreviation of ἐστὶ(ν) or ἐστι(ν) (see fig. 10),²² or that two concurring interpretations are possible, for example περίοδοι(οι|ος).²³ We did not encode this in TEI-XML, because we thought that the use of simple brackets could convey the same information in a less verbose manner than the TEI-XML encoding.

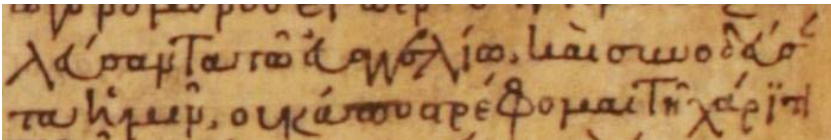


Fig. 9a. M f. 81vb ll. 20-21: συνοδεύσαν|τα, -αν- being expressed at the end of the line by a tachygraphic sign.

22 ἐστὶ(ν) or ἐστι(ν) is a form of the verb 'to be' in present indicative third person singular, with or without an epheletic or movable *nu* – as the verb εἶμι is an enclitic, it may or may not have an accent.

23 περίοδοι and περίοδος are respectively the nominative plural or singular of 'voyage', it is used in the title of the *Acts of Barnabas*, where there is no context to determine the choice between the two possible forms.

καὶ συνοδεύσαν<lb n="21" />τα ἡμῶν,

Fig. 9b. Transcription of the text in fig. 9a.

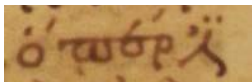


Fig. 10. N f. 60vb l. 24: ὅπερ ἐστὶ(v) - the sign at the end of the line stands for the verb ἐστὶ without an accent, but the sign leaves it open whether there should be a movable *nu* or not at the end of the word (it is followed by a word beginning with a vowel on the next line, and so normally requires an ephelcystic *nu*).

4.4. Corrections by the copyist or by later hands

Three types of correction are encoded: addition <add>, deletion , and replacement <subst>, the latter being considered a deletion followed by an addition. By adding attributes to these tags, we specify, when possible, which hand made the correction, how the deletion was made, and where the addition was placed (see figg. 11a-b). In order to be as precise as possible, we work at the letter level, and not at the word level.

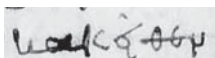


Fig. 11a. Q f. 33v l. 22: the scribe (presumably) scratched two letters and wrote -ακ- above, to form the word κακεῖθεν.²⁴

```
κ<subst>
<del unit="character" rend="scratched" quantity="2" />
<add hand="manus1" place="inline">ακ</add>
</subst>εἶθεν
```

Fig. 11b. Transcription of the text in fig. 11a.

5. Collation

The next step in our process is collating the manuscripts using CollateX. This software compares the transcriptions of each manuscript and produces as output a variant graph (see fig. 13 for an example).²⁵ After a

24 This might not be so easy to see on the microfilm, which is of low quality, but we ascertained this by checking the original manuscript.

25 See Dekker *et al.* 2015, as well as Tara Andrew's section on the collation process in Roelli *et al.* *forthc.*

pre-processing step that automatically fixes some possible issues with the transcriptions, the collation proper takes place in four steps: tokenisation, normalisation or neutralisation, alignment, and visualisation.

5.1. Pre-processing

The pre-processing consists of automatically modifying the TEI-XML file containing the transcriptions before entering them in CollateX.

The first type of pre-processing is the Unicode normalisation encoding, which transforms decomposed characters (for example an omega followed by the character “perispomene before”) into composed characters (omega with perispomene $\omegã$). This allows the comparison between transcriptions which could have been typeset in the computer using different input methods.

As said above (§ 4.3), the *nomina sacra* are automatically expanded using a predefined list of abbreviations and their corresponding expansions.

5.2. Tokenisation

The tokenisation process consists of splitting each TEI-XML file into a list of tokens, which will be automatically aligned by CollateX. A token is normally a single word, possibly followed by a punctuation mark, or accompanied by XML-tags, but the presence of XML-tags may induce a different tokenisation.

If we take the example of the correction made on two words by the copyist as shown in fig. 12a, where in the syntagm $\eta\ \delta\acute{o}\xi\alpha$ (‘the faith’), the article η and the first syllable of the substantive $\delta\acute{o}$ - are written in a space where a larger number of letters (probably five) were scratched out, then in this case the token will contain not just one word, but everything which is included in the <subst>-tag (see fig. 12b).

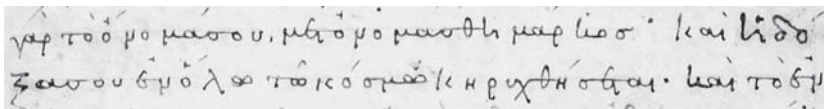


Fig. 12a. Q f. 32r ll. 10-11.

```
<subst>  
<del rend="scratched" unit="character" quantity="5" />  
<add hand="manus1" place="inline">η δό<lb n="11" />ξα</add></subst>
```

Fig. 12b. Transcription of the text in fig. 12a.

Generally, all the contents wrapped between an opening and a closing tag must be considered a single token. The only exceptions are the tags that indicate a title, a paragraph or a textual unit (<head>, <p> and <seg>; see § 3).

5.3. Normalisation / Neutralisation

In order to avoid having too much “noise” (formal differences that are stemmatically irrelevant) in the variant graph, the tokens compared by CollateX are normalised. The type of “normalisation” (in computer terms) that is performed at this stage is not a normalisation in linguistic terms, but rather a “neutralisation” of certain phenomena.

To explain what we mean by normalisation and neutralisation, it is necessary to provide some basic information about the character of mediaeval Greek manuscripts, which is not so different from the character of any other mediaeval tradition. The language of the texts copied by mediaeval Greek copyists was to some extent foreign to them. Between the time of composition of the *Acts of Barnabas* (fifth century in Cyprus) and the time from which the oldest extant manuscripts are preserved (end of the ninth century if we exclude the palimpsest fragment), four centuries elapsed. Moreover, the copies were made in different places of the Byzantine empire, by copyists whose command of the conventions of the written Greek literary language was sometimes low (even though the language of the *Acts* is of poor literary quality). Well-known phenomena in this respect are so-called *iotacism*, i.e. the reduction of the letters *upsilon* (υ) and *ēta* (η), and of most diphthongs containing an *iota* (ει, οι) to the sound [i], and the loss of any perception of the different length of the vowels ο and ὀ (o and ω).²⁶ These phonetic differences were very important in the morphological system, allowing to distinguish, for instance, the endings of the present participle in the neutral or masculine forms: -ov vs. -ωv, or the ending of the present indicative third person singular and the same form in the subjunctive: -ει vs. -η.

Although many of the changes induced by the loss of phonetic differentiation can be introduced or reversed arbitrarily by the copyists, and are therefore, for text-critical purposes, irrelevant, those variants can also

26 Many of those phonetic changes already started in the postclassical (Hellenistic) period; see Holton-Manolessou 2010 for a short presentation of the main linguistic features of mediaeval Greek.

often have a morphological significance and they might even sometimes be the cause of other variants. It is therefore not without danger to “neutralise” them too hastily. Besides, it is impossible to define simple rules for such a neutralisation, as the changes may occur in all directions. We could have added a level of normalised (in linguistic terms) transcriptions to our process of editing, but this would have been too time-consuming and we believe that the “noise” may be less problematic than running the risk of losing some relevant information.

Therefore, we decided to automatically neutralise only a few phenomena for which we could make up some simple rules. For example, two words having the same letters but not the same diacritics (spiritus, accents, iota subscripta – generally absent from the manuscripts – and apostrophes) can be considered, for the collation process, identical, even though for some forms the accentuation is grammatically not insignificant. For the comparison, these differences are neutralised. These neutralised forms of the tokens are compared in CollateX. However, the actual forms are also kept throughout the editing process.

The following phenomena are automatically neutralised:

- diacritics, punctuation and capitalisation are not taken into account;
- supplied forms occurring in unclear parts of the manuscript are compared;
- expanded forms of the *nomina sacra* are compared;
- numerical values of numbers (which were tagged) are compared.

As a working hypothesis, we consider that a correction made by the copyist’s hand was immediate (without resorting to any other source) and therefore does not have any text-critical value. Additions and substitutions made by “manus1” (without any doubt – so manus1-dubia would not be treated the same way) will also be neutralised, unless a special attribute “type” is added. In the case of deletions, the determination of the hand is almost always uncertain.

5.4. Milestones

Comparing many long transcriptions is computationally expensive. In order to facilitate this comparison, we do not compare the transcriptions in one block, but we subdivide them into smaller units. In our transcriptions, each unit is separated from the previous one by a <milestone/> tag

(see fig. 1). The division of the work into milestones is decided upon by the editor, in accordance with some logical structuring of the text. In the case of the Greek *Acts of Barnabas*, we followed the division into paragraphs made by the previous editor (Bonnet 1903).

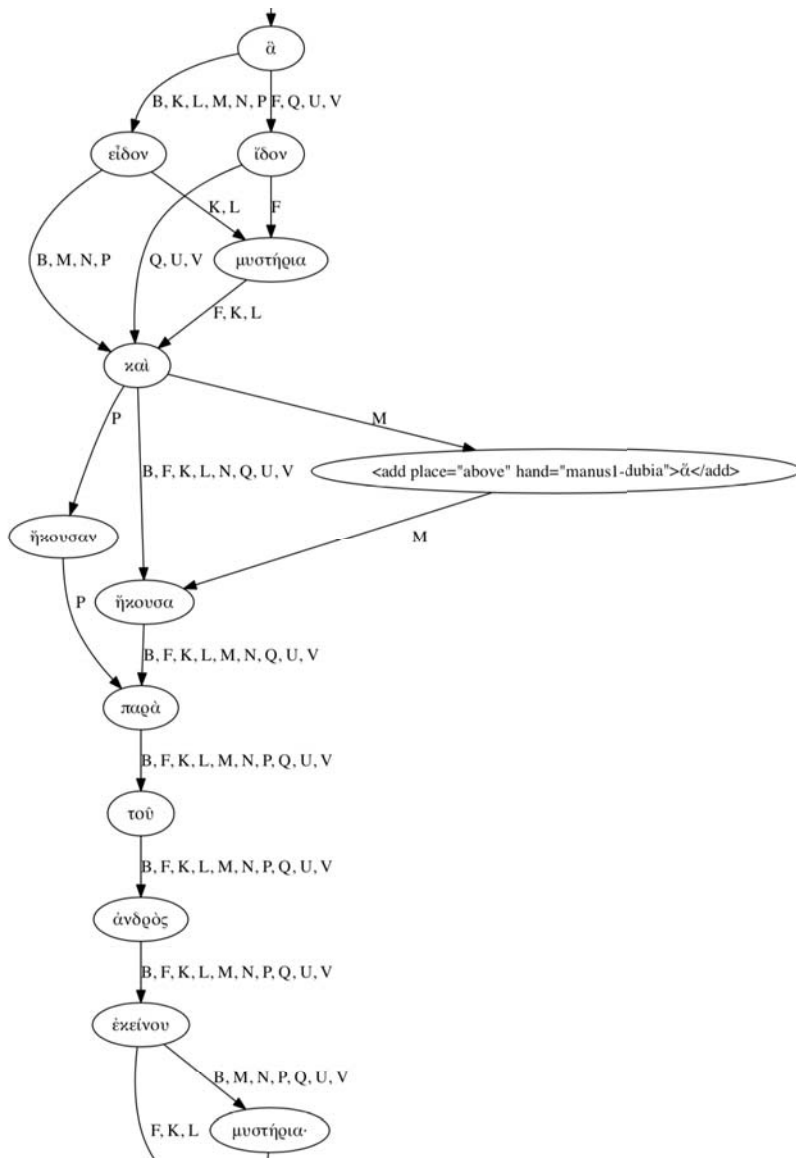


Fig. 13. Example of collation (milestone 4) using CollateX.

6. Conclusion and next steps

In this article, we have described the first two steps of the process of computer-assisted critical editing of works preserved in complex textual traditions, namely, the transcription and the collation. The transcription looks similar to what would be done in the case of a digital documentary edition, for which a methodological and theoretical framework is largely accepted.²⁷ One fundamental difference, however, is that our transcription method is just the first step of a complex process, which, in the digital world, has not yet been so well documented.²⁸ One major difficulty about our transcription method is that it must be designed to serve two different purposes: to be faithful to the manuscripts and to be usable for comparison and text-critical analysis. Each encoding decision must be carefully weighted and a balance must be found between comprehensiveness and relevance.

Typically, in a paper-based editorial project, decisions must also be taken as to the exhaustiveness of the collations, what to record and what to leave out from the wealth of information mediaeval manuscripts contain. These decisions are usually implicit and rarely documented. In our transcription, we deliberately left aside any palaeographical or codicological feature, but we recorded every textual element, without making a preconceived judgement about its text-critical value. Some of these elements may be ignored in the process of comparison, but none will be lost. On the contrary, the digital medium allows for far greater flexibility in terms of content, form and degree of detail displayed. The digital version of the edition makes it possible for a broader category of users to exploit the edition and manuscript data in ways that were previously not imaginable.

As we highlighted above, our goal is not only the digital visualisation, but also the computer-assisted creation of critical editions. We are currently extending the graph model of textual variation in order to provide a modern and convenient formalism for representing, storing and exploiting critical edition data in the form of a hypergraph, with the totality of the edition data being stored as a single graph database. This database will allow for a variety of exports for different purposes, such as online visualisation (in other words, generation of the dynamic form of the dig-

27 See e.g. Pierazzo 2011 and several contributions in Driscoll-Pierazzo 2016.

28 See Andrews 2013.

ital edition), PDF export (i.e., generation of the static form of the digital edition), and RDF export for semantic queries and preservation on a free platform, possibly the Knora platform, the Swiss national platform for storing research data in the humanities.

Bibliographical references

- Andrews 2013 = Tara L. Andrews, *The Third Way: Philology and Critical Edition in the Digital Age*, «Variants», 10 (2013), pp. 61-76.
- Andrews-Macé 2013 = Tara L. Andrews, Caroline Macé, *Beyond the Tree of Texts: Building an Empirical Model of Scribal Variation Through Graph Analysis of Texts and Stemmata*, «Literary and Linguistic Computing», 28, 4 (2013), pp. 504-521.
- Bonnet 1903 = *Acta Philippi et Acta Thomae, accedunt Acta Barnabae*, ed. Maximilianus Bonnet, Lipsiae, apud H. Mendelssohn, 1903.
- Chacon-Straub 2014 = Scott Chacon, Ben Straub, *Pro Git*, 2nd ed., [Berkeley-New York], Apress, 2014, online: <https://git-scm.com/book/en/v2>.
- Dekker *et al.* 2015 = Ronald Haentjens Dekker, Dirk van Hulle, Gregor Middell, Vincent Neyt, Joris van Zundert, *Computer-supported Collation of Modern Manuscripts: CollateX and the Beckett Digital Manuscript Project*, «Literary and Linguistic Computing», 30 (2015), pp. 452-470.
- Driscoll-Pierazzo 2016 = *Digital Scholarly Editing: Theories and Practices*, ed. by Matthew J. Driscoll, Elena Pierazzo, Cambridge, Open Book Publishers, 2016, online: <https://www.openbookpublishers.com/htmlreader/978-1-78374-238-7/main.html>.
- Fasseur-Rochelois 2016 = *Ponctuer l'oeuvre médiévale: des signes au sens*, éd. par Valérie Fasseur, Cécile Rochelois, Genève, Droz, 2016.
- Giannouli 2014 = Antonia Giannouli, *Byzantine Punctuation and Orthography. Between Normalisation and Respect of the Manuscripts. Introductory Remarks*, «Comparative Oriental Manuscript Studies Newsletter», 8 (July 2014), pp. 18-22.
- Holton-Manolessou 2010 = David Holton, Io Manolessou, *Medieval and Early Modern Greek*, in *A Companion to the Ancient Greek Language*, ed. by Egbert J. Bakker, Chichester, Blackwell, 2010, pp. 539-563.
- Macé 2016 = Caroline Macé, *Rules and Guidelines in Book Series and Their Impact on Scholarly Editions*, in *The Arts of Editing Medieval Greek and Latin: A Casebook*, ed. by Elisabet Göransson, Gunilla Iversen, Barbara Crostini, Toronto, PIMS, 2016, pp. 248-267.
- Macé-Andrews 2015 = Caroline Macé, Tara L. Andrews, *Collation*, in *Comparative Oriental Manuscript Studies. An Introduction*, ed. by Alessandro Bausi, Pier Giorgio Borbone, Françoise Briquel-Chatonnet, Paola Buzi, Jost Gippert, Caroline Macé, Marilena Maniaci, Zisis Melissakis, Laura E. Parodi, Witold Witakowski, Eugenia Sokolinski, Hamburg, Tredition, 2015, pp. 331-336.

- Macé-Gippert forthc. = Caroline Macé, Jost Gippert, *Textual Criticism and Editing in the Digital Age*, in *Oxford Handbook of Greek and Latin Textual Criticism*, ed. by Wolfgang de Melo, Scott Scullion, Oxford, O.U.P., forthcoming.
- Norelli 2001 = *Actes de Barnabé*, transl. by Enrico Norelli, in *Écrits apocryphes chrétiens*, II, sous la direction de Pierre Geoltrain et Jean-Daniel Kaestli, Paris, Gallimard, 2005, pp. 618-642.
- Noret 1994 = *Hagiographica Cypria. Vita Sancti Auxibii*, ed. Jacques Noret, Turnhout, Brepols, 1994.
- Noret 2007 = Jacques Noret, *Les règles byzantines de la division en syllabes*, «Byzantion», 77 (2007), pp. 345-348.
- Nury 2018 = Elisa Nury, *Automated Collation and Digital Editions: From Theory to Practice*, PhD thesis, London, King's College London, 2018.
- Paap 1959 = Anton Herman Reiner Everhard Paap, *Nomina Sacra in the Greek Papyri of the First Five Centuries*, Leiden, Brill, 1959.
- Parkes 1998 = Malcolm B. Parkes, *Medieval Punctuation and the Modern Editor*, in *Filologia classica e Filologia romanza: esperienze ecdotiche a confronto. Atti del convegno, Roma 25-27 maggio 1995*, a c. di Anna Ferrari, Spoleto, Centro Italiano di Studi sull'Alto Medioevo, 1998, pp. 351-365.
- Pierazzo 2011 = Elena Pierazzo, *A Rationale of Digital Documentary Editions*, «Literary and Linguistic Computing», 26, 4 (2011), pp. 463-477.
- Robinson 1989 = Peter Robinson, *The Collation and Textual Criticism of Icelandic Manuscripts (1): Collation*, «Literary and Linguistic Computing», 4 (1989), pp. 99-105.
- Roelli et al. forthc. = *Introduction to Stemmatology in Digital Age*, ed. by Philippe Roelli, Tara L. Andrews, Marina Buzzoni, Aidan Conti, Elisabet Göransson, Odd Einar Haugen, Armin Hoenen, Caroline Macé, Joris van Zundert, Berlin, De Gruyter, forthcoming.
- Rouquette 2017 = Maïeul Rouquette, *Étude comparée sur la construction des origines apostoliques des Églises de Crète et de Chypre à travers les figures de Tite et de Barnabé*, PhD thesis, Lausanne-Aix-en-Provence, Universités de Lausanne et d'Aix-Marseille, 2017.
- Rouquette-Seretan forthc. = Maïeul Rouquette, Violeta Seretan, *Using Hypergraphs for Storing Critical Edition Data*, in *Proceedings of the Workshop on Scholarly Digital Editions, Graph Data-Models and Semantic Web Technologies*, forthcoming.
- Snyder 2016 = Glenn E. Snyder, *The Acts of Barnabas. A New Translation and Introduction*, in *New Testament Apocrypha*, vol. 1, ed. by Tony Burke, Brent Landau, Grand Rapids, MI, 2016, pp. 317-336.
- Traube 1907 = Ludwig Traube, *Nomina sacra. Versuch einer Geschichte der christlichen Kürzung*, München, Beck, 1907.

