

## **Historical sources for Early Afrikaans 1675-1925: Small (first) steps towards developing an online historical corpus and archive for linguistic research**

Wannie Carstens and Roné Wierenga

### **Abstract**

The origins of Afrikaans has always interested Afrikaans and Dutch researchers but despite numerous publications on the subject there is still no satisfactory model or theory to discern when Afrikaans stopped 'being Dutch'.

With the advent of new technology and methodologies, it is becoming increasingly possible to base research on this topic on reliable and verifiable data. As new methods for linguistic research are established the need for sufficient linguistic data to apply these methods to is emphasized. The lack of or inaccessibility of language resources for linguistic research often exclude certain topics from being thoroughly researched, e.g. the influence of the Khoi languages on 17th century Cape-Dutch (Early Afrikaans).

*Research and documentation go hand in hand, but both require time. This is also the case with this long-term project which aims to facilitate research on Afrikaans by establishing a large database of historical Afrikaans sources. This database aims to be multifunctional by (a) housing sources that note and or comment on the development of Afrikaans from around 1600-1925, and (b) providing access to various sources that are in themselves language data from this period. This project, therefore, aims to stimulate new and more thorough research on Afrikaans by making more sources and resources freely available.*

This project is largely based on the Digital Library for Dutch Literature (Dutch: Digitale Bibliotheek voor de Nederlandse Letteren) (DBNL) and is therefore not undertaken in isolation but rather aims to incorporate already existing projects (such as the Digital Bibliographies of Afrikaans Linguistics and Literature) and establish cooperation between institutions and other parties working in this environment. As a secondary mission, this project aims to digitalize old Afrikaans texts that are in danger of being destroyed or are no longer accessible to the public and compiling these digital copies in a corpus and electronic archive. The model proposed for this project is designed to be adaptable and therefore instrumental in establishing similar archives and corpora for other indigenous South African languages.

This paper then discusses (a) why such a database (corpus, archive) is needed for Afrikaans, (b) the work that has been completed and is yet to be done to establish such a database (c) the form such a database will take, (d) and the potential it has for other South African languages. *This paper, therefore, reports on an ongoing project that will ultimately have a major impact on existing and future linguistic research on the origins of Afrikaans.*

**Keywords:** general linguistics; historical corpus; historical archive; historical linguistics; language documentation; electronic language research; digitalisation