

**APPLYING TRANSKRIBUS'S ANNOTATION FUNCTIONALITY TO
ÖSTERREICHISCHEN NATIONALBIBLIOTHEK COD. 9737t**

–

A USER REPORT

CHIARA PETROLINI
chiara.petrolini@gmail.com

My five-day (18-23 December 2016) stay in Innsbruck was an invaluable experience that will change my research habits. I had the opportunity and privilege to work daily with the *Transkribus* team. With Professor Günter Mühlberger, and with the researchers Philip Khale, Günter Hackl and Sebastian Colutto I discussed the capability and benefits of *Transkribus* for Early Modern scholars, especially for the project, lead by Thomas Wallnig, “The Oriental Outpost of the Republic of Letters. Sebastian Tegnagel (d. 1636), the Imperial Library in Vienna, and Knowledge about the Orient” (Thoams Wallnig, Paola Molino, Claudia Römer, Hülya Celik, Chiara Petrolini).

Once I was familiar with the tool I learned how to segment documents into areas, lines, baselines, and single words, and how to make structural segmentation, tagging blocks of text as marginalia and notes, folio numbers, lists of books, different hands, and languages such as Arabic or Greek, as well as tagging single works as *person*, *date*, *sic*, *abbreviation*, *place*. Then I began dealing with Sebastian Tegnagel's letters. The Österreichischen Nationalbibliothek Cod. 9737t is now fully uploaded onto the central Transcription and Recognition Platform (TRP) server.

The volume consists of 232 letters (making a total amount of 714 digitized pages as JPEG images) written between 1609 and 1635, mostly to Sebastian Tegnagel (1573-1636) and to Franz Gansneb Tegnagel (1576-1622), by 40 different correspondents.

Since it is a collection of learned correspondence written in many hands, often with a variety of characters, hands and languages within a single document, this *corpus* of letters is an excellent testing ground for *Transkribus*. In future we hope to be able to digitalize and upload the two other volumes of Tengenagel's letters onto the platform.

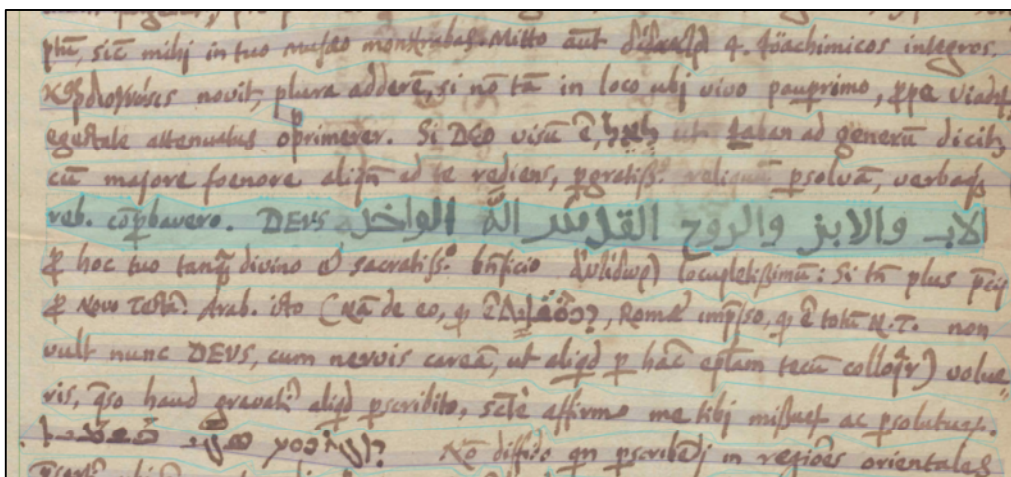
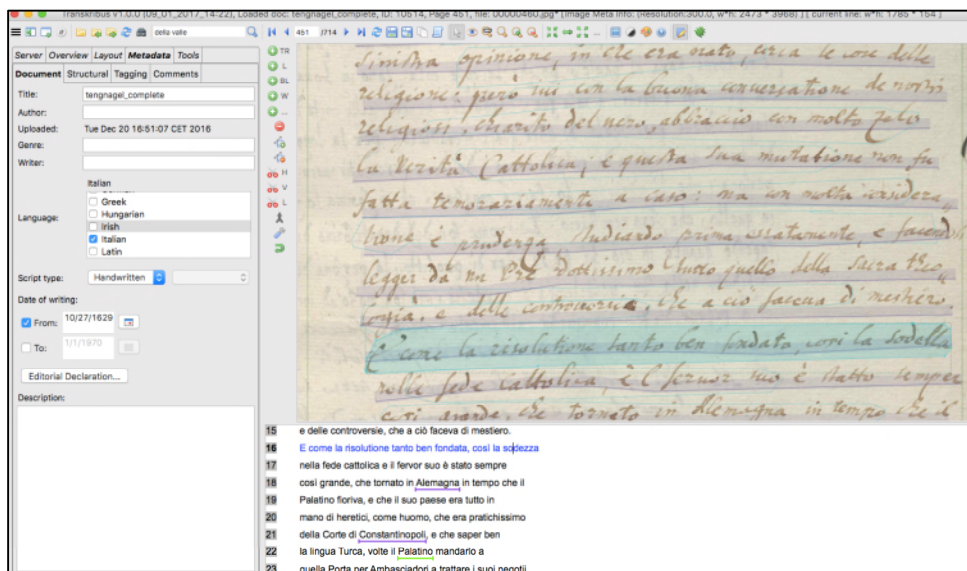


Fig. 1 Elias Opala to Sebastian Tengenagel (f. 320r)

During my stay, I transcribed about thirty letters, and in the coming weeks I will continue this work. The unloading of data for training the HTR (Handwritten Text Recognition) engine will be completed soon, and a model suitable for the three volumes of Tengenagel's correspondence will be created. It will then be possible to run automatic HTR. Due to the variations in hands, languages and styles, and the difficulty of reading some letters, the Tengenagel *corpus* trial is particularly challenging and potentially fruitful for the *Transkribus* team. I will communicate the outcomes of the first trial as soon as the model can start to be used, and I will evaluate the accuracy of the automatic transcription in a further detailed report. I selected for transcription a number of letters that are particularly significant both as regards their formal characteristics, and their content; I alternated relatively readable and clear letters with more challenging

ones, so as to get a clearer idea of the potential and limitations of the digital transcription platform developed by Günter Mühlberger's team. So far, I have transcribed letters in Italian and in Latin, written mainly by Pietro Della Valle, Michel d'Asquier, Sebastian Tengenel himself, and Josephus Barbatius. I will call upon Hülya Celik to insert missing Arabic words, and this will test the capability of *Transkribus* to support that alphabet.



Beyond the good results achieved as far as the automatic recognition of handwritten text is concerned, the work done to date has enabled me to discover the amazing power of *Transkribus* as a tool for manual transcription. Not only does it facilitate the selective annotation of the content of the letters, by identifying keywords, but it is also a powerful and innovative way to manage, store, and work on large and heterogeneous document collections.

As every Early Modern scholar knows, it is extremely hard to maintain a close link between the transcribed text and the materiality of the transcribed manuscript, something that is increasingly important in the age of digital humanities. Moreover, when one is immersed in the process of transcribing, it is

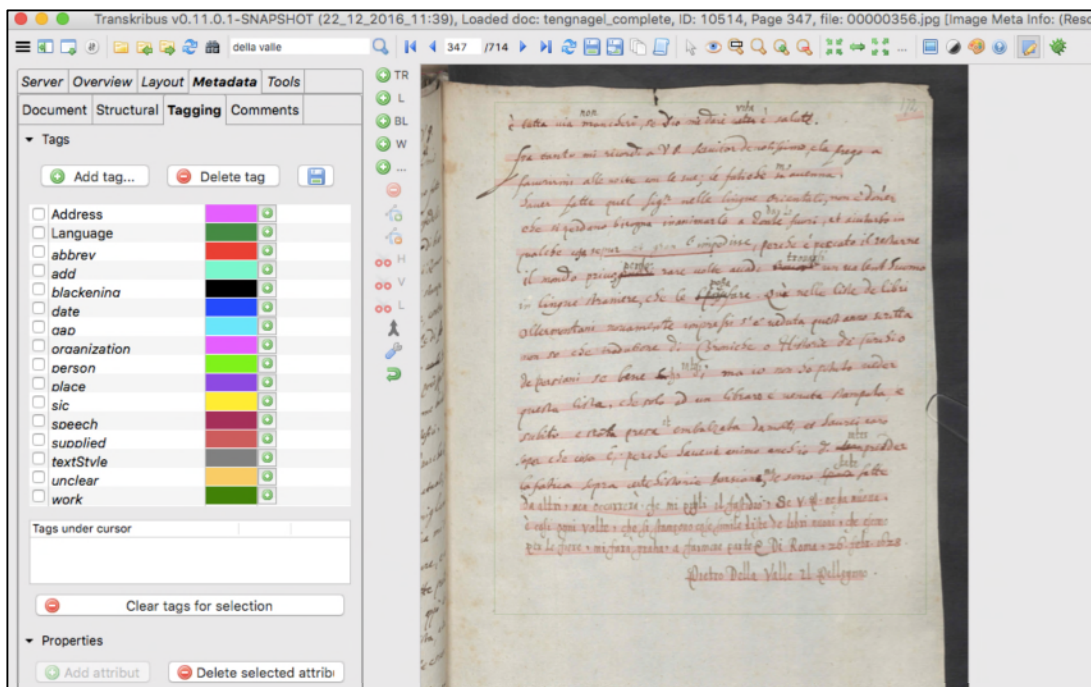
difficult to illustrate the complex and sometimes creative structure of an Early Modern document. *Transkribus* seems perfectly able to meet these challenges. Indeed, it makes it possible to work seamlessly, combining the text and images of the original manuscript, and making the final image fully searchable. Its flexibility allows the document to be separated into multiple areas, fully respecting the original structure and making immediately clear, for instance, the intervention of different hands and a number of chronological layers.

Another key feature of *Transkribus* is that it is a multi-user system. In the case of Tegnagel's letters, this means different scholars with specific skills can work on the same document in different locations, editing the transcription, helping to decipher difficult passages or words – and most importantly – transcribing and translating the passages in Arabic, Hebrew, Ottoman, Greek and other alphabets which often figure in Tegnagel's correspondence. Indeed, a major part of this correspondence concerns the expansion of the collection of manuscripts, including that of Oriental Manuscripts. As court librarian from 1608 to his death in 1636, Tegnagel was an active member of the international community of scholars, especially those with a keen interest in Oriental languages, culture and religion. For this reason for the letters to be studied properly it is essential to assemble people with a number of different skills. For instance, because of the need for close collaboration with Hülya Celik, a scholar of the Ottoman language based in Vienna and a member of the Tegnagel project team, multiple access to the documents uploaded onto the *Transkribus* platform is vital.

A crucial function is tagging. It is extremely quick and easy to create new tags. In Tegnagel's case I have created, among others, tags for book titles, for languages such as Arabic, Greek, and Hebrew, and for quotations. Another

invaluable feature is the “comments” function, which allows one to note down information about any letter, page, or single sentence.

Transkribus makes it possible to export any single page, group of pages or entire document, in various formats: PDF, DOCX, TEI, RTF. The user can choose to export just the images, the transcription, or the tags. I attach just one letters saved in PDF to show how the documents can be shared outside the Transkribus platform and showed to users unable to install or use the software.



As a member of the Tegnagel project, I am interested not in going ‘beyond’ the reading process, but rather in enriching it as much as possible, by critically thinking about metadata creation, pattern discovery, network analysis, and entering into fuller dialogue with scholars of different fields and with computer scientists. *Transkribus* provides a solid instrument for a deeper understanding of traditional research topics and questions, supporting humanities scholars in

their research. Moreover, its cumulative learning ability reshapes access to historical handwritten documents.

Of course, *Transkribus* is not specifically designed for miscellaneous volumes of letters and there are therefore a number of improvements that could be made to adapt it to this kind of material. I talked with the team, and requested the creation of bookmarks and other tools for bringing together several folios in order to browse the whole volume and easily find the required document.

This feature, of course, will also be essential for the *Archilet* project, which aims to create an online database of 16th and 17th century letters, just by registering the sender, recipient, date, location of sender(s) and recipient(s), names quoted in the letter, things of relevance, the incipit and the source. I will be presenting *Transkribus* to the *Archilet* members soon, and I personally intend to use the software in my work on the publication of correspondence between Fulgenzio Micanzio and Dudley Carleton that is to be published by *Edizioni Archilet* (<http://www.archilet.it/>).

I have already talked with the other members of *Coindico* (<http://coindico.com>), and we would like to use *Transkribus* for our collection of accounts of conversions from non-Christian religions (XVI-XVIII). Again, the ability to create subfolders for each story would be enormously useful. I also plan to present the platform to the group researching *Indipetae* – the letters written by young Jesuits applying to join the missions. These date from the sixteenth to the twentieth century and are preserved in the Archivum Historicum Societatis Iesu, Rome.

Finally, I have discussed about *Transkribus* with some members of the German Historical Institute in Rome (DHI), and in particular with Carlo Taviani, who showed great interest in the platform for a project of his that is seeking to build a database on Genoese merchant networks between Africa and the New

World (in the 15th and 16th centuries). Its main source material is notarial deeds from the Genoese archives. The research team will scan and possibly transcribe an entire index of 27 volumes.

I am sure that *Transkribus* would be extremely beneficial for all these projects. However, it would be a great help if a workshop was organized in Italy by the *Transkribus* team. It would be an opportunity to communicate project goals and values not to the Italian scientific community, including archivists and librarians. To use *Transkribus*, it is essential to obtain images of documents, and this process is generally too expensive and complicated to employ in Italy's enormously rich archives and libraries.

However, a joint effort, where computer scientists, archivists, historians, philologists, and linguists work together would undoubtedly be the best way to achieve significant research outcomes. *Transkribus* can be a key factor in such co-operation.