

Mining memories of the Indonesian War of Independence

The research programme *Decolonization, violence and war in Indonesia 1945-1950* is conducted by the Royal Netherlands Institute of Southeast Asian and Caribbean Studies (KITLV), the Netherlands Institute of Military History (NIMH) and the Institute for War, Holocaust and Genocide Studies (NIOD), and aims to investigate the nature and the extent of the war crimes that were perpetrated during the Indonesian War of Independence in between 1945 and 1950.¹ The research is partly based on witness accounts. Geert Oostindie's monograph *Soldaat in Indonesie*, which was first published in 2015, discusses the results of a detailed analysis of about 650 letters, diaries and memoirs containing recollections of the period.²

In 2017, the Leiden University Centre for Digital Scholarship has begun a project to explore whether the insights that were arrived at using traditional scholarly methods could be replicated, and perhaps even enhanced, using computational methods. The ca. 650 egodocuments that had been examined for *Soldaat in Indonesie* had firstly been scanned and converted into machine-readable text via OCR. The texts in the corpus were subsequently subjected to a range of computational analyses. During a first phase of the project, the aim was primarily to identify the words that were used in the egodocuments to characterise the native Indonesian soldiers fighting in for the Dutch army, and to trace relevant historical developments in the attitudes that were expressed. Following the approaches taken by Benjamin Schmidt³ and by Kutuzov et al.,⁴ the words that were associated with the indigenous soldiers were investigated on the basis of word embeddings. Different models were generated for egodocuments written during or directly after the war on the one hand and egodocuments written with a time lag of two or three decades on the other. The attitudes of Dutch soldiers were investigated additionally via techniques in the field of sentiment analysis.

A second objective of the study was to build software for the recognition of specific acts of violence within the digitised texts, to be able to make an assessment of both the nature and the frequency of particular war crimes. It is generally accepted that both the Dutch soldiers and the native Indonesian soldiers faced disproportionate amounts of violence, but it is still unclear whether such act of violence were incidental or more systematic. Using word embeddings, topic modelling and collocation analysis, a lexicon of relevant terms was produced, and this lexicon was used to quantify occurrences of war crimes and to visualise their dispersion, both with individual documents and within the corpus as a whole.

Next to discussing the main results and, this paper reflects on the methodological issues that needed to be addressed and on the overall value of computational methods for research into military memoirs.

¹ <<https://www.ind45-50.org/en>>

² Geert Oostindie. *Soldaat in Indonesië*. Amsterdam: Prometheus, 2015.

³ Benjamin Schmidt. *Vector Space Models for the Digital Humanities*, <<http://bookworm.benschmidt.org/posts/2015-10-25-Word-Embeddings.html>>

⁴ Kutuzov, A., Velldal, E. and Øvrelid, L. "Tracing armed conflicts with diachronic word embedding models". In: *Proceedings of the Events and Stories in the News Workshop*, 2017. <10.18653/v1/W17-27>