Annals of Library and Information Studies
Vol. 62, September 2015, pp. 126-132

The impact of title length and punctuation marks on article citations

Mohammad Reza Falahati Qadimi Fumani[a], Marzieh Goltaji[b], and Pardis Parto[c]

[a]Assistant Professor of Computational Linguistics Research Department, Regional Information Center for Science and Technology (RICeST), Shiraz, Iran,
Email: mrfalahat@yahoo.com

[b]The Islamic World Science Citation Center (ISC), Shiraz, Iran,
Email: marzieh.goltaji@gmail.com

[c]Department of Evaluation and Collection Development, Regional Information Center for Science and Technology (RICeST), Shiraz, Iran,
Email: parto.pardis@gmail.com

Received 27 January 2015, revised 07 August 2015, accepted 18 August 2015

In literature, a number of factors have been reported as affecting citations to articles. This paper is intended to find out if the number of citations made to an article was associated with title length and number of punctuation marks in titles. The article also attempts to determine the most/least frequently used punctuation marks in article titles, and to study titling features in top-30 articles in the dataset examined. To extract the data, ISI Web of Science was searched using the fields, source 'SO = Scientometrics' and publishing year 'PY = 2009-2011'. In all, 650 article titles constituted the sample data of the study. To determine title length in words and in characters (with no spaces), the 'word count' option – 'statistics' – in Microsoft Word was used. The types and number of punctuation marks were extracted using the 'find and replace' option in Microsoft Word. Data analysis revealed that the two variables title length and citations to articles were not correlated; the number of punctuation marks could not be used as a variable to predict an article's citation rate with certainty; 'colon', 'hyphen' and 'coma' were most frequent while 'semi-colon', 'dash' and 'single quotation marks' were least frequent punctuation marks; out of the 30 top titles, 23 included at least one punctuation mark; from the top-10 titles, 9 titles encompassed a punctuation mark and the article receiving the highest amount of citations along with 5 more articles had a question mark in their titles. The average title length in characters with no spaces was 73 in the top-30 titles.

Keywords: citation analysis; title length; punctuation marks; text analysis; scientometrics; citations

Introduction

Scientometric research, a field of information and communication sciences devoted to quantitative studies of science has increased abundantly in recent years[1]. From among different subdomains of the field, scientometric analysis of document titles, in general, and article titles, in particular, has received special attention by researchers. Within this domain, many have scrutinized the impact of variables like title length, type and punctuation marks on citations made to articles. Falagas, Kouranos, Arencibia-Jorge and Karageorgopoulos[2] viewed citation rate as an important factor in determining the scientific impact of a journal. Cheek, Garnham and Quan[3], as cited in Falagas, Zarkali, Karageorgopoulos and Bardakas[4], considered citation as "… a measure of the scientific recognition the study has received".

Jamali and Nikzad[5] observed a significant difference among articles with different types (descriptive, declarative and interrogative) of titles in terms of downloads and citations. They reported, "Articles with interrogative titles tend to be downloaded more but cited less than the others … longer titles are downloaded slightly less than articles with shorter titles … titles with colon tend to be longer and receive fewer downloads and citations".

Unlike Jamali and Nikzad[5], Habibzadeh and Yadollahie[6] observed, "… articles with longer titles receive more citations". They further asserted, "We think that longer titles are mainly those which include the study methodology and/or results in more detail and thus, attract more attention and citations." Similarly, Jacques and Sabire[7] reported that title length and citations were positively correlated. Moore[8] also reported a similar finding. He asserted that longer titles were advantageous because they enjoyed better visibility in databases.

Due to such opposing views, Huggett[9] studied *Cell* scholarly papers between 2006 and 2010, and analyzed their citations within the same window. She concluded:

Overall, there was no direct correlation between title length (measured in number of characters) and total citations … papers with titles between 31 and 40 characters were cited the most. There were also differences in average number of citations per paper depending on the punctuation used in the titles: for instance, the few papers with question marks in their titles were clearly cited less, but titles containing a comma or colon were cited more. There were no papers with a semicolon in their title, and only one (uncited) paper with an exclamation mark in its title … the ten most cited papers in *Research Trends'* case study did not contain any punctuation at all in their titles.

This last piece of result reported by Huggett[9] was against van Wesel, Wyatt and Haaf's[10] finding since the latter researchers viewed punctuation marks, especially 'colon', a contributing factor to the rate of citations made to an article.

Some researchers scrutinized the impact of disciplines on citation rates. Hartley[11,12] observed differences between disciplines in terms of the application of colon in article titles. Nevertheless, unlike Huggett[9] he reported that use of colon in article titles had no impact on their citation rates. Gustavii[13] reported that question titles were more acceptable for review articles than research papers. Similarly, Yitzhaki[14] confirmed such variations within different disciplines.

Some researchers have endeavored to unveil factors contributing to the rate of document citations. Falagas, Zarkali, Karageorgopoulos and Bardakas[4] examined factors associated with the number of citations of published articles. They studied the impact of article length on the number of future citations and concluded that besides journal impact factor, article length and number of authors independently predicted the citation rate. In this regard, Jacques and Sabire[7] emphasized, "The citation rate of an article depends on a large number of factors including the significance and availability of the journal in which it is published, publication type, its subject, etc.". They referred to length of the title and author gender as other contributing factors. Rostami, Mohammadpoorasl and Hajizadeh[15] assessed the association between some features of articles title and number of citations in a volume of *Addictive Behavior*. They found that articles with combinational title (use of a hyphen or a colon separating different ideas within a sentence) and articles with different words in the keywords (at least two different keywords) had higher number of citations. Paiva, Lima and Paiva[16] concluded that articles with shorter titles describing the results were cited more.

Some researchers tackled the issue of what a good title should look like. Soler[17] asserted that the title of an article should clearly indicate the subject, arouse interest, be rather short, informative and attractive. In Kane's[18] terms, as cited in Jamali and Nikzad[5], "It is difficult to balance these qualities and most titles appear to be attractive but not informative or informative but not attractive". According to Manten and Greenhalgh[19], a title should reflect the main content of an article. On features of titles, Ball[20] stated that each title should motivate the reader to read an article and give the readers a summary of the contents. Jacques and Sabire[7] also asserted, "The construction of an article title has a significant impact on the frequency with which a paper is cited".

Some researchers incorporated, in their studies, some control variables as well. de Winter[21], for instance, kept the number of article views constant. He observed that the number of tweets was weakly associated with the number of citations and weakly negatively associated with citations when the number of article views was held constant. The number of tweets was predictive of other social media activity but not of the number of article views on PubMed Central.

Based on the review above, the researchers studied the relationship between title length and punctuation marks, on the one hand, and the rate of citations an article receives, on the other.

Objective of the study

The broad objective of the study is to assess the impact of title length and punctuation marks on article citations. Specific objectives are:

- To investigate the relationship between the rate of citations made to an article and its title length-in-words;
- To study the relationship between the rate of citations made to an article and the number of punctuation marks in its title;
- To study the use of punctuation marks that appear in titles of articles; and
- To examine titles of top 30 highly cited articles.

Method

Data of the study

To collect the data, ISI Web of Science was searched using the source field 'SO = Scientometrics' and publishing year 'PY = 2009-2011' (each year was surfed separately). In all, 650 article titles constituting 192, 233 and 225 titles for the years 2009, 2010 and 2011 respectively were used for the study. For each title, two types of data were collected. The first data type (the number of citations each article had received) was extracted from Web of Science while the second data type was computed by the authors. This latter data included title length in words, title length in characters with no spaces, types of punctuation marks used in titles and the number of punctuation marks used in each title. To determine title length in words and title length in characters with no spaces, the 'word count' option – 'statistics' – in Microsoft Word was used. From among the six features available, two namely 'words' and 'characters (no spaces)' were used to compute title length. Regarding 'type of punctuation marks' and 'number of punctuation marks', they were extracted by the researchers using the 'find and replace' option in Microsoft Word. All the data were retrieved on April 14, 2014.

To reduce the possibility of the raters' mistakes in extracting the type and number of punctuation marks, inter-rater reliability was employed, that is, two raters – the first two authors of the present paper – extracted the data each independently and then compared their lists together. In case, there were discrepancies in the data extracted, the two raters discussed it together and tried to settle their differences. Otherwise, they forwarded the case to the third rater whose judgment was considered final. Having generated the raw data, they were input into SPSS (Version 20) for further analysis.

Analysis

Relationship between rate of citations and title length

For examining the relationship between rate of citations and title length, a hypothesis was formulated as: "The longer the title length the higher the number of citations made to an article".

As indicated in Table 1, the sample data (650 articles) under study received 4144 citations in aggregate. The number of citations ranged between 0 and 72 with a mean of 6 citations per paper (X=6.37, SD=7.35). The median obtained was 4 meaning that titles receiving up to 4 citations comprised 50 percent of the whole titles. Further, titles with two citations were most abundant (85 cases, 13.1%) in the dataset.

Regarding the length in words variable, the total number of words in the 650 titles under study amounted to 8094 (X=12.45, SD=4.79). Titles having 12 words (60 titles, 9.2%) were most abundant in the dataset. The median obtained was 12 meaning that 50 percent of the titles contained 12 or less words.

Pearson correlation was computed between article citation rate and title length in words in the 650 article titles.

As illustrated in Table 2, an article's citation rate (X=6.37, SD=7.35) was not associated with its title length in words (X=12.45, SD=4.79). The P-value observed was .769 which was greater than .05 and hence the hypothesis, "*The longer the title length the higher the number of citations made to an article.*" was rejected.

Similarly, based on the findings in Table 3, when title length was measured by characters with no spaces, rather than words, again no relationship (r=.661, p>0.05) was observed between title length (X=80.39, SD=28.88) and rate of citations (X=6.37, SD=7.35). The analysis based on characters was added to the study due to lack of consensus on definition of 'word'[22].

Table 1—Descriptive statistics on citation rate in 650 articles

N	Valid	650
	Missing	0
Mean		6.3754
Median		4.0000
Mode		2.00
Std. Deviation		7.35396
Variance		54.081
Range		72.00
Minimum		.00
Maximum		72.00
Sum		4144.00

Table 2—Pearson correlation between article citation rate and title length in words in 650 articles from *Scientometrics* between 2009 and 2011

		Citation	Title length in words
Citation	Pearson Correlation	1	.012
	Sig. (2-tailed)		.769
	N	650	650

Table 3—Pearson correlation between article citation rate and title length in characters with no spaces in 650 articles from *Scientometrics* between 2009 and 2011

		Citation	Title length in character with no spaces
citation	Pearson Correlation	1	-.017
	Sig. (2-tailed)		.661
	N	650	650

Relationship between rate of citations and punctuation marks

For studying the relationship between rate of citations and punctuation marks, the hypothesis framed was: "*There is no relationship between the rate of citations made to an article and the number of punctuation marks in its title.*"

As given in Table 4, from among the 650 titles, 173 (26.6%) cases lacked a punctuation mark.

The remaining 477 (73.4%) article titles contained between one and seven punctuation marks (X=1.24, SD=1.1) while those with one punctuation mark (255 titles, 39.2%) were most abundant in the whole dataset. Titles with two (144 cases, 22.2%) and those with three (58 cases, 8.9%) ranked second and third. The largest number of punctuation marks (7 punctuation marks) was observed only in one title which comprised only .2% of the whole dataset. In all, 811 punctuation marks were observed in the titles under study.

Table 5 shows the correlation between the rate of citations made to an article and the number of punctuation marks in its title.

As indicated in Table 5, there was no correlation (r=.944, p>0.05) between the rate of citations made to an article (X=6.37, SD=7.35) and the number of punctuation marks (X=1.24, SD=1.1) in each title. Hence, the number of punctuation marks each title encompassed could not be used as a criterion to predict the rate of citations to an article. Based on this finding the hypothesis, "*There is no relationship between the rate of citations made to an article and the number of punctuation marks in its title.*" could be accepted.

Punctuation marks in article titles

For studying the punctuation marks in the title, the hypothesis "*All punctuation marks appear similarly in titles of articles.*" was formulated.

To answer this question, all punctuation marks appearing in the 650 titles were listed, along with the total frequency with which each appeared in the whole dataset (Table 6).

Table 4—Descriptive statistics on distribution of titles with varying numbers of punctuation marks

		Frequency	Percent	Valid percent	Cumulative percent
valid	.00	173	26.6	26.6	26.6
	1.00	255	39.2	39.2	65.8
	2.00	144	22.2	22.2	88.0
	3.00	58	8.9	8.9	96.9
	4.00	10	1.5	1.5	98.5
	5.00	7	1.1	1.1	99.5
	6.00	2	.3	.3	99.8
	7.00	1	.2	.2	100.0
	Total	650	100.0	100.0	

Table 5—Pearson correlation between the rate of citations made to an article and the number of punctuation marks in its title

		Citation	No. of punctuation marks
Citation	Pearson Correlation	1	.003
	Sig. (2-tailed)		.944
	N	650	650

Table 6—Frequency of punctuation marks in titles of 477 articles studied

Punctuation marks	Frequency	Percentage
:	279	34.4
-	245	30.2
,	92	11.4
?	71	8.8
's or s'	53	6.5
()	23	2.8
.	14	1.7
&	14	1.7
" "	9	1.1
' '	6	0.7
/	3	0.4
;	2	0.3
Total	811	100%

In all, 12 punctuation marks (: - , ? ' () . & "" '' / ;) appeared in the 477 titles encompassing one or more punctuation marks (Table 6). The total frequency with which these punctuation marks appeared in the titles was 811. From among these 12 punctuation marks, 'colon' (34.4%) and 'hyphen' (30.2%) were most frequent and 'coma' (11.4%) ranked the third. In contrast, the punctuation marks 'semi-colon' (0.3%), 'dash' (0.4%) and 'single quotation marks' (0.7%) were least frequent in the titles. Thus, the hypothesis, "*All punctuation marks appear similarly in titles of articles.*" could not be kept.

Table 7—Characteristics of top-30 article titles ranked based on citations received.

Sl. no.	Citations	Length in characters without spaces	Length in words	Punctuation marks	No. of punctuations	No. of authors
1	72	83	13	?	1	2
2	63	89	12	:	1	2
3	49	63	9	: ,	2	2
4	46	80	16	?	1	2
5	41	108	15	: ’	2	6
6	38	87	18	:	1	5
7	37	166	21	: -	2	4
8	35	60	9	:	1	2
9	35	91	17		0	2
10	32	37	7	?	1	3
11	30	64	9		0	2
12	28	106	18	’	1	1
13	28	88	17	:---	4	4
14	27	28	6		0	1
15	26	27	3		0	4
16	26	102	18		0	1
17	22	55	12		0	2
18	22	72	13	: -	2	1
19	22	106	15	: ,	2	2
20	22	38	6	?	1	1
21	22	66	10		0	3
22	22	77	13	“” ,	2	1
23	22	63	8	-	1	2
24	21	43	6	- ’	2	1
25	21	71	10	: ?	2	1
26	21	29	8	? -	2	1
27	21	45	8	:	1	5
28	20	97	17	-	1	1
29	20	87	15	:	1	2
30	20	64	8	( ) - -	3	1

Titles of highly cited articles

The top 30 highly cited article titles were analyzed with regard to title length, application of punctuation marks and citation rates.

Table 7 reveals a number of important findings as follows:

- From among the top 30 titles, 23 had at least one punctuation mark. The mode obtained was 2 meaning that 9 titles encompassed two punctuation marks. Further, 9 of the top-10 titles included at least one punctuation mark. This contrasts Huggett's[9] finding who asserted that the ten most cited papers in *Research Trends*' case study did not contain any punctuation in their titles.
- Six titles included a question mark. More interestingly, the top article (with 72 citations) also contained a question mark. This contrasts Huggett's[9] finding who claimed that titles encompassing a question mark or having an interrogative form would receive less citations compared to those that lack it. These six titles received 214 citations, in the aggregate, with a mean of 36 citations per article, which is much higher than the mean (X=6) obtained for the whole 650 articles.
- Colon and hyphen were the most frequently used punctuation marks.
- In these titles, seven types of punctuation marks were used namely colon, hyphen, double quotation marks, question mark, coma, apostrophe and parentheses.
- The average length of titles in words was 12.
- The average title-length in characters with no spaces was 73.

Discussion

An overview of the literature revealed that there were varying views on the association between title

length and the rate of citations made to an article. Further, the relationship between the application of punctuation marks in titles of articles and citations made to the articles could not be established with ease. This led the present researchers to undertake the present study.

The study found no correlation between title length and citations to articles. It was also found that the number of punctuation marks could not be used as a variable to predict an article's citation rate. Further, punctuation marks were found to be used with varying frequencies in titles – 'colon', 'hyphen' and 'coma' were the most frequent, and 'semi-colon', 'dash' and 'single quotation marks' were the least frequent punctuation marks. Colon is often used to elaborate more on the title which in turn lengthens the title. Colon is often observed in theses and dissertations. Similarly, hyphen is used within compound terms which could possibly justify the high frequency with which it appeared in the sample data. Finally, coma is used to insert pause within the title. Usually this punctuation mark appears in longer titles encompassing multiple information chunks. It was also found that out of the top 30 titles, 23 included at least one punctuation mark. In the top 10 titles, 9 titles encompassed a punctuation mark. Further, the top cited article and 5 more had a question mark in their titles. Here again colon and hyphen were most frequent.

Based on the results, there are, indeed, varying views on the impact of title length and punctuation marks on the rate of citations. At times, the results reported have been quite contradictory. The results we reported here appear to be in line with those introduced by researchers including Jamali and Nikzad[5] but in contrast to those reported by others like Habibzadeh and Yadollahie[6]. The present authors believe that this variance in views could well be rooted in the research design different authors have used in their studies. More specifically, we conjecture that the procedures adopted by researchers to formulate their research variables have been different. In fact, many authors have only scrutinized the association between dependent and independent variables while ignoring the role of an array of other intervening and controlled variables. Of course, some researchers attempted to control some variables in their studies. De Winter[21], for instance, kept the number of article views constant while collecting the data. Similarly, the present researchers attempted to control some factors including variation in data source, field of study and writing style (all data were collected from the journal *Scientometrics*), etc.

Conclusion

Based on the results obtained and the discussions made, it could be concluded that to assess the impact of title length, and punctuation marks on citation rate, multiple variables may be involved. In other words, the variability in the results reported by researchers on the issue may simply mean that authors have not controlled the variables they had to. Some sample variables could be: gender of the authors, author collaboration, nationality of the authors, paper's topic, its accessibility, the journal in which it has been published, the language of the paper, etc. Although we do not mean to imply that all these variables – and even more – should be considered simultaneously, we do believe that based on the context of each study and its objectives, a number of relevant control and intervening variables should also be identified by the researchers. This is crucial since a high quality research work may not be cited well simply because the paper is not easily accessible to the public. Although dealing with such variables was not our prime objective in this paper, we seek to undertake, in the future, another research while keeping such variables into consideration.

References

1 Scharnhorst A and Garfield E, Tracing scientific influence, *Dynamics of Socio-Economics Systems*, 2 (1) (2010) 1-33.

2 Falagas M E, Kouranos V D, Arencibia-Jorge R and Karageorgopoulos D E, Comparison of SCImago journal rank indicator with journal impact factor, *FASEB J*, 22 (2008) 2623-2628.

3 Cheek J, Garnham B and Quan J, What's in number? Issues in providing evidence of impact and quality of research(ers), *Qual Health Res*, 16 (2006) 423-435.

4 Falagas M E, Zarkali A, Karageogopoulos D E and Bardakas V, The impact of article length on the number of future citations: A bibliometric analysis of general medicine journals, *PLoS ONE*, 8 (2) (2013) 1-8.

5 Jamali H R and Nikzad M, Article title type and its relation with the number of downloads and citations, *Scientometrics*, 88 (2011) 653-661.

6 Habibzadeh F and Yadollahie M, Are shorter article titles more attractive for citations? Cross-sectional study of 22 scientific journals, *Croatian Medical Journal*, 51 (2) (2010) 165-170.

7 Jacques T S and Sabire N J, The impact of article titles on citation hits: An analysis of general and specific medical journals, Available at www.shr.sagepub.com (Accessed on 10 February 2014).

8 Moore A, Do article title attributes influence citations? *Wiley-Blackwell Publishing News*, 2010 (2010) 1-4.

9 Huggett S, Heading for success: Or how not to title your paper, *Research Trends*, 24 (2011) 7-8.

10 van Wesel M, Wyatt S and Haaf, J T, What a difference a colon makes: How superficial factors influence subsequent citations, *Scientometrics*, 98 (3) (2014) 1601-1615.

11 Hartley J, Colonic titles, *Journal of the European Medical Writers Association*, 16 (4) (2007a) 147-149.

12 Hartley J, Planning that title: Practices and preferences for titles with colons in academic articles, *Library and Information Science Research*, 29 (4) (2007b) 553-568.

13 Gustavii B, *How to Write and Illustrate Scientific Papers*, 2nd edn (Cambridge University Press; Cambridge), 2008, p. 21-44.

14 Yitzhaki M, Relation of title length of a journal article to the length of the article, *Scientometrics*, 54 (3) (2002) 435-447.

15 Rostami F, Mohammadpoorasl A and Hajizadeh M, The effect of characteristics of title on citation rates of articles, *Scientometrics*, 98 (3) (2014) 2007-2010.

16 Paiva C E, Lima J P S N and Paiva B S R, Articles with short titles describing the results are cited more often, *Clinics*, 67 (2012) 509-13.

17 Soler V, Writing titles in science: An exploratory study, *English for Specific Purposes*, 26 (1) (2007) 90-102.

18 Kane T S, *Oxford Essential Guide to Writing*, 1st edn (Berkley; New York), 2000, p. 10-22.

19 Manten A A and Greenhalgh J F D, Titles of scientific papers, *Animal Feed Science and Technology*, 2 (1) (1977) 1-6.

20 Ball R, 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) (2009) 667-679.

21 de Winter J C F, The relationship between tweets, citations and articles views for PLOS ONE articles, Available at: link.springer.com/article/10.1007%2Fs11192-014-1445-x (Accessed on 07 January 2015).

22 Fasold R and Connor-Linton J, *An Introduction to Language and Linguistics*, 2nd edn (Cambridge University Press; New York), 2006, p. 30-45.