

CZECH LITERATURE STUDIES

PETR PLECHÁČ

Versification and Authorship Attribution

INSTITUTE OF CZECH LITERATURE
KAROLINUM PRESS

This PDF includes a chapter from the following book:
Versification and Authorship Attribution
© Petr Plecháč, 2021

Introduction

Petr Plecháč
Institute of Czech Literature, Czech Academy of Sciences
e-mail: plechac@ucl.cas.cz

This work is licensed under a Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

<https://doi.org/10.14712/9788024648903.1>

Introduction

Contemporary stylometry is one of the fastest-growing fields in the computational study of literature. In recent years, a number of textual characteristics and machine learning techniques have proven highly accurate in distinguishing the texts of different authors. Many of these features like word and character n -gram frequencies amount, however, to what is known as statistical “rare events”, or more precisely, a large number of rare events (LNRE). As a result, their analysis calls for fairly large text samples consisting of thousands or tens of thousands of words. Most theoretical studies in stylometry therefore focus on long novels. Poetry is usually omitted although we might expect to find many more cases of disputed authorship among poetic works.

At the same time, poetry has a number of specific versification features that are essentially Boolean or open to only a limited number of values. Some of these—stanza length and rhyme scheme, for example—are subject to the author’s conscious selection and so unsuitable for authorship recognition. In contrast, others like the preference for certain rhythmic configurations or sound frequencies in rhyme may be outside the author’s rational control. Although these characteristics have traditionally been recognised as author-specific (or at least period-specific), they have rarely featured in authorship attribution studies.

The goal of this book is to examine the applicability of these versification features to authorship attribution projects. To this end, I draw on poetic corpora in three different languages (Czech, German and Spanish) and apply this approach to two real-world cases of disputed authorship.

Chapter 1 gives a brief history of quantitative methods of authorship attribution with special attention to the methods used in this book.

Chapter 2 highlights different ways to capture versification features.

Chapter 3 describes experiments with versification-based attribution and compares the methods commonly used in stylometry.

Finally, Chapter 4 applies these findings to two actual cases of ambiguous authorship involving English- and Russian-language texts respectively. In the first case,

I attempt to determine which parts of the verse play *The Two Noble Kinsmen* were written by William Shakespeare and which were the work of his co-author, John Fletcher. In the second, working together with Artjoms Šeļa, I investigate the potential forgery of numerous 19th-century Russian poems that were originally attributed to Gavriil Stepanovich Batenkov. These poems first appeared in the 1978 edition of the poet's collected works, which was compiled by an established literary scholar—the main suspect in this intrigue.

Previous Publications

Chapter 1 expands on the opening sections of *Versification and authorship attribution. Pilot study on Czech, German, Spanish, and English poetry* (Plecháč, Bobenhausen and Hammerich 2018).

Czech versions of Chapters 1, 2 and 3 were submitted as part of my PhD thesis at Charles University in Prague, Czech Republic in 2019.

Data and Code

The data and code required to reproduce the analyses in this book can be found at <https://doi.org/10.5281/zenodo.4555250>.

HTML version

From early 2022, this book will also be available online at <https://versologie.cz/versification-authorship>.