

Research Article Titles and Disciplinary Conventions: A Corpus Study of Eight Disciplines

Robin L. Nagano
University of Miskolc, Hungary

Abstract

Research articles are clearly influenced by the discipline of the research being reported. Just as disciplinary conventions place constraints on, for example, the moves and language use of abstracts and introductions, they also provide a set of options for title design. This study attempts to identify the title conventions of eight disciplines by focusing on various features that play a part in title design: the use of multiple-unit titles (those with subtitles); the use of noun phrases to form the title; and 'a' or 'the' in initial position. The length of titles is investigated, as is the proportion of substantive words. Data is based on a 3,200-title corpus of titles from research articles published in prestigious journals in four disciplines in the hard sciences (botany, fluid engineering, geology, and medicine) and four in the soft sciences (economics, education, history, and sociology). The data is presented in a visual form that compares title features by discipline, to demonstrate title conventions and to help novice writers understand the features and options available.

Introduction

When introducing learners to academic language use, one vital concept is disciplinary conventions. It is essential that writers realize that, while the writing toolbox is full of options, some are much more likely to be used in certain disciplines than in others. It is thus important that we recognize the most typical patterns and expressions used within the target discipline, that is, the area of study in which the writer works and writes (or intends to, for novice academic writers). We know that just as individual writers have preferences in how they express themselves, areas of study have preferences in how they focus on and present information. Knowledge of the preferences, or conventions, of a field makes it possible to process information more efficiently and to present it in a way that will be familiar and acceptable to its intended audience. This also plays an important role in socialization or representation of a novice author as a member of the discipline (Becher and Trowler 2001).

Hyland (2000) discusses disciplines in terms hard versus soft, as distinguished by structures of knowledge and intellectual inquiry, which tend to be revealed in rhetorical conventions of the discipline. Participants in the hard disciplines see themselves as methodologically constructing their knowledge, building upon earlier work to progress, each step following from a previous one. Since the theoretical framework is well-established, a great deal of knowledge is already shared with readers on the background of a study, its procedure, and the technical lexis involved, and thus the author can focus on describing the results of the study. Soft disciplines, on the other hand, adopt a different approach. Rather than the linear approach of the hard disciplines, writers tend to retrace ideas, adding in new perspectives or knowledge (often from other fields of study). Issues are diverse, and various lines of inquiry are often pursued or integrated. Readers may come from a wide variety of academic backgrounds, and

thus a great deal of care is given to establishing the context of the topic, to citing other authors in order to support the author's claims, and to defining terms (Hyland 2000).

Hyland naturally acknowledges that this distinction is a broad one, unable to reflect the full complexity of differences between disciplines. However, he suggests seeing disciplines as situated somewhere on a scale or continuum between the two extremes of 'soft' and 'hard'. In this way, one discipline in the social sciences, for instance, may be 'harder' than another. Based on analysis of academic texts, Hyland places disciplines on the scale in relation to each other according to discourse features. Biology, for instance, while on the hard side, moves towards soft in terms of its citation practices – previous work is cited more extensively than in the other natural sciences studied – and its likelihood to include an introductory move in research article abstracts (Hyland 2000).

Disciplines thus develop certain conventions that reinforce their approaches to knowledge. For instance, the sections of a research article in the hard sciences tend to be quite predictable, while the headings given to a journal article in the soft sciences may depend largely on the topic(s) discussed. Both present information in a certain order: they introduce a topic, expand upon it, discuss it, and reach some conclusions. Each follows the conventions of the discipline, using a familiar way of sharing knowledge with a group of fellow researchers possessing similar background information and priorities.

At the same time, however, a tremendous amount of variation in usage is possible, and this should not be ignored. Novice writers should not be given the impression that what is commonly used is the only possible way of using the language. Knowledge of the disciplinary conventions is essential, but creativity and breadth of expression are also useful for a proficient writer in the disciplines. This is perhaps especially true of those writing in inter- or multidisciplinary fields. In addition, sometimes a sub-discipline forms its own conventions or a particular journal or funding organization may have preferences that differ from the norm. In such cases, familiarity with the standard conventions is not enough – writers must be able to identify how these particular usages differ from the norm if they wish to conform. Authors need to be able to identify text features and determine which tools should be pulled out of the box for a particular job. They need to be able to teach themselves, and for this, they need to analyze texts to find patterns that they can use.

When introducing novice writers to disciplinary language use, one topic worth considering is research article titles. After all, even novices read quite a lot of titles, and probably recognize the importance of a title in determining whether the reader reads on or instead turns to the next article. Compared to full research articles, which are rather intimidating for people new to academic texts – especially those reading in a second language -- titles are 'bite sized' and therefore more accessible. Still, their significance is considerable: titles represent a full text and yet must stand alone. They should not only inform readers of the topic of the article, but ideally catch their attention as well, to encourage them to read further. In fact, the editor of *BioEssays* points out in an editorial (Moore 2010) that titles play a role not only in attracting potential readers after publication, but in getting a paper to the point of publication, and in helping people find it later online. Peer reviewers are often asked to judge whether to take on a review based only on a title and abstract – and no reviewers means no chance of publication. Additionally, search engines based on keyword searches may fail to find a relevant article if the title does not contain those particular keywords (Moore 2010). The task of designing an appropriate and effective title is thus more important than it may appear at first glance.

How a title can inform and attract – the language chosen to carry out this task – is partly a matter of personal choice, and largely a matter of custom, habit, and copying others. What people in one field find appealing may cause a totally different reaction in readers from another area of study. The use of metaphors and allusion in research articles is strongly discouraged by several editors in the hard sciences (for biomedical journals, e.g., Al-Awqati 2006, Christensen et al. 2009), and yet its use is fairly typical in soft sciences (Haggan 2004).

What works in one field will not always be effective or accepted in another. So how does a writer learn what works in a particular field? It is certainly possible to gain a feeling for typical titles by reading hundreds of them. However, it may be more efficient if certain text features for the sub-genre of research title articles can be identified.

Titles of various genres, including research articles, have been described and analyzed by numerous authors, from a variety of perspectives; pedagogical, linguistic, diachronic, and scientometric approaches all select different features to investigate. One difficulty with applying their findings to a cross-disciplinary context is that many of the studies draw conclusions based on a very small sample, or lump several disciplines together as, for instance, 'science' (as in Haggan 2004). In addition, most studies have focused upon only one or two features, or if the scope is wider, the discipline investigated is limited; Gesuato (2008), for instance, performed a thorough study of titles of journal articles, books, conference proceedings and dissertations, but only within applied linguistics. Scientometric studies tend to survey massive numbers of titles from a huge assortment of journals, but are restricted to a few features such as the the number of words and use of colons (Lewison and Hartley 2005) or the use of question marks (Ball 2009). Diachronic studies have usually looked at the change in the number of words or substantive word rate (e.g., Yitzhaki 1997). This investigation aims to make a more wide-ranging study based on a more balanced corpus than has been typically used in studies of titles.

In this paper I attempt to present some of the results of an eight-discipline corpus analysis, focusing on features that can be quantitatively expressed, and to display the results graphically to show broad disciplinary preferences for different features of title use, as well as to point out examples of variation within disciplines.

Corpus

The corpus is made up of titles of research articles published in English in journals from eight disciplines. Each discipline is represented by four relatively prestigious journals (based on impact factors, ranking lists, etc.) that can be considered general journals, that is, they solicit articles from many branches of the discipline. Since four general journals could not be found for the extremely broad area of engineering, I chose fluid engineering, as it is one area of engineering with wide applications. The source journals are listed in the Appendix.

As these are all international journals, the corpus therefore contains a mixture of titles written by speakers of English as a first and as a second language. Disciplines were chosen to represent one field each from the categories of Biglan (1973), but in this paper I deal only with the soft/hard dimension; the applied/pure and human/non-human dimensions are not discussed.

One hundred titles were collected from each journal by starting from the last issue of 2007 and proceeding back in time, using on-line tables of content, until 100 articles were reached, disregarding texts labeled as editorials, short reports, letters, book reviews, or review articles. This means that each discipline is represented by 400 titles and four journals; other cross-disciplinary studies tend to be more limited in number. For instance, Haggan (2004) studied 307 titles taken from 40 different journals to represent 'science'; Soler's corpus of three social science disciplines and three biological sciences totaled 480 research paper titles (Soler 2007).

The eight disciplines selected for the corpus were: botany, fluid engineering, geology, medicine, economics, education, history, and sociology. The corpus contains 3,200 titles and comes to approximately 40,350 running words.

Methods

Titles were put into a spreadsheet, coded, and multi-unit titles were stored separately as units (normally two but occasionally three). Thus a title consisting of a title and subtitle is handled in the corpus as two units, which are analyzed separately. This means that full data is available for both/all parts of the title, rather than the title being categorized as a compound title, a category of its own, and not analyzed any further, as was done in the studies of, e.g., Haggan (2004) and Soler (2007). Units were considered to be divided by punctuation marks: most typically a colon, although dashes and question marks were also present.

The title features were investigated in a quite low-tech manner. For the title length data, a count of words in each title or title unit was carried out using Microsoft Excel. Syntactic structure was based on classification by the researcher, most often focusing on features at the beginning of the title unit. The investigation into the substantive word rate was done with the help of the vocabulary profiler Web VP Classic v. 3, on the site Compleat Lexical Tutor (www.lextutor.ca), using its standard stop list to distinguish functional from content words. Units with initial articles were identified by hand, double checked by the find function of Microsoft Word. No statistical tests of significance have been employed at this point in the study. Data has been compiled by source journal, although in this paper results are usually given by discipline, i.e., the total data for the four source journals in each discipline.

Results

The corpus was investigated for a number of features. The following features are the focus of this article:

- title length (number of words)
- title style (single-unit or multiple-unit)
- title unit structure (syntactic structure)
- substantive word rate
- initial article use.

These were chosen as the features that are both quantifiable and easily identified even by novice writers with no knowledge of linguistics. Rather than mainly presenting results in detailed tables or graphs, scales and box and whiskers plots are also selected in order to present the findings in a more accessible way to such authors. The aim is to focus on how disciplines compare in these features, based (in most cases) on average values.

Title length – number of words

One obvious feature of titles – and a fairly common question from novices – is the number of words in the title. Journals rarely give an explicit limit in their instructions to authors, instead favouring phrases such as ‘clear and concise’ or ‘brief and specific’, if anything is said at all. The corpus includes a wide variety of title length. Results are presented in Fig. 1 in the form of a box-and-whiskers plot, in which the box itself shows the limits of the first and third quartile, the line inside gives the median value, and the lines or whiskers show the minimum and maximum number of words found in the corpus.

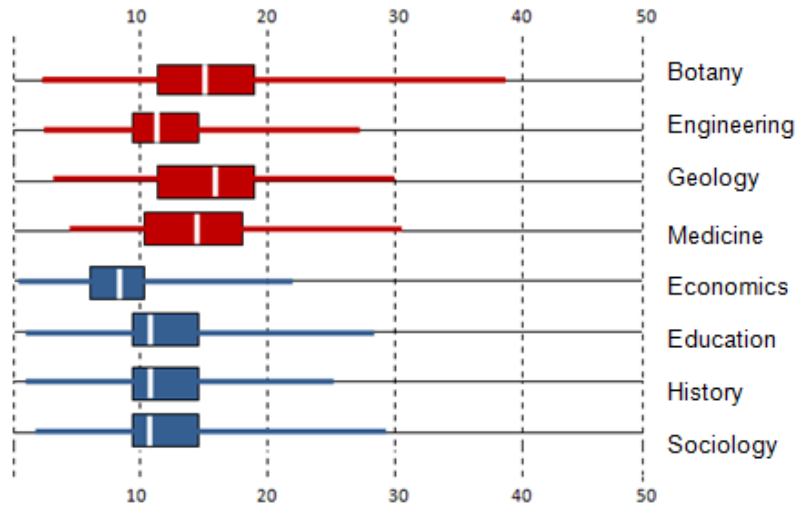


Figure 1. Box and whiskers plot for title length (in words) by discipline

The presentation of data in this form has the advantage of giving a great deal of information – more than, say, a bar chart giving median values – while making comparison among the disciplines relatively simple. After a glance we can see that three of the soft sciences are very similar in their values, while titles in economics are shorter, and an economics title holds the distinction of containing the shortest title (*Dams*, Econ2-75). In general, the hard sciences (indicated by red) tend to be longer than the soft sciences (in blue), although the box for the engineering titles is quite similar to those for education, history, and sociology. The longest title in the corpus, at 38 words, is in botany: *Conditional oxidative stress responses in the Arabidopsis photorespiratory mutant cat2 demonstrate that redox state is a key modulator of daylength-dependent gene expression, and define photoperiod as a crucial factor in the regulation of H2O2-induced cell death* (Bot1-31).

The average results can be compared with those of other studies. For instance, a large-scale diachronic survey of titles in hard science disciplines (Lewison and Hartley 2005) was performed using the Science Citation Index, meaning that titles came from a huge variety and number of journals. This study found that title length tended to increase between 1981 and 2001. Of the disciplines looked at, mathematics had (in 2001) the shortest titles, with around nine words, while biology had the longest, at just under 15 words. Other fields averaged 11–14 words. In another, smaller-scale study, Soler (2007) counted average title length across discipline for articles published 1996–2002, finding that titles in the three hard sciences she investigated (biology, medicine and biochemistry) contained more words – an average of 14.15–15.48 words, quite consistent with my findings -- than titles in the soft sciences (ranging in her data from an average of 7.98 words for linguistics to 12.63 for psychology).

Title style – single or multiple units

Another feature of titles is how many units they are made up of. An example of a single-unit title is *The wisdom of class-size reduction* (Edu1-09). Multiple-unit titles most commonly come in two units, e.g. *Preparing high-quality teachers: Views from the classroom* (Edu1-06), but three-unit titles also occurred, such as *Early literacy instruction and learning in kindergarten: evidence from the early childhood longitudinal study—kindergarten class of 1998–1999* (Edu1-90). Figure 2 gives the results of the corpus study in terms of the average percentage of titles consisting of a single unit, so for instance 88% of Fluid Engineering titles were made up of one unit, and 12% of more than one unit. Note that hard sciences are indicated above the line and are shown with empty circles, while the soft science disciplines are named beneath and denoted with darker circles.

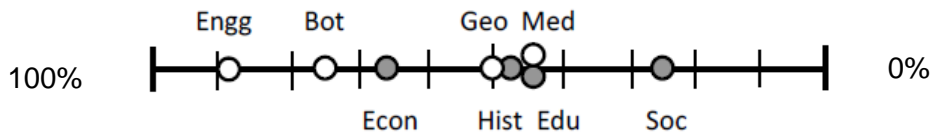


Figure 2. Proportion of single-unit titles by discipline

The data in Fig. 2 is somewhat related to the previous data on title length, since multiple-unit titles tend to be longer than single-unit titles (Lewison and Hartley 2005). There is quite a spread among disciplines, with multiple-unit titles being more common than single-unit in three soft sciences plus medicine. It is interesting to observe that, once again, economics is separated from the other soft sciences. The advantage of a two-unit title allows the author to focus on two types of information in the title. In soft sciences the first unit is often used for some 'catchy' phrase to attract attention (Haggan 2004; Hartley 2007), while in the hard sciences it is often used to add context in terms of further information on the study population, method used, etc. (Haggan 2004, Soler 2007).

However, it is necessary to recognize that factors such as sub-disciplinary preferences and journal policies can affect title design. Figure 3 presents average title length for eight medical journals: four general medical journals and four journals specialized in nephrology. Looking at the four general journals, labelled 1-4, we can find two extremes in terms of the percentage of single versus multiple-unit titles.

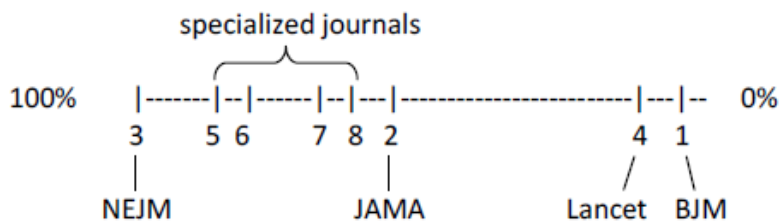


Figure 3. Proportion of single unit titles in eight medical journals: general journals (numbers 1-4) and specialist journals (numbers 5-8)

While all of the 100 titles in the corpus from the *New England Journal of Medicine* are single-unit titles, 92% of the article titles published in *BMJ* have two units, and 6% have three. *The Lancet* also shows a strong preference for multiple-unit titles (94%). In *The Lancet* and *BMJ*, the second unit almost invariably gives information on the type of study, while this information is rather included in the abstract for the other two general biomedical journals. Only the *Journal of the American Medical Association*, with 48%, shows a pattern of occurrence that is even near the averaged figure of 56% for the four general medical journals, which points to one of the dangers of using averaged data: a great deal of variation can be concealed.

The four specialist journals (labeled 5-8 in Fig. 3) show much lower variation (Nagano 2008), indicating that the sub-discipline of nephrology has a more uniform idea of how a title should be formed. A difference between general and specialist journals was also found in the syntactic structure of the title: while titles in all of the general medical journals were almost exclusively formed as extended noun phrases (see Fig. 4), an average of almost 24% of the nephrology journal title units used titles in the form of statements (percentages ranged from 11% to 49%) (Nagano 2008). This great variation is an example of why title writers would be well advised to investigate targeted journals for their individual title style preferences.

Structure of title

The ‘standard’ title, if such a thing can be said to exist, consists of one or more noun phrases. This holds true for all disciplines in this corpus study, to greater or lesser degrees.

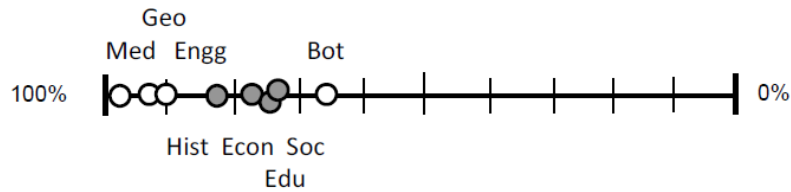


Figure 4. Proportion of noun-phrase titles by discipline

As can be seen in Fig. 4, there are clusters for the soft versus the hard disciplines – with one exception – and although noun phrases (also known as indicative titles) dominate for all of the disciplines, the soft sciences and botany are more likely to use some other form. Such forms include a number of options, as shown below, with examples from the corpus (note: where ‘a’ is used in the title code, this indicates the first unit of a multi-unit title).

- Statement (also known as full sentence or declarative title): *Nitric oxide accumulation is required for molecular and physiological responses to iron deficiency in tomato roots* (Bot1-27)
- Preposition at the beginning: *On the Relations Between Right-Wing Parties and Anti-foreigner Sentiment* (Soc1-17)
- Questions (full or fragment): *Does Money Whiten?* (Soc1-06a)
- -ing: *Blaming the Victims*: (Soc1-67a)
- Other: *How Puerto Rico Became White*: (Soc1-05a)

An example of distribution of title structures among journals within one discipline is given in Table 1. While some variation is found, it is not especially dramatic, indicating that the choice of title structure accepted by the journals for publication is fairly consistent (at least for these particular journals).

Table 1. Distribution of title unit structures for the economics subcorpus

	Econ1	Econ2	Econ3	Econ4	Average
Noun phrase (indicative)	76	75	79	82	78
Question form	10	12	8	9	9
-ing	10	5	10	5	7
Preposition	<1	3	6	3	3
Statement (declarative)	2	0	0	<1	<1
Other	2	5	2	<1	2

Substantive word rate

One feature of titles that has been investigated by several researchers is the substantive word rate, or the proportion of content words in the title. This rate is often considered as an indication of how informative the title is. Many of these studies have been diachronic, and most show that, across fields, the substantive word rate is rising (e.g. Diener 1984; Yitzhaki 1996). Yitzhaki’s latest figures, for 1990, showed 70.2% for natural science titles, 68.9% for social sciences and 66.1% for humanities. While the last two categories roughly match the data from my title corpus (see Fig. 5), my data for the hard sciences show slightly higher proportions; this may be due to the tendency towards a rising rate, as my data is more recent. The percentage in itself is probably not so informative for writers in the discipline, but it may be instructive to consider what differences in title structure and word usage could account for the gap between hard and soft disciplines.

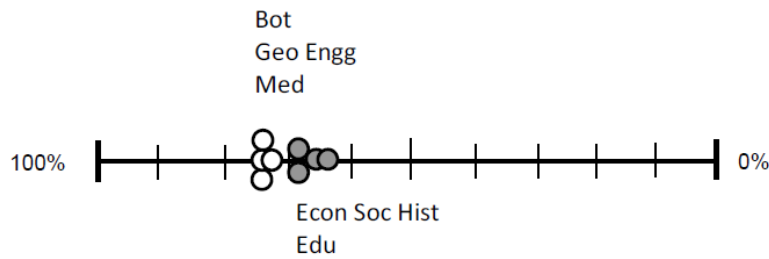


Figure 5. Proportion of substantive words by discipline

One factor may be that the ‘catchy’ titles favoured in some of the softer disciplines tend to have more function words. These can be quotations, metaphors, allusions and questions, which may use a larger proportion of function words. For instance, the first unit of the following title has a low rate of substantive words: *The Ties That Bind and Those That Don't: Toward Reconciling Group Threat and Contact Theories of Prejudice* (Soc2-88). In history, titles often consist of a number of relatively short modifying prepositional phrases, such as *Commercial Conflict and Regulation in the Discourse of Trade in Seventeenth-Century England* (Hist4-69), thus raising the number of prepositions and articles. Also here we see ‘catchy’ title units, as in *‘Waking up to the Fact that there are any Unemployed’: Women, Unemployment and the Domestic Solution in Britain, 1918–1939* (Hist3-71). In contrast, hard sciences tend to employ long strings of nouns and thus need fewer prepositions. A rather extreme example of this is *Self-consistent high-Reynolds-number asymptotics for zero-pressure-gradient turbulent boundary layers* (Engg4-59).

Initial article usage

Another factor contributing to the difference in substantive word rates between hard and soft sciences is article usage at the beginning of the title (Nagano 2013). As Fig. 6 shows, hard and soft sciences tend to use few indefinite articles, and there is no clear distinction between the two groups. The indirect article is almost always used to introduce the type of study: *A Randomized, Placebo-Controlled Trial of Natalizumab for Relapsing Multiple Sclerosis* (Med3-28) or *A life-cycle model of outmigration and economic assimilation of immigrants in Germany* (Econ4-25). Medical title units show slightly more frequency due to the practice of inserting the indirect article before the type of study as a separate unit, as in *Cancer diagnosis, treatment, and survival in Indigenous and non-Indigenous Australians: a matched cohort study* (Med4-65) (though editorial policy seems to play a role here, since BMJ article titles typically omit the article in this case). The discipline of history has a low proportion of indefinite articles primarily because it deals with specific events, people, places, and so on, as in *The Trial of Oscar Slater (1909)* and *Anti-Jewish Prejudices in Edwardian Glasgow* (Hist3-83).

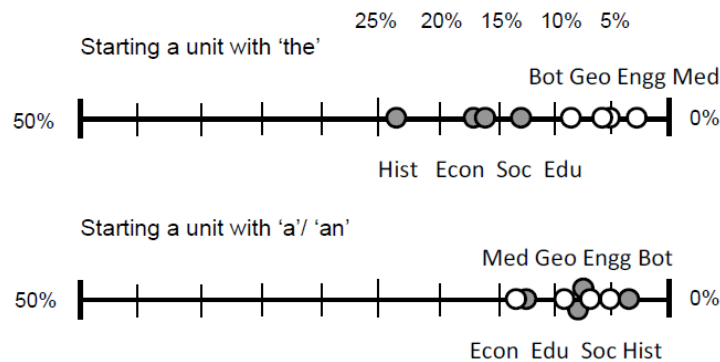


Figure 6. Proportion of title units beginning with the definite article (top) or indefinite article (bottom), by discipline

While indirect article usage shows a mixed picture, the two groups are quite distinct in their patterns of direct article usage. History has the highest proportion for the reason mentioned above: the nature of the discipline means that specific situations are being referred to, such as *The Epidemic of Pneumonic Plague in Manchuria 1910–1911* (Hist1-50). In general, the soft sciences tend to use the definite article as it would be used in a full sentence: the (noun) of..., as in *The risk properties of human capital and the design of government policies* (Econ4-31) or *The effect of grammar teaching on writing development* (Edu4-62). The hard science journals, on the other hand, appear to omit the article in titles: *Effect of Blinded Peer Review on Abstract Acceptance* (Med2-49), *Influence of swirl on the stability of a rod in annular leakage flow* (Engg3-62), *Role of Crustal Contamination in Formation of the Jinchuan Intrusion and Its World-Class Ni-Cu-(PGE) Sulfide Deposit, Northwest China* (Geo3-64). In the hard sciences, an unexpected pattern emerges: the results are reversed. That is, those journals that are more likely to use the definite article are less likely to use the indefinite, and vice versa.

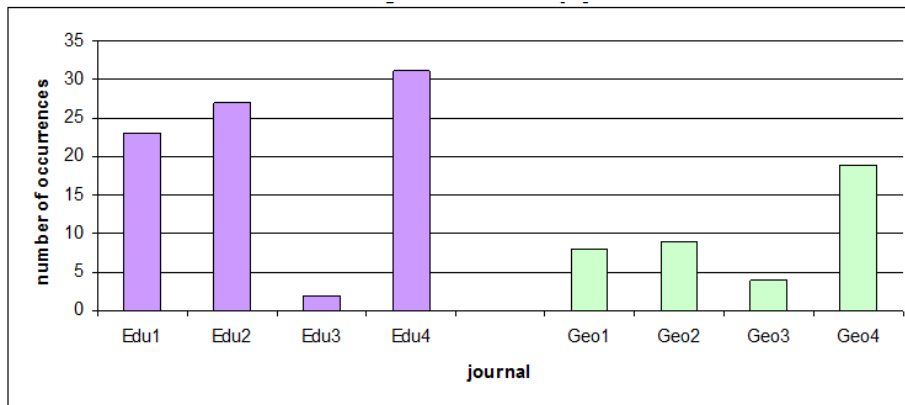


Figure 7. Title units starting with 'the' by journal for education and geology

Once again, however, it is worth noting that not all journals are alike. In Fig. 7 the four source journals of two disciplines are compared in terms of numbers of titles beginning with the definite article. In both cases rather large variation can be observed within the discipline; this seems to be attributable only to the policy or preferences of a particular journal. Incidentally, Edu3 has no guidance regarding titles on its website; the only way to learn of its avoidance of initial definite articles is to observe the titles of published articles.

Conclusion

Even so short a text as the title of a research article displays certain patterns of language use that are shared by other titles in the same discipline. In addition, comparison of cross-disciplinary results shows that there are tendencies for the four disciplines classified as soft sciences to group together, and the same is true of the four disciplines on the hard sciences side. Compared with the soft sciences, the hard sciences tended towards titles with more words, fewer multi-unit titles, more titles consisting of noun phrases, a higher substantive word rate, and lower use of 'the' to start a title unit. More content words are packed in, probably to include key words and information identifying the study as precisely as possible. Soft sciences, with an audience from a variety of academic backgrounds, may try to appeal to a wider variety of readers rather than a smaller set of experts. Interestingly, medicine and botany are often furthest apart from each other within the hard sciences, despite both being essentially biological sciences. Economics also often stands between the hard and other soft sciences, indicating that it may be near the middle of the scale, at least based on the title features investigated.

Clearly the corpus is limited in the number of disciplines and journals represented, and could be both expanded and updated. The choice to concentrate on general journals may also influence title preferences, as indicated by the different conventions used by the specialist

nephrology journals and the general medical journals, and may not represent styles preferred in some sub-disciplines.

The diagrams used here to illustrate disciplinary preferences for four features of title use can probably be applied to other features of titles, such as the use of coordination in titles or modification in noun phrases. It is hoped that the results and the various forms of graphical display can prove useful in introducing novice writers to some features of titles and of language use in the disciplines. Through drawing their attention to these features, we can provide novices with tools to teach themselves. The next step is to test the materials with novice academic writers to determine their usefulness, and to find the most effective means of presenting data that will help build awareness of title features. One task that has proven useful in my experience is to have learners collect titles in their own sub-disciplines, or for a journal that they would like to publish in, and to analyze the features of those titles. It is also important to point out that guidelines or preferences are not set in stone; writers should be aware of other options when designing titles.

References

- Al-Awqati, Q. (2006) 'A general theory of titles!' *Kidney International* 69, 947–948
- Ball, R. (2009) 'Scholarly communication in transition: The use of question marks in the titles of scientific articles in medicine, life sciences and physics 1966-2005'. *Scientometrics* 79 (3), 667–679
- Becher, T. and Trowler, P.R. (2001) *Academic Tribes and Territories*. 2nd edn. Buckingham: Open University Press
- Biglan, A. (1973) 'The characteristics of subject matter in different academic areas'. *Journal of Applied Psychology* 57 (3), 195–203
- Christensen, N.B., Kume, H. and Autorino, R. (2009) 'How to write titles and abstracts for readers'. *International Journal of Urology* 16 (1), 2–3
- Diener, R.A.V. (1984) 'Informational Dynamics of Journal Article Titles'. *Journal of the American Society for Information Science* 35 (4), 222–227
- Dillon, J.T. (1981) 'The Emergence of the Colon: An Empirical Correlate of Scholarship'. *American Psychologist* 36 (8), 879–884
- Gesuato, S. (2008) 'Encoding of information in titles: Academic practices across four genres in linguistics'. In *Ecolingua. The Role of E-corpora in Translation and Language Learning*. ed. by Taylor, C. Trieste: Edizioni Università di Trieste, 127–157
- Haggan, M. (2004) 'Research paper titles in literature, linguistics and science: dimensions of attraction'. *Journal of Pragmatics* 36, 293–31
- Hartley, J. (2007) 'Planning that title: Practices and preferences for titles with colons in academic articles'. *Library and Information Science Research* 29 (4), 553–568
- Hyland, K. (2000) *Disciplinary Discourses: Social Interactions in Academic Writing*. Harlow: Pearson Education
- 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 Reports* 1 (2), 1–5
- Jamali, H.R. and Nikyad, M. (2011) 'Article title type and its relation with the number of downloads and citations'. *Scientometrics* 88, 653–661
- Lewison, G. and Hartley, J. (2005) 'What's in a title? Numbers of words and presence of colons'. *Scientometrics* 63 (2), 341–356
- Moore, A. (2010) 'What's in a title? A two-step approach to optimisation for man and machine'. *BioEssays* 32, 183–184
- Nagano, R.L. (2008) 'Titles of biomedical articles: Data and a debate'. *Porta Lingua* 2008. Szaknyelvoktatók és –Kutatók Országos Egyesülete, 115–123
- Nagano, R. L. (2013) 'The effect of...' or 'Effect of...'? Article use in research paper titles'. *Porta Lingua* 2013. Szaknyelvoktatók és –Kutatók Országos Egyesülete, 277–287
- Soler, V. (2007) 'Writing titles in science: An exploratory study'. *English for Specific Purposes* 26, 90–102
- Yitzhaki, M. (1997) 'Variation in informativity of titles of research paper in selected humanities journals: A comparative study'. *Scientometrics* 38 (2), 219–229

Appendix

Table 2. Title corpus source journals, abbreviations and code numbers

Discipline	Abbr.	No.	Journal
Botany	Bot	1	<i>The Plant Journal</i>
		2	<i>American Journal of Botany</i>
		3	<i>Planta</i>
		4	<i>Annals of Botany</i>
Fluids Engineering	Engg	1	<i>Journal of Fluid Mechanics</i>
		2	<i>Journal of Fluids Engineering</i>
		3	<i>Journal of Fluids and Structures</i>
		4	<i>Physics of Fluids</i>
Geology	Geo	1	<i>Geology</i>
		2	<i>Journal of Geology</i>
		3	<i>International Geology Review</i>
		4	<i>Journal of the Geological Society</i>
Medicine	Med	1	<i>BMJ</i>
		2	<i>Journal of the American Medical Association</i>
		3	<i>New England Journal of Medicine</i>
		4	<i>The Lancet</i>
Economics	Econ	1	<i>American Economic Review</i>
		2	<i>Quarterly Journal of Economics</i>
		3	<i>Review of Economic Studies</i>
		4	<i>European Economic Review</i>
Education	Edu	1	<i>American Educational Research Journal</i>
		2	<i>British Journal of Educational Studies</i>
		3	<i>Journal of Educational Research</i>
		4	<i>British Educational Research Journal</i>
History	Hist	1	<i>Past and Present</i>
		2	<i>Journal of Modern History</i>
		3	<i>History: The Journal of the Historical Association</i>
		4	<i>The Historical Journal</i>
Sociology	Soc	1	<i>American Sociological Review</i>
		2	<i>Social Forces</i>
		3	<i>American Journal of Sociology</i>
		4	<i>British Journal of Sociology</i>