

Article

Tracing the Evolution of Reviews and Research Articles in the Biomedical Literature: A Multi-Dimensional Analysis of Abstracts

Stefano Guizzardi , Maria Teresa Colangelo , Prisco Mirandola and Carlo Galli * 

Histology and Embryology Laboratory, Department of Medicine and Surgery, University of Parma, Via Volturmo 39, 43126 Parma, Italy; stefano.guizzardi@unipr.it (S.G.); mariateresa.colangelo@studenti.unipr.it (M.T.C.); prisco.mirandola@unipr.it (P.M.)

* Correspondence: carlo.galli@unipr.it

Abstract: We previously examined the diachronic shifts in the narrative structure of research articles (RAs) and review manuscripts using abstract corpora from MEDLINE. This study employs Nini's Multidimensional Analysis Tagger (MAT) on the same datasets to explore five linguistic dimensions (D1–5) in these two sub-genres of biomedical literature, offering insights into evolving writing practices over 30 years. Analyzing a sample exceeding 1.2 million abstracts, we observe a shared reinforcement of an informational, emotionally detached tone (D1) in both RAs and reviews. Additionally, there is a gradual departure from narrative devices (D2), coupled with an increase in context-independent content (D3). Both RAs and reviews maintain low levels of overt persuasion (D4) while shifting focus from abstract content to emphasize author agency and identity. A comparison of linguistic features underlying these dimensions reveals often independent changes in RAs and reviews, with both tending to converge toward standardized stylistic norms.

Keywords: abstract; narrativity; scientific publishing



Citation: Guizzardi, S.; Colangelo, M.T.; Mirandola, P.; Galli, C. Tracing the Evolution of Reviews and Research Articles in the Biomedical Literature: A Multi-Dimensional Analysis of Abstracts. *Publications* **2024**, *12*, 2. <https://doi.org/10.3390/publications12010002>

Academic Editor: Yin Ling Cheung

Received: 24 July 2023

Revised: 5 January 2024

Accepted: 10 January 2024

Published: 12 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Biomedical publishing plays a crucial role in the field of biomedicine and has significant importance for the dissemination of scientific knowledge [1]. The release of research articles facilitates the systematic evaluation, synthesis, and analysis of data derived from various studies. The acquisition of robust evidence is imperative for guiding clinical guidelines, shaping treatment protocols, influencing public health policies, and steering healthcare interventions [2]. Moreover, the publication of biomedical literature serves as a conduit through which researchers and scientists globally can share their findings, discoveries, and innovations. This dissemination is indispensable for advancing scientific understanding and fostering collaboration across borders. Beyond its collaborative impact, the act of publishing research findings also champions transparency and accountability within the scientific community. Researchers willingly subject themselves to scrutiny and evaluation by their peers, creating an environment that encourages the conduct of rigorous, well-designed studies. Simultaneously, this system discourages unethical practices and the misrepresentation of data [3].

Biomedical publishing not only serves as a cornerstone for career advancement among researchers, scientists, and healthcare professionals but also plays a pivotal role in shaping their visibility, credibility, and professional reputation. The quantity and impact of publications carry substantial weight in academic promotions, grant applications, and funding decisions [4,5]. This means that there are strong incentives to publish, and the number of papers has steadily grown in the last years, reaching levels that make it very difficult for researchers in any field to keep abreast of the published material [6], a problem made even more acute by the surge in predatory journals [7].

The abundance of literature has given rise to various consequences, including the proliferation of review articles. While these articles, in their traditional narrative form, offer valuable summaries of the literature, they also present a challenge due to their inherent susceptibility to redundancy [8,9]. An additional significant consequence is the increasing importance of effective communication. In an era flooded with thousands of papers, being able to communicate findings, thoughts, and opinions effectively becomes paramount. This is particularly true for abstracts, the concise pieces of text readers encounter after the title, which serve as a critical factor in determining a manuscript's interest. Crafting an abstract that communicates content effectively—complemented by substantive content, of course—can significantly impact how frequently the associated paper is noticed and read amid the extensive background noise. Understanding the evolving landscape of communication practices is crucial for optimizing and maximizing the impact of one's scientific production. Particularly, in the context of abstracts, it is essential to comprehend what makes them effective and how these characteristics may have evolved alongside changes in scientific approaches over the years. This understanding serves as a valuable compass in navigating the ever-developing landscape of knowledge. To delve into the linguistic features characterizing abstracts in research articles and reviews, we utilized the freely available software Multidimensional Tagger (MAT) version 1.3.3, developed by Nini and based on Biber's multidimensional analysis.

Multidimensional text analysis is a comprehensive framework used in corpus linguistics to examine and understand various dimensions of linguistic variation in written and spoken language [10]. Developed by Douglas Biber, this multidimensional analysis aims to identify and describe these variations across multiple linguistic dimensions, including lexical, syntactic, morphological, and discursive features [11]. Biber's multidimensional analysis involved extracting linguistic features from 481 spoken and written texts of contemporary British English, which were then used to compute a series of dimension scores through a factor analysis of the co-occurrences of these features. The texts that Biber used were taken from the Lancaster–Oslo–Bergen Corpus [12] and the London–Lund Corpus [13], which were chosen because they represent over 20 major register categories, including academic writing in many fields (fiction, letters, conversations, etc.). Biber proceeded to cluster his 67 features into five dimensions [14], which he interpreted as listed in Table 1.

Table 1. Outline of the 5 dimensions of Biber's multidimensional analysis that were used in the present paper [11].

Dimension	Feature
1	Involved vs. Informational discourse
2	Narrative vs. Non-Narrative Concerns
3	Context-Independent Discourse vs. Context-Dependent Discourse
4	Overt Expression of Persuasion
5	Abstract and Non-Abstract Information

Biber's multidimensional (MD) analysis has been extensively used to investigate corpora of different kinds and origins, such as abstracts published in different countries [15] or by writers of different origins [16], and has proved very useful because it concisely provides an overview of the general linguistic and rhetorical stances of a text in the broader context of the literature production in many fields and genres. Through the Multi-dimensional Analysis Tagger [17], texts can be added to Biber's MD analysis of English, as it replicates Biber's 67 original features used to compute his dimension scores, on which we have also decided to rely for the present work.

We previously characterized the narrative arcs in research articles and literature reviews in the biomedical field by applying the LIWC 2022 analysis tool [18] to a corpus of abstracts from research articles and reviews obtained from MEDLINE over the course of the last 33 years [19]. Building upon this foundation, we proceeded to apply Biber's analysis to

the same corpus. This analytical approach aims to provide deeper insights into the extant similarities and differences in these two genres as well as linguistic changes that might have occurred in the 1989–2022 period, shedding light on the nuances and divergences in these transformations over time.

2. Materials and Methods

The datasets that we used for the present study were composed of two independent corpora of abstracts of scientific manuscripts obtained from MEDLINE and previously published in [19]. MEDLINE is one of the world's largest and most used repositories for biomedical literature; it is run by the National Institute of Health and is freely accessible through a portal called Pubmed, which works as its search engine. To retrieve data from MEDLINE in a way that was conducive to our analysis, we accessed it via command line through the Pubmed API using Python running on Jupyter notebooks [20], a popular development environment for this programming language. Briefly, we used the Python 3.9 litter-getter library [21] to search and retrieve the abstracts by connecting to Pubmed, performing a search, and automatically downloading the data in a format we could use for subsequent analysis. We relied on the following search terms:

1. #1 'year[dp] NOT Review[pt]';
2. #2 'year[dp] AND Review[pt]'.

Pubmed uses a simple syntax for queries, and search terms can be easily joined by Boolean operators; plus, tags within brackets can be added to limit the search by keywords to the desired field in the record. In this case, the [dp] tag refers to the field 'Date of Publication', while the [pt] tag limits the search for the keyword to the 'Publication Type' field. The word 'year' in our search is not really a keyword but a Python variable, which we set to iterate from 1989 through 2022. These queries retrieved two lists of Pubmed IDs (PMIDs). Search #1 generated a list of abstracts from PMIDs excluding the 'Review' type, and search #2 generated a list of PMIDs exclusively constituting 'Review' abstracts in the same time interval. The reason for this distinction is that Reviews are a genre of scientific article that comprise several peculiar sub-types including 'Narrative reviews' or 'Systematic reviews', each with its distinctive purposes and structure [22], and we hypothesized that reviews may display different linguistic features from research articles, in agreement with the differences in narrativity highlighted by our previous LIWC analysis [19].

As previously explained [19], to balance our corpora, we randomly sorted 20,000 PMIDs out of the total number of retrieved PMIDs for each year, and proceeded to retrieve the data from the PMID list, thus obtaining 2 independent corpora as follows:

1. #1 Abstracts from Research articles (excluding Reviews), published between 1989 and 2022 ($n = 680,000$);
2. #2 Abstracts from Review articles, published between 1989 and 2022 ($n = 680,000$).

To obtain the abstract texts, litter-getter downloaded an XML file for each publication containing all the data in the record, based on its PMID; we then created a Pandas Dataframe [23], which is a special tabular form, not dissimilar from an Excel sheet, and populated it using the BeautifulSoup library [24]. BeautifulSoup is a Python library specifically designed to extract and clean XML data, i.e., recognize XML tags and isolate the desired information. The table was created in such a way that each row contained a publication, and it had columns for the authors, title of the publication, journal, abstract, etc. We then took the abstract column; we lowercased all the abstracts and saved them as a separate text (.txt) file.

This file was passed into the Multidimensional Analysis Tagger v 1.3.3 [17], freely available at <https://sites.google.com/site/multidimensionaltagger> (accessed on 20 June 2023). This tagger is grounded in Biber's (1988) Variation across Speech and Writing tagger for the multidimensional functional analysis of English texts. Unlike Biber's framework, this program is based on the Stanford Tagger [17] and generates both a grammatically annotated version of the text as well as statistics following Biber's method [10]. This tool is

very user-friendly, thanks to its graphical interface, and its output is a series of scores for the 5 dimensions outlined by Biber, plus scores for each of the underlying 67 linguistic features (Appendix A), all expressed as Z scores, in separate comma-separated value files (.csv). A Z score is, simply put, a measure of the distance of a score for a given sample from the mean of that score for a whole population based on Biber's corpus [25], expressed as number of standard deviations from the mean. So, in our case, the MAT software contains the means for each score for a vast corpus of texts from various genres, including conversations, speeches, personal letters, broadcasts, and academic writing [17]. As an example, a Z score of 2 for any linguistic feature means that this score is 2 standard deviations above the mean of that mixed corpus, which is representative of a general literature production. All the analyses were conducted on Jupyter notebooks [20] by importing the .csv files back into Pandas tables in Python. Matplotlib [26] and Seaborn [27] libraries were then used to plot the data [20].

3. Results and Discussion

The two corpora comprised 680,000 abstracts each, without overlap, because of the way they were selected. The selection criteria for corpus #1, however, had a consequence, i.e., that this corpus contained not only RAs, but also a small number of different genres. A post hoc analysis on the corpus showed that 611,450 abstracts out of the total 680,000 in #1 corpus belonged to research articles, and 43,567 abstracts (7.1%) belonged to the comment, letter, and editorial categories, which do not fall within our area of interest, while the remaining 24,983 could be classified as less frequent manuscript types, e.g., news or historical articles [19].

3.1. Dimension 1

Our analysis commenced with Dimension 1 (D1) in MAT, a dimension that, in this context, signifies the level of informational versus involved discourse. The positive pole of D1 is typically linked to dialogues characterized by language rich in interaction and expressive affective content, as can be found in personal correspondence [11]. Conversely, the negative pole of D1 is associated with information-rich and highly edited text, aligning with the expectations from a textbook, or an academic article [28]. Consistently, both research articles (RAs) and reviews in our corpus exhibit negative scores (Figure 1A). While the D1 scores for RAs have remained relatively constant over time, those for reviews have become slightly more negative, i.e., they have become even more aligned to the informational pole of discourse. To delve deeper into how the dimensions evolved within our two document groups, we employed scatterplots depicting the values of different features for RAs and reviews over time. Figure 1B illustrates that RAs with varying D1 scores (and thus with more or less pronounced informational natures) are present across all publication dates. In contrast, however, newer reviews tend to cluster around comparatively more negative values than older ones, suggesting that newer reviews exhibit stronger informational traits. To enhance the transparency of the dimension score and gain further insight into the evolving phenomena in the literature, we turned our attention to analyzing the underlying linguistic features associated with D1.

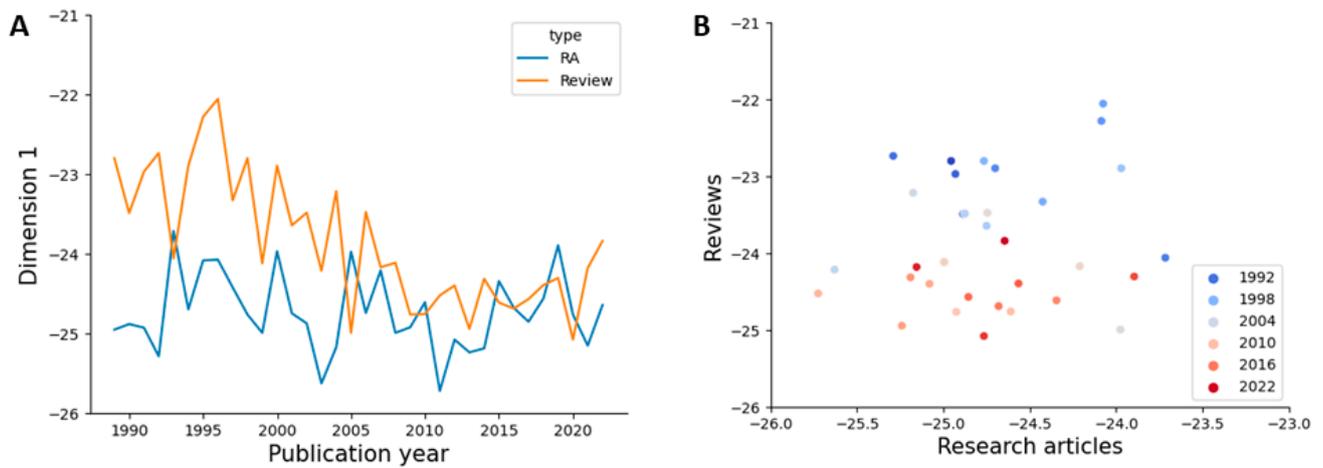


Figure 1. (A) Line plot of Dimension 1 (D1) score over the years for the research article (RA) corpus and the review corpus, in blue and orange, respectively; (B) scatter plot of D1 score for RAs and reviews.

Notably, a frequent use of **nouns** and **long words** is a characteristic feature of information-rich texts, corresponding to negative scores for D1, as these features require careful planning in production and are less frequent in impromptu speeches and dialogues. Consistently with this observation, our analysis reveals that both RAs and reviews have increased their Z scores—i.e., frequency—for these features over the years in a parallel manner (Figure 2A,B). However, no clear trend is discernible for two additional features typical of texts in the negative pole of D1 score: the **type/token** ratio and the frequency of **attributive adjectives** (Figure S1). The former parameter reflects the ratio between different words and the total number of words (tokens), indicating the diversity of language used. The latter parameter is associated with a language rich in adjectives, which, again, is a common feature of planned discourses.

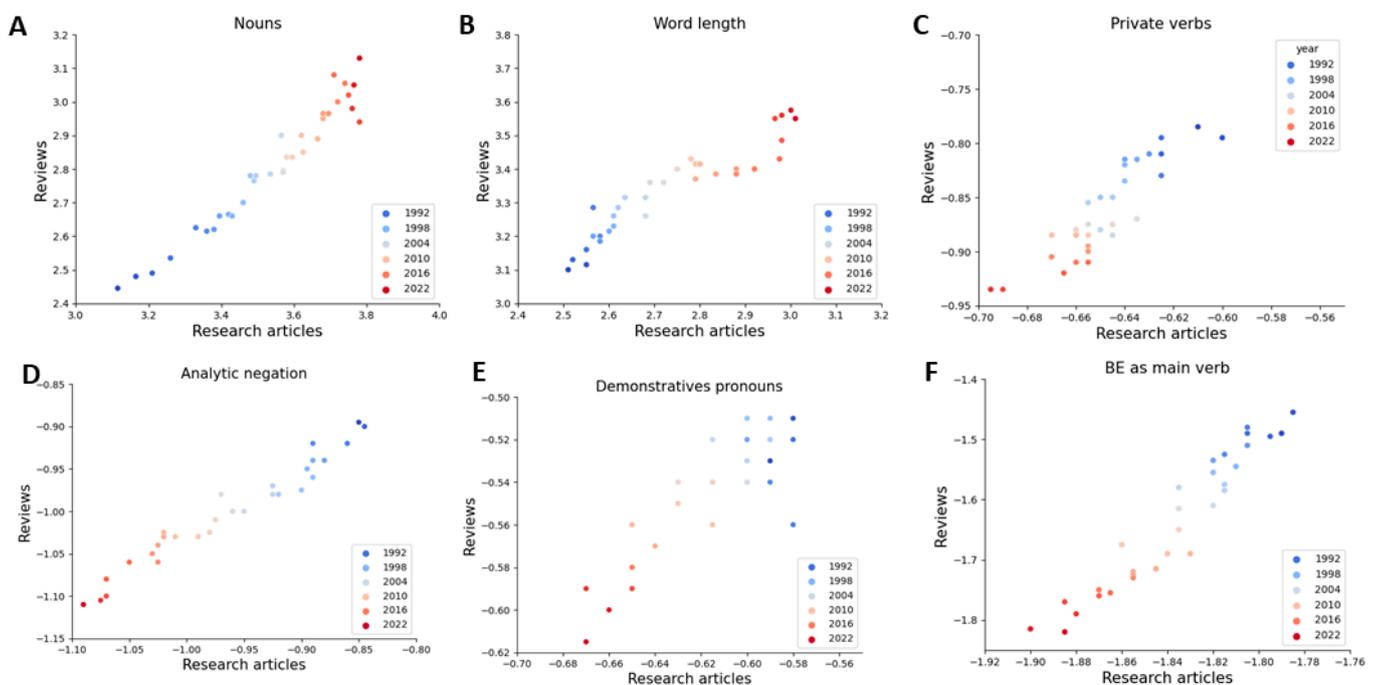


Figure 2. Scatter plots of linguistic features of Dimension 1 in RA and review corpora by publication years. These linguistic features change similarly in the 2 corpora.

A more strictly informational writing style is also evidenced by a decline in other typical features of involved texts. **Analytic negation** (Figure 2D); the use of **demonstra-**

ive pronouns (e.g., This, That, etc.), often employed with a deictic function in spoken language and interaction (Figure 2E); **private verbs** expressing internal cognitive processes (e.g., think, feel, perceive, etc.); and the use of **be as the main verb** (Figure 2F) have all experienced a decrease in frequency. Intriguingly, the use of **prepositions**, typically high in texts with strongly negative D1 scores, decreased in both corpora over time (Figure 3A), while **non-phrasal coordination**, associated with involved writing, has increased in both corpora (Figure 3B).

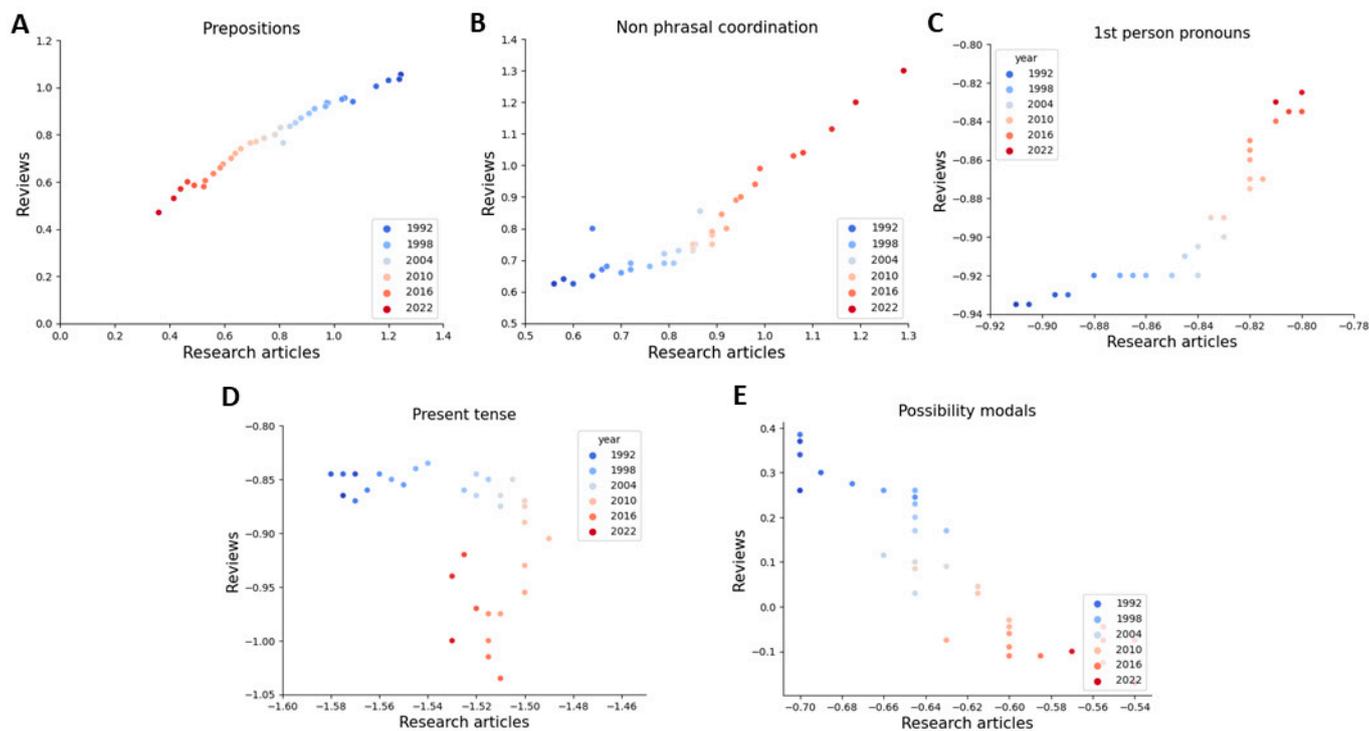


Figure 3. Scatter plots of the linguistic features of Dimension 1 in RA and review corpora by publication years. These features change differently in the 2 corpora.

However, research articles (RAs) and reviews diverged in at least three features. The frequency of **first-person pronouns**, characteristic of dialogue (and involved writing), is generally low in both corpora, aligning with the expected academic style. Yet, there are some notable exceptions, such as the following striking example:

It was my second clinical placement and I was working on a surgical ward when I was asked to accompany a patient to theatre. [29]

Admittedly, this may be an unusual style for academic prose, yet it is found in our corpus.

Notably, the frequency of first-person pronouns increased during the 1990s in RAs but remained relatively constant in reviews. Only in the early 2000s did it start to rise in both text types (Figure 3C). The most likely explanation for this behavior is that, although passive verbs have been used abundantly in academic writing as a rhetorical device to highlight the detachment of the narrator from the events contained in the text and as a sign of objective observation [30], the use of active verbs and first-person pronouns has been advocated in more recent times for the sake of clarity [31] and has been observed to be on the rise in academic writing in biology or life sciences [32]. It may be assumed that RAs were more prone to the use of first-person pronouns, as they often reported on the experimental activity of a research group, as opposed to reviews, which typically summarize the findings of other research groups, and thus this increase occurred earlier.

The use of **present tense** verbs is strongly associated with involved discourse, too (as it is very frequent in interactions between speakers), and, though generally low in

both corpora, our data indicate that their frequency Z score is higher in review papers (Figure 3D). This discrepancy may stem from the nature of reviews, which often encapsulate the current knowledge in a specific area and draw conclusions that are presented as general rules, as in the following:

Primary care clinicians treat patients with cancer and cancer pain. It is essential that physicians know how to effectively manage pain including assessment and pharmacologic and nonpharmacologic treatment modalities. [33]

In such cases, the present tense is aptly employed to convey a sense of lasting value to the conclusions drawn from the literature. Conversely, the purpose of RAs is typically to report on one or more experiments, situated in time and place, often described using past tenses, as in the following:

During 8 observation days (with time delay of 10–14 days between each observation day), all adult patients hospitalized at an internal medicine ward of 4 Belgian participating hospitals were screened for AB use. Patients receiving AB on the observation day were included in the study and screened for signs and symptoms of AAD using a period prevalence methodology. [34]

The use of present tense slightly increased in RAs in the 1990s and remained stable thereafter, while it started to decline in reviews around the same time. The exact explanation is speculative at this point, possibly related to the rise in systematic reviews or a stylistic shift. Notably, the Z score for RAs remains significantly lower than for reviews (Figure 3D). The use of **possibility modals** (i.e., verbs such as *may* or *might*) is associated with an involved style, too, as these are often utilized to express subjectivity or a guess, which is a common situation in a dialogue context. However, they can also be found quite regularly in academic writing [35], usually to express a hypothesis, as in the following:

Administration of thioredoxin may have a good potential for anti-aging and anti-stress effects. [36]

Admittedly, the room for hypotheses, although a common and actually quite essential practice in the scientific method [37], is quite limited in academic literature, given the need for evidence-grounded reasoning, hence their low frequency. Interestingly, the use of possibility modals has moved in opposite directions in the two corpora analyzed. The Z score for this feature was slightly positive in reviews, aligning with a text genre prone to drawing conclusions based on reviewed data. However, in RAs, it was negative, suggesting that assumptions and hypotheses were likely confined to few sentences in such texts. Over time, this index steadily decreased in the review group, reaching negative values in the last decade, possibly in association with the increase in systematic reviews where the extensive use of statistical tools may reduce speculation. Conversely, it increased by almost 30% in the RA corpus, possibly linked to a bolder or more personal style, as observed previously (Figure 3E) [32].

When evaluating the outcomes generated by Nini's MAT software, it is imperative to bear in mind that, while it draws inspiration from Biber's work, it might not entirely capture the nuance of Biber's original analysis. A more in-depth exploration would involve conducting a comprehensive factor analysis. This analytical process aims to delve into the intricate associations between features, weighing the contribution of each feature in each dimension. Moreover, there arises the possibility of redefining these dimensions to better align with the specific characteristics of the corpus under examination. The adoption of a fixed solution, as exemplified by the MAT software, undoubtedly streamlines the presentation of final results and increases the comparability across studies. Yet, it introduces inherent constraints concerning the generalizability of the identified dimensions and their fidelity in reflecting the distinctive attributes of the analyzed documents. However, in the face of these methodological considerations, we maintain the belief that the insights garnered from the MAT software can offer valuable perspectives on the evolutionary trends within academic articles. This assertion gains particular significance when individual Z

scores for the linguistic features in question are examined. By assessing them, it becomes possible to extract more granular insights into how specific linguistic elements contribute to the overarching trends and transformations observed in academic writing.

3.2. Dimension 2

We then proceeded to analyze the second dimension evaluated by MAT, Dimension 2 (D2), associated with narrative discourse. A positive score for D2 indicates a narrative, active, event-oriented nature, while a negative score suggests a more descriptive or static quality [11].

Our corpora of RAs and reviews have negative Z scores for D2, with reviews having a lower score than RAs (Figure 4A). This is not unexpected, as RAs more likely report, by definition, on the execution of one or more experimental procedures, which are usually associated with some sort of activity, as in the following:

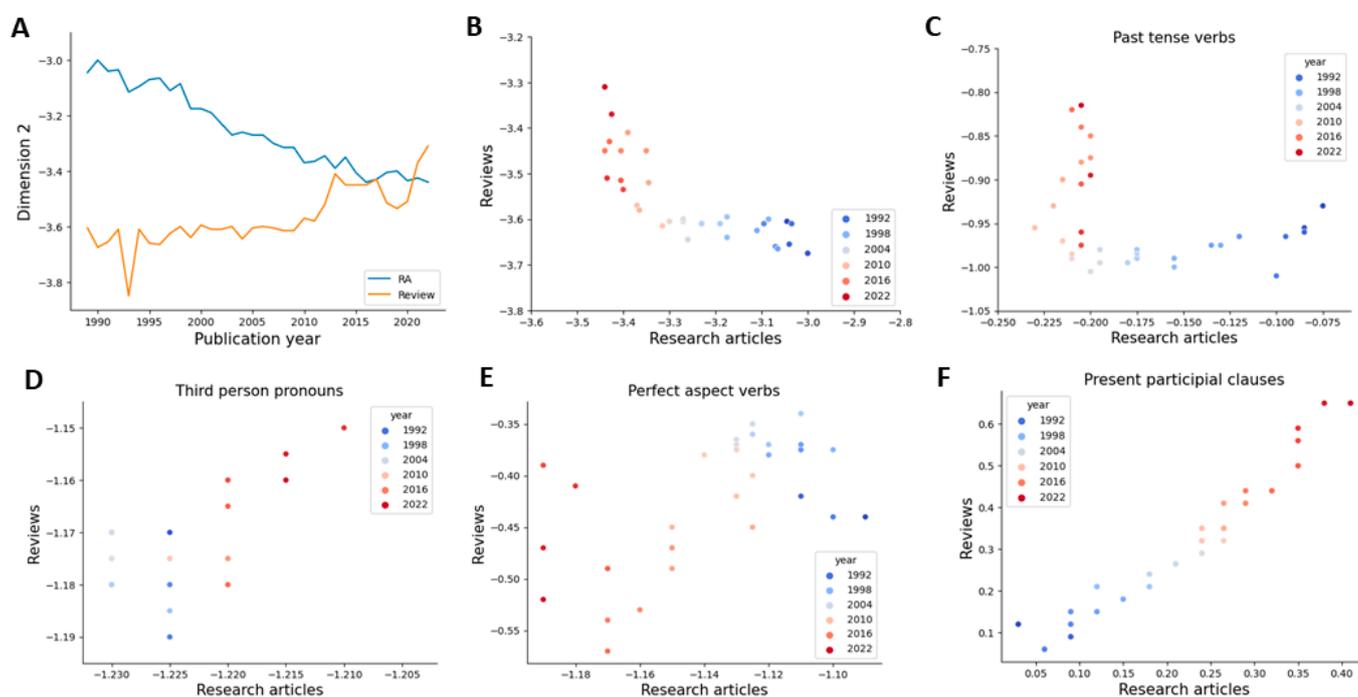


Figure 4. (A) Line plot of Dimension 2 (D2) score over the years for the research article (RA) corpus and the review corpus, in blue and orange, respectively; (B) scatter plot of D2 score for RAs and reviews; (C–F) scatter plots of the linguistic features of D2 in RA and review corpora by publication years.

We investigated expression of the five ssts in various adrenal tumors and in normal adrenal gland. Tissue was obtained from ten pheochromocytomas (PHEOs)... [38]

This passage emphasizes action, doing, selecting, analyzing, and other similar activities that require narration to navigate through them.

Interestingly, the Z score for RAs progressively decreased and became more negative over time, while the D2 score for reviews remained constant and even increased, becoming less negative in the last 5 years (Figure 4A). This trend is reflected in Figure 4B, showing a drop in RAs' D2 score in the 1990s and early 2000s, while reviews' score started to increase independently from RAs in the first decade of the 2000s. This shift might be justified by the change in the Z score for the use of **past tense verbs** [10]. This score, which was and has remained negative in both corpora for the whole timeframe (Figure 4C), decreased in RAs until the first decade of the XXI century, and it was followed by an increase in this score for review articles in the last two decades.

This means that an abstract from a review article in 1989 could more easily contain a passage like the following:

Several lines of evidence indicate that platelet-activating factor (PAF-acether) is implicated in hypersensitivity reactions. Indeed, PAF-acether reproduces the features of asthma in vivo and in vitro, since it induces bronchoconstriction, hypotension, and hemoconcentration and activates platelets and leukocytes. [39]

This passage, rich in present tense verbs, conveys general principles about a phenomenon, such as a disease or a condition. In contrast, a more recent review text might incorporate more past tenses, as in the following:

Mammalian neonates have been simultaneously described as having particularly poor memory, as evidenced by infantile amnesia, and as being particularly excellent learners. [40]

This change could suggest that since the early 2000s, review articles have tended to circumscribe their conclusions to the research papers they use as sources, contextualizing them and possibly being more cautious with generalizations.

Other important linguistic features associated with D2 underwent similar changes in both corpora: the frequency of **third-person pronouns** (i.e., he, she, they) increased for both text types (Figure 4D), as did the use of **present participial clauses** (Figure 4F), while the frequency of **perfect aspect verbs** decreased in both RAs and reviews, although the scores for this feature remained significantly lower in RAs than in reviews (Figure 4E).

Noticeably, these findings are also apparently in contrast with what we reported on the same corpora using LIWC 2022 [19]. In particular, we reported a higher **Narrativity Overall** score for reviews. That score was calculated based on the adherence to particular metrics, i.e., the three fundamental narrative curves that were measured in each abstract, namely Staging, Plot Progression, and Cognitive tension [41]. The theory behind these measures is that a narrative trajectory can be traced in a text which follows Freytag's dramatic arc: first the stage for the action is set, characters and referents are introduced and presented; the action then begins, and as the text progresses, it intensifies as the narrator describes events and activities; and cognitive tension refers to the struggles and conflicts that ensue in the story and that reach a culmination point with the resolution of the crisis that leads to the end of the narration [42]. To obtain an automated measure of these features, Pennebaker et al. decided to rely on grammatical words, which admittedly form a small set of words in English (and any language) [43]. In particular, Boyd et al. proposed measuring the frequency of articles and prepositions as proxies for the staging score, because they can be assumed to be more abundant when new referents are introduced in the text (via articles) and their relations are explained (possibly also through the use of prepositions), while auxiliary verbs and anaphoric pronouns are taken as proxy measures of plot progression, because they can be expected to be used when describing an action. Cognitive tension is measured based on the abundance of verbs in a special dictionary created ad hoc and that includes such words as 'think' or 'believe' (which would be classified as 'private verbs' in Biber's multidimensional analysis). Boyd et al. recommend splitting texts in at least five segments to monitor how these scores vary as the text progresses. It is evident that LIWC 2022 and MAT scores rely on distinct features. Readers should focus on understanding the specific characteristics of the text that these tools measure, rather than becoming fixated on the 'narrativity' label.

3.3. Dimension 3

A positive score for Dimension 3 (D3) is associated with explicit and context-independent references, as opposed to the negative pole of this dimension, i.e., nonspecific, context-dependent content [10]. This means that referents in the text are mentioned and described explicitly, so that there cannot be any doubt about their identity. According to our data, reviews have a higher D3 score than RAs, and both their scores have been progressively increasing over time (Figure 5A,B). Among the features that affect D3, **nominalization** appears to have followed this trend and may be responsible for the visible changes in D3 over time.

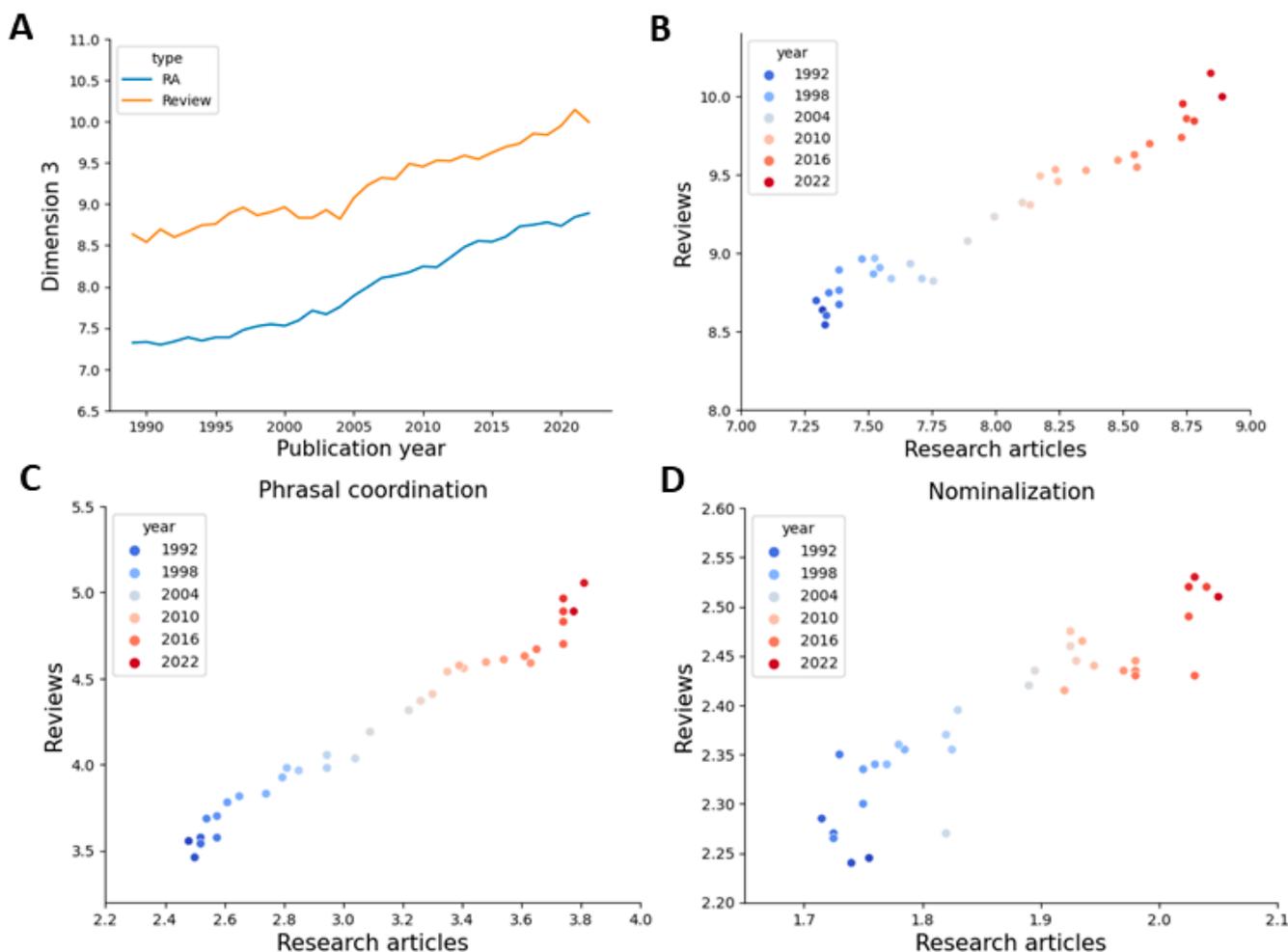


Figure 5. (A) Line plot of Dimension 3 (D3) score over the years for the research article (RA) corpus and the review corpus, in blue and orange, respectively; (B) scatter plot of D3 score for RAs and reviews; (C,D) scatter plots of the linguistic features of D3 in RA and review corpora by publication years.

Nominalization [44] indicates the replacement of a verb with a noun that denotes the same action and is a common feature of technical language [45], which is often used to convey a more impersonal tone, because a noun, by describing an action as an entity, detaches it from the agent and confers it a higher independence [46]. The use of nominalization, albeit often deemed undesirable [47], has been growing in academic writing [48]. An example of nominalization in our corpus could be the following:

Pancreatic cancer (PC) is characterized by high tumor invasiveness, distant metastasis, and insensitivity to traditional chemotherapeutic drugs. . . [49]

Phrasal coordination is also positively associated with D3, as it may be associated with a higher degree of descriptivity and more thorough explanation of textual referents and, similarly to nominalization, displays a similar trend. An example of phrasal coordination in a manuscript with a high score for this feature is the following:

. . . the specific mechanisms are blurry, especially the involved immunological pathways, and the roles of beneficial flora have usually been ignored. [49]

3.4. Dimension 4

Dimension 4 is associated with overt expression (positive pole) or non-overt expression (negative pole) of persuasion [11], not only referring to the writer's opinion, but also the

quality of texts to prompt readers toward a certain course of action. Both our corpora have a negative score (Figure 6A), which indicates that both Ras and reviews from our corpus tend to be non-persuasive, which is in line with the declared function of biomedical literature, as previously stated elsewhere [28]. Unsurprisingly reviews tend to be less negative than Ras in regard to D3 score. This is easily explained by the fact that reviews, by nature, provide readers with an overview of facts and knowledge that can be used to trace recommendations or guidelines. However, the D3 score changed over time, and while Ras have been mostly stable over the years, displaying a slight trend for D3 to increase by about 10% over the course of the last 30 years, reviews have further decreased this score by the same amount in the last decade (Figure 6B), signaling a movement toward a more impartial stance in review papers. Among the factors that may have affected these changes, the use of **infinitives** has been increasing in both corpora in a similar way (Figure 6C), such as in the following:

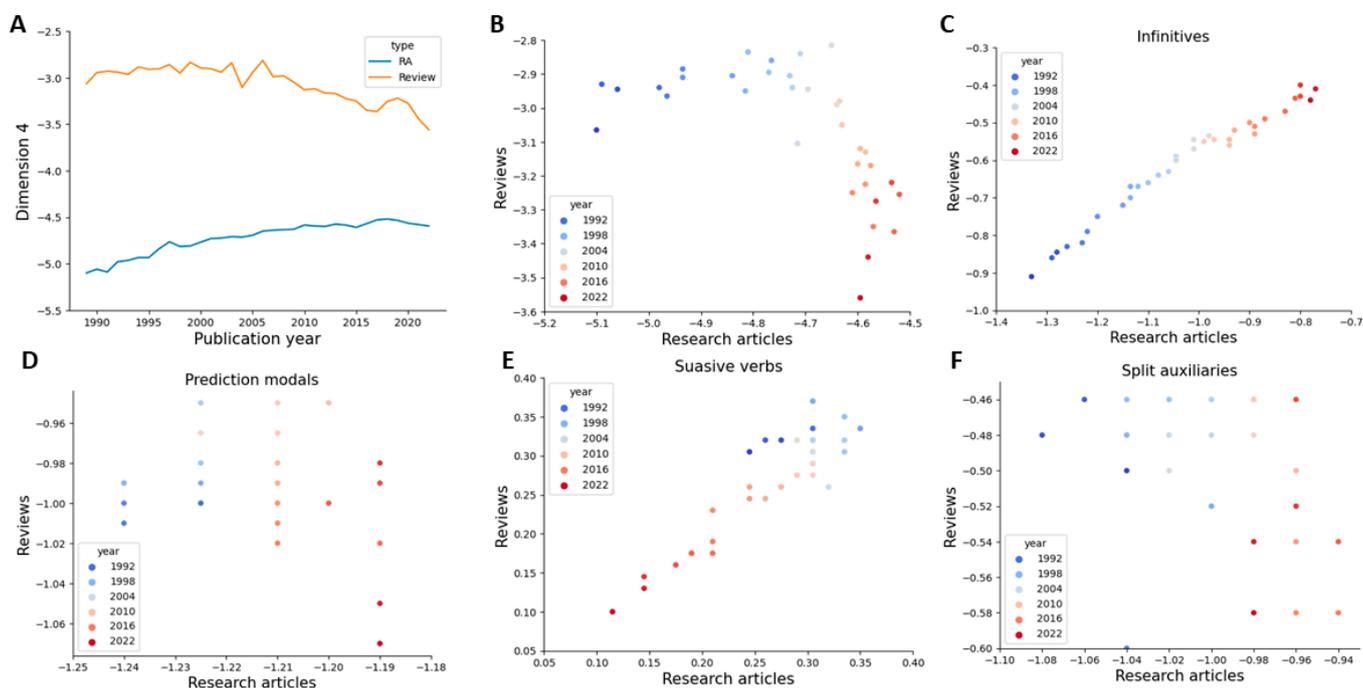


Figure 6. (A) Line plot of Dimension 4 (D4) score over the years for the research article (RA) corpus and the review corpus, in blue and orange, respectively; (B) scatter plot of D4 score for RAs and reviews; (C–F) scatter plots of the linguistic features of D4 in RA and review corpora by publication years.

Understanding the age-dependent neuromuscular mechanisms underlying force reductions ... allows researchers to investigate new interventions to mitigate these reductions. [50]

Suasive verbs are, understandably, another hallmark of overt persuasion, as in the following:

...an ad hoc committee of the American Venous Forum, working with an international liaison committee, has recommended a number of practical changes. [51]

Their frequency, quite similar in both manuscript types, has, however, been decreasing steadily over the years (Figure 6E), which is consistent with that more neutral stance we mentioned above. However, **prediction modals**, which have quite a high bearing on this dimension, despite displaying quite a high variability in our corpora, have mostly changed for RAs (Figure 6D), and a slight increase can be observed. Meanwhile, the use of **split auxiliaries** has changed for reviews only in the last decade (Figure 6F).

Prediction modals include forms like will, should, or must, which indicate the future directions that research or practice should take, as in the following:

*The data suggest that treatment of *H. pylori* infection should be considered in children with concomitant GERD. [52]*

3.5. Dimension 5

Dimension 5 refers to the abstract (positive pole) or non-abstract (negative pole) nature of the information contained in the texts [11]. As already reported, academic texts, including those from the biomedical field, tend to have high scores for D5, as they tend to contain technical, abstract concepts.

In our corpora, review papers score higher than RAs regardless of the publication date (Figure 7A). Although the D5 score decreased for both text types over the years, the gap between the two groups vanished by the mid-second decade of the 2000s (Figure 7A). In the last 5 years, the D5 score appeared to increase again in reviews only (Figure 7B). The frequent use of **passives** is a hallmark of abstract style, as it typically mitigates the action of an agent (even more so if the passive is **agentless**). These two indices—passives with a “by” agent and agentless passives—have been decreasing in both text types (Figure 7D,E), presumably driving the trend of the overall D5 score. The use of **conjuncts**, however, has increased both in reviews and RAs, and this increase has been quite sudden in the last 5 years for reviews, which might explain the surge in D5 score in that timeframe.

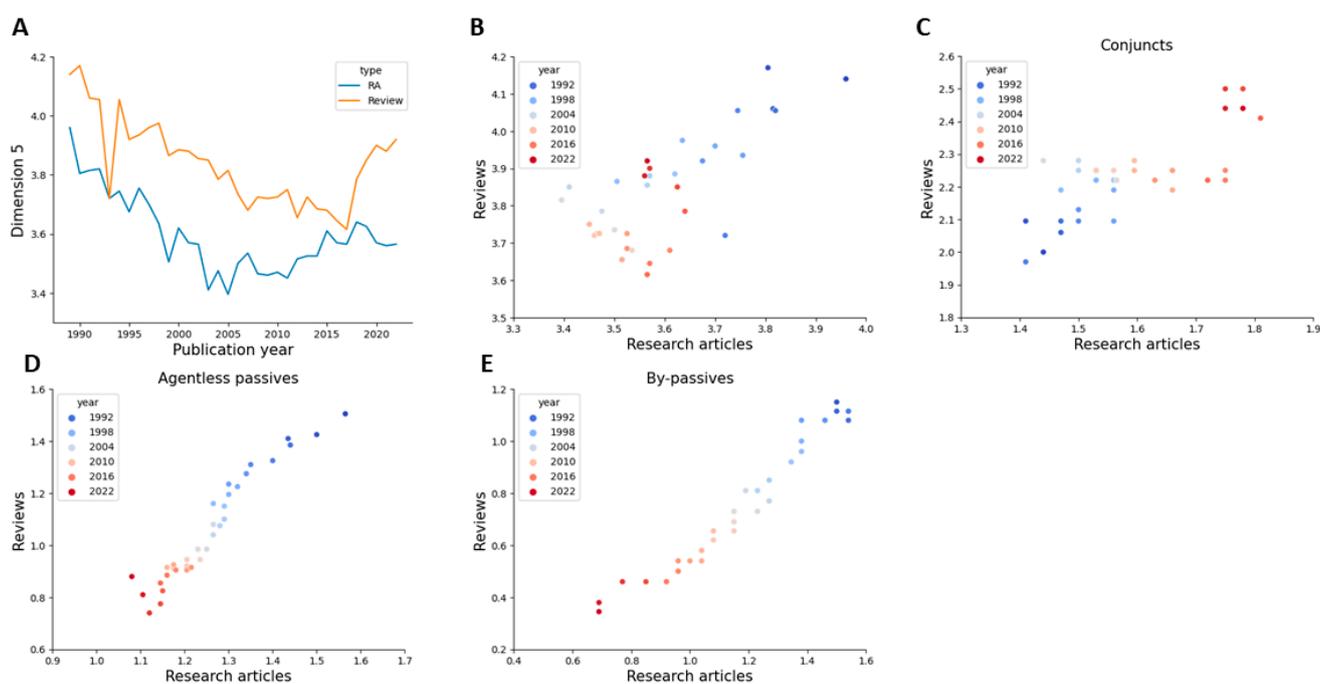


Figure 7. (A) Line plot of Dimension 5 (D5) score over the years for the research article (RA) corpus and the review corpus, in blue and orange, respectively; (B) scatter plot of D5 score for RAs and reviews; (C–E) scatter plots of the linguistic features of D5 in RA and review corpora by publication years.

4. Conclusions

In conclusion, the analysis of over 1.2 million biomedical literature abstracts published in MEDLINE over the last 30 years reveals several noteworthy trends. The consolidation of an informational tone (D1) is observed in both research articles (RAs) and reviews. This is accompanied by a decrease in the use of narrative devices (D2), with this change being more pronounced in the RA corpus. Simultaneously, there is a parallel increase in context-independent stances (D3) in both RAs and reviews. The relative lack of overt

persuasion (D4) in the examined academic texts has remained relatively stable over the years. Additionally, there is a decrease in the degree of abstractness, coinciding with a decline in the use of passive voice constructions. When comparing RAs to reviews, it becomes apparent that RAs used to rely more heavily on narration than reviews. However, RAs have toned down the use of this stylistic device to a level similar to that of reviews. On the other hand, reviews, as a manuscript type, historically exhibited a higher degree of content-independency, overt persuasion, and abstractness. These characteristics have been maintained over the years. This comprehensive multidimensional analysis provides valuable insights into the evolving linguistic and rhetorical characteristics of biomedical literature abstracts, shedding light on how different dimensions have changed over time and distinguishing patterns between RAs and reviews.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/publications12010002/s1>, Figure S1: Scatter plots of additional D1 features.

Author Contributions: Conceptualization, C.G. and S.G.; methodology, C.G. and M.T.C.; formal analysis, C.G.; resources, S.G. and P.M.; writing—original draft preparation, C.G.; writing—review and editing, P.M.; visualization, C.G.; supervision, S.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data are available on request.

Acknowledgments: The authors would like to thank Silvana Belletti for her advice on corpora.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

The following is the list of linguistic features that Biber followed for his multidimensional analysis, after factor analysis, grouped by dimension and sorted by factor loading, from the highest, modified from [11].

Dimension 1: Involved versus informational production

Positive features (involved production)

Private verbs
that-deletions
 Contractions
 Present tense verbs
do as pro-verb
 Analytic negation
 Demonstrative pronouns
 General emphatics
 First-person pronouns
 Pronoun *it*
 Causative subordination
 Discourse particles
 Indefinite pronouns
 General hedges
 Amplifiers
 Sentence relatives
wh- questions
 Possibility modals
 Nonphrasal coordination
wh- clauses
 Final prepositions

Negative features (informational production)

Nouns
 Word length
 Prepositions
 Type/token ration
 Attributive adjectives

Dimension 2: Narrative versus nonnarrative discourse

Positive features (narrative discourse)

Past tense verbs
 Third-person pronouns
 Perfect aspect verbs
 Public verbs
 Synthetic negation
 Present participial clauses

Dimension 3: Situation-dependent versus elaborated reference

Positive features (situation-dependent reference)

Time adverbials
 Place adverbials
 Adverbs

Negative features (elaborated reference)

wh- relative clauses in object positions
 Pied piping constructions
wh- relative clauses in subject positions
 Phrasal coordination
 Nominalizations

Dimension 4: Overt expression of persuasion

Positive features (overt expression of persuasion)

Infinitives
 Prediction modals
 Suasive verbs
 Conditional subordination
 Necessity modals
 Split auxiliaries
 (Possibility modals)

Dimension 5: Nonimpersonal versus impersonal style

Negative features (impersonal style)

Conjuncts
 Agentless passives
 Past participial adverbial clauses
By passives
 Past participial postnominal clauses
 Other adverbial subordinators

References

1. Narin, F.; Pinski, G.; Gee, H.H. Structure of the Biomedical Literature. *J. Am. Soc. Inf. Sci.* **1976**, *27*, 25–45. [[CrossRef](#)]
2. Cartabellotta, A.; Montalto, G.; Notarbartolo, A. Evidence-Based Medicine. How to Use Biomedical Literature to Solve Clinical Problems. Italian Group on Evidence-Based Medicine-GIMBE. *Minerva Med.* **1998**, *89*, 105–115. [[PubMed](#)]
3. Hrynaszkiewicz, I. The Need and Drive for Open Data in Biomedical Publishing. *Serials* **2011**, *24*, 31–37. [[CrossRef](#)] [[PubMed](#)]
4. Sanberg, P.R.; Gharib, M.; Harker, P.T.; Kaler, E.W.; Marchase, R.B.; Sands, T.D.; Arshadi, N.; Sarkar, S. Changing the Academic Culture: Valuing Patents and Commercialization toward Tenure and Career Advancement. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 6542–6547. [[CrossRef](#)] [[PubMed](#)]
5. Rice, D.B.; Raffoul, H.; Ioannidis, J.P.A.; Moher, D. Academic Criteria for Promotion and Tenure in Biomedical Sciences Faculties: Cross Sectional Analysis of International Sample of Universities. *BMJ* **2020**, *369*, m2081. [[CrossRef](#)] [[PubMed](#)]
6. Landhuis, E. Scientific Literature: Information Overload. *Nature* **2016**, *535*, 457–458. [[CrossRef](#)] [[PubMed](#)]
7. Sharma, H.; Verma, S. Predatory Journals: The Rise of Worthless Biomedical Science. *J. Postgrad. Med.* **2018**, *64*, 226. [[CrossRef](#)]

8. Ioannidis, J.P.A. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *Milbank Q.* **2016**, *94*, 485–514. [[CrossRef](#)]
9. Pieper, D.; Antoine, S.-L.; Mathes, T.; Neugebauer, E.A.M.; Eikermann, M. Systematic Review Finds Overlapping Reviews Were Not Mentioned in Every Other Overview. *J. Clin. Epidemiol.* **2014**, *67*, 368–375. [[CrossRef](#)]
10. Biber, D. On the Complexity of Discourse Complexity: A Multidimensional Analysis. *Discourse Process.* **1992**, *15*, 133–163. [[CrossRef](#)]
11. Biber, D. *Variation across Speech and Writing*; Cambridge University Press: Cambridge, UK, 1991; ISBN 0521425565.
12. Stig, J.; Leech, G.N.; Goodluck, H. *Manual of Information to Accompany the Lancaster-Oslo: Bergen Corpus of British English, for Use with Digital Computers*; Department of English, University of Oslo: Oslo, Norway, 1978.
13. Pöldvere, N.; Johansson, V.; Paradis, C. On the London–Lund Corpus 2: Design, Challenges and Innovations. *Engl. Lang. Linguist.* **2021**, *25*, 459–483. [[CrossRef](#)]
14. Biber, D.; Conrad, S.; Reppen, R.; Byrd, P.; Helt, M. Speaking and Writing in the University: A Multidimensional Comparison. *TESOL Q.* **2002**, *36*, 9–48. [[CrossRef](#)]
15. Friginal, E.; Mustafa, S.S. A Comparison of US-Based and Iraqi English Research Article Abstracts Using Corpora. *J. Engl. Acad. Purp.* **2017**, *25*, 45–57. [[CrossRef](#)]
16. Cao, Y.; Xiao, R. A Multi-Dimensional Contrastive Study of English Abstracts by Native and Non-Native Writers. *Corpora* **2013**, *8*, 209–234. [[CrossRef](#)]
17. Nini, A. The Multi-Dimensional Analysis Tagger. In *Multi-Dimensional Analysis: Research Methods and Current Issues*; Bloomsbury Academic: New York, NY, USA, 2019; pp. 67–94.
18. Tausczik, Y.R.; Pennebaker, J.W. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *J. Lang. Soc. Psychol.* **2010**, *29*, 24–54. [[CrossRef](#)]
19. Guizzardi, S.; Colangelo, M.T.; Mirandola, P.; Galli, C. The Evolution of Narrativity in Abstracts of the Biomedical Literature between 1989 and 2022. *Publications* **2023**, *11*, 26. [[CrossRef](#)]
20. Kluyver, T.; Ragan-Kelley, B.; Pérez, F.; Granger, B.; Bussonnier, M.; Frederic, J.; Kelley, K.; Hamrick, J.; Grout, J.; Corlay, S.; et al. Jupyter Notebooks—A Publishing Format for Reproducible Computational Workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas, Proceedings of the 20th International Conference on Electronic Publishing, ELPUB 2016, Göttingen, Germany, 7–9 June 2016*; IOS Press BV: Amsterdam, The Netherlands, 2016; pp. 87–90.
21. Shapiro, A. Litter-Getter. Available online: https://github.com/shapiromatron/litter_getter (accessed on 1 June 2022).
22. Greenhalgh, T.; Thorne, S.; Malterud, K. Time to Challenge the Spurious Hierarchy of Systematic over Narrative Reviews? *Eur. J. Clin. Invest.* **2018**, *48*, e12931. [[CrossRef](#)]
23. McKinney, W. Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010*; pp. 51–56.
24. Richardson, L. Beautiful Soup Documentation. Available online: <https://pypi.org/project/beautifulsoup4/> (accessed on 1 June 2022).
25. Curtis, A.; Smith, T.; Ziganshin, B.; Eleftheriades, J. The Mystery of the Z-Score. *AORTA* **2016**, *4*, 124–130. [[CrossRef](#)]
26. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [[CrossRef](#)]
27. Waskom, M. Seaborn: Statistical Data Visualization. *J. Open Source Softw.* **2021**, *6*, 3021. [[CrossRef](#)]
28. Liu, J.; Xiao, L. A Multi-Dimensional Analysis of Conclusions in Research Articles: Variation across Disciplines. *Engl. Specif. Purp.* **2022**, *67*, 46–61. [[CrossRef](#)]
29. Andrews, R. My Cheerful Attitude Upset an Anxious Pre-Op Patient. *Nurs. Stand.* **2009**, *24*, 27–28. [[CrossRef](#)]
30. Hyland, K. Authority and Invisibility. *J. Pragmat.* **2002**, *34*, 1091–1112. [[CrossRef](#)]
31. Hyland, K. Options of Identity in Academic Writing. *ELT J.* **2002**, *56*, 351–358. [[CrossRef](#)]
32. Hyland, K.; Jiang, F. Is Academic Writing Becoming More Informal? *Engl. Specif. Purp.* **2017**, *45*, 40–51. [[CrossRef](#)]
33. Pathak, S.K.; Salunke, A.A.; Chawla, J.S.; Sharma, A.; Ratna, H.V.K.; Gautam, R.K. Bilateral Radial Head Fracture Secondary to Weighted Push-Up Exercise: Case Report and Review of Literature of a Rare Injury. *Indian. J. Orthop.* **2021**, *56*, 162–167. [[CrossRef](#)]
34. Elseviers, M.M.; Van Camp, Y.; Nayaert, S.; Duré, K.; Annemans, L.; Tanghe, A.; Vermeersch, S. Prevalence and Management of Antibiotic Associated Diarrhea in General Hospitals. *BMC Infect. Dis.* **2015**, *15*, 129. [[CrossRef](#)]
35. Carrió Pastor, M. Cross-Cultural Variation in the Use of Modal Verbs in Academic English. *SKY J. Linguist.* **2014**, *27*, 153–166.
36. Nakamura, H. Experimental and Clinical Aspects of Oxidative Stress and Redox Regulation. *Rinsho Byori* **2003**, *51*, 109–114.
37. Harris, E.E. *Hypothesis and Perception: The Roots of Scientific Method*; Routledge: London, UK, 2014; ISBN 1317851609.
38. Ueberberg, B.; Tourne, H.; Redman, A.; Walz, M.K.; Schmid, K.W.; Mann, K.; Petersenn, S. Differential Expression of the Human Somatostatin Receptor Subtypes Sst1 to Sst5 in Various Adrenal Tumors and Normal Adrenal Gland. *Horm. Metab. Res.* **2005**, *37*, 722–728. [[CrossRef](#)]
39. Pretolani, M.; Lellouch-Tubiana, A.; Lefort, J.; Bachelet, M.; Vargaftig, B.B. PAF-Acether and Experimental Anaphylaxis as a Model for Asthma. *Int. Arch. Allergy Immunol.* **1989**, *88*, 149–153. [[CrossRef](#)] [[PubMed](#)]
40. Wilson, D.A.; Sullivan, R.M. Neurobiology of Associative Learning in the Neonate: Early Olfactory Learning. *Behav. Neural Biol.* **1994**, *61*, 1–18. [[CrossRef](#)] [[PubMed](#)]
41. Boyd, R.L.; Blackburn, K.G.; Pennebaker, J.W. The Narrative Arc: Revealing Core Narrative Structures through Text Analysis. *Sci. Adv.* **2020**, *6*, eaba2196. [[CrossRef](#)] [[PubMed](#)]

42. Freytag, G. *Freytag's Technique of the Drama*; Scott Foresman: Northbrook, IL, USA, 1894.
43. Corver, N.; van Riemsdijk, H. *Semi-Lexical Categories: The Function of Content Words and the Content of Function Words*; Walter de Gruyter: Berlin, Germany, 2013; Volume 59, ISBN 3110874008.
44. Alexiadou, A. Nominalizations: A Probe into the Architecture of Grammar Part I: The Nominalization Puzzle. *Lang. Linguist. Compass* **2010**, *4*, 496–511. [[CrossRef](#)]
45. Khamesian, M. On Nominalization, A Rhetorical Device in Academic Writing. *Armen. Folia Angl.* **2015**, *11*, 42–48. [[CrossRef](#)]
46. Baratta, A.M. Nominalization Development across an Undergraduate Academic Degree Program. *J. Pragmat.* **2010**, *42*, 1017–1036. [[CrossRef](#)]
47. Biber, D.; Gray, B. Challenging Stereotypes about Academic Writing: Complexity, Elaboration, Explicitness. *J. Engl. Acad. Purp.* **2010**, *9*, 2–20. [[CrossRef](#)]
48. Biber, D.; Gray, B. Nominalizing the Verb Phrase in Academic Science Writing. In *The Register-Functional Approach to Grammatical Complexity*; Routledge: London, UK, 2021; pp. 176–198.
49. Wei, X.; Mei, C.; Li, X.; Xie, Y. The Unique Microbiome and Immunity in Pancreatic Cancer. *Pancreas* **2021**, *50*, 119–129. [[CrossRef](#)]
50. Orsatto, L.B.d.R.; Wiest, M.J.; Diefenthaler, F. Neural and Musculotendinous Mechanisms Underpinning Age-Related Force Reductions. *Mech. Ageing Dev.* **2018**, *175*, 17–23. [[CrossRef](#)]
51. Eklöf, B.; Rutherford, R.B.; Bergan, J.J.; Carpentier, P.H.; Gloviczki, P.; Kistner, R.L.; Meissner, M.H.; Moneta, G.L.; Myers, K.; Padberg, F.T.; et al. Revision of the CEAP Classification for Chronic Venous Disorders: Consensus Statement. *J. Vasc. Surg.* **2004**, *40*, 1248–1252. [[CrossRef](#)]
52. Pollet, S.; Gottrand, F.; Vincent, P.; Kalach, N.; Michaud, L.; Guimber, D.; Turck, D. Gastroesophageal Reflux Disease and Helicobacter Pylori Infection in Neurologically Impaired Children: Inter-Relations and Therapeutic Implications. *J. Pediatr. Gastroenterol. Nutr.* **2004**, *38*, 70–74. [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.