

## **Digital Dutch VOC Linguistic Archive: Modeling a social editing platform for historical linguistic research**

*Anna Pytlowany*

### *Goals*

The primary aim of the project 'Digital Dutch East India Company (VOC) Linguistic Archive' is to bring together in one online platform all Dutch documents from 1600-1825, which were written either by VOC employees or under the auspices of the Company, and relate to the languages of the 'newly discovered' territories (Africa and South East Asia). Many of these documents are kept in the VOC archives, yet some are scattered in private and public collections from Paris, to Venice, to Sydney. So far, these documents have never been assembled and arranged into one comprehensive list. An online database would greatly facilitate discovering their provenance and itineraries. It would also allow a better evaluation of the scope and content of Dutch colonial linguistic heritage.

The secondary, yet more ambitious goal is to design and produce scholarly social editions of those texts. Such collaboration would bring together researchers from different countries and languages, and result in opening access to a vast range of unpublished linguistic material.

### *Challenges*

One of the biggest challenges to face in the modeling of such projects is the potential multitude of languages and scripts involved. Although the language of the main body of texts is predominantly Dutch, the grammars and vocabularies, by their very nature, each contain at least one other 'exotic' language, often written in the native, non-Latin script. Inasmuch as one part of the multilayered text may be accessible to a particular researcher, chances are that the remaining levels require further work to make them readily understandable.

This is perhaps the most compelling characteristic of the archival material of this type: it lends itself perfectly to a collaborative, community-sourced online editing. The tasks will include: transcribing and translating existing documents, proofreading and validating transcriptions, making annotations, and adding references.

### *Pilot project*

The pilot study will focus on producing scholarly social editions of the two earliest dictionaries of Hindustani written in Dutch at the turn of 17th and 18th century: 'Instructie of Onderwijsinghe van de Hindustanse en Persiaanse Taale' (Utrecht University Library, Ms. 1478)—a grammar and a thematic vocabulary written by Joan Josua Ketelaar in 1698, and an anonymous Hindustani-Persian-Dutch dictionary, 'Vocabularium Persico-Belgicum' (Leiden University Library, MS 589), written in the late 17th century.

In addition to the standard editorial tasks, the most challenging part will be the interpretation of some phonemically recorded words, and matching them, for comparison purposes, with the entries from later Hindustani and Persian dictionaries (Platts 1884 and Steinglass 1892).

In parallel, an edition of the third unpublished Hindustani dictionary, 'Thesaurus Linguae Indianae', composed by the French Capuchin François-Marie de Tours in 1704, will be undertaken by a team from the University of Uppsala. Together, the three texts form an important historical witness of early modern linguistics; therefore it will be crucial to test, analyse and solve the interoperability issues between the two projects.

The projects will draw on technologies and methods tested in similar undertakings such as tranScriptorium and the Suda On-Line.