**EDER — Magnifying**

<3 images>

**Through the Magnifying Glass: Rolling Stylometry for Collaborative Authorship**

Eder, M.

This poster introduces a new stylometric method that combines supervised machine-learning classification with the idea of sequential analysis. Developed to assess mixed authorship, it can also be used as a magnifying glass to inspect works with an unclear stylometric signal.

Sequential analysis is a very attractive way of assessing linear phenomena. The concept of the moving window is particularly important to measure given elements in their sequential order. It has been introduced to stylometry and extended by van Dalen-Oskam and van Zundert in their study on the medieval Dutch poem titled ‘Roman van Walewein’ (van Dalen-Oskam and van Zundert, 2007). Other notable approaches to visualize stylistic shifts using moving windows include papers on Middle Dutch rhyme words (Kestemont, 2010), on three disputed English prose texts (Burrows, 2010), and on *The Tutor’s Story* by Kingsley and Mallet approached with *t*-tests (Hoover, 2011). Recently, Rolling Delta has been applied to examine collaborative works by Joseph Conrad and Ford Madox Ford (Rybicki et al., 2014).

The new stylometric method as discussed in this study is based on the above approaches. Its general idea is very simple: a text to be analyzed (e.g., an anonymous text to be attributed) is chunked into equal-size blocks (partially overlapping). Then, instead of attributing the text in its entirety, the goal is to perform an independent similarity test for each chunk, and to inspect the results as a sequence of ordered stylistic signals. Arguably, any classification method can be combined with this procedure. In the present study, support vector machines (SVM), nearest shrunken centroids (NSC), and Delta in its classical Burrowsian flavor have been used (Eder, 2015).

The method is designed to detect stylistic takeovers. An example is *The World’s Desire*, a classic fantasy novel written collaboratively by Henry Rider Haggard and Andrew Lang in 1890. It is a story of the hero Odysseus, who returns home to Ithaca after his journey: instead of finding his home at peace, however, he is involved in several new adventures. The plot of the novel as well as its mythological background was set by Lang, while Haggard—the author of several appreciated adventure novels—contributed his imagination and style. It is assumed that Haggard reworked Lang’s drafts and actually wrote most of the novel except the first four chapters, which were written entirely by Lang.

An experiment involving a reference corpus of eight novels by Haggard and eight by Lang corroborates the hypothesis. In Figure 1, the results of the rolling technique in Delta flavor have been shown. When the novel in question is split into sequentially ordered chunks, one can easily observe its development and—more importantly—a stylistic takeover in the first part of the text. The break point takes place roughly after the fourth chapter.

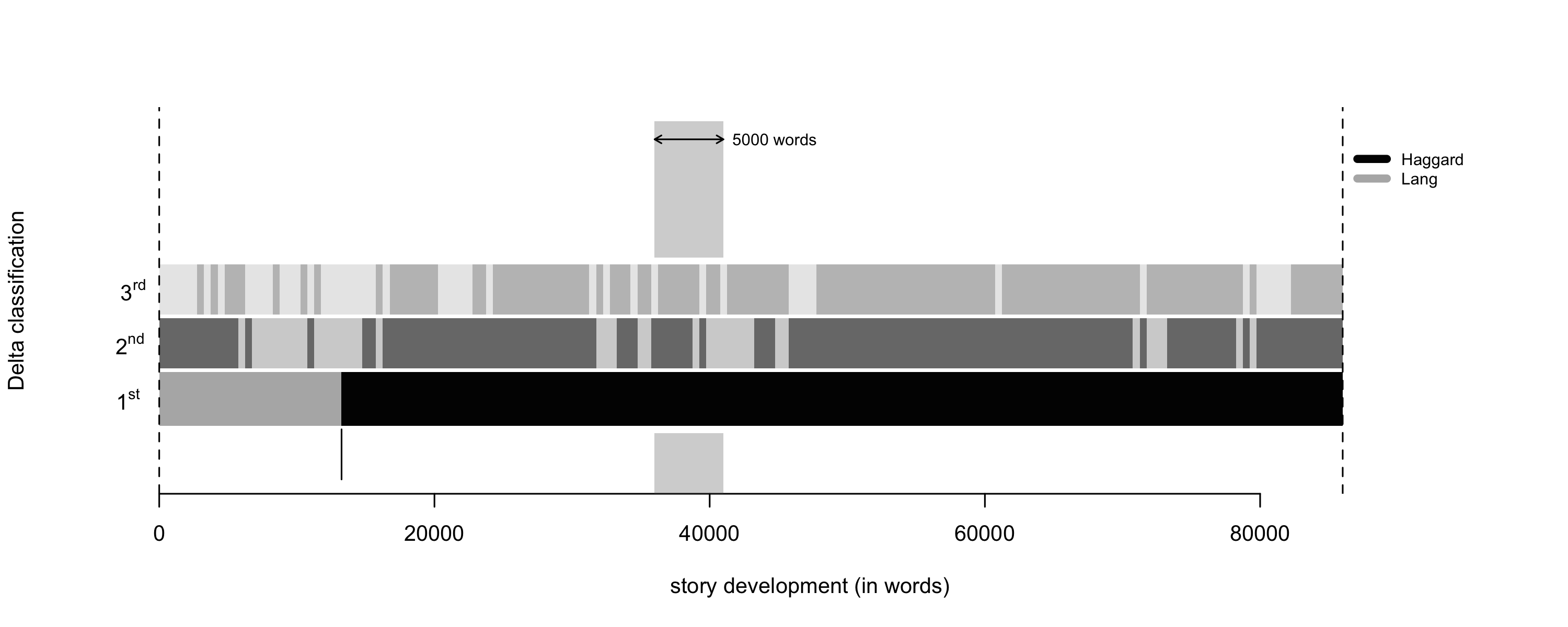


Figure 1. *The World’s Desire* by Haggard/Lang assessed using Rolling Delta and 100 MFWs. Sections recognized to have been written by Haggard are marked black; Lang’s sections are marked gray.

Two other variants of the rolling technique will be demonstrated using the Latin translation of the Bible known as the Vulgate. It is a very interesting example of a multi-level collaborative work. On the one hand, the original biblical text itself has many authors; on the other hand, the Latin translation is partially attributed to St. Jerome, and partially adopted from the already-existing Old Latin version (Vetus Latina). Thus, the authorial signal of the original interferes with translatorial takeovers: it is known that St. Jerome rendered most parts of the Old Testament and adopted the New Testament from earlier Latin versions. Detecting the translatorial voice might be a challenging task, since the author of the original is usually stylistically predominant in a translated text (Rybicki, 2012). Moreover, one should remember that two different languages of the Bible—Hebrew and Greek—might produce yet another stylometric signal.

The testing procedure started with transplanting two training samples: one sample for St. Jerome’s translatorial voice, and one for the Old Latin translation as included in the Vulgate. The Book of Genesis and Acts were chosen, respectively. Next, the subsequent sections of the entire Vulgate were tested against the training samples. The results are shown in Figures 2 and 3. Rolling NSC and Rolling SVM both reveal similar stylistic takeovers. Interestingly, the main change in style takes place roughly between the Testaments (dashed vertical line).

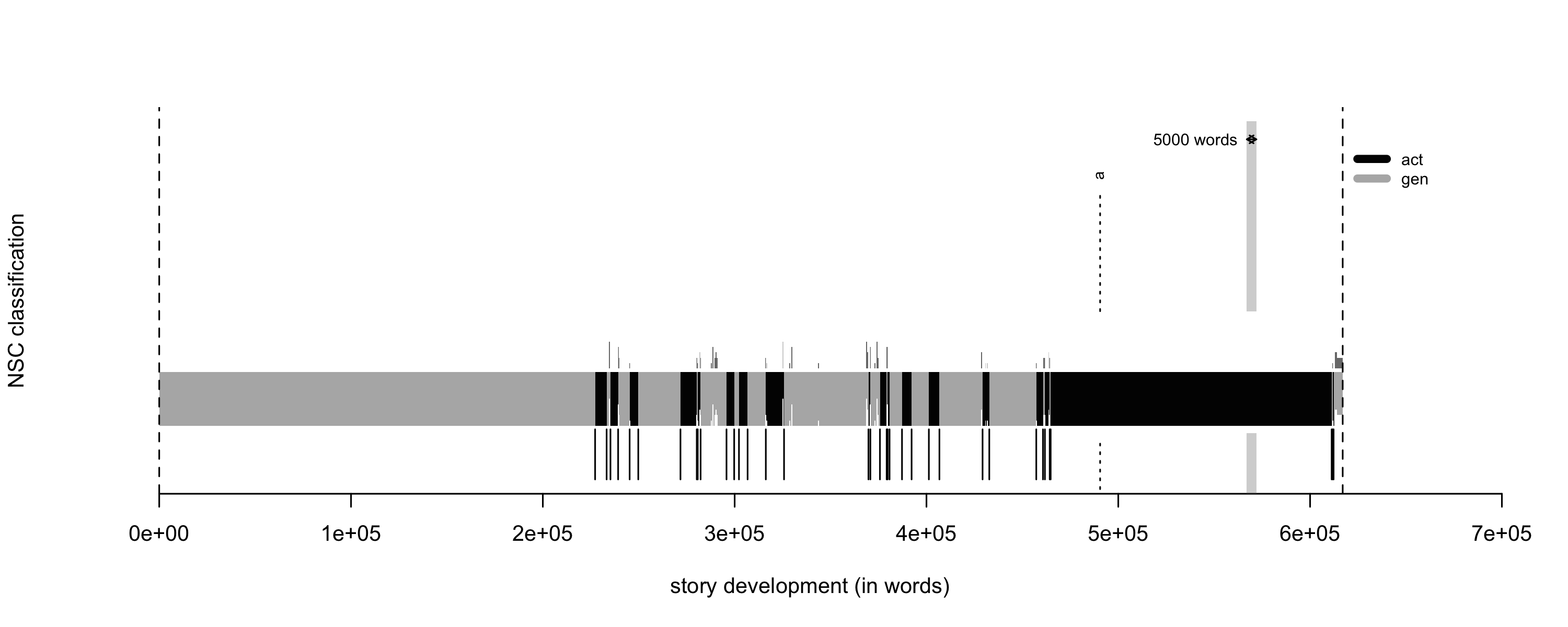


Figure 2*.* The Vulgate assessed using Rolling NSC and 1000 MFWs. Sections recognized to be similar to the Book of Genesis are marked gray*;* sections similar to Acts are marked black.

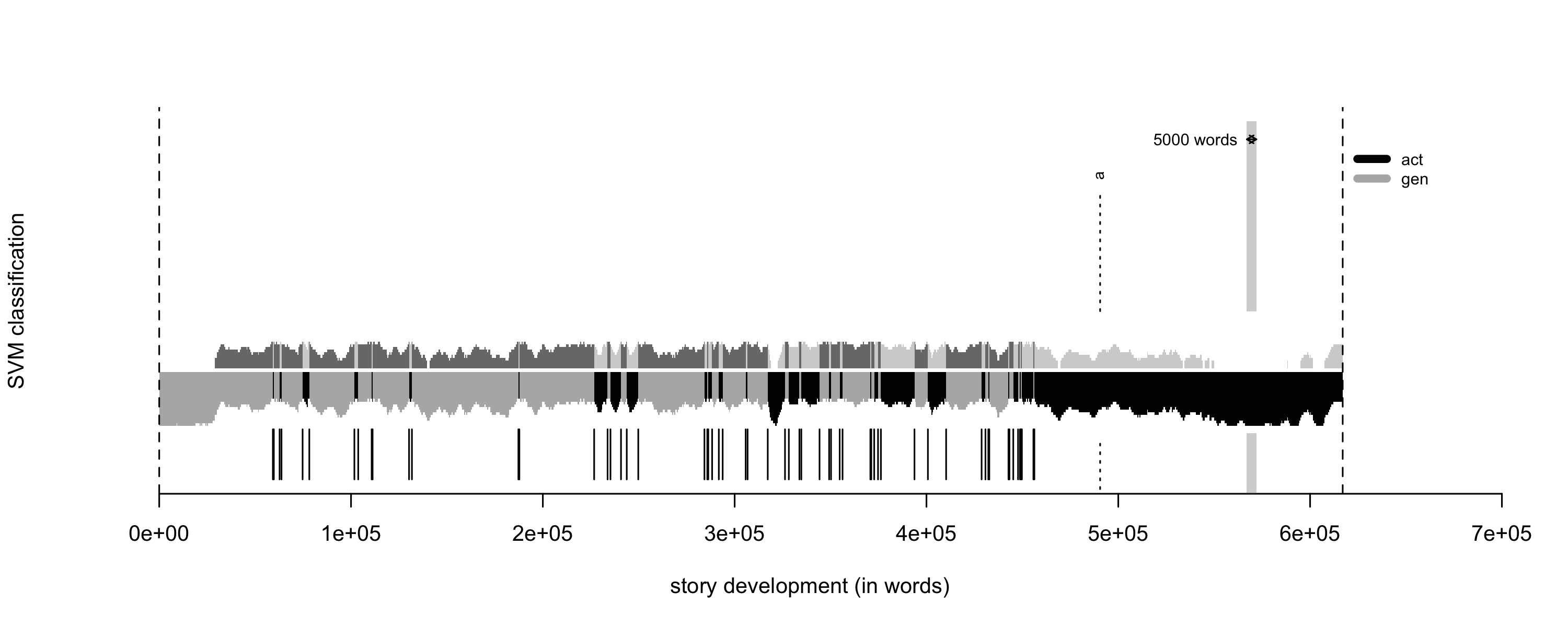


Figure 3. The Vulgate assessed using Rolling SVM and 100MFWs.

## References

**Burrows, J.** (2010). Never Say Always Again: Reflections on the Numbers Game. In McCarty, W. (ed.), *Text and Genre in Reconstruction: Effects of Digitalization on Ideas, Behaviors, Products and Institutions*. Cambridge: Open Book Publishers, pp. 13–36.

**Eder, M.** (2015). Rolling Stylometry. *Digital Scholarship in the Humanities*. 7 April, http://dx.doi.org/10.1093/llc/fqv010.

**Hoover, D. L.** (2011). ‘The Tutor’s Story’: A Case Study of Mixed Authorship. In *Digital Humanities: Conference Abstracts*, Stanford, CA, pp. 149–51.

**Kestemont, M.** (2010). Velthem et al.: A Stylometric Analysis of the Rhyme Words in the Account of the Battle of the Golden Spurs in the Fifth part of the Spiegel Historiael. *Queeste,* **17**: 1–34.

**Rybicki, J.** (2012). The Great Mystery of the (Almost) Invisible Translator. In Oakes, M. P. and Ji, M. (eds), *Quantitative Methods in Corpus-Based Translation Studies*. Amsterdam: John Benjamins, pp. 231–48.

**Rybicki, J., Kestemont, M. and Hoover D.** (2014). Collaborative Authorship: Conrad, Ford, and Rolling Delta. *Literary and Linguistic Computing,* **29**(3): 422–31.

**van Dalen-Oskam, K. and van Zundert, J.** (2007). Delta for Middle Dutch: Author and Copyist Distinction in Walewein. *Literary and Linguistic Computing,* **22**: 345–62.