

Exploring ERPP writing challenges: An investigation into the perceived and analysed difficulties of Spanish EFL university researchers

Natalia Judith Laso Martín & Elisabet Comelles Pujadas

University of Barcelona (Spain)

njlaso@ub.edu, elicomelles@ub.edu

Abstract

Research writing competence in English is essential for EFL communities, who experience increasing pressure to publish their research studies in peer-reviewed international journals. As already stated in the literature (Burgess et al., 2019; Cargill & Burgess, 2008; Cargill & O'Connor, 2013; Corcoran et al., 2019; Lillis & Curry, 2010; Villagran & Harris, 2009), the purpose of the scientific community is to disseminate their results among practitioners in their field.

This paper explores the main challenges that ERPP poses for a group of EFL university researchers as well as whether their perceptions of their weakest areas correlate with the analysis of their drafts. Results show that participants perceive word combinations and grammatical issues as the most challenging aspects in their writing process, followed by misuse of similar pairs of words and, to a much lesser extent, issues regarding punctuation, linking words, and paragraph distribution.

Keywords: ERPP, academic writing, corpus-based resources and tools, EFL researchers' perceptions, research paper drafts.

Resumen

Explorando los desafíos de la escritura en inglés para fines de publicación científica: una investigación sobre las dificultades percibidas y analizadas en inglés como lengua extranjera entre los investigadores universitarios españoles

El manejo de la escritura académica en inglés con vistas a la publicación de artículos es esencial para los hablantes de inglés como lengua extranjera (EFL,

por sus siglas en inglés), quienes experimentan una presión creciente para publicar sus estudios en revistas internacionales con revisión por pares. Tal y como se señala en la bibliografía (Burgess et al., 2019; Cargill y Burgess, 2008; Cargill y O'Connor, 2013; Corcoran et al., 2019; Lillis y Curry, 2010; Villagran y Harris, 2009), el propósito de la comunidad científica es difundir sus resultados entre los profesionales de su campo.

Este artículo explora los principales desafíos que plantea la escritura en inglés para fines de publicación científica (ERPP, por sus siglas en inglés) para un grupo de investigadores universitarios en contextos de EFL, así como si sus percepciones acerca de las áreas más problemáticas se correlacionan con el análisis de las versiones preliminares de sus artículos. Los resultados muestran que los participantes perciben las combinaciones de palabras y los problemas gramaticales como los aspectos más desafiantes en su proceso de escritura, seguidos del uso incorrecto de pares de palabras similares y, en mucha menor medida, cuestiones relacionadas con la puntuación, los conectores y la distribución de los párrafos.

Palabras clave: ERPP, escritura académica, herramientas y recursos basados en corpus, percepciones de investigadores, versiones preliminares de artículos de investigación.

1. Introduction

Research writing competence in English is essential for EFL communities, who experience increasing pressure to publish their research studies in peer-reviewed international journals. As already stated in the literature (Burgess et al., 2019; Cargill & Burgess, 2008/2017; Cargill & O'Connor, 2013; Corcoran et al., 2019; Lillis & Curry, 2010; Villagran & Harris, 2009), the purpose of the scientific community is to disseminate their results among practitioners in their field. Thus, it seems reasonable that in order to gain well-recognized access to a discourse community and, most importantly, acceptance within it, researchers must become familiar with the 'game strategies' (Etherington, 2008) involved in domain-specific research writing (Laso & John, 2017; Pérez-Llantada, 2014).

The pressure applied by the university system on academics has significantly increased in recent decades. There are numerous mechanisms and university policies urging teaching and research staff to publish in English in high-impact journals, which must be appropriately indexed. Therefore, it is not only about the willingness to disseminate scientific results for the scientific

community, but also about the need to do so in English, often in contexts where English is used as a foreign language by non-Anglophone research communities. This pressure often worsens the lack of confidence when drafting a research article, and places EFL scholars at a disadvantage compared to their L1 English peers (Li & Flowerdew, 2020; Mendoza et al., 2021). However, some ERPP research also suggests that L2 English scholars do not always report being disadvantaged compared to their L1 English counterparts, or that any disadvantages may not necessarily be due to language alone (Arnbjörnsdóttir, 2019; Muresan & Pérez-Llantada, 2019).

As a consequence of the aforementioned urgency to have an extensive and high-quality research production in English and the added difficulty of the academic language barrier, an expanding body of research has focused on the different strategies and initiatives that universities are developing to increase scholars' publishing success (Corcoran, 2022; García Landa, 2006; Li & Flowerdew, 2020), as well as the perceptions that EFL academics have regarding the (mainly) lexicogrammatical and rhetorical challenges they face when writing an article in English (Cargill & Burgess, 2017). These challenges often stem from the need to acquire phraseological competence in academic English and develop a good working knowledge of discipline-specific formulaic language (Granger & Meunier, 2008; Ferguson et al., 2011; Laso & John, 2013; Pérez-Llantada, 2014).

Hence, any EFL writer who aspires to get their manuscript accepted for publication needs to master the stylistic conventions characteristic of their research discipline (Cotos, 2016). To this respect, there is mounting evidence in ESP writing research (Ellis, 2007; Granger & Meunier, 2008; Hyland, 2008; Laso & John, 2013; Nesselhauf, 2005; Paquot, 2008, among others) that points to the fact that EFL specialized discourse communities already have a good command of domain-specific terminology. However, they find it tantalizingly challenging to acquire phraseological competence and become familiar with the most prototypical domain-specific collocational patterns in English.

Motivated by this need, lately there has also been a growing number of studies (Boulton & Cobb, 2017; Charles, 2007; Charles & Hadley, 2022; Chen & Flowerdew, 2018; Eriksson, 2012; Lee et al., 2019; Li & Flowerdew, 2020; Sun, 2007) that focus on corpus consultation as a pedagogical tool to improve writing skills in general, and phraseological competence, in particular. Some of these studies (Eriksson, 2012; Lee et al., 2019) stress the

usefulness of corpus linguistic techniques for vocabulary learning. According to Hyland (2004), EFL writers can take advantage of contrasting a corpus of published research articles against their own written production to identify “the preferred way of saying things in a particular discourse” (Wray, 2002).

Sun (2007) argues that students can improve their knowledge of the most prototypical phraseological patterns in academic writing by analysing discourse moves in research paper samples with concordance tools. Similarly, Eriksson (2012) notes that lexical bundles for research writing can also be taught using corpora. More recently, some studies on long-term trends in corpus use like Charles’s and Hadley’s (2022) have explored how graduate students can improve their research writing skills by the autonomous use of a self-compiled domain-specific corpus.

In addition to exploring the linguistic, phraseological, and rhetorical challenges faced by EFL research writers and addressed by ERPP instructors, it is important to analyse the perceptions that EFL research writers have regarding these challenges. In this regard, it is of interest to explore:

- a) how the dominance of English is affecting their research work as scholars whose first language is not English,
- b) the kind of pressures they face in the enterprise of getting their research published in international journals,
- c) what barriers to publishing in English they encounter, and
- d) what areas of English linguistic competence and writing skills they feel themselves less confident about.

Several studies have reported on scholars’ attitudes towards the impact that the prevalence of English as the international language of academic communication has on their scientific production (Bocanegra-Valle, 2013; Pérez-Llantada et al., 2011), as well as the disadvantaged situation they find themselves in. However, they are aware that this is a reality, and that the system requires them to familiarise themselves with the rules of the game in order to contribute to knowledge dissemination and feel engaged in their discourse research community (Laso & John, 2017; Pérez-Llantada et al., 2011; Pérez-Llantada, 2014).

Regarding the perception of the areas that academics find most problematic when writing their articles in English, Giraldo (2019, p. 200) points out that there are few studies exploring “on perceptions and learning of scholars from different disciplines while engaged in English for Research Publication Purposes (henceforth ERPP) courses and on recommendations emerging from such contexts”. However, it is worth noting, for example, the study by Pérez-Llantada et al. (2011), which analyses data from interviews with senior Spanish scholars, who highlight various linguistic aspects (e.g., lexical diversity, sentence syntax) and rhetorical aspects, such as control of modality and hedging devices. Pérez-Llantada et al. conclude that the identification of such difficulties by ERPP writers should be complemented with the textual analysis of their draft papers in order to adequately “diagnose specific linguistic difficulties that could be addressed in instructional programs” (2011, p. 29).

Based on the discussion above, we decided to examine the most problematic areas identified in the research paper drafts of a group of Spanish EFL research writers, as well as their perceptions regarding the difficulties associated with scientific production in English. The value of this research lies in exploring how writers’ perception of their writing difficulties correlates with the challenges noted in their actual research paper drafts.

This study was carried out within the context of an “English for Research Publication Purposes” workshop. On the first day of the workshop, participants in the study were administered a questionnaire regarding their research writing process in English. Additionally, prior to the workshop they had been asked to submit an 800-word draft of a research paper (including the abstract) they were currently working on. All the collected data were used to design corpus-based materials implemented in the workshop. The following research questions (RQs) guided this study:

- RQ1. What areas of research writing pose more difficulties for Spanish EFL university researchers?
- RQ2. How does the writers’ perception of their weakest areas correlate with the analysis of their drafts?

2. Context of the research

In order to provide researchers at the University of Barcelona with the necessary writing skills to produce an academic research article using

appropriate academic language and style, an ESP workshop on “English for Research Publication Purposes (ERPP)” was addressed to a group of thirteen researchers working at the aforementioned university. The ERPP workshop consisted of six 2h-sessions and was held in a computer laboratory. It aimed at a) introducing writers to the language and style characteristic of academic English, b) engaging writers in exercises around the academic language typical of successful academic writing, and c) showing them tools and resources, which they would eventually be able to use independently.

The workshop was part of the training programme offered to academic staff at the University of Barcelona, which usually runs in January and July (two editions). These are non-teaching periods at university, and it is usually when teachers and researchers go on study leave and/or participate in conferences, so a total number of six participants (out of thirteen who had formerly signed up) attended regularly.

One of the aims of the workshop was for participants to familiarise themselves with the use of corpus-based tools. To this end and following Lee’s and Swales’ study (2006), they built up a corpus of already published articles (an average of 5-10) from prestige journals in their research area (i.e., psychology, earth sciences, economy, climatology, and anthropology). The software used during the workshop was AntConc (4.1.2) (Anthony, 2022). Participants were guided through the process of converting the downloaded material into plain texts and were introduced to some of the tools available in AntConc; mainly, search results in KWIC format, collocates and wordlists.

In addition to working on published academic articles, participants had to work on their own pieces of writing and received individual help during the workshop. Thus, prior to the workshop, they had to upload an 800-word draft of one of their research articles to a Moodle platform. The draft consisted of one section of a paper, along with the abstract. It was expected that at the end of the workshop at least one section of the paper would have been revised using the guidelines provided. Participants were equipped with the necessary skills to continue working on the rest of the article on their own.

3. Method and data

The research presented in this study consisted of three different steps. The first step involved the participants sending a draft of their writing to the

researchers before the beginning of the workshop; in the second stage participants were asked to answer a questionnaire once the workshop had already started; and finally, in the last step, workshop facilitators analysed the writings sent by the participants and manually annotated the mistakes they found. After the draft analysis, the results obtained were compared against the participants' answers to the questionnaire. Next, each of these stages is described in detail.

3.1. Stage 1: Draft sending prior to the workshop

Approximately a month before the workshop, participants were sent a welcome message where they were given some information about the training sessions, such as the aims of the workshop and the activities it would involve. They were told that they would be working on published research articles, as well as their own pieces of writing, and that they would receive individual help during the workshop. They were also requested to upload a draft of their writing to the Moodle platform three weeks before the beginning of the workshop. The draft had to be at least 800 words long and had to consist of an abstract and one section of their paper. A total of thirteen participants submitted their drafts, but only six attended the workshop, thus only the data from those six participants will be presented, analysed, and discussed. The participants' submissions belonged to very different domains including psychology, earth sciences, economy, climatology, and anthropology.

Out of the six participants who sent their writings, two sent the abstract and the results section, two of them provided the abstract and the methodology, and the rest submitted the abstract and the introduction. The length of the drafts ranged from 850 words to 912 words. See Table 1 for details about submission type and length:

| Participant ID | Parts submitted | Number of words per section | Total number of words |
|----------------|-------------------------|---|-----------------------|
| ERPP0722_1 | Abstract + results | 331 words (abstract) + 562 words (results) | 893 words |
| ERPP0722_2 | Abstract + methodology | 203 words (abstract) + 677 words (methodology) | 880 words |
| ERPP0722_3 | Abstract + methodology | 389 words (abstract) + 498 words (methodology) | 887 words |
| ERPP0722_4 | Abstract + introduction | 275 words (abstract) + 637 words (introduction) | 912 words |
| ERPP0722_5 | Abstract + results | 252 words (abstract) + 623 words (results) | 875 words |
| ERPP0722_6 | Abstract + introduction | 312 words (abstract) + 538 words (introduction) | 850 words |

Table 1. Participants' submissions details.

3.2. Stage 2: Questionnaire response during the workshop

During the first session of the workshop, participants were asked to answer a questionnaire (see Appendix) through a Google Forms link in order to make them reflect on their writing process and, most specifically, on the use of English when writing. This questionnaire had already been designed for previous rounds of ERPP Workshops conducted by the authors at the University of Barcelona during the years 2014, 2018, 2019, 2020 and 2022 and was published in an earlier investigation (see Laso & John, 2017).

Such information would also make us aware of their main challenges and concerns when facing the task of writing and publishing in English and would help us focus the workshop sessions and activities to meet the participants’ needs.

The questionnaire was divided into 4 main parts: (1) general questions about their experience when writing in English, (2) issues concerning their article writing process in English, (3) difficulties when writing in English and (4) the effect of English in their work as scholars.

The first section, General Questions, included four open questions aimed at getting information on the participants’ reasons for writing in English, their readership, the frequency in which they write papers in English, and the journals they tend to publish in.

The second section focused on the writing process and included four questions. In the first question, participants were asked to choose among

several options related to the pre-writing stage. They could choose as many as they felt appropriate to describe the actions they took before writing their paper. The options presented were:

- I brainstorm ideas
- I read relevant articles from journals
- I write a plan of the paper
- I draft short sections of the paper

The second and third questions were related to the writing stage. In the second question, participants were again provided with several options from which they could choose as many as applicable, being those options the following:

- I write my paper in my mother tongue and then translate it to English
- I write the paper in English and look up a few unknown words and expressions
- I write the paper mostly in one language (either mother tongue or English) and some parts (more than individual words) in the other language
- I check the style of the journal I intend publishing in and try and write in the same style
- I write my paper in a particular order
- I don't write my paper in any particular order

The third question was a follow-up question addressed only to those participants who wrote their paper in a particular order, so that they could specify the sections they tend to start and finish with.

In the last question of this section, participants were given two options regarding the revision process. They could choose either option or both of them, if necessary.

- I check for spelling and grammatical errors
- I ask for an expert to check my writing before submitting my paper

The third section of the questionnaire intended participants to reflect on the difficulties they encounter when writing their papers in English. To that aim, they were asked to rate a set of 5 items according to the level of difficulty by means of a five-point Likert scale, being 1 very easy and 5 very difficult. The items included covered the following language aspects: terminology, word combinations/collocations, distinction between similar pairs of words (e.g., *make/do*, *say/tell*), punctuation, wordiness and paragraphing, and grammar issues (e.g., verb tenses, word order). In addition, an optional box was provided for those participants facing difficulties not included in the items mentioned above.

The last section of the questionnaire consisted of three open-questions for participants to reflect and express their opinion on how English, being the *lingua franca* in their areas of research, affects their life as academics as well as their publishing experience.

3.3. Stage 3: Written production analysis and manual annotation

The writings produced by the workshop participants were analysed by the two facilitators. This analysis involved identifying mistakes and annotating them according to an agreed typology. This typology classified mistakes into 4 categories connected to the last 4 questions in section 3 of the paper: Grammatical Issues (e.g. misuse of verb tenses, wrong word-order, wrong word-forms, etc.); Collocational Patterns, that is, issues with prototypical word combinations of academic writing (e.g. verb + preposition, adjective + noun, etc.); Misuse of similar lexical items (e.g. use of *make* vs. *do*, *say* vs. *tell*, *quick* vs. *fast*, etc.); and Problems with punctuation, paragraphing and wordiness. The six texts produced by the participants were analysed and annotated by both facilitators separately. Later, these annotations were compared and, in cases of disagreement, a discussion was held and a final agreement was reached. The texts were annotated using Atlas.ti, which allowed for easily computing the total number of mistakes per category.

Later, the number of mistakes identified per category were compared against the participants' perceptions to check whether they correlated.

4. Results and discussion

This section presents the results of the questionnaire that participants answered on the first day of the workshop, as well as the findings revealed

by the analysis of their drafts. In addition, both the participants' answers and the draft analysis have been compared in order to find whether the participants' perceptions correlate with the draft findings.

4.1. Results of the questionnaire

In Section 1, most participants pointed out that they write their research in English because English is the publishing language of the best journals and conferences (4 answers). Thus, publishing in these media is key to boost the researchers' academic career (3 answers) and reach a wider audience (2 answers). Interestingly, very few of them (1) only write in English when writing a paper, most of them write in English on a monthly (1) or weekly basis (2).

As regards their potential readership, all the participants agreed that they write for other researchers; although some participants also referred to other types of audiences (e.g. workshop committees and academic authorities (2), educators and teachers (2)).

Section 2 enquired participants about the actual writing of a paper in English. The answers to the pre-writing stage question showed that all participants write a plan of the paper, five participants also read relevant articles from journals, four of them draft short sections of the paper, and half of them (three participants) brainstorm ideas before writing their paper.

In the writing stage question, the answers were more diverse. From the answers related to the language used for writing their papers, it is worth mentioning that half of the participants write their papers in their mother tongue first and they translate them into English later, whereas the other half write their paper directly in English and look up those words they do not know in a dictionary, or they even combine both languages when writing their papers.

In relation to the strategy they use when writing, most of them (five participants) follow a particular order whereas only one does not. Finally, surprisingly, only half of the informants confirmed that they check the style of the journal and try to write the same style.

From the five participants that confirmed writing their paper in a particular order, their answers regarding the section they start with varied considerably; however four of them finish with either the introduction or the abstract.

As for the post-writing stage, three participants follow a double strategy, they first check for spelling and grammatical errors, and ask an expert to check their writing at a later stage. As regards the rest of the participants, two of them check for spelling and grammatical errors without the help of an expert, whereas one participant only relies on an expert's help.

Section 3 asked participants about their difficulties when writing a paper in English. Results showed that terminology only seems to be a problem for two of the participants, whereas four of them did not regard this as a major issue, which confirms findings reported in previous studies (Ellis, 2007; Granger & Meunier, 2008; Hyland, 2008; Laso & John, 2013; Nesselhauf, 2005; Paquot, 2008, among others).

Collocations, however, were definitely one of the most challenging aspects, since they ranged between 4 and 5 (being 5 very difficult).

Regarding the pairs of words with similar meaning but different use, half of the participants stated that they are a difficult aspect, whereas the other half did not seem to consider them particularly challenging.

The answers to the question on punctuation, wordiness and paragraphing were quite distributed, ranging from 2 to 5. Two participants rated them between 4 and 5, whereas the rest of participants did not consider these issues particularly difficult.

Finally, grammar seemed to be problematic for four participants, who rated it from 4 (two participants) to 5 (two participants). The rest of the participants did not seem to find grammar issues particularly problematic.

When asked whether they found other language issues challenging, participants stated that they had problems being clear and concise, sounding English-like (e.g. using typical English structures, shorter sentences) and using suitable academic expressions.

In the last section of the questionnaire, participants were inquired about the effect that English has in their life as scholars. Most participants (four) claimed that the dominance of English strongly affected their everyday lives as scholars and two participants even specified that they had problems when expressing themselves in English (e.g. they cannot convey their ideas fluently). Participants also mentioned that being non-native speakers of English in an English-publishing context meant that getting their articles accepted and published took longer because of the number of revisions required.

4.2. Draft analysis

This section reports on the linguistic areas that have proven to be most challenging in participants’ drafts, as illustrated in Table 2 below. These linguistic areas are related to the last four questions in section 3 of the questionnaire.

| | |
|---|---|
| Collocational patterns | Adverb + Adjective (e.g., <i>essentially significant; frankly poor</i>) Verb + Noun (e.g., <i>represent a (good) indicator; generate uncertainties; express a conclusion; prove blindness; provoke research</i>) Adjective + Noun (e.g., <i>notable errors; significant/appreciable improvements</i>) Adjective + preposition (e.g., <i>responsible 'of</i>) Lexical bundles (e.g., <i>inside a project; in reference to; in the *remainder of this article</i>) |
| Misuse of similar lexical items | Verbs: overuse of light verbs (e.g., <i>give</i>); <i>maintain/hold; make/do (a study)</i> ; <i>focus/centre</i> Nouns: <i>view/vision</i> |
| Punctuation, wordiness and paragraphing | Punctuation Word order Wordy paragraphs Lack of transition between paragraphs Overuse/misuse of linking words (e.g., <i>although, while, therefore</i>) Run-on sentences |
| Grammatical issues | Missing or unclear referents S+V agreement Verb tenses |

Table 2. Challenging areas found in participants’ drafts.

For ease of discussion, we have divided this section into four parts: collocational patterns, misuse of similar lexical items, punctuation, wordiness and paragraphing, and grammatical issues.

4.2.1. Collocational patterns

As already reported in the concerned literature (cf. section 1), domain-specific collocational patterns stand out as one of the most challenging aspects of research writing. Data from the drafts analysed in this study show that participants struggle with certain prototypical word combinations, involving verbs, nouns, adjectives, adverbs, and prepositions. In order to analyse the prototypicality of certain collocational patterns found in the drafts, both the British National Corpus (BNC) and the Academic Collocation List (ACL) (Ackermann & Chen, 2013) have been used as resources. The close analysis of these collocational patterns has yielded interesting findings regarding participants’ command of the following word combinations: adverb + adjective; verb + noun; adjective + noun; and adjective + preposition.

Evidence from participants' drafts shows that certain adjectives like *significant* (example 1) and *poor* (example 2) appear in combination with non-prototypical adverbs, *essentially* and *frankly*, respectively, which can be illustrated in the examples below:

- (1) There are discrepancies in magnitude that are **essentially significant**. IMERG_E and IMERG_L overestimate precipitation by over 20% in almost all the territory with Biases of 160 and 140 mm/year, respectively (ERPP0722_1)
- (2) At sub-daily scales the ability of IMERG to estimate events at different thresholds is **frankly poor**. (ERPP0722_1)

The collocations “essentially significant” and “frankly poor” are not found in the ACL. Similarly, frequently used nouns in academic writing, such as *conclusion*, *research*, and *indicator* are used in co-occurrence with free verb collocates (Wang & Shaw, 2008) like *represent*, *express*, or *provoke*, which result in the following deviant collocations:

- (3) While this significant error reduction could **represent a good indicator** of the improvement in precipitation estimation with IMERG_F, we must consider the limitations of this statistic and its relationship with the cancellation of positive and negative errors (Jolliffe & Stephenson, 2012) between IMERG_F and ground-based observations. (ERPP0722_1)
- (4) [...] ISAE 3000 consider the need to **express a conclusion** about the degree of confidence. (ERPP0722_6)
- (5) The growing interest in conflicts and their management in the field of education, has **provoked** in recent decades **a large amount of research** and approach proposals from different disciplines. (ERPP0722_5)

In the ACL, the verbs *draw* and *lead to* stand out as the most frequently used restricted verb collocates of the abstract noun *conclusion*, whereas the free verb collocate *express* is more limited in its use when reporting results in research writing. Regarding examples (3) and (5), it must be noted that some L1 interference may result from the use of the free verb collocates *represent* (*representar* in Catalan or Spanish) and *provoke* (*provocar* or *suscitar* in Catalan or Spanish), which would be very natural free verb collocates in the verb combinations *representa un buen indicador* (Sp)/*representa un bon indicador* (Cat) and *ha provocado/suscitado una gran cantidad de* (Sp)/*ha provocat/suscitat una gran quantitat de* (Cat), respectively.

There are two other examples of unusual verb-noun collocations; namely *generate uncertainties* and *prove blindness*. Example (6) illustrates the use of a free verb collocate (e.g., *generate*) instead of more restricted options, such as *cause* or *result in*, whereas in example (7) some issues regarding literal translation of metaphorical uses of the abstract noun *blindness* are involved. The construction “blindness of most State courts to cultural factors is proven” seems to recall the Spanish expression “se demuestra una ceguera de los tribunales estatales con respecto a factores culturales”. In English, an SVC construction of the type “blindness is total/severe” (most common adjective collocates of the noun *blindness* in the BNC) or other fixed expressions that highlight the metaphorical use of *blindness* in this context (i.e., the short-sighted approach of State courts in the concerned issues (e.g., “blindness of most State courts to cultural factors is beyond understanding”) would have been preferred.

- (6) The erratic behaviour of the estimates may also be influenced by the variability of precipitation at the point scale and the **uncertainties** that are **generated** in the process of gridding the IMERG products. (ERPP0722_1)
- (7) However, at an empirical level, the **blindness** of most State courts to cultural factors **is proven**. (ERPP0722_4)

Other collocational patterns that seem to have been challenging to participants are related to adjective collocates of nouns. On the one hand, example (8) shows that the noun *error* co-occurs with the adjective *notable* in a predicative position. In this respect, it is relevant that both the adjective *notable* and the adjective *noticeable* are possible collocates of the noun *error*. However, in the ACL only the former appears in combination with the noun *exception*. What seems to be particularly relevant in example (8) is the syntactic construction used. The adjective *notable* is more frequently found in attributive position (e.g., *notable errors*) rather than in predicative constructions of the type “subject + verb + *more notable*”, where the adjective *noticeable* would be preferred.

- (8) These **errors** are more **notable** during the summer period with a systematic error of over 97 mm and MAE and RMSE values around 105 mm and 143 mm, respectively. (ERPP0722_1)

Another observation regarding adjective + noun collocates is the use of the noun *improvement* co-occurring both with the evaluational adjectives *significant*

and *appreciable* in example (9). Both adjectives do collocate with the noun *improvement*, but the former is more prototypically used than the latter. According to the BNC, the adjective *significant* is much more frequent (98 occurrences) than the adjective *appreciable* (2 occurrences) in combination with the noun *improvement* and the latter adjective is only used in an attributive position, unlike example (9) below.

- (9) The analysis of the average error (Bias) reflects a **significant improvement** in the IMERG_F at all scales, although much more **appreciable** at the monthly and annual scales. (ERPP0722_1)

A further point to note about challenging collocational patterns in the analysed drafts concerns the dummy prepositions that collocate with adjectives, such as *responsible*. Data show that the collocation *responsible for* has proven to be problematic, as illustrated by example (10):

- (10) Plus, the fact that financial auditors are **responsible of** this activity denotes a mimetic behaviour with the financial information and also with the big corporations (the prominence of the Big 4). (ERPP0722_6)

Another aspect that needs highlighting concerns lexical bundles, in Biber et al.'s (1999) and Hyland's (2008) words; that is, combinations of words that frequently co-occur in domain-specific texts, which have also proven to be challenging, as can be inferred from the examples below:

- (11) **Inside the App2five project**, redesigning quality educational Apps aimed at early childhood, we interviewed the participants' families in order to find out what factors determine their use of screens. (ERPP0722_5)
- (12) **In reference to** family perceptions, a positive view was raised. (ERPP0722_5)
- (13) **In the remainder of this article**, I intend to describe some paradigmatic sentences on each typology of case, analysing the problematic assessment of Roma culture that judges have made, and offering an alternative view of these cases, based on some ethnographic and anthropological literature. (ERPP0722_4)

Examples (11) and (12) show multiword units of the type "preposition + noun + preposition", introduced by a deviant preposition, which may have

been influenced by L1 interference, i.e., “**inside** the App2five project” vs. “**dins** del projecte App2five”, and “***in** reference to” (English) vs. “**en** referència a” (Catalan). A third prepositional bundle that may be due to L1 interference as a result of translation from Spanish is example (13): “en lo que queda de artículo, [...]” vs. “the remainder of this article, [...]”.

4.2.2. Misuse of similar lexical items

Data drawn from participants’ drafts have also stressed participants’ difficulty in discriminating between similar lexical items. In particular, the current study has revealed that participants overuse certain light verbs, such as *give* (examples (14-17)), which leads to low diversity, and thus less sophisticated, texts:

- (14) This paper is organized as follows: the first section **gives** a brief overview of the previous literature about assurance. (ERPP0722_6)
- (15) Assurance based on the AA1000 AccountAbility Principles is intended to **give** stakeholders assurance on the way an organization manages sustainability performance. (ERPP0722_6)
- (16) The systematic error values in these groups are similar in magnitude to the mean absolute errors and to the mean of the values recorded by the stations in each of the corresponding thresholds, which **gives** us a sign of a more realistic, significant underestimation. (ERPP0722_1)
- (17) Among the three products, IMERG_F **gives** the worst results, while IMERG_L **gives** the best values, although these differences are not marked. (ERPP0722_1)

Apart from frequent repetition of light verbs, some examples of misuse of similar lexical items like *maintain* vs. *hold/take*, *make* vs. *do* and *focus* vs. *centre* were common in the analysed drafts. According to Kaszubski (2000) and Nesselhauf (2005), the inappropriate use of verbs, such as *maintain*, *make*, and *centre* (examples 18, 19 and 20) reflects the widespread confusion of certain pairs of high frequency lexemes (*maintain/hold/take*, *make/do*, *focus/centre*) used as light verbs in verb-noun collocations, as illustrated in the examples below¹:

- (18) They **maintain** the **vision** [*hold the view*] that the consumption of audio-visuals can be less harmful to children than the use of video games, and they use technologies at home in this sense. (ERPP0722_5)

- (19) Recently, IFAC (2021) **made a study** [*conducted a study*] about the state of sustainability assurance in the world and offered some interesting statistics. (ERPP0722_6)
- (20) GRI **centres on** [*focuses on*] the usefulness on the quality of information and ISAE 3000 consider the need to express a conclusion about the degree of confidence. (ERPP0722_6)

Regarding example (18), the misuse of the noun *vision* may be due to some L1 interference with the Spanish expression “mantienen la visión de que el consumo de audiovisuales [...]”. In Spanish, the noun *visión* may be used both in its literal and metaphorical sense, whereas in English, the noun *view* is preferred when expressing a particular way of thinking about something.

4.2.3. Punctuation, wordiness, and paragraphing

During the workshop, we also provided feedback on thesis development, organization of topic sentences and punctuation. Most issues in participants’ drafts were concerned with the misuse of punctuation marks, like commas between the subject and the predicate, as illustrated by example (21):

- (21) The challenges of doing so over areas with complex orography (Chiaravallotti et al., 2018; Mayor et al., 2017) and under different climatic conditions (Sharifi et al., 2016), are factors that influence the performance of IMERG products and therefore the significance of their study. (ERPP0722_1)

Some other deviant structures concern the lack of punctuation marks either preceding or following multiword connective adjuncts like “*according to* + NP”, “(*due to* + NP (*,*))”, and “*such as* + NP”, to name a few:

- (22) R_t/R_0 can be expressed as $(\delta^h E_t + 1) / (\delta^h E_0 + 1)$ **according to** $\delta^h E$ definition, and the uncertainty corresponds to the 95% confidence interval (CI). (ERPP0722_2)
- (23) But this legal culture has been usually misunderstood or poorly described **due to** the impact of some cases reported by the media. (ERPP0722_4)
- (24) Bypassing the examination of these differences, the discussions on possible mitigating factors revolve around figures **such as the prohibition error** (*error de prohibición*) using analogies that are problematic from an anthropological point of view. (ERPP0722_4)

- (25) **At the same time** replicate extraction tests were performed in 500 mL water spiked to 0.025, 0.05 and 0.1 µg/L each analyte (n=3). (ERPP0722_3)
- (26) **For concentration analysis** Divinylbenzene N-vinylpyrrolidone copolymer cartridges (Oasis HLB 500 mg, Waters) were used. (ERPP0722_3)

The use of some of the connective adjuncts illustrated above has also proven to be problematic regarding their prototypical order, that is, “connective adjunct + main clause”, as can be seen in example (27):

- (27) The Spanish judicial system is not an exception, **according to some experts** (Truffin and Arjona 2010). (ERPP0722_4)

Further, other examples of participants’ drafts illustrate misuse of recurrent linking words, such as *although*, *while*, and *therefore*, which may occur several times in the same paragraph. Examples (28) and (29) include concessive clauses introduced by *although*, but the second occurrence of this connector in both examples seems to introduce a contrast and would read best if introduced by the coordinating conjunction *but*. As regards example (29a), the order of the first two clauses, i.e., “IMERG_F [...] scales” and “although [...] scales” should be reversed.

- (28) (28a) **Although** in isolated points, these values are also registered by the rain gauges, there is an overestimation of the precipitation by the IMERG. In the time series of station C9 there is a maximum peak at 12:00 UTC identified by most of the IMERG products, (28b) **although** with a significant overestimation close to 10 mm, which is more noticeable in the IMERG_F. (ERPP0722_1)
- (29) As shown in Fig. 6, IMERG_F has a higher ability to detect negative corrects with values close to 50% at all scales, (29a) **although** IMERG_L and IMERG_E are also very similar at sub-daily and daily scales. The percentage of hits tends to increase at scales higher than daily, while the percentage of misses decreases significantly. According to the selected thresholds, the ability of IMERG to estimate precipitation is affected by the detected false alarms. These usually represent the highest percentage at all scales, (29b) **although** they are higher during the summer period in IMERG_E and IMERG_L. (ERPP0722_1)

As illustrated in example (30), concessive clauses introduced by *while* have also proven to be problematic due to misuse of punctuation marks (i.e., “[...] for entertainment (56%)*. **While**, in the case of the computer, its use is mainly [...]” vs. “[...] for entertainment (56%); **while**, in the case of the computer, its use is mainly [...]”.

- (30) The smartphone is used for family contact (96%), search for information and social networks (85%), management (76%), work (67%) and, to a lesser extent, for entertainment (56%). **While**, in the case of the computer, its use is mainly for work (88%), information (69%) and management (59.5%). (ERPP0722_5)

The use of the consequence connector *therefore* has also posed some difficulties. In example (31) the linking adverbial *therefore* seems to have been misused as it does not introduce a consequence of the previous statement. On the contrary, the clause introduced by “Papadakis et al. (2019) showed how parents’ beliefs [...]” is an illustration of what is stated in the previous clause; that is, “a third factor studied is the perception of parents about ICT”.

- (31) To a lower extent, a third factor studied is the perception of parents about ICT. **Therefore** Papadakis (et al., 2019) showed how parents’ beliefs about technology directly affect the quality, quantity, and use of digital media available to children. (ERPP0722_5)

Closely related to the misuse of the concessive connector *although* in examples (28b) and (29b) is the issue of run-on sentences, which are also prevalent in participants’ drafts, as shown in example (32):

- (32) In the time series of station ‘U6’ [PUNCTUATION MARK MISSING] a false peak of maximum intensity is shown by the IMERG_UnCal 1 hour before the maximum values recorded by the meteorological station and although it is corrected by the IMERG_F, it maintains the overestimation of the real values. (ERPP0722_5)

4.2.4. Grammatical issues

The examination of participants’ drafts has revealed some grammatical issues in their writing, which mainly concern:

a) the absence or misuse of anaphoric reference instanced by the pronouns *this* or *it* (examples 33 and 34):

- (33) This law –a body of unwritten norms that, like other customary laws, is affected by its subordination to the State law (Falk Moore 1973, 1988)– seems to vary a lot in its content and procedures (Weyrauch 2001, Weyrauch and Bell 2001, Acton, Caffrey and Mundy 2001, Caffrey and Mundy 2001, Acton 2005, San Román 1998, Lagunas 2000, Rodríguez 2011). But [ANAPHORIC REFERENCE MISSING] tends to include authority figures that intervene in case of conflict and clear procedures, defining elements of any ‘system of law’ (Nader and Todd 1978). (ERPP0722_4)
- (34) Similarly, for the analysis at monthly and annual scales with cut-off points above 100 mm and 1000 mm, respectively, ***this** provides good guarantees for identifying rainfall events up to this threshold. (ERPP0722_1)

b) verb tense confusion involving perfect and continuous tenses (example 35) or active-passive constructions (example 36):

- (35) In this study, we **are investigating** the potential of using Emulsified Vegetable Oil (EVO) as a long-term electron donor for chloroethenes (CEs) remediation. (ERPP0722_2)
- (36) The experiment **was started** on the same day, and it was prepared in an anoxic chamber, to preserve the conditions of the sampled water. (ERPP0722_2)

c) lack of agreement between the subject and the verb (examples 37-39):

- (37) In addition, **the probability of** occurrence of annual precipitation and Kernel Density Estimation (KDE) curve associated with the distribution of each dataset **are plotted**. (ERPP0722_1)
- (38) **Roma populations has** had a historical presence both in Spanish and Catalan territories. (ERPP0722_4)
- (39) **All data was** analysed using a custom database and also with SPSS software. [...] **The data referring to the social environment show** some variations between families in terms of education and work level of the parents, family situation, immigration, and use of different languages, but without presenting extreme cases. (ERPP0722_5)

Example (39) is particularly interesting as the noun *data* occurs with both a singular and a plural verb (e.g., “all data was”, “the data show”). As stated in the *Macmillan Dictionary* (Rundell, 2007), corpus evidence shows that “*the data* + singular verb” is almost three times as frequent as “*the data* + plural verb”, but both uses are common and equally acceptable. However, the latter is more frequent in technical or scientific writing. What is important to note in the selected examples is the inconsistency found in the drafts; that is, some participants equally use both options in the same text, which may be due to unawareness of the preferred lexicogrammatical environment of this multiword unit in their respective disciplines.

4.3. Correlation between participants’ perceptions and drafts findings

This section explores whether there is a correspondence between the participants’ perception of the most challenging linguistic areas and the linguistic issues found in their drafts. In other words, we seek to analyse if participants are aware of the linguistic areas they are experiencing most problems with by comparing their perception with the linguistic issues identified in their drafts.

As per the results of the questionnaire², collocations seem to be one of the most challenging linguistic areas (see Table 3). This perception definitely correlates with the findings of the draft analysis carried out by the authors, which show that 35.10% (see Fig. 1) of the errors identified are related to collocational patterns. Both findings confirm results reported by research studies on the topic, which state that researchers find it particularly challenging to master the most prototypical domain-specific collocational patterns in English (Granger & Meunier, 2008; Laso & John, 2013; Nesselhauf, 2005; Paquot, 2008, among others).

| Linguistic issues | Mean of the participants’ answers |
|---|-----------------------------------|
| Collocations | 4.3/5 |
| Grammar | 4/5 |
| Misuse of Similar Lexical Items | 3.5/5 |
| Punctuation, wordiness and paragraphing | 3.1/5 |

Table 3. Mean of the participants’ answers to the questions related to linguistic issues.

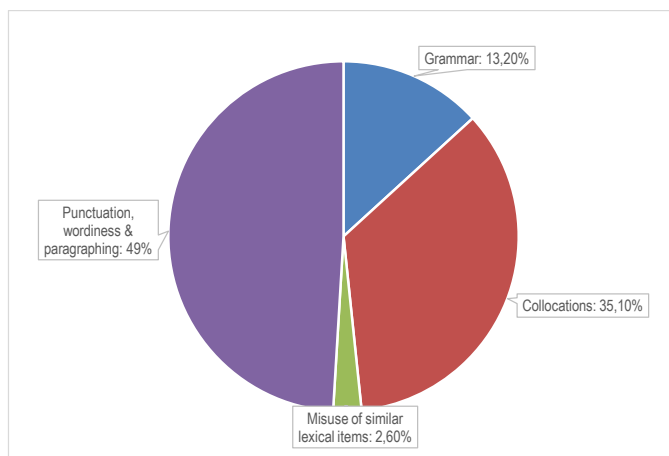


Figure 1. Most challenging linguistic areas identified in the drafts.

Interestingly, though, there is a clear mismatch between how participants perceive their performance regarding punctuation and paragraphing, and grammar and the errors highlighted by the analysis of their written production. As shown in Table 3, participants view grammar as the second most problematic linguistic area, whereas punctuation, wordiness and paragraphing get relegated to the bottom of the ranking. However, the study of the participants' writings reveals that punctuation, linking words, and paragraph organisation pose the most salient difficulties (49%) in their research papers, whereas grammar does not seem to be as problematic as they think (13.20%). Some grammar mistakes were spotted during the analysis, but they were restricted to the absence/misuse of anaphoric references, verb tense confusion, and lack of subject-verb agreement. Thus, the draft findings indicate that although some specific grammar issues seem to be problematic, their occurrence is not as remarkable as perceived by the participants. On the other hand, researchers seem to struggle with punctuation marks, which they either misuse or omit, and with the use of linking devices for intra- and inter-paragraph coherence. Finally, contrary to the participants' perception (3.5/5), misuse of similar lexical items does not seem to be a particularly difficult area, according to the percentage of errors found in the draft analysis (2.6%).

The results presented above do confirm previous studies which indicate that non-native writers perceive lexicogrammatical issues as more problematic

than other structure-related components, such as punctuation or the organization of their writing (Lee & Tajino, 2008; Liming & Lan, 2016; Okumuş, 2019).

5. Pedagogical implications

The predominance of English as the primary language for publishing in prestigious journals and conferences is a critical factor driving researchers to write in English. This requires EAP programmes that focus on helping EFL discourse communities achieve proficiency in the conventions and style of academic writing. Such programs should prioritize training on producing clear, concise, phraseologically competent and stylistically appropriate academic texts to aid researchers in advancing their careers and reaching broader audiences.

This paper leads to a series of specific recommendations, such as rounds of ERPP Workshops that could guide researchers through the writing process, enabling researchers to accurately identify their academic writing needs and areas for improvement and adopt comprehensive pre-writing practices that enhance the coherence and quality of their drafts.

The results of this study indicate that collocational patterns emerge as a major challenge, highlighting the need for focused instruction on domain-specific collocations. Teaching resources should include tailor-made exercises on common academic collocations, using resources such as the British National Corpus (BNC) and the Academic Collocation List (ACL). Instruction should also emphasize the role of L1 interference and provide strategies to mitigate its impact on L2 collocational accuracy.

Additionally, frequent misuse of similar lexical items, such as light verbs and collocates, underscores the need for targeted vocabulary instruction. Teaching should emphasize the nuances of seemingly interchangeable words and their proper contexts. Practical exercises could involve comparing and contrasting usage examples, providing clear distinctions between commonly confused pairs, and reinforcing appropriate contexts through repetition and application in writing tasks.

While participants identified grammar as a significant challenge³, the analysis revealed punctuation, wordiness, and paragraphing as more prevalent issues. This discrepancy suggests that EAP courses need to place a stronger

emphasis on these structural elements. Instruction should cover the correct use of punctuation marks, strategies to avoid wordiness, and effective paragraphing techniques. Practice activities could include editing exercises, peer reviews, and workshops focused on developing coherent and logically structured paragraphs.

The mismatch between perceived and actual difficulties highlights the importance of fostering self-awareness among researchers regarding their writing challenges. EAP training programs should incorporate feedback sessions and one-to-one consultations that could help bridge the gap between perception and reality, leading to more targeted and effective learning strategies.

In conclusion, the findings of this study underscore the necessity for comprehensive and targeted EAP instruction that addresses both linguistic and structural challenges in academic writing. By focusing on these areas, educators and EFL writers can be better equipped with the skills required to produce high-quality academic texts in English, which would ultimately aid their professional development and contribution to the global academic community.

6. Conclusions

This study has identified the scholarly writing challenges faced by a sample of Spanish EFL writers of ERPP, which have been compared against what the same Spanish EFL writers perceive as their weakest areas in relation to English writing production. The mismatch found between EFL writers' perception, and their needs has certainly direct implications not only for individual researchers, but also for EFL and ERPP instruction, which requires further attention.

Despite being a small-scale study, it is reasonable to claim that tailor-made writing seminars that cater to writers' clearly identified needs in relation to research writing skills (e.g. familiarization with prototypical specialized phraseology, inter- and intra-paragraph cohesion) are very useful. Such seminars are effective mechanisms for helping Spanish EFL writers become aware of their needs and cope with the high-pressure environment of academia.

This study has a number of limitations, which need to be taken into account. These preliminary findings, based on a small sample, have stressed the need

to extend this research to a larger number of participants, as well as to explore if targeting a wider range of research disciplines may reveal noticeable differences across disciplines. If so, this would raise awareness about which research fields and particular areas require more ERPP actions. Consequently, it would pave the way for more customised workshops that specifically address the academic writing needs of those scholars and tackle area-related issues.

The identified mismatch between the participants' perceptions and the challenging areas found in the drafts seems to indicate that a needs analysis based on participants' perceptions may not be the most useful tool when planning the teaching of academic writing. On the contrary, further analyses of this kind of participants' drafts is a much better indication of participants' needs. As suggested by Pérez-Llantada et al. (2011, p. 29), future research could gain from conducting textual analyses of successive drafts of papers intended for publication to identify specific linguistic challenges that institutional programs could address.

Article history:

Received 2 May 2023

Received in revised form 25 July 2024

Accepted 18 November 2024

References

- Academic Collocation List (n.d). <https://www.pearsonpte.com/teachers/academic-collocation>. Retrieved 1st May 2023.
- Ackermann, K., & Chen, Y. H. (2013). Developing the Academic Collocation List (ACL) – A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes*, 12(4), 235-247. <https://doi.org/10.1016/j.jeap.2013.08.002>
- Anthony, L. (2022). *AntConc* (Version 4.1.2). Waseda University. <https://www.laurenceanthony.net/software/antconc/>
- Arnbjörnsdóttir, B. (2019). Supporting Nordic scholars who write in English for research publication purposes. In J. N. Corcoran, K. Englander & L. Muresan (Eds.), *Pedagogies and policies for publishing research in English. Local initiatives supporting international scholars* (pp. 77-90). Routledge.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *The Longman grammar of spoken and written English*. Longman.
- Bocanegra-Valle, A. (2013). The perceived value of English for academic publishing among ESP multilingual scholars in Europe. *Journal of English for Specific Purposes at Tertiary Level*, 1(1), 5-25.
- Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348-393. <https://doi.org/10.1111/lang.12224>
- Burgess, S., Martín, P., & Balasanyan, D. (2019). English or Spanish for research publication purposes? Reflections on a critical pragmatic pedagogy. In J. N. Corcoran, K. Englander & L. Muresan (Eds.), *Pedagogies and policies for publishing research in English. Local initiatives supporting international scholars* (pp. 128-140). Routledge.
- Cargill, M., & Burgess, S. (2008). Introduction to the Special Issue: English for research publication purposes. *Journal of English for Academic Purposes*, 7(2), 75-76. <https://doi.org/10.1016/j.jeap.2008.02.006>
- Cargill, M., & Burgess, S. (2017). *Publishing*

research in English as an additional language: Practices, pathways and potentials. University of Adelaide Press.

Cargill, M., & O'Connor, P. (2013). *Writing scientific research articles*. Wiley-Blackwell.

Charles, M. (2007). Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes*, 6(4), 289-302. <https://doi.org/10.1016/j.jeap.2007.09.009>

Charles, M., & Hadley, G. (2022). Autonomous corpus use by graduate students: A long-term trend study (2009-2017). *Journal of English for Academic Purposes*, 56, 101095. <https://doi.org/10.1016/j.jeap.2022.101095>

Chen, M., & Flowerdew, J. (2018). A critical review of research and practice in data-driven learning (DDL) in the academic writing classroom. *International Journal of Corpus Linguistics*, 23(3), 335-369. <https://doi.org/10.1075/ijcl.16130.chen>

Corcoran, J. N. (2022). Reflections on the perceived longer-term impact of an ERPP course. *Journal of English for Research Publication Purposes*, 3(2), 169-197. <https://doi.org/10.1075/jerpp.21015.cor>

Corcoran, J. N., Englander, K., & Muresan, L. (Eds.) (2019). *Pedagogies and policies for publishing research in English: Local initiatives supporting international scholars*. Routledge.

Cotos, E. (2016). Computer-assisted research writing in the disciplines. In S. A. Crossley & D. S. McNamara (Eds.), *Adaptive educational technologies for literacy instruction* (pp. 225-242). Routledge.

Ellis, N. C. (2007). Learned attention in language acquisition: Blocking salience, and cue competition. Paper presented at *The Second European Cognitive Science Conference (EuroCogSci07)*, Delphi, Greece.

Eriksson, A. (2012). Pedagogical perspectives on bundles: Teaching bundles to doctoral students of biochemistry. In J. Thomas & A. Boulton (Eds.), *Input, process and product: Developments in teaching and language corpora* (pp. 195-211). Masaryk University Press.

Etherington, S. (2008). Academic writing and the disciplines. In P. Friedrich (Ed.), *Teaching academic writing* (pp. 26-58). Continuum.

Ferguson, G., Pérez-Llantada, C., & Plo, R. (2011). English as an international language of scientific publication: A study of attitudes. *World Englishes*, 30(1), 41-59. <https://doi.org/10.1111/j.1467-971X.2010.01656.x>

García Landa, L. G. (2006). Academic language barriers and language freedom. *Current Issues in Language Planning*, 7(1), 61-81. <https://doi.org/10.2167/cilp084.0>

Giraldo, F. (2019). An English for research publication purposes course: Gains, challenges, and perceptions. *GIST Education and Learning Research Journal*, 18, 198-220. <https://doi.org/10.26817/16925777.454>

Granger, S., & Meunier, F. (2008). Phraseology in language learning and teaching. Where to from here? In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective* (pp. 247-252). John Benjamins.

Hyland, K. (2004). *Genre and second language writing*. University of Michigan Press.

Hyland, K. (2008). As can be seen: lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21. <https://doi.org/10.1016/j.esp.2007.06.001>

Kaszubski, P. (2000). *Selected aspects of lexicon, phraseology and style in the writing of Polish advanced learners of English: A contrastive, corpus-based approach*. PhD dissertation, Adam Mickiewicz University, Poznań.

Laso, N. J., & John, S. (2013). An exploratory study of NNS medical writers' awareness of the collocational patterning of abstract nouns in medical discourse. *Revista Española de Lingüística Aplicada*, 26, 307-332.

Laso, N.J., & John, S. (2017). The pedagogical benefits of a lexical database (SciE-Lex) to assist the production of publishable biomedical texts by EAL writers. *Ibérica, Journal of the European Association of Languages for Specific Purposes*, 33, 147-172. <https://www.revistaiberica.org/index.php/iberica/article/view/161>

Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, 25(1), 56-75. <https://doi.org/10.1016/j.esp.2005.02.010>

Lee, H., Warschauer, M., & Lee, J. H. (2019). The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis. *Applied Linguistics*, 40(5), 721-753. <https://doi.org/10.1093/applin/amy012>

Lee, S., & Tajino, A. (2008). *Understanding students' perceptions of difficulty with academic writing for teacher development: A case study of the university of Tokyo writing program*. Kyoto University Research Information Repository.

Li, Y., & Flowerdew, J. (2020). Teaching English for

- Research Publication Purposes (ERPP): A review of language teachers' pedagogical initiatives. *English for Specific Purposes*, 59, 29-41. <https://doi.org/10.1016/j.esp.2020.03.002>
- Lillis, T., & Curry, M. L. (2010). *Academic writing in a global context: The politics and practices of publishing in English*. Routledge.
- Liming, L., & Lan, L. (2016). Noun phrase complexity in EFL academic writing: A corpus-based study of postgraduate academic writing. *The Journal of Asia TEFL*, 13(1), 48-65. <http://dx.doi.org/10.18823/asiatefl.2016.13.1.4.48>
- Mendoza, A., Oropeza, V., Rodríguez, D., Sobrevilla, Z., & Martínez, J. (2021). Challenges, feelings, and attitudes towards writing in ERPP in semi-periphery countries: The case of Mexican graduate students. *Journal of English for Research Publication Purposes*, 2(2), 129-159. <https://doi.org/10.1075/jerpp.21008.men>
- Muresan, L. M., & Pérez-Llantada, C. (2019). Research writing in English in a Romanian academic ecosystem: A case study of an experienced multiliterate researcher. In J. N. Corcoran, K. Englander & L. Muresan (Eds.), *Pedagogies and policies for publishing research in English. Local initiatives supporting international scholars*. (pp. 109-127). Routledge.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. John Benjamins.
- Okumuş, N. (2019). Student perceptions of difficulties in second-language writing. *Journal of Language and Linguistic Studies*, 15(1), 151-157.
- Paquot, M. (2008). Exemplification in learner writing: A cross-linguistic perspective. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 100-120). John Benjamins.
- Pérez-Llantada, C. (2014). Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. *Journal of English for Academic Purposes*, 14, 84-94. <https://doi.org/10.1016/j.jeap.2014.01.002>
- Pérez-Llantada, C., Ferguson, G., & Plo, R. (2011). "You don't say what you know, only what you can": The perceptions and practices of senior Spanish academics regarding research dissemination in English. *English for Specific Purposes*, 30(1), 18-30. <https://doi.org/10.1016/j.esp.2010.05.001>
- Rundell, M. (Ed.) (2007). *Macmillan English dictionary for advanced learners*. Macmillan Education.
- Sun, Y.C. (2007). Learner perceptions of a concordancing tool for academic writing. *Computer Assisted Language Learning*, 20(4), 323-343. <https://doi.org/10.1080/09588220701745791>
- Villagran, T., & Harris, P. R. (2009). Some key factors in medical writing. *Revista Chilena de Pediatría*, 80(1), 70-78. <http://dx.doi.org/10.4067/S0370-41062009000100010>
- Wang, Y., & Shaw, P. (2008). Transfer and universality: collocation use in advanced Chinese and Swedish learner English. *ICAME Journal*, 32, 201-232.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.

Natalia Judith Laso Martín is a Serra Hunter fellow in English Linguistics at the University of Barcelona. She holds a PhD in English Philology and is a member of the Lexicology and Corpus Linguistics Research Group (GRELIC). She specialises in learner corpus research and English for Research Publication Purposes (ERPP).

Elisabet Comelles Pujadas is a lecturer in the Department of Modern Languages and Literatures and of English Studies at the University of Barcelona. She holds a PhD in Cognitive Science and Language (UB). Her research interests include Natural Language Processing, Machine Translation, Machine Translation Evaluation and Corpus Linguistics.

NOTES

¹ Prototypical academic combinations are provided in square brackets.

² See range of answers provided in section 4.1.

³ To ensure that the participants shared our understanding of each linguistic issue category, examples of each category were provided and adequately explained when the questionnaire was distributed.

Appendix

English for Research Publication Purposes Workshop

Questionnaire 1

* Indicates required question

General Questions

1. Why do you write your research in English?*
2. How often do you write in English?*
3. Who do you write for? What is the potential readership of your papers?*
4. What journals do you usually publish in?*

When writing a scientific paper in English, what is your writing process like? Tick as many of these statements below which best reflect your writing process:

5. Before writing my paper *

Check all that apply.

- ☐ brainstorm ideas
- ☐ read relevant articles from journals
- ☐ write a plan of the paper
- ☐ draft short sections of the paper

6. In the writing stage*

Check all that apply.

- ☐ write the paper in my mother tongue and then translate it to English
- ☐ write the paper in English and look up a few unknown words and expressions
- ☐ write the paper mostly in one language (either mother tongue or English) and some parts (more than individual words) in the other language
- ☐ check the style of the journal I intend publishing in and try and write in the same style
- ☐ write my paper in a particular order
- ☐ don't write my paper in any particular order

7. If you write your paper in a particular order, which section do you typically begin with and which do you end with? Why?

8. When revising your paper*

Check all that apply.

- ☐ check for spelling and grammatical errors
- ☐ ask for an expert to check my writing before submitting my paper

What are your main difficulties when writing a scientific paper in English? Rank the following aspects in order of difficulty:

9. Specific terminology in English *

Mark only one oval.

Very easy

- 1 ☐
- 2 ☐
- 3 ☐
- 4 ☐
- 5 ☐

Very difficult

10. General word combinations (e.g., v+prep; adj+noun; adv+adj) most commonly used in this type of publications*

Mark only one oval.

Very easy

- 1 ☐
- 2 ☐
- 3 ☐
- 4 ☐
- 5 ☐

Very difficult

11. Distinction between similar pairs of words (e.g., do/ make; say/ tell; sensible/ sensitive; quick/ fast)*

Mark only one oval.

Very easy

- 1 ☐
- 2 ☐
- 3 ☐
- 4 ☐
- 5 ☐

Very difficult

12. Punctuation, wordiness and paragraphing *

Mark only one oval.

Very easy

1 ☐

2 ☐

3 ☐

4 ☐

5 ☐

Very difficult

13. Grammar issues (e.g., verb tenses, word forms, word order, etc.)*

Mark only one oval.

Very easy

1 ☐

2 ☐

3 ☐

4 ☐

5 ☐

Very difficult

14. Others (please specify)

Open questions

15. How is the dominance of English affecting your everyday work as a scholar who uses languages other than English and live/work in non-English dominant contexts?*

16. What pressures do you face in this enterprise of getting your research*

17. What barriers to publishing in English do you encounter?*

