

Summary “Unsilencing the VOC testaments”

Saskia Virginia Noot & Clare Shutt

29 March 2021

In the framework of a Case Study Project as part of the master’s program Archival and Information Studies at the University of Amsterdam, this project’s aim is to unsilence the archive and create a new search tool to make marginalized people within the archive findable. This project focused on constructing ways to create new indexes to find ‘women’ and ‘indigenous people’ in the VOC testaments.

The “oostindische testamenten” provided the basis for the project. This collection is part of the archive of the Dutch East India Company (VOC) at the National Archives in The Hague. The testaments are organized in 51 inventory numbers. There is no logical order in this collection, nor is it clear whether many of the original testaments are missing; it is a collection of surviving testaments. The testaments can be divided in three types:

1. testaments made before notaries in the Netherlands for people who were going to the Dutch East Indies
2. testaments made on ships that were headed towards the Dutch East Indies before the person who was appointed to make the deeds on the ship
3. testaments made before notaries in one of the VOC’s settlements

The current finding aid to search these testaments is based on a 19th century index that only allows for the finding of the testators who were (often) white, male employees of the VOC. Within the current index, ‘women’ and ‘indigenous people’ are virtually untraceable and clouded in obscurity, even though a large part of these documents mention people other than the testators.

This project is situated in the framework that focuses on the decolonization of the archive, which is characterized by the tackling of the dominant narrative and uncovering silences within the archive. Concepts central to the implementation of this project are Valerie Johnson’s “total archive”, “strategic essentialism” explained by Maura Reilly and the “feminist standpoint epistemology” developed by Michelle Caswell. Because classification and categorization are assigned an important role in the implementation of this project, there can be no doubt that a certain amount of bias occurs in this project.

After the digitization of the original testaments, the National Archives used the handwritten text recognition platform Transkribus to transcribe the collection. The output of this process constitutes the training data for this project.

The first step in the project involved a close reading of the testaments. Both the scan of the original document and the transcribed text were analysed. This analysis showed that in addition to the testator, many other individuals are mentioned in the testaments.

In order to be able to extract the 'women' and 'indigenous people' from these testaments, entities referring to these individuals were identified and categorized. In the second step of this project, the transcriptions of the testaments of inventory number 6847 were uploaded into the online annotation tool 'brat' which allows for the creation of custom categories and offers the possibility to identify a relationship between multiple entities. In the course of this research and its results, the categories 'women', 'indigenous' and 'men' were applied. Other categories were also created but proved less applicable. Once the tagging was finished, the annotations made in brat were exported.

Annotating entities in brat is a manual task. In the scope of this project a total of two hundred transcribed documents originating from one inventory number were used for annotation. All of these texts were drawn up in the period around 1770, in and around the city of Batavia. Although this was sufficient to make the concept of this project work, in future research it is advisable to use transcribed texts originating from different inventory numbers; of testaments drawn up in different locations and in different periods of time. Overall, this would result in a more complete and honest representation of the entities mentioned in these texts.

The third step in this project comprised of writing the code for the intended index. For this purpose, the software Jupyter Notebook was used. The written code was directed to import spaCy and the annotated files which were downloaded from brat. Next, the code was directed to scan all of the annotated files to detect the occurrence of the tagged entities. Thereafter, all the entities were extracted from the text files on which the code was executed. In the scope of this paper the text files originating from inventory number 6849 were used.

Finally, the command lines were written to make an export to an Excel document consisting of the occurring entities in these text files, including data comprising of the exact location of the found entity: inventory number, number of the text file (which corresponds with the number of the scan) and the exact character number in the text file where the entity starts. This final step resulted in the creation of a new index for the finding of mentions of 'women' and 'indigenous people' as well as 'indigenous women', in the "oostindische testamenten".

In addition to the creation of a new index the annotated data and the written code offer the possibility to calculate the percentage of documents that have a least one mention of the entities for

'women' or 'indigenous people'. In order to compare the occurrence of entities referring to 'women' and 'indigenous people' in the documents, the percentage of entities referring to 'men' has also been included. For inventory number 6849, this gives the following results:

- one or more mentions of male entities occur in at least 95,7% of all documents
- one or more mentions of women entities occur in at least 61,9% of all documents
- one or more mentions of indigenous people's entities occur in at least 29,5% of all documents

These statistics demonstrate that many more people are mentioned in these documents than just the testators. Moreover, these statistics are proof that a new index to make 'women' and 'indigenous people' findable is useful and a necessary step towards decolonizing the archive.