FORM AND STRUCTURE OF LEXICAL BUNDLES
IN CROWDFUNDING WRITING:
A CORPUS-DRIVEN ANALYSIS

Alberto A. VELA-RODRIGO[1]

Abstract

In the context of science crowdfunding, the technological affordances of digital platforms allow researchers to make scientific contents accessible to lay publics and also engage them in various ways in scientific research. It is, therefore, of interest to uncover how formulaic language operates through the lens of lexical bundle analysis in the emerging crowdfunding proposal genre to know the manner scientists engage their audience. From the analysis of a corpus of scientific projects from Experiment.com, this study provides a description of the form and structure of lexical bundles and the manner they help this genre to reach its communicative goals. Preliminary results show a greater presence of clausal structures (e.g., is one of the; keep up the good) vs. phrasal types (e.g., the end of the; goal of this project), which indicates that the prose of the crowdfunding proposals moves away from the academic register and is closer to the conversational one, being less grammatically compressed. This fact reinforces the idea of the hybrid nature of this new genre.

1. Introduction

The close relationship between lexis and grammar has been discussed in a considerable number of publications on corpus linguistics, highlighting the importance of phraseology when studying the performance of a genre (e.g., Gries, 2008; Hoey, 2005; Hunston, 2002; Hunston and Francis, 2000; Partington, 1998; Römer, 2005 and 2009, among others). Among the phraseological studies of genres developed to date, those related to academic written genres have been the most common, though it must be said that such studies are relatively few compared to studies of other linguistic features in academic prose (Römer, 2009). However, a growing number of researchers in corpus linguistics have been paying attention to phraseology in these genres in the past two decades, with insightful approaches to different phraseological patterns, constructions, and multi-word units (Cortes, 2004; Barbieri, 2007; Hyland, 2008b; Ädel and Erman, 2012). In the context of Citizen Science and Open Science, the members of the scientific community, considered to be a discourse community insofar as its primary determinants are functional (communicative needs and goals) (Swales, 1990), the use of formulaic language has

[1] Alberto A. Vela-Rodrigo, Universidad De Zaragoza, vela@unizar.es.

mainly been investigated in relation to aspects of language and discourse style in academic prose. It is therefore of interest to uncover how formulaic language operates in emerging genres for public communication of science, such as the crowdfunding proposal, where language needs to be accessible and decipherable to non-specialist and lay audiences. Defined as a generic hybrid (Mehlenbacher, 2019), it combines the linguistic conventions of the traditional grant proposal with the affordances of the digital medium. Therefore, when a researcher uses a crowdfunding platform such as Experiment.com (https://experiment.com/) to raise funds, he or she will need to select those terms, words, and expressions and will also need to rely on the 'idiomaticity' or formulaicity of the language to communicate scientific contents clearly and effectively.

Despite the growing attention to emerging digital genres in the past decades, to the best of my knowledge, there are no research studies of phraseology about emerging genres of science communication on the Internet, such as that of crowdfunding proposals (with the exception of Pérez-Llantada, 2021 and Vela-Rodrigo, 2023). This is the research gap that this study aims to fill in. The study of the phraseology of this digital genre is vital in order to describe and understand its discoursal functioning, to characterize the genre and to assess its potential pedagogical implications. Broadly, the communicative purpose of this genre is that of educating citizens in science, informing them about issues of science and favouring crowdfunding (Pérez-Llantada, 2021). Using a corpus driven approach, in this study the aim has been to examine the structural forms of the recurring phraseology in crowdfunding projects, and it is done so through the lens of lexical bundle analysis. Defined as a sequence of three or more words that co-occur frequently in a particular register (Biber et al., 1999; Cortes, 2022), lexical bundles serve important discourse functions in both spoken and written language (Cortes 2004; Biber and Barbieri 2007). A structural description of lexical bundles will thus serve to uncover the formulaicity of the genre and its relationship with the communicative goals that the genre performs. The research questions will be the following:

RQ1. What are the form and structure of the most recurring lexical bundles in crowdfunding projects for science?

RQ2. More broadly, how do lexical bundles help characterizing the crowdfunding genre as a hybrid genre and how do they help reaching its communicative goals?

2. Literature review

The study of lexical bundles as a particular salient approach of formulaic language gained an important space with the affordances of specialized computer programs during the last two decades of the past century. These programs allowed an automatic corpus-driven analysis of recurring expressions (Altenberg and Eeg-Olofsoon, 1990), such as lexical bundles, which were first conceptualized in the *Longman Grammar of Spoken and Written English* (Biber et al., 1999). Since then, traditional written academic genres have been key in the study of lexical bundle use, specially

centred on research articles and their different constituent parts (Cortes, 2004, 2013; Wright, 2019). Other academic genres such as MA theses and doctoral dissertations across different disciplines have also been a subject of study (Biber, 2006; Hyland, 2008b), likewise textbooks and classroom discourse (Biber et al., 2004; Biber and Barbieri, 2007; Römer, 2010) or student/ expert academic writing (Chen and Baker, 2010).

For example, among the works dedicated to the phraseological analysis of research articles, Cortes (2004) compares the use of 4-word bundles in expert and non-expert academic writings for history and biology, classifying them structurally and functionally. Her findings indicate that the use of bundles by professional authors differs from that of non-expert writers, using the former more phrasal bundles and text organizers and avoiding stance expressions. Similar to these results and also following Biber et al.'s (1999) work, Römer (2010), in her analysis of a 4.5-million-word corpus of conversations and a 5.3-million-word corpus of research articles and academic books, realized that most lexical bundles in academic prose were phrasal rather than clausal. For their part, Hyland (2008b) and Pérez-Llantada (2014) have approached the formulaicity of the research article genre from the perspective of expert writing for academic and research publication purposes. The former author explores the forms, structures and functions of lexical bundles in four different disciplines, and together with other authors such as Dontcheva-Navratilova (2012) and her analysis of a corpus of diploma theses written by students conclude that a high degree of systematicity and an ample frequency of occurrence is basic for creating a successful academic communication.

Finally, digital genres have been approached from different perspectives, such as the studies carried out by Scotto di Carlo (2014) about the lexicon of TED talks or the use of lexical bundles in American blogs by Barbieri (2018) shedding light on the use on stance expressions and clausal verb phrases for the creation of self-expression and proximity with readers. This is a feature shared with conversational registers, reinforcing the idea of the existence of an increasing colloquialisation of written texts in digital media, also pointed out by other authors (Biber and Gray, 2016; Pérez-Llantada, 2021).

3. Methods

3.1 Corpus description

For the present study, a small-scale corpus of 50 proposals for crowdfunding environmental research was retrieved from Experiment.com, a New Yorker platform for crowdfunding science across different disciplines. The corpus totalled 140,478 words and considers the information of all sections or tabs in which this website is divided (Overview, Methods, Lab Notes and Discussion) [For a deeper description

of the rhetorical organization of the website see i.e., Mehlenbacher, 2019; Luzón and Pérez-Llantada, 2022; Vela-Rodrigo, 2023. Full details of the corpus of projects can be downloaded at the following link: https://mega.nz/file/CrZkUJiL#lxBgfOyD_o33 MkbdI4uJAx_PUyVIGWR_Lxn0NyfBl34]. The present study has been exclusively based on the analysis of this corpus, avoiding comparisons with other corpora and, therefore, eluding the Zipfian problem discussed by Bestgen (2019).

3.2 Analytical procedures and study design

This study used a 'radical corpus-driven approach' (Biber, 2009: 281), and it is entirely inductive. *Antconc 4.2.2* (Anthony, 2023) was used for the automatic identification of recurring bundles, and two types of frequency lists were created: a list of 3-word bundles and a list of 4-word bundles (available in Appendixes 1 and 2 downloadable at https://mega.nz/file/buJTiRCI#1LjVaDtRCGCJpLUYHKw HVFVZhv6ch4vPsWwcFAYjblU). By this means, overlapping bundles will be easily detected by comparing both lists in search for the number of 3-word bundles which are subsumed in longer bundles (Cortes, 2022). Also, the frequency-first principle was followed as a key criterion for the data collection (Biber et al., 1999; Biber, 2009).

As this is a corpus with a relatively small number of projects, it was necessary to be permissive with the cut-off parameters, since, otherwise, the results would be scarce when conducting the analysis across the different sections of the crowdfunding proposals. Thus, a cut-off point of frequency >5 and a range of at least in 5% of all the corpus texts were established. Raw counts are provided since, as Cortes (2015) and Bestgen (2019) argue, normalization of frequencies does not work well with data-driven formulaic expressions.

The design of the study was motivated by previous taxonomies for lexical bundles by Biber et al. (1999) and Hyland (2008b) for the identification of forms and types of structural units (phrasal/clausal) and types of words forming the bundle. For the analysis of types, this study has focused on 4-word bundles instead of 3-word bundles because the former is the most researched length in this type of studies (Chen and Baker, 2010: 32) and often subsume 3-word bundles.

4. Results

Table 1 shows the most frequent 3 and 4-gram bundles in the corpus. It is evident that many 3-word bundles such as *will be able* and *will be used* often expand into 4-word bundles, for example, *will be able to* and *will be used to*. This fact supports the idea that many four and five-word strings 'hold 3-word bundles in their structure' (Cortes, 2004). This process could be referred to as lexical bundle overlapping if we attend to previous studies (Biber et al., 2004; Hyland, 2008b; Cortes, 2022; Cortes

and Lake, 2023). Table 1 shows the 20 most recurring 3 and 4-word bundles retrieved from the corpus.

Table 1. Top 25 3 and 4-word bundles obtained in the analysed corpus

Freq.	3-word bundle	Freq.	4-word bundle
68	one of the	45	thank you so much
59	will be used	41	will be used to
57	as well as	23	thank you for your
55	be able to	22	will be able to
55	be used to	20	will allow us to
49	in order to	17	thank you very much
45	in the field	16	the end of the
45	thank you so	15	is one of the
45	you so much	13	a better understanding of
43	thank you for	13	of this project is
42	this project is	13	one of the most
38	we will be	13	can't wait to see
36	some of the	13	you so much for
32	for your support	12	so much for your
32	this project will	12	to be able to
31	a lot of	11	can be used to
31	end of the	11	i look forward to
30	i have been	11	thank you all for
29	thank you to	11	thanks so much for
28	the end of	10	as well as the

Results in Table 1 corroborate the trend already observed by Hyland (2008b) in a similar analysis, in his case based on a 3.5-million-word corpus of academic writing in articles, PhD dissertations, and master theses. The highly frequent 3-word lexical bundles typical of these academic genres (*in order to, in terms of, one of the*) are also especially frequent in the corpus of crowdfunding proposals (Table 1) (*in order to [49], in terms of [11] one of the [68]* respectively). [For a complete list of overlapping bundles see https://mega.nz/file/frYR1BpJ#MnzwXzDuy BcvjFv2dzLYETVrTu7lr5eGjWpEM2joys0].

Regarding the structures of the 93 4-word lexical bundles retrieved from the corpus, only 9 are complete structural units [*thank you so much; thank you very much; thank you to all; thanks for your support; i have been working; thank you to everyone; a huge thank you; for the first time; excited about this project*], representing 9.67% of the total (see Table 2). Therefore, and consistent with previous studies on bundles, the majority are either parts of phrases or clauses with other fragments embedded. On the other hand, lexical bundles have strong grammatical correlates. This means that many clausal bundles, for example, can incorporate verb phrase fragments, such as *it's going to be* (Biber, 2009). This allowed me to group them into several basic structural categories too (Biber et al., 1999; Cortes, 2004; Hyland 2008a, 2008b;

Pérez-Llantada, 2021). Grammatical correlates can vary according to the situational register. For example, most lexical bundles in conversation happen to be clausal, of the type (pronoun) + Verb + (complement), such as *thank you so much, I look forward to, you have any questions*. In contrast, in academic prose 60% of the bundles are phrasal, parts of noun phrases or prepositional phrases (Biber et al., 1999), (e.g., *the end of the, of this project is*). The hybrid nature of the crowdfunding genre is also reflected in this aspect, as Table 2 shows.

The results offered after the structural classification of the 4-word-bundle list show there is an ample presence of both types in the corpus, with an interesting greater presence of clausal types (66 bundles, representing 63.45 % of all bundle types) which incorporates verb phrase fragments (e.g., *will allow us to, will be able to, we would like to, I am excited to*). Phrasal types (34 bundles, 36.55% of the total) are mostly formed by parts of noun phrases (e.g., *our understanding of the, the top of the, huge thank you to*) or prepositional phrases (e.g., *over the course of, as a result of, at the same time*). Regarding structural subtypes, 44 correspond to verb phrase fragments (47.31%), 21 to noun phrase fragments (22.58%), also 21 to prepositional phrase fragments (22.58%) and 7 to dependent clause fragments (7.53%).

Table 2. Main structures of 4-word bundles in the corpus

PHRASAL TYPES (36.55%)	the end of the; goal of this project; a better understanding of; of this project is; one of the most, as well as the, has the potential to; over the course of; up the good work; at the end of; by the end of; for all of your; for your support we; to our understanding of; as a result of at the same time; for the first time; for this project will; good luck with the; of the history of; of the project this; of this project we; with the help of; good luck with your; luck with your research; the results of this; goal of this project; the rest of the; the top of the; a part of this; all of your support; better understanding of the; much for your support; our understanding of the; the goal of this.
CLAUSAL TYPES (63.45%)	thank you so much; is one of the; keep up the good, if you have any; stay tuned for more; thanks for your support; thank you to everyone; will be used to; thank you for your, we are able to; will be used for; will be able to; can't wait to see; you so much for; so much for your; to be able to; can be used to; i look forward to; thank you all for; thanks so much for; will allow us to; thank you very much; be a part of; thank you to all; to all of you; to better understand the; we were able to; this project is to; this project will be; we will be able; huge thank you to; i have been working; i would like to; let me know if; this is a great; this project i am; to do this we; to support this project; i am currently a; i am excited to; in the next few; can't wait to hear; that we will be; i am happy to; the project we will; to everyone who has; we would like to; will allow me to; again for your; with your help we; you have any questions; a huge thank you; and i hope you; be used as a; can be used as; again for your support; all of you for

The fact that the percentage of clausal bundles is significantly higher than that of phrasal bundles seems to highlight that the prose of the crowdfunding proposals moves away from the academic and is closer to the conversational register, being less grammatically compressed, that is, less "noun-centric". Therefore, this suggests that the genre of crowdfunding moves away from its traditional antecedent, the research grant proposal, which usually presents high grammatical compression, employing embedded phrases (Biber and Gray, 2010). This digital genre adopts characteristics of conversation, which build a more colloquial tone that creates proximity to a wider and more diversified audience with a more significant use of fuller dependent clauses than embedded ones (e.g., [1] *We have reached 75% support in under a week! This is all thanks to our backers and those that have shared our project to spread the word.* [DOI: 10.18258/11434]). The funders of the project are no longer a team of expert academics from government agencies but an open and global public.

The next stage of this analysis is about exploring the structural characteristics of these lexical bundles according to their grammatical structure. Table 3 displays a categorization adapted from Biber et al. (1999) and Hyland (2008a) with some new categories according to Lorenzo (2011), namely verb phrases + personal pronoun *we*, other noun phrases, other prepositional phrases and other verbal fragments, to collect all those bundles that do not fit in any of the previous classifications.

Table 3. Main structures of bundles (%) in the corpus (adapted from Biber et al., 1999, and Hyland, 2008a)

Noun phrase + *of*	9.7%	the end of (the), a better understanding of (the), the rest of (the), the top of (the), our understanding of (the), the goal of (this project), to our understanding (of), the results of (this)
Noun phrase + other postmodifier fragment	5.27%	thanks so much for, huge thank you to, thanks again for (your), good luck with (the), good luck with (your),
Other noun phrase	8.5%	one of the most, again for your support, all of your support, much for your support, the goal of this, with your help we, a huge thank you, luck with your research
Prepositional phrase + *of*	7.52%	over the course of, at the end of, by the end of, as a result of, of the history of, with the help of, you to all of
Other prepositional phrase	15.12%	of this project is, you so much for, so much for your, as well as the, to all of you, up the good work, for all of your, for your support we, all of you for, at the same time, for the first time, for this project will, of the project is, of this project we, in the next few

Verb or adjective + *to* infinitive clause fragment	24.18%	will be able to, will be used to, will allow us to, to be able to, can be used to, I look forward to, to all of you, to better understand the, we were able to, has the potential to, this project is to, huge thank you to, I would like to, thank you to everyone, I am happy to, we would like to, thank you to everyone, I am happy to, we would like to, will allow me to, you to all of, I am excited to, can't wait to hear, to do this we, to support this project
that + verb/noun phrase fragment	1.07%	that we will be
Anticipatory *it/this* structure	4.3%	this project is to, this project will be, this is a great, this project I am
Verb phrase with personal pronoun *we*	4.3%	we were able to, we are able to, we will be able, we would like to
Other verbal fragment	20.28%	I am currently a, excited about this project, the project we will, can be used as, be used as a, and I hope you, you have any questions, let me know if, I have been working, thanks for your support, stay tuned for more, if you have any, will be used for, keep up the good, be a part of, thank you all for, is one of the, thank you very much, thank you for your, thank you so much

As it can be seen from the above results lexical bundles presenting a *verb or adjective + to infinitive clause fragment* structure are the most frequent, accounting for almost a quarter of all target bundles in the corpus (24.18%). They can be simple *to* clauses or clauses preceded by predicative adjectives or verb phrases (Lorenzo, 2011). In the case of bundles featuring predicative adjectives + *to*-clause, they express ability and/or likelihood, whereas if the *to*-infinitive clause is preceded by a verb phrase it tends to express previous findings or accepted facts, as also reported in previous studies (Biber et al., 1999; Hyland, 2008a; Lorenzo, 2011). As shown in example 2, the lexical bundle *we will be able* is used to express abilities that come as a consequence of previous actions

[2] We will return to Ohio and perform detailed chemical and physical analyses of the rocks and fossils, after which **we will be able** to reconstruct the past environment, climate, cause of death, and mode of preservation of the fossils [DOI: 10.18258/6861]

This result is not unexpected considering the rhetorical structure of the crowdfunding proposal genre, where showing the ability to carry out the project to be financed is vital. At the same time, reference is usually made to achievements throughout the

project (not so much to previous achievements and previous projects), which are usually presented in the Lab Notes tab. On the other hand, without entering into the "mixed bags" of *Other verbal fragments* or *Other prepositional phrases*, it is interesting to highlight that *noun phrase fragments + of* reach roughly 10% of all lexical bundles in the corpus. This typology together with the pattern *Noun phrase + other postmodifier* fragment and *Other noun phrases* comprise almost 25% of all the target bundles which demonstrates an important presence of academic writing features in the crowdfunding genre, with nouns playing not a central role but still significant. In his analysis of lexical bundles on four different scientific disciplines Hyland (2008a) reported a presence of 24.4% of lexical bundles showing the structure of Noun phrase + of, a 10.6% of Noun phrases + other modifications, which adds up 35% of all the bundles he retrieved from his corpus of research articles, doctoral dissertations and Masters' theses. This figure differs partially from the near 25% of the same sort of lexical bundles from my corpus, which in the same way indicates that the crowdfunding genre is less noun-centric, and therefore not so academic. In any case, the substantiation characteristic of the academic record (Swales, 2008) is nuanced by another type of substantiation through the expression of gratitude, which is not related to academic genres but to informal conversational environments (e.g., *a huge thank you, luck with your research, all of your support*, etc.)

Also remarkable is the use of bundles made up of prepositional phrases embedded with the particle *of* (7.52%), since they can signify abstract or logical relationships between propositions and ideas (Biber et al., 1999; Hyland, 2008b; Lorenzo, 2011), such as the preposition *by* associated with the timeline in the Overview section (e.g., *by the end of*), the preposition *over* with processes (*over the course of*) or the preposition *at* to introduce measurements (e.g., *at the end of*), as in the examples below.

[3] Analysis should be completed **by the end of** June with presentation of results by July 2018 and publication in a peer-reviewed journal. [DOI: 10.18258/10428]

[4] We will assemble air quality kits to collect data on PERC and radon levels. We will test 15 homes **over the course of** three weeks. [DOI: 10.18258/5329]

[5] **At the end of** the project, we will have a data set that will provide insight into how effectively greening initiatives cool Baltimore. [DOI: 10.18258/7455]

5. Discussion

My main purpose in this study has been to explore the phraseological profile (or formulaicity) that characterizes the corpus of crowdfunding projects in environment and sustainability. The methodological approach opted for has been the analysis of lexical bundles, which has also allowed me to explore their main forms and

structures. More broadly, it was my aim to understand how lexical bundles help characterizing the crowdfunding genre as a hybrid genre and how they help reaching its communicative goals.

The findings of the present study have shown that the most frequent 4-word bundles coincide, partially, with the most frequent patterns reported in previous studies on academic discourse (Biber, 2009), which demonstrate that the crowdfunding genre has a common background with its printed antecedent, the grant proposal. There are other aspects that point out towards a continuity of a certain traditional academic nature in this new genre, too, such as the presence of many 3-word bundles typical of academic prose and the same sharp drop in frequency already observed for 3 word-units when expanded to 4 and 5 words for research articles (Hyland, 2008a). In this sense, these results also show that the formation of the lexical bundles through their word extension does not influence, in any way, the typology of the genre or the treated discipline.

On the other hand, the fact that almost 2/3 of all analysed lexical bundles were clausal units strongly suggests that the discourse style is not grammatically compressed, or in other words, is less "noun-centric" than academic writing. This was seen in the high frequency of lexical bundles consisting of verb phrase structures, sometimes with several units typical of conversational contexts (*thank you very much, can't wait to see, thanks so much for, I look forward to*). As shown in previous studies, lexical bundles in academic writing are predominantly phrasal (Hyland, 2008b; Biber and Gray, 2010; Ädel and Erman, 2012), therefore since findings here suggest little syntactic elaboration and scarce formality, the crowdfunding genre seems not to rely exclusively on phraseological features especially frequent in academic writing to achieve its communicative goals. At the same time, formulaic language also suggests that these projects do own some features associated with the formal register of academic prose, for example when the researchers wish to objectively show data and their collection criteria, or when the budget planning and the research methodology are presented using descriptive phraseological units. All this seeks to show the importance of the research project and an impression of proficiency in the same way that the traditional grant proposal does in order to obtain funds from funding agencies (Biber et al., 2011; Salazar, 2014; Hyland and Jiang, 2017). Therefore, both features (little grammar compression and the presence of a certain academic prose) seem to support the view of this genre as 'hybrid'.

Regarding the structure of the 4-word bundles, most of them are not complete structural units but a bridge among two different units, that is, the last word of the bundle is often the first element of the following structure. Complete structural units found among these bundles in the corpus seem to be halfway between those reported for academic prose and for conversation in other previous studies (Biber et. al., 1999), also reinforcing the idea of the hybridity of the genre. A genre that, although new, seems to be stabilizing at least at the level of phraseology.

Other particularities that make this genre different from that of the traditional academic genre of grant proposals can also be observed in the grammatical structure of the lexical bundles found in the corpus. With a lower presence of lexical bundles containing noun phrases, the genre is, obviously, less academic in its nature if compared to the traditional grant proposal, or than dissertations and Masters' theses, which are more noun-centric (Hyland, 2008b). The reason is that the audiences of the latter genres (specialists in the field) are different compared to the wider public audience of the crowdfunding proposals. This study has shown that in the crowdfunding proposals the lexical bundles presenting a *verb (or adjective) + to clause fragment* structure are very frequent, helping to express ability to carry out a project or a set of steps for the research, as those present in the Lab Notes section. This is important since researchers need to engage a diversified audience in the project, with many of those people being non-specialists who expect professionality, effort and competence in order to donate funds. Moreover, this is in line with Luzón and Pérez-Llantada (2022: 129), who have already noted that the predominant mode in the Lab Notes section is the verbal mode.

In concluding, this corpus driven approach has shown that phraseology is key to understanding how texts are constructed to reach the communicative goals of the crowdfunding proposals, namely to educate in science and prompt donation. In light of the results of the present analysis, phraseology for crowdfunding projects in Experiment.com points to the merge of two different traditions in the same genre, with bundles typical of formal academic prose co-occurring with bundles that recur in conversation. This fact reinforces the idea of the crowdfunding proposal genre as a hybrid genre. This study also elicits corpus-informed discussion about relevant features of lexical bundles, namely structural constituency and typology, as well as raising awareness to their role for creating coherence in the specific case of Experiment.com.

On the other hand, the comparative analysis of the use of lexical bundles among projects on different crowdfunding websites and, especially, of their functions, can be a prospective area of research. In this way, the suitability of the use of some bundles over others could be established to be more effective in raising funds. In turn, knowing these aspects together with the data we deal with in this study can help researchers learn to use lexical bundles more effectively.

6. Acknowledgements

7. End-Notes

[1] A complete analysis on formulaic variation (bundles with a preceding or following variable slot) was carried out by Biber (2009) in his analysis of a 4.5-million-word corpus of AmE conversation, and a 5.3-million-word corpus of academic prose (research articles and academic books).

[2] Römer (2010) introduced a new measure (variant / p-frame ratio or VPR) as part of a step 2 in her PP model (phraseological profile model), which not only examines the types but also the degrees of internal variation of common word sequences.

References and bibliography

Ädel, A. and B. Erman. 2012. "Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach", in *English for Specific Purposes*, 31 (2): 81-92. https://doi.org/10.1016/j.esp.2011.08.004.

Anthony, L. 2023. *AntConc* (Version 4.2.2) [Computer Software], Tokyo, Japan: Waseda University. Available at https://www.laurenceanthony.net/software.

Altenberg, B. and M. Eeg-Olofsson. 1990. "Phraseology in spoken English: presentation of a project", in Aarts, J. and W. Meijs (Eds.), *Theory and Practice in corpus Linguistics*, Amsterdam: Rodopi: 1-26.

Barbieri, F. 2007. "Lexical bundles in University spoken and written registers", *English for Specific Purposes* 26 (3): 263-286. https://doi.org/10.1016/j.esp.2006.08.003

Barbieri, F. 2018. "'I don't want to and don't get me wrong': Lexical bundles as a window to subjectivity and intersubjectivity in American blogs", in Kopaczyk J. and J. Tyrkkö (Eds.), *Applications of Pattern-Driven Methods in Corpus Linguistics,* Amsterdam: John Benjamins: 251-276.

Bestgen, Y. 2019. "Comparing lexical bundles across corpora of different sizes: The Zipfian problem", in *Journal of Quantitative Linguistics*, 27 (3): 272-290. https://doi.org/10.1080/09296174.2019.1566975.

Biber, D. 2006. *University Language: A Corpus-Based Study of Spoken and Written Registers*, Amsterdam: John Benjamins.

Biber, D. 2009. "A corpus-driven approach to formulaic language in English. Multi-word patterns in speech and writing", in *International Journal of Corpus Linguistics*, 14(3): 275-311. https://doi.org/10.1075/ijcl.14.3.08bib.

Biber, D., Conrad, S. and V. Cortes. 1999. *Longman Grammar of Spoken and Written English,* New Jersey: Pearson Education Ltd.

Biber, D., Conrad, S. and V. Cortes. 2004. "If you look at …: Lexical bundles in University teaching and textbooks", in *Applied Linguistics*, 25, 3: 371-405. https://doi.org/10.1093/applin/25.3.371.

Biber, D., Gray B. and K. Poonpon. 2011. "Should we use characteristics of conversation to measure grammatical complexity in L2 writing development?", in *TESOL Q.* 45(1): 5-35.

Biber, D. and F. Barbieri. 2007. "Lexical bundles in University spoken and written registers", in *English for Specific Purposes,* 26: 263-86. https://doi:10.1016/j.esp.2006.08.003.

Biber, D. and B. Gray. 2010. "Challenging stereotypes about academic writing: Complexity, elaboration, explicitness", in *Journal of English for Academic Purposes,* 9(1): 2-20. https://doi:10.1016/j.jeap.2010.01.001.

Biber, D. and B. Gray. 2016. "The competing demands of popularisation vs. economy: Written language in the age of mass literacy", in Nevalainen T. and E. C. Traugott (Eds.), *The Oxford Handbook of the History of English,* Oxford: Oxford University Press: 314-328.

Chen, Y. and P. Baker. 2010. "Lexical bundles in L1 and L2 academic writing", in *Language Learning and Technology,* 14 (2): 30-49. https://doi.org/ 10125/44213.

Cortes, V. 2004. "Lexical bundles in published and student disciplinary writing: Examples from History and Biology", in *English for Specific Purposes* 23 (4): 397-423. https://doi.org/10.1016/j.esp.2003.12.001.

Cortes, V. 2013. "The purpose of this study is to: Connecting lexical bundles and moves in research article introductions", in *Journal of English for Academic Purposes*, 12 (1): 33-43.

Cortes, V. 2015. "Situating lexical bundles in the formulaic language spectrum". In V. Cortes and E. Csomay (Eds.), *Corpus-based Research in Applied Linguistics: Studies in Honor of Doug Biber*, John Benjamins: 197-216. https://doi.org/10.1075/scl.66.09cor

Cortes, V. 2022. "Lexical bundles in academic writing", in Jablonkai R. and E. Csomay (Eds.), *The Routledge Handbook of Corpora in Language Teaching and Learning,* New York: Routledge: 220-233. https://doi.org/10.4324/ 9781003002901.

Cortes, V. and W. Lake. 2023. "LBiaP: A solution to the problem of attaining observation independence in lexical bundle studies", in *International Journal of Corpus Linguistics,* 28 (2): 263-277. https://doi.org/10.1075/ijcl.21100.cor.

Dontcheva-Navratilova, O. 2012. "Lexical bundles in academic texts by non-native speakers", in *Brno Studies in English*, 38 (2): 37-58.

Gries S. T. 2008. "Phraseology and linguistic theory", in Granger S. and F. Meunier (Eds.), *Phraseology: An Interdisciplinary Perspective*, Amsterdam: John Benjamins: 3-25.

Hoey, M. 2005. *Lexical Priming: A New Theory of Words and Language,* New York: Routledge.

Hunston, S. 2002. *Corpora in Applied Linguistics,* Cambridge: Cambridge University Press.

Hunston, S. and G. Francis. 2000. *Pattern Grammar. A Corpus-Driven Approach to the Lexical Grammar of English,* Amsterdam: John Benjamins. https://doi.org/10.1075/scl.4.

Hyland, K. 2008a. "Genre and academic writing in the disciplines", in *Language Teaching*, 41: 543-562.

Hyland, K. 2008b. "As can be seen: Lexical bundles and disciplinary variation", in *English for Specific Purposes*, 27 (1): 4-21. https://doi.org/0.1017/S0261444808005235

Hyland, K. and K. Jiang. 2017. "Is academic writing becoming more informal?" *English for Specific Purposes*, 45: 40-51. https://doi.org/10.1016/j.esp.2016.09.001

Lorenzo, D. 2011. *Lexical Bundles in Scientific English: A Corpus-Based Study of Native and Non-Native Writing*, Barcelona: Universidad de Barcelona (PhD work / tesis doctoral).

Luzón, M. J. and C. Pérez-Llantada. 2022. *Digital Genres in Academic Knowledge Production and Communication: Perspectives and Practices,* Bristol: Multilingual Matters.

Mehlenbacher, A. R. 2019. *Science Communication Online: Engaging Experts and Publics on the Internet,* Columbus: The Ohio State University Press.

Partington, A. 1998. *Patterns and Meanings. Using Corpora for English Language Research and Teaching,* Amsterdam: John Benjamins.

Pérez-Llantada, C. 2014. "Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage", in *Journal of English for Academic Purposes*, 14: 84-94.

Pérez-Llantada, C. 2021. "Grammar features and discourse style in digital genres: The case of science-focused crowdfunding projects", in *Revista Signos* 54(105): 73-96. http://orcid.org/0000-0003-4052-1321.

Römer, U. 2005. *Progressives, Patterns, Pedagogy: A Corpus-Driven Approach to English Progressive Forms, Functions, Contexts and Didactics,* Amsterdam: John Benjamins.

Römer, U. 2009. "The inseparability of lexis and grammar: Corpus linguistic perspectives", in *Annual Review of Cognitive Linguistics,* 7 (1): 140-162. https://doi.org/10.1075/arcl.7.06rom.

Römer, U. 2010. "Establishing the phraseological profile of a text type: The construction of meaning in academic book reviews", in *English Text Construction*, 3(1): 95-119. https://doi.org/10.1075/etc.3.1.06rom.

Salazar, D. 2014. *Lexical Bundles in Native and Non-native Scientific Writing. Applying a Corpus-based Study to Language Teaching,* Amsterdam: John Benjamins.

Scotto di Carlo, G. 2014. "The role of proximity in online popularisations: The case of TED talks", in *Discourse Studies*, 16 (5): 591-606. https://doi.org/ 10.1177/1461445614538565.

Swales, J. M. 1990. *Genre Analysis, English in Academic and Research Settings*, Cmbridge: Cambridge University Press.

Vela-Rodrigo, A. Á. 2023. "A lexical bundle analysis of art-related crowdfunding projects", in *Ibérica, Journal of the European Association of Languages for Specific Purposes*: 321–349. https://doi.org/10.17398/2340-2784.46.321.

Wright, H. R. 2019. "Lexical bundles in stand-alone literature reviews: Sections, frequencies, and functions", in *English for Specific Purposes*, 54: 1-14.

The author
Alberto A. Vela-Rodrigo is Assistant Lecturer at the Department of English and German Philology, University of Zaragoza, where he has taught courses of History and culture of the United Kingdom, English for oral and written communication in different fields of knowledge and Learning-acquisition of English as a foreign language. His main research interests include genre analysis, English for Academic Purposes and digital genres in science communication. Some of his latest publications can be found in international journals such as *Ibérica: Journal of the European Association of Languages for Specific Purposes* (AELFE), *Complutense Journal of English Studies*, and *Language Value*.