LEGE ARTIS

REDDIT'S LINGUISTIC DIVERSITY EVOLUTION:
A MULTIVARIATE STUDY[1]

Sadia Ali iD

Department of English Language and Literature, College of Science and Humanities,

Prince Sattam bin Abdulaziz University, Al-Kharj, Kingdom of Saudi Arabia

Abstract: The linguistic diversity of Reddit posts is investigated in this study during three different periods: 2009, 2016, and 2023. Reddit posts in the *News, Sports*, and *Stories* categories were chosen to provide a comprehensive overview of communications on Reddit. 4500 posts were analysed using the Multidimensional Analysis Tagger for linguistic analysis. Statistical analyses, including analysis of variance and post-hoc Tukey tests, reveal significant trends and variations, showing a shift towards more informational discourse while maintaining core narrative features.

Keywords: discourse development; linguistic diversity; Multidimensional Analysis; online communication; Reddit.

1. Introduction

As digital communication continues to change, platforms such as Reddit have become important hubs for sharing information and opinions. Though there is much research on online conversations, there is still a lack of knowledge on how language usage on platforms like Reddit has evolved. This study seeks to fill that gap by examining linguistic changes in Reddit posts from 2009, 2016, and 2023.

Frequently called "the front page of the internet," Reddit functions as a microcosm of more general trends in digital communication. Its many subreddits – covering sports, news, and various storytelling communities – offer a rich tapestry of conversation that captures the changing priorities of its user

---

base and the dynamic nature of online interactions. Prior research has mainly concentrated on particular groups or singular incidents, largely ignoring the long-term changes in conversation that can provide a more profound understanding of the platform's development.

In this study, Biber's Multidimensional Analysis framework (1991) is utilised to compare Reddit posts during three different phases in a systematic manner. Analysing linguistic features across various dimensions aims to identify trends and changes that contribute to a thorough comprehension of the evolution of Reddit discourse. The study concentrates on three main categories – *News, Sports,* and *Stories* – that have been chosen to represent a wide range of content on Reddit.

Using historical data retrieved from the Wayback Machine, a service provided by Internet Achieve allowing users to access older versions of webpages, we carefully ranked subreddits according to their popularity level at each stage to guarantee the validity and precision of our analysis. This method allowed us to circumvent the restrictions imposed by Reddit Metrics' 2023 discontinuation and to guarantee that our dataset appropriately represents the most significant subreddits of each time frame. With 1,000 tokens per text, each text was tagged using the Multidimensional Analysis Tagger (MAT), a methodological improvement that boosts the accuracy and comprehensiveness of our linguistic analysis. The MAT ensures the precise and thorough tagging of 150+ linguistic features in the texts.

The use of the Statistical Package for the Social Sciences (SPSS) for in-depth analysis, which includes descriptive statistics, ANOVA, and post-hoc Tukey tests, highlights the statistical rigour of our work. These techniques help us pinpoint important trends and variations between stages and categories, giving us a detailed understanding of the chronological changes in Reddit conversation.

The study's conclusions will significantly impact the understanding of how digital communication has evolved. By presenting the changes in linguistic practices on Reddit, we provide fresh perspectives on how online discourse is influenced by and adapts to larger social and technological developments. This research bridges a significant void in the current body of knowledge and lays the groundwork for future investigations into the intricacies of digital communication on developing platforms.

In the sections that follow, we will present a review of the related literature, our methodology, our findings, and their implications for the study of digital discourse. This study sheds light on how the

language of one of the most popular websites has changed, offering essential insights into how online communication is evolving.

2. Literature review

Using Biber's Multidimensional Analysis framework has greatly advanced the study of linguistic variation in digital communication. This method, which was first created to compare speech and writing (Biber 1991), has been applied to a wide range of discourse forms, including those found on digital platforms. To expand on this body of work, the current study looks at Reddit posts in the *News, Sports*, and *Story* categories from three different time periods: 2009, 2016, and 2023.

Since the introduction of social media, digital communication has undergone significant change. Early research emphasised the highly engaging conversational style prevalent in early Reddit platforms (Herring 2010). As these platforms have developed over time, there has been a discernible shift in discourse toward greater information. This pattern is consistent with findings from other studies (Androutsopoulos 2014; Tagg 2015) that showed a comparable change in other digital communication contexts. Reddit is a well-known social media platform that presents a special chance to research these changes. Reddit encompasses various discourse styles, with its diverse subreddits catering to different interests and topics. Prior studies on Reddit have frequently concentrated on particular groups or occasions, such as analysing political debates or disseminating false information (Medvedev et al. 2009; Starbird 2017; 2019). However, comprehensive studies that monitor linguistic changes over long time spans are comparatively rare (Noble et al. 2021).

Herring et al. (2005) have described in their study that social media discourse can differ dramatically over time and in various contexts. The results also indicate that early internet interactions changed from intimate and personal to more informational and content driven.

Numerous platforms, such as blogs, forums, and social networking sites, have shown this shift (Androutsopoulos 2014). These linguistic trends on Reddit have been studied by looking into particular subreddits. As an illustration, Medvedev et al. (2019) examined political conversations on Reddit. They discovered that, as time passed, the conversation grew less interactive and more structured, reflecting broader movements toward informational content. Starbird (2017) also emphasised how false information is disseminated on Reddit during times of crisis, demonstrating a change in content and the methods by which it is disseminated.

Various interconnected factors have influenced the linguistic development of Reddit. Firstly, the recurring topic of the legalization of marijuana has been influential. Hemmatian et al. (2022) examined discussions on this topic from 2008 to 2019, showing how these dialogues combine personal stories to shape attitudes, as further supported by Ananthasubramaniam (2023). Secondly, political discussions on Reddit, analysed through over 155 million user comments in political subforums, display different forms of engagement—ranging from deliberative to demagogic. Flesch's study (2022) illustrates how diverse forms of discourse foster distinct community interactions. Thirdly, the internet's impact on politics is significant. Papakyriakopoulos et al. (2023) traced word diffusion across subreddits from 2005 to 2017, highlighting how terms associated with the alt-right and internet controversies spread throughout the community.

Finally, the role of writing styles, including the use of emojis and emoticons, is crucial. Willaert et al. (2021) stressed the importance of sociolinguistic studies to consider factors such as gender, age, and ethnicity, emphasizing the complexity of language evolution on digital platforms. These studies demonstrate how Reddit's language and discourse are influenced by culture, society, and politics, reflecting broader societal trends and individual identities.

Biber's Multidimensional Analysis framework has proven essential in comprehending the changes, specifically in the areas of linguistic variation and style shifts across different genres. Through analysing linguistic characteristics in various dimensions, scholars can detect trends that may need to be more readily discernible using conventional techniques. This framework provides insights into how language use changes in response to shifting technological, social, and temporal contexts (Ali 2024a; 2024b; Ali & Abdulhaleem 2024; Ali & Thompson 2022), and it has been successfully applied to various digital communication forms (Biber 1988; Conrad & Biber 2001; Egbert & Biber 2016). This study expands on these findings using the framework of Reddit, a platform well-known for its dynamic and varied user-generated content.

The study's conclusions carry significant implications for understanding digital discourse. The language used in Reddit posts offers an insight into larger patterns in online communication as the platform develops. Users appear to prioritise content delivery over personal engagement, as evidenced by the shift towards a less interactive and informative discourse. This pattern aligns with more general changes in digital communication, where information clarity and efficiency frequently trump interpersonal interaction (Tagg 2015). Even though the medium may change, some fundamental communication patterns seem to hold, because of the consistency of linguistic features

related to narrative and explicit discourse throughout the phases. Future research examining how these constant components interact with changing trends to shape the nature of digital communication can build upon this consistency.

This literature review emphasises the importance of leveraging Biber's Multidimensional Analysis framework to investigate linguistic variations on platforms such as Reddit. Through a thorough examination of previous literature, this study seeks to enhance our comprehension of how language adapts to technological advancements and societal shifts. The results highlight the ever-changing landscape of digital discourse and pave the way for future investigations into these evolving patterns.

3. Methodology

This research conducts a thorough examination of the linguistic diversity found in Reddit posts during three separate time periods: 2009 (Phase 1), 2016 (Phase 2), and 2023 (Phase 3). Three popular categories – *News, Sports*, and *Stories* – were used to select posts in order to guarantee a comprehensive portrayal of Reddit discourse during these times. The selection process was meticulous, using Reddit metrics data from the past, retrievable using the Wayback Machine, to determine which subreddits were most prevalent in each category in those particular years. This method ensured that the selected subreddits reflected the more significant themes and trends common on Reddit at the time.

Two popular subreddits per category were selected to provide a representative discourse sample. In particular, r/worldnews and r/announcements were included in the news category; r/sports and r/sportsToday were included in the sports category; and r/todayIlearned and r/WTF were included in the stories category. This deliberate choice guaranteed the dataset's richness and diversity by enabling a thorough capture of discourse across several topic areas.

In each phase, ten texts were gathered per category, totalling 30 texts per phase and 90 texts for the entire study. Each text contained 50 posts, resulting in a comprehensive dataset of 1500 posts per phase and 4,500 posts in total. The Wayback Machine was essential for accessing Reddit metrics across all years studied, since the Reddit Metrics service was no longer available after 2023. Retrieving this data enabled the precise identification of the most popular subreddits and maintained the accuracy and relevance of the data used in the analysis.

The MAT was utilised to improve the dependability and precision of our linguistic analysis. Using MAT guarantees a high level of accuracy and dependability in capturing the subtleties of Reddit discourse, thanks to its thorough annotation of linguistic features. Compared to the earlier studies, which usually featured fewer tokens, each text was tagged using 1,000 tokens, a significant increase. In order to provide more detailed insights into the differences in discourse, a decision was made to design a more thorough linguistic profile of every text. In order to accurately capture elements pertinent to Biber's Multidimensional Analysis framework, 150+ linguistic features were meticulously annotated during the tagging process. Because of this process, the data was solid and dependable, providing a strong basis for further analysis. This framework aligns with the best practices in quantitative linguistic research, guaranteeing the statistical validity and significance of the conclusions reached (Field 2013).

Statistical analysis was applied to find important patterns and variations between the phases and categories, which was carried out using SPSS. Descriptive statistics were first computed to obtain a basic idea of the distribution of the data and the central tendencies. After that, ANOVA was used to determine whether the observed variations were statistically significant, by evaluating the differences in the 150+ linguistic features between the phases. This allowed for a more detailed understanding of temporal shifts in Reddit discourse, by allowing post-hoc Tukey tests to identify particular phase pairs with significant differences. This multifaceted statistical approach guaranteed a thorough examination of the data by emphasising significant trends and changes over time.

To summarise, the methodological steps involved the following: using Reddit metrics accessed through the Wayback Machine, the researchers selected subreddits based on their historical popularity; collecting ten texts per category per phase, which produced a total of 90 texts and 4,500 posts; using the MAT to tag each text with 1,000 tokens; and conducting a thorough statistical analysis using SPSS, which included ANOVA, post-hoc Tukey tests, and descriptive statistics. This methodical and rigorous approach guaranteed the study's validity and the accuracy of its conclusions, which also offered insightful information about how Reddit discourse has changed over time.

4. Discussion and analysis

Dimension 1

In Biber's Multidimensional Analysis framework, Dimension 1 is labelled as "Involved versus Informational discourse" by Biber (1991). Involved discourse includes personal interactions, personal opinions, and shared knowledge, using first- and second-person pronouns, private verbs, and contractions. On the other hand, informational discourse is characterised by a greater use of nouns,

prepositions, and attributive adjectives, indicating a focus on content rather than interpersonal communication.

Table 1 summarises the descriptive statistics for Dimension 1 across the three phases (2009, 2016, 2023). The means and standard deviations indicate significant variation across phases, suggesting changes like discourse over time.

Table 1. Descriptive Statistics for Dimension 1. Source: own processing

Phase	N	Mean	Std. Deviation	Std. Error
2009	30	-13.6113	9.17673	1.67543
2016	30	-24.4497	3.94878	0.72095
2023	30	-23.6687	4.84080	0.88381
Total	90	-20.5766	8.04916	0.84846

The data clearly shows that Dimension 1 scores have consistently been negative, showing a prevalence of informational discourse features in all phases. A noticeable trend towards more negativity in scores from 2009 to the later phases (2016 and 2023) indicates a growth in informational content over time.

The ANOVA results (Table 2) show a statistically significant difference between the phases on Dimension 1 (F = 26.684, p < 0.001). This suggests that the changes in Dimension 1 scores across the phases are not due to random variation.

Table 2. ANOVA for Dimension 1. Source: own processing

Source	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2192.294	2	1096.147	26.684	0.000
Within Groups	3573.921	87	41.080		
Total	5766.215	89			

The results of Tukey's HSD post-hoc tests (Table 3) show that although the difference between Phase 2 and Phase 3 is not significant, the mean difference between Phase 1 and Phase 2, as well as between Phase 1 and Phase 3, is significant (p < 0.001). This verifies that there was a significant change between Phase 1 (2009) and the phases that followed (2016, 2023).

Table 3. Tukey HSD Post Hoc Comparisons for Dimension 1. Source: own processing

Comparison	Mean Difference	Std. Error	Sig.	95% CI Lower Bound	95% CI Upper Bound
Phase 1 vs 2	10.83833*	1.65488	0.000	6.8923	14.7844
Phase 1 vs 3	10.05733*	1.65488	0.000	6.1113	14.0034
Phase 2 vs 3	0.78100	1.65488	0.885	-3.1650	3.1650

The results show a consistent upward trend in informational discourse over time. This change can be ascribed to a number of things, such as modifications in Reddit communication patterns and the growing focus on content-driven interactions. The more conversational style characterising early social media interactions may be reflected in the 2009 higher involvement features. The posts have shifted in focus from 2016 to 2023, presumably due to the platform's transformation into a more organised and information-focused medium.

These findings will significantly impact the way that discourse develops on digital platforms. The growing emphasis on information delivery over interpersonal interaction implies that users are prioritising content delivery, which may impact how information is shared and consumed. This change aligns with more general developments in digital communication, where interactive engagement is frequently subordinated to information delivery efficiency.

Overall, the analysis of Dimension 1 provides valuable insights into the changing dynamics of discourse on Reddit, highlighting a move towards more informational and less involved communication over time.

Dimension 2

In Biber's Multidimensional Analysis framework, "Narrative" and "Non-Narrative" discourse types are contrasted in Dimension 2. Features like past tense verbs, third-person pronouns, and perfect aspect verbs that communicate temporality and human agency are characteristics of narrative discourse. Conversely, present tense verbs and attributive adjectives characterise non-narrative discourse, suggesting a focus on characterising states and qualities instead of events and actions.

Table 4 summarises the descriptive statistics for Dimension 2 across the three phases (2009, 2016, and 2023). The means and standard deviations indicate moderate variation across phases, suggesting an increase in the narrative nature of discourse over time.

Table 4. Descriptive Statistics for Dimension 2. Source: own processing

Phase	N	Mean	Std. Deviation	Std. Error
2009	30	-4.8970	0.88998	0.16249
2016	30	-4.9417	0.98844	0.18046
2023	30	-5.3063	0.93951	0.17153
Total	90	-5.0483	0.94765	0.09989

These statistics show that the mean scores of Dimension 2 are consistently negative, suggesting a prevalence of non-narrative elements throughout all phases. A subtle uptick in negativity from 2009 to 2023 implies a shift towards increased non-narrative content over time.

The ANOVA results (Table 5) show that the effect of the phases on Dimension 2 scores is not statistically significant ($F = 1.711$, $p = 0.187$). This suggests that the variations in Dimension 2 scores across the phases are not substantial enough to be considered beyond random variation.

Table 5. ANOVA for Dimension 2. Source: own processing

Source	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3.025	2	1.513	1.711	0.187
Within Groups	76.901	87	0.884		
Total	79.926	89			

Post hoc tests using Tukey's HSD (Table 6) confirm no significant differences between any pairs of the phases. This reinforces the conclusion that the narrative characteristics of the posts have remained relatively stable over the three time periods.

Table 6. Tukey HSD Post Hoc Comparisons for Dimension 2. Source: own processing

Comparison	Mean Difference	Std. Error	Sig.	95% CI Lower Bound	95% CI Upper Bound
Phase 1 vs 2	0.04467	0.24275	0.982	-0.5342	0.6235
Phase 1 vs 3	0.40933	0.24275	0.216	-0.1695	0.9882
Phase 2 vs 3	0.36467	0.24275	0.295	-0.2142	0.9435

The results suggest a slight increase in non-narrative discourse on Reddit, although this change is not considered statistically significant. This shift could be due to the changing nature of the platform, with users potentially preferring descriptive and state-oriented content over recounting events. Despite this, the overall narrative characteristics of the subreddits remain stable, indicating that the

core communication practices such as emphasizing storytelling and personal anecdotes and engaging in discussions that follow a clear, narrative thread have not significantly shifted towards a non-narrative focus.

These findings indicate that despite a possible minor rise in the number of linguistic features related to non-narrative elements, Reddit posts' general narrative content has stayed mostly stable over the three phases. The narrative characteristics exhibit consistency, suggesting that users balance narrating events and describing states despite the platform's dynamic changes.

Overall, the Dimension 2 analysis shows that narrative versus non-narrative concerns in Reddit posts have remained stable over time, with only slight, non-significant differences seen in all three phases. This shows that the narrative contents have not changed much over time, indicating a consistent user communication style on the platform.

Dimension 3

"Explicit" and "Situation-Dependent" discourse is contrasted on Dimension 3 of Biber's Multidimensional Analysis framework. *Wh*-relative clauses, pied-piping relative clauses, and nominalisations are examples of characteristics that define explicit discourse and offer precise and unambiguous information. Conversely, situation-dependent discourse depends on immediate context for interpretation; it contains elements such as adverbs that indicate a dependence on the situational context as well as time and place adverbs.

Table 7 provides an overview of the descriptive statistics on Dimension 3 for each of the three phases (2009, 2016, and 2023) of the study. The moderate variation across the phases indicated by the means and standard deviations suggests that the explicitness of the discourse has changed over time.

Table 7: Descriptive Statistics for Dimension 3. Source: own processing

Phase	N	Mean	Std. Deviation	Std. Error
2009	30	2.3193	1.01669	0.18562
2016	30	2.4540	1.11792	0.20410
2023	30	1.9073	0.82038	0.14978
Total	90	2.2269	1.00897	0.10635

These statistics indicate that the mean scores of Dimension 3 are constantly favourable, showing that the explicit discourse features predominate throughout all the phases. The mean score does, however,

show a slight decline from 2016 to 2023, indicating a slight shift from the more explicit toward more situation-dependent discourse.

The ANOVA results (Table 8) show that the effect of the phase on Dimension 3 scores is not statistically significant (F = 2.469, p = 0.091). This suggests that the variations in Dimension 3 scores across the phases are not substantial enough to be considered beyond random variation.

Table 8. ANOVA for Dimension 3. Source: own processing

Source	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4.867	2	2.434	2.469	0.091
Within Groups	85.737	87	0.985		
Total	90.604	89			

Post-hoc tests using Tukey's HSD (Table 9) confirm no significant differences between any of the pairs of phases. This reinforces the conclusion that the explicitness of the posts has remained relatively stable over the three time periods.

Table 9. Tukey HSD Post Hoc Comparisons for Dimension 3. Source: own processing

Comparison	Mean Difference	Std. Error	Sig.	95% CI Lower Bound	95% CI Upper Bound
Phase 1 vs 2	-0.13467	0.25632	0.859	-0.7459	0.4765
Phase 1 vs 3	0.41200	0.25632	0.248	-0.1992	1.0232
Phase 2 vs 3	0.54667	0.25632	0.089	-0.0645	1.1579

The results show a slight trend, though not a statistically significant one, towards more situation-dependent discourse in 2023. This slight change may be explained by how Reddit is developing as a platform, where users may be depending more and more on the situational context to communicate information, reflecting more instantaneous and contextual interactions. Nonetheless, the consistency of explicit features implies that there has not been much of a shift in these subreddits' fundamental communication techniques between the explicit and situation-dependent focuses.

These findings imply that, although situation-dependent features may have slightly increased in recent years, Reddit posts' overall explicitness level has remained mostly stable over the three phases. This consistency in explicit characteristics implies that users still value unambiguous and clear communication despite the changing platform dynamics.

Overall, the Dimension 3 analysis shows that explicit versus situation-dependent concerns in Reddit posts have remained stable over time, with only slight, non-significant differences found in all the three phases. This implies that there has been little change in the basic discourse practices regarding explicit content on the platform.

Dimension 4

In Dimension 4 of Biber's Multidimensional Analysis framework, the "Overt Expression of Argumentation/Persuasion" is contrasted with non-argumentative discourse that lacks these characteristics. Features that collectively signal an intent to influence the reader's beliefs or actions, such as necessity modals, suasive verbs, conditional subordination, and prediction models, are characteristics of persuasive discourse. On the other hand, a discourse that does not perform a persuasive function, i.e., non-argumentative discourse, lacks these argumentative markers and concentrates more on providing facts or descriptions.

Table 10 provides an overview of the descriptive statistics of Dimension 4 for each of the three phases (2009, 2016, and 2023) of the study. The significant variation across the phases indicated by the means and standard deviations suggests that the use of persuasive features has changed over time.

Table 10. Descriptive Statistics for Dimension 4. Source: own processing

Phase	N	Mean	Std. Deviation	Std. Error
2009	30	-5.4243	1.96778	0.35927
2016	30	-6.9147	1.06690	0.19479
2023	30	-7.1697	1.25708	0.22951
Total	90	-6.5029	1.65730	0.17470

These figures indicate that the mean scores of Dimension 4 are continuously low, indicating a lack of persuasive characteristics in all the phases. The mean scores do, however, show a discernible decline between 2009 and 2023, indicating a progressive increase in the absence of persuasive content.

The ANOVA results (Table 11) show the significant effect of the phase on Dimension 4 scores ($F = 12.136$, $p < 0.001$). This suggests that the changes in Dimension 4 scores across the phases are not due to random variation.

Table 11. ANOVA for Dimension 4. Source: own processing

Source	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	53.323	2	26.662	12.136	0.000
Within Groups	191.129	87	2.197		
Total	244.452	89			

The results of Tukey's HSD post-hoc tests (Table 12) show that although the difference between Phase 2 and Phase 3 is not significant, the mean difference between Phase 1 and Phase 2, as well as between Phase 1 and Phase 3, is significant ($p < 0.001$). This verifies that there was a significant change between Phase 1 (2009) and the phases that followed (2016, 2023).

Table 12. Tukey HSD Post Hoc Comparisons for Dimension 4. Source: own processing

Comparison	Mean Difference	Std. Error	Sig.	95% CI Lower Bound	95% CI Upper Bound
Phase 1 vs 2	1.49033*	0.38270	0.001	0.5778	2.4029
Phase 1 vs 3	1.74533*	0.38270	0.000	0.8328	2.6579
Phase 2 vs 3	0.25500	0.38270	0.784	-0.6575	1.1675

The results show a distinct downward trend over time in the overt demonstration of argumentation and persuasion. There are a number of reasons for this change, such as how Reddit users communicate now and how informational and descriptive content is valued more highly than persuasive exchanges. The prevalence of persuasive features in 2009 might indicate a more debate-focused approach, which was common in the early days of social media as users participated in discussions and arguments. The posts' decreased emphasis on persuasion between 2016 and 2023 may be attributed to the platform's transition to a more informational medium.

These findings significantly impact our comprehension of the evolution of discourse on digital platforms. As indicated by the decline in persuasive features, users may favour the delivery of factual content over attempts to sway opinions. This could impact how information is interpreted and used. This change is also consistent with more general patterns in digital communication, where the speed at which information is delivered frequently trumps the ability to engage in debate.

The examination of Dimension 4 sheds light on how Reddit discourse evolves over time, emphasising a shift away from overtly persuasive communication and toward more neutral discourse dynamics.

Dimension 5

In Dimension 5 of Biber's Multidimensional Analysis framework, the discourse types "Impersonal (Abstract)" and "Non-Impersonal (Non-Abstract)" are contrasted. Abstract discourse is distinguished by a depersonalised and formal communication style facilitated by conjuncts, agentless passives, past participial clauses, and other adverbial subordinators. Conversely, context-dependent and more intimate non-abstract discourse lacks these characteristics.

Table 13 provides an overview of the descriptive statistics of Dimension 5 for each of the three phases (2009, 2016, and 2023) of the study. The moderate variation across the phases indicated by the means and standard deviations suggests that the use of the features related to abstract discourse has changed over time.

Table 13. Descriptive Statistics for Dimension 5. Source: own processing

Phase	N	Mean	Std. Deviation	Std. Error
2009	30	-2.4777	1.20480	0.21996
2016	30	-2.7420	1.05626	0.19285
2023	30	-2.9347	0.84833	0.15488
Total	90	-2.7181	1.05190	0.11088

These figures make it clear that the mean scores of Dimension 5 are continuously negative, demonstrating the predominance of non-abstract features throughout all the phases. Nonetheless, the mean score slightly declines from 2009 to 2023, indicating a gradual shift toward a less abstract discourse.

The ANOVA results (Table 14) show that the effect of the phase on Dimension 5 scores is not statistically significant (F = 1.441, p = 0.242). This suggests that the variations in Dimension 5 scores across the phases are not substantial enough to be considered beyond random variation.

Table 14. ANOVA for Dimension 5. Source: own processing

Source	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3.158	2	1.579	1.441	0.242
Within Groups	95.320	87	1.096		
Total	98.478	89			

Tukey's HSD post-hoc tests (Table 15) verify that no significant differences exist between any of the pairs of phases. This supports the finding that the posts' abstractness has mainly stayed consistent over the three periods.

Table 15. Tukey HSD Post Hoc Comparisons for Dimension 5. Source: own processing

Comparison	Mean Difference	Std. Error	Sig.	95% CI Lower Bound	95% CI Upper Bound
Phase 1 vs 2	0.26433	0.27026	0.593	-0.3801	0.9088
Phase 1 vs 3	0.45700	0.27026	0.214	-0.1874	1.1014
Phase 2 vs 3	0.19267	0.27026	0.757	-0.4518	0.8371

The results show a slight trend toward a less abstract discourse in 2023, but this trend is not statistically significant. This slight change may be explained by how Reddit is developing as a platform, as users may favour more direct and intimate communication. However, the consistency of abstract features suggests the fundamental practices of these subreddits.

These findings imply that, although the number of non-abstract features may have slightly increased recently, Reddit posts' overall abstract nature has remained mostly stable over the three phases. Despite the changing platform dynamics, the users balance formal, impersonal communication and more direct, personal discourse, as evidenced by the consistency of the abstract characteristics.

Overall, the Dimension 5 analysis shows that abstract versus non-abstract discourse in Reddit posts has remained stable over time, with only minor, insignificant differences in all the three phases. This implies that the basic discourse practices regarding abstract content have not changed much, indicating a consistent manner in which users interact with one another on the platform.

5. Concluding remarks

The findings of the study indicate a clear trend of Reddit becoming more information-focused over time, with conversational discourse dominating in 2009, giving way to a more content-rich, information-focused medium by 2023. This pattern indicates more general shifts in social media communication, where content distribution is given more importance. Regarding narrative discourse, all the phases showed an overall use of the features related to the narrative discourse of posts, despite a minor increase in the number of non-narrative features. This consistency shows that, despite the platform's evolution, the users maintain the fundamental narrative practices by striking a balance between narrating events and describing states.

The results indicated a slight trend toward situation-dependent discourse. Nonetheless, the users prioritise unambiguous communication, as evidenced by the stability of the explicit characteristics over the phases. This research emphasises the value of clear communication strategies over time and

the transmission of information. A significant decline in the role of persuasive characteristics was observed in the Overt Expression of Persuasion between 2009 and 2023. In line with the general trend in digital communication, where information delivery efficiency frequently triumphs over persuasive efforts, this shift implies a move towards factual and descriptive communication.

Finally, the comparison of abstract versus non-abstract information content showed a minor trend towards a more non-abstract style in 2023. The enduring nature of abstract discourse attributes emphasises how users still combine more direct, personal communication with formal, impersonal communication. This equilibrium shows that users stick to a consistent discourse practice, skillfully combining non-abstract and abstract content, even as the platform's dynamics change.

In summary, this research illuminates the dynamic nature of online communication on Reddit. While maintaining core narrative and explicit discourse practices, the trends show how the platform is evolving to become a more content-driven medium. With regard to the evolving demands of online interactions, these insights underscore the need for adaptive communicative strategies and add to our understanding of how discourse develops on digital platforms. Future studies ought to look into the effects of these developments on digital communication and information distribution as Reddit and related platforms continue to expand.

List of abbreviations

ANOVA – Analysis of Variance

Df – Degrees of freedom

MAT– Multidimensional Analysis Tagger

N – Numbers

Sig. – Significance

SPSS – Statistical Package for the Social Sciences

Std. Deviation – Standard Deviation

Std. Error– Standard Error

Tukey's HSD – Tukey's Honestly Significant Difference

References

1. Abdulhaleem, E. & Ali, S. (2024). Exploring lexico-grammatical patterns in doctoral dissertations: A multidimensional analysis. In *World journal of English language*, 14 (3), p. 162-169. DOI: 10.5430/wjel.v14n3p162

2. 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: 10.5430/wjel.v14n2p452

3. 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: 10.24093/awej/chatgpt.7

4. Ali, S. & Thompson, P. (2022). Reporting local cultures in the globalised world: How indigenised can English be in the free world? In *Corpora*, 17 (2), p. 197-224. DOI: 10.3366/cor.2022.0254

5. Ananthasubramaniam, A., Chen, H., Yan, J., Alkiek, K., Pei, J., Seth, A. & Jurgens, D. (2023). Exploring linguistic style matching in online communities: The role of social context and conversation dynamics. In *arXiv preprint.* DOI: 10.18653/v1/2023.sicon-1.7

6. Androutsopoulos, J. (2014). Languaging when contexts collapse: Audience design in social networking. In *Discourse, context & media*, p. 62-73. DOI: 10.1016/j.dcm.2014.08.006

7. Biber, D. (1991). *Variation across speech and writing*. Cambridge: Cambridge University Press.

8. Biber, D. & Conrad, S. (2005). *Register variation: A corpus approach*. In *The handbook of discourse analysis*, p. 175-196. DOI: 10.1002/9780470753460.ch10

9. Biber, D. & Egbert, J. (2016). Register variation on the searchable web: A multi-dimensional analysis. In *Journal of English linguistics*, 44 (2), p. 95-137. DOI: 10.1177/0075424216628955

10. Biber, D., Conrad, S. & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

11. Borujeni, B.H. (2022). *Taking the high road: A big data investigation of natural discourse in the emerging US consensus about marijuana legalization*. Thesis for the degree of philosophy, Providence: Brown University.

12. Field, A. (2013). *Discovering statistics using IBM SPSS statistics.* London: Sage publications.

13. Flesch, M. (2022). lol thats how reddit talks;): le site américain Reddit comme espace de variation de l'anglais. Étude de corpus intersectionnelle et quantitative d'usages non standard, au prisme du genre, de l'âge et de l'ethnicité. Résumé de thèse. In *GLAD!. Revue sur le langage, le genre, les sexualités*, 138, DOI: 10.4000/glad.2206

14. Herring, S.C. (2010). Computer-mediated conversation Part I: Introduction and overview. In *Language@ internet*, 7(2). Available at:
    https://web.archive.org/web/20240127230705/https://www.languageatinternet.org/articles/2010/2801

15. Herring, S.C., Scheidt, L.A., Wright, E. & Bonus, S. (2005). Weblogs as a bridging genre. In *Information technology & people*, 18 (2), p. 142-171. DOI: 10.1108/09593840510601513

16. Medvedev, P., Stanciu, M. & Brudno, M. (2009). Computational methods for discovering structural variation with next-generation sequencing. In *Nature methods,* 6 (11), p. 13-20. DOI: 10.1038/nmeth.1374

17. Noble, B., Sayeed, A., Fernández, R. & Larsson, S. (2021). Semantic shift in social networks. In Proceedings of "*SEM 2021: The Tenth Joint Conference on Lexical and Computational Semantics", Association for Computational Linguistics, Online, p. 26-37.

18. Papakyriakopoulos, O., Engelmann, S. & Winecoff, A. (2023). Upvotes? Downvotes? No votes? Understanding the relationship between reaction mechanisms and political discourse on Reddit. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. April, 2023, p. 1-28. DOI: 10.1145/3544548.3580644

19. Starbird, K. (2016). Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on Twitter. In *Proceedings of the International AAAI Conference on Web and Social Media.* May, 2017, vol. 11, p. 230-239. DOI: 10.1609/icwsm. v11i1.14878

20. Starbird, K. (2019). Disinformation's spread: Bots, trolls and all of us. In *Nature*, 571 (7766), p. 449-450. DOI: 10.1038/d41586-019-02235-x

21. Tagg, C. (2015). *Exploring digital communication: Language in action*. New York: Routledge.

22. Willaert, T., Van Eecke, P., Van Soest, J. & Beuls, K. (2021). A tool for tracking the propagation of words on Reddit. In *Computational communication research*, 3 (1), p. 117-132. DOI: 10.5117/CCR2021.1.005.WILL

Contact data

name:	**Sadia Ali**
academic title/rank:	Assistant Professor
department:	Department of English Language and Literature
institution:	College of Science and Humanities, Prince Sattam bin Abdulaziz University, Al-Kharj, Kingdom of Saudi Arabia
e-mail:	sad.ali@psau.edu.sa
interest:	ESP, EFL, World Englishes, Corpus Linguistics, AI, and Corpus Development.