# Testing English News Articles for Lexical Homogenization Due to Widespread Use of Large Language Models

**Sarah Fitterer** and **Dominik Gangl** and **Jannes Ulbrich**
Technische Universität Berlin
Berlin, Germany
{fitterer, dominik.gangl, j.ulbrich}@campus.tu-berlin.de

## Abstract

It is widely assumed that Large Language Models (LLMs) are shaping language, with multiple studies noting the growing presence of LLM-generated content and suggesting homogenizing effects. However, it remains unclear if these effects are already evident in recent writing. This study addresses that gap by comparing two datasets of English online news articles – one from 2018, prior to LLM popularization, and one from 2024, after widespread LLM adoption. We define lexical homogenization as a decrease in lexical diversity, measured by the MATTR, Maas, and MTLD metrics, and introduce the LLM-Style-Word Ratio (SWR) to measure LLM influence. We found higher MTLD and SWR scores, yet negligible changes in Maas and MATTR scores in 2024 corpus. We conclude that while there is an apparent influence of LLMs on written online English, homogenization effects do not show in the measurements. We therefore propose to apply different metrics to measure lexical homogenization in future studies on the influence of LLM usage on language change.

## 1 Introduction

Since the release of ChatGPT-3.5 in November 2022, Large Language Model (LLM) powered chatbots have been widely adopted (Hu, 2023), ChatGPT alone currently counting 400 million weekly users (Reuters, 2025). Out of the many functionalities LLMs offer, they are increasingly used as a writing-assistance or co-authoring tool for texts. For instance, their increasing use has been confirmed in scientific writing (Liang et al., 2024b), consumer complaints, corporate communications, job postings, and international organization press releases (Liang et al., 2025). Even though users get unique outputs interacting with LLMs, each output is generated based on the same statistical models (i.e. GPT-3.5, GPT-4o, llama, etc.), whose idiosyncrasies carry over into the “unique” outputs they generate (Sun et al., 2025). Considering the high number of users and the widespread adoption of LLMs, many linguists assume a strong impact on language through their usage, potentially homogenizing it, according to the statistical likelihoods baked into each model. Yakura et al. (2024) provide empirical evidence to this thesis, measuring a significantly increased usage of ChatGPT specific words in spoken language after the chatbot’s release.

The term "linguistic homogenization" stems from the field of sociology, where it is discussed as a side effect of globalization and the general cultural homogenization resulting from it, thereby suppressing pluralistic ethnic identities for the sake of creating homogenous nation states (Bulcha, 1997). It describes the loss of diversity and a simultaneous entrenchment of linguistic hegemony. In the academic field of linguistics, homogenization is increasingly discussed as a possible effect of LLM use in several dimensions: a potential loss of lexical diversity (Reviriego et al., 2024) (Yakura et al., 2024), a homogenization of content and language toward Western-centric language and values (Agarwal et al., 2025), a perpetuation of linguistic discrimination (Fleisig et al., 2024), and an over-representation of hegemonic viewpoints (Bender et al., 2021). All five contributions highlight the importance of maintaining linguistic diversity for the future of AI development and warn of the negative social implications associated with the concept of linguistic homogenization.

Language change, which includes variations of lexical diversity over time, is influenced by many factors reflecting universal trends as well as historical contingencies (Bochkarev et al., 2014). The use of LLMs may not be the only factor contributing to a potential decrease of lexical diversity. Still, Rudnicka (2023) concludes from her research on Grammarly and ChatGPT’s preference of concise language, that while language change is influenced



*Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)*, pages 1239–1245

by many factors, these tools mirror and potentially accelerate language change. She proposes that the rising usage of LLM-driven writing tools might even be a "higher-order process" (Rudnicka, 2018, p. 157) changing language, meaning that their use has a strong, accelerated and system-level influence on the way language changes. Further, LLMs do not need to be actively used in order to exert an influence on human writing. A study by Roemmele (2021) found that automatically generated text, merely shown to the study's participants before they were prompted to write a text, influenced the semantics and sentence structure of the participants' writing.

Several studies investigated whether the use of LLMs has homogenizing effects on language, following Bommasani et al. (2022) who suggest the sharing of foundational models and datasets by distinct actors lead to an algorithmic monoculture, causing a homogenization of AI outputs. On a semantic level, Anderson et al. (2024) found that the users of LLMs may generate a greater number of more detailed ideas, while at a group level different users produced more homogenous, less semantically distinct ideas when using ChatGPT. Padmakumar and He (2023) found that humans writing with the assistance of InstructGPT, an aligned version of ChatGPT-3, produce texts with less lexical and content diversity than humans writing without assistance or the assistance of an unaligned chatbot. Finally, Reviriego et al. (2024) speculate that the increased use of LLMs could contribute to an overall loss of lexical diversity and test their hypothesis by comparing the lexical diversity of human text with that of GPT-generated text, without conclusive results.

Our study continues the search for homogenizing effects on language through the widespread use of LLMs. To summarize, previous studies unveiled the usage of LLMs in text bases (Liang et al., 2024a,b; Kobak et al., 2025), compared the lexical diversity of texts produced by humans to that of texts produced by LLMs (Reviriego et al., 2024), or proved homogenization effects in texts co-authored or fully generated by LLMs (Anderson et al., 2024; Padmakumar and He, 2023; Rudnicka, 2023). What remains unstudied is whether homogenizing effects can already be measured in large corpora of online written English two years after the popularization on LLMs, and whether these effects can be linked to widespread LLM usage. In this study, we address this gap, choosing to focus on one aspect of language: lexis. Lexis defines the body of words used in the sample, in opposition to the meaning or position of the words in sentence structures, etc.). We ask: **To what extent has the lexis of written online English homogenized since the widespread adoption of Large Language Models?**

We examine this question by comparing two sets of texts published at different points in time: Dataset A comprising texts published in 2018, before the popularization of LLM-based chatbots and writing assistants, and dataset B consisting of texts from 2024, when LLMs were already in wide use as writing assistants (Liang et al., 2024b). Following Reviriego et al. (2024), we measure lexical homogenization by a decrease in lexical diversity. In addition, we measure the amount of LLM-style words present in the corpora, following a method by Kobak et al. (2025) in order to link our results to the influence of LLM usage. Accordingly, we test our dataset for two hypotheses:

**H**$_1$: Lexical diversity in dataset A (2018) is significantly higher than in dataset B (2024).

**H**$_2$: LLM-specific vocabulary is significantly more frequent in dataset B (2024) than in dataset A (2018).

## 2 Methods

### 2.1 Compiling the datasets

Our datasets are composed of roughly 30,000 news articles each, taken from a random sample of the News on the Web (NOW) corpus (Davies, 2010). We chose the NOW corpus, as it is one of the largest collections of curated recent English written texts. It comprises data from 37,799,758 texts (at the time of writing) from online magazines and newspapers in 20 different English-speaking countries from 2010 to today. The sample datasets consist of 1/1000 of texts taken completely at random from the full NOW corpus of the selected year.

While we cannot confirm which texts are LLM-generated, news outlets likely contain little LLM-produced content due to reliance on professional journalists and adherence to editorial standards and AI policies (Becker et al., 2025). Additionally, given that news articles follow a fixed style that LLMs can easily mimic, and that an LLM's assigned role affects its lexical output (Martínez et al., 2024), even if there are LLM-generated or co-authored articles within our sample, they are likely to have a similar lexical diversity to human-authored news articles. News articles typically have a broad readership, increasing the influence they might have on language trends. We therefore find our dataset to be suitable for a first exploration analyzing changes in (mainly) human-written language.

### 2.2 Preprocessing

First, we preprocessed the 2 datasets by converting them to lowercase and cleaning them – removing digits, html-tags, punctuation, and stopwords using Python's Natural Language Toolkit (Bird et al., 2009) – so that only the content words remained. Each text was tokenized into words, and both the initial and cleaned word counts were recorded. We then computed the linguistic metrics on the resulting cleaned tokens.

The 2018 sample was composed of 33020 articles with an average of 508 words per article and the 2024 sample was composed of 33326 articles with an average of 574 words per article. Since the 2024 sample thus contained 12.8% more words than the 2018, we reduced the length of each country-specific subset in the 2024 data by this percentage to ensure comparability. This adjustment resulted in two corpora approximately equal in length: the 2018 corpus consists of 33,020 texts with an average of 508 words (totaling 9,445,311 words), and the 2024 corpus contains 29,047 texts with an average of 574 words (totaling 9,469,360 words).

### 2.3 Selecting the right measurements

#### 2.3.1 Measuring lexical diversity

We chose three common metrics to assess lexical diversity in our datasets, following Reviriego et al. (2024): the Maas metric, the Moving Average Type-Token-Ratio (MATTR) and the Measure of Textual Lexical Diversity (MTLD). Each of these measurements compares the total number of words to the total number of distinct words within each text.

The Maas metric (Maas, 1972) uses logarithmic scaling to correct the text-length bias of the Type-Token Ratio (TTR) which is the base measurement for lexical diversity of a text. The lower the score of the Maas calculation, the higher the lexical diversity of the measured text. The MATTR (Covington and McFall, 2010) uses a window (in our case 50 words) that slides through the text one word at a time, calculating the TTR for each window to overcome the TTR method's text length dependency. Higher scores mean higher lexical diversity. The MTLD (McCarthy and Jarvis, 2010) is length independent and sensitive to lexical variation. It creates an expanding window within the text word by word and calculates the running TTR within this window. When the TTR of the active window decreases below 0.72, the window is closed and a new window is started, beginning with the next word. The MTLD score gives the average segment length in number of words. A higher score signifies a higher lexical diversity.

#### 2.3.2 LLM-Style-Word Ratio

To measure potential changes in the frequency of LLM-specific vocabulary, we used a collection of words that Kobak et al. (2025) identified in their study on vocabulary changes in over 15 million biomedical abstracts from 2010 to 2024. Their study demonstrated that the emergence of LLMs led to an abrupt increase in the frequency of certain stylistic words. Based on these words, we developed our own metric, the "LLM-Style-Word Ratio", which we then used for our analysis. This ratio measures the percentage of specific style words commonly used by LLMs (e.g. "delve") across the texts, and thereby approximates the amount of direct or indirect LLM influence on the corpora texts.

### 2.4 Verifying the results

To assess whether the observed changes in lexical diversity between the 2018 and 2024 corpora reflect meaningful shift rather than falling within the range of natural variation, we conducted a control test using a split-sample approach. We divided the 2018 and 2024 corpora into two equally sized subcorpora each and computed the lexical diversity metrics for the halves to establish a baseline for the degree of variation one can expect when no real temporal change is present. We then compared the magnitude of this intra-corpus variation to the differences between the full 2018 and 2024 datasets. If the cross-year differences are comparable to or smaller than the within-year variation, it suggests that any apparent trend may be attributable to random sampling effects rather than significant change due to increased LLM involvement in 2024.

## 3 Results & Discussion

The scores of the lexical diversity measurements and the intra-corpus variations of both datasets are summarized in Table 1.

| Metric | A_2018 | B_2024 | Difference | ICV |
|---|---|---|---|---|
| MATTR | 0.88011 | 0.88121 | 0.00110 | 0.00109 |
| Maas | 0.01469 | 0.01482 | 0.00013 | 0.00016 |
| MTLD | 214.45 | 254.65 | 40.20 | 5.06 |
| SWR | 0.230% | 0.347% | 0.117% | 0.016% |

Table 1: Results of lexical diversity metrics & Style-Word Ratio, difference between scores of Dataset A (2018) and Dataset B (2024), and intra-corpus variation (ICV).

We find no conclusive effect of the use of LLMs on the lexical diversity of our dataset. Therefore, we cannot confirm our first hypothesis. The MTLD score increased by 40.2 points, but this trend was not mirrored in the MATTR and Maas scores: When compared to the changes observed in the same-year split samples, the slight increases in MATTR and Maas values fall within the range of natural variation and therefore do not indicate significant change in lexical diversity. Therefore, we would argue that these changes are negligible. A genuine rise in lexical diversity would typically manifest as increases across all measures.

However, we can confirm our second hypothesis: LLM-specific vocabulary is significantly more frequent in 2024 than in 2018. This suggests either direct use of LLMs in writing or indirect influence on human authors. If LLMs were used, the MTLD rise could stem from their tendency to reduce repetition and promote varied word choices – features often associated with higher-quality writing. Since the MTLD is designed to specifically assess the consistency of lexical variation rather than the absolute level of lexical diversity, this would be reflected in the higher MTLD score. While such tools increase variation within texts, they may also suggest repeated substitutions (e.g. replacing “and” with “as well as”), increasing MTLD without significantly affecting MATTR or Maas.

Assuming some 2024 texts were co-written with LLMs, the negligible variation in lexical diversity we found makes sense. Reviriego et al. (2024) showed that GPT-4 outputs show lexical diversity equal to or exceeding that of human texts. The studied datasets mostly consist of texts that exhibit high lexical diversity through their professional nature (in contrast to other online writing such as informal blog posts) and wide range of topics that require domain-specific vocabulary, attributes can be easily reproduced by LLMs (Martínez et al., 2024). If LLM-generated or co-authored articles were in the dataset, it is unlikely they impacted the lexical diversity of the corpus.

## 4 Conclusion

This study examined whether written online English has become more homogenized since the widespread adoption of Large Language Models. We defined lexical homogenization as a decrease in lexical diversity and introduced the LLM-Style-Word Ratio to measure LLM influence. Comparing news articles from 2018 and 2024, we found a higher MTLD score in 2024, but negligible changes in Maas and MATTR scores. Thus, we could not confirm a decrease in lexical diversity. However, the 2024 dataset showed a significant rise in LLM-specific vocabulary, supporting our second hypothesis. We link the higher MTLD scores in 2024 to LLMs usage, speculating that LLM writing assistants incite users to replace repetitive words for the sake of more lexically diverse, “better” writing, resulting in higher consistency of lexical diversity while not affecting lexical diversity on a corpus level.

We propose to analyze our results within their broader socio-technical context: As more texts influenced by LLMs enter the pool of online writing, the linguistic characteristics of AI systems may become woven into everyday usage, reinforcing certain vocabulary while possibly eroding dialectal (Fleisig et al., 2024) or stylistic variations. Simultaneously, LLMs are continually being updated and retrained, integrating human-authored content, whether AI-influenced or not, back into their models. Analyzing these feedback loops and the co-evolution of technological and social aspects is crucial to understanding how AI tools and human language jointly evolve, and whether such developments might embody a higher-order process in language evolution – leading to the emergence of new linguistic variations and possibly to a broader homogenization of language.

## 5 Outlook

Our findings raise doubts about the effectiveness of traditional lexical diversity metrics in capturing large-scale homogenization effects, as they may not fully reflect subtle shifts in lexical choice or frequency distribution. Indeed, lexical diversity measurements are put into question as in how well they actually measure the phenomenon (Jarvis, 2013; Bestgen, 2025). For example, Fleisig et al. (2024) suggest examining the decline of regionally specific or idiosyncratic vocabulary, which might better be captured by the analysis of individual word frequencies, since increases in diversity within certain domains may obscure losses of rare or context-specific words. Therefore, metrics like proposed LLM-style-word ratio, further refined by incorporating findings from Sun et al. (2025),Liang et al. (2024a), and complemented with a ratio capturing words disfavoured by LLMs, as identified by Kobak et al. (2025) and Fleisig et al. (2024) could be employed in further studies. Moreover, keeping in mind that metrics like MATTR were developed over a decade ago to evaluate then-called long-form texts such as novels (Bestgen, 2025), these tools may require revision when applied to corpora of significantly larger size used in computational linguistics today.

We also recommend including a broader range of text types (e.g., blogs, forums, advertisements, etc.) for a more generalizable analysis. Further, comparing texts produced in a controlled environment without LLM assistance with pre-LLM writing could reveal the indirect influence of LLM usage on language. Finally, an ongoing yearly analysis, repeating the study with datasets from 2025, 2026 and so on, could assess whether homogenizing effects increase as more LLM generated content is published. This would be especially interesting in light of Guo et al. (2024), who found a consistent decrease of linguistic diversity of LLM model outputs when trained with synthetic text created by LLMs.

## Limitations

Our dataset has several limitations. First, it comprises randomly selected news articles with missing metadata, making it unclear how representative it is of different styles and outlets. Second, the NOW corpus has its own limitations, such as 10 out of every 200 words being redacted due to U.S. copyright laws (Davies, 2024), though this likely has minimal impact due to the dataset's size and consistency. Third, the LLM-Style-Word Ratio was derived from Kobak et al. (2025) who extracted them from PubMed articles, which may limit its applicability to news articles due to differences in writing style. Lastly, since the dataset includes only news articles, it excludes other types of online writing, which limits the generalizability of our findings to broader online written English. While our study aimed to investigate the phenomenon of linguistic homogenization, our approach was limited to measuring potential changes in lexical diversity. Thereby, other aspects of linguistic homogenization such as semantics, sentence structure, and so on remain unattended. Moreover, we suspect the lexical diversity methods we applied are inappropriate for revealing a loss of lexical diversity on the scale of a very large text corpus. Therefore, our empirical contribution to the hypothesis that LLM usage is a higher-order process homogenizing language remains highly limited.

## References


Dhruv Agarwal, Mor Naaman, and Aditya Vashistha. 2025. AI Suggestions Homogenize Writing Toward Western Styles and Diminish Cultural Nuances. *Preprint*, arXiv:2409.11360.

Barrett R Anderson, Jash Hemant Shah, and Max Kreminski. 2024. Homogenization effects of large language models on human creative ideation. In *Proceedings of the 16th conference on creativity & cognition*, pages 413–425.

Kim Björn Becker, Felix M. Simon, and Christopher Crum. 2025. Policies in Parallel? A Comparative Study of Journalistic AI Policies in 52 Global News Organisations. *Digital Journalism*, pages 1–21.

Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 610–623.

Yves Bestgen. 2025. Estimating lexical diversity using the moving average type-token ratio (mattr): Pros and cons. *Research Methods in Applied Linguistics*, 4(1):100168.

Steven Bird, Edward Loper, and Ewan Klein. 2009. *Natural Language Processing with Python*. O'Reilly Media Inc.

Vladimir Bochkarev, Valery Solovyev, and Søren Wichmann. 2014. Universals versus historical contingencies in lexical evolution. *Journal of The Royal Society Interface*, 11(101):20140841.

Rishi Bommasani, Kathleen A. Creel, Ananya Kumar, Dan Jurafsky, and Percy Liang. 2022. Picking on the Same Person: Does Algorithmic Monoculture lead to Outcome Homogenization? *arXiv preprint*. ArXiv:2211.13972 [cs].

Mekuria Bulcha. 1997. The politics of linguistic homogenization in ethiopia and the conflict over the status of afaan oromoo. *African affairs*, 96(384):325–352.

Michael A. Covington and Joe D. McFall. 2010. Cutting the Gordian Knot: The Moving-Average Type–Token Ratio (MATTR). *Journal of Quantitative Linguistics*, 17(2):94–100.

Mark Davies. 2010. News on the web corpus (now). `https://www.english-corpora.org/now/`. Accessed May 19, 2025.

Mark Davies. 2024. Limitations and metadata issues in the now corpus. `https://www.corpusdata.org/limitations_now.asp`. Accessed May 19, 2025.

Eve Fleisig, Genevieve Smith, Madeline Bossi, Ishita Rustagi, Xavier Yin, and Dan Klein. 2024. Linguistic bias in chatgpt: Language models reinforce dialect discrimination. *arXiv preprint*.

Yanzhu Guo, Guokan Shang, Michalis Vazirgiannis, and Chloé Clavel. 2024. The curious decline of linguistic diversity: Training language models on synthetic text. *Preprint*, arXiv:2311.09807.

Krystal Hu. 2023. ChatGPT sets record for fastest-growing user base - analyst note. *Reuters*. Accessed on 2025-03-05.

Scott Jarvis. 2013. Capturing the diversity in lexical diversity. *Language Learning*, 63(s1):87–106. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9922.2012.00739.x.

Dmitry Kobak, Rita González-Márquez, Emőke Ágnes Horvát, and Jan Lause. 2025. Delving into chatgpt usage in academic writing through excess vocabulary. *Preprint*, arXiv:2406.07016.

Weixin Liang, Zachary Izzo, Yaohui Zhang, Haley Lepp, Hancheng Cao, Xuandong Zhao, Lingjiao Chen, Haotian Ye, Sheng Liu, Zhi Huang, Daniel A. McFarland, and James Y. Zou. 2024a. Monitoring ai-modified content at scale: A case study on the impact of chatgpt on ai conference peer reviews. *Preprint*, arXiv:2403.07183.

Weixin Liang, Yaohui Zhang, Mihai Codreanu, Jiayu Wang, Hancheng Cao, and James Zou. 2025. The widespread adoption of large language model-assisted writing across society. *Preprint*, arXiv:2502.09747.

Weixin Liang, Yaohui Zhang, Zhengxuan Wu, Haley Lepp, Wenlong Ji, Xuandong Zhao, Hancheng Cao, Sheng Liu, Siyu He, Zhi Huang, Diyi Yang, Christopher Potts, Christopher D Manning, and James Y. Zou. 2024b. Mapping the increasing use of llms in scientific papers. *Preprint*, arXiv:2404.01268.

Heinz-Dieter Maas. 1972. Über den Zusammenhang zwischen Wortschatzumfang und Länge eines Textes. *Zeitschrift für Literaturwissenschaft und Linguistik*, 2(8):73.

Gonzalo Martínez, José Alberto Hernández, Javier Conde, Pedro Reviriego, and Elena Merino-Gómez. 2024. Beware of Words: Evaluating the Lexical Diversity of Conversational LLMs using ChatGPT as Case Study. *ACM Transactions on Intelligent Systems and Technology*, page 3696459.

Philip M. McCarthy and Scott Jarvis. 2010. MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2):381–392.

Vishakh Padmakumar and He He. 2023. Does writing with language models reduce content diversity? *arXiv preprint*.

Reuters. 2025. OpenAI's weekly active users surpass 400 million. *Reuters*. Accessed on 2025-03-05.

Pedro Reviriego, Javier Conde, Elena Merino-Gómez, Gonzalo Martínez, and José Alberto Hernández. 2024. Playing with words: Comparing the vocabulary and lexical diversity of ChatGPT and humans. *Machine Learning with Applications*, 18:100602.

Melissa Roemmele. 2021. Inspiration through observation: Demonstrating the influence of automatically generated text on creative writing. *arXiv preprint*.

Karolina Rudnicka. 2018. Variation of sentence length across time and genre. *Diachronic corpora, genre, and language change*, pages 220–240.

Karolina Rudnicka. 2023. Can grammarly and chatgpt accelerate language change? ai-powered technologies and their impact on the english language: wordiness vs. conciseness. *Procesamiento de Lenguaje Natural*, 71.

Mingjie Sun, Yida Yin, Zhiqiu Xu, J. Zico Kolter, and Zhuang Liu. 2025. Idiosyncrasies in large language models. *Preprint*, arXiv:2502.12150.

Hiromu Yakura, Ezequiel Lopez-Lopez, Levin Brinkmann, Ignacio Serna, Prateek Gupta, and Iyad Rahwan. 2024. Empirical evidence of large language model's influence on human spoken communication. *Preprint*, arXiv:2409.01754.

# A Appendix

accentuates, acknowledges, acknowledging, addresses, adept, adhered, adhering, advancement, advancements, advancing, advocates, advocating, affirming, afflicted, aiding, akin, align, aligning, aligns, alongside, amidst, assessments, attains, attributed, augmenting, avenue, avenues, bolster, bolstered, bolstering, broader, burgeoning, capabilities, capitalizing, categorized, categorizes, categorizing, combating, commendable, compelling, complicates, complicating, comprehending, comprising, consequently, consolidates, contributing, conversely, correlating, crafted, crafting, culminating, customizing, delineates, delve, delved, delves, delving, demonstrating, dependability, dependable, detailing, detrimentally, diminishes, diminishing, discern, discerned, discernible, discerning, displaying, disrupts, distinctions, distinctive, elevate, elevates, elevating, elucidate, elucidates, elucidating, embracing, emerges, emphasises, emphasising, emphasize, emphasizes, emphasizing, employing, employs, empowers, emulating, emulation, enabling, encapsulates, encompass, encompassed, encompasses, encompassing, endeavors, endeavours, enduring, enhancements, enhances, ensuring, equipping, escalating, evaluates, evolving, exacerbating, examines, exceeding, excels, exceptional, exceptionally, exerting, exhibiting, exhibits, expedite, expediting, exploration, explores, facilitated, facilitates, facilitating, featuring, formidable, fostering, fosters, foundational, furnish, garnered, garnering, gauged, grappling, groundbreaking, groundwork, harness, harnesses, harnessing, heighten, heightened, hinder, hinges, hinting, hold, holds, illuminates, illuminating, imbalances, impacting, impede, impeding, imperative, impressive, inadequately, incorporates, incorporating, influencing, inherent, initially, innovative, inquiries, integrates, integrating, integration, interconnectedness, interplay, intricacies, intricate, intricately, introduces, invaluable, investigates, involves, juxtaposed, leverages, leveraging, maintaining, merges, methodologies, meticulous, meticulously, multifaceted, necessitate, necessitates, necessitating, necessity, notable, noteworthy, nuanced, nuances, offering, optimizing, orchestrating, outlines, overlook, overlooking, paving, persist, pinpoint, pinpointed, pinpointing, pioneering, pioneers, pivotal, poised, pose, posed, poses, posing, predominantly, preserving, pressing, promise, pronounced, propelling, realm, realms, recognizing, refine, refines, refining, remarkable, renowned, revealing, reveals, revolutionize, revolutionizing, revolves, scrutinize, scrutinized, scrutinizing, seamless, seamlessly, seeks, serves, serving, shaping, shedding, showcased, showcases, showcasing, signifying, solidify, spanned, spanning, spurred, stands, stemming, strategically, streamline, streamlined, streamlines, streamlining, struggle, substantiated, substantiates, surged, surmount, surpass, surpassed, surpasses, surpassing, swift, swiftly, thorough, transformative, typically, ultimately, uncharted, uncovering, underexplored, underscore, underscored, underscores, underscoring, unexplored, unlocking, unparalleled, unraveling, unveil, unveiled, unveiling, unveils, uphold, upholding, urging, utilizes, varying, versatility, warranting, yielding

Figure A1: Excess style words used for LLM-Style-Word Ratio based on the work of Kobak et al. (2025)