

# The Relationship Between Linguistic Style and Immigration Sentiment: A Case Study of the Flashback Forum

Dominik Künkele

contact@dominik-kuenkele.de

## Abstract

This study examines the relationship between linguistic style and sentiment towards immigration in the Flashback forum, analyzing over 3.5 million posts from authors with different attitudes. Our analysis of various linguistic features reveals no significant relationships with immigration sentiment, except for a weak correlation in vocabulary diversity. This finding challenges the assumption that authors' attitudes towards immigration would be systematically reflected in their general writing style. Instead, linguistic style appears to be more strongly influenced by contextual factors such as the specific topic of discussion and forum norms. These results highlight the complex and context-dependent nature of the relationship between sentiment and linguistic expression in online forums.

## 1 Introduction

Sentiment analysis, also known as opinion mining, is a computational approach to identifying and categorizing opinions expressed in text. It aims to determine the writer's attitude towards a particular topic, product, or service, typically classifying the sentiment as positive, negative, or neutral. The field has seen significant advancements in processing various languages, including Swedish, where research has focused on both general sentiment analysis (Barnes et al., 2022) and domain-specific applications. In recent years, sentiment analysis has become increasingly important in understanding public opinion, particularly in the context of social media and online forums (Liu, 2012; Kokkinakis and Hammarlin, 2022; Rouces et al., 2019). Online forums, such as Flashback, offer a unique window into public opinion, as they allow for relatively unrestricted expression of views and facilitate extensive discussions on

sensitive topics. Previous research has shown that sentiment analysis of immigration-related discourse can reveal patterns in public opinion and help understand the factors influencing attitudes towards immigration (Berdicevskis et al., 2023). However, analyzing sentiment in online forums presents unique challenges, including the informal nature of the language, the use of slang and abbreviations, and the potential for sarcasm and irony.

Linguistic style encompasses the unique ways in which people use language to express themselves. It includes various features such as vocabulary choice, sentence structure, use of figurative language, and even punctuation patterns. Defining linguistic style is challenging because it is inherently subjective and can vary significantly across different contexts, topics, and time periods. As Biber (1988) notes, linguistic style is not a fixed entity but rather a dynamic system that adapts to different communicative situations.

Several approaches have been proposed to analyze linguistic style. These include:

- Statistical analysis of word frequencies and patterns
- Examination of syntactic structures and sentence complexity
- Analysis of discourse markers and pragmatic features
- Study of topic-specific vocabulary and terminology
- Investigation of emotional and sentiment-related language patterns

The challenge in analyzing linguistic style lies in identifying which features are most indicative of an individual's writing style while accounting for the influence of context, topic, and

audience (Schwartz et al., 2017; Pennebaker and King, 2000). This becomes particularly complex when studying online forums, where users may adapt their style based on the specific subforum, topic, or audience they are addressing.

In this paper, we investigate the relationship between sentiment towards immigration and linguistic style in the context of the Flashback forum. We aim to understand how users’ attitudes towards immigration are reflected in their general writing style across different topics and contexts. This research contributes to both sentiment analysis and linguistic style research by examining how these two aspects of language use interact in a large-scale online forum setting.

## 2 Datasets

Flashback is one of Sweden’s largest online discussion forums, established in 2000. It serves as a platform for open discussions on a wide range of topics, from current events and politics to lifestyle and entertainment. The forum is known for its active user base and extensive discussions, making it a valuable source for studying natural language use in online discourse. While the forum maintains a general focus on Swedish society and culture, its discussions often touch upon international topics, including immigration and social issues, which makes it particularly relevant for our study of sentiment analysis. Posts on this forum are collected in the *Flashback* dataset (Språkbanken, 2025). The dataset comprises over 3.521.229 posts collected from 16 subforums. Each sentence in the corpus has been annotated with part-of-speech (POS) tags following the annotation scheme described in (Språkbanken, 2024). The POS tagging provides valuable linguistic information that enables detailed analysis of language patterns and structures across different contexts. We enriched this dataset for this research with information directly scraped from the Flashback forum, to link posts to each other.

The focus of this research lies on sentiments of posts and authors, more specifically sentiment towards immigration. In the *Swedish ABSAbank-Imm v1.1* corpus (Berdicevskis et al., 2023) a subset of 4.872 posts on Flash-

back with the subject-matter of immigration were manually labeled. Hereby, the sentiment is split up into five categories where 1 corresponds to a ‘very negative’ and 5 to a ‘very positive’ sentiment towards immigration. Up to 10 annotators were tasked to label the posts and the final sentiment score is calculated as the average of all worker annotations. From the sentiment-labeled dataset, we identified 619 unique authors. To ensure robust analysis, we focused on authors with more than two posts, resulting in a subset of 361 authors who collectively wrote 1.928 posts. These authors were then labeled based on their average sentiment scores across all their posts. The resulting subset represents a group of users with varying perspectives on immigration, providing a rich dataset for analyzing language patterns across different sentiment orientations.

As the *Swedish ABSAbank-Imm v1.1* corpus is relatively small for an analysis of the linguistic style of an author, we utilize the whole Flashback corpus for our research. More specifically, for all labeled authors based on the sentiment analysis, we derive their linguistic style from all their posts found in the Flashback corpus. To ensure meaningful comparisons, we filtered the sentences to have similar lengths and therefore excluded sentences with fewer than 10 and more than 20 tokens, resulting in a final dataset of 380,323 sentences. The final dataset provides a balanced and representative sample of language use across different sentiment orientations, while maintaining consistent sentence length characteristics. This standardization is particularly important for our analysis of language patterns, as it makes statistical comparisons more stable.

## 3 Methods

The aim of this research is exploring similarities and differences in how authors of different sentiment towards immigration express their opinion. More specifically, we want to test the hypothesis, that authors with a positive sentiment towards immigration utilize a different style in their utterances than authors with a negative sentiment. However, defining and extracting a linguistic style comes with several challenges (Berdicevskis and Erbro, 2023): Linguistic style is not a predefined set of rules

that speakers consciously follow, but rather emerges naturally from their language use. It is the sum of all linguistic choices made in communication. Every aspect of language production contributes to an individual's style, including lexical choices such as word selection and vocabulary, grammatical structures and their complexity, level of descriptiveness and detail, use of figurative language and metaphors, sentence structure and length, and discourse markers and pragmatic features. These choices are often made unconsciously and reflect deeper patterns in how individuals process and express thoughts. The style is therefore not something that speakers impose on their language, but rather something that emerges from their natural language use. Furthermore, this linguistic style is constantly subject to change. It is for example shown that members of online communities adapt their way of communication to the style of their interlocutors [Danescu-Niculescu-Mizil et al. \(2013\)](#); [McPherson et al. \(2001\)](#); [Berdicevskis and Erbro \(2023\)](#). Every turn in a conversation affects the own linguistic style.

To accommodate, we won't define a measure for a general linguistic style of an author, but rather have a look at specific statistical patterns in the expressions. More specifically, we will have a closer look at three properties of the authors' utterances. First, we measure the correlation between the average sentence length an author uses with the sentiment they express towards immigration. Hereby, we include all sentence lengths and don't exclude short or long sentences. Do authors with a negative sentiment use shorter or longer sentences than authors with a positive sentiment?

Secondly, we examine the lexical diversity in the authors' writing by calculating the type-token ratio (TTR) for each author. TTR is a measure of lexical diversity that represents the ratio of unique words (types) to the total number of words (tokens) in a text. A higher TTR indicates greater lexical diversity, while a lower TTR suggests more repetition of words. This measure will help us understand whether authors with different sentiments towards immigration tend to use a more varied or more repetitive vocabulary in their writing. For example, do authors with negative sentiment towards immigration tend to repeat certain words more frequently than those with

positive sentiment?

Lastly, we analyze the distribution of different parts of speech (POS) in the authors' writing. For instance, a higher proportion of nouns might indicate a more descriptive or formal style, while more verbs could suggest a more action-oriented or dynamic communication style. By examining these patterns, we can gain insights into how authors with different sentiments structure their thoughts and express their ideas.

We specifically examine the following POS ratios on linguistic style:

- **Lexical Density:** The ratio of content words (nouns, adjectives, verbs) to the total number of words. This measure helps us understand how information-dense the authors' writing is.
- **Verb Complexity:** The proportion of verbs to participles, which can indicate the complexity of actions described.
- **Nominal Style:** The ratio of nouns to the sum of nouns and verbs, revealing whether authors tend to focus more on objects and concepts rather than actions.
- **Adjectival Richness:** The proportion of adjectives in the text, which can indicate the level of descriptive detail in the writing.
- **Subordination Complexity:** The ratio of subordinating conjunctions to the total number of words, indicating the complexity of sentence structure.
- **Formality Index:** The proportion of prepositions, determiners, and relative pronouns, which can indicate the formality of the writing style.
- **Punctuation Density:** The ratio of punctuation marks to the total number of words, which can reflect the rhythm and structure of the writing.
- **Pronoun Usage:** The proportion of pronouns in the text, which can indicate the level of personal involvement in the discourse.

- **Adverbial Style:** The ratio of adverbs to the total number of words, revealing how authors modify their statements.
- **Conjunction Usage:** The proportion of coordinating and subordinating conjunctions, indicating how authors connect their ideas.
- **Interjection Frequency:** The ratio of interjections to the total number of words, which can indicate emotional expression.

These measures provide a comprehensive view of the authors’ linguistic style across different grammatical and structural dimensions. By comparing these ratios between authors with different sentiments towards immigration, we can identify potential patterns in how they structure their arguments and express their views.

For all three approaches, we calculate the Spearman correlation. The Spearman correlation coefficient ( $\rho$ ) is a non-parametric measure of rank correlation that assesses monotonic relationships between variables. We use it because it is robust to outliers in our dataset, and works well with ordinal data like our linguistic features. For each linguistic feature, we calculate the Spearman correlation with sentiment scores across all authors to identify the strongest associations:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (1)$$

where:

- $d_i$  is the difference between the ranks of corresponding values in the two variables
- $n$  is the number of observations

## 4 Results

Our analysis of the relationship between linguistic features and sentiment towards immigration revealed only one statistically significant correlation. Table 1 presents the Spearman correlation coefficients and their corresponding p-values for all linguistic features examined.

The results show that most examined linguistic features demonstrate no statistically significant correlation with sentiment towards

immigration. The strongest non-significant correlations were observed for Nominal Style ( $\rho = -0.062$ ,  $p = 0.242$ ) and Verb Complexity ( $\rho = 0.051$ ,  $p = 0.336$ ). However, we found a weak but statistically significant positive correlation for Type-Token Ratio (TTR) ( $\rho = 0.150$ ,  $p = 0.048$ ), suggesting that authors with more positive sentiment towards immigration tend to use slightly more diverse vocabulary in their writing.

These findings suggest that authors’ attitudes towards immigration are not systematically reflected in most aspects of their general linguistic style across different topics. The linguistic style of authors appears to be more influenced by the specific context or topic of discussion rather than their general sentiment towards immigration. This could be explained by authors consciously or unconsciously adapting their writing style based on the forum’s norms and expectations. The weak but significant correlation found for TTR might indicate a subtle tendency for authors with positive sentiment to employ a more varied vocabulary, though this relationship is relatively weak and should be interpreted with caution.

## 5 Discussion and Conclusion

Our investigation into the relationship between linguistic style and sentiment towards immigration yielded largely negative results, with only a weak correlation found for vocabulary diversity (TTR). This finding challenges the initial hypothesis that authors’ attitudes towards immigration would be systematically reflected in their general linguistic style. The results suggest that the relationship between sentiment and linguistic style is more complex than previously assumed, and several methodological considerations may help explain these findings.

First, our approach of extracting linguistic style from all posts across different topics and time periods may have masked potential style differences. As noted in the introduction, linguistic style is not a fixed entity but rather a dynamic system that adapts to different communicative situations (Biber, 1988). By aggregating style features across all posts, we may have averaged out topic-specific or temporal variations in writing style. For instance, an author might employ different



Linguistic Feature	Spearman's	p-value
Sentence Length	-0.018	0.726
Type-Token Ratio (TTR)	0.150	0.048
Lexical Density	-0.000	0.993
Verb Complexity	0.051	0.336
Nominal Style	-0.062	0.242
Adjectival Richness	-0.040	0.450
Subordination Complexity	0.026	0.617
Formality Index	-0.021	0.696
Punctuation Density	-0.021	0.695
Pronoun Usage	0.010	0.856
Adverbial Style	0.000	0.994
Conjunction Usage	-0.013	0.803
Interjection Frequency	0.022	0.682

Table 1: Spearman correlation coefficients and p-values for linguistic features with sentiment towards immigration. Only Type-Token Ratio (TTR) shows a significant correlation with immigration sentiment ( $p < 0.05$ ).

linguistic strategies when discussing immigration compared to other topics, or their style might evolve over time. Future research could benefit from analyzing style patterns specifically within immigration-related discussions or within specific time periods.

Second, the context-dependent nature of linguistic style, as discussed in the introduction, suggests that style features might be more strongly influenced by the specific subforum or topic of discussion rather than the author’s general sentiment towards immigration. This aligns with previous research showing that online community members adapt their communication style to their interlocutors (Danescu-Niculescu-Mizil et al., 2013). The weak but significant correlation found for TTR might indicate that vocabulary diversity is one of the few style features that shows some consistency across different contexts, though this relationship is relatively weak and should be interpreted with caution.

The absence of significant correlations for most linguistic features raises important questions about the nature of sentiment expression in online forums. While sentiment analysis has proven valuable for understanding attitudes towards immigration (Berdicevskis et al., 2023), our results suggest that these attitudes may not be consistently reflected in general linguistic style patterns. This could be due to authors consciously adapting their writing style to conform to forum norms and expectations, regard-

less of their sentiment. Additionally, the expression of sentiment might be more dependent on specific lexical choices and topic-specific vocabulary rather than general style features. The relationship between sentiment and style might also be more complex, involving interactions between multiple features or contextual factors that were not captured in our analysis.

Future research could address these limitations by analyzing linguistic style specifically within immigration-related discussions, examining style patterns within specific subforums or time periods, investigating the interaction between multiple linguistic features, and considering topic-specific vocabulary and expressions in addition to general style features. Such approaches would help better understand how sentiment and linguistic style interact in specific contexts and how these relationships might vary across different topics and time periods.

In conclusion, our study suggests that the relationship between sentiment towards immigration and linguistic style is more complex and context-dependent than initially hypothesized. While we found a weak correlation for vocabulary diversity, most linguistic features showed no significant relationship with sentiment. This highlights the importance of considering contextual factors and topic-specific patterns when analyzing the relationship between sentiment and linguistic style in online forums.

## References

- Jeremy Barnes, Laura Oberlaender, Enrica Troiano, Andrey Kutuzov, Jan Buchmann, Rodrigo Agerri, Lilja Øvrelid, and Erik Velldal. 2022. [SemEval 2022 task 10: Structured sentiment analysis](#). In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 1280–1295, Seattle, United States. Association for Computational Linguistics.
- Aleksandrs Berdicevskis, Lars Borin, Jacobo Rouces, and Nina Tahmasebi. 2023. [Swedish absabank-imm 1.1](#).
- Aleksandrs Berdicevskis and Viktor Erbro. 2023. [You say tomato, I say the same: A large-scale study of linguistic accommodation in online communities](#). In *Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 415–424, Tórshavn, Faroe Islands. University of Tartu Library.
- Douglas Biber. 1988. *Variation across Speech and Writing*. Cambridge University Press.
- Cristian Danescu-Niculescu-Mizil, Robert West, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. [No country for old members: user lifecycle and linguistic change in online communities](#). In *Proceedings of the 22nd International Conference on World Wide Web, WWW '13*, page 307–318, New York, NY, USA. Association for Computing Machinery.
- Dimitrios Kokkinakis and Mia-Marie Hammarlin. 2022. Negative vaccine voices in swedish social media. In *Proceedings of the 13th International Conference of Experimental Linguistics*, pages 81–84, Paris. ExLing Society.
- Bing Liu. 2012. *Sentiment Analysis and Opinion Mining*. Springer International Publishing.
- Miller McPherson, Lynn Smith-Lovin, and James M Cook. 2001. [Birds of a feather: Homophily in social networks](#). *Annual Review of Sociology*, 27(Volume 27, 2001):415–444.
- James Pennebaker and Laura King. 2000. [Linguistic styles: Language use as an individual difference](#). *Journal of personality and social psychology*, 77:1296–312.
- Jacobo Rouces, Lars Borin, and Nina Tahmasebi. 2019. Tracking attitudes towards immigration in swedish media. In *CEUR Workshop Proceedings*, volume 2364, pages 387–393, Copenhagen. CEUR Workshop Proceedings.
- Roy Schwartz, Maarten Sap, Ioannis Konstas, Li Zilles, Yejin Choi, and Noah A. Smith. 2017. [The effect of different writing tasks on linguistic style: A case study of the roc story cloze task](#).
- Språkbanken. 2024. [Suc 3.0](#).
- Språkbanken. 2025. [Flashback](#).