



**HAL**  
open science

## Editing New Testament Arabic Manuscripts in a TEI-base: fostering close reading in Digital Humanities

Claire Clivaz, Sara Schulthess, Martial Sankar

### ► To cite this version:

Claire Clivaz, Sara Schulthess, Martial Sankar. Editing New Testament Arabic Manuscripts in a TEI-base: fostering close reading in Digital Humanities. *Journal of Data Mining and Digital Humanities*, 2017, Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages, 10.46298/jdmdh.1390 . hal-01280627v4

**HAL Id: hal-01280627**

**<https://hal.science/hal-01280627v4>**

Submitted on 3 Jun 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

## Editing New Testament Arabic Manuscripts in a TEI-base: fostering close reading in Digital Humanities

Claire Clivaz\*, Sara Schulthess, Martial Sankar

Swiss Institute of Bioinformatics, VITAL-IT, Switzerland

\*Corresponding author: [claire.clivaz@sib.swiss](mailto:claire.clivaz@sib.swiss)

### Abstract

If one is convinced that “quantitative research provides data not interpretation” [Moretti, 2005, 9], close reading should thus be considered as not only the necessary bridge between big data and interpretation but also the core duty of the Humanities. To test its potential in a neglected field – the Arabic manuscripts of the Letters of Paul of Tarsus – an enhanced, digital edition has been in development as a progression of a Swiss National Fund project. This short paper presents the development of this edition and perspectives regarding a second project. Based on the *Edition Visualization Technology* tool, the digital edition provides a transcription of the Arabic text, a standardized and vocalized version, as well as French translation with all texts encoded in TEI XML. Thanks to another Swiss National Foundation subsidy, a new research project on the unique New Testament, trilingual (Greek-Latin-Arabic) manuscript, the Marciana Library Gr. Z. 11 (379), 12th century, is currently underway. This project includes new features such as “Textlink”, “Hotspot” and notes: *HumaReC*.

### Keywords

Digital edition, Arabic TEI, New Testament, Letters, Paul of Tarsus

### INTRODUCTION

The Digital Humanities are often understood as a movement towards big data studies. If one is convinced that “quantitative research provides data not interpretation” [Moretti, 2005, 9], close reading should be considered as not only the necessary bridge between big data and interpretation but also the core duty of the Humanities. To test its potential on a neglected field – the Arabic manuscripts of the Letters of Paul of Tarsus – an enhanced, digital edition has been in development in the framework of a Swiss National Fund 2013-2016 [http1]. This short paper presents the development of this digital enhanced edition based on the EVT tool [http2], a collaborative work that will be further improved in the next research project *HumaReC* [http3].

### I HUMANITIES PROJECT BECOMES DIGITAL

#### 1.1 The case of the Arabic manuscripts of New Testament

Our interest in the Arabic New Testament (NT) manuscripts has largely been piqued by two observations. Firstly, the very low number of publications on the topic is intriguing; the

reference book still used today goes back to 1944 [Graf, 1944]. In fact, despite Western scholars starting to study the Arabic versions very early on [Raimundi, 1590], since the end of the 19th century, biblical textual criticism increasingly lost interest in this tradition [Vollandt, 2013]; only a few orientalist continued to study the field [Kashouh, 2012, 9-33]. Secondly, investigating the digital evolution of New Testament Textual Criticism (NTTC) on the Internet has brought to light Islamic websites that are studying Greek NT manuscripts [Clivaz, 2013; Schulthess, 2013]. These websites provide reference material from the NTTC discipline and use images of manuscripts, which are increasingly accessible online today (e.g. Codex Sinaiticus project [http4], the database of the CSNTM [http5], the NT Virtual Manuscript Room [http6]); this serves to underline the variants and scribal interventions [http7]. This phenomenon is a reinitiation of a long-standing polemical tradition: the concept of the falsification of Scriptures (*at-tahrīf*), now using digital media [Schulthess, 2016a]. In contrast to this, several Evangelical Christian websites attempt to answer the accusations of *tahrīf*, and one can observe interactions between apologists from different communities. Interestingly, Arabic NT manuscripts are also integrated into these discussions. Since the initial project preparations, a renewed global interest has been observable for Arabic NT manuscripts, including numerous new publications and projects [Kashouh, 2012; Arbache, 2012; Moawad, 2014; http8]. Furthermore, certain identity issues are present in this contemporaneous research, for example contemporaneous scholars defending the existence of pre-Islamic translations, a minority point of view (Kashouh, 2012; see 2.1).

## 1.2 The project becomes digital

Despite the starting point of our interest having been digital, we started with a classical Humanities project, and according to the rules still in use in the Faculties in Humanities, the PhD resulting from this project was presented in September 2016 in the form of a book [Schulthess, 2016b], also available online in open access [http9]. We have chosen to study the Arabic manuscripts of the Pauline letters, especially the *Vaticanus Arabicus* 13, a manuscript playing an important role in the controversial question of the existence of pre-Islamic biblical translations into Arabic (see 2.1). With regards to the context previously described, it is clearly important that research be accessible online with defined methodologies, this should encourage broader discussion. The next logical step for this project was then to initiate a digital, scholarly, enhanced edition of the First Letter to the Corinthians in the Vat. Ar. 13.

Today, the acquisition of encoding skills should be one of the duties of a Humanist PhD student working on texts. However at the moment this is not included in Swiss PhD curricula. Therefore, it was a great opportunity to be welcomed at the Swiss Institute of Bioinformatics (VITAL-IT group), where the development of tools for the digital edition has begun. The open source visualization tool *Edition Visualisation Tool* (EVT) [http10] was the base of our work. EVT is software developed for the *Vercelli Book Digitale* project [http11] which deals with XML-encoded text. Most notably, it allows one to edit several layers of the text (image, transcription, standardized text, translation [http2]). Considering all the questions of access and open science [http12], it is all the more necessary to produce an open edition of this manuscript, especially since this neglected field is often misused as a weapon in identity battles. We are developing these issues in the project *HumaReC* (2016-2018), that deals with the only known trilingual (Greek-Latin-Arabic) NT manuscript [http3].

### 1.3 A Digital edition for Early Middle Arabic

Digital editions provide challenging features when it comes to editing Arabic manuscripts. Among digital editions, different projects propose several versions of the same text, a process revealed as “layers” or “levels”. The *Queste del Saint Graal* project [http13] applies this method to Medieval French, while the *Vercelli Book Digitale* project [http11] applies it to Old English, The *Coptic Scriptorium* [http14] does the same to Coptic, and for Arabic there is the *Arabic Papyrology Database* (APD [http15]). As demonstrated by the APD, the “edition by layers” method is particularly relevant for “non-standardized” Arabic, as in the case of so-called “Early Middle Arabic”, used notably by Arabic-speaking Christians in the First Millennium. The text of Vat. Ar. 13 corresponds to the features described as Early Middle Arabic by Joshua Blau [Blau, 1966]. This is a helpful linguistic category despite criticism of its historical development [Kouloughli, 2007]. A digital edition allows for the editing of the text as found in the document, and secondly, the editing of the text with linguistic adaptations: missing diacritical points, *tashkīl* (with or without vocalization), and so forth. In our case, the aim was to edit the First Letter to the Corinthians in the Vat. Ar. 13, one of the oldest Arabic NT codices [Schulthess, 2016b]. Goals for the digital edition were to provide the images of the manuscript, two layers of the text, an accurate diplomatic transcription (respecting graphic peculiarities for example), a standardized and vocalized version, as well as a translation.

## II DIGITAL ENHANCED EDITION OF VAT. AR. 13

### 2.1 The *Vaticanus Arabicus 13*

It can be said with great certainty that the manuscript, held today in the Vatican Library, comes from the Arabicized Greek Orthodox community in the region of Syria-Palestine. It is undated but certain parts of the manuscript seem to date back to the 9th century. Due to its lengthy presence, the manuscript has often been the object of study but it was never edited. It contains parts of Matthew, Mark and Luke and the Pauline letters (with Hebrews at the end). Thanks to the colophon, we know that it originally contained the Psalms, the four Gospels, Acts, the Catholic letters and Pauline letters. This is very unusual for the early date, Graf described it as the possibly the oldest manuscript collection of the NT in Arabic [Graf, 1944, 138]. In his thesis, Kashouh argues that the text of the oldest part of the Gospels goes back to the sixth or early seventh century [Kashouh, 2012, 142-170], an opinion that has been questioned by several researchers [Griffith, 2013, 116-117; Monferrer-Sala, 2015a]. Since research has focused on the Gospel folios (with the exception of Monferrer-Sala [Monferrer-Sala, 2015b]), the project began with a study of the Pauline corpus and focused on the First Letter to the Corinthians [Schulthess, 2016b].

### 2.2 The Edition Visualization Tool

As mentioned previously, the NT project began in a pre-digitized form. The decision was made to provide a transcription of the text already in the printed book (1) as well as a standardized text with vocalization (2). The reader can thus completely follow the interpretations of the translation and has all the information needed to understand the commentary. As a result, one may consider the presence of two versions of the same Arabic text, because the transcription and the standardized text, through the vocalization, constantly present differences. An example:

- (1) بولوس المدعى رسول يسوع المسيح بمسره الله وسناتين الاخ
- (2) بُولُوسُ الْمُدْعَى رَسُولَ يَسُوعَ الْمَسِيحِ بِمَسْرَةَ اللَّهِ وَسُنَّتَانِيْنِ الْآخِ

Thus, the digital edition was already conceptualized as a text with several “layers”: a transcription (1), a standardized and vocalized version (2), as well as a French translation. The *Vercelli Book Digitale* project’s EVT tool provided an interesting starting point. EVT and the *Vercelli Book Digitale* project, much like the *Queste del Saint Graal* project and others, is open source and uses TEI XML for encoding the data, two aspects our team is in support of. Until now there have been very few editions of Arabic manuscripts available that follow the TEI standards [http16; http17; see also in XML: http18]. The goal of the EVT team was to provide an open source software tool to deal with TEI XML-encoded text [Rosselli del Turco et al., 2014-2015]. The EVT software was built on open and modern web technologies such as HTML/CSS3 and JQuery framework. The beta version of the tool was released at the end of 2014 and uploaded under public license; two versions of the software are now available (EVT 1 and EVT 2) [http19]. The project team used EVT 1, the version available at the time the project started, this version is very stable for diplomatic editions. The main benefit of EVT is its ability to automatically render TEI XML encoded text into a modern user-friendly front-end. The project team took advantage of this key feature to wrap together the Arabic text (transcription (1) and normalized text (2)) and the translation of the images of our manuscripts. To make this possible, a few modifications were made to the XSL style sheet template files, mainly for html-rendering of TEI-tags or attributes. Using the manuscript as a starting point, the team used embedded transcription [http20], made possible by the development of XSLT-based TEI XML formatting in EVT [Rosselli del Turco et al., 2014-2015]. Since there were two different versions of the text and not only character or word differences, the element <choice> and the sub-element <orig> for (1) and <reg> for (2) were used for each line. This allows the reader to easily compare (1) and (2) (see Figure 1 below). Furthermore, the biblical verses were marked with the <milestone> element, a “neutral” way of marking something not present in the manuscript. Other elements such as <unclear> and <add> were used in cases of reading difficulty and scribal intervention.

### III FURTHER PERSPECTIVES AND CONCLUSION

This project makes available an edition of the complete First Letter to the Corinthians found in 33 folios of Vat. Ar. 13; the edition is available in open access and offers high quality images, the XML files are also available on github. The EVT tool proved its worth throughout this experience and it will be further utilized in the continuation of the *HumaReC* project. The project aims to test the continuous data and research publication model of a digital, enhanced edition of the unique trilingual NT manuscript (Greek-Latin-Arabic) (12th century, Marciana Library Gr. Z. 11 (379)). Features of the Vat. Ar. 13 edition were adapted to the trilingual manuscript. To offer the best reading experience, the project will use the “Textlink” feature provided by the EVT; this offers the possibility of linking zones in the manuscripts to the corresponding text and highlighting them. Furthermore, commentaries will be added to the edition including “Hotspot” features displayed on the image or notes displayed in the text. We are also currently working on the implementation of an annotation tool [http21]. Finally, we will compare for the Pauline letters the manual transcription and the transcription resulted from the Handwritten Text Recognition tool *Transkribus* [http22], a service funded by the H2020 Project READ [http23].

## IV TABLES AND FIGURES

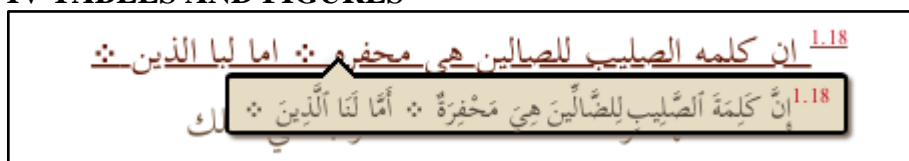


Figure 1. Screenshot of [http2]. Transcription and standardized text.

## V REFERENCES

- Arbache S. *L'Evangile arabe selon Saint Luc: Texte du VIIIe Siècle, copié en 897*. Safran (Bruxelles), 2012.
- Blau, J. *A Grammar of Christian Arabic, Based Mainly on South-Palestinian Texts from the First Millennium* (CSCO 267). Secrétariat du Corpus SCO (Louvain), 1966.
- Clivaz C. Internet Networks and Academic Research: the Example of the New Testament Textual Criticism. *Digital Humanities in Biblical, Early Jewish and Early Christian Studies* (Scholarly Communication 2). Ed. by C. Clivaz, A. Gregory and D. Hamidović. Brill (Leiden), 2013:151-173.
- Graf G. *Geschichte der christlichen arabischen Literatur, Vol. 1*. Biblioteca Apostolica Vaticana (Rome), 1944.
- Griffith S.H. *The Bible in Arabic. The Scriptures of the "People of the Book" in the Language of Islam*, Princeton University Press (Princeton), 2013.
- http1 <http://unil.ch/nt-arabe>
- http2 <http://tarsian.vital-it.ch>
- http3 <https://humarec.org>
- http4 <http://www.codexsinaiticus.org/>
- http5 <http://csntm.org>
- http6 <http://ntvmr.uni-muenster.de>
- http7 [http://alta3b.wordpress.com/2009/04/06/translation\\_jn-1-1/](http://alta3b.wordpress.com/2009/04/06/translation_jn-1-1/)
- http8 <http://biblia-arabica.com>
- http9 <http://hdl.handle.net/2066/159141>
- http10 <http://evt.labcd.unipi.it>
- http11 <http://vbd.humnet.unipi.it>
- http12 <http://snf.ch/en/funding/projects/innovations-project-funding/Pages/default.aspx>
- http13 <http://portal.textometrie.org/demo>
- http14 <http://data.copticscriptorium.org>
- http15 <http://www.apd.gwi.uni-muenchen.de:8080/apd/project.jsp>
- http16 <http://www.ancientwisdoms.ac.uk/library/>
- http17 <http://dare.uni-koeln.de>
- http18 <http://cal-itsee.bham.ac.uk/itseeweb/fedeli/start.xml>
- http19 <http://sourceforge.net/projects/evt-project>
- http20 <http://www.tei-c.org/release/doc/tei-p5-doc/fr/html/PH.html#PHZLAB>
- http21 <https://humarec.org/index.php/continuous-publications-blog/22-annotating-the-manuscript-plans-and-challenges>
- http22 <https://transkribus.eu/Transkribus/>
- http23 <https://read.transkribus.eu/>
- Kashouh H., *The Arabic Versions of the Gospels, The Manuscripts and their Families* (Arbeiten zur neutestamentlichen Textforschung 42). De Gruyter (Berlin/Boston), 2012.
- Kouloughli D. E. Moyen Arabe et questions connexes. 2007. [http://cle.ens-lyon.fr/arabe/moyen-arabe-et-questions-connexes-30683.kjsp?RH=CDL\\_ARA120000](http://cle.ens-lyon.fr/arabe/moyen-arabe-et-questions-connexes-30683.kjsp?RH=CDL_ARA120000).
- Moawad S. *Al-As'ad Abū Al-Farağ Hibat Allāh Ibn Al-'Assāl: Die arabische Übersetzung der vier Evangelien*. Alexandria School (Le Caire), 2014.
- Monferrer-Sala J.P. Translating the Gospels into Arabic from Syriac: Vatican Arabic 13 Restored Section, Strategies and Goals. *Arabica*. 2015a;62:435-458.
- Monferrer-Sala J.P., The Pauline Epistle to Philemon from Codex Vatican Arabic 13 (Ninth Century CE) Transcription and Study. *Journal of Semitic Studies*. 2015b;60(2):341-371.
- Moretti F. *Graphs, Maps and Trees*. Verso (London), 2005.
- Raimundi G. B. *Al-Ingil al Muqaddas. Evangelium sanctum*. Rome, 1590-1591.
- Rosselli Del Turco R., Buomprisco G., Di Pietro C. et al. Edition Visualization Technology: A Simple Tool to Visualize TEI-based Digital Editions. *Journal of the Text Encoding Initiative*. 2014-2015;8. <http://jtei.revues.org/1077>.
- Schulthess S. The Role of the Internet in New Testament Textual Criticism: The Example of the Arabic Manuscripts of the New Testament. *Digital Humanities in Biblical, Early Jewish and Early Christian Studies* (Scholarly Communication 2). Ed. by C. Clivaz, A. Gregory and D. Hamidović. Brill (Leiden), 2013:71-82.
- Schulthess S. Tahriif in the Digital Age. *Ancient Worlds in Digital Culture* (Digital Biblical Studies 1). Ed. C. Clivaz, P. Dilley, D. Hamidović, in coll. with A. Thromas. Brill (Leiden), 2016a ; 214-230.



Schulthess, S. Les manuscrits arabes des lettres de Paul. La reprise d'un champ de recherche négligé, PhD Lausanne University and Radboud University, 2016b. <http://hdl.handle.net/2066/159141>.

Vollandt R. Some Historiographical Remarks on Medieval and Early-Modern Scholarship of Biblical Version in Arabic: A Status Quo. *Intellectual History of the Islamicate World*. 2013;1: 25-42.

#### **ANNEX 1**

The IT support was provided at the Vital-IT (<http://www.vital-it.ch>), Center for high-performance computing of the SIB Swiss Institute of Bioinformatics.

Many thanks to Prof. Ioannis Xenarios, director of VITAL-IT, for his support to our work.

Many thanks to Harley Edwards for his English proof-reading.

#### **ANNEX 2**

All hyperlinks have been checked on 05/30/17.