International Journal of English Linguistics; Vol. 14, No. 5; 2024
ISSN 1923-869X E-ISSN 1923-8703
Published by Canadian Center of Science and Education

# The Theme Progression Patterns in Popular Science Book Writing: A Systemic Functional Linguistics Approach

Ahmad Qassim M Al Darwesh[1]

[1] College of Language Sciences, Department of English, King Saud University, Riyadh, Saudi Arabia.

Correspondence: Ahmad Q. Al Darwesh. EGAB7075 12B-5023, 32613 Tarut, Saudi Arabia.

Received: May 13, 2024 Accepted: July 15, 2024 Online Published: August 14, 2024

doi:10.5539/ijel.v14n5p35 URL: https://doi.org/10.5539/ijel.v14n5p35

**Abstract**

Despite the proliferation of digital mediums such as documentary series, video essays, and science podcasts, popular science books are still the primary medium for promoting science to the public as an epistemic way to understanding and being aware of our natural world. That said, in the domain of systemic functional linguistics, there exists a dearth of studies investigating popular science books. Hence, this study aims to investigate the organization of popular science discourse from the perspective of Theme Progression. A 93,078-word corpus was collected and divided into two main science categories: hard and soft—three disciplines under each main category and six texts under each discipline. The analysis of the corpus followed a mixed-method design where a Theme-counting excel sheet, created by the researcher, was used to calculate the most occurring Theme patterns. The results of the analysis indicate that the hard science disciplines give a logical presentation of the scientific text that intends to unpack and explain the intricacy behind the scientific notion whereas the soft science disciplines focus on expository narrative to connect and relate the scientific text to the target reader. Given these writing behaviors between the two science categories, it is recommended that future SFL studies explore the potentially arising differences within the broader aspect of the popular science genre, for instance, to juxtapose books from articles in the way discoursal features (e.g., coherence and cohesion) are structured in the text.

**Keywords:** popular science genre, systemic functional linguistics, theme progression, writing skills

## 1. Introduction

In the digital age, various mediums such as documentary series, video essays, and podcasts proliferate scientific information with the intent to inform the public about the importance of science (Dijkstra et al., 2020). However, the main source of popularizing science is still popular science books (PSB), which have untangled many complex scientific discussions in fields such as physics, biology, and psychology. Long before the Internet, PSBs have abridged the scientific community with the public and have fostered an appreciative relationship between the two entities.

In higher education, popular science writing (PSW) which includes online articles from scientific magazines and newspapers has been used as teaching materials as treatment for writing instruction. Reportedly, these popularizing scientific materials, delivered in the form of tasks, had helped students at the thesis stage with organizing the content, controlling the degree of abstraction in information, structuring aspects of the text, as well as the scientific writing style (Pelger, 2018). Wu et al. (2018) emphasizes the value of learning to popularize science in writing, as it develops clarity and coherence of the text, in as much as the writer's disposition and stance.

Despite the educational and general-knowledge benefits, as far as the literature of Systemic Functional Linguistics (SFL), there lies a scarcity of interest in the structural organization of PSW discourse that is intently targeted to inform and persuade the non-specialist readers. Only in few investigations was the PSW discourse delved into such as Nwogu and Bloor (2011).

Therefore, investigating the Theme Progression (TP) patterns within a multiplicity of popularized disciplines shall provide further insight into the textual status of the main form of popular science communication. This study aims to explore how PSB writers organize the structure of their discourse (i.e., the type of TP patterns used) in an amassed corpus extracted from various PSBs.

## 2. Literature Review

### *2.1 Theoretical Framework*

In the SFL tradition, three metafunctions construe the variety of language registers: Ideational, Interpersonal, and Textual. Theme Progression (TP) is couched under the Textual metafunction. It is concerned with the language bits and units that organize the utterance—or message—in a spoken or written clause. These bits and units of language—also referred to as textual resources—shape the discourse in line with the variability of the context under which language is used (Halliday et al., 2014).

Halliday (2014) categorizes Theme and Rheme as two textual components that organize the clause. The starting point of Theme is signaled at the departure of the information that mediates the interpretation of the text. Rheme is the remainder of the structural domain of the Theme. Thus, the prominence of the message always comes in the first segment of the clause. The textual system of Theme and Rheme could be semantically related to the system of Information that Halliday (2014) parallelly introduces. The system of Information consists of two interwoven grammatical units: New and Given, which are parallel to Theme and Rheme. Similar to Theme, the beginning of the clause is referenced as New, which marks the departure of the message. What follows this is Given, the parallel to Rheme, which signifies the remainder of the clause that elaborates the context of the message. Though the system of Theme and the system of Information appear to have close textual ties in terms of structure, the former textual system is construed from the perspective of the speaker/writer, as the departure of the message is predetermined by them while the latter textual system, from the perspective of the listener/reader.

In a lengthy written text such as an excerpt of a book chapter, the Theme-Rheme distribution is analyzed on the clausal level, establishing a web of Thematic elements cohesively mapped out throughout the text. This allows to underscore a precisive pattern about the organizational make-up of the text.

Theme Progression initially consisted of three patterns. But the existing literature generally highlights five patterns (Alyousef, 2016, 2021; Carter-Thomas, 1999; Nwogu & Bloor, 2011; Van Thang & Tam, 2020):

(1) Reiteration: Repetition of Theme keeps the text focused by constantly inserting the Theme at the beginning of the clause, thereby effectively creating cohesion.

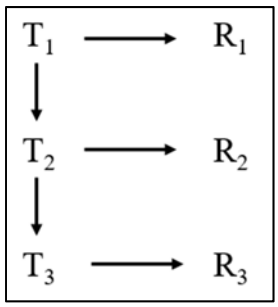


Figure 1. Theme Reiteration

(2) Linear Progression: Also referred to as “Sequential pattern” (Alyousef, 2016), the Theme in the original clause becomes Rheme in the proceeding one, then back to becoming a Rheme in the third clause. In the fourth clause, the Rheme is elevated to Theme from the Rheme of the third clause. This consistent back-and-forth pattern ensures a logical flow and cohesion in the presentation of the information.

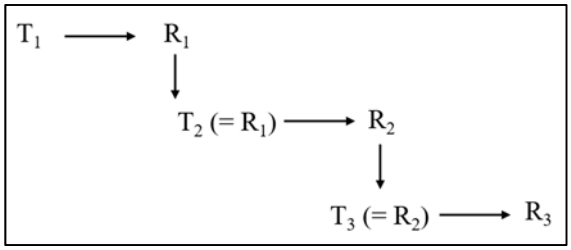


Figure 2. Linear Progression

(3) Zigzag Pattern: A sister to the Linear pattern. Though they have the same cohesion function, the difference between them is that the Rheme introduced in a preceding clause is promoted as the Theme in the subsequent clause. The cohesion of the text is thus slightly differently achieved when this pattern is perpetuated throughout the text (Marfuaty & Wahyudi, 2015).

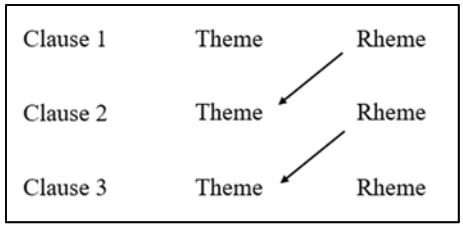


Figure 3. Zigzag Pattern

(4) Multiple-Rheme Pattern: Also referred to as "Fan pattern", a number of pieces of information is introduced as Rheme. From here, the Themes in the subsequent clauses become an extension of the original Rheme that lent the split of the Themes.

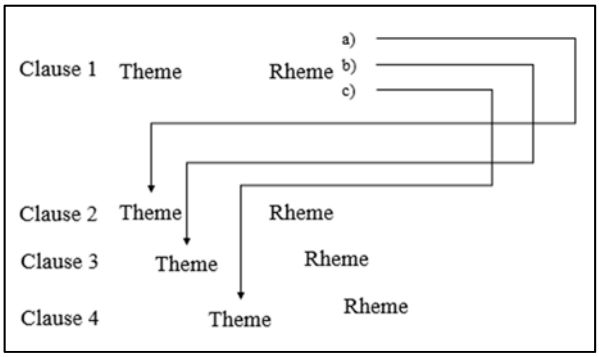


Figure 4. Multiple-Rheme Pattern

(5) Derived Themes: The Theme in the original clause diverges into multiple Themes inside or outside the paragraph. Each derived Theme can be mediated by "various special (mostly extralinguistic) usage of the presentation of subject-matter" (Nwogu & Bloor, 2011).

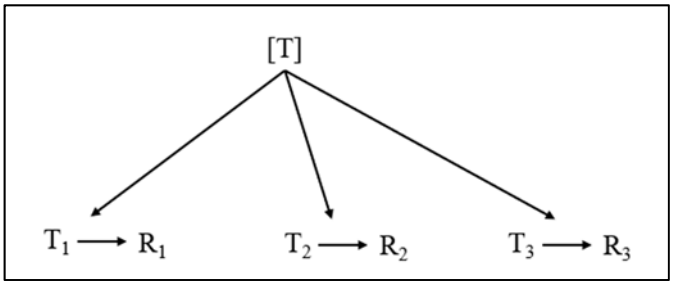


Figure 5. Derived Themes

*2.2. Discourse features in popular science writing*

Presently, Theme Progression (TP) is severely under-investigated with respect to PSW; the handful few found in the SFL literature are by Nwogu and Bloor (2011) and Sharndama and Nwogu (2013). Nwogu and Bloor (2011) annotated a 75,000-word corpus of research papers and popular texts from *New Scientist*, *The Times* and

*Newsweek*. Though the technical and popular texts shared the Linear pattern, which occurred most frequently in their analysis, popularized texts included two more TP patterns: Reiteration and Derived Theme. Nwogu and Bloor attribute the wider variety of Themes in popularized texts as opposed to their technical ones due to the interrelation of the communicative purpose with the target audience. Coinciding with these findings, Sharndama and Nwogu (2013) also found a similar TP pattern projection from annotating ten popular and another ten professional articles in the field of law. Both Reiteration and Derived Theme permeated the popularized texts because the authors repurposed the information of the professional articles to fit a lay audience.

When comparing the TP analyses of PSW to scientific writing, nearly all the cited studies consistently observe that Reiteration and Linear predominate the scientific text, and the reasoning behind this is couched in the nature of scientific writing itself (Alvin, 2015; Alyousef, 2020, 2021; Alyousef & Alsharif, 2020; Babaii et al., 2016; Darmila et al., 2019; Gunawan & Aziza, 2017; Leong et al., 2018; Susilowati et al., 2022; Wei & Yi, 2020). Scientific writing leans favorably to establishing and supporting claims, assumptions and arguments. Structuring an absorbable flow of information and constructing argumentation that negotiates and communicates with a subset of expert readers coincides with the TP choices found in scientific papers. Contrary to this, PSW uses more diverse TP patterns to expound upon the information presented to the general reader.

That said, the textual investigation of PSW makes no distinction between popular articles and popular books. The different types of written medium in PSW currently stand as a blanket genre for articles, books and the like. Studies within this scope have unearthed genre features specific to the discourse of popular science. For instance, the writer's tone is one central discoursal feature to engage with public audiences. The writer's tone could have irony, sarcasm, and/or humor where appropriate. Alternatively, the writer's tone could resemble that of the academic circles—should the need arise. The writer's personality emerges from modulating the tone as the discourse progresses (Kucharski, 2018). Tone and narrative simultaneously construct a popular text as both discoursal aspects shed the writer's personality to the reader. Though a narrative in its fully literary form is realized through dialogue and the protagonist-antagonist parallel—as seen in fiction works like novels and short stories, its adapted form in PSW is stripped from most of these archetypal storytelling elements and is reserved to its functional role in the genre, which is to form trust between the writer and the readers and captivate their attention, as well as to allow them to relate the scientific information to their personal life (Paniukova, 2019; Pilkington, 2016) .

Another discourse feature deals with centralizing the theme of the discussion. This is not to confuse *Theme* and R*heme* from the SFL tradition with the one in Uyun's (2016) study. In the study, theme is defined as the main thesis in the popular science book *The Grand Design*, a book co-authored by the physicist Stephen Hawkins and the cosmologist Leonard Mlodinow, which discusses the history of the universe and examines the theory behind it. To identify the main theme of the book, Uyun placed the word *quantum* and *cosmology* against each other and computed the number of occurrences of each word. The word *quantum* occurred significantly more than *cosmology* and is placed in various topics of the book such as *electrodynamics and chromodynamics*. The constant repetition of *quantum*, suggests the researcher, is a sign of the authors' belief about the nature of the universe, that its make-up is in fact "quantum". This interpretation solely relies on the number of occurrences of *quantum* (140 times) yet glosses over the key observations of its own findings. Nevertheless, this study proves slightly useful in understanding the web of themes that exemplifies the overarching discourse in PSBs.

Most specific to PSW—besides tone, narrative, and thematic wording—are features that highlight research findings and place the reader in the center of the presentation. The credibility of the research findings depends on referencing the scientist's academic position and mentioning their affiliated institution while stressing the importance and implications of the study. As for the latter feature, the writer attempts to stimulate the reader's interest by injecting rhetorical questions and inclusive language (i.e., *you*, *we*) in the narrative (Boynton, 2018).

Insofar, the literature points out that PSW brims with discoursal features that could be described as both structural, as demonstrated by the TP patterns, and interpersonal, demonstrated by the self-imposed personality within the popular science discourse. That said, the literature does not distinguish between types of PSW such as popular articles and popular books, as most investigations are concentrated on the former rather than the latter. For this reason, this analysis shall address this gap and attempt to explore the TP patterns in PSBs.

## 3. Method

The study uses a mixed-method approach to analyze how the PSB discourse writing is organized from the perspective of Theme Progression (TP) patterns. A corpus of 93,078 words was compiled from various popular books.

### *3.1 Data Selection*

The data were collected from two formats: paperback and electronic, in a six-month period, from December 2022 to May 2023. The electronic formats were partially accessible, as some data was taken from Google Books; others were available in Amazon Kindle format. For ethical reasons, the corpus was compiled while considering the fair use of written mediums, as outlined by the UK's Intellectual Property Office (2014).

Each science category contains three disciplines under which are six conspecific book excerpts. An approximation of 2500-word excerpts were typewritten on a plain text file. The corpus size is 93,078 words divided into 46448 words for hard science (See Appendix A) and 46630 words for soft science (See Appendix B). The disciplines under hard science are Physics, Biology, Chemistry. The soft science disciplines are Philosophy, Psychology, and Linguistics. The summary of the corpus is in the table below:

Table 1. Summary of the hard science and the soft science data

| Category | Hard Science | Soft Science |
|---|---|---|
| No. of Disciplines | 3 | 3 |
| No. of Text in a single discipline | 6 | 6 |
| Average word count for each text | 2500 | |
| Total word count by Category | 46448 | 46630 |
| Total word count | 93,078 | |

The books in each discipline were categorized according to the subject matter indicated in the title, the categorization on Goodreads and the author's scientific background. Furthermore, the selection of each text underwent two main criteria: First, the text is selected from the start of a chapter or a subchapter. Second, if any, figures, images, and other charts will be intentionally omitted.

### *3.2 Data Analysis Procedures*

This analysis followed a particular procedure: First, the TP patterns were reported at the subcategory level (e.g., biology, psychology, physics) after which the descriptive and statistical testing are performed. Second, the TP patterns were then reported at the category level (i.e., hard science, soft science) while performing the same statistical procedure. The chi-squared test ran the data on R Studio for each respective annotation tool to demonstrate any categorical differences in the corpus.

To ensure the normality of distribution in the PSB corpus pre-analysis, the one-sample Kolmogorov-Smirnov Test (KS Test) was conducted at two levels of inquiry: The first being by each discipline within the two categories while the second, by the main science category. In the first level, the test shows that the results of the word count totaled for the scientific disciplines is below the significance level ($p > 0.000$***). In the second level, the results are given according to the hard and soft science category under their respective category ($p > 0.000$***). Overall, the test failed to reject that the dataset at the two levels of inquiry do not follow a normality of distribution.

#### 3.2.1 Theme Progression analysis

The Thematic elements were manually annotated by using a researcher-created Excel sheet that counts the number of Themes and Rhemes in each popularized text. The TP patterns were mapped on the clausal and paragraph level. The two levels informed the analysis of results at the text level. At the clausal level, Theme is viewed from three related perspectives: 1) Theme as a message, 2) Theme in the beginning of the sentence structure and 3) Theme in context (Marfuaty, 2015; Syharizal et al., 2018). The Theme and the Rheme in the text followed a list of grammatical criteria introduced by Mauranen (1993): word class relations (e.g., reference, identity, word class change), semantic relations (synonyms, antonyms, etc.), and possessive relations.

When underlining the Theme and the Rheme, the annotation in the SFL tradition identifies the literal first word of the sentence as the Theme while the rest is the rheme, as seen in the figure below.

Table 2. Conventional TP analysis

| Theme | Rheme |
|---|---|
| *One the face of it* | *the monkeys epitomize how some animals survive superbly in conditions for which they were not originally adapted.* |

But this study looks at the Theme in the context of the text, thereby following a slightly modified annotation pattern. The sentence example contains an adverbial phrase, paratactic and hypotactic clause. The following breakdown of the clause could be useful in a more elaborate annotation of the corpus. The label *specifier* is borrowed from Mel'čuk (2001) and Bohnet et al. (2013), who both argue that Theme can be preceded by a contextualizing frame. From this perspective, the Theme is interpreted as the meaning of the utterance; the Rheme is what is stated afterwards.

Table 3. Adapted TP analysis

| Specifier | Theme | Rheme |
|---|---|---|
| *One the face of it* | *the monkeys* | *epitomize how some animals survive superbly in conditions* |
| *for which* | *they* | *were not originally adapted* |

At times, clauses contained empty categories due to the sentence structure. For instance, the biology dataset has this sentence *The cells divided* which lacks a Rheme. In such case, the verb was determined as the Rheme to maintain a consistent flow of the analysis.

The text at the paragraph level could potentially present more than one TP annotation due to the interpretive nature of the context. One example from the biology dataset is this following text, interpreted in two manners. The first follows a Linear progression pattern whereas the second follows a Zigzag.

> **[SP** *It appears***]** *that* **[TP linear** *a change in the hereditary constitution of the cells***]** *has occurred in the soma or body", the biologist T.D.A. Cockrell wrote in 1917, "without having any connection with the process of sexual reproduction." Cockrell called* **[TP Linear** *this change a somatic mutation.***]**
>
> **[SP** *It appears***]** *that a change in the hereditary constitution of the cells has occurred in the soma or body",* **[TP zigzag** *the biologist T.D.A. Cockrell***]** *wrote in 1917, "without having any connection with the process of sexual reproduction."* **[TP zigzag** *Cockrell***]** *called this change a somatic mutation.*

Because the purpose of the study is to denote the TP development of PSB writing, including two annotations in the same paragraph might cause an impediment to the flow of the analysis. An optimal solution to this overlap was to prioritize the broader level of text, taking into account the analysis of the subsequent paragraphs.

## 4. Results

### *4.1 Statistical Analysis*

This study has analyzed the organization of PSB discourse. Table 1 shows the distribution of TP patterns across the subcategories of the science datasets. Under the hard science subcategories, Reiteration is shown to be occurring 625 times in Biology (15.44%), 631 times (15.59%) in Chemistry, and 654 times (16.16%) in Physics. On the flip side, the occurrence of Reiteration is noticeably larger in the soft science subcategories, occurring 739 times (18.26%) in Psychology, 711 times (17.57%) in Philosophy, and 687 times (16.98%) in Linguistics.

Linear is more substantial in the subcategories of hard science than its soft counterpart. The occurrence of Linear in biology is 148 times with the highest percentage distribution, 30.64 percent, followed by Chemistry at 104 times with 21.53 percent. The Linear pattern occurs only 67 times in physics with 13.87 percent. By contrast, the number of occurrences drops significantly in Psychology (n = 56, 11.59%), Philosophy (n = 44, 9.11%), and Linguistics (n = 64, 13.25%).

To some degree, Zigzag patterns fluctuate across the data. The Zigzag patterns in Physics (n = 220, 18.49%) and Psychology (n = 263, 22.10%) have higher occurrences than their respective subcategories. Another observation is: The two hard science subcategories Biology (n = 166, 13.95%) and Chemistry (n = 172, 14.45%) have a slightly lower number of occurrences than the ones in Philosophy (n = 202, 16.97%) and Linguistics (n = 167, 14.03%).

Like Linear, Multiple-R fluctuates proportionately with more occurrences in Biology (n = 30, 23.81%), Psychology (n = 46, 36.51%), and Linguistics (n = 22, 17.46%). However, elsewhere in the data, the occurrences in Chemistry and Physics plummet to an equal value of (n = 10, 7.94%), and even smaller in Philosophy (n = 8, 6.35%).

The last TP pattern, Derived T, shows a similar distribution as Reiteration, where the TP patterns are largest in the soft science subcategories: Psychology (n = 379, 17.50%), Philosophy (n = 441, 19.44%), and Linguistics (n = 458, 20.19%), than the ones in hard science: Biology (n = 317, 13.98%), Chemistry (n = 291, 12.83), and

Physics (n = 364, 16.05).

Table 1. Descriptive Analysis of Science Subcategories

| Science | Reiteration | | Linear | | Zigzag | | Multiple-R | | Derived T | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *n* | % | *n* | % | *n* | % | *n* | % | *n* | % |
| Hard science | | | | | | | | | | |
| Biology | 625 | 15.44 | 148 | 30.64 | 166 | 13.95 | 30 | 23.81 | 317 | 13.98 |
| Chemistry | 631 | 15.59 | 104 | 21.53 | 172 | 14.45 | 10 | 7.94 | 291 | 12.83 |
| Physics | 654 | 16.16 | 67 | 13.87 | 220 | 18.49 | 10 | 7.94 | 364 | 16.05 |
| Soft Science | | | | | | | | | | |
| Psychology | 739 | 18.26 | 56 | 11.59 | 263 | 22.10 | 46 | 36.51 | 397 | 17.50 |
| Philosophy | 711 | 17.57 | 44 | 9.11 | 202 | 16.97 | 8 | 6.35 | 441 | 19.44 |
| Linguistics | 687 | 16.98 | 64 | 13.25 | 167 | 14.03 | 22 | 17.46 | 458 | 20.19 |

Table 2 demonstrates the mean values of each TP pattern across the dataset, which are as follows: on average, Reiteration occurs 674.5 times, making it the largest TP occurrence in the entire data, followed by the Derived T, occurring 378 times. By contrast, Multiple-R has the smallest value, occurring only 21 times in the data. Zigzag and Linear occur 198.33 times and 80.5 times respectively. Interestingly, all the TP patterns have an average occurrence of 16.66 percent despite the variable number of occurrences.

Table 2. Mean of TP Patterns

| Science | Reiteration | | Linear | | Zigzag | | Multiple-R | | Derived T | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *M* | % | *M* | % | *M* | % | *M* | % | *M* | % |
| Hard science | 674.5 | 16.66 | 80.5 | 16.66 | 198.33 | 16.66 | 21 | 16.66 | 378 | 16.66 |

The average occurrence of all the TP patterns implies a normality of distribution. To confirm this observation, the Shapiro test was performed to know whether each TP pattern is distributed equally across the popular science corpus. As seen in Table 3, the p-value of each individual TP pattern exceeds the level of significance ($\alpha = 0.05$), which suggests that every TP pattern is equally distributed across the popular science corpus.

Table 3. Shapiro Test of TP Patterns

| Theme | *W* | *p* |
|---|---|---|
| Reiteration | *0.935* | *0.62* |
| Linear | *0.87* | *0.214* |
| Zigzag | *0.87* | *0.204* |
| Multiple-R | 0.87 | 0.21 |
| Derived T | 0.95 | 0.72 |

*Note*. All the results for W have been approximated.

To examine whether a statistical difference exists across the data, the individual TP pattern was tested against the subcategories of both the hard and soft science (i.e., Biology, Chemistry, etc.) using the chi-squared test. As shown in Table 4, the p-value of each TP pattern is below the significance level from which we could conclude that generally, each TP pattern significantly differs in the organization of discourse within the subcategories and that the differences in the presentation targeted at the general readership is more likely to be related to the science subcategory itself.

Table 4. Chi-squared Test of TP Patterns within the science subcategories

| Theme | $X^2$ | *df* | *p* |
|---|---|---|---|
| Reiteration | *15.43* | *5* | *0.0086* |
| Linear | *93.11* | *5* | *<.001*** |
| Zigzag | *37.23* | *5* | *<.001*** |
| Multiple-R | 53.23 | *5* | *<.001*** |
| Derived T | 58.77 | *5* | *<.001*** |

*Note*. All the results for $X^2$ have been approximated.

Table 5 arranges the results according to their soft and hard science categories instead of their subcategories (e.g., biology, linguistics, etc.). It can be noted that on average, Reiteration occupies approximately fifty percent of the patterns in both soft (49.64%) and hard science (50.14%) and nearly a third of the patterns are Derived T for both respective categories (30.10%, 25.52%). While Zigzag occurs almost fifteen percent in both categories (14.68%, 14.65%), Linear and Multiple-R mark the lowest pattern occurrences in the whole dataset. Linear is between approximately three and eight percents in soft and hard science categories (3.81%, 8.37%); Multiple-R is at no more than two percent in soft and hard science (1.77%, 1.31%).

Table 5. Descriptive Analysis of Science Categories

| Theme | Soft Science | | Hard Science | |
|---|---|---|---|---|
| | *n* | % | *n* | % |
| Reiteration | 2137 | 49.64 | 1910 | 50.14 |
| Linear | 164 | 3.81 | 319 | 8.37 |
| Zigzag | 632 | 14.68 | 558 | 14.65 |
| Multiple-R | 76 | 1.77 | 50 | 1.31 |
| Derived T | 1292 | 30.10 | 972 | 25.52 |

The symmetrical distribution of TPs between the hard and soft science necessitated that the KS one-sample test is run for distribution normality. The result of the test gave no evidence of non-normality in the data. From here, the chi-squared test was used again to draw any statistical differences between the two main categories. The result is below the alpha level 0.05, ($p < .001$***) indicating that the two science categories are statistically different. In line with the previous chi-squared test, this one implies that categorically speaking, hard science and soft science writing may be significantly different in the way each category shapes its Thematic organization of discourse, that of which is used to appropriate the text to their general, non-specialist readership.

### *4.2 Qualitative analysis*

The continuity of the results, that of the subcategories and that of the main science categories bear two observations in the data. First, the TP patterns in both arrangements reveal that hard and soft science authors present similar discoursal structures, as shown in Table 6.

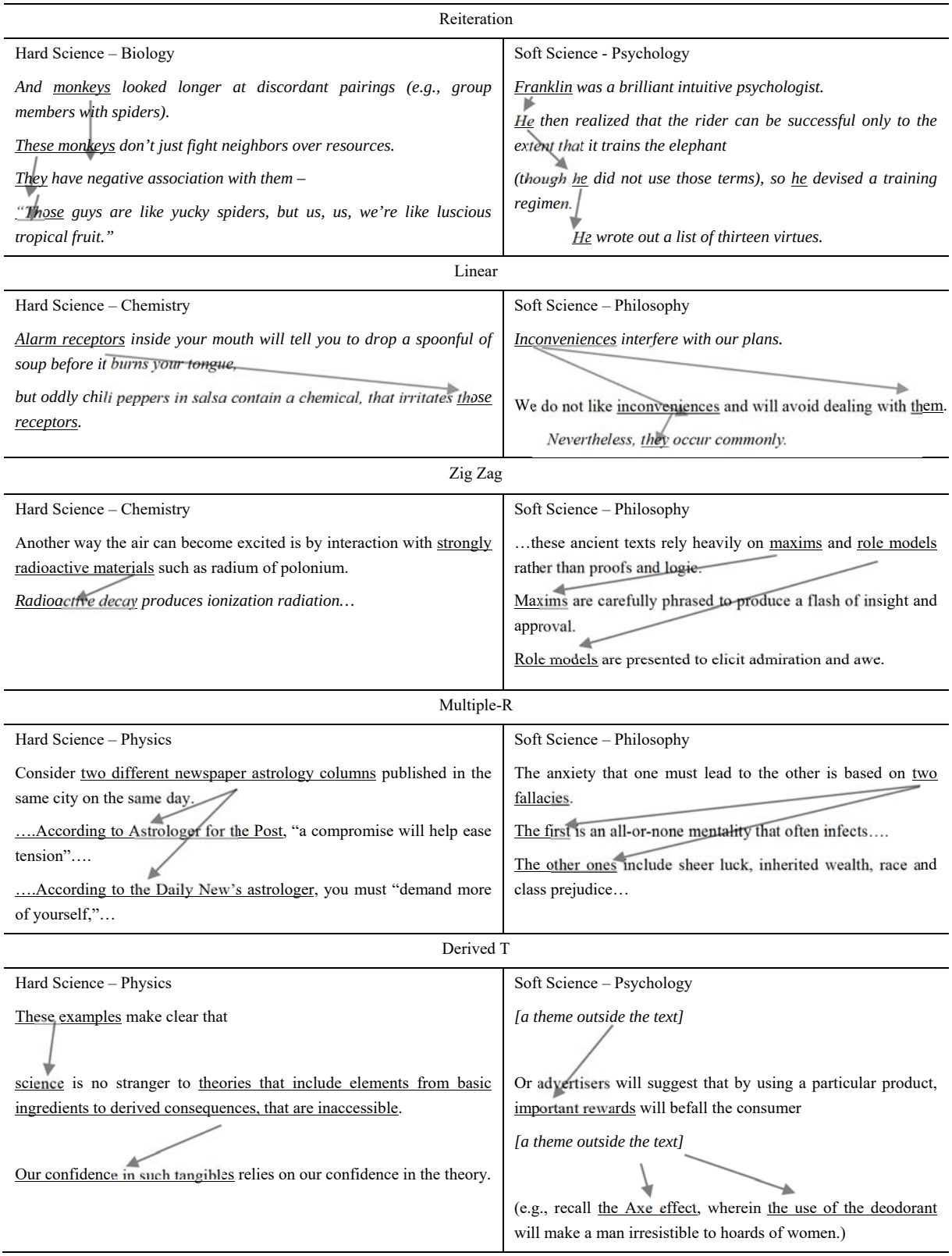


Figure 6. TP Patterns of Hard and Soft Science

What is demonstrated in Table 6 aligns with the second observation: Employing the TP pattern could have a predictable trajectory in the text rather than being due to chance, as suggested by the statistical results. The qualitative analysis yields such akin similarity between the hard and the soft science categories. To begin with, the discoursal techniques mentioned above can be underlined in subcategories such as—but not limited

to—biology (hard science) and psychology (soft science).

The most occurring TP pattern is Reiteration, which is used for different discoursal purposes, for example, to unpack and explain intricate ideas, to tell a story or form a narrative—sometimes a personal one. The following example, excerpted from *The Selfish Genes* (hard science), demonstrates such discoursal behavior:

| **SP** | **Theme** | **Rheme** |
|---|---|---|
| | *We* | *are survival machines* |
| | *'we'* | *does not mean just people* |
| | *It* | *embraces all animals, plants, bacteria, and viruses* |
| | *The total number of survival machines* | *is very difficult to count* |
| *And even* | *the total number of species* | *is unknown* |
| *Taking insects alone,* | *the number of living species* | *has been estimated at around three million* |
| *And* | *the number of individual insects* | *may be a million million million.* |

Unexpectedly, Reiteration is used in anthropomorphization cases—where the author tends to personify inanimate objects or concepts. It is a common narrative device used in PSW, which can be seen in *Stuff Matters* (hard science):

| **SP** | **Theme** | **Rheme** |
|---|---|---|
| | *Ligaments* | *are the elastic bands of the body* |
| *Along with our muscles and tendons, which …* | *they* | *hold our joints together* |
| *and* | *make* | *us springy* |
| *It* | *the ligaments' job* | *to connect one bone to another* |
| | *They* | *are viscoelastic* |
| *which that* | *they* | *will stretch immediately a certain amount* |

Derived themes, the second most frequently occurring Theme, are seen used in the synthesis of informational text by inserting nominalization or abstract concepts from the topical words either within the same paragraph or the previous one. Another Derived function found in the data comes in the form of callbacks of keywords and concepts. Finally, the last Derived function seen in the data is exemplification of conceptual ideas in the text. The following two excerpts from *Because Internet Understanding* (soft science) and *The Hidden Reality* (hard science) demonstrate these functions of Derived themes.

| **SP** | **Theme** | **Rheme** |
|---|---|---|
| *As* | *technology* | *advanced* |
| *So did* | *dialectology* | |
| *In the 1960s* | *The Dictionary of American Regional English* | *sent out fieldworkers in "Word Wagons" (green Dodge vans outfitted with a fold-out bed, an icebox, and a gas stove-top) to record locals in over a thousand communities on briefcase-sized reel-to-reel tape recorders.* |
| *In the 1990s,* | *the creators of the Atlas North American English* | *let their fingers do the biking* |
| *and* | *conducted* | *telephone interviews with 762 random people…* |

| | | |
|---|---|---|
| *In 2002,* | *The Harvard Dialect Survey* | *produced a linguistic questionnaire that anyone could complete online.* |

| **SP** | **Theme** | **Rheme** |
|---|---|---|
| | *This reasoning* | *is borne out by the mathematics* |
| *In the inflationary multiverse* | *universes* | *can collide* |
| *Moreover,* | *a number of research groups (including…)* | *that* |
| *Whereas* | *some collisions* | *may violently disrupt each bubble universe's internal structure…* |

Next is Zigzag, which can also be found abundantly in the data. Although Zigzag patterns create cohesion in numerous ways in the text, in the case of the data, they can be encapsulated in the following: cohesion by synonymous use of words, cohesion by pronouns, and cohesion by topic or subtopic. In *The Righteous Mind* (soft science), the excerpted text elaborates on an ideological movement that took place in the late 2000s, spearheaded by Sam Harris, Richard Dawkins, Daniel Dennett, and Christopher Hitchens. The four of them are known as the horsemen of New Atheism. Structurally speaking, the nominal phrase *the four horsemen of New Atheism* sits as the main cohesive reference to the sequential clauses in the text. The pronoun *he* in the third clause refers to *Hitchens* in the Rheme of the second clause. In the clauses beginning with *Harris*, *Dawkins*, and *Dennett*, all these proper nouns are references to the noun phrase *the four horsemen of New Atheism* in the beginning of the text.

| **SP** | **Theme** | **Rheme** |
|---|---|---|
| | *These four authors* | *are known as the four horsemen of New Atheism* |
| *But* | *I'm* | *going to set Hitchens aside* |
| *because* | *He* | *is a journalist* |
| *whose* | *book* | *made no pretense to be anything other than a polemical diatribe.* |
| | *The other three authors, however,* | *are men of science* |
| | *Harris* | *was a graduate student in neuroscience at the time* |
| | *Dawkins* | *is a biologist* |
| *and* | *Dennett* | *is a philosopher* |
| *who* | *has written* | *widely on evolution* |

The last two TP patterns Linear and Multiple-R, although hardly visible, can yield some insight from the available data. Regarding the Linear themes, they are often represented in a logical progression while Multiple-R themes disperse two Rhemes in the subsequent clauses as Themes. Each excerpted text from *The Selfish Gene* (hard science) demonstrates the respective Theme function of the two TP patterns.

| **SP** | **Theme** | **Rheme** |
|---|---|---|
| | *The paired chromosomes* | *do not spend all their lives physically in contact with each other or even near each other* |
| *In* | *what sense then* | *are they "paired"?* |
| | *In the sense that each volume coming originally from the father* | *as a direct alternative to one particular volume coming originally from the mother* |
| *For instance* | *Page 6 of Volume 13a and Page 6 of Volume 13b* | *might both be 'about' eye color* |

| SP | Theme | Rheme |
|---|---|---|
| | *I* | *seem to be saying two things that contradict each other* |
| *One the one hand,* | *we* | *trust our past knowledge* |
| *And on the other hand* | *we* | *are always ready to modify, in depth, part of our conceptual structure of the world.* |

Overall, the data have an abundance of examples that point to a progression pattern that is used functionally with a reasonable degree of predictability in the text. The similarity demonstrated in the qualitative analysis synchronizes with the quantitative results and the differences are statistically visible. In terms of how the PSB discourse is overall organized, although the hard and soft science datasets identified in this research cannot be regarded as distinct genres but rather should be blanketed under the PSW genre, the differences among the individual scientific category could naturally be underlined.

## 5. Discussion

The TP analysis of popular science book writing aimed to explore how PSW book authors impart information to the intended reader with adequate accessibility and simplicity.

Broadly speaking, the results of both analyses pointed to statistically significant differences between the hard and soft science disciplines. The interpretation of the results is discussed in light of the TP patterns in the popularized hard science and soft science disciplines.

### *5.1 Thematic Progression*

The TP analysis was divided into the science subcategories (e.g., biology, psychology, etc.) and the science categories (i.e., soft, hard). Generally, the study has found a statistically categorical difference in the organized discourse between soft and hard science. More TPs are deployed in the soft science disciplines than their hard science counterparts, which suggests that soft science disciplines are more divergent, expository, and narrative-driven whilst explaining their fields to non-expert audiences. On the other hand, the hard science disciplines are more focused on presenting information in the most logical and progressive way possible to the reader while maintaining entertainment and reader engagement.

### *5.2 Science Subcategories*

The chi-squared test results pointed to significant differences for the five TP patterns. Generally, the popular science genre organizes its discourse using diverse TP patterns, prominently, Reiteration, Zigzag, and Derive themes. Of the six science subcategories, those ones that fall under the soft science category seem to center its discourse around its intended reader, as it engages them in detailed elaboration of the topic. On the flip side, the subcategories under hard science, though having a similar reader-oriented construction, seem to pay more attention to the presentation of its discourse, that is, its progressive and logical structure for the purpose of readability, realized by the higher occurrences of Linear themes. That said, all the subcategories exhibit a similar sensibility when organizing their discourse, exemplified by the aforementioned types of Themes.

In looking closely at the TP patterns, Reiterated, Zigzag and Derived theme patterns exist in larger proportions within the subcategories of soft science, which could be interpreted as a residual of the writing style that of scientific writing. Reiteration in the fields of psychology, philosophy, and linguistics creates coherence that is reader-friendly and relatable, translated into three types of discourses: narrative-driven, exposition, and to some degree, argumentation. While the same is true for the subfields under the hard science category, the underlying explanation for the higher occurrences of coherence in the soft science subcategories lies within the transferred writing conventions from scientific writing.

A possible explanation of this transfer is that these books are written by scientists who are active in their field of expertise and who are also writing to the public. Therefore, because of the author's interjection, shifting the orientation of writing from scientific to popular back and forth is possibly intermixed, which may be one of the reasons why PSW emulates the scientific language at times. With that in mind, because science communication is nearly always imparted by scientists, it is probable that those science-communicator scientists have inadvertently conventionalized the genre through their own writings and other engagements. Another observation of Reiteration themes is the use of storytelling as a rhetorical function in the text. Seemingly, both soft and hard science subcategories leverage the powerful effect of stories and anecdotes, instantiated in Reiteration themes. True that the narrative structure of stories in popular science writing is largely stripped off from the standard elements of storytelling (e.g., dialogues, monologues, etc.). In other words, storytelling in popular science

writing does not present a triumphant tale between two opposing forces, one being the protagonist while the other, the antagonist. Rather, it presents an expository narrative in a structure that renders scientific knowledge in both absorbable and personal way to the reader (Pilkington, 2016).

Likewise, Zigzag patterns appear to be slightly occurring more often in the soft science subcategories, as indicated by the descriptive results. More than that, the Zigzag patterns in the texts create cohesion, that is textually more complex than coherence. To elaborate, coherence is reader-centered while cohesion is text-centered since it does the linking of ideas by synonymous word usage, pronouns, or by important keywords. Another key difference found in the corpus between coherence, instantiated by Reiteration and cohesion, instantiated by Zigzag patterns is the level of operation. Naturally, while coherence ties the entirety of the text together, cohesion is largely limited at the paragraph level. Clearly, there exists a polarity in the discourse realized by Reiteration (i.e., coherence) and Zigzag (i.e., cohesion).

Regarding the noticeably higher occurrence of Derived Themes in the soft science subcategories, it could be explained by the nature of the fields themselves. Psychology, philosophy, and linguistics are densely abstract fields, unlike biology, chemistry, and physics that have more reach to the materialistic things in nature. The abstraction is mainly explained by linking conceptual ideas, and calling back key thematic words that move the discussion forward.

Linear patterns appear to be more proportional in hard science. Again, with this specific TP pattern, this could be interpreted as a residual writing style being transferred by scientists in their popularized works. However, there is more: Linear patterns allow the writer to create a logical progression of the idea, which explains why in this category more care is given to presenting information. The possible interpretation of this is that the hard science disciplines break down the intricate concepts and theory discussions in fields such as physics and chemistry, presenting them in bitesize pieces to the reader rather than bombarding them with information. Having said that, this tendency is less true with respect to the soft science subcategories.

Overall, diverse employment of TPs in PSW seems to diverge its discussion of the topic, as the need to explain and elaborate arises to keep the reader informed at a feasible pace. The divergence of thematic patterns in the corpus is consistent with the results by Nwogu and Bloor (2011) and Shandama and Nwogu (2013). Unfortunately, TP analysis of this genre is still a work-in-progress, which is why there is a scarcity of reflection regarding the TP patterns. The TP patterns underlined in this corpus run counter to scientific writing that appears to primarily contain Reiteration and Linear patterns, as many researchers have pointed out (Alvin, 2015; Alyousef, 2020, 2021; Alyousef & Alsharif, 2020; Babaii et al., 2016; Darrmila et al., 2019; Gunawan & Aziza, 2017; Leong et al., 2018; Susilowati et al., 2022; Wei & Yi, 2020). The main juxtaposing difference between the two writing genres is underlined by the exposition and elaboration characterized in the popular science genre.

### *5.3 Soft and Hard Science Categories*

Broadly speaking, soft and hard science disciplines in PSW mostly constitute Reiteration, Zigzag, and Derive theme patterns, which are in line with the above-mentioned interpretation of the subcategory analysis. Statistically speaking, the chi-squared test for the science categories suggests that there is a categorical difference between the two categories in which the soft science category appears to be proportionally more divergent in its thematic patterns.

This is consistent with the interpretation of the science subcategory analysis that characterizes the organization of the soft science discourse as more personal, relatable, and engaging. Bear in mind, this is also a feature found in hard science, yet not as proportionate. Seemingly, Reiteration, Zigzag and Derived themes in soft science category provide more exposition to the reader whereas the linear themes in the hard science category give a logically organized presentation of the hard science topic.

## 6. Conclusion

This SFL study has investigated the discourse-organizing TP patterns in PSB writing. It deployed a researcher-designed Theme-counting Excel sheet that calculated the TP patterns in the PS corpus and tabulated them afterwards. Generally speaking, the results shone light on the discoursal differences between the hard science disciplines and the soft science disciplines. Simply put, the soft science disciplines were noticed to focus more on exposition that connects with the target reader whereas the hard science disciplines focus on the logical presentation of the discourse to explain the scientific text to the target reader. The results of this study could prove viable to the genre analysis of PSW, particularly between books and online articles, which in turn, could extend to the development of audience-oriented writing skills such as improving coherence and cohesion for students majoring in natural and social sciences, especially at the graduate and postgraduate level. Furthermore,

when juxtaposed with scientific writing, PSW could be used to demonstrate the structural differences between writing for the public and writing for the experts from which an audience-focused writing style emerges in their scientific work.

Certainly, this study is not without its shortcomings, which could shed light on the potential new avenues of investigations within the popular science genre. The TP investigation of PSW is severely lacking in the currently existing literature, and any interpretation attempt to relate to writing pedagogy, for example, may possibly be underwhelmed because of this scarcity. Having said that, the analysis of the corpus in this study shows the potential for future analyses and comparisons within the popular science genre.

**Acknowledgments**

I would like to thank my doctoral supervisor Dr. Hesham Alyousef for his guidance and feedback.

**Authors' contributions**

Not applicable.

**Funding**

Not applicable.

**Competing interests**

Not applicable.

**Informed consent**

Obtained.

**Ethics approval**

The Publication Ethics Committee of the Canadian Center of Science and Education.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

**Provenance and peer review**

Not commissioned; externally double-blind peer reviewed.

**Data availability statement**

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

**Data sharing statement**

No additional data are available.

**References**

Alyousef, H. S. (2016). A multimodal discourse analysis of international postgraduate business students' finance texts: An investigation of theme and information value. *Social Semiotics*, *26*(5), 486–504. https://doi.org/10.1080/10350330.2015.1124518

Alyousef, H. S. (2020). A multimodal discourse analysis of English dentistry texts written by Saudi undergraduate students: A study of theme and information structure. *Open Linguistics*, *6*(1), 267–283. https://doi.org/10.1515/opli-2020-0103

Alyousef, H. S. (2021). The Status of Theme in Research Article Abstracts in Seven Dentistry Subdisciplines: A Text-Based Study of Intradisciplinary Variations and Similarities in Thematic. *Journal of Language and Education*, *1*, 28–45. https://doi.org/10.17323/jle.2021.10574

Alyousef, H. S., & Alsharif, M. H. (2020). A contrastive study of theme types and patterns in medical research article abstracts written in English by native and Saudi scholars. *Umm Al-Qura University Journal for Languages & Literature*, *26*. https://doi.org/10.54940/ll44457307

Babaii, E., Atai, M. R., & Shoja, L. (2016). A comparison of thematic choices and thematic progression patterns in the research articles of well-established and emerging disciplines. *Iranian Journal of Applied Linguistics*, *19*(2), 33–60. https://doi.org/10.29252/ijal.19.2.33

Boynton, S. (2018). The appliance of science: The challenges of undergraduate science students writing popular science. *The Asian Journal of Applied Linguistics*. Retrieved from https://hub.hku.hk/bitstream/10722/265214/1/content.pdf?accept=1

Carter-Thomas, S. (1999). Thematic networks and text types. *Open Edition Journal*, *23*(26), 139–147. https://doi.org/10.4000/asp.2361

Darmila, R. T., Syarif, H., & Amri, Z. (2019). Thematic Progression in Students' Discussion Essays of Undergraduate English Education Program at Universitas Negeri Padang. *International Journal of Educational Dynamics*, *2*(1), 254–262. https://doi.org/10.24036/ijeds.v2i1.253

Dijkstra, A., de Bakker, L., van Dam, F., & Jensen, E. (2020). Setting the Scene. In F. van Dam, L. de Bakker, A. Dijkstra & E. Jensen (Eds.), *Science Communication: An Introduction* (pp. 1–16). https://doi.org/10.1142/9789811209888_0001

Gunawan, W., & Aziza, F. (2017). Theme and thematic progression of undergraduate thesis: Investigating meaning making in academic writing. *Indonesian Journal of Applied Linguistics*, *7*(2), 413–424. https://doi.org/10.17509/ijal.v7i2.8350

Halliday, M. A. K., & Matthiessen, C. (2014). *An Introduction to Functional Grammar*. https://doi.org/10.4324/9780203783771

Intellectual Property Office. (2014). *Exception to copyright*. Retrieved Sept. 30, 2022, from https://www.gov.uk/guidance/exceptions-to-copyright#non-commercial-research-and-private-study

Kucharski, A. J. (2018). Ten simple rules for writing a popular science book. *PLoS computational biology*, 14(2), e1005808. Public Library of Science San Francisco, CA USA. https://doi.org/10.1371/journal.pcbi.1005808

Leong, A. P., Toh, A. L. L., & Chin, S. F. (2018). Examining Structure in Scientific Research Articles: A Study of Thematic Progression and Thematic Density. *Written Communication*, *35*(3), 286–314. https://doi.org/10.1177/0741088318767378

Marfuaty, F. A. & Wahudi, R. (2015). An analysis of thematic progression patterns: Opinion section texts of The Jakarta Post. *International Journal of Language Studies*, *9*(3), 109–130. http://repository.uin-malang.ac.id/5804/

Nwogu, K., & Bloor, T. (2011). Thematic progression in professional and popular medical texts. In *Functional and Systemic Linguistics: Approaches and Uses* (pp. 369–384). https://doi.org/10.1515/9783110883527.369

Paniukova, S. (2019). *Storytelling method in popular science journalism*. Ceeol.Com. Retrieved from https://www.ceeol.com/search/article-detail?id=877526

Pelger, S. (2018). Popular science writing bringing new perspectives into science students' theses. *International Journal of Science Education* (Part B: Communication and Public Engagement), *8*(1), 1–13. https://doi.org/10.1080/21548455.2017.1371355

Pilkington, O. A. (2016). Educating Future Popularizers of Science: Linguistic Insight for Scientists. Conference Proceeding. *New Perspectives in Scienze Education*, 27.

Sharndama, E. C., & Nwogu, K. N. (2013). A Comparative Study of Thematic Progressions in Professional and Popularized Legal Texts. *Taiwan International ESP Journal*, *5*(2), 47–63. https://doi.org/10.6706/TIESPJ.2013.5.2.3

Susilowati, E., Faridi, A., & Sakhiyya, Z. (2022). Thematic Structure and Thematic Progression in Research Articles Published in Scopus-Indexed International Journals. *English Education Journal*, *12*(1), 55–66. https://doi.org/10.15294/eej.v12i1.53229

Uyun, R. M. (2016). A Discourse Analysis Through Corpus Linguistics on the Case of Popular Science Book the Grand Design. *Kajian Linguistik Dan Sastra*, *1*(2), 81–94. https://doi.org/10.23917/kls.v1i2.3643

Van Thang, N., & Tam, N. T. M. (2020). The realization of thematic progression in online IELTS writing task 2 samples in an English-learning website: a case study. *Js. Vnu. Edu. Vn*, *36*(6), 134–149. https://doi.org/10.25073/2525-2445/vnufs.4634

Wei, X., & Yi, Z. (2020). *Coherent Function of Interrogative Mood Metaphor in Academic Discourse: Thematic Progression Pattern Perspective*. https://doi.org/10.13189/lls.2020.080503

Wu, S. M., Lee, S. H., Chun, E., & Chan, Y. (2018). Teaching Academic Literacy Using Popular Science Texts: A Case Study. *Teaching & Learning Inquiry*, *6*(2), 29–49. https://doi.org/10.20343/teachlearninqu.6.2.4

## Appendix A. Datasets of Hard Science Disciplines

| Hard Science | Books | Author | Year | Publisher |
|---|---|---|---|---|
| **Physics** | The Greatest Story Ever Told – So Far: Why Are We Here? | Lawrence M. Krauss | 2017 | Atria Books |
| | The Big Picture: On the Origins of Life, Meaning and the Universe Itself | Sean Carrol | 2016 | Dutton |
| | Time Reborn: From the Crisis in Physics to the Future of the Universe | Lee Smolin | 2014 | Marine Books |
| | Cosmos | Carl Sagan | 1983 | Time Warner Books |
| | The Hidden Reality: Parallel Universes and Deep Laws of the Cosmos | Brian Greene | 2011 | Knopf |
| | The Universe: Leading Scientists Explore the Origin, Mysteries and the Future of the Cosmos | John Brockman | 2014 | Harper Perennial |
| **Biology** | The Selfish Genes | Richard Dawkins | 1976/2016 | Oxford University Press |
| | Behave: The Biology of Humans at Our Best and Worst | Robert M. Sapolsky | 2017 | Penguin Press |
| | She Has Her Mother's Laugh: The Powers, Perversions, and Potential of Heredity | Carl Zimmer | 2018 | Dutton |
| | Genome: the Autobiography of a Species in 23 Chapters | Matt Ridley | 1999 | Harper Perennial |
| | Power, Sex, Suicide: Mitochondria and the Meaning of Life | Nick Lane | 2006 | Oxford University Press |
| | The Story of the Human Body: Evolution, Health, and Disease | Daniel E. Liberman | 2013 | Vintage |
| **Chemistry** | Stuff Matters: Exploring the Marvelous Materials That Shape Our Man-Made World | Mark Miodownik | 2015 | Marine Books |
| | The Disappearing Spoon: And Other True Tales of Madness, Love, and the History of the World from the Periodic Table of the Elements | Sam Kean | 2011 | Back Bay Books |
| | Napoleon's Buttons: How 17 Molecules Changed History | Jay Burreson & Penny Le Couteur | 2004 | Jeremy P Tarcher |
| | Strange Chemistry: The Stories Your Chemistry Teachers Wouldn't Tell You | Steven Farmer | 2017 | Wiley |
| | Roald Hoffmann on the Philosophy, Art, and Science of Chemistry | Roald Hoffmann | 2012 | Oxford University Press |
| | The Alchemy of Air: A Jewish Genuis, A Doomed Tycoon, and the Scientific Discovery That Fed the World but Fueled the Rise of Hitler | Thomas Hager | 2008 | Crown |

## Appendix B

## Datasets of Soft Science Disciplines

| Soft Science | Books | Author | Year | Publisher |
|---|---|---|---|---|
| **Philosophy** | The Blank Slate: The Modern Denialism of Human Nature | Steven Pinker | 2002 | Penguin Books |
| | Speech Matters: On Lying, Morality, and the Law | Seana V. Shiffrin | 2014 | Princeton University Press |
| | The Moral Landscape: How Science Can Determine Human Values | Sam Harris | 2010 | Free Press |
| | Beyond the Hoax: Science, Philosophy, and Culture | Alan Sokal | 2008 | Oxford University Press |
| | The Happiness Hypothesis: Finding Modern Truth in Ancient Wisdom | Jonathan Haidt | 2006 | Basic Books |
| | Maps of Meaning: The Architecture of Belief | Jordan Peterson | 1999 | Routledge |
| **Psychology** | The Righteous Mind: Why Good People Are Divided by Politics and Religion | Jonathan Haidt | 2012 | Pantheon |
| | Thinking, Fast and Slow | Daniel Kahneman | 2011 | Farrar, Straus and Giroux |
| | The Moral Animal: Why We Are the Way We Are – The New Science of Evolutionary Psychology | Robert Wright | 1994 | Vintage |
| | The Better Angels of Our Nature: Why Violence Has Declined | Steven Pinker | 2012 | Penguin Books |
| | The Consuming Instinct: What Juicy Burgers, Ferraris, Pornography, and Gift Giving Reveal about Human Nature | Gad Saad | 2011 | Prometheus Books |
| | The Sociopath Next Door | Martha Stout | 2006 | Harmony |
| **Linguistics** | How Language Began: The Story of Humanity's Greatest Invention | Daniel L. Everett | 2017 | Liveright |
| | The Language Instinct: How the Mind Creates Language | Steven Pinker | 2000 | Harper Perennial Modern Classics |
| | The Unfolding of Language: An Evolutionary Tour of Mankind's Greatest Invention | Guy Deutcher | 2005 | Holt Paperbacks |
| | Because Internet: Understanding the New Rules of Language | Gretchen McCulloch | 2019 | Riverhead Books |
| | Metaphors We Live By | George Lakoff & Mark Johnson | 2003 | The University of Chicago Press |
| | The Language Hoax | John McWhorter | 2014 | Oxford University Press |