

LEGE ARTIS

Language yesterday, today, tomorrow

Vol. X. No 2 2025

**LEXICAL AND GRAMMATICAL PATTERNS IN HIGHLY RATED VS
POORLY RATED DOCUMENTARIES: A COMPARATIVE
MULTIDIMENSIONAL ANALYSIS¹****Sadia Ali** ***Prince Sattam bin Abdulaziz University,
Al-Kharj, Kingdom of Saudi Arabia*****Received:** 1.12.2024 **Reviewed:** 17.07.25 and 29.09.2025**Similarity Index:** 0%

Bibliographic description: Ali, S. (2025). Lexical and grammatical patterns in highly rated vs poorly rated documentaries: A comparative multidimensional analysis. In *Lege artis. Language yesterday, today, tomorrow*. Trnava: University of SS Cyril and Methodius in Trnava, X (2), p. 2-19. ISSN 2453-8035 DOI: <https://doi.org/10.34135/lartis.25.10.2.01>

Abstract: In this paper, the lexical and grammatical patterns of 20 highly- and poorly-rated documentaries of 2023 and 2024 are compared through the Multidimensional Analysis Tagger. Documentaries that are rated highly are those with a narrative discourse style. The poorly-rated ones are more informational and have a higher lexical density. The documentaries with high ratings are also less persuasive and more abstract. Statistical tests, such as ANOVA, support significant lexical and grammatical differences. The results provide an understanding of the role of linguistic choices in influencing the audience and documentary success.

Keywords: audience engagement; lexical density; lexical and grammatical features; linguistic choices; Multidimensional Analysis.

1. Introduction

A documentary is a powerful narration tool. In its capacity, it informs its viewers on a number of issues such as personal stories, history, and society. Interesting stories and the presentation of facts attract listeners. BBC Select and Factual America offer the viewers a chance to watch fascinating and informative documentaries, such as history, nature, and true crime documentaries, among others.

¹ Acknowledgement: This research is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2025/R/1447).



The documentaries to be used in this research are chosen according to the IMDb rating list. IMDb is a well-known portal that collects the reviews and ratings of users on movies and TV series. In the proposed time frame, I categorized the documentaries of 2023 and 2024 based on their rating, including highly rated and poorly rated ones.

The text analysis of the study is done with the help of the MAT tagger, which operates with 400-token chunks. It is a suitable sample size, since it is not too large and can be analyzed in detail. It effectively captures linguistic features (Biber 1991: 89-90). The study focuses on the five dimensions identified by Biber (1991).

This study uses the statistical software (SPSS) for statistical analyses, such as descriptive statistics and analysis of variance (ANOVA) for the calculations of the dimension scores of documentaries to compare the highly and poorly rated documentaries. The multidimensionality of the approach reveals the complex ways in which specific features of language affect documentary audiences and the delivery of the message.

It has also been observed that, in the past, documentaries were used to facilitate language learning by exposing students to real language input (Hertel & Harrington 2015; Rahman et al. 2018). Nevertheless, the current research is more concerned with the linguistic trends that make documentaries more appealing to the general audience and successful.

Some documentaries have been studied outside the language education context, especially in communication and media studies. Scholars of these disciplines have examined how documentaries use rhetorical and persuasive appeals, imaginative narrative methods, and unique forms and functions to persuade audiences and influence social discourse (Nichols 2017; Renov 2004; Winston 2019). These studies underline the documentary's power to inform, persuade, and motivate people to think, and its creative and communicative potential to solve real-life problems.

Therefore, the documentary as a highly rated, useful, and demanded social media product has received much attention. Meanwhile, the lexical and grammatical patterns that determine the success (high IMDb audience ratings) or failure (low IMDb audience ratings) of this genre are not well known. Moreover, the linguistic peculiarities that characterize compelling documentaries are underresearched. This paper will compare the highly and poorly rated documentaries to determine the patterns that are related to audience response.

The study fills the gap in linguistic research on how lexical and grammatical patterns relate to documentary success by examining the use of language in documentaries at different levels of audience ratings. The study seeks to identify tendencies in the use of vocabulary, grammar, and linguistic features, and to offer recommendations for enhancing the linguistic quality of both highly rated and poorly rated documentaries in English. While the current study is limited to English-language production, the identified lexical and grammatical patterns may serve as a basis for exploring whether similar tendencies apply in other linguistic and cultural contexts.

2. Literature review

Documentary filmmaking helps in understanding various topics and subjects in detail. Documentaries can be historical or simply constitute a narrative from a person for whom the advert is meant to be (Nichols 2017; Renov 2004). Filmmakers (Winston 2019) have analyzed the ability of a documentary to teach something to the viewers and make them feel certain emotions. *BBC Select* and *Factual America* produce categorized documentaries, which include history, nature, and true crime.

The narrative content and documentary storytelling have been revisited and analyzed multiple times (Aufderheide 2007; Plantinga 2015), demonstrating how narratives can transform factual content into compelling stories that resonate with audiences. Documentaries such as "The Square" (2013), which portrays the 2011 Egyptian revolution, and "The Hunting Ground" (2015), which portrays sexual assault in the U.S.A., are great examples of social issue documentaries that are inspirational and motivational for change (CODA Filmworks, 2023; Factual America, 2023). The combination of expert interviews, historical re-enactments, and personal accounts, in series like "Digging for Britain" and "Michael Wood's Story of England", from Factual America (2023) showcases the extent and the impact that well-crafted documentaries can reach (Winston et al. 2017).

Most evaluations of documentaries have concentrated on conventional indicators of success, including audience ratings, viewership, or box office revenue. However, there has been limited attention to the lexical and grammatical features that distinguish highly rated and poorly rated documentaries. This gap highlights the need to investigate how specific lexical and grammatical features influence audience reception. Documentaries are an effective method of expressing an opinion on various subjects, including environmental deterioration, ethnic identity, and societal issues. Some of the substantial written communication studies include Abdul et al. (2023), Murat (2022), and Suleimenova (2022). From silent documentaries to sound documentaries and from academic ones to filmmaker documentaries, the form and language of documentary filmmaking have

evolved dramatically in the last century (Deb 2022). Analyzing the language practices applied in the documentaries and independent films, including those of Al Gore's "An Inconvenient Truth", Hubert Sauper's "Darwin's Nightmare", or films by young Kazakhstan director Kanat Beisekeyev, will help us understand what aspects and language mechanisms caused those works to be so popular and effectively reaching the audience (Murat 2022; Suleimenova 2022). Therefore, argumentation, narrative, and stylistic analysis help reveal lexical and grammatical differences that distinguish effective documentaries, providing practical knowledge for both filmmakers and academics.

Biber's Multidimensional (MD) Analysis has become an established framework for examining linguistic variation across registers and genres (Biber 2006). It has been applied in diverse contexts, such as analyzing newspaper reporting patterns (Ali & Ali 2023; Ponder 2023), distinguishing between texts generated by artificial intelligence and human authors (Ali 2024b; Biber 2019; Sardinha 2024), analyzing academic language (Ali 2024a; Ali & Abdulhaleem 2024; Biber et al. 2002), studying human and ChatGPT-generated English translations of films (Ali & Afzal 2024), and examining Reddit's linguistic diversity evolution (Ali 2024c). The use of MD analysis in the context of documentary filmmaking remains a new field of research. In particular, the analysis of highly rated and poorly rated documentaries, along with MD analysis, is still limited.

Based on the MD framework developed by Biber, previous studies have demonstrated its ability to distinguish different lexical and grammatical characteristics in a variety of genres. Nonetheless, the effect of these linguistic features on audience perception and their effectiveness in documentaries has yet to be explored. This paper applies multidimensional analysis to documentary language and compares linguistic patterns of audience-rating levels to demonstrate how textual characteristics match reception.

3. Methodology

This research applies Biber's Multidimensional Analysis to analyze lexical and grammatical features of highly- and poorly rated documentaries from 2023 and 2024. The research material comprises 20 documentaries, divided into sets based on the ratings indicated on the Internet Movie Database (IMDb 2024), a popular website that features viewers' feedback on movies, shows, and documentaries. Considering that IMDb has a vast number of user ratings and reviews, it is a reliable source for selecting documentaries based on audience feedback.

If a documentary has an IMDb rating of 7 or above with the number of votes not less than 5000 votes, it can be considered highly rated. It is considered poorly rated if it has a rating of less than two stars, provided the number of reviews exceeds 5,000. Thus, these criteria help to identify those documentaries that have a significant viewership. This is important for evaluating the factors that define success or lack of success in their undertakings. By singling out documentaries that have had adequate ratings and reviews, it will become easier to evaluate correlations between the kind of writing, on the one hand, and the quality and popularity, on the other, as demonstrated by the data received. The documentaries chosen for this study are:

Highly rated documentaries: *American Nightmare* (2024), *Ashley Madison: Sex, Lies & Scandal* (2024), *Billy and Molly: An Otter Love Story* (2024), *Blue Angels* (2024), *Jim Henson: Idea Man* (2024), *MoviePass, MovieCrash* (2024), *Quiet on Set: The Dark Side of Kids TV* (2024), *The Beach Boys* (2024), *The Roast of Tom Brady* (2024), *What Jennifer Did* (2024).

Poorly rated documentaries: *America in Black* (2023), *Harry & Meghan: All Inclusive* (2023), *Harry & Meghan: What's Next* (2023), *Hollywood Princesses: Grace and Meghan* (2023), *Kurz* (2023), *Making of Meme King – Year Two* (2023), *Meghan Markle: Regretless* (2023), *Meghan & Kate: Women of Windsor* (2023), *Queen Cleopatra* (2023).

Using the MAT tagger (Nini 2019), we labelled the data, analyzed lexical and grammatical patterns, and divided the text into 400-word sections. According to Biber, this method works well because it provides enough context to capture speech patterns and makes it easier to do detailed research. This method guarantees close scrutiny of every component of the documentary discourse through the study of both local and global varieties of speech.

I applied Biber's Multidimensional Analysis to the tagged data, focusing on the first five dimensions: 1) involved vs. informational discourse, 2) narrative vs. non-narrative concerns, 3) explicit vs. situationally dependent reference, 4) overt expression of persuasion, and 5) impersonal (abstract) vs. non-impersonal (concrete) reference (See Appendix A for details). This was the most appropriate model considering the diverse and rich lexical and grammatical features and the overall impact of the documentary discourse on audience reception.

Using IBM SPSS Version 29.0 (IBM Corp. 2024), I automated the calculations for the dimension scores and conducted descriptive statistical analysis. To test the equality of variance across dimension scores for the two groups (highly rated and poorly rated documentaries), I applied Levene's Test for Homogeneity of Variances. Robust tests of equality of means, such as the Welch and Brown-Forsythe

tests, were utilized to validate the ANOVA results, particularly when deviations from homogeneity of variances were suspected. Descriptive statistics were applied to summarize the central tendency and dispersion of the dimension scores for both highly and poorly rated documentaries. The results of the study will enable us to test the hypothesis of the statistically reliable language differences observed between rating tiers.

4. Analysis and discussion

Dimension 1

The analysis in Dimension 1 shows significant differences between the highly- and poorly-rated documentaries in terms of involved and informational discourse. The results, displayed in Table 1, indicate that the mean score of the highly-rated documentaries (16.93) exceeds that of the poorly-rated documentaries (8.89). An ANOVA confirms the statistical significance of this difference: $F(1,58) = 30.171, p < 0.001$.

Table 1. Descriptive statistics for dimension 1. Source: Own processing

Category	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	Minimum	Maximum
Highly Rated	30	16.93	3.60	0.66	15.59 - 18.27	8.50	23.66
Poorly Rated	30	8.89	7.16	1.31	6.22 - 11.57	-3.40	25.32
Total	60	12.91	6.93	0.89	11.12 - 14.70	-3.40	25.32

According to Levene's Test for equality of variances (based on the mean) (Levene Statistic = 12.022, $p = 0.001$), there is a notable difference in variances between the two groups. This indicates that there is more variability in Dimension 1 scores among the poorly-rated documentaries compared to the highly-rated ones.

The greater the score on Dimension 1, the greater the rating of the documentaries in terms of dialogue. The following features of low formalism are typical of these texts: the use of short forms, active verbs in the simple present tense, and first- and second-person pronouns. Linguistic indicators of engagement, including colloquial language, idiomatic expressions that make one feel familiar, and deixis (e.g., *here, now, this*) that establishes immediacy and a common context, are also part of involved discourse. Such texts also tend to use short and simple sentences, ellipses, and direct address through second-person pronouns (*you, your*) that make the viewer feel that he/she is involved. Adjectives and adverbs of evaluation (e.g., *remarkable, deeply, incredibly*, etc.) are used to appeal to emotion and keep the audience empathizing.

In order to differentiate it from Dimension 3, it is necessary to mention that Dimension 1 concerns interpersonal involvement rather than contextual reference. In this case, deixis does not serve to place events in time or space but to create the illusion of proximity between the speaker and the viewer, the sense of dialogue and experience. It is concerned with human relationships and emotional closeness rather than representing a real physical scenario.

Such linguistic devices as deixis, colloquial expressions, and first-and second-person pronouns make the conversation interactive and understandable, strengthening the connection and improvised communication. This makes the documentaries easy to follow, thereby enhancing the associations that the viewers make and the flow of the content.

On the other hand, the low rating value documentaries have a discourse style that is portrayed by low scores of Dimension 1. It is thus more informative and less engaged, with a greater number of nominalizations, longer words, and higher lexical density. As a result, the content seems more formal and detached, and this influences how people view such documentaries.

The qualitative differences between the highly rated documentaries and the poorly rated ones are significant, which proves the inconsistent use of features related to the involved discourse. The poorly-rated documentaries are more incoherent and less engaging to the audience, perhaps due to the fact that they do not reach the same level of involvement and interpersonal connection as more successful ones.

In this respect, the results highlight the importance of the discourse style in defining the effects of the documentaries. Again, the focus is more on the documentaries that include elements of engaged argumentation, which may be explained by the fact that these elements enable the audience to become more engaged with the topic. This follows the general concept that a media strategy is comprised of some narrative, which is usually included in the context of relevance.

In the case of the results of Dimension 1, compared to the poorly rated documentaries, the highly rated ones are more likely to be described as having a more engaged style of discourse. On the whole, the use of this communicative strategy that fosters a feeling of intimacy and solidarity is the key to the overall performance achieved by the documentaries. These results indicate that the filmmakers who want to achieve a better rating from the audience should take into consideration the use of a

balanced discourse style that would allow combining the ease of the language used with the structural and thematic complexity required to maintain the attention of the viewers.

Dimension 2

Despite the lack of statistical significance, the analysis of Dimension 2, distinguishing narrative from non-narrative concerns, reveals notable differences between the highly- and poorly-rated documentaries. The descriptive statistics shown in Table 2 indicate that the mean score of the highly-rated documentaries (-0.32) is higher (less negative) than that of the poorly-rated documentaries (-0.68). The ANOVA findings, however, show that this difference is not statistically significant ($F(1,58)=0.907$, $p=0.345$).

Table 2. Descriptive statistics for dimension 2. Source: Own processing

Category	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	Minimum	Maximum
Highly Rated	30	-0.32	0.91	0.17	-0.66 - 0.02	-1.82	1.67
Poorly Rated	30	-0.68	1.87	0.34	-1.38 - 0.01	-3.82	3.20
Total	60	-0.50	1.47	0.19	-0.88 - -0.12	-3.82	3.20

There is a significant variance between the two groups according to Levene's test for homogeneity of variances (Levene Statistic = 21.231, $p < 0.001$). This indicates that there may be greater variability in the narrative versus non-narrative scores among the poorly-rated documentaries compared to the highly-rated ones.

Comparing the highly-rated documentaries to the poorly-rated ones, Dimension 2 scores for the former indicate a stronger tendency towards narrative features. It is often a conversational style of narration, combining factual reporting with elements of storytelling, such as personal anecdotes, dialogue excerpts, and first-person narration or viewpoints of participants or interlocutors (*I, we*).

The temporal sequencing, along with the causal connectors (*then, because, as a result*) and the evaluative language, helps the audience in their interpretation and contributes to the narrative cohesion as well. The involvement and empathy are increased with the help of emotionally connoted verbs and adjectives (e.g., *struggled, astonishing, tragic*).

All these features contribute to the storytelling and the generation of narrative flow. Even the inclusion of a few of these elements can increase the relatability and interest of the content. Conversely, the poorly-rated documentaries tend to be founded on the non-narrative discourse, and the more expository or report-based documentaries, which that are not framed in the first person, lack dialogue

and are not emotional. They are informative but not graphic or engaging because they use present-tense exposition and attributive adjectives.

The significant difference in the ratings of the poorly-rated documentaries suggests that the elements of the narration were not used consistently, and the storyline can be unclear or uninteresting. Even though the discrepancy in the scores on Dimension 2 is not that significant, the results still indicate the significance of storytelling in increasing the appeal and effectiveness of the documentaries. Narrative elements such as past-tense narrations, third-person reference, and time order, helps to enhance the coherence and to spark the interest of the viewer, which makes the material easier to remember and more emotional. On the whole, the discussion of Dimension 2 indicates that though the narrative elements do not indicate statistically significant differences between the highly and poorly rated documentaries, they are the factors that contribute to the audience satisfaction in the case of their successful combination.

Dimension 3

The highly-rated documentaries and the poorly-rated ones have significant differences in Dimension 3 (explicit versus situation-dependent discourse). According to the descriptive statistics in Table 3, the mean score for the highly-rated documentaries is lower (-0.57) compared to the poorly-rated ones (0.59). The ANOVA results confirm that this discrepancy is statistically significant ($F(1,58)=5.350$, $p = 0.024$).

Table 3. Descriptive Statistics for Dimension 3. Source: Own processing

Category	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	Minimum	Maximum
Highly Rated	30	-0.57	1.44	0.26	-1.11 - -0.04	-3.24	2.32
Poorly Rated	30	0.59	2.35	0.43	-0.29 - 1.47	-4.18	5.19
Total	60	0.01	2.02	0.26	-0.51 - 0.53	-4.18	5.19

The variances between the two groups are significantly different, as indicated by Levene's test for homogeneity of variances (Levene Statistic = 6.292, $p = 0.015$). This suggests that there is more variation in the explicit versus situation-dependent scores among the poorly-rated documentaries compared to the highly-rated ones.

The dimension with the highest positive mean score was associated with the documentaries that received lower audience ratings, suggesting a stronger viewer preference for situation-dependent

discourse in the texts. Such features, including adverbs of time and place, known as temporal and spatial deixis, provide the discourse with a situational frame and are frequently employed in situation-dependent discourse. The efficiency of this type of discourse in enhancing the areas of relatability and engagement is due to the number of explicit references to the situation aimed at the audience.

In contrast to Dimension 1, which indicates interpersonal involvement and dialogic tone, Dimension 3 is concerned with the explicit location of information in context, i.e., the explanation of when, where, and under what conditions the events take place. In this case, deixis and referential markers are not relational, but informational: they lead the viewer through actual time-space coordinates (e.g., in 2019, at the factory, on this day) rather than create the illusion of a conversation. Contextual accuracy and clarity rather than emotional proximity are therefore captured in this dimension.

The perspectives corresponding to the lower rating of the documentaries contain more explicit information. In this context, the data may seem less interesting or less relevant to the audience. Nominalizations and use of *wh*-pronouns in relative clauses mark this style. These linguistic features reflect less context-specific material. Consequently, the documentaries with such linguistic patterns display a more explicit and informational discourse style, evident in the positive mean scores in Table 3.

The poorly rated films show some inconsistent use of the explicit and situation-dependent characteristics, which leads to a substantial gap among the films. This inconsistency likely results from the difficulty for the screenwriters and producers to vary the tone of communication while balancing the explicit and situation-dependent references. Such inconsistencies may cause viewers to lose interest or become disengaged.

The research is concerned primarily with the specificity of the content of documentaries. The results show that including more situation-dependent discourse increases the odds of audience connection. The reason is that communication is more engaging and more accessible when supported by explicit situational contextualization, which enhances the relevance and immediacy of the message. This situation is frequently mentioned in communication and media theory and practice, whereby situated messages are more easily understood and more engaging (Lu et al. 2025).

The results of the analysis of Dimension 3 reveal the differences between the highly- and the poorly-rated documentaries. It was also discovered that the highly rated documentaries are more inclined to a more concrete and perceptually based subject matter, as compared to the poorly rated documentaries, which are more inclined to use simpler and more direct language. Nevertheless, the

success of a given documentary is frequently associated with its intertextual interaction with other documentaries in terms of similar stylistic, thematic, or narrative patterns. This implies that documentary filmmakers ought to use a communicative approach where they emphasize real-life situations and real-life experiences so that the film can attain a high degree of attachment to the audience.

Dimension 4

In analyzing Dimension 4, the existence and non-existence of overt expression of persuasion, it is evident that there is a difference between the highly rated and poorly rated documentaries. As shown in Table 4, the average score for the highly rated documentaries is -0.94, whereas the poorly rated ones received an average rating of 0.89. The results obtained from the ANOVA test give a value of $F(1,58)=10.571$, $p = 0.002$, suggesting that this differentiation is statistically significant.

Table 4. Descriptive statistics for dimension 4. Source: Own processing

Category	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	Minimum	Maximum
Highly Rated	30	-0.94	1.77	0.32	-1.61 - -0.28	-5.27	3.03
Poorly Rated	30	0.89	2.53	0.46	-0.06 - 1.83	-3.46	6.25
Total	60	-0.03	2.35	0.30	-0.64 - 0.58	-5.27	6.25

Thus, the result of Levene's test of equality of variances is as follows: Levene Statistic = 3.591; $p = 0.063$. Hence, the result shows that there is no significant difference between the variances of the two groups. This implies that, while responding to Dimension 4, the differentiation in scores between the highly rated documentaries and the poorly rated ones is comparable.

The highly-rated documentaries have little prompting or overtly persuasive language in Dimension 4. This finding implies that they do not force the audience into a particular opinion. Instead, such documentaries aim to present facts and real-life stories in a balanced, objective manner. Their objectivity makes them more credible and appealing to emotions without trying to convince the audience directly.

It can be postulated that those documentaries that have received the highest rating do not attempt to influence the viewers to a great extent. On the other hand, the poorly-rated documentaries use more forceful language and may come across as manipulative or less trustworthy. This could lead to a hostile reception from the audience.

The mean score between the two documentary groups is different. It means that the same audience has rated the documentaries differently depending on the extent of overt persuasion or objectivity of their language. The results indicate that excessive use of persuasive appeals will decrease the level of audience approval, and a more factual and balanced method is more acceptable. In other words, less persuasive and more informative documentaries are more popular among the audience.

The analysis of Dimension 4 concludes that the highly- and poorly-rated documentaries differ significantly. The documentaries that are rated highly contain few persuasive elements; they present the facts clearly and objectively to the audience. On the other hand, the poorly-rated documentaries are based more on the strong, direct, persuasive techniques, which may be viewed as biased or manipulative. Therefore, the success of documentaries depends on their capability to be objective and balanced, using restrained and respectful persuasive methods that do not compromise the audience.

Dimension 5

Dimension 5 indicates a difference between the highly rated documentaries and the poorly rated ones. In this case, the comparison is made between the impersonal (abstract) style and the non-impersonal (non-abstract) style. Table 5 indicates that the mean score of the highly rated documentaries is equal to -1.94 and the mean score of the poorly rated documentaries is equal to -0.89. The table ANOVAs indicate that the difference is significant ($F(1,58) = 7.728, p = 0.007$).

Table 5. Descriptive statistics for dimension 5. Source: Own processing

Category	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	Minimum	Maximum
Highly Rated	30	-1.94	0.83	0.15	-2.25 - -1.63	-3.53	-0.09
Poorly Rated	30	-0.89	1.89	0.35	-1.59 - -0.18	-3.50	4.36
Total	60	-1.41	1.54	0.20	-1.81 - -1.01	-3.53	4.36

In the case of the test of homogeneity of the variances, the Levene test indicates that the variances of the two groups differ significantly (Levene Statistic = 12.760, $p = 0.001$). This observation shows that the scores in Dimension 5 of the poorly-rated documentaries are more dispersed than those of the highly-rated documentaries.

The documentaries with high ratings are more inclined towards the non-abstract style, which is clear and rather limited in objectivity without being too impersonal. Non-abstract language is, however, well used in such productions to communicate their messages, making them more intellectual and credible.

The mean scores of the poorly rated documentaries have gone down, which is more towards the less personal style. As a result, the explanations that are regarded as non-abstract are easier to understand by the addressees since more definite words are used. The information is also easily organized in this structure.

The poor framing of the aspects of the films that received the lowest ratings demonstrates that non-abstract and non-impersonal concepts were not applied properly. The abstract language was sometimes used in these documentaries, and this made it difficult to get a clear and coherent presentation.

The highly-rated documentaries exhibit a non-abstract style, which is objective and at the same time understandable, clear, and concrete without being too detached from the audience. Thus, this paper shows that the effectiveness of documentaries is best achieved when the style is non-abstract and moderately impersonal, yet still approachable. In this case, impersonality does not imply distance or detachment; it means being objective and analytically precise without sacrificing accessibility. The results show that the documentaries that are rated highly have less complicated sentence structures and clear presentation, which is credible and easy to comprehend for the viewers. This balance means that the audience is entertained by thought-provoking and emotionally engaging documentaries.

In general, Dimension 5 shows the difference between the highly rated documentaries and the poorly rated ones. The viewers' reactions show that the most successful productions are objective and well-organized in tone, neither too abstract nor too informal. This restrained application of an impersonal style, based on clarity and accuracy, increases audience interest and quality of perception. Therefore, filmmakers seeking a higher rating should strive for a language style that is both analytical and communicative, so that their material is not only informative but also enjoyable.

5. Conclusion

This research aimed to provide a comprehensive overview of the lexical and grammatical patterns of the highly and poorly rated documentaries for the years 2023 and 2024. This was achieved by using Biber's MD Analysis. Consequently, the aim of classifying the documentaries according to their IMDb ratings led to the conclusion that there are certain features of the language that can be the cause of the film's success or failure. Thus, it is possible to reveal the importance of the linguistic style in making documentaries, which is beneficial for further research and the practical work of filmmakers.

The results show that features of involved discourse are more prominent in the highly rated documentaries. This kind of involvement in style creates rapport with the audience, thereby making the content more familiar. Meanwhile, the poorly rated documentaries were characterized by a much less involved approach and a discourse style that makes them less appealing to the audience.

With reference to Dimension 2, there is a slightly higher frequency of the use of the aspects related to the non-narrative features in the poorly rated documentaries. The documentaries which include a mix of the components of the narrative and non-narrative into their framework are more appealing and, therefore, more engaging to viewers.

The results of Dimension 3 indicate that the discourse of the highly rated documentaries is closer to situation-dependent, and therefore includes the time and place adverbs that fix the discourse in a particular context. Such situational grounding increases the film's realism and, therefore, audience appeal. The poorly rated ones, on the other hand, can be very provocative and even abstract, which could be a reason for the audience perceiving the documentary as less appealing.

On Dimension 4, the highly rated documentaries are less exposed to the attempt of direct persuasion. They present information and data without bias, which enables the audience to draw their own conclusions. This is in contrast to the more forceful style seen in the poorly rated documentaries, which can be assertive or, in some cases, biased, thereby compromising credibility.

Lastly, Dimension 5 finds that the preference for the highly rated documentaries is related to a non-abstract style. Conversely, the poorly rated documentaries, which are less non-abstract, might not be as refined as the audience anticipates from high-quality productions.

To sum up, this study confirms the opinion that lexical and grammatical patterns are significant to the effectiveness of documentary films. Using the MD Analysis of Biber, the paper has revealed the characteristics of the corpora that led to the high level of impact of the documentaries. This knowledge can inform documentary filmmakers on how to structure their narrative form and language use to increase audience consumption.

Based on this, future studies can expand on this study by applying other types of linguistic analysis or by utilizing a larger dataset. Moreover, in case the traditional content analysis is performed on the answers of the audience, more delicate patterns can be identified in the context of the influence of

lexical and grammatical decisions on the viewer. This research can create a possibility of conducting further studies on the correlation between language and media in order to enhance the future of documentaries.

List of abbreviations

ANOVA: Analysis of Variance

IMDb: Internet Movie Database

MAT: Multidimensional Analysis Tagger

SPSS: Statistical Package for Social Sciences

MD: Multidimensional

References

1. Abdul, R.G., Jemilatu, S.D. & Morolake, A. (2023). A contextual analysis of documentary film as a product and tool for academic exercise. In *European journal of communication and media studies*, 2 (4), p. 25-35. DOI: <https://doi.org/10.24018/ejmedia.2023.2.4.22>
2. Ali, M. & Ali, S. (2023). Exploring lexico-grammatical patterns in Saudi press reportage: A multidimensional analysis. In *Arab world English journal*, 14 (1), p. 502-512. DOI: <https://doi.org/10.24093/awej/vol14no1.32>
3. Ali, S. (2024a). A multidimensional analysis of academic writing: A comparative study of Saudi and British University students' writing. In *World journal of English language*, 14 (2), p. 452-461. DOI: <https://doi.org/10.5430/wjel.v14n2p452>
4. Ali, S. (2024b). Locating the intersection of generative artificial intelligence and human English writing skills: A comparative study. In *Arab world English journal*, 1 (1), p. 112-123. DOI: <https://doi.org/10.24093/awej/ChatGPT.7>
5. Ali, S. (2024c). Reddit's linguistic diversity evolution: A multivariate study. In *Lege artis. Language yesterday, today, tomorrow*. Trnava: University of SS Cyril and Methodius in Trnava, 9 (1), p. 2-19. DOI: <https://doi.org/10.34135/lartis.24.9.1.01>
6. Ali, S. & Abdulhaleem, E. (2024). Multidimensional analysis of differences in lexico-grammatical patterns in Saudi and British linguistics and English language teaching research articles. In *Arab world English journal*, 15 (2), p. 52-72. DOI: <https://doi.org/10.24093/awej/vol15no2.4>
7. Ali, S. & Afzal, N. (2024). A multidimensional analysis of human and ChatGPT-generated English translations of Arabic film. In *World journal of English language*, 14 (4), p. 632. DOI: <https://doi.org/10.5430/wjel.v14n4p632>
8. Aufderheide, P. (2007). *Documentary film: A very short introduction*. Oxford University Press.
9. Biber, D. (2019). Multi-dimensional analysis: A historical synopsis. In *Multi-dimensional analysis: Research methods and current issues*. Sardinha, T.B. & Pinto, M.V. (eds.). London: Bloomsbury, p. 11-26. DOI: <https://doi.org/10.5040/9781350023857.0009>
10. Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins Publishing.
11. Biber, D. (1991). *Variation across speech and writing*. Cambridge: Cambridge University Press.

12. Biber, D., Conrad, S., Reppen, R., Byrd, P. & Helt, M. (2002). Speaking and writing in the university: A multidimensional comparison. In *TESOL quarterly*, 36 (1), p. 9-48. DOI: <https://doi.org/10.2307/3588359>
13. Deb, J. (2022). The development of documentary film language from the silent era to the present. In *International journal of research publication and reviews*, 3 (10), p. 1224-1226. DOI: <https://doi.org/10.55248/gengpi.2022.3.10.47>
14. Hertel, T.J. & Harrington, S. (2015). Promoting cultural and linguistic competence with documentary film in Spanish. In *Hispania*, 98 (3), p. 549-569. DOI: <https://doi.org/10.1353/hpn.2015.0094>
15. IBM Corp. (2024). *IBM SPSS Statistics for Windows, Version 29.0*. Armonk, NY: IBM Corp.
16. IMDb. (2024). *Internet Movie Database*. May 19, 2025. Available at: <https://www.imdb.com>.
17. Lu, Y., Xie, Z., Chiu, D. K. & Kafeza, E. (2025). Digital exposure and emotional response: Public discourse on mandatory IP location disclosure in Chinese social media. In *Systems*, 13 (11), 975. DOI: <https://doi.org/10.3390/systems13110975>
18. Murat, K. (2022). Inconvenient truth and Darwin's nightmare movies for environmental discourse: A comparative analysis. In *IDEALKENT*, 13 (37), p. 1566-1588. DOI: <https://doi.org/10.31198/idealkent.1116634>
19. Nichols, B. (2017). *Introduction to documentary* (3rd ed.). Bloomington: Indiana University Press.
20. Nini, A. (2019). *Multidimensional analysis tagger (MAT)* [Computer software]. Available at <https://sites.google.com/site/multidimensionaltagger>
21. Plantinga, C. (2015). *Rhetoric and representation in nonfiction film*. Grand Rapids, MI: Schuler Books.
22. Ponder, M.L., Kasymova, S., Coleman, L.S., Goodman, J.M. & Smith, C. (2023). COVID-19 frames the news: An examination of race and pandemic frames in newspaper coverage. In *Howard journal of communications*, 35 (2), p. 166-182. DOI: <https://doi.org/10.1080/10646175.2023.2263737>
23. Rahman, T.A.F.T.A., Chik, A.R. & Sahrir, M.S. (2018). The use of documentary film text in developing Arabic writing skills: What the experts say. In *LSP international journal*, 4 (2), p. 59-71. DOI: <https://doi.org/10.11113/lspi.v4n2.56>
24. Renov, M. (2004). *The subject of documentary*. Minneapolis: University of Minnesota Press.
25. Sardinha, T. B. (2024). AI-generated vs human-authored texts: A multidimensional comparison. In *Applied corpus linguistics*, 4 (1), p. 100083. DOI: <https://doi.org/10.1016/j.acorp.2023.100083>
26. Suleimenova, A. E. (2022). Narrative analysis of documentary films by Kanat Beisekeyev as an example of new media publicism. In *Herald of journalism*, 65 (3), p. 77-85. DOI: <https://doi.org/10.26577/HJ.2022.v65.i3.08>
27. Winston, B. (2019). *Claiming the real II: Documentary: Grierson and beyond*. London: Bloomsbury.
28. Winston, B., Vanstone, G., & Chi, W. (2017). *The act of documenting: Documentary film in the 21st century*. London: Bloomsbury.

Contact data

	<i>name:</i> <i>academic title/rank:</i> <i>department:</i> <i>institution:</i> <i>e-mail:</i> <i>interest:</i>	Sadia Ali Assistant Professor Department of English Language and Literature College of Science and Humanities, Prince Sattam bin Abdulaziz University, Al-Kharj, Kingdom of Saudi Arabia sad.ali@psau.edu.sa ESP, EFL, world Englishes, corpus linguistics, AI, and corpus development.
---	---	---

APPENDIX A

Co-occurring linguistic features on the five textual dimensions of the 1991 MD analysis (p. 89-90)

Dimension 1: Involved vs. Informational Discourse Positive features <i>That</i> deletion Verb (uninflected present, imperative & third person) Second person pronoun/possessive Verb <i>do</i> Demonstrative pronoun Adverb/Qualifier – emphatic (e.g., <i>just</i>, <i>really</i>) First-person pronoun/possessive Verb <i>be</i> (uninflected present tense, lexical and auxiliary) Subordinating conjunction – causative Discourse particle Nominal pronoun Adverbial – hedge Adverbial/Qualifier – amplifier <i>Wh</i>-question Modals of possibility Coordinating conjunction - clausal connector <i>Wh</i>-clause Negative features Nouns (excluding -ing forms functioning as gerunds) Preposition Attributive adjective Dimension 2: Narrative vs. Non-Narrative Concerns Positive features Past tense verb Third-person pronoun (except <i>it</i>) Verb - perfect aspect 	Dimension 3: Explicit Vs. Situation-Dependent Discourse Positive features <i>Wh</i>-pronoun-relative clause-object position <i>Wh</i>-pronoun-relative clause-subject position <i>Wh</i>-pronoun-relative clause-object position with prepositional fronting (pied-piping) Nominalization Coordinating conjunction – phrasal connector Singular noun-nominalization Negative features Adverb of time Adverb of place Adverb – Other Dimension 4: Overt Expression of Argumentation/ Persuasion Positive features Infinitive verb Modal of prediction Persuasive verb Subordinating conjunction-conditional Modal of necessity Adverb within auxiliary Negative features Private verb Public verb Third-person pronoun Hedge Dimension 5: Impersonal (Abstract) VS. Non-Impersonal (Non-Abstract Style) Positive features Adverbial conjuncts
--	--

Negative features Present tense verb Place adverbial <i>That</i> deletion Pronoun <i>it</i>	Agentless passive verb Passive verb + <i>by</i> Passive post-nominal modifier Subordinating conjunction - Other Negative features (none)
--	---