Titles of Scientific Letters and Research Papers In Astrophysics: A Comparative Study of Some Linguistic Aspects and Their Relationship with Collaboration Issues

David I. Méndez[1]*, M. Ángeles Alcaraz[2]

[1]*Department of Physics, System Engineering and Signal Theory, Polytechnic University College, University of Alicante, Spain*

[2]*Department of English Studies, Faculty of Arts, University of Alicante, Spain*

Corresponding Author: David I. Méndez , E-mail: david.mendez@ua.es

ABSTRACT

In this study we compare the titles of scientific letters and those of research papers published in the field of astrophysics in order to identify the possible differences and/or similarities between both genres in terms of several linguistic and extra-linguistic variables (length, lexical density, number of prepositions, number of compound groups, number of authors and number of countries mentioned in the paper bylines). We also carry out a cross-genre and cross-journal analysis of the referred six variables. Our main findings may be summarized as follows: (1) When compared to research paper titles, scientific letter titles are usually shorter, they have a lower lexical density, they include a higher number of prepositions per number of words and a lower number of compound groups per number of words, although they have more up to 4-word compound groups, i.e. the simplest ones. As a consequence, scientific letter titles include less information, which is also less condensed, than research paper titles. (2) The predominance of compound adjectives over compound nouns in the titles of both genres highlights the scientificity of astrophysical discourse. (3) In general terms, our data show a positive correlation between title length and the number of countries mentioned in the bylines for both genres. The positive correlation between title length and number of authors is only met in the case of research papers. In light of these findings, it may be concluded that scientific letters are a clear example of a timeliness and more "immediate" science, whereas research papers are connected to a more timeless and "elaborate" science. It may also be concluded that two different collaboration scenarios are intertwining on the basis of three separate geographic and linguistic publication contexts (Mainland Europe, The United Kingdom and The United States of North America).

INTRODUCTION AND PURPOSE

Apart from providing keywords and index terms in electronic databases so as to trace any type of manuscript (Wang & Bai, 2007; Beel & Gipp, 2009; Jamali & Nikzad, 2011; Fox & Burns, 2015), titles summarizes the contents of the papers in a limited number of words (Haggan, 2004; Hartley, 2008; Gesuato, 2009), thus helping prospective readers decide (or not) to go on reading the documents that follow (Yitzhaki, 1994; Ball, 2009). The importance and pivotal role of titles in literature searching within the academic world has thus provoked that titling practices have been the object of a significant and diversified amount of research. In all these studies, the variables most frequently analyzed have been title length, lexical density, syntactic encoding, frequency of punctuation marks, semantic relations, structural organization, sub-phrasal syntax, content analysis and information sequencing (for a comprehensive review of the vast and heterogeneously rich literature on the subject, see Jaime-Sisó, 2009; Soler, 2011 or Alcaraz & Méndez, 2016).

Of all the documents where titles have been the most analyzed, either from a mono- and a multi-disciplinary approach or from a multi-generic and a multi-linguistic standpoint, the research article, as the main channel not only for the continuous training of scientists, but also for the distribution of new knowledge within the scientific and academic community all over the world (Leventhal, 2011; Publishing Research Consortium, 2011) has been dedicated the maximum attention in the past decades. Titles have also been examined in further scholarly genres such as books, case reports, conference presentations, dissertations and review papers, albeit to a lower degree (see Alcaraz & Méndez, 2016). There is however another type of academic, scientific and technical document, the titles of which have never been studied. We are referring to "scientific letters", also known as "scientific communications". Scientific letters (SLs), which may be categorized as a "primary source"[1] like the research article, are short descriptions (4-5 pages) of important current research findings which allow researchers to rapidly publish (4-6 weeks) the results of their investigation[2]. Like research papers (RPs), 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

Titles of Scientific Letters and Research Papers In Astrophysics:A Comparative Study of Some Linguistic Aspects and Their Relationship with Collaboration Issues

129

the former" (see the scope of *Astrophysical Journal Letters*). A field where SLs are very important is that of astrophysics as they are one of the media used to publish "spectacular developments in astronomy" (Chandrasekhar, 1967, p. 1). The importance given to this genre when advancing scientific knowledge was reinforced by the fact that two of the most prestigious astrophysical journals such as *The Astrophysical Journal* and *Monthly Notices of the Royal Astronomical Society* decided to launch separate issues publishing SLs exclusively.

It may be added that the discipline of astrophysics has seldom been approached from a linguistic and/or rhetorical point of view. A few exceptions would be the study of the passive voice in research articles (Tarone et al., 1998) or the analysis of titles in research and popular science articles (Méndez et al., 2014a; Méndez and Alcaraz, 2015a, 2015b; Alcaraz and Méndez, 2016). Somehow more numerous and varied are the bibliometric papers published by astronomers on the relation of citations to impact factors (Trimble and Zaich, 2006) or on the relationship between the number of articles published yearly in the main astrophysical journals and the number of authors per paper (Butler Burton, 2007; Henneken, 2012). Other articles also dealing with astrophysics were those dedicated to authorship and collaboration patterns by Abt (2000, 2007a, 2007b, 2007c), Méndez et al. (2014b), Méndez & Alcaraz (2015c, 2016), or Smith (2016).

Since, on the one hand, SLs have not been approached so far and, on the other hand, astrophysics has not been dedicated many linguistic studies, it is our intention to " kill two birds with one stone" by examining the titles of a series of SLs published in astrophysics. More precisely, a first aim in our study is to identify possible similarities and/or differences between SL and RP titles in terms of length, informativeness and linguistic complexity. A second aim in our research is to study the possible relationship between the size of SL titles and the number of authors mentioned in the bylines of both genres in order to see whether our results are in line with those previously obtained by authors such as, for instance, Kuch (1978), White (1991), Yitzhaki (1994) and Hudson (2016), although their studies were conducted in fields other than astrophysics.

Kuch (1978) discovered a positive correlation between the number of authors and the number of words of the titles of papers published in four biological journals and the *Journal of American Society for Information Science*, whereas a fifth biological journal showed no correlation. In his study with six social sciences journals, White (1991) observed that an increase in title length generally led to an increase in the number of authors. Yitzhaki (1994) measured title informativity (the number of substantive words included in the titles) and its possible correlation with the number of authors of papers published in the hard sciences, social sciences and humanities journals. He found that in most scientific journals (excluding mathematics) there was a moderate positive correlation between the number of authors and the number of content words in the titles. By contrast, in the social sciences, the correlation was found to be rather low and relevant to a few titles only. As for the humanities, he found no cor-

relation between the number of authors and the number of content words. In his research on journal articles published in 36 different disciplines submitted for assessment in the UK's four-year Research Evaluation Framework (the REF), Hudson (2016) noticed that title length increased with the number of authors in almost all disciplines.

Moreover, something that is new in our research when compared to previous studies is that we are introducing two further analysis variables: 1) the possible impact of the number of countries on the size of SL and RP titles; and 2) the possible impact of the combined effect of the number of authors and the number of countries on title length in both genres. We are therefore formulating the following questions:

1. How does the number of authors mentioned in the bylines of SLs and RPs affect title length?
2. How does the number of countries mentioned in the bylines of SLs and RPs affect title length?
3. How does the combined effect of the number of authors and countries mentioned in the bylines of SLs and RPs affect title length?

We are also addressing further purely linguistic questions related to cross-genre issues:

4. Is there any relationship between the number of prepositions and compound groups[3] in SL and RP titles?
5. Are there any variations in the size of compound groups in SL and RP titles?
6. Do SL titles convey more or less information than RP ones?

We are also formulating a final question concerning all the afore-mentioned matters:

7. Are there any relevant variations among the different journals studied, and if so, which and why?

To answer all these questions, we carried out our investigation in three phases. First, we counted all the words making up the titles as well as the number of authors and of countries mentioned in the bylines of the SLs and the RPs included in our sample. Second, we registered the number of prepositions and compound groups included in the titles. Third, we carried out a cross-genre and cross-journal analysis of the referred variables in order to find out the differences and similarities between letter and RP titles and among the journals analysed in this research.

CORPUS

We started to compile our SL titles in the year 2000 from highly reputed journals, i.e. with high impact factors[4]. The journals selected are *The Astrophysical Journal Letters* (ApJLs), *Monthly Notices of the Royal Astronomical Society Letters* (MNRASLs), *Monthly Notices of the Royal Astronomical Society* (MNRAS), and *Astronomy and Astrophysics* (A&A). Since MNRASLs was launched in the year 2005, we also had to use MNRAS to complete our corpus, although for the sake of easier reading we will always use the abbreviation "MNRASLs" when referring to any of both journals. As for A&A, it has no separate section for SLs and publishes them together with RPs, etc. There is a fourth well-reputed astrophysical journal, the *Astronomical Journal*, but it does

not publish SLs. This is why we could not include it in our sample. In order to have a more diversified corpus, we randomly collected 40 SL titles per journal from four different periods (years 2000, 2005, 2010, and 2015), i.e. our corpus amounts to 480 SL titles.

With respect to our RP title corpus, we used the same one we dealt with in our previous studies on the matter, but for the journal AJ because it does not publish SLs as was mentioned above. Summing up, the titles analysed are those drawn from *The Astrophysical Journal* (ApJ), *Monthly Notices of the Royal Astronomical Society* (MNRAS), and *Astronomy and Astrophysics* (A&A) (see Méndez, Alcaraz & Salager-Meyer, 2014a, 2014b and Méndez &Alcaraz, 2016 for a detailed account of the three different corpora). Since both samples are not similar in size (480 SL titles and 200 RP titles), we base all our calculations on average values in order to be able to compare the obtained data.

METHODOLOGY

As formerly done in our research on titles in the field of astrophysics, we define the concept of "word" as the unit occurring between two spaces or separated by a hyphen from the following word. We counted capitalized abbreviations according to the number of their semantic components. For instance, the abbreviation "PSPC" (< Position Sensitive Proportional Counters) was counted as four different words, while acronyms and shortenings such as "pulsar" (< pulsating star) and "Oph" (< Ophiuchi) were counted as an only item each. We then manually recorded all the words included in all the titles, as well as the number of authors and of countries mentioned in the SL and RP bylines.

In order to see if the differences between the titles included in our sample are statistically significant, we computed the following ten variables and calculated their mean number with respect to all the analysed titles and bylines:
1) Number of words per title or 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).
7) Number of prepositions per number of words (PTL)
8) Number of compound groups per number of words (CGTL)
9) Number of compound nouns per number of words (CNTL)
10) Number of compound adjectives per number of words (CATL)

With respect to TC (variable 3), a clarification is in order here: whenever a given country is indicated more than once in the bylines of a single SL/RP, we counted them as a unique item; by contrast, whenever the same country is indicated in the bylines of different SLs/RPs, we counted them as different items. Regarding TLA (variable 4), it was studied in terms of the mean number of words per author. The numerical indicator refers to the mean quotient between the number of words and the number of authors per SL/RP title, i.e. it tries to measure the average influence of each au-

thor of each SL/RP on the size of each title. With regards to TLC (variable 5), it was studied in terms of the mean number of words per country. The numerical indicator refers to the mean quotient between the number of words and the number of countries per SL/RP title, i.e. it tries to measure the average influence of each country of each SL/RP on the length of each title. As for TLAC (variable 6), it was studied in terms of the mean number of words per author and country. The numerical indicator refers to the mean quotient between the number of words and the product of the number of authors and the number of countries per SL/RP title, i.e. it tries to measure the combined average influence of each author and country on the size of each title.

It is important to make clear that all the proposed numerical indicators are always sample-affected: TL, TA and TC are singly affected by the sample, TLA and TLC are doubly affected by it and TLAC is three times affected. While TLA and TLC consider only two single variables grouped together (title length and the number of authors or countries per SL/RP), TLAC includes three single variables grouped together (title length, the number of authors and the number of countries per SL/RP).

With reference to variable 7 (PTL), we registered all the prepositions included in the SL/RP titles. In relation to compound groups (CGTL, variable 8), we divided all them into compound nouns (CNTL, variable 9), i.e. groups of words built with at least two nouns, and compound adjectives (CATL, variable 10), i.e. groups of words built with at least two nouns and one or more adjectives.

Moreover, and in order to determine whether the paired two-sample differences observed in the above mentioned ten variables were statistically significant or not, we analysed our data by means of the Student's *t*-test. The alpha value was set at 0.05.

Regarding all the found compound groups, we divided them in five ranges which read as follows: 1) up to three words; 2) four words; 3) five words; 4) six words and 5) more than six words. In the case of compound nouns the established ranges were as follows: 1) two words; 2) three words; 4) four words and 5) more than four words. As for compound adjectives, the chosen ranges were the following ones: 1) three words; 2) four words; 3) five words; 4) six words and 5) more than six words.

Finally, and in order to determine the amount of information conveyed by the titles, i.e. their lexical density, we divided the words found in our corpus into lexical or content words (nouns, adjectives, adverbs, past and present participles, mathematic and chemical symbols, conjugated and infinitive verbs) and grammatical or function words (auxiliary verbs, determiners—definite and indefinite articles, possessives— conjunctions, prepositions, pronouns, and wh-words)[5].

RESULTS

SL Titles

TL, TA, TC, TLA, TLC and TLAC variables

Table 1 displays the mean numbers of words in SL titles, as well as the mean number of authors and countries men-

Titles of Scientific Letters and Research Papers In Astrophysics:A Comparative Study of Some Linguistic Aspects and Their Relationship with Collaboration Issues

131

tioned in the bylines of the three journals analysed in our study.

As can be seen, the highest TL is found in A&A letters, although the differences with the other samples are not statistically significant. Here-below are some of the longest and shortest SL titles found in each journal:

(1) "Towards DIB mapping in galaxies beyond 100 Mpc A radial profile of the λ5780.5 diffuse interstellar band in AM1353-272 B" = 28 words (A&A letters, 2015)
(2) Is Sedna another Triton? = 4 words (A&A letters, 2005)
(3) A periodically varying luminous quasar at z = 2 from the pan-starrs1 medium deep survey: A candidate supermassive black hole binary in the gravitational wave-driven regime = 34 words (ApJLs, 2015)
(4) Delayed Recombination = 2 words (ApJLs, 2000)
(5) Do star formation rates of galaxy clusters depend on mass? Blue/late-type fractions and total star formation rates of 115 galaxy clusters as a function of cluster viral mass = 30 words (MNRASLs, 2005)
(6) Weak microlensing = 2 words (MNRASLs, 2010)

Regarding TA, the lowest value is observed in MNRASLs, which presents statistically significant differences with A&A letters (p=0.00017) and APJLs (p=0.003). Moreover, it is worth pointing out that TA standard deviations are very high in APJLs and A&A letters due to the fact that an APJL is signed by 169 authors and an A&A letter is signed by 99 authors, whereas the highest number of authors in a MNRASL only amounts to 14. With respect to TC, the highest value is found in A&A letters, with statistically significant differences with APJLs (p=0.0024) and MNRASLs (p=0.000025), the difference also being statistically significant between APJLs and MNRASLs (p=0.015). If we consider the absence of large TL variations in the three journals, the TLA reverse behaviour as regards TA should not surprise. In this sense, statistically significant differences are found between A&A letters and MNRASLs (p=0.004) and between APJLs and MNRASLs (p=0.0013). The same phe-

nomenon is given in TLC in comparison with TC. Similarly to the TLA case, statistically significant differences in TLC are found between A&A letters and MNRASLs (p=0.00022) and between APJLs and MNRASLs (p=0.0055). Finally, the combined effect of TL, TA and TC is responsible for the TLAC behaviour. As a result, the highest value is found in MNRASLs, showing besides statistically significant differences with respect to A&A letters (p=0.0001) and APJLs (p=0.039).

TL, PTL, CGTL, CNTL and CATL variables

Table 2 illustrates the different linguistic variables analysed in the whole letter corpus.

The only statistically significant difference is found in CATL between A&A letters and ApJLs (p=0.014). However, it is interesting to note that ApJLs is the journal with the highest CGTL value, as well as the lowest PTL and, surprisingly, CNTL values. In addition, it is worth pointing out that the number of compound adjectives is much higher than the number of compound nouns in the three journals, albeit the differences between CATL and CNTL are only statistically significant in the case of MNRASLs (p=0.014) and APJLs (p=0.0000015). If we focus on the average values in the three journals, a statistically significant difference is also found (0.000003).

Compound groups

Figure 1 displays the size distribution of compound groups in SL titles per journal.

As figure 1 clearly illustrates, groups built with less than four words top the frequency scale of compound groups, the highest percentage corresponding to MNRASL titles and the lowest one to A&A letter titles. 4-word groups rank second, the highest percentage corresponding to ApJL titles and the lowest one to MNRASL titles. 5-word groups rank third, the

Table 1. Mean numbers of words, authors and countries in SL titles and bylines per journal

Journal	A&A letters	ApJLs	MNRASLs	TOTAL
Mean number of words per title (TL)	12.86	12.34	12.33	12.51
Mean number of authors per title (TA)	6.29	6.62	3.42	5.44
Mean number of countries per title (TC)	2.44	1.95	1.68	2.02
Mean number of words per author (TLA)	3.85	3.97	5.24	4.35
Mean number words per country (TLC)	7.02	8.64	9.02	8.22
Mean number of words per author and country (TLAC)	2.74	3.18	4.36	3.43

Table 2. Linguistic variables in SL titles per journal

Journal	A&A letters	ApJLs	MNRASLs	TOTAL
Mean number of words per title (TL)	12.86	12.34	12.33	12.51
Mean number of prepositions/TL (PTL)	0.142	0.136	0.142	0.140
Mean number of compound groups/TL (CGTL)	0.104	0.111	0.110	0.108
Mean number of compound nouns (CNTL)	0.047	0.037	0.045	0.043
Mean number of compound adjectives (CATL)	0.057	0.074	0.065	0.065

highest percentage corresponding to A&A letter titles and the lowest one to ApJL titles. The 6-word groups come after, with the highest percentage found in ApJL titles and the lowest one in A&A letter titles.

As for the groups built with more than six words, the highest percentage is found in A&A letter titles and the lowest one in ApJL titles. From a global standpoint, the percentage of compound groups built with more than six words is higher than that of compound groups built with six words. This result should come as no surprise if we consider that some compound groups range from seven to even 12 words, especially in the case of compound adjectives including abbreviations within their linguistic components. The same situation is reproduced in the case of A&A letter titles.

Table 3 illustrates the compound group breakdown into nouns and adjectives in SL titles per journal.

As can be seen, the 2-word structure tops the frequency scale of compound nouns, far away from 3-, 4- and more than 4-word ones. The percentages of 2-word nouns are very similar in ApJL and MNRASL titles, while A&A letter titles contain the lowest one. A&A letter titles also include the highest percentage of 3-, 4- and more than 4-word nouns.

As for compound adjectives, ApJL titles contain the highest 3 and 4-word percentages, while the highest 5-word one is found in A&A letter titles, closely followed by MNRASL ones.

Furthermore, ApJL and A&A letter titles contain the highest percentages of 6-word groups, while the highest percentage of more than 6-word groups is clearly found in A&A letter titles. From a global standpoint, it is worth pointing out that the 6-word adjective percentage is lower than the more than 6-word adjective one. This situation occurs in A&A letter and MNRASL titles as well. The following structures, two from each of the three journals, exemplify different types of compound groups (they have been underlined):

(7) Constraints on the gas content of the Fomalhaut debris belt = 2- and 3-word compound nouns (A&A Letters, 2015)

(8) XMM-Newton observations of two transient millisecond x-ray pulsars in quiescence = 5- and 6-word compound adjectives (A&A Letters, 2015)

(9) Ab initio pulsar magnetosphere: three-dimensional particle-in-cell simulations of oblique pulsars = 4-word compound noun + 6-word compound adjective (ApJLS, 2015)
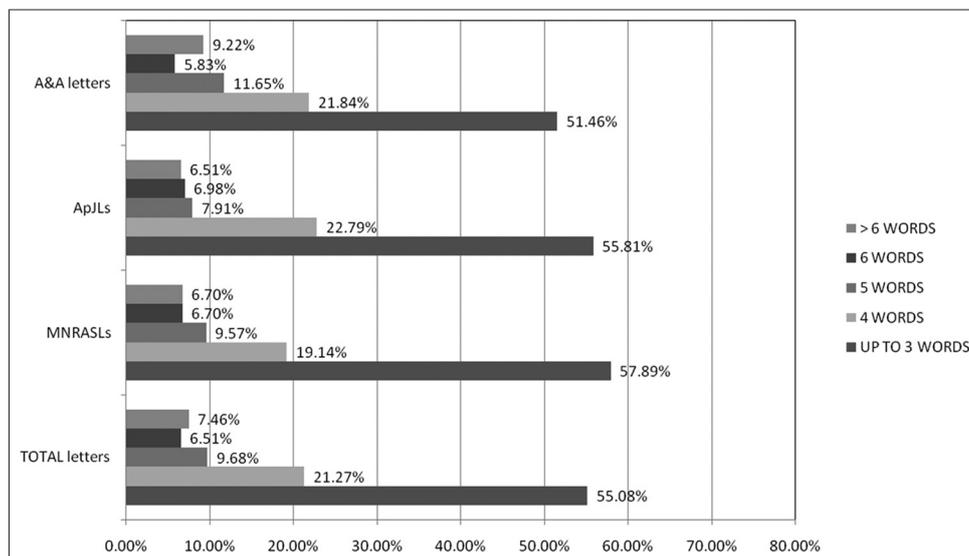

Figure 1. Size distribution of compound groups in SL titles per journal

Table 3: Breakdown of compound groups into nouns and adjectives in SL titles per journal

Journal		A&A letters	ApJLs	MNRASLs	TOTAL
Compound nouns (CNs)	2 WORDS	52.22%	61.97%	62.5%	58.51%
	3 WORDS	28.89%	26.76%	27.5%	27.80%
	4 WORDS	11.11%	7.04%	5%	7.88%
	> 4 WORDS	7.78%	4.23%	5%	5.81%
Compound adjectives (CAs)	3 WORDS	28.45%	39.58%	37.98%	35.73%
	4 WORDS	30.17%	30.55%	27.91%	29.56%
	5 WORDS	15.52%	10.42%	14.73%	13.37%
	6 WORDS	10.34%	10.42%	8.53%	9.77%
	> 6 WORDS	15.52%	9.03%	10.85%	11.57%

Titles of Scientific Letters and Research Papers In Astrophysics:A Comparative Study of Some Linguistic Aspects and Their Relationship with Collaboration Issues

133

(10) Discovery of the hard x-ray features around the hotspots of Cygnus A = 4-word compound adjective (ApJLS, 2000)

(11) Maximum speed of hypervelocity stars ejected from binaries = 2-word compound noun (MNRASLs, 2015)

(12) Nature or nurture of coplanar Tatooines: the aligned circumbinary Kuiper belt analogue around HD 131511 = 5-word compound adjective (MNRASLs, 2015)

Lexical density

Figure 2 displays the percentages of content and function words found in our sample of SL titles.

As figure 2 clearly illustrates, content words outnumber by far function words in all the titles under study. From a cross-journal standpoint, the percentages of content and function words are very similar.

RP Titles

TL, TA, TC, TLA, TLC and TLAC variables

Table 4 displays the mean numbers of words, authors and countries in RP titles and bylines per journal.

According to the obtained data, and like in the case of SLs, A&A RPs include the highest TL value, although the difference with the other samples is not statistically significant (some of the longest and shortest RP titles were already mentioned in Méndez, Alcaraz & Salager-Meyer, 2014b). Regarding TA, it follows the same pattern as TL, namely the highest value is found in A&A RPs. Moreover, the lowest TC value is found in ApJ RPs, which present statistically significant differences with MNRAS (p=0.0022) and A&A RPs (p=0.0052).

With respect to TLA, the highest value is found in ApJ RPs, although there are no statistically significant differences among the three journals. Similarly to TLA, the highest TLC value is shown in ApJ RPs. Nevertheless, and in parallel to the TC case, we may find statistically significant differences with MNRAS RPs (p=0.0021) and with A&A RPs

(p=0.047). As for TLAC, the highest value is found once again in ApJ RPs, but with no statistically significant difference with the other journals.

TL, PTL, CNTL, CATL and CGTL variables

Table 5 illustrates the different linguistic variables analysed in the whole RP title corpus.

The only statistically significant differences are found in PTL and CGTL between A&A and ApJ RP titles (p=0.038 and p=0.010, respectively). It is also worth stressing that MNRAS RP titles show nearly statistically significant differences in CNTL with respect to A&A RPs (p=0.054) and in CATL with respect to ApJ RP titles (p=0.053). In addition, PTL and CGTL values are inversely correlated in the three journals.

Besides, and similarly to SLs, the number of compound adjectives is much higher than the number of compound nouns, the differences between CATL and CNTL being statistically significant in MNRAS RPs (p=0.00004), ApJ RPs (p=0.029) and A&A RPs (0.026). The difference is also statistically significant (p=0.000004) in global terms.

Compound groups

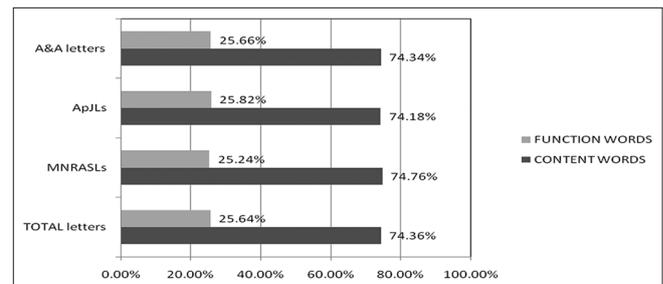Figure 3 displays the size distribution of compound groups in RP titles per journal.

Figure 2. Lexical density in SL titles per journal

Table 4. Mean numbers of words, authors and countries in RP titles and bylines per journal

Journal	A&A RPs	ApJ RPs	MNRAS RPs	TOTAL RPs
Mean number of words per title (TL)	13.49	12.81	13.01	13.10
Mean number of authors per title (TA)	4.99	4.39	4.6	4.65
Mean number of countries per title (TC)	2.24	1.65	2.27	2.05
Mean number of words per author (TLA)	4.40	4.51	4.42	4.44
Mean number words per country(TLC)	8.19	9.67	7.26	8.37
Mean number of words per author and country (TLAC)	3.19	3.73	2.98	3.30

Table 5. Linguistic variables in RP titles per journal

Journal	A&A RPs	ApJ RPs	MNRAS RPs	TOTAL RPs
Mean number of words per title (TL)	13.49	12.81	13.01	13.10
Mean number of prepositions/TL (PTL)	0.129	0.147	0.132	0.136
Mean number of compound groups/TL (CGTL)	0.134	0.110	0.122	0.122
Mean number of compound nouns (CNTL)	0.055	0.042	0.038	0.045
Mean number of compound adjectives (CATL)	0.079	0.067	0.084	0.077

As can be seen, the groups built with less than four words top the frequency scale of compound groups, the highest percentage corresponding to A&A RP titles and the lowest one to MNRAS RP titles, i.e. a reverse situation to the one observed in the case of SL titles. 4-word and 5-word groups rank second and third in ApJ and MNRAS RP titles. With respect to A&A RP titles, the 5-word groups rank second, where the 4- and more than 6-word groups show the lowest percentages. A point worth noting is that the 5- and 6-word group percentages in ApJ RP titles are considerably low. Globally speaking, and similarly to the case of SL titles, the percentage of compound groups built with more than six words is higher than that of compound groups built with six words. The same situation is even more evident in the case of ApJ RP titles, where the difference between both percentages amounts to 7%.

Table 6 illustrates the compound group breakdown into nouns and adjectives in RP titles per journal.

In relation to compound nouns, the 2-word structure tops the frequency scale, far away from the 3-, 4- and more than 4-word ones. The percentages of 2-word nouns are lower in ApJ and MNRAS RP titles than in A&A RP titles. By contrast, A&A RP titles contain the lowest percentages of 3- and 4-word nouns. With respect to the more than 4-word nouns, their percentage is lower than that of the 4-word one in MNRAS RP titles, but it is higher in A&A and ApJ RP titles, which is a characteristic feature of the RP title sample taken as a whole.

As for compound adjectives, A&A RP titles contain the highest 3-, 5- and 6-word percentages, while the highest 4-word and more than 6-word percentages are found in ApJ RP titles. From a global standpoint, it is worth pointing out that the 6-word adjective percentage is lower than the more than 6-word one.

Lexical density

Figure 4 displays the percentages of content and function words found in our sample of RP titles.

As figure 4 clearly illustrates, and similarly to the case of SLs, content words outnumber by far function words in all the RP titles under study. From a cross-journal standpoint, the highest percentage of content words, and consequently the lowest one of function words, is found in A&A RP titles. In any case, the percentage differences between the different journals never amount to more than 2.5%.
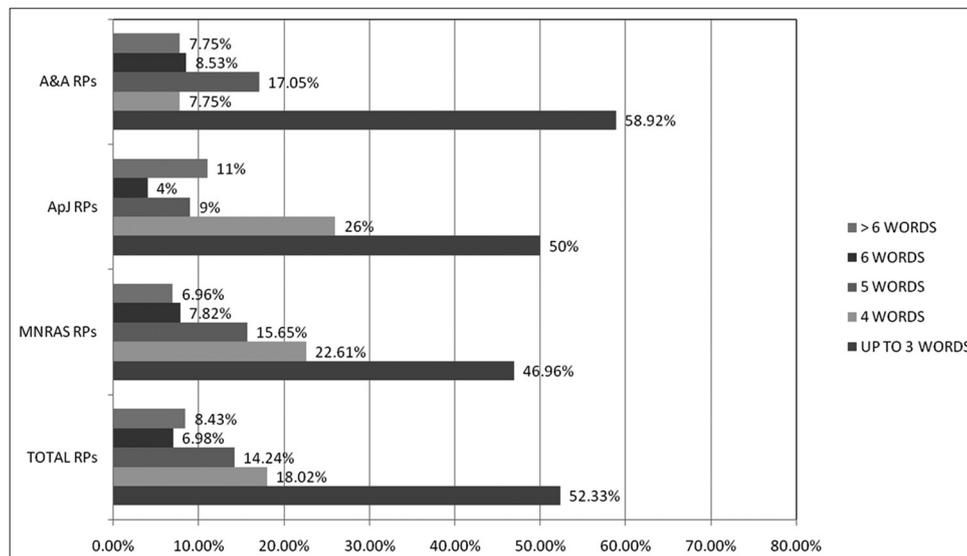

Figure 3. Size distribution of compound groups in RP titles per journal

Table 6. Breakdown of compound groups into nouns and adjectives in RP titles per journal

Journal		A&A RPs	ApJ RPs	MNRAS RPs	TOTAL RPs
Compound nouns (CNs)	2 WORDS	64.58%	48.48%	45.95%	54.24%
	3 WORDS	25.01%	30.30%	32.43%	28.81%
	4 WORDS	2.08%	9.10%	13.51%	7.63%
	> 4 WORDS	8.33%	12.12%	8.11%	9.32%
Compound adjectives (CAs)	3 WORDS	40.74%	35.82%	32.05%	36.28%
	4 WORDS	11.11%	34.32%	26.92%	23.45%
	5 WORDS	23.46%	8.96%	20.51%	18.14%
	6 WORDS	13.58%	4.48%	10.26%	9.74%
	> 6 WORDS	11.11%	16.42%	10.26%	12.39%

est one in SL titles, contrary to the global as well as ApJ and MNRAS behaviours. Moreover, A&A presents the lowest 3- and 4-word percentages in RP titles and the highest ones in SL titles, showing once again a reverse pattern with respect to ApJ and MNRAS.

As for compound adjectives, A&A presents the highest 3-word percentage in RP titles and the lowest one in SL titles, contrary to the patterns showed by ApJ and MNRAS. In the case of the more than 6-word range, A&A behaves differently from ApJ and the overall results since it shows a higher percentage in SL titles than in RP ones.

Lexical density (Figures 2 and 4)

In general, as well as in individual terms, lexical density is slightly higher in RP titles than in SL titles, the largest variation taking place in A&A.

DISCUSSION AND CONCLUSIONS

In order to obtain more conclusive statements, we propose to divide our discussion in two different steps: an individual cross-genre analysis and a global cross-genre analysis.

Individual Cross-Genre Analysis

A&A letters and RPs

On the one hand, A&A letter titles are shorter than A&A RP titles. On the other hand, PTL is higher in SLs than in RPs, whereas CGTL and lexical density are lower in SLs than in RPs. As a consequence, the information supplied by SL titles is smaller and less compressed when compared to RP titles, where the higher amount of information is presented in a more processed and elaborate way. This finding is in agreement with the fact that RP titles include a higher percentage of more than 4-word compound groups (see figures 1 and 3).

Moreover, and although research in astrophysics is mainly carried out within the so-called "Big Science" scenario (de Solla Price, 1963), which involves team work requiring large personnel, facilities and financial support, the obtained data indicate the co-existence of two slightly different sub-scenarios in A&A: one referred to SLs, where many authors and countries participate in the investigation (a higher collaboration scenario, HCS), and another one related to RPs, where a lower number of authors and countries work on the same project (lower collaboration scenario, LCS). The union of all these factors (shorter and less elaborate titles, less information, bigger teams of experts) would suggest that only some of the authors engaged in the investigation would take part in the composition of SL titles. Conversely, longest titles, a higher degree of information and condensation in RP titles, as well as smaller investigation teams, would imply that more researchers would leave their personal imprint on the writing of RP titles.

ApJLs and ApJ RPs

In ApJ, TL, TA, TC and lexical density yield similar results to those disclosed in A&A, although TL and lexical density

show minor discrepancies between letters and RPs. What is really relevant is that PTL is lower and CGTL is higher in SL titles than in RP titles (although this is not the case of CNs taken separately). This result might indicate that, contrary to A&A, there are no extreme differences between RP and SL titles with respect to the condensation of the information. This idea is also reinforced by the fact ApJ shows the smaller title length difference between SLs and RPs. Moreover, the percentages of up to 4-word compound groups are very similar in SL and RP titles (slightly higher in the case of SL titles). Nevertheless, the higher values of TA and TC in SLs with respect to RPs suggest again a HCS for SLs and a LCS for RPs.

MNRASLs and MNRAS RPs

Like in A&A and ApJ, MNRASL titles are shorter than MNRAS RP titles. With respect to lexical density, PTL and CGTL, SL and RP titles follow a similar trend to A&A RP titles, i.e. they include more information, which is also more condensed.

However, and unlike A&A letters and ApJLs, the results obtained in MNRASLs clearly state that, in average, authors and countries, separately and together, contribute more words to the composition of titles. This finding comes as no surprise since MNRAS, contrary to the other journals, present higher values in the number of authors and countries in the case of RPs.

These data would indicate that MNRASLs portray a completely reverse scenario with respect to A&A letters and ApJLs since they seem to be the outcome of a LCS, where authors and countries contribute the most in terms of number of words to the creation of titles, whereas MNRAS RPs are related to a HCS, where multiple authors from multiple countries are interacting.

Cross-Journal Analysis

A&A letters and RPs contain the longest titles. A&A letters also show the highest TC and the lowest TLA, TLC, TLAC and CGTL, whereas A&A RPs show the lowest PTL, as well as the highest TL, CGTL and lexical density. A&A RP titles also include the highest percentage of more than 4-word compound groups. In this sense, it can be assured that CGTL is always inversely proportionate to PTL, both in SL and RP titles, not only in A&A but also in all the journals.

Furthermore, in all the titles, the highest percentages of CN and CA always correspond to the shortest compound groups (2- and 3-words, respectively), except in the case of CA in A&A letter titles, where the percentage of the 4-word compound group is slightly higher than the percentage of the 3-word compound group. It is interesting to point out that the highest CN and CA percentages in RP titles correspond to A&A, while the situation is completely opposite in A&A letter titles. In other words, A&A authors tend to use fewer, albeit more complex, compound groups in SL titles and more, albeit less complex, compound groups in RP titles. In general terms, it should be stressed that A&A is the journal with the largest TL and the highest percentage of more than

Titles of Scientific Letters and Research Papers In Astrophysics:A Comparative Study of Some Linguistic Aspects and Their Relationship with Collaboration Issues

137

4-word compound groups, i.e. the most complex, and most difficult to decode, titles are found in this journal. It could be speculated that A&A authors, who are usually non-native English speakers, try to emulate native English authors by being "more papist than the pope" with their creation of even longer and more elaborate titles. This discrepancy could be attributed to their lower ability to communicate more effectively.

Another interesting point to remark is the negative correlation found between TA and CGTL in A&A letters and RPs. By contrast, in MNRASLs, MNRAS, ApJLs and ApJ, whose vast majority of authors are usually native English speakers[6], the correlation between both variables is positive, i.e. the more numerous the authors in the bylines of a paper, the more numerous the compound groups per number of words, hence the presence of fewer prepositions per number of words, in its title. This result would imply that A&A is the journal in which the lowest percentage of authors would take part in the writing of SL and RP titles.

Regarding ApJLs, they show the highest TA value as well as the lowest lexical density. ApJ RPs show the lowest TC and the highest TLC and PTL values, i.e. the minor degree of international collaboration in the whole sample. This finding would be in agreement with the previous results by Méndez & Alcaraz (2016) on the different types of collaboration practices in RPs. Furthermore, ApJL and ApJ RP titles include the lowest percentages of more than 4-word compound groups. In this sense, and according to the information provided in their titles and their degree of condensation, the titles recorded in both journals seem to follow an opposite trend to A&A since they contain less information, which is in addition less condensed. The higher impact factors of the two journals[7] and a more widespread audience can be clearly responsible for this fact.

As for MNRASLs and MNRAS, it is interesting to remark that the former is the journal with the lowest TL and TA and the highest TLA and TLAC in the whole sample. Besides, if we consider exclusively the SL sample, MNRASLs is also the journal with the lowest TC and the highest TLC. MNRAS RPs also show the second lowest TLAC and the lowest lexical density, whereas MNRASLs show the highest lexical density. These findings could only be understood within a noteworthy conceptual difference between RPs and SLs, where the SL scenario is a clear reminiscence of the primeval SL situation which is closely related to a LCS: to publish spectacular results in a rapid and immediate way. On the contrary, the RP scenario is clearly related to an intense international collaboration situation, i.e. an HCS.

Indeed, these findings agree with the fact that MNRASL titles have a high up to 4-word compound group percentage (the most basic one), and the highest PTL value in the whole letter sample. Conversely, MNRAS RP titles show a much lower up to 4-word compound group percentage and a low PTL (very close to the lowest one, which is found in A&A RPs). With respect to the basic CN groups, the highest and lowest percentages are found in MNRASL and MNRAS RP titles, respectively. In the case of the basic CA groups, MNRAS RP titles have the lowest percentage, whereas ApJL titles have the highest one. In the light of the striking linguistic discrepancies between RP and SL titles, it is clear that MNRAS authors follow distinct patterns when creating either an RP or a letter title. Even though, the different lexical density values between MNRAS RP and letter titles are the smallest in the whole sample.

Global Cross-Genre Analysis

From a global point of view, SL titles are shorter than RP titles. In addition, SLs are characterized by a higher presence of authors and RPs by a slightly higher presence of countries, which implies that taken separately authors and countries contribute fewer words to SL titles, i.e. the TLA and TLC values are always higher in RP titles. However, the combined contribution of authors and countries (TLAC) is slightly higher in SL titles. Namely, if we choose two titles with the same number of words and of authors, one belonging to a SL and the other belonging to an RP, there is a better chance of finding a higher number of countries, and so a major degree of international collaboration, in the RP. A parallel reasoning applies if we fix the number of words and countries. Furthermore, it is interesting to remark that A&A and ApJ follow the general trend previously described in terms of TLC and TLA, while MNRAS show a reverse pattern. As for TLAC, MNRAS is the only journal that tags along the general trend. In the light of these results, it would be extremely hazardous to establish watertight compartments between HCSs and LCSs.

Anyway, we find that the positive correlation between TL and TA previously established by Kuch (1978), White (1991), Yitzhaki (1994) and Hudson (2016) in some of the samples they analysed is only met in the case of RP titles, but not in SL titles. As for TL and TC, the positive correlation between them is found both in the RP and the SL samples, i.e. the higher number of countries mentioned in the bylines of the referred documents, the longer their titles, either from a general or cross-journal standpoint. However, since previous literature did not make any distinction between genres and never took into account the number of countries, it is impossible to perform a direct comparison between our results and the afore-mentioned ones.

From a purely linguistic standpoint, our analysis has shown that SL titles contain less information, which is also less condensed. Both findings come as no surprise if we consider that the main aim of letters is to publish direct and spectacular results in a rapid and immediate way, and this is the reason why their titles always tend to be less complex than RP ones. In other words, SLs are a clear example of a timeliness and more "immediate" science, whereas RPs are connected to a more timeless and "elaborate" science. Moreover, this idea is also reinforced by the fact that the percentage of up to 4-word compound groups, i.e. the easiest to understand, is higher in SL titles (76.35%) than in RP titles (70.35%).

Nevertheless, the prevalence of shorter compound groups over longer ones both in SL and RP titles is, with no doubt, due to the main purpose of titles which is to inform in a clear, accurate and precise way (Day, 1995; Haggan, 2004; Ball, 2009; Hartley, 2008; Gesuato, 2009; Swales & Feak 2012)

in order to conform to the principles of informativeness and economy (Bush-Lauer, 2000). The predominance of longer compound groups would definitely imply a lack of attention, or even a rejection, on the readers' part because of the effort required to fully decode the information supplied. In this sense, it is interesting to point out that the more abundant compound groups are always the shortest ones (2-word compound nouns and 3-word compound adjectives, respectively) in both genres. This result is in agreement with the one found by Entralgo, Salager-Meyer & Luzardo Briceño (2015). Likewise, it is worth stating that, contrary to the case of titles of popular science articles (Méndez & Alcaraz, 2015c), compound adjectives always outnumber compound nouns in the titles of both genres, which is a characteristic feature of scientific discourses. The high lexical density found in both samples also accounts for the scientificity and informativeness needed in astrophysical titles.

In addition, we find quite interesting to remark the positive correlation disclosed between TL and CGTL. This should come as no surprise since, from a statistical point of view, compound groups are evidently more frequent in longest titles. By contrast, we have found a negative correlation between TA and CGTL, both variables being inversely proportionate, i.e. if there are more authors in the bylines of a paper, its title will likely contain fewer compound groups per number of words, hence more prepositions per number words. Generally speaking, the condensation of information seems to be inversely correlated to the number of authors, i.e. the more numerous the authors in the bylines of a paper, the lower the chance for its title to be more complex and condensed. In other words, it seems that when more people are involved in a given RP, its resulting title tends to be more neutral and aseptic, i.e. less linguistically complex, in an attempt to satisfy all the authors, some of whom may be non- English native speakers.

Figure 5 (p.135) summarizes the most important issues discussed previously.

In order to get a more comprehensive panorama of the relationship between titles and collaboration patterns in astrophysics, we think that it would be interesting to complete this study with a further analysis of other genres such as, for example, dissertations, review papers or conference proceedings, to name just a few. Likewise, a diachronic study would provide valuable historical data on title construction processes in astrophysics along time. Finally, 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.

Endnotes

i   Secondary sources would be, for example, review papers.
ii In the ―Instructions for authors‖ section, the journals Astronomy & Astrophysics, Astrophysical Journal Letters and Monthly Notices of the Royal Astronomical Society Letters explicitly inform about the main characteristics of the letters they publish.
iii Compound groups are compressed structures where information is condensed through the juxtaposition of content words without any function word.

iv Although impact factors (IFs) are a somewhat dubious measure for the many problems involved in the matter, i.e. invisibility of research published in new, innovative and specialized journals, underrepresentation of many sub-disciplines in the databases used to calculate the IF, which actually measures influence, not quality, the mandatory three year ―waiting period‖ for all new journals, etc., etc. (Adler and Harzing 2009), they however serve a useful purpose for grouping papers together.
v   Other word class items were not found in our corpus.
vi ApJ is US-based and is currently published on behalf of the American Astronomical Society by the University of Chicago Press. MNRAS is published on behalf of the Royal Astronomical Society by Blackwells Synergy and is often the journal of choice for astronomers from the UK and the Commonwealth. A&A is a European journal published on behalf of Édition Diffusion Presse (EDP) Sciences.
vii A&A IF: 5.185, ApJ IF: 5.909, MNRAS IF: 4.952 (the information was found in each journal home page).

REFERENCES

Abt, H. A. (2000). Astronomical Publication in the Near Future. *Publications of the Astronomical Society of the Pacific*, *11*(777), 1417-1420. doi:10.1096/317692.

Abt, H.A. (2007a). The publication rate of scientific papers depends only on the number of scientists. *Scientometrics*, *73*(3), 281-288. doi:10.1007/s11192-007-1807-8.

Abt, H. A. (2007b). The frequencies of multinational papers in various sciences, among them astronomy. *Scientometrics*, 72, 105-115. doi:10.1007/s11192-007-1686-z.

Abt, H. A. (2007c). The future of single-authored papers. *Scientometrics*, 73, 353-358. doi:10.1007/s11192-007-1822-9.

Adler, N. J., & Harzing, A.-W. (2009). When knowledge wins: transcending the sense and nonsense of academic rankings. *Academy of Management Learning and Education*, *8*(1), 72-95. doi: 10.5465/AMLE.2009.37012181.

Alcaraz, M. Á., & Méndez, D. I. (2016). When astrophysics meets lay and specialized audiences: titles in popular and scientific papers. *Journal of Language and Communication*, *3*(2), 109-120.

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. doi: 10.1007/s11192-007-1984-5.

Beel, J., & Gipp, B. (2009). Google Scholar's ranking algorithm: an introductory overview. In B. Larsen, & J. Leta, (Eds.). *Proceedings of the 12*th *International Conference on Scientometrics and Informetrics* (pp. 230–241, Vol. 1). Rio de Janeiro (Brazil), 14-17 July.

Butler Burton, W. (2007). Some statistics regarding the *Astrophysical Journal* as published between 1990 and 2006, including comparisons with sister journals. In A. Heck, & L. Houziaux (Eds.). *Future Professional Communication in Astronomy* (pp. 31-59). Mémoires de l'Académie Royale de Belgique.

Titles of Scientific Letters and Research Papers In Astrophysics:A Comparative Study of Some Linguistic Aspects and Their Relationship with Collaboration Issues

139

Busch-Lauer, I. A. (2000). Titles of English and German research papers in medicine and linguistics theses and research articles. In A. Trosborg (Ed.). A*nalysing Professional Genres* (pp. 77-94). Amsterdam/Philadelphia: John Benjamins.

Chandrasekhar, S. (1967). Editorial. *The Astrophysical Journal*, *148*(1).

Day, R. A. (1995). *How to write and publish a scientific paper* (5th). Cambridge (UK): Cambridge University Press.

Entralgo, J., Salager-Meyer, F., & Luzardo Briceño, M. (2015). ¿Cuán gramaticalmente complejos son los títulos de los artículos científicos en las ciencias naturales? *Revista de Lenguas para Fines Específicos*, *21*(2), 70-97. doi: 10.20420/rlfe.2015.0012.

Fox, C. H., & Burns, C. S. (2015).The relationship between manuscript title structure and success: editorial decisions and citation performance for an ecological journal. *Ecology and Evolution*, *5*(10), 1970-1980. doi: 10.1002/ece3.1480.

Gesuato, S. (2009). Encoding of information in titles: practices across four genres in linguistics. In C. Taylor (Ed.). *The role of E-corpora in translation and language learning* (pp. 127-157). Edizioni Università di Trieste (Italy): Ecolingua.

Haggan, M. (2004). Research paper titles in literature, linguistics and science: dimensions of attraction. *Journal of Pragmatics,* 36(2), 293-317. doi: 10.1016/S0378-2166(03)00090-0.

Henneken, E. A. (2012). Publication trends in astronomy: the lone author. Retrieved from arXiv:1202.4646.

Hartley, J. (2008). *Academic Writing and Publishing: A Practical Handbook*. Abingdon (UK)/New York (USA): Routlege.

Hudson, J. (2016). An analysis of the titles of papers submitted to the UK REF in 2014: authors, disciplines, and stylistic details. *Scientometrics*, *109*(2), 871-889. doi: 10.1007/s11192-016-2081-4.

Jaime-Sisó, M. (2009). Titles or headlines? Anticipating conclusions in biomedical research article titles as a persuasive journalistic strategy to attract busy readers. *Miscelánea*, *39*, 29-54.

Jamali, H. R., & Nikzad, M. (2011). Article type and its relation with the number of downloads and citations. *Scientometrics*, *88*(2), 653-661. doi: 10.1007/s11192-011-0412-z.

Kuch, T. d. C. (1978). Relation of title length to number of authors in journal articles. Journal of the American Society for Information Science, 29(4), 200-202. doi: 10.1002/asi.4630290408.

Leventhal. P. (2011). Introducing the manuscript writer's column. *The Write Stuff*, *20*(3), 199.

Méndez, D. I., Alcaraz, M. Á., & Salager-Meyer, F. (2014a). Evaluating authorship variation in English-medium Astrophysics research papers: an across journal and diachronic study (1998-2012). *Revista Canaria de Estudios Ingleses*, *69*, 51-63.

Méndez, D.I., Alcaraz Ariza, M. Á, & Salager-Meyer, F. (2014b). Titles in English-medium Astrophysics research articles. *Scientometrics*, *98*(3), 2331-2351. doi: 10.1007/s11192-013-1174-6.

Méndez, D.I, & Alcaraz, M. Á. (2015a). Exploring acknowledgment practices in English-medium astrophysics research papers: implications on authorship. *R e - vista de Lenguas para Fines Específicos*, *21*(1), 132-159. doi: 10.20420/rlfe.2015.0007.

Méndez, D. I., & Alcaraz, M. Á. (2015b). The use of abbreviations in English-medium astrophysics research paper titles: a problematic issue. *Advances in Language and Literary Studies*, *6*(3), 185-196. doi: 10.7575/aiac.alls.v.6n.3p.185.

Méndez, D.I, & Alcaraz, M. Á. (2015c). Astrophysics titles in Scientific American magazine (1990-2014): linguistic and discourse practices. *International Journal of Applied Linguistics & English Literature*, 4(6), 39-51. doi: 10.7575/aiac.ijalel.v.4n.6p.39.

Méndez, D. I., & Alcaraz, M. Á. (2016). Simultaneous and inter-comparable numerical indicators of collaboration practices in English-medium Astrophysics research papers. *Journal of Information Research*, *21*(3), paper 719. Retrieved from http://InformationR.net/ir/21-3/paper719.html.

Publishing Research Consortium (2011). *European Science Editing*, *37*(1), 21. http://www.ease.org.uk/sites/default/files/february_2011_371.pdf.

Smith, G. H. (2016). Trends in multiple authorship among papers in astronomy. *Publications of the Astronomical Society of the Pacific*, *128*(970), 124502. doi: 10.1088/1538-3873/128/970.

Soler, M. V. (2011). Comparative and contrastive observations on scientific titles written in English and Spanish. *English for Specific Purposes*, *30*(2), 124-137. doi: 10.1016/j.esp.2010.09.002.

de Solla Price, D. J. (1963). *Little Science, Big Science*. New York (USA): Columbia University Press.

Swales, J., & Feak, C. (20123rd). *Academic Writing for Graduate Student: Essential Tasks and Skills*. Ann Arbor (MI): University of Michigan Press.

Tarone, E., Gillette, S., Dwyer, S, & Icke, V. (1998). On the use of the active and passive voice in astrophysics journal articles: With extensions to other languages and other fields. *English for Specific Purposes*, 17, 113–132. doi: 10.1016/S0889-4906(97)00032-X.

Trimble, V. (1984). Postwar growth in the length of astronomical and other scientific papers. *Publications of the Astronomical Society of the Pacific*, 96(586), 1007-1016. doi: https://doi.org/10.1086/131467.

Wang, Y. & Bai, Y. (2007). A corpus-based syntactic study of medical research article titles. *System*, *35*(3): 38-399. doi: 0.1016/j.system.2007.01.005.

White, A. (1991). A further exploration of title size and author number. *Journal of the American Society of Information Science*, *42*(5), 384-385. doi: 10.1002/(SICI)1097-4571(199106)42:5<384:AID-ASI9>3.0.CO;2-G.

Yitzhaki, M. (1994). Relation of title length of journal articles to number of authors. *Scientometrics*, *30*(1), 321-332. doi: 10.1007/BF02017231.