



Jack Unterweger An authorship analysis of the notorious killer's autobiography “Fegefeuer oder die Reise ins Zuchthaus”

Karoline Marko ¹ and Alesia Locker ²

¹University of Graz, karoline.marko@uni-graz.at

²Royal Danish Defence College, ales@fak.dk

Abstract

Jack Unterweger, a notorious killer and celebrated author, was a highly skilled manipulator and was found more than once to have plagiarized poems and short stories (Herwig, 2022; Leake, 2010). Praised for his work as an author, and particularly for his autobiography “Fegefeuer oder die Reise ins Zuchthaus”, Unterweger was released from prison in 1990. Almost 40 years after the publication of his autobiography, rumors have begun circulating that Unterweger might not have written his autobiography by himself and that Sonja von Eisenstein, one of Unterweger’s benefactors and supporters, might have been involved in the writing of it (Herwig, 2022). Thus, based on a corpus of nine books by Unterweger and von Eisenstein, this study sets out to investigate whether these rumors are potentially true. The analysis is carried out with the help of a combination of quantitative (HCA, PCA) and qualitative (stylistic) methods. The analysis shows that Unterweger’s writing style is similar to that of his supposed autobiography, but that another individual’s writing style is also present.

Keywords: Principal component analysis, N-grams, Stylistics, Function words, Questioned authorship.

Resumo

Jack Unterweger, famoso assassino e célebre autor, era um manipulador muito habilidoso, tendo sido descoberto, mais do que uma vez, que plagiou vários poemas e contos. Elogiado pelo seu trabalho como autor e, em particular, pela sua autobiografia “Fegefeuer oder die Reise ins Zuchthaus”, Unterweger foi libertado da prisão em 1990. Quase 40 anos após a publicação da sua autobiografia, começaram a circular rumores de que Unterweger poderia não ter escrito a sua autobiografia sozinho e que Sonja von Eisenstein, uma das benfeitoras e apoiantes de Unterweger, poderia ter estado envolvida na sua redação.

*Assim, com base num corpus de nove livros de Unterweger e von Eisenstein, este estudo pretende investigar se estes rumores são potencialmente verdadeiros. A análise é efetuada com a ajuda de uma combinação de métodos quantitativos (HCA, PCA) e qualitativos (estilísticos). A análise mostra que o estilo de escrita de Unterweger é semelhante ao da sua suposta autobiografia, mas que o estilo de escrita de outro indivíduo também está presente. **Palavras-chave:** Análise de Componentes Principais, N-gramas, Estilística, Palavras funcionais, Autoria questionada.*

1. Introduction

Jack Unterweger was a notorious killer, malignant narcissist, psychopath (Haller in Leake, 2010; Widler, 2019), and celebrated author of a variety of books, plays, poems, and children's stories (Holzer & Reibenwein, 2022). Growing up in Wimitztal, he was a speaker of a Carinthian dialect, but he also gained rudimentary skills in English. In the late 1970s, he received a life sentence for the murder of an 18-year-old woman. During his 15 years in prison, Unterweger began to take writing classes and caught up on education he had missed¹. Unterweger also initiated a biannual journal called "Wortbrücke" (Wortbrücke, n.d.), which he edited and published while in prison (Leake, 2010). He reportedly had "the charm of a psychopath" (Haller in Leake, 2010, p. 401) and was a highly skillful manipulator. More than once, he was caught selling other people's work as his own, such as a poem by Herman Hesse and children's stories by Sonja von Eisenstein (Leake, 2010). Sonja von Eisenstein, who was an author herself and who initially fostered and encouraged Unterweger's education, was one of several benefactors he managed to attract the attention of during his time in prison (Herwig, 2022).

In the late 1980s, a petition was established to release Unterweger from prison as an example of successful rehabilitation (Holzer & Reibenwein, 2022). The petition fell on sympathetic ears and Unterweger was released in 1990 (Busch, 2019; Widler, 2019). Despite not having received any therapy during his time in prison, he was judged to have been completely rehabilitated (Leake, 2010). Upon his release, Unterweger began to work as a freelance journalist, gave readings of his books throughout Austria, and was a welcome guest on TV and various radio channels (Herwig, 2022; Leake, 2010). However, a series of murders of prostitutes in various parts of Austria that began shortly after his release started to attract the attention of some investigators. Because the *modus operandi*² in these murders was similar to the way Unterweger had killed his first victim, he soon became the prime suspect. In line with his psychopathic and narcissistic character traits, Unterweger, now working as a freelance journalist, began to report on the murders and even held interviews with the leading police officers to receive inside information on the status of the investigation (Leake, 2010). Upon finding out that he

¹It is reported that he graduated from 'Hauptschule' during his time in prison, a level of education usually obtained by 13-14 year-olds in Austria.

²The *modus operandi* he was known for is the strangling of his victims with pieces of their own clothing using a very specific and rare type of knot, and he used the same knot on himself when he committed suicide (Leake, 2010).

had been identified as the prime suspect in the killings, Unterweger fled to the United States but was soon extradited to Austria. In Graz, he was tried for the murder of eleven women and was convicted of the murder of nine of them³ in and around Austria as well as in the Los Angeles area. The conviction, however, was never spent, as Unterweger committed suicide in his cell after the announcement of the verdict (Busch, 2019; Leake, 2010).

Unterweger’s path to freedom was largely paved by the success of his 1983 autobiography called “Fegefeuer oder die Reise ins Zuchthaus”, which was also made into a movie in 1989 (Leake, 2010). Almost 40 years after the publication of the book, doubts about its authorship have arisen, and suspicions have emerged that Sonja von Eisenstein might have been involved in the writing of it (Herwig, 2022). As these doubts have so far largely been based on intuition and gut feeling, the present paper sets out to explore this issue from a more scientific standpoint and determine whether Unterweger’s autobiography is really his own.

1.1. Authorship Analysis

In order to investigate the proposed research question, this paper takes an authorship comparison approach. Forensic authorship comparison aims to analyze the writing styles of two or more authors with the objective of determining whether the authors of the investigated texts are likely to be the same or different (Ehrhardt, 2018). This can either be done largely qualitatively (e.g., Fobbe, 2020, 2022; Larner, 2014; Turell, 2010; Turell & Gavalda, 2013; Wright, 2013), largely quantitatively (e.g., Argamon, 2008, 2018; Argamon & Levitan, 2005; Belvisi, Muhammad, & Alonso-Fernandez, 2020; Johnson & Wright, 2014; Juola, 2021; Koppel, Schler, & Argamon, 2009, 2011; Stamatatos, 2009, 2013; Swain, Mishra, & Sindhu, 2017; Wright, 2017), or through a combination of both these approaches (Locker, 2019; Nini & Grant, 2013; Wright, 2013).

Qualitative approaches to authorship analysis focus on stylistic aspects of individuals’ writing styles, such as features relating to text format, the use of numbers and symbols, abbreviations, punctuation, capitalization (or lack thereof), spellings and misspellings, word formation, syntax, discourse, and mistakes and errors (McMenamin, 2002). Approaches based on a stylistic tradition, however, suffer from one disadvantage, namely that they are often considered to be too subjective (Solan, 2013). Quantitative approaches, on the other hand, are generally considered to be more objective compared to qualitative approaches, yet they frequently lack explicability (Locker, 2019). Such approaches consider the frequencies with which features appear rather than their actual use in context. It needs to be mentioned, however, that neither approach can definitively be said to be superior to the other (Ainsworth & Juola, 2018; Solan, 2013). In fact, the advantages of one method might be considered disadvantages of the other – thus, they are rather complementary, and the choice of one or the other as well as their application depends almost entirely on the situational and circumstantial elements of the respective case. That is to say, even though some automated approaches reportedly have low error rates (Koppel et al., 2009), their success rate largely depends upon the length of the available textual material (Eder, 2013, 2017). In contrast, qualitative approaches with

³Two of the bodies were too decayed, so not enough evidence was found to convict Unterweger (Leake, 2010)

detailed manual analyses are rather impractical to apply to long texts. Further, quantitative methods can miss qualitative differences and/or similarities, while qualitative methods can miss larger patterns or distributions of linguistic features (Solan, 2013). Therefore, we have chosen to apply a combination of both quantitative and qualitative methods to our dataset (see also Locker, 2019).

2. Data Methodology

The dataset for this analysis consists of five books by Jack Unterweger in addition to his supposed autobiography, and of three books by Sonja von Eisenstein. Prior to the analysis, these books were digitalized and subsequently proofread to correct OCR (Optical Character Recognition) errors, as these could add random noise to the dataset.

In total, the corpus comprises 103,606 words in the Unterweger corpus, 96,862 words in the von Eisenstein corpus, and 66,724 words of the Q-text, which all together amount to 267,192 words. Table 1 provides an overview of the dataset.

Table 1. Overview of the whole dataset.

Author	Date of Publication	Title	Short title for in-text references	Wordcount	Genre
Q	1983	Fegefeuer oder die Reise ins Zuchthaus	Fegefeuer	66,724	Autobiography
Unterweger	1982	Tobendes Ich	Tobendes Ich	6,023	Poems
Unterweger	1987	Reflexionen	Reflexionen	855	Poems
Unterweger	1988	Wenn Kinder Liebe leben	Kinder	28,303	Short stories for children
Unterweger	1990	Kerker	Kerker	49,752	Novel
Unterweger	1990	Mare adriatico	Mare	18,673	Novel
				103,606	
von Eisenstein	1992	Verschlafete eure Träume nicht	Verschlafete	51,605	Advice book
von Eisenstein	1993	Wenn die Seele Märchen erzählt	Seele	27,712	Fairy tales for children
von Eisenstein	2014	Ein Dickkopf mit zwei Flügeln	Dickkopf	17,545	Children's story
				96,862	

To establish the authorship of the supposed autobiography, a two-stage analysis was undertaken. Stage 1 involved the statistical identification of strong co-occurring lexico-grammatical patterns in each of the texts. Due to the length of the texts, we split the individual files into chunks of 5,000 words in order to test whether our model could identify linguistic features that differentiate the writing styles of the texts known to be written by the two authors before using the model on the text whose authorship is questionable. In our analysis, we focused on function words, which are described in more detail below. After presenting the statistical identification of patterns, we provide

a graphical visualization of the relationship between the texts based on the similarities identified. In Stage 2 we then conducted a qualitative inspection of the patterns identified in Stage 1 in order to interpret the functional and idiosyncratic bases underlying the identified patterns. To investigate the use of the identified features in context, we used the corpus program AntConc (Anthony, 2022) to generate concordance lines⁴, MaxQDA for the qualitative annotation of the texts, and #LancsBox (Brezina, Weill-Tessier, & McEnery, 2002) for the comparison of n-grams (lockgrams). This combination of quantitative and qualitative methods allows us to describe the author’s writing styles in great depth and to support our findings with linguistic theories.

2.1. N-Grams Function Words

N-grams are strings of either words or characters of varying length (Wright, 2017). For example, word 4-grams are strings of four consecutive words, and character 4-grams are strings of four consecutive characters. Even though it is still unclear why character n-grams appear to be successful in some authorship analyses (Kestemont, 2014), the successful use of word n-grams in authorship analyses can potentially be explained. It has been argued, for instance, that the linguistic input that individuals are exposed to differs from other individuals because of their social backgrounds, including their families, network of friends, the acquisition of first and additional languages, education, profession, hobbies, etc. (see, e.g., Ehrhardt, 2018; Nini, 2018b). Thus, due to their sociolinguistic histories (Grant & MacLeod, 2018), individuals will subconsciously be primed for the combination of particular words (i.e., collocations) if they occur with high frequencies in their linguistic surroundings (Larner, 2014; Wright, 2017).

It has been suggested that the “formulaic sequences” (Wray, 2002, p. 9) that individuals are exposed to are stored as “holistic units” (Larner, 2014, p. 2) in the minds of the speakers and can thus be useful indicators of an individual’s style. This results in what Coulthard (2004, 2007) has termed “preferred co-selections”: the potentially individuating preference for the co-selection of particular words, the combination of which can differentiate one person’s style from that of others. Evidence for this theory comes from different linguistic disciplines, such as psycholinguistics (Wray, 2002), sociolinguistics (Coulmas, 1979), corpus linguistics (Moon, 1998), and language acquisition (Peters, 1984). Such formulaic sequences become a part of an individual’s vocabulary, and, with that, a part of their idiolect (Coulthard, 2004). Schmitt, Grandage, and Adolphs (2004), for instance, say that

every person has their own unique idiolect made up of their personal repertoire of language, and as part of that idiolect, it seems reasonable to assume that they will also have their own unique store of formulaic sequences based on their own experience and language exposure. (p. 138)

In the present paper, we have chosen to focus on the use of word n-grams, because their use is explicable and can complement the qualitative analysis of texts. As Wright (2017) has pointed out, however, it is possible that different lengths of n-grams will have different success rates for different authors, and that for some authors, n-grams generally work better than for others.

⁴“A concordance is (...) a list of all of the occurrences of a particular search term in a corpus, presented within the context that they occur in” (Baker, 2006, p. 71).

Function words such as determiners, conjunctions, and prepositions have been found to be less topic-dependent than content words (i.e., nouns, verbs, adjectives, and adverbs) and can technically be considered unigrams. An individual's use of function words is more likely to be indicative of their writing style rather than the content of the texts they produce (e.g., Argamon & Levitan, 2005; Miranda García & Calle Martín, 2007; Pennebaker, 2011). Importantly, the use of function words is less likely to be consciously controllable by an individual, while content words seems to be reflected upon more by the author and chosen more deliberately (Argamon & Levitan, 2005). Further, function words particularly lend themselves to authorship analysis because they appear with high frequencies in language, which also makes them easily quantifiable (Kestemont, 2014). The inclusion of pronouns in the analysis, in contrast, is less widely agreed upon, as evidence has been found that some of them are highly influenced by topic and genre (Biber, Conrad, & Reppen, 2006; Burrows, 1987a; Herring & Paolillo, 2006; Hoover, 2004), which might make them less reliable as markers of authorship, even though evidence of linkage between the author's personality and pronoun choices also exists (Newman, Groom, Handelman, & Pennebaker, 2008; Pennebaker, 2011; Rude, Gortner, & Pennebaker, 2004).

The use of the relative occurrence of function words as input for statistical analysis is among the oldest and best-known computational authorship analysis methods. Studies that have used this technique to investigate authorship date back to 1984 (Mosteller & Wallace, 1984). An advantage of using function words is that they have very little semantic content, and therefore they are less dependent on the topic of the text. On the other hand, function words express grammatical relationships among the words within a sentence and can consequently give us some insight into the speaker's or writer's grammatical preferences. The use of function words is also grounded in functional linguistics (Biber & Conrad, 2009), which gives our analysis interpretability. The relative stability of the frequency of function words has also been demonstrated in other studies. An interesting study in this respect and one which resembles our case to some extent was undertaken by Binongo (2003). Binongo compared the styles of Frank Baum, the creator of the Land of Oz, to the style of his successor, Ruth Thompson, who continued the book series after Baum's death. Interestingly, although Thompson's early Oz novels are often considered faithful to Baum's canon, the Principal Component Analysis revealed that the styles of the two authors were very different in their use of function words.

2.2. Principal Component Analysis (PCA) Hierarchical Cluster Analysis (HCA)

Principal Component Analysis (PCA) is a statistical method that helps an analyst to "clarify a complex mass of data" (Grant & Baker, 2001, p. 77). The method was first adapted and introduced to authorship attribution tasks by Burrows (1987b, 1992). Through the extraction of correlations between linguistic features (Gries, 2022), PCA reduces the list of features for the comparison of potential authorship markers to only those markers that differ in the writing styles of the authors investigated (Grant & Baker, 2001). In this paper, PCA thus facilitates the identification of features in the writing styles of the two authors that separate them from each other before the document whose authorship is questionable is introduced into the analysis. Once the features that differentiate Unterweger's writing style from von Eisenstein's style have been identified, the document under investigation is added in order to determine whether the

writing style in Unterweger’s supposed autobiography is more similar to Unterweger’s style or to von Eisenstein’s style. PCA has previously been used by other scholars such as Holmes (1992), and Baayen, van Halteren, and Tweedie (1996) for authorship attribution as well.

A Hierarchical Cluster Analysis (HCA) is frequently used in identifying similarities and dissimilarities in large datasets. As Gries and Stefanowitsch (2010) point out:

this is done by (i) comparing pairwise (dis)similarities between the items in a (dis)similarity matrix, (ii) successively amalgamating all items into clusters such that the items within a cluster are highly dissimilar to each other and at the same time highly dissimilar from all other items and clusters, and (iii) representing the resulting structure in the form of a tree-like dendrogram (p. 77).

Both methods are utilized in the present analysis in order to obtain an overview of potential linguistic patterns in the text, which can be used as a starting point for further, more in-depth investigations.

2.3. Limitations

The main limitation of our dataset is that even though we have gathered and collected all the available books by the two authors under investigation, some books are out of print. Thus, the books that we include in the analysis are of different genres, which can pose a problem for the analysis (see, e.g., Overdorf & Greenstadt, 2016). Interestingly, however, even though the document examined in our study is officially an autobiography, it contains several elements of fiction, as it has been found that not all the reported events happened in the way they are portrayed in the narrative (e.g., Herwig, 2022). This makes the text under investigation similar to Unterweger’s known novels, which in turn also contain some biographical elements and references⁵. Also, the oldest book in the corpus dates from 1982, while the newest one dates from 2014, which represents a 32-year difference. This is particularly an issue for Sonja von Eisenstein, whose books date from 1992, 1993, and 2014 – the time span between the oldest and the newest book can potentially explain some intra-author differences. For Unterweger, the dating of the books is less of an issue, as they were all published within an eight-year period and during and shortly after his time in prison, which means that the situational aspects remain relatively constant. Lastly, even if the results of the analysis were to show some similarities to Sonja von Eisenstein’s writing style, the presence of a third possible author cannot entirely be ruled out (see also Grant & Baker, 2001), and it is also possible that other people, such as Unterweger’s teacher(s), had a strong influence in this writing. Similarly, it is possible, as with any book, that the texts have been proofread and edited by the publishers, which can also interfere with the author’s writing style (Fobbe, 2011). However, this is perceived to be less of an issue in the present case, as we believe that editing and proofreading was kept to a minimum with Unterweger’s texts⁶. This argument is supported by the presence of multiple spelling mistakes and grammatical errors in his books, as will be discussed in more detail in the qualitative analysis.

⁵In one interview, Unterweger even refers to them as sequels to the autobiography (Mose, 2015).

⁶In a podcast by Herwig (2022), one of Unterweger’s editors states that Unterweger’s style was retained in the books and that only minor changes were introduced by the editor.

3. Analysis

The analysis section is separated into three parts. First, we will provide some general information about the dataset before we proceed to an overview of the quantitative examination of the investigated material. Based on these findings, we have identified words and phrases of particular interest, which are examined in a follow-up qualitative analysis. The qualitative analysis also focuses on potentially idiolectal markers, as well as on mistakes and errors.

3.1. Stage 1: Pre-evaluation of the texts

First, we consider sentence and word lengths as well as sentence complexities measured using the Flesch Readability Ease score (Bachmann, 2014), which provides some indication of the complexity of the texts. Table 2 shows the results of this analysis:

Author	Short title	Readability Ease (RE)	Average word length	Average sentence length
Questioned	Fegefeuer	70	4.9	13.1
Unterweger	Tobendes Ich	n/a*	6.3	n/a*
Unterweger	Reflexionen	n/a*	5.5	n/a*
Unterweger	Kinder	69	4.8	15.1
Unterweger	Kerker	61	5.1	12.1
Unterweger	Mare	60	5.2	15.9
von Eisenstein	Verschlafet	46	5.4	18.4
von Eisenstein	Seele	71	4.8	13.9
von Eisenstein	Dickkopf	63	5.1	13.4

Table 2. Overview of Readability Ease (RE), average word lengths, and average sentence lengths. Low RE scores indicate higher complexities; higher scores indicate less complexity.

*For the collection of poems, no RE scores and average sentence lengths are provided, as the poems contain many sentences without punctuation. Imposing punctuation marks would skew the results to a similar degree to calculating the values and scores with the texts as they are.

As shown in Table 2, the RE score of almost all the books in the corpus is between 60 and 71, which indicates that the sentence structures in all of the texts are rather simple. The only exception is von Eisenstein's "Verschlafet eure Träume nicht", which, with a score of 46, is more complex than the others. No scores could be obtained for Unterweger's books of poems ("Tobendes Ich" and "Reflexionen"), as the lack of punctuation and incomplete sentences would have skewed the analysis. In terms of word length, both Unterweger's and von Eisenstein's fairy tales contain shorter words; von Eisenstein's children's stories also show decreased sentence lengths compared to her advice books, whereas Unterweger's fairy tales contain the third-longest sentences of his corpus, with sentences similar in length to "Mare", and it is the novel "Kerker" that contains the shortest sentences. After this initial evaluation of the texts, a quantitative examination was conducted.

3.2. Stage 2 – Quantitative examination

Our statistical analysis investigates frequency patterns of function words and their co-occurrence in the books of Jack Unterweger and Sonja von Eisenstein. Subsequently, the results of the PCA and HCA will be presented.

3.2.1. Pre-tests with Known Texts

The statistical AA was carried out in R (R Core Team, 2022), a programming language for statistical analysis, utilizing the Stylo package for computational stylistics (Eder, Rybicki, Kestemont, & Pielstrom, 2022). To identify the frequency patterns which distinguish the styles of Jack Unterweger and Sonja von Eisenstein, a list of the 200 most frequently occurring words in all of their known texts was created. In computational AA, the length of the feature list is not agreed upon, with research suggesting numbers from the 50 most frequent words to all the words in the corpus (see Stamatatos, 2009, pp. 540–541). The number in our study was chosen to include the majority of frequently occurring function words. It is noteworthy that AA based on frequency patterns is susceptible to the modelling of the topic, since many of the lexical features are highly topic dependent. The wordlist was reviewed to exclude content words and pronouns, as it has been demonstrated that the topics favored by authors can significantly influence the authorship attribution accuracy and result in artificially high scores (Layton, Watters, & Dazeley, 2015). The words which occur in fewer than 50% of the texts were excluded from the analysis, since sparse data would lead to poor statistics. The removal of the content words and rare words shortened our feature list to 114 words.

The statistical AA was carried out in R (R Core Team, 2022), a programming language for statistical analysis, utilizing the Stylo package for computational stylistics (Eder et al., 2022). To identify the frequency patterns which distinguish the styles of Jack Unterweger and Sonja von Eisenstein, a list of the 200 most frequently occurring words in all of their known texts was created. In computational AA, the length of the feature list is not agreed upon, with research suggesting numbers from the 50 most frequent words to all the words in the corpus (see Stamatatos, 2009, pp. 540–541). The number in our study was chosen to include the majority of frequently occurring function words. It is noteworthy that AA based on frequency patterns is susceptible to the modelling of the topic, since many of the lexical features are highly topic dependent. The wordlist was reviewed to exclude content words and pronouns, as it has been demonstrated that the topics favored by authors can significantly influence the authorship attribution accuracy and result in artificially high scores (Layton et al., 2015). The words which occur in fewer than 50% of the texts were excluded from the analysis, since sparse data would lead to poor statistics. The removal of the content words and rare words shortened our feature list to 114 words.

The length of the books allowed them to be split into blocks of 5000 words. This provided us with a substantial number of samples of undisputed authorship to serve as a baseline against which we could assess the success of the AA while still leaving enough data for the computational analysis to identify patterns in the stylistic preferences of Unterweger and von Eisenstein.

In order to further investigate the extent of the influence of genre, an analysis of the whole corpus was undertaken. The dendrogram in Figure 1 shows that Unterweger’s children’s stories are similar in the use of function words to von Eisenstein’s texts. As

mentioned before, there are rumors that some of these stories might have been stolen from von Eisenstein; however, to avoid exaggerating the implications of our findings, it has to be emphasized that the genre of children's stories could play a part in this categorization as it could, to a certain extent, affect the distribution of function words. Due to the potentially disputable authorship of "Wenn Kinder Liebe leben", we exclude this book from further analysis.

The rest of the books by Unterweger can be successfully clustered together despite differences in genre and topic. Even though "Verschlafet eure Träume nicht" by von Eisenstein is a self-help book aimed at the adult reader, it is revealed by the analysis to be more similar to the children's books by the same author than the books for adults by Unterweger.

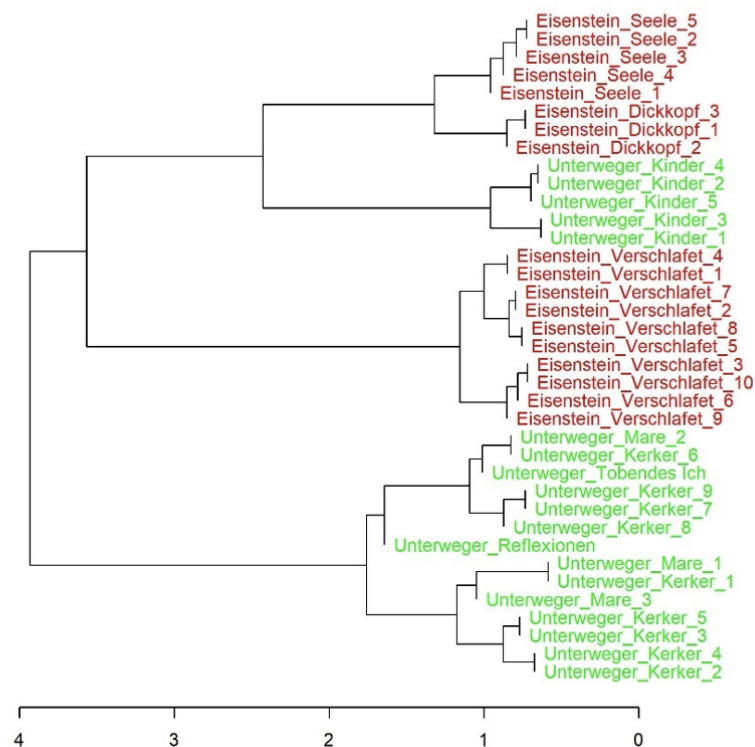


Figure 1. Cluster analysis of the known texts by Unterweger and von Eisenstein.

3.2.2. Adding the Q-Text

After having established that the list of 114 functional features distinguished Unterweger's writing style from von Eisenstein's style, the autobiography was added to the analysis. While the style of the books known to be written by Unterweger is very consistent, as indicated by the blending of the text blocks from several books (see Figure 2), it is remarkable that the autobiography forms a separate tight cluster of its own. This means that the frequency profiles of this text's blocks are consistently distinct from the books known to be by Unterweger, although they are still more similar to Unterweger's style than to that of von Eisenstein.

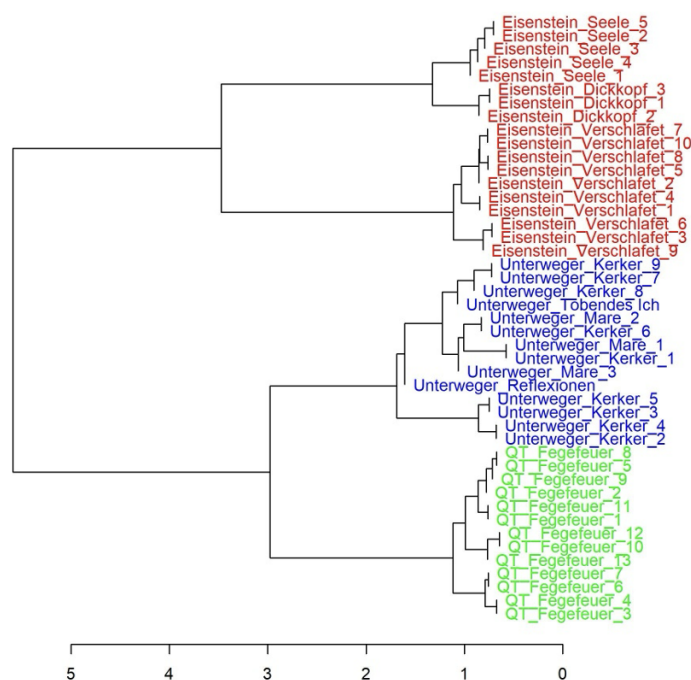


Figure 2. Cluster analysis: the questioned autobiography and the known texts by Unterweger and von Eisenstein.

Since we cannot exclude the possibility that Unterweger’s later books were influenced by authors other than himself, seven of his longer personal letters used in a different study (Marko & Leibetseder, 2023) were added to the analysis for the following reason: these private letters were addressed to the same person, and it is known that they were authored by Unterweger himself. Thus, this analysis aimed at finding out whether the books known to be by Unterweger or his supposed autobiography is more similar to the letters in style, or whether these letters would form a separate cluster like the text under investigation. The letters were not included in the initial analysis, because they are of a very different genre. However, since the cluster analysis did not identify either Unterweger or von Eisenstein as the author of the text under examination, and because we wanted to investigate whether this text or Unterweger’s known books were closer to his “true” writing style, we added the personal letters. Figure 3 illustrates that Unterweger’s letters cluster with his known texts. Despite being prose, the distribution of function words in his letters is more similar to Unterweger’s poem collection “Reflexionen” than to his supposed autobiography. This provides support for the hypothesis that the style of the autobiography is different from the style of his other books and personal letters.

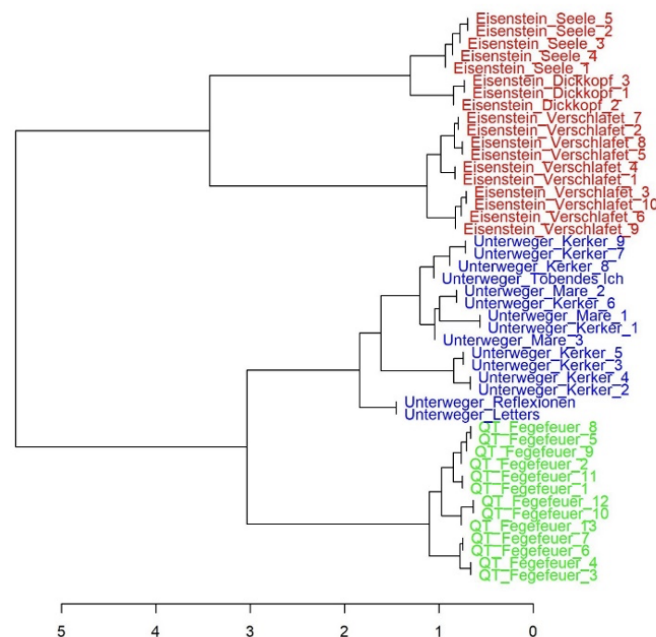


Figure 3. Dendrogram with personal letters.

To further understand which of the function words are most informative in differentiating between the styles of the authors, we refer to the exploratory statistical technique of PCA. This analysis boils down the frequencies of the 114 function words from our feature list to a single point on the plot that represents a specific text (see Figure 4a). The texts with similar frequency profiles are grouped together, while the loadings (see Figure 4b) help us more deeply examine how strongly specific function words contribute to this process and which function words frequently occur together in these groups. It is clearly evident that the analysis separates the authors quite well. The words with higher values in Figure 4b, positive or negative, have a larger effect on the results of the separation. Generally, the words on the right of this plot are the ones that are more indicative of Unterweger's writing style and the words on the left are characteristic of von Eisenstein's writing style.

“to”) all occur more frequently compared to the books by von Eisenstein. If we inspect the data, we notice that the relative frequencies of these features are higher in the supposed autobiography than in the texts known to be by Unterweger (see Table 3). Interestingly, a similarity between the autobiography and von Eisenstein’s style of children’s stories can be found in the frequent appearance of the conjunction “Und” (“And”) and adverb “Dann” (“Then”) at the beginning of the sentence, which does not seem to be the case for the texts known to be by Unterweger.

	Und	und	Dann	dann	weil	am	ins	zurück	beim	muß	mußte	wurde	wollte
Eisenstein_Dickkopf	0.60	2.99	0.16	0.18	0.04	0.11	0.05	0.05	0.03	0.00	0.00	0.16	0.11
Eisenstein_Seele	0.79	3.57	0.14	0.21	0.14	0.18	0.06	0.09	0.03	0.04	0.04	0.16	0.10
Eisenstein_Verschlafet	0.10	2.67	0.07	0.34	0.10	0.14	0.06	0.07	0.03	0.08	0.07	0.11	0.11
QT_Fegefeuer	0.37	2.75	0.18	0.37	0.25	0.31	0.30	0.20	0.13	0.07	0.30	0.26	0.30
Unterweger_Kerker	0.15	3.58	0.04	0.26	0.17	0.15	0.14	0.13	0.10	0.13	0.07	0.11	0.10
Unterweger_Mare	0.24	3.48	0.05	0.22	0.19	0.18	0.17	0.09	0.07	0.12	0.13	0.12	0.14
Unterweger_Tobendes Ich	0.15	3.20	0.08	0.42	0.13	0.30	0.13	0.17	0.10	0.08	0.03	0.12	0.05
Unterweger_Reflexionen	0.00	3.17	0.00	0.23	0.12	0.12	0.00	0.35	0.00	0.23	0.00	0.00	0.12

Table 3. Characteristic features of the autobiography and their relative frequencies.

On its own, each of these features may have little significance for AA, but their combined weight pulls the autobiography into a separate cluster in the plots. While it is difficult to make any claims about authorship at this stage, we can argue that the autobiography, for whatever reason, was not written in the same style as Unterweger’s other books. Subsequently, we will move on to the qualitative part of the analysis.

3.3. Stage 3 – Qualitative description

The qualitative analysis presented below proceeds as outlined in Fobbe (2021), and thus begins with an error analysis followed by a stylistic analysis.

3.3.1. Error Analysis

As mentioned previously, several mistakes were identified in the texts. These errors are largely orthographical, but the texts also contain some syntactic mistakes and some relating to idiomatic language use. This indicates only minimal interference of an editor. Sonja von Eisenstein’s books, in contrast, do not contain any of these errors or issues with idiomatic language use. Some examples of Unterweger’s texts are shown in Table 4.

	German – Error	German - Corrected	English translation	Source Text
1	Wörthsee	Wörthersee	Lake Wörth	Fegefeuer (QT)
2	Fibrieren	Vibriieren	Vibrate	Fegefeuer (QT)
3	Fridhof	Friedhof	Cemetery	Fegefeuer (QT)
4	Radstatt	Radstadt	<i>Name of a city</i>	Fegefeuer (QT)
5	Steh’ns auf	Stehn’S auf	Stand up	Fegefeuer (QT)
6	Artbeitskittel	Arbeitskittel	Work coat	Kerker
7	Agression	Aggression	Aggression	Kerker

Table 4. Examples of orthographical errors in Unterweger’s books.

Table 4 above shows orthographical errors, but, as mentioned, other mistakes were also identified, for example in relation to idiomatic language use. These mistakes might point towards Unterweger’s basic level of education and his relative inexperience with written language up to the point at which he decided to become an author. Examples of his unidiomatic use of words and expressions can be found in Table 5 below:

	Example	Intended	Translation (literal)	Source Text
8	Minute auf Minute wird Vergangenheit		Minute after minute becomes the past	Kerker
9	steirischer Österreicher		Styrian Austrian	Kerker
10	schönwörtlicher Neubeginn	Sprichwörtlicher Neubeginn	Proverbial fresh start	Kerker
11	Vorsichtsabstand	Sicherheitsabstand	Safe distance	Kerker & Mare
12	ich rede Worte		I talk words	Mare
13	hassende Wut		Hating anger	Kerker & Mare

Table 5. Examples of unidiomatic language use in Unterweger’s known texts.

Similar salient words and expressions can be found in Unterweger’s supposed autobiography (see Table 6), but they do not appear in von Eisenstein’s book.

	Example	Intended	Translation (literal)	Source Text
14	In dieser Nacht <u>erbrach</u> ich alle Geschäftsstände (...)	In dieser Nacht brach ich in alle Geschäftsstände (...)	In this night, I <u>vomited</u> all of the booths (...) [intended: broke into]	Fegefeuer (QT)
15	Versprechungen waren <u>zerronnen</u>		Promises melted away	Fegefeuer (QT)
16	<u>hämmernde Vergewaltigung</u>		Hammering rape	Fegefeuer (QT)
17	<u>Rauflustigkeit</u>	Rauflust	Rowdiness	Fegefeuer (QT)

Table 6. Examples of unidiomatic language use in the autobiography.

After this overview of errors and mistakes, we will subsequently examine the stylistic features detected using PCA and HCA in more detail, in order to find out if there is an explanation for their use in the text under investigation.

3.3.2. Stylistic analysis

The next step in the qualitative part of the analysis is based on the PCA and HCA described above. As mentioned, the features that were significantly more common in Unterweger’s supposed biography than in his known texts are the words “zum” (“to”), “weil” (“because”), “Und” (“And”), “ins” (“into”), “beim” (“with”), and “am” (“on”), which are prepositions and conjunctions. These features were investigated with AntConc, i.e., concordance lines were created to see how the words are used in context in order to find an explanation for why they appear so much more frequently in the autobiography

compared to Unterweger's known texts. Table 7 shows an example of the concordance lines for the preposition "zum" ("to"):

#	File		KWIC	
182	Q.txt	Ich schwieg und fuhr mit ihm	zum	Bezirksgericht der Stadt. "Wir hatten Glück, der
		I kept silent and drove him	to	the circuit court of the city. "We were lucky, the
183	Q.txt	Kaffee. "Ist er das?" sagte der Mann	zum	Amtsrat. "Stimmt, Herr Rat." Der zweite setzte
		coffee. "Is that him?" said the man	to	the councilor. "Yes, Mr. Councilor." The second sat
184	Q.txt	und Jeans, kichernd und flüsternd, trippelnd vorbei	zum	Greifen nahe und doch unerreichbar.
		and jeans, giggling and whisperingly, tripping past	within	reach and yet unattainable.
185	Q.txt	der Küche haben sie sicher noch was	zum	Essen", sagte der Cheferzieher und schob mich
		the kitchen, you sure have something	to	eat", said the teacher and pushed me
186	Q.txt	die Küche, die geben ihm noch was	zum	Essen", sagte der älteste, ruhigste, kleine,
		the kitchen, they give him something	to	eat", said the oldest, quietest, short
187	Q.txt	älteste, ruhigste, kleine schnauzbärtige Erzieher vom Dienst	zum	Blondgelockten. Ich müßte vor ihm gehen,
		oldest, quietest, short mustachioed teacher	to	the one with the blond curls. I would have to walk in front of him,
188	Q.txt	sagte mir der Erzieher, "da hast Zeit	zum	Nachdenken!" Ich dachte nach, in meinem Kopf
		said the teacher to me, "there you will have time	to	ponder!" I thought, in my head
189	Q.txt	hat... In der zweiten Woche kam ich	zum	Arzt. Nachdem ich eine gute Stunde
		had... in the second week I came	to	the doctor. After an hour
190	Q.txt	ich als gesund registriert. Anschließend kam ich	zum	Psychologen. Zusammen mit sieben anderen
		I was registered as healthy. Afterwards I came	to	the psychologist. Together with seven others
191	Q.txt	hat, brauchen wir drei neue Häuser." Und	zum	Vorführbeamten: "Die haben ihm drei Tage auf
		has, we need three new houses". And	to	the clerk: "they gave him three days

Table 7. Examples of the use of the preposition "zum" ("to"), original and translation.

After the extraction of concordance lines, the concordance lines were imported into MaxQDA, which allows for the easy coding of datasets. Using MaxQDA, we attempted

to identify patterns in the use of these words which would help us find an explanation for their frequent occurrence in the text under investigation. The patterns that were found are related to physical localizations and movements, the description of moods and states, and to temporal information. Many of the prepositions are used with nominalized constructions, which suggests that the writing style employed in the autobiography is more nominal compared to Unterweger’s known texts, at least in respect to the description of movements. The difference in the usage of these words across the autobiography and known texts is of interest insofar that all three of the texts (the autobiography and the novel-like books) focus on travelling, movement and people. Thus, the fact they occur significantly more frequently in the autobiography is a remarkable stylistic difference. Whether or not this difference is due to differences in authorship is yet unclear.

Another interesting finding was discovered in relation to the conjunction “Und” (“And”) with a capital letter, as has been pointed out above. The spelling of this word with a capital letter indicates that it is used sentence-initially. A close examination of the texts further reveals that it is used for the incorporation of direct speech (i.e., turn-taking sequences in conversations) in the autobiography, which apparently is not the case in the texts known to be by Unterweger.

A close reading of the texts has further shown that Unterweger has a very distinctive writing style characterized by words he creates on the fly, such as compounds (mostly nouns but also adjectives), and he uses nominalizations that are not part of standard German language. Examples can be seen in Table 8 below and are also provided in Table 4 above. In order to establish how common these words are in a general reference corpus, the DWDS 1900-1999 (DWDS Kernkorpus, 2023), which comprises 121,494,429 words from a variety of genres, was consulted. A corpus search revealed that Unterweger’s word creations are indeed very rare⁷ in the corpus, if they appear at all. While these words serve as illustrations for Unterweger’s creative language use, this does not mean that these exact words appear in more than one of the books. If they appeared in more than one of his writings but rarely occurred in the reference corpus, this would lend support to the idea that he is indeed the author of all his books. Thus, in order to investigate whether any of his salient uncommon words appear in the autobiography and at least one other book, a 1-gram analysis with #LancsBox was undertaken (see below).

	German word	English translation (literal)	Occurrences in the DWDS reference corpus
18	Antwortgebung	Answerprovision	0
19	Blondhaar	Blondhair	49 (i.e., 0.4 times per million words)
20	Augenkrieg	Eyewar	0
21	Freundenmorgen	Delightmorning	0
22	Einsammachung	lonely-making	0

Table 8. Examples of salient uncommon words in Unterweger’s writings.

⁷(Nini, 2018a, p. 628) refers to frequencies of 10-18 times per million words as “quite rare”.

As mentioned, nominalizations such as those in Table 8 are frequently found in Unterweger's texts. In addition to nominalizations of compounds, temporal nominalizations of adverbs such as "das Jetzt" ("the now"), "das Gestern" ("the yesterday"), and "das Irgendwann" ("the anytime"), which provide the text with the air of a literary register, are also present. In addition, nominalized constructions using verbs and adjectives, such as "ein Verjagtwerden" ("the chasing away"), "die Hinterfragung" ("the scrutinization"), and "Herbstlichkeit" ("autumn-ity") can be detected.

Before moving to the 1-gram analysis of uncommon words that appear in more than one of Unterweger's books, it has to be pointed out that the qualitative analysis shows a large lexical overlap between "Kerker" and "Mare", which, upon closer reading, can easily be explained: large parts of these two books are identical with only slight modifications, and can thus be considered as instances of self-plagiarism. Based on this finding, the lexical similarities between these two texts must not be overrated; this, however, does not affect the comparison of the autobiography and Unterweger's known texts. The following examples in Table 9 serve as an illustration:

	Mare	Kerker	Translation
23	Ich war dreiundzwanzig und gierte nach Leben, Lust. Hot-Love! In den Gedanken: Aggression , Wut, Ohnmacht und irreale Allmacht . Das Erwachen, vermischt mit wilden Gefühlen, in der Zeit der Morgendämmerung. An diesem Morgen, dieser späten Nachtstunde, kreischten meine Nerven wie ein rostiges Sägeblatt im nassen Buchenholz und verhinderten ein weiteres Bettspiel. Später ertönte die Sirene dreimal, grell, Gänsehaut erzeugend und befehlend: Tagwache!	Ich war 23 und gierte nach Leben, Lust, Hot Love! In den Gedanken: Agression [sic], Wut, Ohnmacht und Allmacht . Das Erwachen vermischt mit wilden Gefühlen in der Zeit zur Morgendämmerung. An diesem Morgen, dieser späten Nachtstunde, kreischten meine Nerven und verhinderten ein weiteres Bettspiel. Später ertönte die Sirene dreimal, grell, Gänsehaut erzeugend und befehlend: Tagwache!	I was 23 and craved for life, passion, hot love! In my mind: aggression, anger, impuissance, and omnipotence. The awakening mixed with wild feelings during dawn. On this morning, at this late time of night, my nerves screeched and prevented another bed game. Later, the siren sounded three times, gratingly, giving me the shivers, and commanding: a wake-up call!
24	Für mich war es wieder einmal ein letztes Mal . Ich wartete auf meine siebente Entlassung aus einem Gefängnis.	Für mich wieder einmal zum letzten Mal . Für diesmal . Ich erwartete meine sechste Entlassung , (...)	For me, it was another last time. For this time. I was awaiting my sixth release, (...)

Table 9. Examples of self-plagiarism in "Kerker" and "Mare adriatico"

As mentioned, we will now consider which salient uncommon words appear in the autobiography and in at least one of the books known to be written by Unterweger.

First of all, the use of the adjective “berockt” (“skirted”, as in “dressed in a skirt”) in the autobiography and both of Unterweger’s other books is of interest. A search in the DWDS corpus reveals only two hits (0.016 times per million words) of this word, which indicates that it is very rare indeed. Similarly, the adjective “tränenblind” (“tearblind”), which only has three hits (0.02 times per millions words) in the DWDS corpus, appears twice in the autobiography and once in “Kerker”; “Vergangenheitsbilder” (“pastimages”) only has one hit (0.008 times per million words), “Erinnerungsbilder” (“memoryimages”) has no hits in the DWDS, and “schmutziggrau” (“dirtygray”) has 20 hits (i.e., 0.16 per million words), but all of them appear in both the autobiography and in at least one of Unterweger’s known texts. The same is true for the expressions “jovialer Ton” (“jovial tone”), which has only one hit (0.008 times per million words) in the DWDS, and “Au-Schrei” (“ouch-cry”) and “Au-Rufe” (“ouch-call”), which do not appear in the DWDS at all, appear in the autobiography and at least one of Unterweger’s known texts. Importantly, none of these appear in von Eisenstein’s book. The appearance of all these lexical items in the autobiography and at least one of Unterweger’s known texts provides evidence for Unterweger being the author of all of them.

As with the use of nouns and compounds, Unterweger is similarly creative in his use and creation of adjectives, many of which end in -los (-less), and most of which are also not part of the standard vocabulary of German. Examples thereof can be seen in Table 10 below:

	German word	English translation (literal)	Occurrences in the DWDS reference corpus
25	blicklos	Gazeless	68 (0.6 per million words)
26	fluchtlos	flightless	0
27	gegenwartslos	presentless	2 (0.016 per million words)
28	kollegenlos	colleagueless	0

Table 10. Examples of adjectives ending in -los (-less) in Unterweger’s books.

Additionally, Unterweger’s use of color terms is a salient feature of his writing style. He appears to have a preference for the combination of two colors, or a color plus another adjective, but, as he does with compound nouns, he combines these words largely without hyphens. To conclude this part of the stylistic analysis, it is remarkable that Unterweger employs a very nominal writing style generally characterized by nominalizations, compounds, attributions, subordinate attributions, and extended participial attributes (Hennig, 2020), as well as a high level of creativity.

3.3.2.1. Lockgrams

The stylistic analysis in the previous section can be complemented by an examination of lockgrams in order to investigate potential overlap between n-grams of different lengths across all the books in the corpus. Using the corpus program #LancsBox, lists of n-grams were generated and compared among the books. As mentioned before, Lockgrams are n-grams that appear with similar frequencies in the investigated texts (Brezina, 2018, p. 79-87), which means that they might indicate common authorship. The resulting list of lockgrams for each of the books was then checked in the DWDS to

determine how frequent they are in the reference corpus. An overview of the number of lockgrams in the books is provided in Table 11.

First, the autobiography and Sonja von Eisenstein's book were compared. Although 800 words that appear in the autobiography also appear in von Eisenstein's book, none of these are rare, salient, or non-standard (unlike Unterweger's non-standard words, as discussed previously). However, 2,221 2-grams from the Q text, 869 3-grams, 74 4-grams, nine 5-grams, and even one 6-gram appear in the autobiography and von Eisenstein's book. An investigation of these word sequences in the DWDS reveals only two hits for two of the 5-grams (i.e., 0.016 per million words).

	1-grams	2-grams	3-grams	4-grams	5-grams	6-grams	7-grams
Fegefeuer + Verschlafet (von Eisenstein)	800	2,221	869	74	9	1	-
Fegefeuer + Kerker (Unterweger)	1,314	4,399	1,447	237	46	7	1
Fegefeuer + Mare (Unterweger)	332	517	74	1	-	-	-

Table 11. Lockgrams in the dataset. The following settings in #LancsBox were used: relative frequencies, dispersion = Coefficient of Variation (CV), and Simple Math statistic.

The overlap between Unterweger's books and the autobiography in terms of n-grams is as follows: similarities between the autobiography and "Kerker" are more extensive than between the autobiography and "Mare"; however, as discussed above, some 1-grams are particularly uncommon in the reference corpus but appear in the autobiography and in Unterweger's known texts. The n-gram analysis is of importance insofar that it points towards the possibility of von Eisenstein having had some influence (be it active or passive) on the production of Unterweger's autobiography. The implications of these findings will be discussed below.

4. Discussion

This paper has attempted to investigate the questioned authorship of Jack Unterweger's autobiography. We were able to show that although Unterweger and von Eisenstein have different writing styles, the writing styles in the text under investigation is different from Unterweger's and von Eisenstein's other books, as it forms a separate cluster in the PCA and HCA. Since Unterweger's novels and poems cluster together, and his supposed autobiography is similar in genre to the novels, an influence of genre on the analysis can largely be excluded. Based on this, we attempted to identify features that make the autobiography different from texts known to be written by a particular author. Even though individual words were identified, no explanations for their appearance in the autobiography were found, other than that they contribute to the nominal writing style, which might be related to Unterweger's attempts to sound more formal and professional, as will be discussed below.

An influence of Sonja von Eisenstein in the use of n-grams might be indicated by the lockgram analysis. This analysis reveals several longer strings of words that appear

in von Eisenstein’s text and in the autobiography which do not appear in the reference corpus. The nature of Sonja von Eisenstein’s influence on the production of the autobiography remains unclear; it is also possible that the elements that point towards her involvement might have come from passive influence or from collaboration. It is known that she was an early supporter of Unterweger, and thus the close interaction between them might have left some traces in Unterweger’s own writing style. On the other hand, it is also known that Unterweger took some of her children’s stories and sold them as his own; consequently, some of her wording might have been integrated into his autobiography as well. Lastly, it is also a possibility that Sonja von Eisenstein played a role in the editing or revision of the book, which would also explain the presence of some of her stylistic features in the book. The results, however, remain inconclusive with regard to her involvement.

Our qualitative analysis has revealed that Unterweger employs a very distinctive writing style characterized by the creative use of words, particularly with respect to nouns and adjectives. This style is consistent across his novel-like books, and it can also be found in the autobiography. Examples illustrate his linguistic creativity in terms of non-standard words and expressions, colloquial, and informal expressions. In many instances, these words and expressions collide with an otherwise formal, literary, or intellectual style typified by a strong tendency towards nominalizations, which is generally characteristic of German “*officialese*” (Hennig, 2020). The incorrect use of some words, the use of unidiomatic expressions, and the preference for a nominal style might therefore indicate an attempt at appearing more educated than Unterweger actually was. This strategy might be considered a form of disguise, which is difficult to maintain in a longer text and often results in errors (Dern, 2008), such as those found in Unterweger’s books.

Based on our analyses, we conclude that the probability that Unterweger’s books and his supposed autobiography were produced independently is rather small. The lexical similarities provide strong evidence that Unterweger is indeed the (main) author of his autobiography. However, the autobiography still forms a neat separate cluster in the PCA and HCA, which indicates some stylistic differences. The nature of these differences cannot be conclusively explained, but a possible explanation is that Unterweger’s writing style had changed in the seven years between the publication of his autobiography in 1983 and his other novels in 1990. This difference could possibly be accounted for by the education he obtained while in prison, which included several writing classes. Moreover, it is also a possibility that, given that the autobiography was one of his first books (and definitely his most important), he put more effort into it than into the other books which were published when he had already become well-known. On the other hand, his poems, published in 1982, a year before his autobiography, are more stylistically similar to his later novels despite being of a different genre – a fact that again highlights that the autobiography is different.

Another promising direction for further research is the use of pronouns. As mentioned, all pronouns were removed from the analysis prior to testing, since they can be highly register and topic dependent and thus lead to biased results. However, since the use of pronouns can partially reveal people’s conceptualization of themselves and others (e.g., Orvell, Gelman, & Kross, 2022), they might be particularly interesting in the present context, as Unterweger was diagnosed with malignant narcissism and psy-

chopathy (Leake, 2010; Widler, 2019; see also Marko & Leibetseder, 2023). Additionally, the use of modal verbs and the perspectives taken in the books (past tense vs. present tense) appear to provide interesting grounds for future research.

5. Conclusions

This paper aimed at investigating whether or not there is linguistic evidence for the hypothesis that Jack Unterweger is not the (sole) author of his well-known autobiography “Fegefeuer oder die Reise ins Zuchthaus”. By using a combination of quantitative and qualitative methods, we were able to show that while differences between Unterweger’s other books and his autobiography do exist, many similarities can also be detected. The quantitative investigation has suggested stylistic differences between the autobiography and Unterweger’s novel-like books; the qualitative analysis has revealed lexical similarities, the latter of which provides support for the argument that Unterweger did in fact write the autobiography. The differences in style might be explained by the time that had passed between the writing of the autobiography and the writing of the other books. This finding, however, raises important (forensic) linguistic questions such as: how much does (or can) a person’s writing style change over a period of seven years? How much can writing classes (or other types of education) influence a native speaker’s language competence within seven years? The answers to these questions have practical consequences in terms of the comparability of texts produced several years apart. The analyses in this paper have also shown that the inclusion of pronouns in the analysis, even though they are often dependent on the genre, might actually be useful for certain authors and certain writing styles. In this specific context, it might even be argued that the inclusion of pronouns would be highly valuable, since Unterweger was diagnosed with psychopathy and malignant narcissism (Leake, 2010; Widler, 2019) – two psychological disorders that have been shown to be reflected in general language use (e.g., Fine, 2006; Viding, 2019), and in pronoun usage (e.g., Pennebaker, 2011) in particular. Therefore, another important avenue for future research in authorship analysis is how psychological disorders are reflected in written language, whether or not psychopathic individuals with high manipulative skills are also adept at manipulating their written language (e.g., in terms of adapting to others’ styles), and how pronoun use is influenced by genre, topic, audience, and psychological disorders.

To conclude, Unterweger’s writing style is characterized by distinctive and salient configurations of register, words and expressions. Whether or not his style can be considered “good” or “literary” is beyond the scope of this paper, but as Bandelow (2013, p. 24, translation) puts it, it is unlikely that “his [Jack Unterweger’s] books would [...] ever have had the slightest chance hadn’t there been the dark side to them that obviously bewitched the public.”

References

- Ainsworth, J., & Juola, P. (2018). Who Wrote This?: Modern Forensic Authorship Analysis as a Model for Valid Forensic Science. *Washington University Law Review*, 96(5). Retrieved from https://openscholarship.wustl.edu/law_lawreview/vol96/iss5/10
- Anthony, L. (2022). *AntConc*. Waseda University. Retrieved from <https://www.laurenceanthony.net/software/antconc/>
- Argamon, S. (2008). Interpreting Burrows’s Delta: Geometric and Probabilistic Foundations. *Literary and Linguistic Computing*, 23(2), 131–147. Retrieved 2024-03-02, from <https://academic.oup.com/dsh/article-lookup/doi/10.1093/llc/fqn003> doi: 10.1093/llc/fqn003
- Argamon, S. (2018). Computational Forensic Authorship Analysis: Promises and Pitfalls. *Language and Law / Linguagem e Direito*, 5(2), 7–37.
- Argamon, S., & Levitan, S. (2005). Measuring the Usefulness of Function Words for Authorship Attribution. In *Proceedings of the Joint Conference of the Association for Computers and the Humanities and the Association for Literary and Linguistic Computing* (pp. 1–3).
- Baayen, H., van Halteren, H., & Tweedie, F. (1996). Outside the Cave of Shadows: Using Syntactic Annotation to Enhance Authorship Attribution. *Literary and Linguistic Computing*, 11(3), 121–131.
- Bachmann, C. (2014). *Leichtlesbar.ch*. Retrieved from <http://www.leichtlesbar.ch/html/>
- Baker, P. (2006). *Using corpora in discourse analysis*. London & New York: Bloomsbury.
- Bandelow, B. (2013). *Wer hat Angst vorm bösen Mann?* Hamburg.
- Belvisi, N. M. S., Muhammad, N., & Alonso-Fernandez, F. (2020). Forensic Authorship Analysis of Microblogging Texts Using N-Grams and Stylometric Features. , 1–6.
- Biber, D., & Conrad, S. (2009). *Register, genre and style*. Cambridge: CUP.
- Biber, D., Conrad, S., & Reppen, R. (2006). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Binongo, J. (2003). Who wrote the 15th Book of Oz? An application of multivariate analysis to authorship attribution. *Chance*, 16(2), 9–17.
- Brezina, V. (2018). *Statistics in corpus linguistics. A practical guide*. Cambridge: Cambridge University Press.
- Brezina, V., Weill-Tessier, P., & McEnery, A. (2002). *#LancsBox*. Retrieved from <http://corpora.lancs.ac.uk/lancsbox>.
- Burrows, J. F. (1987a). *Computation into criticism: a study of Jane Austen’s novels and an experiment in method*. Oxford.
- Burrows, J. F. (1987b). Word-Patterns and Story-Shapes: The Statistical Analysis of Narrative Style. *Literary and Linguistic Computing*, 2(2), 61–70. Retrieved 2024-03-02, from <https://academic.oup.com/dsh/article-lookup/doi/10.1093/llc/2.2.61> doi: 10.1093/llc/2.2.61
- Burrows, J. F. (1992). Not Unless You Ask Nicely: The Interpretative Nexus Between Analysis and Information. *Literary and Linguistic Computing*, 7(2), 91–109. Retrieved 2024-03-02, from <https://academic.oup.com/dsh/article-lookup/doi/10.1093/llc/7.2.91> doi: 10.1093/llc/7.2.91
- Busch, V. (2019). Jack Unterweger. In *Psychologie des Guten und Bösen* (pp. 359–369). Berlin/Heidelberg: Springer.
- Coulmas, F. (1979). On the sociolinguistic relevance of routine formulae. *Journal of Pragmatics*, 3(3-4), 239–266. Retrieved 2024-03-02, from <https://linkinghub.elsevier>

- .com/retrieve/pii/037821667990033X doi: 10.1016/0378-2166(79)90033-X
- Coulthard, M. (2004). Author identification, idiolect and linguistic uniqueness. *Applied Linguistics*, 25(4), 431–447.
- Coulthard, M. (2007). By their words shall ye know them: on linguistic identity. In C. R. Caldas-Coulthard & R. Iedema (Eds.), *Identity trouble* (pp. 143–155). Palgrave Macmillan.
- Dern, C. (2008). Wenn zahle nix dann geht dir schlecht“. Ein Experiment zu sprachlichen Verstellungsstrategien in Erpresserbriefen. *ZFGL*, 36(2), 240–265. Retrieved 2024-03-02, from <https://www.degruyter.com/document/doi/10.1515/ZGL.2008.018/html> doi: 10.1515/ZGL.2008.018
- DWDS Kernkorpus. (2023). *DWDS Kernkorpus*. Retrieved from <https://dwds.de/d/korpora/kern>
- Eder, M. (2013). Does size matter? Authorship attribution, small samples, big problem. *Digital Scholarship in the Humanities*, 30(2), 167–182. Retrieved 2024-03-02, from <http://academic.oup.com/dsh/article/30/2/167/390738/Does-size-matter-Authorship-attribution-small> doi: 10.1093/llc/fqt066
- Eder, M. (2017). Short samples in authorship attribution: a new approach. *DH*, 1–5.
- Eder, M., Rybicki, J., Kestemont, M., & Pielstrom, S. (2022). “*Stylo*”: a package for stylometric analyses. Retrieved from https://github.com/computationalstylistics/stylo_howto
- Ehrhardt, S. (2018). Authorship attribution analysis. In J. Visconti & M. Rathert (Eds.), *Handbook of Communication in the Legal Sphere* (pp. 169–200). Berlin: De Gruyter. Retrieved 2024-03-02, from <https://www.degruyter.com/document/doi/10.1515/9781614514664-010/html> doi: 10.1515/9781614514664-010
- Fine, J. (2006). *Language in psychiatry. A handbook of clinical practice*. London: Equinox.
- Fobbe, E. (2011). *Forensische Linguistik*. Tübingen: Narr.
- Fobbe, E. (2020). Text-Linguistic Analysis in Forensic Authorship Attribution. *International Journal of Language & Law (JLL)*, Vol 9: Forensic Linguistics: New Procedures and Standards. Retrieved 2024-03-02, from <https://www.languageandlaw.eu/jll/article/view/78> (Publisher: [object Object]) doi: 10.14762/JLL.2020.093
- Fobbe, E. (2021). Forensische Linguistik. Eine kriminaltechnische Disziplin in Deutschland. *SIK-Journal – Zeitschrift für Polizeiwissenschaft und polizeiliche Praxis*(4), 18–27. Retrieved 2024-03-02, from https://doi.org/10.7396/2021_4_B doi: 10.7396/2021_4_B
- Fobbe, E. (2022). Authorship identification. In V. Guillén-Nieto & D. Stein (Eds.), *Language as evidence. Doing forensic linguistics* (pp. 185–218). Palgrave Macmillan.
- Grant, T., & Baker, K. (2001). Identifying reliable, valid markers of authorship: a response to Chaski. *International Journal of Speech, Language & the Law*, 8(1), 1350–1771.
- Grant, T., & MacLeod, N. (2018). Resources and constraints in linguistic identity performance: a theory of authorship. *Language and Law / Linguagem e Direito*, 5(1), 80–96.
- Gries, S. (2022). How to use statistics in quantitative corpus linguistics. In A. O’Keeffe & M. J. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 168–181). London/New York: Routledge.
- Gries, S., & Stefanowitsch, A. (2010). Cluster analysis and the identification of collexeme classes. In S. Rise & J. Newman (Eds.), *Empirical and experimental methods in*

- cognitive/functional research (pp. 73–90). CSLI Publications.
- Hennig, M. (2020). *Nominalstil*. Tübingen: Narr.
- Herring, S. C., & Paolillo, J. C. (2006). Gender and genre variation in weblogs. *Journal of Sociolinguistics*, 10(4), 439–459. Retrieved 2024-03-02, from <https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9841.2006.00287.x> doi: 10.1111/j.1467-9841.2006.00287.x
- Herwig, M. (2022). *JACK. Gier frisst Schönheiten*. Retrieved from <https://www.ardaudiothek.de/sendung/jack-gier-frisst-schoenheiten/10385179/>
- Holmes, D. I. (1992). A Stylometric Analysis of Mormon Scripture and Related Texts. *Journal of the Royal Statistical Society*, 155(1), 91–120.
- Holzer, E., & Reibenwein, M. (2022). Jack Unterweger: Der Popstar unter den Serienmördern. Retrieved from <https://kurier.at/chronik/oesterreich/jack-unterweger-der-popstar-unter-den-serienmoerdern/401888249>
- Hoover, D. L. (2004). Testing Burrow’s Delta. *Literary and Linguistic Computing*, 19(4), 453–475.
- Johnson, A., & Wright, D. (2014). Identifying idiolect in forensic authorship attribution: an n-gram textbite approach. *Language and Law / Linguagem e Direito*, 1(1), 37–69.
- Juola, P. (2021, August). Verifying authorship for forensic purposes: A computational protocol and its validation. *Forensic Science International*, 325, 110824. Retrieved 2024-03-02, from <https://linkinghub.elsevier.com/retrieve/pii/S0379073821001444> doi: 10.1016/j.forsciint.2021.110824
- Kestemont, M. (2014). Function Words in Authorship Attribution From Black Magic to Theory? In *Proceedings of the 3rd Workshop on Computational Linguistics for Literature* (pp. 59–66).
- Koppel, M., Schler, J., & Argamon, S. (2009). Computational methods in authorship attribution. *Journal of the American Society for Information Science and Technology*, 60(1), 9–26. Retrieved 2024-03-02, from <https://onlinelibrary.wiley.com/doi/10.1002/asi.20961> doi: 10.1002/asi.20961
- Koppel, M., Schler, J., & Argamon, S. (2011). Authorship attribution in the wild. *Language Resources and Evaluation*, 45(1), 83–94. Retrieved 2024-03-02, from <http://link.springer.com/10.1007/s10579-009-9111-2> doi: 10.1007/s10579-009-9111-2
- Larner, S. (2014). A preliminary investigation into the use of fixed formulaic sequences as a marker of authorship. *International Journal of Speech Language and the Law*, 21(1), 1–22. Retrieved 2024-03-02, from <https://journal.equinoxpub.com/IJSLL/article/view/5898> doi: 10.1558/ijssl.v21i1.1
- Layton, R., Watters, P. A., & Dazeley, R. (2015). Authorship analysis of aliases: Does topic influence accuracy? *Natural Language Engineering*, 21(4), 497–518. Retrieved 2024-03-02, from https://www.cambridge.org/core/product/identifier/S1351324913000272/type/journal_article doi: 10.1017/S1351324913000272
- Leake, J. (2010). *Der Mann aus dem Fegefeuer: das Doppelleben des Jack Unterweger*. Heyne.
- Locker, A. (2019). “Because the computer said so!” Can computational authorship analysis be trusted? *Journal of Language Works*, 4(1), 23–37.
- Marko, K., & Leibetseder, I. (2023). Linguistic Indicators of Psychopathy and Malignant Narcissism in the Personal Letters of the Austrian Killer Jack Unterweger. *Forensic Sciences*, 3(1), 45–68. Retrieved 2024-03-02, from <https://www.mdpi.com/>

- 2673-6756/3/1/6 doi: 10.3390/forensicsci3010006
- McMenamin, G. R. (2002). *Forensic stylistics*. CRC Press.
- Miranda García, A., & Calle Martín, J. (2007). Function Words in Authorship Attribution Studies. *Literary and Linguistic Computing*, 22(1), 49–66. Retrieved 2024-03-02, from <https://academic.oup.com/dsh/article-lookup/doi/10.1093/lc/fql048> doi: 10.1093/lc/fql048
- Moon, R. (1998). *Fixed expressions and idioms in English: a corpus-based approach*. Clarendon Press.
- Mose, J. (2015). *Jack Unterweger Interview im Gefängnis Stein Part 1*. Retrieved from <https://www.youtube.com/watch?v=kjYNVQjggto>
- Mosteller, F., & Wallace, D. L. (1984). The Federalist Papers As a Case Study. In *Applied Bayesian and Classical Inference* (pp. 1–15). New York, NY: Springer New York. Retrieved 2024-03-02, from http://link.springer.com/10.1007/978-1-4612-5256-6_1 (Series Title: Springer Series in Statistics) doi: 10.1007/978-1-4612-5256-6_1
- Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender Differences in Language Use: An Analysis of 14,000 Text Samples. *Discourse Processes*, 45(3), 211–236. Retrieved 2024-03-02, from <http://www.tandfonline.com/doi/abs/10.1080/01638530802073712> doi: 10.1080/01638530802073712
- Nini, A. (2018a). An authorship analysis of the Jack the Ripper letters. *Digital Scholarship in the Humanities*, 33(3), 621–636. Retrieved 2024-03-02, from <https://academic.oup.com/dsh/article/33/3/621/4824843> doi: 10.1093/lc/fqx065
- Nini, A. (2018b). Developing forensic authorship profiling. *Language and Law / Linguagem e Direito*, 5(2), 38–58.
- Nini, A., & Grant, T. (2013). Bridging the gap between stylistic and cognitive approaches to authorship analysis using Systemic Functional Linguistics and multidimensional analysis. *International Journal of Speech Language and the Law*, 20(2), 173–202. Retrieved 2024-03-02, from <https://journal.equinoxpub.com/IJSLL/article/view/5925> doi: 10.1558/ijssl.v20i2.173
- Orvell, A., Gelman, S. A., & Kross, E. (2022). What “you” and “we” say about me: How small shifts in language reveal and empower fundamental shifts in perspective. *Social and Personality Psychology Compass*, 16(5), e12665. Retrieved 2024-03-02, from <https://compass.onlinelibrary.wiley.com/doi/10.1111/spc3.12665> doi: 10.1111/spc3.12665
- Overdorf, R., & Greenstadt, R. (2016). Blogs, Twitter Feeds, and Reddit Comments: Cross-domain Authorship Attribution. *Proceedings on Privacy Enhancing Technologies*, 2016(3), 155–171. Retrieved 2024-03-02, from <https://petsymposium.org/popets/2016/popets-2016-0021.php> doi: 10.1515/popets-2016-0021
- Pennebaker, J. (2011). *The Secret Life of Pronouns: What Our Words Say about Us*. Bloomsbury.
- Peters, A. (1984). *The Units of Language Acquisition*. Cambridge: Cambridge University Press.
- R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. Viena, Austria: R Foundation for Statistical Computing. Retrieved 2022-08-14, from <https://www.R-project.org/>
- Rude, S., Gortner, E.-M., & Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, 18(8), 1121–

1133. Retrieved 2024-03-02, from <http://www.tandfonline.com/doi/abs/10.1080/02699930441000030> doi: 10.1080/02699930441000030
- Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing, use* (pp. 127–151). John Benjamins.
- Solan, L. M. (2013). Intuition versus Algorithm: The Case of Forensic Authorship Attribution. *Journal of Law and Policy*, 21(3), 551–576.
- Stamatatos, E. (2009). A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology*, 60(3), 538–556. Retrieved 2024-03-02, from <https://onlinelibrary.wiley.com/doi/10.1002/asi.21001> doi: 10.1002/asi.21001
- Stamatatos, E. (2013). On the Robustness of Authorship Attribution Based on Character N-gram Features. *Journal of Law and Policy*, 21, 421–439.
- Swain, S., Mishra, G., & Sindhu, C. (2017). Recent approaches on authorship attribution techniques — An overview. In *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)* (pp. 557–566). Coimbatore: IEEE. Retrieved 2024-03-02, from <http://ieeexplore.ieee.org/document/8203599/> doi: 10.1109/ICECA.2017.8203599
- Turell, M. T. (2010). The use of textual, grammatical and sociolinguistic evidence in forensic text comparison:. *International Journal of Speech, Language and the Law*, 17(2), 211–250. Retrieved 2024-03-02, from <https://journal.equinoxpub.com/IJSLL/article/view/6001> doi: 10.1558/ijssl.v17i2.211
- Turell, M. T., & Gavalda, N. (2013). Towards an index of idiolectal similitude (or distance) in forensic authorship analysis. *Journal of Law and Policy*, 21(2), 495–514.
- Viding, E. (2019). *Psychopathy: A Very Short Introduction* (1st ed.). Oxford University Press. Retrieved 2024-03-02, from <https://academic.oup.com/book/28395> doi: 10.1093/actrade/9780198802266.001.0001
- Widler, Y. (2019). Psychiater Haller: “Gefährlich ist nicht der finstere Wald, sondern das eigene Heim”. *kurier.at*. Retrieved from <https://kurier.at/chronik/oesterreich/profiler-haller-gefaehrlich-ist-nicht-der-finstere-wald-sondern-das-eigene-heim/400637114>
- Wortbrücke. (n.d.). Retrieved 2022-08-08, from <https://www.onb.ac.at/oe-literaturzeitschriften/Wortbruecke/Wortbruecke.htm>
- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge University Press.
- Wright, D. (2013). Stylistic variation within genre conventions in the Enron email corpus: developing a textsensitive methodology for authorship research. *International Journal of Speech Language and the Law*, 20(1), 45–75. Retrieved 2024-03-02, from <https://journal.equinoxpub.com/IJSLL/article/view/5944> doi: 10.1558/ijssl.v20i1.45
- Wright, D. (2017). Using word n-grams to identify authors and idiolects: A corpus approach to a forensic linguistic problem. *International Journal of Corpus Linguistics*, 22(2), 212–241. Retrieved 2024-03-02, from <http://www.jbe-platform.com/content/journals/10.1075/ijcl.22.2.03wri> doi: 10.1075/ijcl.22.2.03wri