

# Encoding Strategies and the Ethiopic Literary Heritage: The *Physiologus* as a Case Study

Massimo Villa, Universität Hamburg

Producing the Clavis of Ethiopic literature while making the data computer readable for extensive indexing and research purposes is among the aspirations of the project *Beta maṣāḥəft*. Here, we illustrate the challenges faced with and the solutions offered by the project on the example of the *Physiologus*, a literary work, translated into Ethiopic from Greek during the Late Antiquity.

*Beta maṣāḥəft* is a manuscript-centred research environment, and in this initial phase is predominantly focused on codicology. This notwithstanding, one of the project's main spheres of interest is obviously the Ethiopic textual heritage.<sup>1</sup>

Working with texts in a TEI/XML hierarchical structure one faces manifold key issues. A basic terminology-related question which has hitherto received very little scholarly attention in the field of Ethiopian studies is: 'what is a work?'<sup>2</sup> The solution proposed by the team is that, for the purposes of the project, we consider a work any text with an independent circulation.<sup>3</sup> According to this principle we create a new XML-based file for each work. Any work file is associated with a univocal fist-level ID, which is a string consisting of three elements: a fixed sequence LIT that identifies the type of entity, a progressive numerical sequence which is arbitrarily assigned, and an alphabetical part (generally a six-letter sequence) added to help recognize the work. The work ID of the *Wəddāse Māryām* or 'Praise of Mary' is, for instance, LIT2509Weddas. On the contrary, texts with no independent circulation are not considered as works. Being systematically found as sub-units of other works. They can be given their own structure and IDs (e.g. LIT2509Weddas#Monday), and be referred to by pointing to the `@xml:id` anchor inside the file. Any textpart is 'upgraded' to a work as soon as it is found independently.

- 1 *Beta Maṣāḥəft: Die Schriftkultur des christlichen Äthiopiens und Eritreas: Eine multimediale Forschungsumgebung* is a long-term project headed by Prof. Alessandro Bausi, funded within the Academies' Programme, and coordinated by the Union of the German Academies of Sciences and Humanities, under survey of the Akademie der Wissenschaften in Hamburg (2016–2040). The structures and the strategies of encoding illustrated in the present paper are the result of joint efforts by the entire project team (see <https://www.betamasahft.uni-hamburg.de/en/team/projectteam.html>). See also the project note by Dorothea Reule in this issue.
- 2 See also the contribution by Tito Orlandi in this issue.
- 3 See project guidelines at <http://betamasahft.eu/Guidelines/?id=definitionWorks>.

To attribute a given text its own appropriate status, i.e. to state whether it is a work or a textpart, has consequences at various levels. First, it encourages researchers to monitor the ways of circulation of the text and to find out systematic correlations among other texts. Secondly, it is interlinked with the creation of a full repertory of Ethiopic texts, i.e. a Clavis, which is one of the declared goals of the project. Any Clavis requires a numbering; the *Clavis Aethiopica* (CAe) numbering stems from the numerical sequence of the record ID (see above). Consequently, fulfilling the relevant requirements, particularly that of independent circulation, is crucial to the choice to include the text into the Clavis.

The creation of work records is currently underway for the following types of textual items: i. texts with an independent circulation; ii. compilations (multiple-text *corpora*) and parts of compilations which can circulate independently at various stages of granularity;<sup>4</sup> iii. each recension of a multiple-recension text plus an extra record for the general abstract work. The record for the general work does not represent the *Urtext* or the genetic ancestor of the textual tradition, yet an ‘architext’ in which the distinctive features of the individual recensions are ideally neutralized, and to which witnesses not properly specified in the existing catalogues are assigned. The relations between the general work and the individual recensions are expressed by means of relation elements (see below).

A case study which proves to be particularly representative of the challenges met by the project team in a multiple-recension tradition is the textual tradition of the Ethiopic *Physiologus*.

4 These include e.g. the hagiographical collection *Gadla Samā' tāt* (‘Acts of Martyrs’) and the single Acts which are traditionally transmitted within it; the cycle of Marian miracles *Ta' amməra Māryām* and the individual miracles; a stable hagiographical dossier of a saint and its parts (*Life, Miracles, History* of the translation of his body, etc.). On the *Gadla Samā' tāt* as a ‘corpus organizer’ see Bausi 2010. Manuscript evidence shows that multiple-text corpora such as the *Acts of the Martyrs* and the *Miracles of Mary* are considered as single units in the local literary and scribal culture. Nonetheless their textual components (the single acts or miracles) can be missing, arranged in a different order, or transmitted separately outside the main corpus. The project records such a documented fluidity in the transmission by treating both the corpus and the single components as independent works. The historical process which led to the development of some multiple-text corpora still lies beyond our knowledge: for instance, a collection such as the *Acts of the Martyrs* appears as one single compilation, with its own traditional label and independent circulation, since the earliest available documentation (thirteenth to fourteenth century). Other compilations, or at least their earliest core, were received as translated works, such as the *Miracles of Mary* and the *Sənkəssār* (‘Synaxarion’), and further enriched with additional original pieces.

**Case study**

The *Physiologus* is a popular literary work, composed in Greek in Alexandria and translated into several Christian oriental languages. The Ethiopic version was produced during the Aksumite Age (fifth–sixth centuries). It counts a variable number of chapters, each providing a legendary description of the natural properties of a species of animal, plant, or mineral, and the explanation of these properties in a moralizing Christian context.

The Ethiopic tradition is somewhat multifaceted (fig. 1). It consists of three recensions: a ‘Homily of the blessed Physiologus’ (*Dərsān za-bəṣu’ Fisālgos*, in short *Phys. α*), a ‘History of the similitudes of the wise Physiologus’ (*Zenā məssāleyāt za-ṭabib Fisə’algos*, in short *Phys. β*), and a ‘History of the wisdom of the wise Physiologus’ (*Zenā ṭəbab za-Fisə’algos ṭabib*, in short *Phys. γ*).<sup>5</sup> Such a scenario entails the creation of four records, one for the general work (LIT1401Physio), and one for each recension: LIT4915PhysA for *Phys. α* (CAe 4915), LIT4916PhysB for *Phys. β* (CAe 4916), and LIT4917PhysC for *Phys. γ* (CAe 4917). The three recensions are disambiguated by the numerical string of their IDs, which is assigned arbitrarily. A printed edition is also extant: it was published by Fritz Hommel in 1877 on the basis of three manuscripts belonging to *Phys. α*. The texts of *Phys. β* and *Phys. γ* are still unpublished.

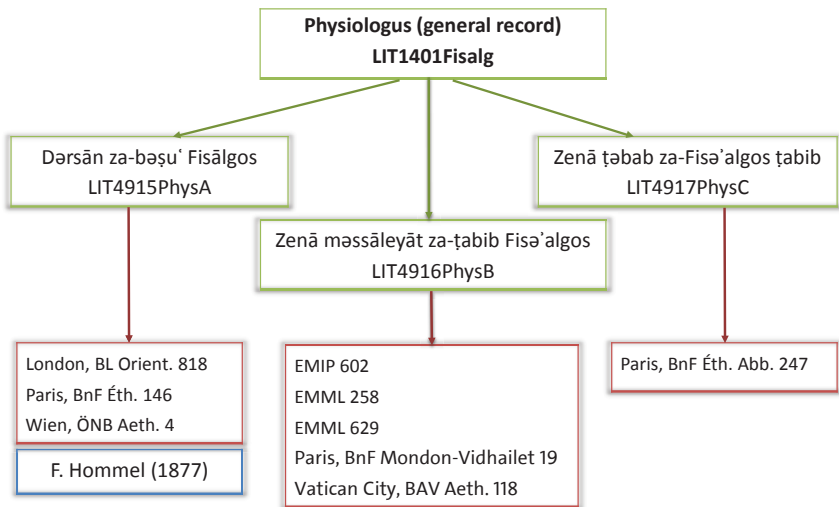


Fig. 1. Diagram illustrating the textual tradition of the Ethiopic *Physiologus*.

5 A detailed picture of the manuscript and textual tradition of the Ethiopic *Physiologus* will be presented in forthcoming publications by the present writer.

Work records are structured through a variety of metadata, in which the relevant pieces of information are properly entered. Some of these components are illustrated hereby.

**Title(s).** The main title of the work is given by a threefold set of elements, which contains different ways of identifying the work: in Ethiopic script (e.g. ‘ድርሳን ፡ ዘብጹዕ ፡ ፊሳልጎስ ፡’ for LIT4915PhysA), in transliteration (‘Dərsān za-bəṣu ‘Fisālgos’), and in English translation (‘Homily of the blessed Physiologus’). This basic set can be supplemented with additional tags for alternative variant titles, short titles (e.g. ‘Phys. α’), and parallel titles in other traditions.

**Text witnesses.** The list of manuscripts which are actually used in the text edition is explicitly given, while a *recensio* of all encoded witnesses is visible online since the app can retrieve and display all manuscripts whose contents point to the ID of the work in question.<sup>6</sup> The manuscript IDs can be further associated with an `@xml:id` that shortly identifies the witness used in the text edition, and can be identical to the siglum used in the printed edition. For LIT4915PhysA, they are L (= London, BL Orient 818), P (= Paris, BnF Éth. 146), and W (= Wien, ÖNB Aeth. 4).

**Claves.** In addition to the generated CAe number, existing repertoires of claves have also been incorporated.<sup>7</sup> Individual works can be therefore searched for their number in one of the Claves through an appropriate filtered search.

**Keywords.** Keywords help categorize the literary corpus under different typologies. They specify the genre (Bible, Liturgy, Hagiography, etc.) or the age (Aksumite, Gondarine, etc.) of a given work. Keywords allow users to filter their search and automatically extrapolate all works belonging to a certain age or genre. An *ad-hoc* query on the Aksumite texts, for instance, generates a constantly updated repertory of the earlier textual heritage of Ethiopic literature.

**Relations.** Relation elements contain our formal description of the relationships between different entities, in the present case abstract works.<sup>8</sup> Each relation expresses a statement with a subject, a predicate, and an object (technically, a ‘triple’). Relations can or cannot involve items belonging to the same type of entity: in our case we use them to define a relationship between two abstract

6 See guidelines at <http://betamasaheft.eu/Guidelines/?id=work-teiHeader>.

7 Among these, the *Clavis Apocryphorum Veteris Testamenti* (CAVT), the *Clavis Apocryphorum Novi Testamenti* (CANT), the *Bibliotheca Hagiographica Orientalis* (BHO), the *Bibliotheca Hagiographica Graeca* (BHG), the Kinefe-Rigb Zelleke’s repertory of hagiographical texts (KRZ), and others.

8 See <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-relation.html>.

works. For multiple-recension texts a relation named `saws:isVersionOf` is used.<sup>9</sup> Therefore, a statement like

```
<relation active="LIT4915PhysA"
  name="saws:isVersionOf" passive="LIT1401Physio"/>
```

is a formal way to claim that *Phys. α* is a version of the *Physiologus* as abstract notion of the work. Other relationships among works can also be stated: for multiple-text corpora a relation `saws:formsPartOf` is employed, and for parallel versions in other languages `saws:isVersionInAnotherLanguageOf`.<sup>10</sup> A diagram generated in the website displays all the relations entertained by a given record.

**Text editions.** Being a multiple-recension work, the *Physiologus* is not represented by a unitary text. Each recension is supplied with its own edition or, if existing, more than one. Since any TEI/XML work record points to the abstract notion of a given work, it can host as many text editions as needed. *Phys. α* hosts two editions, the *editio princeps* by Fritz Hommel and a new improved edition accommodating different text-critical choices. Each of them is identified by its own local `@xml:id`. A further challenge has been addressed by the modularity of the work. The *Physiologus* is subdivided into chapters, each obviously identified by a univocal value (e.g. ‘ed2ch3’). However, this reference system, focused on the chapter arrangement and not content-oriented, does not take into consideration that the chapter numbering is variable in the three recensions (*Phys. α* counts 49 chapters, *Phys. β* 57, and *Phys. γ* 58), and a certain subject is attributed different chapter numbers in the distinct recensions. As an example, the chapter on the caladrius (ገራድዮን ፣ *karādyon*, a legendary bird with diagnostic and medical powers) is numbered as 3 in *Phys. α* and as 5 in the remaining recensions. Consequently, a narrative unit with its own ID ‘NAR0014caladrius’ is needed to specify one and the same subject regardless of its numbering.<sup>11</sup>

Once transcribed, the ‘plain text’ can be annotated with tags providing paratextual, text-critical and content-related information. Various scribal phe-

9 Property names are inherited from the ontology developed by the *Sharing Ancient Wisdoms* project; see <<http://www.ancientwisdoms.ac.uk/>>. See also Tupman and Jordanous 2014.

10 Many other relations are used in the *Beta maṣāḥaft* data. They are beyond the scope of this contribution and can be consulted at <<http://betamasaheft.eu/Guidelines/?id=relation>>.

11 The notion of ‘narrative unit’ is indebted to, and further develops, the terminology proposed by Tito Orlandi (2013, 93) for Coptic literary documentation. It is adopted in the project to refer to text portions (paragraphs, chapters, miracles performed by a saint, stanzas, etc.) sharing the same narrative content, no matter if extant in different versions, in multiple recensions of the same work, or even in different works; see <<http://betamasaheft.eu/Guidelines/?id=narrativeUnits>>.

nomena can be recorded, e.g. erasures, omissions, and marginal or interlineal additions. The critical apparatus makes use of the `<xm1:ids>` previously associated with the witnesses. Finally, quotations from the Scriptures and other texts can be marked-up and referenced to the precise verse or line of the source text.

The digital edition is visualized on the website in a three-column structure: the left column outlines the paragraph-based arrangement of the text, the central column displays the text, the right column the critical apparatus. Quotations are clickable and open a pop-up window displaying the source text. Once data from the multiple versions have been entered and properly marked-up, it is desirable to have them displayed together for later comparison. This can be done by enabling a specific function which detects all the existing parallel versions, extrapolates from each the exact portion of text identified by the same narrative unit, and visualizes all relevant portions together (fig. 2). It is important to highlight that the functioning of such a tool arises from a precise encoding strategy, which combines the presence of a relation `saws:isVersionOf` and the reference to a narrative unit. Parallel-version out-

The screenshot displays a digital edition interface for a chapter of an Ethiopian text. It is organized into three main columns:

- Left Column:** Contains edition metadata including the location (London, Paris, Vienna), library names (British Library, Bibliothèque nationale de France, Österreichische Nationalbibliothek), and a 'Compare' button.
- Middle Column:** Displays the text in Amharic script, organized into four sentences. Each sentence is preceded by a green 'Ugari Alignment' button and a 'Quotations' button. The text includes various diacritical marks and some interlineal additions.
- Right Column:** Features a 'Editions Bibliography' section with a list of references, including Hommel (1877), Conti Rossini (1951), Hommel (1890), and Weninger (2005).

Below the main text area, there are two highlighted boxes providing additional context:

- Version Zenā ṭəḅab za-Fis'alǧos ṭəḅib (LIT4917PhysC):** Edition: Paris, Bibliothèque nationale de France, BnF Éthiopian d'Abbadie 247. This version includes interlineal additions (ገፍታ) and corrections (ገፍታ) to the original text.
- Version Zenā massāllēyāt za-ṭəḅib Fis'alǧos (LIT4916PhysB):** Edition: Massimo Villa. This version includes interlineal additions (ገፍታ) and corrections (ገፍታ) to the original text.

Fig. 2. Parallel versions of chapter 4 of the Ethiopian *Physiologus* as visualized on the <<http://betamasaheft.eu/>> portal.

puts are particularly desirable to keep track of the discrepancies and the similarities in multiple-recension works (e.g. some *Lives* of most venerated saints) or in works preserved in different oriental traditions.

Finally, text is also linked to the online version of Dillmann's *Lexicon linguae Aethiopicae*. Any individual word is clickable and redirects to the corresponding entry in the *Lexicon*, further complemented by a list of all occurrences of the same lemma in the *Beta maṣāḥaft* corpus.<sup>12</sup> Many opportunities of investigation arise in this way. Search potentialities can be enhanced maximizing the benefits of the data pool at disposal, and reducing the time needed to put them into correlation.

### References

- Bausi, A. 2010. 'A Case for Multiple Text Manuscripts being 'Corpus-Organizers'', *Manuscript Cultures Newsletter*, 2 (2010), 34–36.
- 2016. 'Introduction. 150 Years After Dillmann's Lexicon', in A. Bausi and E. Sokolinski, eds, *150 Years after Dillmann's Lexicon: Perspectives and Challenges of Gə'əz Studies*, Supplement to *Aethiopia*, 5 (Wiesbaden: Harrassowitz Verlag, 2016), 3–10.
- Dillmann, C.F.A. 1865. *Lexicon linguae Aethiopicae, cum indice Latino. Adiectum est vocabularium tigre dialecti septentrionalis compilatum a W. Munziger* (Lipsiae: T.O. Weigel, 1865).
- Hommel, F. 1877. *Die aethiopische Uebersetzung des Physiologus nach je einer Londoner, Pariser und Wiener Handschrift* (Leipzig: J. C. Hinrichs'sche Buchhandlung, 1877).
- Orlandi, T. 2013. 'A Terminology for the Identification of Coptic Literary Documents', *Journal of Coptic Studies*, 13 (2013), 87–94.
- Tupman, C. and A. Jordanous 2014. 'Sharing Ancient Wisdoms across the Semantic Web Using TEI and Ontologies', in T. L. Andrews and C. Macé, eds, *Analysis of Ancient and Medieval Texts and Manuscripts: Digital Approaches*, *Lectio: Studies in the Transmission of Texts & Ideas*, 1 (Belgium: Brepols, 2014), 213–228.

12 See Bausi 2016. The tool is developed by the *TraCES* project and is accessible at <<http://betamasahft.eu/Dillmann/>>.

