A lexical and syntactic study of research article titles in Library Science and Scientometrics

Junli Diao[1]

Abstract

Title of a research article is an abstract of the abstract. Titles play a decisive role in convincing readers at first sight whether articles are worth reading or not. Not only do research article titles show how carefully words are chosen by authors, but also reflect disciplinary differences in terms of title words and structure between hard sciences and soft sciences. This study examined the lexical density and syntactic structure of 690 research article titles chosen from five Library Science and Scientometrics journals, aiming to reveal disciplinary differences. The result suggested both Library Science and Scientometrics have almost the same title length and the prevalent usage of Nominal Phrase (NP) to govern the title structure. The result also stated some disciplinary differences: Library Science demonstrates more punctuation complexity, particularly a greater frequency in using colons; but Scientometrics shows more involvement of words related to research methods, which is an indicator to papers' scientific value, and more usage of declarative Full Sentence (FS) structure, which were mostly discovered in the research articles in hard sciences.

Keywords Library Science · Scientometrics · Title · Lexical density · Syntactic structure

Introduction

Titles of academic publications summarize and represent the content. Although brief, they are "serious stuff" (Swales, 1990, p. 144). Therefore, titles should be clear and accurate so as to reflect what the publications are about. Meanwhile, they have to be effective enough to establish instant communication with readers and attract them to read.

The history of using titles to represent the whole of a literary work can be dated back to the Bronze Age, when the first line of clay tablet texts were grouped together as a list in the library of the ancient city of Hattusas (Casson, 2001). However, titles treated as a field of study is fairly modern. It was conceived in the articulation of *Titrologie* in French scholars' literary critique in the 1970s (Baicchi, 2003). Baicchi, hence, underscored the English term "titlelogy" in the review of studies on titles that were carried out in the twentieth century. In the past 3 decades, the role that titles play in academic publications, including journal

✉ Junli Diao
  jdiao@york.cuny.edu

1 York College Library, The City University of New York, New York, USA

articles, conference papers, dissertations, and research reports, are attracting an increasing number of researchers' attention. The rise of the study of titles in academic publications was not an isolated, independent, self-growing phenomenon. On the contrary, it was strongly influenced by genre-based textual analysis in the field of English for Specific Purpose (ESP), ever since John Swales published his milestone monograph *Genre Analysis: English in Academic and Research Setting* (Moattarian & Alibabaee, 2015; Morales et al., 2020). Titlelogy has been examined at the language (Busch-Lauer, 2000; Soler, 2011), cultural (Xie, 2020; Yakhontova, 2002), and format level (Morales et al., 2020; Slougui, 2018). Irrespective of different findings in various aspects, these studies lend themselves to being a strong statement of how crucial a role titles play in the whole text.

Research articles are a major academic publication through which scholars share their research results and/or contributions to a given field. The importance of research article titles have been increasingly investigated in various knowledge disciplines. Some studies concentrate on individual discipline, such as Computer Science (Anthony, 2001), Medicine (Goodman, Thacker, & Siegel, 2001; Wang & Bai, 2007), Linguistics (Cheng, Kuo, & Kuo, 2012); some have a comprehensive coverage of multiple disciplines and examined research article titles through a comparative perspective (Appiah et al., 2019; Haggan, 2004; Moattarian & Alibabaee, 2015; Nagano, 2015). A number of diachronic studies investigated the patterns and types of information provided in titles of academic publications as the passage of time (Sahragard & Meihami, 2016; Salager-Meyer, Ariza, & Marianela, 2013). Numerous studies done on research article titles demonstrate that titles have the lexical, syntactic, and semantic complexity in academic writing, which calls for continued effort to step into the discipline that is neglected, such as Library and Information Science (LIS). Therefore, this research attempts to fill the blind spot and to provide a preliminary analysis of lexical density and syntactic structure of research article titles in this discipline.

Literature review

Linguistic models

Linguistic models, which were formulated by researchers and applied to studies on titles, usually display a conflation or synthesization of lexical, syntactic, and semantic parameters. Lexical parameters check the attributes of words; syntactic parameters examine the features of sentence structure; semantic parameters look into factors pertaining to language meaning. Buxton and Meadows (1977)'s study is the earliest study on research article titles that the author can find. It analyzed hundreds of titles from English, French, and German periodicals. Parameters involved in this study included year (1946–1973), all words per title, substantive words per title, propositional substantive words, and characters per substantive words. The involvement of time range and the prominence of words among all parameters show that this study used a diachronically-based and lexically-oriented research model.

In the 1980s, the usage of colons in the titles of scholarly publications caught researchers' attention. Dillon (1981, 1982) hypothesized and then Perry (1985) evidenced a link between colons in the titles of academic publications and authors' scholarly productivity. Since the 1980s, punctuation marks in titles, in particularly the colon, have been specifically investigated in a series of studies (Diers & Downs, 1994; Hartley, 2007; Lewison & Hartley, 2005; Michelson, 1994; Ziebland & Pope, 1995, cited in Hartley, 2007) or as

an indispensible component of comprehensive studies on titlelogy in various disciplines (Appiah et al., 2019; Haggan, 2004; Salager-Meyer et al. 2013, just to name a few).

In the 1990s, the observed studies on titles of academic publications expanded to a moderately broader and deeper scope, with a comprehensive coverage of words, punctuation marks, verb forms, articles, and patterns of phrase coordination (Fortanet, Coll, Palmer, & Posteguillo, 1997; Fortanet et al., 1998; Yitzhaki, 1997). At the same time in the 1990s, genre-based analysis of different types of texts increasingly attracted researchers' attention as the consequence of information explosion and the drastic increase of scholarly communication (Moattarian & Alibabaee, 2015, p. 28). Readers of academic journals tended to treat research article titles like newspaper headlines to grab instant information and keep up with literature (Trosborg, 2000, viii). Therefore, the linguistic models used to analyze the dynamic aspects of titles, as the opening, leading structural component of academic articles, gradually prefer integrating with the analysis of titles' social-cognitive functions. For instance, Haggan (2004) categorized titles into three basic types: full sentence, compound, and a remaining group. Such broad categorizations leave the researcher much room to explore and interpret titles' pragmatic functions, such as advertising and information packaging. Soler (2007) categorized the occurrence of a total number of 660 titles in social science and biological science into nominal-group construction, compound construction, full-sentence construction, and question construction, anticipating the model to show how authors expect to communicate and interact with readers through research article titles.

Gesuato (2008) analyzed 1000 English titles of publications in Applied Linguistics from four different publication genres: books, dissertations, journal articles, and proceedings papers. The researcher developed a comprehensive, thorough, sophisticated analytic model. In addition to measurement of title length, Gesuato divided all titles into single-unite titles and multi-unit titles (two-unit, three-unit and four-unit) (See examples 5–14). Multi-unit titles were exhaustively subdivided by the usage of full sentences, noun phrases, verb phrases, prepositional phrases, and adverb phrases, etc. The syntactic structure of two-unit titles, which were dominantly distributed among four genres, was further subdivided into 24 different categories. The structure of nominal heads was analyzed into two categories: pre-modification, consisting of five subtypes, and post-modification and its coordination, comprising four subgroups. Although this research was conducted within Applied Linguistics only, Gesuato's comprehensive analysis of the complexity of linguistic characteristics of titles was influential. Its impact can be traced directly or indirectly in a number of succeeding studies in the past few years (Appiah et al., 2019; Morales et al., 2020; Nagano, 2015; Slougui, 2018). This research was influenced by Gesuato's study as well.

Disciplinary differences

The extent to which titles are informative is measured by title length. Generally speaking, the longer titles are, the more information they contain. The surveyed literature demonstrates that titles in hard sciences tend to be longer than ones in soft sciences (Buxton & Meadows, 1977; Fortanet et al., 1997; Nagano, 2015; Soler, 2007). Yitzhaki (1997) believed that titles in *harder* sciences required more terminological, substantive words for title-based indexing and retrieval purposes, leading to longer, more informative titles; however, titles in *softer* sciences tended to use shorter, freer, more flexible title presentation. Not only did title length bear the mark of a disciplinary difference between hard sciences and soft sciences, so did the usage of punctuation marks, in particular the colon. Through

reviewing 17 studies, Hartley (2007) noticed that there was a gradual increase of the percentages of colonic titles from natural sciences to social sciences.

No matter whether in single-unit titles or multiple-unit titles, there is a major group that gives preference to the use of the Nominal Phrase (NP) (See example 6). NP titles comprise of at least one noun serving as the leading head of the whole title structure. A very interesting finding in the structural organization of research article titles is that the nominal type is dominant across both soft sciences and hard sciences (Busch-Lauer, 2000; Fortanet et al., 1998). The prevalence of nominal title construction in both soft and hard sciences suggests the possibility of disciplinary difference to be small. However, Full Sentence (FS) title structure is a different story. Berkenkotter and Huckin (1995) pointed out FS titles are a trait of science papers, particularly Biology. Haggan (2004) examined research article titles in Literature, Linguistics, and Sciences, which evidenced that FS titles dominantly occurred in research papers related to Biology. Among six FS titles identified in this study, five came from three different Biology journals but only one from a Psychology journal. Soler (2007) supported Haggan's argument and observed 92 instances of FS titles in sampled journals, with 13 from Medicine, 41 for Biology, 37 for Biochemistry, but only 1 for Anthropology. No FS titles emerged from Linguistics and Psychology. Milojević (2017) discovered that FS titles came into existence in the journals in Astronomy, Ecology, Economics, Mathematics, and Robotics since the middle of the 1990s. The instances of FS titles observed in the literature above suggest that they were preferably used in hard sciences rather than in soft sciences. In the last decade, the definition of FS has been interpreted differently in subsequent studies. The conclusive, declarative FS titles are continued to be observed as a feature of science papers (Moattarian & Alibabaee, 2015; Nagano, 2015; Salager-Meyer et al., 2013; Soler, 2011). At the same time, FS titles have been "expanded" to a broader scope that includes interrogative sentences and clauses (Archibald, 2017; Cheng et al. 2012; Morales et al., 2020). This study will follow the line of research discussed above and take all types of FS titles into consideration.

What Makes the Library and Information Science (LIS) Special?

The existing studies, which target LIS article titles, were largely conducted from the perspective of classification, citation, indexing, and information retrieval (Ávila-Argüelles et al., 2010; Adams, 1967; Arsenault & Ménard, 2011; Jahoda & Stursa, 1969; Maiti & Dutta, 2013; O'Connor, 1964). Lexical density and syntactic structure of research article titles published in the journals of LIS have never been researched specifically.

The appellation of *Library and Information Sciences* seems to suggest that this field is composed of two branches: Library Science and Information Science. However, Milojević et al. (2011)'s cognitive, co-word analysis revealed that Library and Information Science is actually formed by three branches: Library Science, Information Science, and Bibliometrics and Scientometrics Sciences (Hereafter, Scientometrics will be used to cover both bibliometrics and scientometrics). This study indicated that the traditionally-recognized Library Science is considered as a *softer* area which includes the studies of librarianship, services, policy, and publishing. Scientometrics, which deals with performance assessment, author productivity, citation studies, and metric analysis, is recognized as a *harder* field. Hence, it leads to a logical question: If Milojević, Sugimoto, Yan, and Ding's argument about Library Science as a softer field and Scientometrics as a harder one is examined under the lens of lexical density and syntactic structure, would there be any disciplinary

differences between these two fields? In light of literature on linguistic models and disciplinary differences, this study attempts to answer the following questions:

1. What are the average research article title lengths for Library Science and Scientometrics? Does Library Science tend to have shorter titles than Scientometrics?
2. Could the lexical density mark a disciplinary difference between Library Science and Scientometrics?
3. Could the usage of punctuation marks outline a disciplinary difference between Library Science and Scientometrics?
4. Is NP, as a title structure, prevalently used in both Library Science and Scientometrics, or one has more usage than the other?
5. Could the declarative FS title structure, which was preferably used in hard sciences in the literature discussed above, be possibly used in both Library Science and Scientometrics, or just used in one of them?

Methods

Selection of journals and articles

Journals used in this research are selected from the list compiled by Milojević et al. (2011) (p. 1936). This list was built on the recommendation of directors of the American Research Libraries (ARL) and deans of LIS programs accredited by the American Library Association (ALA). After Information Science journals and journals that cover both Library Science and Information Science were removed from this list, six pure Library Science journals were retained, including *College and Research Libraries*, *Journal of Academic Librarianship*, *Library Quarterly*, *Library Resources and Technical Services*, *Reference & User Services Quarterly*, and *Library Trends*. *College and Research Libraries* covers both academic libraries and research libraries; hence, *Journal of Academic Librarianship* was not selected for this study. The researcher is working in a technical services librarian's position; therefore, *Library Resources and Technical Services* was not selected so as to eliminate personal favor. Finally, this study selected four journals to represent Library Science, which are *College & Research Libraries, Library Trends, Library Quarterly,* and *Reference & User Services Quarterly*. *Scientometrics* was the only journal on the original list; therefore, it was inherited in this study to represent Scientometrics.

For the purpose of this study, research article titles were taken from an individual journal's website. Research articles technically refer to the publications usually aggregated under the section termed as "Articles" or "Features," or papers individually labeled as "Original Paper." Therefore, articles published under "Announcements," "Annual Reports," "Bibliographies," "Brief Communications," "Book Reviews," "Columns," "Correspondence," "Notes," and "Perspective," are not included in this study. As for special bilingual issues, only articles and titles written in English were considered for data collection.

Corpus of the study

The author went to the homepage of each journal and copied and pasted article information in a spreadsheet, which was coded as Journal Title, Year, Volume, Issue, and Article Title. The text corpus in this study consisted of a total number of 690 research article titles,

(See Table 1). Library Science includes 345 titles, spanning from 2017 and 2019. 145 titles come from *College & Research Libraries*, 65 from *Library Quarterly*, 99 from *Library Trends*, and 36 from *Reference & User Services Quarterly*. Scientometrics includes 345 titles from the journal *Scientometrics*, ranging from 2018 to 2019. It is easy to see that the number of articles that *Scientometrics* produced within 2 years is equivalent to what four library journals put together in 3 years. Apparently, *Scientometrics* is a highly productive journal, attracting more scholars' attention.

Data analysis

A total number of 690 research article titles were collected from individual journal's official website and coded in an Excel spreadsheet for further analysis. To ensure the reliability of this study, data was examined, analyzed, and then reviewed twice at different points of time by the author. Titles in question were picked out and native English speakers with backgrounds in literature and linguistics were consulted.

Each title was first measured by calculating the length, namely the number of words. Title length was counted typographically, not semantically. This means the concept of word is defined as a string of letters occurring between spaces or punctuations marks. By such a definition, an abbreviations (both capitalized and uncapitalized) was counted as one word and a hyphenated compound as multiple individual words (See example one and two below).

1. Access provision for sight impaired students (SISs) in Nigerian University Libraries (11 words)
2. The Brazilian academic genealogy: Evidence of advisor–advisee relationships through quantitative analysis (12 words)

In order to calculate types and numbers of punctuation marks, the corpus of titular texts were copied and pasted into separate Microsoft Word documents so as to take advantage of the search function keys *Ctrl + F*. Individual punctuation marks were typed in the search box and the total number of punctuation marks was given after *Highlight All* was selected. Punctuation marks identified in this study include colon, comma, hyphen, apostrophe, quotation marks, question mark, period, parentheses, exclamation point, and dash.

The informativeness of titles was measured by counting the lexical words. Lexical words refer to ones that have meanings, namely nouns, verbs, adjectives, and adverbs. Function words are ones that bind text together, such as articles, conjunctions, and prepositions. Lexical density, an indicator of text informativeness, is the ratio of lexical words to

Table 1 Information about the title corpus

	Journals	Titles from individual journal *n*	Total number in each discipline *n*
Library Science	College & Research Libraries	145	
	Library Quarterly	65	
	Library Trends	99	345
	Reference & User Services Quarterly	36	
Scientometrics	Scientometrics	345	345
Total		690	690

the total number of words. In this study, the corpus consisting of lexical words was analyzed by taking a bottom-up approach. Each lexical word was coded by its nature and then classified into nine broad categories, including Topic, Research, Context, Domain, Action, Spatial, Temporal, Numeric, and Others (See example three and four below). Topic refers to the matter that research deals with, such as *resource sharing*, *research trends*, or *journal choice*. Research alludes research sample, process, methods, or results, for instance, *effect*, *comparison*, or *altmetrics analysis*. Context is the setting where the research was conducted, for example, *public library system* or *open access*. Domain is considered as the area the research points to, such as *LIS education* or *blockchain study*. Action refers to words that described doing something, for example, *investigation, mapping* or *predict*. Spatial contains words indicating space, which could be either explicit (*China* or *Fukushima*) or vague (*regional* or *national*). Temporal includes words relating to time, which could be specific (*1992* or *1932*) or ambiguous (*digital era* or *decades*). The rest is grouped as Others, which includes, but is not limited to quotations, metaphors, and rhetoric sentences. The categorizations are personal interpretation, which is subject to criticism. At the semantic level, words could mean both a research topic and method, and clear boundaries between context and domain are difficult to define, too.

> 3. Four decades of fuzzy sets theory in operations management: Application of life-cycle, bibliometrics and content analysis (Topic: fuzzy sets theory; Domain: operation management; Research: life cycle, bibliometrics and content analysis; Action: application; Temporal: four decades.
> 4. Don't call it a comeback: Popular reading collections in academic libraries (Topic: popular reading collections; Context: academic libraries; Others: don't call it a comeback)

The syntactic structure of titles was analyzed by taking a top-down approach. First, the whole titles were classified into three broad groups: single-unit group, two-unit group, and three unit group. Four-unit group, as Gesuato (2008) observed, did not occur in the collected data. Single-unit group means titles embody syntactic wholeness as phrases or sentences, including NP and FS. The two-unit and three-unit groups are categorized by NP's coordination with adjacent phrases, which include V-ing Phrase (VP), Propositional Phrase (PP), and FS. The following titles (See example 5–14) serve as illustration of NP coordination in single-unit, two-unit, and three-unit title groups:

> 5. Is the library's online orientation program effective with English language learners? (single unit; FS)
> 6. A hybrid approach to detecting technological recombination based on text mining and patent network analysis (single unit; NP)
> 7. Disability, the silent D in diversity (two unit; NP + NP)
> 8. The ISSAS model: Understanding the information needs of sexual assault survivors on college campuses (two unit; NP + VP)
> 9. Antisemitism and Islamophobia: What does a bibliometric study reveal? (two unit; NP + FS)
> 10. Twenty years of statistical learning: From language, back to machine learning (two unit; NP + PP)
> 11. Negotiating borders: Librarianship and twenty-first-century politics (two unit; VP + NP)
> 12. Who reads international Egyptian academic articles? An altmetrics analysis of Mendeley readership categories (two unit; FS + NP)

13. On the bibliometric nature of a foreseeable relationship: Open access and education (two unit; PP + NP)

14. Software survey: ScientoPy, a scientometric tool for topics trend analysis in scientific publications (three unit, NP + NP + NP)

Results and discussion

Title length

As is shown in Table 2, the results of the two-independent Welch t-test demonstrates that the difference of title length between Library Science (M = 12.83, SD = 4.28) and Scientometrics (M = 12.72, SD = 4.41) at the 0.05 level of significance (t = −0.0.33, *df* = 687.32, p > 0.05) is not statistically significant. Therefore, the null hypothesis that titles in Library Science tend to be shorter than ones in Scientometrics is rejected. The results suggest that both Library Science and Scientometrics have the equivalent title length, indicating that there is no disciplinary difference between them. 12.83 words in Library Science and 12.72 words in Scientometrics fall below 14.15–15.48 words, which is the average range of numbers of research article titles in Biology, Medicine, and Biochemistry discovered in Soler (2007). However, title lengths in Library Science and Scientometrics are more or less close to Psychology (12.63 words) in Nagano (2015) or Business (12.88 words) in Appiah et al. (2019). Therefore, both Library Science and Scientometrics fall in the softer science side in terms of title length. Whether the phenomenon of concise titles is positively influenced by the instructions for authors outlined by journals needs a separate research with a large number of journal samples. At least, in this study, *Reference & User Services Quarterly* clearly states "give the article a brief title" and *Scientometrics* requires "the title should be concise and informative," in their author guidelines. Although *College and Research Libraries* does not give a specific instruction on article titles, its author guidelines recommend that "clear, simple prose enhances the presentation of ideas and opinions." Apparently, this recommendation also applies to titles because they are the opening but overarching text of articles, where authors' fundamental ideas and opinions lie.

Lexical density and lexical words

Lexical density is measured by the ratio of lexical words to the total number of words (See Table 3). Library Science and Scientometrics have a total number of 4424 and 4428 words respectively. Library Science has 3152 lexical words (9.14 words per title) and 1272 function words (3.69 words per title). Scientometrics has a total number of 4428 words, which

Table 2 Results of descriptive statistics and welch t-test for title length

	Title length						
	Shortest	*Longest*	*M*	*SD*	*n*	t	*df*
Library Science	3	29	12.83	4.28	345	−0.33	687.32
Scientometrics	3	31	12.72	4.41	345		

*p > .05

Table 3 Lexical density

	Total words *n*	Lexical words *n*	Lexical words per title *n*	Function words *n*	Function words per title *n*	Lexical density %
Library Science	4424	3152	9.14	1272	3.69	71.25
Scientometrics	4428	3101	8.99	1327	3.85	70.03

are made up of 3101 lexical words (8.99 words per title) and 1327 function words (3.85 words per title). Library Science has a 71.25% lexical density and Scientometrics has a 70.03%. Therefore, Library Science and Scientometrics demonstrate almost equal value in lexical density, total lexical words, and lexical words per title.

Lexical words were further analyzed as a separate category since they were a reflection of title informativeness in various areas (See Table 4). Library Science carries more weight in words related to Topic (1232 word and 39.09% in Library Science; 978 words and 31.54% in Scientometrics) and Context (470 words and 14.91% in Library Science; 300 words and 9.67% in Scientometrics). This finding concurs with Milojević et al. (2011)'s argument that Library Science's topics contain information retrieval, web search, catalogs, and databases in the context of academic librarianship, public librarianship, information literacy, school librarianship, and policy, etc. Both topics and contexts require more description and elaboration, leading to bigger number of words. However, Scientometrics has a considerably higher usage of words related to Research (436 words and 14.06% in Scientometrics; 340 words and 10.79% in Library Science), Domain (806 words and 26.00% in Scientometrics; 537 words and 17.04% in Library Science), and Spatial (151 words and 4.87% in Scientometrics; 63 words and 2.00% in Library Science). If it is the involvement of research related words that help the brief titles generate an impression that articles would carry concrete scientific evidence and credibility, this category of words merits further analysis (See Table 5). Instead of counting the number of individual words, research related words were

Table 4 Categories of lexical words

Type	Library Science		Scientometrics	
	n	%	*n*	%
Topic	1232	39.09	978	31.54
Research	340	10.79	436	14.06
Context	470	14.91	300	9.67
Domain	537	17.04	806	26.00
Action	234	7.42	231	7.45
Spatial	63	2.00	151	4.87
Temporal	58	1.84	61	1.97
Numeric	14	0.44	15	0.48
Other	204	6.47	123	3.97
Total	3152	100	3101	100

Table 5 Types of research related

Type	Library Science		Scientometrics	
	n	*%*	*n*	*%*
Research General	38	11.01	22	6.38
Research Method	83	24.06	137	39.71
Research Result	44	12.75	53	15.36
Total (Research)	165	47.82	212	61.45
Total (Other)	180	52.18	133	38.55
Total	345	100	345	100

further examined by their semantic meaning. Words like *review*, *study*, *analysis*, and *exploration* suggest a general research, which is 38 titles (11.01%) in Library Science and 22 titles (6.38%) in Scientometrics. *Case study*, *bibliometric analysis*, and *systematic review* indicate an involvement of a specific research method, which comprises of only 83 titles (24.06%) in Library Science but 137 titles (39.71%) in Scientometrics. *Impact*, *relationship*, and *factors* imply research results, which is 44 titles (12.75%) in Library Science and 53 titles (15.36%) in Scientometrics. In addition, 180 titles (52.18%) in Library Science do not have research related words in them. Only 133 titles (38.55%) in Scientometrics belong to this category.

Appiah et al. (2019) considered the general, research-related expressions, such as *investigation of*, *study of*, or *observation on*, as ineffective content words in titles. They believed that those words indicating research in general make lengthy titles and create ambiguity and redundancy. They argued that the general expression should be avoided in title construction, especially in science. Salager-Meyer et al. (2013) pointed out that "the more precise and accurate the title is, the easier it is for bibliographers to compile data for indexing, abstracting and other documentation purposes" (p. 258). Haggan (2004) also specified that titles for scientific papers should have "an up-front, straightforward presentation of information" (p. 313). Therefore, when article titles are stated with more clarity and specificity regarding what methods are involved and what results come out, they will have more chances to be effectively classified and indexed in the system by indexers and bibliographers. The involvement of research methods and results will increase the probability that articles will be more easily identified and selected by users due to their research-driven demeanor and scientific relevance.

Library Science and Scientometrics contain similar number of words related to Action (234 words and 7.42% in Library Science; 231 words and 7.45% in Scientometrics), Temporal (58 words and 1.84% in Library Science; 61 words and 1.97% in Scientometrics), and Numeric (14 words and 0.44% in Library Science; 15 words and 0.48% in Scientometrics). The rest of words are categorized in Others (204 words and 6.47% in Library Science; 123 words and 3.97% in Scientometrics). Considering the slight difference in lexical words (3152 in Library Science vs. 3101 in Scientometrics) and lexical density (71.25% in Library Science vs. 70.03% in Scientometrics), substantive word rate cannot be used to draw a line that defines Library Science as a softer science and Scientometrics as a harder one. The striking finding is that, in contrast to Library Science, Scientometrics titles contain much more substantive words to indicate specific research methods, which enhance the articles' scientific outlook.

Table 6 Titles and punctuation marks

	Library Science		Scientometrics	
	n	*%*	*n*	*%*
Titles with punctuation	293	84.93	250	72.46
Titles without punctuation	52	15.07	95	27.54
Total	345	100	345	100

Table 7 Usage of punctuation marks

Punctuation marks	Library Science		Scientometrics	
	n	*%*	*n*	*%*
Colon	230	42.67	159	37.95
Comma	91	16.88	48	11.46
Hyphen	72	13.36	104	24.82
Apostrophe	63	11.69	33	7.88
Question mark	35	6.49	38	9.07
Quotation marks	22	4.08	14	3.34
Period	13	2.41	4	0.95
Parentheses	7	1.30	17	4.06
Exclamation point	3	0.56	0	0.00
Dash	3	0.56	2	0.48
Total	539	100	419	100

Punctuation marks

Table 6 offers an overview of the usage of punctuation marks: Library Science has 293 titles (84.93%) with punctuation and 52 titles (15.07%) without; Scientometrics has 250 titles (72.46%) that use punctuation and 95 titles (27.54%) that do not use punctuation. Overall, a considerably higher number of titles in Library Science use punctuation marks than Scientometrics. Punctuation marks are used in research article titles to coordinate structures, negotiate text space, and express authors' intention and emotions. The usage of punctuation mark is an indication of titular complexity. In terms of overall percentage of using punctuation marks, Library Science outshines Scientometrics without question.

Specifically speaking, ten punctuation marks were identified from the title corpus, which are colon, comma, hyphen, apostrophe, question mark, quotation mark, period, parentheses, exclamation point, and dash (see Table 7). In comparison to Scientometrics, Library Science has a considerably higher frequency of using colons (230 titles and 42.67% in Library Science; 159 titles and 37.95% in Scientometrics), commas (91 titles and 16.88% in Library Science; 48 titles and 11.46% in Scientometrics), apostrophes (63 titles and 11.69% in Library Science; 33 titles and 7.88% in Scientometrics), quotation marks (22 titles and 4.08% in Library Science; 14 titles and 3.34% in Scientometrics), and exclamation points (3 titles and 0.56% in Library Science; none in Scientometrics). Library Science significantly surpasses Scientometrics on the usage of a number of punctuation marks, in particular colons. If Dillon (1981, 1982)'s and Perry (1985)'s arguments, which stated that colonic titles were an in indicator of scholarly productivity and intelligent distinction, were

still effective and convincing, then the result seems to suggest that research article titles in Library Science display more scholarly outlook than Scientometrics. However, after Perry's empirical support in 1985, Dillon's hypothesis about correlation between colonic titles and scholarly productivity was rarely tested or pursued over years. Perhaps colons are the easiest and the most common way to construct multi-unit titles, so that titles could offer authors the capacity to package more information across the disciplinary difference, either soft sciences or hard sciences.

In addition to the similar frequency of using dashes, Scientometrics tends to have fairly more usage of hyphens (104 titles and 24.82% in Scientometrics; 72 titles and 13.36% in Library Science) and parentheses (17 titles and 4.06% in Scientometrics; 7 titles and 1.30% in Library Science). Hyphens are joiners, which combine different words together to indicate a new meaning, for instance, "advisor–advisee relationships" in example two. Hyphens are most commonly used in the situation that Scientometrics authors are in a need to create a compounded new word that may not exist in the dictionary. Parentheses are wrappers, which enclose abbreviated information in titles to represent the whole phrases, for instance "sight impaired students (SISs)" in example one. More use of hyphens and parentheses could be interpreted as an indicator of lexical complexity, which means Scientometrics authors are more frequently engaged in the circumstances to meet emerging language needs through creating new compounds, or save text space and avoid redundant and lengthy repetition by using parentheses for abbreviations.

In summary, punctuation complexity marks a disciplinary difference between Library Science and Scientometrics in terms of the overall usage. Particularly, Library Science outweighs Scientometrics in the use of colons; however, Scientometrics does demonstrate a preference for hyphens and parentheses.

NP in Single-Unit, Two-Unit, and Three-Unit Titles

Table 8 shows the complexity of NP coordination in single-unit, two-unit, and three-unit titles. NP is semantically coordinated together with other NP, VP, FS, and PP, either at the beginning, middle, or rear position. The striking finding is that NP enjoys the overall prevalence and dominance in the whole title corpus (287 titles and 83.19% in Library Science; 286 and 82.90% in BBS).

Library Science demonstrates slightly higher numbers in a few NP coordination types. Specifically speaking, in terms of the two-unit structure, Library Science demonstrate a little bigger number of NPs than Scientometrics titles in the title coordination of NP + NP (107 titles and 31.01% in Library Science; 103 titles and 29.86% in Scientometrics), NP + VP (29 titles and 8.41% in Library Science; 8 titles and 2.32% in Scientometrics), NP + PP (4 titles and 1.16% in Library Science; 3 titles and 0.87% in Scientometrics), VP + NP (46 titles and 13.33% in Library Science; 28 titles and 8.12% in Scientometrics), FS + NP (26 titles and 7.54% in Library Science; 11 titles and 3.19% in Scientometrics), and PP + NP (8 titles and 2.32% in Library Science; 5 titles in 1.45% in Scientometrics). However, there is one NP coordination type, which is NP (See example six) in the single-unit titles, marks a big, contrastive disciplinary disparity. As is shown in Table 8, 116 (33.62%) single-unit titles in Scientometrics take a single NP to lead the titular sentences, which is approximately twice as many as that in Library Science (59 titles and 17.10%)!

The above results provide a strong evidence that Library Science and Scientometrics in general favor various NP coordinations as the dominate way to construct titles, reaffirming the finding discovered in Busch-Lauer (2000) and Fortanet et al. (1998)

Table 8 NP in single-unit, two-unit and three-unit titles

Syntactic structure	Coordination	Library Science		Scientometrics	
		n	%	n	%
Single-unit	NP	59	17.10	116	33.62
Two-unit	NP + NP	107	31.01	103	29.86
	NP + VP	29	8.41	8	2.32
	NP + FS	6	1.74	7	2.03
	NP + PP	4	1.16	3	0.87
	VP + NP	46	13.33	28	8.12
	FS + NP	26	7.54	11	3.19
	PP + NP	8	2.32	5	1.45
Three-unit	NP + NP + NP	0	0	3	0.87
	FS + NP + NP	2	0.58	0	0
	NP + FS + NP	0	0	1	0.29
	FS + NP + NP	0	0	1	0.29
Total (NP)		287	83.19	286	82.90
Total (Other)		58	16.81	59	17.10
Total (Title)		345	100	345	100

that nominal phrase titles prevalently occur across various disciplines. However, the contrast of NP as single-unit titles between Library Science and Scientometrics merits further discussion. Gómez, Gómez, García, and Silveira (1998) observed a disciplinary variation that more usage of single nominal heads on the harder sciences side (Chemistry and Computer Science) than the softer sciences side (Linguistics and Business/Economics). Wang and Bai (2007) observed single head nominal groups were used in medical research article titles more frequently than bi-head nominal groups and multi-head nominal groups.

As is shown in the example six, this type of title structure is made up of a noun(s) as the head(s) leading the sentence, with appropriate modifier(s) before and/or after. Theoretically, the grammatically centered head may not necessarily mean that the head should be positioned in the middle of the whole title. Either the nominal head is put in the middle, the front, or the rear of titles, the position does not decrease its articulation of a concentrated semantic expression. Wherever it is located, the nominal head could be supported by a variety of pre and post-modifiers to deliver the key information to users what this article is about. Empirically, Wang and Bai (2007) elaborated the grammatical capability of how information is packaged through prepositional phrases, to-infinitive clauses, past participles, and present participle clauses. The comprehensive grammatical analysis provides practical implications of how effective titles could be constructed for authors who were engaged in medical research, practice, and learning. However, the diversity of pre and post-modifiers closely tied to nominal heads are not clear in this study. In light of the theoretical elaboration and Wang and Bai's practical suggestion, the structure, grammatical components, and functions of modifiers in the nominal heads of single-unit titles, which were not explored in this research, calls for a future study.

Declarative FS in Single-Unit Titles

Overall, four types of FS in single-unit titles are identified in this study: Interrogative (2 titles and 0.58% in Library Science; 11 titles and 3.19% in Scientometrics), Declarative (1 title and 0.29% in Library Science; 5 titles and 1.45% in Scientometrics), Imperative (none in Library Science; 1 title and 0.29% in Scientometrics), and Clause (none in Library Science; 2 titles and 0.58% in Scientometrics) (See Table 9). Declarative FS in single-unit titles, which are believed as titular notation in hard sciences, as a matter of a fact, do exist in the title corpus of both Library Science (1 title) and Scientometrics (5 titles). Scientometrics only has four more titles; however, this small difference is even more significant if the titles' rarity is considered in the whole corpus.

15. Is science driven by principal investigators? (Interrogative; Scientometrics)
16. Revitalizing scholarly reference for digital research requires a redoubled commitment to quality and community (Declarative; Library Science)
17. The author's ignorance on the publication fees is a source of power for publishers (Declarative; Scientometrics)
18. Cited text spans identification with an improved balanced ensemble model (Declarative; Scientometrics)
19. Measures of linear type lead to a characterization of Zipf functions (Declarative; Scientometrics)
20. The open access citation premium may depend on the openness and inclusiveness of the indexing database, but the relationship is controversial because it is ambiguous where the open access boundary lies (Declarative; Scientometrics)
21. Few research fields play major role in interdisciplinary grant success (Declarative; Scientometrics)
22. Re-examine the determinants of market value from the perspectives of patent analysis and patent litigation (Imperative; Scientometrics)
23. How to measure the performance of a Collaborative Research Center (Clause; Scientometrics)

The current literature shows that a fairly small number of declarative FS in the single-unit research article titles were dominantly used by research articles in hard sciences, such as Biology and Medicine. The result of this study expands such evidences into the

Table 9 Full sentence in single-unite titles

Type	Library Science		Scientometrics	
	n	*%*	*n*	*%*
FS				
Interrogative	2	0.58	11	3.19
Declarative	1	0.29	5	1.45
Imperative	0	0	1	0.29
Clause	0	0	2	0.58
Total (FS)	3	0.87	19	5.51
Total (Other)	342	99.13	326	94.49
Total	345	100	345	100

field of LIS, in particular its branch Library Science and Scientometrics. Using declarative sentences as titles is a very special phenomenon in the literature. Declarative FS titles are interpreted as a feature of scientific papers in many ways. First, this conclusive, self-reporting title type helps authors of scientific studies to deliver a pragmatic, non-flirtatious, authoritative demeanor with assurance about approaches or results. Secondly, by taking this title structure, the whole research results are clearly summarized and delivered to readers in condensed sentences, which "make confident, unqualified assertions, presented as statements of fact" (Haggan, 2004, p. 296). Haggan further noted that the use of the simple present tense in declarative FS may not be given equivalent status to attention-grabbing news headlines because they frequently omit articles and the verb "to be." However, it does underscore "the note of optimism being projected by the writer that what he is reporting stands true for all time or is not simply a one-off occurrence" (p. 297). The advantage of using such title structure is that it advocates the statement as a fact; the downside is that the attempting leaves no room for other possibilities or rejects a need of elaboration. Therefore, Soler (2007) warned that presenting results in an assertive way in full sentences could lead to the research seen as attenuated evidentials.

Conclusion

Through the analysis of title length, lexical density, punctuation marks, and syntactic structure of research article titles, this study attempted to identify whether disciplinary differences existing between soft sciences and hard sciences could also be found between Library Science and Scientometrics. The findings reveal that both Library Science and Scientometrics have equivalent title length. However, between Library Science and Scientometrics, there does exist interesting disciplinary differences in some elements of lexical density, punctuation marks, and title types. Findings can be concluded as below:

1. Both Library Science and Scientometrics titles demonstrate similar lexical density in terms of lexical words, function words, lexical words per title, and function words per title. However, Scientometrics titles contain much more lexical words regarding specific research methods involved in the articles, which makes it on the hard science side. The usage of lexical words stating research methods could be considered as an indicator to instantly evaluate whether a research paper has scientific value or not at the first sight. But, a proper caution should also be taken when whether a paper is scientific or not is only judged by its title without further examining the content.
2. Overall, Library Science demonstrates punctuation complexity in terms of total number of punctuation marks employed. Library Science has much more use of colons, however, which were once considered as a symbol of scholar productivity in literature published in the1970s and the 1980s only. More involvement of hyphens and parentheses in Scientometrics suggest its lexical complexity and authors' need of negotiation for new meaning and space.
3. Although NP is overall dominantly used to govern the structural coordination of titles in both fields, Scientometrics has twice as many of NPs in single-unit titles as Library Science. This finding suggests Scientometrics titles are more likely to have a whole, concise, non-broken syntactic structure. This finding also correspond with more usage of colons in Library Science, leading to broken, multi-unit title structure. Future studies need to further investigate what types of pre and post modifiers are specifically involved

in the titles of both fields, which will help generate a full grammatical and semantic picture.

4. Instances of the conclusive, declarative FS in single-unit titles, which are dominantly found in the hard sciences, are also evidenced in both Library Science and Scientometrics. However, Scientometrics has more instances in comparison to Library Science. In this aspect, it is safe to say Scientometrics is a harder science than Library Science.

In light of the above findings, it can be concluded that Library Science and Scientometrics demonstrate disciplinary differences in individual preference of lexical words, punctuation marks, and title types, even though both fields are nested under the same big umbrella of the Library and Information Science. Clearly, the title corpus is limited in a number of journals in Library Science and Scientometrics, which did not include the third branch Information Science. With the inclusion of Information Science in future studies, a comprehensive, full picture of lexical density and syntactic structure of the Library and Information Science will be captured.

References

Ávila-Argüelles, R., Calvo, H., Gelbukh, A., & Godoy-Calderón, S. (2010). Assigning Library of Congress Classification codes to books based only on their titles. *Informatica, 34*, 77–84

Adams, W. M. (1967). Relationship of keywords in titles to references cited. *American Documentation, 18*, 26–32

Anthony, L. (2001). Characteristic features of research article titles in computer science. *IEEE Transactions on Professional Communication, 44*, 187–194. https://doi.org/10.1109/47.946464

Appiah, K. R., Ankomah, C., Osei, H. Y., & Hattoh-Ahiaduvor, T. (2019). Structural organisation of research article titles: A comparative study of titles of business, gynaecology and law. *Advances in Language and Literary Studies, 10*(3), 145–154. https://doi.org/10.7575/aiac.alls.v.10n.3p.145

Archibald, A. J. B. (2017). A linguistic analysis of conference titles in Applied Linguistics. *International Journal of Foreign Language Teaching & Research, 5*(18), 11–25

Arsenault, C., & Ménard, E. (2011). Searching titles with initial articles in library catalogs: A case study and search behavior analysis. *Library Resources & Technical Services, 51*, 190–203. https://doi.org/10.5860/lrts.51n3.190

Baicchi, A. (2003). Relational complexity of titles and texts: A semiotic taxonomy. In L. Merlini Barbaresi (Ed.), *Complexity in language and text.* (pp. 319–341). Edizione Plus-Universidad de Pisa.

Berkenkotter, C., & Huckin, T. N. (1995). *Genre knowledge in disciplinary communication: Cognition, culture, power*. Lawrence Erlbaum Associates.

Busch-Lauer, I. A. (2000). Titles of English and German research papers in Medicine and Linguistics. *Analysing Professional Genres, 74*, 77–94

Buxton, A. B., & Meadows, A. J. (1977). The variation in the information content of titles of research papers with time and discipline. *Journal of Documentation, 33*, 46–52. https://doi.org/10.1108/eb026633

Casson, L. (2001). *Libraries in the ancient world*. Yale University Press.

Cheng, S. W., Kuo, C.-W., & Kuo, C.-H. (2012). Research article titles in Applied Linguistics. *Journal of Academic Language and Learning, 6*(1), A1–A14

Diers, D., & Downs, F. S. (1994). Colonizing: A measurement of the development of a profession. *Nursing Research, 43*, 316–318

Dillon, J. T. (1981). The emergence of the colon: An empirical correlate of scholarship. *American Psychologist, 36*, 879–884. https://doi.org/10.1037/0003-066X.36.8.879

Dillon, J. T. (1982). In pursuit of the colon: A century of scholarly progress: 1880–1980. *The Journal of Higher Education, 53*, 93–99. https://doi.org/10.1080/00221546.1982.11780427

Fortanet, I., Coll, J. F., Palmer, J. C., & Posteguillo, S. (1997). The writing of titles in academic research articles. In R. M. Chamorro & A. R. Navarrete (Eds.), *Lenguas aplicadas a las ciencias y la tecnología. Aproximaciones.* (pp. 155–158). Universidada de Extremadura, Servicio de Publicaciones.

Fortanet, I., Posteguillo, S., Coll, J. F., & Palmer, J. C. (1998). Linguistic analysis of research articles: Disciplinary variations. In I. Vazquez & I. Camilleu (Eds.), *Perspectivas praguietices en linguistica aplicada, zaragoza.* (pp. 443–447). Anubar Ediciones.

Gómez, I. F., Gómez, S. P., García, J. F. C., & Silveira, J. C. P. (1998). Linguistic analysis of research article titles: Disciplinary variations. In I. V. Orta & I. G. Galve (Eds.), *Perspectivas pragmáticas en lingüística aplicada.* (pp. 443–448). Anubar Ediciones.

Gesuato, S. (2008). Encoding of information in titles: Academic practices across four genres in linguistics. In C. Taylor (Ed.), *Ecolingua: The role of e-corpora in translation and language learning.* (pp. 127–157). EUT.

Goodman, R. A., Thacker, S. B., & Siegel, P. Z. (2001). What's in a title? A descriptive study of article titles in peer-reviewed medical journals. *Science Editor, 24*, 75–78

Haggan, M. (2004). Research paper titles in literature, linguistics and science: Dimensions of attraction. *Journal of Pragmatics, 36*, 293–317. https://doi.org/10.1016/S0378-2166(03)00090-0

Hartley, J. (2007). Planning that title: Practices and preferences for titles with colons in academic articles. *Library & Information Science Research, 29*, 553–568. https://doi.org/10.1016/j.lisr.2007.05.002

Jahoda, G., & Stursa, M. L. (1969). A comparison of a keyword from title index with a single access point per document alphabetic subject index. *American Documentation, 20*, 377–380. https://doi.org/10.1002/asi.4630200422

Lewison, G., & Hartley, J. (2005). What's in a title? Numbers of words and the presence of colons. *Scientometrics, 63*, 341–356. https://doi.org/10.1007/s11192-005-0216-0

Maiti, D. C., & Dutta, B. (2013). Comparative study between words in titles and keywords of some articles on knowledge organisation. *DESIDOC Journal of Library & Information Technology, 33*, 498–508

Michelson, G. (1994). Use of colons in titles and journal status in industrial relations journals. *Psychological Reports, 74*, 657–658. https://doi.org/10.2466/pr0.1994.74.2.657

Milojević, S. (2017). The length and semantic structure of article titles—Evolving disciplinary practices and correlations with impact. *Frontiers in Research Metrics and Analytics, 2*(2), 1. https://doi.org/10.3389/frma.2017.00002

Milojević, S., Sugimoto, C. R., Yan, E., & Ding, Y. (2011). The cognitive structure of library and information science: Analysis of article title words. *Journal of the American Society for Information Science and Technology, 62*, 1933–1953. https://doi.org/10.1002/asi.21602

Moattarian, A., & Alibabaee, A. (2015). Syntactic structures in research article titles from three different disciplines: Applied linguistics, Civil Engineering, and Dentistry. *Journal of Teaching Language Skills, 34*, 27–50

Morales, O. A., Perdomo, B., Cassany, D., Tovar, R. M., & Izarra, É. (2020). Linguistic structures and functions of thesis and dissertation titles in Dentistry. *Lebende Sprachen, 65*, 49–73. https://doi.org/10.1515/les-2020-0003

Nagano, R. L. (2015). Research article titles and disciplinary conventions: A corpus study of eight disciplines. *Journal of Academic Writing, 5*(1), 133–144. https://doi.org/10.18552/joaw.v5i1.168

O'Connor, J. (1964). Correlation of indexing headings and title words in three medical indexing systems. *American Documentation, 15*, 96–104. https://doi.org/10.1002/asi.5090150207

Perry, J. A. (1985). The Dillion hypothesis of titular colonicity: An empirical test from the ecological sciences. *Journal of the American Society for Information Science, 36*, 251–258. https://doi.org/10.1002/asi.4630360405

Sahragard, R., & Meihami, H. (2016). A diachronic study on the information provided by the research titles of Applied Linguistics journals. *Scientometrics, 108*, 1315–1331. https://doi.org/10.1007/s11192-016-2049-4

Salager-Meyer, F., Ariza, M. A. A., & Briceño, M. L. (2013). Titling and authorship practices in medical case reports: A diachronic study. *Communication & Medicine, 10*(1), 63–80. https://doi.org/10.1371/journal.pone.0197775

Soler, V. (2007). Writing titles in science: An exploratory study. *English for Specific Purposes, 26*, 90–102. https://doi.org/10.1016/j.esp.2006.08.001

Soler, V. (2011). Comparative and contrastive observations on scientific titles written in English and Spanish. *English for Specific Purposes, 30*, 124–137. https://doi.org/10.1016/j.esp.2010.09.002

Swales, J. (1990). *Genre analysis: English in academic and research setting*. Cambridge: Cambridge University Press.

Slougui, D. (2018). Dissertation titles in EFL and UK-based contexts: How much do they differ? *ASp, 74*, 135–161. https://doi.org/10.4000/asp.5466

Trosborg, A. (2000). Introduction. In A. Trosborg & J. Benjamins (Eds.), *Analysing professional genres.* (pp. vii–xvi). ProQuest Ebook Central.

Wang, Y., & Bai, Y. (2007). A corpus-based syntactic study of medical research article titles. *System, 35*, 388–399. https://doi.org/10.1016/j.system.2007.01.005

Xie, S. (2020). English research article titles: Cultural and disciplinary perspectives. *SAGE Open, 10*(2), 1–12. https://doi.org/10.1177/2158244020933614

Yakhontova, T. (2002). *Titles of conference presentation abstracts: A cross-cultural perspective*. Retrieved on October 9, 2020 from, https://www.academia.edu/36296603/Titles_of_conference_presentations_abstracts_A_cross_cultural_perspective.

Yitzhaki, M. (1997). Variation in informativity of titles of research papers in selected humanities journals: A comparative study. *Scientometrics, 38*, 219–229. https://doi.org/10.1007/BF02457410

Ziebland, S., & Pope, C. (1995, September/October) Use of the colon in titles of British Medical Sociology Conference papers, 1970 to 1993. *Annals of Improbable Research*, 7–9.