



# A diachronic comparative study of research article titles in linguistics and literature journals

Xuechun Xiang<sup>1</sup> · Jing Li<sup>1</sup>

Received: 19 February 2019 / Published online: 14 December 2019  
© Akadémiai Kiadó, Budapest, Hungary 2019

## Abstract

By investigating a total number of 1364 article titles extracted from seven SSCI-indexed linguistics journals and seven A&HCI-indexed literature journals between 1980 and 2018, this study attempts to present the prominent form features and content attributes of the titles and their diachronic variations, as well as the similarities or differences between the two disciplines regarding title features. Results show that article titles in both subject domains demonstrate consistency in the long-range trend of title length and informativity, enjoying a steady and linear growth in length and an increase with fluctuations in informativity. Nominal and compound structures are both frequently adopted in linguistics and literature journals, but the proportion of compound constructions has always been higher than that of nominal ones in literature journals over the past 39 years. Titles containing topical information enjoy the highest percentage from the 1980s to the 2010s in both linguistics and literature journals. However, in the recent decade, content information provided by linguistic article titles tends to be more diversified, presenting not only research topics, but also results, dataset and methods of the study. No significant variations between the two disciplines have been found in lexical diversity, both of which experienced an overall downward trend.

**Keywords** Titles · Diachronic approach · Linguistics · Literature

## Introduction

Titles, although they are the smallest composition of academic publications, play a critical role in bringing about desirable communication between the writer and the potential reader (Haggan 2004; Ball 2009; Cheng et al. 2012). The significance of titles has been well recognized. They serve as an “eye catcher” and may possibly influence the reader’s decision about whether to continue reading, or that of the editor about whether to consider for publication. In some situations, titles, especially those of articles in scientific journals, are self-explanatory and mirror a set of requisites which are crucial to the construction,

---

✉ Jing Li  
lijing@hust.edu.cn

<sup>1</sup> School of Foreign Languages, Huazhong University of Science and Technology, 1037 Luoyu Rd., Wuhan 430074, China

communication, and progress of new knowledge (Soler 2007). Besides, a well-phrased title enables an article to be more “findable” by incorporating keywords and index terms required in online databases (Fox and Burns 2015). The study of titles was firstly termed in French by Duchet (1973) as *titrologie*. (cited in Salager-Meyer and Alcazar-Ariza 2013, p. 298). It was later formally introduced in English as “titleology” by Baicchi (2003). He further elaborated the relational complexity of titles and texts through a semiotic taxonomy and declared “titleology” as a new field of study. Early title research during the 1970s mainly concentrated on written literary texts (Soler 2011; Salager-Meyer and Alcazar-Ariza 2013) and barely extended to other domains (Netterville and Hirsch 1958; Wilsmore 1987). However, scholarly attention began to be paid to titles of academic publications in 1990 when Swales mentioned in his book *Genre analysis: English in academic and research settings* that titles in academic genres were not fully addressed and were worth being further explored. Since then, extensive research has been conducted on academic English (Swales 1990; Yitzhaki 1994, 1997), and specifically on English titles in various academic genres. For example, titles of medical literature, including medical case reports, were broadly investigated (e.g. Wang and Bai 2007; Jacques and Sebire 2009; Habibzadeh and Yadollahie 2010; Salager-Meyer and Alcazar-Ariza 2013). He (2000) examined the information transfer in the translation of medical article titles between English and Chinese. Furthermore, conference paper titles have also been studied (Afful 2017). Within the academic genres, unsurprisingly, journal article titles have received the most attention among researchers in recent years (Ball 2009; Guo et al. 2015). Title research has so far been conducted in a wide range of disciplines. However, while some disciplines, such as medicine and engineering, have been adequately explored, others like linguistics and literature, still call for further investigation. Therefore, the present study intends to examine the long-range trend of the past four decades and discipline variations, concerning title features of research articles published in linguistics and literature journals.

## Literature review

Title research has covered both form features and content attributes of journal article titles. Form features are generally classified into three types—title length, syntactic structure and the use of punctuation marks, while content attributes include prominent information types provided in titles, title informativity and lexical diversity.

Previous research on title length has mainly focused on the variations in title length over time (Whissell 2013) and the correlations between title length and other title features, such as article length (Yitzhaki 2002), author numbers (Kuch 1978; White 1991; Yitzhaki 1994) and citation rate (Nair and Gibbert 2016; Gnewuch and Wohlrabe 2017). All diachronic studies of title length seemed to conclude that titles had become increasingly longer with time (Yitzhaki 1994; Méndez et al. 2013), though title length might vary across disciplines. For example, Méndez et al. (2013) found that titles in natural sciences articles were longer than those in social sciences, therefore presenting more extensive information content. On the other hand, results have been inconsistent in correlational research, not only across disciplines but within sub-disciplines belonging to the same broad area of knowledge. For example, Rostami et al. (2013) found that title length of articles published in *Addictive Behaviors Journal* had no obvious relation to citation rate while Paiva et al. (2012) came to the conclusion that shorter titles gathered more citations in Biomed Central (BMC) journals. However, a few studies did confirm

the results of previous studies or even supplemented further findings. The study conducted by Jacques and Sebire (2009) suggested a strong positive correlation between title length and citation rate in medical research journals, which was in line with the finding in Habibzadeh and Yadollahie's study (2010). They concluded that longer titles of medical research articles had a positive impact on the number of citations.

Another high-profile form feature is syntactic structure. Several models for analyzing syntactic structure based on different taxonomies have been widely employed. Soler (2007) identified four main syntactic structures used in titles, including nominal, full-sentence, compound and question constructions. Cheng et al. (2012) classified syntactic structure into compound structure, nominal structure, V-ing phrase structure, full sentence structure and prepositional phrase structure. Moattarian and Alibabae (2015) adapted the existing model proposed by Dietz in 1995 (as cited in Busch-Lauer 2000) and examined the frequency of nominal, verbal and prepositional construction of titles in three disciplines. Their results revealed a marked dominance of nominal constructions over other structures. Meanwhile, some researchers focused on the subdivision of only one syntactic structure type. Wang and Bai (2007) grouped the nominal structure into three types, which are Uni-head nominal group, Bi-head nominal group, and Multi-head nominal group, and they counted the frequencies of each type. Existing research on syntactic structure seemed to show that nominal construction is the most recurrent structural construction with an overwhelming majority. Various disciplines were examined, in which articles in medicine accounted for the majority (Wang and Bai 2007; Ball 2009). A contrastive approach was often employed in the study of syntactic structure, featuring either several subject fields belonging to different disciplines or two broad-spectrum disciplines. For example, Moattarian and Alibabae (2015) compared syntactic structure of article titles from applied linguistics, civil engineering, and dentistry. Soler's (2007) study contrasted syntactic structure of journal article titles between sciences, including medicine, biology and biochemistry, and social sciences, including linguistics, psychology and anthropology. However, most investigations were synchronic studies without observing the long-range trends.

The third category of research on form features focuses on punctuation marks, which are used to establish sentence boundaries, thereby identifying different types of titles. Haggan (2004) defined three basic types of titles, including full sentence (e.g. *Was Spenser a Republican?*), compound (e.g. *Circling the spheres: A Dialogue*) and a remaining group made up largely of noun phrases with or without postmodification (e.g. *Neo-Petrarchan Kitsch in Romeo and Juliet*). According to the taxonomy proposed by Méndez et al. (2013), compound titles comprised a general heading and a specific theme generally separated by a colon, a dash, a full stop, or written in two different lines, while simple titles only consisted of a general heading. Lewinson and Hartley (2005) reported that titles with colons were longer and more informative than those without. Hartley (2007) found that there was a greater use of colons in arts and humanities than in sciences. His finding was also verified in Nagano's (2015) investigation. Furthermore, researchers were interested in the relationship between the use of punctuations and citations (Rostami et al. 2013; Paiva et al. 2012; Nair and Gibbert 2016) and the change of punctuations in titles with time (Ball 2009; Hartley and Cabanac 2015). Rostami et al. (2013) discovered that articles (from *Addictive Behaviors*) with titles using hyphens or colons had a higher number of citations while Nair and Gibbert (2016) summarized that the use of punctuations and citations did not seem to correlate in management science articles. Studies by Ball (2009) and Hartley and Cabanac (2015) both came to the conclusion that there was a growing inclusion of question marks in scientific article titles and in Hartley's academic papers.

Content-related title research has mainly investigated information types provided in titles based on different taxonomies. Anthony (2001), as one of the earliest scholars who studied the content of academic article titles, proposed a set of categories named “rhetorical combination” based on Hamp-Lyon’s (1987) language assessment criteria. In his taxonomy, Name-Description, Description-Name, Topic-Description, Topic-Scope, and Topic-Method were used to describe the content of compound titles. Based on Anthony’s model, Cheng et al. (2012) developed a more precise classification, consisting of eleven information types. Salager-Meyer and Alcazar-Ariza (2013) identified titles that contain a statement of purpose, method and/or outcome as Research Procedure Titles, and explored the long-range trend of this attribute in titles extracted from the *British Medical Journal*. Sahragard and Meihami (2016) established a framework by adding definition to each feature of the prototype of Goodman et al. (2001). They identified five elements conveyed by titles: topic only, method/design, dataset, result and conclusion, and applied this framework to their diachronical study on the research titles of applied linguistics journal. McGowan and Tugwell (2005) as well as Gjersvik and Nylenna (2014) examined whether article titles in the medical field were “declarative”, i.e., whether titles revealed the research conclusion. The research on title content revealed information trends across different journals or disciplines, whether they showed similar or different patterns, and provided suggestions to novice writers on how to formulate their article titles.

Since Diener (1984) investigated the informational dynamics in journal article titles, an increasing number of scholars have begun to study the informativity of article titles. Title informativity is of great importance, and highly informative titles tend to perform their functions more effectively (Yitzhaki 1994). This acknowledgement thus promoted increased attention to the study on the informativity of journal article titles (Buxton and Meadows 1977; Nagano 2015). Related research commonly employed the method of counting the number of “substantive” words. Tocatlian (1970) conducted a diachronic analysis of informativity of chemical paper titles. He defined non-substantive words as words that convey little or no information, such as articles, prepositions, conjunctions, pronouns, and auxiliary verbs. This objective approach was then extensively used by scholars to study title informativity. Journals of various subject fields or disciplines, such as ecology (Rodríguez and Moreira 1996), humanities (Yitzhaki 1997) and multi-disciplines (Buxton and Meadows 1977) have been investigated. Existing studies on the informativity of journal article titles were all diachronic, consistently indicating the increase of the number of substantive words in titles with time.

Lexical diversity is also an important parameter in the content of titles. According to Bérubé et al. (2018), lexical diversity is an indication of whether academic discourse develops towards a more disparate or concentrated vocabulary. On the one hand, it seems that the advancement of a scientific discipline may bring about a diversification of its terminology. On the other, stabilizing the vocabulary is necessary to ensure effective transmission of knowledge within a subject field (Jacob 2004). This contradiction naturally would call for empirical studies on the variation trend of lexical diversity of scholarly discourse, particularly journal article titles. However, only a few studies can be found addressing this question. In 2015, Milojević track the evolution of vocabulary diversity in three scientific fields, namely, physics, astronomy and biomedicine, to quantify the extents of cognitive domains of different bodies of scientific literature independently from publication volume. Results showed that while the publication rates continued to grow exponentially, the number of distinct noun phrases expanded on linear scales “within a factor of few” (Milojević 2015). Later in, Bérubé et al. (2018) examined the trend of lexical diversity of scholarly titles. The findings of this study revealed that

titles in the fields of natural sciences & engineering and social sciences & humanities incorporated the use of an increasingly restricted and cross-disciplinary set of words, which was indicated by the slight decrease of lexical diversity. Up to now, the most frequently-used and intuitive way of measuring lexical diversity on the basis of word repetition patterns is type-token ratio (hereafter TTR) (Bérubé et al. 2018). Unfortunately, TTR is a function of sample size, which means a larger sample size could lead to a lower TTR. Other commonly used measures derived from TTR are also problematic although they are claimed to be independent of sample size. (Malvern and Richards 2002). In 2004, Malvern et al. proposed Measure of Textual Lexical Diversity (MTLD) to deal with the sample size dependency of TTR through fixed size sampling procedures. Later Milojević (2015) proceeded by calculating the number of different noun phrases within consecutive segments of 1500 noun phrases, which was similar in kind to the method advocated in Malvern et al. (2004)'s study. However, according to Bérubé et al. (2018), such fixed size sampling techniques did not solve the sample size dependency problem at all, but merely transformed it into a sampling scheme dependency problem. In Bérubé et al. (2018)'s study, the authors adopted a new indicator based on zipfian frequency-rank distribution tail fits, which was claimed to be more independent of corpus size than other lexical diversity indicators. Still, in the same study, Bérubé et al. (2018) pointed out that there is no universally agreed method of standardizing samples to avoid the limitation of TTR yet (Malvern et al. 2004). But they also clarified that lexical diversity variables measured by TTR would still be valid in cases where language segments are of relatively small size. Nevertheless, they did not define what is “small enough”. Given the considerably smaller sample size in the present study compared to the previously-mentioned studies on lexical diversity of titles (e.g. Milojević 2015; Bérubé et al. 2018), TTR will be used to measure lexical diversity of linguistics and literature article titles.

Decades of title research has yielded fruitful results, but limitations still exist. First, some disciplines received more scholarly attention than others. Medical journals were explored the most comprehensively with an uncontested conclusion that medical article titles tend to be long, precise, and informative (Méndez et al. 2013). In contrast, a rather limited number of studies were conducted on language-related journals, such as applied linguistics journals (Cheng et al. 2012; Sahragard and Meihami 2016) and literature journals (Diener 1984; Yitzhaki 1994; Haggan 2004; Cook and Plourde 2016). Even within the same discipline, imbalanced attention was paid to different sub-fields. For example, in linguistics, unlike applied linguistics journals, theoretical linguistics journals have seldom been examined. Second, in comparative title studies, researchers usually chose to compare sciences or engineering with social sciences or humanities (e.g. Buxton and Meadows 1977; Haggan 2004; Soler 2007; Méndez et al. 2013; Moattarian and Alibabae 2015), but seldom between humanities and social sciences (Yitzhaki 1994). However, differences in title features may still exist among these two closely related disciplines or subject fields although social sciences and humanities both study the human aspects of the world. Third, most of the existing title research concentrated on the correlation between various title features and citations. However, the results of these studies were sometimes inconsistent (Haggan 2004; Hartley 2008; Soler 2011). A change of approach then might reveal some interesting and new findings. Hence, the present study intends to investigate research article titles in linguistics (social sciences) and literature (humanities) journals between 1980 and 2018 from a diachronic and comparative perspective. Specifically, we will study both form features and content attributes by examining titles from seven leading journals belonging to either of the two disciplines. Our study will therefore address the following two research questions:

1. Are there any diachronic changes in title features of the two disciplines respectively between 1980 and 2018?
2. Do title features in linguistics journals and literature journals exhibit a similar pattern?

## Methodology

### Data

We adopted a two-step approach to compiling the corpus. Firstly, we selected journals following the criteria below:

1. The linguistics journals are all SSCI-indexed (Social Sciences Citation Index) and the literature journals are all A&HCI-indexed (Art & Humanities Citation Index);
2. The journals should have begun publishing since 1980.
3. The journals publish articles not limited to only one particular research area;
4. Considering the differences between linguistics and applied linguistics in terms of definition, principles, scope, focus, etc., the linguistics journals we selected generally publish research on language itself rather than applied studies, such as language teaching.

Taking the above four criteria into consideration, we selected the following fourteen journals. (see Table 1).

In the second phase, we divided the decade years between 1980 and 2018 into four periods (A: 1980–1989; B: 1990–1999; C: 2000–2009; D: 2010–2018). To mitigate sampling bias, we collected research article titles from all the volumes in 1988, 1998, 2008 and 2018 respectively to represent each period, excluding book reviews, review articles, announcements, notes, squibs and discussions, remarks and replies, as well as forum articles. Articles written in languages other than English were not included either. The number of titles collected from linguistics journals and literature journals are 679 and 685 respectively. Therefore, our corpus contained a total number of 1364 titles. Table 2 presents the specific information of this corpus.

**Table 1** The 14 selected journals

Linguistics journals	Literature journals
Language	College literature
Theoretical linguistics	Cambridge quarterly
Linguistics and philosophy	Comparative literature
Linguistic inquiry	Journal of literary studies
Journal of linguistics	Journal of modern literature
Lingua	Essays in criticism
Language science	Modern fiction studies

**Table 2** The corpus data

Journals	1988	1998	2008	2018	Total
Language	17	15	14	36	82
Theoretical linguistics	11	12	20	2	45
Linguistics and philosophy	17	16	21	18	72
Linguistic inquiry	24	13	10	18	65
Journal of linguistics	12	12	15	19	58
Lingua	30	25	84	86	225
Language science	19	24	28	61	132
Total	130	117	192	240	679
Journals	1988	1998	2008	2018	Total
College literature	22	34	31	50	137
Cambridge quarterly	11	14	20	22	67
Comparative literature	13	13	20	26	72
Journal of literary studies	24	21	19	31	95
Journal of modern literature	12	28	44	59	143
Essays in criticism	12	10	8	19	49
Modern fiction studies	29	35	31	27	122
Total	123	155	173	234	685

### Data analysis methods

The present study will specifically investigate form features (average title length and syntactic structure) and content attributes (information types, lexical diversity and informativity) of journal article titles. Unlike Méndez et al. (2013) and Nair and Gibbert (2016), we adopted both manual counting and automated calculation using statistical software WordCounter to measure title length. The two results were then manually checked for consistency. Title length refers to the number of words in a title. A word, in our research, refers to a unit occurring with space or punctuation on either side, while a hyphenated word counts as one. Acronyms or abbreviations combining capital letters and figures are counted based on the number of capitalized initials (as well as the number of figures). For example, “EFL” (English as Foreign Language) is counted as three words; “L2” (Second Language) is counted as two words. Considering that the semantic components of hyphenated words in the disciplines of linguistics and literature are usually interpreted as a whole, a hyphenated word in this corpus is counted as one word.

For the purpose of the present study, existing frameworks were employed to examine the other title features manually. Specifically, Sahragard and Meihami (2016) framework (see Table 3) adapted from Goodman et al. (2001) was used to investigate the information type of titles. The types include *topic only*, *method/design*, *dataset*, *result*, and *conclusion*. It is worth mentioning that many titles contain more than just one information type. For instance, the title *Evaluating S(c)illy Voices: The Effects of Saliency, Stereotypes, and Co-present Language Variables on Real-time Reactions to Reginal Speech* includes both method and result. To investigate the syntactic structure of titles, we adopted the framework proposed by Cheng et al. (2012). They identified five types

**Table 3** Framework of information types (Sahragard and Meihami 2016, adapted from Goodman et al. 2001)

Categories	Description	Example
Topic only	Title indicates a subject but does not include information on other categories listed below, or it seems ambiguous. Titles referring to second language acquisition in general are categorized in this category	Storytelling and Legal Legitimacy; Lexical and Syntactic Causatives in Oromo
Method/design	Title specifies the topic and an approach to study design, data management, or analysis (such as case-control, cohort, effectiveness, efficacy, frequency, incidence, prevalence, trend, or validity study; meta-analysis; randomized), or provides a vague description of a method (such as assessment, evaluation, or comparison) or refers to the statistical tests used in a study (ANOVA, ANCOVA, MANOVA). It may contain information about the method of the study but not its design (for example randomization), or vice versa	Biafra and the Question of Genocide: Examining the Dimensions of Violence in Heroes and Toads of War; Balancing the communication equation: An outreach and engagement model for using sociolinguistics to enhance culturally and linguistically sustaining K-12 STEM education
Dataset	Title indicates the topic and name or acronym of a specific study, which can reveal the origin of the dataset considered or the population sample consulted	The Walls that Emancipate: Disambiguation of the “Room” in A Room of One’s Own; What Are You Cookin’ On a Hot?: Movement Constraints in the Speech of A Three-Year-Old Blind Child
Result	Title includes the topic and quantitative information (a specific value), semi quantitative or ordinal information (such as increased, decreased, high, or low), or some other specification of a relation (such as association, change, correlation, determinants, effect, evidence, impact, influence, outcomes, predictors, relation, remission, risk, variability, or variation) about the findings. Other titles may contain terms such as role, effectiveness, relationship etc	The Role of Water Image In: Uncle Tom’s Children”; The Role of Diffusion in the Genesis of Hawaiian Creole
Conclusion	Title includes the topic and an unequivocal statement based on the analysis of the reported evidence such as the implications of a study or the overall conclusion	The time course of individuals’ perception of coarticulatory information is linked to their production: Implications for sound change



of syntactic structures: nominal, compound, full sentences, V-ing phrases and prepositional phrases (see Table 4).

To measure informativity, we followed the existing approach used by Tocatlian (1970) and Buxton and Meadows (1977) by counting the number of “substantive” words in each title manually with articles, prepositions, conjunctions, pronouns, and auxiliary verbs excluded. For example, the number of substantive words in the title *Emphasis Harmony in a Modern Aramaic Dialect* is five and the informativity (the number of substantive words divided by the total) of this title is 71.4%.

In the present study, we adopted TTR to measure lexical diversity. It is calculated by dividing the number of type count in a title by the number of token count in it. The token count of a language segment represents the total number of words it contains, while the type count refers to the number of different words in it. For example, lexical diversity of the title *Inversion and finiteness in Spanish and English: Developmental evidence from the optional infinitive and optional inversion stages* is 76.5% (the total number of words in this title is seventeen, while the number of different words in it is thirteen).

## Results and discussion

The present study intends to unveil the long-range trends of research article title features published in linguistics and literary journals between 1980 and 2018 and to examine possible disciplinary variations in title features. In this section we will firstly present the features in different time periods and analyze the diachronic changes in linguistics journals and literature journals respectively. Then we will describe the similarities and differences in title features between the two disciplines.

### Diachronic description of title features in linguistics journals

#### Average title length

Table 5 shows that the average title length of linguistics journals enjoyed a trend for each decade year, with an increase from 6.6 words per title in 1988–11.2 words per title in 2018. The increase for the first decade-year was gradual but not significant, with the exception of the latest decade which saw a rather dramatic increase of 2.3 words per title, suggesting that linguistic researchers are more likely to use longer titles than before. In general, from 1980 to 2018, the title length has increased by 4.6 words, an increase of 70% on average.

#### Syntactic structure

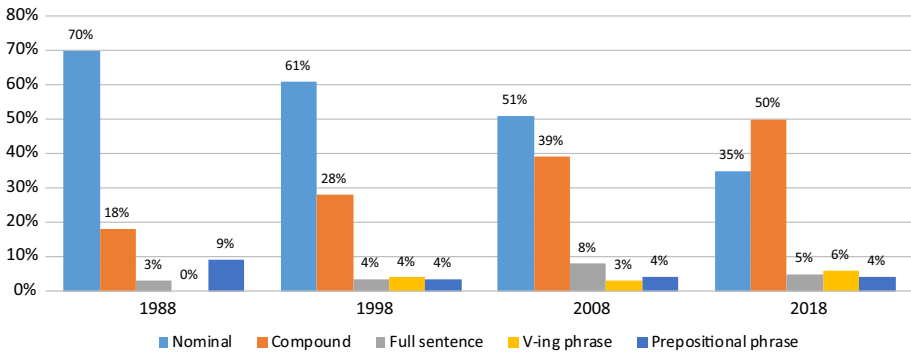
Figure 1 shows the information on syntactic structure of titles in linguistics journals. On the whole, nominal structure and compound structure are the dominant structures used, taking up almost ninety percent of the entirety. However, the use of nominal structure has suffered a continuous decline while compound structure has enjoyed a steady 10% growth for each decade year. In the latest period, the popularity of compound structure surpassed nominal structure and reached its peak (50%) in the four periods. This result in the 2010s is consistent with the finding in Cheng et al. (2012) study, in which compound structure stood at 54% and nominal structure took up about 39% of the applied linguistic journal article titles between 1999 and 2008. The other three structures, including full sentence, V-ing

**Table 4** Framework of syntactic structures (Cheng et al. 2012)

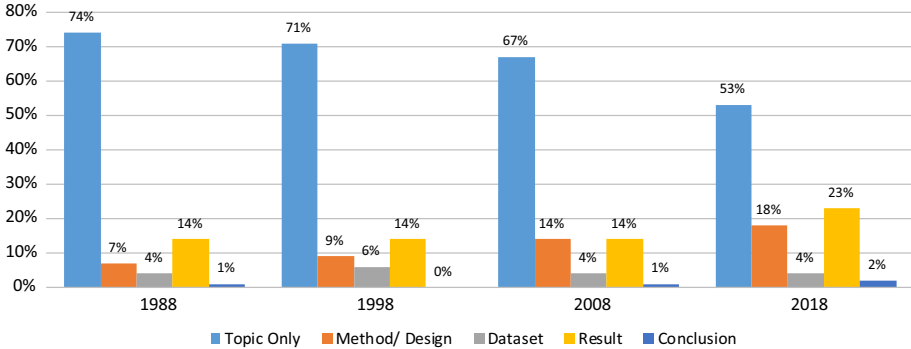
Categories	Description	Example
Nominal	It consists of one or more nouns (that is, a head or heads) that can be preceded by pre-modifiers or followed by post-modifiers	Meals and Mourning in Woolf's "The Waves"; Experimental Evidence for the Productivity of Total Reduplication in Japanese Ideophones and Ordinary Vocabulary
Compound	It consists of two parts separated by a colon, dash, question mark, or period	Economic Hauntings: Wealth and Class in Edith Wharton's Ghost Stories Context and Information Structure Constrains on Factivity: The Case of Know
Full sentences	It comprises a complete sentence or clause, including both declarative structures	What is Good? How Children's Relatives Solve a Problem for Minimalism?
V-ing phrase	It is the V-ing form of one or more verbs, followed by objects or modifiers of the verb(s)	Redefining Family in Jennifer Johnston's Foolish Mortals Predicting L2 Vowel Identification Accuracy From Cross-language Mappings Between L2 English and L1 Korean
Prepositional phrase	It starts with a preposition that is followed by the object of the preposition	Between the O'Neills On the Questions of Acehnese "Passive"

**Table 5** Average title length in linguistics journals

	1988	1998	2008	2018
Total words/numbers of titles	854/130	829/117	1708/192	2690/240
Average length	6.6 words/title	7.1 words/title	8.9 words/title	11.2 words/title



**Fig. 1** The trends of syntactic structures of titles in linguistics journals



**Fig. 2** The trends of information types in titles in linguistics journals

phrase and prepositional phrase were also adopted in linguistics journals, but they made up a very small proportion, less than 20% combined for each decade year.

### Information types

In general, titles containing the information on topic only predominated over the past four periods with no exception, followed by result, method/design, dataset and conclusion in sequence (see Fig. 2). However, from 1980 to 2018, topic only titles have experienced a dramatic decrease from 74 to 53%, while titles with information about result and method/

design have both enjoyed relatively substantial increases, from 14 to 23% and 7 to 18% respectively. Although topic only titles in the 2010s still accounted for the largest portion, the absolute predominance was gradually weakened in this period of time. Meanwhile, it is worth noticing conclusion titles were the least favored with a very slight fluctuation between 0 and 2%. This finding is basically in line with Sahragard and Meihami's (2016) results, which showed that applied linguistics journals tended to provide the least information on conclusion of the studies. However, their research revealed a decline in titles containing information about conclusion in *Modern Language Journal*, *Language Learning*, and *Foreign Language Annals*, while our findings suggest a weak increase of conclusion titles between 2010 and 2018. As the proportion of titles containing methods/design and result both rise steadily, it is not difficult to make the case that more dimensions of research content are being included in linguistics article titles. That means that journal article titles in the field of linguistics have become increasingly information-diversified and can better serve as the "compass" to provide initial information for readers. Another important finding about information type is that the proportion of dataset did not witness much change and is relatively lower than that presented in Sahragard and Meihami's (2016) study. A possible explanation might be the difference in our data source. The journals chosen by Sahragard and Meihami (2016) mainly publish applied linguistics articles, which focus more on the subjects and specific groups of people; while the journals that we selected all center around pure linguistic topics. Another explanation could be the differences in journal guidelines between disciplines. Researchers normally formulate titles in accordance with the journal policies and discourse conventions to a great extent (Gesuato 2009).

## Informativity

Informativity of titles can greatly determine whether the title is catchy or not (Nagano 2015). The figures shown in Table 6 reveal that informativity of linguistics article titles reached a low of 59.8% in the 2000s and peaked at 70% for the last decade year. Although the number of substantive words in titles increased in the 2000s, its growth rate could not catch up with that of average title length, resulting in a slight decline in the 2000s. We can see from Table 6 an overall growing trend of informativity of titles in linguistics journals, which is consistent with findings of the previous research conducted by Yitzhaki (1997), who claimed that the increase in informativity in humanities journal article titles from one decade to the following was not always significant and not necessarily linear even if the long-range trend indicated an increase. More specifically, Yitzhaki (1997) examined the number of substantive words in article titles in the journal *Language*, which also contributes to the corpus of our research, between 1930 and 1990 in his study. He came to the conclusion that the average number of substantive words in titles published in *Language* in 1980 and 1990 were 4 and 5 respectively, which was exactly consistent with our results.

**Table 6** The trends of informativity of titles in linguistics journals

	1988	1998	2008	2018
Substantial words/total words	563/854	569/829	1021/1708	1879/2690
Informativity	65.9%	68.6%	59.8%	70.0%
Substantial words/per title	4.3/title	4.9/title	5.3/title	7.8/title

**Table 7** The trends of lexical diversity of titles in linguistics journals

Period	1988	1998	2008	2018
TTR	432/854	438/829	739/1708	1200/2690
Lexical diversity	50.6%	52.8%	43.3%	44.6%

**Table 8** Average title length in literature journals

Period	1988	1998	2008	2018
Total words/numbers of titles	1067/123	1602/155	1895/173	2617/234
Average length	8.7 words/title	10.3 words/title	11.0 words/title	11.2 words/title

## Lexical diversity

Table 7 shows the trend of lexical diversity of titles in linguistics journal articles. From the 1980s to the 2000s, lexical diversity went through an overall decrease from 50.6 to 44.6%, with slight increases of 2.2% and 1.3% for the second and the last decade year respectively. A more significant decline of nearly ten percent was registered for the third period from 52.8 to 43.3%. Nevertheless, variations in lexical diversity of titles seem to be small in linguistics journals.

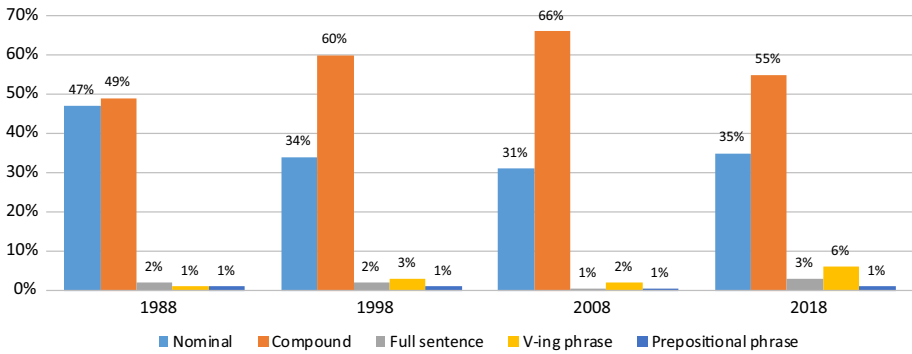
## Diachronic description of title features in literature journals

### Title length

The average title length of literature journals has kept increasing steadily during the four periods, from 8.7 words per title to 11.2 words per title, which is in line with the previous studies on the variation of title length (Yitzhaki 1994; Méndez et al. 2013). Furthermore, there is only a slight increase after the 2000s, indicating that the average title length of literary articles has been staying around 11 words per title for the past two decades (Table 8).

### Syntactic structure

Results show that compound structure is the most widely used syntactic structure in literature journals throughout the four periods, followed by nominal structure (see Fig. 3). Since most of the literature journal articles concentrate on one specific topic, such as a book, an author, a literary figure, a particular phenomenon, etc., the authors seem to prefer using compound structure, especially using colons to give an abstract concept first and then further introduce the real topic after the colon. For instance, *Woman in a Trap: Pope and Ovid in “Eloisa to Abelard”* (published in *College Literature*) can be a good illustration in point. Our finding is consistent with that of Haggan’s (2004), which concluded that in literature journals compound titles had a much higher percentage than nominal titles. The proportions of the other three syntactic structures remained very small and witnessed few

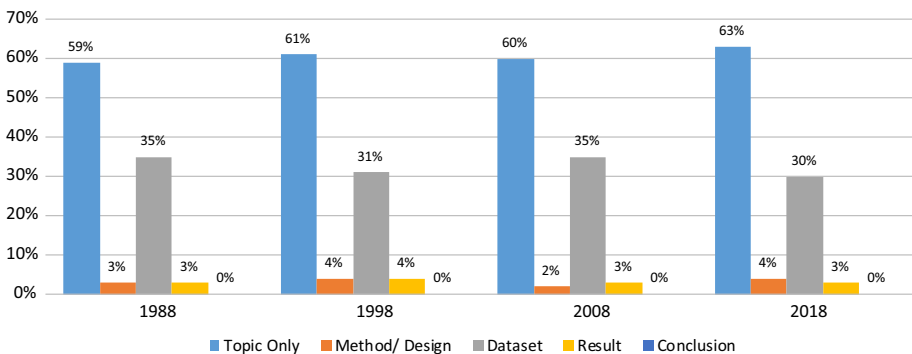


**Fig. 3** The trends of syntactic structures of titles in literature journals

noticeable changes across the periods examined. However, there was a slight increase of 4% in V-ing phrase in the last decade year. This change might suggest that researchers tend to include the analysis process of literary works in their titles.

### Information types

Figure 4 shows that titles containing the information on topic only are the most favored and frequently-used in literature journals from 1980 to 2018. Although there were small fluctuations between different periods, the long-term trend of increase in topic only titles is still significant. Given that most literature journal articles do not require experimental designs or fixed and objective results, result and method/design especially were not provided in most titles throughout the four periods. Since a great number of literature journal articles focus on a literary work or even a narrator, the authors would sometimes include the name of the work or even one specific quotation in the title, thus revealing part of the source of the dataset. For example, the title *Back to the Future: Late Modernism in J.G. Ballard’s The Drowned World* (published in *College Literature*) directly revealed the research subject is the book *The Drowned World* written by J.G. Ballard. Therefore, titles providing



**Fig. 4** The trends of information types in titles literature journals

**Table 9** The trends of informativity of titles in literature journals

Year	1988	1998	2008	2018
Substantial words/total words	735/1067	1130/1602	1275/1895	2011/2617
Informativity	68.9%	70.5%	67.3%	76.8%
Substantial words/title	6.0/title	8.4/title	7.4/title	8.6/title

**Table 10** The trend of lexical diversity of titles in linguistics journals

Period	1988	1998	2008	2018
TTR	586/1067	891/1602	1011/1895	1261/2617
Lexical diversity	54.9%	55.6%	53.4%	48.2%

information on dataset account for more than one third of the whole proportion, which is much higher than that of method/design and result.

### Informativity

The trend of informativity in literature article titles is presented in Table 9. It can be seen that, in general, informativity increased with periods investigated, which is basically consistent with the results of other studies (Buxton and Meadows 1977, Yitzhaki 1994). Informativity remained high though with occasional fluctuations. The proportion fell suddenly in the 2000s to 67.3% before peaking at about 76.8% in the 2010s. Though the chosen literature journals in this paper were not incorporated in Yitzhaki’s study (1997), our finding confirms its conclusion that the pace of increase in the humanities journals has been much slower, and there are some decreases at certain points in time as for the long-range trend.

### Lexical diversity

The trend of lexical diversity of titles in literature journal articles is presented in Table 10. Average lexical diversity decreased from 54.9% in the 1980s to 48.2% in the 2010s, with a minor increase of 0.7% for the second decade year. Since the 2000s, the downward trend became more prominent with a decrease of 5.2%. Nevertheless, the whole process seems to be a gradual fall without a trough.

### Comparison between linguistics journals and literature journals in terms of title features

Titles in linguistics journals and literature journals show consistency in the long-range trend of title length and informativity, both enjoying a steady and linear growth in length and an increase with fluctuations in informativity, which is basically consistent with the findings of previous studies. For example, Buxton and Meadows (1977) came to the conclusion that article titles of most subjects showed a significant increase in informativity from 1947 to 1973. Although they did not examine article titles in linguistics and literature,

our findings show consistency with theirs and verify the validity of their conclusion. Compared with the 1980s, titles of both disciplines achieved an immense increase in length in the 1990s. While the length had been growing steadily for the third decade year, titles of both disciplines encountered a sharp decrease in informativity in the same period. Although longer titles may imply the increase of substantive words, our finding suggests that word increase does not necessarily equal the growth of information content. Moreover, when title lengths of both disciplines were enjoying a rapid increase from the 1980s to the 1990s, the proportions of compound structures also experienced a sudden increase in the same period, from 18 to 28% and 49 to 60% respectively. The average length of compound structures is the longest in general since the compound structure can serve the role of the other three while playing its unique part (Wang and Bai 2007), therefore the growth in the proportions of compound structure may lead to that in title lengths. As far as lexical diversity is concerned, both disciplines experienced a downward trend. Starting from the 1980s, lexical diversity increased slightly and apparently experienced a decline since the 2000s. This might be the result of the increasing use of the same words for important concepts and problems. The downward trend is also generally consistent with the finding in Bérubé et al. (2018)'s research on the lexical diversity of humanities and social sciences titles, which experienced an 8% decrease over the period from 1975 to 2014. However, a difference can still be identified between the two subject domains. Even though lexical diversity of linguistics article titles enjoyed a slight increase for the latest period, it was still about four percentage points lower than that of literature article titles. A possible reason for this might be the difference in information types carried by titles between these two fields. While incorporating terminology in the description of problems, methods and results may be a common practice in formulating linguistics article titles, literary works or main characters would often be mentioned in literature article titles, contributing to a relatively higher level of lexical diversity.

Nevertheless, significant variations between the two disciplines do exist in several aspects. First, compound structure has remained predominant in literature journals. While nominal structure had been the most widely adopted syntactic structure of titles in linguistics journals for the first three periods, it was surpassed by compound structure in the most recent decade. While Wang and Bai (2007) recognized the heads in nominal titles usually function to inform readers of the general focus of study and often need further specification, Soler (2007) in the same year also pointed out nominalization is the materialization of informativity via the piling up of pre- and post-modifiers. For example, in *The Role of Diffusion in the Genesis of Hawaiian Creole* (published in *Language*), the subject *Diffusion*, is an abstraction of the research object. However, the specific research result and the origin of data is revealed through both the pre-modifier, *The Role* and the post-modifier *Genesis of Hawaiian Creole*. The pre-modifier *The Role* is an elicitation of the result of the research, while the post-modifier *Genesis of Hawaiian Creole* gives a specific statement about in which group of subjects this research is conducted. This may account for the high proportion of nominal structure used in linguistics article titles. However, as titles also serve the function of an eye catcher, the use of punctuation marks in titles, especially question marks can greatly arouse the interests of readers in further reading. Though nominal titles effectively and coherently summarize the essential information of research articles (Cheng et al. 2012), the same group of research also shows clearly the advantages of compound structures in their study. Titles with a compound structure can better present the indication of a specific research focus in titles (Cheng et al. 2012) and can realize more functions. The most irreplaceable function of compound titles is its clever format, which allows the general topic and specific content of the research to coexist. Specifically, as either part of



a compound title can be a nominal, V-ing phrase, full sentence or prepositional phrase, a compound title can achieve a maximized function compared to the other three structures. For example, in *What Are You Cookin' on a Hot?: Movement Constraints in the Speech of A Three-Year-Old Blind Child* (published in *Language*), the author(s) combined a full sentence with a nominal statement modified by a post-modifier. This title achieved the functions of interrogative sentence and nominal structure at the same time, arousing the interests of readers as well as revealing the research subject and result. It is understandable linguistics article titles tend to explore the complicated relationships between a number of key elements such as social context, data source, participants, method, scope. However, while the researchers try to squash more information into the title, the old-fashioned nominal structure is proceeding from glory to decline and will be replaced by emerging compound structures. Because of the literariness and subjectivity of literary research, colons are employed in a great number of literary titles so as to retain the romance and mystery before the colon and supply subjective elaboration after the colon. Therefore, in literature journals, compound structure has always been the most popular across the four periods.

As far as information type is concerned, the proportion of topic only titles has always been overwhelmingly high (around 60%) in literature journals. On the contrary, topic only titles in linguistics journals, though still dominant compared with other information types, suffered a linear and consistent decrease from 74 to 53%. Specifically, linguistic article titles have gradually covered more information types by incorporating the description of result and method/design. The variation can be accounted for from the following aspects. One is that the empirical and quantitative approach has been increasingly popular in linguistic research, which calls for rational experimental designs and is expected to present objective results. In contrast, literary research usually focuses on literary analysis of specific literary figures or phenomena based on subjective interpretation. The other possible explanation might be that literature research does not necessarily generate specific and objective results, let alone specifying information on result in titles. Furthermore, dataset takes up the second highest proportion in literary article titles while it is seldom included in linguistics article titles. This might be related to the particularity of literary research. A large amount of literature research will focus on a specific literary work, therefore, its title, a type of dataset, would normally be included in the article title.

## Conclusions

This paper investigated five title features of research articles in seven linguistics journals and seven literature journals. Since linguistics and literature are both language-related disciplines, it might be interesting to conduct a diachronic study on their respective prominent features, long-range trends, as well as the similarities and differences between the two disciplines. Article titles in both disciplines demonstrated a similar trend: title lengths enjoyed a steady increase while informativity went up with fluctuations and lexical diversity experienced a modest decline. Nominal structure and compound structure combined have the largest proportion, accounting for more than ninety percent, which is in line with Cheng et al (2012) results. As for information types, in our study, the proportion of method/design in linguistics journal article titles is not as high as that of topic only, which is overwhelmingly high in the first three decades and still relatively high during the latest decade. We also found that topic only is the information type provided most frequently by titles in both

linguistics and literature journals, which is consistent with the conclusion of Siegel et al. (2006) and Goodman et al. (2001) drawn from articles in the field of medicine.

The present study has enriched the existing title research and may bear some implications for future studies. First, since titles in literature and linguistics journals were not studied as comprehensively as other science subjects investigated before, this study has verified the research results of predecessors and served as a solid supplement. In addition, the findings of this article can provide valuable information on the formulation of a research article title for those who have the intention to publish in the fourteen selected journals in this study. Considering the research result that compound article titles are more prevailing than the other four types, EAP teachers ought to encourage linguistic and literary students to use compound structures more often when formulating research article titles in academic English writing class.

However, the present study still has some limitations. First, the results of lexical diversity in this research may still need further validation, since no empirical assessment of the sample size robustness of the TTR on our corpus was conducted. Second, information types and syntactic structure of article titles in this study are manually analyzed. Although two authors went through the same procedures and checked for consistency, the classification may still be subjective. In addition, a larger corpus of titles covering a wider time span would probably offer more significant variations. With the increase of sample size, a different lexical diversity measure could also be employed to resolve the sample size dependency problem, thus hopefully yielding more convincing results in lexical diversity. Moreover, this study mainly focused on the diachronic variation of each of the five features studied. Not much attention was paid to the possible correlations among the five features, which may call for further investigation as well. Future studies can also be conducted on the specific types of compound structure, since a comprehensive model has already been developed by Anthony (2001) for further analysis.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

- Afful, J. B. A. (2017). A linguistic analysis of conference titles in applied linguistics. *International Journal of Foreign Language Teaching*, 5(18), 11–25.
- Anthony, L. (2001). Characteristic features of research article titles in computer science. *IEEE Transactions on Professional Communication*, 44(3), 187–194.
- 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). Pisa: PLUS.
- Ball, R. (2009). Scholarly communication in transition: The use of question marks in the titles of scientific articles in medicine, life sciences and physics 1966–2005. *Scientometrics*, 79(3), 667–679. <https://doi.org/10.1007/s11192-007-1984-5>.
- Bérubé, N., Sainte-Marie, M., Mongeon, P., & Larivière, V. (2018). Words by the tail: Assessing lexical diversity in scholarly titles using frequency-rank distribution tail fits. *PLoS ONE*, 13(7), e0197775.
- Busch-Lauer, I. (2000). Titles in english and German research papers in medicine and linguistics. In A. Trosborg (Ed.), *Analysing professional genres* (pp. 77–97). Amsterdam: John Benjamins.
- Buxton, A. B., & Meadows, A. J. (1977). The variation in the information content of titles of research papers with time and discipline. *Journal of Documentation*, 33(1), 46–52.
- 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.

- Cook, M. J., & Plourde, D. (2016). Do scholars follow Betteridge's Law? The use of questions in journal article titles. *Scientometrics*, 108(3), 1119–1128.
- Diener, R. A. V. (1984). Informational dynamics of journal article titles. *Journal of the American Society for Information Science*, 35(4), 222–227.
- Duchet, C. (1973). La Fille abandonnée et La Bête humaine, éléments de titrologie romanesque. *Littérature*, 12, 49–73.
- Fox, C. W., & Burns, C. S. (2015). The relationship between manuscript title structure and success: Editorial decisions and citation performance for an ecological journal. *Ecology and Evolution*, 5(10), 1970–1980.
- Gesuato, S. (2009). Encoding of information in titles: Practices across four genres in linguistics. In C. Taylor (Ed.), *Ecolingua: The role of e-corpora in translation and language learning* (pp. 125–157). Trieste: EUT.
- Gjersvik, P., & Nylenna, M. (2014). Declarative titles in original research papers in dermatologic journals 1974–2014. *Journal of the European Academy of Dermatology and Venereology*, 30(1), 124–125.
- Gnewuch, M., & Wohlrabe, K. (2017). Title characteristics and citations in economics. *Scientometrics*, 110(3), 1573–1578.
- Goodman, R. A., Stephen, B. T., & Paul, Z. S. (2001). What's in a title? A descriptive study of article titles in peer-reviewed medical journals. *Science Editor*, 24(3), 75–78.
- Guo, S., Zhang, G., Ju, Q., Chen, Y., Chen, Q., & Li, L. (2015). The evolution of conceptual diversity in economics titles from 1890 to 2012. *Scientometrics*, 102(3), 2073–2088.
- Habibzadeh, F., & Yadollahie, M. (2010). Are shorter article titles more attractive for citations? Cross-sectional study of 22 scientific journals. *Croatian Medical Journal*, 51(2), 165–170.
- Haggan, M. (2004). Research paper titles in literature, linguistics and science: Dimensions of attractions. *Journal of Pragmatics*, 36(2), 293–317.
- Hamp-Lyons, L. (1987). *Study writing: A course in written english for academic and professional purposes*. Cambridge, UK: Cambridge University Press.
- Hartley, J. (2007). Planning that title: Practices and preferences for titles with colons in academic articles. *Library and Information Science Research*, 29(4), 553–568.
- Hartley, J. (2008). *Academic writing and publishing: A practical handbook*. Abingdon: Routledge.
- Hartley, J., & Cabanac, G. (2015). An academic odyssey: Writing over time. *Scientometrics*, 103(3), 1073–1082.
- He, S. (2000). Translingual alteration of conceptual information in medical translation. *Journal of the American Society for Information Science*, 51(11), 1047–1060.
- Jacob, E. K. (2004). Classification and categorization: A difference that makes a difference. *Library Trends*, 52(3), 515.
- Jacques, T. S., & Sebire, N. J. (2009). The impact of article titles on citation hits: An analysis of general and specialist medical journals. *Journal of the Royal Society of Medicine Short Reports*, 1(2), 1–5.
- Kuch, D. C. (1978). Relation of title length to number of authors in journal articles. *Journal of the American Society for Information Science*, 29(4), 200–202.
- Lewinson, G., & Hartley, J. (2005). What's in a title? Number of words and the presence of colons. *Scientometrics*, 63(2), 341–356.
- Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, 19(1), 85–104.
- Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. New York: Palgrave Macmillan.
- McGowan, J., & Tugwell, P. (2005). Informative titles described article content. *Journal of the Canadian Health Libraries Association*, 26(3), 83–84.
- Méndez, D. I., Ángeles Alcaraz, M., & Salager-Meyer, F. (2013). Titles in english-medium astrophysics research articles. *Scientometrics*, 98(3), 2331–2351.
- Milojević, S. (2015). Quantifying the cognitive extent of science. *Journal of Informetrics*, 9(4), 962–973.
- Moattarian, A., & Alibabae, A. (2015). Syntactic structures in research article titles from three different disciplines: Applied linguistics, civil engineering, and dentistry. *Journal of Teaching Language Skills*, 34(1), 27–50. <https://doi.org/10.22099/jtls.2015.3530>.
- Nagano, R. L. (2015). Research article titles and disciplinary conventions: A corpus study of eight disciplines. *Journal of Academic Writing*, 5(1), 133–144.
- Nair, L. B., & Gibbert, M. (2016). What makes a good title and (how) does it matter for citations? A review and general model of article title attributes in management science. *Scientometrics*, 107(3), 1331–1359.
- Netterville, V. S., & Hirsch, B. L. (1958). Piracy and privilege in literary titles. *Southern California Law Review*, 32, 101.

- Paiva, C. E., Lima, J. P., et al. (2012). Articles with short titles describing the results are cited more often. *Clinics (Sao Paulo)*, 67(5), 509–513.
- Rodríguez, K., & Moreiro, J. A. (1996). The growth and development of research in the field of ecology—as measured by dissertation title analysis. *Scientometrics*, 35(1), 59–70.
- Rostami, F., Mohammadpoorasl, A., & Hajizadeh, M. (2013). The effect of characteristics of title on citation rates of articles. *Scientometrics*, 98(3), 2007–2010.
- Sahragard, R., & Meihami, H. (2016). *A diachronic study on the information provided by the research titles of applied linguistics journals*. New York: Springer.
- Salager-Meyer, F., & Alcazar-Ariza, M. A. A. (2013). Titles are “serious stuff”: A historical study of academic titles. *JAHHR*, 4(7), 257–271.
- Siegel, P. Z., Thacker, S. B., Goodman, R. A., & Gillespie, C. (2006). Titles of articles in peer-reviewed journals lack information on study design: A structured review of contributions to four leading medical journals, 1995–2001. *Science Editor*, 29(6), 183–185.
- Soler, V. (2007). Writing titles in science: An exploratory study. *English for Specific Purposes*, 26(1), 90–102.
- Soler, V. (2011). Comparative and contrastive observations on scientific articles written in English and Spanish. *English for Specific Purposes*, 30(1), 124–137.
- Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Tocatlian, J. J. (1970). Are titles of chemical papers becoming more informative? *Journal of the American Society for Information Science*, 21(5), 345–350.
- Wang, Y., & Bai, Y. (2007). A corpus-based syntactic study of medical research article titles. *System*, 35(3), 388–399.
- Whissell, C. (2013). Titles in highly ranked multidisciplinary psychology journals 1966–2011: More words and punctuation marks allow for the communication of more information. *Psychological Reports*, 113(3), 969–986.
- White, A. (1991). A further exploration of title size and author number. *Journal of the American Society for Information Science*, 42(5), 384–385.
- Wilsmore, S. J. (1987). The role of titles in identifying literary works. *The Journal of Aesthetics and Art Criticism*, 45(4), 403–408.
- Yitzhaki, M. (1994). Relation of title length of journal articles to number of authors. *Scientometrics*, 30(1), 321–332.
- Yitzhaki, M. (1997). Variations in informativity of titles of research papers in selected humanities journals: A comparative study. *Scientometrics*, 38(2), 219–229.
- Yitzhaki, M. (2002). Relation of the title length of a journal article to the length of the article. *Scientometrics*, 54(3), 435–477.