The Length and Semantic Structure of Article Titles—Evolving Disciplinary Practices and Correlations with Impact

Staša Milojević*

Center for Complex Networks and Systems Research, School of Informatics and Computing, Indiana University, Bloomington, IN, USA

In this paper, we examine the characteristics of titles (average length, proportion of titles with subtitles, proportion of interrogatory, and indicative titles) and how they changed over a substantial period of time (half a century). We consistently analyze core literature in five diverse fields, in order to probe the usage of titles as a disciplinary identity building tool. In addition, we study whether different types of titles are used differently by authors depending on their academic age and productivity and collaboration levels, which has not been studied before. Finally, we revisit the connection between title characteristics and impact. We found that belonging to some discipline is the strongest determinant for the length of the titles and the occurrence of different forms of titles. This suggests that authors try to comply with the norms set in their fields. However, these norms are not fixed in time. Over the decades, the practices have changed, some of them quite abruptly. Individual groups of authors most often did not differ in their practices regarding the use of titles. We find that using titles posed as questions or titles stating the result do not lead to citation benefit.

Keywords: article titles, research impact, scholarly communication, discourse analysis, computational linguistics

INTRODUCTION

Scientific documents are "the central medium for the dissemination and exchange of ideas" (Bowker, 2005, p. 126), and therefore represent the key element in the practice of science (Callon et al., 1983; Latour and Woolgar, 1986). In many scientific fields the principal type of scientific document is a research article ("paper"). Titles and the choice of words in them play two important roles: they inform the reader about the content of the paper, while at the same time attempting to trigger their attention (Bazerman, 1985, 1988; Ball, 2009). For the first role, the words are chosen to "convey credible information for a given population of producer-readers" (Callon et al., 1983, p. 199). Thus, the choice of words is a negotiation process reflecting both the behavior of individual authors and scholarly collectives (Hyland, 2004, 2012; Nagano, 2015). Authors choose particular words to denote their alliance with particular communities, sub-communities, and lines of thought. Thus, titles serve a very important third role, which is the denotation of academic identity. According to Hyland (2012) "we use language as the raw materials for the presentation of ourselves to the world and that what we say and write aligns us with or separates us from other people and their positions" (p. x). Identity is socially and historically constituted by individuals through their social relations and is

thus something that does not belong *within* an individual, but *between* persons (Vygotsky, 1978).

Titles are co-constructed within the "conceptual frames" (Hyland, 2012) of disciplines (Nagano, 2015). The usage of the acceptable discourse is a sign that a novice author is being socialized into a discipline (Becher and Trowler, 2001). We see that as early as dissertation work (Demarest and Sugimoto, 2015), where it was possible to discern among three disciplines (physics, psychology, and philosophy) based on a limited set of terms they used. Nagano (2015) claimed that the choices made in title construction are "largely matter of custom, habit, and copying others" (p. 134). The great influence of individual disciplines was supported by a number of studies that found greater differences in title characteristics between rather than within disciplines (van Wesel et al., 2014). This was also supported by the finding that time is the most important factor related to title length (White, 1991).

In relation to their role as attention triggers, titles are perceived to be an integral part of a strategy to attract an audience (Thelwall, 2017). Many studies from researchers from a number of fields: scientometrics, medicine, ecology (ECL), and linguistics, among others, focused on identifying title characteristics that may be correlated with high impact. Thus, these studies take a normative stance, trying to assist researchers in choosing title characteristics that have been found to lead to more citations. These studies range in the size of the corpus they use: from 50 articles (Jacques and Sebire, 2010) to nearly 20 million articles (Ball, 2009). The following title characteristics have been most studied: length, presence of non-alphanumeric characters, syntax/type, and word analysis.

The title length has been considered important for the retrieval of articles, which is an essential step toward reading, and eventually citing them. Namely, longer titles contain more words and, therefore, more potential keywords, thus increasing the potential for retrieval. On the other hand, longer titles may be "more difficult to digest and may reduce the attraction factor" (Hudson, 2016, p. 878). Shorter titles may also indicate a narrower scope. The results of the studies examining title lengths did not universally support the hypothesis of the greater impact of longer titles. Positive correlation between length and impact has been found for 25 most cited articles in medical journals published in top four journals in 2005 (Jacques and Sebire, 2010), and general and internal medicine articles (van Wesel et al., 2014). However, a very large number of studies found the negative relationship in biology (Didegah and Thelwall, 2013), social sciences (Didegah and Thelwall, 2013), sociology (van Wesel et al., 2014), and psychology (Subotic and Mukherjee, 2014). Others have found no correlation; e.g., in chemistry (Didegah and Thelwall, 2013), management science (Nair and Gibbert, 2016), ECL (Fox and Burns, 2015), or in articles published in six PLoS journals (Jamali and Nikzad, 2011). Anthony (2001) found that the lengths of titles within computer science literature varied widely, attributing it to different subdisciplines, claiming that "an adequate description depends more on the type of study or problem being investigated than the discipline itself" (p. 193).

Titles containing non-alphanumeric characters, such as colons, hyphens, dashes, and periods, indicating titles with a subtitle, or a multicomponent title, have been found to be common across the disciplines and generally associated with higher citation (Buter and van Raan, 2011). Specifically, the usage of colon has become more prevalent, although it differed across the fields (Lewison and Hartley, 2005). For example, a study of the medical literature found that 70% of highly cited articles included a colon (Jacques and Sebire, 2010), while a study of PLoS articles found that articles with colon had both fewer downloads and citations (Jamali and Nikzad, 2011).

Turning to title syntax, three types of article titles have been identified as most common: descriptive (or indicative), declarative (or informative), and titles posed as a question (or interrogatory) (Jamali and Nikzad, 2011; Wager et al., 2016). Descriptive (or indicative) titles describe the subject or topic of the article. Declarative (or informative) titles, in addition to describing the topic also state the main conclusion, or findings, of the study. Thus descriptive titles consist only of noun phrases, while declarative titles include action verbs. Rosner (1990), who calls such titles "the assertive sentence titles," found that in biological sciences they first appeared in the 1970s and have been on the rise ever since, being especially prominent in molecular biology, making up to 45% of all the titles in *Cell* since 1986. Goodman (2010) also found a trend toward the usage of longer words and active verbs in medical literature over the period between 1970 and 2009. Usage of declarative titles has been somewhat controversial, especially in medical literature (Rosner, 1990; Aronson, 2010). While some journals ban the use of declarative titles (e.g., *Microbiology*), others (e.g., *The Journal of Clinical Epidemiology*) have been enforcing their usage (McGowan and Tugwell, 2005). While their usage has increased, the study that used randomized trial on doctors and senior medical and dental students found that declarative titles did not have significant effect on "readers' perceptions of the conclusions" (Wager et al., 2016, p. 3).

Although the usage of question marks in titles has increased over time, they still account for a very small proportion of all the titles (Cook and Plourde, 2016). Anthony (2001) found interrogatory titles to be rare in computer science literature. Fox and Burns (2015) found that although the proportion of submitted articles to the journal *Functional Ecology* including questions in titles has decreased, the percentage of such articles among the published articles has increased over the 10-year period they had studied. Hudson (2016) found that the titles that include question marks were longer than other titles. The increase in the usage of questions in titles has been attributed to marketing purposes (Ball, 2009). The study of over 2,000 articles published in 6 PLoS journals found that articles with question titles were more often downloaded, but less frequently cited (Jamali and Nikzad, 2011).

Of all the title characteristics, the intellectual content of article titles, as expressed in terms or concepts, should be the one most closely related to their success and impact. Some studies tried to characterize the usage of concepts (terms) in titles. For example, a recent study of word frequency distribution in nanoscience/nanotechnology confirmed the power law distribution of the frequency of words used in the titles (Bartol and Stopar, 2015). Another study quantified lexical diversity in titles using title phrases rather than individual words (Milojević, 2015) and found that the fields expanded their cognitive extent over time even when normalized for increased productivity. Other studies,

focused more closely on the sematic aspects of the words used in the titles. For example, Nagano (2009) analyzed word usage in four disciplines [history, sociology, economics (ECN), and education] and found history to be different from the others in terms of high usage of personal names, while sociology did not include a large number of unique terms. At the same time, medical literature titles that referred to a specific country (Jacques and Sebire, 2010), all the articles published by Italian researchers that included a country's name in the title (Abramo et al., 2016), and ECL articles that referred to specific names of study organisms in the titles (Fox and Burns, 2015) fared poorly in terms of impact, or in the case of ECL even in terms of being accepted for publication. Although amusing titles might be perceived as a good strategy to boost attention, studies have shown that they either did not have any citation advantage (Sagi and Yechiam, 2008; Subotic and Mukherjee, 2014). By analyzing word frequency of over 800,000 article titles from 18 different Scopus categories (excluding social sciences, arts and humanities), Thelwall (2017) found that usage of obscure (not frequently used) words in titles is associated with below average citation. The analysis of the diversity of ECN words in titles based on the ECN dictionary showed cyclical patterns (of varying length—30 or 40 years) of word usage (Guo et al., 2015). A study of 420 titles in medical journals found that the largest number of titles included topics related words, followed by the methods, with most articles lacking information on research design, methods, and results (Goodman et al., 2001).

Contemporary science is becoming more collaborative. This trend has naturally led some researchers to examine whether the change in authorship had an effect on title construction (e.g., Hudson, 2016). And in general, studies have found some differences in titles depending on the number of authors. For example, coauthored papers by UK researchers published in 2014 use colons and question marks less, which was explained by a difficulty to reach consensus about their usage (Hudson, 2016). At the same time, the titles of coauthored papers were found to be longer, which was attributed to a need to accommodate all the authors' views and needs (Hudson, 2016). Hudson also found that the titles of papers authored by a very large number of authors started resembling titles of papers authored by a small number of authors. This might be due to the fact that in, hyperauthorship scenario, only very small number of authors were actually responsible for making decisions regarding the paper. Namely, Hudson (2016) has found that the single-authored papers have the shortest titles, with the title length increasing with the increase of number of authors on a paper up to 25–49 authors and then declining. However, the positive correlation between title length (measured as the number of substantive words per title) and number of authors on papers was found not to be universal—while common in sciences, it was not present in social sciences, and there was a negative correlation in humanities (Yitzhaki, 1994).

Some studies have looked into the nationality of authors and have found that it was not always decisive in the construction of titles. While Lewison and Hartley (2005) found that the nationality of authors did not affect title characteristics, Ball (2009) found national and regional differences in the use of question marks in article titles over a 40-year period he studied. These discrepancies might be described by the different disciplines studied, because

researchers have found, analyzing the full text of articles, that disciplinary identity in medicine transcends national differences, while in ECN and linguistics the national writing tradition is more visible (Breivega et al., 2002; Dahl, 2004).

The aforementioned studies paint a complex and sometimes contradictory picture of the usage of titles in science. Few studies have analyzed multiple, diverse fields *and* multiple title characteristics with consistent methodology. Furthermore, most studies focused on possible correlations between titles and impact, paying less attention to the role that may be played by different characteristics of the authors. This study aims to disentangle the effects of time and discipline with those of the characteristics of the authors. We examine the characteristics of the titles and whether they have changed over a substantial period of time (half a century). We do so consistently for five diverse fields, in order to probe the usage of titles as a disciplinary identity building tool. The secondary objective is to establish whether different types of titles are used differently *today* by authors depending on their academic age, productivity, and collaboration levels, potentially important determinants that have not been explored before. Finally, we revisit the connection between title characteristics and impact.

DATA AND METHODS

We analyzed title characteristics in five scientific fields: astronomy (AST), mathematics (MAT), robotics (ROB), ECL, and ECN. These diverse fields have been chosen in order to allow us to capture a wide range of practices and discern whether certain practices are field-specific. AST is representative of a classical field with a range of knowledge production modes (small and large teams). MAT is different from other fields by being done individually or in small teams. ROB and ECL are representative of fields with more recent history and higher interdisciplinarity. Finally, ECN is representative of a field with a social component, unlike the other four.

The data consist of research articles published in up to 10 core journals in each of the five fields between 1961 and 2010 (from 1984 for ROB). The list of journal titles and the details of the selection criteria are given in Milojević (2012), which did not analyze the titles. Research articles are defined as publications classified as "article" or "conference paper." In several cases in which the journal changed its title, we collected data corresponding to all predecessor titles. In total, the data set contained records for 213,756 articles. **Table 1** shows breakdown by field. To study the

TABLE 1 | Summary of the data used in this study.

Field	Papers 1960–2010	Papers 2006–2010	Citations in 2013 to 2006–2010 papers	Unique lead authors 2006–2010
Astronomy	154,221	30,968	748,650	13,224
Ecology	33,052	5,409	129,038	3,965
Economics	18,807	1,097	17,195	885
Mathematics	21,602	1,754	14,664	1,470
Robotics	8,350	2,616	28,263	2,018

practices with respect to authors' characteristics, such as academic age, productivity, and number of collaborators and their relationship with title characteristics, we focused on the papers published in the most recent 5-year period (2006–2010). To explore correlations with impact, we obtained citations to these 2006–2010 articles received as of the end of 2013. All data were obtained from the Web of Science database.

Title Characteristics

In this study, we focus on five title characteristics: title length, titles with subtitles, titles posed as questions, titles with active verbs, and the cognitive extent (conceptual diversity) of titles.

Title Length

We define title length as a number of individual words [string of characters separated by space(s)] in a title. Hyphenated words count as one.

Titles with Subtitles (Multicomponent Titles)

We have identified the presence of subtitles automatically. The title is considered to have a subtitle if it has one of the following non-alphanumeric symbols: colon, period, or dash, anywhere in the title. Dash and hyphen use the same character, so we require it to be preceded and followed by a space, in order to differentiate it from hyphenated words.

Titles Posed As Questions

We have identified such titles automatically by detecting the presence of question mark in the title. However, many interrogative titles omit the question mark, so we additionally include all titles starting with: how, where, what, why, or which.

Titles Containing Active Verbs

We have automatically identified articles containing active verbs by matching them to a list of 1,000 common verbs, from which we have manually excluded 370 that are more often used in the same form as nouns or adjectives (for example: study, change, open). To this list, we added "has." The original list was obtained from an online resource.[1]

Note that a given title can fall in more than one of the three above categories.

Cognitive Extent

Conceptual diversity of an ensemble of titles is a new measure designed to quantify the cognitive extent of scientific literature (Milojević, 2015). It is defined as the number of unique phrases appearing in titles of a statistically large unit quota of literature (thousands of articles). It is similar to the lexical diversity used in computational linguistics to study richness of verbal expression (McKee et al., 2000; McCarthy and Jarvis, 2007; Koizumi, 2012), but uses phrases (concepts) instead of individual words. Bodies of literature that have more diverse concepts in titles will have a higher fraction of unique phrases and could be considered to cover larger cognitive extents. When possible, we use literature

---

[1]http://www.enchantedlearning.com/wordlist/verbs.shtml.

quota containing 10,000 title phrases, which is large enough not to suffer from the effects of non-linearity. Otherwise, we use 1,000 or 3,000 phrases and apply a statistical correction to relate it to the number of unique phrases in a quota of 10,000 phrases (Milojević, 2015).

Author Characteristics

Here, we wish to examine whether author characteristics impact the choices regarding titles. After all, titles are one of the tools for building professional identity and may reflect different characteristics of authors. Furthermore, it is reasonable to assume that in the case of coauthored articles, which are the norm in most disciplines, it is the lead author (first-listed, or corresponding author) who is expected to have had the greatest influence on the choice of the paper title. Therefore, in this study, we associate each author only to papers that he or she has led. We exclude articles that do not permit the identification of a lead author, such as when the authors are listed alphabetically. In order to examine the relationship between the characteristics of authors and the characteristics of their paper titles quantitatively, we follow (Milojević, 2012) and classify recent authors (those who have published between 2006 and 2010) by the following three properties.

Academic Age

Academic age is defined as the span in years, for a given author, between the first and the most recent article in the dataset, regardless of the role in the article (lead author or coauthor).

Individual Productivity

Number of articles published between 2006 and 2010 (5 years) by an author in the role of a lead author. Focusing on recent individual productivity decouples it from the effects of collaboration, which we explore separately.

Collaboration Level

Collaboration level is defined as the number of different lead authors on recent articles on which the author was listed as a coauthor. To decouple this measure from productivity, we do not count coauthors on articles on which the author him/herself was the lead author. Note that this measure is very different from the number of coauthors on a paper (team size), which would be the characteristic of a paper, not of the lead author.

We disambiguated author names using the hybrid method (Milojević, 2013), which considers only the first initial for infrequent last names (thus avoiding splitting due to the inconsistent reporting of the middle initial), and both initials for more common last names (thus providing better disambiguation for cases when the mix-up is more likely).

We have performed and will discuss the analysis regarding titles and author characteristics for all five fields, but in order to make the graphs more legible, we will show in them only for AST, where the current dataset is the largest and the trends are best revealed. Trends are fit with a linear function and the resulting slopes and their errors are presented in order to evaluate the significance of trends (assuming they are linear). Trends are considered significant if the absolute values of slopes are more

than three times their errors. We will mention any disciplinary particulars if warranted by the analysis.

Paper Characteristics

Previous studies have looked extensively at a relation between the characteristics of the papers (e.g., the number of authors) and their titles and impact, with mixed results. Here, we revisit that question examining the relation between title characteristics and the impact of a paper.

Impact

Impact is defined as all citations received by recent papers (2006–2010) as of the end of 2013. The citation window varies from 4 to 9 years, which is not ideal, but is still reasonable, and does not produce spurious relations.

RESULTS

We analyzed five title characteristics: length, proportion of sub-titles, proportion of question titles, proportion of assertive titles, and cognitive extent over the period of 50 years (1961–2010) in five disciplines. We also analyzed the same characteristics against author characteristics, such as academic age, productivity and collaboration level, and paper citation for the most recent 5-year period (2006–2010).

Title Length

As discussed above, the title length is an important characteristic in that it needs to strike a balance between informativeness (i.e., providing enough keywords for retrieval purposes) and attractiveness (with shorter titles being able to keep the attention on them longer). **Figure 1A** shows that the average length of titles is strongly discipline dependent. Currently, the average length ranges from 8 words in MAT and economy, to 13 words in ECL. Trends show that, on average, the titles have been increasing in length in AST, MAT, ROB, and ECL. ECN is the only field where the titles have not increased in size in recent decades. In relative sense, the greatest expansion has been seen in AST, with the most recent titles being 40% longer, on average, than the titles in 1960.

To test whether author characteristics have strong influence on self-representation *via* titles, we examine the characteristics of titles for the period 2006–2010. We first look at the academic age of the lead author and see that in AST (**Figure 1B**), on average, authors of different academic ages use titles of the same length. A small downward trend is present in ECL and MAT. We do not find a correlation between the productivity and title length in AST (**Figure 1C**) or other fields. However, there is a statistically significant positive correlation between the level of collaboration and title length, at least in AST where the collaboration is most extensive (**Figure 1D**). There, the authors who collaborate with large number of diverse individuals tend to have longer titles (13 versus 11).

Also, we found that titles with more citations tend to have, on average, longer titles, in AST (**Figure 1E**), and to some extent in economy, but not in other three fields. Even for fields where there is a correlation on average, the range of title lengths at any citation level is large (90 percentile range is between 6 and 20 words

almost irrespective of citation), which casts doubt on the ability to use extensiveness of the title as an attention grabbing strategy. It has been shown that papers having more authors tend to have higher citation rates. To separate possible influence of this factor, we also look at the trend between title length and citations based only on papers having three to five authors, and find that in that case the trend is diminished.

To summarize, our analysis has shown that between-field differences dominate over the within-field differences at a given time, whereas the trends over time in some cases reach the level of between-field difference.

Titles with Subtitles or Multiple Components

Previous studies have not directly examined the prevalence of titles with subtitles. However, the studies have examined the usage of different non-alphanumeric symbols which in most cases (except the usage of question mark at the end of the title) corresponds to the presence of subtitles. One of the most common ways to introduce subtitles is *via* a colon. Previous studies have found the usage of colons to be more prevalent than the usage of other non-alphanumeric symbols, but different across the fields. We find that subtitles are least common in MAT (currently 5%) and most common in AST and ECL (around 30%, **Figure 2A**). The usage of subtitles in different subfields followed different trends over the last 50 years. It has increased in AST and ECL (from ~20 to ~30%), while it stayed more constant in other fields. One may wonder if the increased usage of subtitles is what drives the overall lengthening of titles in AST and ECL. To this end, we have calculated the trends like the ones shown in **Figure 1A**, but only keeping the longest single component of a title. We found that the length of titles in AST and ECL has been increasing nevertheless.

When we examine presence of subtitles for different groups of authors, focusing on the most recent 5-year period, we see that the usage of subtitles in AST is pretty stable, around 33% for authors of all ages, except for the ones just starting, where it is 25% (**Figure 2B**). There is also no overall correlation between author productivity and the usage of subtitles in AST (**Figure 2C**), which is similar to the situation in other fields. However, in AST, the usage of subtitles is positively correlated with the number of unique collaborators a researcher has had, reaching 70% of all titles for the authors who had 100 collaborators in the 5-year period (**Figure 2D**).

We have also found that the usage of subtitles is positively correlated with number of citations in AST, with almost 50% of the most highly cited papers having subtitles, while less than 30% of less highly cited papers having them (**Figure 2E**). The trend is weaker, but remains significant when only papers with three to five authors are selected. Positive correlations are seen in other fields, except in ECL.

As in the case of title length, we conclude that between-field differences in the usage of subtitles dominate over the trends with time, and over some of the author characteristics. However, the trends with respect to the level of collaboration and with respect to citedness can be as strong, at least in some fields.
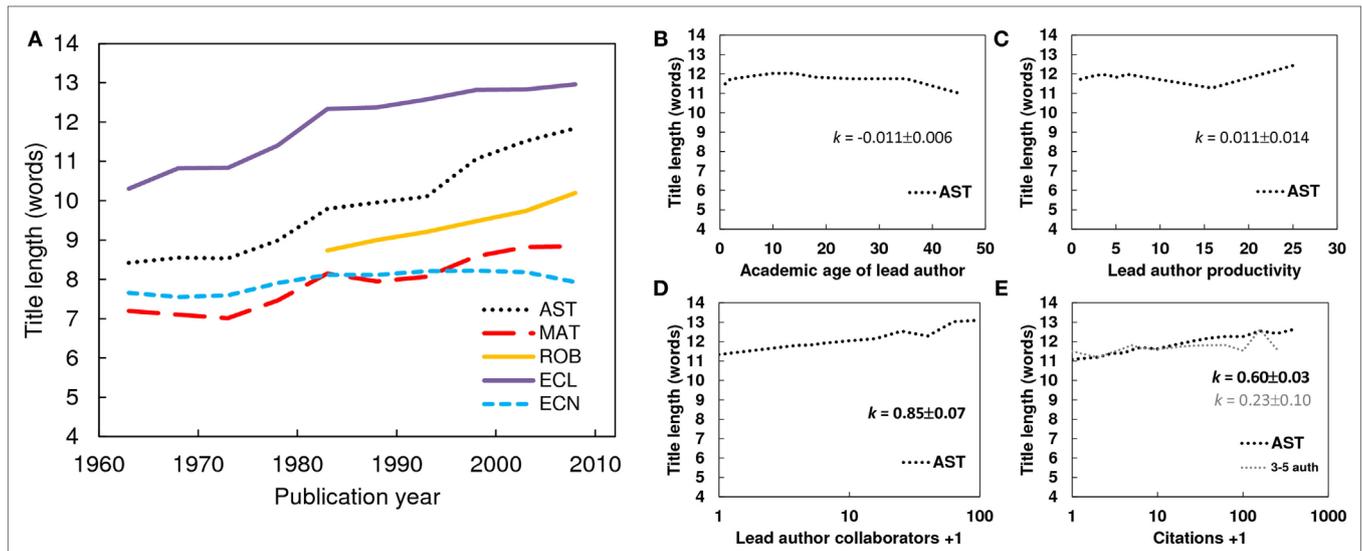
FIGURE 1 | Trends involving title length (measured in words). (A) Average title length for five disciplines [astronomy (AST), mathematics (MAT), robotics (ROB), ecology (ECL), and economics (ECN)] from 1961 to 2010. Data for ROB start in 1983. Data points are averaged in bins of 5 years. **(B)** Average title length for authors of different academic age (number of years spent in the field) in AST from 2006 to 2010. Data points are averaged in bins of ranging from 1 to 10 years (bin size increases with age). **(C)** Average title length for authors of different recent productivity in AST from 2006 to 2010. **(D)** Average title length for authors of different recent collaboration level in AST from 2006 to 2010. **(E)** Average title length for papers of different impact, measured by citations received through the end of 2013. For panels **(C–E)**, data are binned in logarithmic intervals of 0.2 decades. Slopes of linear fits and their errors are given for trends in panels **(B–E)**. Significant trends are emphasized by bold script.
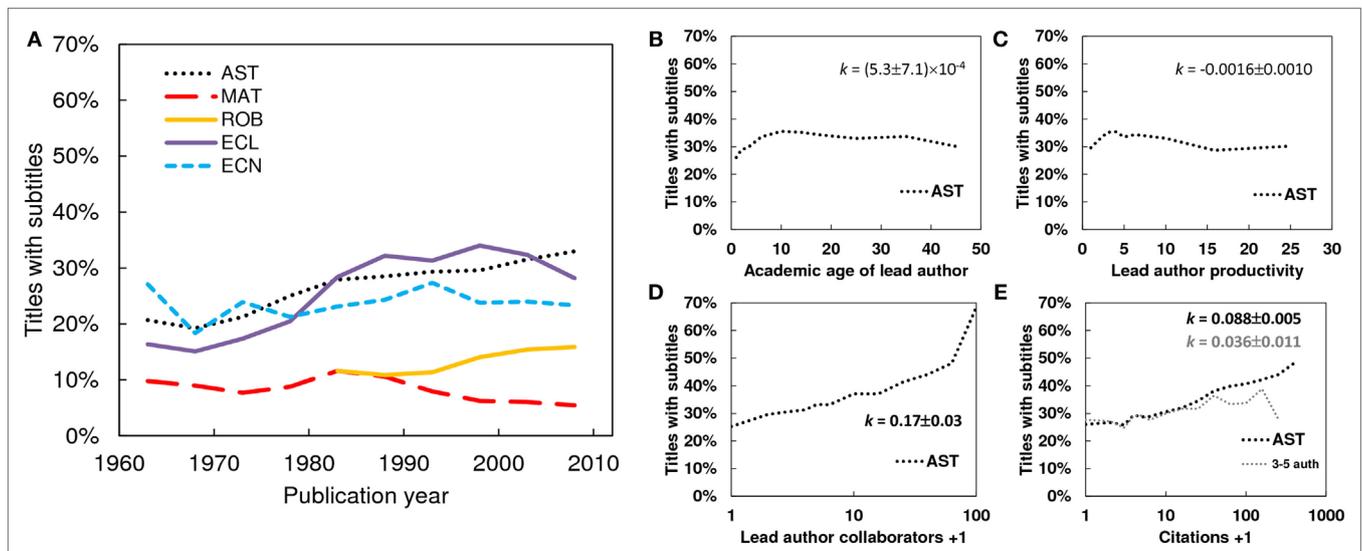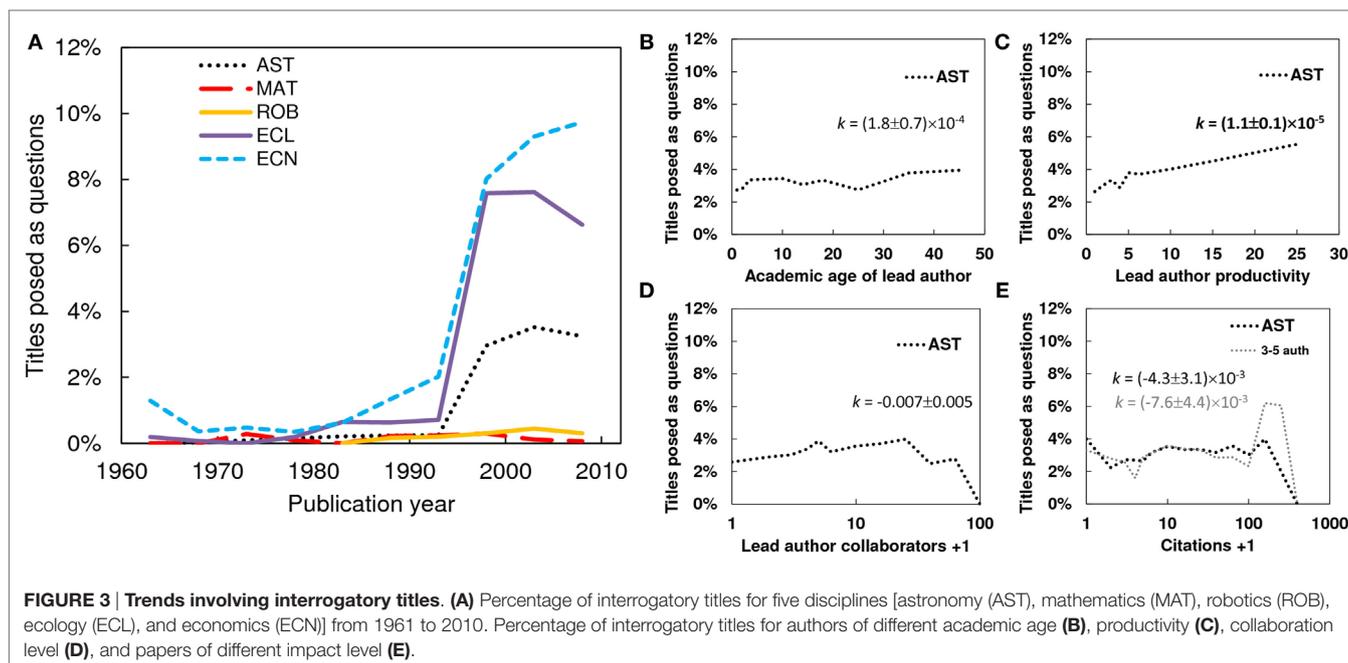

FIGURE 2 | Trends involving the presence of subtitles in titles. (A) Percentage of titles with subtitles for five disciplines [astronomy (AST), mathematics (MAT), robotics (ROB), ecology (ECL), and economics (ECN)] from 1961 to 2010. Percentage of titles with subtitles for authors of different academic age **(B)**, productivity **(C)**, collaboration level **(D)**, and papers of different impact level **(E)**.

Interrogatory Titles

Previous studies have found the increase of interrogatory titles over time, and yet, these titles still constituted only a small portion of all titles. For example, such titles were not found to be common in computer science (Anthony, 2001). Our study confirms these findings (**Figure 3A**). In the period between 1960

and mid-1990s, interrogatory titles were almost non-existent in all the fields we examined except in ECN, where close to 1% of titles posed a question. Today, ECN is the field with the largest proportion of titles posed as questions (~10%), followed by ECL (~7%) and AST (~3%). Interrogatory titles remain non-existent in MAT and ROB. It is interesting that in the three fields that

FIGURE 3 | Trends involving interrogatory titles. **(A)** Percentage of interrogatory titles for five disciplines [astronomy (AST), mathematics (MAT), robotics (ROB), ecology (ECL), and economics (ECN)] from 1961 to 2010. Percentage of interrogatory titles for authors of different academic age **(B)**, productivity **(C)**, collaboration level **(D)**, and papers of different impact level **(E)**.

found usage of questions in titles acceptable, the surge in their usage occurred at the same time, in the mid-1990s. The growth in their usage has mostly leveled just a few years after the surge, and there may be some decline most recently.

When it comes to how different groups of authors use interrogatory titles, in AST we find no correlation between academic age and using questions (**Figure 3B**), and a positive correlation between usage of questions and productivity, with ~6% of titles of most productive authors having questions compared to ~3% of the least productive ones (**Figure 3C**). There are indications of such correlation other fields as well. We find no significant correlation with respect to the level of collaboration in AST (**Figure 3D**) or other fields. Interestingly, posing title as a question does not seem to be more prevalent in highly cited articles (**Figure 3E**) in any field.

We conclude that when it comes to titles posed as questions, it is a recent practice, but one which has taken hold at various levels in various disciplines, from none to ~10%. Inter-field differences among different groups of authors or different impact levels are small.

Declarative (or Informative) Titles

The usage of declarative titles, the ones that place the result in the title in the form of a sentence (thus containing an active verb), has been somewhat controversial, especially in medical literature. Such titles appeared in biological sciences in 1970s and have been on the rise (Rosner, 1990). We find that declarative titles appeared in the five fields that we examined here much later, in the mid-1990s (**Figure 4A**). Interestingly, this is the same time that saw the rise of interrogatory titles. Declarative titles have experienced very strong rise in usage in ECL, which currently contains the largest percentage of such titles (14%) and is still on the rise, and ECN, which at the peak usage had ~11%, but experienced a

drop starting in 2000, to be at 8% at the moment. Both ROB and MAT (which never use interrogative titles) have experienced an increase in usage of declarative titles, although at a smaller scale than ECL and ECN, and both saw the usage of declarative titles decrease in mid 2000s. Declarative titles were the least prevalent in AST, never reaching even 1% of titles.
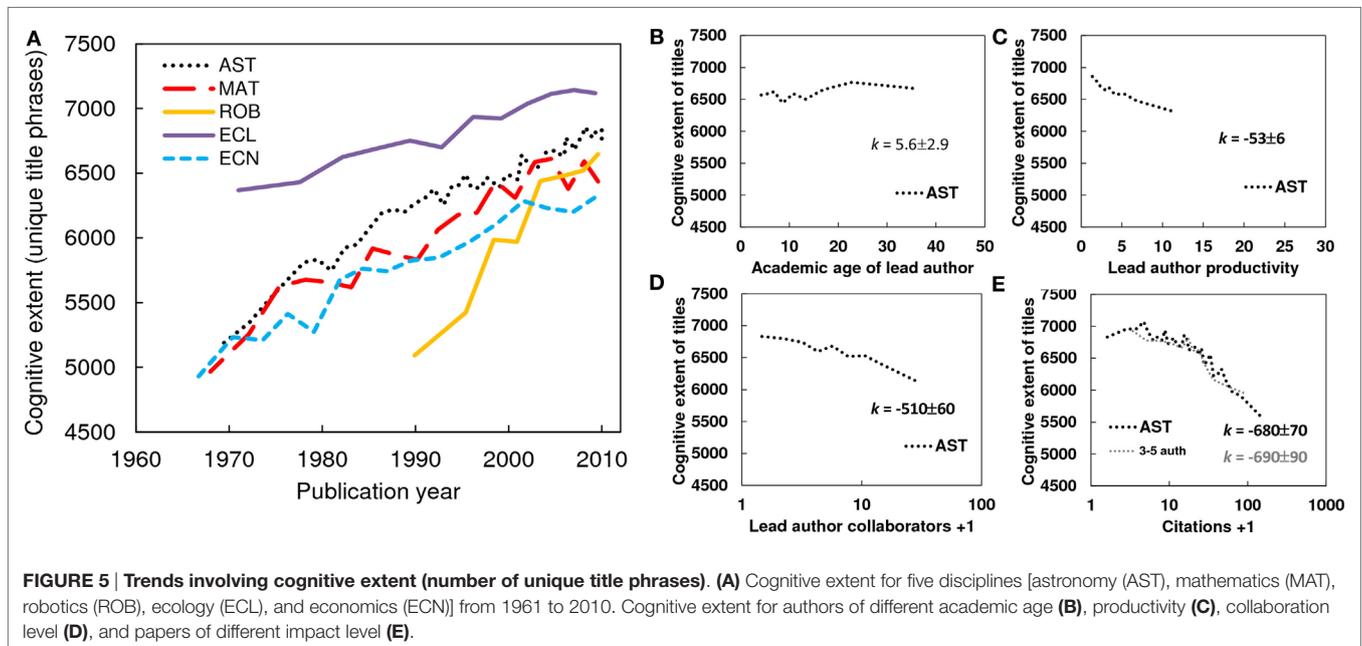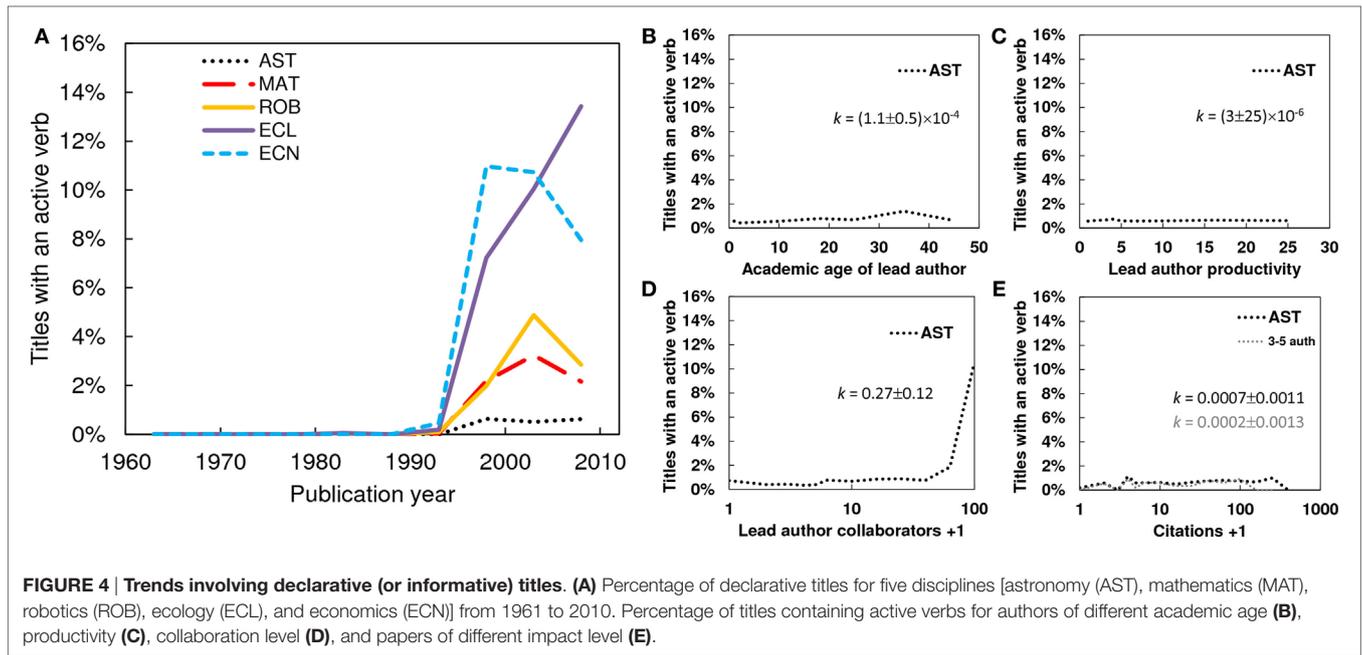
When it comes to differences in usage among different groups of authors in AST, we find no significant correlations with the academic age (**Figure 4B**) or a level of productivity (**Figure 4C**). This also holds for other fields. We also find that the biggest users of this type of titles are authors with very large number of unique collaborators (**Figure 4D**) (the like of which are not present in other fields). Even 10% of articles written by authors who had 100 unique collaborators in the 5-year period we studied used declarative titles.

While the usage of declarative titles may appear to offer a special advantage in grabbing readers' attention, we do not find that the prevalence of such titles is different for articles of different citation level, in AST (**Figure 4E**), or in other fields.

Again, we conclude that once the practice of declarative titles has entered fields other than biomedicine, it has established itself at varying levels depending on the field, with little intra-field difference.

Cognitive Extent

Cognitive extent is a measure of lexical diversity of titles, measuring the number of unique title phrases over set quotas of text (Milojević, 2015). As such, this is not the characteristic of any given title (as in the case of previous properties studied here) but the characteristic of an ensemble of titles. It is interesting that all five fields have been experiencing the expansion of their cognitive extents, although not at the same rate (**Figure 5A**). ECL has had the largest cognitive extent in all the periods, although it had the

FIGURE 4 | Trends involving declarative (or informative) titles. (A) Percentage of declarative titles for five disciplines [astronomy (AST), mathematics (MAT), robotics (ROB), ecology (ECL), and economics (ECN)] from 1961 to 2010. Percentage of titles containing active verbs for authors of different academic age **(B)**, productivity **(C)**, collaboration level **(D)**, and papers of different impact level **(E)**.

FIGURE 5 | Trends involving cognitive extent (number of unique title phrases). (A) Cognitive extent for five disciplines [astronomy (AST), mathematics (MAT), robotics (ROB), ecology (ECL), and economics (ECN)] from 1961 to 2010. Cognitive extent for authors of different academic age **(B)**, productivity **(C)**, collaboration level **(D)**, and papers of different impact level **(E)**.

slowest growth rate. AST has the second largest cognitive extent, but it has experienced faster growth in 1970s than today. ROB, which now has the third largest cognitive extent has undergone the period of fastest expansion since 1990s (when we start following it). MAT and ECN have had similar cognitive extents and MAT, much like AST, has experienced accelerated growth in the 1970s.

When it comes to the cognitive extent and author characteristics, in AST (and in other fields) we find no significant correlation between the academic age and cognitive extent. Thus, authors of all ages maintain the same diversity of topics that they study (**Figure 5B**). There is a negative correlation between the cognitive extent and productivity in all fields including AST (**Figure 5C**), in the sense that very productive authors tend to work on a more focused set of concepts. The same is true for authors who collaborate more, in AST (**Figure 5D**) and other fields. Finally, we find the negative correlation between cognitive extent and citations papers accrue, in AST (**Figure 5E**) and other fields.

For the case of cognitive extent, we find that intra-field and between-field variations are comparable to each other and to changes that occur over time.

DISCUSSION AND CONCLUSION

Article titles are a rich source of data that can be used not only to elucidate the development of disciplines, fields, and research areas, identify trendy topics or subfields, but to help us better understand the processes of knowledge creation. Namely, a large number of studies have found titles to be the product of interaction between both individual and disciplinary identity building. While there were a number of studies that explored a particular aspect of titles for particular fields at particular times, mostly focusing on correlations with impact, the present study has simultaneously and consistently explored (a) a range of disciplines over long periods of time, (b) a suite of title characteristics, and (c) a number of variables related to author characteristics.

We found that belonging to some discipline is the strongest determinant for the average length of the titles and the occurrence of different forms of titles. This suggests that authors try to comply with the norms set in their fields. However, these norms are not fixed in time. Over the decades, the practices have changed. The increase in length over time has been well documented and should come as no surprise. As the field evolves, the studies become more detailed, and require more words to distinguish them with respect to previous studies, which were probably more general in scope. We have shown that this is the result of more detailed/specific description of article contents, and not the mere inclusion of subtitles.

The most drastic example of change in practices is the nearly simultaneous appearance, across different fields, of titles of interrogatory and declarative types in the mid-1990s, suggesting a common origin, perhaps influenced by the practices in prestigious multidisciplinary journals. We are in the epoch of increased overall assertiveness in title-giving practices.

Overall, we confirmed the importance of both the discipline and time period on the construction of titles. Interestingly, we found that not only the authors follow one another in their title choices (Nagano, 2015) within disciplines but that certain trends emerged in a whole range of disciplines at the same time. This is worth further study.

While individual groups of authors often did not differ in their practices regarding the use of titles, occasionally they did. Most notably, we find that authors with many collaborators tend to use more subtitles. Authors with many collaborators are often members of large teams, whose papers tend to be published in installments, which thus necessitate the use of subtitles. Also, the authors with a lot of collaborators tend to work on a smaller set of topics, as a group, than the ones who tend to work in smaller collaborations or alone. More productive authors tend to use interrogatory titles more often, suggesting that more productive authors tend to publish on more speculative topics. In some fields, older authors use shorter titles, perhaps reflecting the fact that these authors have started their careers when titles used to be shorter.

It would be interesting to further explore whether these differences are truly due to different author characteristics or due to authors with certain characteristics being more prevalent in particular subfields which may have differing practices.

A topic that has received much interest is whether the titles of articles with greater impact tend to have special characteristics. We find that the prevalence of what may be considered more aggressive or assertive titles (the ones posed as questions or which state the result) is actually similar among the articles having vastly diverse impacts (from zero to hundreds of citations), suggesting that the use of such strategies does not confer any citation benefit. More highly cited papers do have longer titles and more often contain subtitles in AST and ECL, but this does not necessarily imply a causal connection. Longer titles are the characteristic of articles produced by larger teams, which tend to be cited more. Furthermore, no such correlation is seen in MAT, ECN, and ROB. In any case, it would be interesting to see whether in instances when certain title characteristics were correlated with impact, such a correlation was due to either paper or title characteristics, as opposed to the a fact that certain subfields or modes of doing science may have particular title preferences or even constraints, and that at the same time these subfields are moving the research front, therefore accruing more citations quicker.

Interesting, in all the fields, we find that the higher the impact, the narrower the diversity of title concepts. This means that while the papers covering all the topics can do poorly, a more restricted set of topics does very well in terms of citations. This may be the effect of very active subfields or a small number of "hot" topics pushing the research frontier.

AUTHORS NOTE

AUTHOR CONTRIBUTIONS

SM designed the study, performed the analyses, and wrote the manuscript.

REFERENCES

Abramo, G., D'Angelo, C. A., and Di Costa, F. (2016). The effects of a country's name in the title of a publication on its visibility and citability. *Scientometrics* 109, 1895–1909. doi:10.1007/s11192-016-2120-1

Anthony, L. (2001). Characteristic features of research article titles in computer science. *IEEE Trans. Prof. Commun.* 44, 187–194. doi:10.1109/47.946464

Aronson, J. (2010). When I use a word… Declarative titles. *Q. J. Med.* 103, 207–209. doi:10.1093/qjmed/hcp084

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, 667–679. doi:10.1007/s11192-007-1984-5

Bartol, T., and Stopar, K. (2015). Nano language and distribution of article title terms according to power laws. *Scientometrics* 103, 435–451. doi:10.1007/s11192-015-1546-1

Bazerman, C. (1985). Physicists reading physics, schema-laden purposes and purpose-laden schema. *Writ. Commun.* 2, 3–23. doi:10.1177/0741088385002001001

Bazerman, C. (1988). *Shaping Written Knowledge: The Genre and Activity of the Experimental Article in Science*. Madison: University of Wisconsin Press.

Becher, T., and Trowler, P. R. (2001). *Academic Tribes and Territories: Intellectual Enquiry and the Culture of Disciplines*. Buckingham: The Society for Research into Higher Education & Open University Press.

Bowker, G. C. (2005). *Memory Practices in the Sciences*. Cambridge, MA: The MIT Press.

Breivega, K. R., Dahl, T., and Fløttum, K. (2002). Traces of self and others in research articles. A comparative pilot study of English, French and Norwegian research articles in medicine, economics, and linguistics. *Int. J. Appl. Linguist.* 12, 218–239. doi:10.1111/1473-4192.00032

Buter, R. K., and van Raan, A. F. J. (2011). Non-alphanumeric characters in titles of scientific publications: an analysis of their occurrence and correlation with citation impact. *J. Informetrics* 5, 608–617. doi:10.1016/j.joi.2011.05.008

Callon, M., Courtial, J. P., Turner, W. A., and Bauin, S. (1983). From translations to problematic networks: an introduction to co-word analysis. *Soc. Sci. Inf.* 22, 191–235. doi:10.1177/053901883022002003

Cook, J. M., and Plourde, D. (2016). Do scholars follow Betteridge's Law? The use of questions in journal article titles. *Scientometrics* 108, 1119–1128. doi:10.1007/s11192-016-2030-2

Dahl, T. (2004). Textual metadiscourse in research articles: a marker of national culture or of academic discipline? *J. Pragmatics* 36, 1807–1825. doi:10.1016/j.pragma.2004.05.004

Demarest, B., and Sugimoto, C. R. (2015). Argue, observe, assess: measuring disciplinary identities and differences through socio-epistemic discourse. *J. Assoc. Inf. Sci. Technol.* 66, 1374–1387. doi:10.1002/asi.23271

Didegah, F., and Thelwall, M. (2013). Which factors help authors produce the highest impact research? Collaboration, journal and document properties. *J. Informetrics* 7, 861–873. doi:10.1016/j.joi.2013.08.006

Fox, C. W., and Burns, C. S. (2015). The relationship between manuscript title structure and success: editorial decisions and citation performance for an ecological journal. *Ecol. Evol.* 5, 1970–1980. doi:10.1002/ece3.1480

Goodman, D. J. (2010). Novel tool constitutes a paradigm: how title words in medical journal articles have changed since 1970. *J. Eur. Med. Writers Assoc.* 19, 269–271.

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

Guo, S., Zhang, G., Ju, Q., Chen, Y., Chen, Q., and Li, L. (2015). The evolution of conceptual diversity in economics titles from 1890 to 2012. *Scientometrics* 102, 2073–2088. doi:10.1007/s11192-014-1501-6

Hudson, J. (2016). An analysis of the titles of papers submitted to the UK REF in 2014: authors, disciplines, and stylistic details. *Scientometrics* 109, 871–889. doi:10.1007/s11192-016-2081-4

Hyland, K. (2004). *Disciplinary Discourses: Social Interactions in Academic Writing*. Ann Arbor: The University of Michigan Press.

Hyland, K. (2012). *Disciplinary Identities: Individuality and Community in Academic Discourse*. Cambridge: Cambridge University Press.

Jacques, T. S., and Sebire, N. J. (2010). The impact of article titles on citation hits: an analysis of general and specialist medical journals. *JRSM Short Rep.* 1, 2. doi:10.1258/shorts.2009.100020

Jamali, H. R., and Nikzad, M. (2011). Article title type and its relation with the number of downloads and citations. *Scientometrics* 88, 653–661. doi:10.1007/s11192-011-0412-z

Koizumi, R. (2012). Relationship between text length and lexical diversity measures: can we use short texts of less than 100 tokens? *Vocabulary Learn. Instr.* 1, 60–69. doi:10.7820/vli.v01.1.koizumi

Latour, B., and Woolgar, S. (1986). *Laboratory Life: The Construction of Scientific Facts*. Princeton: Princeton University Press.

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

McCarthy, P. M., and Jarvis, S. (2007). vocd: a theoretical and empirical evaluation. *Lang. Test.* 24, 459–488. doi:10.1177/0265532207080767

McGowan, J., and Tugwell, P. (2005). Informative titles described article content. *J. Can. Health Libr. Assoc.* 26, 83–84. doi:10.5596/c05-029

McKee, G., Malvern, D., and Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Lit. Linguist. Comput.* 15, 323–337. doi:10.1093/llc/15.3.323

Milojević, S. (2012). How are academic age, productivity and collaboration related to citing behavior of researchers? *PLoS ONE* 7:e49176. doi:10.1371/journal.pone.0049176

Milojević, S. (2013). Accuracy of simple, initials-based methods for author name disambiguation. *J. Informetrics* 7, 767–773. doi:10.1016/j.joi.2013.06.006

Milojević, S. (2015). Quantifying the cognitive extent of science. *J. Informetrics* 9, 962–973. doi:10.1016/j.joi.2015.10.005

Nagano, R. L. (2009). Lexical comparison of journal article titles in soft disciplines. *Porta Lingua* 2009, 111–117.

Nagano, R. L. (2015). Research article titles and disciplinary conventions: a corpus study of eight disciplines. *J. Acad. Writing* 5, 133–144. doi:10.18552/joaw.v5i1.168

Nair, L. B., and 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, 1331–1359. doi:10.1007/s11192-016-1937-y

Rosner, J. L. (1990). Reflections of science as a product. *Nature* 345, 108. doi:10.1038/345108a0

Sagi, I., and Yechiam, E. (2008). Amusing titles in scientific journals and article citation. *J. Inf. Sci.* 34, 680–687. doi:10.1177/0165551507086261

Subotic, S., and Mukherjee, B. (2014). Short and amusing: the relationship between title characteristics, downloads, and citations in psychology articles. *J. Inf. Sci.* 40, 115–124. doi:10.1177/0165551513511393

Thelwall, M. (2017). Avoiding obscure topics and generalising findings produces higher impact research. *Scientometrics* 110, 307–320. doi:10.1007/s11192-016-2159-z

van Wesel, M., Wyatt, S., and ten Haaf, J. (2014). What a difference a colon makes: how superficial factors influence subsequent citation. *Scientometrics* 98, 1601–1615. doi:10.1007/s11192-013-1154-x

Vygotsky, L. S. (1978). *Mind in Society*. Cambridge, MA: Harvard University Press.

Wager, E., Altman, D. G., Simera, I., and Toma, T. P. (2016). Do declarative titles affect readers' perceptions of research findings? A randomized trial. *Res. Integr. Peer Rev.* 1, 11. doi:10.1186/s41073-016-0018-3

White, A. (1991). A further exploration of title size and author number. *J. Am. Soc. Inf. Sci.* 42, 384–385. doi:10.1002/(SICI)1097-4571(199106)42:5<384::AID-ASI9>3.0.CO;2-G

Yitzhaki, M. (1994). Relation of title length of journal articles to number of authors. *Scientometrics* 30, 321–332. doi:10.1007/BF02017231