



The banner features a row of six icons: a globe, a book, a handshake, a money bag with a Euro symbol, a scale of justice, and a bicycle. Below the icons, the text reads: **AIUCD 2021**, **DH per la società: e-guaglianza, partecipazione, diritti e valori nell'era digitale**, and **10° congresso annuale PISA 19-22 gennaio**. On the right side, a list of topics is displayed in colored text: **DIGITAL PUBLIC HUMANITIES** (red), **OPEN CULTURE** (orange), **RETI SOCIALI** (yellow), **TECH ECONOMY** (green), **E-PARTICIPATION** (blue), and **TECNOLOGIE ASSISTIVE** (purple). The background includes binary code and a classical building facade.

Versione PROVVISORIA del contributo presentato al Convegno Annuale

DISCLAIMER

Questa versione dell'abstract non è da considerarsi definitiva e viene pubblicata esclusivamente per facilitare la partecipazione del pubblico al convegno AIUCD 2021

Il Book of Abstract contenente le versioni definitive e dotato di ISBN sarà disponibile liberamente a partire dal 19 gennaio sul sito del convegno sotto licenza creative commons.

Who wrote the erotic novel *Josefine Mutzenbacher*? Speculations, theories, and stylometric analyses

Simone Rebora¹, Massimo Salgaro²

¹ Università degli Studi di Verona, Italy - simone.rebora@univr.it

² Università degli Studi di Verona, Italy - massimo.salgaro@univr.it

ABSTRACT

The paper proposes a series of stylometric analyses aimed at attributing the erotic novel *Josefine Mutzenbacher* (1906) to one of its possible authors: Felix Salten, Arthur Schnitzler, Hermann Bahr, Peter Altenberg, Hugo von Hofmannsthal, Ernst Klein, and Willi Handl. After reviewing the discussion among literary scholars, two analyses have been performed: one that combines 480 different methods to compare the seven candidate authors; one that verifies the attribution using the "impostors" method. Results show how the most probable author is Felix Salten, more commonly known for his children's book *Bambi*.

KEYWORDS

Josefine Mutzenbacher, Austrian literature, Stylometry, Authorship attribution.

1. INTRODUCTION

Josefine Mutzenbacher oder Die Geschichte einer Wienerischen Dirne von ihr selbst erzählt (engl. *Josefine Mutzenbacher* or the story of a Viennese whore as told by herself), published in Vienna in 1906, has surely constituted one of the most fascinating cases of authorship attribution in German literature. First of all, because this novel is "the only so-called pornographic novel by a German writer that has achieved the rank of world literature" (Wiener, 1990). This singular *Bildungsroman* (novel of development) depicts the childhood of the prostitute Mutzenbacher from the age of five to fourteenth years. Due to its massive success, it was followed by two sequels, widely translated, reprinted, and adapted into several films.

Secondly, because it has generally been attributed to Felix Salten, the author of the world-famous *Bambi*, but his authorship has never been proven. Other authors have been suspected of writing this erotic masterpiece. Among them is also Arthur Schnitzler who, in 1902, had published *Reigen*, a cycle of amorous adventures that caused scandal in the Viennese society. In addition, both *Reigen* and *Josephine Mutzenbacher* were published by the same publisher, Wiener Verlag. Some scholars maintain that this scandalous text was the result of a "bet" by the Jung-Wiener writers Hermann Bahr, Peter Altenberg, and Hugo von Hofmannsthal (Liebrand, 2006); others state that it was written by less-known writers such as Ernst Klein (Weichinger, 1991; Hall, 2015) and Willi Handl (Englisch, 1990); even the publisher Fritz Freund has been included in the list of suspects (Englisch, 1990; Schidrowitz, 1990).

Thirdly, this controversy on authorship attribution became a juridical case. In 1988, a German court rejected Salten's heirs claim for royalties, since the documents and witness accounts they delivered were unable to demonstrate without fail Salten's authorship of the *Mutzenbacher* novel (Farin, 1990).

This unsolved quest for the author of the novel is the starting point of our analysis.

2. STYLOMETRIC ANALYSIS

2.1 CORPUS CONSTRUCTION

In order to test all theories and speculations proposed so far, we built a corpus of digitized texts, by collecting material from multiple sources such as [Project Gutenberg](#), [Projekt Gutenberg-DE](#), and [Kolimo](#). In four cases, we manually scanned and OCR-ed documents using the *Transkribus* platform (Kahle et al., 2017): manual evaluation of randomly-selected passages showed a character-error rate of about 2%, much below the minimum threshold for reliable authorship attribution (Eder, 2012; Franzini et al., 2018).

To compare the candidates for authorship in a genre-coherent setup, we collected novels and short stories by seven authors: Felix Salten, Arthur Schnitzler, Hermann Bahr, Peter Altenberg, Hugo von Hofmannsthal, Ernst Klein, and Willi Handl. The only candidate we had to leave out of the analysis is Fritz Freund, of which no novels or short stories (but just some lyrics and a collection of letters) are available.

Finally, we normalized any possible typographic discrepancy in the texts by transforming into digraphs all Umlauts and the Eszett (ä -> ae; ö -> oe; ü -> ue; ß -> ss).

An overview on the composition of the corpus is provided by Table 1.

Author	Number of texts	Number of words	Source
Arthur Schnitzler	40	777,234	<i>Kolimo</i>
Felix Salten	12	421,000	<i>Projekt Gutenberg-DE</i>
Hugo von Hofmannsthal	18	203,656	<i>Kolimo</i>
Hermann Bahr	4	203,482	<i>Kolimo</i>
Ernst Klein	3	169,377	OCR
Peter Altenberg	195	92,276	<i>Kolimo</i>
Willi Handl	1	52,441	OCR

Table 1. Corpus composition¹

2.2 ANALYSIS SETUP

All stylometric analyses² were performed by using the functions of the R package *Stylo* (Eder et al., 2016), which enables an efficient application of Burrows’s Delta (Burrows, 2002) and its numerous variants. To maximize the reliability of results through multi-methodology (Juola, 2015), a total of 480 different analyses were performed on the basis of word frequency, modifying respectively:

- the number of most frequent words (MFW) to be taken into consideration (from 50 to 2,000, with 50-word steps);
- the measure to calculate the distances between texts (Burrows’s Delta, Cosine Delta, Eder’s Delta, and Canberra);
- the culling percentage (0%, 20%, and 50%), to tentatively counter-balance the possible influence of thematic similarities.

For each of the 480 analyses, then, a quality metrics was produced by adopting the “simple score” described by (Evert et al., 2017): this, in order to preserve the potential of multi-methodology, acknowledging at the same time the varying efficiency of different approaches.

Before running the analyses, all texts of the candidate authors were split into 5,000-word-long chunks (being it the minimum text length for reliable authorship attribution, cf. Eder, 2013): this produced between 10 and 155 measurements for each author. Of all these measurements (i.e. stylometric distances between Josefine Mutzenbacher and the other texts in the corpus), only the shortest three were compared for each candidate author. This approach (inspired by the “rolling.classify” function in the *Stylo* package) offered the possibility to identify with the maximum level of precision the passages with a style closest to that of the *Mutzenbacher*.

2.3 RESULTS

Results of the analyses are presented in Table 2, which shows the percentage of attributions for each candidate, balanced through the “simple score” quality metrics.

Author	Percentage of attributions
Felix Salten	78.41 %
Arthur Schnitzler	12.26 %
Ernst Klein	9.33 %
Peter Altenberg	0 %
Hermann Bahr	0 %
Willi Handl	0 %
Hugo von Hofmannsthal	0 %

Table 2. Final results of the analysis

¹ The text of *Josefine Mutzenbacher* was downloaded from [Project Gutenberg](https://www.gutenberg.org/).

² All scripts and corpora used for the analyses are available on [Github](https://github.com/).

Felix Salten clearly dominates in the attribution, followed at a long distance by Arthur Schnitzler and Ernst Klein. This stylometric analysis seems to confirm the dominant (but never-confirmed) theory among scholars, that the author of the scandalous *Mutzenbacher* is also the future inspirer of Disney’s children’s classic. This is confirmed also by a curious coincidence, when *Bambi* (in a separated analysis ran on the non-split corpus) clusters closely to the *Mutzenbacher* (see Figure 1).

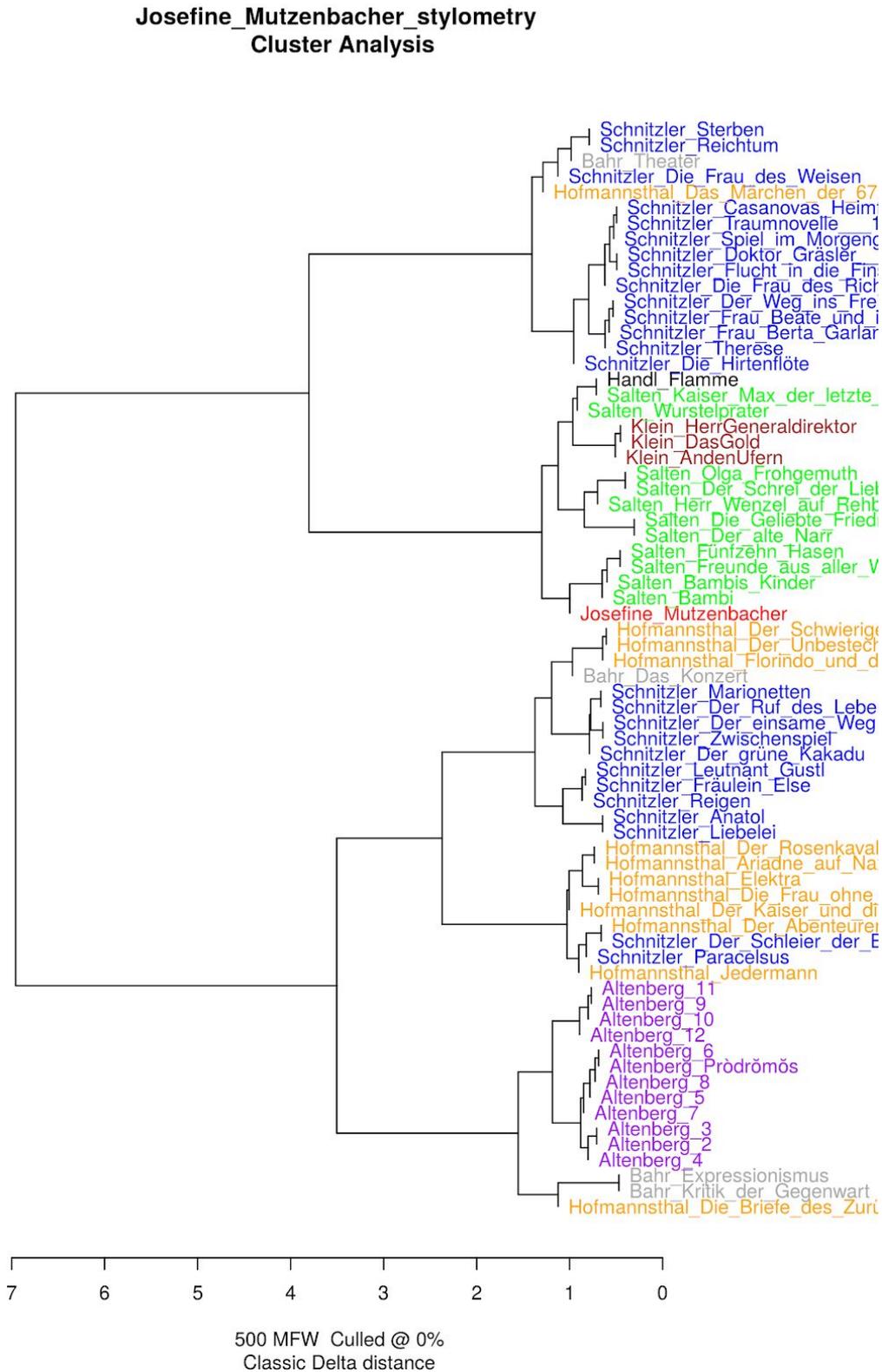


Figure 1. Dendrogram showing a close connection between *Bambi* and *Mutzenbacher*

However, a more detailed analysis of the results (see Figure 2) also shows how thin the attribution threshold can be. As good practice suggests in these cases, a thorough verification of the results becomes necessary.

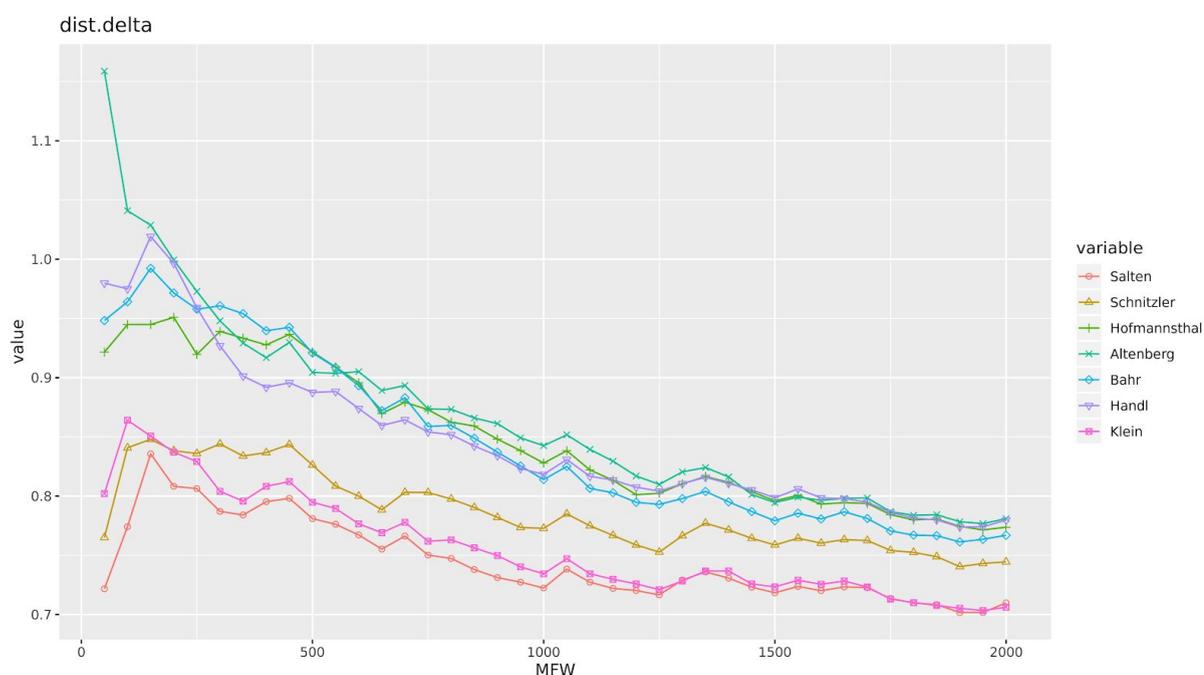


Figure 2. Distances between candidate authors and the Mutzenbacher (with Burrows’s Delta and 0% culling). Notice how the lines of Salten and Klein get closer for high MFW

3. AUTHORSHIP VERIFICATION

Validation of the results was performed via the “impostors” procedure, described by (Koppel and Winter, 2014) and recently implemented in the *Stylo* package (Eder, 2018). To determine the impostors (i.e. writers who cannot be the authors of the text under examination, but who can “disturb” the stylistic signal of the main candidate and thus confirm/deny the validity of the attribution), the entire *Kolimo* database was used, by selecting a total of 1,875 works published after the year 1880. Based on these works, 200 authors were selected that showed closest to the style of the *Mutzenbacher* (via a simple analysis with Cosine Delta, 2,000 MFW, and 0% culling) and compared with the seven candidate authors through the “impostors” function in *Stylo*. In this case, texts were not split into 5,000-word chunks, but preserved in their entirety. Texts shorter than 5,000 words were simply excluded from the analysis.

The results, shown in Table 3, confirm Salten’s authorship but also raise some doubts. A probability of 53%, in fact, is too close to the threshold of 50%, under which the impostors would have dominated over Salten. However, all the other candidate authors are entirely overshadowed by the impostors, thus clearly excluding that they could be the authors of the *Mutzenbacher*.

Author	Authorship probability
Felix Salten	53 %
Peter Altenberg	0 %
Hermann Bahr	0 %
Willi Handl	0 %
Hugo von Hofmannsthal	0 %
Ernst Klein	0 %
Arthur Schnitzler	0 %

Table 3. Authorship probability verified with the “impostors” method. Each percentage indicate the probability of the candidate author against the impostors

4. CONCLUSION

Through the combination of two stylometric approaches, our analysis confirmed how Felix Salten is the most probable author of the erotic novel *Josefine Mutzenbacher*. In methodological terms, we have shown how computational methods can be fruitfully paired with traditional research practices, by building on top of their critical theories and proposals; by testing, verifying, and eventually disproving them.

However, we cannot but consider the conclusion of this work as provisional. Future research should be devoted for example to finding ways to include the publisher Fritz Freund in the analysis; or to verify theories according to which the ending of the *Mutzenbacher* was written by a different hand (Schidrowitz, 1990; Englisch, 1990). All these possibilities can still open new scenarios for this attributive problem. Still, with our work for the first time we have provided scientific evidence in support of a hypothesis that, despite having been debated for a century, had never found an actual confirmation.

REFERENCES

- [1] Burrows, John. “Delta: A Measure of Stylistic Difference and a Guide to Likely Authorship.” *Literary and Linguistic Computing* 17, no. 3 (2002): 267–87. <https://doi.org/10.1093/lc/17.3.267>.
- [2] Eder, Maciej. “Authorship Verification with the Package Stylo.” *Computational Stylistics Group Blog*, 2018. <https://computationalstylistics.github.io/blog/imposters/>.
- [3] ———. “Does Size Matter? Authorship Attribution, Small Samples, Big Problem.” *Digital Scholarship in the Humanities* 30, no. 2 (November 14, 2013): 167–82. <https://doi.org/10.1093/lc/fqt066>.
- [4] ———. “Mind Your Corpus: Systematic Errors in Authorship Attribution.” In *Digital Humanities 2012: Conference Abstracts*, (Hamburg, Grmany), 181–85. Hamburg: Hamburg Univ. Press, 2012. https://sites.google.com/site/computationalstylistics/preprints/m-eder_mind_your_corpus.pdf?attredirects=0.
- [5] Eder, Maciej, Jan Rybicki, and Mike Kestemont. “Stylometry with R: A Package for Computational Text Analysis.” *The R Journal* 8, no. 1 (2016): 107–21.
- [6] Englisch, Paul. “Geschichte der erotischen Literatur (1927).” In *Josefine Mutzenbacher oder Die Geschichte einer wienerischen Dirne von ihr selbst erzählt*, by Felix Salten, 507–8. edited by Michael Farin, Ungekürzter Nachdr. der Erstausg. aus dem Jahr 1906. Die erotische Bibliothek. München: Schneekluth, 1990.
- [7] Evert, Stefan, Thomas Proisl, Fotis Jannidis, Isabella Reger, Steffen Pielström, Christof Schöch, and Thorsten Vitt. “Understanding and Explaining Delta Measures for Authorship Attribution.” *Digital Scholarship in the Humanities*, 2017. <https://doi.org/10.1093/lc/fqx023>.
- [8] Farin, Michael. “Die letzten Illusionen. Josefine Mutzenbacher vor Gericht.” In *Josefine Mutzenbacher oder Die Geschichte einer wienerischen Dirne von ihr selbst erzählt*, by Felix Salten, 511–45. edited by Michael Farin, Ungekürzter Nachdr. der Erstausg. aus dem Jahr 1906. Die erotische Bibliothek. München: Schneekluth, 1990.
- [9] Franzini, Greta, Mike Kestemont, Gabriela Rotari, Melina Jander, Jeremi K. Ochab, Emily Franzini, Joanna Byszuk, and Jan Rybicki. “Attributing Authorship in the Noisy Digitized Correspondence of Jacob and Wilhelm Grimm.” *Frontiers in Digital Humanities* 5 (2018). <https://doi.org/10.3389/fdigh.2018.00004>.
- [10] Hall, Murray. “Josef Und Josefine Mutzenbacher. Oder Recycling Der Pornographie.” In *Erotisch-Pornografische Lesestoffe: Das Geschäft Mit Erotik Und Pornografie Im Deutschen Sprachraum Vom 18. Jahrhundert Bis Zur Gegenwart*, edited by Christine Haug, Johannes Frimmel, Anke Vogel, Teresa Lang, and Franz Adam, 159–81. *Buchwissenschaftliche Beiträge, Band 88*. Wiesbaden: Harrassowitz Verlag, 2015. HQ472.G3 E76 2015.
- [11] Juola, Patrick. “The Rowling Case: A Proposed Standard Protocol for Authorship Attribution.” *Digital Scholarship in the Humanities* 30, no. suppl. 1 (2015): 100–113. <https://doi.org/10.1093/lc/fqv040>.
- [12] Kahle, P., S. Colutto, G. Hackl, and G. Mühlberger. “Transkribus - A Service Platform for Transcription, Recognition and Retrieval of Historical Documents.” In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 04:19–24, 2017. <https://doi.org/10.1109/ICDAR.2017.307>.
- [13] Koppel, Moshe, and Yaron Winter. “Determining If Two Documents Are Written by the Same Author: Determining If Two Documents Are Written by the Same Author.” *Journal of the Association for Information Science and Technology* 65, no. 1 (January 2014): 178–87. <https://doi.org/10.1002/asi.22954>.
- [14] Liebrand, Claudia. “‘Josefine Mutzenbacher’. Die Komödie Der Sexualität.” In *Felix Salten. Schriftsteller – Journalist – Exilant*, edited by Siegfried Mattl and Werner Michael Schwarzer, 82–89. Wien: Holzhausen, 2006.
- [15] Schidrowitz, Leo. “Bilderlexikon der Erotik (1929), Band 2, S. 692.” In *Josefine Mutzenbacher oder Die Geschichte einer wienerischen Dirne von ihr selbst erzählt*, by Felix Salten, 508–10. edited by Michael Farin, Ungekürzter Nachdr. der Erstausg. aus dem Jahr 1906. Die erotische Bibliothek. München: Schneekluth, 1990.
- [16] Weichinger, Robert. “‘Schluß Mit Genuß! Wer Schrieb Die ‘Mutzenbacher’? Wer Kriegt Die Tantiemen? Felix Saltens Erben Wohl Nicht.’” *Die Presse* 2/3 (1991): 6.
- [17] Wiener, Oswald. “Der obszöne Wortschatz Wiens. Beiträge zu einer Ädöologie des Wienerischen (1969).” In *Josefine Mutzenbacher oder Die Geschichte einer wienerischen Dirne von ihr selbst erzählt*, by Felix Salten, 361–462. edited by Michael Farin, Ungekürzter Nachdr. der Erstausg. aus dem Jahr 1906. Die erotische Bibliothek. München: Schneekluth, 1990.