

Karina van Dalen-Oskam¹, Bram Oostveen¹, Joris van Zundert¹, Henny Brandhorst²,
Menno Metselaar², Marco Streefkerk²

¹ Huygens Instituut (KNAW), Amsterdam

² Anne Frank Stichting, Amsterdam

Stylistic Influences in Anne Frank’s Writings

Introduction

Using computational methods for the analysis of writing style (stylometry), this paper will examine which of the books Anne Frank read influenced her own writing style in a measurable way. We especially focus on the manuscript in which she started rewriting her diary into a novel. In Table 1 we list Anne Frank's original texts as we used them. In some cases, we combined files A1, A2 and A3 – based on the red and white checkered diary and two other notebooks respectively – in one file A. File B is the version that Anne Frank wrote in 1944 and intended to become the novel *Het Achterhuis*, which is translated into English in the new online edition by De Bruijn et al. (2024) as *The Back House*. Elsewhere, the works are also referred to as *The (Secret) Annex* and *The Diary of a Young Girl*.

Table 1: The works of Anne Frank

Text	Filename	Tokens
Diary	A1	23,293
	A2	37,923
	A3	21,200
<i>Het Achterhuis</i> (<i>The Back House</i> , the rewritten version of the diary, which Anne Frank hoped to publish as ‘a novel about the Back House’.)	B	49,532
Cady’s leven (Cady’s Life)	CL*	6,730
Cady’s leven, fragmenten (Cady’s Life, fragments)	CLF*	1,647
Mooie-zinnenboek (Book of Beautiful Sentences)	MZ	8,800
Verhaaltjes (Tales, and Occurrences from the Back House)	V**	33,030
Overige verhaaltjes (Other tales)	OV**	3,093

* Combined in file CL-CLF, with 8,377 tokens.

** Combined in file V-OV, with 36,123 tokens.

Case study 1

For our first case study we compared the texts written by Anne Frank with a random selection of books that she mentions to have read, including the four Joop ter Heul volumes by Cissy van

Marxveldt that are often mentioned as most influential on her work (e.g. Broos. 2000, Patterson Iskander 2000, Soeting 2018). For this first exploration we used texts from different genres: novels for adults by Jo van Ammers-Küller and Ina Boudier-Bakker, novels for young adults by F.J. de Clercq Zubli, Cissy van Marxveldt, Nanda, Willy Pétillon, and Nico van Suchtelen, a selection of stories for old and young by Cor Bruin and others, a biography by Willy Corsari, and a literary anthology edited by Johan Winkler.

We made use of the Stylo package for R (Eder et al, 2018), focusing on words (unigrams). A cluster analysis using the 1000 most frequent words in the corpus resulted in Figure 1. What immediately stands out is that the novels by Van Marxveldt, which Anne Frank read in 1942, are not the closest ones to her own works, but those by De Clercq Zubli, which she read more recently, in 1944. This is confirmed by the principal components analysis of the 1000 most frequent unigrams, see Figure 2.

Figure 1: Cluster analysis, cosine delta, 1000 MFW.

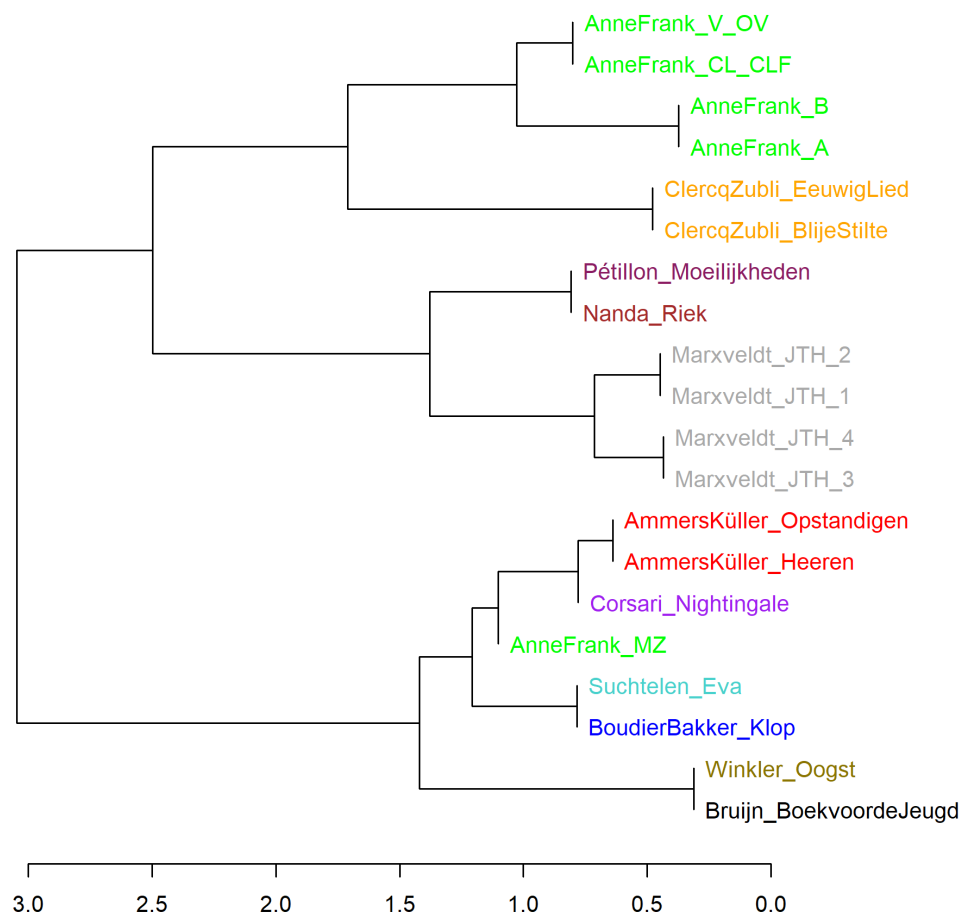
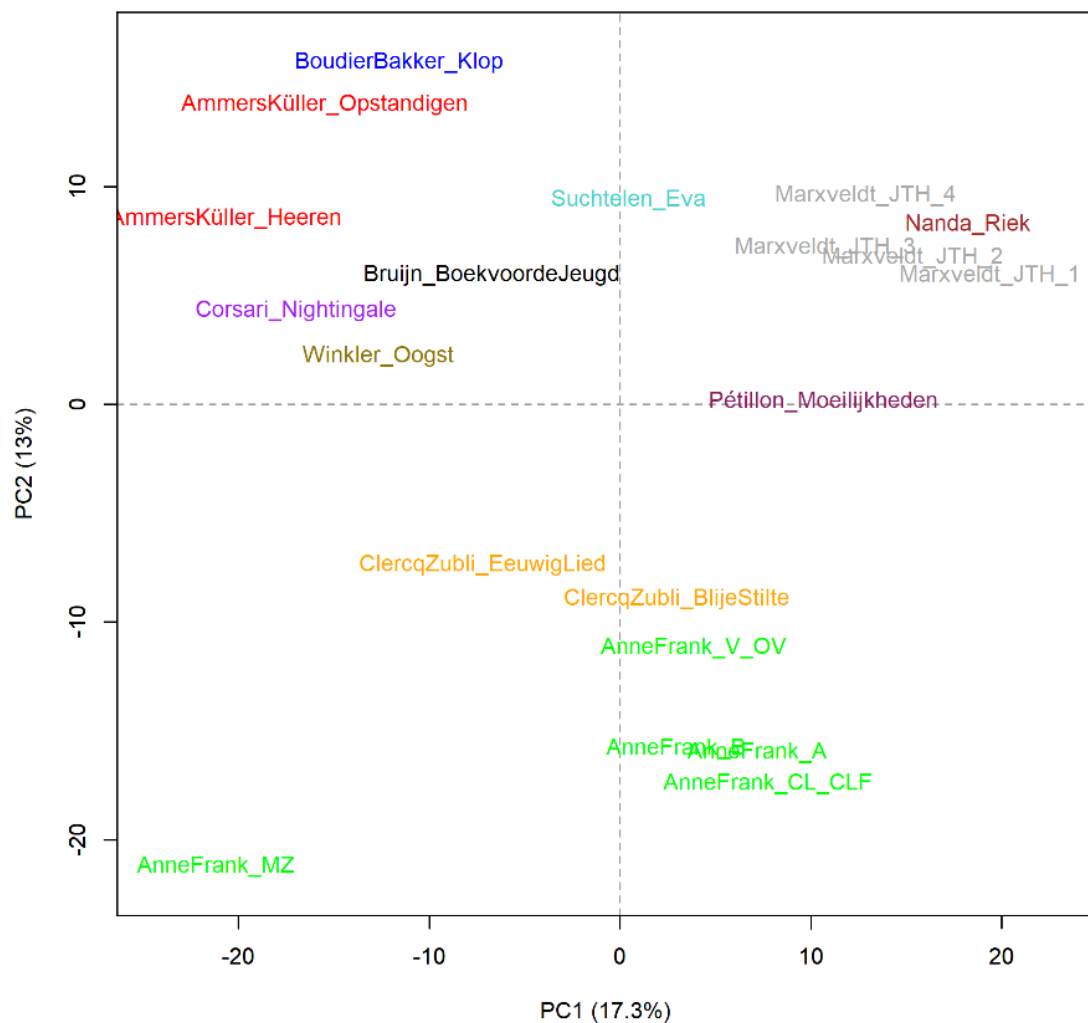


Figure 2: Principal components analysis, 1000 MFW.



Additionally, we applied Rolling stylometry to manuscript B, the rewritten version of the diary, using as reference set the novels from the same main cluster in Figure 1: the four novels by Van Marxveldt, the two by De Clercq Zubli, and the novels by Pétillon and Nanda. We use svm as measure, based on cosine delta, the combination currently considered to be the most precise (Evert et al. 2017). When we take the 100 most frequent words (Fig. 3), which are usually all function words, the two novels by De Clercq Zubli are the most prominent influence – the red part of the lower bar, only very occasionally interrupted by some echoes from Van Marxveldt’s works (in green). On the second tier, the top bar, the start of manuscript B shows similarities in the frequencies of most frequent words with the novel by Pétillon. Nanda’s novel remains invisible. With 1000 most frequent words (Fig. 4), content words do start to come in. And then the picture changes significantly: here, finally, Van Marxveldt’s influence becomes visible, which therefore seems to reside more in content words than in function words. Additionally, Pétillon’s influence on the start of text B also seems to have grown.

Figure 3: Rolling stylometry, 100 MFW, svm on cosine delta, slice size 5,000, slice overlap 4,500 tokens.

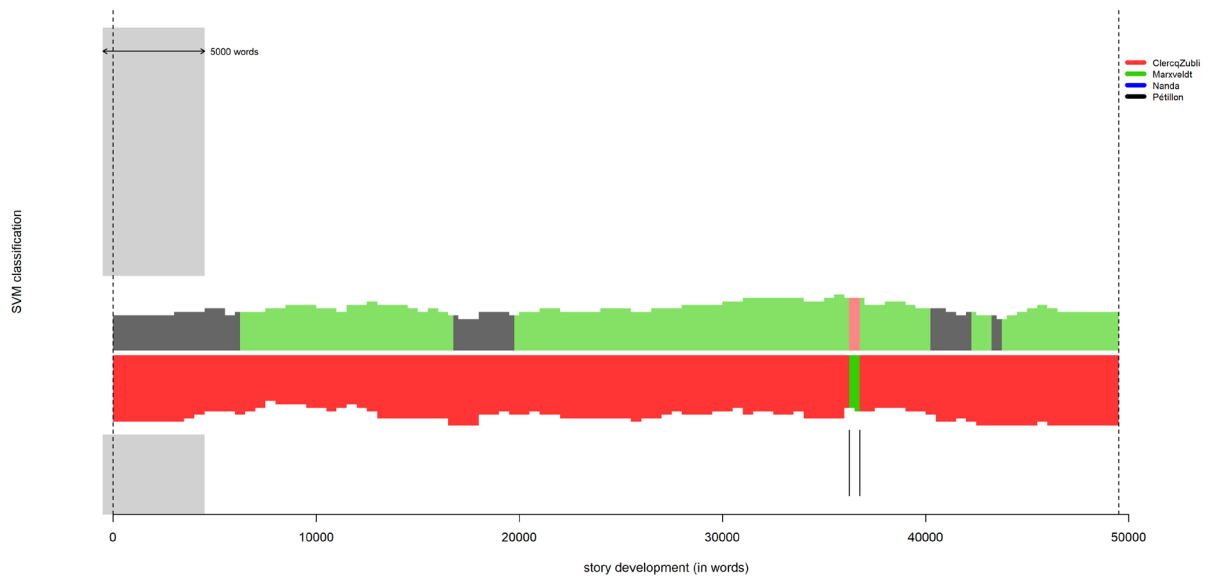
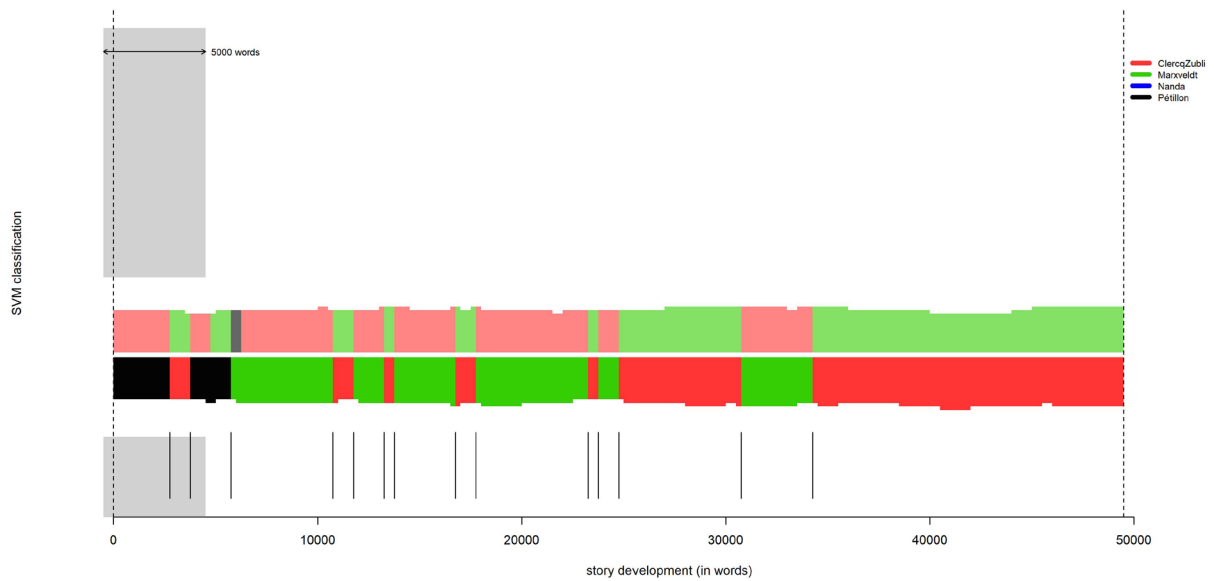


Figure 4: Rolling stylometry, 1000 MFW, svm on cosine delta, slice size 5,000, slice overlap 4,500 tokens.



Case study 2

The first case study used only a small set of texts. In our second case study, we will replicate the research on a larger corpus, to see whether the picture then changes again. We made a list of all works that Anne Frank may have read during her time in hiding. Because stylometric analyses usually pick up on genre differences, we will now limit ourselves to narrative fiction and narrative non-fiction. For this, we need to add around 40 other texts to our corpus. Half of these are already available in digital form and only need to be cleaned up for computational analysis. The other half needed to be digitized first and then cleaned. By now, they have been digitized, and we are currently cleaning all texts. In our presentation at the conference, we will be able to present the results of this replication study on a larger corpus and provide an answer to our question which of the books Anne Frank read influenced her writing style the most.

We also aim to explore which concrete linguistic forms this influence took: can we pinpoint which phrases, which grammatical structures, or which other style-related features reflect which of the works read by Anne Frank the most? For this, we will apply various feature space models for which we will operationalize features based on, amongst others, the findings of De Bruijn et al. (2024). Furthermore, because the corpus is comparatively small, it will be possible to apply a brute force approach by which we compare each sentence written by Anne Frank with each sentence in the other novels. Using Levenshtein distance on both lemmatized and non-lemmatized sentences we will be able to produce a list of all similar sentences and generate a network visualization. We hypothesize that these will reflect the influence of the books Anne Frank read on her own writing style.

References

- Broos, Ton J. (2000), Anne Frank's literary connections. – In: Deleu, Josef (chief ed.), *The Low Countries, arts and society in Flanders and the Netherlands*. – Rekkem : Stichting Ons Erfdeel, p. 177-189 . URL https://www.dbnl.org/tekst/_low001200001_01/_low001200001_01_0026.php
- Peter de Bruijn (ed.), with the cooperation of Elli Bleeker and Marielle Scherer *Anne Frank – Manuscripts*. Dutch/English edition, translated by Sam Garrett et al. Under the auspices of the Huygens Institute for History and Culture of the Netherlands (KNAW) & Anne Frank Stichting [Anne Frank House]. Online publication, provided by the Association for Research and Access to Historical Texts. Brussels, 2021. <https://annefrankmanuscripts.org>. Version: June 2024.
- Maciej Eder (2016). Rolling Stylometry. In: *Digital Scholarship in the Humanities*, Volume 31, Issue 3, September 2016, Pages 457–469, <https://doi.org/10.1093/lc/fqv010>
- Maciej Eder, Mike Kestemont, and Jan Rybicki (2016). Stylometry with R: A package for computational text analysis. In: *The R Journal*, 8(1): 1–15, <https://journal.r-project.org/articles/RJ-2016-007/>
- Stefan Evert, Thomas Proisl, Fotis Jannidis, Isabella Reger, Steffen Pielström, Christof Schöch, Thorsten Vitt (2017). Understanding and explaining Delta measures for authorship attribution. In: *Digital Scholarship in the Humanities*, Volume 32, Issue suppl_2, December 2017, Pages ii4–ii16, <https://doi.org/10.1093/lc/fqx023>

- Sylvia Patterson Iskander (2000). Anne Frank's reading: a retrospective, in: Enzer, Hyman A. & Solotaroff-Enzer, Sandra (eds), *Anne Frank : reflections on her life and legacy*. - Urbana, IL [etc.]: University of Illinois Press, p. 199-109.
- Monica Soeting (2018). Dear diary, dear comrade: fiction and non-fiction in the diaries of Setske de Haan, Joop ter Heul and Anne Frank. - In: *European Journal of Life Writing*, 7 (2018), p. CP183-CP199, <http://dx.doi.org/10.5463/ejlw.7.270>.