Titles in research articles

Ken Hyland a,*, Hang (Joanna) Zou b

a School of Education and Lifelong Learning, University of East Anglia, UK
b School of Foreign Languages, East China Normal University, China

ARTICLE INFO

Keywords:
Research articles
Titles
Academic writing
Disciplinary differences
Corpus analysis

ABSTRACT

Titles are a key part of every academic genre and are particularly important in research papers. Today, online searches are overwhelmingly based on articles rather than journals which means that writers must, more than ever, make their titles both informative and appealing to attract readers who may go on to read, cite and make use of their research. In this paper we explore the key features of 5070 titles in the leading journals of six disciplines in the human and physical sciences to identify their typical structural patterns and content foci. In addition to proposing a model of title patterns, we show there are major disciplinary differences which can be traced to different characteristics of the fields and of the topics of the articles themselves. Our findings have important implications for EAP and ERPP teachers working with early career academic writers.

1. Introduction

Titles are the most ubiquitous of academic texts, being a key part of every research paper, thesis, conference presentation and blog post. They not only name the work they accompany and act to pull in potential readers, but their importance in research articles has rapidly increased in the age of ‘article-based publishing’. Today almost all the major publishers have prioritised the quicker delivery and easier internet searching of individual papers and relegated the volume/issue shell of journals to archiving. Publication is thus geared towards discrete articles, available online as they are ready without waiting for an issue to be compiled, allowing faster publication and citation. This makes titles a central element of research visibility as readers typically find papers through keyword-generated online search results rather than by scanning the contents tables of favourite journals (e.g. Hsin et al., 2016). As a result, writers have an interest in making their titles both informative and appealing to attract readers who may go on to read, cite and make use of their research. As Belcher observes (2021: 282), “Your title is the highway billboard of your article”.

In an age where careers are dominated by performance metrics, a good title can therefore contribute to the reputation and professional advancement of its writer. The mega Open Access science publisher PLOS, for instance, reinforces the idea:

Once readers find your article, they’ll use the title as the first filter to decide whether your research is what they’re looking for. A strong and specific title is the first step toward citations, inclusion in meta-analyses, and influencing your field.

Given the prominence of titles, advanced EAP students, and those studying English for Research and Publication Purpose courses, would benefit from an awareness of this importance and a better understanding of their features. We hope the findings will also benefit...
those who assist and instruct novice writers with their efforts to publish.

The purpose of this paper is to explore the most frequent patterns and features of titles in leading journals in the human and physical sciences to identify what, through successfully published article titles in top journals, academics regard as ‘strong and specific’ about titles. We focus on 65 papers from each of 13 highly ranked journals in six fields, a total of 5070 titles, to answer the following questions:

1. What are the typical formal patterns of titles in published articles?
2. Are there disciplinary differences in the structure and length of these titles?
3. Are there disciplinary differences in the focus of these titles?
4. How might we explain these differences?

2. Titles in research writing

Researchers, of course, want their work to be read and, perhaps equally, to be cited, and this is increasingly difficult with the tsunami of papers affecting every discipline. Competition to be noticed is intense, as eight million academics in 17,000 universities around the world seek to get their voices heard in English-language journals each year (UNESCO, 2017). One of the largest journal publishers, Elsevier, for example, reports over two million articles submitted and one billion read in 2019 (RELX, 2020). Authors, then, need to leverage every advantage they can to hook potential readers, especially as search patterns now involve combing internet databases rather than individual journals. The titles of academic papers thus reflect how authors believe they will best attract an audience for their work, and various means are used to achieve this, including statements, questions, puns, cultural allusions, quotations, compounds, etc. It is unclear, however, how authors achieve their goal in different fields.

What we do know is that the ways readers use titles have changed as technologies have come to determine the search and, by extension, reading patterns of academics. Thirty years ago, for example, Bazerman (1988: 239) spoke of biophysicists in fast-moving fields staying current by visiting the library once a week to search the tables of contents of newly arrived journals. Scanning these tables can involve rapid processing based on the topics or approaches in titles, which ‘triggered their attention’. Even 15 years ago, title-scanning was still:

> a regular activity within the scientific community for whom part of its daily routine is to search titles in libraries, catalogues, periodical indexes, references, databases, and tables of contents of edited books, reports, and proceedings.

(Soler, 2007: 91)

Today, academics are more likely to search Google Scholar, rely on email alerts from key journals or use tools such as PubCrawler, which automatically searches online publication databases such as PubMed using key search terms and then sends weekly emails highlighting potentially relevant papers (Pain, 2016). All these methods, however, rely on the information writers provide in their titles.

To better accommodate these internet searches, numerous science journals have introduced detailed reporting guidelines. In the biomedical sciences, for example, journals introduced title requirements 20 years ago, and this may have influenced both the form and content of titles in that field. One major shift has been longer titles and the growing mention of methods in clinical Research Article (RA) titles, although this trend seems to differ by journal (Kerans et al., 2020). There has also been a substantial and consistent trend towards more concrete and definitive titles in psychology since the mid-1980s (Whissell, 2012) and, despite advice not to use them, an increasing use of question marks in the social sciences (Ball, 2009). These requirements and trends potentially create problems for authors using English as an additional language as longer, richer titles mean having to deal with greater phrasing complexity.

There is, however, no shortage of advice for learners on how to write enticing titles. Most sources recommend that writers keep their titles short, avoid irony, jargon and humour and aim to make them clear, informative and specific (e.g. Grant, 2013; Gustavil, 2012; Hays, 2010; Hartley, 2005). Such discussions of titles, however, tend to be short, prescriptive and intuitive. The widely used books by Belcher (2021), Day (2008) and Cargill and O’Connor (2013), for example, devote only a few pages at most to titles and offer general advice such as making the title specific, relevant and packed with searchable keywords. Paltridge and Starfield (2016) and Swales and Feak (2012), however, encourage readers to analyse the titles they find in target journals and disciplines, asking if they get a clear sense from them of what the article is about. Advice also appears in authoritative style guides, so, for example, the APA (2020) manual advises writers to keep their title focused, no more than 12 words long and without ‘abbreviations or words that serve no purpose’. In the sciences, PLOS also recommends brevity, informativeness and incorporation of keywords; it also admonishes writers to avoid sensationalism and questions.

Some journals also specify title requirements, especially in the sciences, some being very specific:

Nature: Titles must fit on two lines in print and should avoid technical terms, abbreviations and active verbs.

Cell: The title should capture the conceptual significance for a broad audience. The most effective titles are no more than 10 to 12 words and provide an overall view of the paper’s significance rather than the detailed contents of the paper.

---

3 https://pubcrawler.gen.tcd.ie/.
5 https://www.nature.com/nature/for-authors/initial-submission.
6 https://www.cell.com/pb-assets/journals/research/matter/Matter_IFA.PDF.
BMJ: The title should be informative and, for research papers, a subtitle with the study design (for example, “a phase III clinical trial” or “a systematic review and meta-analysis”).

Communications on Pure and Applied Mathematics: A short informative title that contains the major key words. The title should not contain abbreviations.

It is unclear, however, the extent to which writers actually observe these instructions. Kerans et al. (2020), for example, found that fewer than half the titles in their corpus of medical journals actually complied with instructions to mention a result, although some might have implied results to an insider. About a third of the titles they coded as including results, moreover, only did so implicitly with phrases like “improved [outcome] with [treatment]”. In Goodman et al.’s (2001) medical corpus results appeared in only 18% of titles. Similarly, Gastel and Day’s (2017 p. 42) admonishments to avoid compound titles (those divided into two parts), as pedantic and confusing to indexes, seems to be largely ignored in the sciences with an increase in recent years (e.g. Milojevic, 2017). Soler (2007) also found a heavy use of compound constructions in social science disciplines while Lewis and Hartley (2005) and Hartley (2007) found compound constructions, especially those where the parts were separated by a colon, to be very common in titles.

Most research studies have taken a normative stance, seeking to advise authors what they should do to gain attention, by identifying title features which correlate with high citation counts (e.g. Paiva et al., 2012; Thelwall, 2017). So, longer titles are cited more often in many disciplines (Jacques & Sebire, 2010, pp. 2–3; van Wesel, Wyatt & ten Haaf, 2014), although this might not be the case in pure sciences (van Wesel et al., 2014). Titles that contain a colon may be more likely to attract citations (Buter & van Raan, 2011; Jacques & Sebire, 2010), while those with question marks had poorer citation rates (Paiva et al., 2012).

In the largest study to date, Hudson (2016) examined the impact of multiple authorship on the titles of papers submitted to the UK’s four-yearly Research Evaluation Framework (the REF) in 2014. This involved 155,500 articles in 36 disciplines with the main findings being that: (i) the lengths of the titles increased with the number of authors in almost all disciplines, (ii) the use of colons and question marks tended to decline with increasing author numbers in most fields, (iii) citations to papers were higher when the titles were shorter and contained colons, (iv) citations were lower when they contained question marks. The author observes, however, that these results varied considerably between the disciplines, a finding which is repeated for other features of titles.

Hartley (2007) showed disciplinary differences in the use of colons, for instance, with a greater use in the arts than in the sciences and also in the work of single authors than multiple authors. Haggan (2004) concurs, showing colons to comprise 61% of titles in literature journals compared with less than 30% in linguistics and sciences. Nagano (2015) and Milojevic (2017) argue that the length of titles also differs between disciplines, being longer in the hard sciences such as medicine than the discursive sciences such as sociology. Terminology has also attracted attention. Anthony (2001), for example, found considerable use of discipline and field-specific lexis in 600 papers from six IEEE computer science journals. Looking at the form of titles in a small corpus of 574 titles in three disciplines, Appiah et al. (2019) found that noun phrases dominated the titles across the three fields, but business titles were longer than those in gynaecology and law and more likely to contain compound units with a colon. Soler (2007) found a prevalence of nominal group titles in the science titles in her sample of 480 RA titles and that full-title sentence constructions were a disciplinary peculiarity of Biology titles. Wang and Bai (2007) also found that 99% of titles in the New England Journal of Medicine were nominal groups. Kerans et al. (2020) noted substantial variation in the mention of methods in titles and the mention of results differed between general and more specific journals.

Overall, the relative lack of research into titles may lead us to suppose that practices are homogeneous across the academy, but what research there is shows different interests and foci with titles differentiated by topic, journal and discipline. This is because, we assume, titles aim to both attract and inform and these goals are achieved in different ways in different contexts. Titles have to be sufficiently appealing to grabbing attention and appropriately descriptive of the article’s content to be detected by information retrieval systems. However, there is a tension here as longer titles may increase information content but reduce the amount of interest they create. These variations can confuse novice writers and their teachers and so their analysis may assist those engaged in writing for publication courses. Here we explore the use of titles in six contrasting fields to identify their main characteristics and the ways that discipline can influence them.

3. Corpus and methods

In order to compile as representative a corpus as possible we aimed for both adequate breadth and a large size. We selected six disciplines offering a wide spread of disciplinary practices from the sciences, social sciences and humanities using categories recognised by Scimago:\footnote{https://www.scimagojr.com/journalrank.php.}\footnote{https://onlinelibrary.wiley.com/page/journal/10970312/homepage/forauthors.html.} applied linguistics (AL), education (Edn), history (His), mathematics (Math), mechanical engineering (Eng) and biochemistry (Bio). From each discipline we selected 13 journals from among the highest rated journals on the Scimago journal rankings (list in Appendix). From the tables of contents on each individual journal website we used stratified random sampling to identify the titles of 65 original full research articles published over the last 8 years to ensure greater comparability, eliminating reviews, correspondence, case reports and theoretical papers. This produced a corpus of 5070 article titles (Table 1) which exceeded almost all previous studies of titles.

The corpus was searched both manually and automatically, using AntConc (Anthony, 2019), to gain some understanding of the titles, identifying non-alphanumeric characters, such as colons, hyphens, question marks, etc, as well as words with the potential to
mark content. After several attempts, we devised the structure described in section 5 below, distinguishing ‘indicative’ and ‘interrogative’ according to syntax, the use of punctuation to create two-part titles, and whether the examples referenced methods or results. Using this structure, each author independently coded a 30% sample of the corpus manually and an inter-rater agreement of 96% was achieved through discussion. Intra-reliability tests were also conducted by one author re-categorising 20% of the cases two weeks after the initial coding with full agreement between the two. All titles were then categorised according to the model.

We chose to focus on these formal features as objective and measurable means of distinguishing the preferences of different disciplines for particular kinds of titles. They are the characteristics that are most often discussed in the research literature and have tended to occur most frequently in style guides. Some ‘how to’ authors admonish their readers to avoid ‘broad’, ‘dense’, ‘ambiguous’ or ‘clever’ titles (e.g. Belcher, 2021), but we feel ill-equipped to make judgements which might not correspond with those of others – particularly those of disciplinary insiders. How explicit does a title on endocrinology or electromagnetic fields need to be before it becomes ‘dense’, for example? Without clear identifying criteria these are problematic categories to study in a large corpus such as ours. We also assume these transgressions will occur only infrequently in these successfully published papers. All our numeric data were imported into MATLAB (Version: R2020b (9.9.0. 1467703), a programming and numeric computing platform for analyzing data and creating models. We used this to calculate the length of titles, grouping them by disciplines, and the frequencies of each target feature per 100 words. The results are discussed in the following sections.

4. At some length: how long is a title?

We first looked at the length of titles, finding considerable disciplinary variability. Title length is considered important for the retrieval, and thus the eventual citation, of articles. Because longer titles contain more words they offer more scope for search words but this can also make them inelegant and “more difficult to digest and may reduce the attraction factor” (Hudson, 2016, p. 878). Shorter titles may also indicate a more specific scope and can attract a higher number of views and citations (Paiva et al., 2012), although other studies have found the opposite. Jacques and Sebire (2010) and van Wesel et al. (2014), for example, found a positive correlation between length and impact in medical articles and Milojevic (2017) found the same in astronomy and ecology but not in mathematics, robotics and economics. There also seems to be a negative relationship in biology and social sciences (Didegah & Thelwall, 2013), sociology (van Wesel et al., 2014), and psychology (Subotic & Mukherjee, 2014). So the jury is out on the advantages of longer titles and a causal connection remains to be established.

As we noted above, the average length of titles seems to be strongly discipline-dependent (Milojevic, 2017; Nagano, 2015; Soler, 2007), with longer titles produced by larger teams (White, 1991) and now averaging 40% longer than 60 years ago (Milojevic, 2017). Speculatively this is a result of the greater detail needed to distinguish them from previous, possibly more general, studies as they evolve. Kerans et al.’s (2020) study of RA titles in eight top-ranked journals in clinical medicine notes that exceptionally long titles appear consistently only in some clinical journals and that the trend towards longer titles may have stabilised.

Using the technical computing language MATLAB, we grouped all 5070 titles into four categories depending on their length. These results, shown in Table 2, confirm disciplinary variation in title length, revealing the surprising finding that longer titles tended to be more common in the more discursive social science disciplines than the STEM papers. This contradicts Nagano (2015), but as she focused on articles in just four ‘relatively prestigious’ journals in each field, the results may have been influenced by the conventions of particular journals. Similarly, Soler (2007) found average title lengths higher in three hard sciences, but this was a smaller-scale study and refers to articles published between 1996 and 2002, and so does not take account of changes in the last 20 years.

As can be seen, applied linguists, educationalists and historians seem to prefer titles over 12.5 words long while mathematicians, biochemists and engineers had much lower averages. Like our study, Lewison and Hartley (2005), looking at half a dozen STEM fields, found that mathematicians employed the shortest titles, with around nine words, while biology had the longest, averaging 15. In fact, 70% of the titles in our maths corpus were under 10 words. The reason for these disciplinary differences is unclear, but the numbers certainly suggest definite disciplinary preferences.

One explanation for different title lengths may lie in the nature of the specific research undertaken. Thus in newly emerging areas, those with little activity or in overviews, shorter titles are able to capture the broad scope of the paper in an immediate and engaging way:

1) Lumpen Politics? A Day in “El Hueco” (Hist)
2) Additive manufacturing of multi-material structures (Eng)
3) Does College Influence Sociopolitical Attitudes? (Edn)

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Titles</th>
<th>Words</th>
<th>Sub-totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applied Linguistics</td>
<td>845</td>
<td>11312</td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>845</td>
<td>12016</td>
<td></td>
</tr>
<tr>
<td>Engineering</td>
<td>845</td>
<td>10700</td>
<td>34,028</td>
</tr>
<tr>
<td>Mathematics</td>
<td>845</td>
<td>7581</td>
<td></td>
</tr>
<tr>
<td>History</td>
<td>845</td>
<td>9379</td>
<td></td>
</tr>
<tr>
<td>Biochemistry</td>
<td>845</td>
<td>8157</td>
<td>25117</td>
</tr>
<tr>
<td>Total</td>
<td>5070</td>
<td>59,145</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Corpus size and composition.
Such succinctness, of course, runs the risk of promising more than can be delivered within the confines of a short paper, or, equally, of condensing information to the point of excluding all readers but the most informed specialist:

4) Riemannian hyperbolization (Maths)
5) Regular supercuspidal representations. (Maths)
6) A Thermally Reformable Protein Polymer (Bio)

The ability to compress a topic to its bare bones in this way therefore risks narrowing the potential audience for a paper and so minimising its impact in terms of readership and citations.

In contrast, authors working in more densely explored and contested areas are more likely to spell out what is distinctive about their study to distinguish it from the herd and attract the larger number of authors working in that field:

7) A unified methodology for calculation of compliance and stiffness contribution tensors of inhomogeneities of arbitrary 2D and 3D shapes embedded in isotropic matrix - open access software (Eng)
8) Do Schools Reduce or Exacerbate Inequality? How the Associations Between Student Achievement and Achievement Growth Influence Our Understanding of the Role of Schooling. (Edn)

Such explicit labelling, however, lacks the catchy impact of shorter versions and may be missed by the algorithms which search for research topics.

An alternative explanation for differences in title lengths is that unlike physical scientists, social scientists work in areas which often span more disparate audiences which may lack the background needed to see the importance and coverage of the work. Research is often less linear and more disparate, interdisciplinary and less bound to a specific context of knowledge, which means titles may require greater explicitness, and so length, to create such a recognizable context:

9) Where’s the vision? Rescuing integrativeness to understand the language learning motivation of English-speaking EFL instructors living in South Korea. (AL)

Similarly, longer titles help spell out the topic to specify what might be new, encouraging readers to explore the text further and discover results relevant to their own work:

11) Do Schools Reduce or Exacerbate Inequality? How the Associations Between Student Achievement and Achievement Growth Influence Our Understanding of the Role of Schooling (Edn)
12) Camp Composition and the Rocky Mountain Fur Trade: Estimating the Native Presence at the Early Rendezvous and Winter Camps, 1825–1830 (Hist)

Whatever the reason, it is clear that some disciplines have a greater tolerance or need for longer titles and this allows authors to offer a more explicit representation of their work. We now turn to the ways authors structure their titles.

5. Representing research: title structures

The literature offers almost as many structures to describe titles as there are studies of them. Hartley (2008) lists 13 different formats. These range from titles that announce the general subject (e.g. On writing academic titles), state findings (e.g. titles grow with number of authors) or indicate a controlling question (e.g. How do academics write titles?) through to those that offer descriptions or options (e.g. Seven types of title). Kerans et al. (2020) focus on the parts of medical journal articles included in their titles, together with the number of parts of the titles and their punctuation, adding up to ten sub-types. Soler (2007) gives a description of four types: nominal group, full sentence, compound, and question while, in a later study of 260 titles in one education journal, Hartley (2017) suggests three: those with a colon, single sentence and question. Jamali and Nikzad (2011) also identify three common types:
declarative titles, which state the main findings, descriptive titles, which simply describe the subject of the article, and interrogative titles, which present the topic as a question.

As can be seen, however, these categorisations often tend to confuse form with content, fail to show the connections between them and obscure the fact that very few of the described styles are actually used, thus giving undue prominence to infrequent types. Some prioritise positioning or grammatical patterning, and others punctuation or content. We sought to create a robust and comprehensive model to help illuminate the options available to writers and reveal the preferences of those working in different disciplines. Drawing on the earlier models and after several abortive attempts, we settled on a tree diagram which categorises the titles according to their form and then their content. This meant coding all 5070 cases as having either an Indicative or Interrogative structure, and identifying these as either a single sentence, conveying the topic, or a compound, comprising two parts separated by a colon or other punctuation. Each Compound can be descriptive, simply stating the topic, or detailed, including the approach or results. This is shown diagrammatically as Fig. 1 together with the raw figures for each branch to indicate its overall use. We elaborate these below.

5.1. Syntactic types: indicative or interrogative?

Titles are either indicative or interrogative: posed as a statement or a question. Fig. 1 shows that they are overwhelmingly of the former type in all fields. The latter type comprise just 8.5% of all titles in our corpus. Table 3, however, shows clear disciplinary differences, with less than 1% of interrogative titles in the three STEM disciplines – which show a strong preference for indicative titles.

Clearly, indicative titles are able to express the topic of the paper in a straightforward, no-nonsense way, either briefly or with elaboration to include not only the topic (13), but also the approach (14) or the result (15):

(13) Geometric criteria for overtwistedness (Maths)
(14) A Longitudinal Study of Voice Onset Time Development in L2 Spanish Stops (AL)
(15) The Hannes hand prosthesis replicates the key biological properties of the human hand (Eng)

These writers are intent on presenting the topic of their paper in a clear and accessible way which minimizes processing effort and will not confuse algorithmic searches. The preference for indicative titles in the sciences might be a consequence of this and the firm admonishments of key journals, publishers and style guides to ensure brevity and informativity and avoid questions, which might be seen as distracting and gimmicky.

Although other studies confirm this, it is, perhaps, surprising to find so few interrogatives in the corpus. After all, academic research is governed by questioning. Papers are written with questions in mind or problems to solve, but they seem to rarely emerge explicitly in titles. Despite increasing over time, they are still rare in the sciences (Cook & Plourde, 2016; Fox & Burns, 2015). Hudson (2016) found that titles which included question marks were longer than other titles, helping to explain the longer titles in the soft sciences. While it is hard to establish causality, a study of over 2000 articles published in 6 PLoS science journals found that those with question titles were more often downloaded, but less frequently cited (Jamali & Nikzad, 2011).

Questions, however, exploit the interactivity of familiar conversational discourses and have the impact to grab the reader at the outset with an arresting directness and to demand attention with a striking expression. This is all the more provocative when found in a discipline which normally shuns interrogatives:

![Fig. 1. A categorization of titles.](image-url)
Questions in titles, however, are never a naïve display of ignorance, nor do they simply signal the topic of a paper. They both promote the article itself to the reader and represent the writer as someone with an insider’s understanding of what constitutes a real issue and who has a plausible response to it.

5.2. Formats: single or compound?

Beyond selecting either an indicative or interrogative title, authors have the option of a single format, describing the article in one sentence, or as a compound of two or more parts separated by punctuation, usually a colon. Tables 4 and 5 show that single formats predominated, but only marginally, at 55% of the total, with compounds dominating the interrogative forms. More striking, perhaps, are the disciplinary contrasts, with the hard sciences overwhelmingly favouring single forms and the soft sciences preferring compounds.

Single titles, as we have noted, generally offer straightforward representations of what the reader will find in the accompanying article. Their effectiveness lies in their transparency and this can be disarmingly direct and effectively minimalistic:

(19) Soft sensors that can feel it all (Eng)
(20) A game-theoretic analysis of the Waterloo campaign (Hist)

Often, however, their impact lies in their detail, providing more targets for automated searches and suggesting both potentially relevant material for readers and a credible indication of the author’s competence:

(21) Multistage distributionally robust mixed-integer programming with decision-dependent moment-based ambiguity sets (Maths)
(22) Roles of MACl in Sequentially Deposited Bromine-Free Perovskite Absorbers for Efficient Solar Cells (Eng)

Interrogative forms can be particularly disarming, apparently posing real and enticing questions that are almost journalistic in their effect of capturing attention and arousing interest:

(23) How Does the Ribosome Fold the Proteome? (Bio)
(24) How many rushed during the Oklahoma land openings? (Hist)

Single titles are therefore typically unadorned and explicit, speaking to the community directly and through the search engines they use. This, then, may be one important reason why the STEM authors favour this form as in some ways their audience is relatively better defined than the social science readers. The way that scientific knowledge is created, step-by-step in a more linear fashion, helps define a specific context of research to which the current work is a contribution: a problem being addressed within an established framework of knowledge. This means that scientists can presuppose a certain amount of shared knowledge with the potential audience and can refer to this in a direct and straightforward way. The social scientists, on the other hand, often need to elaborate what readers will find in the paper to attract an audience who may be unfamiliar with the area (e.g. Becher & Trowler, 2001; Hyland, 2004).

Compound titles, containing non-alphanumeric characters, such as colons, hyphens, dashes, and periods, allow authors to expand the topic in this way and are extremely commonplace across the soft knowledge disciplines. Only 15% of the hard science titles in our corpus comprised compound types and the differences across fields, in fact, are considerable: ranging from 82% in History to 5% in

### Table 3
Distribution of syntactic forms in titles.

<table>
<thead>
<tr>
<th></th>
<th>App Ling</th>
<th>Education</th>
<th>History</th>
<th>Maths</th>
<th>Biochemistry</th>
<th>Engineering</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. %</td>
<td>No. %</td>
<td>No. %</td>
<td>No. %</td>
<td>No. %</td>
<td>No. %</td>
</tr>
<tr>
<td>Indicative</td>
<td>724</td>
<td>85.7</td>
<td>646</td>
<td>76.4</td>
<td>769</td>
<td>91.0</td>
</tr>
<tr>
<td>Interrogative</td>
<td>121</td>
<td>14.3</td>
<td>199</td>
<td>23.6</td>
<td>76</td>
<td>9.0</td>
</tr>
<tr>
<td>Total</td>
<td>845</td>
<td>100.0</td>
<td>845</td>
<td>100.0</td>
<td>845</td>
<td>100.0</td>
</tr>
</tbody>
</table>

### Table 4
Distribution of Indicative patterns (number and %).

<table>
<thead>
<tr>
<th></th>
<th>App Ling</th>
<th>Education</th>
<th>History</th>
<th>Maths</th>
<th>Biochemistry</th>
<th>Engineering</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. %</td>
<td>No. %</td>
<td>No. %</td>
<td>No. %</td>
<td>No. %</td>
<td>No. %</td>
</tr>
<tr>
<td>Single</td>
<td>263</td>
<td>36.3</td>
<td>189</td>
<td>29.3</td>
<td>140</td>
<td>18.2</td>
</tr>
<tr>
<td>Compound</td>
<td>461</td>
<td>63.7</td>
<td>457</td>
<td>70.7</td>
<td>629</td>
<td>81.8</td>
</tr>
<tr>
<td>Total</td>
<td>724</td>
<td>100.0</td>
<td>646</td>
<td>100.0</td>
<td>769</td>
<td>100.0</td>
</tr>
</tbody>
</table>
Maths. Hartley (2007) also found greater use of compounds in the Arts than in the sciences while Lewison and Hartley (2005) observed that single authors used significantly more colonic titles than groups of authors, suggesting why they might be dominant in the soft fields. Overall, however, the literature has been preoccupied with discovering whether compound titles are associated with higher citations. Jacques and Sebire (2010) and Buter and van Raan (2011), for example, believe they are while Hartley (2007) found the opposite. Thus, once again, disciplinary preferences influence practices. As can be seen in our examples, the prevalent punctuation is the colon, which seems to have increased in popularity in recent years (Lewison & Hartley, 2005) although Jamali and Nikzad (2011) found that articles with a colon had both fewer downloads and fewer citations.

Using punctuation to provide a subtitle or a multicomponent title enables authors to pack more information into their title and so may be associated with the longer titles in the social sciences. Compounds allow authors to pack more information into a title by either elaborating the topic (25) or by specifying it more precisely (26).

(25) Recipes for Respect: African American Meals and Meaning (Hist)
(26) Simulation-Based Learning in Higher Education: A Meta-Analysis (Edn)

Compounds also provide authors with the opportunity to enliven their title, sometimes by a pop culture reference (27–28) or more often by a quotation (29–30), which are particularly favoured by applied linguists and historians:

(27) Space Oddities: The Acquisition of Agreement Verbs by L2 Learners of Sign Language of the Netherlands (AL)
(28) Sympathy for the devil? A defence of EAP (AL)
(29) “We are all republicans”: Political Articulation and the Production of Nationhood in France’s Face Veil Debate. (Hist)
(30) “I believe the findings are fascinating”: Stance in 3-min theses (AL)

Interrogative compounds, as Table 5 shows, are almost non-existent in the science titles, representing the two patterns least favoured by scientists. A desire to convey the content direct and concisely, the admonishment of style guides and the proscriptions of journals may combine to restrict this form to the soft knowledge journals. We turn to these compound types in more detail in the next section.

Table 5

<table>
<thead>
<tr>
<th></th>
<th>App Ling</th>
<th>Education</th>
<th>History</th>
<th>Maths</th>
<th>Biochemistry</th>
<th>Engineering</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>%</td>
<td>No.</td>
<td>%</td>
<td>No.</td>
<td>%</td>
</tr>
<tr>
<td>Single</td>
<td>22</td>
<td>18.2</td>
<td>23</td>
<td>11.6</td>
<td>10</td>
<td>13.2</td>
</tr>
<tr>
<td>Compound</td>
<td>99</td>
<td>81.8</td>
<td>176</td>
<td>88.4</td>
<td>66</td>
<td>86.8</td>
</tr>
<tr>
<td>Total</td>
<td>121</td>
<td>100.0</td>
<td>199</td>
<td>100.0</td>
<td>76</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Table 6

<table>
<thead>
<tr>
<th></th>
<th>Descriptive (Topic only)</th>
<th>Detailed (Approach or result)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indicative</td>
<td>76.8</td>
<td>23.2</td>
<td>100</td>
</tr>
<tr>
<td>Interrogative</td>
<td>78.9</td>
<td>21.1</td>
<td>100</td>
</tr>
</tbody>
</table>

K. Hyland and H.(J. Zou

5.3. Compound content: descriptive or detailed?

The final level of our hierarchy of author choices concerns the content of the title and whether to opt for a descriptive title or one adorned with greater detail. Because single types principally concern the topic of the paper, we focus here on how authors exercise the options that compound titles make available. Compounds, as we noted, assist authors in presenting an informative, yet concise, statement of their paper. The interrelationship between the two parts can succinctly present the object of study in different ways. They may offer the topic in general and specific terms, a problem followed by a solution, a topic with a method, a topic with a result, and so on. We suggest that the author has two main options: should the title be descriptive, that is, indicate only the subject of the article, or detailed, stating the main findings or approach? Interestingly, while the latter is more informative, over three quarters of our 5070 titles were descriptive (Table 6).

(31) Breaking barriers: Quebec platelet disorder (Bio)
(32) “I’m a Teacher, I’m Gonna Always Protect You”: Understanding Black Educators’ Protection of Black Children (Edn)

Alternatively the second part can function to generate interest from a recognizable context (33) or offer a solution to known problems (34):

Table 6

<table>
<thead>
<tr>
<th></th>
<th>Descriptive types (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>i Descriptive types</td>
<td>can clarify the topic following a general, attention-grabbing opening (31), or quotation (32):</td>
</tr>
<tr>
<td>Indicative</td>
<td>76.8</td>
</tr>
<tr>
<td>Interrogative</td>
<td>78.9</td>
</tr>
</tbody>
</table>
In interrogative formats, descriptive titles can make a powerful impression by arresting the reader with an eye catching and stimulating problem, following up the question with a clearer statement of the topic:

(35) Can pandemics affect educational attainment? Evidence from the polio epidemic of 1916 (Hist)
(36) Should We Listen or Read? Modality Effects in Implicit and Explicit Knowledge (AL)

Such strategies present readers with an issue that they realise they may want to learn more about, recognising that the question is rhetorical and the title promises an answer in the paper. Interestingly, biochemists tend to favour a reversal of this pattern, with the question concluding a statement of the topic, thereby problematising the familiar and raising an issue which challenges current thinking on a topic:

(37) Lipid nanovesicles for biomedical applications: ‘What is in a name’? (Bio)
(38) Plant-based sterols and stanols in health & disease: “Consequences of human development in a plant-based environment?” (Bio)
(39) An appraisal of meta-analysis guidelines: how do they relate to safety outcomes? (Bio)

Detailed titles, on the other hand, are more informative. They not only describe the topic but also offer greater detail about either the main findings of the study or, more often, the methodology used to reach them. Thus, in (40) a question informs the reader what the paper is about and then how this will be answered using a comparative methodology of legislative proposals. Example (41), an indicative case, informs potential readers that the article contains a study of a current problem using a cutting-edge methodology.

(40) Exclusion or exemption from risk regulation? A comparative analysis of proposals to amend the EU GMO legislation (Bio)
(41) Laser writing of electronic circuitry in thin film molybdenum disulfide: A transformative manufacturing approach (Eng)

All disciplines show a clear preference for descriptive titles, preferring to hold their fire on the details of their studies. Table 7 shows that only education and engineering flout this pattern in any major way, presenting about one-third of their titles as detailed types. Titles in history and mathematics offer readers very little additional information about the papers they accompany.

When authors chose to provide a detailed title they overwhelmingly did so by adding the approach or methodology adopted in the study, rather than the results, so aiding searches and encouraging readers to read on. This may be especially useful in attracting readers who may be using the same method or looking for innovative approaches to a shared problem:

(42) Developing and Validating an Academic Vocabulary List in Russian: A Computational Approach. (AL)
(43) Assessing the security of electricity supply through multi-scale modeling: The TIMES-ANTARES linking approach (Eng)
(44) Avant-garde: an automated data-driven DIA data curation tool (Bio)

Combining an approach with an interrogative in a detailed title can be particularly effective in making a strong impression on potential readers by suggesting that the writer has a methodological solution to a common problem:

(45) “It has no meaning to me.” How do researchers understand the effectiveness of literature searches? A qualitative analysis and preliminary typology of understandings (Edn)
(46) Thriving? Or Surviving? An Approach-Avoidance Perspective on Adult Language Learners’ Motivation (AL)

The inclusion of results in titles seems to be on the increase in biological science journals such as Cell (Rosner, 1990) but is somewhat controversial, especially in medical literature, with some journals banning mention of results and others insisting on it (McGowan & Tugwell, 2005). In Kerans et al.’s (2020) small-scale study of titles in medical papers, only those in the dermatology journals mentioned results and Wager et al. (2016) found readers were not overly influenced by them. In our corpus they are relatively rare, presumably because authors don’t want to give too much away before they get a click recorded on the paper itself. Where they do

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Indicative</th>
<th>Interrogative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Descriptive</td>
<td>Detailed</td>
</tr>
<tr>
<td>App Ling</td>
<td>71.8</td>
<td>28.2</td>
</tr>
<tr>
<td>Education</td>
<td>57.5</td>
<td>42.5</td>
</tr>
<tr>
<td>History</td>
<td>96.7</td>
<td>3.3</td>
</tr>
<tr>
<td>Mathematics</td>
<td>88.9</td>
<td>11.1</td>
</tr>
<tr>
<td>Biochemistry</td>
<td>76.3</td>
<td>23.7</td>
</tr>
<tr>
<td>Engineering</td>
<td>67.2</td>
<td>32.8</td>
</tr>
</tbody>
</table>
occur, they are often outsider-opaque, enticing further investigation of the paper:

(47) Repeat after us: Syntactic alignment is not partner-specific (AL)
(48) Nothing to lose: why early career scientists make ideal entrepreneurs (Bio)
(49) Does an Option to Review Instructional Explanations Enhance Example-Based Learning? It Depends on Learners’ Academic Self-Concept (AL)

An interesting variation, found almost exclusively in the applied linguistics papers, is the writer’s decision to front the result in the title, thereby challenging title norms and forcing readers to reflect for a moment:

(50) Statistical Regularities Affect the Perception of Second Language Speech: Evidence from Adult Classroom Learners of Mandarin Chinese (AL)
(51) Second and third language learners’ sensitivity to Japanese pitch accent is additive: an information-based model of pitch perception (AL)
(52) Pronunciation Can Be Acquired Outside the Classroom: Design and Assessment of Homework-Based Training (AL)

This strategy therefore achieves its impact by creating a dissonance in the reader’s expectations, presenting what is ‘new’, or what might be thought to be the bottom line of the paper, in the thematic ‘given’ position. This then encourages the reader to pause in scanning potential titles to focus on this one and perhaps be drawn to the rhyme of the title which contains information about the way the result was obtained.

6. Conclusions

Article titles are a rich source of information about disciplines, fields, journals and the proclivities of individual authors. They reveal something of fashionable topics and emerging sub-fields, but for EAP researchers they are important for what they tell us about authors’ perceptions of disciplinary practices and how they think they might best reach potential readers. While most studies have focused (inconclusively) on correlations of particular aspects of titles with impact in specific fields, we have cast our net more widely and sought to discover patterns and characteristics of titles across a range of disciplines. Such information is useful for understanding more about the processes of knowledge creation, about publishing practices, and about how we might usefully advise novice researchers.

Returning to our research questions, we have found some contradictions between the advice provided by style guides and journal requirements and actual instances of academic titles. While Gastel and Day’s (2017) caution to avoid questions is largely adhered to, especially in the sciences, authors make considerable use of compound titles to elaborate their topics and, occasionally, to indicate their methods. Unfortunately, much of the advice, while well-intentioned and sensible, often seems vague or contradictory. Students, for example, are likely to have difficulties in making practical use of admonishments such as “provide as much relevant information as possible but be concise” (Cargill & O’Connor, 2013: 61), “have an argumentative title” (Belcher, 2021, p. 289), or to make their titles “stand out in some way … but also tell the reader what the paper is about” (Hartley, 2008, p. 23). We have also observed that titles in the more discursive soft fields are generally longer, contain more interrogative and compound forms and, aside from history, are more likely to mention the approach or findings.

These differences are hard to explain given the need for hard scientists to clearly distinguish their work from others in the heavily crowded lines of inquiry pursued in the STEM disciplines. Theirs is a highly competitive environment of considerable investment and fast-paced publishing to establish early priority, and where researchers inhabit a relatively discrete and clearly identifiable area of study. The social scientists, however, have their own issues to address. While their research might be more diverse and interdisciplinary, the fact that it follows less linear tracks and seeks to appeal to wider, more heterogeneous audiences, means that it is harder to assume a coherent and expectant cohort of readers. It is therefore possible, because they cannot presuppose the same degree of research background, knowledge of prior work and theoretical understandings among readers, their titles have to be catchier and more explicit to create rather than assume an audience for the work.

In other words, titles are not produced from an infinite range of options but from a relatively restricted sub-set which reveal how their authors understand their communities. Any title anticipates a reader’s response and addresses wider expectations, so that choices are disciplinary practices as much as individual decisions. Titles, then, like the articles which follow them, are co-constructed within the “conceptual frames” (Hyland, 2012) of disciplines and so represent something of how members understand what is likely to be effective and familiar to colleagues. In the classroom this means, like most features of academic writing, that one size does not fit all and that teachers mentoring novice researchers might be advised to focus on both the discipline, and possibly the journals, that students are keen to target. Teachers might find it useful, for example, to ask students to ponder the purpose of titles and consider a sample with different structures on the same topic in their own area in order to judge their effectiveness. Alternatively, they might edit draft titles to improve their impact or compare the features of titles in different fields to discover how they work.

Finally, we hope our description of titles proves useful to teachers and that researchers feel it is worth developing further, perhaps through investigating users’ preferences and perspectives. The title is often an unconsidered aspect of academic discourse research, but it plays an increasingly important role in knowledge construction by encouraging readers to notice, and perhaps subsequently read and cite, new research. In Swales’ (1990: 224) memorable words, “Titles consist of only a few words, but they are serious stuff.”
Author statement

None

Appendix. Source of titles

Applied Linguistics

Modern Language Journal
Studies in second language acquisition
Applied linguistics
Journal of memory and language
Language teaching research
Language learning
Annual review of applied linguistic
TESOL quarterly
Journal of English for Academic Purposes
Journal of Second Language Writing
Computer assisted language learning
Language, Cognition and Neuroscience
System

Education

Review of Educational research
Internet and higher education
Sociology of education
Educational researchers
Research Synthesis methods
Journal of engineering education
Journal of educational psychology
American Educational research journal
Educational research review
Education finance and policy
Educational evaluation and policy analysis
Journal of teacher education
Developmental review

History

American Historical Review
Clometrics
Comparative Studies In Society and History
Environmental History
Journal of Global History
Journal of Victorian Culture
Journal of Modern History
History and Anthropology
Critical historical studies
Historical Methods: A Journal of Quantitative and Interdisciplinary History
Memory Studies
Journal of the History of Sexuality
History Workshop Journal

Mathematics

Publications Mathématiques
Journal of the American Mathematical Society
Inventiones Mathematicae
Acta Mathematica
Duke Mathematical Journal
Communications on Pure and Applied Mathematics  
Journal of the European Mathematical Society  
Memoirs of the American Mathematical Society  
Archive for Rational Mechanics and Analysis  
American Journal of Mathematics  
Mathematical Programming  
Biometrika  
Proceedings of the London Mathematical Society

Engineering

Nature Materials  
Advanced Materials  
Materials Today  
Materials Science and Engineering: R: Reports  
Nano Letters  
International Materials Reviews  
Science Robotics  
International Journal of Machine Tools and Manufacture  
NPJ 2D Materials and Applications  
International Journal of Engineering Science  
Applied Energy  
Archive for Rational Mechanics and Analysis  
International Journal of Robotics Research

Biochemistry

Nature Methods  
Annual Review of Biochemistry  
Nature Catalysis  
Trends in Biochemical Sciences  
Annual Review of Biophysics  
Journal of the American Chemical Society  
Chem  
PLoS Medicine  
Blood  
Current Protocols in Bioinformatics  
Progress in Lipid Research  
EMBO Reports  
Critical Reviews in Biochemistry and Molecular Biology

References


Hang (Joanna) Zou is a lecturer (Minyuan Chenhui Scholar) in the School of Foreign Languages, East China Normal University where she gained her PhD under the supervision of Prof. YANG Yanning. She was also a joint Ph.D. student at the University of East Anglia under the supervision of Prof. Ken Hyland. Her recent publications have appeared in Journal of English for Academic Purposes, Discourse Studies, Journal of Pragmatics, Applied Linguistics etc. Email: hzou@english.ecnu.edu.cn


Ken Hyland is Visiting Professor of Applied Linguistics at the University of East Anglia. He was previously a professor at UCL, the University of Hong Kong and the UEA. He is best known for his research into writing and academic discourse, having published over 260 articles and 28 books on these topics with 68,000 citations on Google Scholar. A collection of his work was published by Bloomsbury in 2018. Email: K.Hyland@uea.ac.uk

Hang (Joanna) Zou is a lecturer (Minyuan Chenhui Scholar) in the School of Foreign Languages, East China Normal University where she gained her PhD under the supervision of Prof. YANG Yanning. She was also a joint Ph.D. student at the University of East Anglia under the supervision of Prof. Ken Hyland. Her recent publications have appeared in Journal of English for Academic Purposes, Discourse Studies, Journal of Pragmatics, Applied Linguistics etc. Email: hzou@english.ecnu.edu.cn