Scientometrics, Vol. 30, No. 1 (1994) 321-332

RELATION OF TITLE LENGTH OF JOURNAL ARTICLES TO NUMBER OF AUTHORS*

M. YITZHAKI

Department of Information and Library Studies, Bar-Ilan University, Ramat-Gan, 52900 (Israel)

(Received January 18, 1994)

The great importance of titles being highly informative is almost unanimously accepted in literature, assuming that the more informative titles are, the more effectively they serve their functions. The most common measure of title "informativeness" has been the number of *"substantive"* words included in it, and one of the factors which might be associated with it is the number of authors. The present study attempted to test, in a large group of journals from *different* areas, and over six decades, the hypothesis that a paper signed by a larger number of authors will have more substantive words in its title. Large samples of original research papers were drawn from each decade year of fourteen leading journals. For each paper, the number of substantive words in the title was correlated with the number of authors. Findings indicate a difference between the scientific journals on the one hand, and the social sciences and humanities journals on the other. A moderate positive correlation was found in most scientific journals (excluding mathematics) for many periods. In the social sciences journals, and to a greater extent, in the humanities journals, a significant positive correlation was limited to only a few periods, while the rest showed a very low correlation, or even a negative correlation. The different findings for the sciences may be somehow associated with their higher rate of multiple authorship.

Introduction

As is well known, the title is a very important element of any scientific or scholarly paper. Its primary functions are to draw a reader's attention to a paper and to indicate its content in a short glimpse, thus contributing to its initial selection or rejection. Titles constitute the most concise statement of a document's content, surrogating it in bibliographies, databases, indexes and reference lists. Information retrieval systems and services, used by research workers, make considerable use of titles in the process of storing, searching and retrieving information.[2,3,6,7,9] Many of these systems depend heavily on indexing by means of automated, computerized selection of words from the author's title.

*Paper presented at the Fourth International Conference on Bibliometrics, Informetrics and Scientometrics in Berlin (Germany), September 11-15, 1993.

Elsevier, Amsterdam - Oxford - New York - Tokyo
Akadémiai Kiadó, Budapest

It is not surprising that the great importance of highly informative titles is almost unanimously accepted in literature and has been heavily emphasized by many writers, such as *Luhn*,[12] *Feinberg*[7] and others,[3,13,20] as well as by authors of guidance books for scientific and professional authors.[14,17] Their common underlying assumption has been that if titles became more informative, they would perform their functions more effectively. The most common measure of title "informativeness" has been the number of "substantive" words included in it, assuming that the more "substantive" words in a title, the more informative it is.

The central role played by titles in the processes of knowledge organization and information retrieval calls for identifying possible factors which might be associated with the number of "substantive" words in the title. Some of these factors could be certain formal features of a paper, such as the number of its authors.

Literature review

Buxton[3] has already reviewed in detail the broader topic of evaluating the value of titles for retrieval purposes. A related question, concerning the variation in "informativeness" of titles of research papers with time and fields was recently reviewed and studied by *Yitzhaki*.[21] Unlike the question of variations in the "informativity" of titles, the possibility of associating it with the number of authors assigned to a paper has never been thoroughly investigated. *Kuch*,[11] *Balog*,[1] and *White*,[23] were probably the only ones to treat, albeit partially and incompletely, the association between the number of authors of a paper and the number of substantive words in its title.

Kuch's study of five journals published in the 1970's led him to conclude that for at least some journals title length is positively correlated with the number of co-authors. He offered two possible explanations: (1) Multi-author papers are based on more extensive or intensive studies than single-author papers, hence communicate more information, hence tend to have longer titles; (2) Interaction among co-authors acts to increase title length; each may have a different idea of why the paper is significant and each adds a particular term or phrase that suits him.

Checking an agricultural journal, *Balog*[1] found that in series papers the proportion of substantive words in a title was greater for multiple-author papers than for single author papers. Besides the fact that both Kuch's and Balog's studies were limited to the sciences, both also had certain drawbacks which rendered their findings ambiguous and inconclusive. Kuch's study was limited mainly to one or two

years for each journal checked. Consequently, no long-range trend could be traced for any of the five journals studied. Balog's samples for each year seem too small and no attempt was made to calculate Pearson correlation in titles of each sample. Since he considered all multiple-author papers as *one* subsample for which *one* average figure was calculated, his findings seem irrelevent to our question.

White,[23] who was the first to study the social sciences fields, found for six journals correlations between author number and title size, ranging from 0.07 to 0.18. Similarly, the mean sizes of titles with more than one author were larger than those with one author only. Thus, he concluded that his findings supported those of *Kuch*. White's findings and conclusions, however, should be treated cautiously, since his aforementioned correlations were seemingly computed for the whole period as a single sample, although it extended over long periods of time. Thus, the time factor was inevitably introduced as an intervening variable which could be a third "Hidden Factor", affecting both increases in number of authors and in size of titles as conceded by White himself. Former studies have already established the tendencies towards more authors per paper and longer titles, with the passing of time,[21,22,24] probably due to the increasing complexity of research.

In summation, it seems that all three studies failed to produce solid clear-cut findings concerning the hypothesis that, controlling for the time variable, papers with more authors tend to have longer titles with more substantive words.

Purpose of the study

The objective of the present study was to test empirically, in a large group of journals in different fields over a prolonged period, the aforementioned hypothesis. More specifically, the aims of our study were to obtain empirical answers to the following questions:

(1) Is Kuch's[11] finding applicable to other journals too?
(2) Is there a unique pattern in journals belonging to the same discipline?
(3) Do the three broad areas of knowledge (the sciences, social sciences and the humanities) differ with respect to the association in question?
(4) What has been the long-range trend concerning this association?

Methodology

Fourteen leading English-language journals were selected from the following fields: sciences and medicine, mathematics, psychology, economics, sociology, philosophy and law. Large samples of original research papers were drawn from the appropriate volumes of each tenth calendar year, in the last sixty years. All in all, 88 periods (i.e. decade years) were checked for the fourteen journals under study, giving an average of at least six consecutive periods for each journal. Usually, the *entire annual* population of all "regular" research papers published in those volumes was checked. In many cases adjacent volumes were also checked so that the sample usually included about 150-200 articles.

"Regular research papers" were defined as ordinary full-length original research articles, published in the main part of the journal, *excluding* review papers as well as various *short* publications, such as: "Communications", "Brief Communications", "Research Comments", etc.

The two basic approaches regarding the exact operational definition of the concept of "informativeness" of titles, a subjective approach and an objective one were already described by us elsewhere.[21] The most common variation of the objective approach, which has been followed here, is based on a count of *substantive* words in the title, and sometimes its total number of words too.

The count of "substantive" words in the title depends, of course, on compiling a "stoplist" of "meaningless" or "non-significant" words. Following *Tocatlian*,[20] *Feinberg*,[7] *Buxton* and *Meadows*[4] and *Diodato*[5] it was decided, in the present study, to confine the stoplist to articles, prepositions, conjunctions, pronouns and auxiliary verbs. The basic problems encountered in compiling such a list and our own procedure were already described elsewhere.[21] For each decade year, the number of authors of each paper was correlated with the number of substantive words included in its title. Pearson r's were computed and t-test was employed to test their significance. A linear correlation was the most likely assumption to be followed, in view of the hypothesis under study, as well as Kuch's[11] explanations and methodology.

Findings and discussion

Tables 1 to 3 present the Pearson Product-Moment Correlation Coefficients between the two aforementioned parameters of papers published in fourteen journals

in decade (and adjacent) years, throughout the past 60 years, from 1930 until 1990. It is quite evident from these Tables that a consistent long-range trend is difficult to trace. Contrary to Kuch's findings showing a positive correlation, mostly significant, for four of five journals for a specific period, our data indicates that his finding does not hold for most journals checked here, representing a wide spectrum of various fields over a lengthy period. Moreover, our data shows that for a considerable number of decade years checked (about one-third) the correlation was even *negative*.

Table 1
Pearson correlation coefficients between number of authors of a paper and number of substantive words in its title for five scientific journals

Journal		1930	1940	1950	1960	1970	1980	1990
Journal of	years checked		1942–46	1950–51	1960–61	1970	1980	1990
Animal	Pearson r		0.458	0.026	0.062	0.156	0.176	0.271
Science	P <		0.001	***	**	0.05	0.02	0.001
	N		200	200	200	200	200	200
Biochimica	years checked		1947–49	1950–51	1960	1970	1980	1989
et	Pearson r		0.089	0.096	0.119	0.129	0.113	0.105
Biophysica	P <		**	*	*	*	*	*
Acta	N		120	120	120	120	120	120
American	years checked		1940	1950	1960	1970–71	1979–80	1990
Heart	Pearson r		0.247	−0.183	−0.012	0.213	0.242	0.251
Journal	P <		0.01	0.05	***	0.01	0.001	0.01
	N		150	150	160	218	260	160
Journal	years checked	did	did	1949–52	1959–60	1969–70	1979–80	1989
of	Pearson r	not	not	0.125	0.293	0.251	0.108	0.137
Gerontology	P <	exist	exist	*	0.01	0.01	*	*
	N			119	116	115	119	129
American	years checked	1929–31	1939–41	1949–51	1959–61	1969–71	1979–82	1987–91
Journal	Pearson r	0.076	−0.053	−0.176	−0.124	−0.027	0.028	−0.048
of	P <	**	**	0.02	0.1	***	***	**
Mathematics	N	173	219	200	177	176	206	196

* P = 0.2–0.3.
** P = 0.4–0.5.
*** P = 0.6 or more.

Table 1 suggests that, except for the *American Journal of Mathematics*, the other four journals do have a common unique pattern of *positive* correlation for most periods checked. But, apparently, this common pattern has several exceptions of periods with very low r values, statistically insignificant, and even a few negative ones. Generally speaking, however, the figures presented in Table 1 seem to support Kuch's[11] findings.

Table 2

Pearson correlation coefficients between number of authors of a paper and number of substantive words in its title for six social science journals

Journal		1930	1940	1950	1960	1970	1980	1990
Journal of Social Psychology	years checked	1930–34	1939–41	1949–53	1959–61	1969–71	1979–81	1988–91
	Pearson r	0.064	0.172	0.042	0.057	0.001	−0.019	0.044
	P <	**	0.05	**	**	–	***	**
	N	99	135	183	227	261	267	254
Journal of Counseling Psychology	years checked		1939–41	1949–51	1960–61	1970	1980–81	1990–91
	Pearson r		−0.083	0.042	0.260	−0.091	−0.034	0.022
	P <		**	***	0.02	**	***	***
	N		87	95	93	99	120	95
Journal of Abnormal Psychology	years checked	1927–34	1938–41	1950–52	1960	1970	1980	1990
	Pearson r	0.115	−0.066	0.141	0.082	0.00	0.093	0.011
	P <	*	**	*	**	–	**	***
	N	120	120	120	120	120	120	120
American Economic Review	years checked		1940–42	1950–54	1959–64	1969–73	1979–84	1986–90
	Pearson r		−0.065	0.023	0.051	0.154	0.004	0.092
	P <		***	*	*	0.1	***	*
	N		100	145	155	158	273	239
Journal of Marketing	years checked		1938–42	1948–52	1958–61	1968–71	1979–81	1988–92
	Pearson r		−0.160	0.352	−0.294	−0.314	0.178	0.068
	P <		0.1	0.001	0.001	0.001	0.05	**
	N		135	125	143	125	130	112
American Sociological Journal	years checked		1938–41	1948–51	1958–61	1968–71	1978–82	1988–91
	Pearson r		0.079	0.091	0.237	0.014	0.00	0.169
	P <		*	*	0.001	***	–	0.01
	N		211	228	233	189	220	251

* P = 0.2–0.3.
** P = 0.4–0.5.
*** P = 0.6 or more.

As for the social sciences, Table 2 seems to indicate the lack of any common unique pattern, neither for any certain journal along the sixty years checked, nor for any group of journals in a given period. One can readily observe for the same journal a wide range of r values, over zero and below, with statistical significance and without, during the period studied.

Table 3
Pearson correlation coefficients between number of authors of a paper and number of substantive words in its title for three humanities and law journals

Journal		1930	1940	1950	1960	1970	1980	1990
Philosophy	years checked	1932–35	1938–43	1948–53	1958–63	1968–73	1978–83	1988–91
	Pearson r	0.155	−0.135	−0.080	−0.149	−0.063	−0.066	−0.022
	P <	*	*	**	*	**	**	–
	N	90	110	107	114	114	138	92
Philosophy and Phenomenological Research	years checked		1940–45	1948–52	1958–63	1968–73	1978–83	1988–91
	Pearson r	–	−0.057	−0.117	0	0.120	−0.131	−0.171
	P <		***	*	–	0.1	0.1	0.1
	N		138	134	148	207	184	95
Yale Law Journal	years checked	1929–34	1939–44	1949–54	1959–64	1969–74	1979–84	1989–91
	Pearson r	0.296	−0.050	−0.011	0.136	0.044	−0.035	0.054
	P <	0.01	***	–	*	***	***	***
	N	161	135	120	121	131	148	89

* P = 0.2–0.3.
** P = 0.4–0.5.
*** P = 0.6 or more.

The most remarkable findings in the humanities Table (Table 3) are the *negative* values of r found for *most* periods (13 out of 20!), mainly in the philosophical journals, but also for the law journals. In only four periods (20%), r values were greater than 0.12, but only *one* was found to be highly significant (at $P < 0.01$). With all due caution stemming from the fact that the humanities subsample comprised only three journals, one may conclude that Kuch"s[11] finding does not hold for these journals and perhaps not for other humanities journals either.

Ignoring the time factor, in order to convey an overall view, Tables 4 and 5 summarize Tables 1 to 3 according to different levels of r found in the various periods to enable comparison between the three broad areas of human knowledge.

Table 4
Frequency distribution of Pearson r values according to groups of journals (in %)

Pearson correlation coefficient	Sciences with Math.	Sciences without Math.	Social sciences	Humanities
Positive	76.7	91.3	71.0	30.0
0	–	–	5.3	5.0
Negative	23.3	8.7	23.7	65.0
Total	100%	100%	100%	100%
N (number of periods checked)	30	23	38	20

Due to the grouping, the considerable differences between the sciences and social sciences journals on the one hand, and the humanities journals on the other hand, are obvious. While the vast majority (71% to 77%, or even reaching 91% if mathematics is excluded- see below) of r values found for the former ones were positive, the opposite is true for the humanities journals, in which 65% of the correlations were *negative*.

In view of the essential differences between mathematics and other scientific fields, it seemed appropriate to also present the proportions among the sciences when the mathematics journal is excluded. The figures thus obtained indicate that the sciences do differ significantly from the social sciences, having a positive r in *91%* of the periods, as compared to only 71% for the social sciences. Similarly, the proportion of periods with a negative r is about three times in the latter ones (23.7% *vs* 8.7%). Needless to say, the differences between the sciences and the humanities are now much higher: over 91% of positive correlations *vs* only 30% for the latter ones.

Table 5
Frequency distribution of Pearson r values according to groups of journals (in %)

Pearson correlation coefficient	Sciences with Math.	Sciences without Math.	Social sciences	Humanities
$r \geq 0.1$	56.7	73.9	23.7	20.0
$0.1 > r > -0.1$	33.3	21.7	68.4	55.0
$r \leq -0.1$	10.0	4.3	7.9	25.0
Total	100%	100%	100%	100%
N (number of periods checked)	30	23	38	20

Table 5 presents the frequencies of r values according to a different grouping of categories, in which all r values greater than -0.1 and smaller than 0.1 were added to the zero category. The rationale for enlarging the zero category was that such low r values ($-0.1 < r < 0.1$) do not indicate any real correlation, and as already shown above (in Tables 1 to 3), these values are *statistically insignificant*, their P values being as high as 0.5 to 0.9. The $r = 0.1$ cut-off point also enables comparison with Kuch's findings for four journals, whose lowest r found was 0.102.

This form of presentation in Table 5 re-emphasizes the considerable differences between the three broad areas of knowledge (indicated in Table 4). It is quite obvious that, taken together, the sciences journals checked differ considerably from the social sciences group, not to mention the humanities. Relatively speaking, the discrepancies revealed between the sciences and the social sciences are much greater than those found between the latter ones and the humanities, especially if mathematics is excluded. Needless to say, this exclusion greatly increases the discrepancies found between the sciences on the one hand and the social sciences and humanities on the other. The proportion of sciences journals in which $r \geq 0.1$ is more than double (or *triple* if mathematics is excluded) that proportion among the social sciences group (56.7% or 73.9% *vs* only 23.7%, respectively). Similarly, the proportion of journals with r values equal or close to zero, indicating no correlation, is twice (or three-times) higher for the social sciences group as compared to the sciences (68.4% *vs* 33.3% or 21.7%). As a group, the humanities journals reveal the lowest r values: in 25% of their periods $r \leq -0.1$, while for the other two groups of journals the equivalent proportion was considerably lower.

We may conclude that as far as the sciences journals studied (excluding mathematics) are concerned, in most cases a positive correlation, sometimes significant, was found between the number of authors signed on a paper and the number of substantive words included in its title. This conclusion supports Kuch's[11] findings for three (out of four) scientific journals he checked during one period only. Kuch's[11] two possible explanations for this finding are mentioned in the beginning of this paper. However, both explanations fail to adequately explain why these causes affect mainly the sciences, but not the social sciences and the humanities, as revealed above.

Apparently, this positive correlation is somehow related to the prevalence of multiple authorship in a given field. Thus the factors Kuch mentions occur mainly in subject fields with a relatively high degree of multiple authorship. As already demonstrated by *Yitzhaki* and *Ben-Tamar* elsewhere[22] in a detailed comparative table, most scientific fields have the highest rate of authors/paper, followed by the social sciences and then the humanities, whose rate has been as low as 1.05. Consequently, the aforementioned correlation is usually found in the sciences, rather than the social sciences and the humanities. The vast majority of papers published in the humanities (and to a lesser extent in the social sciences) are still by a *single* author, drastically lowering the chances of this correlation to manifest. When the average number of authors per paper is relatively low, and the range is only between 1 and 2, the chances of obtaining a high or even moderate correlation are fairly low.

The case of mathematics may serve to validate this explanation. The field of mathematics is one of the *last* in the descending list of subject fields in the aforementioned comparative table,[22] since its authors/paper rate was only around 1.3 in 1980. This same field is a prominent exception among the sciences (in Table 1 above), with its mostly negative or very low correlation coefficients.

The correlation found above for scientific journals as well as a former finding[21] of a relatively higher number of substantive words in scientific titles may both be related to Storer's[19] and Price's[18] classic distinction between "hard", "soft" and non-sciences. Thus, one may speculate that the "harder" the field, the higher the correlation between number of authors of a paper and the number of substantive words in its title. However, the limited scope of our data does not enable us to thoroughly study this interesting hypothesis.

Conclusions

1. *Sciences*: Titles of journal articles in the four scientific fields (excluding mathematics) studied, show for most periods, a moderate degree of positive correlation between number of authors of a paper and number of substantive words in its title.

2. *Social Sciences*:A significant positive correlation was found in only seven out of 38 periods checked for the six journals. In the remaining periods the correlation was very low and statistically insignificant.

3. *Humanities*: A significant correlation was found in only *one* period out of twenty, for the three journals checked. In the rest the correlation was mostly *negative*, and statisically insignificant.

4. The fact that most *positive* correlations were in the sciences, rather than the social sciences and the humanities, indicates that it is somehow associated with the fact that scientific papers have a much higher rate of multiple authorship.

5. Caution should be, however, exercised in view of the relatively small number of journals included in each group in our samples. Further research is needed on many more journals, from all three broad areas of human knowledge, over a prolonged period, in order to confirm the aforementioned findings and conclusions.

References

1. C. BALOG, The information content of titles of papers in an agricultural journal, *Journal of Research Communication Studies*, 2 (1979-1981) 263-270.
2. R. T. BOTTLE, C. I. PREIBISH, The proposed KWIC index for psychology: an experimental test of its effectiveness, *Journal of the American Society for Information Science*, 21 (1970) 427-428.
3. A. B. BUXTON, *The Bibliographical Infomation Content of Research Papers*, Unpublished Ph.D. Dissertation, University of Leicester, 1979.
4. A. B. BUXTON, A. J. MEADOWS, The variation in the information content of titles of research papers with time and discipline, *Journal of Documentation*,33 (1977) 46-52.
5. V. DIODATO, The occurence of title words in parts of research papers: variation among disciplines, *Journal of Documentation*, 38 (1982) 192-206.
6. V. DIODATO, K. PEARSON, Source indexing in science journals and indexing services: a survey of current practices, *Science & Technology Libraries*, 6 (1985) 103-118.
7. H. FEINBERG, *Title Derivative Indexing Techniques : A Comparative Study*, Scarecrow Press, Metuchen, N.J.,1973.
8. J. S. GHOSH, Content representation in document titles: a case study with Prostaglandin literature, *Aslib Proceedings* 26 (1974) 83-86.
9. P. R. HODGES, Keyword in title indexes: effectiveness of retrieval in computer searches, *Special Libraries*, 74 (1983) 56-60.
10. M. KOLL, T. NOREAULT, Relation of title length to number of authors, *Journal of the American Society for Information Science*, 30 (1979) 175-176.

11. T. D. C. KUCH, Relation of title length to number of authors in journal articles, *Journal of the American Society for Information Science*, 29 (1978) 200-202.
12. H. P. LUHN, Keyword-in-Context Index for technical literature (KWIC Index), *American Documentation*, 11 (1960) 288-295.
13. A. A. MANTEN, J. F. D. GREENHALGH, Titles of scientific papers, *Animal Feed Science and Technology*, 2 (1977) 1-6.
14. J. H. MITCHELL, *Writing for Professional and Technical Journals*, Wiley, N.Y., 1968.
15. C. MONTGOMERY, D. R. SWANSON, Machine-like indexing by people, *American Documentation*, 13 (1962) 359-366.
16. J. O'CONNOR, Correlation of indexing heading and title words in three medical indexing systems, *American Documentation*, 15 (1964) 96- 104.
17. M. O'CONNOR, F. P. WOODFORD, *Writing Scientific Papers in English*, Elsevier, Amsterdam, 1975.
18. D. J. PRICE, Citation measures of hard science, soft science, technology and nonscience, In: *Communication Among Scientists and Engineers*, C. E. NELSON, D. K. POLLOCK (Eds),Heath, Lexington, Mass.,1970,pp.3-22.
19. N. W. STORER, The hard sciences and the soft: some sociological observations, *Medical Library Association Bulletin*, 55 (1967) 75-84.
20. J. J. TOCATLIAN, Are titles of chemical papers becoming more informative?, *Journal of the American Society for Information Science*, 21 (1970) 345-350.
21. M. YITZHAKI, The variation in informativity of titles of research papers with time and field, *Cognitive Paradigms in Knowledge Organisation*, Second International ISKO Conference, Madras 26-28 August 1992, Sarada Ranganathan Endowment for Library Science, Bangalore,1992, pp.401-418.
22. M. YITZHAKI, D. BEN-TAMAR, Multiple authorship in biochemistry and other fields; a case study of the Journal of Biological Chemistry throughout 1905-1988, *Informetrics 89/90*: Second International Conference on Bibliometrics, Scientometrics and Informetrics, London (Ontario), July 5-7,1989, Elsevier, Amsterdam, 1990, pp.373-389.
23. A. WHITE, A further exploration of title size and author number, *Journal of the American Society for Information Science*, 42 (1991) 384-385.
24. A. WHITE, N. R. HERNANDEZ, Increasing field complexity revealed through article title analyses, *Journal of the American Society for Information Science*, 42 (1991) 731-734.