

GHOST IN THE ARCHIVE: USING COMPUTATIONAL STYLOMETRY TO RECOVER LOST VOICES IN DIGITIZED MARGINALIZED LITERATURE

Ilyar Kuzminsk

Staatliche Akademie der Bildenden Künste Stuttgart
Germany

Abstract

Traditional literary archives are structural artifacts of power, historical contingencies, and systematic biases, frequently erasing or silencing marginalized voices. While large-scale digitization efforts have successfully preserved vast physical corpora, simply converting paper to pixel does not solve the underlying exclusions of the archival record. Unattributed, pseudonymous, or collectively published prose often hides the idiosyncratic hand of minor writers, women, queer creators, and racialized individuals within dominant, mainstream structures. This review paper synthesizes the burgeoning cross-disciplinary field of computational stylometry as a radical method of archival recovery. By leveraging algorithmic pattern recognition—including lexical, syntactic, character n-gram, and advanced deep learning vector frameworks—scholars can extract unique stylistic markers to identify, classify, and re-attribute prose written by suppressed or anonymous authors within large unmapped datasets. Framed within structural digital humanities, poststructuralist archival theory, and the political economy of data, this article maps the technical workflow of algorithmic text attribution, analyzes its primary methodological and mathematical underpinnings, critically reviews established case studies, and outlines the deep ethical boundaries governing machine-assisted text reassignment. Ultimately, we argue that computational stylometry serves as an indispensable hermeneutic intervention, resurrecting the 'ghosts in the archive' to permanently correct and democratize the literary canon.

Keywords: Keywords: Computational Stylometry, Archival Recovery, Digital Humanities, Authorship Attribution, Marginalized Literature, Machine Learning, Postcolonial Archives.