



Stability Amid AI Disruption: A Diachronic Linguistic Analysis of MA Thesis and PhD Dissertation Titles in Humanities, 2015–2025

Maha Abdulghafar Alayyash

Assistant Professor of Applied Linguistics, Department of English, College of Languages and Translation, University of Jeddah, Jeddah, Saudi Arabia

Email: maalayyash@uj.edu.sa

ORCID: <https://orcid.org/0000-0002-0809-8028>

ABSTRACT

The integration of large language models such as ChatGPT has raised concerns about stylistic homogenization in scholarly writing. While scientific literature shows clear LLM-driven shifts, e.g., increased lexical markers and reduced cohesion (Bao et al., 2025; Kousha & Thelwall, 2024), this study examines whether similar changes appear in humanities thesis and dissertation titles. Drawing on 8,631 unique MA and PhD titles from ProQuest in History, Religion, Literature, Philosophy, and Musicology, linguistic features were compared between 2015 (pre-AI) and 2025 (post-AI stabilization). Five dimensions were analyzed: word length, informativity, lexical diversity, syntactic structure, and semantic content. Results reveal remarkable stability across most metrics (title length ~12–13 words, informativity ~67%, lexical diversity near 100%). Only a modest increase in compound structures (70% to 74%) occurred, reflecting amplification of existing humanities conventions rather than disruption. The brevity of titles and extended human supervision appear to limit deep LLM intervention. These findings contrast with scientific fields and highlight the resilience of disciplinary norms in graduate scholarship.

Keywords: Large Language Models, thesis titles, humanities disciplines, ChatGPT, academic writing style, stylistic stability, diachronic analysis.



1. Introduction

With the development of Large Language Models (LLMs) like ChatGPT, academic writing has undergone notable changes, sparking widespread discussion about their effect on scholarly communication (Liang et al., 2024; Bao et al., 2025; Kousha & Thelwall, 2024; Juzek & Ward, 2025). Introduced by OpenAI in November 2022, ChatGPT has been described as one of the most powerful and widely used LLMs so far (Bao et al., 2025), capable of producing high-quality text in response to user prompts (Cao et al., 2023). This accessibility has led to its integration into various stages of academic production, from drafting manuscripts to summarizing abstracts (Geng & Trotta, 2024; Huang & Tan, 2023). However, as Thorp (2023) warns, the growing dependence on LLMs for academic writing has sparked debate regarding their implications for academic integrity and originality, and possible shifts in academic writing styles. These developments challenge longstanding norms in which academic disciplines have cultivated distinctive writing conventions that shape how knowledge is communicated within their respective fields (Song, Chen & Zhao, 2023).

Existing research has begun to quantify these transformations, revealing linguistic changes in academic texts post-ChatGPT. For instance, studies analyzing abstracts from arXiv datasets have identified a marked increase in the proportion of words preferred by LLMs, alongside rises in lexical complexity and sentiment, but declines in syntactic complexity, cohesion, and readability (Bao et al., 2025). This suggests LLMs may introduce a broader range of new vocabulary and simplify sentence structure, while making texts harder to read because of reduced use of connecting words. Similarly, full-text analyses of over 2.4 million PubMed Central publications from 2021 to 2025 show dramatic surges in LLM-associated terms like 'delve' (+1,500%), 'underscore' (+1,000%), and 'intricate' (+700%), with their co-occurrence patterns intensifying post-2022 (Kousha & Thelwall, 2024). Such patterns are more pronounced in fields like Computer Science, where scholars demonstrate more noticeable changes in writing style, whereas Mathematics shows only minimal variation (Bao et al., 2025). In biomedical sciences, terms like 'delves', 'showcasing', and 'underscores' rose to 10% prevalence in PubMed abstracts by 2024 (Kobak et al., 2024), emphasizing the growing impact of generative AI on the way article titles are formulated (Comas-Forgas, Koulouris & Kouis, 2025).

Titles, as the face of academic works (Lebrun, 2011), are particularly susceptible to these influences, serving as condensed summaries that must be attractive, informative, and concise (Ball, 2009; Hartley, 2005). Genre analysis highlights titles as a subgenre shaped by disciplinary conventions, with variations in length, syntactic structure, and information content across fields (Swales, 1990; Bhatia, 1993). Diachronic studies reveal evolving trends: for example, applied linguistics titles from 1975 to 2015 increasingly emphasized method/design information, with fluctuating patterns in topic, result, and dataset details (Sahragard & Meihami, 2016). In pragmatics, compound titles incorporating rhetorical elements like quotations or metaphors have surged, reflecting a shift toward post-academic writing style and rhetorical marketization in



competitive publishing contexts (Dillon, 1982; Haggan, 2004; Chen & Liu, 2023). Biomedical article titles have also become longer, incorporating more complex structures and persuasive features such as boosters to highlight findings (Soler, 2007; Jaime-Sisó, 2009; Martín & León Pérez, 2024). Cross-disciplinary comparisons often contrast ‘hard’ and ‘soft’ sciences, noting longer, method-focused titles in the former (Nagano, 2015; Soler, 2007).

Amid these developments, non-native English speakers (NNESs) may disproportionately rely on LLMs to overcome language barriers, potentially amplifying stylistic changes (Flowerdew, 1999; Huang, 2010; Bao et al., 2025). Surveys indicate LLMs are commonly used for editing, proofreading, and drafting, with 76% of researchers reporting AI tool adoption and 67% noting benefits in content discovery (OUP, 2024; Elsevier, 2024). Yet, evaluations of AI-generated titles show they often outperform human ones in perceived alignment with content and appeal, with raters favoring AI versions in 61.8% of cases (Sebo et al., 2026). Blinded reviews of abstracts further reveal that ChatGPT can generate believable scientific abstracts, but with entirely fabricated data, and they are often more vague and formulaic (Gao et al., 2023). Large-scale analyses confirm rising LLM-modification rates, peaking at 17.5% in Computer Science abstracts by 2024, associated with shorter papers and crowded research areas (Liang et al., 2024).

This study fills a gap by investigating the impact of AI on thesis and dissertation titles in humanities fields, i.e., History, Religion, Literature, Philosophy, and Musicology, comparing those produced by MA and PhD students in 2015 (pre-AI) and 2025 (post-AI stabilization). Drawing from a corpus of 8,631 titles sourced from ProQuest, the analysis explores potential linguistic shifts attributable to AI tools. To guide this investigation, the following research questions are proposed:

RQ1: How has the word length of MA and PhD titles changed between 2015 and 2025, and does this vary across disciplines?

RQ2: To what extent has the informativity of MA and PhD titles evolved from 2015 to 2025, and are there disciplinary differences?

RQ3: What changes in lexical diversity are evident in MA and PhD titles pre- and post-AI, and how do these differ by field?

RQ4: Have syntactic structures in MA and PhD titles shifted between 2015 and 2025, and what variations exist across disciplines?

RQ5: How has the semantic content of MA and PhD titles transformed from 2015 to 2025, and are there field-specific patterns?

2. Related Work

Academic papers and theses are key mediums for communicating and disseminating scholarly knowledge, typically emphasizing precision, clarity, and a balance between structural complexity and concise expression (Biber & Gray, 2010). Unlike many other genres, academic writing often combines structural and phrasal complexity with relatively simple sentence structures to facilitate clear communication of research



findings (Atkinson, 1998; Gross, Harmon & Reidy, 2002). Writing style is influenced by several factors, including authors' linguistic background, training, and intended audience, yet it plays a crucial role in shaping readability, reader engagement, dissemination reach, downloads, citation impact, and perceptions of scholarly innovation (Lu et al., 2019; Chen et al., 2020; Hu, Wang & Deng, 2021). With the rapid development of LLMs, particularly following the public release of ChatGPT, new tools have become widely available to assist academic writing through functions such as grammar correction, machine translation, text generation, summarization, and stylistic refinement, thereby enhancing accessibility for a wider range of researchers, particularly those who are non-native English speakers (Altmäe, Sola-Leyva & Salumets, 2023; Gao et al., 2023; Lozić & Štular, 2023; Liu & Bu, 2024).

Meanwhile, the expanding presence of LLMs in academia has sparked critical concerns. While LLM-generated text often exhibits high fluency and rhetorical polish, it may lack originality, factual reliability, and authentic authorial voice, prompting debates about transparency, plagiarism risks, authorship attribution, and broader issues of academic integrity (Uzun, 2023; Thorp, 2023). Quantitative studies conducted after the release of ChatGPT have also documented measurable stylistic shifts in academic writing, including increased frequency of LLM-favored lexical items such as "delve," "underscore," and "intricate," as well as higher lexical complexity and sentiment polarity, accompanied by declines in syntactic complexity, textual cohesion, and readability scores across abstracts and full-text corpora (Bao et al., 2025; Geng & Trotta, 2024; Kousha & Thelwall, 2024; Liang et al., 2024; Liu & Bu, 2024).

Linguistic analyses of academic writing have examined features at word, sentence, and paragraph levels, including lexical density (proportion of content words), lexical sophistication (vocabulary depth), syntactic complexity (clauses, T-units), cohesion, readability, and adherence to structures like IMRaD (Plavén-Sigray et al., 2017; Song, Chen & Zhao, 2023; Greenhalgh et al., 2004; Wang et al., 2023; Dong et al., 2024). Tools such as L2SCA for syntactic complexity (Lu, 2010), TAACO/TAACS for cohesion (Crossley, Kyle & McNamara, 2016; Kyle & Crossley, 2015), and custom analyses have supported these investigations. Recent LLM-era studies have expanded metrics to include sentiment scores and frequencies of LLM-preferred words (e.g., 'delve', 'underscore', 'intricate'), revealing surges in such terms post-ChatGPT, alongside increased lexical complexity but decreased syntactic complexity, cohesion, and readability in abstracts and full texts (Bao et al., 2025; Kousha & Thelwall, 2024; Liang et al., 2024).

LLMs have been applied for grammar correction, summarization, abstract generation, and literature reviews, often producing fluent, coherent output with satisfactory rhetorical quality (Altmäe, Sola-Leyva & Salumets, 2023; Gao et al., 2023; Lozić & Štular, 2023). However, generated content frequently lacks authenticity, originality, and factual accuracy, raising ethical concerns about plagiarism, transparency, and integrity (Uzun, 2023; Thorp, 2023). Quantitative studies have tracked word frequency shifts, LLM-modification probabilities in preprints, and correlations with author



characteristics, showing higher adoption in established fields or by frequent publishers (Geng & Trotta, 2024; Liang et al., 2024; Liu & Bu, 2024). Evaluations indicate AI-generated titles often score higher in perceived alignment, appeal, and preference than human ones (Sebo, Nie & Wang, 2026).

Research on academic titles, as condensed yet critical ‘front matter’ elements, has explored form features (length, syntactic structure, punctuation) and content attributes (informativity via substantive words, lexical diversity, information types like topic, method, results) (Swales & Feak, 1994; Haggan, 2004; Hartley, 2007; Soler, 2007, 2011; Hyland & Zou, 2022). Diachronic trends show lengthening titles with more complex structures, informative elements, and promotional language (boosters, hyperbole) in fields like biomedicine (Jaime-Sisó, 2009; Vinkers, Tjindink & Otte, 2015; Martín & León Pérez, 2024). Compound titles with colons or rhetorical parts (quotations, questions, metaphors) have increased, driven by competitive publishing and ‘rhetorical marketization’ (Dillon, 1982; Chen & Liu, 2023). Cross-disciplinary comparisons reveal variations: humanities and soft sciences favor longer, compound, descriptive titles, while hard sciences prefer informative, nominal ones (Haggan, 2004; Hartley, 2007; Nagano, 2015). Informativity has generally risen over time, though lexical diversity trends vary (Sahragard & Meihami, 2016; Bérubé et al., 2018).

Despite extensive work on journal articles and some on theses/dissertations (e.g., genre analyses of acknowledgments or abstracts), studies rarely compare MA and PhD titles linguistically, and diachronic examinations in humanities fields remain limited. Most title research focuses on journal articles in sciences or applied linguistics, with fewer diachronic or comparative analyses in pure humanities like history, religion, literature, philosophy, or musicology. No comprehensive study has yet examined LLM influences on thesis titles, particularly contrasting pre-AI (2015) and post-AI stabilization (2025) periods across degree levels. This gap is significant, as theses represent foundational scholarly training, where stylistic shifts may reflect broader AI adoption, especially among students (including non-native speakers) using LLMs for polishing or generation. MA titles, often shorter and less original in scope than PhD ones, may show distinct responses to AI tools compared to doctoral works requiring deeper novelty.

This study addresses these gaps by analyzing 8,631 thesis titles from ProQuest in History, Religion, Literature, Philosophy, and Musicology, comparing MA and PhD outputs from 2015 (pre-AI) and 2025 (post-AI). It examines word length, informativity, lexical diversity, syntactic structure, and semantic content to assess potential AI-driven linguistic shifts and disciplinary/degree-level variations, providing empirical insights into how stabilized AI may be reshaping humanistic scholarly titling practices.

3. Methodology

3.1. Data Source and Collection

The data for this study were electronically collected from the ProQuest Dissertations & Theses Citation Index, accessed via Web of Science and integrated with ProQuest Dissertations & Theses Global (PQDT Global), the largest curated collection of



multidisciplinary dissertations and theses in the world, offering extensive global coverage of graduate-level works with metadata including titles, authors, institutions, degree types, and completion years.

For each target period, the relevant completion year was selected (January 1 to December 31, 2015, for the pre-AI benchmark; or January 1 to December 31, 2025, for the post-AI stabilization period). Results were filtered to dissertations and theses only; the Dissertation and Thesis Subject was restricted to History, Religion, Literature, Philosophy, and Musicology using appropriate subject headings and classification codes; the language was set to English; and all qualifying records, including titles, author names, degree types (Master's/MA or Doctoral/PhD), and other metadata, were exported in Excel format.

These core humanities disciplines were chosen because their titling conventions are typically more descriptive, interpretive, and rhetorically nuanced than those in STEM fields. The strict 2015 vs. 2025 comparison ensured a clear temporal contrast with a 10-year gap, avoiding transitional years (2023–2024) where AI adoption was still emerging and less stabilized. Although an expanded dataset including 2023–2024 was initially considered to boost sample size, the decision was made to focus exclusively on 2025 for the post-AI period to minimize confounding factors from early, limited exposure to generative AI tools. Table 1 presents the distribution of MA and PhD titles across the five disciplines for 2015 and 2025, showing a total of 8,631 unique records after cleaning.

Table 1: MA and PhD Titles (2015 vs 2025)

| Discipline | MA 2015 | MA 2025 | PhD 2015 | PhD 2025 | Total MA | Total PhD | Total |
|------------|---------|---------|----------|----------|----------|-----------|-------|
| History | 342 | 383 | 538 | 480 | 725 | 1018 | 1743 |
| Religion | 268 | 188 | 973 | 912 | 456 | 1885 | 2341 |
| Literature | 196 | 296 | 278 | 279 | 492 | 557 | 1049 |
| Philosophy | 251 | 128 | 755 | 447 | 379 | 1202 | 1581 |
| Musicology | 384 | 157 | 792 | 584 | 541 | 1376 | 1917 |
| Totals | 1441 | 1152 | 3336 | 2702 | 2593 | 6038 | 8631 |

Duplicates (e.g., identical titles under multiple records due to institutional variations, embargo releases, or indexing errors) were identified and removed. The final cleaned corpus comprises 8,631 unique English-language thesis and dissertation titles from the selected fields. This systematic electronic process yielded a focused, high-quality dataset while preserving the database's comprehensive scope and metadata richness.

3.2. Data Processing and Analysis Overview

All 8,631 unique titles were exported into a structured digital dataset (CSV format) for systematic coding and quantitative analysis. Titles were tokenized using standard rules adapted for academic titling: hyphenated compounds (e.g., post-colonialism) counted as separate words where appropriate for density measures, while acronyms (e.g., USA) and proper nouns were treated as single tokens. Punctuation such as colons, dashes, and



question marks was preserved for syntactic classification but excluded from word counts.

Coding and measurements were conducted primarily through a combination of automated scripts (Python with NLTK and spaCy for tokenization, part-of-speech tagging, and basic metrics) and targeted manual review to resolve ambiguities, especially in complex humanities phrasing. A random subset of 15% of titles (~1,295) was independently double-coded by two reviewers to establish reliability; inter-coder agreement reached >95% across all variables after resolving minor discrepancies through discussion.

The analyses addressed the five research questions through the following specific measures, tailored to the interpretive and often rhetorically elaborate nature of humanities thesis titles (which tend to prioritize thematic depth, nuance, and descriptive richness over strict methodological signaling common in sciences):

3.2.1. Word Length (RQ1)

Title length was calculated as the total number of orthographic words (space-separated tokens). This measure followed conventions in bibliometric and genre studies of academic titles. Means, medians, standard deviations, and distributions were computed overall, by degree level (MA vs. PhD), by discipline, and across the two time periods (2015 vs. 2025). Comparisons tested for significant changes potentially attributable to AI stabilization.

3.2.2. Informativity (RQ2)

Informativity was operationalized as the percentage of substantive (content-carrying) words relative to total words, following the classic approach of Buxton and Meadows (1977) and its extensions in title research. The analysis focused on substantive words, such as nouns (both common and proper), lexical verbs, adjectives, and adverbs, while function words, including articles, prepositions, conjunctions, pronouns, auxiliary verbs, and determiners, were excluded. The formula was: (number of substantive words / total words) × 100. Averages were derived per title and aggregated by discipline, degree, and time period to detect shifts toward greater or lesser informational density, which might reflect AI tendencies toward polished but sometimes less dense phrasing.

3.2.3. Lexical Diversity (RQ3)

Lexical diversity was measured using the Type-Token Ratio (TTR = number of unique word types / total tokens), supplemented by Moving-Average Type-Token Ratio (MATTR) to mitigate length sensitivity in short texts like titles. These metrics captured vocabulary variation and potential increases in novel or AI-favored terms (e.g., interpretive boosters or abstract nouns common in humanities). Discipline-specific terminology was retained to preserve authentic field variation.

3.2.4. Syntactic Structure (RQ4)



Syntactic patterns were classified using an adapted hierarchical scheme from Cheng, Kuo & Kuo (2012) and Xiang and Li (2020), suitable for humanities theses, which frequently employ descriptive, compound, or rhetorically layered structures. Categories were mutually exclusive, with ambiguities resolved by prioritizing the most structurally prominent feature (e.g., compound structures took precedence). Compound subtitles were further noted for rhetorical function (e.g., interpretive vs. specifying). The scheme is detailed in Table 2 below.

Table 2: Syntactic Structure Categories

| Category | Description | Examples |
|---------------|--|--|
| Compound | Colon, dash, or divider separating segments | 'Empire's Echoes: Colonialism in Victorian Literature' |
| Nominal | Complex noun phrase, no finite verb | 'Postmodern Identity in American Fiction' |
| V-ing Phrase | Begins with gerund (exploratory/process focus) | 'Exploring Musical Motifs in Baroque Music' |
| Prepositional | Starts with preposition | 'On the Origins of Modern Philosophy' |
| Full Sentence | Complete clause with finite verb | 'How Does Memory Shape Historical Narratives?' |

Ambiguities were resolved by prioritizing the most structurally prominent feature. Compound subtitles were further noted for rhetorical function (e.g., interpretive vs. specifying).

3.2.5. Semantic Content (RQ5)

Semantic features were coded using an adapted version of the framework from Sahragard and Meihami (2016), originally drawing on Goodman, Thacker & Siegel (2001). Five primary categories were used, with a hierarchical approach: titles defaulted to Topic if no more specific elements were present; otherwise, priority was given to Method, Dataset, Result, or Conclusion. These latter categories were non-exclusive, allowing co-occurrence (e.g., Method + Dataset), with percentages potentially exceeding 100% for multi-component titles. Coding relied on lexical indicators, applied sensitively to humanities conventions (e.g., 'Method' as theoretical approaches, 'Dataset' as cultural artifacts, 'Result/Conclusion' as interpretive insights). The framework is detailed in Table 3 below.

This framework was applied with sensitivity to humanities conventions, where 'Method' might encompass theoretical frameworks or interpretive approaches, 'Dataset' could refer to texts, archives, or cultural artifacts, and 'Result' or 'Conclusion' often highlights insights or reinterpretations rather than empirical findings.

**Table 3: Semantic Content Categories**

| Category | Description / Focus | Lexical Indicators | Examples (Humanities) |
|------------|--|--|--|
| Topic | General subject/phenomenon (default) | Broad nouns (identity, empire...) | 'Victorian Literature and Social Reform' |
| Method | Analytical/theoretical approach | analysis, framework, approach | 'A Feminist Analysis of Shakespearean Tragedies' |
| Dataset | Specific sources, texts, archives, artifacts | texts, archives, based on, from | 'Religious Symbols in Medieval Manuscripts' |
| Result | Outcomes, impacts, evidence | impact, effect, evidence, role | 'The Impact of Colonialism on Literary Expression' |
| Conclusion | Strong interpretive claims/insights | challenges, reinterprets, implications | 'Rethinking Identity: Implications for Ethics' |

3.2.6. Statistical Procedures

Descriptive statistics (means, standard deviations, medians, frequencies, and percentages) were computed in IBM SPSS Statistics (Version 31) to summarize patterns in word length, informativity, lexical diversity, syntactic structures, and semantic content across time periods (2015 vs. 2025), disciplines, and degree levels (MA vs. PhD). Due to non-normal distributions in all continuous variables (confirmed via preliminary normality checks), non-parametric tests were used throughout for inferential analyses. Mann–Whitney U tests compared 2015 versus 2025 and MA versus PhD titles on continuous measures (title length, informativity percentages, TTR and MATTR scores). Kruskal–Wallis tests examined disciplinary differences, with pairwise follow-ups where appropriate. Chi-square tests of independence assessed distributional shifts in categorical variables (syntactic categories and semantic emphases) across time, degree levels, and fields. All tests were two-tailed ($\alpha = 0.05$). Effect sizes were reported where relevant (Cramér's V for chi-square tests) to evaluate practical significance. Considering the study's exploratory design and the high number of comparisons involved, no correction for multiple testing was applied; significant results should be interpreted cautiously. Visualizations (boxplots for continuous measures, grouped bar charts for categorical distributions, and line graphs for temporal trends) were generated to illustrate patterns and support interpretation of potential AI-influenced linguistic changes in humanities thesis titles.

4. Findings and Discussion

The following subsections present the core empirical findings from the analysis of 8,631 humanities thesis and dissertation titles, addressing each research question in sequence. Subsection 4.1 examines changes in word length (RQ1), 4.2 informativity



(RQ2), 4.3 lexical diversity (RQ3), 4.4 syntactic structure (RQ4), and 4.5 semantic content (RQ5). Results focus on temporal shifts (2015 pre-AI vs. 2025 post-AI stabilization), degree-level patterns (MA vs. PhD), and disciplinary variation (History, Religion, Literature, Philosophy, Musicology), with discussion connecting findings to potential LLM influence and established genre conventions.

4.1. Word Length

RQ1 investigated how word length of MA and PhD thesis titles changed between 2015 and 2025, and whether changes varied across disciplines. Overall title lengths showed high stability across the decade (see Table 4). Mean word counts remained in the narrow 12.3–12.9 range, medians stayed fixed at 12 words, and standard deviations were comparable. Neither MA nor PhD titles exhibited statistically meaningful change over time (both Mann-Whitney U tests $p > 0.3$).

Table 4: Overall Word Length by Degree and Year

| Degree | Year | Mean | Median | SD | N |
|--------|------|-------|--------|------|-------|
| MA | 2015 | 12.37 | 12 | 5.18 | 1,441 |
| MA | 2025 | 12.34 | 12 | 5.15 | 1,152 |
| PhD | 2015 | 12.79 | 12 | 4.85 | 3,336 |
| PhD | 2025 | 12.89 | 12 | 4.80 | 2,702 |

Discipline-level patterns revealed only modest directional shifts: slight lengthening in History and Religion, slight shortening in Literature, Musicology, and Philosophy. When examined by degree and discipline, the large majority of 2015–2025 comparisons were statistically non-significant. The only exceptions were PhD titles in Religion (which lengthened) and Philosophy (which shortened), as shown in the selected pairwise tests in Table 5.

Table 5: Key Pairwise Comparisons of Word Length (2015 vs. 2025)

| Comparison | Test Type | Statistic | p-value | Interpretation |
|---|----------------|-----------|--------------|-----------------------|
| MA overall | Mann-Whitney U | 826,536 | 0.854 | No change |
| PhD overall | Mann-Whitney U | 4,439,704 | 0.317 | No change |
| PhD Religion | Mann-Whitney U | 413,848.5 | 0.011 | Significant increase |
| PhD Philosophy | Mann-Whitney U | 181,937 | 0.023 | Significant decrease |
| All other discipline, degree combinations | Mann-Whitney U | — | > 0.05 | No significant change |

Tests for differences among disciplines (Kruskal-Wallis) were highly significant in every degree-year group (all $p < 0.001$), confirming that Philosophy consistently produced the shortest titles (typically 10–11 words) while Religion produced the longest (13–14 words). These disciplinary contrasts persisted across both time periods with little alteration (see Figure 1 for a visual comparison of title word counts by discipline across 2015 and 2025, and Figure 2 for the minimal change by degree level).

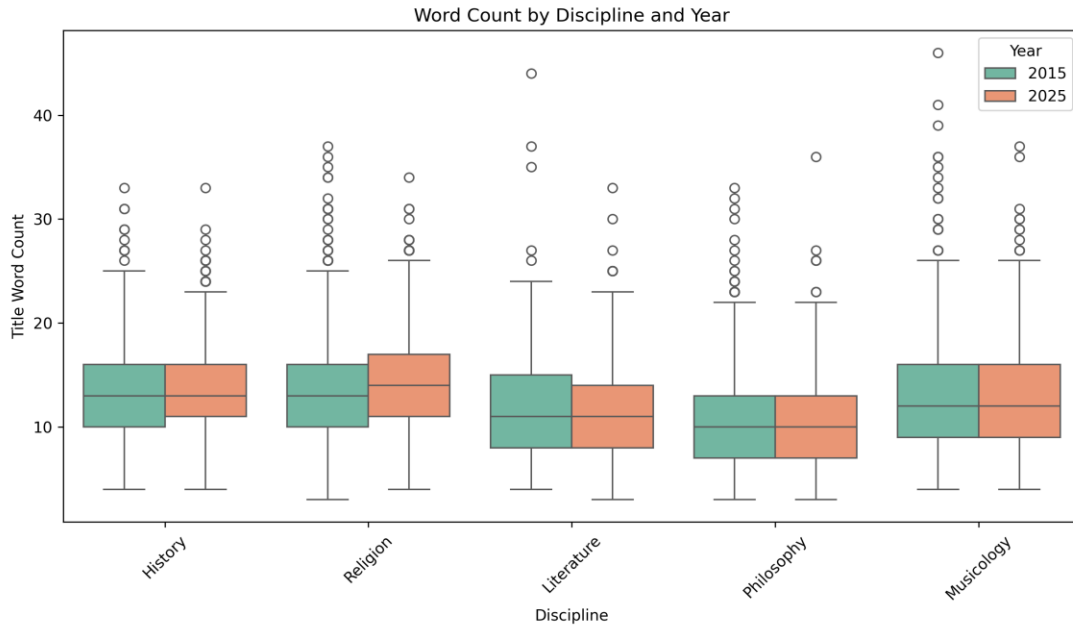


Figure 1. Boxplot of thesis title word counts by discipline (History, Religion, Literature, Philosophy, Musicology) for 2015 and 2025, showing stable lengths and persistent disciplinary differences.

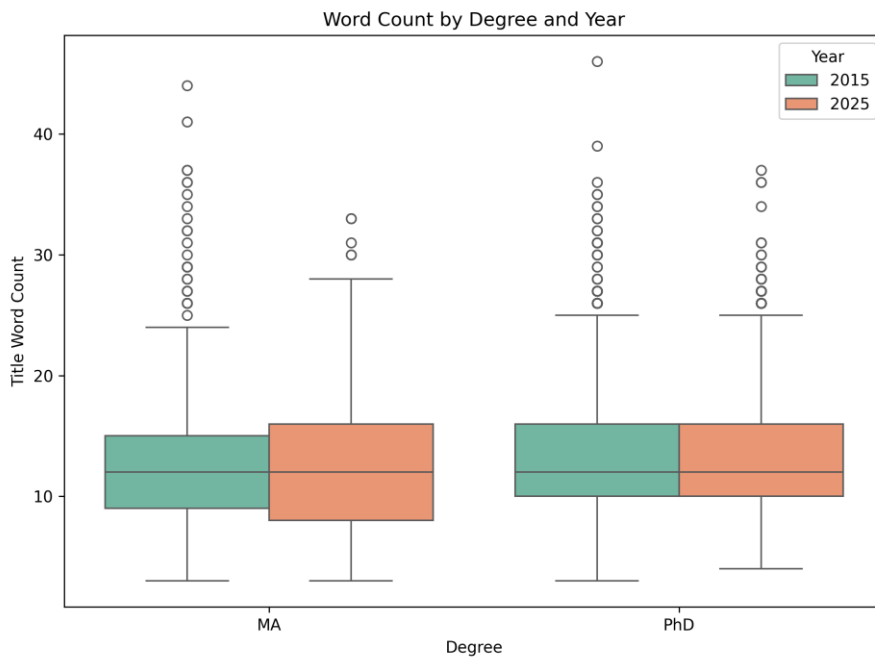


Figure 2. Boxplot of thesis title word counts by degree level (MA vs. PhD) for 2015 and 2025, showing minimal change over time.



Taken together, the findings indicate that the stabilization and widespread adoption of LLMs by 2025 did not trigger any broad or uniform change in humanities thesis title length. The observed stability contrasts with lengthening trends documented in some scientific fields post-ChatGPT and aligns more closely with the persistence of long-standing disciplinary titling norms in the humanities (Haggan, 2004; Nagano, 2015). The isolated significant shifts in PhD Religion (longer) and PhD Philosophy (shorter) titles may reflect subtle, field-specific amplification of existing rhetorical preferences, i.e., elaborative interpretation in Religion versus analytic conciseness in Philosophy, potentially facilitated by LLM polishing, especially among doctoral students. Overall, however, AI appears to have exerted minimal transformative pressure on this particular dimension of titling practice.

4.2. Informativity

RQ2 examined the extent to which the informativity of MA and PhD thesis titles (percentage of substantive/content-carrying words) evolved from 2015 to 2025, and whether disciplinary differences were evident. Overall informativity levels remained highly stable across the decade (see Table 6). Mean percentages of substantive words fluctuated narrowly between ~66.7% and ~68.2%, with medians consistently at 66.67% in all four degree-year groups. Neither MA nor PhD titles showed a statistically significant change between 2015 and 2025 (both Mann-Whitney U tests $p > 0.06$).

Table 6: Overall Informativity (%) by Degree and Year

| Degree | Year | Mean | Median | SD | N |
|--------|------|-------|--------|-------|-------|
| MA | 2015 | 67.58 | 66.67 | 11.14 | 1,441 |
| MA | 2025 | 66.72 | 66.67 | 11.35 | 1,152 |
| PhD | 2015 | 67.78 | 66.67 | 9.88 | 3,336 |
| PhD | 2025 | 68.24 | 66.67 | 10.04 | 2,702 |

Discipline-level patterns showed only minor directional variation: slight increases in History and Musicology, very small decreases in Literature and Philosophy, and near-zero change in Religion. When disaggregated by degree and discipline, almost all 2015–2025 comparisons were non-significant. The single exception was PhD Musicology titles, which exhibited a modest but statistically significant increase in informativity ($p = 0.020$), as highlighted in the pairwise tests in Table 7.

Table 7: Key Pairwise Comparisons of Informativity (2015 vs. 2025)

| Comparison | Test Type | Statistic | p-value | Interpretation |
|--|----------------|-----------|--------------|------------------------|
| MA overall | Mann-Whitney U | 854,615.5 | 0.193 | No change |
| PhD overall | Mann-Whitney U | 4,385,038 | 0.070 | No change (borderline) |
| PhD Musicology | Mann-Whitney U | 214,350.5 | 0.020 | Significant increase |
| All other discipline/ degree combinations | Mann-Whitney U | — | > 0.05 | No significant change |



Kruskal-Wallis tests confirmed significant differences among disciplines in every degree-year group (all $p < 0.001$), with Musicology consistently showing the highest informativity (69–70%) and Literature and Religion among the lower values (66–67%). These disciplinary contrasts were largely stable over time (see Figure 3 for a visual comparison of informativity percentages by discipline across 2015 and 2025).

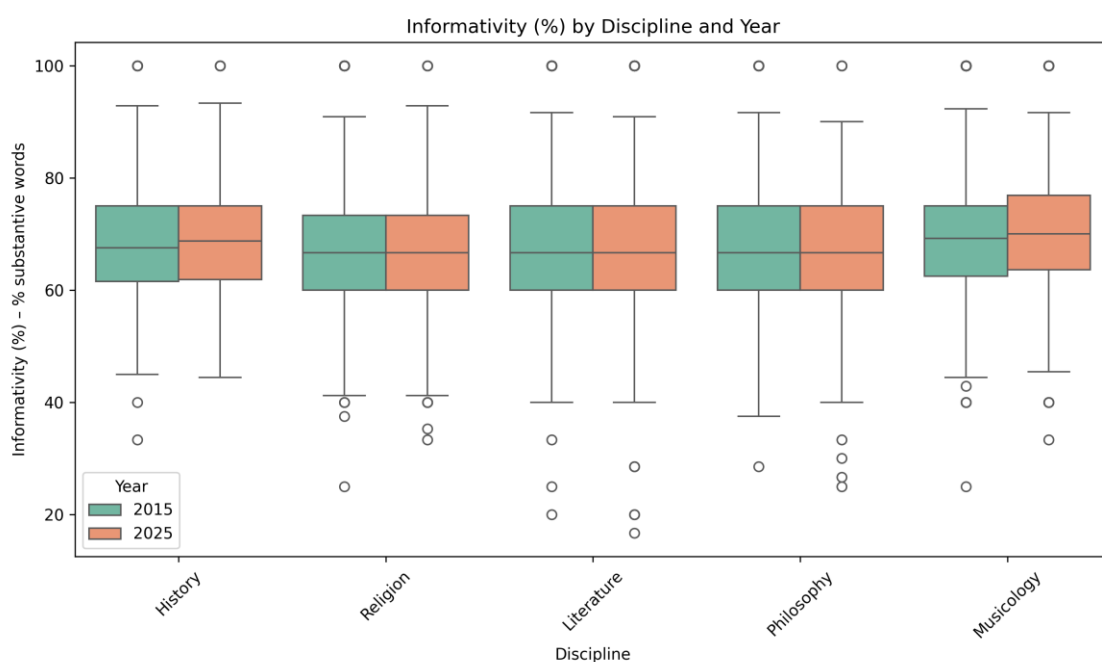


Figure 3. Boxplot of informativity percentage (substantive words) by discipline and year (2015 vs. 2025).

The overall picture reveals no broad or systematic shift in title informativity attributable to AI stabilization by 2025. The high degree of stability and the absence of widespread decline contrasts with some observations in LLM-era journal abstracts and full texts, where reduced cohesion or denser function-word usage sometimes appeared (Bao et al., 2025; Kousha & Thelwall, 2024). The isolated increase in PhD Musicology informativity may reflect field-specific tendencies toward descriptive precision (e.g., naming musical works, periods, or theoretical concepts) being subtly reinforced by LLM-assisted polishing, without evidence of the “polished but less dense” phrasing sometimes associated with generative AI. Disciplinary persistence remains the dominant pattern, consistent with genre conventions in humanities titling that prioritize thematic and interpretive substance (Haggan, 2004; Soler, 2007).

4.3. Lexical Diversity

RQ3 investigated changes in lexical diversity (e.g., Type-Token Ratio or MATTR, expressed here as percentage of unique words) in MA and PhD thesis titles from 2015 to 2025, and whether these changes varied across disciplines. Overall lexical diversity remained exceptionally high and stable across the decade (see Table 8). Mean values



ranged narrowly between 93.8% and 94.7%, with medians consistently at or very near 100% in all degree-year groups (indicating many titles had no or minimal word repetition). Neither MA nor PhD titles showed statistically significant change between 2015 and 2025 (both Mann-Whitney U tests $p > 0.18$).

Table 8: Overall Lexical Diversity (%) by Degree and Year

| Degree | Year | Mean | Median | SD | N |
|--------|------|-------|--------|------|-------|
| MA | 2015 | 94.32 | 100 | 7.37 | 1,441 |
| MA | 2025 | 94.53 | 100 | 7.27 | 1,152 |
| PhD | 2015 | 94.43 | 100 | 7.25 | 3,336 |
| PhD | 2025 | 94.73 | 100 | 6.98 | 2,702 |

Discipline-level patterns revealed only small directional differences: slight increases in Literature, Musicology, and Philosophy, and very minor decreases in History and Religion. When broken down by degree and discipline, the large majority of 2015–2025 comparisons were non-significant. The sole exception was PhD Philosophy titles, which showed a modest but statistically significant increase in lexical diversity ($p = 0.009$), as highlighted in the pairwise tests in Table 9.

Table 9: Key Pairwise Comparisons of Lexical Diversity (2015 vs. 2025)

| Comparison | Test Type | Statistic | p-value | Interpretation |
|--|----------------|-----------|--------------|-----------------------|
| MA overall | Mann-Whitney U | 817,411.5 | 0.469 | No change |
| PhD overall | Mann-Whitney U | 4,425,280 | 0.187 | No change |
| PhD Philosophy | Mann-Whitney U | 155,680.5 | 0.009 | Significant increase |
| All other discipline/ degree combinations | Mann-Whitney U | — | > 0.05 | No significant change |

Kruskal-Wallis tests indicated significant differences among disciplines in every degree-year group (all $p \leq 0.0001$), with Philosophy and Musicology consistently exhibiting the highest lexical diversity (often 95–96% mean) and Religion and History the lowest (93–94%). These disciplinary contrasts persisted with only minor shifts over time (see Figure 4 for a visual comparison of lexical diversity by discipline across 2015 and 2025).

The near-ceiling levels of lexical diversity (medians at 100% in most groups) and the overall temporal stability suggest that thesis titles in these humanities fields already maximize vocabulary uniqueness, i.e., a hallmark of interpretive, concept-rich titling in History, Philosophy, Literature, and related disciplines (Haggan, 2004; Bérubé et al., 2018). The lack of widespread decline or surge contrasts with some LLM-era patterns in journal abstracts, where lexical complexity sometimes increased but cohesion or originality concerns arose (Bao et al., 2025; Geng & Trotta, 2024). The isolated increase in PhD Philosophy titles may reflect greater use of diverse abstract nouns or interpretive terms in doctoral work, potentially amplified by LLM polishing without introducing repetition. Overall, however, lexical diversity appears largely unaffected by



AI stabilization, underscoring the resilience of field-specific titling conventions rather than broad technological reshaping.

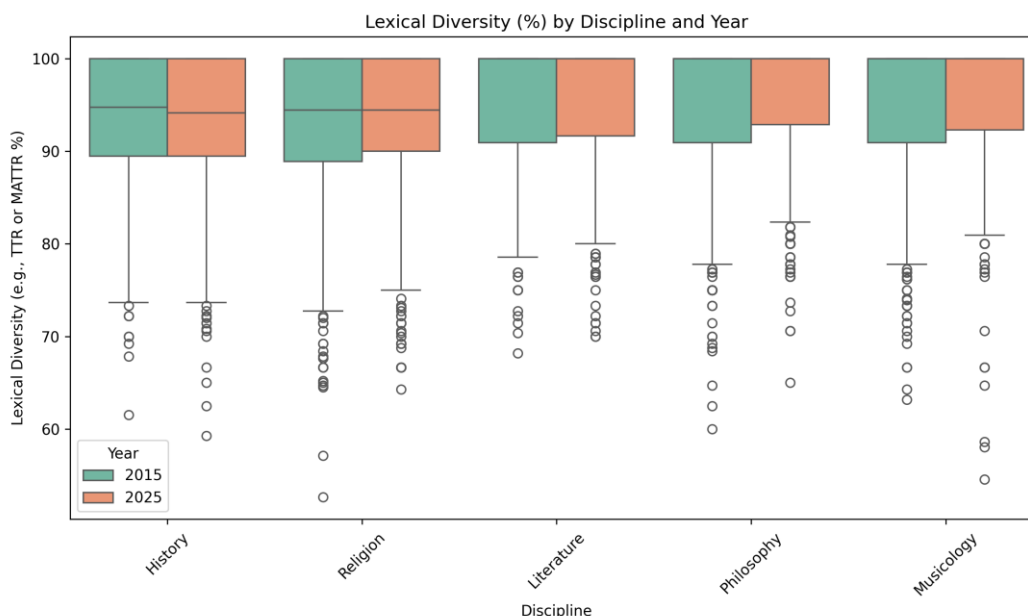


Figure 4. Boxplot of lexical diversity by discipline and year (2015 vs. 2025).

4.4. Syntactic Structure

RQ4 examined whether syntactic structures in MA and PhD thesis titles shifted between 2015 and 2025, and whether variations existed across disciplines. The dominant syntactic pattern in humanities thesis titles was Compound structures (titles with a colon, dash, or other divider separating segments), accounting for 70–74% overall. Nominal phrases ranked second (20–25%), while V-ing Phrase, Full Sentence, and Prepositional structures remained rare (each <5%). From 2015 to 2025, the overall distribution shifted modestly but significantly (chi-square = 39.305, $df = 4$, $p < 0.001$, Cramér's $V = 0.067$), driven mainly by an increase in Compound titles (70.02% → 73.92%) and a corresponding decrease in Nominal titles (24.99% → 19.93%) (see Table 10).

Table 10: Overall Distribution of Syntactic Categories by Year (%)

| Year | Compound | Full sentence | Nominal | Prepositional | V-ing phrase | Total |
|------|----------|---------------|---------|---------------|--------------|--------|
| 2015 | 70.02 | 1.19 | 24.99 | 0.75 | 3.04 | 99.99 |
| 2025 | 73.92 | 1.14 | 19.93 | 0.65 | 4.36 | 100.00 |

By degree level, both MA and PhD titles showed significant distributional changes over time (MA: chi-square = 24.295, $p < 0.001$, $V = 0.097$; PhD: chi-square = 32.437, $p < 0.001$, $V = 0.073$), as detailed in Table 11. Compound structures increased in both groups (MA: 66.07% → 73.00%; PhD: 71.73% → 74.32%), while Nominal structures



declined (MA: 28.38% → 20.83%; PhD: 23.53% → 19.54%). V-ing phrases showed a slight rise among PhD titles (2.79% → 4.77%).

Table 11: Distribution of Syntactic Categories by Degree and Year (%)

| Degree | Year | Compound | Full sentence | Nominal | Prepositional | V-ing phrase |
|--------|------|----------|---------------|---------|---------------|--------------|
| MA | 2015 | 66.07 | 1.67 | 28.38 | 0.28 | 3.61 |
| MA | 2025 | 73.00 | 1.82 | 20.83 | 0.95 | 3.39 |
| PhD | 2015 | 71.73 | 0.99 | 23.53 | 0.96 | 2.79 |
| PhD | 2025 | 74.32 | 0.85 | 19.54 | 0.52 | 4.77 |

Disciplinary patterns revealed persistent variation, with History showing the highest proportion of Compound titles in both years (81.02% in 2015 → 86.79% in 2025), Philosophy the highest Nominal usage (33.90% → 29.57%), and Religion a notable rise in V-ing phrases (4.27% → 8.00%). Most discipline-specific shifts were not individually significant (e.g., Philosophy: chi-square = 5.084, $p = 0.279$), except where driven by overall trends (e.g., increased Compound use in History and Musicology) (see Figure 5 for a visual breakdown of syntactic structure categories by discipline and year).

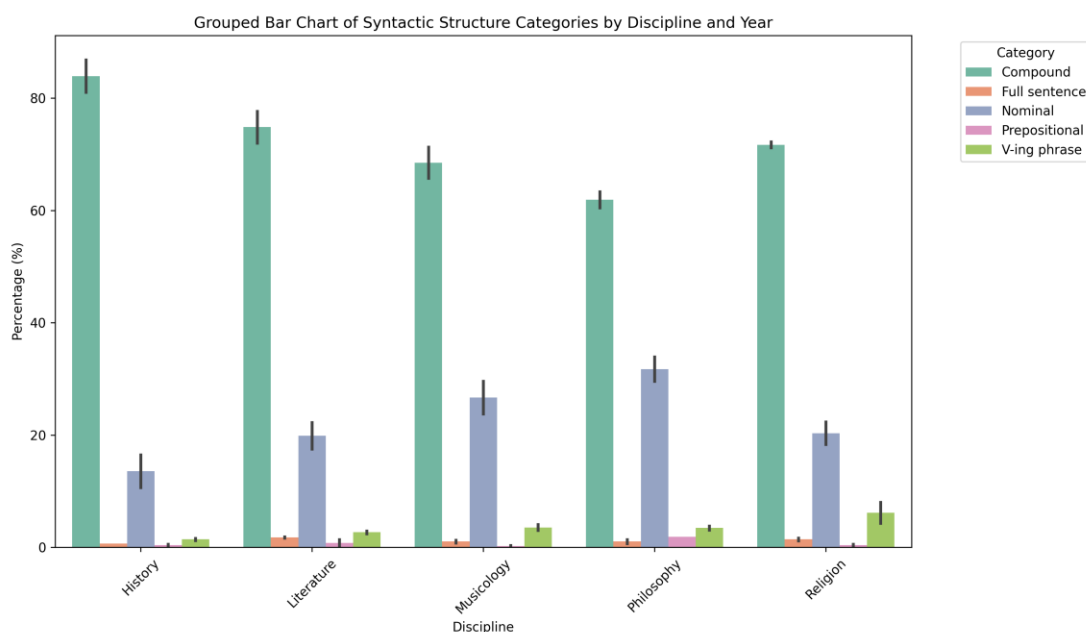


Figure 5. Grouped bar chart showing the distribution of syntactic structure categories by discipline and year (2015 vs. 2025).

The modest but statistically detectable increase in Compound structures (often colon-separated: main title + interpretive/subtitle) aligns with broader trends toward rhetorical marketization and descriptive elaboration in humanities titling (Chen & Liu, 2023; Haggan, 2004). This shift was more pronounced among MA titles, possibly reflecting greater LLM reliance for polishing shorter, less original works into more



conventionally attractive formats. The decline in simple Nominal phrases and slight rise in V-ing forms (especially in PhD Religion and Musicology) may indicate subtle reinforcement of exploratory or process-oriented phrasing, i.e., features that LLMs can readily generate or enhance (Bao et al., 2025). However, the changes remain small in magnitude (Cramér's V values ~ 0.07 – 0.10 indicate weak-to-moderate effect sizes), and disciplinary differences persisted strongly across both periods. Overall, AI stabilization appears to have modestly amplified existing genre preferences for compound, rhetorically layered titles rather than fundamentally altering syntactic conventions in humanities theses.

4.5. Semantic Content

RQ5 explored how the semantic content of MA and PhD thesis titles transformed from 2015 to 2025, and whether field-specific patterns emerged. Categories were non-exclusive (percentages can exceed 100% due to multi-component titles), with Topic as the default/general focus. Overall, Topic dominated semantic emphasis (≈ 48 – 49% of titles), followed by Dataset ($\approx 28\%$), Method ($\approx 15\%$), Result (≈ 5 – 6%), and Conclusion (≈ 2 – 3%). From 2015 to 2025, the distribution showed only minor, non-significant shifts (chi-square = 2.139, $df = 4$, $p = 0.710$, Cramér's V = 0.015). Topic slightly increased ($47.89\% \rightarrow 49.03\%$), while Dataset and Method showed small declines (see Table 12 for the overall distribution by year).

Table 12: Overall Distribution of Semantic Categories by Year (%)

| Year | Conclusion | Dataset | Method | Result | Topic |
|------|------------|---------|--------|--------|-------|
| 2015 | 2.53 | 28.31 | 15.56 | 5.70 | 47.89 |
| 2025 | 2.66 | 27.74 | 14.74 | 5.84 | 49.03 |

Degree-level patterns were similarly stable. Neither MA nor PhD titles exhibited significant distributional change (MA: chi-square = 3.199, $p = 0.525$; PhD: chi-square = 2.203, $p = 0.699$). Topic remained the most frequent category in both groups (≈ 49 – 50% for MA, ≈ 47 – 49% for PhD), with Dataset consistently second (≈ 26 – 29%). Minor fluctuations (e.g., slight rise in Topic for PhD titles) lacked statistical support (see Table 13 for the breakdown by degree and year).

Table 13: Distribution of Semantic Categories by Degree and Year (%)

| Degree | Year | Conclusion | Dataset | Method | Result | Topic |
|--------|------|------------|---------|--------|--------|-------|
| MA | 2015 | 1.72 | 26.21 | 14.82 | 7.38 | 49.87 |
| MA | 2025 | 2.20 | 27.52 | 12.89 | 7.55 | 49.84 |
| PhD | 2015 | 2.89 | 29.23 | 15.88 | 4.97 | 47.03 |
| PhD | 2025 | 2.86 | 27.84 | 15.53 | 5.10 | 48.67 |

Disciplinary differences were pronounced and largely persistent. History heavily emphasized Dataset (≈ 49 – 51%), reflecting archival and source-focused titling. Literature and Philosophy prioritized Topic (≈ 55 – 65%), consistent with interpretive and conceptual framing. Musicology and Religion showed balanced Topic and



Dataset/Method usage. No discipline exhibited a statistically significant shift between 2015 and 2025 (all $p > 0.05$; highest borderline: Musicology $p = 0.070$). Small increases in Topic were observed in Literature, Musicology, and Philosophy, while Dataset declined slightly in most fields (see Figure 6 for a visual comparison of semantic content distributions by discipline across 2015 and 2025).

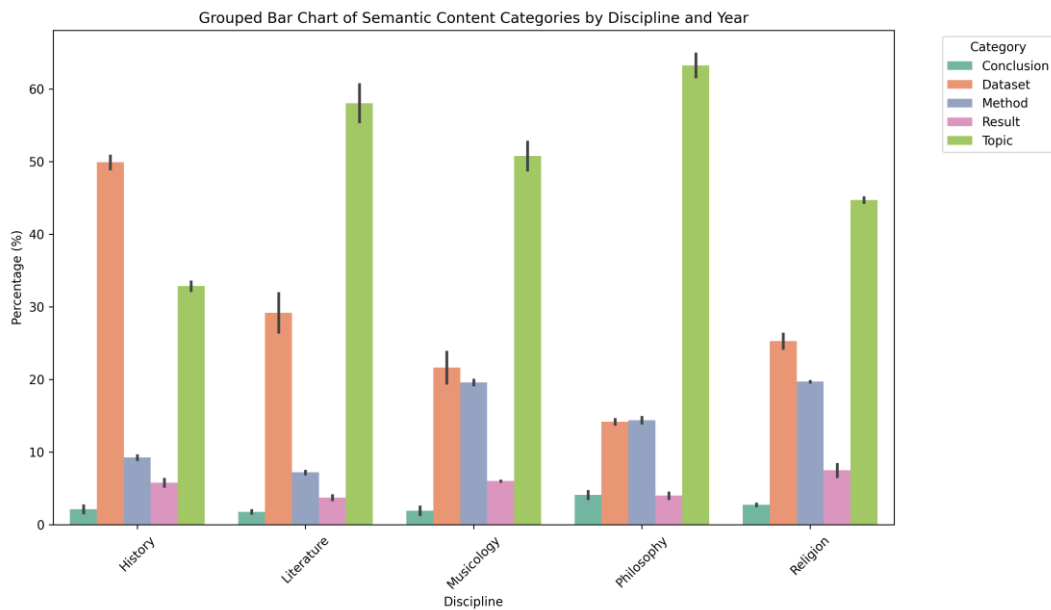


Figure 6. Stacked bar chart showing the distribution of semantic content categories by discipline and year (2015 vs. 2025).

The overall stability in semantic emphasis, dominated by broad Topic descriptions with frequent Dataset and Method specification, aligns with established humanities titling conventions that favor thematic breadth, source grounding, and theoretical framing (Sahragard & Meihami, 2016; Haggan, 2004). The absence of major shifts contrasts with some LLM-era observations in scientific abstracts, where Result-oriented or booster-heavy phrasing sometimes increased (Bao et al., 2025; Comas-Forgas, Koulouris & Kouis, 2025). The minor, non-significant rise in Topic emphasis may reflect subtle reinforcement of descriptive, interpretive titles, possibly aided by LLM polishing, without evidence of formulaic or conclusion-heavy over-representation often associated with generative AI. Disciplinary persistence remained the strongest pattern, suggesting AI tools, if used, primarily amplified existing genre norms rather than reshaping semantic priorities in humanities theses.

4.6. Summary

Across all linguistic dimensions analyzed, humanities thesis and dissertation titles exhibited remarkable stability from 2015 (pre-AI) to 2025 (post-AI stabilization), with minimal broad-scale shifts attributable to LLMs. Word length, informativity, and lexical diversity remained consistent (means ≈ 12 – 13 words, ≈ 67 – 68% substantive



words, $\approx 94\%$ unique words), showing no overall significant temporal changes and persistent disciplinary variations (e.g., longer titles in Religion, higher diversity in Philosophy). Syntactic structures displayed a modest increase in Compound forms ($70\% \rightarrow 74\%$), particularly in History and PhD titles, suggesting subtle amplification of rhetorical layering. Semantic content was dominated by Topic and Dataset emphases ($\approx 48\text{--}49\%$ and $\approx 28\%$), with no significant distributional shifts, underscoring interpretive and source-focused conventions. Isolated significant changes (e.g., longer PhD Religion titles, higher informativity in PhD Musicology, increased lexical diversity in PhD Philosophy) were field- and degree-specific, potentially reflecting targeted LLM polishing amid existing norms. Overall, AI appears to have reinforced rather than revolutionized titling practices, highlighting the resilience of humanities genre conventions amid technological adoption. Figure 7 illustrates the stability of mean title word length across disciplines and years, confirming the lack of broad temporal change. The modest rise in compound structures, visible across most disciplines, is further highlighted in Figure 8.

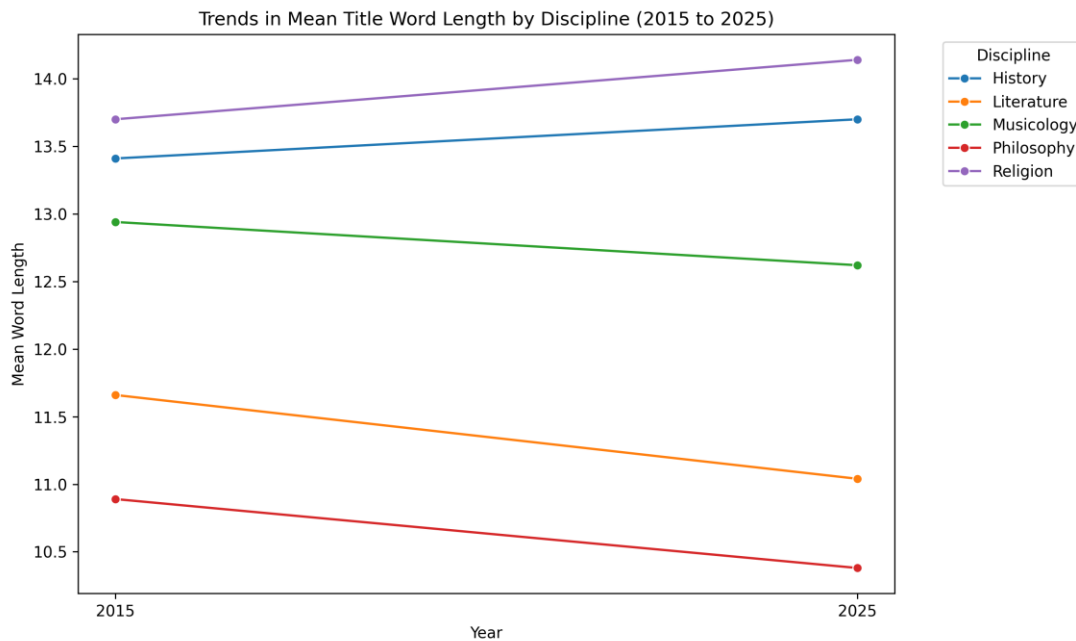


Figure 7. Line graph of mean title word length trends by discipline and year (2015 vs. 2025).

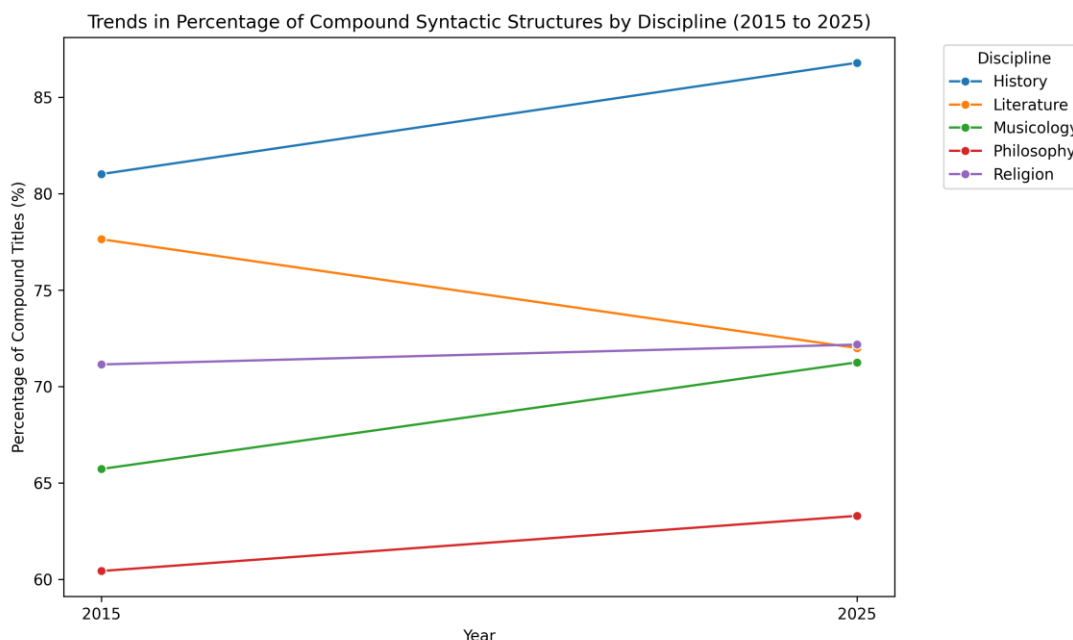


Figure 8. Line graph of percentage of Compound syntactic structures trends by discipline and year (2015 vs. 2025).

In the bigger picture, these findings indicate that while LLMs have permeated academic writing workflows, especially for NNES students and polishing tasks (Bao et al., 2025; Huang, 2010), their impact on foundational scholarly outputs like thesis titles remains limited and non-disruptive in humanities fields. This stability may stem from strong disciplinary gatekeeping, ethical guidelines against over-reliance (Thorpe, 2023; Uzun, 2023), or AI's tendency to mimic rather than innovate genre-specific styles. Future studies could extend this to full abstracts or manuscripts to assess broader transformations, informing policies on AI transparency and integrity in graduate education.

5. Closing Remarks

The comprehensive analysis of 8,631 MA and PhD thesis and dissertation titles across five core humanities disciplines, i.e., History, Religion, Literature, Philosophy, and Musicology, demonstrates a remarkable pattern of linguistic and structural stability between the pre-AI benchmark year of 2015 and the post-AI stabilization period of 2025. Despite the widespread adoption of LLMs such as ChatGPT in academic workflows, core metrics including title length (stable at approximately 12–13 words), informativity (approximately 67% substantive words), lexical diversity (medians consistently near 100%), and semantic emphasis (dominated by Topic and Dataset categories) exhibited no broad, statistically significant shifts. The sole modest change, a statistically detectable rise in compound syntactic structures from 70% to 74% (primarily colon-separated titles), represents an amplification of longstanding humanities conventions rather than a disruptive innovation. Disciplinary differences



persisted robustly (e.g., longer and more dataset-oriented titles in History and Religion; greater nominal conciseness and lexical diversity in Philosophy), underscoring the resilience of field-specific genre norms (Haggan, 2004; Sahragard & Meihami, 2016; Nagano, 2015).

These findings stand in notable contrast to the pronounced LLM-driven transformations documented in scientific and biomedical literature. Studies of arXiv preprints, PubMed Central articles, and abstracts have reported sharp increases in LLM-preferred vocabulary (e.g., ‘delve’, ‘underscore’, ‘intricate’), elevated lexical complexity coupled with reduced syntactic cohesion and readability, and higher modification probabilities peaking at 17.5% in computer science (Bao et al., 2025; Kousha & Thelwall, 2024; Liang et al., 2024; Kobak et al., 2024). AI-generated titles have even been rated as more appealing and content-aligned than human-written counterparts in blinded evaluations (Sebo et al., 2026). The absence of such effects in humanities thesis titles negates the notion of uniform stylistic convergence across academia and instead supports the view that interpretive, rhetorically nuanced fields may be less susceptible to formulaic homogenization (cf. Geng & Trotta, 2024; Comas-Forgas, Koulouris & Kouis, 2025).

A key factor contributing to this muted impact in thesis titles, particularly compared to longer-form elements like abstracts or full manuscripts, is the inherent brevity and high-stakes nature of titles themselves. Titles are short, condensed artifacts (typically 12–13 words here) that serve as the primary “face” of a multi-year scholarly project. Unlike extended texts that benefit substantially from AI-assisted drafting, editing, polishing, or expansion, where iterative LLM refinements can introduce detectable stylistic markers, titles demand precision, originality, and alignment with the thesis’s core interpretive contribution. The effort required for meaningful AI intervention on such a compact element is often disproportionate to the gain: a student or supervisor can craft or refine a title manually in minutes, with direct human oversight ensuring it captures nuanced disciplinary voice and avoids generic phrasing. This contrasts sharply with full papers or abstracts, where time-intensive editing of hundreds or thousands of words makes LLM assistance more “worthwhile” for efficiency, leading to greater detectable traces (Kousha & Thelwall, 2024; Kobak et al., 2024). The extended timeline of thesis development (often 1–7 years, with close supervisory guidance) further reinforces human intervention at critical framing stages like titling, limiting opportunities for unchecked AI overhauls. Thus, the observed stability likely reflects not only disciplinary resilience but also practical barriers to deep LLM penetration in this specific subgenre. This resilience likely stems from strong supervisory gatekeeping, theses’ emphasis on original scholarly voice, and LLMs’ tendency to mimic rather than override established disciplinary conventions, i.e., factors less operative in high-volume, empirical publishing contexts (Thorpe, 2023; Uzun, 2023; Huang, 2010).

The present study’s limitations warrant recognition. First, the study focuses solely on titles, i.e., short, highly conventionalized texts that may conceal subtler LLM influences



detectable in longer-form elements such as abstracts or full chapters (as seen in full-text PubMed analyses; Kousha & Thelwall, 2024). Second, reliance on ProQuest data, while comprehensive, introduces potential coverage and indexing biases that could underrepresent certain institutions or non-English-dominant contexts. Third, the study employs indirect inference rather than direct LLM-detection models (e.g., those used by Liang et al., 2024 or Kobak et al., 2024), rendering causal attribution to AI probabilistic rather than definitive. Finally, the snapshot design (2015 vs. 2025) omits transitional dynamics in 2023–2024 and cannot fully isolate AI effects from concurrent changes in graduate training or publishing pressures.

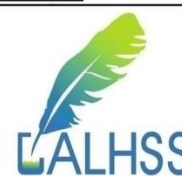
Future research directions are therefore essential. Longitudinal extensions incorporating full thesis abstracts, literature reviews, or complete manuscripts would permit finer-grained tracking of lexical, syntactic, and semantic shifts using advanced LLM-probability tools. Comparative studies across STEM disciplines, cross-cultural corpora (especially NNES student work), and post-2025 cohorts would clarify whether the observed humanities resilience generalizes or erodes over time. Mixed-methods designs combining quantitative metrics with surveys or interviews on actual AI usage patterns among graduate students could further illuminate adoption disparities by degree level and discipline. Such work would directly inform institutional policies on AI transparency, authorship attribution, and integrity training (building on calls by Thorp, 2023 and Uzun, 2023).

In conclusion, the stabilization of generative AI by 2025 has reinforced rather than revolutionized titling practices in humanities graduate scholarship. By amplifying existing rhetorical preferences without eroding disciplinary distinctiveness, LLMs appear to function primarily as efficiency-enhancing tools within well-established genre boundaries, particularly where the brevity and human-supervised nature of titles limit extensive intervention. These results temper broader anxieties about AI-induced homogenization in scholarly communication and highlight the interpretive robustness of humanistic fields. As generative technologies continue to evolve, sustained empirical scrutiny, paired with ethical vigilance, will remain critical to preserving the authenticity and diversity of academic voice across all disciplines.



References

1. Altmäe, S., Sola-Leyva, A., & Salumets, A. (2023). Artificial intelligence in scientific writing: A friend or a foe? *Reproductive BioMedicine Online*, 47(1), 3–9. <https://doi.org/10.1016/j.rbmo.2023.04.009>
2. Atkinson, D. (1998). *Scientific discourse in sociohistorical context: The Philosophical Transactions of the Royal Society of London, 1675-1975*. Routledge. <https://doi.org/10.4324/9781410601704>
3. Ball, R. (2009). Scholarly communication in transition: The use of question marks in the titles of scientific articles in medicine, life sciences and physics 1966–2005. *Scientometrics*, 79(3), 667–679. <https://doi.org/10.1007/s11192-007-1984-5>
4. Bao, T., Zhao, Y., Mao, J., & Zhang, C. (2025). Examining linguistic shifts in academic writing before and after the launch of ChatGPT: A study on preprint papers. *Scientometrics*, 130, 3597–3627. <https://doi.org/10.1007/s11192-025-05341-y>
5. Bérubé, N., Sainte-Marie, M., Mongeon, P., & Larivière, V. (2018). Words by the tail: Assessing lexical diversity in scholarly titles using frequency-rank distribution tail fits. *PLoS ONE*, 13(7), Article e0197775. <https://doi.org/10.1371/journal.pone.0197775>
6. Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings*. Routledge. <https://doi.org/10.4324/9781315844992>
7. Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, 9(1), 2–20. <https://doi.org/10.1016/j.jeap.2010.01.001>
8. Buxton, A. B., & Meadows, A. J. (1977). The variation in the information content of titles of research papers with time and discipline. *Journal of Documentation*, 33(1), 46–52. <https://doi.org/10.1108/eb026626>
9. Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., & Sun, L. (2023). A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to ChatGPT (arXiv:2303.04226). *arXiv*. <https://doi.org/10.48550/arXiv.2303.04226>
10. Chen, B., Deng, D., Zhong, Z., & Zhang, C. (2020). Exploring linguistic characteristics of highly browsed and downloaded academic articles. *Scientometrics*, 122(3), 1769–1790. <https://doi.org/10.1007/s11192-020-03361-4>
11. Chen, X., & Liu, H. (2023). Academic “click bait”: A diachronic investigation into the use of rhetorical part in pragmatics research article titles. *Journal of English for Academic Purposes*, 66, Article 101306. <https://doi.org/10.1016/j.jeap.2023.101306>
12. Cheng, S. W., Kuo, C.-W., & Kuo, C.-H. (2012). Research article titles in applied linguistics. *Journal of Academic Language and Learning*, 6(1), A1–A14. Retrieved from <https://journal.aall.org.au/index.php/jall/article/view/178>
13. Comas-Forgas, R., Koulouris, A., & Kouis, D. (2025). ‘AI-navigating’ or ‘AI-sinking’? An analysis of verbs in research articles titles suspicious of containing AI-generated/assisted content. *Learned Publishing*, 38(1), Article e1647. <https://doi.org/10.1002/leap.1647>
14. Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text



- cohesion. Behavior Research Methods, 48(4), 1227–1237. <https://doi.org/10.3758/s13428-015-0651-7>
15. Dillon, J. T. (1982). In pursuit of the colon: A century of scholarly progress: 1880-1980. Journal of Higher Education, 53(1), 93–99. <https://doi.org/10.2307/1981541>
16. Dong, S., Mao, J., Ke, Q., & Pei, L. (2024). Decoding the writing styles of disciplines: A large-scale quantitative analysis. Information Processing & Management, 61(4), Article 103718. <https://doi.org/10.1016/j.ipm.2024.103718>
17. Elsevier. (2024). Insights 2024: Attitudes toward AI – Full report. Elsevier. <https://www.elsevier.com/insights/attitudes-toward-ai>
18. Flowerdew, J. (1999). Problems in writing for scholarly publication in English: The case of Hong Kong. Journal of Second Language Writing, 8(3), 243–264. [https://doi.org/10.1016/S1060-3743\(99\)80116-7](https://doi.org/10.1016/S1060-3743(99)80116-7)
19. Gao, C.A., Howard, F.M., Markov, N.S. et al. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. NPJ Digital Medicine, 6, 75. <https://doi.org/10.1038/s41746-023-00819-6>
20. Geng, M., & Trotta, R. (2024). Is ChatGPT transforming academics' writing style? (arXiv:2404.08627). arXiv. <https://doi.org/10.48550/arXiv.2404.08627>
21. Goodman, R. A., Thacker, S. B., & Siegel, P. Z. (2001). What's in a title? A descriptive study of article titles in peer-reviewed medical journals. Science Editor, 24(3), 75–78.
22. Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of innovations in service organizations: Systematic review and recommendations. The Milbank Quarterly, 82(4), 581–629. <https://doi.org/10.1111/j.0887-378X.2004.00325.x>
23. Gross, A. G., Harmon, J. E., & Reidy, M. S. (2002). Communicating science: The scientific article from the 17th century to the present. Oxford University Press.
24. Haggan, M. (2004). Research paper titles in literature, linguistics and science: Dimensions of attraction. Journal of Pragmatics, 36(2), 293–317. [https://doi.org/10.1016/S0378-2166\(03\)00090-0](https://doi.org/10.1016/S0378-2166(03)00090-0)
25. Hartley, J. (2005). To attract or to inform: What are titles for? Journal of Technical Writing and Communication, 35(2), 203–213. <https://doi.org/10.2190/nv6e-fn3n-7ngn-twqt>
26. Hartley, J. (2007). Planning that title: Practices and preferences for titles with colons in academic articles. Library & Information Science Research, 29(4), 553–568. <https://doi.org/10.1016/j.lisr.2007.05.002>
27. Hu, H., Wang, D., & Deng, S. (2021). Analysis of the scientific literature's abstract writing style and citations. Online Information Review, 45(7), 1290–1305. <https://doi.org/10.1108/OIR-05-2020-0188>
28. Huang, J. C. (2010). Publishing and learning writing for publication in English: Perspectives of NNES PhD students in science. Journal of English for Academic Purposes, 9(1), 33–44. <https://doi.org/10.1016/j.jeap.2009.10.001>



29. Huang, J., & Tan, M. (2023). The role of ChatGPT in scientific communication: Writing better scientific review articles. *American Journal of Cancer Research*, 13(4), 1148–1154.
30. Hyland, K., & Zou, H. J. (2022). Titles in research articles. *Journal of English for Academic Purposes*, 56, 1-13. <https://doi.org/10.1016/j.jeap.2022.101094>
31. Jaime-Sisó, M. (2009). Titles or headlines? Anticipating conclusions in biomedical research article titles as a persuasive journalistic strategy to attract busy readers. *Miscelánea: A Journal of English and American Studies*, 39, 29–54.
32. Juzek, T. S., & Ward, Z. B. (2025). Why does ChatGPT “delve” so much? Exploring the sources of lexical overrepresentation in large language models. In *Proceedings of the 31st International Conference on Computational Linguistics* (pp. 6397–6411).
33. Kobak, D., González-Márquez, R., Horvát, E. Á., & Lause, J. (2024). Delving into ChatGPT usage in academic writing through excess vocabulary (arXiv:2406.07016). arXiv. <https://doi.org/10.48550/arXiv.2406.07016>
34. Kousha, K., & Thelwall, M. (2024). How much are LLMs changing the language of academic papers after ChatGPT? A multi-database and full text analysis (arXiv:2509.09596). arXiv. <https://doi.org/10.48550/arXiv.2509.09596>
35. Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, 49(4), 757–786. <https://doi.org/10.1002/tesq.194>
36. Lebrun, J. (2011). Title: The face of your paper. In J. Lebrun (Ed.), *Scientific writing 2.0: A reader and writer’s guide* (pp. 115–133). World Scientific Publishing.
37. Liang, W., Zhang, Y., Wu, Z., Lepp, H., Ji, W., Zhao, X., Cao, H., Liu, S., He, S., Huang, Z., Yang, D., Potts, C., Manning, C. D., & Zou, J. Y. (2024). Mapping the increasing use of LLMs in scientific papers (arXiv:2404.01268). arXiv. <https://doi.org/10.48550/arXiv.2404.01268>
38. Liu, J., & Bu, Y. (2024). Towards the relationship between AIGC in manuscript writing and author profiles: Evidence from preprints in LLMs (arXiv:2404.15799). arXiv. <https://doi.org/10.48550/arXiv.2404.15799>
39. Lozić, E., & Štular, B. (2023). ChatGPT v Bard v Bing v Claude 2 v Aria v human-expert. How good are AI chatbots at scientific writing? *Future Internet*, 15(10), Article 336. <https://doi.org/10.3390/fi15100336>
40. Lu, C., Bu, Y., Wang, J., Ding, Y., Torvik, V., Schnaars, M., & Zhang, C. (2019). Examining scientific writing styles from the perspective of linguistic complexity. *Journal of the Association for Information Science and Technology*, 70(5), 462–475. <https://doi.org/10.1002/asi.24126>
41. Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4), 474–496. <https://doi.org/10.1075/ijcl.15.4.02lu>
42. Martín, P., & León Pérez, I. (2024). The evolution of sub-disciplinary linguistic trends: A diachronic study of biomedical research article titles. *Journal of English for Research Publication Purposes*, 5(1/2), 60–81. <https://doi.org/10.1075/jerpp.00027.mar>



43. Nagano, R. L. (2015). Research article titles and disciplinary conventions: A corpus study of eight disciplines. *Journal of Academic Writing*, 5(1), 133–144. <https://doi.org/10.18552/joaw.v5i1.168>
44. OUP (2024). How are researchers responding to AI? Oxford University Press, May 23, 2024, <https://corp.oup.com/news/how-are-researchers-responding-to-ai/>
45. Plavén-Sigray, P., Matheson, G. J., Schiffler, B. C., & Thompson, W. H. (2017). The readability of scientific texts is decreasing over time. *eLife*, 6, Article e27725. <https://doi.org/10.7554/eLife.27725>
46. Sahragard, R., & Meihami, H. (2016). A diachronic study on the information provided by the research titles of applied linguistics journals. *Scientometrics*, 108(3), 1319–1335. <https://doi.org/10.1007/s11192-016-2049-4>
47. Sebo, P., Nie, B., & Wang, T. (2026). Can ChatGPT write better scientific titles? A comparative evaluation of human-written and AI-generated titles. *F1000Research*, 14, Article 1470. <https://doi.org/10.12688/f1000research.173647.2>
48. Soler, V. (2007). Writing titles in science: An exploratory study. *English for Specific Purposes*, 26(1), 90–102. <https://doi.org/10.1016/j.esp.2006.08.001>
49. Soler, V. (2011). Comparative and contrastive observations on scientific articles written in English and Spanish. *English for Specific Purposes*, 30(1), 124–137. <https://doi.org/10.1016/j.esp.2010.09.002>
50. Song, N., Chen, K., & Zhao, Y. (2023). Understanding writing styles of scientific papers in the ISLS domain: Evidence from abstracts over the past three decades. *Journal of Informetrics*, 17(1), Article 101377. <https://doi.org/10.1016/j.joi.2023.101377>
51. Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.
52. Swales, J. M., & Feak, C. B. (1994). *Academic writing for graduate students* (pp. 155–156). University of Michigan Press.
53. Thorp, H. H. (2023). ChatGPT is fun, but not an author. *Science*, 379(6630), 313. <https://doi.org/10.1126/science.adg7879>
54. Uzun, L. (2023). ChatGPT and academic integrity concerns: Detecting artificial intelligence generated content. *Language Education and Technology*, 3(1), Article 1. <http://www.langedutech.com/letjournal/index.php/let/article/view/49>
55. Vinkers, C. H., Tjeldink, J. K., & Otte, W. M. (2015). Use of positive and negative words in scientific PubMed abstracts between 1974 and 2014: Retrospective analysis. *BMJ*, 351, Article h6467. <https://doi.org/10.1136/bmj.h6467>
56. Wang, G., Wang, H., Sun, X., Wang, N., & Wang, L. (2023). Linguistic complexity in scientific writing: A large-scale diachronic study from 1821 to 1920. *Scientometrics*, 128(1), 441–460. <https://doi.org/10.1007/s11192-022-04550-z>
57. Xiang, X., & Li, J. (2020). A diachronic comparative study of research article titles in linguistics and literature journals. *Scientometrics*, 122(2), 847–866. <https://doi.org/10.1007/s11192-019-03329-z>