EUROPHRAS 2022

MALAGA

Computational and Corpus-based Phraseology

Proceedings of the International Conference
EUROPHRAS 2022

(short papers, posters and MUMTTT workshop contributions)

28-30 September, 2022
Malaga, Spain

ORGANISERS

SPONSORS

ISBN 978-954-452-080-9

2022. INCOMA Ltd. Shoumen, BULGARIA

This document is available at http://europhras.com/2022/publications/

Editors of the Proceedings

Gloria Corpas Pastor
Maria Kunilovskaya
Rocío Caro Quintana
Ruslan Mitkov

Organisers:

Europhras 2022 was jointly organised by the European Association for Phraseology (Europhras), the University of Malaga (Research Group in Lexicography and Translation), Spain, the University of Wolverhampton (Research Group in Computational Linguistics), United Kingdom, and the Association for Computational Linguistics, Bulgaria.

Conference Co-Chairs:

Gloria Corpas Pastor, University of Malaga, Spain
Ruslan Mitkov, University of Wolverhampton, UK

Programme Committee:

Margarita María Alonso Ramos, University of A Coruña, Spain
María Belén Alvarado Ortega, University of Alicante, Spain
Verginica Barbu Mititelu, Romanian Academy, Romania
Ignacio Bosque, Complutense University of Madrid, Spain
María Luisa Carrió-Pastor, Polytechnic University of Valencia, Spain
Anna Čermáková, University of Cambridge, United Kingdom
Parthena Charalampidou, Aristotle University of Thessaloniki, Greece
Ken Church, Baidu
Jean-Pierre Colson, Université Catholique de Louvain, Belgium
Dmitrij Dobrovolskij, Russian Language Institute, Russian Federation
Peter Ďurčo, University of St. Cyril and Methodius, Slovakia
Natalia Filatkina, University of Hamburg, Germany
Elizaveta Goncharova, National Research University Higher School of Economics, AIRI
María Isabel González Rey, University of Santiago de Compostela, Spain
Stefan Gries, University of California, United States of America
Enrique Gutiérrez Rubio, Palacký University Olomouc, Czech Republic
Kleanthes K. Grohmann, University of Cyprus, Cyprus
Amal Haddad Haddad, University of Granada, Spain
Miloš Jakubíček, Sketch Engine
Eva Lucía Jiménez-Navarro, University of Cordoba, Spain
Cvetana Krstev, University of Belgrade, Servia
Natalie Kübler, Université Paris Cité, Grance
Maria Kunilovskaya, University of Wolverhampton, United Kingdom
Ljubica Leone, Lancaster University, United Kingdom
Óscar Loureda Lamas, Heidelberg University, Germany
Elvira Manero Richard, University of Murcia, Spain
Ramón Martí Solano, University of Limoges, France
María del Carmen Mellado Blanco, University of Santiago de Compostela, Spain
Flor Mena Martínez, University of Murcia, Spain
Pedro Mogorrón Huerta, University of Alicante, Spain

Johanna Monti, “L’Orientale” University of Naples, Italy
Esteban Tomás Montoro del Arco, University of Granada, Spain
Inés Olza Moreno, University of Navarra, Spain
Adriane Orenha Ottaiano, São Paulo State University, Brazil
Antonio Pamies Bertrán, University of Granada, Spain
Rozane Rebechi, Federal University of Rio Grande do Sul, Brazil
María Ángeles Recio Ariza, University of Salamanca, Spain
Ute Römer, Georgia State University, United States of America
Leonor Ruiz Gurillo, University of Alicante, Spain
Kathrin Steyer, University of Mannheim, Germany
Joanna Szerszunowicz, University of Bialystok, Poland
Yukio Tono, Tokyo University of Foreign Studies, Japan
Agnès Tutin, University of Grenoble Alpes, France
Aline Villavicencio, Federal University of Rio Grande do Sul, Brazil, and University of Sheffield, United Kingdom
Tom Wasow, Stanford University, United States of America
Eric Wehrli, University of Geneva, Switzerland
Michael Zock, Laboratoire d’Informatique Fondamentale de Marseille, France

Additional Reviewers:

Dayana Abuin Rios, University of Wolverhampton, United Kingdom
Rocío Caro Quintana, University of Wolverhampton, United Kingdom
Isabel Durán, University of Malaga, Spain
Richard Evans, University of Wolverhampton, United Kingdom
Emma Franklin, University of Wolverhampton, United Kingdom
Carlos Manuel Hidalgo Ternero, University of Malaga, Spain
Nieves Jiménez Carra, University of Malaga, Spain
Alfiya Khabibullina, University of Wolverhampton, United Kingdom
Lilit Kharatyan, University of Wolverhampton, United Kingdom
Ruslan Mitkov, University of Wolverhampton, United Kingdom
Daria Sokova, University of Wolverhampton, United Kingdom

Invited Speakers:

Aline Villavicencio, Federal University of Rio Grande do Sul, Brazil, and University of Sheffield, United Kingdom
Jean-Pierre Colson, Université Catholique de Louvain, Belgium
María del Carmen Mellado Blanco, University of Santiago de Compostela, Spain
Miloš Jakubíček, Sketch Engine

Organising Committee:

University of Malaga

Presentación Aguilera Crespillo
Marta Alcaide Martínez
Rosario Bautista Zambrana
Isabel Durán Muñoz
J. Alejandro Fernández Sola

Mahmoud Gaber
Rut Gutiérrez Florido
Carlos Manuel Hidalgo Ternero
Hanan Saleh Hussein
Adriana Iglesias Lara
Francisco Javier Lima Florido
Gema Lobillo Mora
Araceli Losey León
Jorge Lucas Pérez
Luis Carlos Marín Navarro
Desiré Martos García
Laura Noriega Santiáñez
Laura Parrilla Gómez
Míriam Pérez Carrasco
Encarnación Postigo Pinazo
María del Pilar Rodríguez Reina
Juan Antonio Sánchez Muñoz
Fernando Sánchez Rodas
Míriam Seghiri Domínguez
Cristina Toledo Báez

University of Wolverhampton

Dayana Abuin Rios
Isuri Anuradha
Anastasia Bezobrazova
Rocío Caro Quintana
Ana Isabel Cespedosa Vázquez
Amal El Farhmat
Suman Hira
Alfiya Khabibullina
Lilit Kharatian
Maria Kunilovskaya
Gabriela Llull
Kamshat Saduakassova
Kanishka Silva
Daria Sokova

Association for Computational Linguistics (Bulgaria)

Nikolai Nikolov

MUMTTT 2022 Workshop Chairs

Gloria Corpas Pastor, Universidad de Málaga, Spain
Ruslan Mitkov, University of Wolverhampton, United Kingdom
Johanna Monti, Università degli Studi di Napoli “L’Orientale” Italy
Maria Pia di Buono, Università degli Studi di Napoli “L’Orientale” Italy

MUMTTT 2022 Programme Committee

Giuseppe Attardi, University of Pisa
Verginica Barbu Mititelu, Romanian Academy Research Institute for Artificial Intelligence
Jean-Pierre Colson, Université catholique de Louvain
Anna Beatriz Dimas Furtado, University of Wolverhampton
Federico Gaspari, University for Foreigners "Dante Alighieri"
Amal Haddad Haddad, University of Granada
Philipp Koehn, The Johns Hopkins University
Judyta Mężyk, Paris-Est Créteil University and University of Silesia in Katowice
Pavel Pecina, Charles University
Éric Poirier, Université du Québec à Trois-Rivières
Carlos Ramisch, Aix Marseille University
Max Silberztein, Université de Franche-Comté
Kathrin Steyer, Institut für Deutsche Sprache, Mannheim
Beata Trawinski, Institut für Deutsche Sprache, Mannheim
Agnes Tutin, Université Grenoble Alpes

MUMTTT 2022 Organising Committee

Gennaro Nolano, Università degli Studi di Napoli "L'Orientale" Italy
Giulia Speranza, Università degli Studi di Napoli "L'Orientale" Italy
Khadija Ait ElFqih, Università degli Studi di Napoli "L'Orientale" Italy

Table of Contents

Multi-word Units in Machine Translation and Translation Technology (MUMTTT 2022) workshop contributions

Long Word Sequences in the Discourse of Adventure Tourism

Eva Lucía Jiménez-Navarro[1[0000-0001-9377-6921]] and Isabel Durán-Muñoz[2[0000-0002-6795-498X]]

[1,2] Department of English and German Studies, Universidad de Córdoba, Córdoba, Spain
lucia.jimenez@uco.es

Abstract. Tourism discourse as a domain-specific discourse is characterized by a set of linguistic, pragmatic, and function features that make it different from other discourses and the general language. One of its essential elements is the usage of appealing, innovative, exotic-sounding words in order to attract potential tourists by "persuading, luring, wooing and seducing" [6]. In this context, formulaic language plays a key role. To date, research into chunks of language used in tourism have mostly focused on collocations [1, 8, 23], with a few works on longer sequences [11, 12, 13].
Bearing this in mind, this paper aims to contribute to the analysis of 4-word bundles in this domain, more specifically, in the segment of adventure tourism. To do so, a corpus-driven analysis was undertaken. As for our methodology, a specialized corpus containing English promotional texts was compiled. After that, the software Sketch Engine was used to extract a list of potential 4-word bundles. Next, manual verification was performed to ensure the validity of the units. Finally, the resulting list was classified according to their structural framework and their function in the text. The findings show that, in terms of the structure, the most typical sequences were verbal bundles; on the other hand, in terms of the function, a significant amount of the units was mainly used to address readers directly.

Keywords: Adventure Tourism, 4-Word Bundle, Function, Structure.

1 Long Word Sequences in Specialized Discourse

Traditionally, phraseological units have been categorized according to their degree of fixedness and compositionality [5, 14, 21]. Thus, collocations are found at the end of one continuum and idioms at the other end. It means that the former are less structurally fixed and more semantically transparent than the latter. However, another criterion commonly set to identify typical word sequences has been frequency of use. This has been possible thanks to corpus linguistics and automatic software that allows the exploration of corpora.

A typical focus of corpus and phraseological studies has been the specialized discourse. In this context, not only has the emphasis been placed on collocations, but

Proceedings of the International Conference EUROPHRAS 2022, pages 8–16,
Malaga, Spain, 28-30 September, 2022.
https://doi.org/10.26615/978-954-452-080-9_002

research has also delved into longer sequences of words. For instance, structures of 3, 4, and 5 words have been analyzed in the field of applied linguistics [17]; 4-grams have been explored in scientific research articles [19]; complex nominals have been covered in the specialized domain of the environment [4]. As to the discourse of tourism, recurrent lexical bundles and phrase frames have been examined in hotel websites [11, 12, 13], concluding that the flexible elements of these sequences are content words which fill the slot in frames such as *will be [required, charged] to* or *we are [happy, delighted] to*. Regarding the subdomain of adventure tourism, two-word combinations have been covered both in English and Spanish [8, 18, 20], but longer sequences have not been examined yet.

Having said that, the main aim of this study is to contribute to the linguistic description of this field by analyzing the usage of 4-word bundles focusing on two aspects, their structure and their function in the discourse, which is where the contribution of this paper lies in. These multi-word combinations can be defined as "sequences of [four] words that show a statistical tendency to co-occur" [2]. The underlying hypothesis is that the discourse of adventure tourism can display an extensive range of phraseological units which evidence its degree of specialization, to clarify, its being regarded as a specialized discourse. In order to test this hypothesis, two are the stated objectives: first, we will identify the structural frameworks of these sequences of words, and second, we will address their function in the text.

This paper is organized according to the following sections: Section 2 describes the methodology employed to achieve our objectives; Section 3 explains and discusses the main results obtained; Section 4 presents the conclusions drawn as well as some lines of further research.

2 Methodology

This section will explore the methodological steps followed in order to attain the objectives of this study, which are: (1) the compilation of a specialized corpus, (2) the extraction of 4-word bundles, (3) their structural classification, and (4) their functional categorization.

2.1 Compilation of ADVENCOR EN

The first step to perform a linguistic study is the compilation of a reliable corpus, given that "The results are only as good as the corpus" [24]. For this reason, this paper presents a corpus-driven analysis of 4-word sequences extracted from a specialized 1,005,480-word English corpus about adventure tourism, which was automatically compiled using Sketch Engine. The texts selected were originally written in English, contemporary, and recently published in electronic format by public or private institutions, registered tourist companies, or travel agencies from English-speaking countries all over the world, such as the United Kingdom, the United States, and Ireland. The texts included were full texts, since they represent the genre under study better than samples of a certain length would [10]. Regarding the level of specialization, these promotional texts represent a specialized/non-specialized communicative situation (from expert to non-expert), for their primary purpose was to woo tourists interested in adventure tourism (in general) and adventure activities (in particular).

ADVENCOR EN has already proved to be representative of the domain of adventure tourism and shed new light on the linguistic description of this segment. For instance, the keyness of adjectives has been examined and it has been discovered that they can be descriptive (e.g., *aerial, complimentary*) and evaluative (e.g., *lovely, pleasant*), being their aim to persuade the reader by contributing to the creation of mental representations of destinations [7]. On the other hand, motion verbs have been analyzed from a lexico-semantic perspective and it has been found that they explain how knowledge is expressed in this tourism segment [9]. Last but not least, collocations of motion verbs have also been studied and the main findings have been that collocates represent semantic roles of the argument structures [8, 18, 20].

2.2 Extraction of 4-word bundles

The second step of this study was the extraction of 4-word bundles typical of our specialized corpus. At this point, the 'N-grams' function available at Sketch Engine was used. The reason for exploring 4-word sequences rather than 3-/5-word sequences is that the former often subsume 3-word sequences [22]; in addition to that, they are much more frequent than 5-word sequences, offering a clearer range of structures and functions [15]. A frequency threshold of 20 tokens per million words was set [16], which means that 4-word bundles occurring at least 20 times in ADVENCOR EN were retrieved. This step produced a list of 234 items with a total frequency of 8,236 tokens. Nevertheless, we had to manually weed out some troublesome chunks for the following reasons:

1. They belonged to the name of a document included in the corpus, for instance, *activity tourism in wales, paragliding and hang gliding.*
2. They had been wrongly annotated, such as *m ore likely to*.
3. They only occurred in one specific context, not being representative of the whole corpus, for example, *price is per adult, for gift certificate redemptions*.
4. They made no sense in this study, such as *m o u n, n ta i n, av i n g*.

After this manual work, 76 items were discarded, so the final list of 4-word bundles amounted to 158 sequences.

2.3 Structural categorization of 4-word bundles

The next step in this investigation was the categorization of the final list of the units according to their structure. For this task, we contemplated the following classes based on Biber *et al.*[1] [3]: (1) nominal bundles, whose head is a noun (e.g., *his bristly short*

[1] These are classes which could embrace sequences containing a number of words other than four; in fact, the examples provided are taken from the authors and do not specifically show 4-word bundles.

hair, the journey back); (2) verbal bundles, whose head is a verb (e.g., *was walking, can see*); (3) adjectival bundles, whose head is an adjective (e.g., *so lucky, subject to approval by*); (4) adverbial bundles, whose head is an adverb (e.g., *fortunately enough, hardly ever*); (5) prepositional bundles, whose head is a preposition (e.g., *to him, in a street*). Additionally, we considered two more classes, conjunctions and full phrases (when they were registered in a dictionary as such).

2.4 Functional categorization of 4-word bundles

The final step of our methodology was the categorization of the 4-word bundles according to their function in the text. Thus, three broad categories along with their own subcategories were considered [15]:

1. Research-oriented sequences, used to structure the information:
   a. Location, which indicate time and place (e.g., *at the same time*).
   b. Procedure, concerning methods and processes (e.g., *the role of the*).
   c. Quantification, related to quantities (e.g., *a wide range of*).
   d. Description, used to describe facts (e.g., *the structure of the*).
   e. Topic, connected to the field of research (e.g., *the currency board system*).
2. Text-oriented sequences, which concern the organization of the text and the meaning of its elements as a message or argument:
   a. Transition signals, establishing additive or contrastive links between elements (e.g., *in addition to the*).
   b. Resultative signals, which mark inferential or causative relations (e.g., *as a result of*).
   c. Structuring signals, defined as text-reflexive markers which organize stretches of discourse or direct reader elsewhere in text (e.g., *in the next section*).
   d. Framing signals, used to specify limiting conditions (e.g., *in the case of*).
3. Participant-oriented sequences, focused on the writer or the reader of the text:
   a. Stance features, which convey the writer's attitudes and evaluations (e.g., *are likely to be*).
   b. Engagement features, addressing readers directly (e.g., *it should be noted*).

3 Results and Discussion

As it has been previously mentioned, the final list of 4-word bundles amounted to 158 items. The most recurrent units were *one of the most* (253 tokens), *is one of the* (243 tokens), and *one of the best* (108 tokens). Some of the least recurrent units (i.e., occurring 20 times in ADVENCOR EN) were *at the bottom of, is famous for its, the great barrier reef*. The following subsections show the results obtained in this study in terms of the structural framework and function of the sequences selected.

3.1 Structural features of 4-word bundles in adventure tourism

The first specific objective outlined in this research was the structural classification of the 4-word bundles selected. Table 1 displays this classification and shows the different structures identified organized according to the number of items, along with their overall frequency in the corpus (i.e., the total number of tokens), the percentage they occupy, and some examples:

Table 1. Structural classification of the 4-word bundles selected

Structure	No. of sequences	Overall frequency	Percentage	Examples
Verbal bundle	61	2,168	38.6	*to book your trip, you are looking for*
Nominal bundle	48	2,128	30.4	*impact of outdoor activity, the heart of the*
Prepositional bundle	33	1,227	20.9	*in the middle of, for the first time*
Adverbial bundle	6	191	3.8	*off when you spend, all over the world*
Adjectival bundle	4	93	2.5	*likely to participate in, are more likely to*
Conjunction	3	92	1.9	*but not limited to, so that you can*
Full phrase	3	99	1.9	*as well as a, thank you so much*
Total	158	5,998	100	

As shown in Table 1, there is a big difference between the three most frequent structural categories (verbal, nominal, and prepositional bundles, whose representation is over 20%) and the four least recurrent categories (adverbial and adjectival bundles, conjunctions, and full phrases, whose recurrence is below 5%).

Regarding the most frequent category, verbal bundles, more than a third of the items (26 sequences) incorporate a subject pronoun into the sequence, such as *you are interested in* and *we look forward to*, which makes emphasis on the potential tourist as well as the adventure activity's provider. With respect to the nominal bundles, one of the most recurrent structures consists of a noun phrase plus a preposition, especially *of*, for instance, *the base of the, a full day of*; other prepositions are *to* (e.g., *a departure date to, the best way to*) and *in* (e.g., *via ferrata in the, a dip in the*). As to the most common prepositions introducing prepositional bundles, we found *at* (7 tokens, e.g., *at the foot of, at the bottom of*), *in* (6 tokens, e.g., *in the middle of, in the united states*), and *of* (5 tokens, e.g., *of the most beautiful, of the world's most*), among others.

3.2 Functions of 4-word bundles in adventure tourism

The second specific objective stated in this study was the classification of the 4-word bundles selected according to the function they perform in the text. Table 2 represents this classification, showing the specific categories/subcategories identified, the number of sequences, their overall frequency, and the percentage they occupy in the corpus:

Table 2. Functional classification of the 4-word bundles selected

Category/Subcategory	No. of sequences	Overall frequency	Percentage
Research-oriented	**96**	**4,101**	**60.8**
1. Location	32	1,238	33.4
2. Procedure	0	0	0
3. Quantification	23	1,321	24
4. Description	18	499	18.6
5. Topic	23	1,043	24
Text-oriented	**8**	**259**	**5**
1. Transition signals	2	77	25
2. Resultative signals	1	39	12.5
3. Structuring signals	0	0	0
4. Framing signals	5	143	62.5
Participant-oriented	**54**	**1,638**	**34.2**
1. Stance features	18	488	33.3
2. Engagement features	36	1,150	66.7
Total	158	5,998	100

As it can be observed in Table 2, the “research-oriented” category contains more than half (60.8%) of the units analyzed. These items are classified into four distinct subcategories, being the largest one “location” (33.4%), which includes units referring to time and place, such as *at the end of, from the top of.* The second place is occupied by two subcategories, given that both “quantification” and “topic” incorporate 24% of the sequences, for instance, *one of the largest* and *there are plenty of* (“quantification”), *please select another departure* and *experience the thrill of* (“topic”). Finally, “description” includes 18.6% of the units, such as *speeds of up to, had a great time*. Regarding the “procedure” subcategory, no 4-word bundles were identified.

In the second place, the “participant-oriented” category contains over a third (34.2%) of the chunks selected. This category is divided into two subcategories: (1) “engagement features” represents more than half (66.7%) of the units, probably because they are used to address readers directly, for example, *you will need to* and *if you wish to*, which makes sense considering that ADVENCOR EN comprises tourism promotional texts; (2) “stance features” entail sequences used to voice the writers of the texts’ opinions, and occupy 33.3% of the structures included in the “participant-oriented” category, such as *can’t wait to, we look forward to*.

Last but not least, the “text-oriented” category represents only 5% of the 4-word sequences. Most of them (62.5%) are used to specify limiting conditions in the “framing signals” subcategory, for instance, *with the help of* and *including but not limited*. After that, “transition signals” occupy 25% of these units and are used to describe addition, such as *as well as the*. Finally, only one unit (12.5%) was found to show result: *as a result of.* No structuring signals were identified in the corpus.

On the other hand, Table 3 represents the relation between the structures and the functions performed by the 4-word bundles selected:

Table 3. Structural frameworks used in terms of the functional classification

Structure	Research-oriented		Participant-oriented		Text-oriented	
Nominal bundle	42	43.8%	6	11.1%	0	0
Prepositional bundle	26	27%	4	7.5%	3	37.5%
Verbal bundle	22	23%	38	70.4%	1	12.5%
Adverbial bundle	4	4.2%	1	1.8%	1	12.5%
Adjectival bundle	1	1%	3	5.6%	0	0
Conjunction	1	1%	1	1.8%	1	12.5%
Phrase	0	0	1	1.8%	2	25%
Total	96	100%	54	100%	8	100%

Table 3 shows that each broad functional category is mostly characterized by a different structural framework. To put it differently, nominal bundle (43.8%) is the most recurrent structure identified in "research-oriented" sequences (e.g., *the edge of the, queensland adventure activity standards*), the head describing location, the topic of the texts, quantities, among others. On the other hand, verbal bundle (70.4%) is the most typical structure of "participant-oriented" bundles (e.g., *if you have any, give us a call*), for the verbs help to engage the readers of the texts and render the writers' opinions. Finally, prepositional bundle (37.5%) is the most common structure found in "text-oriented" sequences (e.g., *as a result of, in the event of*), being useful to organize the text.

4 Conclusions and Further Research

The current investigation has explored the structural and functional features of 4-word bundles in the specialized discourse of adventure tourism. In total, 158 sequences were selected after their automatic extraction and manual verification.

As for our first objective, the most common structure was verbal bundle (38.6%). This result may be surprising, as it is not closely related to the findings revealed in the achievement of our second objective, that is, the functions performed by the bundles. To explain, the vast majority of items (60.8%) were included in the "research-oriented" category and subcategorized into "location", "quantification", "topic", and "description", and most of the structures in these groups were nominal (43.8%) and prepositional bundles (27%). Nevertheless, it must be highlighted that 34.2% of the units were classified as "participant-oriented" sequences, from which the largest amount referred to "engagement features" (66.7%) and were verbal bundles (70.4%). It means that the most recurrent structure does not represent the most typical function of the bundles. However, it makes sense considering that the texts of the corpus were promotional texts about adventure tourism which aimed to attract tourists, therefore, a wide range of the units address the readers directly. This fact also demonstrates the specificity of this domain, thus confirming our hypothesis.

All in all, the objectives of this study have been successfully achieved. Future research may focus on shorter/longer bundles and other languages, which may allow contrastive studies. Additionally, this methodology may be applied to other segments of the tourism discourse (e.g., eco-tourism, sun-and-beach tourism) or other specialized domains (e.g., the environment or the academic discourse).

References

1. Baynat Monreal, M. E.: El léxico de la gestión turística en lengua francesa en el Diccionario Multilingüe de Turismo: Análisis contrastivo con la lengua inglesa. Çédille, Revista de Estudios Franceses 13, 53–82 (2017).
2. Biber, D., Conrad, S.: Lexical bundles in conversation and academic prose. In: Hasselgård, H., Oksefjell, S. (eds.) Out of corpora: Studies in honour of Stig Johansson, pp. 181–189. Rodopi, Amsterdam/Atlanta, GA (1999).
3. Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E.: The Longman grammar of spoken and written English. Longman, London (1999).
4. Cabezas-García, M., Faber, P.: Phraseology in specialized resources: An approach to complex nominals. Lexicography 5(1), 55–83 (2018).
5. Cowie, A. P.: The treatment of collocations and idioms in learners' dictionaries. Applied Linguistics 2(3), 223–235 (1981).
6. Dann, G.: The language of tourism. A sociolinguistic perspective. CAB International, Wallingford (1996).
7. Durán-Muñoz, I.: Adjectives and their keyness: A corpus-based analysis of tourism discourse in English. Corpora 14(3), 351–378 (2019).
8. Durán-Muñoz, I., Jiménez-Navarro, E. L.: Colocaciones verbales en el turismo de aventura: Estudio contrastivo inglés-español. In: Corpas Pastor, G., Bautista Zambrana, M. R., Hidalgo-Ternero, C. M. (eds.) Sistemas fraseológicos en contraste: Enfoques computacionales y de corpus, pp. 121–142. Comares, Granada (2021).
9. Durán-Muñoz, I., L'Homme, M.-C.: Diving into English motion verbs from a lexico-semantic approach. A corpus-based analysis of adventure tourism. Terminology 26(1), 33–59 (2020).
10. Flowerdew, L.: The argument for using English specialized corpora to understand academic and professional language. In: Connor, U., Upton, T. A. (eds.) Discourse in the professions. Perspectives from corpus linguistics, pp. 11–33. John Benjamins Publishing Company, Amsterdam/Philadelphia (2004).
11. Fuster-Márquez, M.: Lexical bundles and phrase frames in the language of hotel websites. English Text Construction 7(1), 84–121 (2014).
12. Fuster-Márquez, M.: The discourse of US hotel websites: Variation through the interruptibility of lexical bundles. In: Gotti, M., Maci, S., Sala, M. (eds.) Ways of seeing, ways of being: Representing the voices of tourism, pp. 401–420. Peter Lang, Bern/Berlin/Brussels/Frankfurt am Main/New York/Oxford/Wien (2017).
13. Fuster-Márquez, M., Pennock-Speck, B.: Target frames in British hotel websites. International Journal of English Studies 15(1), 51–69 (2015).
14. Howarth, P. A.: Phraseology in English academic writing. Some Implications for language learning and dictionary making. Niemeyer, Tübingen (1996).
15. Hyland, K.: Academic clusters: Text patterning in published and postgraduate writing. International Journal of Applied Linguistics 18(1), 41–62 (2008).
16. Jalali, Z. S., Moini, M. R., Arani, M. A.: Structural and functional analysis of lexical bundles in medical research articles: A corpus-based study. International Journal of Information Science and Management 13(1), 51–69 (2015).
17. Jalilifar, A., Ghoreishi, S. M.: From the perspective of: Functional analysis of formulaic sequences in Applied Linguistics research articles. International Journal of English Studies 18(2), 161–186 (2018).

18. Jiménez-Navarro, E. L.: Treatment and representation of verb collocations in the specialized language of adventure tourism. Doctoral dissertation (Universidad de Córdoba, Cordoba, Spain) (2020).
19. Jiménez-Navarro, E. L.: A corpus-based study of 4-grams in the research article genre. ELUA 38, 241–262 (2022).
20. Jiménez-Navarro, E. L., Durán-Muñoz, I.: Collocations of fictive motion verbs in adventure tourism: A corpus-based study of the English language. RESLA (2022/forthcoming).
21. Mel'čuk, I. A.: Phraseology in the language, in the dictionary, and in the computer. Yearbook of Phraseology 3(1), 31–56 (2012).
22. Pérez-Llantada, C.: Formulaic language in L1 and L2 expert academic writing: Convergent and divergent usage. Journal of English for Academic Purposes 14, 84–94 (2014).
23. Piccioni, S., Pontrandolfo, G.: La construcción del espacio turístico a través de la fraseología metafórica. Linguistik Online 94(1/19), 137–153 (2019).
24. Sinclair, J.: Corpus, concordance, collocation. Oxford University Press, Oxford (1991).