Titles of Scientific Letters in Astrophysics (2000-2015): A Diachronic Study of Type Distribution and the Relationship between Title Length and Collaboration Issues

David Israel Méndez¹, María Ángeles Alcaraz²

¹Department of Physics, System Engineering and Signal Theory, Polytechnic University College, University of Alicante, SPAIN.
²Department of English Studies, Faculty of Arts, University of Alicante, SPAIN.

ABSTRACT
This paper examines how some linguistic and extra-linguistic features of scientific letters published in well-reputed journals in the field of astrophysics (types and length of titles, as well as the relationship between title length and authorship patterns and collaboration practices) have changed over time. Our main results may be summarized as follows: 1) simple and nominal titles significantly outweigh compound, question and verbal titles, although the latter are on the rise; 2) the colon is the most frequently used punctuation mark; 3) the frequency of appearance of colons, full stops and commas increases over time; 4) there is a steady upward trend in longer titles, number of authors and countries; 5) Although over time authors contribute fewer words and countries involved in the research provide more words to the writing of scientific letter titles, authorship variations seem to be more relevant in relationship to the evolution of title length, i.e. title length is more author-driven than country-driven. A final diachronic cross-journal analysis confirms the co-existence of two different collaboration scenarios as already disclosed in a previous synchronic study on the same topic.

Keywords: Titles, Scientific Letters, Astrophysics, Diachronic, Collaboration.

INTRODUCTION
Today, it has been widely recognized that titles play a capital role, not only to provide keywords and index terms in electronic databases so as to trace any type of document,[1-3] but also to give a summary of their content in a limited number of words[4-6] and help prospective readers decide (or not) to go on reading the manuscript that follows. The growing acknowledgement of the importance of titles in scientific research as the reader’s first encounter with a document, whether it is a research article, a thesis, a conference paper, a review paper, etc., has thus provoked that the issue has been the object of a study that has dealt with their titles is the comparative (SL versus research paper titles) and synchronic study[11] carried out in astrophysics, a field where SLs are very important as they are one of the media used to publish “spectacular developments in astronomy”.[12]

SLs allow researchers to rapidly publish (4-6 weeks) short descriptions (4-5 pages) of important current research findings that are expected to have a significant impact on the development of research in all fields of astronomy. Like research papers, SLs are peer-reviewed and must meet the same high standard of quality with the addition of timeliness and brevity, “although they may be more speculative and less rigorous than the former” (see the scope of Astrophysical Journal Letters).

The importance given to this genre to further the progress of science was reinforced by the fact that two of the most pres-
tigious astrophysical journals such as *The Astrophysical Journal* and *Monthly Notices of the Royal Astronomical Society* decided to launch online separate issues publishing SLs exclusively.

Since SL titles have been approached on only one occasion, it is our intention here to expand our knowledge on them by presenting the results of a diachronic analysis (2000-2015) of a corpus of SL titles drawn from three well reputed astrophysics journals (see ‘Corpus’ below). More precisely, our main aim in the present study is to find out whether there have been any changes along time in relation to the following three questions: 1) calculate the frequency of occurrence of the different types of SL titles found in our sample and examine the punctuation marks that differentiate them; 2) see whether there exists any relationship between title length and authorship patterns and collaboration practices, both from a global and cross-journal standpoint; 3) try to provide possible explanations for the differences observed, if any.

**CORPUS AND METHODOLOGY**

We compiled our corpus of SL titles from large circulation, authoritative and prestigious astrophysical journals such as *The Astrophysical Journal Letters* (ApJLs) and *Monthly Notices of the Royal Astronomical Society Letters* (MNRASLs). Since MNRASLs was launched in the year 2005 and our diachronic analysis covered the period 2000-2015, we also had to draw SL titles published in the year 2000 in Monthly Notices of the Royal Astronomical Society (MNRAS), although for the sake of easier reading we will always use the abbreviation “MNRASLs” when referring to any of both journals. We also selected SL titles from a third journal, Astronomy and Astrophysics (A&A), which has no separate section for SLs and publishes them together with research papers, etc. There is a fourth well-reputed astrophysical journal, *The Astronomical Journal* (AJ), but it does not publish SLs and this is the reason why we could not include it in our sample.

In order to have a more diversified corpus, we randomly collected our titles from four different periods comprising 120 titles each: Block A (year 2000), Block B (year 2005), Block C (year 2010) and Block D (year 2015). The 120 titles per block comprise 40 titles per journal and block, i.e. 160 titles per journal and a grand total of 480 titles.

Then we differentiated simple titles from compound titles. Simple titles consist of a general heading that pack the information without any sub-division, whereas compound titles or “colonic titles”[11] split the title focus into two parts: the topic of the study embedded in the general heading and a more specific theme encoded in the second part usually separated from the first one by a punctuation mark. A compound title may even be written on two different lines.

We also established a second title type distinction, which is non-excluding with the previous one, within nominal and verbal titles. A nominal title is a more or less expanded nominal phrase that gives a straightforward presentation of the object of the study. On the contrary, a verbal title contains an active verb with a full sentence that usually states the findings or the conclusion of the research being reported, very much along the lines of newspaper headlines. Nominal and verbal constructions may also be phrased as questions in an attempt to arouse readers’ curiosity and/or let them “find an answer to the question raised in the title”.[9] Here below are some examples of the different types of titles commented here-above:

(1)Simple nominal title

Near-Infrared Imaging Polarimetry of the GG Tauri Circumbinary Ring (ApJLs)

(2) Compound nominal title

Linear spectropolarimetry: A New Diagnostic Tool for the Classification and Characterization of Asteroids (MNRASLs)

(3) Compound nominal question title

G337.2+0.1: A new X-ray supernova remnant? (A&A SLs)

(4) Simple verbal title

The x-ray source at the core of NGC 300 is designated NGC 300 X-1 (MNRASLs)

(5) Simple verbal question title

Is Galactic Structure Compatible with Microlensing Data? (ApJLs)

(6) Compound verbal question title

The early stage of a cosmic collision? XMM-Newton unveils two obscured AGN in the galaxy pair ESO509-IG066 (A&A SLs)

In order to measure the evolution of the length of the SL titles in relation to the number of authors and countries mentioned in the bylines of a single SL and in the whole corpus, we established the following numerical indicators:

1) Title length (TL)

2) Number of authors per title (TA)

3) Number of countries per title (TC)

4) Number of words per author (title length–author, TLA),

5) Number of words per country (title length–country, TLC),

6) Number of words per author and country (title length–author–country, TLAC).

The procedures used to estimate all these variables are the same as those followed in.[11]
Finally, so as to assess whether the observed differences were statistically significant or not, we analysed our results by means of the Student’s test. The alpha value was set at 0.05.

RESULTS AND DISCUSSION

Global diachronic analysis (Table 1, Table 2 and Table 3)

Table 1 displays the distribution over time of the different types of titles found in our study.

As can be seen in Table 1, simple and nominal titles outnumber by far compound and verbal titles. This result confirms the findings of previous research conducted on titles in different types of documents in a variety of disciplines that found that these types of titles are a predominant characteristic of today’s scholarly publication.[]

The predominance of simple titles is due to the main function of titles, which is to provide accurate information on what articles are about in a limited number of words in order to conform to the principles of informativeness and economy.[]

Neither does surprise the prevalence of nominal titles, especially in a discipline such as astrophysics that mainly deals with distant objects that can only be seen through images or spectra, which implies a non-generalization of the results obtained. As for question titles, their low presence in SLs indicates that they are not usually favoured in scientific texts since they are better suited for editorials and/or oral communications.[]

From a temporary point of view, compound titles steadily increase from Block A to Block D. By contrast, verbal and question titles behave in an uneven manner, although their percentages are much higher in Block D than in Block A, especially in the case of verbal titles that grow nearly four times as much in Block D as in Block A. The shift along time towards a higher number of compound, verbal and question titles had already been noticed, not only in astrophysics but also in other scientific fields.[]

The growing number of verbal and question titles may be either interpreted as a desire to generalize the ideas presented or connected to the need for a strong communicative imprint that resembles the journalistic style, thus showing the evident influence of the mass media style already disclosed in the titles of popularized science articles.[25] In other words, titles need to be not only informative, but also appealing. This is the reason why the increasing use of sensationalist language in titles has been found across the full scientific literature since the 1950s.[]

Table 2 shows the distribution over time of the different punctuation marks used in the compound titles registered in our research.

The most frequently used punctuation mark in compound titles is the colon and this is probably due to the diverse roles that the colon plays: it may introduce something that explains or amplifies what has preceded it; it may announce the reader that some important information is coming; it may emphasize previous information; it is the easiest and more direct way of combining different issues in a same title; like verbal and question titles, it is also used to give a journalistic style to titles in order to capture a wider audience.

From a temporary point of view, colons, full stops and commas show an increase from Block A to Block D. It is interesting to highlight that full stops and commas are absent in Block A and Block B and only appear in Block C and Block D. Regarding colons, their increasing use had already been remarked in previous studies on titles in disciplines such as medicine, education, psychology or literature and was attributed to a progressive complexity of scientific research.[14,25,26]

As for the two-lines, the dash and the question marks, they behave in an uneven manner. For example, the two-lines increases from Block A to Block C but decreases in Block D, while the dash, absent in Block A, rises from Block B to Block C and falls again in Block D. The question mark, on its side, increases in Block B, then decreases in Block C and finally disappears in Block D.

The higher frequency of appearance of punctuation marks in Block D has with no doubt to be attributed to the higher presence of compound SL titles in this block (see Table 1). It also has a direct bearing on title length in the sense that titles with

<table>
<thead>
<tr>
<th>Table 1: Evolution of the different types of titles.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title type</strong></td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td>Simple titles</td>
</tr>
<tr>
<td>Compound titles</td>
</tr>
<tr>
<td>Nominal titles</td>
</tr>
<tr>
<td>Verbal titles</td>
</tr>
<tr>
<td>Question titles</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2: Evolution of the different types of punctuation marks.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Punctuation Marks</strong></td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>Colon</td>
</tr>
<tr>
<td>Two-lines</td>
</tr>
<tr>
<td>Dash</td>
</tr>
<tr>
<td>Full stop</td>
</tr>
<tr>
<td>Question mark</td>
</tr>
<tr>
<td>Comma</td>
</tr>
</tbody>
</table>
Table 3: Evolution of the variables analysed in the whole corpus.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Block A</th>
<th>Block B</th>
<th>Block C</th>
<th>Block D</th>
</tr>
</thead>
<tbody>
<tr>
<td>TL</td>
<td>11.47</td>
<td>12.7</td>
<td>12.83</td>
<td>13.05</td>
</tr>
<tr>
<td>TA</td>
<td>4.05</td>
<td>5.28</td>
<td>6.76</td>
<td>5.68</td>
</tr>
<tr>
<td>TC</td>
<td>1.72</td>
<td>2.03</td>
<td>2.13</td>
<td>2.23</td>
</tr>
<tr>
<td>TLA</td>
<td>4.71</td>
<td>4.65</td>
<td>3.85</td>
<td>4.20</td>
</tr>
<tr>
<td>TLC</td>
<td>8.16</td>
<td>8.40</td>
<td>7.84</td>
<td>8.51</td>
</tr>
<tr>
<td>TLAC</td>
<td>3.83</td>
<td>3.66</td>
<td>2.83</td>
<td>3.37</td>
</tr>
</tbody>
</table>

Table 4: Evolution of the variables analysed in A&A SLs.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Block A</th>
<th>Block B</th>
<th>Block C</th>
<th>Block D</th>
</tr>
</thead>
<tbody>
<tr>
<td>TL</td>
<td>12.45</td>
<td>12.98</td>
<td>13.28</td>
<td>12.75</td>
</tr>
<tr>
<td>TA</td>
<td>4.88</td>
<td>7.03</td>
<td>6.78</td>
<td>6.48</td>
</tr>
<tr>
<td>TC</td>
<td>2.00</td>
<td>2.48</td>
<td>2.52</td>
<td>2.75</td>
</tr>
<tr>
<td>TLA</td>
<td>4.81</td>
<td>4.18</td>
<td>3.32</td>
<td>3.09</td>
</tr>
<tr>
<td>TLC</td>
<td>8.08</td>
<td>7.08</td>
<td>6.12</td>
<td>6.81</td>
</tr>
<tr>
<td>TLAC</td>
<td>3.77</td>
<td>3.03</td>
<td>1.97</td>
<td>2.18</td>
</tr>
</tbody>
</table>

punctuation marks are usually longer on average and contain more information than titles without them (see Table 3).

Table 3 illustrates the distribution over time of the remaining variables analysed in this study.

As Table 3 illustrates, TL grows steadily from Block A to Block D and shows statistically significant differences (from Block A to Block B, p=0.020, from Block A to Block C, p=0.016 and from Block A to Block D, p=0.006). This result is in accordance with those of previous studies carried out on titles in a variety of disciplines according to which rising title words along time may reflect more detailed information about the type of research being performed and the increasing complexity of scientific disciplines.\[27\]

TA and TC also rise from Block A to Block D, which is a direct consequence of large research projects involving great collaboration in an increasingly globalized world. The trend toward a growing number of authors in scientific papers had already been associated with title length because when a greater variety of specialties and authors is taking part in research, titles tend to be longer,\[13,17,28-31\] although this may vary among journals.\[18\]

By contrast, the overall decline observed in TLA implies that over time SL authors contribute fewer words to the composition of SL titles, while the overall growth noticed in TLC indicates that the countries involved in the research contribute more words to the composition of SL titles. As for TLAC, it follows the same pattern as TL since in the combined action of TA and TC, TA seems to be the major contributing factor to the evolution of title length. In other words, title length is more author-driven than country-driven. If we take into account the positive correlation between TL and TC found by\[41\] in their SL and research paper titles, this new result would imply that at least in SL title writing authorship variations are more relevant to the detriment of country variations.

Diachronic cross-journal analysis (A&A SLs Table 4, ApJLS Table 5 and MNRASLs Table 6)

Table 4 discloses the distribution over time of the variables analysed in A&A SLs.

Although the variations are not statistically significant, TL increases continuously from Block A to Block C and then decreases in Block D, standing at an intermediate level between Block A and Block B.

TA goes up from Block A to Block B and then starts to go down till Block D. All in all, TA displays an overall increase from Block A to Block D although the difference is not statistically significant. TC rises steadily from Block A to Block D (p=0.017), the difference between Block A and Block C being statistically significant (p=0.022). Also statistically significant is the decrease (p=0.008) between Block A and Block D in TLA. In the case of TLC, the statistically significant fall happens to be between Block A and Block C (p=0.008). Statistically significant decreases are also noticed in TLAC between Block A and Block D (p=0.002) and between Block A and Block C (p=0.005).

If we compare the A&A SL analysis with the global one, TLC is the only value that displays an overall different pattern since it decreases from Block A to Block D in the former and increases in the latter in the same time period.

Table 5 illustrates the distribution over time of the variables analysed in ApJLS.

As displayed in Table 5, TL increases continuously from Block A to Block D (p=0.006), the rise between Block A and Block B also being statistically significant (p=0.025). TA dramatically increases from Block A to Block C and strongly decreases in Block D, although remaining at a higher level than in Block B. The overall increase between Block A and Block D is the only statistically significant difference (p=0.039), whereas the difference between Block A and Block B is almost significant (p=0.051). The differences between Block C and the other ones are not statistically significant because the standard deviation associated to Block C is very high due to the presence of a SL signed by 169 authors. TC goes up steadily from Block A to Block D (p=0.033). Contrary to A&A SLs, TLC, TLA and TLAC show no statistically significant differences among any of the four blocks in ApJLS. When compared to the global analysis, the ApJL one shows an identical overall variation pattern in all the variables.
Table 6 illustrates the distribution over time of the variables analysed in MNRASLs.

As can be appreciated in Table 6, TL shows an irregular pattern from Block A to Block D since it rises in Block B, falls in Block C and rises again in Block D to the same level of Block B. It is worth stressing that only between Block A and Block B and Block D is the TL difference nearly statistically significant (p=0.051).

TA also discloses an erratic behaviour, albeit with some differences when compared to TL: it decreases from Block A to Block B and then grows steadily to Block D. In any case, only the increase between Block B and Block D is statistically significant (p=0.034). Like TL and TA, TC presents an irregular pattern as it increases from Block A to Block C (p=0.045) and decreases from Block C to Block D (with no statistically significant difference). Similarly to ApJLs and contrary to A&A SLs, TLA, TLC and TLAC differences between blocks are never statistically significant in MNRASLs.

If we compare the MNRASL analysis with the global one, we can see that TL, TA, TC and TLC also show an overall increase from Block A to Block D although their variation patterns are different. On the contrary, TLA and TLAC display an overall reverse pattern since the values increase from Block A to Block D in MNRASLs and decrease in the corpus analysed as a whole in the same time span.

Although research in astrophysics is mainly carried out within the so-called “Big Science” scenario,[32] which involves teamwork requiring large personnel, facilities and financial support, it could be speculated that the fall of TA in Block D in A&A SLs and ApJLs in comparison to Block C could be attributed to the financial cut-backs in research because of the 2008 worldwide economic crisis that affected all sectors of society. In ApJLs, TA shows a dramatic increase in Block C when compared to Block B, probably because the SLs published in that time included the results of big projects funded before the crisis. By contrast, the decrease of TA from Block B to Block C noticed in A&A SLs could imply that the budget cuts would affect more rapidly and severely smaller-scaled research projects, albeit participated by many countries, which are characteristic of mainland European research.

Besides, the fall of TA in MNRASLs does not occur in Block D but in Block B and we could hypothesize that this phenomenon would be rather due to some political and social reasons. Indeed, the climate of political and social instability in the United Kingdom following the general election in May 2005, when the Labour Party was the lowest of any majority government in British Electoral History, and the terrorist attacks of July 2005, would slow down research. TA rises again from Block B on, which would suggest that MNRASLs are not directly related to large and/or expensive research, i.e. the economic crisis did not affect it in the same way as the investigation done within the scope of A&A SLs and ApJLs.

All in all, these results once again support the co-existence of two different collaboration scenarios, as already described by[10] in their synchronic study on SL titles. In this sense, A&A SLs and ApJLs fall within the scope of a “higher collaboration scenario” (HCS) with a high number of authors and countries involved in research, whereas MNRASLs lie within that of a “lower collaboration scenario” (LCS), where a low number of authors and countries work on the same project. The idea of a LCS characteristic of MNRASLs is also reinforced by the fact that TC shows the smallest increase from Block A to Block D when compared to the other two journals.

In ApJLs and MNRASLs, TLC follows the general pattern observed in the whole sample, i.e. it increases from Block A to Block D. By contrast, TLC decreases in the same time span in A&A SLs, the journal with the highest values for TC. In this sense, it should be mentioned that title length is and will always be finite because of title characteristics.

As for TLAC, the overall pattern, as well as those of A&A SLs and ApJLs, shows that the increasing number of authors and countries contribute fewer words to SL titles, whereas MNRASLs behave in an opposite manner. If we take into account that TLA follows the same reverse pattern in MNRASLs in relation to the global pattern and to the A&A SL and ApJL ones, we see once again that authors are the major contributing factor to SL title length.
CONCLUSION

In this paper, we have conducted a diachronic study of SL titles published in three of the most prestigious journals in astrophysics. The material presented here supports the following general conclusions:

1) From a global standpoint, simple and nominal titles outnumber by far compound and verbal titles.

2) There is a shift along time towards a higher number of compound, verbal and question titles.

3) Globally speaking, the most frequently used punctuation mark in compound titles is the colon.

4) From a temporary point of view, colons, full stops and commas show an increase over time, whereas the two-lines and dash marks behave in an uneven manner.

5) There is a trend for longer titles, with a higher number of authors and countries over time.

6) From a diachronic standpoint, authors contribute fewer words to the composition of SL titles, whereas countries involved in the research provide more words.

7) In the combined action of authors and countries in SL title composition, authorship variations are more relevant along time in comparison with country variations.

8) Our diachronic cross-journal analysis confirms the coexistence of two different collaboration scenarios, as already highlighted in a previous synchronic study on SL titles. [11]

9) More specifically, A&A SLs and ApJLs fall within the scope of a HCS with a high number of authors and countries involved in research, whereas MNRASLs lie within that of a LCS with a lower number of authors and countries working on the same projects.

10) The different social, economic and geographical contexts characteristic of each of the three journals may account for the variation patterns observed in the variables analysed in this diachronic study.

To better identify the general trend along time in SL titles in astrophysics and their relationship with authorship patterns and collaboration practices, it would be interesting to complete this study with further analyses of longer time spans. This could be an avenue of research in its own right. Moving to other issues, the proposed methodology could also be applied to other scientific fields different from astrophysics with the purpose of finding any possible differences and/or similarities.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ABBREVIATIONS


REFERENCES

24. Jasienski M. Garfield’s demon and “surprising” or “unexpected” results in sci-


