UNIVERSIDAD DE CÓRDOBA

PROGRAMA DE DOCTORADO EN LENGUAS Y CULTURAS

DEPARTAMENTO DE FILOLOGÍAS INGLESA Y ALEMANA

ÁREA DE LINGÜÍSTICA INGLESA

TESIS DOCTORAL

TREATMENT AND REPRESENTATION OF VERB COLLOCATIONS IN THE SPECIALIZED LANGUAGE OF ADVENTURE TOURISM

EL TRATAMIENTO Y LA REPRESENTACIÓN DE LAS COLOCACIONES VERBALES EN EL LENGUAJE ESPECIALIZADO DEL TURISMO DE AVENTURA

Directora: Dra. Isabel Durán Muñoz

Autora: Eva Lucía Jiménez Navarro

Noviembre 2020

TITULO: *TREATMENT AND REPRESENTATION OF VERB COLLOCATIONS IN THE SPECIALIZED LANGUAGE OF ADVENTURE TOURISM*

AUTOR: *Eva Lucía Jiménez Navarro*

© Edita: UCOPress. 2021
Campus de Rabanales
Ctra. Nacional IV, Km. 396 A
14071 Córdoba

https://www.uco.es/ucopress/index.php/es/
ucopress@uco.es

TÍTULO DE LA TESIS: *Treatment and representation of verb collocations in the specialized language of adventure tourism*

DOCTORANDA: EVA LUCÍA JIMÉNEZ NAVARRO

INFORME RAZONADO DEL/DE LOS DIRECTOR/ES DE LA TESIS
(se hará mención a la evolución y desarrollo de la tesis, así como a trabajos y publicaciones derivados de la misma).

La presente tesis doctoral cumple con todos los requisitos legales exigibles a un trabajo de investigación para la obtención del título de Doctor con Mención Internacional.

La doctoranda aborda, desde una perspectiva teórica y práctica, las colocaciones especializadas en un campo de especialidad específico, el turismo de aventura, partiendo de la siguiente hipótesis inicial: "Las colocaciones verbales en el lenguaje especializado del turismo de aventura expresan significados especializados que merecen ser recopilados en productos terminológicos". Para dar respuesta a su tesis, establece su objetivo principal, que consiste en la realización de un estudio exhaustivo de las colocaciones verbales en este campo de especialidad y su implementación en las entradas de los verbos de movimiento en *DicoAdventure*, un diccionario especializado del turismo de aventura, a partir de un corpus monolingüe inglés especializado. Para ello, establece una serie de objetivos teóricos y prácticos secundarios que va alcanzando en los diferentes capítulos de su trabajo.

El trabajo se estructura en seis capítulos bien diferenciados entre sí. Por un lado, se encuentra la parte teórica, que se desarrolla en los tres primeros capítulos y donde se realiza una revisión de la literatura con respecto a la Terminología y la Fraseología, se define el concepto de *colocación especializada* y se examina una serie recursos lexicográficos y terminológicos con respecto a las colocaciones verbales para proponer un modelo propio que se utilizará para incluir esta información en *DicoAdventure*. Por otro lado, la parte práctica presenta el material y la metodología utilizados, los resultados obtenidos y el análisis elaborado. En esta parte, la doctoranda realiza una serie de tareas clave, como son la extracción de los verbos de movimiento del corpus y su clasificación por su significado de movimiento real o figurado (o metafórico), así como sus colocaciones léxicas, que le permiten ofrecer un análisis exhaustivo de los resultados obtenidos. Finalmente, aporta una serie de conclusiones y da respuesta a la hipótesis inicial, además de abrir nuevas líneas de trabajo futuro. La bibliografía empleada por la doctoranda es completa y actual, sin desdeñar otra de mayor tradición en el ámbito del trabajo, por lo que se puede afirmar que la fundamentación teórica del trabajo es sólida.

Como resultados de esta investigación, esta tesis doctoral también ha dado lugar a varias publicaciones, tanto en revistas de impacto y volúmenes editados como en congresos internacionales, entre las que cabe destacar las siguientes:

- Jiménez-Navarro, E. L. (2019). "Nominal collocations in scientific English: A frame-semantic approach", en G. Corpas Pastor y R. Mitkov (eds.). *Proceedings of Computational and Corpus-based Phraseology. Third International Conference, Europhras 2019*. Universidad de Málaga. 25-27 de septiembre de 2019 (pp. 187-199). Cham: Springer. ISBN: 978-3-030-30134-7.

- Jiménez-Navarro, E. L. (2020). "A comparative analysis of collocational information in monolingual learner's dictionaries". *XII CILC 2020 International Congress (Corpus-Assisted Specialized Discourse Analysis and Translation Studies)*. Universidad de Bérgamo (Italia). 10-12 diciembre 2020 (pospuesto a mayo de 2021).

- Jiménez-Navarro, E. L. (2020). "Where language and knowledge conflate: The case of syntagmatic patterns in the discourse of science". *IV Congreso Internacional Ciencia y Traducción: "Puentes interdisciplinares y difusión del conocimiento científico"*. Universidad de Córdoba. 1-3 abril 2020 (pospuesto a enero de 2021).

- Durán Muñoz, I. y Jiménez-Navarro, E. L. (2020). "Advencor, the corpus of adventure tourism: Design, compilation and management". *LexESP 2020. VI International Seminar on English and ESP Lexicology and Lexicography: The Language of Tourism, Sports and Outdoor Leisure Activities*. Universidad de Alicante, 7-8 de mayo 2020 (pospuesto sin fecha).

- Durán Muñoz, I. y Jiménez-Navarro, E. L. (2021/en revisión). "Colocaciones verbales en el turismo de aventura: Estudio contrastivo inglés-español". En G. Corpas Pastor, R. Bautista y C. Hidalgo. *Sistemas fraseológicos en contraste: enfoques computacionales y de corpus*. Berlín: Peter Lang.

Por todo ello, se autoriza la presentación de la tesis doctoral.

Córdoba, 2 de noviembre de 2020

Firma de la directora

Fdo.: Isabel Durán Muñoz

To my mother, and to the memory of my father

No vayas por donde el camino te lleve.

Ve en cambio por donde no hay camino y

deja rastro.

Ralph Waldo Emerson.

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my advisor Professor Isabel Durán Muñoz for providing support when it was most needed. Thank you, Isa, for giving me the opportunity to work with you, for your guidance and motivation, you opened my eyes to a wonderful world of knowledge and I am eternally indebted. Thank you for sharing your time with me during such an important stage in your life, and most importantly thank you for making it possible.

Special thanks to Professor Olga Blanco Carrión for her encouragement throughout the years we worked together. Thank you, Olga, for sparking my interest in collocations and thank you for your advice, without which I would not be in the place I am now.

I would like to show my greatest appreciation to Professor Antonio Barcelona Sánchez for trusting me and allowing me to become part of the *Cognitive and Functional Linguistics Research Group*. Thank you, Antonio, for expressing your empathy when I had to take a landmark decision.

I truly thank Professor Marie-Claude L'Homme for permission to use *DicoAdventure* and for giving insightful comments and suggestions. My grateful thanks also to Benoît Robichaud for the detailed guidance and guidelines offered.

Let me also thank my colleagues from the Department of English and German Studies for their support during this time. Also, a hearty thanks to everyone for making me feel part of the team.

My thanks also go to my super girls, for sticking by me through thick and thin and for the big meals we have enjoyed together, and the MBLA Team, for sharing their valuable knowledge and for wanting to shine together. Thank you, guys.

I would like to show my greatest appreciation to my friends, those to whom I could not devote the time they deserve. A special mention must go to Chero, Cris, Illa, María, Reichel, Solecito, Sophia and Twin.

My deepest heartfelt gratitude goes to my family, who I could not love any more. Thank you, mum and dad, for bringing me up the way you did, you will always be present

in every step that I take. Thank you, Moi, David and sister, for your everyday messages and for always being by my side, I could not imagine my life without you. A big thanks to Amanda, for showing me her true love despite not being able to dedicate more time to playing with her, she makes me feel the luckiest auntie in the world. And my special thanks also go to the other part of the family, for they will not understand these words, but their unconditional love is beyond compare.

Finally, I owe a very important debt to Alejandro, my faithful companion. Thank you for half a life together and for your sympathy and understanding when we could not spend any more time exploring. Adventure awaits us.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BBI	The BBI Combinatory Dictionary of English
BNC	British National Corpus
CDE	Cambridge Dictionary of English
CECL	Centre for English Corpus Linguistics
CED	Collins English Dictionary
COBUILD	Collins Birmingham University International Language Database
COCA	Corpus of Contemporary American English
CTT	Communicative Theory of Terminology
DiCoEnviro	*Dictionnaire fondamental de l'environnement*
DICOFE	*Dictionnaire contextuel du français économique*
DiCoInfo	*Le dictionnaire fondamental de l'informatique et de l'Internet*
DIN	*Deutsches Institut für Normung*
DS	Danish Standards
EAGLES	Expert Advisory Group on Language Engineering Standards
ECD	Explanatory Combinatorial Dictionary
ÉCLECTIK	*Équipe de Recherché en Combinatoire Lexicale, Terminologie et Informatique*
enTenTen13	English Web 2013 Corpus
FBT	Frame-Based Terminology
FE	Frame element
FN	FrameNet
GL	General language
GRELEP	*Groupe de Recherche en Lexicographie Pédagogique*
GTT	General Terminology Theory
HASK	HASK: Collocation Databases
IATE	Interactive Terminology for Europe
ICM	Idealized cognitive model
IPA	International Phonetic Alphabet
ISO	International Organization for Standardization
LDOCE	Longman Dictionary of Contemporary English
LLC	London-Lund Corpus of Spoken English

LF	Lexical Function
LSP	Language for Special/Specific Purposes
LU	Lexical unit
LVC	Light-verb construction
L_1	First language
L_2	Second language
MCD	Macmillan Collocations Dictionary for Learners of English
MED	Macmillan English Dictionary
MTM	Meaning-Text Model
MTT	Meaning-Text Theory
MW	Multi(-)word
MWE	Multi(-)word expression
MWU	Multi(-)word unit
NC	Noun compound
NLP	Natural Language Processing
OAAD	Oxford Advanced American Dictionary
OALD	Oxford Advanced Learner's Dictionary
OCD	Oxford Collocations Dictionary for Students of English
OLST	*Observatoire de Linguistique Sens-Texte*
POS	Part-of-speech
PP-D	Determinerless-prepositional phrase
PU	Phraseological unit
PV	Prepositional verb
RP	Received Pronunciation
SEU	Survey of English Usage
SL	Specialized language
SSE	Survey of Spoken English
STT	Sociocognitive Terminology Theory
TU	Terminological unit
UNWTO	World Tourism Organization
VNIC	Verb-noun idiomatic combination
VPC	Verb-particle constructions
XML	eXtensible Markup Language

ABSTRACT

Treatment and representation of verb collocations in the specialized language of adventure tourism

A *collocation* is considered a frequent co-occurrence of two words which hold a syntactic relationship and whose elements enjoy a different status. Given their perception as a unit in language, access to the prominent word (*base*) involves immediate access to the other item (*collocate*). In terms of meaning, some combinations tend to be more transparent than others. The pervasiveness of these word associations in language has sparked a strong research interest in the last decades. A compelling reason for this approach may be the fact that they are naturally produced by native speakers but must be actively learned by non-native individuals. Not only has this reality led to their treatment in the general language, but it has also become a legitimate field of study in a wide range of specialized languages, such as the environment, computing, law or tourism, which is our object of study. As a consequence, specialized knowledge resources covering this type of word combinations have seen the light with the primary purpose of offering some extra help to people who deal with this type of language, for example, translators, linguists or other professionals. Nevertheless, there is still much to do in this respect.

Taken this into account, it is hypothesized that verb collocations in the specialized language of adventure tourism convey specialized meaning that is worth being collected in terminological products. Therefore, this work endeavors, as its main purpose, to perform a deep analysis of verb collocations in this specialized domain and their implementation in the entries for motion verbs in *DicoAdventure*, a specialized dictionary of adventure tourism, whose inspirational idea was to highlight the significant role of verbs in the linguistic expression of concepts. Accordingly, the following theoretical objectives were set: first, to cover the linguistic branches which influence specialized lexicography; second, to define the concept of *specialized collocation*; and third, to examine a vast number of lexicographical and terminological resources so as to discover the items of information that would make an adequate representation of collocations in a specialized dictionary and, then, design a model for such task. Furthermore, the following practical objectives were formulated: first, to extract the motion verbs which would be

the bases of the collocations implemented; second, to retrieve the lexical collocations of these verbs; and third, to classify the resulting list of collocations according to the meaning expressed, that is, actual motion or fictive (or metaphorical) motion. The practical steps taken in this research were based on the English monolingual specialized corpus ADVENCOR, which contains promotional texts about adventure tourism, and the use of corpus management software.

The results of the theoretical work can be summarized as follows: (1) the specialized language of adventure tourism must be considered as specialized as any others; (2) collocations are not usually encoded in verb entries in dictionaries; and (3) a specialized collocation carries specialized knowledge which must be covered in terminological products. On the other hand, regarding the practical work, 12% of the verbs extracted were selected, as they were the ones expressing motion. However, only 46.61% of them produced collocations according to the extraction criteria established. Last, after applying more strict criteria for the collocation classification, only 25.42% of the verbs along with their collocations were collected in the dictionary. In addition to these results, the theory of Frame Semantics proved useful to understand the meaning of the verbs and their collocates. As for their implementation, which was the primary objective of this doctoral dissertation, the inclusion of verb collocations was of paramount importance for the identification of distinct meanings expressed by one verb in different contexts, as collocates conveyed subtle nuances of meaning. Finally, it was concluded that the incorporation of explanations about the combinations in lay terms facilitates the comprehension of the entries to any type of user, from experts to laypersons, which makes *DicoAdventure* a terminological product that can render valuable assistance to individuals with distinct specialized expertise.

Keywords: adventure tourism, collocation, specialized corpus, specialized dictionary, specialized language.

RESUMEN

El tratamiento y la representación de las colocaciones verbales en el lenguaje especializado del turismo de aventura

Una *colocación* es una coaparición frecuente de dos palabras que mantienen una relación sintáctica y cuyos elementos alcanzan un estatus diferente. Puesto que se perciben como una unidad del lenguaje, el acceso al elemento prominente (*base*) conlleva el acceso inmediato al otro componente (*colocativo*). Con respecto a su significado, algunas combinaciones tienden a ser más transparentes que otras. La constante presencia de las colocaciones en el lenguaje ha despertado gran interés por su investigación en las últimas décadas. Una razón convincente de este acercamiento podría ser el hecho de que los hablantes nativos las producen de forma natural, mientras que los no nativos deben aprenderlas de manera activa. Esta realidad no solo ha llevado a su tratamiento en el lenguaje general, sino también a que se hayan convertido en un campo de estudio legítimo en una amplia gama de lenguajes especializados, como son el medio ambiente, la informática, el derecho o el turismo, que es el objeto de estudio de esta investigación. Como consecuencia, se han creado recursos de conocimiento especializado con el propósito fundamental de ofrecer ayuda a las personas que interactúan con este tipo de lenguaje, por ejemplo, traductores, lingüistas u otro tipo de profesionales. No obstante, aún queda mucho por hacer en este aspecto.

Teniendo esto en cuenta, la hipótesis de este trabajo se basa en la idea de que las colocaciones verbales en el lenguaje especializado del turismo de aventura expresan significados especializados que merecen ser recopilados en productos terminológicos. Por lo tanto, este trabajo tiene como principal objetivo el estudio exhaustivo de las colocaciones verbales en este campo de especialidad y su implementación en las entradas de los verbos de movimiento en *DicoAdventure*, un diccionario especializado del turismo de aventura, cuyo punto de partida fue la intención de destacar el importante papel que juegan los verbos en la expresión lingüística de los conceptos. Por consiguiente, se establecieron los siguientes objetivos teóricos: primero, revisar las ramas de la lingüística que ejercen una influencia en la lexicografía especializada; segundo, definir el concepto de *colocación especializada*; y tercero, examinar un gran número de recursos

lexicográficos y terminológicos para descubrir qué tipo de información conformaría una representación adecuada de colocaciones en un diccionario especializado y, a continuación, diseñar un modelo para esta tarea. Además, se propusieron estos objetivos prácticos: primero, extraer los verbos de movimiento que serían las bases de las colocaciones implementadas; segundo, extraer las colocaciones léxicas de estos verbos; y tercero; clasificar la lista resultante de colocaciones según su significado, es decir, movimiento real o movimiento figurado (o metafórico). Los pasos prácticos que se dieron en esta investigación se llevaron a cabo mediante la gestión del corpus especializado monolingüe en inglés ADVENCOR, que contiene textos promocionales sobre el turismo de aventura, y el uso de software de gestión de corpus.

Los resultados de la parte teórica del trabajo se pueden resumir de la siguiente manera: (1) el lenguaje especializado del turismo de aventura debe considerarse tan especializado como otros; (2) las colocaciones no suelen codificarse en las entradas de verbos en los diccionarios; y (3) una colocación especializada contiene conocimiento especializado que debe aparecer en productos terminológicos. Por otro lado, con respecto al trabajo práctico, se seleccionó el 12% de los verbos extraídos, ya que eran los que expresaban movimiento. Sin embargo, solo el 46,61% de ellos produjeron colocaciones según los criterios de extracción establecidos. Por último, después de aplicar criterios más estrictos para la clasificación de las colocaciones, solo el 25,42% de los verbos con sus colocaciones fueron recogidos en el diccionario. Además de estos resultados, se demostró la utilidad de la teoría de la Semántica de Marcos para entender el significado de los verbos y sus colocativos. En cuanto a su implementación, que era el objetivo principal de esta tesis doctoral, la inclusión de colocaciones verbales fue de suma importancia para la identificación de los distintos significados expresados por un verbo en diferentes contextos, puesto que los colocativos aportaban sutiles matices de significado. Finalmente, se concluyó que la incorporación de explicaciones sobre las combinaciones en términos legos favorece la comprensión de las entradas por parte de cualquier tipo de usuario, desde expertos a personas no especialistas, lo cual hace de *DicoAdventure* un producto terminológico que puede proporcionar valiosa ayuda a personas con diversa formación especializada.

Palabras clave: turismo de aventura, colocación, corpus especializado, diccionario especializado, lenguaje especializado.

CHAPTER 0. INTRODUCTION

> *A specific set of **combinations** in which terms can be found, i.e. **collocations**, are particularly interesting in terminology and resources have started to record them […]. Non-experts might experience problems for producing proper collocates in specialized texts.*
>
> (L'Homme, 2020, p. 32)

0.1. Motivations for this study

It is fully acknowledged that collocations are pervasive in language.[1] Nevertheless, while they may be naturally produced by native speakers, they tend to pose serious problems for second language learners. In fact, the latter are not always consciously aware of their existence and of some aspects of their behaviour. From my experience as a teacher with students interested in developing their skills in English, I have developed an intense interest in this type of word combinations, for works examining ways of teaching and/or learning phraseology seem to be still scarce.[2] Moreover, my experience as a supervisor of university students having different tastes and undertaking their Bachelor's Thesis has allowed me to be willing to explore the role of collocations in specialized languages, such as sports, economy, music or movies. By the same token, it has opened my eyes to the enormous difficulties encountered when aiming to translate collocations into their mother tongue or to find an equivalent in their second language. Given these points, I think that the present work will help to pursue the abovementioned interests, which can be summarized in one sentence: the correct production of collocations in specialized languages.

[1] See Alonso Campos, Williams, & DeCesaris (2017, p. 45), Barnbrook, Mason, & Krishnamurthy (2013, p. 93), Handl (2009, p. 70), Sinclair (1991, p. 111) and Wray (2002, p. 13), to mention but a few.
[2] Some authors who have dealt with phraseology teaching and learning are Boers & Lindstromberg (2008), Cervero & Pichardo Castro (2000), Lewis (2000) and Meunier & Granger (2008).

0.2. Introduction

The current doctoral dissertation aims to make a useful contribution to *DicoAdventure. An online dictionary of adventure tourism* (freely available at http://olst.ling.umontreal.ca/dicoadventure/).[3] More specifically, the study presented here investigates how the effective implementation of collocations in the entries for English motion verbs may be secured. With this idea in mind, the theoretical framework of this research revolves around the following three focal points: (1) the status of specialized languages, (2) the concepts of *collocation* and *specialized collocation*, and (3) the phraseological information encoded in lexicographical and terminological resources.

First, specialized languages are considered to be used for the dissemination of specialized knowledge. However, it may not reach every person in a society if their degree of specialization is extremely high, for different levels of specialization are found (Durán-Muñoz, 2012; Pearson, 1998). To put it differently, the vocabulary employed in communicative settings where these languages play a prominent role may be unknown to laypersons who have never learned them, given that they are not naturally acquired as the words used in the general language are (Cabré Castellví, 1999, 2003).

On the other hand, the position of specialized languages with respect to the general language has been addressed in different ways (e.g., Bergenholtz & Tarp, 1995; De Beaugrande, 1987; Hoffmann, 1979; Rey, 1976; Rondeau, 1983), being the most embraced opinion the one which regards specialized languages as subsystems which share some features with the general language. However, they are characterized by having a specific set of words that can define the subject field of the language, which is its *terminology*. According to Alcaraz Varó (2000) and Picht and Draskau (1985), this is the feature that defines a specialized language, but this criterion is not enough.

As a consequence, Terminology has shown different ways of dealing with specialized languages in the last century, as it started taking a prescriptive view in the 1930s (Wüster, 1931) and has evolved to become more flexible since then (e.g., Boulanger 1991, 1995; Gambier 1991, 1993; Gaudin 1993, 2003). Additionally, the various approaches to terminology developed in the last decades have pointed out

[3] This dictionary is the result of research which has been carried out so far and which is part of an ongoing project (cf. Durán-Muñoz, in press; Durán-Muñoz & L'Homme, 2020).

essential features of specialized languages (Cabré Castellví, 2003; Faber Benítez, 2011, 2015; Temmerman, 1997, 1998), such as their interdisciplinarity, the synonymy of terms, language evolution or the importance of context.

Last but not least, the study of specialized languages has been facilitated by the emergence of three specific branches of Applied Linguistics, namely, Cognitive Linguistics, Corpus Linguistics and Computational Linguistics. The first one regards language meaning as a module in our mind which is influenced by our experience of the world (Geeraerts, 2006). The second one proposes the use of naturally-occurring texts to perform linguistic analyses (Tognini-Bonelli, 2001; Viana, Zyngier, & Barnbrook, 2011). The last one provides tools to carry out these analyses (Grishman, 1986; Kay, 2003). By the same token, these theories have significantly influenced the investigation into phraseology.

Second, collocations are a part of a language's phraseology and have been extensively explored since the 1940s, when Phraseology was firmly established as a discipline. Pioneering work by Vinogradov (1947) inspired other linguists who followed his steps and decided to categorize word combinations too (Amosova, 1963; Cowie, 1981, 1998; Howarth, 1996, 1998). It was done according to two criteria, semantic opacity and combinatory flexibility, thus, they have normally been regarded as word associations located in a continuum in which free combinations are found at one end and idioms are found at the other (e.g., Cowie, 1981; Howarth, 1996; Melčuk, 1998, 2012).

In addition to the abovementioned criteria, collocations have been defined in terms of other aspects in the last years (Moreno Jaén, 2009; Seretan, 2011). On the one hand, they are characterized by features that affect their form, that is, the number (Martin, 2008; Pearce, 2001) and type of elements (Greenbaum, 1974; Halliday, 1966; Mitchell, 1971) included in a collocation and the number of words that separate these elements (Sinclair, 1991; Vargas Sierra, 2012a). On the other hand, other features distinguish them from distinct word combinations (e.g., idioms, lexical bundles, collostructions), such as the motivation behind a combination of words (Benson, Benson, & Ilson, 2009; Smadja, 1993), the frequency with which they are used (Bartsch, 2004; Cruse, 1986; Kjellmer, 1994) and their survival in a speaker's mind after having learned them (Palmer, 1933).

Aside from that, it is strongly agreed that collocations are omnipresent, therefore, their role in specialized languages has also been carefully examined (e.g., Bevilacqua,

2004; Pavel, 1993; Picht, 1987; Sanz Vicente, 2011). It has led to the emergence of the concept of *specialized collocations*, which are defined in the same way as collocations belonging in the general language but have also some distinguishing features. For example, they are used in specific communicative settings, such as medicine (Bieliaieva, Lysanets, Znamenska, Rozhenko, & Nikolaieva, 2017), law (Biel, Biernacka, & Jopek-Bosiacka, 2018) or the environment (Alonso Campos & Torner Castells, 2010).

Moreover, they have been analyzed in the specialized language of tourism, which is the focus of attention in this work. For instance, it has been discovered that this language shows specific phraseological preferences that are not shared with other specialized languages (Fuster-Márquez, 2014). Furthermore, the phraseology retrieved from tourism texts can be semantically connected to the content of the texts and can be classified into semantic fields depending on their reference (Manca, 2008). Finally, this phraseology can also be interpreted figuratively, since metaphors may underlie some meanings (Piccioni & Pontrandolfo, 2019).

The third fundamental aspect discussed in this dissertation is the phraseological information encoded in lexicographical and terminological resources, which should always been considered (Bergenholtz & Tarp, 1995; Hudson, 1988). Nevertheless, in the context of specialized resources, few of them actually include word combinations (L'Homme & Leroyer, 2009) and, when they do, tools are not systematic and they lack an effective method of representation (Buendía Castro, Montero Martínez, & Faber Benítez, 2014). In fact, there seems to lack a total agreement on the type of word combinations that should be encoded in dictionaries and how they should be described.

Added to that, when this information is included, it is common practice to do it in noun entries, despite the fact that specialized verbs should also be the object of a lexicographical description (L'Homme, 1998) and considering their co-occurrents can help to identify the verbs' meaning (L'Homme, 2003). Without a doubt, it poses a serious problem for users who are interested in finding word combinations that contain verbs, as these linguistic units will probably not be accessed. A feasible solution to this problem may be the implementation of word combinations, such as collocations, in the entries for this word type in specialized resources. To this end, the use of a template that incorporates valuable information and is within reach of all types of users is recommended, in other

words, one that is free of difficult metalanguage. In fact, this is the prime aim furthered in this work.

0.3. Hypothesis

The research conducted in this dissertation is based on the following hypothesis:

- There is a strong need for the implementation of verb collocations in terminological products for three sound reasons: first, they are not naturally produced by non-native speakers, that is, they must be learned; second, they convey specialized meanings; and third, these tools are greatly used by translators and other types of professionals, given their wide range of practical applications.

0.4. Objectives

The ultimate objective set in this study is the following:

- To implement collocations in the entries for motion verbs in the specialized dictionary *DicoAdventure*.

To this end, other core objectives are established:

1) To explore the theories which influence specialized lexicography;
2) To define the concept of (specialized) collocation;
3) To discover the items of information that would make an adequate representation of collocations in a specialized dictionary;
4) To extract the motion verbs from a specialized corpus consisting of promotional texts about adventure tourism;
5) To retrieve the lexical collocations of the motion verbs selected by applying stringent criteria; and
6) To classify the resulting list of collocations in terms of the meaning conveyed into two categories: actual motion and fictive motion.

0.5. Structure of this work

The current doctoral dissertation is organized into two parts, namely: *Part I: Theoretical framework* and *Part II: Methodology and empirical analysis*. The former will try to reach objectives 1), 2) and 3), whereas the latter will endeavor to achieve objectives 4), 5) and 6).

Chapter 1 addresses a set of linguistic theories which provide insight into specialized lexicography. The first section discusses the distinction between *general language* and *specialized language* (§1.1.) and delves into the second, which is explained according to three factors: linguistic, pragmatic and functional (§1.1.1.). These two types of language are the object of study of Lexicology and Terminology, respectively, a dichotomy examined in Section 1.2. Section 1.3. revolves around the harsh criticism prompted by the traditional view of the latter, in fact, it encouraged the emergence of new approaches to terminology, such as Socioterminology (§1.3.1.), Communicative Theory of Terminology (§1.3.2.), Sociocognitive Terminology Theory (§1.3.3.) and Frame-Based Terminology (§1.3.4), which have been relied on to create general and specialized resources, supporting the work carried out in the fields of Lexicography and Terminography, both dealt with in Section 1.4. Additionally, other disciplines which have enhanced a better understanding of terminology and the development of linguistic applications are treated in the last section (§1.5.), such as Cognitive Linguistics (and its offshoot Frame Semantics) (§1.5.1. and §1.5.1.1.), Corpus Linguistics (§1.5.2.) and Computational Linguistics (§1.5.3.).

Chapter 2 is devoted to the principal object of study of this work, collocations. The first section outlines the birth of Phraseology as a discipline in the 1940s (§2.1.) and highlights some of the major phraseological categorizations from then to the twenty-first century (§2.1.1.–§2.1.4.). Next, Section 2.2. endeavors to define the concept of *collocation* as a phraseological unit. To this end, the most significant formal and functional features which characterize them are examined (§2.2.1.). Added to that, this section describes one of the most influential typologies of collocations (§2.2.2.), and then explores other recurrent word combinations, such as idiomatic expressions, lexical bundles, collostructions and phrasal verbs (§2.2.3.), before providing our own definition of the concept (§2.2.4.). The closing section of this chapter (§2.3.) analyzes the idea of *specialized collocations* and investigates some pieces of work which have conducted

research into their role in the domain of tourism, for these are the main concern in this study.

Chapter 3 aims to give a revealing insight into an accurate representation of verb collocations in a specialized dictionary. With this objective in mind, Section 3.1. presents the pieces of information whose inclusion in a general language dictionary, on the one hand, and a specialized dictionary, on the other, is encouraged. After that, Section 3.2. is focused on the representation of collocations in lexicographical resources, more specifically, in five big English general dictionaries (§3.2.1.) as well as in five English collocation dictionaries (§3.2.2.). The following section analyzes the representation of these word combinations in five terminological resources (§3.3.) and highlights the lack of agreement on methodologies for listing and representing collocations in the entries for terms. Finally, Section 3.4. assesses the information gathered in the previous sections so as to summarize the strengths and weaknesses of every resource and be able to share our fresh idea for implementing collocations in a specialized dictionary.

Chapter 4 concentrates on the materials and methods used to carry out this work. It first specifies the characteristics of the specialized language of tourism according to three aspects: linguistic, pragmatic and functional (§4.1.), and then delves into the fundamental aspects of the adventure tourism domain (§4.1.1.). Section 4.2. explains *DicoAdventure*, the online concept-based dictionary in which collocations will be implemented. A sample entry is displayed along with examples of the distinct data categories already integrated in the tool in order to show where the 'Collocations' tab will be incorporated. Later, several aspects of a corpus are explored before explaining the one used in this dissertation. Therefore, Section 4.3. starts defining the concept of *corpus* and outlines the main types in terms of several elements, such as the specificity, the size or the language (§4.3.1.). It then continues to explain some of the factors that are key to the design of a corpus, more specifically, representativeness and balance, size and codification (§4.3.2.). These characteristics should be considered before a corpus compilation (§4.3.3.). The section closes with the description of the ADVENCOR corpus, the specialized corpus of adventure tourism which has made possible the current research (§4.3.4.). Furthermore, Chapter 4 deals with the computer software used to explore this corpus, *Sketch Engine* (§4.4.), and explains the steps taken for the attainment of objectives 4), 5) and 6), that is, the extraction of candidate motion verbs (§4.4.1.), the extraction of collocations (§4.4.2.) and the classification of the collocations according to the type of

motion represented, actual or fictive (§4.4.3.). Moreover, other corpus management tools, *TermoStat Web 3.0* and *WordSmith Tools 7.0*, are evaluated (§4.4.4.). Last but not least, Section 4.5. proposes a template for the implementation of collocations in *DicoAdventure*.

Chapter 5 presents the results obtained in this study. First, the resulting list of candidate motion verbs which were potential bases of the collocations implemented is shown in Section 5.1. Second, the extraction of the collocations is explained (§5.2.). Third, the verbs which produced collocations according to all eligibility criteria are displayed (§5.3.), as well as the verbs whose collocations conveyed more than one meaning. Thus, the five verbs which represented actual motion and fictive motion are further explained in Section 5.4., which contains five other sections analyzing both meanings of each verb. The closing section of this chapter addresses the implementation of collocations in the specialized dictionary of adventure tourism (§5.5.).

Chapter 6 concentrates on the extensive data obtained in this research and performs a comprehensive analysis in terms of several aspects. First, Section 6.1. studies the keyness and collocation production of the motion verbs. Second, Section 6.2. focuses on the strength and frequency of the collocations selected. Third, the lexical types of these word combinations are examined in Section 6.3. Finally, this chapter is closed with a thorough analysis of the meaning of the bases and their co-occurrents in terms of the theory of Frame Semantics. To this end, the semantic frames evoked by the bases are addressed as well as the semantic roles expressed by their collocates. Ultimately, the span between bases and collocates is discussed (§6.4.).

Last of all, Chapter 7 outlines the conclusions that can be drawn from this research and identifies the limitations revealed. Furthermore, it suggests a perspective of future work.

PART I:

THEORETICAL FRAMEWORK

CHAPTER 1. TERMINOLOGY AND SPECIALIZED LANGUAGES: A STATE-OF-THE-ART OVERVIEW

Terminology or specialized language is more

than a technical or particular instance of

general language.

(Faber Benítez & López Rodríguez, 2012, p. 9)

Chapter 1 presents part of the theoretical framework of the research conducted in this doctoral dissertation and concentrates on several linguistic theories. To begin with, we focus on the difference between *general language* and *specialized language* (§1.1.) and cover the three components which characterize the latter, namely, linguistic, pragmatic and functional. Section 1.2. deals with the dichotomy between *Lexicology* and *Terminology* and concludes that there are two perspectives upon the status of Terminology, which are derived from the distinction between *word* and *term*. After that, Section 1.3. is motivated by the adverse reactions against the Traditional Terminology and briefly explores four theories which aimed to improve the approach to terminology,[4] namely, Socioterminology, Communicative Theory of Terminology, Sociocognitive Terminology Theory and Frame-Based Terminology. Section 1.4. centers on the applied branches of Lexicology and Terminology, that is, *Lexicography* and *Terminography*, and describes the birth of the so-called *specialized lexicography*. Finally, the last section of this chapter (§1.5.) outlines other linguistic branches which can improve the products of specialized lexicography, such as Cognitive Linguistics and its offshoot Frame Semantics, Corpus Linguistics and Computational Linguistics.

1.1. The dichotomy between general and specialized language

Specialized language (SL) is usually described in opposition to general language (GL) (Bowker & Pearson, 2002, p. 25; Bajčić, 2017, p.28). It is generally acknowledged that

[4] It must be noticed that *Terminology* (upper case) designates the theory underlying the study of *terminology* (lower case), which is the vocabulary of a specialized language.

GL is the language used in everyday conversation to talk about ordinary things in diverse common situations. In contrast, SL is the language used to discuss specialized fields of knowledge (if there can be different specialized fields of knowledge, there can be more than one SL), which consists of specialized vocabulary that a native speaker of a language is not required to know. Since specialized vocabulary is an essential feature of SLs, there is a discipline aimed at collecting and describing it, that is, Terminology (cf. §1.2.).

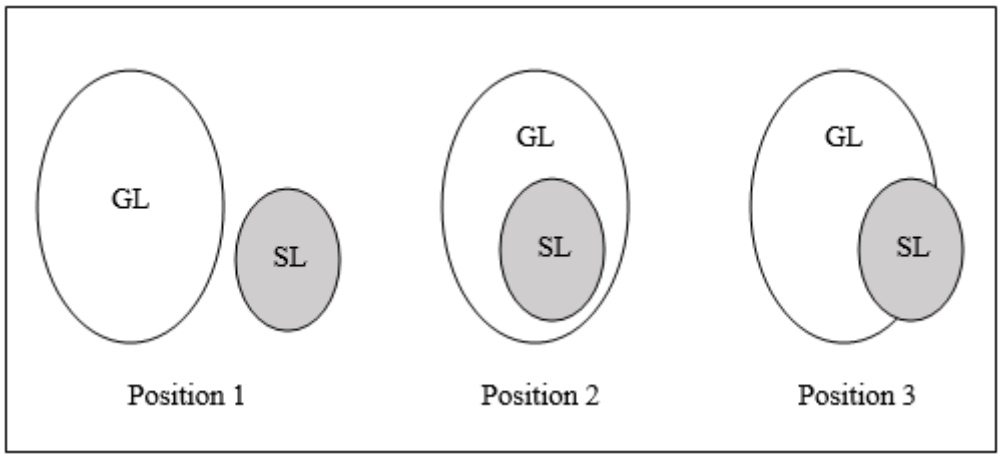
Despite this basic distinction, descriptions of SLs as opposed to the GL abound, given that the two concepts resist clear-cut definitions. According to Cabré Castellví (1999, p. 59), the vast majority of speakers of a language share a common set of rules, units and restrictions, which constitute the GL. In contrast, languages also consist of multiple sets, called *subcodes*, which are characterized by specific elements, such as the sort of participants, their intentions, the situation, the context, the type of communicative situation, and so on. These subcodes are what generate an SL. Therefore, we agree with Cabré Castellví (1999), Alcaraz Varó (2000) and Pérez Hernández (2002) that SLs cannot be defined only in linguistic terms, but they are also characterized by pragmatic and extralinguistic features.

To put it differently, communicative situations shape SLs, as the features of the latter depend on the former. Thus, SLs focus on restricted thematic areas, which normally escape the general knowledge of speakers. However, it may happen that the topic of the situation was dealt with in everyday communication, for this reason, not all subject fields exhibit the same degree of specialization or abstraction. To this respect, Rondeau (1983) distinguishes three groups: (1) highly SLs, found in chemistry, physics, mathematics and architecture; (2) languages with a small degree of specialization, for example, the language found in hairdressing, sports and restoration; and (3) languages whose degree of specialization is between those two, for example, the one detected in the areas of banking, law, economics and stock exchange. Regarding the users of SLs, they are fewer in number when compared to standard language users and have a more complete knowledge of the main concepts belonging to the field. Finally, in terms of the channel where specialized communication can happen, it can be in written form (e.g., specialized books or scientific publications) and in oral form (e.g., workshops, seminars, conferences or research institutions).

Following De Beaugrande (1987), there are three approaches to the concept of SLs, as it is illustrated in Figure 1. First, authors like Hoffmann (1979) state that they are different from the GL because they consist of specific rules and units. Furthermore, they are limited by specific subjects and intentions, thus considering extralinguistic factors. Second, SLs are considered a lexical variant of the GL, such as sociolects or dialects are. More specifically, this perspective contends that SLs acquire their status because specific sets of vocabulary exist (Rey, 1976; Rondeau, 1983). The third view believes that SLs are derived from the GL but used in special contexts in terms of the field to be covered, which implies the consideration of pragmatic features (Picht & Draskau, 1985; Sager, Dungworth & McDonald, 1980).

Figure 1

Positions on the Concept of Specialized Language With Respect to General Language (De Beaugrande, 1987)



In addition to these three positions, a fourth possibility is acknowledged by Bergenholtz and Tarp (1995, p. 17), who explain that linguists adopting a lexically oriented approach consider all GL expressions to be found in SLs, thus, the GL is a subset of SLs. However, we agree with the third position suggested by De Beaugrande, since we believe that an SL will make use of part of the GL, for example, some grammatical constructions and GL words, but it will also incorporate special technical terms and even register-specific structures, so these are the reasons for their intersection. A well-known author who also addresses the issue of SLs in accordance with this view is Cabré Castellví (1999, p. 64). In her opinion, two factors define *special languages* (this is the term she uses to refer to

SLs): first, the topic to be explored (according to Picht & Draskau (1985), this is the feature that defines an SL), whose degree of specialization can vary, and second, the pragmatic circumstances in which an SL is used. The author (p. 66) perceives SLs as sets which share features with the GL, as it is shown in Figure 2:

Figure 2

Intersection Between General Language and Specialized Languages (Cabré Castellví, 1999)

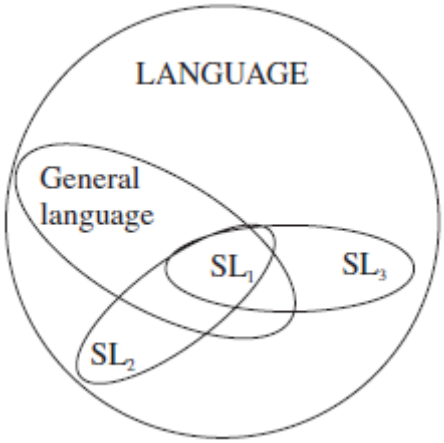


Figure 2 displays several SLs intersecting with the GL. It can be explained if we consider that a wide range of subject fields can be found, therefore, we can talk of special languages in plural.[5] In addition, different SLs can share some common features, for this reason, they also cross at a particular point. Our stance on this matter is that the GL is a system that consists of a multitude of functional subsystems, that is, the SLs, located at the same level. Accordingly, the language of tourism, which is the SL under study in this dissertation, must be regarded as a functional subsystem that occupies a place (it being as special as the one hold by any other SL) within the GL. Then, the next section will perform a more detailed analysis of the concept of SLs and will present the characteristics they share with the GL as well as the aspects that make them specialized.

[5] This idea is also supported by Bergenholtz & Tarp (1995, p. 18).

1.1.1. Specialized languages

First thing to remember is that the GL is part of the knowledge of native speakers of a language. It is used in everyday situations and consists of rules and units which users know. On the other hand, SLs make use of aspects of the GL (e.g., grammar, vocabulary), but they are used within specific pragmatic frameworks to provide specialized information. In other words, the purpose of SLs is to facilitate communication when discussing a specialized subject field. However, users do not need to be experts on the field, as an SL is voluntarily acquired, contrary to the GL, which is inherited and learned at the core of any linguistic community. With this in mind, we agree with Cabré Castellví (1999, p. 64) that an SL must be defined according to the following factors: the subject field and its pragmatic criteria.

With respect to the subject field, Kocourek (1982) contemplates five: (1) pure sciences, (2) experimental sciences, (3) applied sciences and engineering, (4) subjects viewed from the production standpoint, and (5) subjects viewed from the consumer standpoint. On the other hand, in Varantola's (1986) view prototypical subject fields are science, technology, law and religion. The author highlights that, although SLs differ from each other, there are interface areas where they fuse (p. 11). For this reason, we must consider a continuum formed by different SLs which share a central common core, as it is displayed in Figure 2. Then, the core expands in all directions depending on the levels of specialization and abstraction of each SL. To put it another way, we can distinguish various levels of abstraction which result in different types of specialized texts,[6] which are also characterized by their topic or user in a particular communicative situation.

Regarding the central common core that all SLs share, Picht and Draskau (1985) list a series of characteristics which are supposed to belong to this common core, and hence, that are typical of all SLs: (1) they are independent of the GL because when variation occurs in an SL, the GL is not affected; (2) they have a limited number of users, given that they are not naturally learned; (3) users choose to learn these languages, so this is a voluntary acquisition; and (4) they have a single purpose because they are used for communication in specific social settings. Nevertheless, in our opinion, these features are

[6] We concur with Halliday (1978, p. 40, as cited in Wang, 2018, p. 298) that *text* refers to "what the speaker actually says," either in written or spoken form. By the same token, Brezina (2018, p. 15) states that "In general linguistics, the word 'text' is often used as a general term to represent both written and oral texts." Therefore, in this work, *text* encompasses both written and spoken texts.

more noticeable in the highest levels of specialization of an SL, that is, in an expert-expert communicative situation. In situations where the addressee is a layperson, the language used is also specialized, but in a lower degree, as we will see below. For instance, regarding (1), variation exists because less technical terms (synonyms) are used, and with respect to (3), this language is somehow naturally acquired by addressees, since they do not study them on purpose, such as a patient talking to a doctor in a medical setting. From these features, we can conclude that SLs are related via three factors: linguistic (they undergo linguistic changes) pragmatic (type of user) and functional (their primary function is informative), aspects that will be explained in more detail below.

1.1.1.1. Linguistic component

Experts agree that the distinctive feature of SLs is the specific vocabulary they use, known as *terminology*, since this is the element responsible for representing the reality (Alcaraz Varó, 2000, p. 24). Thus, the lexicon of an SL displays specific (and sometimes exclusive) elements that constitute its terminology, that is, lexical units (LUs) that are specific to a specialized field of knowledge, as well as neologisms, loanwords, learned expressions, non-linguistic elements, among others. Terminologies increase SLs' specificity and allow differentiation from other SLs and the GL. However, as it has been previously emphasized, two or more SLs can intersect, which will result in a terminological transfer from one SL to a different SL. For instance, terms belonging to the language of art are frequent in the language of tourism because art is intimately related to activities in the field of tourism, such as the description of monuments, paintings, and so on.

Unlike the lexicon, which is specific to every SL, the grammatical constructions used in SLs are similar to those of the GL, although SLs are characterized by a distinctive style which makes frequent use of some conventional patterns. As an illustration, it is common to find peculiar structures with scarce punctuation in the language of law, for example, *The Policyholder having made to the Corporation a written proposal and declaration which shall be the basis of this contract and having paid or agreed to pay on demand the premium stated in the Schedule the Corporation will provide insurance hereinafter contained in respect of events happening within the Territorial Limits or in the course of transit by sea or air (including hovercraft) between places within the Territorial Limits (including processes of loading and unloading) during the period of*

insurance stated in the Schedule or during any period for which the Corporation may accept payment for the renewal of this Policy,[7] modifiers such as *the said, the aforementioned*, for example, *It recites that it is based on a judicial order, dated 13th February 2004, requiring the arrest and remand in custody of the requested person for the purpose of bringing him before the aforementioned court*,[8] and phrasal verbs in a quasi-technical sense, for instance, *She is seeking a capital sum to enable her to start her life again and allow her to put down a deposit on a house*.[9]

1.1.1.2. Pragmatic component

The pragmatic factor of SLs is the most accepted feature of their anatomy. In this sense, Cabré Castellví (2002, pp. 30-31) suggests two different axes: (1) a horizontal axis determined by the subject field, and (2) a vertical axis determined by the degree of specialization and abstraction of the content. In accordance with the first axis, textual typologies have been constructed according to the field of knowledge that a text is about. Most of these specialized subject fields are not part of language users' common knowledge, for this reason, they are the object of specialized learning, for example, medicine, engineering and law. In addition, it is difficult sometimes to propose a detailed classification of a given text, since it may belong to more than one field of knowledge.

Regarding the second axis, the degree of specialization of a text is a relevant criterion for its classification. This axis consists of two variables: (1) a text's users, namely, the sender and the receiver, and (2) the communicative situation. On the one hand, users of a text may range from experts and knowledgeable people to learners and amateurs. Senders of the message usually belong to the first group, whereas it is common to find receivers in both of them. We can also find professional people who make a living from their specialized knowledge. On the other hand, communicative situations may be in the form of journal publications or academic research as well as newspapers aimed at experienced professionals.

In brief, the combination of these three aspects, that is, subject field, users and communicative situation, represents the pragmatic viewpoint of SLs. They determine the

[7] Example extracted from the British National Corpus (BNC), document ID: HB5 (01/07/2019).
[8] Example extracted from the British Law Report Corpus, document ID: NIHCQB12 (01/07/2019).
[9] Example extracted from the British Law Report Corpus, document ID: SCSHC50 (01/07/2019).

degree of specialization of a text, being the type of user the most decisive factor. Consequently, authors share a tiny or a vast amount of information about terms depending on their degree of specialized knowledge and that of their readers. For instance, highly specialized texts are likely to include more terms than simpler texts, although it is important to realize that the latter can be classified as specialized texts too. Another key point is that the higher the number of terms in a text, the more precise and concise the SL. On the other hand, we have previously mentioned that senders of a message are usually experts on the subject field, whereas receivers can belong to different groups: experts, semi-experts and laypersons. Bearing this in mind, three different situations are suggested (Durán-Muñoz, 2012, pp. 31-33; Pearson, 1998, pp. 36-38), which are summarized in Table 1:

1) *Expert-expert communication*: It is the most specialized type of communication, therefore, the terminology used is also the most specialized and the texts are the lexically densest. The terms (or jargon) are used in a precise manner depending on their specific meaning, which is known and shared by all members of the communicative interaction, that is, they share a common language. These texts are characterized by several features but do not normally include the meaning of the terms, given that both the sender and the receiver are expected to understand the whole lexicon, so when certain terms or phrases are used each person understands what is meant. This communication tends to be established in specialized journals, books and meetings;

2) *Expert to initiates*: In this case, the sender of the message is an expert working within a subject domain, whereas the receiver does not have the same level of expertise, that is, he/she is a lay person with previous knowledge of this or other subject fields, for example, advanced students in third level institutions or people working within the same area but with a different training background. The expert uses the same terminology as he/she would use if communicating with his/her peers, but, in all probability, he/she explains the terms that may be unknown to receivers or that may cause problems to get the message. For this reason, this type of communication is likely to be interspersed with explanations that include the use of more general

vocabulary, that is, non-terms, if necessary. It is found in subject-specific textbooks, conferences, debates and discussion groups; and

3) *Relative expert to the uninitiated*: The term *uninitiated* refers to "adults with a general education who are not necessarily involved, either professionally or through their leisure interests, in a particular subject field" (Pearson, 1998, p. 37). This kind of communication has a much lower term density, although terms are clearly and properly explained when used. Additionally, it is assumed that receivers do not have any prior subject-specific knowledge, so using terms may prevent communication. For this reason, it is common to: (1) use a GL word to describe a concept, and (2) use synonyms for terms (e.g., *stroke* for *ictus* in medicine). This situation is established in informative booklets, specific sections of magazines (e.g., general medicine, psychology, advisory service), science journals and explanatory leaflets.

Table 1

Classification of Specialized Texts and Their Features

Degree of specialization	Sender	Receiver	Channel
High	Expert	Expert	Specialized journals / Books / Meetings
Moderate	Expert	Laypersons with previous knowledge / Students	Subject-specific textbooks / Conferences / Debates / Discussion groups
Low	(Relative) Expert	Laypersons without previous knowledge	Informative booklets / Magazines / Science journals / Explanatory leaflets

From the communicative situations above, it can be inferred that users of an SL acquire two broad types of knowledge (Bowker & Pearson, 2002, pp. 30-31). On the one hand, linguistic knowledge concerns the terms typical of an SL, its phraseology (e.g., collocations; cf. Chapter 2), grammatical structures and stylistic features. On the other hand, conceptual knowledge consists of information about the subject field that is being covered. This type of knowledge is responsible for effective communication, since

making use of a few specialized phrases does not allow communication, but the concepts behind the terms must be understood.

1.1.1.3. Functional component

This is the third element that defines an SL. According to Jakobson (1960, pp. 353-356), the functions of language are six, each of them determined by a different factor, consequently, six are the factors that constitute any act of verbal communication. It is important to realize that few verbal messages fulfill only one function, but it can be stated that the verbal structure of a message depends primarily on the predominant function. Having said that, the functions of language are: (1) referential, centered on the context; (2) emotive, focused on the addresser; (3) conative, oriented toward the addressee; (4) phatic, focused on the contact; (5) metalingual, centered on the code; and (6) poetic, oriented toward the message. Table 2 shows more information about these functions:

Table 2

Jakobson's (1960) Functions of Language

Function	Focus	Aim	Examples
Referential	Context	Leading task of messages; to denote, to provide contextual information	· *Our opening hours are 9am-2pm, Monday through Friday.*
Emotive	Addresser	A direct expression of the speaker's attitude toward what he/she is speaking about	· *Goodness me!* · *Amazing!*
Conative	Addressee	To express our desire so that the addressee does something	· *Drink!* · *Get out of here!*
Phatic	Contact	To establish, prolong or discontinue communication, to check whether the channel works, to attract the attention of the interlocutor	· *Are you listening?* · *Hello, do you hear me?*
Metalingual	Code	To use language to discuss language	· *And what is 'sophomore'?* · *What do you mean by 'plucked'?*
Poetic	Message	To select the correct words	· *"Why do you always say 'Joan and Margery' yet never 'Margery and Joan'? Do you prefer Joan to her twin sister?"*

Given this classification, the primary function of all specialized texts is the referential function, since they aim at transmitting information objectively, subordinating aesthetics and expression to communicative efficiency. However, as it has been aforementioned, not only is one function fulfilled in one verbal message, but other functions of language are also involved, for example, emotive (transmission of addresser's emotions), conative (aiming at the addressee's answer) and poetic (form of the message), which can be considered secondary functions. As a result, the secondary functions of a specialized text will depend on the pragmatic factors explained above, that is, subject field, users and communicative situation.

Overall, we can say that SLs intersect with the GL due to the fact that they include features of the latter but also incorporate new ones, for this reason, no rigid boundaries between both concepts must be established. All SLs are located at the same level with respect to the GL, therefore, our object of study, that is, the SL of adventure tourism, can be approached in the same way as any other SL. Additionally, SLs are described as a composition of three different elements: linguistic, pragmatic and functional, and their major function is referential, although they may serve other functions in terms of the subject field that a specialized text is about (e.g., law, engineering, veterinary), the type of interlocutors involved (e.g., experts, laypersons) and the channel where communication takes place (e.g., academic journal, conference, informal meeting).

At this point, we must also distinguish *specialized language* from *specialized discourse*. The former has just been defined in terms of three aspects: linguistic, pragmatic and functional, and is materialized in texts. The latter, according to Pritchard and Jaworski (2005, p. 4), can be approached from a range of methodological perspectives, starting from a textual view (e.g., Stubbs, 1983; Fasold, 1990; it explains the use of both terms interchangeably) to the conception of discourse as *social practice* (Fairclough, 1992). The last viewpoint includes non-verbal elements (Foucault, 1972), body image (Goodwin, 1981), music (Van Leeuwen, 1993), face-to-face interactions (Coupland & Jaworski, 2001), social communication (Wang, 2018), among others. However, this work does not approach these spheres of communication, but we focus on SLs, understood as functional subsystems which play a bridging role in communication between people and whose essential feature is their vocabulary, that is, their terminology. Thus, the next section will explore the discipline that deals with it.

1.2. The dichotomy between Lexicology and Terminology

Lexicology is a term derived from the Greek word *lexiko*, meaning *word*, and *logos*, meaning *learning* or *science*. The etymological meaning of this word speaks for itself when it comes to the scope of this branch of Linguistics. That is, Lexicology is the science of the word and entirely depends on language being written down. In this context, we consider a *word* the unit of the written language which occurs between two spaces, for example, in *It always rains in England in winter* we find seven words. This science was born "as a way of keeping alive ancient texts whose meanings were beginning to be lost as the language continued to change" (Halliday, 2004, p. 11). Therefore, it describes the general vocabulary of a language in terms of its origin, development and current use, that is, the *lexicon*.

However, not only is the lexicon defined as the vocabulary of a language, but it is also the set of LUs containing information of different types, such as phonological, morphological, syntactic and semantic, which means that Lexicology is inseparable from other linguistic branches, for example, Phonology, Morphology, Syntax and Semantics. Added to this linguistic information, speakers usually store other items of information which determine the real usage of each word, namely, paralinguistic, communicative and extralinguistic. The first type of information can be found in dictionaries, whereas the second usually belongs to speakers' own knowledge. In fact, it is the latter which helps a speaker to decide which word/s he/she will (not) use in a given context.

According to Halliday (2004, pp. 11-13), lexicological studies originated in India as early as the third to second century BC, when glossaries were compiled to explain the difficult words found in the Vedas.[10] They were considered the predecessor of what we would recognize today as dictionaries and, after that, other works were produced. For example, China compiled a thesaurus of 3,500 words in the third century BC, and two centuries later a Homeric lexicon was written. Also, the first dictionary of Farsi-dari, the literary language used in Persia, was published in the ninth to tenth centuries. Nevertheless, one of the greatest works is said to be a tenth-century dictionary which

[10] The Vedas are the holiest books of Hinduism and are considered among the oldest religious works in the world. Their content was not written down at first, but it was passed down orally from generation to generation until 1500-500 BC. Therefore, by the time glossaries to study the language were created, it was already a thousand years old. [https://www.ancient.eu/The_Vedas/] (Last accessed: 25/10/2020).

aimed to explain the etymology and meaning of words. It included around 30,000 entries from literary works in Ancient, Hellenistic and Byzantine Greek and in Latin.

Several theoretical models emerged in the twentieth century to deal with the study of the lexicon.[11] At the same time, interest in the field of Terminology started to grow. This science was characterized for focusing on *concepts* and *terms* of SLs, in comparison to Lexicology, which was centered on words of the GL. According to Cabré Castellví (2000, p. 37), Terminology emerged with specialists feeling the need to enhance communication and transfer their knowledge. The first studies on this matter date back to the eighteenth century, when Antoine-Laurent de Lavoisier and Arnold Adolph Berthold created lists of specialized vocabulary in chemistry, and Carl von Linné created a glossary of terms related to botany and zoology. These works are considered the basis for Terminology. After that, scientific and technical developments in the nineteenth and twentieth centuries encouraged the gradual emergence of new terms to name new concepts (Tebé, 2003, as cited in Durán-Muñoz, 2012, p. 35).

Nevertheless, most experts agree that Terminology as a discipline of study was born in the 1930s with Eugen Wüster. He came from the field of engineering and justified a systematic method to cover the treatment of terminological data in his PhD thesis *Internationale Sprachnormung in der Technik, Besonders in der Elektrotehnik* (1931). In addition, he was the founder of the first theoretical model in Terminology, the General Terminology Theory (GTT), which was developed on the basis of his terminographic experience in compiling *The Machine Tool. An Interlingual Dictionary of Basic Concepts* (1968), a piece of work "intended as a model for future technical dictionaries" (Cabré Castellví, 2003, p. 165). This dictionary gathered 1,401 concepts and showed terms along with definitions in English and French, but only terms in German. Terms and definitions were mainly taken from standardized sources (Campo, 2012, p. 48).

Wüster's conception of terminology was undoubtedly influenced by his experience as an engineer according to two major aspects: first, he belonged to a world where new physical objects, procedures and measurements were introduced in different fields of engineering, and second, he was involved in national and international terminology standardization. It is important to realize that Wüster's theory, commonly known as Traditional Terminology, was based on five central tenets (Temmerman, 2000,

[11] See De Miguel Aparicio (2009, pp. 218-368) for a careful examination.

p. 4): (1) its object of study are concepts without considering language; (2) concepts are distinct and specific and have their own place within a concept system, so they should not be studied in isolation; (3) traditional definitions should be given to concepts, which can be of three types: intensional[12] (ideally), extensional[13] and part-whole; (4) concepts are assigned terms permanently and univocity[14] prevails; and (5) terms and concepts are considered solely from a synchronic point of view.

After having briefly introduced how Lexicology and Terminology emerged, it must be acknowledged that they share three main characteristics (Cabré Castellví, 1999, p. 35): (1) they both use words, (2) they have a theoretical and an applied side, and (3) they are concerned with dictionaries. However, there are other distinguishing features that allow the understanding of Terminology as an autonomous discipline, which are compared and summarized in Table 3 (Durán-Muñoz, 2011, p. 73):

Table 3

Comparison Between Lexicology and Traditional Terminology

	Lexicology	**Traditional Terminology**
Object of study	Words	Concepts
Form	Oral and written	Written
Field of study	Analysis and description of the lexicon	How concepts are perceived, independently of their name
Aim	To understand lexical behaviour	To name concepts and standardize new terms; to improve specialized communication
Point of departure	Form and meaning	Concept, regardless of the form
Approach	Semasiological	Onomasiological
Kind of study	Synchronic and diachronic	Synchronic

As it can be observed, the differences between Lexicology and Terminology revolve around seven aspects. To begin with, the priority in Terminology are the concepts, while

[12] An intensional definition lists all the essential properties of a term, for instance, "a ship is a vehicle used on water."
[13] An extensional definition provides instances of the thing being defined, such as *passenger ship*, *cargo ship* and *battleship* for *ship*.
[14] *Univocity* means that "one concept is referred to by one term (no synonymy) and one term can only refer to one concept (no polysemy)" (Temmerman, 1997, pp. 54-55).

Lexicology focuses on words. As a result, the type of approach followed in each discipline is different, that is, onomasiological (the concept comes first and, then, the term that names that concept) versus semasiological (a word comes first and, then, its meaning), respectively. Additionally, Lexicology is interested in studying the phonic and written dimensions of words from both a synchronic and diachronic perspective, whereas Terminology prefers centering just on the written form from a synchronic perspective. Finally, they pursue different goals. On the one hand, Lexicology aims to comprehend how the lexicon of a language behaves by means of its analysis and description. On the other hand, Terminology strives for understanding how concepts are perceived and finding a name for them so that it will be standardized.

Given these differences between Lexicology and Terminology, the latter is regarded as an independent discipline with its own and original tenets. This belief is derived from the dichotomy between word and term, which seems logical if one thinks of the dichotomy between GL and SL. To put it differently, words belong to the GL and are examined by lexicologists, whereas terms belong to SLs and are studied by terminologists, thus, Terminology can exist without any other discipline. Sager (1990) offers the following distinction between these two concepts:

> The items which are characterized by special reference within a discipline are the 'terms' of that discipline, and collectively they form its 'terminology;' those which function in general reference over a variety of sublanguages are simply called 'words,' and their totality the 'vocabulary.' (p. 19)

Despite Sager's statement, it must be emphasized that, in the same way as the boundaries separating the GL from SLs are not crystal-clear, the boundaries between words and terms can be blurred. More specifically, some lexical items belonging to the GL can acquire a specialized meaning and be incorporated into an SL, such as *window* and *mouse* in the field of computing (Durán-Muñoz, 2012, p. 41); this process is known as *terminologization* (Cabré Castellví, 1993, p. 168). On the other hand, a lexical item that was once attributed a fixed meaning within a specialized domain can be used in the GL, such as *virtual* in computing and *recycling* in environmental sciences; this process is called *de-terminologization* (Meyer & Mackintosh, 2000, p. 112). Nevertheless, authors

hold the opinion that, contrary to Wüster's GTT, the distinction between word and term can only be drawn in context, that is, by observing the real use of the LU in a specific communicative setting. This perspective is a prelude to more modern theories of Terminology which developed as a reaction to the Traditional Terminology. They are the topic of the following section.

1.3. Modern theories of Terminology

According to Campo (2012, p. 72), Wüster's Traditional Terminology provoked both positive and negative reactions. In this section, we focus on the latter, since they are the ones that proposed new approaches to terminology. In general terms, the negative reception of the author's work can be summarized in five points (Campo, 2012, p. 139):

1) First, new theorists disapproved the onomasiological approach to the study of terminology because it disregarded the context of communication. In addition, the boundaries between words and terms were fuzzy, given that the same term may appear in different linguistic contexts;

2) Second, Wüster regarded concepts as elements belonging to a concept system. However, stating whether a given term belongs to an SL or the GL may be confusing, thus, the different dimensions of concepts should be considered as well as their flexibility to be assigned new meanings or to be synonyms for others;

3) Third, the author contemplated three ways of defining concepts, namely, intensional, extensional and part-whole. In his opinion, intensional definitions were the most comprehensive definitions because they showed the position of the term within the concept system. Nevertheless, lay users need encyclopedic definitions which provide them with semantic and lexical information to understand a list of terms;

4) Fourth, modern theories did not accept univocity, which means that one concept is referred to by one term and one term can only refer to one concept, because concepts are not independent but hold a relationship with language and the communicative act. As a result, cases of synonymy and polysemy exist in SLs, especially in activities such as Translation; and

5) Finally, synchronic studies of terms were rejected, given that concepts evolve and their meanings vary. In this sense, the arrangement of concepts within an SL is changeable, so diachronic studies were necessary.

On balance, the disapproval of Wüster's ideas did not try to refute his theory, but to highlight its limitations. In addition, Cabré Castellví (2005, p. 4) emphasizes that new approaches to terminology in the twentieth century were also encouraged by the social changes of the moment, which affected the systems of communication, the status of languages and the specialized terminology in the spreading of knowledge. For example, some of the aspects that highly influenced the emergence of new theoretical and practical perspectives on the field of terminology were: (1) communicative interaction in plurilingual contexts where users from different backgrounds participated, (2) new requirements needed to be met due to the emergence of new specialized fields, (3) specialized knowledge was to be spread worldwide, and (4) new technologies appeared in different professional domains.

A thorough analysis of the various approaches to terminology from Wüster onwards can be found in Campo (2012), Fernández Fernández (2011), Sanz Vicente (2011) and León Arauz (2009). To our mind, the most influential theories are four, which depend on the perspective adopted when trying to improve the GTT, namely: (1) a sociolinguistic perspective: Socioterminology, (2) a communicative perspective: Communicative Theory of Terminology, and (3) a cognitive perspective: Sociocognitive Terminology Theory and Frame-Based Terminology. They will be more carefully examined in the following sections.

1.3.1. Socioterminology

Socioterminology can trace its origin to the 1970s in France and Quebec (Canada). Its name was first used at the beginning of the 1980s by Jean-Claude Boulanger and the discipline matured in the 1990s, when some of the most relevant publications appeared (Boulanger 1991, 1995; Gambier 1991, 1993; Gaudin 1993, 2003). The strongest criticism coming from Socioterminology against the GTT was that terminology is considered a phenomenon that occurs in specialized communication. Authors were greatly concerned about the real use of language influenced by social factors, so this approach was directed toward the study of terms in connection with the contexts in which

they appeared, that is, linguistic, pragmatic, social and historical. In this regard, Boulanger (1995) points out that there is "l'obligation de tenir compte de l'existence de l'**usage reel**"[15] (p. 196).

On the other hand, two main ideas promoted by Socioterminology are: (1) they adopt a descriptive rather than prescriptive perspective of languages, that is, they aim to describe languages rather than standardize terms; however, it must be emphasized that they try a prescriptive approach to dictionary making, and (2) they develop a semasiological rather than onomasiological approach, in other words, they base their analyses on the term (the form), which they study first, instead of the concept (the meaning). Apart from that, this theory aims to confront some major problems created by Wüster's GTT (Boulanger, 1995; Campo, 2012):

1) Unlike the GTT, it favours a diachronic study of language, given that it is in constant change and evolves to describe new realities. Thus, it affirms the existence of terminological variation;

2) The main consequence of the previous idea is that concepts are not static, therefore, standardization is a chimera;

3) Concepts do not belong to an only term and terms do not belong to an only concept, as a result, synonymy and polysemy need to be studied;

4) In some cases, specialized subject fields may lack absolute limits, thus, the distinction between words and terms is not clear-cut; and

5) Not only is terminology found in written language, but oral language is also used to communicate ideas, therefore, both dimensions of language must be observed.

Overall, one could say that the major issue highlighted in Socioterminology is the incorporation of social and communicative factors into the description of a language, given that it opened the door for other theories of Terminology, such as the Communicative Theory of Terminology.

[15] "the obligation to take into account the existence of **real use**" (translation mine).

1.3.2. Communicative Theory of Terminology

The Communicative Theory of Terminology (CTT) emerged in Spain in the 1990s with the publication of works by Cabré Castellví (1993, 1999, 2000, 2003, 2005). In her opinion, her theory starts from two assumptions: first, terminology is, at the same time, "a set of needs, a set of practices to resolve these needs, and a unified field of knowledge" (Cabré Castellví, 2003, p. 182), and second, the terminological units (TUs) are the elements of terminology. Regarding the first assumption, terminology is, firstly, absolutely necessary to broadly disseminate specialized knowledge; secondly, terminology allows the development of products targeted at all professions so that this specialized knowledge can be shared, and thirdly, Terminology is a discipline whose object of study is the TU. Thus, turning to the second assumption, TUs belong to terminology and have three different sides: cognitive, linguistic and communicative, given that they are, at the same time, units of knowledge, units of language and units of communication.

The cognitive component of TUs is observed in the following features (Cabré Castellví, 2003, p. 184):

a) They depend on a thematic context;

b) They belong to a conceptual structure and their meaning, which is fixed and considered one of their properties, is given by the place they occupy inside this structure; and

c) They are fixed and spread thanks to professional users.

The linguistic component of TUs is present in the following characteristics (Cabré Castellví, 2003, p. 184; 2005, pp. 8-9):

a) They are LUs, which explains the processes of terminologization and de-terminologization. In fact, they are units of the GL that become TUs when they spread specialized knowledge, given the pragmatic features of the discourse in which they are used;

b) Given the previous idea, TUs may be similar to LUs analyzed from a phonological, morphological or syntactic perspective, however, they are different in terms of their semantic and pragmatic dimensions (Cabré Castellví, 2003, p. 189);

c) They can have lexical and syntactic structure;

d) They can be nouns, verbs, adjectives or adverbs and go through word formation processes; in addition, they can combine with other units, such as collocations;

e) They allow denominative variation, therefore, polysemy and synonymy can take place; and

f) They are not independent from the lexicon that a language's user has, but they own a specialized value that is activated in some communicative contexts.

Finally, the communicative component of TUs is found in the following features (Cabré Castellví, 2003, pp. 184-185):

a) They occur in specialized discourse, either oral or written, which is an organized structure of knowledge;

b) They adapt to a given specialized discourse[16] according to their thematic and functional characteristics;

c) They are learned rather than naturally acquired, so specialists in their field are the ones who can handle them; and

d) They are basically denotative, although connotations are not excluded.

These components of TUs demonstrate that the author is especially concerned with the pragmatic dimension of SLs, that is, the sphere where the cognitive (concept), the linguistic (term) and the communicative (situation) functions of Terminology converge. It represents what she calls *the theory of doors*, a model that explains the plural access to the TU, as whether starting from the concept or the term or the situation the central object (the TU) is directly addressed (Cabré Castellví, 2003, p. 186). More specifically, Cabré Castellví approaches the units through the door of language, that is, the term, which explains her semasiological approach to terminology (vs. the onomasiological approach defended by Wüster). Finally, another key point, in this case related to phraseology, is that lexical combinations in specialized domains are considered units of specialized knowledge (Cabré Castellví, 2003, p. 189), which supports the notion of specialized collocations embraced in this dissertation (cf. §2.3.).

[16] It is important to realize that the author uses the word *discourse* instead of *language*.

1.3.3. Sociocognitive Terminology Theory

Unlike the CTT, the Sociocognitive Terminology Theory (STT) approaches the study of terminology from a sociocognitive perspective, that is, it benefits from the findings of cognitive semantics and considers the social setting where language is used. It was born in the late 1990s with the publications by Rita Temmerman. The author aimed to question the principles and methods of the Traditional Terminology through the study of the SL of the life sciences, such as biology, genetics, microbiology, genetic engineering, biotechnology, and so forth. Her methodology consisted in analyzing a corpus of written scientific texts published by subject specialists and targeted at both experts and non-experts on the subject field, for example, specialized articles, dictionaries and books. The reasons for choosing this topic were two, namely: (1) life sciences was a recent and rapidly growing domain in science, and (2) its results were obtained from interdisciplinary approaches (Temmerman, 2000, p. 45).

The cognitive perspective underlying the STT helped to explore the semantic triangle containing the world (reality), the language (a means of communication and of reality construction) and the human mind (the centre of reasoning to understand both the world and the language), and how the three elements interact. More specifically, Termmerman's studies on the life sciences allowed to draw the following conclusions: (1) the world is experiential, in other words, what we know and understand is embodied and results from our sensory perceptions; (2) language has a cognitive, textual and communicative function; and (3) the world is not just perceived objectively, but humans can create *categories* in the mind. At this point, it must be clarified that categories are *units of understanding* (concepts in Wüster's words) which show prototype structure (cf. §1.5.1.). Temmerman's conclusions were a reaction against the way that the GTT regarded the semantic triangle (world, language and human mind), which the author (1997, p. 54; 2000, p. 60) explains as follows:

> The **world** was reduced to what could be objectified, **language** was only considered in its naming capacity, and the **human mind** was only given credit for its capacity to classify individual objects on the basis of recognizing characteristics which were common to all the members of the class representing a concept.

However, according to Temmerman (1998, p. 78), the GTT's idea which deserved more unfavorable criticism was its primary objective, that is, the standardization of terminology, which constricted language. In particular, univocity ideal in the Traditional Terminology was not valid in the analysis of the SL of the life sciences, given that both polysemy and synonymy were functional (in all SLs, in general) and language was used figuratively too (Temmerman, 1997, p. 67), so the idea of one concept-one term was wrong. The main reason for these phenomena was language evolution, that is, the language used by a community evolves at the same time as new things appear, so the meaning of already existing words changes simultaneously; it also explains the fact that language must be studied from a diachronic perspective. Apart from that, some other observations by the STT in contrast to the principles formulated by Wüster and applied by his followers are summarized in the next lines (Temmerman, 1998, p. 79; 2000, p. 16):

1) Language must be considered in Terminology, so the concept should not be the starting point (semasiological vs. onomasiological approach);

2) Categories are not sharply delineated, so clear differences between concepts may not exist;

3) Definitions are not likely to be intensional, since a definition may vary in accordance with a number of parameters, such as the type of category or the level of specialization of the message's sender and receiver;

4) Univocity does not exist, that is, monosemy is not a *sine que non* in SLs, given that polysemy and synonymy frequently occur;

5) Categories, concepts and terms evolve over time, so they must be studied diachronically; and

6) The relationship between concept and term is not arbitrary, but terms are motivated, as cognitive models play a role in the development of new ideas; more specifically, relations between concepts are represented in the form of idealized cognitive models (ICMs) (Lakoff, 1987).

One of the most significant contributions of the STT to the study of terminology is the *termontography* (a term coming from terminology, ontology and terminography), defined as "a multidisciplinary approach in which theories and methods for multilingual

terminological analysis of the sociocognitive approach [...] are combined with methods and guidelines for ontological analysis" (Temmerman & Kerremans, 2003, p. 4). To put it differently, it is considered a methodology to explain ontologies with (multilingual) terminological data in order to manage knowledge and build comprehensive lexical resources which help to face communication problems. A positive aspect of this methodology is that it can be employed in any specialized domain, such as in the context of welfare professionals in the European Union (De Baer, Kerremans, & Temmerman, 2006) and the field of competencies and learning strategies (De Baer, Meersman, & Temmerman, 2009).

1.3.4. Frame-Based Terminology

Like the STT, Frame-Based Terminology (FBT) (e.g., Faber Benítez, 2011, 2015; Faber Benítez & León Arauz, 2010, 2019; Faber Benítez, Márquez Linares, & Vega Expósito, 2005; Faber Benítez et al., 2006; Ureña Gómez Moreno & Buendía Castro, 2017) presents a cognitive view of terminology, but, in contrast to the previous approach, specialized knowledge is organized in semantic frames (cf. §1.5.1.1.) instead of ICMs. The potentiality of frames and their power to create representations which are independent from the language under study have been extensively proved by Pamela Faber Benítez in recent years, leader of the *LexiCon Research Group* at the University of Granada (Spain), and the members of her research group, who have applied this theory to the design of EcoLexicon,[17] a multilingual terminological knowledge base dealing with the specialized domain of the environment (cf. §3.3.).

Some of the reactions of the FBT against the GTT are (Faber Benítez & López-Rodríguez, 2012): (1) the GTT did not take into account the syntax and pragmatics of SLs, (2) concepts were regarded as independent from terms, (3) language variation was not contemplated, (4) standardization involved the synchronic study of terms, and (5) the reference between terms and concepts was considered to be monosemic. On the other hand, the FBT shares many of the same premises as the CTT and STT. For instance, one of its basic assumptions is that words and terms belong to a continuum, so no absolute distinction can be drawn between them. Additionally, context is necessary to study specialized knowledge units, so texts, which are used to spread knowledge, are the main

[17] http://ecolexicon.ugr.es/en/index.htm (Last accessed: 25/10/2020).

source to observe their behaviour. Moreover, the FBT incorporates other psychological and linguistic models and theories, such as the Lexical Grammar Model (Faber Benítez & Mairal Usón, 1999), Frame Semantics (cf. §1.5.1.1.), the Generative Lexicon (Pustejovsky, 1995) and Situated Cognition (Barsalou, 2003, 2008), and is interested in the combinatorial value of terms, which is also the prime goal of this dissertation.

The fundamental aspects explored within this theory are: (1) conceptual organization; (2) multidimensional nature of TUs, that is, communicative, linguistic and cognitive dimensions (Cabré Castellví, 2003, p. 183; cf. §1.3.2.); and (3) the use of dictionaries and specialized corpora for the extraction of semantic and syntactic information. Regarding the first item, the FBT organizes knowledge in frames, which are understood as representations of elements and entities associated with a scene, situation or event that are based on our experience. These frames belong to categories that encompass specific concepts whose semantic relations are organized in a network and are linked by both vertical (hierarchical) and horizontal (non-hierarchical) relations. This way, frames "become large-scale representations that link categories by means of semantic relations" (Faber Benítez, 2015, p. 15). This sort of organization can fit specialized knowledge areas, such as the environment in the EcoLexicon knowledge base.

With respect to the second focus of the FBT, the multidimensional nature of TUs stems from the three views represented in a term, that is, symbolic, conceptual and referential, which allow to make a distinction among three aspects: communicative, linguistic and cognitive. First, terminology is used to transfer specialized knowledge in one or more languages. Second, from a linguistic point of view, a term consists of three dimensions: form, meaning and referent; to explain, (1) form is related to the appearance of a word and the processes and rules of word formation, (2) the meaning evoked by a term is linked to other meanings with which it holds a relationship, and (3) referents of terms are represented in reality by objects. Third, the cognitive aspect of terms is the most complex item to deal with, since different people understand objects and build concepts in different ways.

The third essential aspect in the FBT is the use of dictionaries and specialized corpora for the extraction of semantic and syntactic information. On the one hand, dictionaries provide definitions of terms, but more than one should be used in order to compare the conceptual relations activated in each definition and, after that, craft a well-

structured definition encoding the relations and attributes of the concept under study. Besides, dictionaries offer conceptual parameters to create frames for conceptual categories and their members. On the other hand, linguistic clues can be used with corpora to retrieve conceptual information on semantic relations. Next, this information helps to place the concept within the conceptual framework of a knowledge domain. Additionally, word lists can be extracted from corpora to know which LUs are more recurrent in a text and, then, concordance lines can be checked to obtain the context and investigate patterns of combination.

Some of the most recent works following the FBT's premises are Faber Benítez and Cabezas García (2019) and Rojas García and Faber Benítez (2019).[18] The use of frames for developing conceptual representations in specialized domains can also be observed in Durán-Muñoz (2012, 2016), Peruzzo (2014) and Pimentel (2014).

Section 1.3. has explained the reasons why some modern theories of Terminology (i.e., Socioterminology, CTT, STT and FBT) emerged as a reaction to the GTT postulated by Wüster. Thus, it must be highlighted that the main responses focused on: (1) a semasiological approach to the study of terminology, that is, the starting point is the term; (2) the status of terms within the GL, given that they are words which become terms when occurring in SLs; (3) definitions of terms containing encyclopedic information (vs. being intensional), since terms can be better understood if the user accesses his/her encyclopedic knowledge; (4) the rejection of univocity and the presence of synonymy and polysemy in SLs; and (5) synchronic as well as diachronic studies due to the existence of language variation.

On balance, modern Terminology is placed half-way between Lexicology and the Traditional Terminology, since it combines aspects of both. At this point, we must get back to Table 3 and include modern Terminology, as it is displayed in Table 4:

[18] For more publications by the members of the *Lexicon Research Group*, visit http://lexicon.ugr.es/publications (Last accessed: 25/10/2020).

Table 4

Comparison Between Lexicology, Traditional Terminology and Modern Terminology

	Lexicology	**Traditional Terminology**	**Modern Terminology**
Object of study	Words	Concepts	Terms
Form	Oral and written	Written	Oral and written
Field of study	Analysis and description of the lexicon	How concepts are perceived, independently of their name	Analysis and description of terms and concepts
Aim	To understand lexical behaviour	To name concepts and standardize new terms; to improve specialized communication	To study the relationship between terms and concepts; harmonization of terms; to improve specialized communication
Point of departure	Form and meaning	Concept, regardless of the form	Term and concept
Approach	Semasiological	Onomasiological	Semasiological
Kind of study	Synchronic and diachronic	Synchronic	Synchronic and diachronic

As we can see, modern Terminology coincides with Lexicology in the object of study, that is, the written form. Consequently, the type of approach followed is also semasiological. Similarly, it is interested in oral and written language as well as synchronic and diachronic studies. Finally, the goal of modern Terminology is the harmonization of terms rather than their standardization, although it also endeavors to improve specialized communication like in the Traditional Terminology. One way of doing so is to provide terminological resources which help to understand the vocabulary of SLs so that both experts and non-experts in a subject field can communicate. This is the applied part of Terminology, which is called Terminography, and is usually viewed in contrast with Lexicography, which is the applied part of Lexicology. Thus, the dichotomy Lexicography/Terminography is the focus of the following section.

1.4. The applied branches: Lexicography/Terminography and the birth of specialized lexicography

[...] la lexicología es a la terminología lo que la lexicografía a la terminografía.[19]

(Pérez Hernández, 2002, Section 3.3)

Most authors usually make a parallelism between Lexicography and Terminography in the same way as Lexicology and Terminology. In fact, it is commonly accepted that the former pair refers to the applied branch of the latter, respectively. Hence, we can state that Lexicology is the discipline that analyzes and describes the lexicon of a given language and Lexicography deals with its application in the form of general resources. On the other hand, Terminology studies TUs and Terminography deals with the compilation of specialized resources. From this perspective, it could be said that Lexicography and Terminography aim at the same objective, that is, the description of usage. For instance, Lexicography deals with the compilation and editing of general dictionaries as well as wordlists, glossaries, thesauruses and computerized word banks, and Terminography deals with the documentation of the terminology of specific subject fields, for example, physics, mathematics, chemistry, veterinary, and so on, in order to provide terminological or technical resources. Consequently, the approach of Lexicography is word-oriented, whereas the approach of Terminography is subject-oriented (Alberts, 2001, pp. 75-76).

In the 1980s, when the fundamental assumptions of the Traditional Terminology were still accepted, Lexicography and Terminography were considered two dissimilar areas of research. More specifically, Rondeau (1983) drew four basic distinctions: (1) their object of study, which were words for lexicographers and terms for terminographers; (2) Lexicography was considered to adopt a descriptive approach because it aimed at the documentation, description and preservation of a language in all its facets and registers, whereas Terminography was said to be prescriptive, since it aimed at the standardization of terms so as to avoid any ambiguity, polysemy or metaphorical meaning within a specialized subject field; (3) the methodology followed in Lexicography was semasiological, while the methodology followed in Terminography at that time was

[19] "[...] lexicology is to terminology what lexicography is to terminography" (translation mine).

onomasiological; and (4) lexicographers, contrary to terminographers, did not see multi-word units (MWUs) as a whole. Needless to say, these differences are intimately linked to the differences stressed between Lexicology and the Traditional Terminology (they were summarized in Table 3, cf. §1.2.). Additionally, Bergenholtz and Kaufmann (1997, pp. 92-93) and Bergenholtz and Tarp (2010, p. 28) explain other differences made between Lexicography and (traditional) Terminography, which are gathered in Table 5:

Table 5

Differences Between Lexicography and (Traditional) Terminography

	Lexicography	**(Traditional) Terminography**
Aim	To decode texts	To encode texts
Dictionary format	Printed	Electronical
Dictionary macrostructure	Alphabetic	Systematic
Dictionary purpose	Text reception and translation	Text production
Target group	Layman	Expert
Method	Diachronic	Synchronic
Polysemy	Yes	No

As it can be observed, these differences are largely related to the types of resource developed in both disciplines. Thus, lexicographers usually provide printed dictionaries which have an alphabetic macrostructure, include polysemous words and are used to decode and translate texts, while terminographers design electronical dictionaries which have a systematic (e.g., terms can be hyperlinked to related entries) macrostructure, exclude polysemous words and are used to produce texts. Apart from that, the target group in Lexicography is the layman, whereas it is the expert in Terminography. Finally, lexicographers work with diachronic methods for dictionary compilation, but terminographers work with synchronic ones. However, Bergenholtz, Kaufmann and Tarp present these arguments with the aim of showing that they can be easily rejected.

In 1996, the *Centre for Lexicography – Research into Needs-Adapted Information and Data Access*[20] was established at Aarhus University (Denmark). Its goal was to conduct research within the field of lexicography so as to develop lexicographical tools targeted at all types of users. Currently, the centre director is S. Tarp and his co-workers are H. A. Christensen, H. Bergenholtz, P. Leroyer and S. Nielsen. The group is remarkably productive in terms of international publications, being the *Manual of Specialized Lexicography* (Bergenholtz & Tarp, 1995) one of the most significant. In the first chapter of the book, the authors recognize that both "*terminology*" and "*terminography*" might have been used in the title "without the manual thereby becoming essentially different in terms of contents and structure" (p. 10). It means that, in their opinion, specialized lexicography and terminology/terminography have a great deal in common. More specifically, they regard terminology as an integral part of the former (p. 11) and they "see terminography as a synonym [for] specialized lexicography" (Bergenholtz & Tarp, 2010, p. 29) (the last idea is also supported by Bergenholtz, 1995, and Wiegand, 1995, as cited in Bergenholtz & Kaufmann, 1997, p. 91).

Accordingly, the rapprochement between Lexicography and (traditional) Terminography is called specialized lexicography and is explained as follows (Bergenholtz & Tarp, 1995, p. 11): (1) it works with terms, which are special words of lexicography; (2) it works with both systematic and alphabetic macrostructures; (3) it needs to be both descriptive and prescriptive; (4) it addresses itself to laypersons and experts alike; and (5) it compiles dictionaries for both encoding and decoding purposes. The ideas of this new Terminography are supported by the modern theory of lexicographic functions, developed by the research group of the Centre for Lexicography since the early 1990s and whose two main postulates are: first, lexicography is an independent scientific discipline whose subject field is dictionaries, and second, dictionaries are utility products created to satisfy certain human needs in specific-user situations (Bergenholtz & Tarp, 2003, p. 172).

The Function Theory of Lexicography (Bergenholtz, 1996, 1998; Bergenholtz & Kaufmann, 1997; Bergenholtz & Nielsen, 2002; Bergenholtz & Tarp, 2002; Tarp, 1992, 1994, 1995, 1998, 2000, 2001, 2002, 2009) endeavors to formulate the basic principles underlying lexicographical work so as to achieve successful outcome in the final product.

[20] https://cc.au.dk/en/Centre-for-Lexicography/ (Last accessed: 25/10/2020).

The theorists suggest that users do not need general information to be included in the different resources, but they need information that address their needs. In this way, accurate information must be supplied in terms of the potential users and the communicative situations where the resources will be used, given that dictionary consultation aims to meet needs, as it is sustained by Tarp (2009): "Dictionary consultation takes place when users with a specific type of need occurring in a specific type of extra-lexicographical situation think that this type of need can be satisfied by consulting a dictionary" (p. 278).

In this work, we adopt this stance and use the term specialized lexicography to refer to the discipline which deals with the creation of terminological resources that aim at effective communication in specialized situations. In addition, we agree with the assumptions made by the authors of the Centre for Lexicography, therefore, we think that the approach to terminology must be descriptive and prescriptive, depending on the purpose of the project. Furthermore, resources must be targeted at any user, from experts to laypersons, so that communication can be established in all subject-specific situations. Finally, we believe that the transition from general to specialized knowledge is necessarily a gradual transition with no sharp boundaries, so "a certain grey area may exist between general and specialized lexicography" (Fuertes Olivera & Tarp, 2014, p. 8). Last but not least, it must be emphasized that specialized lexicography can benefit from distinct areas of research for the creation of tools that satisfy the users' needs, such as Cognitive Linguistics, Corpus Linguistics and Computational Linguistics. They are the central point of the last section of this chapter.

1.5. Linguistic branches which enhance specialized lexicography

The twentieth century witnessed the development of several linguistic branches which allowed new approaches to terminology. First, some of the principles of Cognitive Linguistics were embraced across the world and reached a large variety of disciplines, in particular, specialized lexicography found a powerful frame to develop knowledge structures and forms of representation. Second, Corpus Linguistics was of supreme importance as it permitted analyses based on corpora and the selection of materials showing real language use. Finally, Computational Linguistics represented a real breakthrough for being an interdisciplinary field where linguistics and computing were

involved, thus, multiple tools for a variety of branches were developed, for example, automatic translation, language teaching and specialized lexicography. The following sections will delve into each of these linguistic branches.

1.5.1. Cognitive Linguistics

Cognitive Linguistics emerged in the late 1970s and early 1980s from the work of a few researchers who were interested in the relationship between language and the human mind. The ideas that promoted this science quite differed from the ones postulated by structuralists in the 1920s, such as the arbitrariness of the sign, and generativists in the 1950s, who stated that grammar is a system of rules which allows combinations of words. The gist of Cognitive Linguistics is that linguistic knowledge is "part of general cognition and thinking" (Ibarretxe-Antuñano, 2004, p. 3), so language belongs to our cognitive capacities. This principle is opposed to Generative Grammar's belief that language is an autonomous cognitive component, and it is characterized for considering that human beings use the same cognitive skills both in the representation of linguistic knowledge (knowledge of meaning and form) and in other domains outside language, such as visual perception, reasoning or motor activity (Croft & Cruse, 2004, p. 2).

It must be emphasized that Cognitive Linguistics is not one clearly delimited domain, but an amalgamation of diverse areas of linguistic research into semantics, syntax, morphology, language acquisition, phonology, and so on.[21] Although inspirational ideas started in the late 1970s, the publication of two scholarly pieces of work in 1987 is considered a milestone in the birth of this field of study, namely, *Women, Fire and Dangerous Things* by George Lakoff and *Foundations of Cognitive Grammar* by Ronald Langacker (first volume in 1987 and second volume in 1991). On the one hand, Lakoff argued that not only was metaphor a formal aspect of language, but it was also a mechanism that allowed people to structure concepts according to their bodily experience of the world. Consequently, there are central metaphors that influence people's lives. On the other hand, Langacker was the founder of the Cognitive Grammar theory, which considers that the basic unit of language is the pairing of a semantic structure plus a phonological label.

[21] For more information on cognitive linguistic theories, see Evans, Bergen, & Zinken (2007).

Apart from these two authors, Cognitive Linguistics is greatly indebted to Charles J. Fillmore ("Frame Semantics," 1982) and Leonard Talmy ("Force Dynamics in Language and Cognition," 1988). On the one hand, Fillmore developed the theory of Frame Semantics (cf. §1.5.1.1.), according to which words do not relate to each other but to frames that provide speakers with the conceptual base to determine words' meanings. These frames entail the description of a type of event, relation or entity and the participants in it. On the other hand, Talmy was concerned with the properties of conceptual organization and cognitive theory and was particularly interested in natural language semantics, namely, typologies and universals of semantic structure, the relationship between semantic structure and formal linguistic structures (lexical, morphological and syntactic) and their relation to discourse, development, impairment, culture and evolution.

Although Cognitive Linguistics covers different areas of research, scholars find themselves on common ground on the concept of language. In their opinion, language is an instrument "for organizing, processing, and conveying information" (Geeraerts, 2006, p. 3), which turns it into a basic tool in the cognitive process. Taking this idea as a starting point, the central tenets of Cognitive Linguistics can be determined as follows (Cuenca & Hilferty, 1999, p. 181; Langacker, 1987):

1) Language is not an autonomous cognitive faculty, but it relates to other human cognitive faculties. For example, linguistic structures are closely connected with knowledge and thinking and must be understood in relation to the communicative function of language (this idea was implemented in Cabré Castellví's CTT, cf. §1.3.2., and the FBT, cf. §1.3.4.);

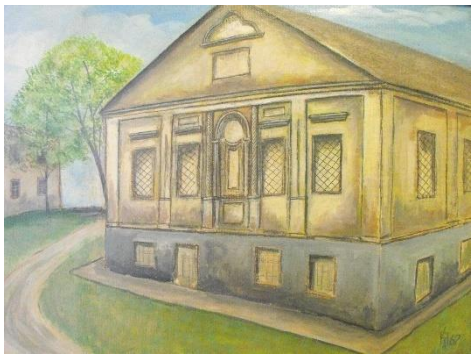
2) Language is dynamic, so it can change in terms of how it is used, that is, depending on the communicative situation, users, context, and so forth (it explains why terms should not be defined intensionally, i.e., it opposes Wüster's GTT). In addition, language is diffuse, so it has some special semantic and syntactic characteristics that cannot be readily distinguished; and

3) As language is dynamic, the different levels of language cannot be clearly differentiated, for instance, semantics and pragmatics, semantics and

grammar, grammar and lexis. Added to that, the distinctions between linguistic/non-linguistic aspects and denotation/connotation are not clear-cut.

On the other hand, language is a form of knowledge and is considered to be primarily semantic, so it must be analyzed with a focus on meaning. According to Geeraerts (2006, pp. 3-6), the four specific characteristics of linguistic meaning are:

1) Perspectival: Language meaning is used to construe the world and it can embody different perspectives, that is, the same objective situation can be expressed linguistically in different ways. For example, given the following picture,



you can verbalize what you see as *The house is in front of the tree* or *The tree is behind the house*. In the first statement, you are looking at the house, which is the focus of your attention, so its position is what your words are describing. In the second statement, your gaze is directed toward the tree, so the house is an obstacle along that direction and the focus of your attention, the tree, is behind that obstacle;

2) Dynamic: Language meaning is used to shape the world, thus, if the world changes, our use of the language changes too. From this view, language is not regarded as a rigid and fixed structure, but it is adapted to transformations (this assumption explains why terminology should be studied diachronically);

3) Encyclopedic: Language meaning is not a separate module in our mind, but it is influenced by our experience of the world. Likewise, our cultural and social identities can influence how language is used. For instance, Eskimos can perceive hundreds of different shades of white,[22] such as the color of the

[22] https://www.culturamas.es/2010/05/10/los-mil-blancos-de-los-esquimales/ (Last accessed: 25/10/2020).

bears, of falling snow, of icebergs, of clouds, and so on, because snow is omnipresent all year round. However, somebody living in western areas of Europe will say that a polar bear, falling snow, icebergs and clouds are just white; and

4) Usage-based: Languages usually have abstract structures, such as subject + verb + indirect object + direct object in the sentence *Mary sent Peter a message*. But these structures are not observable sometimes, since we only notice a string of words at a more concrete level. Then, our experience of language is based on language usage, not on isolated words or grammar patterns.

One of the main concerns of Cognitive Linguistics are the problems arising from conceptual categorization and the ways in which linguistic knowledge and world knowledge relate. As a result, we find an extensive range of cognitive theories that aim to provide clues about the organization and distinction of concepts. According to Croft and Cruse (2004), *categorization* is one of the most basic cognitive activities and is defined as follows: "Categorization involves the apprehension of some individual entity, some particular of experience, as an instance of something conceived more abstractly that also encompasses other actual and potential instantiations" (p. 74). On the other hand, Cohen and Lefebvre (2005) define this concept as "the mental operation by which the brain classifies objects and events […] is the basis for the construction of our knowledge of the world [...] the most basic phenomenon of cognition" (p. 2). However, a consensus on the way of categorizing has not emerged yet, that is, researchers do not agree on how we order knowledge in our mind or the internal structure of the resulting categories.

The classical model, known as Aristotelic, regards conceptual categories as fixed cognitive entities that associate with linguistic expressions in a stable way. According to this model, conceptual categories are determined by a set of delimited features which are necessary for an entity to be considered a member of that category; in fact, an entity having all those features is guaranteed membership of that category. This view is very close to the one held in Wüster's GTT, since he states that one concept is referred to by one term and one term can only refer to one concept (univocity). It may be the reason why this model also received unfavourable criticism. First, concepts are usually defined in multiple ways and their defining characteristics may not coincide. Second, some

concepts are more representative of a given category than others, so all concepts are not positioned at the same level. Finally, the classical model does not explore the boundaries between categories, which are fuzzy.

Contrary to this model, Eleanor Rosch (1973, 1975, 1978) developed the prototype theory in the 1970s, which has become the most influential theory in Cognitive Linguistics regarding conceptual categorization (in fact, this theory was implemented in Temmerman's STT, cf. §1.3.3.). The underlying fundamental principles are: (1) a category includes members of different status, and (2) the necessary features shared by all the members of a category are not enough to define that category. The prototype theory endeavors to explore differences and similarities among the members of a category as well as to create *prototypes*, which are defined as the most representative members of a category (Cuenca & Hilferty, 1999, p. 35). Therefore, the establishment of new categories involves the creation of prototypes, which are used for the classification of the remaining members according to their goodness or typicality, that is, how good an example of a category is (Ungerer & Schmid, 1996, pp. 16-17). In this respect, Rosch (1975) states:

> This study has to do with what we have in mind when we use words which refer to categories. Let's take the word *red* as an example. Close your eyes and imagine a true red. Now imagine an orangish red ... imagine a purple red. Although you might still name the orange red or the purple red with the term *red*, they are not as good examples of red (as clear cases of what *red* refers to) as the clear "true" red. In short, some reds are redder than others. The same is true for other kinds of categories. Think of dogs. You all have some notion of what a "real dog," a "doggy dog" is. To me a retriever or a German shepherd is a very doggy dog while a Pekinese is a less doggy dog. Notice that this kind of judgment has nothing to do with how well you like the thing [...]. (p. 198)

However, the prototype theory by Rosch is said to present some limitations. First, she does not consider the contextual information and is based on a list of properties, moving away from the communicative situation and ignoring extralinguistic data, which may be relevant for the organization of concepts into categories. As a reaction, other theories emerged to overcome these limitations and considered the contexts where the concepts

appear, such as the theory of exemplar models (Smith & Medin, 1981) and the theory of conceptual coherence (Murphy & Medin, 1985), but they have passed almost unnoticed.

Apart from the theories centered on categorization, Cognitive Linguistics has witnessed the emergence of other theories aiming at the explanation of knowledge organization and world comprehension through the relationships between categories and concepts. Some examples that are worth mentioning are the Idealized Cognitive Model theory by Lakoff (1987), the Schema theory by Schank and Abelson (1977) and the Frame Semantics theory (cf. §1.5.1.1.) by Fillmore (1977, 1982, 1985). These theories share their interest in the context where a concept appears (like the modern theories of Terminology, cf. §1.3.) and the dynamics of conceptual structures established between categories.

In general, we can say that theories stemming from Cognitive Linguistics have exerted a significant influence on the recent approaches to terminology, particularly regarding the conceptualization of reality in a given communicative setting, both in terms of conceptual categorization (e.g., the FBT, cf. §1.3.4.) and world knowledge organization. Concrete evidence is the fact that methods of knowledge representation based on cognitive theories have been incorporated, such as ontologies (e.g., in the STT, cf. §1.3.3.), which help to establish a link between mental knowledge representation and language. More specifically, the theory of Frame Semantics by Fillmore has been of the first importance in order to create a more coherent version of the relationships between categories and to represent the organization of the world. Therefore, this is the focus of the following subsection.

1.5.1.1. Frame Semantics

The central tenet of Frame Semantics (Fillmore, 1976, 1977, 1982, 1985) is that words and phrases do not relate to each other, but they belong to frames which provide speakers with the conceptual base to determine the concept that the word or phrase encodes. To put it another way, the conceptual structures underlying the meanings of linguistic entities are produced by semantic frames which contain elements of the kind of situation described, such as the participants in the situation or the processes that take place. These *frame elements* (FEs) can be classified as core and non-core elements. The former refer to elements that are essential to the meaning of a frame, such as a speaker in frames

connected with communication; the latter are elements which a frame can dispense with, for example, expressions of time, place and manner. Atkins and Rundell (2008) define *frame* as follows:

> A semantic frame is a schematic representation of a situation type (e.g., speaking, eating, judging, moving, comparing, etc. – activities and situations which make up our everyday life) together with a list of the typical participants, props, and concepts that are to be found in such a situation; these are the semantic roles, or 'frame elements' (FEs). (p. 145)

The idea behind this theory is that "people have in memory an inventory of schemata for structuring, classifying, and interpreting experiences, and that they have various ways of accessing these schemata and various procedures for performing operations on them" (Fillmore, 1976, p. 25). Particular words are associated in memory with particular frames, therefore, exposure to a specific linguistic form in a given context activates or "evokes" (Fillmore, 2006, p. 378) a frame in the perceiver's mind. Words that evoke a frame are called *lexical units*,[23] understood as pairings of a word with a meaning or sense. This activation allows access to the rest of linguistic material which is associated with the same frame. Fillmore (2006) mentions this activation in the following definition of frame:

> By the term 'frame' I have in mind any system of concepts related in such a way that to understand any one of them you have to understand the whole structure in which it fits; when one of the things in such a structure is introduced into a text, or into a conversation, all of the others are automatically made available. (p. 373)

An illuminating example of a cognitive frame is the COMMERCIAL_EVENT. Fillmore's (2006) intention when outlining this frame was "to show that a large and important set of English verbs could be seen as semantically related to each other by virtue of the different ways in which they 'indexed' or 'evoked' the same general 'scene'" (p. 378). Therefore, the author (1976) describes it in the following terms:

[23] LUs.

> The frame for such an event has the form of a scenario containing roles that we can identify as the buyer, the seller, the goods, and the money; containing subevents within which the buyer surrenders the money and takes the goods and the seller surrenders the goods and takes the money; and having certain institutional understandings associated with the ownership changes that take place between the beginning and the end of each such event. Any one of the many words in our language that relate to this frame is capable of accessing the entire frame. Thus, the whole commercial event scenario is available or "activated" in the mind of anybody who comes across and understands any of the words "buy," "sell," "pay," "cost," "spend," "charge," etc., even though each of these highlights or foregrounds only one small section of the frame. (p. 25)

The practical application of Frame Semantics is FrameNet (FN),[24] an ambitious project running at the International Computer Science Institute in Berkeley, California (United States), since 1997. It aims at the creation of an online lexical resource for English based on this theory and supported by corpus evidence (mainly, the BNC) (Ruppenhofer et al., 2016, p. 7), and the data is freely available for download. FN has been extended to a number of languages and some other projects are also currently going on, for example, Chinese FrameNet,[25] FrameNet Brasil,[26] German FrameNet,[27] Spanish FrameNet[28] and Swedish FrameNet.[29]

It is important to realize that FN is a huge database which provides valuable information for any linguist. This is certainly true in the case of Natural Language Processing (NLP), since the more than 200,000 manually annotated sentences linked to more than 1,200 semantic frames provide a unique training dataset for semantic role labeling, used in applications such as information extraction, machine translation, event recognition, sentiment analysis, etc. However, according to Buendía Castro (2013, p. 260), it also has some limitations, for instance: (1) there is an infinite number of FEs and

[24] https://framenet.icsi.berkeley.edu/fndrupal/framenet_search (Last accessed: 25/10/2020).
[25] http://nlg.csie.ntu.edu.tw/nlpresource/FrameNet/CFN-Lex/ (Last accessed: 25/10/2020).
[26] https://www.ufjf.br/framenetbr/ (Last accessed: 25/10/2020).
[27] http://www.laits.utexas.edu/gframenet/ (Last accessed: 25/10/2020).
[28] http://spanishfn.org/ (Last accessed: 25/10/2020).
[29] https://spraakbanken.gu.se/eng/swefn (Last accessed: 25/10/2020).

semantic roles; (2) some frames provide all the senses for a predicate, such as *believe*, but others include more peripheral ones, such as *antagonize*; (3) there is no explanation of how lexical-semantic information can condition the syntactic realization of a predicate; and (4) there is no proposal of a metalanguage, which would enhance its consistency.

One thing that must not be overlooked is that the activation of a frame and its elements through LUs is possible because human beings have experience of the world. As it is stated by Fillmore and Atkins (1992), "An implicit awareness of this particular organization of our physical and social world provides the conceptual background for a fairly large body of lexical material" (p. 77). Additionally, this specific schematization of experience is responsible for holding words together, thus, when a word is encountered, a frame is evoked and many other words that relate to the same frame are accessed (in fact, this is what happens with collocations, where the access to the most prominent word of the pair activates the access to its accompanying elements). Regarding experience, it depends on several factors (e.g., the culture group to which a person belongs), which lead to stereotypical situations and stereotyped knowledge that are necessary in order to understand concepts and meanings. According to Minsky (1975), "A frame is a data-structure for representing a stereotyped situation" (p. 211). Nevertheless, a semantic frame evoked by the same situation may be understood differently depending on a person's experience. A representative example is the LU *bride*, which will activate different elements and situations depending on whether the person accessing the frame, called WEDDING, is Catholic, Muslim or Buddhist.

1.5.2. Corpus Linguistics

Corpus linguistics is thus a methodology,
comprising a large number of related methods
which can be used by scholars of many different
theoretical leanings.
(Lindquist, 2009, p. 1)

Corpus Linguistics is the second field which, to our mind, can substantially improve the products of specialized lexicography, as it involves the use of real language examples to study a language with the help of software tools. The *Conference on the Use of Computer*

Corpora in English Language Research was held in Nijmegen (The Netherlands) in 1983 and one year later a collection of papers was published: *Corpus Linguistics: Recent Developments in the Use of Computer Corpora in English Language Research* (Aarts & Meijs, 1984). It was the first use of *corpus linguistics* and the person who coined it, Jan Aarts, said: "I've never felt the need to argue that corpus linguistics is a discipline in its own right, as some people do; for me it's always been a means to an end" (as cited in Johansson, 2009, p. 34). Since then, the debate concerning whether corpus linguistics is a theory or a methodology has been present, although the latter has more followers. For instance, three of the most prolific authors in this area, Mark Davies, Stephan Th. Gries and Mike Scott, agree that it is a methodological paradigm rather than a branch of linguistics (Viana, Zyngier & Barnbrook, 2011, pp. 65, 83, 214).

In brief, corpus linguistics facilitates the study of authentic language from different perspectives by providing tools for searching, retrieving, annotating and analyzing it. Many have been the fields using corpus linguistics as a methodology. To name but a few, it has become an effective way to access "data suitable for quantitative analysis of social and linguistic variables" in sociolinguistics (Romaine, 2009, p. 97). Also, corpus linguistics has become a powerful tool in language teaching[30] and has proved itself to be a reliable method for the study of figurative language (Stefanowitsch & Gries, 2006), cognitive semantics (Glynn & Fischer, 2010), phraseology (Evert, 2009; Pęzik, 2018) and translation studies (Corpas Pastor, 2008; Mitkov, Monti, Corpas Pastor, & Seretan, 2018). Finally, the field of study which is greatly indebted to corpus linguistics is lexicography, both general and specialized, as corpora are the source from which lexicographers and terminographers get the material to create dictionaries (cf. Heid, 2009).

The starting point in Corpus Linguistics is the *corpus* (this concept will be fully explored in §4.3.1.), which is defined as "a collection of texts assumed to be representative of a given language put together so that it can be used for linguistic analysis" (Tognini-Bonelli, 2001, p. 2). That is to say, a corpus contains real language in context and allows researchers to examine how this language is used, how words combine with each other, how reality is construed in speech, the differences between written and spoken language, and so forth. The presence or absence of context is what makes a

[30] See Römer (2009) for an accurate description of pedagogical applications of corpora.

language to be studied *in vivo*, that is, within a real communicative context, as opposed to *in vitro*, that is, out of context, such as in lists of words, dictionaries, glossaries, and so on. Nowadays, corpus refers to electronic corpora, which can be explored using computers. However, it has not always been the case. Although pre-electronic corpora date back to the thirteenth century, when a Latin concordance of the Bible was written,[31] the founding fathers of this paradigm are two (Leech, 2011, p. 156): the British Randolph Quirk (1920-2017) and the American W. Nelson Francis (1910-2002).

On the one hand, in the European continent Quirk founded the *Survey of English Usage* (SEU) in 1959, which was the first center in Europe to conduct research on corpus linguistics. Influenced by the Firthian notion that context determines language use, he aimed to compile spoken and written English samples from a range of different types of genres (Meyer, 2009, p. 11). The resulting corpus, known as *SEU Corpus*, had approximately one million words divided into 5,000-word samples. This project was followed by the *Survey of Spoken English* (SSE), started by Jan Svartvik at Lund University (Sweden) in 1975. He endeavored to computerize the SEU corpus material which originally consisted of 87 texts of spoken British English. However, 13 more spoken texts were added afterwards, making a corpus of 100 spoken texts and 500,000 words, which was released in 1990 (Svartvik, 1990) under the name of *The London-Lund Corpus of Spoken English* (LLC).

On the other hand, in the American continent the compilation of the *Brown Corpus* arose in response to the linguistic fashions of the early 1960s led by Noam Chomsky.[32] Francis and Henry Kučera aimed to provide linguists with actual written language despite not finding support among their colleagues from the Brown University, as they "had a predictable fear of the new 'calculating machines' and […] those who dared to commit the treason of joining the scientists' camp of […] binary numbers" (Kučera, 1992, pp. 402-403). However, in spite of the setbacks that the project experienced, it was carried out and the Brown Corpus was available in 1961. It contained more than one million words and was divided into 500 samples of more than 2,000 words each from different prose genres (verse and drama were excluded) published in the United States. The main

[31] For a detailed description of pre-electronic corpora, see Meyer (2009).
[32] See Karlsson (2009) for a comprehensive description.

sources for extracting the texts were the Brown University Library, the Providence Athenaeum and the New York Public Library.

Nevertheless, in the field of corpus linguistics there is one project that deserves special mention. Investigations on word combinations led John Sinclair (cf. §2.2.) to a revolution in the field of lexicography in the 1980s. The COBUILD project, whose name stands for *Collins Birmingham University International Language Database*, involved the creation of a new dictionary based on an electronic corpus of contemporary texts. This corpus was called the *Collins Corpus* and contained both written and spoken material, becoming the largest corpus of English language in the world (Lavid, 2007, p. 9). In fact, new texts were added and in 1991 it was launched as the *Bank of English*. The result was the *COBUILD Dictionary*, which was regarded as innovative in many respects, since it provided information on how words were used, which words were used together (e.g., collocations) and how often they did. The COBUILD project was one of the projects that prepared the ground so that corpus linguistics could become mainstream in the 1990s (Johansson, 2009, p. 49).

Today, the number and type of corpora available in several languages is huge and widely used by experts from various disciplines. The number of studies based on corpora has increased due to several factors (Durán-Muñoz, 2012, p. 60): (1) researchers have started to rely on authentic language included in corpora rather than their intuition to gain an insight into lexical, grammatical and semantic knowledge, the relationship between words and their context, and so forth; (2) researchers have become interested in providing a more descriptive account of real language use; and (3) researchers have witnessed a great advance in technology, so now they are able to collect and handle a full set of data by using computers, which makes the analysis of information easier. As a result, both generic and specialized corpora have become indispensable to linguistic studies. Added to that, their use offers other advantages, such as objectivity in confirming or refuting hypotheses, rapidness of processing huge amounts of data in short periods of time, access to information which can be representative of a language as a whole, among others (Villayandre Llamazares, 2008, pp. 335-336). But the major advantages of corpora over manual investigations are speed and reliability, since linguists "can investigate more material and get more exact calculations of frequencies" by using a corpus (Lindquist, 2009, p. 5).

However, the use of corpora also has some disadvantages. For example, in the field of pragmatics, large corpora may be impoverished both textually and contextually (Rühlemann, 2010, as cited in Clancy & O'Keeffe, 2015, p. 243), since corpus builders exclude important contextual features of spoken language, such as prosody. Similarly, in the field of dialectology and sociolinguistics, it is difficult to obtain high-quality recordings of spoken language and to gather extensive data about the social background of informants (Grieve, 2015, p. 380). On the other hand, it is important to realize that syntactic parsers in corpora are not perfect (cf. §4.3.2.), so relevant information may be missed when extracting specific word combinations, for this reason, corpus-based grammar research has been "less impressive than advances in other areas" (Leech, 2015, p. 149). Other main drawbacks may have to do with corpus compilation, corpus size, part-of-speech (POS) tagging,[33] license price and search tools. In addition, language is in constant change, so corpora need to be frequently updated and supplemented with new data (Birkner, 2015, p. 14).

Despite the general arguments against the use of corpora, we must say that, on balance, and for the purposes of this study, their role is of the utmost importance in the field of specialized lexicography. For instance, they have provided the methods and tools to better deal with the different units and uses and have allowed the development of novel approaches to the lexicographical treatment of meaning, grammar and pragmatics (Paquot, 2015, p. 460). By the same token, specialists have a mine of information at their disposal, thus, they can use real SLs as a base for their work, whose results would be objective and systematic. Another key point is the revolutionary effect that corpora have had on phraseology. This can be illustrated briefly by substantial evidence showing that words do not occur isolated but combine with each other in preferred syntagmatic patterns, such as collocations (e.g., Sinclair, 1991; Hanks, 2012). In fact, a wide range of software tools and resources are available today to assist in the identification and analysis of TUs, word combinations, concordance lines, and so forth, for example, *TermoStat Web 3.0* (cf. §4.4.4.),[34] *Sketch Engine* (cf. §4.4.),[35] *AntConc*,[36] *WordSmith Tools* (cf. §4.4.4.),[37]

[33] *Part-of-speech tagging* is the process of labeling each word in a text corpus to indicate the part of speech and often also other grammatical categories, e.g., tense, number (singular/plural), case, etc. [https://www.sketchengine.eu/pos-tags/] (Last accessed: 25/10/2020).
[34] http://termostat.ling.umontreal.ca/ (Last accessed: 25/10/2020).
[35] http://www.sketchengine.eu (Last accessed: 25/10/2020).
[36] https://www.laurenceanthony.net/software.html (Last accessed: 25/10/2020).
[37] https://www.lexically.net/wordsmith/ (Last accessed: 25/10/2020).

WMATRIX (Rayson, 2001), *ConcGram 1.0* (Greaves, 2009), to mention but a few. To our view, the main reason for relying on corpus linguistics is that it provides a large collection of real language and quantitative methods as the basis for any analysis, consequently, research findings have more validity.

1.5.3. Computational Linguistics

Computational Linguistics is the last area we will cover in connection with specialized lexicography. It is an interdisciplinary field concerned with language and computational approaches, that is, comprehensive computer systems are implemented for understanding and generating natural language. It was born in the 1950s with the first conference on machine translation, although the phrase *computational linguistics* started to appear in the 1960s, when the *Association for Machine Translation and Computational Linguistics* was formed (Kay, 2003, p. xvii). Therefore, machine translation has been central in the development of this discipline as well as information retrieval and man-machine interfaces (Grishman, 1986, pp. 4-5).

Regarding information retrieval, it must be remembered that much of the information we use appears in natural language form, for example, in books, journals and reports, for this reason, an application able to retrieve information from these texts is necessary. In the 1970s success was limited in this area because the texts were quite complex. However, in the early 2000s automatic information retrieval systems started to be studied (El Emary & Atwan, 2005; Nuray & Can, 2006; Vicente-López et al., 2015). Man-machine interfaces refer to the use of natural language as a means of communication with interactive systems. This application was more advantageous and successful than the previous ones because the input to such systems used to be simpler than the texts involved in machine translation and information retrieval. In fact, most computational linguistic work has included interactive interfaces since the early 1970s. One of the leading figures in this area is Martin Kay (2005), who defines this discipline as follows:

> Computational linguistics is not a natural language processing. Computational linguistics is trying to do what linguistics do in a computational manner, not trying to process text, by whatever methods, for practical purposes. Natural language processing, on the other

> hand, is motivated by engineering concerns. I suspect that nobody would care about building probabilistic models of language unless it was thought that they would serve some practical end. (p. 429)

Computational Linguistics consists of two differentiated frameworks. On the one hand, the theoretical perspective, which encompasses the theoretical basis and methodologies of the discipline. On the other hand, the applied perspective, also known as Linguistic Engineering, which deals with technology and is focused on the design and development of computer systems that can understand, produce and translate spoken and written language into any natural language (Lavid, 2005, p. 76). According to Corpas Pastor (2008, p. 10) and Orăsan, Ha, Evans, Hasler, & Mitkov (2007), corpora have played a key role in both perspectives and have exerted a great impact on the creation of systems aimed at morphosyntactic analysis, computational lexicons, computerized dictionaries, and so forth. It was in the 1990s when this impact was more powerful for two reasons (which are very much connected to the development of Corpus Linguistics): (1) technological advances allowed the use of vast corpora, and (2) researchers realized that systems based on their knowledge showed some limitations, so using methods based on empirical evidence was more reliable. McEnery (2003) highlights the "happy symbiosis" between Computational and Corpus Linguistics:

> Corpora have placed a useful role in the development of human language technology to date. In return, corpus linguistics has gained access to ever more sophisticated language processing systems. There is no reason to believe that this happy symbiosis will not continue ─to the benefit of language engineers and corpus linguistics alike─ in the future. (p. 460)

Moreover, Corpus Linguistics has made use of the applications and resources developed by Computational Linguistics, given that corpora need to be explored with a computer so that information can be retrieved. In Leech's (1991) words: "It is widely acknowledged today that a corpus needs the support of a sophisticated computational environment, providing software tools both to retrieve data from the corpus and to process linguistically

the corpus itself" (p. 22). As there is a broad range of software tools available, it is the linguist's responsibility to choose which tools better suit his/her requirements. To this regard, terminographers are in charge of selecting the most appropriate software tools that will help them design their terminological resources.

Additionally, Computational Linguistics aims to study natural language by using computer tools, as it is stated by Bolshakov and Gelbukh (2004): "the main task of computational linguistics is just the construction of computer programs to process words and texts in natural language" (p. 25). This discipline allows the analysis of large volumes of data contained in corpora to identify relevant pieces of information (e.g., Chien & Chen, 2001; Daille, 2001; Maynard & Ananiadou, 2001), which can be done by using social, cultural and geographic contexts. Thus, all terms and expressions can be detected, that is, not only can the most common terms to explain a concept be retrieved by using computational methods, but hidden and unusual terms, such as expressions in slang or terms with a higher level of specialization, can also be detected. Consequently, terms used by a particular group of specialists can be pinpointed, which guarantees accurate translations.

However, term extraction is only one of the many tasks that Computational Linguistics can do. Other pieces of information about terms can be extracted. For example, an increasing amount of research is devoted to extracting semantic information on terms from specialized texts. It is generally accepted that texts provide clues to the meaning of TUs, thus, if they are automatically or semi-automatically uncovered, users will gain a better insight into the meaning of terms. Added to that, it can lead to the construction of conceptual hierarchies or semantic networks, which display different kinds of relation between terms, such as hypernymic/hyponymic, meronymic/holonymic or synonymic/quasi-synonymic relationships (Condamines & Rebeyrolles, 2001).

Computational Linguistics also allows syntactic parsing, which tries to find structural relationships between units in a sentence. As various kinds of ambiguity can become apparent when doing so, Hobbs and Bear (1994) suggest two general principles of parse preference. The first one is called *The Most Restrictive Context Principle* and can be explained as follows: "Where a constituent can be placed in two different structures, favor the structure that places greater constraints on allowable constituents" (p. 505). The second one is called *The Attach Low and Parallel Principle* and can be

defined in the following terms: “Attach constituents as low as possible, and in parallel with other constituents if possible” (p. 508). Nevertheless, these authors recognize that both principles can be overridden by more than just semantics and pragmatics, for example, by commas in written discourse and pauses in spoken discourse.

Another task that this discipline can performed is lemmatization, which consists in grouping all the inflected forms of a unit together so that they can be analyzed as a single item, for example, *work, works, working, worked*. This is one of the most important stages of text preprocessing and has also been an active area of research, especially for morphologically rich languages (Chrupała, 2014; De Pauw & De Schryver, 2008; Sorokin & Shavrina, 2016). In general terms, two major approaches can be distinguished (Dereza, 2019, p. 114): (1) rule-based approach, which requires manual intervention but produces good results for being language-specific, such as with Swedish (Cinková & Pomikálek, 2006) and Slovene (Plisson, Lavrac, & Mladenic, 2004), and (2) statistical approach, which is computationally expensive and requires a large annotated corpus; besides, it does not produce impressive results. Apart from that, neural networks may also be used (Kestemont, De Pauw, Van Nie, & Daelemans, 2017).

Finally, Computational Linguistics allows the extraction of concordances, collocations and frequencies from corpora. First, a concordance is a screen display of a chosen term or phrase in its different contexts in such a way that the words coming before and after it to the left and right are presented (cf. §4.4.3.). Second, collocations show words that habitually co-occur together in a language (cf. §4.4.2.). Third, frequency lists display the words that are most often used in a language (cf. §4.4.1.). These lists can be organized in different ways, for example: (1) in terms of the frequency of all types of words; (2) in terms of the frequency of every type of word, such as nouns, verbs, prepositions, and so forth; and (3) in terms of the frequency of words in a given genre, for example, spoken, fiction, academic. Needless to say, all the applications of Computational Linguistics can be used either for analyzing the GL or an SL.

Summing up, it must be emphasized that there is a strong connection between the three linguistic theories explained above, that is, Cognitive Linguistics, Corpus Linguistics and Computational Linguistics, and the field of specialized lexicography. First, the principles of Cognitive Linguistics are applied to analyze and understand how our mind organizes and categorizes concepts in an SL as well as their semantic

representation so as to facilitate terminology definitions. Second, Corpus Linguistics allows the access to terminology in its real communicative context, since corpora include authentic language. Third, Computational Linguistics is the link between the two previous disciplines, given that it provides the necessary tools to perform the tasks each field is faced with, such as terminology candidate extraction from corpora (Corpus Linguistics) and their organization in the mental lexicon (Cognitive Linguistics). In general terms, it can be stated that Corpus Linguistics and Computational Linguistics are present in every terminological work, although Cognitive Linguistics is not.

On the other hand, Cognitive Linguistics, Corpus Linguistics and Computational Linguistics have strengthened their relationship with phraseology in the last decades. For instance, cognitive linguistic theories, such as the Metaphor theory (Lakoff & Johnson, 1980) and the Conceptual Integration theory (Fauconnier, 1994; Fauconnier & Turner, 1998, 2002, 2003), have proved themselves to be effective in the analysis of phraseological material, since they provide insights into how it is created, organized and creatively adapted (Langlotz, 2001; Omazić, 2005). Regarding Corpus Linguistics, it has been heavily relied on in phraseological studies, given that it provides researchers with a method of identification, extraction and classification of word combinations contained in corpora (e.g., Anderson, 2006; Corpas Pastor, 2016; Gray & Biber, 2015; Hunston & Francis, 2000). Finally, Computational Linguistics, whose influence on phraseological studies has been more recently exerted, has allowed phraseologists to address MWUs from a computational perspective so as to identify MWUs automatically, extract and translate them and develop NLP applications for their study (Corpas Pastor & Colson, 2020; Corpas Pastor & Mitkov, 2019; Mitkov, 2017). The field of phraseology will be thoroughly explored in the following chapter.

CHAPTER 2. PHRASEOLOGY AND SPECIALIZED COLLOCATIONS

[…] the knowledge of a language, whether it is the mother tongue or a foreign language, is not enough: it is also necessary to acquire a command of the particular and peculiar phraseology unique to a subject field.

(Patiño García, 2014, p. 125)

The second chapter of this dissertation introduces our object of study, that is, collocations, and is divided into three sections. First, Section 2.1. provides a general outline of Phraseology as a discipline. It briefly mentions the first works on groupings of words and then focus on the most representative authors in this field, who were interested in categorizing word combinations and were deeply inspiring. This description ranges from the 1940s to the present day (§2.1.1.–§2.1.4.), for it concludes with a glance at some of the most relevant phraseological taxonomies of the current century. Next, Section 2.2. is motivated by a desire to define collocations within a framework of phraseological units. After providing a brief sketch of some approaches to the study of these word combinations, it looks at collocations in greater detail, since their features will be adequately addressed (§2.2.1.). Also, we will explore some typologies of collocations (§2.2.2.) and will compare collocations to other multiword units, such as idioms, lexical bundles or collostructions (§2.2.3.). All the aspects covered in this section will help to provide a definition of the term (§2.2.4.). Finally, Section 2.3. will deal with specialized collocations and their role in the specialized language of tourism, for these are the main concern in this dissertation. Therefore, we will examine their own characteristics as well as those shared with general language collocations and will close this section with an analysis of investigations on this topic.

2.1. Phraseology: An introduction

In the venerable field of phraseology, people have always been aware that language is full of units of meaning larger than the single word.

(Teubert, 2004, p. 84)

Phraseology is a term derived from Greek *phrasis* and *logia*. The former refers to a group of words with some unity and the latter refers to a doctrine, theory or science. Accordingly, Phraseology can be described as the linguistic discipline which studies groups of words. Interest in this field has grown exponentially in the last decades, probably because 80% of the combinations of words are made through co-selections (Sinclair, 2000, p. 197). In addition, this is a dynamic discipline in terms of the different reasons that motivate its study, for example, types of words that combine with each other, psychological factors behind the way we recall them from memory, steps for their teaching and learning in second language acquisition, their semantic status in SLs, and so forth.

It can be stated that the idea of classifying groups of words was first advanced by the Swiss linguist Charles Bally, disciple of Ferdinand de Saussure. In his *Traité de Stylistique Française,* whose first volume was published in 1909 (the second volume was published in 1919), [38] he mentioned combinations of words that we use on a regular basis and that become inseparable units: "Ces groupements peuvent être passagers, mais, à force d'être répétés, ils arrivent à recevoir un caractère *usuel* et à former même des *unités* indissolubles"[39] (1951, p. 66). In addition, he identified three types of syntagmatic combinations in terms of how fixed they are: (1) *les associations libres* −or *occasionnelles*− [free −or occasional− combinations] (less fixed), (2) *les series phraseologiques* or *groupements usuels* [phraseological series or usual groupings], and (3) *les unités indissolubles* [indivisible units] (more fixed).

On the other hand, Harold E. Palmer, a British linguist, used the term *collocation* to refer to words that co-occur. He was the author of the "Second Interim Report on English Collocations" (1933), submitted to the *Tenth Annual Conference of English*

[38] The second edition of this piece of work was published in 1951.
[39] "These groups can be temporary, but by dint of being repeated they manage to become usual and even to form indissoluble units" (translation mine).

Teachers in Tokyo, and realized that word combinations should be regarded as a unit. As Palmer was involved in language teaching, he developed a more pedagogical approach and dealt with how these units could be better learned, suggesting that "A collocation is a succession of two or more words that must be learnt as an integral whole and not pieced together from its component parts" (p. 1). His perspective became particularly influential in works on the teaching of the lexicon in foreign languages (e.g., Lewis, 1993, 2000; Willis, 1990).

However, it was not until the late 1940s that Phraseology was firmly established as a discipline. A group of Russian scholars became very active in this field and aimed at categorizing all set expressions as well as describing their varying degrees of fixedness. It was led by Victor V. Vinogradov (disciple of Bally), considered the father of the Russian phraseology (Cowie, 1998b, p. 213), whose lines were closely followed by Natalia N. Amosova (cf. §2.1.1.). In the 1960s and 1970s, non-Russian speaking scholars motivated by Vinogradov's work, such as Klappenbach (1968), Weinreich (1969) and Lipka (1974), also dealt with the classification of word combinations. Finally, some authors working in the 1980s and 1990s deserve special attention, such as Anthony P. Cowie and Peter A. Howarth (cf. §2.1.2.), on the one hand, and Igor A. Mel'čuk (cf. §2.1.3.), on the other, given that their purposes were more practical than theoretical. The following sections will cover some of the ideas promoted by these authors.

2.1.1. Vinogradov and Amosova, the Russian scholars

Without a doubt, the Russian phraseology has played a leading role in both theoretical works and their applications, such as the design of dictionaries and other lexicographical resources. As it has been previously stated, their major objective was to group MWUs into categories according to two aspects: their degree of fixedness and their idiomaticity. Vinogradov (1947) was the first author in suggesting a tripartite scheme of subcategories within the most inclusive category, the *phraseological unit* (PU). This category is used to refer to a group of words or a sentence whose meaning is fully or partially transformed and with stability at the phraseological level. The underlying principle of the semantic classification of PUs is the degree of motivation (idiomaticity), that is, the relation that the meaning of the whole expression bears to the meaning of its components. For this

reason, the three subcategories he mentions are located along a continuum from less motivated to more motivated combinations:

1) *Phraseological fusions* are found at one end of the continuum, since they are groupings of unmotivated units, that is, no relationship exists between the meaning of the whole sequence and the meaning of each of its elements. They are characterized by being completely stable at the semantic and grammatical levels, so they are, in general, structurally fixed, for example, *once in a blue moon, to be on the carpet*;

2) *Phraseological unities* are located in the middle of this continuum because they are partially non-motivated, that is, the whole combination can have a literal reading but can also be interpreted figuratively. Besides, they are semantically indivisible, such as *to lose one's head, a fish out of water*; and

3) *Phraseological combinations* are placed at the other end of the continuum since they are clearly motivated. They consist of an element which preserves its direct meaning and an element which is used figuratively, then, the variability of their components is limited. For example, in the verb + noun combination *to meet the demand*, the verb *to meet* has a non-literal meaning which would remain the same even if the noun changed into *the necessity* or *the requirements.* Other examples are *to make a mistake* and *to pay a compliment.*

Amosova (1963) accepted the ideas developed by her predecessor. However, unlike Vinogradov, she distinguished two groups within the so-called phraseological combinations: *phrasemes* and *phraseoloids*. The former includes more restricted combinations where the elements cannot be substituted; the latter is a less restricted category. For instance, a phraseme is *grind one's teeth*, since it has a figurative meaning determined by its context and the context accepts a single item, that is, no other noun can be used instead of *teeth* if we want to keep the meaning of the verb. On the other hand, the adjective + noun combination *small talk* is a phraseoloid because the noun might be substituted for another noun, for example, *small hours, small change*, and the figurative meaning of the adjective would remain constant. With regards to the other word combinations proposed by Vinogradov, that is, phraseological unities and phraseological fusions, Amosova did not really explain the difference between them. Instead, she used

the term *idioms* to refer to semantically and grammatically inseparable units, characterized by the lack of relation between the meaning of the whole and the meaning of the components taken in isolation.

Other authors that were motivated by the Russian linguists in the 1960s and 1970s are the following. First, Klappenbach (1968) contemplated five different MWUs: (1) verbal groups, (2) attributive phrases and technical terminology, (3) adverbial groups, (4) pairs of words, and (5) syntactic patterns. Second, Weinreich (1969) made a distinction between *idioms* and *stable collocations* in terms of figurativeness. Thus, both are co-occurrences of words, but the former involve a figurative or metaphorical usage, such as *pull someone's leg*, whereas the latter do not, for instance, *part and parcel*. Third, Lipka (1974) used the term *idiom* to refer to complex linguistic units whose meaning is not deductible from that of their component parts and they can be of three types: (1) compounds, such as *holiday*; (2) fixed collocations, for example, *black market*; and (3) complex expressions, for instance, *kick the bucket*. Other authors inspired by Vinogradov and Amosova came later and they are explained in the following sections.

2.1.2. Cowie and Howarth and applied phraseology

The second pair of authors who need to receive a special mention in this dissertation are Cowie and his disciple Howarth, given the practical intentions of their work on phraseology (such as ours). On the one hand, research conducted by Cowie (1981, 1998b) aimed at the compilation of dictionaries, since his approach was lexicographical. In addition, he tried to unveil the influence that British and Russian phraseologists had exerted on dictionaries of collocations and idioms targeted at students of English as a foreign language. On the other hand, Howarth (1996, 1998) was concerned with the analysis of phraseology in academic written English of foreign students, so it meant an approach to the study of SLs, more specifically, the SL of academic writing. He aimed to investigate how native speakers used the language and how foreign learners deviated from such norms.

Both authors were heavily influenced by the Russian phraseologists. As an illustration, Cowie (1998b) recognized that Vinogradov's concept of phraseological fusion or idiom represented "the standard approach to idiomaticity at both the popular and technical levels in Britain and the USA" (p. 214). Thus, its definition emphasizing

the lack of relation between the meaning of the whole combination and its component parts was key in his work. Likewise, Howarth (1996) acknowledged that he had borrowed "the systematic categorization of all set expressions" (p. 25) from the Russian lexicology. However, one difference with their predecessors was that, while the Russian scholars were more concerned with the semantic aspect of PUs, Cowie and Howarth were more interested in combination restrictions and strived to separate restricted from open expressions to discover what was not free.

Regarding their categorization of MWUs, both agreed that these could be arranged according to two groups. Firstly, *formulae* are combinations which recur unchanged in normal spoken and written communication. They are described as structures being quite stable in form and having a primarily pragmatic rather than syntactic function. Two sub-categories are distinguished (Cowie, 2001): (1) *routine formulae*, which perform speech-act functions, for example, *How are you?, Good morning*, and (2) *speech formulae*, which are used to organize messages and show speakers' attitudes, for instance, *Would you like to come?, Are you following me?* Nevertheless, it was not the first time that formulaic language of this type was mentioned, since Cowie (1988, pp. 133-134) explained that Keller (1979) used the term *gambit* and Coulmas (1981) used the phrase *routine formula* to refer to the same concept.

Secondly, the term *composites* derives from Mitchell's (1971, p. 57) *composite elements* and refers to word combinations that are not stable and can undergo semantic changes, such as idioms, collocations and compounds. This type embraces four kinds of word combinations according to two criteria (Nesselhauf, 2005, p. 14): (1) transparency, that is, the literalness of the elements of the combination and the combination itself, and (2) commutability, that is, the degree of substitution of the elements involved in the combination. Table 6 illustrates the classification of composites in terms of these two criteria (Cowie, 1981, pp. 226-230):

Table 6

Classification of Composites (Cowie, 1981)

	Transparency criterion	**Commutability criterion**	**Examples**
Open collocations	All its elements have a literal meaning	Possible substitution for other elements semantically similar	*Drink tea*
Restricted collocations	Transparent combinations; one of their elements is used literally while the other has a figurative meaning	Possible substitution with arbitrary restrictions	*Capture somebody's imagination*
Figurative idioms	Combinations with both a literal interpretation and a figurative reading	Seldom possible substitution	*Catch fire*
Pure idioms	Combinations having only a figurative interpretation	Impossible substitution	*Blow the gaff*

The interesting fact about Table 6 is the use of the general categories *collocations* and *idioms*. Although the so-called *open collocations* could also be named *free combinations*, the reason for choosing the term *collocation* in this case was to permit their inclusion in the "scale or continuum from unmotivated and formally invariable idioms to partially motivated and partially variable collocations" (Cowie, 1998a, p. 6), or in the "continuum from, at one extreme, the most freely co-occurring lexical items and transparent combinations to, at the other, the most cast-iron and opaque idiomatic expressions" (Howarth, 1996, p. 32).

The first group, *open collocations*, has also been referred to as *free constructions* (Weinreich, 1969), *free phrases* (Arnold, 1973) and *free word-combinations* (Aisenstadt, 1979). It encompasses the least restricted combinations allowing a broader range of substitution of their constituents for other words when they are semantically similar (they were not included in Vinogradov's and Amosova's taxonomies). According to Cowie (1981), the number of open collocations may be massive, thus, they would need to be included in a separate dictionary instead of in a dictionary which contains word

combinations: “the sheer bulk of such information would be overwhelming. There is clearly a place for separate dictionaries of ‘open’ collocations” (p. 226).

The second category is *restricted collocations,* also called *semi-fixed combinations,* which allow some substitution of their component elements for other words. However, there is extreme restriction collocability because one element is used with a literal meaning and the other is interpreted figuratively; the latter determines which elements can collocate with it. The context usually helps to choose the special meaning of the less transparent element. They are similar to Vinogradov and Amosova’s phraseological combinations.

Finally, there are two types of composites which are named idioms. On the one hand, *figurative idioms* do not allow much substitution given that, in addition to preserving a literal meaning, they can also be understood figuratively. They are reminiscent of Vinogradov and Amosova’s phraseological unities. On the other hand, *pure idioms* are immutable and semantically opaque, so they can only be read non-literally. They are the type of combination that Vinogradov and Amosova called phraseological fusions. These similarities reveal a considerable influence of the Russian scholars upon the work by Cowie and Howarth, which is shared by the author examined in the following section.

2.1.3. Mel’čuk and the Meaning-Text Theory

The third phraseologist who is worth mentioning at this point is Igor Aleksandrovič Mel’čuk. He was a follower of the Russian phraseology theory and his motivation was also lexicographical (cf. §3.2.2.). In Cowie’s words (1998a), the classical Russian theory “has been developed and applied with great rigour and sophistication” (p. 2) to the design and compilation of dictionaries by Mel’čuk and his associates. Mel’čuk is the author of the Meaning-Text Theory (MTT) (1981, 1988), which “assigns to the lexicon a central place, so that the rest of linguistic description is supposed to pivot around the lexicon” (Mel’čuk & Polguère, 1987, p. 261). In other words, he maintains that semantics has priority over syntax. This theory was launched in Moscow in the late 1960s and early 1970s and provides an elaborate basis for linguistic description that can be used in

different computational applications, including machine translation, phraseology and lexicography, both general and specialized.[40]

The first postulate of the MTT is that a natural language establishes correspondences between a series of possible meanings, which are the linguistic contents to be communicated, and an array of possible texts, which are phonetic representations, hence the name of the theory. Between these two there can be more than one correspondence, since a given meaning can correspond to many texts (e.g., synonymy) and a given text can correspond to many meanings (e.g., polysemy). This is contrary to univocity, one of the tenets of Wüster's Traditional Terminology. The second postulate is called the Meaning-Text Model (MTM) and refers to a symbolic model which includes a finite set of rules that define the correspondence between a semantic and a phonetic representation, that is, between a meaning and a text. The levels of representation are three: semantic (meaning), syntactic (sentence level) and morphological (word).

According to Mel'čuk (1998, p. 25), text production involves three steps. First, the speaker creates in his/her mind the conceptual representation of the situation that he/she wants to verbalize. Second, the speaker constructs the semantic representation (meaning) of the utterance that he/she aims to give. Third, the speaker constructs the phonetic representation (text) of that utterance. Thus, the utterance is considered acceptable when it depicts the correspondence of the semantic representation and the phonetic representation, which coincides with the speaker's mental portrait of reality. One key point is that the MTT tries to describe the linguistic aspects of the utterance, leaving aside all its extralinguistic aspects, for example, pragmatic, encyclopaedic (Milićević, 2006, p. 7).

With respect to a taxonomy of PUs, the author makes a basic distinction between two families which are just the opposite: *set phrases* (or *phrasemes*) and *free phrases*. The first group describes combinations of words which are non-compositional and of "'irregular' semantic and syntactic nature" (p. 24). In addition, these phrases are "selected by the speaker in a linguistically constrained way" (Mel'čuk, 2012, p. 33). The second group, which is identical to Cowie and Howarth's open collocations, defines combinations of words in which their elements are freely chosen and combined and

[40] A couple of specialized lexicographical resources that apply the MTT from a terminological point of view are the DiCoInfo and the DiCoEnviro dictionaries (cf. §3.3.).

adhere to the semantic and syntactic rules of a language. In the author's (2012) opinion, a straightforward corollary of the concept of free phrase is the following:

> Each lexical component of a free phrase can be replaced by any of its (quasi-)synonyms without affecting its meaning and grammaticality. In the phrase *select the word freely*, you can replace any component with its synonym and the meaning is preserved: *choose the lexeme without constraint*. (p. 33)

We will explain in more detail the first group, that is, set phrases or phrasemes, given that they are the word combinations that pose a problem for both theoretical linguistics and general and specialized lexicography. Mel'čuk (1998, pp. 28-29) mentions, on the one hand, ready-made expressions which are restricted by the situational context. They are called *pragmatic phrasemes* or *pragmatemes* and include greetings, proverbs, typical phrases used in letters or conversational formulae. As an illustration, a *No talking please* sign in a library is a pragmateme; in other words, "A pragmatically constrained cliché is a pragmateme" (Mel'čuk, 2012, p. 41). To our mind, this category is akin to Cowie and Howarth's formulae. On the other hand, the linguist presents the concept of *semantic phraseme*, which, in our view, is reminiscent of Cowie and Howarth's composite. This term designates MWUs which are restricted along the syntagmatic axis due to the semantic transparency or opacity of their elements, that is, the degree to which the meaning of the whole combination includes the meaning of its constituent elements. In this sense, three sorts of word combinations emerge (pp. 37-38):

1) *Full phrasemes* or *full idioms*: Word combinations where the meaning of the whole does not include the meaning of any of its lexical components. The general combination rule is A + B = C. That is, a combination consisting of A + B conveys a meaning, C, which includes neither A nor B. For example, *to spill* (A) *the beans* (B) means "to reveal secret information unintentionally or indiscreetly"[41] (C); *to go* (A) *ballistic* (B) means "to suddenly become very angry"[42] (C). In Vinogradov and Amosova's terms, these word combinations are phraseological fusions, and Cowie and Howarth call them pure idioms;

[41] https://en.oxforddictionaries.com/definition/spill_the_beans (Last accessed: 25/10/2020).
[42] https://www.ldoceonline.com/es-LA/dictionary/go-ballistic (Last accessed: 25/10/2020).

2) *Semi-phrasemes*, *semi-idioms* or *collocations*: Whereas an element of the word grouping, A, is freely chosen by the speaker and preserves its own meaning, the second element, B, is semantically limited and is contingent on A; besides, it includes an additional meaning, C. The combination formula is A + B = A + C. For instance, *strong* (B) *coffee* (A) refers to coffee (A) which has a very noticeable taste or smell[43] (C); *high* (B) *wind* (A) means fast or strong (C) wind (A).[44] Vinogradov and Amosova name this group phraseological combinations, and Cowie and Howarth restricted collocations; and

3) *Quasi-phrasemes* or *quasi-idioms*: The meaning of A + B includes the meaning of both constituent elements, A and B, neither as the semantic pivot, and it also includes an additional meaning, C, as its semantic pivot. The combination rule is A + B = A + B + C. For example, *to give the breast* means to feed a baby with the milk produced by a woman's breasts; *to start a family* means to have your first child. This type is somehow similar to Vinogradov and Amosova's phraseological unities and Cowie and Howarth's figurative idioms, although it must be emphasized that Mel'čuk does not contemplate a literal reading of the combination.

As we can see, all the set phrases suggested by this author are a type of idiom. According to Mel'čuk (2012, p. 37), an idiom is a lexical phraseme which is non-compositional (i.e., completely opaque). To put it more simply, (1) a full idiom does not include the meaning of any of its lexical components, for example, *by heart* = remembering verbatim; (2) a semi-idiom includes the meaning of only one of its lexical components and is attributed an additional meaning different from the other lexical component, for instance, *private eye* = private detective; and (3) a quasi-idiom incorporates the meaning of both of its lexical components, but neither of them being the central meaning of the combination, which is an additional meaning, for example, *barbed wire* = "a type of strong wire with sharp points on it, used to prevent people or animals from entering or leaving a place, especially a field."[45]

[43] https://www.ldoceonline.com/dictionary/strong (Last accessed: 25/10/2020).
[44] https://dictionary.cambridge.org/dictionary/english/high?q=high+winds (Last accessed: 25/10/2020).
[45] https://dictionary.cambridge.org/dictionary/english/barbed-wire (Last accessed: 25/10/2020).

Another key point in Mel'čuk's (1996) theory is the existence of *lexical functions* (LFs), which are used to describe syntagmatic and paradigmatic relations of words. They are regarded as semantic correspondences that associate a given LU (the element which is freely chosen) with a set of lexical items (the elements that are contingent on the LU), which express a specific meaning of the former. These semantic correspondences are divided into *standard* and *non-standard*. A standard LF "corresponds to a very general and abstract meaning that can be lexically expressed in a large variety of ways" (p. 40) and satisfies two conditions: (1) the meaning can be expressed by a vast number of LUs, and (2) these LUs can be accompanied by a considerable number of words. An illuminating example is the LF Magn: 'intense(ly), 'very,' for its meaning is "very general" and has "a very large number of expressions" (p. 43), for instance, *close shave, infinite patience, thin as a rake, to condemn strongly*. On the other hand, a non-standard LF fails to fulfill at least one of these conditions, such as 'without addition of a dairy product,' since this meaning is "too specific" and "applicable only to a few names of beverages" (p. 43), for example, *black coffee, coffee without milk*. In total, Mel'čuk suggests 25 paradigmatic LFs (pp. 47-55) and 36 syntagmatic LFs (pp. 56-72).[46]

Before proceeding to examine the most recent phraseological taxonomies, it is necessary to show a comparison of the classification of MWUs by the Russian scholars and their followers (Table 7). Although the categories included in each column may not enter into a direct correspondence with each other, it is important to consider them along a continuum in terms of their degree of compositionality and fixedness (since these are the criteria established by the authors to categorize word combinations), that is, from more semantically transparent and less restricted combinations to less semantically transparent and more restricted combinations. In Cowie's words (1998a): "categories […] are ranged along a scale or continuum from unmotivated and formally invariable idioms to partially motivated and partially variable collocations" (p. 6) (the table depicts them in reverse order):

[46] Some syntagmatic LFs can be used to encode collocations in a combinatory dictionary, as it has been proved in the DiCoInfo and the DiCoEnviro specialized dictionaries (cf. §3.3.).

Table 7

Comparison of the Classification of Multiword Units by the Russian Scholars and Their Followers

	General category	**Compositionality and fixedness**		
		+ motivated - opaque + variable - restricted ⟷		**- motivated + opaque - variable + restricted**
Vinogradov (1947)	Phraseological unit	Free phrases	Phraseological combinations	Phraseological unities	Phraseological fusions
Amosova (1963)	Phraseological unit	Free phrases	Phrasemes (more restricted) and phraseoloids (less restricted)	Idioms	
Cowie (1981) and Howarth (1996)	Composite	Open collocations	Restricted collocations	Figurative idioms	Pure idioms
Mel'čuk (1998, 2012)	Semantic phraseme	[Free phrases]	Semi-phrasemes / semi-idioms / collocations	Quasi-phrasemes / quasi-idioms	Full phrasemes / full idioms

To sum up, four distinct types of PUs can be distinguished. Given a continuum ruled by the principles of semantic compositionality and restriction, *free combinations* are found in the first place, since they are 100% compositional and unrestricted PUs. In the second place, *collocations* are 50% compositional and unrestricted PUs, as one of their elements is usually interpreted literally and the other figuratively, and these elements can be substituted for others if conforming to some language norms. In the third place, *figurative idioms* are combinations of words which allow two readings, one literal and one figurative; their elements allow scarce substitution. Finally, *pure idioms* can only be interpreted figuratively, so they are 100% non-compositional, and the substitution of any of their elements is impossible.

A steep rise in the interest in phraseology continued after the work of the Russian scholars and their followers and enthusiasm for this field has been growing since then. For this reason, the following section will cover some relevant phraseological taxonomies

suggested in the last few years by some of the most prolific authors in this subject, which we think is an appropriate way to complete the introduction to Phraseology presented in this doctoral dissertation.

2.1.4. Phraseological taxonomies in the twenty-first century

The most recent taxonomies in phraseology are greatly indebted to the authors mentioned in the previous pages. The basic idea originated by the Russian scholars and phraseologists from other countries of Eastern Europe is that phraseology is a continuum along which word combinations are situated. As we have previously mentioned, their position in this continuum depends on two fundamental aspects: (1) the degree of fixedness of the elements involved, and (2) their semantic transparency. As a result, the most fixed and opaque word combinations are located at one end (e.g., pure idioms), whereas the most variable and transparent ones are located at the other end (e.g., free combinations). However, other authors have also suggested phraseological typologies based on different criteria. In this work, we will explore those by Sag, Baldwin, Bond, Copestake, and Flickinger (2002), Granger and Paquot (2008), Heid (2008) and Baldwin and Kim (2010).

To begin with, the term *multiword expression* (MWE)[47] is used by Sag et al. (2002) to refer to combinations of words or *words-with-spaces*. Their work is oriented toward the field of NLP and they recognize that "Specialized domain vocabulary, such as terminology, overwhelmingly consists of MWEs" (p. 2). It means that MWEs appear in all text types, however, they would pose a problem if they were covered by general methods of linguistic analysis. On the one hand, the overgeneration problem is caused when a generation system is uninformed and produce incorrect combinations of words, for example, **telephone cabinet*,[48] **telephone closet* versus *telephone booth* (American) or *telephone box* (British/Australian). On the other hand, idiomatic expressions that do not conform to grammar rules might not be recognized by syntactic parsers, bringing the idiomaticity problem, for instance, *kick the bucket*.

[47] We will use the terms *phraseological unit* (PU), *multiword expression* (MWE) and *multiword unit* (MWU) interchangeably in this dissertation to refer to any combination of words, such as collocations, idioms, compounds, and so forth.
[48] We use an asterisk (*) before a linguistic expression to show that it is incorrect.

Regarding their taxonomy of MWEs, they adapt the terminology from Bauer (1983) and distinguish two general categories. First, lexicalized phrases have idiosyncratic syntax or semantics, or contain words that do not occur in isolation. They are further classified into three subcategories according to their degree of rigidity (or fixedness, as the Russian scholars did): (a) fixed expressions, which defy the conventions of grammar and are entirely rigid, not allowing any kind of morphosyntactic variation or internal modification; (b) semi-fixed expressions, whose word order and composition is fixed, but they can show some degree of lexical variation, and (c) syntactically-flexible expressions, which permit a greater degree of syntactic variability. Second, institutionalized phrases are syntactically and semantically compositional units which occur frequently[49] in a given context, thus becoming conventionalized. Table 8 (Sag et al., 2002, pp. 4-8) summarizes this typology and provides some examples:

Table 8

Typology of Multiword Expressions by Sag et al. (2002)

Lexicalized phrases	Fixed expressions	*by and large, in short, every which way*		
	Semi-fixed expressions	Non-decomposable idioms	Semantically opaque; syntactic variability is not possible, but inflection and variation in reflexive form are	*kick the bucket, shoot the breeze, trip the light fantastic*
		Compound nominals	Syntactically unalterable but inflect for number	*car park, attorney general, part of speech*
		Proper names	Syntactically highly idiosyncratic	U.S. sports team names, e.g., *the San Francisco 49ers, the Oakland Raiders*
	Syntactically-flexible expressions	Verb-particle constructions	Either semantically idiosyncratic or compositional	*brush up on, break up*

[49] The factor of frequency when defining word combinations is especially relevant, as we will see in §2.2.1.2.

		Decomposable idioms	Syntactically flexible to some degree	*let the cat out of the bag, sweep under the rug*
		Light-verb constructions	Highly idiosyncratic but fully syntactically variable; the meaning of the noun is literal whereas that of the verb is not	*make a mistake, give a demo*
Institutionalized phrases	Anti-collocations	Potential lexical variants of a given institutionalized phrase which are observed with zero or markedly low frequency		**traffic lamp* (vs. *traffic light*), **happy occasion* (vs. *auspicious occasion*)
	Collocations	Any statistically significant cooccurrence		*sell + house, drink + wine, eat + banana*

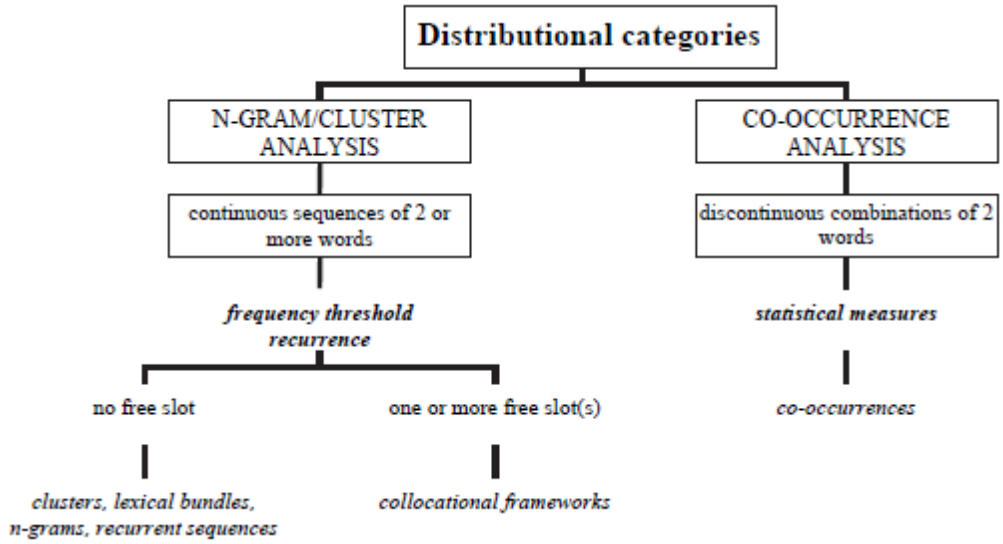
As we can see, this classification is much broader than those examined above, therefore, we will briefly comment on those categories which were also present in the taxonomies proposed by the Russian linguists and their followers. For example, Sag et al. distinguish two types of idioms: *non-decomposable idioms*, which are 100% non-compositional and rigid (also contemplated by the previous authors), such as *kick the bucket*, and *decomposable idioms*, which are syntactically-flexible, for instance, *let the cat out of the bag*. On the other hand, they also suggest two kinds of collocations: first, the so-called *collocations* are institutionalized phrases characterized for being recurrent in a language, such as *sell + house*; second, lexical variants of true collocations which are not sufficiently recurrent are called *anti-collocations*, given that they are incorrect word combinations, such as **traffic lamp* versus *traffic light*. Although the factor of frequency did not appear in the previous taxonomies, the concept of collocation is similar to the ones presented above.

The second typology of PUs covered in this section is that by Granger and Paquot (2008), who admit that the use of corpora can be a big help to the elaboration of a typology of lexical co-occurrences. They call this perspective frequency-based (p. 29) because linguists do not aim to distinguish between different linguistic categories and subcategories of word combinations, but to extract MWUs that can be either pervasive or

infrequent in language.[50] This approach emerges from distributional studies and suggests a two-fold typology of types of units (Figure 3) based on the extraction procedures used (p. 39). Nevertheless, we will not explain the concepts included in this categorization now, since some of them (e.g., frequency threshold, statistical measures, lexical bundles) will be addressed later.

Figure 3

Typology of Distributional Categories by Granger & Paquot (2008)



Apart from the distributional categories, these authors also propose a linguistic classification of PUs. For so doing, they follow Burger (1998), who based his classification on the function of units in discourse, namely, referential, communicative or structural, defining three broad categories:

1) Referential PUs encompass two types according to a syntactic-semantic criterion: (a) nominative PUs, which refer to objects, phenomena or facts of life (Cowie's composites) and is further classified into idioms, partial idioms and collocations, and (b) propositional PUs, which function at sentence or text level and refer to a statement or an utterance about objects or phenomena. The latter also include proverbs and idiomatic sentences, which belong to Cowie's formulae and Mel'čuk's pragmatic phrasemes;

[50] They use the criterion of frequency to identify word combinations, just like Sag et al. (2002).

2) Communicative PUs are used to fulfill an interactional function, such as initiating, maintaining or closing a conversation, for example, *Good morning* (they are also like Cowie and Howarth's formulae and like Mel'čuk's pragmatemes); and

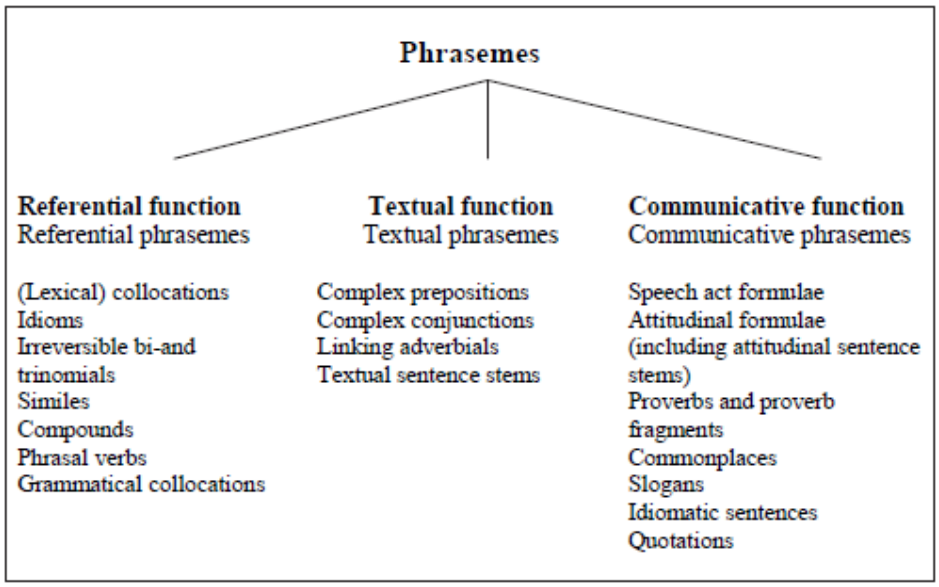
3) Unlike Cowie and Mel'čuk, Burger proposed a new category which he considered the least interesting. It is called structural and includes word combinations that establish grammatical relations, for example, *in Bezug auf [concerning], sowohl ... als auch [as well … as …]*.

Inspired by this typology, Granger and Paquot (2008) distinguish three categories: (1) *referential phrasemes*, used to convey a message, such as collocations (e.g., *heavy rain, closely linked*), idioms (e.g., *to spill the beans, to bark up the wrong tree*) or compounds (e.g., *goldfish, blow-dry*); (2) *textual phrasemes*, used to structure and organize the content, for example, complex prepositions (e.g., *with respect to, apart from*) or linking adverbials (e.g., *to conclude, in other words*); and (3) *communicative phrasemes*, used to express feelings or beliefs toward a propositional content or to explicitly address interlocutors, such as speech act formulae (e.g., *Happy birthday!, How do you do?*) and slogans (e.g., *Make love, not war*).[51] Figure 4 represents these phraseological categories and the types of PUs that they include (p. 42):

[51] For a more exhaustive definition of these categories and their subtypes, see Granger & Paquot (2008, pp. 42-44).

Figure 4

Typology of Phraseological Categories by Granger & Paquot (2008)



Another author who deserves a brief mention is Heid (2008). His work is on computational phraseology and encompasses two kinds of computational activities regarding PUs: (1) computing and computational linguistic methods applied to phraseology, and (2) methods of automatic language processing that can handle PUs in analyses of texts. Like Sag et al. (2002), he uses the phrase *multi-word expressions*[52] to refer to combinations of words because "Researchers working on aspects of this field prefer to speak of *multi-word expressions* (MWE)" (p. 339). Regarding his classification, he does not suggest his own taxonomy, but a list of MWEs that have been analyzed by computational phraseologists. Unlike the categories suggested in the Russian tradition, no order is imposed on this list, which is elaborated based on formal (e.g., binomials), semantic (e.g., idioms) and pragmatic (e.g., stereotyped comparisons) aspects. Table 9 displays this list, which includes thirteen categories, along with examples (p. 340):

[52] Unlike Sag et al. (2002), Heid hyphenates the term *multi-word*.

Table 9

Typology of Multiword Expressions by Heid (2008)

Multiword expression (MWE)	Examples
MW named entities	*New York, George W. Bush*
MW function words	· prepositions: *in spite of* · adverbs: *by and large* · conjunctions: *neither … nor*
MW adjectives	*be up to date, be excited*
Verb + particle constructions	*take off, call (someone) up*
Nominal compounds	*bookshelf, board game*
MW verbal expressions and complex predicates	*be part of, do one's best*
Collocations, including light verb constructions	*take a walk, give a speech*
Stereotyped comparisons	*to eat like a horse, to fit like a glove*
Idiomatic expressions	*to pull one's leg, to rain cats and dogs*
Binomials	*salt and pepper, bed and breakfast*
Proverbs, quotes and sayings	*An apple a day keeps the doctor away, Actions speak louder than words*
Spoken language particle clusters	*gerade mal ('just')* (German)
Routine formulae of conversation	*never mind, my pleasure*

Despite the heterogeneity of Heid's list of MWEs, some similarities with the categories proposed by the previous twenty-first-century authors (i.e., Sag et al., 2002, and Granger & Paquot, 2008) can be found. For instance, his *MW named entities* are the same type as Sag et al.'s proper names, which they consider semi-fixed. Another example is his *MW function words*, akin to Sag et al.'s fixed expressions and Granger and Paquot's textual phrasemes. Also, his *nominal compounds* are similar to Sag et al.'s compound nominals and Granger and Paquot's compounds. Furthermore, Heid's *collocations*, which is the PU covered in the current research, are present in Sag et al.'s and Granger and Paquot's taxonomies, which are called light verb constructions, anti-collocations and collocations by the former and lexical collocations by the latter. An unsurprising PU suggested by Heid is *idiomatic expressions* (already present in the Russian scholars' categorizations, it was the least compositional PU), which Sag et al. subdivide into two types, non-decomposable and decomposable idioms, and Granger and Paquot name simply idioms. Another category which is shared with the previous authors is *verb + particle*

constructions, called the same by Sag et al. and phrasal verbs by Granger and Paquot. Finally, other categories which are present in Granger and Paquot are *binomials* (irreversible bi- and trinomials), *proverbs, quotes* and *sayings* (proverbs, slogans), and *routine formulae of conversation* (speech act formulae, attitudinal formulae).[53]

The last categorization examined in this section comes under the influence of Sag et al. (2002) and is suggested from the perspective of NLP. We refer to the work by Baldwin and Kim (2010), who recognize that PUs can occur in a wide range of lexical and syntactic configurations and they point out that their semantics can produce different effects. For instance, in some cases the component elements preserve their original meaning (e.g., *serial number*, which are reminiscent of the Russian linguists' free phrases or open collocations), whereas in other cases the meaning of one of them has no bearing on the meaning of the whole (e.g., *in a nutshell*). In addition, they believe that an MWE must fulfill the requirement of (lexical, syntactic, semantic, pragmatic and/or statistical) idiomaticity, defined as "markedness or deviation from the basic properties of the component lexemes" (p. 269) (it is important to realize that semantic idiomaticity was considered a criterion to define word combinations by the Russian scholars). The authors define three broad categories in terms of the type of word around which the MWE emerges, that is, a noun, a verb or a preposition, and they include several subcategories.

The first category is *nominal MWEs*, which is one of the most common types in the world's languages. The most representative subtype is the noun compound (NC), where two or more nouns combine, being the last noun in the combination the head and the other the modifier/s, for example, *golf club, computer science department*. The head of an NC is always a noun, but the modifier/s can also be a verb or an adjective, for instance, *connecting flight*, *open secret*. According to the authors (p. 275), NCs have received special attention in NLP because of two aspects: (1) their semantics is underspecified, thus, *nut tree*, *clothes tree* and *family tree* share the same head, *tree*, but they show little semantic commonality, and (2) they need to be syntactically disambiguated when containing three or more words, such as *glass window cleaner*, which may be interpreted as "a window cleaner made of glass" (*glass [window cleaner]*) or as "a cleaner of glass window" (*[glass window] cleaner*).

[53] Some of the categories proposed by these authors, for instance, compounds, idioms and phrasal verbs, may cause confusion to language users when coming across word combinations. For this reason, they will be explained in more detail and compared to collocations in §2.2.3.

The second category suggested is *verbal MWEs*, within which the authors distinguish four subcategories:

1) Verb-particle constructions (VPCs), highly frequent in general English, which consist of a verb and an obligatory particle, for example, an intransitive preposition (e.g., *play around, take off*), an adjective (e.g., *cut short, let fly*);

2) Prepositional verbs (PVs), which contain a verb and a transitive preposition that can be fixed or mobile, depending on whether it needs to be adjacent to the verb (e.g., *come across, grow on*) or it can undergo limited syntactic alternation (e.g., *refer to the book and to the DVD*), respectively. This type of MWE is extremely frequent in English text and it may present problems when distinguishing it from VPCs and simple verb-preposition combinations;

3) Light-verb constructions (LVCs), consisting of a light verb (whose meaning contributes little to the meaning of the combination) and a noun complement. The major light verbs in English are *do* (e.g., *do a drawing, do a demo*), *give* (e.g., *give a kiss, give a sigh*), *have* (e.g., *have a drink, have a rest*), *make* (e.g., *make an offer, make an attempt*) and *take* (e.g., *take a bath, take a photograph*). This type of MWE is very frequent in English language; and

4) Verb-noun idiomatic combinations (VNICs), made up of a verb and a noun in direct object position, for example, *kick the bucket, shoot the breeze*. Their distinctive feature is their idiomaticity, which produces two subtypes: (1) decomposable VNICs, where the elements involved can match the distinct elements of the VNIC interpretation, such as *spill the beans*, meaning "reveal a secret," where *spill* matches "reveal" and *beans* matches "secret" through a process of figuration, and (2) non-decomposable VNICs, which do not allow such semantic decomposition, for instance, *kick the bucket, get the hang (of)*. This type of combination occurs with low frequency.

The last broad category along with nominal and verbal MWEs is *prepositional MWEs*, where Baldwin and Kim distinguish two subcategories. On the one hand, determinerless-prepositional phrases (PP-Ds), which rarely occur, consist of a preposition and a singular noun without a determiner. Some of these phrases display syntactic modifiability, producing different combinations, such as *by car/bus/bike*, in contrast with others that do not allow any modification, for instance, *on top* cannot be separated by a

noun, **on table top*, and are not productive, for example, *on top* versus **on bottom*. On the other hand, complex prepositions can be fixed, such as *in addition to*, or semi-fixed, for example, allowing internal modification: *with (due/particular/special) regard to*, or determiner insertion: *on (the) top of*.

Table 10 compiles the MWE taxonomy suggested by Baldwin and Kim (2010, pp. 274-279):

Table 10

Typology of Multiword Expressions by Baldwin & Kim (2010)

Nominal MWEs	Noun compound (NC)	Two or more nouns combine (or a noun + an adjective or a verb)		*golf club, connecting flight*
Verbal MWEs	Verb-particle constructions (VPCs)	A verb and an obligatory particle, e.g., an intransitive preposition or an adjective		*play around, take off, cut short, let fly*
	Prepositional verbs (PVs)	A verb and a transitive preposition	Fixed preposition (adjacent to the verb)	*come across, grow on*
			Mobile preposition (can undergo limited syntactic alternation)	*refer to*
	Light-verb constructions (LVCs)	A light verb and a noun complement (the verb is semantically light and contributes little to the meaning of the combination)		*do a drawing, give a kiss, have a drink, make an offer, take a bath*
	Verb-noun idiomatic combinations (VNICs)	A verb and a noun in direct object position; idiomatic	Decomposable VNICs (the elements involved can match the distinct elements of the VNIC interpretation)	*spill the beans*
			[Decomposable idioms in Sag et al., 2002]	

			Non-decomposable VNICs (do not allow semantic decomposition) [Non-decomposable idioms in Sag et al., 2002]	*kick the bucket*
Prepositional MWEs	Determinerless-prepositional phrases (PP-Ds)	A preposition and a singular noun without a determiner	They display syntactic modifiability and can produce different combinations	*by car/bus/bike*
			They do not allow any modification, they are not productive	*on top*
	Complex prepositions	Fixed (do not allow any kind of modification) [Fixed expressions in Sag et al., 2002]	*in addition to*	
		Semi-fixed (allow internal modification or determiner insertion)	*with (due/particular/special) regard to, on (the) top of*	

As it has been previously stated, the taxonomy suggested by Baldwin and Kim (2010) is influenced by the taxonomy proposed by Sag et al. (2002). For instance, the type of combination that the latter call compound nominals are referred to as NCs by the former. On the other hand, the VPCs and LVCs receive the same name by both authors. Finally, what Sag et al. call verb-particle constructions is named PVs by Baldwin and Kim. Regarding idioms and collocations, the current authors contemplate the two types of idioms suggested by Sag et al., namely, non-decomposable and composable VNICs. Finally, in spite of the fact that Baldwin and Kim do not use the name collocations to refer to any of the subcategories, we believe that LVCs can be regarded as such.

All in all, we can highlight that the twenty-first-century taxonomies of word combinations included in this work are inspired by the phraseological classifications

proposed by the Russian scholars. More specifically, the most recent taxonomies also apply the criteria of fixedness and idiomaticity in some cases. For instance, the first criterion is considered to distinguish MWEs that allow modifiability from those which do not, and the second criterion is used to distinguish MWEs whose meaning can be inferred from the meaning of their elements from those whose meaning goes beyond the semantics of their elements. In addition to that, Sag et al. (2002) and Granger and Paquot (2008) contemplate the factor of frequency when defining word combinations, which is of the first importance in this research, as it will be shown later. Regarding the type of combinations proposed, all the authors, except for Baldwin and Kim (2010), include the so-called collocations, which is the focus of this work and the topic of the next section.

2.2. Collocations in a framework of phraseological units

You shall know a word by the company it keeps!

(Firth, 1962, p. 11)

As it has been outlined in §2.1.1., when the Russian scholars categorized PUs, they placed collocations half-way along a continuum where freely chosen expressions were found at one end and fixed combinations were found at the other, as the criteria of fixedness and idiomaticity of the integrating elements were applied. Nevertheless, apart from these aspects, others have been discussed when defining collocations. For instance, linguists like John R. Firth (1957, 1962), Michael A. K. Halliday (1966, 1994, 1996, 2002) and John Sinclair (1991, 1996/2004, 1998, 2000) described these PUs on the basis of their frequency in texts,[54] which was possible thanks to the development of methodologies based on corpora (cf. §1.5.2.). From these authors, Sinclair merits special mention for two reasons.

On the one hand, Sinclair's contributions to the fields of corpus linguistics and lexicography materialized in the COBUILD project (of which he was the leading figure), something that would not have been possible without the technological advances that took place in the 1970s and 1980s. To this respect, the author acknowledged: "the ability to

[54] As it has been explained above, this criterion was considered by Sag et al. (2002) and Granger & Paquot (2008) for their categorization of MWEs.

examine large text corpora in a systematic manner allows access to a quality of evidence that has not been available before" (1991, p. 4).

On the other hand, Sinclair's studies on collocations resulted in notable success for defining fundamental concepts. For example, the elements of a collocation were named *node* and *collocate*, both enjoying the same status; the distinction between them is that the former is the element under study and the latter is the accompanying word, thus, the same word can be the node or the collocate, although not at the same time. Furthermore, he introduced the notion of measurement between these elements, which he called *span*, and recognized that "A span of -4, +4 […] four words on either side of the node word will be taken to be its relevant verbal environment" (p. 175). In addition, he distinguished between a *collocation*, the syntagmatic attraction at the lexical level (e.g., *hard + work/luck/facts*), and a *colligation*, the syntagmatic attraction at the grammatical level (e.g., *set about* + vb-*ing*), although he recognized that these terms were Firth's (1957) terms (Sinclair, 1998, p. 15).

Taking this into account, many are the authors (e.g., Anderson, 2006; Gablasova, Brezina, & McEnery, 2017; Granger & Paquot, 2008; Moreno Jaén, 2009; Nesselhauf, 2005) who have stated that two views on the study of collocations can be distinguished: (1) the phraseological approach, which refers to the one developed by the Russian authors, that is, the one distinguishing collocations from other MWUs in terms of the degrees of fixedness and compositionality of their elements, and (2) the statistical approach, based on the frequency of co-occurrence of two words. Nevertheless, despite offering advantages, they also have shortcomings if considered individually, as it is summarized in Table 11:

Table 11

Positive and Negative Aspects of the Phraseological and Statistical Approaches to the Study of Collocations

	Positive aspects	Negative aspects
Phraseological approach	- Aims to establish a set of criteria to define collocations, such as fixedness and idiomaticity	- Authors do not always agree on which criterion is given more attention - Does not offer a systematic method to identify and compile collocations, so manual labour seems to be the only available option
Statistical approach	- Offers a practical method of automatic corpus-based extraction and analysis	- Does not mention how frequent a word combination should be to be called a collocation - Does not explore the idiosyncratic elements of collocations compared to other recurrent MWEs

As it can be observed, while the phraseological approach concerns the semantics of the components of a collocation, the statistical approach offers a method to identify them. It means that a problem-solving approach to defining collocations may be to combine both perspectives, which has been defended by linguists like Kjellmer (1984, 1987, 1994) and Gyllstad (2007). The result is a more integrated approach that combines "the best of the two worlds" so that it can give phraseology "the place it deserves in linguistic theory and practice" (Granger & Paquot, 2008, p. 45). We will call this perspective the mixed or hybrid approach to collocations, which is the stance adopted in this work.

Nevertheless, the fact of studying collocations from different perspectives has led to a lack of agreement when it comes to their definition. For this reason, this section aims to comprehensively define them by examining their typical features (§2.2.1.). After that, a summary of typologies of collocations suggested by several authors will be provided (§2.2.2.), highlighting that by Benson, Benson and Ilson (1986b). Next, one thing that we consider necessary is to distinguish these PUs from others which may look alike, so the concept of collocation will be compared to that of *idiom, lexical bundle, collostruction, compound, phrasal verb* and *collocational catena* (§2.2.3.). Finally, this overview on

different aspects concerning collocations will help to provide the definition of the concept proposed in this study (§2.2.4.).

2.2.1. Features of collocations

In our opinion, collocations can be defined in terms of a set of formal and functional features (Moreno Jaén, 2009; Seretan, 2011). The former are related to their form and encompass three aspects: (1) number of elements, (2) type of elements, and (3) word span. The latter recognize collocations as PUs and allow their differentiation from other MWUs; they entail four aspects: (1) arbitrariness, (2) semantic compositionality, (3) frequency of use (institutionalization), and (4) prefabrication. Each of these issues will be briefly addressed in the following pages.

2.2.1.1. Formal features of collocations

Number of elements

It is generally accepted that collocations are composed of two lexical elements (Martin, 2008, p. 56; Pearce, 2001, p. 42), which do not have the same status within the combination (Granger & Paquot, 2008, p. 43; Heid & Gouws, 2006, p. 980; Moreno Jaén, 2009, p. 52) (it must be remembered that this idea is contrary to Sinclair's). Therefore, one of the elements enjoys a privileged status, is freely chosen by the speaker and determines the choice of the other. On the contrary, the second element is dependent on and restricted by the first one. Franz J. Hausmann (1979) was the first person who pointed out the different status attached to the constituent elements of a collocation, being the *base* the main element and the *collocator* contingent on it.

However, experts have not reached a consensus about the number of elements of a collocation yet. In this regard, Sinclair (1991) states: "In most of the examples, collocation patterns are restricted to pairs of words, but there is no theoretical restriction to the number of words involved" (p. 170). Similarly, Seretan (2011) highlights: "there is no length limitation for collocations" (p. 16). This lack of agreement may be due to "the recursive nature of collocational properties" whereby "the components of a collocation can again be collocational themselves" (Heid, 1994, p. 232). To explain, the elements of

a collocation can be part of other collocations at the same time, such as *make a great mistake*, where two collocations are identified: *make* + *mistake* and *great* + *mistake*.[55]

Type of elements

Another controversial subject when considering the composition of a collocation is the sort of elements that it contains. Some authors (Greenbaum, 1974; Halliday, 1966; Mitchell, 1971) believe that these elements are lexemes, that is, inflections and derivations coming from the same lexeme originate the same collocation, consequently:

> *[S]trong, strongly, strength* and *strengthened* can all be regarded as the same item; and *a strong argument, he argued strongly, the strength of his argument* and *his argument was strengthened* all as instances of one and the same syntagmatic relation. (Halliday, 1966, p. 151)

There is a second group of authors (Gyllstad, 2007) who believe that inflections from the same lexeme originate the same collocation, but derivations do not, for example, *implement a method* and *implementing methods* are examples of the same collocation; however, *the implementation of a method* is not. Finally, the third group presents a more restricted viewpoint and considers that all the elements within a collocation are lexical forms (words), thus *confirmed bachelor* is a collocation different from *confirmed bachelors* (Kjellmer, 1994; Nesselhauf, 2005).

Word span

There is not a formal agreement on how many words can occur between the elements involved in a collocation. According to Sinclair (1991, p. 175), we can find relevant collocations within a span of -4/+4, which means four words on either side of the node. A similar view has been supported by other authors (e.g., Vargas Sierra, 2012a, p. 183), who have applied spans of -4/+4 (Lindquist, 2009, p. 73; Louw, 1993, as cited in

[55] This phenomenon has been named as *collocational catena* (Pęzik, 2018), although Hill (2000) considers it a collocation itself (cf. §2.2.3.).

Barnbrook et al., 2013, p. 162; Mehler, 2009, p. 351) and -5/+5 (Culpeper & Demmen, 2015, p. 94; Evert, 2009, p. 1218; Patiño García, 2014, p. 124) in their studies. However, the components of a collocation, as well as occurring within the same clause or the same sentence, can appear in different sentences or even in different paragraphs. For example, Pęzik (2018, p. 71) found out that a base and its collocate can be separated from each other even by 16 words, which does not mean that the collocation is less significant: ***Failures*** *such as his Middle East peace process and dealing with Iran and North Korea have simply* ***led*** *to resignation and inattention*. What is more, O'Halloran (2010) chose a span of 25 words to both the left and the right of the node in his investigations on *But* "so as to capture the maximum possible co-occurrences" (p. 572).

2.2.1.2. Functional features of collocations

Arbitrariness

Collocations are regarded as arbitrary combinations of words, that is, it is widely considered that there is not an underlying motivation for two words to co-occur in the text. For example, Benson et al. (2009) emphasize that "collocations are arbitrary and non-predictable" (p. xxxii); Smadja (1993) agrees that collocations are "recurrent combinations of words that co-occur more often than by chance and that correspond to arbitrary word usages" (p. 143); finally, Moreno Jaén (2009, p. 57) notes that collocations are restricted in an arbitrary way in terms of the words involved. Additionally, Seretan (2011, p. 15) contends that they do not conform to rules. Other authors who agree with collocations' arbitrariness are Fontenelle (1992, p. 225), Van der Wouden (1994, p. 19) and Kheirzadeha and Marandi (2014, p. 942).

Semantic compositionality

Semantic compositionality of collocations implies that the meaning of the whole incorporates the meaning of its constituent elements. Nevertheless, we agree with Choueka (1988, p. 612) that compositionality is not a predominant feature of collocations, given that they are not entirely transparent in meaning, as it is also stated by Evert (2005): "[their] semantic […] properties cannot be fully predicted from those of [their]

constituents" (p. 17). Similarly, in Philip's (2011) words, they "may express untruths if compared to a literal reading" (p. 24). The author illustrates it with the following examples: first, it is utterly impossible to find the height in *highly dangerous*; second, rarely could we identify the having in *have a look*; and third, there is no way of getting ink or a vehicle in a *pen drive*.

To this regard, the Russian scholars and their followers, who considered compositionality as one criterion to categorize MWUs, suggested the existence of combinations of words containing one element interpreted literally and one element interpreted figuratively, which may be the most embraced idea when dealing with collocations nowadays. The names chosen for this type of PUs were phraseological combinations according to Vinogradov and Amosova, restricted collocations in Cowie and Howarth's terms and collocations as postulated by Mel'čuk. By contrast, those non-compositional PUs whose meaning does not derive from the literal meaning of their components were called phraseological fusions (Vinogradov), idioms (Amosova), pure idioms (Cowie and Howarth) and full phrasemes or idioms (Mel'čuk), while fully compositional combinations have normally been referred to as free combinations or free phrases.

Frequency of use (institutionalization)

The high frequency of use of a collocation leads to its institutionalization or conventionalization, since the more frequently collocations are used, the quicker they become institutionalized (Corpas Pastor, 1996, p. 21). According to Seretan (2011, p. 16), frequent recurrence is the most widely accepted characteristic when it comes to defining collocations, as it is expressed by many well-known linguists too, for instance, "sequences of lexical items which habitually co-occur" (Cruse, 1986, p. 40), "frequently repeated" (Sinclair, 1991, p. 170), "very often repeated" (Smadja, 1993, p. 147), "recurring sequences of items" (Kjellmer, 1994, p. 5), "a group of words which occur repeatedly" (Carter, 1998, p. 51), "recurrent expression" (Gledhill, 2000, p. 7), "recurrent co-occurrences" (Bartsch, 2004, p. 76).

This feature explains why the number of studies based on corpus linguistics has increased exponentially in the last decades. Quantitative methods used in this discipline allow to analyse massive blocks of text and gather accurate objective information on the

frequency of co-occurrence (one word co-occurring with another) and recurrence (the frequency of two words co-occurring together in a language) of words. As a result, linguists can know which word combinations are considered statistically significant. In fact, it must be highlighted that Sag et al. (2002) "reserve the term collocation to refer to any statistically significant cooccurrence" (p. 7) (cf. §2.1.4.).

Prefabrication

Collocations are regarded as prefabricated phrases that we learn and store as chunks of language, that is, they become entrenched in speakers' minds as a single unit. This idea was first put into effect by Palmer (1933) (cf. §2.1.), since his approach was pedagogical and promoted the learning of collocations as an integral whole. After that, other authors who have supported this view are Sinclair (1991): "[collocations are] semi-preconstructed phrases that constitute single choices" (p. 170), and Corpas Pastor (1996): "los hablantes […] no van creando sus propias combinaciones originales de palabras al hablar, sino que utilizan combinaciones ya creadas y reproducidas repetidamente en el discurso"[56] (p. 22), to name but a few. This aspect of collocations is crucial in translation studies, since these word combinations should not be translated as single words (Štefic, Mravak-Stipetić, & Borić, 2010, pp. 177-178).

As it has been previously stated, all these formal (number and type of elements and word span) and functional (arbitrariness, semantic compositionality, frequency of use and prefabrication) features of collocations will help to formulate a definition of the concept as it is seen in this study. Before that, the next section will focus on one of the most accepted typologies of collocations that has been suggested so far.

2.2.2. Types of collocations

As well as collocations have been defined in different ways, they have been classified according to diverse criteria. For example, linguists like Hill (2000) distinguish two types of collocations, *strong* and *weak*. The former refers to collocations whose constituent elements show a very close semantic relationship, thus the possibilities of association are

[56] "speakers [...] do not create their own original combinations of words when they speak, but they use ready-made combinations that are repeatedly reproduced in speech" (translation mine).

limited. For instance, the adjective *auspicious* does not collocate with many words, although it co-occurs in the collocations *auspicious occasion, auspicious moment* and *auspicious start*. The latter refers to collocations which allow a wider range of different combinations, such as the noun *moon* in *crescent moon, full moon, half moon, new moon* and *harvest moon*.

Other authors have classified these PUs according to the type of word of their components (e.g., noun, verb, adjective, adverb, etc.). One of the most well-known and revolutionary classifications was proposed by Morton Benson, Evelyn Benson and Robert Ilson in 1986. They were influenced by the Russian tradition and endeavored to categorize word combinations. For so doing, they (Benson, Benson, & Ilson, 1986a) identified six large groups in terms of their degree of cohesiveness, that is, their ability to relate in a reasonable way to form a whole, namely: (1) *free combinations*, (2) *idioms*, (3) *proverbs* and *sayings*, (4) *collocations*, (5) *transitional combinations*, and (6) *compounds* (pp. 252-254). These word combinations have already been addressed in the previous sections, except for the transitional combinations, which these authors place between idioms and collocations.[57]

Influenced by Cowie, Howarth and Mel'čuk, Benson et al. (1986a) suggest the existence of a continuum where MWUs are located, positioning collocations between free combinations and idioms, given that collocations are more restricted than the former but not as frozen as the latter. In their opinion, the meaning of the whole can normally be inferred from the meaning of the parts, which are entrenched as a unit and used together frequently, for example, *to commit murder* (p. 253). Without a doubt, these PUs are the most interesting to these authors, since they are the central point of *The BBI Combinatory Dictionary of English* (cf. §3.2.2.), which they compiled in 1986. This dictionary is based on the authors' classification of collocations into two main categories, *grammatical* and *lexical* (1986b). In the former group, a lexical word (noun, adjective, verb, adverb) is combined with a preposition or grammatical structure, such as a *that*-clause or an infinitive, for instance, *an agreement that, necessary to* (this combination is referred to as colligation by Sinclair). Eight broad types of grammatical collocations are identified,

[57] Like the Russian scholars, Benson et al. apply the criteria of fixedness and compositionality to distinguish MWUs. Thus, they state that transitional combinations "are more *frozen* than ordinary collocations," but "unlike idioms these phrases seem to have a meaning close to that suggested by their component parts" (1986a, p. 254).

each of which consists of a significant number of subtypes (e.g., G8 consists of nineteen English verb patterns) (pp. x-xxii):

- G1: noun + preposition (except *of* and *by*), for example, *blockade against, apathy toward*;
- G2: noun + *to* + infinitive, such as *a pleasure to, an attempt to*; this group excludes: (1) infinitives associated with the sentence where the phrase *in order* may be inserted (e.g., *she closed the window (in order) to keep the flies out*), and (2) infinitive constructions that can be replaced by a relative clause (e.g., *a book to read = a book that should be read, a procedure to follow = a procedure that is to be followed*);
- G3: noun + *that*-clause, for example, *we reached an agreement that she would represent us in court*; however, this group does not include nouns followed by relative clauses introduced by *that*, that is, when *that* can be replaced by *which*: *we reached an agreement that/which would go into effect in a month*;
- G4: preposition + noun, such as *by accident, in agony*;
- G5: adjective + preposition occurring in the predicate, for example, *they were angry at everyone;* adjective + *of*-constructions are included when the subject of the construction is animate: *boys are afraid of him*; however, derived prepositions (e.g., *regarding, concerning*) and past participles + *by* (e.g., *abandoned by, absolved by*) are excluded from G5;
- G6: predicate adjective + *to* + infinitive, for instance, *it was necessary to work* (this example is empty subject and a prepositional phrase with *for* can be inserted: *it was necessary for her to work*), *she (the girl) is ready to go* (the subject is real and usually animate). However, past participles that can be followed by a *to* + infinitive phrase of purpose (e.g., *the text was proofread (in order) to eliminate errors*) and past participles that are used in passive constructions (e.g., *she was appointed to serve as our delegate*) are excluded;
- G7: adjective + *that*-clause, for example, *she was afraid that she would fail the examination*; and

- G8: this group consists of nineteen English verb patterns, designated by the capital letters A to S:[58]
    - G8-A: verbs that allow the dative movement transformation, that is, when the direct and indirect objects are nouns, the indirect object can be moved before the direct object with deletion of *to*: *he sent the book to his brother → he sent his brother the book*;
    - G8-B: transitive verbs that do not allow the dative movement transformation as explained in G8-A, for example, *they described the book to her → *they described her the book*;
    - G8-C: transitive verbs using the preposition *for* that allow the dative movement transformation, that is, the indirect object can be moved before the direct object with deletion of *for*, for instance, *she bought a shirt for her husband → she bought her husband a shirt*;
    - G8-D: verb + preposition (+ object), such as *to act as (a robber), to serve as (a servant)*; compound verbs + preposition are included (e.g., *to break in on, to catch up to*). However, combinations of verb + *by/with* when the latter denote *means* or *instrument* are excluded, for example, *they came by train, we cut bread with a knife*;[59]
    - G8-E: verb + *to* + infinitive, for instance, *they began to speak, she continued to write*; however, verbs which are usually used in a phrase showing purpose (*in order* can be inserted) are excluded: *he was running (in order) to catch a train*;
    - G8-F: verb + infinitive without *to*, for example, *we must work*; this group of verbs are called *modals* and exclude *dare, help* and *need*;
    - G8-G: verb + verb-*ing*, for instance, *they enjoy watching television, he kept talking*. Some of these verbs also belong to G8-E (e.g., *they began speaking, she continued writing*) and some of these pairs of

[58] This is the group which would encompass the verbs extracted in this work. However, it must be clarified that grammatical collocations will not be considered for their implementation in *DicoAdventure*, the specialized resource addressed in this dissertation, as they are regarded as *government patterns*, that is, the valency of the verb term (cf. §4.4.2.).

[59] Some of the combinations included in this group may be considered phrasal verbs, a type of MWU present in Granger & Paquot's (2008) taxonomy (cf. §2.1.4.) and further explained in §2.2.3.

verbs have a similar meaning, while there are others which do not. For instance, *he remembered to tell them* means that "he intended to tell them and told them" and *he remembered telling them* means that "he remembered the act of telling them;"

- G8-H: transitive verb + object + *to* + infinitive, for example, *she asked me to come, they challenged us to fight*; most of these verbs can be passivized, although not all of them (e.g., *bring, intend, like, prefer, thank, trouble, want, write*);
- G8-I: transitive verb + object + infinitive without *to*, for example, *she heard them leave, he felt his heart beat*;
- G8-J: verb + object + verb-*ing*, for instance, *I caught them stealing apples, we found the children sleeping on the floor*. Some of these verbs also belong to G8-I (e.g., *she heard them leaving, he felt his heart beating*) and most of them can usually be passivized (e.g., *they were caught stealing apples, the children were found sleeping on the floor*);
- G8-K: verb + possessive (pronoun or noun) + verb-*ing*, for example, *please excuse my waking you so early, they love his clowning*;
- G8-L: verb + *that*-clause, for instance, *they admitted that they were wrong, she believed that her sister would come*. Some verbs always take a noun or pronoun object before the *that*-clause (e.g., *she assured me that she would arrive on time*), other verbs can be used with or without a nominal object (e.g., *he bet that it would rain / he bet me that it would rain*), some verbs allow the insertion of *the fact* with little or no change in meaning (e.g., *he acknowledged (the fact) that he was guilty*) and some verbs regularly have an *empty* subject with *it* (e.g., *it appears that they will not come*);
- G8-M: transitive verb + direct object + *to be* + adjective/past participle/noun/pronoun, for example, *we consider her to be very capable, we found the roads to be cleared of snow, the court declared the law to be a violation of the Constitution*; this group includes verbs that normally take *to be* after the direct object,

however, those verbs which take a different infinitive belong to G8-H;

- G8-N: transitive verb + direct object + adjective/past participle/noun/pronoun, for example, *he made his meaning clear, the soldiers found the village destroyed, her friends call her Becky*; some patterns are similar to those in G8-M: *we consider her (to be) very capable, we found the roads (to be) cleared of snow, the court declared the law (to be) a violation of the Constitution;*
- G8-O: transitive verbs that can take two objects, but neither of which can normally be used in a prepositional phrase with *to* or *for*, for example, *the teacher asked the pupil a question, we bet her ten pounds*. They can be used with only one of their objects alone (e.g., *the teacher asked the pupil, the teacher asked a question*) and they can usually be passivized because at least one of the objects can become the subject of the passive construction (e.g., *a question was asked, ten pounds were bet*). Some structures are close to G8-N, for example, *they called him a fool*, but, in this case, the noun (*a fool*) is a predicate complement of the object (*him*);
- G8-P: intransitive/reflexive/transitive verb + adverb/prepositional phrase/noun phrase/clause, for example, *he carried himself well, Tuesday comes after Monday*; the basic idea is that an adverbial is required to form a complete sentence. Intransitive verbs requiring adverbials of duration or measurement are included (e.g., *the meeting will last all day, the trunk weighs thirty pounds*), however, phrasal verbs are excluded (e.g., *hang around, well up*);[60]
- G8-Q: verb + interrogative word (*wh*-word), for example, *he wants what I want*, or verb + object + interrogative word (*wh*-word), for example, *we told them what to do*; a few verbs can be used with or without an object (e.g., *she asked why we had come / she asked us why we had come*);

[60] Unlike in G8-D, the authors specify that phrasal verbs are not included in this group.

- G8-R: *it* + transitive verb (expressing emotion) + *to* + infinitive, or *it* + transitive verb (expressing emotion) + *that*-clause, for example, *it surprised me to learn of her decision, it surprised me that our offer was rejected*; and
- G8-S: intransitive verb + predicate noun/predicate adjective, for example, *she became an engineer, she was enthusiastic*.

On the other hand, the second group of collocations proposed by Benson et al. (1986b) is lexical collocations, which are constituted by two content words, that is, noun, adjective, verb, adverb, for example, *warmest regards, to commit murder*. The authors recognize seven major types of this kind (pp. xxiv-xxiii):

- L1: transitive verb denoting *creation* or/and *activation* + noun/pronoun/prepositional phrase, for example, *make an impression, come to an agreement*; some other collocations without meaning *creation* or *activation* are included (e.g., *do the laundry, resist temptation*);
- L2: verb meaning *eradication* or/and *nullification* + noun, for instance, *break a code, annul a marriage*. However, combinations of the verb *destroy* plus nouns are not included because this verb can be used with a very large number of nouns, as a result, they are free combinations (e.g., *destroy a bridge, destroy a document, destroy a school*);
- L3: adjective + noun, such as *strong tea, a chronic alcoholic*; combinations of noun + noun when the first noun is used as an adjective are included (e.g., *jet engine, land reform*);
- L4: noun + verb: the verb denotes an action characteristic of the person or thing designated by the noun, for example, *adjectives modify, blood circulates*; predictable combinations are excluded (e.g., *bakers bake, dancers dance*);
- L5: $noun_1$ + *of* + $noun_2$: in this group two types are distinguished: (1) $noun_1$ denotes the larger unit to which a single member belongs ($noun_2$), for example, *a colony of bees, a pack of dogs*, and (2) $noun_1$ denotes the specific, concrete, small unit of something larger, more general ($noun_2$), for instance, *a piece of advice, an act of violence*;

- L6: adverb + adjective, for example, *strictly accurate*, *sound asleep*; and
- L7: verb + adverb, such as *affect deeply, apologize humbly*.

This detailed classification of lexical collocations inspired authors like Hausmann (1989), Corpas Pastor (1996), Hill (2000) and Koike (2001). The first two suggest the same sorts of lexical combinations in French and Spanish, respectively. Hill is focused on collocation teaching and highlights seven kinds of lexical collocations as the most useful for students, namely: (1) adjective + noun (e.g., *huge profit*), (2) noun + noun (e.g., *pocket calculator*), (3) verb + adjective + noun (e.g., *learn a foreign language*), (4) verb + adverb (e.g., *live dangerously*), (5) adverb + verb (e.g., *half understand*), (6) adverb + adjective (e.g., *extremely clean*), and (7) verb + preposition + noun (e.g., *be in mind*). Regarding Koike, the classification proposed is reminiscent of Benson et al.'s, but we must highlight a significant contribution, which is the verb + adjective collocation, for example, *stay awake*.

From our point of view, the categorization of collocations suggested by Benson et al. (1986b) can be regarded as the most complete description of collocational types to date, since it gives full details of both grammatical and lexical collocations. Nevertheless, we find it necessary to add Koike's (2001) new type, as we think verb + adjective collocations are also common in languages, more specifically, in English, for instance, *remain critical, fall ill, stay single*. Therefore, the taxonomy proposed in this dissertation is inspired by these two pieces of work and is shown in Table 12:[61]

[61] We use "Gr" for grammatical collocations and "Le" for lexical collocations.

Table 12

Collocation Taxonomy Proposed in This Dissertation

Type of collocation		Examples
Grammatical	Gr1-Noun + preposition	*dream about, harm from*
	Gr2-Noun + *that*-clause	*idea that, news that*
	Gr3-Preposition + noun	*below the level, within an hour*
	Gr4-Adjective + preposition	*late for, happy with*
	Gr5-Adjective + *that*-clause	*afraid that, glad that*
	Gr6-Verb + preposition	*apologize for, arrive at*
	Gr7-Verb + *that*-clause	*suggest that, believe that*
Lexical	Le1-Adjective/noun + noun	*serious problem, company boss*
	Le2-Verb + noun	*break a consensus, pose a risk*
	Le3-Noun + verb	*a war breaks out, his writings reflect*
	Le4-$Noun_1$ + *of* + $noun_2$	*the world of politics, country of origin*
	Le5-Adverb + adjective	*completely wrong, heavily reliant*
	Le6-Adverb + verb/verb + adverb	*honestly think, sleep soundly*
	Le7-Verb + adjective	*become tired, smell delicious*

As it can be seen, we identify seven types of grammatical collocations (Gr1–Gr7) and seven types of lexical collocations (Le1–Le7).[62] In the first group, we include combinations of a lexical word plus a function word, that is, noun/adjective/verb + preposition/*that*-clause, as well as combinations of a preposition followed by a noun. It is important to mention that relative clauses introduced by *that* are excluded. In the second group, we place collocations of two lexical words (nouns, verbs, adjectives, adverbs). Moreover, word order matters in some cases. For example, we consider different collocations the verb + noun and the noun + verb collocations, where the noun is the object in the former and the subject in the latter. However, word order in adverb + verb/verb + adverb collocations is not relevant, so they belong to the same category. Another key point is that, following Benson et al., noun + noun collocations where the first noun is used as an adjective fall into the same category as adjective + noun collocations.

[62] However, only lexical collocations will be contemplated for their implementation in *DicoAdventure*. More specifically, as we are concerned with verb collocations, the relevant types to this research are Le2, Le3, Le6 and Le7 shown in Table 12 (cf. §4.4.2.).

Although we have mainly focused on the classification of collocations proposed by Benson et al. (1986b), it must be said that other authors have suggested different kinds of collocations (Nesselhauf, 2005, pp. 34-36; Torner Castells & Bernal, 2017, pp. 166-168; Van der Wouden, 1997, pp. 6-7). This means that there is a lack of consensus among phraseologists on the typology of these PUs. In some cases, they are even referred to with other names, such as compounds or idioms. For this reason, we think it is indispensable to explore some basic differences between collocations and other PUs, which is the purpose of the following section.

2.2.3. Collocations versus other phraseological units

> *Categorization is notoriously difficult in phraseology because of the bristling array of variables ─syntactic, pragmatic, stylistic, semantic─ which the material is constantly throwing up.*
>
> (Cowie, 1998b, p. 210)

Collocations are difficult to define if we think about the broad range of aspects involved. Therefore, it is hardly surprising that phraseologists do not adopt a standard definition. In addition, they are usually compared to other concepts that also encompass combinations of words, which complicates the issue (Vargas Sierra, 2012a, p. 182). Thus, the following lines will provide a brief explanation of some PUs which are quite differentiated from, and closely related to at the same time, our object of study, namely, *idiomatic expressions*, *lexical bundles*, *collostructions*, *compounds*, *phrasal verbs* and *collocational catenae*.

To begin with, *idiomatic expressions* (Howarth, 1996; Zgusta, 1971) or idioms (e.g., Amosova, 1963; Cowie, 1998a; Mel'čuk, 1998) are present in every categorization of word associations, as it has been explained in previous pages of this doctoral dissertation. They generally refer to the most restricted and least compositional combinations in a continuum where free combinations are located at the other end. Thus, idioms are frozen groupings of semantically and grammatically inseparable units, characterized by a lack of semantic transparency when aiming to unveil their meaning,

since it bears no relation to the meaning of the individual parts. Apart from semantic compositionality, there are other features which allow the distinction between idioms and other PUs:

> Fortunately, a number of features set out the parameters for idiomaticity, making it possible to differentiate idioms (e.g., *not the end of the world*) from lexical bundles (e.g., *at the end of*) or collocations (e.g., *world view*). These are compositionality/analysability, salience, semantic transparency, and adherence to truth conditions. The first three are phrase-internal, pertaining to the relationship between the discrete words within the idiom, while the fourth is identified by contrasting the idiom with its context. The four features function together, not in isolation, and represent clines or gradations of idiomaticity rather than discrete categories. (Philip, 2011, p. 17)

What Philip tries to communicate above is that idioms are non-compositional and semantically opaque, since their constituent elements cannot be taken apart and understood independently so as to unveil the meaning of the whole. However, collocations work differently, given that it is widely recognized that at least one of the elements keeps its literal meaning. Additionally, idioms are more salient than collocations because their meaning tends to be coded in the mental lexicon more easily than that of collocations. Finally, truth conditions are determined by context. In the case of idioms, there is usually a mismatch between the contextual situation and the truthfulness of the lexical item, that is, a literal reading does not actually correspond with reality. For instance, in *wave a red flag in front of a bull*, there is no red flag or bull.

Another PU which has been compared to collocations is *lexical bundle*,[63] mentioned in Philip's statement. Biber and Conrad (1999) try to draw a distinction between idioms, collocations and lexical bundles. When comparing this concept to that of idioms, the authors highlight idioms' well-known properties of fixedness and non-compositionality, visible in the examples *kick the bucket* ("die") and *bear in mind* ("remember"). When talking about collocations, they focus on the fact that their elements

[63] This PU is included in the typology of distributional categories suggested by Granger & Paquot (2008) (cf. §2.1.4.).

are associated as well as their high degree of compositionality. To this respect, they compare the adjectives *little* and *small*, which may seem synonyms out of context but show a tendency to co-occur with different nouns, for example, *little + baby/devil/dog/girl, small + amount/letters/piece/size*. Thus, they regard lexical bundles as "extended collocations: sequences of three or more words that show a statistical tendency to co-occur," such as *do you want me to, I said to him, in the case of the* (p. 183). Moreover, Biber and Barbieri (2006) highlight the function of these PUs in a language: "[lexical bundles] serve important discourse functions in both spoken and written texts" (p. 264).

Next, the concept of *collostruction* was coined by Stefanowitsch and Gries in 2003.[64] This term refers to the combination of lexemes and a construction mutually attracted. Thus, they deal with *collexemes*, that is, "Lexemes that are attracted to a particular construction," and *collostructs*, that is, "construction[s] associated with a particular lexeme" (p. 215). They base their concept of construction on Lakoff (1987) and Goldberg (1995). Furthermore, they propose a method called *collostructional analysis*, which they say to be a type of collocational analysis (since collostructions are characterized by an association between collexemes and collostructs), which "always starts with a particular construction and investigates which lexemes are strongly attracted or repelled by a particular slot in the construction" (p. 214). For instance, a study on the construction [N *waiting to happen*] revealed that the slot [N] is usually occupied by a noun with a negative semantic prosody, such as *accident* or *disaster*. On the other hand, they define collocations as "(purely linear) co-occurrence preferences and restrictions pertaining to specific lexical items" (p. 210), that is, no constructions are involved.

Another category of PU that must be compared to collocations is *compounds*,[65] which Mitchell (1975) describes as "bipartite and uninterruptable [associations of words], except by *and* in the special case of coordinate compounds" (p. 129). However, these features do not seem to be enough when defining a compound, so Donalies (2004) mentions some others: (1) they are complex, (2) they are formed without word-formation affixes, (3) they are spelled together, (4) they have a specific stress pattern, and (5) they are inflected as a whole. Scalise and Vogel (2010, pp. 1-4) stress the importance of

[64] They carried out more work on collostructions in 2004 and 2008, to mention but a few.
[65] This category is included in the taxonomies suggested by Sag et al. (2002), Granger & Paquot (2008), Heid (2008) and Baldwin & Kim (2010) covered in §2.1.4.

compounds in all areas of linguistics for several reasons: (1) it is the only process of word formation in some languages; (2) they are *words* but, at the same time, exhibit a type of *internal syntax*; (3) they are a contact point between syntagmatic and paradigmatic relationships; (4) the meaning of a compound may have a range of possible meanings; and (5) they provide insight into the early stages of language evolution. Roughly speaking, compounds differ from collocations in that the former function as a single unit of meaning, such as *toothbrush*, *farmhouse*, whereas the latter are combinations of, at least, two words where each of the constituent elements contributes to the meaning of the whole combination.

The following PU which needs to be differentiated from collocations is *phrasal verbs*. They were considered by Benson et al. (1986a, 1986b) and Granger and Paquot (2008) and have normally received attention in linguistic studies, as they are defined as a non-compositional group of words (like idioms), more specifically, a combination of a verb plus one or two adverbs or prepositions (Mart, 2012, p. 114, based on Koprowski, 2005), such as *carry out, hand in, look down on*. Additionally, it is the sum of both parts what gives the meaning to a phrasal verb, which is different from the meaning of any of those parts. By contrast, verb collocations including the same type of items are more semantically transparent than phrasal verbs and both of their integrating elements contribute to the meaning of the whole.

Finally, a *collocational catena* is described as "an independently recurrent, multiword dependency subtree which is formally and/or contextually stereotyped" (Pęzik, 2018, p. 96). To put it differently, several words which establish dependency relations combine to form longer structures. For example, the sequence *run + long + fingers* in *She ran her long fingers over a carving of a boat* is a collocational catena because "it is recurrent as well as formally and contextually stereotyped; it usually instantiates the verb *run* in its past form and it is found in narrative descriptions of people more often than in other text genres" (p. 97). Therefore, a collocational catena can be regarded as the cluster of several collocations, such as *run + fingers*, *long + fingers*. In this regard, it must be emphasized that the combination of two collocations is contemplated by Hill (2000), but this author also calls it a collocation. More specifically, the author regards this type of word combination as a verb + adjective + noun collocation, for instance, in *learn a foreign language* the collocations *learn + language* and *foreign + language* are identified.

Having said that, Table 13 summarizes the features of the distinct PUs explained in this section in terms of four criteria: compositionality, fixedness, number of elements and frequency. In short, collocations satisfy these criteria in the same way as collocational catenae do, given that they share the essential features, the only difference is that the latter is an amalgam of collocations, so they contain more than two elements. The rest of the PUs differ from collocations in, at least, one of these three criteria: compositionality, fixedness and number of elements, given that they all are frequent in English, as it has been proved by corpus-based studies carried out by authors dealing with each of these PUs. The next section will explain in more detail the concept of collocation formulated in this study.

Table 13

Collocations Versus Other Phraseological Units

	Compositionality	**Fixedness**	**Number of elements**	**Frequency**
Collocation	50%	50%	2	Yes
Idiom	No	100%	+2	Yes
Lexical bundle	100%	50%	+2	Yes
Collostruction	100%	50%	+2	Yes
Compound	50%	100%	2	Yes
Phrasal verb	No	100%	(+)2	Yes
Collocational catena	50%	50%	+2	Yes

2.2.4. The definition of *collocation* in this study

Having discussed some key aspects of the concept of collocation, it will be convenient here to take stock of these discussions and summarize the view of collocation held in this research. Regarding the approach to tackle collocations, two perspectives have been previously mentioned, phraseological and statistical. On the one hand, we believe that collocations can be defined according to two main criteria: fixedness and idiomaticity, as postulated by the Russian scholars Vinogradov (1947) and Amosova (1963) and their followers Cowie (1981), Howarth (1996) and Mel'čuk (1998). The first criterion refers to the combinatory restrictions on the parts of a collocation, whereas the second criterion

refers to the extent to which the elements in a collocation have a literal or a non-literal meaning. However, we consider that these criteria are not sufficient and require to be complemented with a statistical perspective.

To our mind, frequency of co-occurrence must be considered when defining collocations, as it is contended by Firth (1957), Halliday (1966) and Sinclair (1991). Generally, when speakers access a word, lexical activation occurs, thus the access to other words is activated both on the syntagmatic axis (e.g., collocations) and the paradigmatic axis (e.g., synonyms, antonyms) (Jenkins, 1970; Wolter, 2001). As for collocations, this process is only possible because words are very frequently combined. Taking this into account, it could be highlighted that the phraseological approach offers fine-grained linguistic analyses of word combinations, from which the statistical approach can benefit. In addition, the statistical approach offers a method of collocation identification and extraction based on corpora, which can be of great help to phraseologists. For this reason, and influenced by Kjellmer (1984), Gyllstad (2007) and Granger and Paquot (2008), we consider a hybrid perspective a favourable position on the matter.

With respect to a definition of this PU, we believe that both the formal and functional aspects explained in Section 2.2.1. must be considered.[66] Thus, we state that collocations are restricted combinations of, at least, two words (Martin, 2008; Pearce, 2001) which hold a syntactic relationship. Sometimes more than two words co-occur, given that the two lexical words making a collocation may be linked by a grammatical word (Vargas Sierra, 2012a, p. 183), such as the type Le4-Noun$_1$ + *of* + noun$_2$ in Table 12 (cf. §2.2.2.), or in cases of collocations containing a phrasal verb or a compound. In our view, the lexical integrating elements do not enjoy the same status (Hausmann, 1979), given that the base is freely chosen and its collocates are contingent on it.[67] Additionally, we believe that, at first, these elements combine in an arbitrary way.

Furthermore, as it has been aforementioned, one prominent feature of collocations is that their parts co-occur frequently in a speech community (Cruse, 1986; Seretan, 2011; Smadja, 1993), which can be known by using association measures in corpus-based studies, as it will be shown in Chapter 4. Two consequences of their high recurrence in a

[66] Formal aspects: (1) number of elements, (2) type of elements, and (3) word span; functional aspects: (1) arbitrariness, (2) semantic compositionality, (3) frequency of use (institutionalization), and (4) prefabrication.

[67] We borrow the term *base* from Hausmann (1979) to refer to the privileged word and *collocate* from Sinclair (1991) to designate the accompanying and dependent element.

language are: (1) collocations' entrenchment (prefabrication), that is, they are memorized as groups of words which are considered single choices (Pęzik, 2018), therefore, access to the base involves immediate access to its collocates, and (2) collocations' institutionalization (Corpas Pastor, 1996, pp. 21-23; Vargas Sierra, 2012a, p. 183), that is, they become readily accepted by speakers and other word combinations are blocked.

With regard to their meaning, we believe that they are mostly monosemic (i.e., they have only a single meaning) and display varying degrees of transparency (Choueka, 1988), depending on whether their elements are used with a literal or a specialized (e.g., figurative, delexical or technical) meaning. Nevertheless, we agree with Vinogradov (1947), Cowie (1998a) and Mel'čuk (2012), to name but a few, that there is every likelihood that one of the elements of a collocation must be interpreted figuratively. More specially, the meaning of the collocate is normally influenced, or even imposed, by the meaning of the base, so it is the base the one which usually keeps its literal meaning. Evert (2005) also holds this view and sustains that, in some cases, the meaning of a collocation cannot be clearly derived from the meaning of its parts (p. 17).

Another significant point to mention is the types of collocations which can be found in terms of their integrating words. As we explained in Section 2.2.2., where we included our own taxonomy of collocations, we agree with Benson et al. (1986b) that they can be grammatical, that is, a dominant word (noun, adjective, verb, adverb) combines with a preposition or grammatical structure, or lexical, that is, two dominant words occur together. In cases of inflections and derivations, the creation of different collocations depends on the syntactic type of the integrating elements. For example, *hunt a murderer* is a verb + noun collocation, while *a hunted murderer* is an adjective + noun collocation. In other words, they are distinct collocations from a syntactic point of view, despite knowing that they are combinations of the same lexemes.

Finally, the last important issue arises from the span between these integrating parts. To clarify, we support the view that the constituents of a collocation do not necessarily co-occur in immediate textual slots but within a varying span of words (Culpeper & Demmen, 2015; Sinclair, 1991), consequently, they can appear in the same sentence or even in different sentences or paragraphs. The implementation of this aspect in the current research will be explained in Section 4.4.2.

As it has been previously mentioned, this work is focused on lexical collocations whose base is a verb and can enter in a combination with a noun, an adverb or an adjective. Since this study is corpus-based, two key aspects that we will cover are: (1) the frequency of co-occurrence of a collocation's integrating elements, and (2) their association score (cf. §4.4.2.). The collocations under study are characterized for being regarded as specialized collocations, for they will be extracted from a specialized corpus containing texts on the SL of adventure tourism. For this reason, the following section will address the issue of specialized collocations and will outline previous studies on these word combinations in the domain of tourism.

2.3. Specialized collocations and their role in the language of tourism

As with general language, specialized

collocations describe regular associations

between terms in a particular field.

(Orliac, 2008, p. 379)

As it was mentioned at the beginning of this chapter, groupings of words were first analyzed by Bally (1909), who applied the criterion of fixedness for their classification. Then, Palmer (1933), more concerned with language teaching, adopted a pedagogical approach and supported the view of collocations as an integral whole. After that, research on PUs became more extensive when a group of Russian scholars and their followers, such as Vinogradov (1947), Howarth (1996) and Mel'čuk (1998), aimed to categorize all existing MWUs, making Phraseology emerge as a discipline. However, these authors' investigations chiefly centered on PUs found in the GL, ignoring their role in SLs, which led to the treatment of specialized PUs within the discipline of Terminology (Bevilacqua, 2004; Pavel, 1993; Picht, 1987; Sanz Vicente, 2011). In addition to that, Cabré Castellví, Estopà Bagot, and Lorente Casafont (1996, pp. 5-6) explain that specialized phraseology has become the object of study of terminologists for four reasons: (1) it allows the identification and conceptual delimitation of terms, given that they contextualized; (2) the description of phraseology in every SL and its representation in terminological resources make speakers use specialized PUs more; (3) the development of computer engineering and corpus linguistics has facilitated the study of specialized phraseology; and (4) the

establishment of criteria for selecting and including phraseology in specialized resources has been necessary.

Additionally, it is not strange that specialized phraseology has been covered by terminologists if one thinks that both general and specialized PUs share dominant features. For instance, Lorente Casafont (2002, p. 175) lists the following: (1) they are complex combinations of lexical and grammatical units; (2) they are syntagmatic structures; (3) they are not rigidly fixed, but allow some variation; and (4) their meaning may or may not be inferred from their parts, depending on the degree of compositionality of the PU. By the same token, the author defines specialized phraseology as the set of PUs of a language that have a specialized meaning. In her opinion, they are usually non-lexicalized, but they are stereotyped and display some degree of fixedness; moreover, they contain at least one term. With this in mind, she establishes a distinction between general PUs and specialized PUs based on five criteria (p. 176):

Table 14

Distinction Between General and Specialized Phraseological Units (Lorente Casafont, 2002)

General PUs	Specialized PUs
They are combinations of words found in the GL	They are combinations of terms (or, at least, they contain one term)
They have a connotative function	They have a denotative function
They carry general meaning, related to our knowledge and experience of the world	They carry specialized knowledge which is voluntarily acquired
They are frequent in speech and literary expressions	They are mainly found in written texts, since oral texts are not commonly analyzed in specialized domains
They represent a linguistic, cultural, religious community or a social group	They represent an academic community that is specialist in a given topic

As we can observe in Table 14, general PUs contain words used in the GL with a connotative function. In other words, they belong to our knowledge of the world and the knowledge gained from experience, so they can carry an array of positive and negative associations. They are usually used in speech and literary contexts and are characteristic of different groups, such as linguistic, cultural, religious and social groups. On the other hand, specialized PUs consist of terms (or they contain at least one) that perform a

denotative function (i.e., they carry a precise meaning) in the texts in which they occur. Furthermore, they are conventional (Vargas Sierra, 2012a, p. 185), for they have been created and are employed by a community of experts in a specialized domain, that is, by people who have acquired specialized knowledge by choice, and then they are used to spread this knowledge.

Among the great variety of PUs that can be detected in SLs (e.g., idioms, compounds, lexical bundles, etc.), this section focuses on specialized collocations, given that this is the object of study of this dissertation. In order to define this concept, we follow Patiño García (2014), who coincides with Lorente Casafont (2002), Lorente Casafont, Martínez-Salom, Santamaría, and Vargas Sierra (2017) and L'Homme (2017). Patiño García's (2014) definition is reminiscent of the concept of collocation formulated in this study (cf. §2.2.4.) and specifies the following:

> A specialized collocation is a type of multiword expression composed of at least a term that serves as the node. Its collocates can be nouns, verbs, adjectives or adverbs in a direct syntactic relation with the node. These constituents make a lexical combination that can be unpredictable and semicompositional and have an internal and statistical tendency of preference. This definition implies that the constituents of a specialized collocation are not necessarily adjacent to each other. (pp. 125-126)

In short, a specialized collocation includes a term, that is, "a unit with special reference within a specialized subject field" (L'Homme & Bertrand, 2000, p. 497). This term plays a prominent position within the collocation (it is the base) because it is freely chosen, and it holds a syntactic relation with a lexical word (a noun, a verb, an adjective or an adverb), which is chosen on the basis of the base to express a specific meaning (it is a collocate). They are usually restricted and unpredictable, and their meaning is half-way between compositionality and non-compositionality. To put it differently, the meaning of a collocation is easier to deduce than the meaning of an idiomatic expression (which is non-compositional), although less directly than that of a free combination (which is fully compositional), since one of its elements, the base, keeps its literal meaning whereas the other element, the collocate, is partially obscure and more idiomatic. In addition, the

elements of a specialized collocation tend to co-occur together, which can be proved by using corpora and corpus management software, although they do not necessarily co-occur in immediate text, that is, intervening words may be found. Some examples of verb collocations in the field of computer science are: *[to] run a programme, [to] press a key, [to] execute a command* (Orliac, 2008, p. 379).

Another name given to specialized collocations has been *specialized lexical combinations*. L'Homme and Bertrand (2000, p. 498) consider that they are two different concepts and cannot be described in the same way for one reason: "specialized lexical combinations cannot truly be described as prototypical collocations, since many lexemes defined as co-occurrents can combine with groups of semantically-related terms." Nevertheless, this *semantic preference*, understood as the preference of one unit for a set of co-occurrents belonging to the same lexical field (e.g., *large* combining with words referring to quantities and sizes, like *number, scale, part, amounts*), was suggested by Sinclair (1996/2004) in the *extended unit of meaning*. He discovered that lexical meaning is not so much a question of meaning isolated in the lexeme, but rather in an extended unit of meaning which encompasses four relations: (1) collocation, (2) colligation, (3) semantic preference, and (4) semantic prosody. As a result, a word can be defined in terms of: (1) its co-occurrence with other word forms, (2) its grammatical choices, (3) the semantic features of the lexical set with which it tends to co-occur, and (4) the positive or negative connotations it can carry, respectively. It is for this reason that the concept of specialized lexical combination can be considered similar to that of specialized collocation.

According to Patiño García (2014), several authors have placed specialized collocations within the so-called "LSP (Language for Special[/Specific] Purposes) Phraseology" (p. 125), which is at the crossroads between Terminology and Phraseology. Nevertheless, the author recognizes that little research has been done on the field of LSP Phraseology because it requires in-depth knowledge of the science or profession in question. In other words, a researcher cannot analyze the language of an LSP text without considering the subject matter treated in the text. Despite that, we can still highlight some work dealing with word combinations in SLs. Many of these studies agree that a substantial amount of words found in these fields are taken from the GL, but they display a terminological sense when they are used in specific contexts. For example, research has been done on the language of medicine (Bieliaieva et al., 2017), computer science

(Curado Fuentes, 2001), law (Biel et al., 2018), the environment (Alonso Campos & Torner Castells, 2010), economics (Tognini-Bonelli, 2002) and scientific writing (Jiménez-Navarro, 2019).

However, this section is focused on research into specialized collocations in the language of tourism, given that it is the SL addressed in this work.[68] We agree with Fuster-Márquez (2014, p. 85) that this specialized domain is characterized by a series of PUs, in fact, the author demonstrates that "hotel websites share highly distinctive phraseological preferences which are not shared by other registers or text types" (p. 105). His work is on lexical bundles, which are 4-word sequences (cf. §2.2.3.) but are also considered a type of (extended) collocation, and it is based on a specialized corpus of about 240,000 words. He discovers that the flexible elements of these sequences are content words (nouns, adjectives, adverbs or lexical verbs) and fill the slot in frames such as *will be [required, charged] to, we are [happy, delighted] to, in the [heart, centre] of.*

Another author who is concerned with this topic is Corbacho Sánchez (2005), who performs a contrastive study of specialized collocations containing the bases *Tour(ismus)* [tourism], *Reise* [trip] and *Fahrt* [trip] in German and their translations into Spanish. For so doing, he selects 87 structures from monolingual and bilingual dictionaries and terminological resources, regardless of their frequency. He concludes that metaphors play a significant role in specialized phraseology, given that the collocations found for *Tour(ismus)* do not hold a relationship with tourism, and recognizes that the number of collocations analyzed is not high enough to draw conclusions about tourism and trips (p. 73). However, a more recent study on verb + noun collocations (Corbacho Sánchez, 2017) proves that the language of German tourism is characterized by an array of specific collocations which give it the status of SL, an aspect that is illuminated by the existence of exact equivalents in Spanish (p. 103). Some remarkable examples are *ein Angebot ablehnen = rechazar una oferta*, *einen Aufschlag erheben = cargar un suplemento*, *eine Buchung bestätigen = confirmar una reserva*, which also have a direct equivalent in English: *withdraw an offer, charge a supplement* and *confirm a booking*, respectively.

Another relevant study is carried out by Manca (2008), who aims to know the influence of culture and context of situation in phraseology. It centers on specialized

[68] An analysis of its features along with a description of the language of adventure tourism, which is the content of our corpus, is offered in §4.1.

collocations containing qualifying adjectives, which are considered the base for being the elements under analysis (cf. Sinclair, 1991). The word list obtained from a British farmhouse holiday corpus reveals that adjectives can be classified into three semantic fields, namely, *description of rooms, description of surroundings* and *description of food*. The results shown are largely related to the description of rooms and a key point is that adjectives refer to *size* (e.g., *spacious, large*), *equipment* (e.g., *equipped, furnished*) and *beauty* (e.g., *attractive, beautiful*). However, the comparative analysis performed with an Italian comparable agritourism corpus indicates that no adjectives refer to the beauty of the rooms in this language. The author's conclusions are that the texts in the British corpus are content-oriented, that is, they give detailed and explicit descriptions of what a holiday in a farmhouse can offer, whereas the texts in the Italian corpus are form-oriented, that is, what remains unsaid or implicit is what counts more and the idea of a good holiday is left to people's imagination (p. 382).

One more piece of research that must be mentioned is authored by Baynat Monreal (2017). She regards specialized collocations as compositional groups of words where the collocate is contingent on the base. Her work examines simple LUs and MWUs retrieved from a three-million corpus containing web pages for private lodgings and accommodations in three languages: English, Spanish and French. These words are the entries of the *Multilingual Dictionary of Tourism* (COMETVAL Group, 2014) and the author focuses on simple and compound units belonging to the lexical field of hotel management in French. Nevertheless, 50% of these units are nouns, for example, *voucher, demande, annulation*, so her work is mainly on noun collocations, which are categorized into ten different semantic fields, such as documentation (e.g., *fiche technique, coupon promotionnel*), booking management (e.g., *mentions légales, politique d'annulation*) or invoicing (e.g., *modes de paiement, conditions tarifaires*). A deeper analysis on the base *forfait* shows that its collocates add subtle nuances of meaning and make it semantically richer (Baynat Monreal, 2017, p. 71).

The last piece of research covered in this section is conducted by Piccioni and Pontrandolfo (2019). Their first aim is to identify and classify metaphorical collocations related to space in a bilingual (Spanish-Italian) corpus which contains all types of text used in tourism, such as guides, web pages, blogs, contracts. Moreover, their ultimate objective is to know how metaphorical phraseology can help tourists to create an image of their tourist destination (p. 138). Their results reveal that metaphors based on body

parts are behind some collocations, for example, *pulmón verde, arteria urbana, boca de la ría*. However, one major aspect is linked to fictive motion verbs. To explain, metaphors are present in motion verbs whose subject is represented as if it were moving but actually is not, such as *la llanura manchega se alza, Un excelente y amplio golfo se extiende*, helping the tourist to perceive the movement (p. 149).[69]

To conclude, it must be highlighted that the analyses described above, except for Piccioni and Pontrandolfo's (2019), focus on collocations containing nouns, probably because nouns are considered the most productive category in terminology (Vargas Sierra, 2012b, p. 71). For this reason, the current investigation aims to shed more light on the role of verb collocations in the SL of tourism, more specifically, it concentrates on collocations of motion verbs in the language of adventure tourism. The pivotal role of verbs in SLs is emphasized by L'Homme (1998, 2003, 2012), López Rodríguez (2007), Lorente Casafont (2007), Buendía Castro (2012) and Casademont (2014), despite having been usually relegated in comparison to studies on nouns (L'Homme, 1998, 2003, 2012). An adverse consequence of this fact is that there are scarce terminological resources which incorporate verb collocations. Furthermore, collocations are not normally encoded under the verb but under the noun (L'Homme & Leroyer, 2009, p. 258), which also happens in lexicographical resources (Buendía Castro & Faber Benítez, 2014; cf. §3.2.). For this reason, we hope this work also contributes to this field, which is the topic explored in the following chapter.

[69] These findings are closely connected with the results obtained in this dissertation, where fictive motion events are of great relevance (cf. Chapter 5).

CHAPTER 3. COLLOCATIONS IN GENERAL AND SPECIALIZED RESOURCES

Phraseology presents a major challenge for lexicographical practice. [...] Lexicographers confront phraseology word by word, sense by sense, and make decisions about the status and significance of recurrent patterns [...].

(Moon, 2008, p. 313)

Chapter 3 focuses on the representation of collocations in general and specialized resources. As our ultimate objective is to provide a list of items which should be considered for encoding collocations in an SL dictionary, this chapter is organized as follows. To start with, Section 3.1. will briefly explain the contents that make a dictionary complete, either general or specialized. Then, the focus will be on how phraseology has been represented in these types of dictionaries so far. Thus, Section 3.2. will deal with GL dictionaries and will be divided into two sections: the first one will provide a description of collocational information in English general dictionaries (§3.2.1.) and the second one will address how this information is collected in English collocation dictionaries (§3.2.2.). After that, Section 3.3. is motivated by the aforementioned objective, therefore, it will fully describe the implementation of collocations in specialized resources. Finally, Section 3.4. will assess the conclusions reached in the previous sections and will present a model for an ideal specialized dictionary, in general, and an adequate representation of collocations, in particular.

3.1. The contents of lexicographical and terminological resources

According to Fuertes-Olivera and Tarp (2014), lexicographical works (either covering the GL or an SL) are, or should be, utility tools "designed with the purpose of satisfying specific types of human needs" (p. 45). It implies that lexicographical tools are social creations representing a relation between at least two people, the user and the

lexicographer/terminographer. Accordingly, when the latter aims to compile a dictionary,[70] they should bear two aspects in mind (Bergenholtz & Tarp, 1995, p. 20): first, the type of user that will be using the dictionary, and second, the situation/s in which he/she will be using the product.

In terms of the first aspect, dictionaries can be, for example, general-purpose (aimed at native speakers of a language), scholarly (aimed at students of a subject area), for children (aimed at children acquiring their first language, L_1) or for learners (aimed at students of a second language, L_2) (Jackson & Amvela, 2007, p. 188). With respect to the second aspect, a large range of situations can arise. For instance, for native-language and foreign-language production and reception, a monolingual (covering one language) dictionary can be used, although for foreign-language production and reception a bilingual (covering two languages, e.g., the user's L_1 and L_2) dictionary can be used too. Bilingual as well as multilingual (covering more than two languages) dictionaries can also be present in translation studies. On the other hand, users who are concerned with word combinations can consult combinatory dictionaries, and those who are interested in SLs can use specialized dictionaries.[71] A key point is that one dictionary may address different situations at the same time, for example, it can be multilingual, combinatory and specialized all in one.

Nonetheless, dictionaries aimed at the same target user or the same target situation of use may include distinct items of information, given that deciding which sort of information must be implemented in a dictionary is in the hands of the lexicographer/terminographer, and it goes without saying that different people can make different decisions. For this reason, this section will try to shed some light on this aspect and will show a comprehensive list of the contents that lexicographical and terminological resources, more specifically, dictionaries, should include from the perspective of influential authors. It will be done in the hope of providing future

[70] This work is particularly focused on dictionaries (as opposed to glossaries or vocabularies), since the terminological resource we address, *DicoAdventure*, is of this type. To our mind, a dictionary is a "product [that] aims to record the lexicon of a language [either general or specialized] in order to provide the user with an instrument with which he can quickly find the information he needs to produce and understand [the] language" (Van Sterkenburg, 2003, p. 8).

[71] The last two situations, which involve word combinations and SLs, are central to the work developed in this research, since we aim to incorporate collocations into a specialized dictionary which covers the SL of adventure tourism.

lexicographers/terminographers with some help in the fields of both GL and SL dictionary compilation, but we recognize that other suggestions may be equally valid.

On the one hand, Hudson (1988) suggests a list of eight items "for inclusion in an all-inclusive lexicon" (p. 310). He recognizes that, in all probability, all linguists will not agree with such a list, but he justifies his decision by saying that all the information proposed constitutes facts about words. This list (pp. 310-312) is aimed at general-purpose monolingual dictionaries:[72]

1) *Phonology*. Dictionaries tend to give a transcription of the pronunciation of words and include the accentual pattern of those items which are not monosyllabic. This information normally appears very close to the headword and normally use the International Phonetic Alphabet (IPA).[73] Lexicographers may be faced with the problem of whose pronunciation must be represented in the dictionary but *Received Pronunciation* (RP) (Strong & Dyer, 2009) is the version generally accepted. In some cases, regional variations are shown;

2) *Morphology*. Two main aspects of the morphology of words must be considered, that is, inflections and derivations. On the one hand, inflections are usually regular and do not need to be included in a dictionary, except for those which are irregular, for example, *spy – spies, travel – traveled.* On the other hand, Jackson and Amvela (2007, p. 195) state that derivations should be specified when their origins are different from the simple root, such as *mind – mental, lung – pulmonary*;

3) *Syntax*. Dictionaries can incorporate some basic items of syntactic information, being the most common the word class (e.g., noun, verb, adjective). In addition, verbs can be classified as transitive or intransitive and they can give information about their grammatical patterns, for instance, a *that*-clause or an infinitive clause. In the case of nouns, apart from saying whether they are singular or plural, specifying whether they are countable or uncountable may help users;

[72] According to the author, "The structure of the list is not meant to have much significance" (p. 311).
[73] http://www.internationalphoneticalphabet.org/ (Last accessed: 25/10/2020).

4) *Semantics*. The primary function of monolingual dictionaries is to provide words' meanings. The selection of meanings to show is up to the lexicographer, for which he/she can check a corpus and find evidence of examples of a word in use. In addition to definitions, Jackson and Amvela (2007, p. 198) believe that this section must also include semantic relations in the paradigmatic axis, like synonymy and antonymy, as well as syntagmatic associations, that is, collocations of a word. Nevertheless, these authors believe that, while the former have usually been included systematically, collocations have been incorporated accidentally, despite the fact that they "should be included" (Benson et al., 2009, p. xix);

5) *Context*. Contextual information is systematically recorded in dictionaries as well as example sentences that include the word in question. Its function is to let users know whether words are restricted to types of text or discourse, or social contexts and occasions. Besides, it includes information on the level of formality of a word, for example, formal, informal, slang, colloquial. In some cases, the lexicographer provides information on the subject domain which a word belongs to, such as *intravascular* – Anatomy, *software* – Computing, *peritoneoscopy* – Medicine;

6) *Spelling*. As it is stated by Jackson (1988, p. 194), after meaning spelling is the second reason for consulting a dictionary. It makes much more sense in languages where the pronunciation of words does not normally correspond to the way they are spelled, such as English. This information is extremely useful when the addition of a suffix to a root causes a spelling change in the root, for example, *lorry – lorries, have – having*. Additionally, when words have alternative spellings, they all should be included, for instance, *centre* versus *center*, *favour* versus *favor*, or even when they exist in the same variety of the language (e.g., *chamomile* vs. *camomile*);

7) *Etymology*. As it was pointed out in Section 1.2., one of the first greatest lexicographical works was a dictionary compiled in the tenth century aiming to explain the etymology and meaning of words in Ancient, Hellenistic and Byzantine Greek and in Latin. Thus, we can say that one of the motivations that led to the development of dictionaries was the recording of etymologies.

Nevertheless, it must be recognized that dictionaries are not particularly concerned with this aspect nowadays; and

8) *Usage*. This is the last item of information that Hudson suggests. It must not be confused with the *Context* label, since this label refers to snippets of information that are rarely provided in dictionaries (Jackson & Amvela, 2007, p. 202). Specifically, the author mentions: (1) the frequency of use of a word (the development of corpus studies has facilitated the inclusion of this item), (2) the age of acquisition of a word, (3) particular occasions of use of a word (e.g., in memorable sayings), (4) clichés containing the word, and (5) taboos on words.

On the other hand, SL dictionary compilation is not so far from GL dictionary compilation. In view of their successful publications on the field of lexicography, we will follow the authors of the Function Theory of Lexicography (cf. §1.4.) to explain the items that should be included in a specialized dictionary. More specifically, this list is based on the *Manual of Specialized Lexicography* by Bergenholtz and Tarp (1995), given that this book is intended to provide a foundation for practical specialized lexicography (p. 8) and, to our mind, it performs a thorough analysis on the matter. In their opinion, a specialized dictionary should be corpus-based (p. 95) and include two types of information, linguistic and encyclopedic:

1) Linguistic information (pp. 111-142): According to the authors, monolingual specialized dictionaries do not normally include linguistic information, such as the grammatical use of terms or their combination possibilities, for example, collocations (p. 111).[74] A consequence of this is that terminographers refer their users to the simultaneous use of a GL dictionary. In total, they identify six items of information in this group:

    i. *Grammar*. Specialized dictionaries should emulate the procedure performed by GL dictionaries and provide information on word class, inflection, syntax and word formation. In the case of verbs,

[74] Bergenholtz & Tarp's (1995, p. 111) opinion as to the lack of explicit information on collocations in specialized lexicography chimes with Jackson & Amvela's (2007, p. 198) idea about general-purpose monolingual dictionaries.

which are the type of terms collected in *DicoAdventure*,[75] irregularities should be shown in the entries when they are common in the language of the dictionary. The model proposed by Bergenholtz and Tarp implies the preparation of a "differential grammar" (p. 116), that is, one describing the special grammatical features of the SL of the dictionary when compared to the GL;

ii. *Word combinations.* In the authors' opinion, specialized dictionaries are characterized by a lack of information on collocations, thus, they suggest a model that includes a selection of these PUs by applying two criteria: frequency and relevance,[76] that is, collocations belonging in the intersection between the SL and the GL may be unnecessary. Regarding the number of collocations to be provided, it depends on the dictionary functions and the user competence; generally, the need for collocations will be greatest in translation. They summarize it as represented in Figure 5 (p. 123):

Figure 5

The Need for Collocations in Specialized Dictionaries on the Basis of User Types and Dictionary Functions (Bergenholtz & Tarp, 1995)

	USER TYPE 1 **low** level of **both** encycl. and foreign-language competence	USER TYPE 2 **low** level of encycl. and **high** level of foreign-language competence	USER TYPE 3 **high** level of encycl. and **low** level of foreign-language competence	USER TYPE 4 **high** level of **both** encycl. and foreign-language competence
Reception of L_1	few	few	—	—
Production in L_1	(many)	(many)	fewer	fewer
Translation L_1-L_2	(many)	many	many	fewer / many
Translation L_2-L_1	many	many	many	fewer / many
Reception of L_2	few	few	few	— / few
Production in L_2	(many)	(many)	many	fewer / many

[75] *DicoAdventure* is the specialized resource addressed in this work and will be fully described in Section 4.2.

[76] As we will see in Section 4.4.2., these two criteria were considered for the extraction of the collocations to be implemented in *DicoAdventure*.

iii. *Synonyms and antonyms.* Although it is widely acknowledged that total synonymy in GL does not exist, it can be established between two terms in normatively classified SLs. However, grammatical or collocational differences may arise, so it can be admitted that two synonymous terms may not be immediately interchangeable in a given context. The model proposed includes synonyms and antonyms in the encyclopedic notes (explicitly or implicitly), such as cross-references or as double lemmata, among others;

iv. *Linguistic labeling.* This item provides information on regional or dialectical expressions, their frequency of use and restrictions on use (e.g., jargon, informal, colloquial), among others. Again, the number of linguistic labels in specialized dictionaries "is extremely modest" (p. 131);

v. *Pronunciation.* As most specialized dictionaries are targeted at written language, information on pronunciation, either in the form of phonetic transcription or by indication of stress pattern, is obviated. Nevertheless, Bergenholtz and Tarp (p. 135) suggest a model of an SL dictionary which incorporates this item for two reasons: (1) the pronunciation of a term may have a mnemonic value, and (2) SLs can also be used in oral communication; and

vi. *Examples.* While the inclusion of examples in bilingual specialized dictionaries is strongly supported, the authors contend that their inclusion in monolingual specialized dictionaries is also suitable in terms of function, given that "the dictionary will be a more valuable aid in foreign-language production and translation" (p. 142). When the examples contain collocations, they should be classified as such,[77] rather than providing them as implicit information in the examples selected;

2) Encyclopedic information (pp. 143-166): Bergenholtz and Tarp (p. 111) also use the term *encyclopedia* to refer to monolingual specialized dictionaries,

[77] Section 4.2. will explain the section of *DicoAdventure* into which collocations will be incorporated.

given that encyclopedic information should be provided in these products. To their mind, it can be of great benefit in several situations, such as SL text reception, native-language and foreign-language production, translation and as an introduction to the dictionary subject matter. It can be supplied in three forms:

i. *Encyclopedic notes.* The model suggests their inclusion in individual dictionary entries and they should accompany all SL terms (words that the SL has in common with the GL need not be addressed). As aforementioned, in the cases of synonymy and antonymy, homonymy and hyponymy, cross-references can be used so as to save space, and double lemmata are also accepted. Some dictionaries opt for longer encyclopedic notes rather than brief, and they can use full or incomplete sentences;

ii. *Encyclopedic labeling*. Encyclopedic labels make brief comments on lemmata or equivalents, since their primary purpose is to help users to find a particular lemma or to select the correct equivalent. Also, information on the subject field associated with the term can be made available, specially where a dictionary covers more than one subject field or sub-field. A particular type of encyclopedic label is the standard, for example, DIN,[78] DS,[79] ISO,[80] as they inform the user that the term has been standardized;

iii. *Encyclopedic section.* This section aims to introduce the subject field covered by the dictionary and can be of tremendous help to the layman user. The model highlights three main types, which can be combined: (1) a brief introduction, (2) a more systematic exposition, and (3) a comparative description. Type (2) is "the most appropriate for satisfying all dictionary user types and functions" (pp. 155-156); and

[78] Deutsches Institut für Normung (German Institute for Standardization). [https://www.din.de/en/about-standards/din-standards] (Last accessed: 25/10/2020).
[79] Danish Standards. [https://www.ds.dk/en] (Last accessed: 25/10/2020).
[80] International Organization for Standardization. [https://www.iso.org/home.html] (Last accessed: 25/10/2020).

iv. *Illustrations*. They are common in some specialized dictionaries which use them as a supplement or to substitute encyclopedic information. Furthermore, they have an additional aesthetic function. The model does not issue explicit guidelines on illustrations, since they depend on the subject field and the nature of the terminology, although it mentions different options for the placing of illustrations, such as the individual entries (recommended), in particular pages in the middle part of the dictionary or at the end.

On balance, we think that the ideas promoted by Hudson (1988) for compiling general-purpose monolingual dictionaries flashed across Bergenholtz and Tarp's (1995) mind for the design of specialized dictionaries for two reasons: first, several fundamental aspects are discussed by both, and second, the latter firmly believe that SL dictionaries may learn from GL dictionaries (p. 112). Table 15 compares and summarizes the items of information that these two proposals suggest being included in GL dictionaries, on the one hand, and specialized dictionaries, on the other, to see to what extent they agree:

Table 15

Items of Information That Should Be Included in General Language Dictionaries and Specialized Language Dictionaries

	GL dictionaries (Hudson, 1988)	**SL dictionaries (Bergenholtz & Tarp, 1995)**
Pronunciation and accentual patterns	●	●
Word formation	●	●
Syntactic information	●	●
Definition	●	●
Word combinations	●	●
Contextual information	●	●
Spelling changes	●	
Etymology	●	
Frequency of use	●	●
Age of acquisition	●	
Examples		●
Illustrations		●

As it can be observed, these proposals share clear similarities. For instance, the points in common are related to information on phonological, morphological, syntactic, semantic and contextual aspects of words and terms. In addition, they agree that information on syntagmatic associations should be provided (e.g., collocations) as well as on the frequency of use of words/terms. However, some dissimilarities are related to spelling changes and etymology. Regarding the latter, Bergenholtz and Tarp (1995, p. 111) indicate that etymologies are sometimes included, but they do not mention that it is an item of their model. Another aspect that Hudson mentions but is not relevant to specialized lexicography is the age of acquisition of a word, for SLs, unlike the GL, are not naturally learned. Finally, the terminographers suggest the inclusion of examples and illustrations, which do not appear in Hudson's list but are usually common in GL dictionaries, as we will see in the following section.

It goes without saying that a clear view shared by the aforementioned authors which directly concerns our work is related to the presence of collocations in either GL or SL dictionaries. While both agree that they should be included in all resources, they admit that only very sporadically do lexicographers/terminographers supply collocations of words and terms. In this context, Sections 3.2. and 3.3. will describe how phraseology is represented in both lexicographical and terminological resources, as we expect that "The treatment of phraseological units differs significantly across dictionaries both in terms of coverage and access" (Paquot, 2015, p. 468). The tools chosen to fulfil this aim are of three types: (1) English general dictionaries, (2) English collocation dictionaries, and (3) terminological resources. This analysis will allow us to know how collocations have been covered in general and specialized resources so far and will help to suggest a model for their implementation in the specialized dictionary *DicoAdventure*.

3.2. Collocations in lexicographical resources

According to Benson et al. (1986a), "The critical problem for the lexicographer has been, heretofore, the treatment of collocations" (p. 256). Thus, this section deals with the representation of PUs, in general, and of collocations, in particular, in two different sorts of resources. First, the focus is on five English general dictionaries which are available online, as "many specialists predict the disappearance of paper dictionaries in the near

future" (Granger, 2012, p. 2). Second, five English collocation dictionaries will be examined.[81] The procedure followed is the selection of one word to investigate how the entry in the different dictionaries under analysis encodes collocations.

Despite the fact that this dissertation examines verb collocations, the entries for the noun *bottle* are the examples set in our analysis of the English general dictionaries. The reason for choosing a noun instead of a verb is that most resources do not provide phraseological information under verb entries, but they rather opt for including them under the element with which they combine, for example, the noun. On the other hand, the reason for choosing *bottle* is that we intuitively feel that it can display a wide range of combinations with adjectives, verbs and other nouns.

Nevertheless, our analysis of the English collocation dictionaries will focus on the entries for the verb *improve*. The reason for choosing a verb is that we believe that dictionaries which are focused on word combinations will be more prone to encode them under verb entries than general dictionaries, which is deeply significant for the objective set in this work. On the other hand, the reason for choosing *improve* is that its grammatical category can only be a verb, and we foresee that it can be used in several verb patterns and collocations.

3.2.1. Collocations in English general dictionaries

Section 3.1. has explained what type of information should be included in English monolingual dictionaries. First of all, they are mainly used for obtaining information about the meaning and spelling of words. In addition, they usually include information on the pronunciation of words, their morphemic relations, their syntactic structures, their context of use and their origin. In a few rare cases, they also contain other items of information related to the usage of words. However, dictionaries should also show collocational relationships between words as part of their semantic information.

In the next pages, we will examine how the following English general dictionaries, nicknamed the "Big Five," that is, "the five major electronic monolingual learners'

[81] Romero Aguilera (2015) performs an analysis of Spanish collocation dictionaries. Additionally, Torner Castells & Bernal (2017) examine the information related to the collocations of four nouns in three Spanish general dictionaries and two Spanish collocation dictionaries.

dictionaries of English" (Paquot, 2015, p. 470), represent collocations:[82] (1) *Collins English Dictionary* (CED),[83] (2) *Longman Dictionary of Contemporary English* (LDOCE),[84] (3) *Oxford Advanced Learner's Dictionary* (OALD),[85] (4) *Cambridge Dictionary of English* (CDE),[86] and (5) *Macmillan English Dictionary* (MED).[87] A brief introduction to each dictionary will be also provided.

Collins English Dictionary (CED)

Section 1.5.2. made a special mention of the work carried out by Sinclair in the field of corpus linguistics. This work came to fruition in 1987 (cf. Sinclair, 1987) with the release of the *Cobuild Dictionary*, which was based on a corpus of about eight million words of contemporary English texts called the *Bank of English*. This corpus is a subset of 650 million words of the *Collins Corpus*,[88] which contains over 4.5 billion words and includes written and spoken material from different sources, such as websites, newspapers, books, radio programmes or everyday conversations. It is updated every month to help lexicographers include the latest information in the dictionary and, nowadays,[89] the CED still uses the corpus to show how people really use words. An advantage of relying on corpora to compile dictionaries is that it allows the inclusion of real information about words' meaning, word combinations, frequency of words, example sentences, among others. Without a doubt, the CED was the forerunner of a new generation of corpus-based dictionaries.

Figure 6 shows the entry for *bottle* in the CED,[90] which provides collocational information as follows:

i. Corpus sentences are given in italics under the different definitions, from which some collocations can be inferred, such as adjective + noun (e.g., *empty*

[82] Other studies which compare different types of dictionaries are Bogaards (1996), Mittmann (1999), Moon (2008), Walker (2009), Buendía Castro & Faber Benítez (2014) and Paquot (2015).
[83] https://www.collinsdictionary.com/ (Last accessed: 25/10/2020).
[84] https://www.ldoceonline.com/ (Last accessed: 25/10/2020).
[85] https://www.oxfordlearnersdictionaries.com/ (Last accessed: 25/10/2020).
[86] https://dictionary.cambridge.org/ (Last accessed: 25/10/2020).
[87] https://www.macmillandictionary.com/ (Last accessed: 25/10/2020).
[88] https://collins.co.uk/pages/elt-cobuild-reference-the-collins-corpus (Last accessed: 25/10/2020).
[89] The latest 13th edition of the CED was launched in 2018. [https://www.collinsdictionary.com/about] (Last accessed: 25/10/2020).
[90] https://www.collinsdictionary.com/dictionary/english/bottle (Last accessed: 25/10/2020).

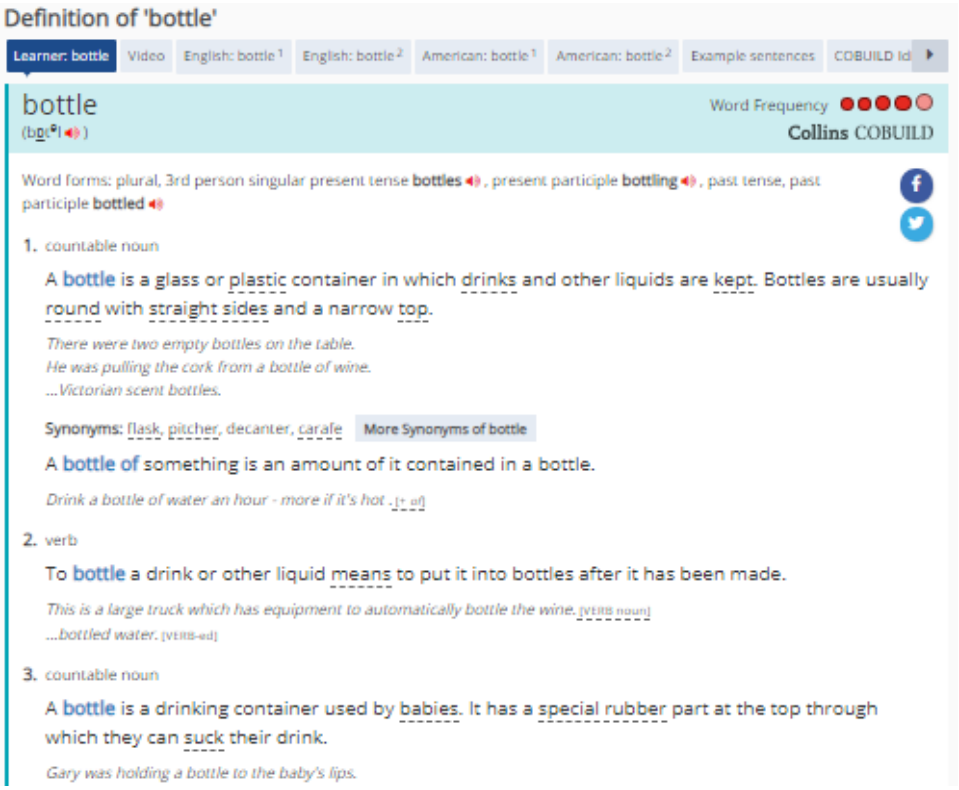
bottles, scent bottles), noun$_1$ + *of* + noun$_2$ (e.g., *a bottle of wine, a bottle of water*), verb + noun (e.g., *holding a bottle, have the bottle*) collocations;

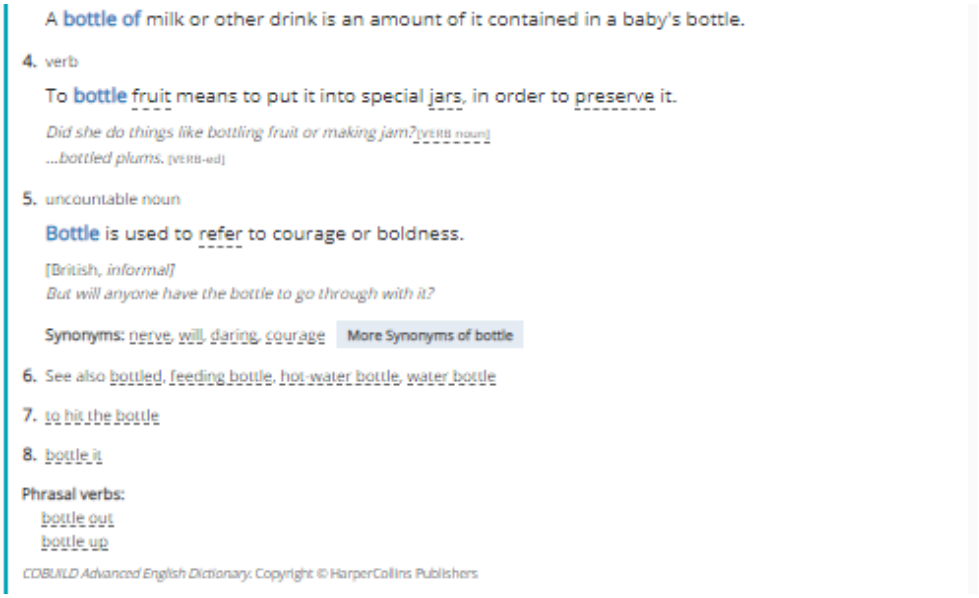
ii. Some other phraseological information is readily available, as in 6.: *feeding bottle, hot-water bottle, water bottle*, and 7.: *to hit the bottle*, which take the user to their definitions by clicking on them; and

iii. Phrasal verbs are shown separately: *bottle out* and *bottle up*.

Figure 6

Entry for bottle *in the Collins English Dictionary*





A **bottle of** milk or other drink is an amount of it contained in a baby's bottle.

4. verb

To **bottle** fruit means to put it into special jars, in order to preserve it.

Did she do things like bottling fruit or making jam? [VERB noun]

...bottled plums. [VERB-ed]

5. uncountable noun

Bottle is used to refer to courage or boldness.

[British, informal]

But will anyone have the bottle to go through with it?

Synonyms: nerve, will, daring, courage More Synonyms of bottle

6. See also bottled, feeding bottle, hot-water bottle, water bottle

7. to hit the bottle

8. bottle it

Phrasal verbs:

bottle out

bottle up

COBUILD Advanced English Dictionary. Copyright © HarperCollins Publishers

This description shows that the treatment of phraseological information is not special in the CED, given that, except for the PUs included in 6. and 7., the rest of the collocations are implicit in the corpus sentences. Regarding 6. and 7., we believe these two sections are separated because, while the PUs in 6. are collocations, *to hit the bottle* is an idiom, since its meaning is non-compositional (= to drink a lot of alcohol). Finally, it is important to realize that phrasal verbs are explicit and indicated as such.

Longman Dictionary of Contemporary English (LDOCE)

The LDOCE is also based on updated corpora, more specifically, on the *Longman Corpus Network*.[91] It contains 330 million words extracted from an extensive range of authentic sources, both written and spoken, so the example sentences included in the LDOCE are a representation of real English. The written material is taken from books, magazines and newspapers as well as students' essays and exam scripts. The spoken material is taken from a collection of recordings of over 3,000 British and American English speakers. It also makes use of the *Longman/Lancaster Corpus*[92] and the *BNC Spoken Corpus*.[93] Relying on corpora allows the LDOCE: (1) to know which words are more popular in English language, (2) to use 2,000 common words in the definitions to make

[91] http://www.pearsonlongman.com/dictionaries/corpus/ (Last accessed: 25/10/2020).
[92] http://www.pearsonlongman.com/dictionaries/corpus/lancaster.html (Last accessed: 25/10/2020).
[93] http://www.pearsonlongman.com/dictionaries/corpus/spoken-BNC.html (Last accessed: 25/10/2020).

understanding easy, (3) to highlight the 9,000 most relevant words with three red circles, and (4) to present the most common meanings of a word first.

Figure 7 shows the entry for *bottle* in the LDOCE,[94] which provides collocational information as follows:

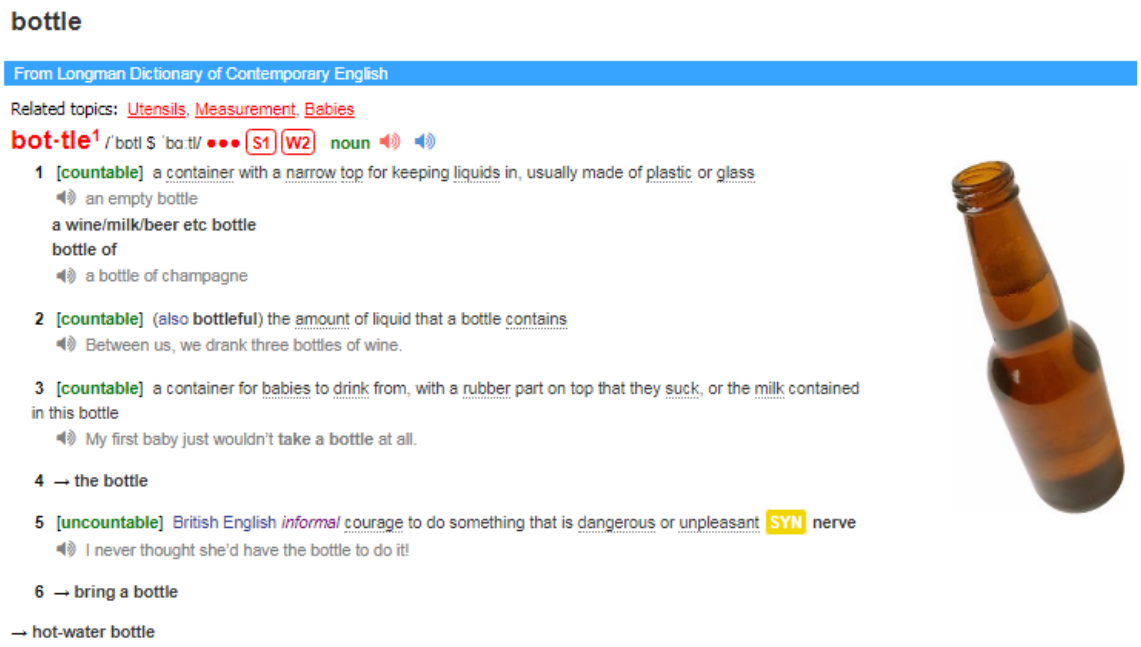
i. Collocations are specified under the first definition: *a wine/milk/beer bottle, bottle of*, and the corpus sentence in 3 also shows a collocation highlighted in bold: *take a bottle*;

ii. Some other PUs are available in 6: *bring a bottle* and *hot-water bottle*, which take the user to their definitions by clicking on them; and

iii. More example sentences from the corpus are provided at the bottom of the entry, which allow users to know other collocations, such as adjective + noun (e.g., *oil bottles, a whole bottle*) and $noun_1$ + *of* + $noun_2$ (e.g., *a bottle of schnapps, a bottle of Trazadone*) collocations.

Figure 7

Entry for bottle *in the Longman Dictionary of Contemporary English*



[94] https://www.ldoceonline.com/dictionary/bottle (Last accessed: 25/10/2020). As a novelty, this entry incorporates an illustration of a bottle, as suggested by Bergenholtz & Tarp (1995) in their model for specialized dictionaries (cf. §3.1.).

Examples from the Corpus

bottle

- Egon managed to drink half a **bottle** of schnapps that night.
- A **bottle** of Trazadone, an anti-depressant medicine, was in the cabin.
- Do you want me to give Kayla her **bottle**?
- On the table in the milk **bottle** was the daffodil he'd picked.
- It was time for a bit of **bottle**, not for knocking knees.
- They found several oil **bottles** in his car of the same brand as one found at the starting point of the fire.
- We took our **bottles** of beer.
- We ended up with about a dozen assorted gin, whisky and stout **bottles**.
- The genie is out of the **bottle**.
- I only want one glass, not a whole **bottle**.

a wine/milk/beer etc bottle

- I had my cape over my shoulder and **a milk bottle**.
- The incident leading to the court case had been sparked by **a beer bottle** being thrown at him the previous night.
- The lilacs would stand in **a milk bottle**.
- I hear a scraping of **a beer bottle** base across a wooden surface.
- The water trough was empty except for a layer of scum, **a beer bottle** and some sweet wrappers.
- Doyle was flung back across the table, **a milk bottle** exploding in the bag he held across his chest.
- I stopped in my tracks, **a beer bottle** in each hand.

take a bottle

- Lock up here, but **take a bottle** of good brandy with you.
- He had **taken a bottle** of cough mixture with him but it hadn't helped much.
- When he came home, he **took a bottle** of whisky from a cupboard and began to drink.
- She **took a bottle** of oil from the side of the bed and leaned over him.
- Dann looked at Maxim's glass, then **took a bottle** of gin from a desk drawer and refilled his own.
- In April last year a worker was told to **take bottles** to a handling bay for disposal.

This coverage of collocations and that of the CED have very similar features. For instance, they make some collocations explicit, but there are others that must be inferred from the corpus sentences. Additionally, they take the user to other entries in the dictionary by clicking on the phrases. On the other hand, the LDOCE adds some different collocations, for example, *beer bottle*, *wine bottle*, *take a bottle*, and the phrase *bring a bottle*, which can have a literal and a figurative reading.[95] However, it does not include any phrasal verbs. Regarding the use of corpora, it can be said that the LDOCE exploits the corpus more than the CED, given that, apart from making the aforementioned uses of it, it extracts more example sentences to show the headword in context.

Oxford Advanced Learner's Dictionary (OALD)

Again we are dealing with a dictionary that relies on corpus data. The OALD is based on the BNC,[96] which is considered a well-balanced corpus containing over 100 million words of modern British English. This feature means that this dictionary will be more familiar to those people who use this variety of English (the American variety is found in

[95] As it was shown in the previous chapter of this dissertation, MWUs which can be interpreted literally and figuratively are contemplated by some authors, for example, Vinogradov (1947) and Cowie (1981).
[96] https://www.oxfordlearnersdictionaries.com/bnc.html (Last accessed: 25/10/2020).

the *Oxford Advanced American Dictionary*, OAAD). The BNC covers both written and spoken material, although the former is much more abundant (90%) than the latter (10%). The lexicographers of the OALD use this corpus to confirm their intuitions about words' meaning, grammatical behaviour and phraseology, and to learn things about words that they did not know before. They also use sentences from the corpus to exemplify the meanings of the words as they represent real English, the one that appears in books, journals, brochures, lectures, job interviews, and so forth.

Figure 8 shows the entry for *bottle* in the OALD,[97] which provides collocational information as follows:

i. PUs are highlighted in bold under the definitions 1 and 3: *a wine/beer/milk/water bottle, message in a bottle, hit the bottle*; the specific meaning of the latter is provided given its non-compositionality ("= started drinking heavily");

ii. Other collocations are implicit in the sentences in 2, for example, *a whole bottle, a bottle of wine/beer*;

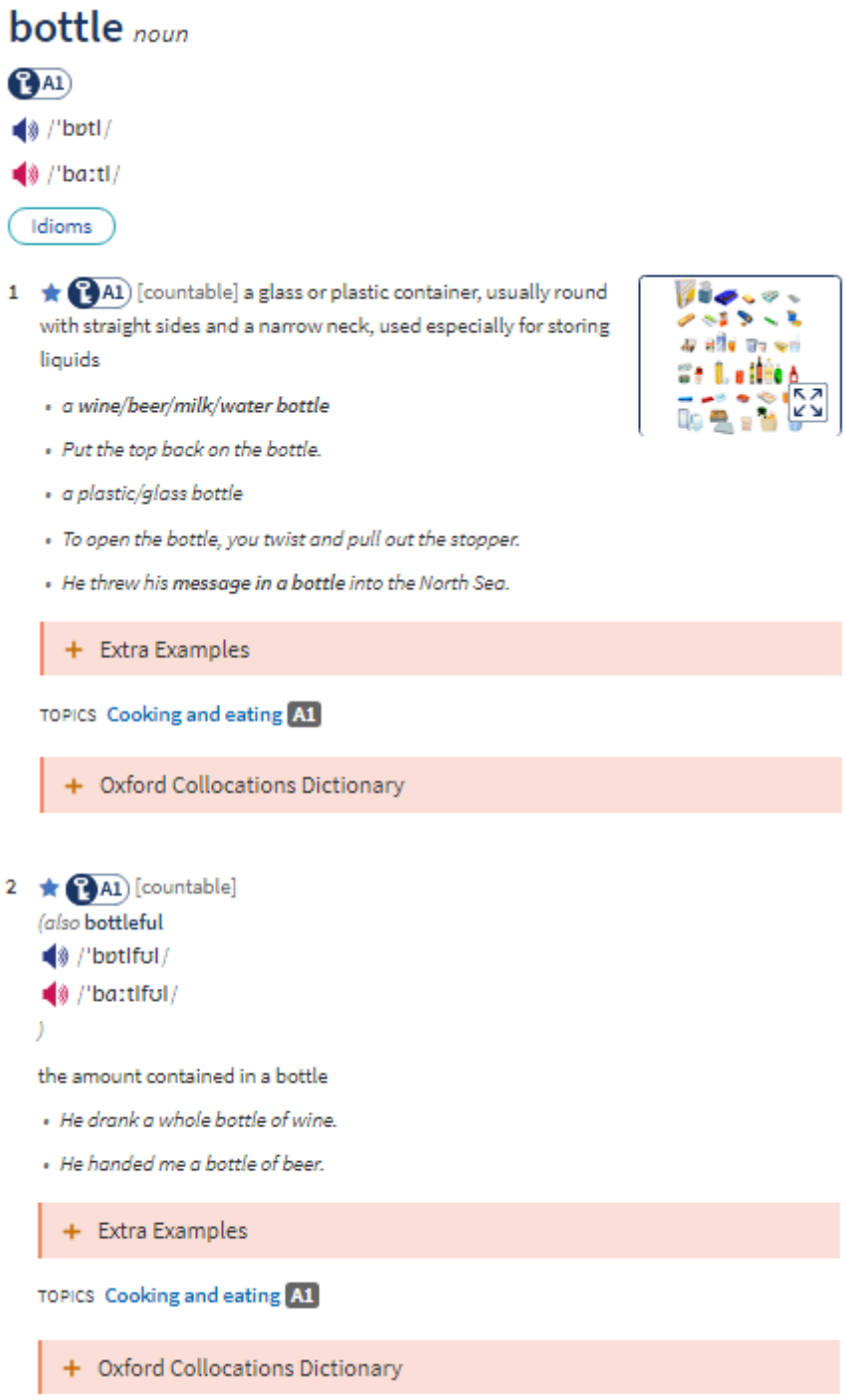
iii. Definitions 1, 2 and 3 include a small foldout (all the foldouts contain the same information) which displays a summary of the PUs that the word can be part of, retrieved from the *Oxford Collocations Dictionary*, as it can be seen in Figure 9; and

iv. The corpus sentences in definition 5 incorporate two more PUs, namely, *take bottle* and *have the bottle*, which can be regarded as idioms given their semantic opacity. Moreover, one more idiom is explicit in a separate section under 5: *let the genie out of the bottle*.

[97] https://www.oxfordlearnersdictionaries.com/definition/english/bottle_1?q=bottle (Last accessed: 25/10/2020).

Figure 8

Entry for bottle *in the Oxford Advanced Learner's Dictionary*



bottle *noun*

A1

/ˈbɒtl/

/ˈbɑːtl/

Idioms

1 ★ A1 [countable] a glass or plastic container, usually round with straight sides and a narrow neck, used especially for storing liquids

- *a wine/beer/milk/water bottle*
- *Put the top back on the bottle.*
- *a plastic/glass bottle*
- *To open the bottle, you twist and pull out the stopper.*
- *He threw his* ***message in a bottle*** *into the North Sea.*

+ Extra Examples

TOPICS Cooking and eating A1

+ Oxford Collocations Dictionary

2 ★ A1 [countable]
(also **bottleful**
/ˈbɒtlfʊl/
/ˈbɑːtlfʊl/
)

the amount contained in a bottle

- *He drank a whole bottle of wine.*
- *He handed me a bottle of beer.*

+ Extra Examples

TOPICS Cooking and eating A1

+ Oxford Collocations Dictionary

3 ★ **the bottle** [singular] *(informal)* alcoholic drink

- *After his wife died, he really **hit the bottle*** (= started drinking heavily).

+ Oxford Collocations Dictionary

4 ★ [countable, usually singular] a bottle used to give milk to a baby; the milk from such a bottle (used instead of mother's milk)

- *It's time for her bottle.*

TOPICS Life stages B2

5 ★ [uncountable] *(British English, informal)* courage or confidence, for example to do something that is dangerous or unpleasant

SYNONYM nerve

- *It took a lot of bottle to do that.*
- *I didn't think she'd have the bottle to ask him.*

TOPICS Personal qualities C2

+ Word Origin

Idioms

let the genie out of the bottle

★ to do something that has a big effect and after which it is very difficult or impossible to go back to how things were before

- *When guns were invented, the genie was let out of the bottle.*

This idiom can be expressed in various different ways. Other versions include **the genie is out of the bottle** and opposite form **put the genie back in the bottle**. *Many people realized the genie was out of the bottle and that free music and movie downloads were here to stay.* • *The world has failed to put the nuclear genie back in the bottle.*

Figure 9

Summary of the Collocations of bottle *in the* Oxford Collocations Dictionary

The representation of collocations in the OALD reminds those in the CED and the LDOCE for two reasons: first, some collocations are explicitly specified, and second, others must be retrieved from the corpus sentences. Nevertheless, it can be stated that the OALD is more systematic than the previous ones, since it provides specific information on PUs extracted from a collocation dictionary. Additionally, not only does this extract show collocations, but it also displays other word associations under the 'phrases' label (i.e., idioms). As a novelty, this extract offers *bottle* + noun collocations, for example, *bottle cap, bottle opener*. By the same token, the OALD is the first dictionary in explicitly addressing a preposition + noun combination, such as *over the bottle*. From this analysis, it can be deduced that the OALD seems more concerned about effectively incorporating collocations into its entries than the CED and the LDOCE.

Cambridge Dictionary of English (CDE)

The CDE is based on the *Cambridge English Corpus*,[98] a multi-billion-word collection of texts that contain written and spoken material, compiled from an enormous range of sources, such as newspapers, novels, radio, the Internet. It is constantly updated so that

[98] https://www.languagesciences.cam.ac.uk/news/cambridge-english-corpus-available-academic-use (Last accessed: 25/10/2020).

all the different teaching resources can guarantee that they use the latest information and show helpful and relevant material. The *Cambridge English Corpus* comprises several smaller corpora which allow linguists to conduct research into a variety of subject areas. For instance, the *Cambridge Learner Corpus* contains 50 million words from exam scripts; the *Cambridge Reference Corpus* is a multi-billion corpus of written and spoken *expert* English; the *Cambridge Academic Corpus* includes 400 million words of written and spoken material used in the academic sphere, such as lectures, seminars, student presentations and essays; and the *Cambridge Spoken Corpus* consists of 75 million words of transcriptions from everyday conversations, radio broadcasts, meetings, among others. Added to that, this corpus also incorporates a number of specialized corpora, such as business English, legal English and financial English.

Figure 10 shows the entry for *bottle* in the CDE,[99] which provides collocational information as follows:

i. Collocations are explicitly mentioned (*a milk bottle, a wine bottle, a bottle of beer/whisky*) as well as included in the sentence (*plastic bottles*) under the first definition;

ii. The third definition[100] provides only one collocation and it is implicit in the sentence: *take bottle*;

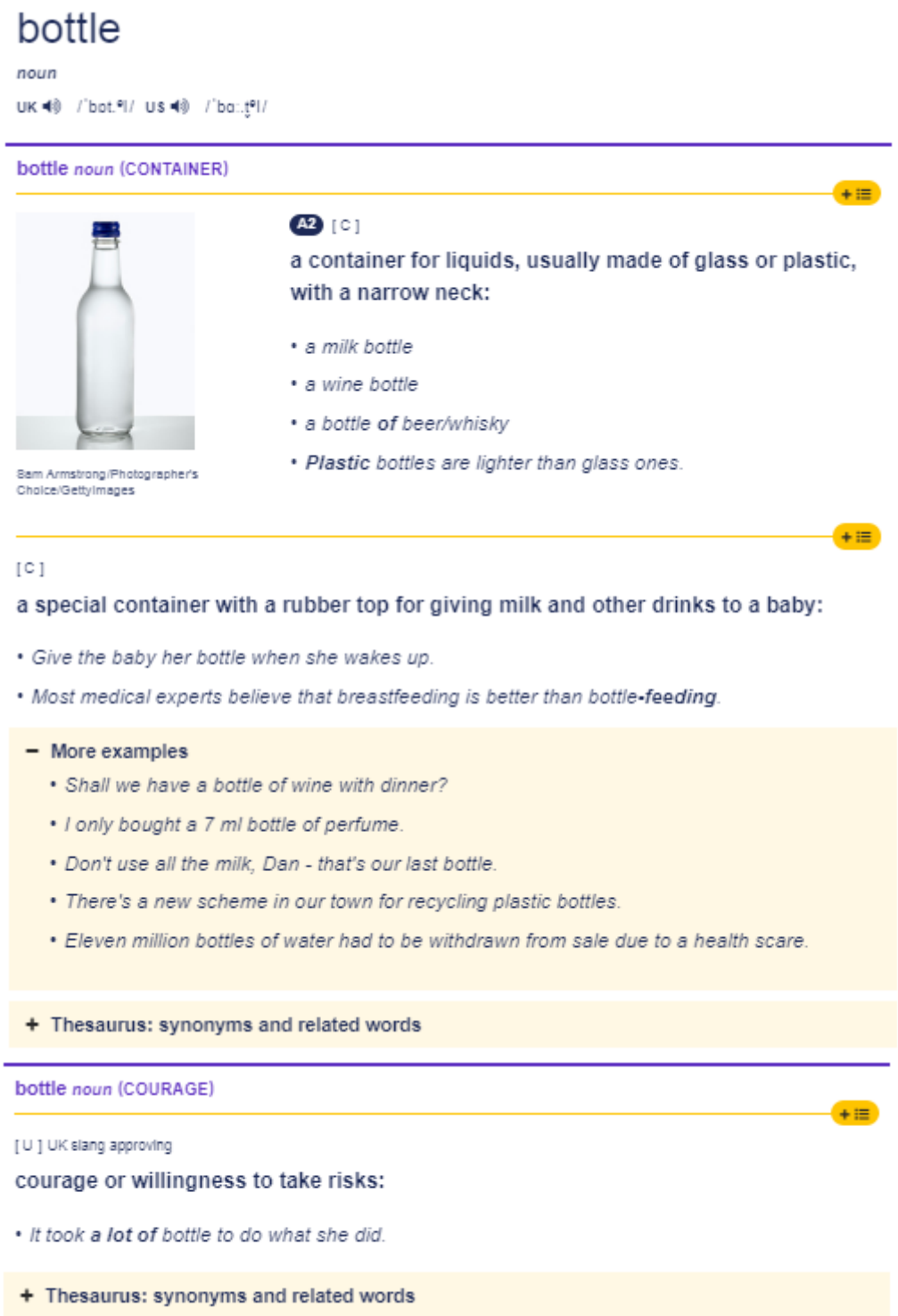
iii. The entry includes more sentences under the 'More examples' label from which other collocations can be inferred, for example, *There's a new scheme in our town for **recycling** plastic **bottles***; and

iv. A separate section gives more information on other PUs containing *bottle*, either as a noun or as a verb, such as collocations (e.g., *hot water bottle, bottle cap*), phrasal verbs (e.g., *bottle out*) and idioms (e.g., *hit the bottle*), as it can be seen in Figure 11.

[99] https://dictionary.cambridge.org/dictionary/english/bottle (Last accessed: 25/10/2020).
[100] It must be emphasized that the CDE makes a distinction between the two different meanings of *bottle*, "container" and "courage."

Figure 10

Entry for bottle *in the Cambridge Dictionary of English*



bottle

noun

UK /ˈbɒt.ᵊl/ US /ˈbɑː.ţᵊl/

bottle *noun* (CONTAINER)

Sam Armstrong/Photographer's Choice/GettyImages

A2 [C]

a container for liquids, usually made of glass or plastic, with a narrow neck:

- *a milk bottle*
- *a wine bottle*
- *a bottle **of** beer/whisky*
- ***Plastic** bottles are lighter than glass ones.*

[C]

a special container with a rubber top for giving milk and other drinks to a baby:

- *Give the baby her bottle when she wakes up.*
- *Most medical experts believe that breastfeeding is better than bottle-**feeding**.*

− **More examples**

- *Shall we have a bottle of wine with dinner?*
- *I only bought a 7 ml bottle of perfume.*
- *Don't use all the milk, Dan - that's our last bottle.*
- *There's a new scheme in our town for recycling plastic bottles.*
- *Eleven million bottles of water had to be withdrawn from sale due to a health scare.*

+ **Thesaurus: synonyms and related words**

bottle *noun* (COURAGE)

[U] UK slang approving

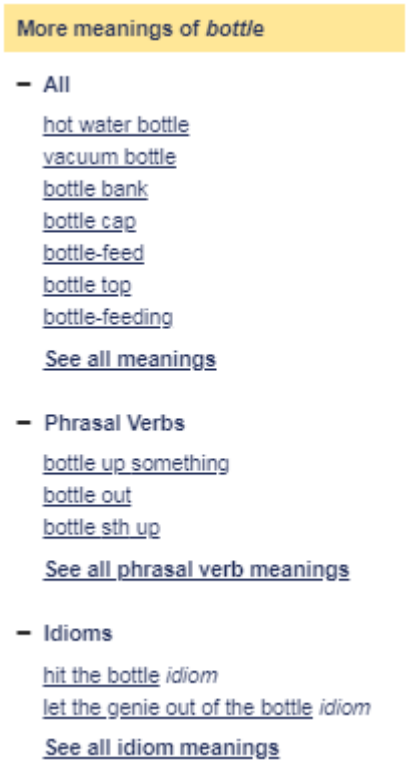
courage or willingness to take risks:

- *It took **a lot of** bottle to do what she did.*

+ **Thesaurus: synonyms and related words**

Figure 11

Extra Information on Phraseological Units Included in the Entry for bottle *in the Cambridge Dictionary of English*



In general, it can be said that the representation of PUs in the CDE is more systematic than the previous ones, for it provides a separate section showing extra information about MWUs containing the word *bottle*. Although this section is also offered in the OALD, the CDE clearly specifies the type of PU in the case of phrasal verbs and idioms (the 'All' section includes both collocations, e.g., *bottle top*, and compounds, e.g., *bottle-feeding*). Aside from that, this representation is reminiscent of those in the CED, the LDOCE and the OALD, given that some collocations are explicitly mentioned, whereas others are implicit in the corpus sentences.

Macmillan English Dictionary (MED)

The MED is the last of the English general dictionaries analyzed in this section. First thing to highlight is that it is also based on corpora.[101] The resources at Macmillan include

[101] http://www.macmillandictionaries.com/features/from-corpus-to-dictionary/ (Last accessed: 25/10/2020).

general and specialized corpora. On the one hand, the general corpus contains almost 1.6 billion words of written and spoken English, extracted from books, newspapers, TV programmes, radio meetings, among others. This is the corpus that the lexicographers use the most. On the other hand, the specialized corpus includes three subcorpora: (1) the *Macmillan Curriculum Corpu*s, with 20 million words of academic English; (2) the *Corpus of Environmental Science*, with 60 million words (the first of a planned series of new corpora for specific domains); and (3) the *Learner Corpora* or, more specifically, the *World English Corpus*. The latter is the result of Macmillan team's collaboration with the *Centre for English Corpus Linguistics* (CECL) at the Université Catholique de Louvain in Belgium and includes oral and written English produced by non-native speakers who are learning the language. As in the previous dictionaries, the compilers of the MED use corpora to find the meaning of words in context, their grammatical patterns, information about collocations and regional varieties, to name but a few.

Figure 12 shows the entry for *bottle* in the MED,[102] which provides collocational information as follows:

i. A collocation is made explicit only under definition 1: *bottle of*. In the rest of the definitions, they must be inferred from the sentences, for example, *an empty bottle, a whole bottle*, *have the bottle, lose one's bottle*;

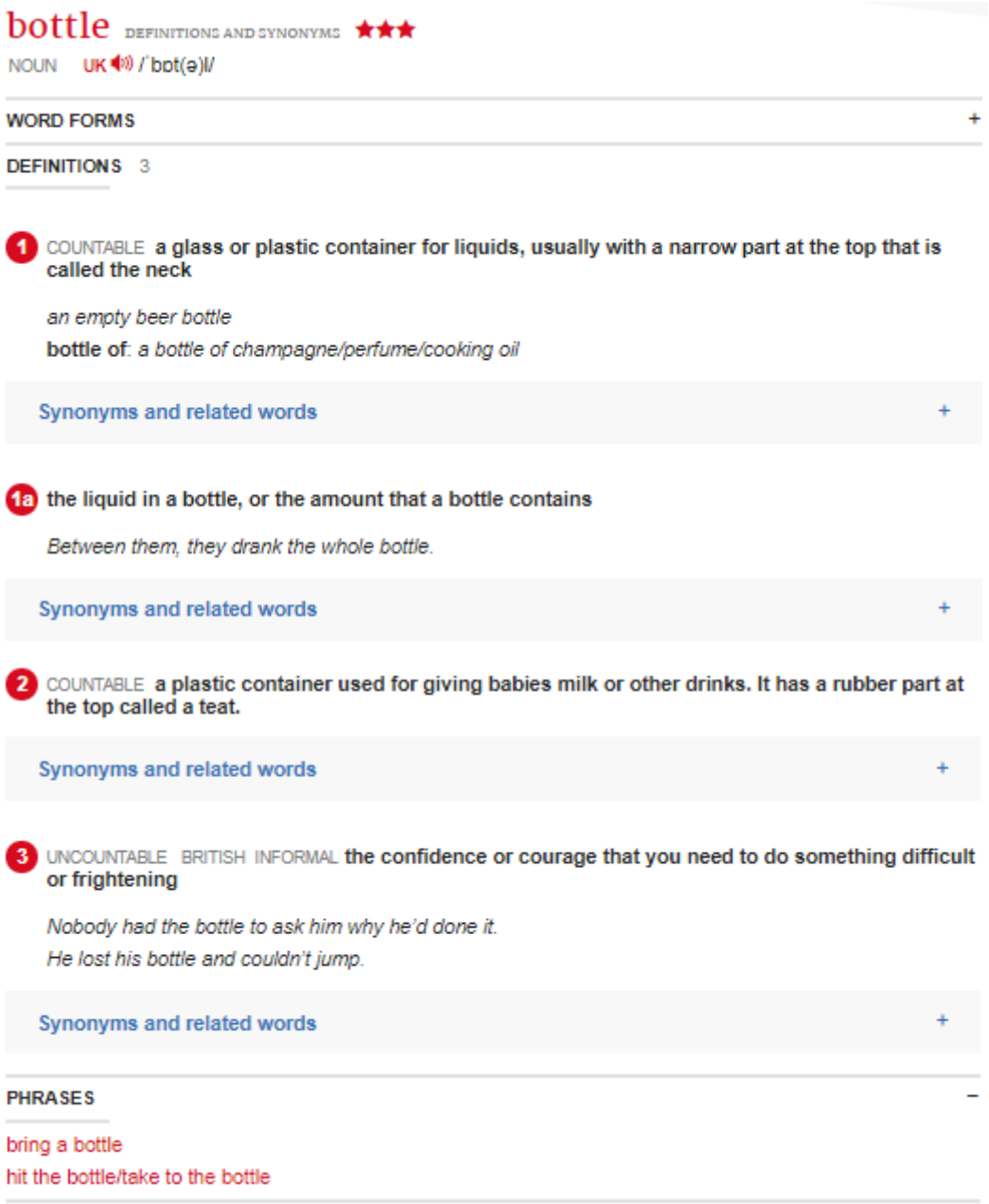
ii. At the bottom of the entry, two idioms are provided: *bring a bottle* and *hit the bottle/take to the bottle*. Their opaque meaning justifies the links to their definition in a separate entry; and

iii. The entry includes a little column called 'Other entries for this word,' which allows the access to other collocations (e.g., b*ottle shop*), phrasal verbs (e.g., *bottle up*), compounds (e.g., *bottle-feed*) and idioms (e.g., *let the genie out of the bottle*). An extract from this section is shown in Figure 13.

[102] https://www.macmillandictionary.com/dictionary/british/bottle_1 (Last accessed: 25/10/2020).

Figure 12

Entry for bottle *in the MED*



bottle DEFINITIONS AND SYNONYMS ★★★

NOUN UK /ˈbɒt(ə)l/

WORD FORMS +

DEFINITIONS 3

1 COUNTABLE **a glass or plastic container for liquids, usually with a narrow part at the top that is called the neck**

an empty beer bottle

bottle of: *a bottle of champagne/perfume/cooking oil*

Synonyms and related words +

1a **the liquid in a bottle, or the amount that a bottle contains**

Between them, they drank the whole bottle.

Synonyms and related words +

2 COUNTABLE **a plastic container used for giving babies milk or other drinks. It has a rubber part at the top called a teat.**

Synonyms and related words +

3 UNCOUNTABLE BRITISH INFORMAL **the confidence or courage that you need to do something difficult or frightening**

Nobody had the bottle to ask him why he'd done it.

He lost his bottle and couldn't jump.

Synonyms and related words +

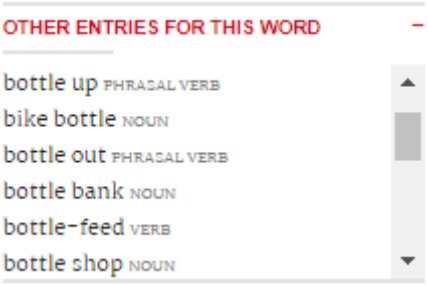
PHRASES −

bring a bottle

hit the bottle/take to the bottle

Figure 13

Extract From the Section Containing Information on Multiword Units From the Entry for bottle *in the Macmillan English Dictionary*



In our opinion, the treatment of collocations in the MED is as systematic as that in the OALD and the CDE, given that it includes specific references to MWUs in a separate entry. On the other hand, it is similar to all the previous dictionaries, that is, a collocation is explicitly mentioned, but the others need to be retrieved from the corpus sentences. In this respect, the MED includes a limited number of examples of the headword in context, so less collocations than in the previous analyses can be inferred.

In summary, we can say that all the English general dictionaries examined in this section offer a similar treatment of collocations in their entry for *bottle* in terms of their encoding, given that they can be explicitly incorporated under the definition/s of the headword or they can be implicit in the corpus sentences. However, the behaviour of the OALD, the CDE and the MED toward this matter must be highlighted, since they contain specific references to collocations and other MWUs, such as compounds, phrasal verbs and idioms, in a separate section in their entries. On the other hand, despite Walker's (2009, p. 287) idea that dictionaries tend to use a wide definition of collocation (including PUs that range from fully compositional to more fixed and non-compositional) and the different opinions on MWU typologies (cf. §2.1.), we believe that the categorization of the PUs contemplated by the compilers of the previous dictionaries is very much alike. By way of example, they agree on categorizing *hot-water bottle* as a collocation, *bottle up* as a phrasal verb and *hit the bottle* as an idiom, to mention but a few.

3.2.2. Collocations in English collocation dictionaries

The growing importance of phraseology and how words combine with each other has recently encouraged the publication of dictionaries that exhaustively examine

collocations, albeit it is not "an easy task" (Martin, 2008, p. 54). However, the origins of combinatory lexicographical works can be traced to the 1960s, as a result of Mel'čuk's investigations in the field of phraseology (cf. §2.1.3.), which inspired other works. Thus, his *Explanatory Combinatorial Dictionary* (ECD) (Žolkovskij & Mel'čuk, 1965, 1966, 1967) will be the starting point in this section, where we will analyze the representation of collocations containing the verb *improve*. After that, the following English collocation dictionaries will be also examined: *The BBI Combinatory Dictionary of English* (BBI) (Benson et al., 2009), *Oxford Collocations Dictionary for Students of English* (OCD) (Crowther, Dignen, & Lea, 2002; McIntosh, Francis, & Poole, 2009), *Macmillan Collocations Dictionary for Learners of English* (MCD) (Rundell, 2010)[103] and *HASK: Collocation Databases* (HASK)[104] (Pęzik, 2014). A brief introduction to each dictionary will be also provided.

Explanatory Combinatorial Dictionary (ECD)

Section 2.1.3. analyzed the work carried out in the field of phraseology by Mel'čuk, whose motivation was lexicographical. In fact, it reached fruition with the publication of the ECD (Žolkovskij & Mel'čuk, 1965, 1966, 1967). The objective underlying this combinatorial dictionary was to offer a complete linguistic description of language by including two main aspects: first, the lexicon, and second, a grammar which covers the inflections and derivations of words and the semantic, syntactic, morphological and phonological rules that can be applied to individual words and sets of words, such as collocations. To put it simply, the authors' main goal was to provide a systematic and formal description of the entire set of linguistic properties of LUs, where the lexicon enjoyed a privileged status: "A formalized semantically-oriented and cooccurrence-centered lexicon must be one of the central components of a linguistic description of any language" (Mel'čuk, 2006, p. 226).

The ECD has been one of the most influential and inspirational combinatory dictionaries. According to Mel'čuk (2006), this dictionary is characterized by three "general" properties (pp. 227-230): (1) it is theory-oriented because it has been compiled

[103] See Fuertes-Olivera (2011) for a full review on Rundell's (2010) book and Nurmukhamedov (2011) for an objective description of the MCD.
[104] http://pelcra.pl/hask_en/Home (Last accessed: 25/10/2020).

within a specific linguistic theory, that is, the MTT; (2) it is a formalized dictionary, so nothing can be left to the user's intuitions; and (3) all entries must be complete, that is, all the information about an LU must be fully presented. On the other hand, the ECD is also characterized by five "specific" properties (pp. 230-242): (1) it aims at production, which is the reason for providing the user with the maximum of linguistic information about an LU; (2) semantics has priority over syntax, thus, the ECD is semantically-based and definitions are more complex than in a conventional dictionary (this is the reason why the ECD is called *explanatory*); (3) this dictionary is called *combinatorial* because it endeavors to cover restricted lexical co-occurrences of each LU, for example, collocations; (4) both compounds and idioms/quasi-idioms are treated as single LUs, therefore, their representation in the ECD is homogeneous; regarding collocations, they are encoded in the entry for their base, which is normally a noun,[105] such as *pay a **visit***, *decline an **invitation***; and (5) each lexical entry is a monosemous LU, so polysemous words have separate entries.

Figure 14 shows the entry for *improve* in the ECD (Mel'čuk, 2006, p. 241), which provides collocational information as follows:

i. Despite stating that the ECD aims to cover lexical co-occurrences of words, the entry for *improve* is not a good example, given that it displays none;

ii. A couple of grammatical collocations, that is, verb + preposition, are provided in definitions IMPROVE$_{I.3}$: *improve at*, and IMPROVE$_{III}$: *improve upon*;

iii. Different types of lexical collocations can be inferred from the example sentences, such as noun + verb (e.g., *the weather improves*, *a system improves*), verb + noun (e.g., *improve a system*, *improve a house*) and adverb + verb (e.g., *drastically improve*, *steadily improve*) collocations; and

iv. The entry explicitly mentions the quasi-idiom [*to*] *improve oneself*, which is accompanied by a definition ("to educate oneself, to improve one's culture") but lacks an example in context.

[105] L'Homme (2003, p. 404) reveals that this is a shared feature of several terminological resources. However, it must be highlighted that *DicoAdventure* encodes collocations in verb entries.

Figure 14

Entry for improve *in the Explanatory Combinatorial Dictionary*

IMPROVE, verb

IMPROVEL1a *X improves* ≡ 'The value of the quality of X becomes higher'
[*The weather suddenly improved*; *The system will improve over time*]

IMPROVEL1b *X improves Y* ≡ 'X causes1 that Y improvesL1a'[9]
[*The most recent changes drastically improved the system*]

IMPROVEL2 *X improves* ≡ 'The health of a sick person X improvesL1a'
[*Jim is steadily improving*]

IMPROVEL3 *X improves at Y* ≡ 'X's execution of Y improvesL1a, which is caused1 by X's having practiced or practicing Y'
[*Jim is steadily improving at algebra*]

IMPROVEII *X improves Y by Z-ing* ≡ 'X voluntarily causes2 that the market value of a piece of real estate Y becomes higher by doing Z-ing to Y'
[*Jim improved his house by installing indoor plumbing*]

IMPROVEIII *X improves upon Y* ≡ 'X creates a new Y' by improvingL1b Y'
[*Jim has drastically improved upon Patrick's translation*]

(The phrase [*to*] *improve oneself* is a set expression—actually a quasi-idiom, meaning 'to educate oneself, to improve one's culture',—and should be considered here.)

It can be stated that the entry for *improve* in the ECD does not give a comprehensive description of the collocations which the verb can be part of. In other words, it shows a couple of grammatical collocations, although they are not specified as such, but all lexical collocations must be retrieved from the sentences. As these PUs are not explicitly indicated, the application of Mel'čuk's (1996) LFs (cf. §2.1.3.) to encode collocations in a combinatory dictionary cannot be examined. However, the ECD specifically addresses another type of PU, that is, the quasi-idiom [*to*] *improve oneself*, which comes with a definition, as the meaning of this phrase is non-compositional. Finally, the fact of using separate lines for different meanings of the verb is a positive aspect of the dictionary, for different collocates can provide the same base with different nuances of meaning (Baynat Monreal, 2017, p. 71).

The BBI Combinatory Dictionary of English (BBI)

The compilation of the BBI (Benson et al., 1986b)[106] was determined by two significant factors: first, the influence of the ECD, and second, the authors' pedagogical approach to PUs. Thus, this dictionary was compiled to help students of English as an L_2 avoid errors when combining words, given that these products "must do more than describe words that stand alone" (Benson et al., 1986a, p. 252). The dictionary is referred to as *combinatory* rather than *collocational* because, although it only contains collocations, it also provides relevant information on complementation patterns of nouns, verbs and adjectives (as well as 'Usage Notes,' which offer full explanations for language problems, and a workbook with exercises). Therefore, this dictionary is not only about phraseology but also valency (Benson et al., 2009):

> To use language you must be able to combine words with other words to form phrases and to combine words into grammatical patterns to form clauses and sentences. Traditionally, the combination of words with words has been called collocation and its result has been called phraseology. Traditionally, the combination of words into grammatical patterns has been called colligation or complementation or construction (though in BBI it is called collocation, too) and its result has been called valency. A dictionary that provides both phraseology and valency is a dictionary of word combinations; or, in the terminology of Igor Mel'čuk, whose work has inspired us, a combinatory or combinatorial dictionary. BBI is a combinatory dictionary. (p. i)

As it can be seen, the authors make a distinction between the combination of words with words and the combination of words with grammatical patterns. The former is known as lexical collocation (i.e., the combination of two lexical words: noun, adjective, verb, adverb) and the latter is called grammatical collocation (colligation in Sinclair's terms) (i.e., a lexical word plus a grammatical structure), and this taxonomy (cf. §2.2.2.) forms the basis for the BBI, whose third edition (2009) is evidence of its success and great acceptance. It represents an extensive revision of the 1997 edition, which included 18,000

[106] The BBI was first launched in 1986. Then, the second edition came in 1997 and the third edition, which is the one that will be analyzed in this section, was released in 2009.

entries and roughly 90,000 collocations, being 20 to 25% larger and which incorporates new collocations from the field of computing and the Internet. It also pays close attention to differences between American and British English.

Figure 15 shows the entry for *improve* in the BBI (Benson et al., 2009, p. 198), which provides collocational information as follows:

i. The verb's collocates are listed using different numbers, apparently according to the meaning of the verb within the combinations. However, the verb is not defined, despite the authors' statement: "Definition of senses of verbs [...] are given in double quotation marks and parentheses" (p. xxxvii). Groups 2. and 3. inform that the verb collocates with a preposition ("D/d") and is used intransitively ("intr.") in these collocations;

ii. Number 1. offers examples of the lexical collocation verb + adverb/adverb + verb (type L7), for example, *improve* + *dramatically, greatly* and *noticeably*. However, the grammatical collocation (type G8-D) *to improve beyond all recognition* is also included here;

iii. Numbers 2. and 3. also provide grammatical collocations with the pattern verb + preposition + object: *to improve in*, *to improve on/upon*, along with example sentences; and

iv. Number 4. is the category of miscellaneous entries ("misc.") and includes the same type of grammatical collocation: *to improve with age* (no example showing the phrase in context is given), on the one hand, and an example of the verb passivized, included in the sentence *your work is much improved!*, on the other.

Figure 15

Entry for improve *in the BBI Combinatory Dictionary of English*

improve *v.* 1. to ~ beyond all recognition, dramatically, greatly, noticeably, significantly, very much; gradually, slowly; rapidly 2. (D; intr.) to ~ in (she has ~d in English) 3. (d; intr.) to ~ on, upon (I cannot ~ on her performance) 4. (misc.) to ~ with age; your work is much ~d!

Despite knowing that the taxonomy of collocations suggested by Benson et al. (1986b) was of paramount importance for the compilation of the BBI, it must be highlighted that its influence on the entry for *improve* is not overwhelming. Although the arrangement of collocations is made according to the compilers' (2009, p. xvi) idea, that is, lexical collocations (L7) coming before grammatical collocations (G8-D), the inclusion of *to improve beyond all recognition* in the first group is confusing. Then, sections 2. and 3. are separated, in spite of including the same type of collocation with the same use of the verb (intransitive), which remains unexplained. To our mind, it may be due to the fact that the verb has a different meaning in each of the cases, given that in *to improve in* the verb keeps its literal meaning, but *to improve on/upon* is a phrasal verb (cf. §2.2.3.), as it is indicated by the LDOCE,[107] the OALD,[108] the CDE[109] and the MED.[110] In this regard, two aspects must be explained: (1) the authors (1986b, p. xxi) clearly state that phrasal verbs are excluded from collocations of the type G8-P (cf. §2.2.2.) but do not say anything about including them in G8-D, and (2) the authors (1986a, p. 254) place phrasal verbs within the category compounds (vs. the category collocations), so there is no reason to mention them in their taxonomy of collocations. Given these points, the inescapable conclusion drawn is that the border between grammatical collocations and phrasal verbs is a bit blurred for the authors of the BBI.

On the other hand, compared to the ECD, the entry in the BBI represents the collocations of *improve* more accurately, as they are made explicit. Nevertheless, the latter could learn from the former in two aspects: first, the BBI lacks more and longer examples of the collocations in context, and second, an explanation of the meaning of the verb in the different collocations may have resolved the considerable uncertainty over the concept and role of phrasal verbs in the dictionary.

[107] https://www.ldoceonline.com/dictionary/improve-on-upon (Last accessed: 25/10/2020).
[108] https://www.oxfordlearnersdictionaries.com/definition/english/improve-on (Last accessed: 25/10/2020).
[109] https://dictionary.cambridge.org/dictionary/english/improve-on-upon-sth (Last accessed: 25/10/2020).
[110] https://www.macmillandictionary.com/dictionary/british/improve-on (Last accessed: 25/10/2020).

Oxford Collocations Dictionary for Students of English (OCD)

The OCD was first published in 2002 (Crowther et al., 2002) and a second edition was published in 2009 (McIntosh et al., 2009).[111] The main reason for creating a collocation dictionary[112] was that "No piece of natural spoken or written English is totally free of collocation" and the objective was to help English language learners develop their competence, as using collocations "produce natural-sounding speech and writing" (Crowther et al., 2002, p. vii). Just like its monolingual counterpart, that is, the OALD, the OCD is based on the BNC, which allows the compilers to check the frequency of recurrence of a collocation in a language and in particular contexts. The 2009 edition of the OCD contains 250,000 collocations and about 9,000 headwords (nouns, verbs and adjectives), along with 75,000 example sentences extracted from the corpus. When the headword is polysemous, a definition of each meaning is provided.

For the selection of the collocations, the introductions to both editions of the dictionary (Crowther et al., 2002, pp. viii-ix; McIntosh et al., 2009, p. v) state that they are selected on the basis of the answers to the following questions: (1) "Is this a typical use of language?," (2) "Might a student of English want to express this idea?," and (3) "Would a student look up this entry to find this expression?" The first question requires that all collocations be obtained from reliable data, that is, authentic English (i.e., the corpus); the second question leads to a focus on current English, more specifically, British English; and the third question leads to the exclusion of nominal collocates from verb and adjective entries, since a user would rather look up the entry for *rain* than the entry for *heavy* to know the collocation *heavy rain*.

Figure 16 shows the entry for *improve* in the OCD (McIntosh et al., 2009, p. 4375), which provides collocational information as follows:

i. Three types of collocations are clearly displayed in separate sections. As no definitions are provided, it can be guessed that the verb has the same meaning in all the combinations;

ii. First, a large number of adverbs that can combine with the base forming the lexical collocation verb + adverb/adverb + verb are provided, for example,

[111] The most recent edition is the one considered for the current analysis.
[112] See Lea (2007) for a detailed explanation of the process and principles behind the selection and presentation of collocations in the OCD.

considerably, noticeably, marginally. They seem to be organized in terms of how much (e.g., *greatly* vs. *slightly*) or how quickly (e.g., *rapidly* vs. *steadily*) something improves; no example sentences are given;

iii. Second, verbs that tend to precede the headword are shown, such as *continue to, strive to, help to*. Only two sentences show the collocations in context, one of which is not complete; and

iv. Finally, the last section includes different types of phrases containing the verb, for instance, *an attempt/effort to improve sth, much improved*, together with some example sentences.

Figure 16

Entry for improve *in the Oxford Collocations Dictionary for Students of English*

improve *verb*

ADV. **considerably, dramatically, greatly, immeasurably, materially, radically, significantly, substantially, vastly | markedly, noticeably | marginally, slightly | rapidly | steadily**

VERB + IMPROVE **continue to** *The weather should continue to improve over the weekend.* | **strive to, try to** *trying to improve their working conditions* | **help to** | **be designed to**

PHRASES **aimed at improving sth** *measures aimed at improving government efficiency* | **an attempt/effort to improve sth, an incentive to improve sth** *Workers need to be given an incentive to improve their performance.* | **much improved** *We now offer a much improved service to our customers.*

The presentation of collocations in the OCD looks more effective than that in the BBI (and, obviously, that in the ECD) for three main reasons: (1) the colors and fonts used make it more visually appealing; (2) the collocations are grouped in terms of their type, and differences in meaning within the same type are expressed by means of a vertical line "|;" and (3) the number of collocations displayed is larger and they are more varied, since combinations of the headword with other verbs are included. Nevertheless, a major drawback is that the semantic organization of collocations within the same section is not efficient, since the semantic relationship between the base and the collocate is not stated

and must be deduced by the user. To this respect, we agree with McGee (2012) that collocate entries are "semantically organised, but this organisation is not actually labeled" (p. 332). Apart from that, we miss some corpus sentences in the first section. Finally, it can be emphasized that the arrangement of collocational types is in accordance with Benson et al.'s (2009) thought, that is, lexical collocations (L7) appear before grammatical collocations (G8-E) and other phrases. However, the OCD's compilers have not revealed the influence of the BBI on their product, so we think it may be purely incidental.

Macmillan Collocations Dictionary for Learners of English (MCD)

The MCD was published in 2010 (cf. Rundell, 2010) and, like its monolingual counterpart, that is, the MED, and the OCD, is based on corpora. To put it more simply, the information collected is retrieved from a two-billion-word corpus containing written and spoken English, which allows to know the frequency of co-occurrents, given that the MCD regards collocations as "words that are often used together."[113] The compilers of the MCD recognize that their dictionary may include fewer entries than other collocation dictionaries, for they mostly concentrate on headwords which present a good range of strong collocates, thus ignoring those which present no real collocational problems, such as *food, house* or *goalkeeper*. The novelty of this dictionary lies in its methodology to extract collocations from the corpus,[114] since it uses *word sketches*, corpus-derived summaries which show a word's collocational behaviour.[115] The advantage of using word sketches is that they offer reliable and valuable information. For instance, Kilgarriff, Kovář, Krek, Srdanovic, & Tiberius (2010) discovered that two thirds of the collocations yielded in word sketches could be included in a collocation dictionary according to users of different languages, such as Dutch, English, Japanese and Slovene.

Figure 17 shows the entry for *improve* in the MCD (Rundell, 2010, pp. 392-393), which provides collocational information as follows:

[113] http://www.macmillandictionaries.com/about/collocations-dictionary/ (Last accessed: 25/10/2020).
[114] http://www.macmillandictionaries.com/features/from-corpus-to-dictionary/ (Last accessed: 25/10/2020).
[115] The 'Word Sketch' function is incorporated into the Sketch Engine corpus query system and is of great relevance to this work, as we will explain in §4.4.

i. Two types of collocations are separately displayed and signaled with this symbol "●" in red color;

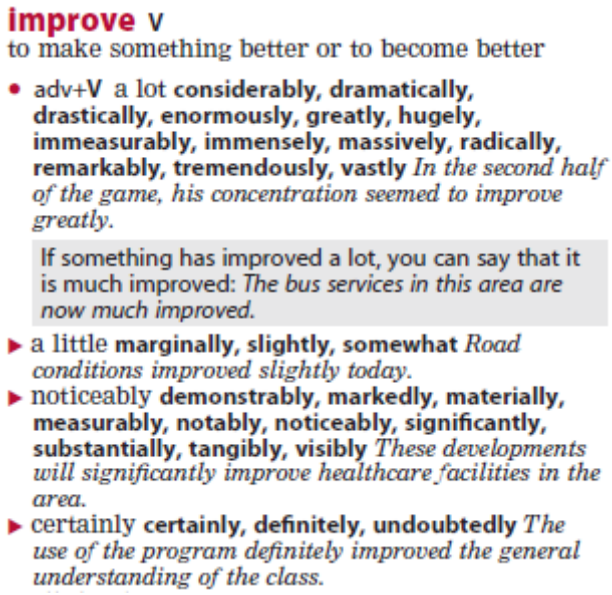
ii. The first of the types shown is adverb + verb collocations, which includes a huge number of collocates organized semantically. For instance, these groups are arranged in terms of the magnitude of the improvement, such as "a lot:" *enormously, massively* (this section includes a box explaining a different way of expressing this meaning), versus "a little:" *marginally, slightly*, and the speed at which the improvement is made, for example, "slowly or gradually:" *progressively, slowly*, versus "quickly:" *fast, rapidly*, among others;

iii. The second type of collocation is grammatical and contains the verb plus the preposition *with*; then, several nouns co-occurring in this structure are given, for example, *age, experience, time*; and

iv. Each of the collocates of the verb is highlighted in bold and full example sentences are given in italics.

Figure 17

Entry for improve *in the Macmillan Collocations Dictionary for Learners of English*



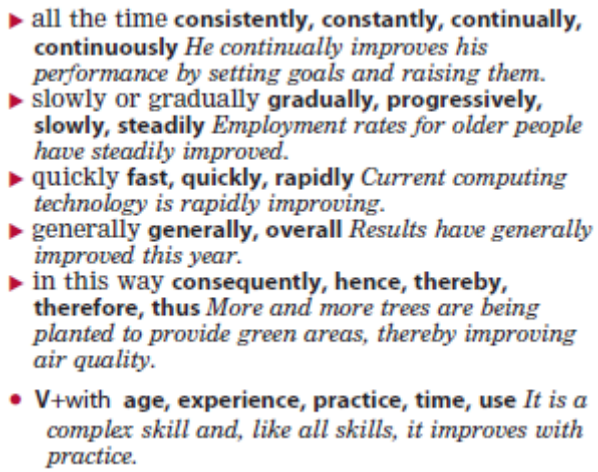
improve v

to make something better or to become better

- adv+V a lot **considerably, dramatically, drastically, enormously, greatly, hugely, immeasurably, immensely, massively, radically, remarkably, tremendously, vastly** *In the second half of the game, his concentration seemed to improve greatly.*

  If something has improved a lot, you can say that it is much improved: *The bus services in this area are now much improved.*

- ▸ a little **marginally, slightly, somewhat** *Road conditions improved slightly today.*
- ▸ noticeably **demonstrably, markedly, materially, measurably, notably, noticeably, significantly, substantially, tangibly, visibly** *These developments will significantly improve healthcare facilities in the area.*
- ▸ certainly **certainly, definitely, undoubtedly** *The use of the program definitely improved the general understanding of the class.*



Of course, the representation of collocations in the MCD is more accurate and comprehensive than those in the ECD, the BBI and the OCD. On the positive side, distinct aspects must be emphasized: (1) the definition of the verb is provided under the headword; (2) the MCD first classifies the different types of collocations (lexical vs. grammatical) and then groups the collocates semantically; (3) at least one full example sentence is provided in each of the semantic groupings; (4) the number of collocations supplied is higher than in the previous dictionaries, although they are of the same type and phrases are excluded; (5) the entry is more visually attractive because of the many colors, fonts and symbols used; and (6) an explanatory box is included to explain how something can be paraphrased. Furthermore, like in the OCD, lexical collocations (L7) are shown before grammatical collocations (G8-D), which agrees with the order suggested by Benson et al. (2009). However, since no references to their work are dropped in the MCD, we think it may be again incidental.

On the negative side, it must be pointed out that we miss the inclusion of nominal collocates, but we think it is due to the fact that they are normally excluded from verb entries, as it is mentioned by the OCD's compilers (e.g., Crowther et al., 2002, p. ix); in fact, they also lack in all the previous dictionaries. Despite that, we still expected to find them for the following reason. McGee (2012) and Buendía Castro and Faber Benítez (2014) carried out a comprehensive comparative analysis on the encoding of collocations in lexicographical resources, more specifically, the former examined the *LTP Dictionary of Selected Collocations*, the BBI, the OCD and the MCD, and the latter explored the BBI, the OCD and the MCD. Their conclusions were that none of the dictionaries included nominal collocates in verb entries, except for the MCD (McGee, 2012, p. 333;

Buendía Castro & Faber Benítez, 2014, p. 215), which broke "with collocation dictionary tradition in providing nominal collocates for both adjective and verb entries" (McGee, 2012, p. 333). Nevertheless, our analysis denies this statement and demonstrates that the entry for *improve* in the MCD still continues the collocation dictionary tradition regarding the inclusion of nominal collocates in verb entries.

HASK: Collocation Databases **(HASK)**

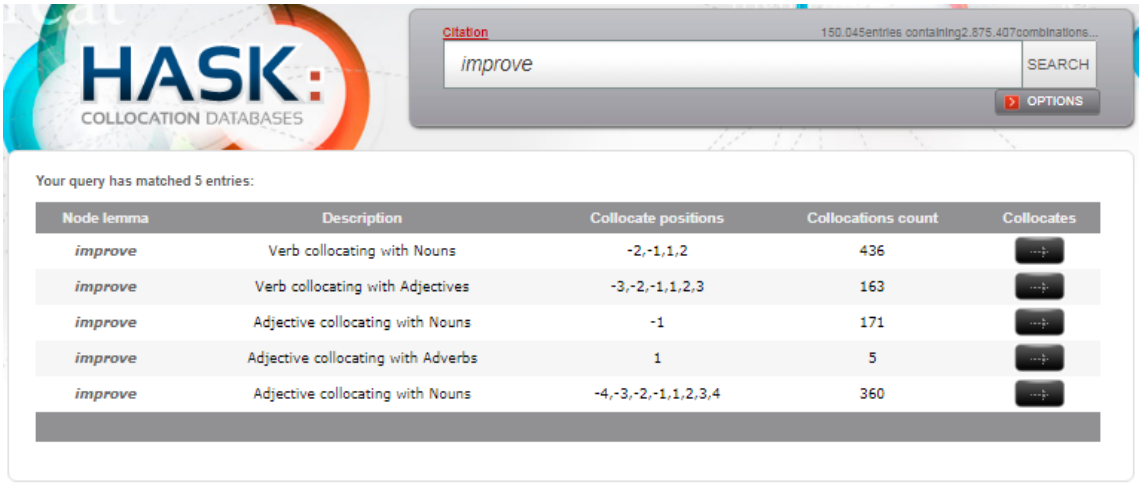
The last English collocation dictionary analyzed in this section is HASK (Pęzik, 2014), a bilingual phraseological database of English and Polish language targeted at linguists, lexicographers, language teachers, language material developers and, basically, anyone who wants to use a dictionary of this type. Its browser gives instant access to lists of word combinations and predefined patterns found in reference corpora; in the case of the English dictionary, information is extracted automatically from the BNC. Thus, HASK finds potential collocates of nouns, verbs, adjectives and adverbs. The English version of HASK contains over 150,000 entries with a total of 2.8 million recurrent word combinations (p. 229). The tool is freely available online.[116]

Figure 18 shows the results obtained when *improve* is entered in the search box. As this dictionary makes searches based on lemmata, a list of the combinations containing the verb *improve* and the adjective *improving* is provided. This list also gives information about the span existing between the base and the collocates ('Collocate positions') and the number of collocations found in each group ('Collocations count'), together with a link to the collocates ('Collocates') in each type of combination.

[116] http://pelcra.pl/hask_en/Home (Last accessed: 25/10/2020).

Figure 18

Results Obtained After Searching for the Lemma improve *in HASK*



Since our analysis is on the representation of collocations of the verb *improve*, Figure 19 shows an extract from the results from the 'Verb collocating with Nouns' section,[117] which provides collocational information as follows:

i. A long list of nouns that collocate with the verb *improve* in the positions -2/+2 is generated;

ii. This list is arranged according to the t-score[118] of the collocations in the reference corpus, although other values are also provided, for example, MI3 (Daille, 1994; Fano, 1961) and JD (Juilland, Brodin, & Davidovitch, 1971); and

iii. Each of the collocates offers access to more 'Details,' such as other statistical indicators and corpus sentences (Figure 20), and 'Equivalents' of the collocations in Polish (Figure 21).

[117] We have decided to show the results from this section because it cannot be guaranteed that the word associations listed in 'Verb collocating with Adjectives' contain the verb *improve* modified by adjectives, given that these adjectives seem to rather collocate with the nouns combining with the verb, for instance, *We are helping to* ***improve public*** *administration [...]*, or even with words in other clauses, such as in *They liked Mr Major, whose performance visibly* ***improved*** *as his* ***public*** *assurance grew [...]*.

[118] The t-score is one of the preferred association measures in corpus linguistics (Clear, 1993; Gries, 2010). It considers both the strength and the significance of association between two words along with the corpus size. It can be admitted that the higher the score is, the higher the confidence is in assuming that the combination is not random and "we can claim that there is some association" (Clear, 1993, p. 281). According to Xiao (2015, p. 110), a t-score of 2.576 or above is considered to be statistically significant.

Figure 19

Extract From the Results From the 'Verb collocating with Nouns' Section of the Lemma improve *in HASK*

Display collocates 1 - 50 from 436

Your query has fetched 436 combinations of *improve* + *Noun*

Show 50 entries Filter Ok

Node frequency: 10829

entries

#	Collocate	POS	A	TTEST	MI3	CHISQ	JD	R	AWT	fAWT	Details	Equivalents
1.	quality	Noun	426	19,85	22,17	10.300,43	0,91	90	462.008	117,66		
2.	performance	Noun	252	15,07	20,27	4.518,58	0,88	76	799.580	67,99		
3.	efficiency	Noun	176	13,02	20,66	9.091	0,88	65	1.073.644	50,63		
4.	condition	Noun	183	12,01	18,18	1.284,3	0,91	74	859.379	63,25		
5.	standard	Noun	166	11,92	18,49	1.907,52	0,88	63	1.086.952	50,01		
6.	situation	Noun	159	11,25	17,84	1.177,01	0,9	72	852.166	63,79		
7.	service	Noun	194	10,74	17,33	505,48	0,86	69	980.202	55,46		
8.	effort	Noun	97	8,69	16,28	639,63	0,88	56	1.291.752	42,08		
9.	health	Noun	107	8,27	15,8	343,19	0,86	60	1.152.370	47,17		
10.	safety	Noun	82	8,23	16,17	745,08	0,8	38	2.468.842	22,02		

Figure 20

Extract From the Results From the 'Details' Section From the Entry improve [Verb] + situation [Noun] *in HASK*

improve [Verb] + situation [Noun]

Strength		Strength		Dispersion	
A:	159	CHISQ:	1.177,01	AWT:	852.166
TTEST:	11,25	G2:	431,95	FAWT:	63,79
MI:	3,21	LOGLOG:	23,5	JD:	0,9
MI3:	17,84	DICE:	1,06	RANGE:	72

1.	But we acknowledge that things are n't perfect and we are seeking to improve the situation.	A4W
2.	It shows how little the FBI really understood about film production that their proposals to improve the situation focused on increasing the quantity of production, rather than stimulating a few high-quality pictures that might promote British prestige abroad.	A7L
3.	If, however, other countries take similar measures to cut pollution, the situation obviously improves.	AB6
4.	It is doubtful if its situation has improved since then.	AB6
5.	No, I ca n't see a weekend in Paris improving the situation ... with all that it would obviously lead to.	ACE
6.	You should find that with regular watering and feeding until the foliage dies down naturally the situation will improve next year.	ACY

Figure 21

Extract From the Results From the 'Equivalents' Section From the Entry improve [Verb] + quality [Noun] *in HASK*

improve [verb] + quality [noun]

Show 5 entries

#	Hask	Node PoS	Collocate PoS	Node	Collocate	Cooc	Dice
1	jakość [rzeczownik] + poprawić [czasownik]	czasownik	rzeczownik	poprawić	jakość	230	0.09
2	poprawić [czasownik] + jakość [rzeczownik]	rzeczownik	czasownik	jakość	poprawić	230	0.09
3	członkowski [przymiotnik] + państwo [rzeczownik]	rzeczownik	przymiotnik	państwo	członkowski	228	0.09
4	państwo [rzeczownik] + członkowski [przymiotnik]	przymiotnik	rzeczownik	członkowski	państwo	228	0.09
5	być [czasownik] + cel [rzeczownik]	rzeczownik	czasownik	cel	być	155	0.06

Showing 1 to 5 of 20 entries

Previous 1 2 3 4 Next

Without a doubt, the approach to the representation of collocations developed in HASK is entirely different from those in the ECD, the BBI, the OCD and the MCD, so direct comparison is not invited. Despite this, we can still highlight that HASK includes reliable statistics to know the significance of the collocations retrieved, which leads to an efficient organization of the collocates. Obviously, it is possible because the tool is available on the Internet, a fact that allows the user to choose different ways of displaying the results (e.g., on a diagram) and to download them onto his/her computer (e.g., in a spreadsheet). A novelty if we compare HASK with the previous dictionaries is that it includes PUs where the verb *improve* collocates with a noun. However, unlike them, it does not retrieve verb + adverb/adverb + verb collocations or grammatical collocations.

The conclusions drawn from this analysis on English collocation dictionaries will focus on the results obtained from the ECD, the BBI, the OCD and the MCD, as we feel HASK cannot be compared with them in view of its avant-garde style and sophisticated features (for this reason, we think that it can become useful only to people who can profitably exploit this resource). All in all, it can be noted that the representation of collocations in the aforementioned tools has been progressively improving from the first dictionary, the ECD, to the last one, the MCD. Thus, the ECD does not explicitly address collocations, which certainly need to be inferred from the sentences. Then, the BBI specifies the collocations of the verb *improve*, but they are not properly organized or defined and they lack examples in context. After that, the OCD offers an enhanced version, for it uses colors, different sections to arrange the different types of collocations

and displays more collocates; however, semantic information as well as more corpus sentences are missing. Finally, we really believe that all the drawbacks mentioned are overcome by the MCD, whose strengths are the inclusion of the definition of the verb, a more detailed classification of collocations in terms of types and meaning, the presence of plenty of example sentences, a higher number of collocations extracted and the use of several colors, fonts and symbols that make the entry easier to understand. The only weakness mentioned is that the MCD does not include nominal collocates (either in subject or object position) (cf. Buendía Castro & Faber Benítez, 2014; McGee, 2012), as they will probably be encoded in the noun entry, which is the common practice in lexicography. Thus, the following section will investigate whether it also happens in specialized lexicographical tools.

3.3. Collocations in terminological resources

Like lexicographers, terminographers have also acknowledged the importance of including collocations in specialized dictionaries (e.g., Bergenholtz & Tarp, 1995; Heid & Freibott, 1991; Schneider, 1998; Vargas Sierra, 2010). Nevertheless, the fact that there are still subject fields which have not been covered makes it an area insufficiently exploited. It may be the reason for the inconsistent treatment given to collocational information in these tools, as contended by L'Homme and Leroyer (2009, p. 259), whereby one could say that there is no agreement on methodologies for listing and representing collocations in entries, or for which of them should be included and how they should be described and classified (Buendía Castro et al., 2014, p. 61). Although it may be true, there seems to be some common ground to this respect, as a number of similarities when encoding collocations in specialized dictionaries or term banks has been discovered (L'Homme, 2009a, p. 239): (1) these PUs are listed under a headword that has already been defined as a term in a specialized subject field, (2) the base of the collocation is usually a noun or a noun phrase, and (3) the collocates are usually content words, such as verbs, nouns and adjectives.

In this section, a complete description of five valuable terminological resources that include collocational information about the English language will be provided.[119] Where possible, this analysis will comment on how collocates are encoded in verb entries,

[119] This description will focus on other aspects of the tools aside from the representation of collocations.

although different terms will be selected as these tools deal with varied specialized domains, namely: (1) the domain of economy: *Dictionnaire contextuel du français économique* (DICOFE) (Verlinde & Binon, 1994; Verlinde, Binon, & Van Dyck, 2000; Verlinde, Folon, Binon, & Van Dyck, 2000, 2003a, 2003b); (2) the domain of computing and the Internet: *Le dictionnaire fondamental de l'informatique et de l'Internet* (DiCoInfo);[120] (3) the domain of the environment: *Le dictionnaire fondamental de l'environnement* (DiCoEnviro)[121] and *EcoLexicon: Terminological Knowledge Base on the Environment* (EcoLexicon);[122] and (4) the European Union's terminology database: *Interactive Terminology for Europe* (IATE).[123]

Dictionnaire contextuel du français économique (DICOFE)

The DICOFE (Verlinde & Binon, 1994; Verlinde, Binon, et al., 2000; Verlinde, Folon, et al., 2000; Verlinde et al., 2003a, 2003b) was the first attempt to make a production-oriented dictionary of Business French with an onomasiological organization.[124] Its compilation was the work of the *Groupe de Recherche en Lexicographie Pédagogique* (GRELEP), a research group whose members are interested in pedagogical lexicography and most of whom teach Business French at university level. It was motivated by the lack of information on word combinations in specialized dictionaries, which were mainly focused on terminology. The project was conducted at the Institute for Modern Languages of the Catholic University of Leuven (Belgium) and was based on two sources of information: first, the compilers' didactic experience, and second, a corpus of about 1.1 million words of business texts covering a wide range of topics. Therefore, the DICOFE is composed of four volumes dealing with the subject fields of *l'enterprise* [company], *le commerce* [trade], *les finances* [finances] and *l'emploi* [employment].

Given the pedagogical approach to the DICOFE compilation, it is hardly surprising that its target users are intermediate learners. For this reason, an exercise book with key is provided. It contains more than twenty different tasks which are classified

[120] http://olst.ling.umontreal.ca/cgi-bin/dicoinfo/search.cgi (Last accessed: 25/10/2020).
[121] http://olst.ling.umontreal.ca/cgi-bin/dicoenviro/search.cgi (Last accessed: 25/10/2020).
[122] https://ecolexicon.ugr.es/en/index.htm (Last accessed: 25/10/2020).
[123] https://iate.europa.eu/home (Last accessed: 25/10/2020).
[124] The onomasiological organization proposed by the DICOFE's compilers chimes with Siepmann's (2005) belief that an "onomasiological rather than semasiological dictionary constitutes the ideal repository for the collocational and colligational units required by active users" (p. 3).

according to the type of exercises: (1) exercises on single words, word combinations and words in the text, (2) written and oral exercises, and (3) yes/no questions and open exercises. This variety, with different levels of difficulty, is remarkable, since the exercises are aimed to encourage autonomous learning. This is indeed one valuable and distinctive aspect of the DICOFE, which is considered a comprehensive dictionary for some other reasons too (Verlinde & Binon, 1994, pp. 524-525):

1) It contains the basic vocabulary of Business French;
2) The headwords are mainly nouns and they include their lexical combinations with verbs and adjectives;
3) The entries along with their word combinations are translated into Dutch;
4) Word families, instead of isolated words, are included and they are presented in alphabetical order;
5) It offers an introductory text to the world of business;[125] and
6) The dictionary is user-friendly and well-organized.

Access to the information gathered in the DICOFE can be gained by choosing a specific subject field, as Figure 22 (Verlinde et al., 2003a, p. 1) shows, or by choosing a word from the table of contents, as Figure 23 (Verlinde, Binon, et al., 2000, p. 68) displays. In the latter case, four items of information are provided: (1) the frequency of the word in the corpus is shown in the form of little stars, thus "****" means that it is high, "***" means that it is normal, "**" means that it is low and "*" means that it is very low; (2) grammatical information is given to the right of the word, for example, feminine noun ("nf"), adjective ("adj"); (3) reference to pages where the word appears is included (bold italics is used for pages where the word family is presented); and (4) collocations are listed under the headword and they are systematically organized: noun + adjective, noun + complement, verb + noun and derived noun + noun.

[125] It concurs with the idea promoted by Bergenholtz & Tarp (1995) of a specialized dictionary, since the authors suggest that the encyclopedic section should provide the user with an introduction to the subject field covered by the dictionary (pp. 155-156) (cf. §3.1.).

Figure 22

Extract From the Subject Fields of the DICOFE

Figure 23

Extract From the Alphabetical Table of Contents of the DICOFE

Odd and even pages of the DICOFE have different functions. On the one hand, the odd pages offer information about the world of business so that the user can learn about the economic reality and the language used in this domain. For so doing, it makes use of easy-to-read and straight-to-the-point sentences; an illustrating example is *Trafalgar Square est connu pour sa multitude d'enseignes lumineuses* in Figure 24 (Verlinde & Binon,

1994, p. 527). In addition, special attention is paid to groupings of words which have similar meanings, such as *la propriété de terres agricoles* and *le métayage* in Figure 25 (Verlinde et al., 2003a, p. 17), and to collocations, for example, *una unité de fabrication* and *complexe industriel* in Figure 26 (Verlinde, Binon, et al., 2000, p. 29).

Figure 24

Extract From an Odd Page of the DICOFE Showing Information About the Domain of Business

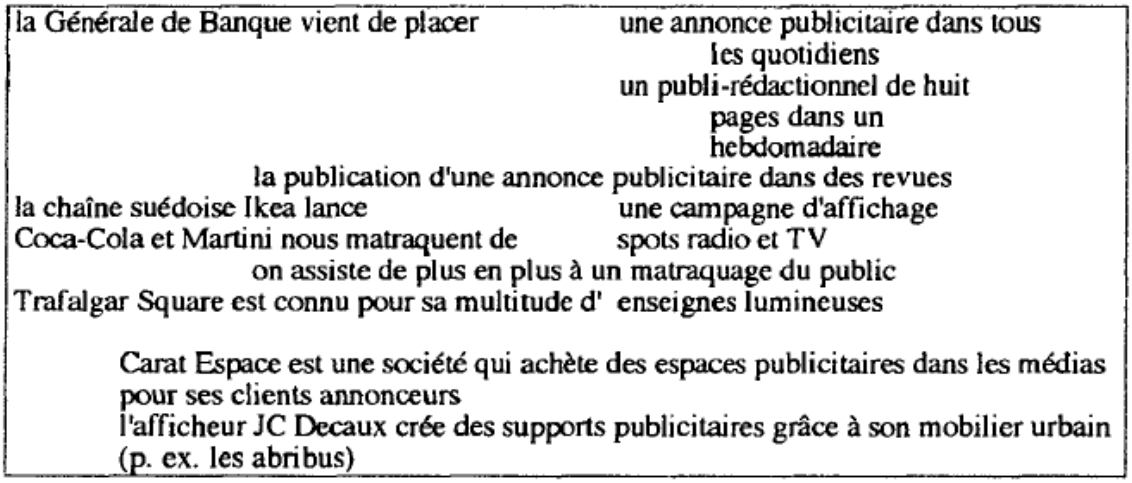


Figure 25

Extract From an Odd Page of the DICOFE Showing Synonyms

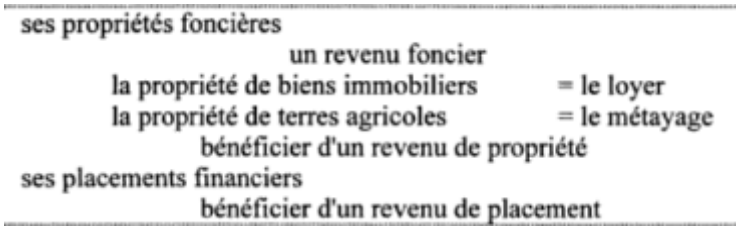
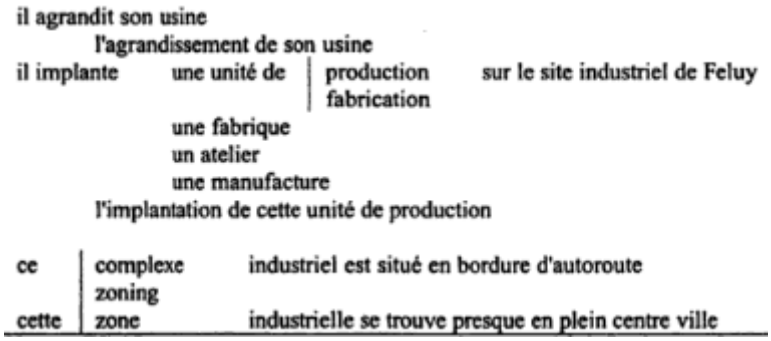


Figure 26

Extract From an Odd Page of the DICOFE Showing Collocations



On the other hand, the even pages deal with the derivatives that appear on the odd pages, including orthographical difficulties. In addition, they provide a definition of some members of the family which may pose a problem for learners, such as *une action* in Figure 27 (Verlinde & Binon, 1994, p. 528), along with example sentences and relevant word combinations that were excluded from the odd pages, for example, *(r)acheter des actions*:

Figure 27

Extract From an Even Page of the DICOFE

substantif/nom quoi	qui	verbe	adjectif *adverbe*
une action un actionnariat	un actionnaire	-	-

une action
1. Titre délivré par une société de capitaux qui donne à son détenteur la propriété d'une partie du capital avec tous les droits attachés à cette propriété. (...) Son revenu, appelé dividende, est touché par l'actionnaire (d'après Silem 1989:9).
1.1. un portefeuille d'actions
1.2. une action nominative <-> une action au porteur
1.3. émettre des actions — l'émission d'actions
coter une action en Bourse — la cotation (d'une action) en Bourse
négocier des actions:
céder des actions <-> (r)acheter des actions
la cession d'actions — le (r)achat d'actions

un actionnariat
1. L'ensemble des actionnaires

As we see it, a significant development took place in the field of business education and specialized lexicography with the publication of the DICOFE. According to Verlinde and

Binon (1994, p. 529), the most appealing aspect of the dictionary was "the detailed information of the combinatorics of words combined with the introduction to economics." Thus, collocations appear in three different parts of the dictionary: first, they are listed under the term (which is normally a noun) in the table of contents; second, they are included in the odd pages; and third, they can also appear in the even pages. In our opinion, the dictionary might be improved in four distinct ways: (1) by indicating the lexical type of the collocations when they are encoded in the odd pages; (2) by displaying some more examples of the collocations in context, which sounds easy as the project is corpus-based; (3) by providing their meaning when it cannot be inferred from the context; and (4) by making it available online so that it can reach a larger audience.

Le dictionnaire fondamental de l'informatique et de l'Internet (DiCoInfo)

The DiCoInfo[126] is the result of work carried out by the *Équipe de Recherché en Combinatoire Lexicale, Terminologie et Informatique* (ÉCLECTIK) of the Meaning-Text Linguistics Observatory (*Observatoire de Linguistique Sens-Texte*, OLST) of the University of Montreal, Canada. It mainly draws on the theoretical principles of the MTT (cf. §2.1.3.) and the ECD (cf. §3.2.2.) by Mel'čuk and its main objective is to provide "rich lexico-semantic information on terms" (L'Homme, 2010, p. 145) from various domains of computer science and Internet usage. For this reason, this dictionary relies on a corpus of 1 million words that comprises texts about the Internet, computer networks and programming, micro-computing and operating systems, most of them taken from didactic material (L'Homme, 2009b, p. 6).

The DiCoInfo can be described as a multilingual online specialized tool which provides detailed semantic, grammatical and phraseological data of English, French and Spanish terms related to computing and the Internet. It includes approximately 1,100 entries in French and 850 in English; the Spanish version is under development. Its compilers rely on NLP for the selection of terms, which are lemmata of four types: nouns, verbs, adjectives and adverbs, and can describe objects (e.g., *button*), human participants (e.g., *hacker*), representations (e.g., *data*), activities (e.g., *corrupt*), properties (e.g., *volatile*) and units of measurement (e.g., *megabyte*). It also contains proper names and

[126] An in-depth analysis of this resource is performed by L'Homme (2010).

acronyms represented as common nouns (e.g., *Internet*, *web*) (L'Homme, 2009b, pp. 3-4).

This specialized tool is considered innovative because it is the first computer science dictionary that deals with the linguistic behaviour of terms. In fact, it covers paradigmatic as well as syntagmatic lexical relations and provides a short description of the precise lexical relation between the headword and the related term. When using the dictionary, two different levels of details can be shown: (1) with technical metalanguage, and (2) without technical metalanguage. The former version can be useful to linguists, lexicographers and terminographers, since it offers information that they can apply to their own work. However, the version without metalanguage is user-friendlier and more recommended for people who have limited linguistic knowledge.

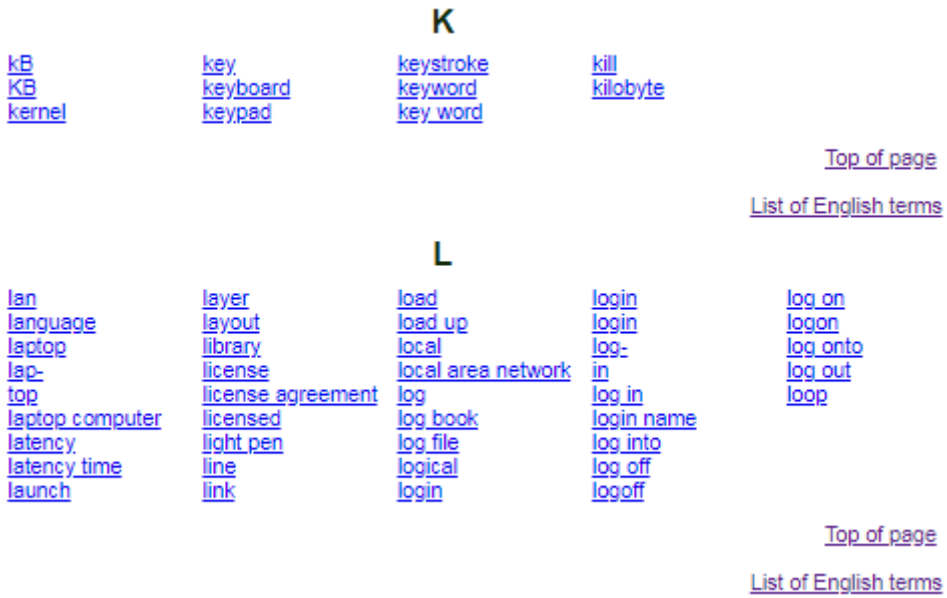
The DiCoInfo can be searched in two ways. First, there is a window which offers several advanced search options, such as 'Language,' 'Mode' and 'Precision,' as shown in Figure 28. Choosing one language will limit the results to that language, whereas choosing 'Trilingual' will return terms in any of the three languages. Regarding the search mode, it can be for isolated terms, lexical relations and expressions. Finally, the list generated can contain three types of results according to the search precision: (1) terms with exactly the same spelling as in the search box, (2) terms starting with the character sequence entered, and (3) terms containing that sequence. Second, a different way to search the dictionary is by using an alphabetical list of terms available in any of the three languages, which allows users a quick direct access to the description of terms, such as the alphabetical list of English terms displayed in Figure 29.

Figure 28

The DiCoInfo Search Parameters (English Version)

Figure 29

Extract From the Alphabetical List of English Terms of the DiCoInfo

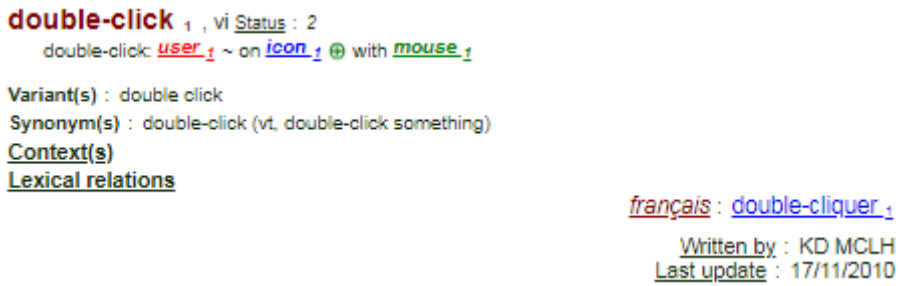


Regarding the entries, each lemma in the dictionary is associated with one or various LUs, given that, influenced by Mel'čuk's (2006, p. 231) MTT and ECD, each lexical entry

corresponds to one meaning (L'Homme, 2010, p. 147). For instance, Figure 30 displays the only entry for the verb *double-click*$_1$, in which most of the default categories can be observed: (1) the headword (MWUs can also appear as headwords if their meaning is non-compositional); (2) grammatical information (e.g., "vi," intransitive verb); (3) status, that is, degree of completion of the entry ("2" means that the entry is still being developed, "1" means that the entry is in an advanced stage of editing and "0" means that the entry is completed); (4) actantial structure, which describes the typical actants activated by the verb and their semantic role (further information can be obtained by clicking on each of the actants); (5) definition of the term (only in those whose state is 0); (6) spelling variants (e.g., the nonhyphenated version *double click*) and synonyms (e.g., *double-click* used as a transitive verb); (7) contexts, that is, corpus sentences including the source which they were extracted from; (8) lexical relations between the headword and other LUs held in the paradigmatic and syntagmatic axes, such as antonyms and collocations; (9) correspondences in the other languages (e.g., *double-cliquer* in French); and (10) additional information, more specifically, administrative data on the person who added this entry (e.g., "KD MCLH") and the date of the last update.

Figure 30

Entry for double-click$_1$ *in the DiCoInfo*

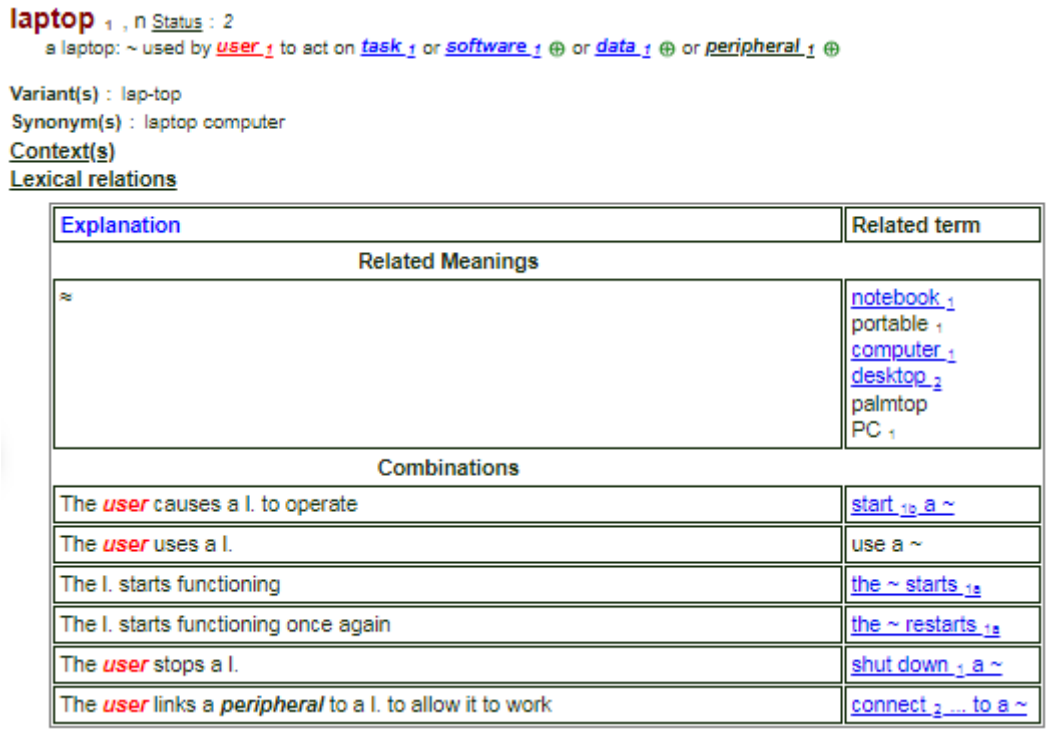


With respect to the representation of phraseology in the DiCoInfo, we have just mentioned that collocations are included in the 'Lexical relations' label. However, the DiCoInfo only encodes lexical collocations and nouns are considered the bases, therefore, collocations are normally included in noun entries (M.-C. L'Homme, personal communication, July 27, 2020). By way of example, Figure 31 shows the lexical relations of the entry for *laptop*$_1$. The information is offered in a table with several rows and two columns. First, semantically-related (on the paradigmatic axis) terms under the 'Related

Meanings' label are supplied, such as *notebook*$_1$, *desktop*$_2$ (terms which do not have a meaning number are not encoded in the dictionary). Second, collocations (relations on the syntagmatic axis) are listed under the 'Combinations' label and are accompanied by an explanation which uses different fonts and colors, for example, *start*$_{1b}$ *a laptop* is defined as "The ***user*** causes a l. to operate."

Figure 31

Lexical Relations of the Entry for laptop$_1$ *in the DiCoInfo*



laptop $_1$, n Status : 2
a laptop: ~ used by ***user*** $_1$ to act on ***task*** $_1$ or ***software*** $_1$ ⊕ or ***data*** $_1$ ⊕ or ***peripheral*** $_1$ ⊕

Variant(s) : lap-top
Synonym(s) : laptop computer
Context(s)
Lexical relations

Explanation	Related term
Related Meanings	
≈	notebook $_1$ portable $_1$ computer $_1$ desktop $_2$ palmtop PC $_1$
Combinations	
The ***user*** causes a l. to operate	start $_{1b}$ a ~
The ***user*** uses a l.	use a ~
The l. starts functioning	the ~ starts $_{1a}$
The l. starts functioning once again	the ~ restarts $_{1a}$
The ***user*** stops a l.	shut down $_1$ a ~
The ***user*** links a ***peripheral*** to a l. to allow it to work	connect $_2$ ... to a ~

Unlike the DICOFE, the DiCoInfo does not introduce the subject field covered by the dictionary, but, to our mind, the recent advances in technology make it unnecessary, as everyone can access this kind of information with a click by using the Internet. Regarding the representation of phraseology, the DiCoInfo offers a wide variety of combinatorial information, that is, paradigmatic and syntagmatic relations of headwords. More specifically, collocations come with an explanation that uses lay terms, which can make understanding easier for laypersons and people who do not have a background in linguistics. This is indeed the result of the profound influence of Mel'čuk's MTT and ECD (cf. §3.2.2.) on the dictionary. Nevertheless, despite being a reference tool, we believe that improvements in this area might follow three directions: (1) encoding

collocations in verb entries would result in a more comprehensive specialized dictionary, (2) incorporating corpus annotated sentences together with the collocations would help to better understand their use in context, and (3) giving correspondences for collocations in the other two languages would be an additional advantage. The consideration of these aspects, specially the last one, might accelerate the achievement of the ultimate goal set by the ÉCLECTIK, which is to convert the DiCoInfo into a learner's dictionary targeted at translation students (Alipour, Robichaud, & L'Homme, 2015, p. 51).

Le dictionnaire fondamental de l'environnement (DiCoEnviro)

The DiCoEnviro is an incipient online specialized dictionary related to the field of the environment, also developed by the ÉCLECTIK. For this reason, it falls under the influence of Mel'čuk's MTT and ECD. Unlike the previous resource, it covers a wider variety of languages, namely, French, English, Spanish, Italian, Portuguese and Chinese. The DiCoEnviro is based on a corpus of 2,500,000 words found in didactic, popular and scientific texts (L'Homme & Laneville, 2009, pp. 5-6) and currently contains more than 2,500 terms (over 1,300 in French, around 1,000 in English, 170 in Spanish, about 40 in Italian and Portuguese and 23 in Chinese). Like in the DiCoInfo, terms can be accessed in two ways: (1) through the query page (Figure 32), and (2) through an alphabetical list (Figure 33).

Figure 32

The DiCoEnviro Search Parameters (English Version)

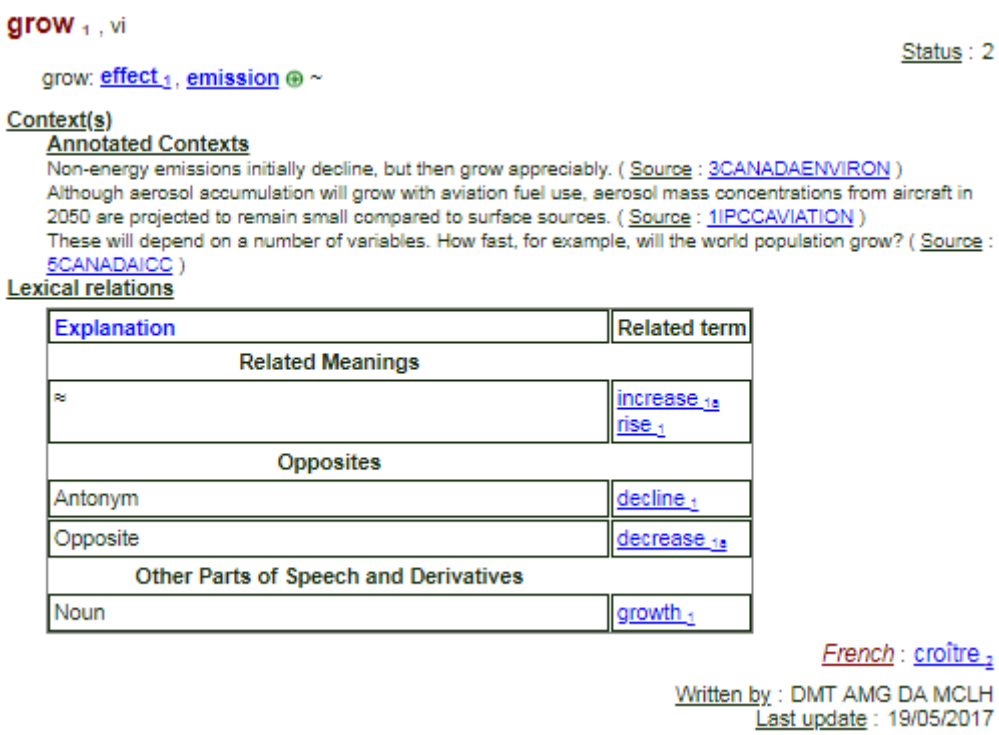
Figure 33

Extract From the Alphabetical List of English Terms of the DiCoEnviro

The kind of information displayed in the entries is the same as in the DiCoInfo. An illustrating example is the entry for the verb *grow*$_1$, shown in Figure 34. The headword is accompanied by grammatical information, that is, it is an intransitive verb ("vi"), and the entry does not include a definition because it is still under development ("Status: 2"). Regarding its actantial structure, it allows the user to know that *effect*$_1$ and *emission* can act as subjects of the verb. After that, corpus sentences (which are not annotated despite saying 'Annotated Contexts') and lexical relations (e.g., synonyms, antonyms and derivatives) of the term are displayed, although combinations of the headword with other LUs are not, for they are encoded in noun entries, such as in *efficiency*$_1$ in Figure 35 (like in the DiCoInfo, they are accompanied by an explanation in lay terms). In addition, a French equivalent (*croître*$_2$) is provided and, finally, administrative information (authors of the entry and date of the last update) completes the entry.

Figure 34

Entry for grow$_1$ *in the DiCoEnviro*



grow $_1$, vi

Status : 2

grow: effect $_1$, emission ⊕ ~

Context(s)

Annotated Contexts

Non-energy emissions initially decline, but then grow appreciably. (Source : 3CANADAENVIRON)
Although aerosol accumulation will grow with aviation fuel use, aerosol mass concentrations from aircraft in 2050 are projected to remain small compared to surface sources. (Source : 1IPCCAVIATION)
These will depend on a number of variables. How fast, for example, will the world population grow? (Source : 5CANADAICC)

Lexical relations

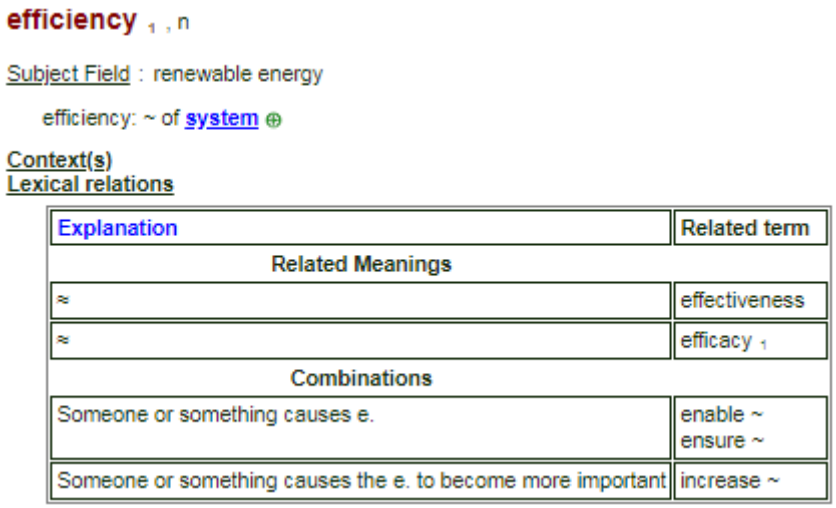
Explanation	Related term
Related Meanings	
≈	increase $_{1a}$ rise $_1$
Opposites	
Antonym	decline $_1$
Opposite	decrease $_{1a}$
Other Parts of Speech and Derivatives	
Noun	growth $_1$

French : croître $_2$

Written by : DMT AMG DA MCLH
Last update : 19/05/2017

Figure 35

Lexical Relations of the Entry for efficiency$_1$ *in the DiCoEnviro*



efficiency $_1$, n

Subject Field : renewable energy

efficiency: ~ of system ⊕

Context(s)
Lexical relations

Explanation	Related term
Related Meanings	
≈	effectiveness
≈	efficacy $_1$
Combinations	
Someone or something causes e.	enable ~ ensure ~
Someone or something causes the e. to become more important	increase ~

In general, one could say that the DiCoEnviro has the same advantages and disadvantages as the DiCoInfo, since its compilers belong to the same research team. As a result, the

entries are quite comprehensive and the dictionary can be of practical use to any type of user. Furthermore, a positive aspect is that collocations are accompanied by an explanation in lay terms. On the other hand, aspects that could be improved on are: (1) the inclusion of collocations in verb entries, (2) corpus annotated sentences together with the collocations, and (3) collocation equivalents in other languages. However, we understand that the DiCoEnviro is still young and further improvement might glance in these or other directions.

EcoLexicon: Terminological Knowledge Base on the Environment (EcoLexicon)

Another multilingual specialized resource dealing with the domain of the environment is the EcoLexicon, a terminological knowledge base developed by the *LexiCon Research Group* at the University of Granada, Spain (it was first mentioned in Section 1.3.4.). It hosts more than 24,000 terms and 4,400 concepts in a wide range of languages: English, Spanish, Greek, German, French, Arab, Russian and Dutch.[127] It also relies on corpus information and is targeted at different user groups, such as translators, environmental experts or students. The objective of the tool is to organize the contents according to different levels, from more general to more specific. For so doing, it follows the theoretical premises of the FBT (cf. §1.3.4.), consequently, each concept appears in the context of a specialized frame that highlights its relation to other concepts. The most generic level is the ENVIRONMENTAL_EVENT, which provides a frame for the organization of all concepts in the knowledge base.

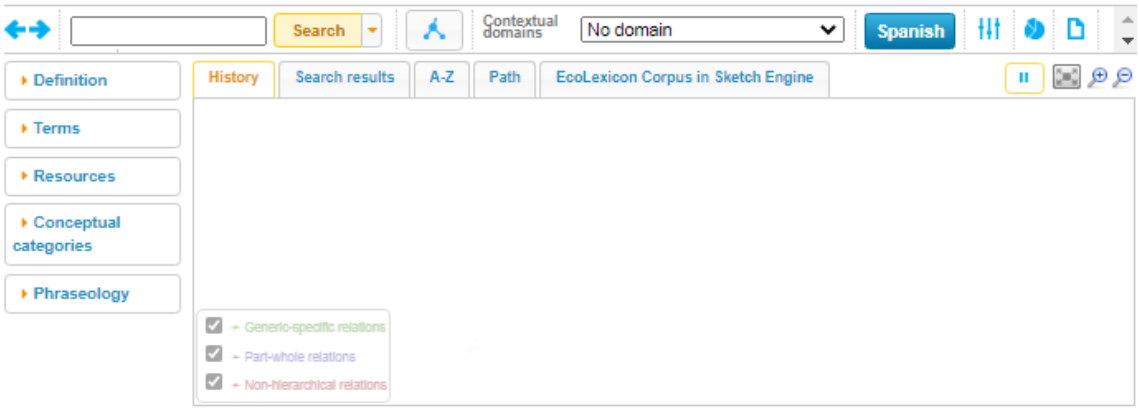
The search interface of the EcoLexicon is shown in Figure 36. As it can be observed, the data stored in this dictionary is more extensive. When a term is entered in the search box, several pieces of information are displayed: (1) definition; (2) equivalent terms in the other languages; (3) other resources, which can be of four different types: images, videos, web pages and documents; (4) conceptual categories with which the term is associated (they are hierarchically classified); and (5) phraseology, which displays the phraseological information about the headword. Furthermore, the user can choose the type/s of relation which he/she would like to see, that is, generic-specific, part-whole

[127] http://ecolexicon.ugr.es/visual/statistics.html (Last accessed: 25/10/2020).

and/or non-hierarchical relations. In addition to searching a term via the search box, the EcoLexicon also provides an alphabetical list of terms under the 'A-Z' label.

Figure 36

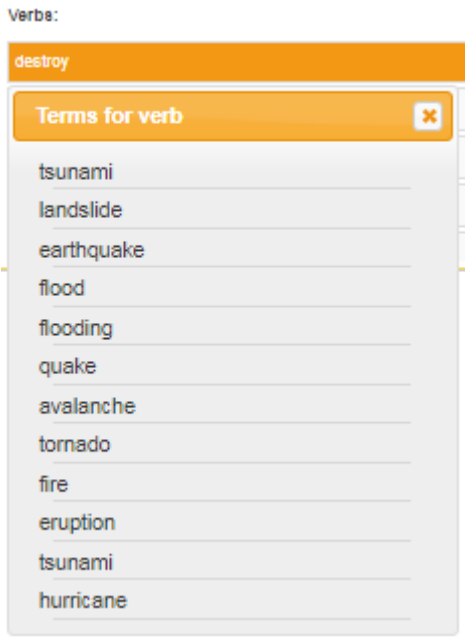
Search Interface of the EcoLexicon



With respect to the representation of phraseological information, it is encoded in noun entries, as the creators of the tool consider that the noun is the base and the verb is the collocate in the noun + verb and verb + noun patterns (Buendía Castro et al., 2014, p. 73; it chimes with the idea behind the DiCoInfo's and the DiCoEnviro's compilation). Thus, when a verb, for example, *destroy*, is entered in the search box, a list of nouns that collocate with the headword is provided, as it can be seen in Figure 37:

Figure 37

List of Terms That Collocate With the Verb destroy *in the EcoLexicon*



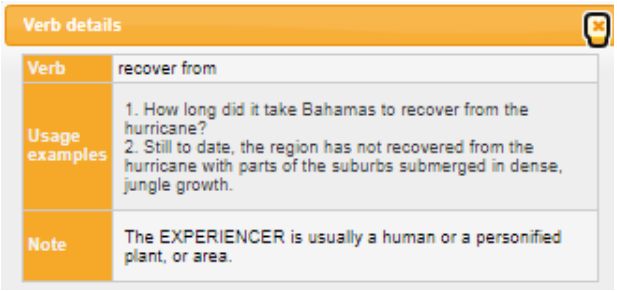
By clicking on the nouns, access to their phraseological entry is gained. As a way of example, an extract from the phraseological entry for the term *hurricane* is displayed in Figure 38. As it can be seen, verb collocates of the noun are grouped according to their nuclear meaning first, for example, EXISTENCE, and in terms of the frame activated within the lexical domain ('Meaning dimension') afterwards. Then, the phraseological pattern is described (the verb arguments are capitalized) and verbs are specified in a blue box. In the case of *hurricane*, two patterns are possible, that is, noun + verb, for example, followed by *happen* or *take place*, and verb + noun, for example, following *survive* or *recover from*. If the user is interested in examples of the collocation in context, a click on the verbs can take him/her to the verb details, which show usage examples and a brief explanatory note, as it is shown in Figure 39.

Figure 38

Extract From the Phraseological Entry for hurricane *in the EcoLexicon*

Figure 39

Verb Details of recover from *in the EcoLexicon*



Overall, it can be said that the EcoLexicon is a representative example of a comprehensive electronic dictionary for providing valuable pieces of information on terms, namely: (1) definition, (2) equivalents in other languages, (3) resources with extra information, (4) corpus sentences showing the collocations in context (incorporating annotated examples would be a remarkable feature), (5) explanatory notes on the collocations, and (6) description of phraseological patterns, among others.[128] Nonetheless, we agree with Reimerink and León Araúz (2017, p. 176) that EcoLexicon needs some improvements regarding phraseological representation. For instance, like in the DiCoInfo and the DiCoEnviro, collocations are accessible only via the noun, which is an inconvenient to

[128] For a more detailed description of the entries in the EcoLexicon, see Buendía Castro (2013), Buendía Castro et al. (2014) and Buendía Castro & Faber Benítez (2017).

users interested in verb collocations. A negative consequence is that information on verbs is not as accurate as that on nouns because they lack an entry, thus, their definition, equivalent terms in other languages, collocations and other sort of information are missing. Although collocations are not encoded in verb entries in the DiCoInfo and the DiCoEnviro either, these entries do offer the sort of information similar to the one found in noun entries. Finally, we think that the EcoLexicon's interface is not genuinely user-friendly and some linguistic knowledge is necessary to be able to use all the functions available.[129]

Interactive Terminology for Europe (IATE)

The last of the terminological resources analyzed in this section is the IATE, the shared multilingual terminology database of the institutions of the EU, supported by the European Parliament, the European Commission, the European Central Bank and the Translation Centre for the Bodies of the EU, to name but a few.[130] It was created in 2004 for the collection, dissemination and management of EU-specific terminology and was the result of importing some legacy databases that had existed since 1999, such as Eurodicautom, Euterpe and Euroterms. The current version was released in 2018 and shows a more visually attractive system implemented with state-of-the-art technologies and the latest software development standards. Its target user is not a specific one, but EU linguists, translators, staff and the public in general, since it covers any domain that has appeared in the EU's texts, for example, financial crisis, trade, agriculture, migration, which make the corpus that forms its basis. It takes a concept-based approach, like the DiCoInfo and the DiCoEnviro, which means that there is a separate entry for every concept. The IATE is regularly updated and information on the last build date appears in the bottom left corner of the screen.

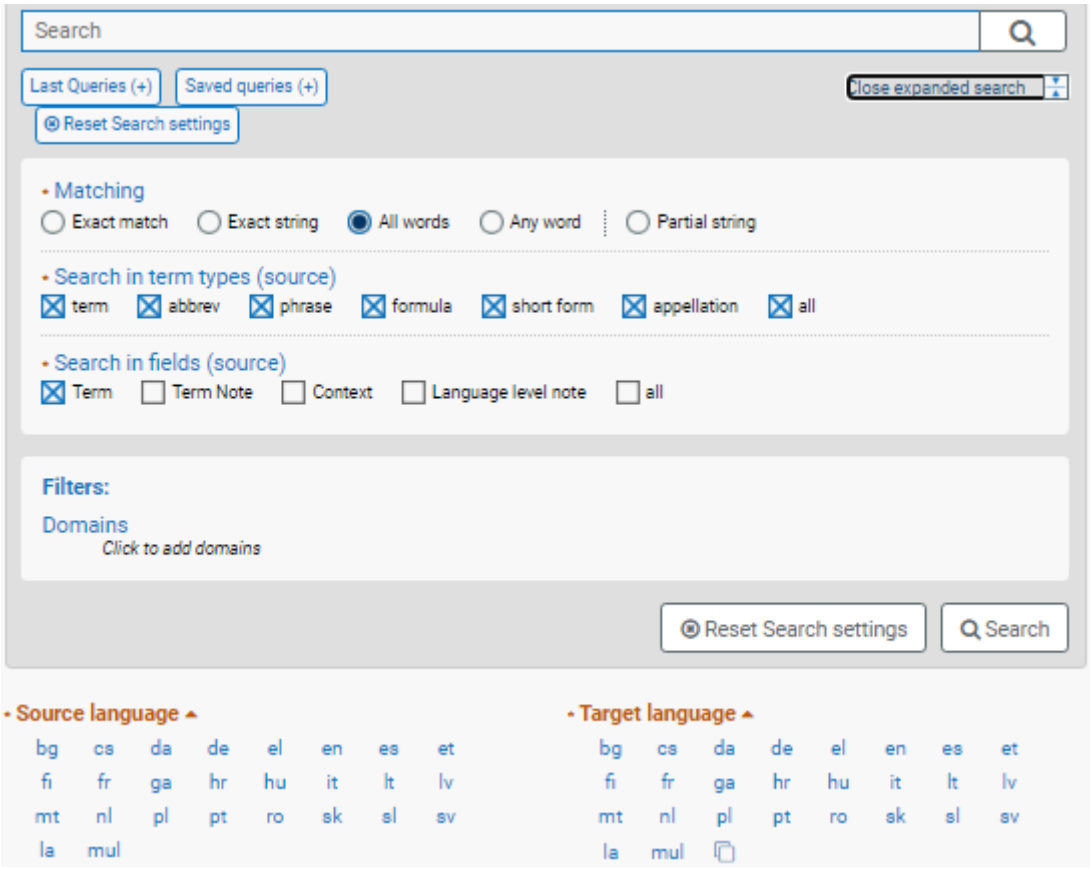
Searching IATE is not a difficult task thanks to the user's handbook which is freely available on the web.[131] The expanded search of the resource, displayed in Figure 40, includes four filters that allow users to personalize their search. First, the 'Matching'

[129] This perception chimes with L'Homme's (2014) idea that "lexicographers or linguists who have a strong background in lexical semantics [may] represent meanings and connections between words in unconventional formats (semantic networks, hierarchies, graphs, frames)" (p. 335).
[130] https://iate.europa.eu/about (Last accessed: 25/10/2020).
[131] https://iate.europa.eu/assets/IATE_Handbook_public.pdf (Last accessed: 25/10/2020).

of the term entered can produce five types of results: (1) exact sequence of characters, (2) exact sequence of characters within longer strings, (3) results containing all the words (this option is set by default), (4) results containing at least one of the words, and (5) part of the exact sequence of characters within longer strings. Second, the terms can be searched for according to six different types: the term itself, abbreviations, phrases, formulas, short forms and appellations (they are all selected by default). The third filter is related to the place where the term appears, such as in the headword, in the term note, in the context or in the language level notes (all of them can be chosen). Finally, the tool allows the user to filter the results by selecting only one or several specific domains. In addition, a source and a target language must be selected (for searches in one language, it must be selected in both the source and the target languages).

Figure 40

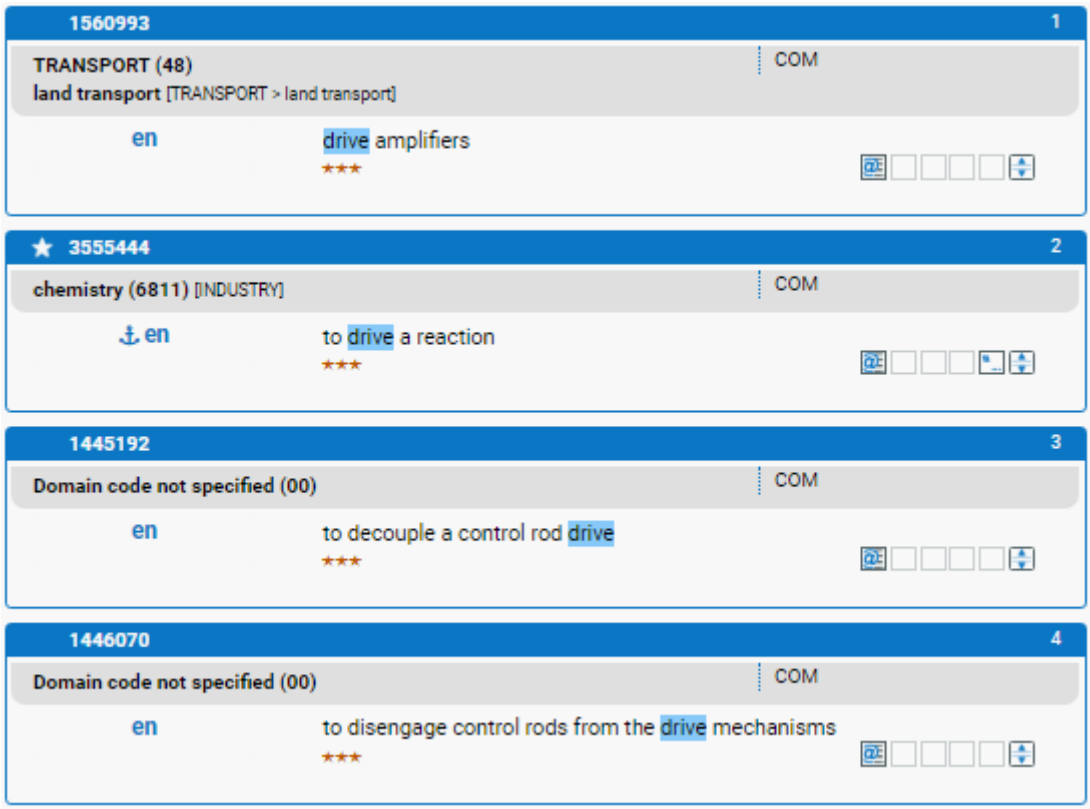
Expanded Search Interface of the IATE



Regarding phraseological information, it must be emphasized that the IATE is not especially interested in this area. To put it more simply, it aims to provide relevant data to support multilingual drafting of EU texts (legal texts in particular), so gathering PUs is not the designers' specific objective. Despite that, collocations can be obtained by selecting "Any word" from the matching label in the search interface, "phrase" from the term types and "Term" from the source. For example, Figure 41 displays an extract from the results achieved for the lemma *drive*. As we can see, the tool does not distinguish between different grammatical forms, therefore, collocations where *drive* is the adjective (e.g., *drive amplifiers*), the verb (e.g., *to drive a reaction*) and the noun (e.g., *a control rod drive*) must be deduced from the results. Entries also include other pieces of information: (1) the entry ID (e.g., "1560993"), (2) the domain and subdomain (e.g., "TRANSPORT," "land transport"), (3) an abbreviation of the language used for the search (e.g., "en" for English), (4) the phrase along with a number of stars ("***" means that the result is reliable, i.e., it has been thoroughly checked and manually assigned to the term, and "**" means that it is not entirely reliable, i.e., it has been automatically assigned to terms entered by native speakers and needs to be more carefully checked by the IATE terminologists), (5) the institution owner (e.g., "COM"), and (6) the reference of the phrase can be obtained by clicking on  .

Figure 41

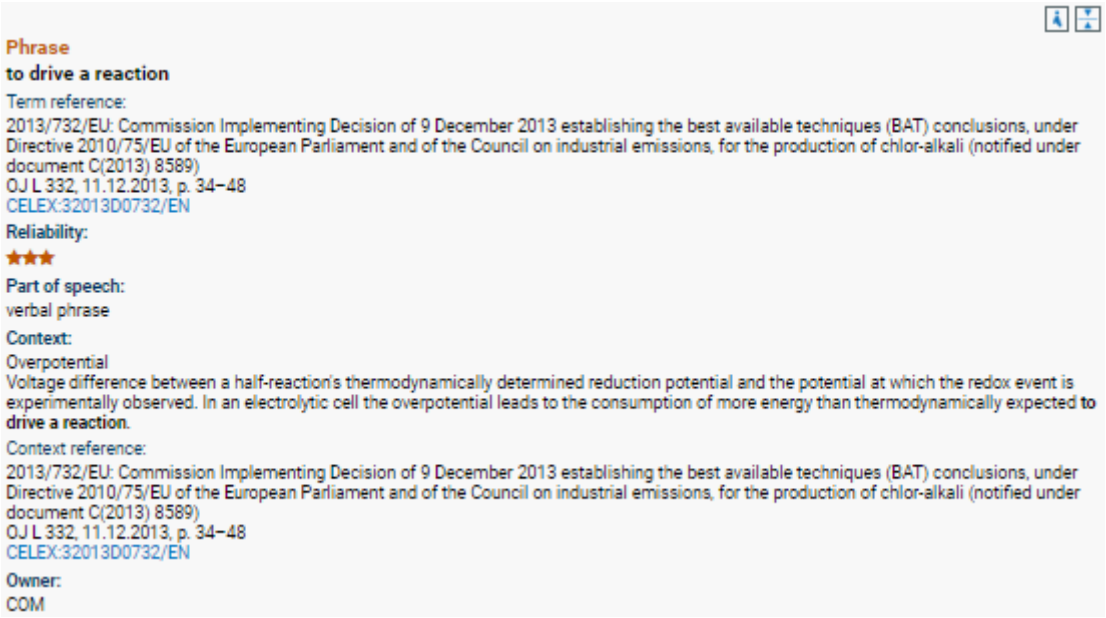
Entry for Phrases Containing the Lemma drive *in the IATE*



A key point is that some phrases contain more information, accessible by clicking on the abbreviation of the language explored, such as *to drive a reaction* in Figure 42. From this information, only the 'Context' seems useful to people interested in phraseology, as it shows a full sentence extracted from authentic language containing the collocation. The rest of the items look rather informative and are not expected to play a significant role on linguistic-oriented works.

Figure 42

Entry for to drive a reaction *in the IATE*



Phrase

to drive a reaction

Term reference:

2013/732/EU: Commission Implementing Decision of 9 December 2013 establishing the best available techniques (BAT) conclusions, under Directive 2010/75/EU of the European Parliament and of the Council on industrial emissions, for the production of chlor-alkali (notified under document C(2013) 8589)
OJ L 332, 11.12.2013, p. 34–48
CELEX:32013D0732/EN

Reliability:

★★★

Part of speech:

verbal phrase

Context:

Overpotential
Voltage difference between a half-reaction's thermodynamically determined reduction potential and the potential at which the redox event is experimentally observed. In an electrolytic cell the overpotential leads to the consumption of more energy than thermodynamically expected **to drive a reaction**.

Context reference:

2013/732/EU: Commission Implementing Decision of 9 December 2013 establishing the best available techniques (BAT) conclusions, under Directive 2010/75/EU of the European Parliament and of the Council on industrial emissions, for the production of chlor-alkali (notified under document C(2013) 8589)
OJ L 332, 11.12.2013, p. 34–48
CELEX:32013D0732/EN

Owner:

COM

To our mind, the IATE can be a valuable and reliable resource in terminology, for it is a huge database supplying a gold mine of information. However, it is not a resource oriented toward the study of phraseology, therefore, it does not offer direct access to any types of PUs and information on them is highly limited.

The conclusions reached after this analysis on the representation of collocations in terminological resources will focus on the DICOFE, the DiCoInfo, the DiCoEnviro and the EcoLexicon, as we believe the IATE contributes little to this research field. Regarding the DICOFE, the access to collocations via three ways (the alphabetical list, the odd pages and, in some cases, the even pages) is a positive aspect, although they are not clearly highlighted or classified and they lack examples in context. Apart from that, a major limitation is that an electronic version is not available. Moving on to the DiCoInfo and the DiCoEnviro (they can be treated as siblings), they are more comprehensive because they provide the user with collocations (as well as paradigmatic associations, e.g., synonyms and antonyms) accompanied by an explanatory sentence in lay terms. Nevertheless, they lack corpus annotated sentences and equivalents in other languages; in addition, they are encoded in noun entries, so verb collocations are not provided. With respect to the EcoLexicon, we think that it is a large multilingual database collecting interesting pieces of information, such as definitions, translations, concordance lines, and

so forth. However, it needs some improvements in the field of phraseology. For instance, we suggest the encoding of collocations in verb entries (verb entries would also need to be improved on), the inclusion of annotated corpus sentences and the representation of collocation equivalents in other languages. All things considered, the DiCoInfo, the DiCoEnviro and the EcoLexicon share the same drawbacks, but a main difference makes the first two user-friendlier, that is, their interface. Finally, this analysis demonstrates that in specialized resources, like in lexicographical resources (cf. §3.2.), it is common practice to encode collocations in noun entries rather than in verb entries.

3.4. Assessment of the representation of collocations in lexicographical and terminological resources

As it was mentioned at the beginning of this chapter, our aim was to perform an analysis on the representation of collocations in both general and specialized lexicography so as to identify the right decisions made by the compilers and the dictionaries' weaknesses. This analysis would allow us to propose a model for the implementation of these PUs in the verb entries of the specialized dictionary *DicoAdventure*. With this objective in mind, the following paragraphs will describe the main observations made and will conclude with a set of guidelines for a comprehensive specialized dictionary, in general, and an ideal phraseological entry, in particular, which we hope can be of great help to future terminographers.

The starting point of this analysis was a description of the items of information that GL and SL dictionaries must include, which was addressed in Section 3.1. In short, we explained that Hudson (1988) highlights eight aspects of a word that should appear in every GL dictionary entry, namely: (1) pronunciation and accentual pattern, (2) inflections and derivations, (3) syntactic information, (4) meaning as well as paradigmatic and syntagmatic relations, (5) contextual information, (6) spelling and spelling variations, (7) etymology, and (8) other aspects related to usage, such as frequency of use, age of acquisition and clichés or taboos. With respect to SL dictionaries, we stated that Bergenholtz and Tarp (1995) seem to be inspired by Hudson for their design of a model of a specialized dictionary, as they suggest some of the aspects proposed by the former, such as (1), (2), (3), (4), (5) and (8), and add two more: examples of use and illustrations. These similarities reveal that general and specialized lexicography are not so distant from

each other. Furthermore, a valuable idea shared by both authors is that collocations need to be encoded in dictionaries, but our analysis demonstrates that it has not become a matter of routine yet.

On the one hand, our description of collocational information in English general dictionaries implies that the "Big Five" (Paquot, 2015, p. 470) encode collocations in similar ways. Thus, some of these PUs are explicitly addressed under the definitions of *bottle* and other are implicit in the corpus sentences. However, the OALD, the CDE and the MED also include a separate section of collocations, so they may be more aware that these word associations need to be specified. In general, all of them seem to place PUs along a continuum and distinguish between more compositional (e.g., collocations) and less compositional (e.g., phrasal verbs and idioms) combinations; the latter normally include separate entries in the dictionary providing the user with a definition. A point in common is that the five dictionaries are corpus-based, so they supply information about current usage of language and examples in context. Additionally, the LDOCE uses frequency information to display the most common meanings of a word first.

On the other hand, our review of the representation of collocations in English collocation dictionaries (we will omit HASK given its rather statistical approach to the encoding of collocations) indicates a scale where the ECD is located at one end and the MCD is at the other. It may be the result of having presented these tools in chronological order, as it can be admitted that notable technological advances have been made and a deeper knowledge in the field of phraseology has been gained from the 1960s, when the ECD was launched, to the current days. Therefore, one could say that all the inconveniences found in the ECD, the BBI and the OCD are remedied by the MCD. For instance, the main drawbacks perceived in the first three dictionaries relate to: (1) a lack of systematic organization of the collocates, (2) a lack of semantic information about the verb in the different combinations and about the collocates within their groups, and (3) the scarcity of sentences showing the collocations in context. Nevertheless, there is still a problem that the MCD has not solved, which is the incorporation of nominal collocates into verb entries, as opposed to McGee's (2012) and Buendía Castro and Faber Benítez's (2014) conclusions.

The last scope of this analysis has covered specialized tools and their representation of collocational information. Our focus will be on the DiCoInfo, the

DiCoEnviro and the EcoLexicon, as we think the DICOFE and the IATE provide a limited coverage of phraseology. In our opinion, the fact that the other three are more comprehensive specialized dictionaries targeted at any type of users, ranging from experts to laypersons, may be because they are theory-based, so they aim to compile a formalized dictionary whose entries encode as much information as possible. Therefore, a distinctly positive aspect is that collocations are accompanied by an explanation which is comprehensible to all users. Nonetheless, we sustain that they still require some modest improvements to be a complete guide to specialized phraseology, such as annotated corpus sentences displaying the collocations in context, encoding collocations in verb entries and offering collocation equivalents in other languages.

Ultimately, we will endeavor to suggest a set of guidelines for a comprehensive specialized dictionary and collocational description in its entries. First of all, this tool should be electronic (freely available if possible) and with a user-friendly interface. Furthermore, it should be corpus-based, concept-based (i.e., each LU representing a specific meaning), targeted at any type of user (from laypersons to more specialists), conceived for almost any type of situation (either for encoding or decoding) and without any difficult metalanguage. Moreover, inspired by Hudson (1988) and Bergenholtz and Tarp (1995), every entry in this resource should include the following items: (1) definition of the headword, (2) pronunciation, (3) grammatical information (e.g., type of word, inflections and derivations), (4) word associations in the paradigmatic (e.g., synonyms and antonyms) and syntagmatic (e.g., collocations) axes, (5) annotated corpus sentences, (6) cross-references when necessary (e.g., in cases of synonymy), (7) equivalents in other languages if more than one language is covered in the dictionary, and (8) a useful illustration which helps the user visualize the object or the action. The inclusion of some of these items, such as the pronunciation and grammatical information on the headword, would prevent users from making a simultaneous use of a GL dictionary. To our mind, an introduction to the topic covered in the dictionary is not necessary, as rapid access to this kind of information can be gained by searching the web.

Moving on to the matter of greatest importance in this work, that is, collocational representation, the terminographer should decide first: (1) which types of collocations are to be encoded in the dictionary (e.g., lexical and/or grammatical), (2) which method will be used to retrieve and select them from the corpus and to organize them in the entries, (3) which place they will take in the microstructure (the internal structure of an entry) or

macrostructure (the list of the LUs entered in the dictionary) of the dictionary, and (4) which sorts of information will be included. Regarding (1), we obviously disagree with the tendency of encoding collocations only in noun entries and consider that they should also appear in verb entries, as in *DicoAdventure*. In this case, lexical collocations are more common, as grammatical collocations are regarded as *government patterns*, that is, the valency of the verb term. With respect to (2), we believe that using word sketches, like the MCD, can facilitate the extraction and selection of collocations, for it can help the terminographer save time and collect more accurate and detailed information. Since word sketches rely on statistical measures dealing with the significance of the collocations, their score may be used to decide which collocations will be included (most of the times it is necessary to set a threshold, given that the amount of data produced can be huge). Additionally, they may be arranged in descending order in terms of their score, that is, displaying strongest collocations first (in fact, frequency is the basis of the arrangement of word's meanings in the LDOCE). Turning to (3), it is a tendency to offer access to collocations via the base, in whose entry they should be explicitly represented (vs. having to be inferred from the context). Last but not least, (4) is summarized in the following lines:[132]

1) The lexical type of the collocation should be identified, for example, verb + noun, verb + adverb, noun + verb;

2) An explanation of the collocation in lay terms should be offered, as different collocates can add different nuances of meaning to the base. If they are ordered in terms of relevance, meaning correspondences between collocates can be explained;

3) Corpus annotated sentences showing the collocation in context should be provided next to the collocation;

4) In the case of bilingual and multilingual dictionaries, equivalents of the collocations in the other/s language/s should be supplied.[133]

As we see it, the analysis performed in this chapter, focused on the contents of GL and SL dictionaries and the representation of collocations in these tools, has eventually

[132] These items of information are highly reminiscent of the ones proposed by L'Homme (1998, pp. 76-81) for a lexicographical description and representation of specialized verbs.
[133] This idea is promoted by Vargas Sierra (2010, p. 37), who proposes a list of categories that should be included in any bilingual specialized combinatory dictionary.

led to the development of the aforementioned guidelines, which can offer real help to future SL dictionary terminographers. Personally, they will be applied to the implementation of collocations in the verb entries in *DicoAdventure*. The methodology employed for the extraction, selection and arrangement of these collocations will be explained in the following chapter, as well as the template used to this end. However, an introduction to the SL of adventure tourism, covered in *DicoAdventure*, and the description of the ADVENCOR corpus, the corpus used in this dissertation, will be given first.

PART II:

METHODOLOGY AND EMPIRICAL ANALYSIS

CHAPTER 4. MATERIALS AND METHODS

A study may claim path-breaking results, but if the methodology used to achieve those results is vague in its details, the scientific community would not readily accept it.

(Gupta, 2018, p. 142)

The present chapter encompasses five broad sections which describe the materials used in this study and the method employed in order to achieve our practical objectives. First, Section 4.1. will explore the specialized language of tourism in general in terms of three components: linguistic, pragmatic and functional, and then it will focus on the description of the segment of adventure tourism and some of the features of the language used in this subdomain (§4.1.1.). After that, Section 4.2. will present *DicoAdventure*, the online dictionary of adventure tourism which aims to include verb collocations. Section 4.3. will define what a corpus is and will cover different aspects of corpora, such as types (§4.3.1.), design features (§4.3.2.) and compilation (§4.3.3.). This section will be closed with the analysis of the specialized corpus used in this research, that is, the ADVENCOR corpus (§4.3.4.). Next, Section 4.4. will deal with the computer software which has made this investigation possible, *Sketch Engine*, and will explain the methodological steps taken for the achievement of our objectives (§4.4.1.–§4.4.3.). Furthermore, it will assess two other well-known corpus management tools, *TermoStat Web 3.0* and *WordSmith Tools 7.0* (§4.4.4.). The last section of this chapter (§4.5.) will suggest a template for the implementation of lexical collocations in the verb entries of *DicoAdventure*.

4.1. The specialized language of tourism

Via static and moving pictures, written texts and audio-visual offerings, the language of tourism attempts to persuade, lure, woo and seduce millions of human beings, and, in so doing, convert them from potential into actual clients.

(Dann, 1996, p. 2)

As it was stated in Chapter 1 (cf. §1.1.), SLs intersect with the GL because they make use of part of the latter (e.g., some grammatical constructions and words), but at the same time they are characterized by the presence of distinguishing features, such as special technical terms and linguistic structures. An illustrative example is the SL of tourism,[134] which exploits strategies (e.g., metaphorization) to convey specialized meanings through GL words and uses certain syntactic structures more frequently than in the GL (Maci, 2018, pp. 29, 31). However, there is not an only type of SL of tourism, given that it can be found in three different textual genres (Calvi, 2010, pp. 18-19): (1) texts describing the theoretical knowledge of the phenomenon of tourism and its characteristics, which can be addressed from an economic, geographical, sociological or anthropological perspective, among others; this type uses the language in a more scientific and interdisciplinary way; (2) texts dealing with the management of distinct activities, such as travel agencies, catering, means of transport, travel insurances, conferences, and so on; a big part of the vocabulary used is specific; and (3) texts concerning with the description and promotion of tourist destinations; the terminology employed comes from multiple areas and disciplines, such as geography, history, art or gastronomy, although special uses are made.

Without a doubt, the third group of texts is the one which gives rise to a more original use of language. In fact, most authors seem to agree that the SL of tourism is a

[134] The World Tourism Organization (UNWTO) defines *tourism* as follows: "Tourism is a social, cultural and economic phenomenon which entails the movement of people to countries or places outside their usual environment for personal or business/professional purposes. These people are called visitors (which may be either tourists or excursionists; residents or non-residents) and tourism has to do with their activities, some of which involve tourism expenditure" [https://www.unwto.org/glossary-tourism-terms] (Last accessed: 25/10/2020).

language for promotion (Dann, 1996; Federici, 2018; Mănescu, 2020; Nigro, 2006), for it tries to persuade, attract and seduce potential customers/buyers by using special words, images and symbols. It is normally found in advertisements, travel catalogues, hotel brochures and web sites, and is within everyone's reach thanks to the extraordinary development of information technology, which has also contributed to the creation of both new audiences and new texts (Maci, 2018, p. 39). For this reason, much of the vocabulary employed is comprehensible to laypersons. Regarding its distinctive features, they can be divided into three categories: linguistic, pragmatic and functional (cf. §1.1.1.),[135] which can permit the establishment of effective communication when properly combined (Ruiz-Garrido & Saorín-Iborra, 2013, p. 1).

First, the linguistic component of this SL[136] comes in many forms, being the terminology used its identifying characteristic. Terms may range from more (e.g., *rack rate, inbound operator*) to less (e.g., *check-in, booking*) specialized and are monoreferential and transparent; the use of short forms (e.g., acronyms) is encouraged so that the reader can achieve maximal specificity with minimal effort. Regarding keywords, they are informative and carefully chosen in order to draw the customer's attention, such as evaluative adjectives with positive connotations (e.g., *amazing, enjoyable*) and superlatives (e.g., *largest, finest*) (Durán-Muñoz, in press). Furthermore, it makes a special use of personal pronouns, modal verbs, future tenses and imperative forms to attract prospective tourists, for example, *So **book** now and **strap yourself in** for some of the most fun **you will have had** all year*.[137] On the other hand, not only is verbal language employed, but visual language is also common. Pictures may be positioned between the headline and the text with the intention of selling products, since they help the reader to visualize a perfect holiday; for this reason, colorful images are the most suitable.

Second, the pragmatic component of the SL of tourism emerges from the diverse communicative situations that can develop: (1) they can happen between tourism professionals at different locations, such as travel agencies, fairs and professional meetings; (2) communication can be established between an expert and a non-professional person (e.g., a tourist) in the form of a conversation in a travel agency or a

[135] See Durán-Muñoz (2011, pp. 491-508) for a deeper analysis.
[136] Dann (1996), Calvi (2010), Cappelli (2006), Gotti (2006) and Maci (2018) carry out a more extensive exploration of the topic.
[137] Example retrieved from the ADVENCOR corpus.

hotel; and (3) communication between non-specialists is becoming more common, given that those who were addressees are becoming addressers and use blogs and forums to provide information on their trips. Depending on the communicative situation that arises, the linguistic elements can vary. For instance, the vocabulary used in the first situation tends to be more specialized, whereas messages in the last situation are likely to include subjective descriptions and tend to be related with organizational aspects of trips, such as transport, insurance, currencies, among others (Calvi & Bonomi, 2008, p. 188).

Third, regarding the functional component (cf. §1.1.1.3.) of tourism texts, they are referential and convey denotative meanings, given that they endeavor to transmit accurate and precise information. However, the emotive function is also involved, especially in contexts where the addresser offers his/her personal opinion in his/her attempt to help the addressee decide. Accordingly, the conative function emerges, since the addresser's own opinion can induce the addressee "to respond in a particular way" (Nord, 2006, p. 138), after giving recommendations and advice or persuading. What is more, this function and the poetic function go hand in hand, as the form of the message is highly significant to convince tourists to choose a particular destination. Therefore, stylistic devices, such as alliteration, rhyme and metaphors, are common in these texts, as well as the use of imperative forms and modal verbs (e.g., *must*). In a word, it can be stated that the SL of tourism is multifunctional.

Finally, the "multifaceted character" (Mănescu, 2020, p. 223) of the SL of tourism has given rise to a variety of linguistic studies on different tourism segments (cf. §2.3.). For example, Corbacho Sánchez (2005) explores terminology related with trips in German and Spanish; Manca (2008) examines the role of qualifying adjectives in the description of farmhouse holidays in English and Italian; Raţă and Petroman (2012) and Raţă, Petroman, and Petroman (2012) deal with English borrowings in Romanian sites of travel agencies and Romanian agritourism; Fuster-Márquez (2014, 2017) and Fuster-Márquez and Pennock-Speck (2015) analyze the language used in British and American hotel websites; Baynat Monreal (2017) focuses on the information contained in web pages for private lodgings and accommodations, and Piccioni and Pontrandolfo (2019) centre on metaphors used in several sources, such as guides, web pages, blogs and contracts. In this work, we are keenly interested in the segment of adventure tourism, which will be described in the following section.

4.1.1. Description of adventure tourism

As it has been previously mentioned, the language of tourism is characterized for being used for promotion of tourist destinations, which will be chosen depending on the tourist's motivations for travelling. For instance, Lea (1988, pp. 24-25) notices four types of motivation behind travelers' decisions, namely, physical, cultural, personal and prestige. Consequently, some traditional market segments of tourism are: (1) health tourism, whereby travelers organize a trip outside their local environment with the intention of maintaining and enhancing their wellbeing in mind and body; (2) cultural tourism, whereby people are motivated by a deep desire for visiting local attractions, meeting new people and observing ancient customs; (3) religious tourism, which is conceived as the activity of going on journeys to places of worship (Trono, 2015, p. 21), also called pilgrimages; and (4) business tourism, which is concerned with people whose principal purpose is related to their job.

Nevertheless, these traditional segments are being superseded by new types of tourism that aim to allure travelers yearning for a sustainable and flexible kind of tourism. An excellent example is adventure tourism, a sort of *alternative tourism*[138] that strays from the masses and takes care of the environment, which may be the reasons for it gaining a pre-eminent position. Moreover, it has recently become a significant component of the mainstream tourism sector because it has developed new products to satisfy travelers that pursue alternative ways of enjoyment in unusual, exotic, remote or unconventional destinations. Furthermore, it offers physical outdoor activities for everyone, since they range from "soft" to "hard" (Page, 2007, pp. 108-109), which usually encompass high levels of risk, adrenaline rushes, excitement and personal challenge. Thus, unexperienced tourists can opt for soft activities, which entail low risk and little knowledge, are less physically demanding and are perceived to be safer (e.g., bicycle touring, wildlife watching), and braver tourists can choose hard experiences, since they entail high risk and are more physically and mentally demanding (e.g., bungee jumping, whitewater rafting).

[138] According to Fennell (2003), *alternative tourism* "is a generic term that encompasses a whole range of tourism strategies (e.g., 'appropriate,' 'eco-,' 'soft,' 'responsible,' 'people to people,' 'controlled,' 'small-scale,' 'cottage' and 'green' tourism), all of which purport to offer a more benign alternative to conventional mass tourism in certain types of destinations" (p. 5).

Other terms used both in the literature and in the industry to refer to adventure tourism are "adventure travel," "adventure recreation" and "hard and soft adventure" (Swarbrooke, Beard, Leckie, & Pomfret, 2003, p. 4). However, regardless of the name given, most experts agree on the description of this kind of tourism, which incorporates danger, risk, challenge, novelty, excitement, exploration and discovery. Added to that, it is referred to as active tourism because it is engaging and entails action, that is, it is not a passive experience but requires an active participation with one's body. To this respect, Cloke and Perkins (1998) point out: "adventure tourism is fundamentally about active recreational participation, and it demands new metaphors based more on 'being, doing, touching *and* seeing' rather than just 'seeing'" (p. 189). Therefore, an adventure tourist must channel his/her efforts into a particular activity which may require some mental and/or physical preparation or training. Some of the definitions given for adventure tourism are:

1) "[it is] the sum of the phenomena and relationships arising from the interactions of adventure touristic activities with the natural environment away from the participant's usual residence area and containing elements of risk in which the outcome is influenced by the participation, setting, and the organizer of the tourist's experience" (Sung, 1996, as cited in Sung, Morrison, & O'Leary, 1996, p. 53);

2) "[it is] any activity trip close to nature that is undertaken by someone who departs from known surroundings to encounter unfamiliar places and people, with the purpose of exploration, study, business, communication, recreation, sport, or sightseeing and tourism" (Addison, 1999, p. 417);

3) "[it is] characterized by its ability to provide the tourist with relatively high levels of sensory stimulation, usually achieved by including physically challenging experiential components with the (typically short) tourist experience" (Muller & Cleaver, 2000, p. 156);

4) "[it] involves travel and leisure activities that are contracted into in the hope that they will produce a rewarding adventure experience. An adventure experience will be of a heightened nature and involve a range of emotions, of which excitement will be key. It will entail intellectual, spiritual, physical or emotional risks and challenges. The *vehicle* or *product* that encompasses the

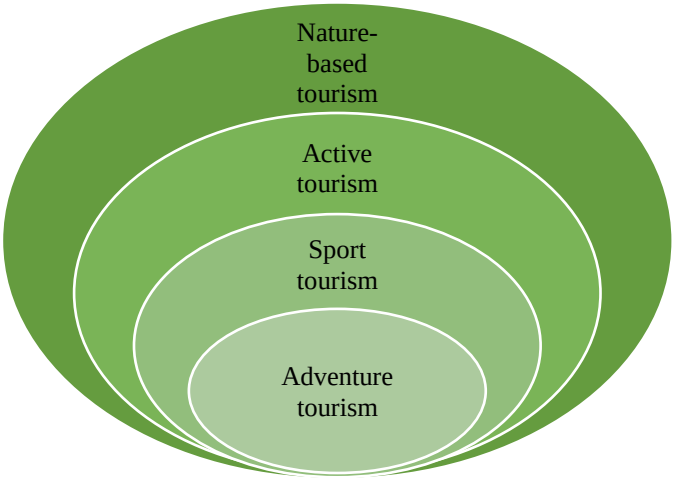
adventure tourism experience will be constructed from the basic constituents of a tourism experience –environmental setting, core activities and transportation– and some or all of these may contribute the stimulus for adventure" (Swarbrooke et al., 2003, p. 27); and

5) "[it] is a broad term which encompasses all types of commercial outdoor tourism and recreation with a significant element of excitement" (Buckley, 2010, p. 4).

In short, adventure tourism contemplates the participation in mental activities and physical exercises which may require a level of fitness, such as in sport. Thus, the tourist has an active role and is responsible for making this activity a real adventure experience, given that a higher degree of involvement will produce the desired outcome, which usually arouses a feeling of excitement. The major characteristic of these activities is that they are done outdoors, so the tourist is in contact with nature, which provides a perfect setting for self-development activities, making it a type of sustainable tourism. Additionally, they usually contain elements of risk, which enhance life experiences, and tend to be short, as some degree of physical effort is required when undertaking the activities. According to these aspects, Durán-Muñoz (2011, p. 464) proposes the conceptual frame of adventure tourism represented in Figure 43, that is, its inclusion in sport, active and nature-based tourism:

Figure 43

Conceptual Frame of Adventure Tourism (Durán-Muñoz, 2011)



Given that the central feature of adventure activities is that they are nature-based, it is not strange that "adventure tourism has a well-defined geography" (Buckley, 2010, p. 28). To put it differently, the geographical characteristics of a place are the determining factor when it comes to offering a list of activities to attract tourists. In fact, some authors have examined the strengths and weaknesses of specific countries and areas to this respect, for instance: Nepal may be famous for river-based activities, such as whitewater rafting (Zurich, 1992); South Africa may offer water-based activities, for instance, diving and snorkeling (Rogerson, 2007); the polar regions may be frequented by tourists interested in mountain-based activities and expedition cruises (Snyder & Stonehouse, 2007), and the states of Australia may draw tourists who like raft trips on rivers or skiing in the Australian Alps (Buckley, 2010). Considering this, geography seems a practical aspect to a taxonomy of adventure activities, thus, many experts have categorized these activities according to the place where they are undertaken, that is, land, water or air (Durán-Muñoz, 2011, in press; Olivera Betrán & Olivera Betrán, 1995; Peñalver Torres, 2004), as it is shown in Table 16:

Table 16

Taxonomy of Adventure Tourism Activities According to the Place Where They Are Undertaken

Land-based	Water-based	Air-based
Abseiling	(White-water) Canoeing	Ballooning
Archery	Canyoning / Canyoneering	Base jumping
Backpacking	Caving	Bungee jumping
Birdwatching	Coastal safari	Free falling
Camping	Cruising	Gliding
Caving	Diving	Paragliding
(Ice/Rock) Climbing	(Ice) Fishing	Skydiving
Cycling	Glass-bottom boat rides	Space travel
Dogsledding	Jet-skiing	Zip-lining / Zip-wiring
Hiking	(White-water) Kayaking	
Horseback riding	Paddling	
Hunting	Parasailing	
Karting	Powerboating	
Mountaineering	(White-water) Rafting	
Mountain biking	Sailing	
Orienteering	Scuba-diving	
Overlanding	Snorkelling	
Quad biking	Surfing	
Rappelling	Swimming	
Safari	Underwater photography	
Skiing	Wakeboarding	
Snowboarding	Water-skiing	
Snow-shoeing	Windsurfing	
(Desert) Trekking		
Wildlife spotting		

As we can see, land-based and water-based adventure tourism activities are more numerous than air-based activities. It goes without saying that all the terms naming these activities belong to the terminology used in the SL of adventure tourism, which is also characterized by the linguistic, pragmatic and functional features described in the previous section. Nevertheless, it has some core components which make it different from other tourism segments, such as the sun-and-beach or health-and-beauty segments. For

instance, Durán-Muñoz (2013) analyzes a corpus of over 250,000 words of Spanish promotional texts and finds out that the main terms extracted can be classified into four semantic categories, namely: (1) people organizing, guiding or doing the adventure activities; (2) activities to be done; (3) place where they are done; and (4) safety equipment required to undertake the activities, such as clothes and shoes, vehicles and safety and other types of material. In addition, she discovers that the corpus' terminology largely consists of: (1) terms designating adventure sports (e.g., *barranquismo, espeleología*) and equipment that is needed, (2) foreign words (e.g., *rafting, team building*), (3) neologisms (e.g., *zorbing, aquaseiling*), and (4) polysemous words (e.g., *ala delta*, which refers to the vehicle used in the activity and to the activity itself). In addition, the coexistence of Spanish terms and borrowings promotes the incorrect use of terms as if they were synonyms, such as *senderismo* and *trekking*.[139]

On the other hand, the language used in English promotional texts of adventure tourism has also been examined. First, Durán-Muñoz (2019) sheds new light on the importance of adjectives in this domain, since she measures their keyness by comparing their usage in a specialized corpus to two of the most well-known reference corpora, that is, the COCA[140] and the BNC. Her findings are surprising because the frequency of 50% of the adjectives analyzed is not significantly high compared to the results obtained from the reference corpora, whereas the frequency of the remaining half of the adjectives is. Thus, the terms under investigation can be categorized as descriptive (e.g., *aerial, complimentary*) and evaluative (e.g., *lovely, pleasant*), although the latter type is more common (61%). Additionally, superlative forms are slightly more frequent in the specialized corpus, such as *of the most* + adjective and *the best*. In short, the author concludes that adjectives which aim to persuade the reader or the hearer because they contribute to the creation of mental representations of destinations (or services) are "a key part of the lexico-semantic characterization" (p. 365) of this SL, for instance, *magnificent, stunning, pristine, towering*.

Second, Durán-Muñoz and L'Homme (2020) rely on Fillmore's Frame Semantics and the FN project (cf. §1.5.1.1.) to perform a lexico-semantic analysis of motion verbs

[139] According to the author, *senderismo* involves a long walk in which the tourist faces low levels of difficulty and is shorter than one day, whereas *trekking* entails a more strenuous walk which is normally longer than one day and for which the tourist must stay overnight.
[140] *Corpus of Contemporary American English*.

in the SL of adventure tourism in order to "show how these verbs and their participants are connected to a structure that could correspond to the representation of knowledge in this specialized field" (p. 54). With this objective in mind, they first select 40 motion verbs and then develop a method to annotate the contexts extracted from a specialized corpus of English promotional texts and define the argument structures of the verbs. This procedure is explained in terms of the fulfilment of two criteria: (1) the motion verbs selected describe the displacement of tourists as active participants, that is, verbs in which the first argument was an inanimate object (e.g., *trail*) were discarded, and (2) these motion verbs occur in at least 20 different contexts in the specialized corpus. The authors acknowledge that their study on motion verbs has proved to be a perfect example to explain how knowledge is expressed in the field of adventure tourism. The results of this study are implemented in the *DicoAdventure* database, which is the focus of the next section.

4.2. *DicoAdventure. An online dictionary of adventure tourism*

DicoAdventure is a specialized terminological database on adventure tourism freely available at http://olst.ling.umontreal.ca/dicoadventure/ and still under construction. In fact, this is the terminological resource in which collocations will be implemented after the research conducted in this work. It is conceived for both encoding and decoding and is specifically targeted at professional translators, albeit any type of user who linguistically interacts in the domain of tourism, such as interpreters or tourist guides, can benefit from this resource. The underlying motivations for developing this product were two: first, the terminology used in the field of adventure tourism is varied and rich, and second, no tools dealing with this type of SL exist so far (Durán-Muñoz, in press). The collection of information gathered in this resource is based on the study conducted by Durán-Muñoz (2012), whereby an entry would include: a definition, equivalents, derivatives and compounds, subject field, real examples of use, phraseological information, abbreviations and acronyms. On the other hand, the design of the product relies on the lexico-semantic analysis carried out by Durán-Muñoz and L'Homme (2020). Thus, the entries in *DicoAdventure* provide a description of the lexico-semantic properties of motion verbs (currently, these are the only entries encoded) by means of different labels

and sorts of information which make understanding easy for those who are non-experts in this specialized field.

Regarding the macrostructure of the dictionary, the entries can be accessed through an alphabetical entry list located at the top of the interface (Figure 44). However, hyperlinks included in the different entries also provide access to others. Added to that, other search options are being implemented at the moment, such as a search box, but it will be available when more entries are incorporated (Durán-Muñoz, in press).

Figure 44

Alphabetical List of Terms of DicoAdventure

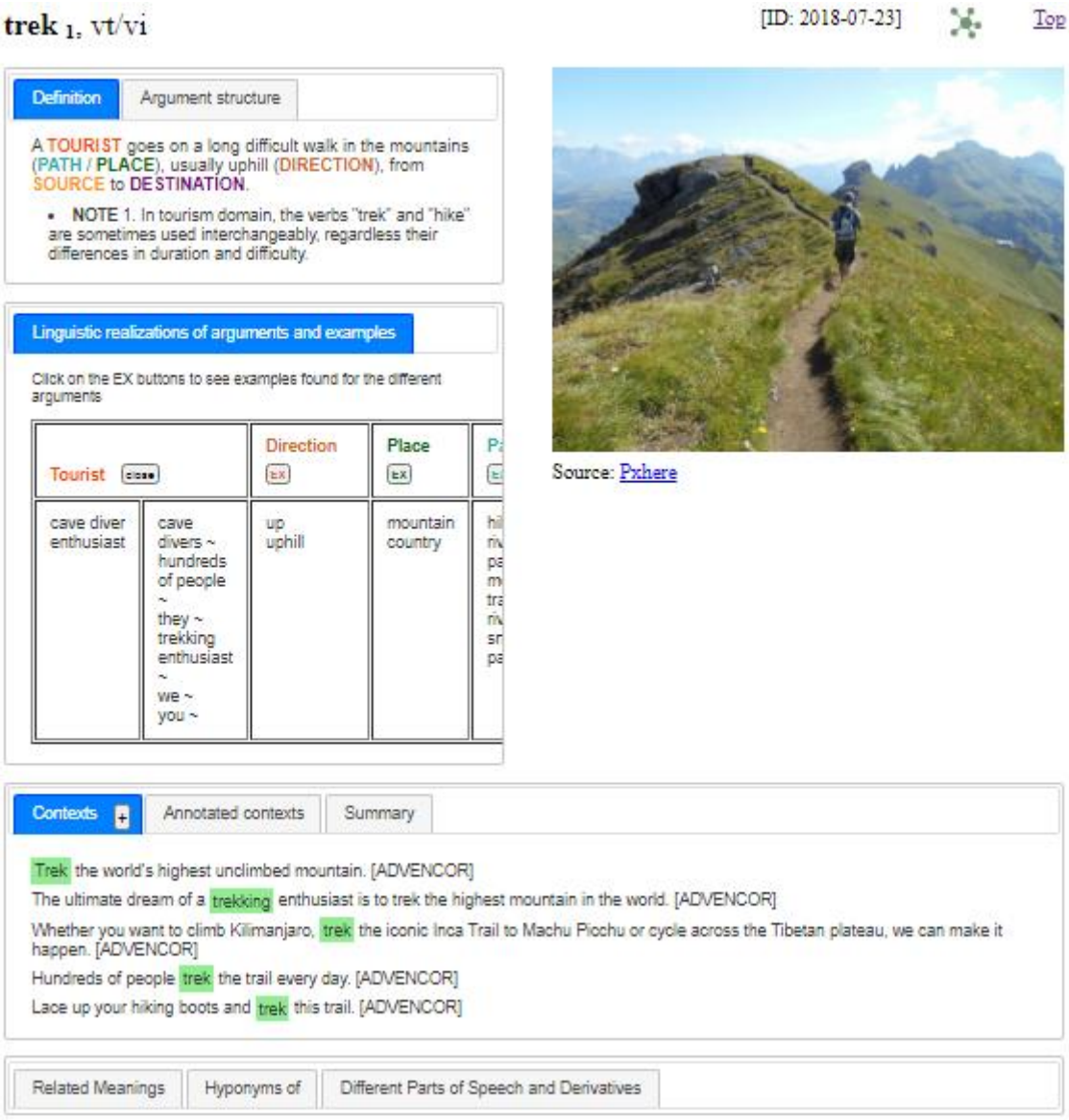
Entry List:

- abseil 1
- arrive 1
- ascend 1
- ascend 2
- bike 1
- climb 1
- cross 1
- depart 1
- descend 1
- dive 1
- drive 1
- enter 1
- exit 1
- fall 1
- fall 2
- float 1
- fly 1
- glide 1
- head 1
- hike 1
- jump 1
- kayak 1
- land 1
- launch 1
- navigate 1
- paddle 1
- pass 1
- pull 1
- raft 1
- rappel 1
- reach 1
- ride 1
- ride 2
- scale 1
- scramble 1
- skydive 1
- slide 1
- soar 1
- swim 1
- traverse 1
- trek 1
- venture 1
- walk 1
- zip 1

With respect to the microstructure of the dictionary, it offers linguistic, pragmatic and semantic information that can be easily accessed. Additionally, the arrangement of this information is so suitable that it allows users to clearly see and interpret the comprehensive data stored, as it can be seen in the sample entry shown in Figure 45:

Figure 45

Entry for trek$_1$ *in* DicoAdventure



All the entries in *DicoAdventure* have the same structure. Therefore, the first item of information found is the term itself along with its grammatical category. Although now all the entries are for verbs, this tab will be of crucial importance when new categories are implemented, since it will help to distinguish between verbs and nouns which are spelled the same, such as *raft*. Additionally, the type of verb is provided, that is, *trek$_1$* can be used transitively or intransitively ("vt/vi"). Moreover, it must be emphasized that entries have a number because *DicoAdventure* is a concept-based dictionary (like the DiCoInfo and the DiCoEnviro, cf. §3.3.), thus, an entry is created for each meaning of

the same term. Furthermore, administrative data is included on the upper right corner of the entry, which informs about the person who added this entry (i.e., "ID") and the date when it was created (i.e., "2018-07-23"). It also incorporates a link to the top of the page (i.e., "Top"). After these pieces of information, four sections can be found (which are based –although adapted– on L'Homme, Robichaud, & Subirats Rüggeberg, 2014, as cited in Durán-Muñoz, in press).

First of all, the 'Definition' and the 'Argument structure'[141] of the term are provided. As we can see in Figures 46 and 47, both tabs display the same semantic roles, albeit they are arranged differently in order to attain their objective. On the one hand, the definition aims to describe the meaning of the term in the context of adventure tourism; on the other hand, the argument structure shows the relationship between the term and the main semantic roles found in the structure of the verb. In addition, if further clarification is needed, notes are included.

Figure 46

Definition of trek$_1$ *in* DicoAdventure

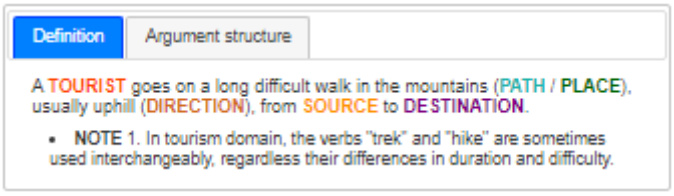
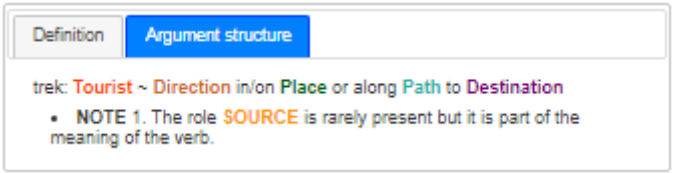


Figure 47

Argument Structure of trek$_1$ *in* DicoAdventure



[141] L'Homme (2017, p. 223) suggests that a verb's actantial (argument) structure be accounted for in its entry in a terminological database. In her opinion (2010), it is "an efficient and elegant method for referring to actants that share the same relationship with different predicates but that might occupy a different syntactic position" (p. 149).

Without a doubt, the most valuable information collected in *DicoAdventure* relates to the argument structure of the verbs, whose core components are based on a refinement of the categorization framework suggested by Durán-Muñoz (2016, p. 236; Figure 48), whereby the access to one category activates the other four. As a result, the categories <agent>, <location> and <instrument> were subdivided and gave rise to a total of 12 potential participants (semantic roles) in the argument structures of the motion verbs, which are explained in Table 17 (Durán-Muñoz & L'Homme, 2020, p. 46).

Figure 48

Categorization Framework on Adventure Tourism (Durán-Muñoz, 2016)

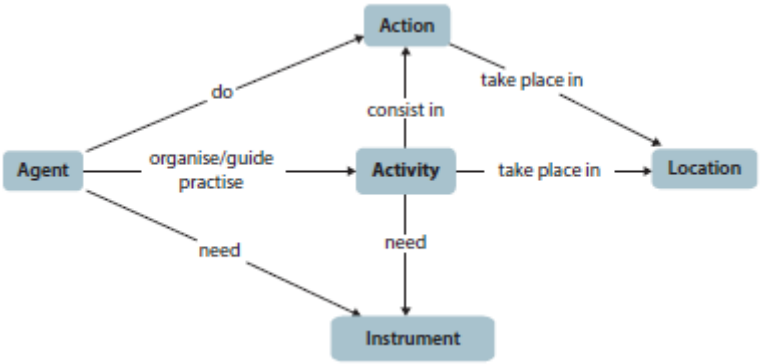


Table 17

Participants (Semantic Roles) in the Argument Structures of the Motion Verbs in DicoAdventure *(Durán-Muñoz & L'Homme, 2020)*

Broad conceptual category	**Semantic role**	**Definition**
<agent>	TOURIST	Person practicing the adventure activity (e.g., *hiker*)
	RESPONSIBLE	Person in charge of taking care of, guiding or briefing the TOURIST (e.g., *guide*)
<location>	PLACE	Location or area where the adventure activity takes place (e.g., *mountain*)
	DIRECTION	Direction of the action (e.g., *upwards*)
	SOURCE	Starting point of the adventure activity (e.g., *cave*)
	DESTINATION	End point of the adventure activity (e.g., *ground*)
	PATH	Path along/through which the adventure activity is developed (e.g., *trail*)
<instrument>	INSTRUMENT	Object employed as instrument in a specific adventure activity (e.g., *paddle*)
	VEHICLE_WITH_ENGINE	Vehicle with engine employed in a specific adventure activity (e.g., *car*)
	VEHICLE_WITHOUT_ENGINE	Vehicle without engine employed in a specific adventure activity (e.g., *canoe*)
	SAFETY_INSTRUMENT	Gear required in an adventure activity to guarantee TOURIST's safety (e.g., *helmet*)
	CLOTHING	Pieces of clothes required in an adventure activity (e.g., *wetsuit*)

Durán-Muñoz and L'Homme (2020, p. 47) consider the participants included in Table 17 more central to adventure activities and divide them into *arguments* and *circumstantials*, which are defined in the following terms: "arguments [...] are core components of the meanings of motion verbs in adventure tourism, and circumstantials [...] add peripheral

information but are optional as far as characterizing the meaning is concerned" (p. 45). For instance, some circumstantials that were detected during the analysis of the contexts are DISTANCE, DURATION, FREQUENCY, METHOD, PURPOSE or MANNER (p. 47).

Turning back to the argument structure of *trek*$_1$ displayed in Figure 47, the arguments found are TOURIST, PATH, PLACE, DIRECTION, SOURCE and DESTINATION. As it can be observed, different colors are used to distinguish the distinct semantic roles. For example, TOURIST appears in red, PATH in greenish blue, PLACE in green, DIRECTION in light brown, SOURCE in yellowish orange and DESTINATION in purple. A significant point is that these colors are reproduced in other data categories "to maintain consistency in the entry" (Durán-Muñoz & L'Homme, 2020, p. 51), as it will be observed below.

The second section of the entry shows the 'Linguistic realizations of arguments and examples.' The number of columns displayed will depend on the arguments found in the verb. For instance, the entry for *trek*$_1$ includes six columns because six are the arguments of the verb, as we have previously described; it is illustrated in Figure 49. The examples represented are the ones identified in the contexts extracted from the specialized corpus. Furthermore, more examples found for the different arguments can be accessed by clicking on the 'Ex' button in every column, as it is shown in Figure 50 for TOURIST.

Figure 49

Linguistic Realizations of Arguments and Examples of trek$_1$ *in* DicoAdventure

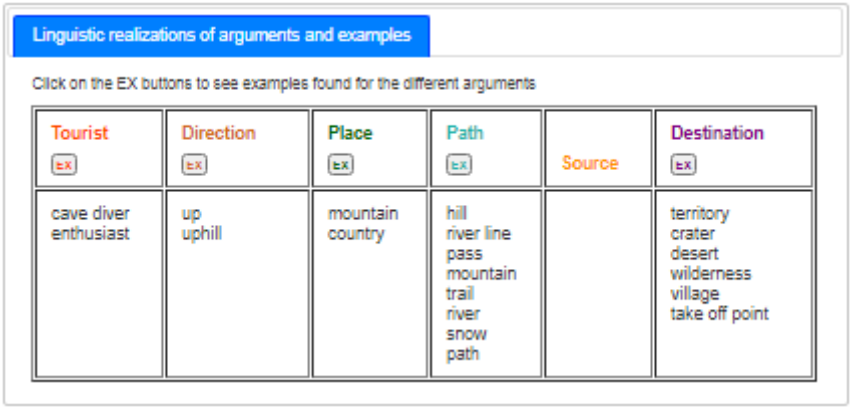
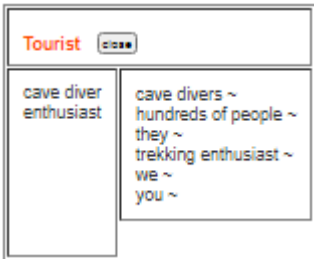


Figure 50

Examples of the Argument TOURIST *of* trek$_1$ *in* DicoAdventure



The third section of the entry contains three subsections: (1) 'Contexts,' (2) 'Annotated contexts,' and (3) 'Summary.' The first item shows real contexts retrieved from the specialized corpus, in which the term is highlighted in green color (more contexts are displayed by clicking on the "+" button) (Figure 51). The second item presents the same contexts but are annotated syntactically and semantically (this part of the entry sparks a great interest among users, for it allows them to observe the different arguments and circumstantials in context) (Figure 52). To explain, the motion verb is capitalized in black color and the different semantic roles of the structure of the verb use their corresponding color (they can be identified by placing the cursor on the colored elements). The third item summarizes the elements annotated (both arguments and circumstantials) (Figure 53).

Figure 51

Contexts of trek$_1$ *in* DicoAdventure

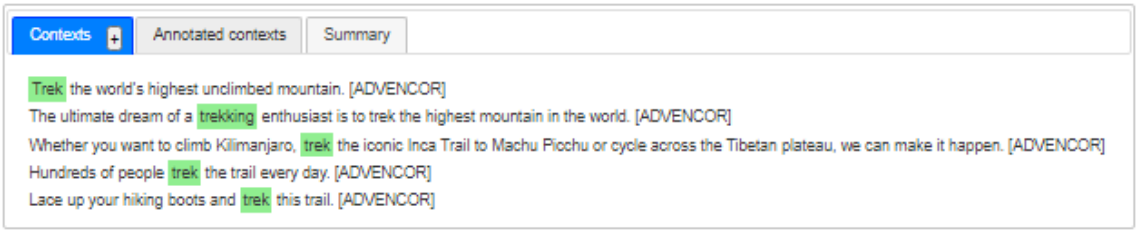


Figure 52

Extract From the Annotated Contexts of trek$_1$ *in* DicoAdventure

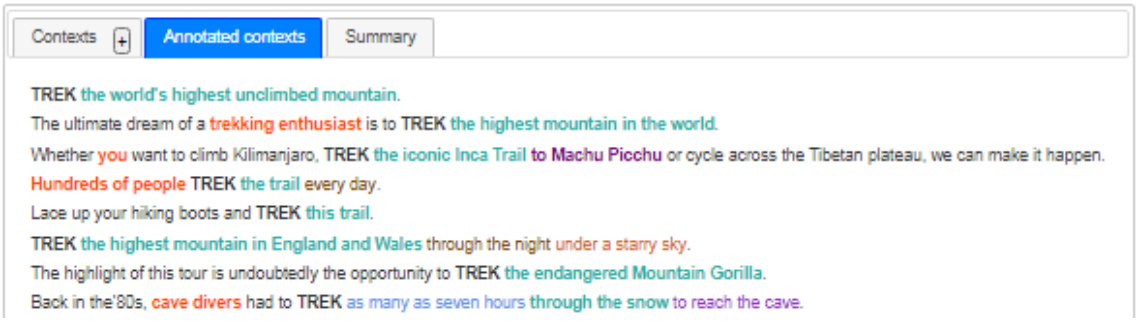
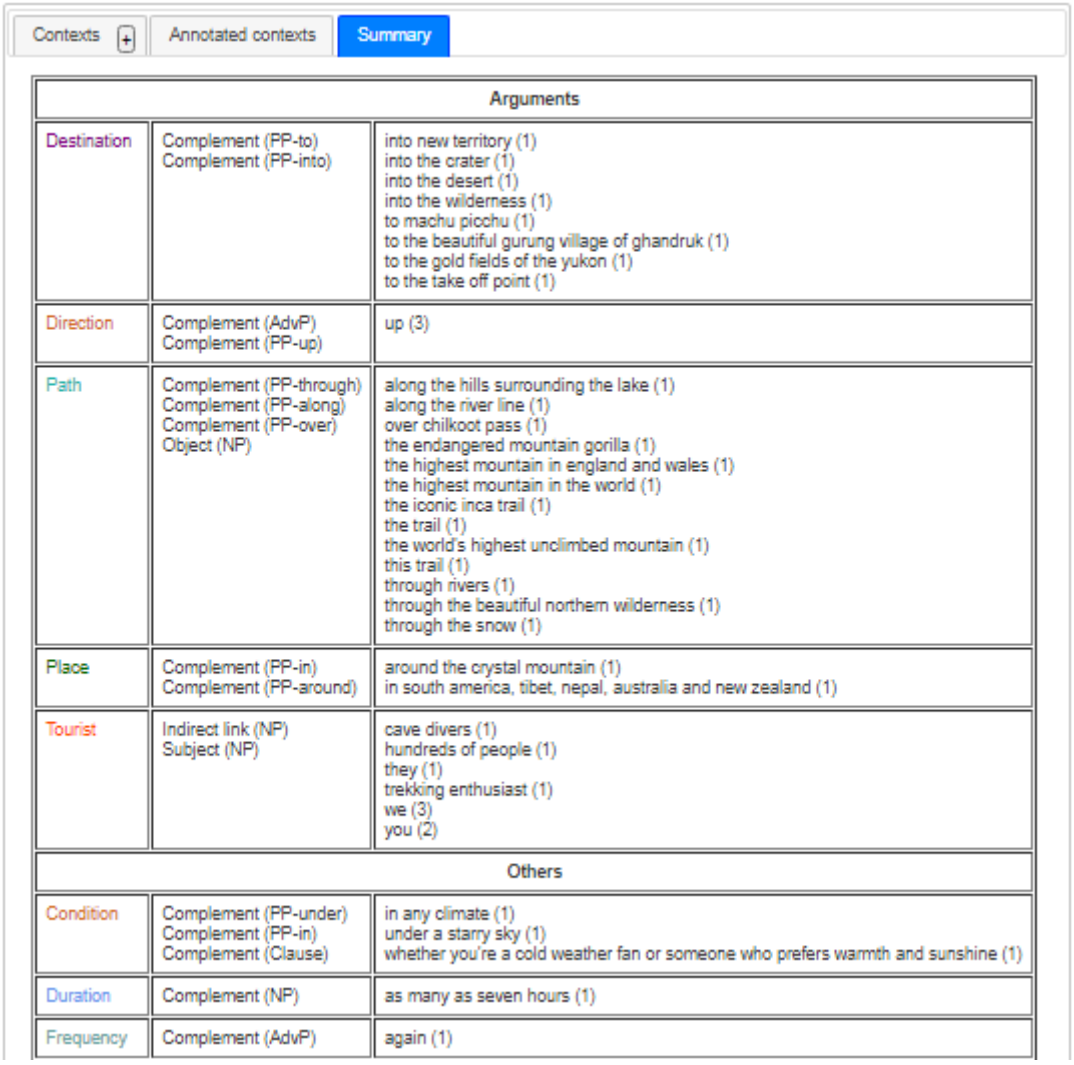


Figure 53

Extract From the Summary of trek$_1$ *in* DicoAdventure



Arguments		
Destination	Complement (PP-to) Complement (PP-into)	into new territory (1) into the crater (1) into the desert (1) into the wilderness (1) to machu picchu (1) to the beautiful gurung village of ghandruk (1) to the gold fields of the yukon (1) to the take off point (1)
Direction	Complement (AdvP) Complement (PP-up)	up (3)
Path	Complement (PP-through) Complement (PP-along) Complement (PP-over) Object (NP)	along the hills surrounding the lake (1) along the river line (1) over chilkoot pass (1) the endangered mountain gorilla (1) the highest mountain in england and wales (1) the highest mountain in the world (1) the iconic inca trail (1) the trail (1) the world's highest unclimbed mountain (1) this trail (1) through rivers (1) through the beautiful northern wilderness (1) through the snow (1)
Place	Complement (PP-in) Complement (PP-around)	around the crystal mountain (1) in south america, tibet, nepal, australia and new zealand (1)
Tourist	Indirect link (NP) Subject (NP)	cave divers (1) hundreds of people (1) they (1) trekking enthusiast (1) we (3) you (2)
Others		
Condition	Complement (PP-under) Complement (PP-in) Complement (Clause)	in any climate (1) under a starry sky (1) whether you're a cold weather fan or someone who prefers warmth and sunshine (1)
Duration	Complement (NP)	as many as seven hours (1)
Frequency	Complement (AdvP)	again (1)

The last section of the entry illustrates the lexical relations of the term, therefore, this is the section in which collocations will be implemented. Again, the number of columns displayed will depend on the lexical relations identified, which are presented in the following order: (1) 'Synonyms of,' (2) 'Related Meanings,' (3) 'Types of,' (4) 'Collocations,'[142] (5) 'Hyponyms of,' (6) 'Opposites,' and (7) 'Different Parts of Speech and Derivatives.' For instance, the lexical relations of *trek*$_1$ are 'Related Meanings' (Figure 54), 'Hyponyms of' (Figure 55) and 'Different Parts of Speech and Derivatives' (Figure 56).

Figure 54

Related Meanings of trek$_1$ *in* DicoAdventure

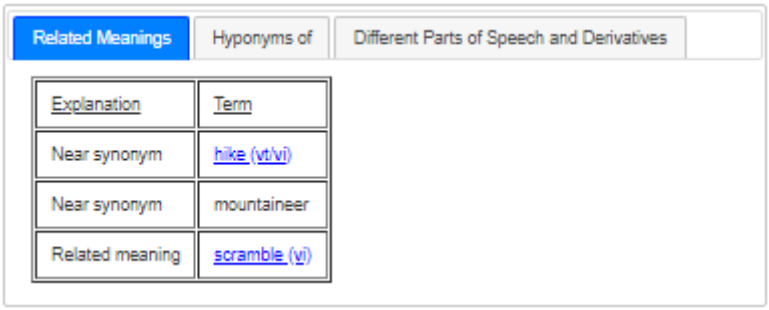
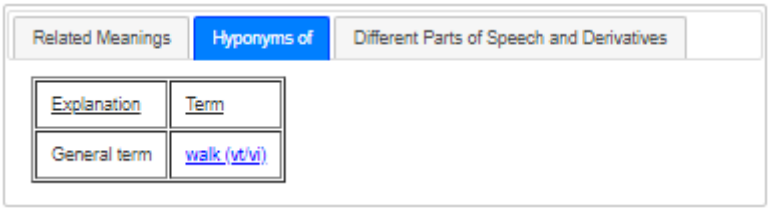


Figure 55

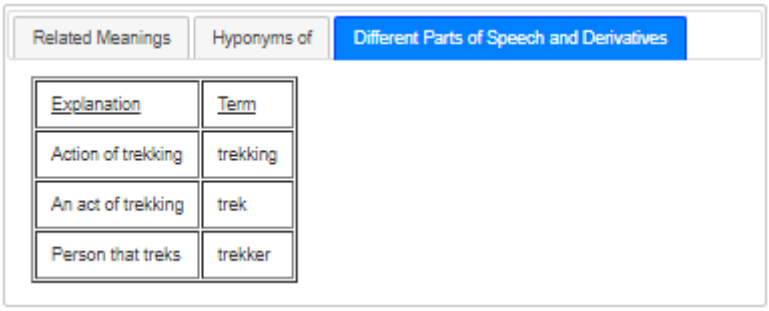
Hyponyms of trek$_1$ *in* DicoAdventure



[142] The research conducted in this work aims to complete this tab.

Figure 56

Different Parts of Speech and Derivatives of trek$_1$ *in* DicoAdventure



In addition to the four sections described above, every entry in *DicoAdventure* includes a picture (together with the source from which it was retrieved), which supports the definition of the term and facilitates the interpretation of its meaning. In the case of *trek$_1$*, it shows a man trekking in a mountain. It must be highlighted that this visual support in the form of a descriptive picture chimes with the idea of a comprehensive specialized dictionary put forward in Section 3.4., as it is suggested by Bergenholtz and Tarp (1995). Other features which give it this status are its electronic format, its user-friendly interface, being corpus-based and concept-based and the fact of not using any difficult metalanguage.

Aside from *DicoAdventure*, other resources have been created to provide information on the domain of tourism. For example, the *Dictionary of Travel, Tourism and Hospitality* (Medlik, 2003) was first conceived in the 1990s and its third edition is a revised and expanded edition which includes more than 4,000 entries explaining terms, acronyms and abbreviations. It is divided into sections describing some 300 international and national organizations, profiling 100 outstanding individuals and providing data on over 200 countries, all of this connected in some way with travel, tourism and hospitality. Another resource dealing with the SL of tourism is the *Diccionario de Términos de Turismo y de Ocio. Inglés-Español, Spanish-English* (Alcaraz Varó, Hugues, Campos Pardillos, Pina Medina, & Alesón Carbonell, 2006), which was first published in 2000 and its second edition provides an update with the inclusion of terms belonging to new fields within the domain of tourism, such as health, security or ecology. The dictionary contains over 16,000 entries of Spanish and English terms used in four sectors: leisure, travel, catering and lodging, and arranged around several topics, such as art and culture,

weather, sports, marketing or landscape. Finally, the latest edition of the *Tourism Society's Dictionary for the Tourism Industry* (Reily Collins, 2008) gives an overview of industry terminology, explanations of industry specific terms, useful travel related facts on over 130 countries, practical words and idioms for guides and tour managers and airline and airport codings. Nevertheless, two essential differences between *DicoAdventure* and these dictionaries must be stressed, aside from the topics covered by each of them. First, *DicoAdventure* is an online tool, and second, it is corpus-based. For this reason, the following section will delve into the concept of corpus, its design and compilation in order to finish with the description of the specialized corpus used for the creation of this terminological database.

4.3. Design, compilation and description of the specialized corpus

The beauty of a corpus is that it places no prior constraints on the imagination or curiosity of the investigator.

(Leech, 2011, pp. 161-162)

As it was stated in Chapter 1 (cf. §1.5.2.), corpus linguistics is considered a methodology rather than a theory, given that it facilitates the study of authentic language and language usage in different subject fields. Researchers who rely on this method can approach the corpus in two ways: first, it can be used as evidence to expound, test or exemplify a theoretical statement that has been previously made, and second, it can be used to formulate a theoretical statement about an idea that has been previously considered. The first of the methodologies is known as *corpus-based* and the second as *corpus-driven* (Tognini-Bonelli, 2001, pp. 10-11).

Albeit corpus linguistics has proved itself to provide a useful method for plenty of purposes, lexicography (both general and specialized) is profoundly indebted to it, as the information found in dictionaries is directly extracted from corpora. Additionally, "Corpus data is the great facilitator for the description of phraseology" (Moon, 2008, p. 314). Considering this, a corpus is the starting point in any research of this type. Thus, the third section of this chapter will provide an overview on the concept of corpus and types of corpora that can be found and will explain aspects related to their design and

compilation. Finally, the characteristics of the ADVENCOR corpus, the monolingual specialized corpus used for the creation of *DicoAdventure* and in the current work, will be examined.

4.3.1. The concept of corpus and typology

It is generally accepted that a *corpus* is a collection of texts which should accurately represent a language, given that it is assembled with the intention of conducting linguistic analyses (Biber & Reppen, 2015, p. 1; Lindquist, 2009, p. 3; Meyer, 2002, p. xi; Tognini-Bonelli, 2001, p. 2). According to McCarthy and O'Keeffe (2010, p. 3), it is necessary to go back to the thirteenth century to find the antecedents of corpora, when a group of biblical scholars aimed to manually index the words of the Christian Bible, line by line, page by page, in alphabetical order. This process of language arrangement influenced the way that linguists would work in the future, since they committed themselves to gathering real language data in the same way as the biblical scholars had done. Nevertheless, nowadays there is a radical difference in the way of compiling corpora, for they are built electronically and stored in digital media. In other words, it has been technology the major enabling factor in the growth of corpus compilation, as it allows to collect the maximum amount of data possible to satisfy researchers' investigation needs. As a result, different types of corpora can be identified in terms of their specificity, their size or their language, to mention only a few.

The first distinction is to be made between reference (generic) and specialized (also called special or special-purpose) corpora. The former are expected to comprise between 50 and 500 million words of standard language on varied topics extracted from written and/or spoken sources. They can exist for many different languages and are used for a multitude of purposes. A classic example is the BNC, a corpus of 100 million words of British English compiled in the 1990s. One of the purposes that reference corpora serve is to be useful benchmarks for specialized corpora. The latter are compiled when researchers are interested in a special phenomenon of language, that is, they are assembled for one particular investigation, for this reason, they are smaller, maybe under a million words (Fuster-Márquez & Pennock-Speck, 2015, p. 53; Teubert & Čermáková, 2004, p. 119). Both reference and specialized corpora can be identified as sample or monitor corpora (Sinclair, 1991, pp. 23-26), which is the second distinction to be made. While a

sample corpus only includes extracts of texts and is static (it was common in the past when computers had limited storage capacity), such as the Brown Corpus, a monitor corpus is one that includes complete texts and monitors language change, so it is regularly updated and open-ended, for example, the COCA.

The third distinction that must be drawn is between parallel and comparable corpora, which are usually specialized. Although both are collections of electronic texts that are closely related to each other, these relations are borne in different ways. First, the prototypical parallel corpus consists of a set of texts in language A and their translations in language B or more languages, so it can be bilingual or multilingual and the relationship lies in shared meaning. However, the links among the texts in comparable corpora are established because they have been gathered according to the same type of criteria (e.g., size, topic, text type, time span, author characteristics) in the same language, so they are monolingual. Thus, two corpora can be comparable if, for example, they are similar in size and their composition is similar in terms of genres included. According to Kenning (2010, p. 489), the issue of quality is decided in terms of different criteria. On the one hand, the focus in parallel corpora is on translation quality. On the other hand, the focus in comparable corpora is on authenticity, representativeness and the need to check the credentials of authors and sources.

All the types of corpora previously mentioned, that is, reference, specialized, sample, monitor, parallel and comparable, can include electronic material of two different modes: spoken and/or written. First, spoken corpora consist of recordings and/or transcriptions of a speech event in written form and endeavor to include speakers of all ages, socio-economic groups and regions of a country in order to represent a wide range of interaction types and social situations, such as casual conversations, seminars, meetings and interviews (Hunston, 2009, p. 158). Second, written corpora contain material from a variety of sources, such as newspapers, magazines or novels, and "are purportedly easier to create than spoken, largely because of the problems of spoken language transcription" (Nelson, 2010, p. 63). Normally, the source from which material is retrieved is the Web and there are two distinct processes: (1) the user accesses the data stored in the millions of servers around the world as if they were a giant corpus, which is known as *Web as Corpus* methodology (Bernardini, Baroni, & Evert, 2006), and (2) a selection of texts can be automatically or manually downloaded from the Web and

compiled in a special corpus, which is known as *Web for Corpus* methodology (De Schryver, 2002).[143]

Table 18 summarizes the parameters that define every type of corpus. In our opinion, these aspects should be considered before designing one, which is the topic of the following section.

Table 18

Parameters Which Define Corpus Types

Specificity	General (it represents the totality of a given language)
	Specialized (it represents a part/domain of a language)
Size	Large
	Small
Language/s	Monolingual (it covers one language)
	Multilingual (it covers two or more languages)
Mode of text	Written
	Spoken
Sample length	Full texts
	Extracts of texts

[143] The Web for Corpus methodology is usually preferred because it has more positive aspects: (1) the websites used for extracting the corpus texts can be checked by the compiler; (2) more careful linguistic analyses can be performed, since the sources of the texts have been verified; and (3) software tools have been created in the last years to facilitate the automatic creation of corpora (Buendía Castro & López Rodríguez, 2013; Fletcher, 2004, 2007, 2013).

4.3.2. Corpus design

The beginning of any corpus study is the creation of the corpus itself. The decisions that are taken about what is to be in the corpus, and how the selection is to be organized, control almost everything that happens subsequently. The results are only as good as the corpus.

(Sinclair, 1991, p. 13)

Without a doubt, the corpus is an essential element in corpus-based and corpus-driven studies, since the results obtained will solely depend on its characteristics and, of course, on the researcher's skills in obtaining and evaluating these results. For this reason, its design is of supreme importance and compilers must consider several factors that contribute to a top-quality corpus, namely, representativeness and balance, size and codification (Hunston, 2009; McEnery & Wilson, 2001; Rayson, 2015; Reppen, 2010), which will be briefly explored below.

Representativeness and balance

The representativeness of a corpus is the extent to which a corpus can be regarded as a sample of the variety of the language under study, for example, American English, legal English and academic English, as it is "una muestra finita de una población infinita"[144] (Vargas Sierra, 2006, p. 6; 2012b, p. 69). According to Biber (1993, p. 243), this sample aims to include the full range of variability in a population, which can be measured in terms of two aspects: situational and linguistic. The first one encompasses the types of texts, registers and genres to be selected, and the second one involves the number of words per text sample and the number of texts per text type. Obviously, the latter depends on the former, since, if a corpus does not represent the range of text types in a population, it will not represent its linguistic features either. Therefore, these two aspects correlate with the two criteria that can be established when selecting the texts that will make a corpus (Biber, 1993, p. 245; Sinclair, 2003a, pp. 170-171): (1) external criteria, that is, features

[144] "a finite sample of an infinite population" (translation mine).

of the non-linguistic environment or society in which the texts occurred, which produce text types, genres and registers (e.g., sports, psychology), and (2) internal criteria, which have to do with the linguistic choices within the texts.

One concept that is closely connected with representativeness is balance. In Atkins, Clear, and Ostler's (1992) terms, a balanced corpus is "so finely tuned that it offers a manageably small-scale model of the linguistic material which the corpus builders wish to study" (p. 6). It depends on the range of genres that constitute a corpus, the length and number of individual text samples and the population that will supply the texts. For instance, Meyer (2002, pp. 30-32) takes the BNC as an illustrative example of a balanced corpus, given that it contains approximately 100 million words and has the following characteristics: (1) it is composed of both spoken (10%, collected on the basis of two criteria, *demographic* and *context governed*) and written (90%, selected using three criteria: *domain, time* and *medium*) material; (2) this material includes a range of genres, such as arts and humanities, medicine, social sciences or education; (3) each genre is divided into text samples, each of which does not exceed 40,000 words; (4) for the written part of the corpus, a variety of variables were considered, such as the place and the time in which the texts were written or published, the type of text (e.g., books, articles, manuscripts, etc.) and their target audience; and (5) for the spoken part of the corpus, recordings were made from individuals representing the extant dialects in Great Britain and the various social classes found in these regions.

Size

The second factor that may influence the quality of a corpus is its size, that is, the number of words that it contains. However, this is still a subject of debate and experts do not agree on what size a corpus should be. First of all, there is the dichotomy between finite and non-finite size. Thus, static corpora will be of a finite size, whereas monitor corpora will be of a non-finite size, given that the latter keep updated and new words are added regularly.[145] Consequently, the debate is about the size of static corpora. For example, Davies (2015, p. 11) makes a categorization in terms of corpus size: (1) small corpora,

[145] In this connection, Vargas Sierra (2006, p. 13) states that, when a corpus is compiled with a lexicographical purpose in mind, it must be static at some point in order to extract the terminology typical of the SL under analysis as well as other linguistic information.

containing between one and five million words; (2) moderately sized corpora, containing around 100 million words; and (3) large corpora, containing about 450 million words. While Tognini-Bonelli (2010, p. 20) agrees with the author that 100 million words is the typical size of a corpus, Reppen (2010, based on Carter & McCarthy, 2001) states that "Corpus size is certainly not a case of one size fits all," therefore, "There is not a specific number of words that answers this question" (p. 31). For instance, what for Davies is a small corpus (one million words), for Biber (1993), Borja-Albi (2000) and McEnery and Wilson (2001) is the ideal number of words that any corpus should reach.

On the other hand, mottos such as *more data is better data* or *the bigger the corpus the better* have been used over the years (Corpas Pastor & Seghiri Domínguez, 2010, p. 120). Nevertheless, Leech (1991, p. 10) explains that corpora of a larger size are not necessarily better, since a large number of texts does not mean that they are representative. By the same token, Clear (1994) authored the article "I Can't See the Sense in a Large Corpus," whose title speaks for itself. Other authors have followed the same line of thought, particularly when it comes to specialized domains, where it is broadly agreed that smaller corpora give optimum results (Ahmad & Rogers, 2001, p. 735; Corpas Pastor & Seghiri Domínguez, 2010, p. 124; Murison-Bowie, 1993, p. 50; Bowker & Pearson, 2002, p. 48). In this context, Baker (2006) attaches, at least, the same status as quantity to quality: "when building a specialized corpus [...], we may want to be more selective in choosing our texts, meaning that the quality or content of the data takes equal or more precedence over issues of quantity" (p. 29), while Vargas Sierra (2006) stoutly contends that "la cantidad es un aspecto menor en importancia que la calidad"[146] (p. 11).

Another key point is that the size of a corpus highly depends on the phenomenon that is being investigated (Bravo Gonzalo & Fernández Nistal, 1998, p. 216; Flowerdew, 2004, p. 26). To this regard, Ahmad and Rogers (2001, p. 594) suggest that tens of thousands of words can already shed light on key terms in a specific domain. What is more, Biber (1995, p. 131) reduces these figures still further and maintains that one thousand words can represent practically the totality of elements of a specific genre, and Fillmore (1992) concurs on this point and believes that smaller corpora are not meant to be less relevant: "every corpus that I've had a chance to examine, however small, has

[146] "quantity is a less important aspect than quality" (translation mine).

taught me facts that I couldn't imagine finding out about in any other way" (p. 35). In this context, we agree with Ahmad and Rogers (2001, p. 594), Vargas Sierra (2006, p. 11) and Corpas Pastor and Seghiri Domínguez (2010, p. 124) that specialized texts are denser than GL texts when it comes to terminology. Consequently, a specialized corpus does not need to be as large as it would be desirable for a reference corpus, since positive results can be obtained if the corpus texts are representative of the SL under investigation in terms of terminology and technicality.

Codification

The last factor which can determine the quality of a corpus is its codification. It is generally accepted that electronic form is a sine qua non for corpus building (e.g., Hunston, 2009, p. 158; Lee, 2010, p. 107; McEnery & Wilson, 2001, p. 31), as electronic corpora have advantages unavailable to their paper-based equivalents. First, texts can be quickly searched and selected for language studies, so data can be easily manipulated as well as accurately and consistently processed. Second, since these processes are automatic, human bias is eliminated from analyses, thus making results more reliable. Finally, machine-readable corpora can be swiftly and easily enriched with additional information. This last advantage is particularly significant because the documents comprising a corpus are just a representation of the original documents. For this reason, much of the contextual information may be lost when these texts are rendered in a corpus-friendly form. Therefore, having the opportunity to add this information to corpora is more than welcome. Given the prominence of the computer in corpus linguistics, it is unsurprising that corpora are typically computerized corpora.

Electronic corpora may exist in two forms: unannotated, that is, in their existing raw state of plain text (*pure corpus*), and annotated/marked-up, that is, enhanced with linguistic information. Sinclair (2003b, pp. 186-187) distinguishes two kinds of information that can be annotated: (1) information about the origin of the text, such as the author, the place of publication or the year, and (2) information about the words found in the text, such as their POS or semantic information. In his opinion, the former can be contained in a simple reference to the database from which the text was extracted, whereas adding the latter can pose dangers, given that annotation is subjective and the annotator may be wrong. For this reason, the author (p. 183) believes that annotations may corrupt

original texts and oblige users to view the text through the filter of annotations, so he contends that plain texts are more appropriate. In fact, he argues that computational linguistics is bringing pressure to bear on corpus builders to annotate texts, for information scientists find plain texts useless.

Be that as it may, linguists broadly agree that the utility of a corpus is considerably increased by the provision of annotation. According to Meyer (2002), "For a corpus to be fully useful to potential users, it needs to be annotated" (p. 81). The author highlights that annotations can fit into three categories: (1) "structural" mark-up provides descriptive information about the texts (e.g., bibliographic citation for a written text or ethnographic information about the participants in a spoken dialogue), (2) "part-of-speech" mark-up assigns a POS designation to every word in a corpus (e.g., noun, verb, adjective) and is done automatically by a software program called a *tagger*,[147] and (3) "grammatical" mark-up assigns labels to grammatical structures beyond the level of the word (e.g., phrases, clauses) and is inserted by a software program called a *parser*. The important point about annotating a corpus is that linguistic information is made explicit, thus easier to grasp in analysis.

Nonetheless, despite being generally acknowledged that corpora should be annotated, it is advisable to have both versions of the corpus at the linguist's disposal, as argued by Sinclair (2003b): "a lexicographer should have access simultaneously to a corpus with tags, and an untagged version of the same corpus" (p. 187), and maintained by Leech (2011): "it is a cardinal principle of annotation that it should always be possible to separate the corpus and the annotated material" (p. 167). As a matter fact, Leech (1993) suggests seven maxims that need to be considered when a corpus builder purports to add annotation to his/her corpus. More specifically, these maxims need to be applied when evaluating the skills of the experts at annotating a corpus and when checking the usefulness of the annotative scheme that they have adopted. These seven maxims may be paraphrased as follows (Leech, 1993, p. 275; McEnery & Wilson, 2001, pp. 33-34):

1) It should be possible to recover the original corpus. In other words, annotations should be easily removable so that the annotated corpus could

[147] In Vargas Sierra's (2012b, p. 71) view, this is the most common tagging. In fact, it has become the canonical form of tagging for two main reasons: first, it is an easy task that can be performed automatically, and second, it is highly useful (for instance, it is the first step toward automatic lemmatization in the field of lexicography) (2010, p. 26).

revert to the raw corpus. For instance, if an annotated sentence is shown as *Claire_NP1 collects_VVZ shoes_NN2*, the user should be able to remove the annotation in order to get the original sentence *Claire collects shoes*;

2) The annotations should be extractable from the text so that they could be stored elsewhere, for instance, in a relational database or in an interlinear format, that is, on a separate line below the relevant line of running text. It means that the annotated corpus should allow flexibility for manipulation by the user;

3) The annotation scheme, which consists of the symbols used, their definitions and the guidelines for their application, should be available to the end user. Therefore, it would not be necessary to resort to guesswork when an instance of annotation could be given more than one interpretation;

4) The corpus user should be able to know who made the annotations and how. For instance, they can be made manually by one person or more than one, or they can be made automatically by a computer program whose output may or may not be manually checked afterwards;

5) The end user must be informed that an annotated corpus is not absolute truth, that is, the possibility of containing mistakes should not be discounted. For this reason, an annotated corpus should be regarded simply as a useful and valuable tool;

6) Annotation schemes should be based on widely agreed and theory-neutral principles; and

7) Annotation standards are not a sine qua non, although they may exist. To this regard, Lehmberg and Wörner (2009) sustain: "Standards in the linguistic context have to comprise a wide range of linguistic theories, languages, research areas and text types" (p. 485).

Finally, it must be mentioned that texts may be manually annotated with the use of any standard word processor, albeit automatic annotation is preferred for several reasons. For instance, when a big group of manual annotators is involved, objectivity is jeopardized and ensuring consistency may pose a problem (Nivre, 2009, p. 233). Additionally, when the amount of data to be annotated is huge, manual annotation can be

the most-consuming approach (Wichmann, 2009, p. 199). Nevertheless, compilers must bear in mind that automatic taggers and parsers are not absolutely perfect (Pęzik, 2018, p. 66; Vargas Sierra, 2010, p. 29).[148] In fact, Meyer (2002, p. 91) recognizes that automatic parsing has low accuracy rates (70%-80% at best), so a solution to this problem may be adopting a hybrid approach whereby an automatic process as well as manual intervention are involved. It is known as semi-automatic annotation (Zinsmeister, Hinrichs, Kübler, & Witt, 2009, p. 765) and is characterized for being a process in which a program suggests annotations to be approved or corrected by the human annotator and/or presents the annotation in such a way that compilers can detect errors. In any case, whichever approach is developed, a corpus must have been built beforehand, so the following section will explore the task of corpus compilation.

4.3.3. Corpus compilation

Since corpus compilation may demand considerable effort, it is reasonably assumed that researchers compile a new corpus when there is not an extant corpus that can fulfill their needs. This task is the result of the corpus design and the decisions made regarding the abovementioned aspects, namely, types and features of the texts, registers and genres to be selected, size and codification of the corpus. Additionally, it largely depends on the purpose that the product will serve, for it will determine the kind of information to be gathered. First of all, the compiler must decide whether it will be a corpus of written or spoken material, since the collection of samples of each type is made differently. For instance, with respect to written texts,[149] they can be found in three formats: (1) handwritten (e.g., a letter), (2) word-processed (e.g., a book), and (3) electronic (e.g., an article published online). The first type is the most problematic, given that it needs to be keyed in, which is a laborious and time-consuming process. The second type can be converted into a digital electronic form by a scanner and, depending on the quality of the image, the scan copy will be more or less accurate. The last type is the optimum way to build a corpus of written texts, given that, as it was previously mentioned, nowadays corpora are built in electronic format.

[148] It will be shown in Chapter 5 of this work.

[149] This section describes the collection of written texts, since the corpus used in this dissertation is comprised of this material. For a detailed analysis of the collection of speech samples, see Meyer (2002, pp. 56-61) and Adolphs & Knight (2010).

A crucial aspect when collecting written texts is their authorship. Data can be obtained from two basic sources, publicly available texts and privately available sources. Publicly available data can be found in a variety of sources, such as newspapers, magazines, journals or a number of sites on the Internet. Nevertheless, their copyright may pose a problem for corpus compilers, since some texts may not be available for private/research usage. Therefore, compilers need to be aware of the laws concerning this matter, not only in their own country but also in the country from where the texts are taken. Nelson (2010, p. 61) recommends using texts available in open source text archives on the Internet, such as Project Gutenberg.[150] On the other hand, private data is more difficult to access. As it is not in the public domain, corpus builders need to contact the author/s of the documents and obtain informed consent for the material to be used. According to Nelson (p. 61), the compilers who know the author/s beforehand have more chances of being assisted in the data gathering process, since this is common practice that authors tend to refuse to help, being compilers forced to persist.

Once the texts have been collected, they must be classified. To this respect, Sinclair (2003a, p. 172) states that choosing the topic of texts as the criterion for classification can be problematic for two reasons: first, specialized domain texts should be classified in terms of their social role rather than what they are about, and second, there is no broadly agreed classifications of topics that compilers can trust. Then, he suggests (based on an EAGLES[151] Report) a classification according to external criteria (e.g., author, audience, aim), which he expects can help people without training in the field. This classification includes three categories: (1) *Origin*, covering matters related to the origin of the text which may affect its structure or contents; (2) *State*, relating to the appearance of the text, its layout and other aspects of non-textual nature; and (3) *Aims*, which has to do with the reason for making the text and the intended effect that it is expected to have. Not only does the author emphasize that these criteria may not be applicable to all texts, but he also recognizes that a reference corpus does not adopt them. Table 19 summarizes the main parameters included in each category (pp. 173-177):

[150] https://www.gutenberg.org/ (Last accessed: 25/10/2020).
[151] Expert Advisory Group on Language Engineering Standards.

Table 19

Classification Criteria for Corpus Texts (Sinclair, 2003a)

Origin	*People* (involved in shaping the text)	*Author*	The writer of the original text
		Editor	Anyone who alters the text after being produced
		Publisher	The person responsible for publication
		Rights holder	Anyone with a legal right in the published work
		Translator	A person who translates a text into another language
		Adapter	A person who alters a text to make it suitable for another artistic genre
	Processes	Production processes which may have influenced the text	
	Circumstances	Other circumstances that may have influenced the structure or contents of the text	
	Timing	Matters of dating and timing concerning the text	
State	*Mode*	Mode of transmission, that is, spoken, written or electronic	
	Relation to the medium	Information that will not be retained (e.g., paper of a written text, acoustic conditions of spoken material)	
	Relation to non-linguistic communicative matter	Non-linguistic elements (e.g., diagrams, figures, tables)	
	Appearance	Aspects of presentation that may influence the language, such as the design in advertising leaflets	
Aims	*Audience*	*Size*	How many people are included
		Constituency	What kind of people are included
		Author-audience relationship	There is personal acquaintance (*close*), there is not personal acquaintance (*distant, neutral*)
	Intended outcome	For example, information, discussion, recommendation, recreation, instruction, ceremonial	

We agree with Sinclair that these classification criteria cannot be adopted by all texts. For instance, the origin of the texts included in the specialized corpus used in this dissertation,

the ADVENCOR corpus, lies in public and private sources that had recently published them on the Web at the time of the corpus compilation. Consequently, they were available in electronic form and written mode, but no information was stored with respect to their state, as it was not considered to be relevant to the linguistic analyses that the corpus would serve. Regarding the aims of the texts, since they are promotional, they endeavored to reach as many audiences as possible, given that their intended outcome was to market products and help to sell them. The next section will examine this corpus in more detail.

4.3.4. Description of the ADVENCOR corpus

[...] without suitable corpora, lexicographers

are at a loss when it comes to collocations.

(Teubert, 2004, p. 91)

The ADVENCOR corpus is the specialized corpus used in this dissertation and is a source of information with no profit motive in mind, but rather for academic and research purposes. It was compiled to undertake the research described in Durán-Muñoz (2019) and Durán-Muñoz and L'Homme (2020) and is the one from which the contents of *DicoAdventure* were extracted. It must be highlighted that the need for compiling a new corpus arose because no reference corpora may serve the researchers' purpose in mind, that is, the linguistic analysis of the SL of adventure tourism with a terminological intention. Before its design, the five parameters displayed in Table 18 (cf. §4.3.1.) were considered: (1) the specificity of the corpus stems from being a specialized corpus representing the SL used in the domain of adventure tourism, (2) its size is smaller than that of a reference corpus, (3) it is monolingual and the language described is English, (4) the material selected consisted of electronic written texts, and (5) the texts included were full texts. These features indicate that the ADVENCOR corpus falls into the category of specialized corpora. Regarding the abovementioned factors that would influence the quality of the corpus, that is, representativeness and balance, size and codification, they will be mentioned in the following paragraphs. The first step of the corpus' compilation was the selection of the material, which was made by applying a set of external criteria (Biber, 1993; Sinclair, 2003a):

1) *Time of publication*: the selected texts had been recently published on the Internet (they are contemporary), given that we aim to perform a terminological and phraseological analysis that shows the most current linguistic characteristics of the SL of adventure tourism;

2) *Geography*: the texts chosen were not geographically limited and had been published in websites hosted in English-speaking countries all over the world, such as the United Kingdom, the United States and Ireland, as we endeavor to use a corpus that is representative of the genre under investigation, so containing texts from a variety of authors would prevent any authorial style or typical dialectical patterns;

3) *Text type*: all the texts belong to the promotional genre, given that, as it was mentioned in Section 4.1., it is widely agreed that the SL of tourism is a language for promotion, thus we think they are representative of the texts usually found in the domain under analysis (cf. Vargas Sierra, 2006, p. 7). The texts included aim to promote destinations and were published by (public or private) registered tourist companies or travel agencies located in different parts of the world, such as Scotland,[152] England,[153] the United States[154] and Ireland.[155] More specifically, the purpose of the texts was to woo tourists interested in adventure tourism (in general) and adventure activities (in particular);

4) *Text length*: complete and original English written texts were selected (as opposed to fragments of texts), since we agree with Flowerdew (2004, p. 26) that it is necessary to collect full texts rather than samples of a certain length in order to adequately represent the genre; and

5) *Level of specialization*: the publishers of the texts are specialized in the segment of adventure tourism. However, the target readers are laypersons, such as potential tourists and prospective travelers, who are not expected to know anything about what the texts are telling. For this reason, the corpus is

[152] https://www.visitscotland.com/ (Last accessed: 25/10/2020).
[153] https://www.visitpeakdistrict.com/ (Last accessed: 25/10/2020).
[154] https://www.rei.com/adventures (Last accessed: 25/10/2020).
[155] http://dingleadventurerace.com/ (Last accessed: 25/10/2020).

semi-specialized and represent a communicative situation from experts to non-experts.

The second step was searching for the texts to be included. As compiling corpora from texts that are already in electronic form is easier (Flowerdew, 2004, p. 26), the 'New Corpus' function integrated in the Sketch Engine software (cf. §4.4.) was used. Its interface is quite intuitive and it allows the corpus builder to download the material directly from the Internet. After choosing a name for the corpus (i.e., ADVENCOR) and clicking on the option of finding texts on the Web (manual uploading of pre-selected texts is also possible), the extraction of the material can be done in three ways (Figure 57): (1) using (at least three) words or phrases that define the topic of the corpus to find documents containing them ('Web search'), (2) using URLs to download texts directly from specified locations ('URLs'), and (3) using a single web domain to get the data from the whole site ('Website'). The ADVENCOR corpus was compiled in accordance with the first option and 33 keywords and phrases (Table 20) for the domain of interest, that is, adventure tourism, were selected (Durán-Muñoz, 2019, p. 358).

Figure 57

Interface for the Extraction of the Corpus Texts From the Web in Sketch Engine

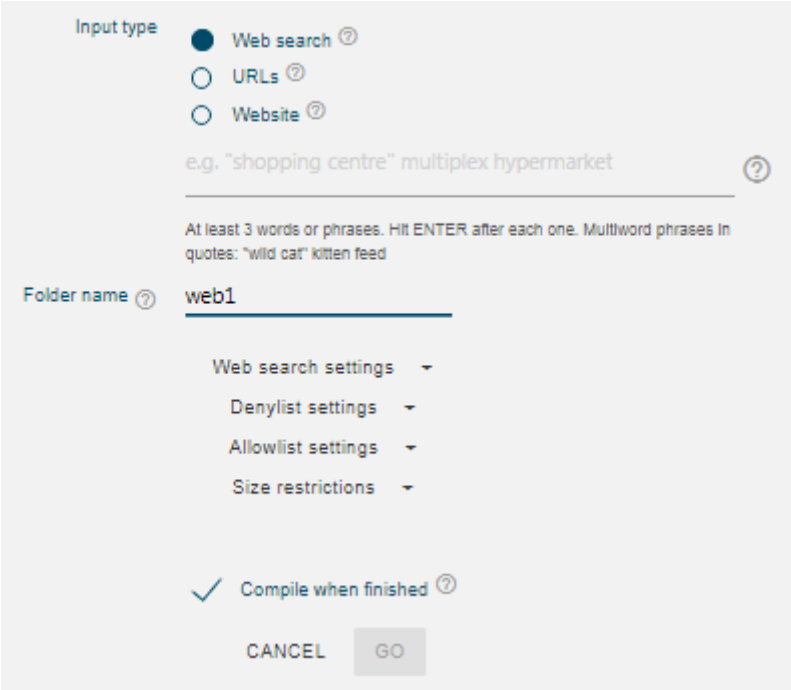


Table 20

Keywords and Phrases Used for the Extraction of the Corpus Texts in Sketch Engine

1. adventure	12. climbing	23. paragliding
2. adventure activity	13. cycling	24. parasailing
3. adventure activities	14. dogsledding	25. potholing
4. adventure sport	15. hang gliding	26. rafting
5. adventure tour	16. hike	27. rappel
6. adventure tourism	17. hiking	28. riding
7. bungee jumping	18. kayak	29. skydiving
8. canoeing	19. kayaking	30. speleology
9. canyon	20. mountain biking	31. trek
10. canyoning	21. mountaineering	32. trekking
11. caving	22. parachuting	33. zip lining

Once the keywords and phrases had been keyed in, a list of potentially relevant webpages was produced, which was manually checked to discard: (1) those considered irrelevant, such as *Wikipedia*, *Amazon*, social networks, *YouTube*, *Scribd* and *eBay*; (2) those that were not originally written in English; (3) those which were not published by public or private institutions (e.g., blogs, articles); and (4) duplicates in order to improve representativeness. In total, 30% of the URLs suggested by the tool was discarded. After confirming the webpages that better addressed the corpus' purpose, the last step was its compilation, which was automatically built from the selected webpages by means of the Web for Corpus methodology. At this stage, the corpus was loaded into Sketch Engine and was available to be downloaded in plain text (.txt) into the computer. It was also annotated using the same software, since we agree that annotated corpora can provide explicit linguistic information that may go unnoticed in plain text format (despite recognizing that it may not be 100% accurate).

The resulting corpus contains over one million words (i.e., tokens). Although there is no optimum size for a specialized corpus, we believe that this size is appropriate for the terminological and phraseological analyses that will be performed in this dissertation. Additionally, the fact that a series of criteria were established when selecting the texts, such as the expertise of the authors, the time of publication, the degree of specialization or the subject field, makes it representative of the genre under investigation. By the same token, it can be regarded as a balanced corpus because of the distinct variables that were

considered when extracting the texts, such as the length and number of individual text samples, the language in which they were written or the sources from which they were obtained. Table 21 provides a summary of the main characteristics of the ADVENCOR corpus:

Table 21

Defining Characteristics of the ADVENCOR Corpus

Number of tokens	1,179,210
Number of types	49,988
Type of corpus	Specialized / electronic
Size	Small
Mode	Written
Language	English
Domain	Tourism
Subdomain	Adventure tourism
Genre	Promotional
Text length	Complete texts
Purpose	Terminological and phraseological analyses
Communicative situation	Semi-specialized
Publication date	Recent
Source of texts	Websites
Publishers	English-speaking public or private institutions

Last but not least, it is important to realize that the major advantage of automatic corpus compilation is building large corpora quickly and effortlessly. Although the ADVENCOR corpus was thoroughly and manually revised in order to suppress inaccurate or useless information that could exert a negative effect on the analyses, and thus guarantee successful results, the whole activity was not much time-consuming. Moreover, one software was enough to extract the texts from the Web and to compile and annotate the corpus, that is, Sketch Engine, which was also used in the methodology applied in the current research. Therefore, the following section will examine the different functions that this tool offers before delving into the methodological steps that made it possible the achievement of the goals established in this dissertation.

4.4. Computer software: *Sketch Engine*

Sketch Engine[156] is a corpus query system which performs two main functions: (1) being a web service including corpus building and management, and (2) being a software tool for corpus exploration.[157] It is developed by Lexical Computing Limited, a company founded in 2003 by the lexicographer Adam Kilgarriff (Kilgarriff et al., 2014; Kilgarriff, Rychlý, Smrž, & Tugwell, 2004), and is aimed at the exploration of authentic language, given that it works with text corpora which gather real material. The tool trusts algorithms to identify what is typical in language and what is rare, and this is the reason why the first users were lexicographers, more specifically, Macmillan was the first dictionary that extracted word sketches[158] (cf. §3.2.2.) and Oxford University Press was the first user of Sketch Engine. The strong link between lexicography and this software program is blatantly obvious if one considers the fact that four[159] of the five main dictionary publishers in the United Kingdom use Sketch Engine extensively (Kilgarriff et al., 2014, p. 15).[160]

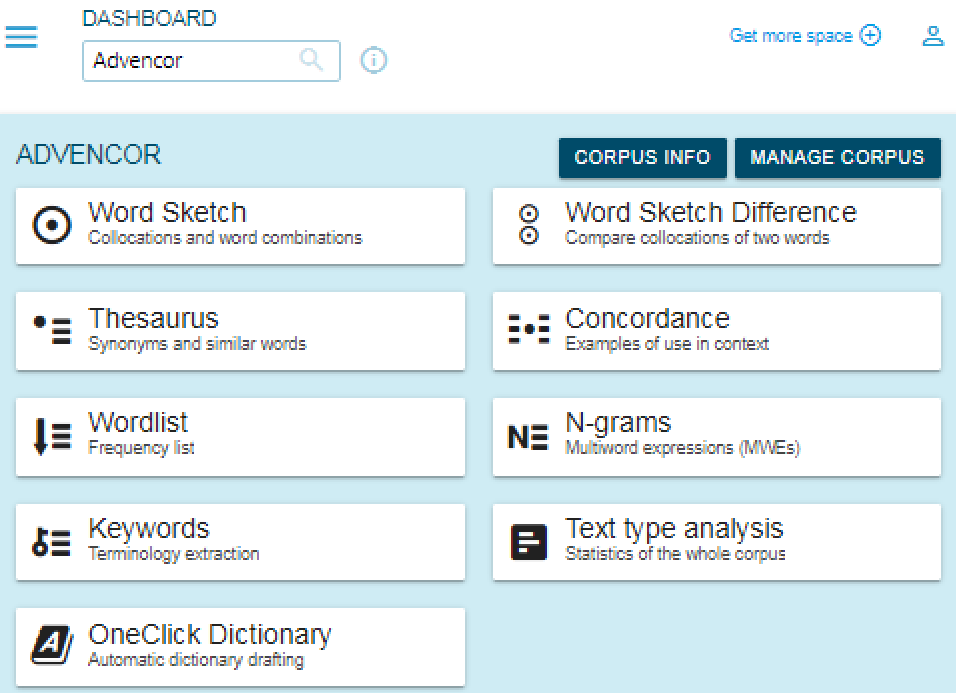
This tool endeavors to cover all the large languages in the world and the only prerequisite to using it is a corpus. Accordingly, it offers 500 *preloaded* (i.e., ready-to-use) corpora in more than 90 languages from all continents, which are regarded as representative samples of language for containing up to 30 billion words.[161] As this vast quantity of material covering a wide range of text types is difficult to collect, in most cases these are web corpora which are regularly updated. However, as it has been described in the previous section, if users cannot find a corpus that fulfills their needs from the great variety offered by the Sketch Engine Team, they can compile their own. Regardless of whether linguists use their own corpus or preloaded corpora, they can access the same functions,[162] for which registration[163] is necessary. After logging in and

[156] http://www.sketchengine.eu (Last accessed: 25/10/2020).
[157] Both functions were vital for the purpose of this dissertation, given that the first one allowed us to build and annotate the ADVENCOR corpus and the second one allowed us to explore it linguistically.
[158] It was for the preparation of the first edition of the MED in 1998, prior to the launch of Sketch Engine. The process is described in Kilgarriff & Rundell (2002).
[159] Cambridge University Press, Harper Collins, Macmillan and Oxford University Press.
[160] It goes without saying that this fact truly justifies the use of Sketch Engine in the work carried out here.
[161] https://www.sketchengine.eu/#blue (Last accessed: 25/10/2020).
[162] A complete user guide is freely available at https://www.sketchengine.eu/guide/ (Last accessed: 25/10/2020).
[163] Thirty-day free trials are available and allow access to all functions, featured corpora for all languages and corpus building for all languages. Access funded by the European Union is also available through the

choosing a corpus, the ‘Dashboard’ of the tool shows the huge variety of functions that allow users to explore corpora and, thus, languages, as it is displayed in Figure 58:

Figure 58

Dashboard of Sketch Engine


DASHBOARD
Advencor
Get more space
ADVENCOR
CORPUS INFO
MANAGE CORPUS
Word Sketch
Collocations and word combinations
Word Sketch Difference
Compare collocations of two words
Thesaurus
Synonyms and similar words
Concordance
Examples of use in context
Wordlist
Frequency list
N-grams
Multiword expressions (MWEs)
Keywords
Terminology extraction
Text type analysis
Statistics of the whole corpus
OneClick Dictionary
Automatic dictionary drafting

The first function that the dashboard offers is ‘Word Sketch,’[164] the one from which the tool took its name. It tells how a word is used at a glance by providing a one-page summary of its grammatical and collocational behaviour, providing “a feast of information on the word” (Kilgarriff et al., 2014, p. 8). Furthermore, each of the collocations includes links to examples in context (‘Concordance’ “ ”), to the word sketch of the collocation (‘Word Sketch’ “ ”) and to synonyms for the collocate (‘Thesaurus’ “ ”). The search is done by lemma and it is possible to choose the POS, the minimum frequency and the minimum score of the target word in the Advanced tab, among other options. The results can be downloaded onto one’s computer in four different

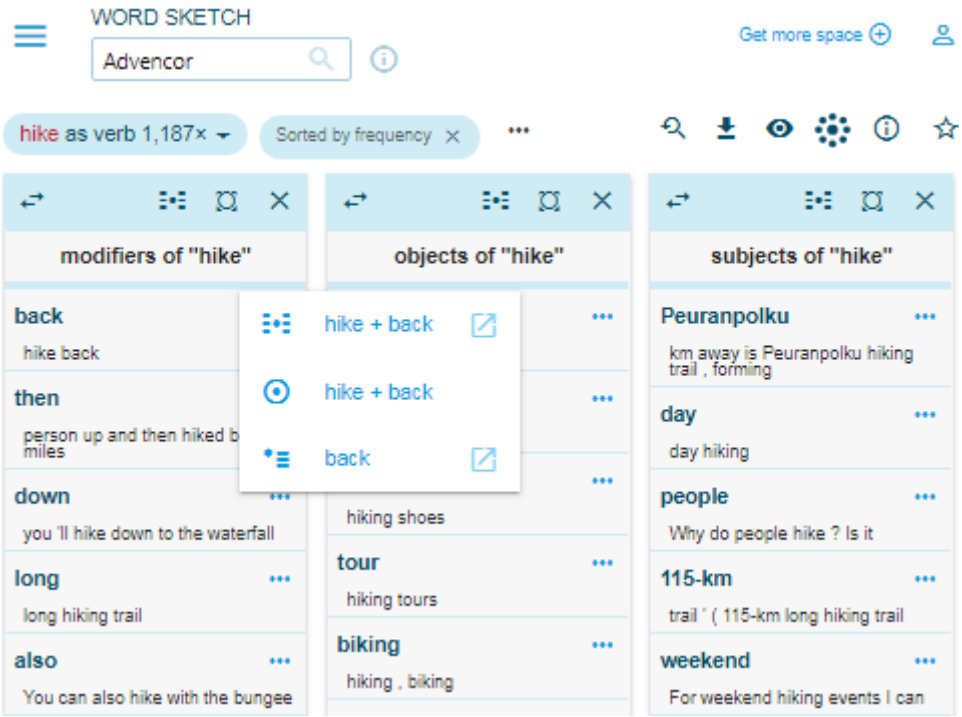
ELEXIS infrastructure project ([https://elex.is/] Last accessed: 25/10/2020) between 2018 and 2022, thus, all users belonging to one of the 305 universities or academic institutions listed in the project can access the tool for free. Other types of access are possible for a fee, depending on whether this is academic or non-academic, single or multiple.

[164] The Word Sketch function will be explained in more detail when we describe the methodology employed in this dissertation.

formats (.csv, .xls, .xml and .pdf) by clicking on the download icon "⬇."[165] By way of illustration, Figure 59 provides an extract of the word sketch of the lemma *hike* in the ADVENCOR corpus. After keying in the lemma and selecting the POS ("verb") and the minimum frequency ("auto"), the tool generates a list with separate columns according to the function fulfilled by the collocates of the lemma, for example, modifier, object, subject.

Figure 59

Extract From the Word Sketch of the Lemma hike *in the ADVENCOR Corpus; Sketch Engine, Word Sketch Function*



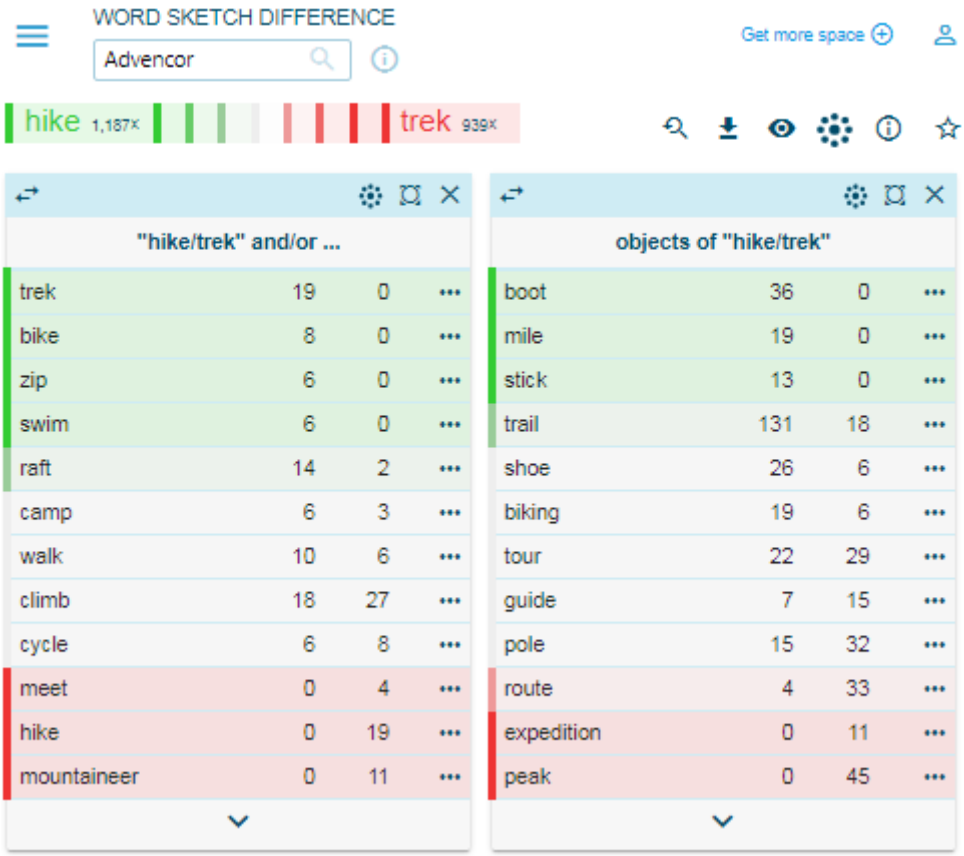
Word sketches of two words can be compared through the 'Word Sketch Difference' function. It may happen that a user wants to know how two words apparently similar behave in other words' company, so this option allows to observe differences in usage. The feature is especially useful for close synonyms, antonyms and words from the same semantic field. The POS of the target lemma and the minimum frequency in the corpus can be chosen. As an illustration, we compare the word sketches of *hike* and *trek*, verbs with a close meaning that implies a walk in the mountains. As it can be observed in Figure

[165] This option is available through all the functions of the software.

60, the result assigns red and green colors to each lemma, therefore, the collocates in green tend to combine with the green lemma, the collocates in red tend to combine with the red lemma and the white collocates combine with both lemmata. Bolder colored shades indicate stronger collocations.

Figure 60

Extract From the Word Sketch Difference of the Lemmata hike *and* trek *in the* ADVENCOR *Corpus; Sketch Engine, Word Sketch Difference Function*

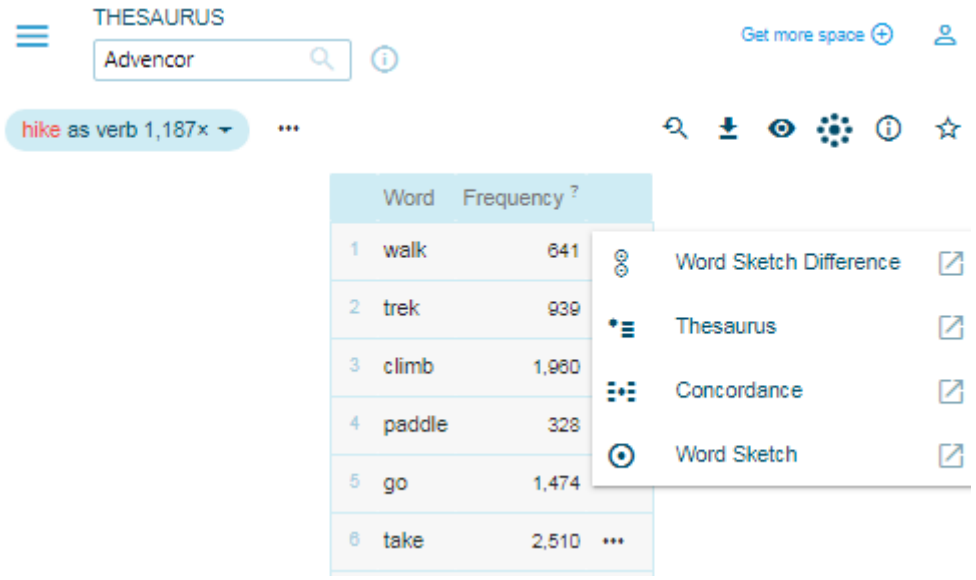


trek	19	0		boot	36	0
bike	8	0		mile	19	0
zip	6	0		stick	13	0
swim	6	0		trail	131	18
raft	14	2		shoe	26	6
camp	6	3		biking	19	6
walk	10	6		tour	22	29
climb	18	27		guide	7	15
cycle	6	8		pole	15	32
meet	0	4		route	4	33
hike	0	19		expedition	0	11
mountaineer	0	11		peak	0	45

The third function to be discussed is 'Thesaurus,' which retrieves synonyms for a chosen word. The lemma and the POS can be selected and the results show the synonyms' frequency in the corpus. Besides, each synonym comes with a link to other functions, as it is displayed in Figure 61:

Figure 61

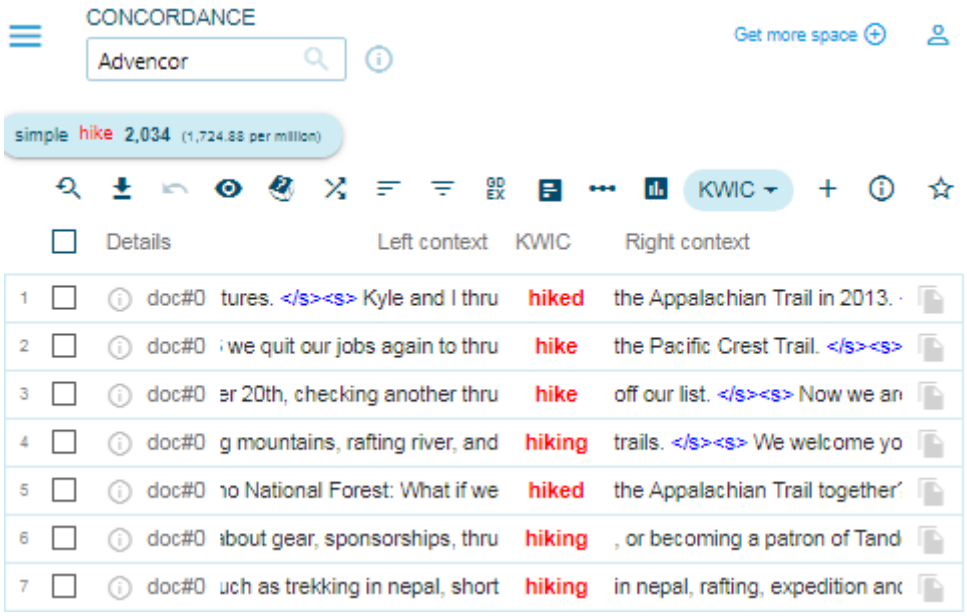
Extract From the Thesaurus Function of the Lemma hike *in the* ADVENCOR *Corpus; Sketch Engine*



The following function, 'Concordance,' is the one that allows users to find examples of a word, lemma, phrase, and so forth, in context. Thus, this search can contain only a single item (e.g., *hike*) or more than one (e.g., *hike through*). Apart from the query type (e.g., simple, lemma, phrase), the POS can be chosen. The results are given in concordance lines in the order in which they appear in the corpus, as it is shown in Figure 62:

Figure 62

Extract From the Concordance Lines of the Lemma hike *in the* ADVENCOR *Corpus; Sketch Engine, Concordance Function*



Next, the 'Wordlist' function generates frequency lists of all kinds: lemmata, nouns, verbs, tags, words containing certain characters, and so forth. Complex filtering criteria are available via the Advanced tab and prove extremely convenient when the user wants to obtain fine-grained data. Figure 63 shows an extract from the results obtained from the ADVENCOR corpus:

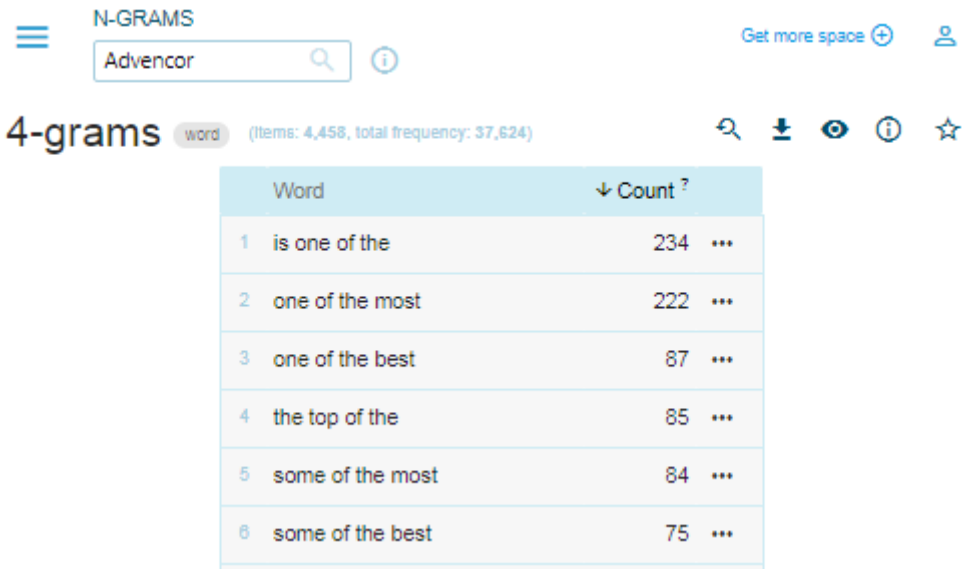
Figure 63

Extract From the Wordlist in the ADVENCOR *Corpus; Sketch Engine, Wordlist Function*

After the Wordlist function, the 'N-grams' function can be used, which searches for MWEs. The user can choose the length of the n-gram, which can range from two to six words. The Advanced tab provides the user with fine-grainer data. Frequent 4-grams in the ADVENCOR corpus along with their frequency are represented in Figure 64:

Figure 64

Extract From the 4-Grams List in the ADVENCOR Corpus; Sketch Engine, N-Grams Function

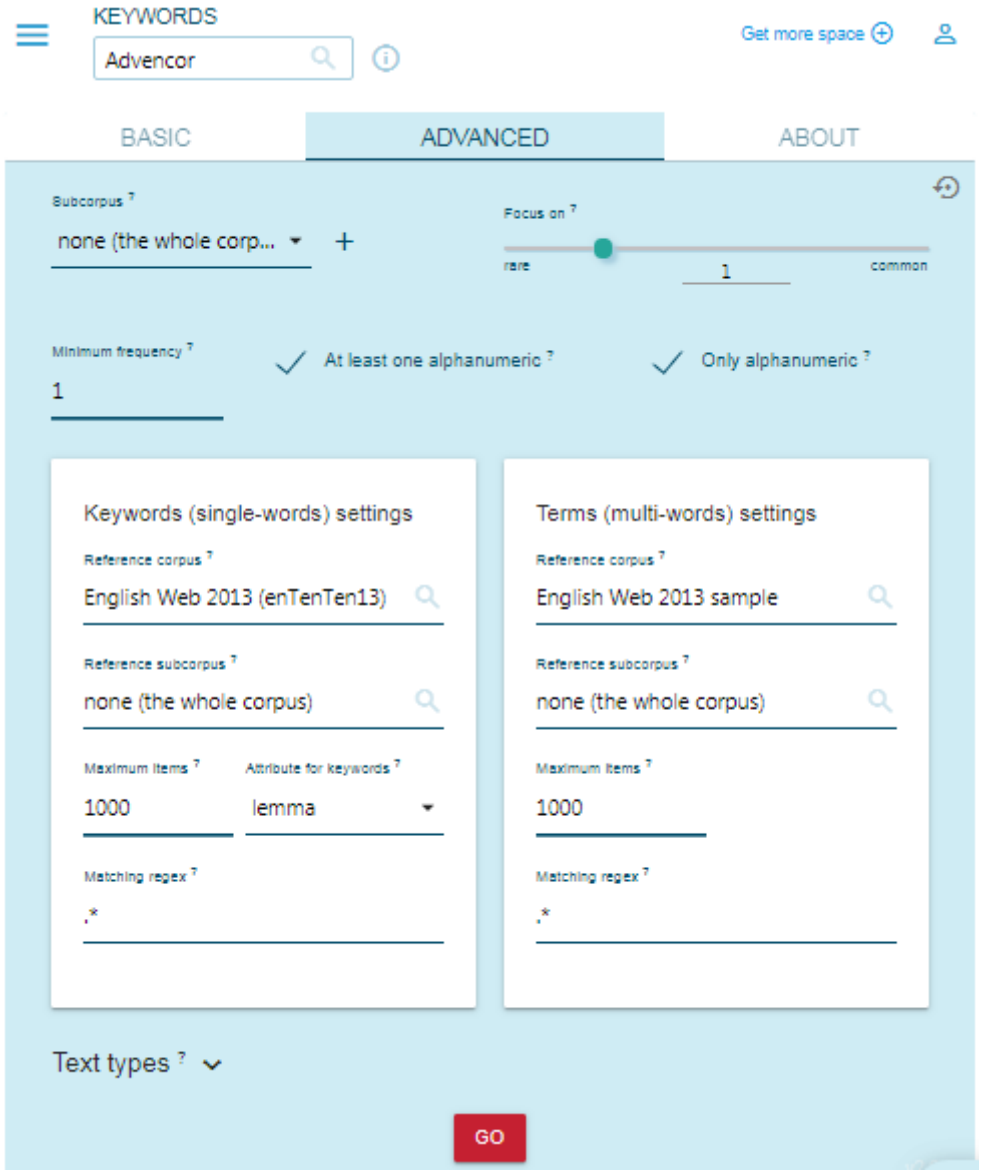


The last function to be described is 'Keywords,' which allows users to extract single units and MWUs combining statistics with linguistic criteria. Since these words are typical of the corpus under analysis because they occur more frequently than they would in the GL, a reference corpus is used. Sketch Engine sets the *English Web 2013 Corpus* (henceforth 'enTenTen13') by default for English language analyses, as it is the largest corpus available in this language. Extraction criteria can be changed in the Advanced tab, as it is displayed in Figure 65, but the Sketch Engine Team recommends that only expert users do so.[166]

[166] As we will see in the following section, this function was used to fulfil the first objective of this study, that is, the extraction of candidate motion verbs for the collocations to be implemented in *DicoAdventure*, so the criteria set by default had to be modified.

Figure 65

Advanced Tab of the Keywords Function of Sketch Engine



There are several other functions available in Sketch Engine. For instance, the 'Text type analysis' function provides statistics on the whole corpus, and the 'OneClick Dictionary' function creates dictionary drafts to be exported into the Lexonomy dictionary writing system.[167] Furthermore, the 'Parallel concordance' function can be used with a parallel corpus in order to explore words or phrases and their translations in context. Finally, the 'Trends' function requires a corpus whose documents are annotated with a date, given

[167] https://www.lexonomy.eu/ (Last accessed: 25/10/2020).

that it detects words which undergo changes in the frequency of use in time, that is, it makes diachronic analyses.

Moving on to the methodology employed in the current work, it must be highlighted that the most useful functions from the ones described above were Keywords, Word Sketch and Concordance. Since the primary objective established in this dissertation is the implementation of collocations of motion verbs into the SL database *DicoAdventure*, the methodological steps taken were three: (1) the extraction of candidate motion verbs which would be the bases of the collocations (Keywords function),[168] (2) the extraction of lexical collocations containing those verbs (Word Sketch function), and (3) the classification of those collocations into two categories according to their meaning: actual motion or fictive motion (Concordance function). The following sections will fully describe each of these steps.

4.4.1. Extraction of candidate motion verbs

The first step of our methodology involved the extraction of candidate verbs representing motion, as they would be the bases of the collocations implemented in *DicoAdventure*. For so doing, the Keywords function of Sketch Engine was used, as it compares the frequency and relative frequency of the same words in the specialized corpus and the reference corpus selected (in this case, we chose the one set by default, i.e., 'enTenTen13') in order to obtain the candidate terms which are specific to the former, demonstrating their significance in the domain under study. Additionally, it employs a hybrid method whereby statistical plus linguistic information is considered, which is the ideal situation for extraction (Vargas Sierra, 2010, p. 33). The keyness score is *simple math* (Kilgarriff, 2009) and is calculated according to the following formula:

$$\frac{fpm_{rmfocus} + N}{fpm_{rmref} + N}$$

where $fpm_{rmfocus}$ is the normalized frequency (per million) of the word in the specialized corpus and fpm_{rmref} is the normalized frequency (per million) of the word in the reference corpus; N is the so-called *smoothing parameter* ($N = 1$ is the default value). It can be said that this measure is an effect-size statistic which tells the importance of the difference

[168] In SLs, the bases of the collocations are normally the terms and the collocates are the words that combine with them, regardless of their syntactic category (Vargas Sierra, 2010, pp. 36-37).

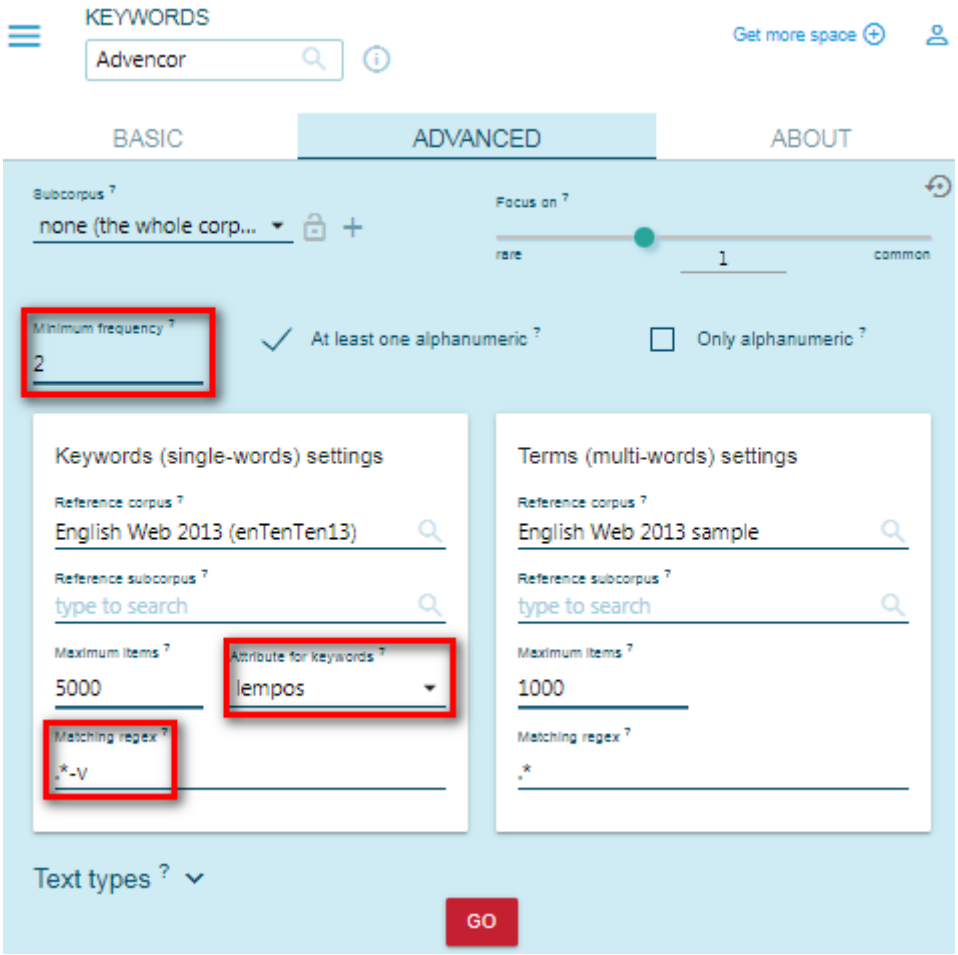
between keyness in the two corpora (as opposed to a significance statistic, which reveals the level of certainty to claim that the difference exists).[169]

Regarding the keyness score, no threshold was set because simple math produces real keywords, so we were confident that they were typical of the specialized corpus. The only difference between higher values (e.g., 100) and lower values (e.g., 1) is that the former focus on high-frequency words, that is, more common words, whereas the latter focus on low-frequency words, that is, rarer words. Nevertheless, a minimum frequency of the candidate verbs in the corpus was set, more specifically, the verb should occur at least two times, as it would be the minimum frequency for a word combination to be considered a collocation, which will be explained in the following methodological step. Additionally, the parameters were adjusted to extract lemmata whose POS was "verb," as it is illustrated in Figure 66:[170]

[169] See Gabrielatos & Marchi (2012) for a deeper analysis of the appropriate metric for keyness.
[170] The parameters in the 'Terms (multi-word) settings' column were not changed, since the focus in this step was on single words.

Figure 66

Parameters Set for the Extraction of Candidate Motion Verbs in this Dissertation; Sketch Engine, Keywords Function, Advanced Tab



This step produced a list of 978 items. However, manual work was needed to evaluate the results and to achieve desirable final outcomes for two reasons: first, because automatic extraction can be imperfect due to taggers that are not 100% correct, and second, because we were interested in verbs representing motion. Consequently, 860 verbs were discarded, which is explained below:

1) They belonged to POSs other than verb. Despite the very accurate performance of the tool, some candidates were wrongly proposed, for example, *NZD*, *pp*, *ncludes*, *ie*, *un*;

2) They were abbreviations, for instance, *yrs*, *lbs*, *gwt*;

3) They bore no relation to the SL of adventure tourism, such as *sandwich*, *tinkle*, *sip*, *apportion*, *allay*;

4) They were not regarded as motion verbs, for example, *beep*, *enroll*, *classify*, *substitute*, *staff*; and

5) They displayed distinct lemmata of the same motion verb, such as *paraglided*, *paraglides*, *paragliding* for *paraglide*.

The final list of candidate motion verbs amounted to 118 and were considered potential bases of collocations in the ADVENCOR corpus, which was checked in the second step of our methodology.

4.4.2. Extraction of collocations

After having explored the terminology (in particular, the verbs representing motion) of the ADVENCOR corpus through the Keywords function of Sketch Engine, the second methodological step of this work focused on the extraction of collocations containing the verbs selected, since they would be the bases of the collocations. Although the Keywords function also extracts MWEs, it was not used at this stage because it does not produce verb phrases in the patterns (M. Cukr –Documentation Copywriter at Sketch Engine–, personal communication, February 3, 2020). Therefore, the Word Sketch function was profitably exploited and potential collocates for the verbs under investigation were supplied. As it was previously mentioned, the results are organized in columns according to the function of the collocates. In this case, we focused on the columns showing modifiers, subjects and objects of the verbs, since only lexical collocations were contemplated for their implementation in *DicoAdventure*, that is, the types Le2-Verb + noun, Le3-Noun + verb, Le6-Adverb + verb/verb + adverb and Le7-Verb + adjective included in Table 12 (cf. §2.2.2.).[171] In this step, two criteria were applied. Nevertheless, no word span was specified, since it entirely depends on the function, which retrieves collocations whose integrating elements co-occur within the same sentence.

The first criterion in the extraction of collocations was motivated by the association measure used in Sketch Engine, that is, *logDice*. This measure tells how

[171] It must be remembered that grammatical collocations were not considered because they are seen as government patterns of the verbs.

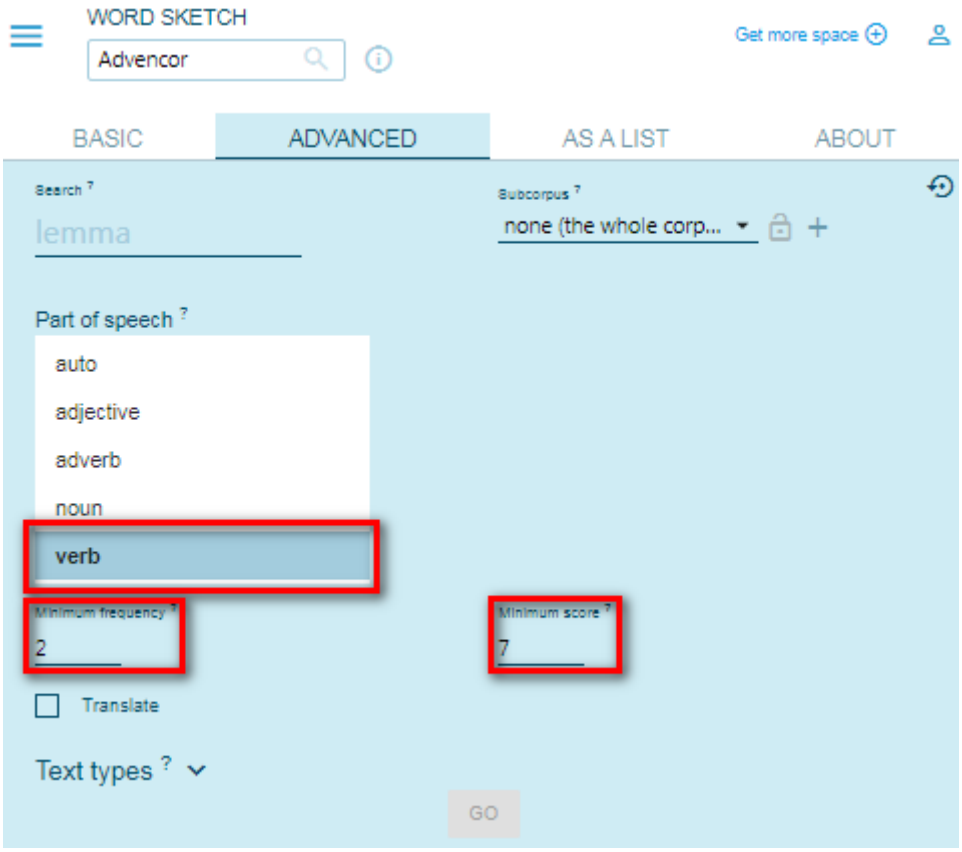
strong an association is and its implementation is recommended because it "has a reasonable interpretation, scales well on a different corpus size, is stable on subcorpora, and the values are in reasonable range" (Rychlý, 2008, p. 9). Moreover, this score is related to typicality, which means that the collocations retrieved are those which are not (completely) predictable, so it refers to collocations useful for learning and teaching or for inclusion in a dictionary.[172] The resulting values can be negative (no statistical significance) or positive, and theoretical maximum is 14. Despite the fact that this measure is extensively used, there seems not to be a total agreement on the minimum value for a word combination to be regarded as a collocation.[173] Nevertheless, it was determined that half this value, that is, 7 points, could serve as a useful indicator of a significant association, which could be considered typical of the corpus under analysis (Aldhubayi & Alyahya, 2014, p. 428). For this reason, the first criterion established for the extraction of collocations was a *logDice* score of ≥ 7 points.

The second criterion was encouraged by Evert's (2009, p. 1215) opinion that the best results are obtained by an association measure combined with a frequency threshold. Thus, we followed a mixed approach which integrated the ranked list of collocations in terms of their *logDice* score and the application of a frequency cutoff threshold, which was decided to be two.[174] Figure 67 shows the criteria established in the Advanced tab of the Word Sketch function of Sketch Engine for the extraction of collocations:[175]

[172] https://www.sketchengine.eu/blog/most-frequent-or-most-typical-collocations/ (Last accessed: 25/10/2020).
[173] For instance, Cao & Deignan (2019), Frankenberg-García (2018) and Lew, Frankenberg-García, Rees, Roberts, & Sharma (2018) set a threshold of 4, 3 and 5 points, respectively.
[174] According to Smadja (1993), "These thresholds have to be determined by the experimenter and are dependent on the use of the retrieved collocations" (p. 155). Since the main aim of this work was the implementation of verb collocations in a specialized dictionary, enabling the extraction of a higher number of collocations helped to achieve our goal.
[175] The lemma of the motion verb whose collocations were to be extracted was keyed in under the 'Search' label.

Figure 67

Criteria Established for the Extraction of Collocations in this Dissertation; Sketch Engine, Word Sketch Function, Advanced Tab



In brief, for the collocation extraction the criteria established were a minimum *logDice* score of 7 points and a minimum frequency of two tokens.[176] Nevertheless, we will see that a higher *logDice* score does not imply a higher frequency of the collocation in the corpus, since these two figures do not increase or decrease simultaneously. In other words, while *logDice* scores inform the user about the degree of association strength between two elements, frequencies tell how many times a collocation recurs in the corpus. On the other hand, although the fulfillment of both criteria suggested that the collocations selected were true collocations representing the SL of adventure tourism, one more step was taken in order to ensure that they could be regarded as specialized collocations (cf. §2.3.). Following Flowerdew (2004), who states that "A reference corpus, [...] which is usually much bigger than the specialized corpus and of a general nature, is often used as

[176] The application of these two criteria chimes with the model suggested by Bergenholtz & Tarp (1995) for the selection of word combinations, that is, relevance (typicality) and frequency (cf. §3.1.)

a yardstick for comparison purposes" (p. 27), the *logDice* values obtained in the ADVENCOR corpus were matched against the *logDice* values obtained for the same collocations in the reference corpus 'enTenTen13,' the largest corpus available in English in Sketch Engine and the one also used for the extraction of the verb terms. When the former were higher than the latter, it could be safely assumed that the collocations selected were specialized collocations.

At this point, it must be mentioned that, in addition to the Word Sketch function, the 'Collocations' option available in the Concordance function can also be used to explore collocations of terms. It allows the user to retrieve syntagmatic combinations within a span of -5/+5 from the lemma, whose elements may occur within the same sentence or in different sentences. The results can also be organized in terms of their *logDice* score. However, this option was not used for the whole collocation extraction process in this work for one strong reason: unlike the Word Sketch function, it does not identify collocations based on grammar rules (M. Cukr –Documentation Copywriter at Sketch Engine–, personal communication, July 3, 2020), so the query system is less smart and less sophisticated. Moreover, the concept of collocation defined in this study entails two words holding a syntactic relationship (cf. §2.2.4.), thus ignoring grammar rules was not acceptable. Despite this, when new collocations seemed to emerge (i.e., they had not been extracted through the Word Sketch function because they did not satisfy one of the established criteria or maybe neither of them, but they were perceived as a strong association between two words), this option was used and the concordance lines retrieved with the Concordance function were checked so as to dispel our doubts and, in some cases, it allowed us to identify new collocations worthy of inclusion in *DicoAdventure*. Likewise, these concordance lines were also examined to classify the collocations selected according to the meaning represented, as it will be explained in the following section.

4.4.3. Classification of collocations according to the meaning represented

The last step of our methodology centered on the classification of the collocations selected into two categories according to their meaning, namely: (1) verbs denoting actual motion, that is, they express the displacement of people in an adventure activity, so the entity is

actually moving, for example, *There was the danger of getting a head injury if **you ascend** from the water too quickly...*, and (2) verbs denoting fictive motion, that is, an inanimate object is depicted as if it were moving, but it is completely static in the real world, for instance, *A number of **trails ascend** to the top....*[177] Accomplishing this task was possible thanks to the examples in context (concordance lines, accessible through the Concordance function of Sketch Engine) retrieved from the ADVENCOR corpus.

At this point, two criteria were established. First, the collocations should describe the same meaning in at least two different contexts in the specialized corpus in order to be included in *DicoAdventure*. To put it differently, if one collocation was selected in the previous step for fulfilling the two collocation extraction criteria (minimum *logDice* score and minimum frequency), but from the two tokens found one represented actual motion and the other one represented fictive motion, it was not incorporated in any of the categories, given that two tokens representing the same meaning were necessary. Second, the contexts of the collocations should contain a specific reference to the domain of adventure tourism. When it was not case, the collocations were discarded, such as *head home, (a) road head*.[178]

On the other hand, it must be remembered that different entries were created for the same lemma when it represented both types of motion, given that *DicoAdventure* is a concept-based dictionary (like the DiCoEnviro and the DiCoInfo, cf. §3.3.) and each LU represents a different meaning. Generally, verbs occurring in contexts where actual motion is depicted belong to type one (e.g., *ascend*$_1$), whereas verbs occurring in contexts where fictive motion is depicted belong to type two (e.g., *ascend*$_2$). However, this is not an absolute rule, since new entries for the same lemma can be created as new meanings emerge.[179] Moreover, in order to create new entries, it would be appropriate to identify about 15-20 different contexts in the specialized corpus, although in some cases it may be reduced to 10 as long as they include precise information on the behavior of the verb.[180]

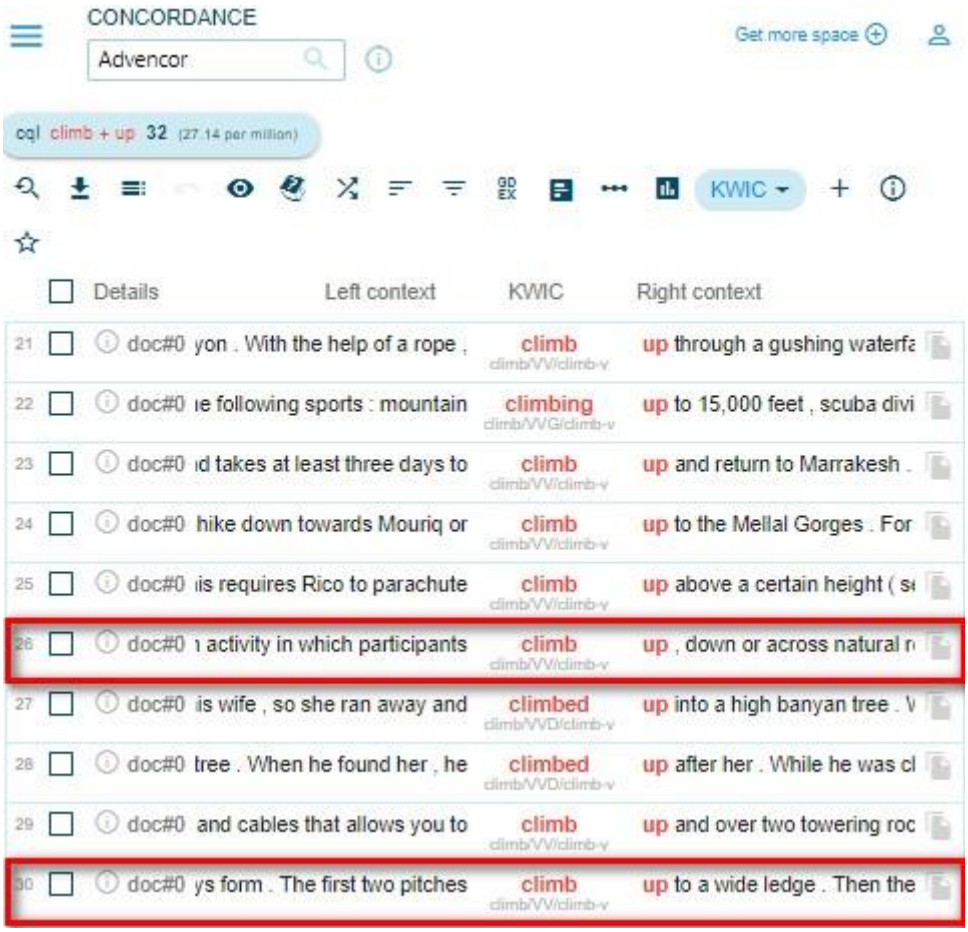
[177] Both examples were retrieved from the ADVENCOR corpus.
[178] In this work, the collocations will be described using the lemmata of both elements.
[179] For instance, the entries for *fall*$_2$ and *ride*$_2$ represent a different meaning from the one included in types one, but it is not a kind of fictive motion.
[180] According to L'Homme (2004, p. 12), the observation of the interactions of a term with other LUs makes it possible the delimitation of meanings.

If the meaning of a verb could not be attested in at least this number of contexts, no entry would be created in *DicoAdventure*.[181]

In short, the task of classifying the collocations in terms of the type of motion represented required a closer manual inspection of each of the concordance lines extracted from the ADVENCOR corpus so as to identify the participants in the argument structures of the verbs, since the contexts found for a given collocation may depict both types of motion. For example, Figure 68 shows an extract from the concordance lines containing the collocation *climb up*, in which examples of the two different types of motion are highlighted (line 26 describes actual motion, whereas line 30 describes fictive motion):

Figure 68

Extract From the Concordance Lines of the Collocation climb up *in the ADVENCOR Corpus Representing Actual and Fictive Motion*



[181] For example, it happened with the verbs *reach* and *traverse* representing fictive motion events, so the entries for *reach*$_2$ and *traverse*$_2$ could not be created.

To sum up, it must be stated that Sketch Engine was proved itself to be a valuable corpus compilation and management tool in the research conducted here. The use of the Keywords, Word Sketch and Concordance functions allowed us to go through the different methodological steps taken in this study in order to successfully attain our goals, that is, (1) the extraction of candidate motion verbs that would be the bases of the collocations implemented in *DicoAdventure*, (2) the extraction of collocations containing these verbs, and (3) the classification of these collocations according to the meaning represented (actual motion or fictive motion). However, we are aware that other tools are available which can also perform multiple functions when it comes to the exploration of corpora with terminological and phraseological intentions.[182] Thus, the next section will present two of them, *TermoStat Web 3.0* and *WordSmith Tools 7.0*, and will explain their main functions and why they could not serve the purposes of this study. After that, the template used for the implementation of collocations in the different verb entries in *DicoAdventure* will be presented.

4.4.4. Assessment of other corpus management tools

TermoStat Web 3.0

TermoStat Web 3.0[183] is an online tool designed for term extraction developed by Patrick Drouin (Drouin, 2003) in the OLST at the University of Montreal, Canada. It uses a hybrid approach, that is, statistical plus linguistic, to identify candidate terms in domain-specific (specialized) corpora by matching against a general reference corpus. On the one hand, potential candidate terms are extracted according to their linguistic structure (a POS tagger is used to identify nouns and adjectives and complex structures that contain these items). On the other hand, the extraction process considers the relative frequencies of these potential candidate terms in the corpus under analysis and in the reference corpus. This hybrid method allows TermoStat to find not only multi-word but also single-word

[182] See Vargas Sierra (2010, pp. 28-32; 2012b, pp. 90-91) for a list of programs used for terminological, concordance and combinatory extraction.
[183] http://termostat.ling.umontreal.ca/ (Last accessed: 25/10/2020).

candidate terms in a single extraction process.[184] The online version of TermoStat supports five different languages: English, Portuguese, French, Italian and Spanish.

Before using the tool, registration is necessary, which is completely free. Then, a corpus in plain text format (.txt) must be uploaded onto the platform so that the system can generate a list of candidate terms. The functions available are: (1) the extraction of single terms (adjectives, adverbs, nouns and verbs; more than one category can be selected), which will appear in the 'List of terms' tab;[185] (2) the extraction of multi-lexical terms (only nominal MWUs can be retrieved), which will be represented in the 'Bigrams' tab;[186] (3) the 'Cloud' tab displays the 100 candidate terms with the highest specificity scores and are displayed with varying font sizes according to this score, the more frequent looking bigger; (4) 'Stat' shows the statistics of the pattern/s under study; and (5) 'Structuration' contains a table of single candidate terms along with their frequency in the specialized corpus. All the results can be saved as a file in .txt format and users can open it afterwards in a spreadsheet program, such as Microsoft Excel.

All in all, the results provided by TermoStat are highly accurate. However, it must be remembered that they are obtained automatically according to formal aspects, which makes human evaluation necessary. Thus, the user of the tool will need to determine whether the terms extracted are indeed terms whose meaning is relevant to the specialized domain of interest (like with any other automatic term extractor, including Sketch Engine). An important aspect that moved us to refuse to use this tool is that verb-based MWUs cannot be extracted, which is the second step of our methodology, in addition to other functions that Sketch Engine can perform but TermoStat cannot, like setting a frequency threshold or expanding the concordance lines.

WordSmith Tools 7.0

WordSmith Tools 7.0[187] (Scott, 2001, 2016) is a software package aimed at the analysis of words' behaviour in texts. It has been developed by Lexical Analysis Software and

[184] As it was previously mentioned, this is the type of approach adopted by Sketch Engine to extract single units and MWUs through the Keywords function.
[185] This task can be undertaken by the Keywords function of Sketch Engine.
[186] This task can be undertaken by the Word Sketch function of Sketch Engine.
[187] https://www.lexically.net/wordsmith/downloads/ (Last accessed: 25/10/2020). The current version is 8.0.

Oxford University Press since 1996 and its leading author is Mike Scott. Access to all functionalities is gained through an online purchase, so different types of licenses are available depending on the number of users. WordSmith contains a diverse set of features, such as MWU conversion, POS tag filtering, lemma file building, subcorpora building and chargram patterns, among others, but the most common ones are 'WordList,' 'KeyWord' and 'Concord.' Firstly, the Wordlist option shows how often each distinct word occurs in the chosen text files; information can be arranged in alphabetical or descending frequency order.[188] Secondly, the KeyWord function helps to locate and identify keywords in the corpus under analysis after their matching against a reference corpus.[189] Thirdly, the Concord function extracts all the occurrences in context of a given word in the corpus.[190]

Although WordSmith offers a wide range of choices to know how words behave in concrete text files, beginners may not find the tool easy enough to use, given that it is not very intuitive. Apart from that, it seems like owning a corpus is not sufficient to make use of all the functions available, since a word list obtained from a reference corpus and saved by WordSmith is necessary if one aims to use the KeyWord function. To our mind, it does not make any sense for one cogent reason: unlike the Sketch Engine, the tool does not provide access to any reference corpora from which you can obtain this word list. Based on all this, we decided not to use WordSmith. Additionally, like TermoStat, the number of functions that WordSmith can perform is more limited than that in Sketch Engine.

4.5. Template for the implementation of collocations in *DicoAdventure*

As it was mentioned in Section 4.2., the 'Collocations' tab will appear in the section devoted to the lexical relations of the verbs in the entries in *DicoAdventure*[191] (that is, they will be accessible via the base of the collocations, i.e., the verb). After clicking on

[188] This function is similar to the Wordlist function of Sketch Engine.
[189] This function is similar to the Keywords function of Sketch Engine.
[190] This function is similar to the Concordance function of Sketch Engine.
[191] More specifically, these lexical relations appear in the following order: (1) 'Synonyms of,' (2) 'Related Meanings,' (3) 'Types of,' (4) 'Collocations,' (5) 'Hyponyms of,' (6) 'Opposites,' and (7) 'Different Parts of Speech and Derivatives.'

the tab, a table with three columns, including as many rows as collocations selected for the verb, will be displayed:

1) The first column, called 'Collocation,' will include the collocations themselves and will be organized according to their *logDice* score in the specialized corpus, thus, collocations with a higher score (i.e., stronger collocations and collocations more typical of the SL of adventure tourism) will be displayed first;[192]

2) The second column, named 'Type,' will identify the lexical type of the collocation, for example, verb + adverb, noun + verb; and

3) The third column, called 'Explanation,' will provide an explanation of the collocation in terms of the semantic roles identified in the argument structure of the motion verb that is part of the collocation (cf. Table 17, §4.2.), such as PLACE, PATH, DESTINATION.[193] Moreover, the 'Explanation' column will include an 'Ex' button that will give access to a fourth column called 'Examples,' which will show annotated examples of the collocation in context selectively extracted from the ADVENCOR corpus.[194]

Figure 69 and Figure 70 represent the template used for the implementation of collocations in *DicoAdventure*. The former shows the 'Examples' column unexpanded and the latter shows this column expanded.

[192] As it was described in Chapter 3, the compilers of the LDOCE also trust quantitative data to arrange the meanings of words within the entries, using the frequency parameter to show the most common meanings first.
[193] This idea is conceived by L'Homme (2010, p. 143). Furthermore, we believe that the tags used are comprehensible to laypersons.
[194] The inclusion of annotated examples was one of the suggestions offered in Section 3.3. for the improvement of specialized resources.

Figure 69

Template for the Implementation of Collocations in DicoAdventure *('Examples' Column Unexpanded)*

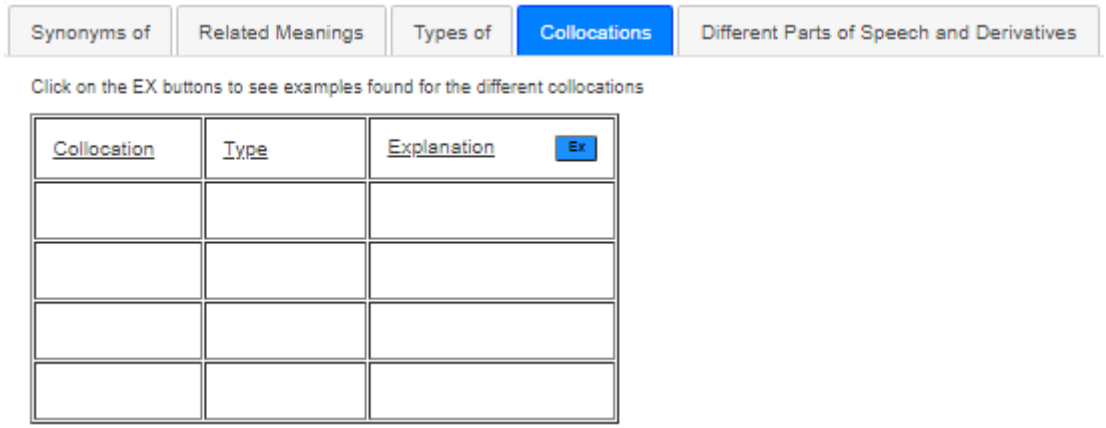
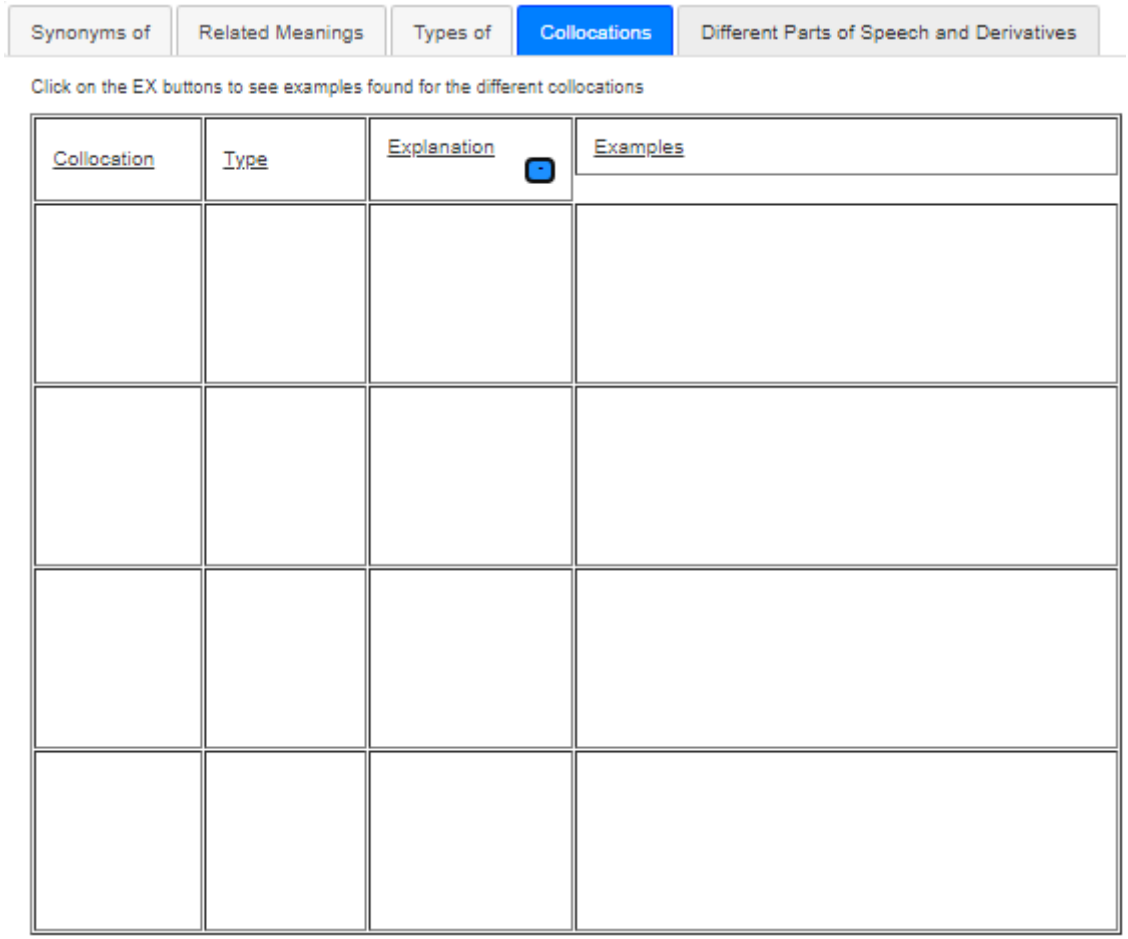


Figure 70

Template for the Implementation of Collocations in DicoAdventure ('Examples' Column Expanded)



As it can be observed, the model suggested for the implementation of collocations in the verb entries of *DicoAdventure* incorporates the items of information described in Section 3.4., in which a set of guidelines for collocational description in the entries of a comprehensive specialized dictionary was proposed, namely: (1) the lexical type of the collocation (e.g., verb + noun, verb + adverb), (2) an explanation of the collocation that laypersons can understand, and (3) corpus annotated sentences showing the collocation in context. This template will be used in Chapter 5 to show the results obtained in this dissertation.

CHAPTER 5. RESULTS

Since terms are highly subject matter specific, it is possible to identify single-worded terms on the bases of their frequencies of occurrence and distribution. Multiworded terms are identified on the basis of their collocational behaviour.

(Yang, 1986, p. 93)

Chapter 5 will examine the results obtained after the employment of the methodology described in Chapter 4 and will be organized into the following five sections. Section 5.1. will deal with the verbal terminology extracted from the ADVENCOR corpus and will show the final list of candidate motion verbs that were potential bases of the collocations implemented in *DicoAdventure*, whose extraction will be covered in Section 5.2. Then, Section 5.3. will provide the list of verbs which produced collocations meeting all eligibility criteria and will explain the classification of the selected collocations in terms of the meaning expressed. After that, Section 5.4. will perform an in-depth analysis of the five motion verbs which produced collocations representing actual motion and fictive motion. Finally, Section 5.5. will focus on the implementation of the collocations in the specialized dictionary of adventure tourism.

5.1. Terminology extraction: Candidate motion verbs

As it was explained in Chapter 4 (cf. §4.4.1.), terminology extraction was the first step of the methodology applied in this dissertation. It consisted in retrieving all candidate verbs representing motion from the ADVENCOR corpus, given that they would be the bases of the collocations to be implemented in *DicoAdventure*. With this objective in mind, we made use of the Keywords function of Sketch Engine, since it compares the same words in the specialized corpus and the reference corpus (the 'enTenTen13' was chosen) in order to demonstrate that the extracted verb terms are typical of the SL under study (i.e., adventure tourism). Thus, after setting the parameters for the extraction, that is, lemmata

whose POS was "verb" and occurred at least two times in the corpus, the resulting list of verbs contained 978 items; an extract from the results is shown in Figure 71:

Figure 71

Extract From the List of Candidate Motion Verbs Extracted From the ADVENCOR Corpus

	Word	Frequency		Relative freq.		Score
		Focus	Reference	Focus	Reference	
1	skydive-v	499	11,746	423.165	0.517	279.65
2	raft-v	756	58,426	641.107	2.571	179.83
3	rappel-v	244	6,607	206.918	0.291	161.09
4	trek-v	939	95,972	796.296	4.223	152.67
5	canyoning-v	163	912	138.228	0.04	133.86
6	abseil-v	188	4,984	159.429	0.219	131.58
7	glide-v	340	83,103	288.329	3.656	62.14
8	mountaineer-v	97	7,938	82.258	0.349	61.71
9	canoe-v	190	37,054	161.125	1.63	61.64
10	hike-v	1,187	375,630	1,006.606	16.527	57.49
11	parachute-v	97	10,874	82.258	0.478	56.32
12	paddle-v	328	96,253	278.152	4.235	53.33
13	paragliding-v	53	1,361	44.945	0.06	43.35
14	climb-v	1,960	864,430	1,662.13	38.033	42.61
15	snowshoe-v	56	11,305	47.489	0.497	32.38
16	exhilarate-v	148	66,304	125.508	2.917	32.3
17	belay-v	42	4,274	35.617	0.188	30.82
18	cave-v	93	41,048	78.866	1.806	28.46
19	ziplining-v	32	354	27.137	0.016	27.71
20	zip-v	109	67,419	92.435	2.966	23.56
21	canyoneering-v	22	223	18.657	0.01	19.47
22	soar-v	183	185,186	155.189	8.148	17.07
23	book-v	565	636,826	479.134	28.019	16.55
24	spelunk-v	20	2,244	16.961	0.099	16.35
25	adventure-v	33	21,774	27.985	0.958	14.8
26	parasailing-v	17	1,166	14.416	0.051	14.66
27	jump-v	853	1,131,798	723.366	49.796	14.26
28	camp-v	176	218,982	149.252	9.635	14.13
29	snorkel-v	63	69,122	53.426	3.041	13.47
30	bouldering-v	16	1,990	13.568	0.088	13.4
31	Canyoning-v	14	106	11.872	0.005	12.81
32	traverse-v	63	78,274	53.426	3.444	12.25
33	sightsee-v	21	12,598	17.809	0.554	12.1
34	mushing-v	13	640	11.024	0.028	11.7
35	sled-v	14	2,593	11.872	0.114	11.55
36	ascend-v	89	128,573	75.474	5.657	11.49
37	roof-v	24	19,886	20.353	0.875	11.39
38	canopy-v	16	6,396	13.568	0.281	11.37
39	descend-v	176	284,354	149.252	12.511	11.12
40	boat-v	50	67,857	42.401	2.986	10.89
41	reschedule-v	35	42,622	29.681	1.875	10.67
42	pilot-v	39	53,511	33.073	2.354	10.16
43	bike-v	36	48,158	30.529	2.119	10.11
44	ski-v	52	80,263	44.097	3.531	9.95
45	snorkelling-v	13	4,968	11.024	0.219	9.87
46	scramble-v	70	119,360	59.362	5.252	9.66
47	guide-v	476	935,097	403.66	41.142	9.6
48	snowmobiling-v	11	2,083	9.328	0.092	9.46
49	fly-v	772	1,611,018	654.676	70.88	9.12
50	dive-v	123	250,441	104.307	11.019	8.76

Rows per page: 50 1–50 of 978 1 / 20

Nevertheless, the manual work carried out to evaluate these results and to achieve desirable final outcomes led us to discard 860 units because:

1) They belonged to POSs other than verb, such as *unique* (adjective), *guaranty* (noun), *difficult* (adjective);[195]
2) They were abbreviations, for instance, *yrs*, *lbs*, *gwt*;
3) They had no relation to the SL of adventure tourism, for example, *pride*, *roof*, *gorge*;
4) They were not regarded as motion verbs,[196] such as *reserve*, *unsubscribe*, *cancel*; and
5) They displayed distinct lemmata of the same motion verb, for instance, *paraglided*, *paraglides*, *paragliding* for *paraglide*.

[195] It demonstrates that automatic taggers are not 100% accurate (Pęzik, 2018, p. 66; Vargas Sierra, 2010, p. 29), as it was stated in Chapter 4.
[196] We define motion verbs as "verbs describing a displacement of an entity, either a person or a thing, in space" (Durán & L'Homme, 2020, p. 43).

Therefore, the final list of candidate motion verbs totaled 118 items, which are displayed in Table 22 in descending order according to their keyness score (simple math). It also shows the frequency of these verb terms in the ADVENCOR corpus ('Freq') and their frequency in the reference corpus ('Ref freq'), the 'enTenTen13:'

Table 22

Final List of Candidate Motion Verbs Extracted From the ADVENCOR Corpus

	Term	Keyness score	Freq	Ref freq
1.	*skydive*-v	279.650	499	11,746
2.	*raft*-v	179.830	756	58,426
3.	*rappel*-v	161.090	244	6,607
4.	*trek*-v	152.670	939	95,972
5.	*abseil*-v	131.580	188	4,984
6.	*glide*-v	62.140	340	83,103
7.	*canoe*-v	61.640	190	37,054
8.	*hike*-v	57.490	1,187	375,630
9.	*parachute*-v	56.320	97	10,874
10.	*paddle*-v	53.330	328	96,253
11.	*climb*-v	42.610	1,960	864,430
12.	*snowshoe*-v	32.380	56	11,305
13.	*cave*-v	28.460	93	41,048
14.	*zip*-v	23.560	109	67,419
15.	*soar*-v	17.070	183	185,186
16.	*jump*-v	14.260	853	1,131,798
17.	*snorkel*-v	13.470	63	69,122
18.	*traverse*-v	12.250	63	78,274
19.	*sled*-v	11.550	14	2,593
20.	*ascend*-v	11.490	89	128,573
21.	*descend*-v	11.120	176	284,354
22.	*boat*-v	10.890	50	67,857
23.	*pilot*-v	10.160	39	53,511
24.	*bike*-v	10.110	36	48,158
25.	*ski*-v	9.950	52	80,263
26.	*scramble*-v	9.660	70	119,360

	Term	Keyness score	Freq	Ref freq
27.	*guide*-v	9.600	476	935,097
28.	*fly*-v	9.120	772	1,611,018
29.	*dive*-v	8.760	123	250,441
30.	*swim*-v	8.460	189	410,502
31.	*explore*-v	7.050	732	1,983,118
32.	*zipline*-v	6.920	7	68
33.	*dogsled*-v	6.900	7	116
34.	*meander*-v	6.640	24	50,355
35.	*cycle*-v	6.630	47	117,391
36.	*ford*-v	6.320	9	8,316
37.	*surf*-v	6.100	57	160,999
38.	*cascade*-v	6.030	14	25,807
39.	*pedal*-v	5.980	20	45,581
40.	*retrace*-v	5.690	13	25,342
41.	*tour*-v	5.600	96	311,535
42.	*ride*-v	5.570	371	1,266,100
43.	*depart*-v	5.320	91	311,150
44.	*criss-cross*-v	5.230	8	11,086
45.	*trail*-v	4.840	50	180,898
46.	*tube*-v	4.790	11	26,292
47.	*tramp*-v	4.780	7	10,221
48.	*cross*-v	4.610	252	1,035,186
49.	*tow*-v	4.200	27	106,501
50.	*capsize*-v	4.190	7	14,861
51.	*immerse*-v	4.070	34	144,039
52.	*circumnavigate*-v	3.970	5	7,246
53.	*clamber*-v	3.850	7	18,173
54.	*navigate*-v	3.830	77	371,212
55.	*float*-v	3.800	73	353,404
56.	*plummet*-v	3.770	14	54,941
57.	*exit*-v	3.420	41	215,237
58.	*transport*-v	3.300	71	399,169
59.	*ramble*-v	3.260	12	55,187
60.	*board*-v	3.250	36	197,567

	Term	Keyness score	Freq	Ref freq
61.	*disembark*-v	3.190	6	20,678
62.	*race*-v	3.170	71	415,472
63.	*walk*-v	3.140	641	3,913,303
64.	*cruise*-v	3.130	44	255,399
65.	*sledge*-v	3.130	3	3,023
66.	*slide*-v	3.070	78	474,016
67.	*land*-v	2.980	95	599,449
68.	*travel*-v	2.980	369	2,374,201
69.	*wade*-v	2.920	14	77,395
70.	*paraglide*-v	2.680	2	119
71.	*flow*-v	2.680	98	691,335
72.	*wind*-v	2.680	65	454,073
73.	*scale*-v	2.660	33	225,359
74.	*leap*-v	2.540	26	183,912
75.	*plunge*-v	2.500	18	125,251
76.	*shuttle*-v	2.470	4	17,744
77.	*toboggan*-v	2.430	2	2,485
78.	*sail*-v	2.370	32	247,076
79.	*crawl*-v	2.370	27	206,575
80.	*head*-v	2.350	235	1,917,808
81.	*embark*-v	2.330	28	218,878
82.	*bicycle*-v	2.310	6	37,141
83.	*arrive*-v	2.180	266	2,334,730
84.	*reach*-v	2.140	452	4,058,294
85.	*venture*-v	2.120	21	178,835
86.	*steer*-v	2.120	30	261,078
87.	*skate*-v	2.040	8	63,805
88.	*hop*-v	1.970	17	154,727
89.	*anchor*-v	1.920	15	139,419
90.	*jog*-v	1.920	15	139,727
91.	*journey*-v	1.810	8	75,042
92.	*row*-v	1.760	4	34,125
93.	*hover*-v	1.680	12	128,185
94.	*rush*-v	1.640	49	565,664

	Term	Keyness score	Freq	Ref freq
95.	*transfer*-v	1.620	80	945,060
96.	*accelerate*-v	1.470	29	373,197
97.	*drive*-v	1.460	288	3,794,676
98.	*trickle*-v	1.430	4	47,116
99.	*loop*-v	1.420	5	61,023
100.	*lag*-v	1.370	6	78,030
101.	*speed*-v	1.350	28	394,681
102.	*edge*-v	1.330	8	110,167
103.	*buoy*-v	1.330	2	23,357
104.	*stroll*-v	1.310	9	127,234
105.	*launch*-v	1.290	117	1,747,258
106.	*dunk*-v	1.280	4	55,107
107.	*curve*-v	1.270	11	162,069
108.	*recede*-v	1.260	4	56,706
109.	*pace*-v	1.250	11	165,476
110.	*pass*-v	1.250	267	4,123,840
111.	*budge*-v	1.240	2	26,585
112.	*overtake*-v	1.230	7	105,025
113.	*flutter*-v	1.230	3	42,650
114.	*run*-v	1.160	535	8,901,493
115.	*enter*-v	1.150	192	3,218,337
116.	*roll*-v	1.150	66	1,106,044
117.	*fall*-v	1.070	215	3,864,378
118.	*drop*-v	1.030	109	2,030,253

Table 22 represents the completion of the first step of our methodology, that is, terminology extraction. As the selected motion verbs were to be the bases of the collocations to be implemented in *DicoAdventure*, the following section explains how collocations were extracted from the ADVENCOR corpus.

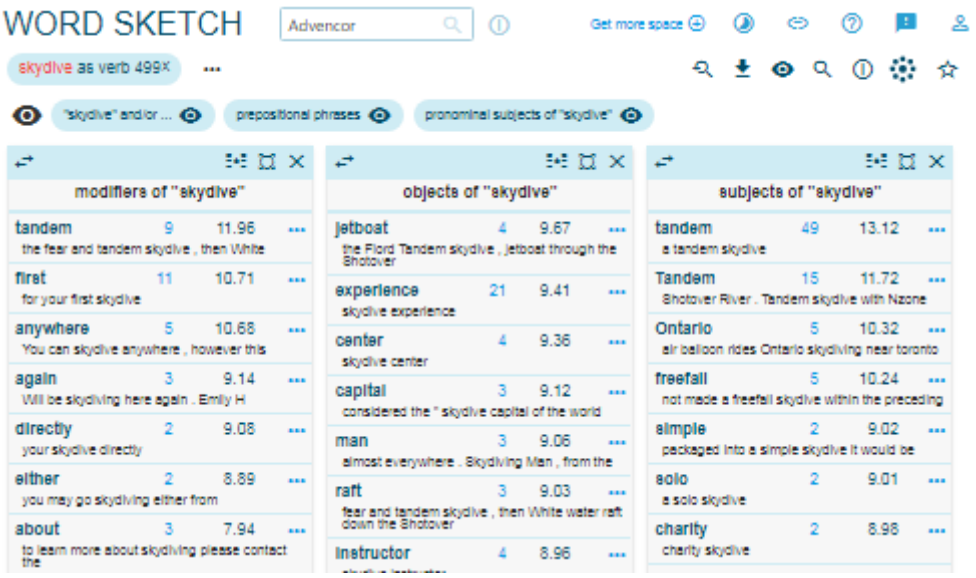
5.2. Phraseological extraction: Collocations

After the selection of the motion verbs typical of our specialized corpus, the second step of our methodology (cf. §4.4.2.) involved the extraction of the collocations produced by

each of these verbs. For so doing, the Word Sketch function of Sketch Engine was relied on, given that it automatically shows the collocational behavior of a lemma by adopting a hybrid approach:[197] (1) statistical, because collocations are detected by implementing the *logDice* measure, which tells how significant the association between the two elements (base and collocate) is, and (2) linguistic, because it extracts potential collocates in terms of the syntactic relationship held with the base; for this reason, the results are organized according to the function performed by the collocates, for example, modifier, subject, object. Nevertheless, we did not consider all the results achieved for each verb, since we were interested in lexical collocations, that is, the types Le2-Verb + noun, Le3-Noun + verb, Le6-Adverb + verb/verb + adverb and Le7-Verb + adjective included in Table 12 (cf. §2.2.2.) (it must be highlighted that no Le7 collocations were retrieved from the specialized corpus); thus, combinations with prepositions were ignored. Figure 72 shows an extract from the word sketch of the verb with the highest keyness score in the ADVENCOR corpus, *skydive*-v:

Figure 72

Extract From the Word Sketch of the Verb skydive-*v in the ADVENCOR Corpus*



At this point, it must be remembered that two were the criteria established for the phraseological extraction in this work: first, the *logDice* score of the collocations should

[197] According to Vargas Sierra (2010, p. 33), this is the ideal approach when automatically extracting combinations of terms.

be of ≥ 7 points, for this score means that the collocation is significant and can be considered typical of the corpus under analysis, thus, it is worthy of implementation in *DicoAdventure*, and second, the collocation should have at least two tokens in the ADVENCOR corpus. This mixed approach toward the extraction of collocations, whereby an association measure and a frequency cutoff threshold were used, helped to obtain better results and discard arbitrary combinations (Evert, 2009, p. 1215). Nevertheless, although Sketch Engine works according to reliable parameters and is usually quite effective, a major error detected after the extraction of collocations was that, in some cases, the verb did not function as a verb (e.g., *helicopter skydive*, where *skydive* functions as a noun), so manual intervention was also required at this step. Apart from that, additional human work was needed so as to make sure that the collocations fulfilling both extraction criteria were specialized collocations of the SL of adventure tourism. Therefore, their *logDice* score was compared with the score obtained for the same collocation in the reference corpus, that is, the 'enTenTen13.' When it was higher in the ADVENCOR corpus, we could readily admit that the collocation was typical of the domain under study and, then, it could be regarded as a specialized collocation.[198]

As it was previously mentioned, the list of motion verbs selected in step number one of our methodology contained 118 verbs that could be regarded as candidates for the bases of the collocations to be extracted in this step and implemented in *DicoAdventure*. Nevertheless, not all of them became successful candidates, given that the number of verbs that produced collocations fulfilling the established criteria was 55, which are displayed in Table 23 in alphabetical order:

[198] About 5% of the collocations selected achieved a higher *logDice* score in the reference corpus. However, they were still considered relevant to this study for several reasons, as we will explain in §6.2.

Table 23

Motion Verbs Which Produced Collocations in the ADVENCOR Corpus

1. *abseil*-v	**12.** *enter*-v	**23.** *hike*-v	**34.** *raft*-v	**45.** *slide*-v
2. *arrive*-v	**13.** *exit*-v	**24.** *jump*-v	**35.** *rappel*-v	**46.** *steer*-v
3. *ascend*-v	**14.** *explore*-v	**25.** *land*-v	**36.** *reach*-v	**47.** *swim*-v
4. *board*-v	**15.** *fall*-v	**26.** *launch*-v	**37.** *retrace*-v	**48.** *transfer*-v
5. *cascade*-v	**16.** *float*-v	**27.** *leap*-v	**38.** *ride*-v	**49.** *transport*-v
6. *climb*-v	**17.** *flow*-v	**28.** *navigate*-v	**39.** *run*-v	**50.** *travel*-v
7. *cross*-v	**18.** *fly*-v	**29.** *paddle*-v	**40.** *rush*-v	**51.** *traverse*-v
8. *depart*-v	**19.** *ford*-v	**30.** *pass*-v	**41.** *scale*-v	**52.** *trek*-v
9. *descend*-v	**20.** *glide*-v	**31.** *pedal*-v	**42.** *scramble*-v	**53.** *venture*-v
10. *drive*-v	**21.** *guide*-v	**32.** *plunge*-v	**43.** *ski*-v	**54.** *walk*-v
11. *drop*-v	**22.** *head*-v	**33.** *race*-v	**44.** *skydive*-v	**55.** *wind*-v

After having extracted the collocations whose bases were the 55 motion verbs represented in Table 23 from the specialized corpus, their concordance lines were examined with different intentions, which will be described in the following section.

5.3. Collocation classification: Meaning represented

The last step of the method employed in this study (cf. §4.4.3.) focused on the classification of the collocations extracted into two categories according to the meaning conveyed by the base of the collocation, that is, the verb: (1) actual motion, where the verb represents the displacement of people in an adventure activity, so the entity is actually moving in space, for instance, ***You*** *also* ***ascend*** *more slowly, walking up the long valley*, and (2) fictive motion, where the verb describes an inanimate object performing a motion, but it is completely static in the real world,[199] for example, *Then the* ***route ascends*** *several pitches of vertical ice up a narrow curtain*.[200] The Sketch Engine function used at this point was Concordance, as it allows the user to access examples of the collocation in context taken from the corpus under analysis. The criteria established when checking the contexts of each collocation were two: first, the collocation should occur in

[199] Fictive motion verbs were noticed by Piccioni & Pontrandolfo (2019), who concluded that metaphors are present in motion verbs whose inanimate subject is represented as if it were moving (cf. §2.3.).
[200] Both examples were retrieved from the ADVENCOR corpus.

at least two contexts describing the same type of motion, and second, these two contexts should drop a specific reference to the domain of adventure tourism. These two criteria led us to discard almost half (45%) of the 55 motion verbs whose word sketches had produced collocations in step number two (cf. Table 23). In other words, the final list of motion verbs whose collocations fulfilled all eligibility criteria and are covered in this dissertation totals 30 items, which are represented in Table 24:

Table 24

Motion Verbs Whose Collocations Met all Eligibility Criteria

1. *abseil*-v	**7.** *drive*-v	**13.** *glide*-v	**19.** *paddle*-v	**25.** *slide*-v
2. *arrive*-v	**8.** *enter*-v	**14.** *head*-v	**20.** *reach*-v	**26.** *swim*-v
3. *ascend*-v	**9.** *exit*-v	**15.** *hike*-v	**21.** *ride*-v[201]	**27.** *traverse*-v
4. *climb*-v	**10.** *fall*-v	**16.** *jump*-v	**22.** *scale*-v	**28.** *trek*-v
5. *cross*-v	**11.** *float*-v	**17.** *land*-v	**23.** *scramble*-v	**29.** *venture*-v
6. *descend*-v	**12.** *fly*-v	**18.** *navigate*-v	**24.** *skydive*-v	**30.** *walk*-v

In short, 30 motion verbs were finally selected as the bases of the collocations to be implemented in *DicoAdventure*, for the collocations that they produced satisfied all the criteria established through the whole methodological process which took place in this research. Nonetheless, this chapter will centre on the motion verbs whose collocations represented both types of motion, that is, actual and fictive,[202] which are: *ascend*-v, *climb*-v, *cross*-v, *descend*-v and *head*-v,[203] whose entry number one (e.g., *ascend*$_1$) includes collocations which display actual motion performed by an animate entity, such as *ascend slowly*, whereas their entry number two (e.g., *ascend*$_2$) includes collocations which convey fictive motion performed by an inanimate entity, for instance, *(a) path climb*.[204]

Aside from that, the Concordance function also allowed us to discard verbs, and hence collocations, whose meaning could not be tested in at least 10 different illustrative contexts. To put it differently, in order to create a new entry for a verb depicting fictive

[201] Both meanings of the verb *ride*-v included in *DicoAdventure*, that is, *ride*$_1$ and *ride*$_2$, produced collocations.
[202] It must be remembered that *DicoAdventure* is a concept-based dictionary, therefore, each distinct meaning of the same term is described in a different entry.
[203] The rest of the verbs together with their collocations are included in the Appendix in alphabetical order.
[204] It must be emphasized that, in this work, the collocations extracted will be represented by the lemmata of both elements.

motion in *DicoAdventure*, 10 distinct knowledge-rich contexts representing this meaning should be identified in the ADVENCOR corpus, that is, contexts showing the behaviour of the verb and its main arguments.[205] For example, it happened with the verbs *reach*-v and *traverse*-v, whose entries belonging to type two could not be created. Additionally, when the concordance lines were manually checked and new collocations which had not been identified in the previous steps seemed to emerge from the contexts (e.g., noun + verb collocations where the noun was the inanimate subject of a fictive motion event but had not been retrieved as such[206]), a deeper analysis using the Collocations option included in the Concordance function was performed.[207] This process made it possible the identification of collocations which deserved their inclusion in the dictionary, given that they fulfilled the criteria established for the collocation extraction (minimum *logDice* score and frequency) and were identified in two different contexts conveying the same meaning and making a specific reference to the domain of adventure tourism, as we will explain in the following section.

5.4. Verbs whose collocations represent actual and fictive motion

As we have previously mentioned, this chapter will focus on those verbs which produced collocations representing both types of motion, that is, actual and fictive, in the ADVENCOR corpus, namely: *ascend*-v, *climb*-v, *cross*-v, *descend*-v and *head*-v. This choice is mainly due to the semantic richness of these verbs, which allows us to illustrate the methodology employed in this work and the analysis performed for the implementation of collocations in *DicoAdventure*. Nevertheless, it must be highlighted that the rest of the verbs (i.e., *abseil*-v, *arrive*-v, *drive*-v, *enter*-v, *exit*-v, *fall*-v, *float*-v, *fly*-v, *glide*-v, *hike*-v, *jump*-v, *land*-v, *navigate*-v, *paddle*-v, *reach*-v, *ride*-v,[208] *scale*-v, *scramble*-v, *skydive*-v, *slide*-v, *swim*-v, *traverse*-v, *trek*-v, *venture*-v and *walk*-v) and their collocations, included in the Appendix, will be also considered to carry out the analysis presented in Chapter 6. The results provided below for the five verbs selected are organized in the following way:

[205] It must be remembered that, in the case of verbs depicting actual motion (whose entries had already been created in *DicoAdventure* at the time of this research), 20 contexts were needed (Durán-Muñoz & L'Homme, 2020, p. 44).
[206] These special cases were also checked in the Word Sketch function of Sketch Engine.
[207] Further details of this option were given in §4.4.2.
[208] The results for *ride*$_1$ and *ride*$_2$ are included.

1) First, a table containing the collocations retrieved along with their *logDice* score and frequency in both corpora, that is, the specialized corpus (the ADVENCOR corpus) and the reference corpus (the 'enTenTen13'), is shown, which will help to unveil whether the collocations selected for their implementation in *DicoAdventure* are typical or not of the SL of adventure tourism. This table will include all the collocations that each verb produced in the specialized corpus, therefore, different fonts have been used to indicate the type of motion represented, namely: (1) collocations depicting actual motion (belonging to type one) appear in normal font, (2) collocations portraying fictive motion (belonging to type two) are in **bold type**, and (3) collocations which represent both types of motion are underlined; and

2) Second, the different meanings of the verbs, that is, actual motion (type one) and fictive motion (type two), will be covered separately, since our dictionary is concept-based and contains one entry for every verb which has a different argument structure. Each of the sections devoted to each of the meanings of a verb will show a table with four columns representing: (1) the collocation evoking the meaning of the term;[209] (2) its lexical type, for instance, verb + adverb, noun + verb; (3) a brief explanation of the meaning of the collocation, which uses the semantic roles in the argument structure of the verb; and (4) corpus sentences extracted from the ADVENCOR corpus and representing the collocation in context; these examples will show the base and the collocate of the collocation in **bold type** and, in the cases of the collocations included in types two of the verbs (i.e., those depicting fictive motion), the inanimate subject of the verb which seems to be moving is underlined in the sentence. These four items make the information that will be implemented in *DicoAdventure*, as it was explained in Section 4.5.

The order chosen to present the five selected verbs is alphabetical, that is, *ascend*-v (§5.4.1.), *climb*-v (§5.4.2.), *cross*-v (§5.4.3.), *descend*-v (§5.4.4.) and *head*-v (§5.4.5.).

[209] The collocations are presented in descending order according to their *logDice* score in the ADVENCOR corpus, that is, from more to less strong/specialized collocations.

5.4.1. *ascend*-v

The results obtained for *ascend*-v in the Word Sketch function of Sketch Engine were an object, a modifier and three subjects of the verb. The collocations selected together with their *logDice* score and frequency in both corpora, specialized and reference, are shown in descending order in Table 25:

Table 25

logDice *Score and Frequency of the Collocations Containing* ascend-*v*

Collocation	***logDice* in ADVENCOR**	**Frequency in ADVENCOR**	***logDice* in 'enTenTen13'**	**Frequency in 'enTenTen13'**
ascend (a) slope	11.3	5	7.05	446
ascend slowly	10.16	2	4.29	558
(a) climb ascend	10.11	2	3.54	16
(a) trail ascend	9.48	8	6.41	602
(a) route ascend	8.43	2	4.37	118

The collocations selected for *ascend*-v were five. The most common lexical type is noun + verb (three types), where the noun is the subject of the action. Then, there is one verb + noun and one verb + adverb collocations. As it can be observed, the *logDice* scores in the specialized corpus are much higher than those in the reference corpus, with a difference of 5.96 points on an average. In fact, the *logDice* scores of *ascend slowly* and *(a) climb ascend* in the ADVENCOR corpus double those in the 'enTenTen13,' and the *logDice* of *(a) route ascend* in the specialized corpus almost doubles that in the reference corpus. These results support the idea that the collocations extracted from the ADVENCOR corpus are highly specialized, despite being apparently less frequent than in the 'enTenTen13.'

The strongest collocation extracted from the ADVENCOR corpus is *ascend (a) slope* (11.3 points), which describes fictive motion. However, it does not imply that it has also the highest number of tokens, since *(a) trail ascend* is more recurrent (5 vs. 8 tokens, respectively). On the other hand, *ascend slowly* presents a *logDice* score of 10.16 points with only two tokens in the specialized corpus. This fact is also appreciated in the results achieved in the reference corpus and proves the claim that *logDice* scores and frequencies of collocations are not interrelated, for collocations assigned the highest scores are not

necessarily the most recurrent ones, such as *ascend (a) slope* (7.05 points, 446 tokens) versus *ascend slowly* (4.29 points, 558 tokens). At this point, it must be remembered that the *logDice* score measures typicality, that is, it extracts collocations which are not (completely) predictable and are considered stronger, while frequency produces the most common (and, probably, weak) collocations, such as collocations that are widely known.

Regarding the meaning of the verb, *ascend*-v describes vertical motion, more specifically, upward direction of motion, which involves a subject moving to a higher position. Its use can be transitive or intransitive. For instance, in the former case, the path through which the movement takes place or the height reached can be mentioned, for instance, *Go into the sea to witness the underwater creatures with snorkelling or challenge yourself to* ***ascend*** *tall granite hills with rock climbing*; *A three to four-hour hike in which time you* ***ascend*** *about 2,000 metres*. In the latter case, the speed at which the motion is carried can be specified, for example, *We* ***ascend*** *gradually so that we have optimum time to acclimatize to the high-altitude environment*. With respect to the subject performing the motion, it can be an animate entity (type one), like in the collocation *ascend slowly*, or an inanimate entity perceived as if it were moving (type two), like in *ascend (a) slope, (a) trail ascend* and *(a) route ascend*. As a result, two entries, *ascend*$_1$ and *ascend*$_2$, are available in *DicoAdventure*.

ascend$_1$

ascend$_1$ includes collocations representing actual motion performed by an animate entity and is defined in the following terms: "A TOURIST moves in an upward DIRECTION along a PATH from SOURCE to DESTINATION."[210] The only collocation illustrating this meaning is *ascend slowly:*

Collocation	Type	Explanation	Examples
ascend slowly	verb + adverb	~ at a specific SPEED	(1) You also **ascend** more **slowly**, walking up the long valley. (2) Roam the wide-open skies and taste your own personal rainbow as you **ascend slowly** into the Florida Skies!

[210] The definitions provided are those gathered in *DicoAdventure*.

The collocation *ascend slowly* represents an intransitive use of the verb in combination with an adverb. The collocate tells the SPEED at which the motion is carried and is modified by *more* in context (1), which means that this speed can be described along a scale. Finally, it must be said that the Concordance function showed no concordance lines for *slowly ascend*, so it can be inferred that the adverb only occurs after the verb in the ADVENCOR corpus.

ascend$_2$

ascend$_2$ displays fictive motion performed by an inanimate entity and is defined in the following terms: "A PATH has an upward course (DIRECTION) that departs from a SOURCE through a PLACE to a DESTINATION." The collocations illustrating this meaning are *ascend (a) slope, (a) climb ascend, (a) trail ascend* and *(a) route ascend:*

Collocation	Type	Explanation	Examples
ascend (a) slope	verb + noun	~ a specific PLACE	(3) The first pitch goes up a lower angled WI3 gully to the bottom of the second pitch. It then **ascends** this snow **slope** to the beginning of the crux pitch. (4) The Moraine Trail from Snowshed Lodge to the top of Snowshed Slope **ascends** a gentle **slope** from Snowshed Pond, through wooded and glacial moraine before joining the Wildlands Trail.
(a) climb ascend	noun + verb	A PATH ~s	(5) The scenery becomes more remarkable as the **climb ascends** toward Larkya La. (6) The **climb ascends** bulges for over 300 feet and then splits into two gullies.
(a) trail ascend	noun + verb	A PATH ~s	(7) A number of **trails ascend** to the top, including the most popular, the 13-mile Barr Trail. (8) The **trails ascend** step by step till Ulleri, then the trails lead you gradually up to Ghorepani through the rhododendron forest and vegetation where you can see birds and Fish Tails mountain.

Collocation	Type	Explanation	Examples
(a) route ascend	noun + verb	A PATH ~s	(9) Then the **route ascends** several pitches of vertical ice up a narrow curtain. (10) Pitch 1 and 2 climb straight up the ice flow. Then the **route ascends** a WI3 snow gully.

Two are the lexical types identified in the four collocations containing *ascend*-v representing fictive motion. First, the verb + noun collocation, *ascend (a) slope*, expresses motion taking place through a PLACE, that is, the verb is used transitively and the collocate corresponds with the object of the sentence. It occurs three positions to the right of the verb, as it can be observed in examples (3) and (4). The intervening words add distinguishing features of the slope, for example, the material: *snow*, and the steepness of the slope: *gentle*. Second, the noun + verb collocations *(a) climb ascend*, *(a) trail ascend* and *(a) route ascend* describe a PATH performing the motion, that is, the collocates correspond with the subjects of the sentences, which occur right next to the verb (-1). An interesting fact is that the second collocate, *trail*, is also the subject in example (4) of *ascend (a) slope*, although in this case the subject is located within a span of -10 from the verb. On the other hand, the subject in context (3) can only be identified if a larger context is examined. To explain, the subject occurs within a span of -2 from the verb, but it remains unspecified due to the use of the personal pronoun *it*. Thus, it is necessary to extend the span up to -17 words from the base so as to know that *it* corresponds to *pitch*, which occurs in a different sentence. A key point is that *pitch* was not identified as a collocate of *ascend*-v either in the word sketch of the verb or through the Collocations option available in the Concordance function.

Summing up, five were the collocations extracted for *ascend*-v. One of them represents actual motion: *ascend slowly*, and four of them represent fictive motion: *ascend (a) slope, (a) climb ascend (a) trail ascend* and *(a) route ascend*. In terms of the lexical types, the most common one is noun + verb (three types, 60%), where the noun describes a PATH moving. Then, we find the combinations verb + adverb (actual motion, one type) and verb + noun (fictive motion, one type). With respect to the span between the elements of these collocations, it is slightly shorter in noun + verb (-1) and verb + adverb (+1 and +2) than in verb + noun (+3) collocations. It means that nominal collocates

acting as the subject of *ascend*-v co-occur in an immediate position of the verb, while when they act as the object they are modified by adjectives.

5.4.2. *climb*-v

The results obtained for *climb*-v in the Word Sketch function of Sketch Engine were objects, modifiers and subjects of the verb. The collocations selected together with their *logDice* score and frequency in both corpora, specialized and reference, are shown in descending order in Table 26:

Table 26

logDice *Score and Frequency of the Collocations Containing* climb-*v*

Collocation	***logDice* in ADVENCOR**	**Frequency in ADVENCOR**	***logDice* in 'enTenTen13'**	**Frequency in 'enTenTen13'**
climb (a) wall	10.34	40	9.09	13,731
climb steeply	10.16	5	8.4	1,428
climb up	10.11	32	5.98	34,922
climb (a) route	9.83	35	7.2	3,314
gradually climb	9.64	4	5.82	990
climb (a) mountain	9.52	24	10.09	16,288
climb uphill	9.43	3	7.18	606
climb down	9.21	4	5.89	6,883
climb (a) peak	8.92	16	8.25	3,892
(a) pitch climb	8.81	8	4.02	121
climb (a) face	7.89	7	4.69	945
(a) trail climb	7.76	6	7.07	1,635
climb (a) waterfall	7.57	6	4.73	267
climb (a) cliff	7.4	5	6.6	982
climb (a) tree	7.09	4	8.85	12,738

The collocations selected for *climb*-v were 15. The most common lexical type is verb + noun (eight types), followed by adverb + verb/verb + adverb (five types) and noun + verb (two types). As we can see, the *logDice* scores are, in general, higher in the specialized corpus than in the reference corpus, with a difference of 1.99 points on an average, so it

can be admitted that they are all specialized collocations. Nevertheless, there are two exceptions, *climb (a) mountain* and *climb (a) tree*, whose *logDice* scores are higher in the reference corpus (the difference is of 1.16 points on an average), showing that these two collocations are stronger in the 'enTenTen13.' Despite this, we think that they must be included in this analysis for three reasons: (1) both are closely linked to adventure tourism and the activity of climbing, (2) they are more typical of the SL of adventure tourism than the GL based on the normalized frequency per million in each corpus,[211] and (3) they satisfy the criteria established for the extraction of collocations.

The collocation showing the highest *logDice* is *climb (a) wall* (10.34 points), which is the most frequent in the ADVENCOR corpus (40 tokens). However, it does not imply that both figures increase (or decrease) simultaneously. For instance, *climb steeply*, which is very close in *logDice* score (10.16 points) to *climb (a) wall*, presents an extremely low frequency in the specialized corpus (five tokens). Conversely, *climb (a) peak* obtains a lower *logDice* score (8.92 points) but a higher frequency (16 tokens). Again, from these results we can infer that *logDice* scores and frequencies of collocations are not interrelated, in other words, they do not have an effect or depend on each other. A more illustrative example is *climb up* in the reference corpus, which being the most frequent collocation (34,922 tokens) displays one of the lowest *logDice* scores (5.98 points). It is, in all probability, due to the fact that *climb up* is a predictable collocation, given the meaning of the verb.

Climb-v expresses vertical motion, it being normally upward direction of motion, although downward direction is also possible. Thus, the verb describes a subject moving up or down and implies considerable difficulty. It can be used transitively and show the path that is climbed, for example, *You could **climb** the Sydney Harbour Bridge or take in the incredible views from the SkyPoint Observation Deck on the Gold Coast*, or intransitively and include the place that is climbed on, for instance, *Though the techniques used in ice climbing is the same as in rock climbing, the difference is that you need few more tools which will help you **climb** on vertical ice sheets*. The path where the movement

[211] This score allows linguists to compare frequencies in corpora of different sizes and is calculated by dividing the overall frequency of the collocations in the corpus by the number of words contained in the corpus and, then, the result is multiplied by 1,000,000. Thus, while the results in the ADVENCOR corpus are 20.35 and 3.39 for *climb (a) mountain* and *climb (a) tree*, respectively, the results obtained in the reference corpus are 0.82 and 0.64, respectively, which shows that these collocations would occur much more frequently in the specialized corpus.

is made is always construed as vertically oriented. Regarding the type of subject moving, it can be animate (i.e., a human being), or inanimate. In the former case, the motion is real and is represented by 10 of the collocations above, namely: *climb (a) wall, climb up, climb (a) route, climb (a) mountain, climb down, climb (a) peak, climb (a) face, climb (a) waterfall, climb (a) cliff* and *climb (a) tree.* In the latter case, there is no actual motion and is identified in seven of the collocations: *climb steeply, climb up, gradually climb, climb uphill, (a) pitch climb, climb (a) face* and *(a) trail climb*. Accordingly, *climb up* and *climb (a) face* are included in both entries for *climb*-v in *DicoAdventure*.

Before moving on to the analysis of *climb$_1$*, it must be mentioned that the strongest collocations retrieved from the ADVENCOR corpus through the Word Sketch function were *rock climb* and *ice climb*, with a *logDice* score of 13 and 11.76 points and 204 and 69 tokens, respectively. Nevertheless, they are not included in this work because we believe that they rather belong to the category of compounds for two reasons: first, they are "uninterruptable" associations of words (Mitchell, 1975, p. 129; cf. §2.2.3.), that is, unlike collocations, the elements in compounds always occur in the immediate positions, and second, both elements contribute to the meaning of the whole combination, thus, they describe two different types of climbing which concern specific places (i.e., rock and ice). Our decision about categorizing it as a compound chimes with the entry for *rock climb* in the CED,[212] which also spells it hyphenated: *rock-climb*.

climb$_1$

climb$_1$ includes collocations which portray an animate entity moving and is defined as follows: "A TOURIST follows a PATH on an inclined surface (PLACE), usually in an upward DIRECTION and with effort, from the ground (SOURCE) to a higher point (DESTINATION). Special safety equipment (SAFETY_INSTRUMENT) is often required." The collocations illustrating this meaning are *climb (a) wall, climb up, climb (a) route, climb (a) mountain, climb down, climb (a) peak, climb (a) face, climb (a) waterfall, climb (a) cliff* and *climb (a) tree:*

Collocation	Type	Explanation	Examples
climb (a) wall	verb + noun	~ a specific PATH	(11) As well we offer single climbing where you can use our

[212] https://www.collinsdictionary.com/dictionary/english/rock-climb (Last accessed: 25/10/2020).

Collocation	Type	Explanation	Examples
			auto-belay system allowing you to **climb** the **walls** without someone belaying you. (12) The vertical walls of fixed rope routes offer us the possibility to **climb** high **walls** under optimal safety conditions, due to the continuous fixation to a rope, the so called "line of life" supported by clamps and metallic hoops.
climb up	verb + adverb	~ in an upward DIRECTION	(13) The rock scramble is an easy 35 ft. rock face that participants will need to **climb up** using a rope as a hand line for assistance. (14) With the help of a rope, **climb up** through a gushing waterfall to find more secluded hot pools.
climb (a) route	verb + noun	~ a specific PATH	(15) He's **climbed** difficult **routes** on the Eiger, Matterhorn and fabled peaks in Europe and South America. (16) Using the famous via ferratas (Italian for "iron roads") we **climb** awesome **routes** on the craggy heights of Italy's dramatic Dolomites usually reserved only for expert rock climbers.
climb (a) mountain	verb + noun	~ a specific PATH	(17) Guests are attached to a safety cable system as they **climb** the **mountain**, enjoying the views of the canyon and Ogden on their way up. (18) When you go with us, you won't just **climb** the **mountain**; you'll meet the people who live on and around it and come away with a greater sense of alpine life.
climb down	verb + adverb	~ in a downward DIRECTION	(19) After you **climb down**, keep going about a quarter to a half a mile just past a big sand dune until you hit another drainage to the left. (20) I recommend that you bring a pair of gloves (cheap gardening gloves are fine) to protect your hands, especially when **climbing down**.

Collocation	Type	Explanation	Examples
climb (a) peak	verb + noun	~ up to the top of a PLACE	(21) It is not a good idea to venture out into the mountains alone – unless you happen to be a distant relative of the mythical Himalayan yeti (or an aspiring Reinhold Messner, the first person in the world to have **climbed** all fourteen 8000m **peaks**, including the first oxygen-less ascent and later the first solo ascent of Everest). (22) The last day we **climb** the highest **peak** of the Dolomites, the Punta Penia in Marmolada.
climb (a) face	verb + noun	~ a specific PATH	(23) Really hardcore rock climbers will often spend a number of days **climbing** a single **face**. (24) Some climbers have been known to **climb** the 5000-foot **face** in less time than most hikers.
climb (a) waterfall	verb + noun	~ a specific PATH	(25) We will provide you with the canyoning skills you need to **climb waterfalls**, jump into pristine pools and abseil down normally inaccessible waterfalls. (26) The Lowes started climbing in the rugged cliffs above their house when they were kids. Years later, they began **climbing** the steep **waterfalls** and snow couloirs.
climb (a) cliff	verb + noun	~ a specific PATH	(27) Remember the movie Lakshya, with actor Hrithik Roshan **climbing** a **cliff** as erect as a spine? Now imagine the same cliff cushioned with ice! (28) You start by **climbing** a **cliff** with the help of well-anchored bars in the cliff, you find yourself on obstacle course setup to work on agility, such as rope bridges, zip-lines, pendular trunks, suspension tunnels, hanging platforms...
climb (a) tree	verb + noun	~ a specific PATH	(29) Were you always **climbing trees** as a kid? Then get back to nature and branch out on a new adventure with a high ropes tree

Collocation	Type	Explanation	Examples
			climbing course or tree surfing experience. (30) Replicating the daily work of macaw researchers, we will **climb** a 30-meter **tree** with a rope, harness and jumar.

In total, *climb*$_1$ encompasses 10 collocations representing actual motion performed by an animate subject. The most recurrent lexical type is verb + noun (eight types), where the verb is used transitively and accompanied by a collocate revealing the PATH followed in the activity, for instance, *climb (a) wall, climb (a) route, climb (a) tree*. The remaining two collocations consist of the verb modified by an adverb describing the DIRECTION followed in the motion event, namely, upward: *climb up*, and downward: *climb down*. At this point, we must refer to the previously mentioned idea that the movement involves in *climb*-v usually develops upward rather than downward, which is evidenced in the number of tokens of every collocation retrieved from the specialized corpus (32 vs. 4, respectively).

A more in-depth analysis of the nominal collocates acting as the objects of the sentences confirms that the paths followed in the climbing activity are parts of the nature, such as a mountain, a waterfall or a cliff. Moreover, a remarkable feature is that the collocation *climb (a) peak* activates the PART FOR WHOLE metonymy, since it highlights the highest point of a mountain, so the motion refers to the whole process of climbing it. Furthermore, a closer look at the examples displaying the verb + noun collocations reveals that the usual span between the verb and the collocate is +2 (eight tokens), followed by +3 (five tokens), +1 (two tokens) and +4 (one token). In more than half of the examples (nine contexts), the collocate is modified by an adjective referring to the height (e.g., *high walls*), the level of difficulty (e.g., *difficult routes*) and the impressiveness (e.g., *awesome routes*) of the path, to name but a few. Additionally, measures are used in three of the contexts to describe the height of the path (e.g., *5000-foot face, 30-meter tree*). On the other hand, the span in *climb up* and *climb down* is +1, so there are no intervening words between the verb and the adverbial collocates.

climb₂

climb₂ displays fictive motion performed by an inanimate entity and is defined in the following terms: "A PATH usually follows an upward course (DIRECTION) that departs from a SOURCE to a DESTINATION in PLACE." The collocations illustrating this meaning are *climb steeply, climb up, gradually climb, climb uphill, (a) pitch climb, climb (a) face* and *(a) trail climb:*

Collocation	Type	Explanation	Examples
climb steeply	verb + adverb	~ in a specific MANNER	(31) The trail **climbs steeply** requiring participants to be hands free as they pull on trees and rocks to make their way up to the rock scramble. (32) The trail continues, sometimes **climbing steeply** up a series of rocky switchbacks over the north quarter dome.
climb up	verb + adverb	~ in an upward DIRECTION	(33) The first two pitches **climb up** to a wide ledge. (34) The approach goes past the Aspen Grove Falls and **climbs up** to the final switchback on the trail to the cirque above Aspen Grove.
gradually climb	adverb + verb	~ in a specific MANNER	(35) Look for a path that **gradually climbs** down from the ridge, starting to the right. (36) After your complete acclimatization, the trail heading you **gradually climbs** up until Duglha offering stunning views of Tawachee and Cholatse.
climb uphill	verb + adverb	~ in an upward DIRECTION	(37) Cross the bridge and turn right onto a small track, which **climbs uphill**. (38) Near the first house in the village (the one on the photo) you need to turn left, onto the dirt track **climbing uphill**.
(a) pitch climb	noun + verb	A PATH ~s	(39) The first two **pitches climb** up to a wide ledge.[213] (40) The first pitch goes up some small pillars and then the second **pitch climbs** over a 60-foot pillar

[213] Examples (33) and (39) show the same context, but the focus is on different collocations.

Collocation	Type	Explanation	Examples
			and a few shorter ones to the top of the route.
climb (a) face	verb + noun	~ a specific PLACE	(41) The route **climbs** the northwest **face** of the Tofana di Rozes by a series of natural rock ledges, gullies and open face scrambles.
			(42) On the small wall, we did an 80-foot route that **climbs** a short **face** to gain a finger crack and other intermittent cracks, has two bolts, takes gear up to 1.5 inches, could be 11b/c, and has NO bolted anchor.
(a) trail climb	noun + verb	A PATH ~s	(43) The **trail climbs** steeply requiring participants to be hands free as they pull on trees and rocks to make their way up to the rock scramble.[214]
			(44) At this spot the **trail climbs** a small hillside that has fairly unobstructed views of the river.

Seven collocations of *climb*-v depict fictive motion, being the most recurrent the adverb + verb/verb + adverb collocation (four types). In these cases, the collocates can modify the verb in two ways: (1) alluding to the MANNER that the movement adopts: *climb steeply* (in a difficult way) and *gradually climb* (slowly and in small stages), and (2) emphasizing the upward DIRECTION of the motion: *climb up* and *climb uphill*. In all the examples, the elements of the collocations co-occur next to each other and always in the same order, that is, there are no concordance lines showing *steeply climb* or *climb gradually*, for example.

As for the subjects of the verb in this type of collocation (which are found in spans ranging from -1 to -8), there are five different items. The most common one is *trail*, which is identified in three examples, followed by *track*, occurring in two examples, and, finally, *pitches*, *approach* and *path* have one token each. A significant point is that two of these subjects were retrieved as collocates of the verb, that is, *(a) pitch climb* and *(a) trail climb*, with a *logDice* score of 8.81 and 7.76, respectively. The rest of the subjects were checked

[214] Examples (31) and (43) show the same context, but the focus is on different collocations.

in the Word Sketch function, but they were not included in the resulting list. Nevertheless, two of them were extracted as collocates through the Concordance function:

1) *(a) path climb*: six tokens of *path* + *climb* were retrieved from the specialized corpus, with a *logDice* of 6.30 points. Their concordance lines were manually checked and we discovered that this word combination was a true collocation in three of the contexts, that is, in sentence (35) above and in the following examples: (1) *This is another beautiful day trip heading through a rocky loose **path** that follows the stream and **climbs** up to some Mani walls*, and (2) *Follow the **path**, it slowly **climbs** above the lake and then circumvents nameless hill from the right side.* The same sort of search in the reference corpus produced this collocation with a *logDice* score of 3.99 points, which leads to the conclusion that *(a) path climb* is a specialized collocation; and

2) *(a) track climb*: four tokens of *track* + *climb* were detected in the ADVENCOR corpus within a span of -5, with a *logDice* of 5.73 points. After manually checking their concordance lines, we discovered that it was a true collocation in three of the contexts, that is, in sentences (37) and (38) above and in the following: *For more adventures continue upwards along the wadi or take another **track climbing** out*. The *logDice* score of this word combination in the reference corpus (after setting identical parameters in the Concordance function) was 3.67 points, so we can state that *(a) track climb* is a specialized collocation of the SL of adventure tourism.

Nonetheless, neither *(a) path climb* nor *(a) track climb* were included in *DicoAdventure*, given that they did not fulfill the *logDice* criterion, that is, their *logDice* score was lower than 7 points. Regarding *approach*, we think that it was not retrieved as a collocate of *climb*-v because of the span between the two elements in (34), where the subject occurs in a range of -8 (it must be remembered that the Concordance function allows the user to make search within a span of -5/+5 at the maximum).

The second most recurrent collocation in $climb_2$ is noun + verb (two types). As it has been mentioned above, the word sketch of *climb*-v identified two nominal collocates acting as the subjects of the verb, included in the collocations *(a) pitch climb* and *(a) trail climb*. Thus, each collocation describes a different PATH carrying the motion. The span between both elements of the collocation is -1 in all contexts.

The last collocation we will comment on also presents a nominal collocate, but this time in the object position: *climb (a) face*. It describes the PLACE where the motion event takes place and co-occurs within a range of +3, since the words in the middle provide a feature of the noun, which is related to its direction (*northwest*) and its length (*short*). As for the subject of the sentences, *route* can be identified within a span of -1 (example (41)) and -2 (example (42)) from the verb. In order to know whether Sketch Engine recognized *route* as a nominal collocate of *climb*-v in the subject position, the Word Sketch and the Concordance functions were further explored, and the latter produced a result of 11 tokens of the word combination *route + climb* with a *logDice* score of 7.03 points. A manual check of the concordance lines unveiled one new context, added to the ones in (41) and (42), where *(a) route climb* was a true collocation: *This **route** goes up the main flow of ice. It is to the right of the main falls and to the left of the Bridal Veil Falls Right route. It **climbs** over several bulges and a steep final pitch*. As it can be observed, this collocation co-occurs within a span of -29 but is still significant (cf. §2.2.1.1.), as its *logDice* score is higher than 7 points and higher than in the reference corpus (where it got 5.59 points), which means that *(a) route climb* can be regarded as a specialized collocation. Furthermore, we think that it must be included in the entry $climb_2$, as it satisfies both criteria for the collocation extraction established in this dissertation, namely, *logDice* score and frequency.

To sum up, 15 collocations were selected for *climb*-v, from which ten represent actual motion performed by an animate entity: *climb (a) wall, climb up, climb (a) route, climb (a) mountain, climb down, climb (a) peak, climb (a) face, climb (a) waterfall, climb (a) cliff* and *climb (a) tree,* and seven depict fictive motion: *climb steeply, climb up, gradually climb, climb uphill, (a) pitch climb, climb (a) face* and *(a) trail climb*. It means that the collocations *climb up* and *climb (a) face* are included in $climb_1$ and $climb_2$.

As for the frequency of the lexical types extracted, the verb + noun collocation is the most common one, making more than half of the overall results (eight types, 53.33%). Except for one, which is found in actual and fictive motion examples (*climb (a) face*), all the verb + noun collocations represent actual motion and, in all cases, the noun refers to the PATH that is climbed. Second, the adverb + verb/verb + adverb collocation is identified in 33.33% of the selected collocations (five types) and, unlike the verb + noun collocation, this type is more common in fictive motion events (four types, 66.67%) than in actual motion events (two types, 33.33%). The collocates specify the DIRECTION of the

movement when it is real (e.g., *down*) and the MANNER and the DIRECTION when it is not real (e.g., *steeply, up*). Finally, the noun + verb combination occupies 13.34% of all types (two types) and only represents fictive motion events, where the collocate alludes to the PATH performing the action (i.e., *pitch* and *trail*). However, three more noun + verb collocations representing fictive movement were identified by using the Concordance function of Sketch Engine: *(a) path climb, (a) track climb* and *(a) route climb.* They can be regarded as specialized collocations, which was discovered by matching their *logDice* scores in the ADVENCOR corpus against the reference corpus. Moreover, as the *logDice* score of *(a) route climb* was 7.03 points and the collocation was identified in three different contexts, we believe that it deserves its inclusion in the dictionary.

In terms of the word span between the elements of the collocations, the following conclusions are drawn. In verb + noun collocations, the span ranges from +1 to +4, where the intervening words express a feature of the path that is climbed, such as height (e.g., *highest*) or direction (e.g., *northwest*). On the contrary, there are no words occurring between the elements of the adverb + verb/verb + adverb and noun + verb collocations, where the spans are -1 or +1.

5.4.3. *cross*-v

The results obtained for *cross*-v in the Word Sketch function of Sketch Engine were objects and subjects of the verb. The collocations selected together with their *logDice* score and frequency in both corpora, specialized and reference, are shown in descending order in Table 27:

Table 27

logDice *Score and Frequency of the Collocations Containing* cross-*v*

Collocation	***logDice* in ADVENCOR**	**Frequency in ADVENCOR**	***logDice* in 'enTenTen13'**	**Frequency in 'enTenTen13'**
cross (a) bridge	11.13	20	9.7	23,833
cross (a) river	11	22	9.89	27,022
cross (a) stream	10.57	10	7.31	4,431
(a) bridge cross	10.23	4	8.09	3,627
(a) road cross	10.21	4	7.22	3,195

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
cross (a) meadow	8.42	2	--	--
(a) trail cross	8.39	4	7.53	2,716
(a) trek cross	8.34	2	--	--
(a) route cross	8.24	2	6.14	939

The collocations selected for *cross*-v were nine, whose syntactic structure is noun + verb (five types) and verb + noun (four types). As it can be observed, all the collocations achieved a *logDice* score higher in the specialized corpus than in the reference corpus. What is more, it must be highlighted that two of them are not identified in the reference corpus, namely, *cross (a) meadow* and *(a) trek cross*, which means that they are highly specialized collocations in the SL of adventure tourism. The average difference in points between the *logDice* scores obtained in both corpora is 3.41. Once again, the numbers reveal that *logDice* scores and frequencies are independent and no connection can be established between them, given that high *logDice* scores do not imply high frequencies, or vice versa. For instance, *cross (a) stream* presents 7.31 points and 4,431 tokens in the 'enTenTen13,' while *(a) bridge cross* presents 8.09 points and 3,627 tokens.

The strongest collocations retrieved from the ADVENCOR corpus are *cross (a) bridge, cross (a) river* and *cross (a) stream*, with a *logDice* score of 11.13, 11 and 10.57 points, respectively. Curiously enough, they are the collocations which represent both types of motion, actual motion performed by an animate entity and fictive motion carried by an inanimate entity. As a result, these collocations are encoded in the two entries for *cross*-v in *DicoAdventure*. In both entries, the verb focuses attention on a path that goes from one side to the other, therefore, it is normally found in transitive constructions, for instance, *On some of our longer tours we* ***cross*** *the Arctic Circle*, although it can also appear in intransitive constructions, for example, *We* ***crossed*** *into Canada on September 20th, checking another thru hike off our list.*

cross$_1$

cross$_1$ shows collocations whose examples in context portray an animate subject performing the motion. The verb is defined as follows: "A TOURIST goes across something (PATH) from one side (SOURCE) to the other (DESTINATION)" and the

collocations illustrating this meaning are *cross (a) bridge, cross (a) river, cross (a) stream* and *cross (a) meadow:*

Collocation	Type	Explanation	Examples
cross (a) bridge	verb + noun	~ a specific PATH	(45) We **cross** several suspension **bridges**, pass by several monasteries and villages before reaching Lukla. (46) Soon you are **crossing** the suspension **bridge** to Phillim.
cross (a) river	verb + noun	~ a specific PATH	(47) Everyone will be **crossing** the San Gabriel **River** and getting their feet wet. (48) After **crossing** the **river**, we descended through a high and steep trail.
cross (a) stream	verb + noun	~ a specific PATH	(49) Be prepared to **cross** at least 5-6 small water **streams**, so wear shoes accordingly and, if possible, get an extra pair if you don't prefer jumping barefoot. (50) Valley here bends to the east, you will have to **cross** several smaller **streams** on your way.
cross (a) meadow	verb + noun	~ a specific PATH	(51) From the cable car, head to the east. Walk between trees, **cross** small **meadows** and you will come to bigger dirt road. (52) You will **cross** several other **meadows** and belts of trees and after amazing two kilometers you will come to a spot where trail turns north and dives into the forest.

As it can be seen, the four collocations depicting actual motion of *cross*-v have the syntactic structure verb + noun, where the noun, object of the verb, reveals the PATH that is crossed by the subject, namely, *bridge, river, stream* and *meadow*. A detailed analysis of the contexts reveals that 75% of them show the verb and the collocate separated by other words providing specific information about the paths crossed. For instance, examples (45) and (46) specify the type of bridge: *suspension*, in a span of +3, and example (47) gives the name of the river, *San Gabriel*, using a span of +4. Finally, the small size of the streams and the meadows is provided in (49), (50) and (51), which represent the collocates in spans of +6, +3 and +2 words from the base, respectively. On

the other hand, the broader span in context (49) finds room for some more details about the streams, such as the material: *water*, and the quantity: *5-6*.

cross₂

cross₂ includes collocations depicting an inanimate entity performing fictive motion and is defined in the following terms: "A PATH1 traverses a PATH2 to DESTINATION."[215] The collocations illustrating this meaning are *cross (a) bridge, cross (a) river, cross (a) stream, (a) bridge cross, (a) road cross, (a) trail cross, (a) trek cross* and *(a) route cross:*

Collocation	Type	Explanation	Examples
cross (a) bridge	verb + noun	~ a specific PATH	(53) Then the trail continues down a hot steep path, **crosses** the suspension **bridge** on Phawa Khola (1430m) and up steeply to the scattered bamboo and bananas settlement of Kunjari (1800m). (54) This long itinerary **crosses** the well-known 27m suspension **bridge** and then follows a magnificent and airy line along ledges used during the First World War.
cross (a) river	verb + noun	~ a specific PATH	(55) A South Island trek leading up the Hooker Valley to Mount Cook, **crossing** the **river** and passing the glacial lake, is a very accessible walk with some of the most spectacular views in the South Island. (56) The trail also **crosses rivers** and leads to a glistening 10m high waterfall.
cross (a) stream	verb + noun	~ a specific PATH	(57) The right path is shorter and recommended – it dives into the forest, **crosses** 4 smaller **streams** and, after few kilometers, comes to the big flat meadow with a shepherd's hut with roof made of blue plastic bags (abandoned at the time of our visit), ideal place for camping.

[215] As the argument structure of *cross₂* contains two distinct PATHS, numbers are used to distinguish them, being PATH1 the inanimate entity that seems to perform the motion, found in the subject position, and PATH2 the path along/through which the movement is developed, found in the object position.

Collocation	Type	Explanation	Examples
			(58) This is a country designed around the use of the bicycle, the narrow lane ways and small <u>bridges</u> **crossing streams** and rivers abound in the country.
(a) bridge cross	noun + verb	A PATH ~s	(59) This is a country designed around the use of the bicycle, the narrow lane ways and small **bridges crossing** streams and rivers abound in the country.[216]
			(60) Trail trailhead is located next to Junction Creek and is fairly easy for the first few miles until you get the first **bridge crossing** and then becomes a bit steeper.
(a) road cross	noun + verb	A PATH ~s	(61) In one tenth of a mile keep straight when a woods **road crosses** over and across a wooden footbridge in another tenth of a mile.
			(62) Then, few miles before Kveda Vedi comes one last obstacle. The **road** here **crosses** Urashi river twice, but there is no bridge.
(a) trail cross	noun + verb	A PATH ~s	(63) Then the **trail** continues down a hot steep path, **crosses** the suspension bridge on Phawa Khola (1430m) and up steeply to the scattered bamboo and bananas settlement of Kunjari (1800m).[217]
			(64) The **trail** also **crosses** rivers and leads to a glistening 10m high waterfall.[218]
(a) trek cross	noun + verb	A PATH ~s	(65) Today a long **trek crossing** the famous Thorang La (5416m/17764ft) high pass is going to be our lifetime memorable experience.
			(66) A South Island **trek** leading up the Hooker Valley to Mount Cook, **crossing** the river and passing the glacial lake, is a very accessible walk with some of the

[216] Examples (58) and (59) show the same context, but the focus is on different collocations.
[217] Examples (53) and (63) show the same context, but the focus is on different collocations.
[218] Examples (56) and (64) show the same context, but the focus is on different collocations.

Collocation	Type	Explanation	Examples
			most spectacular views in the South Island.[219]
(a) route cross	noun + verb	A PATH ~s	(67) The **route crosses** the South-Eastern wall of Col Rosa, rising like a lonely giant in the valley between Tofana and Pomagagnon. (68) The **route** now **crosses** through the focella to the other side of the ridge, and traverses along toward the Forcella Lavaredo.

In total, eight collocations of *cross*-v represent fictive motion, which belong to two lexical types: noun + verb (five collocations) and verb + noun (three collocations). The collocations included in the former type incorporate the inanimate subject of the action into the combination, which is a PATH: *bridge, road, trail, trek* and *route*. In half of the examples, the noun co-occurs right next to the verb in a -1 span. In (62), (64) and (68), the span is -2 and the intervening words refer to place: *here*, show addition: *also*, or allude to time: *now*. Finally, in examples (63) and (66), the spans are wider: -7 and -9, respectively, since additional information on the subjects in distinct clauses is supplied.

The second group of collocations, verb + noun, is slightly less recurrent, although their elements are more strongly associated, given their higher *logDice* scores. The examples accompanying these collocations depict an inanimate entity crossing a PATH, which can be a *bridge*, a *river* or a *stream*. These objects are located within a span of +1 (two examples), +2 (one example), +3 (two examples) and +5 (one example) from the verb. The words in between identify the length and/or the type of bridge, such as *suspension* and *27m suspension* in examples (53) and (54), and the quantity and size of the streams: *4 smaller* in (57).

The inanimate entities functioning as subjects in the contexts of the verb + noun collocations (located within spans ranging from -1 to -10) are *trail* ((53) and (56)), *itinerary* ((54)), *trek* ((55)), *path* ((57)) and *bridge* ((58)). Most of them (i.e., *trail, trek* and *bridge*) have been previously examined because they were identified as nominal collocates of *cross*-v in the subject position. The remaining subjects, *itinerary* and *path*, were not extracted in the word sketch of the verb. It can be explained because the criteria

[219] Examples (55) and (66) show the same context, but the focus is on different collocations.

established were a minimum *logDice* score of 7 points and a minimum frequency of two tokens in the ADVENCOR corpus, and neither of the nouns satisfied the latter, since the only recurrences in the specialized corpus are in contexts (54) and (57). However, they fulfilled the first criterion and the results were 8.69 points for *(an) itinerary cross* and 8.35 points for *(a) path cross*. These high *logDice* scores show that they are indeed strong associations of words and very significant collocations. Moreover, it must be pointed out that *(an) itinerary cross* was not retrieved from the reference corpus after doing a similar search and *(a) path cross* was given 8.35 points (the same score as in the ADVENCOR corpus). As a result, we can state that these two collocations are also specialized in the domain of adventure tourism, despite the fact that the latter is identified as a strong collocation in general English too.

In summary, nine collocations of *cross*-v have been covered in this study and 88.89% of them represents fictive motion events: *cross (a) bridge, cross (a) river, cross (a) stream, (a) bridge cross, (a) road cross, (a) trail cross, (a) trek cross* and *(a) route cross*. On the other hand, 44.44% of the collocations selected describe actual motion events: *cross (a) bridge, cross (a) river, cross (a) stream* and *cross (a) meadow*, which means that three collocations are included in $cross_1$ and $cross_2$, that is, *cross (a) bridge, cross (a) river* and *cross (a) stream.*

The first entry is characterized for including collocations of only one lexical type, verb + noun (this structure belongs to 44.44% of the collocations), where the noun specifies the PATH that is crossed. However, the second entry contains collocations of two types, verb + noun and noun + verb (this structure belongs to 55.56% of the collocations), where the noun in the first type has the same function as in $cross_1$ and the noun in the second type designates the inanimate subject of the action, that is, a PATH. Furthermore, it must be highlighted that two other noun + verb collocations were identified through the Word Sketch function of the software when the frequency criterion was not laid down, namely, *(an) itinerary cross* and *(a) path cross*. They both represent fictive motion and their degree of specialization has been proved by checking their *logDice* score in the ADVENCOR corpus (8.69 and 8.35 points, respectively) and by matching it against their *logDice* in the 'enTenTen13' (0 and 8.35 points, respectively).

Turning to the span between the elements of the collocations selected, when the motion is real, the path that is crossed occurs within a range of +2 (25%), +3 (50%), +4

(12.5%) and +6 (12.5%) from the verb; when the motion is fictive, it occurs within a range of +1 (33.33%), +2 (16.67%), +3 (33.33%) and +5 (16.67%) from the verb. On the other hand, the path performing the action in fictive motion events occurs within a range of -1 (50%), -2 (30%), -7 (10%) and -9 (10%) words from the base.

5.4.4. *descend*-v

The results obtained for *descend*-v in the Word Sketch function of Sketch Engine were modifiers, an object and a subject of the verb. The collocations selected together with their *logDice* score and frequency in both corpora, specialized and reference, are shown in descending order in Table 28:

Table 28

logDice *Score and Frequency of the Collocations Containing* descend-*v*

Collocation	***logDice* in ADVENCOR**	**Frequency in ADVENCOR**	***logDice* in 'enTenTen13'**	**Frequency in 'enTenTen13'**
gradually descend	11.58	7	5.83	859
descend (a) gully	10.97	3	5.24	85
descend steeply	10.82	3	8.9	831
gently descend / descend gently	10.41	3	5.27	447
slowly descend / descend slowly	10.24	3	6.01	1,885
(a) trail descend	9.48	8	7.1	1,197
descend down	8.18	2	4.2	1,112
descend back	7.37	3	--	--

The collocations selected for *descend*-v were eight. The most frequent syntactic structure is adverb + verb/verb + adverb (six types); aside from that, there is one verb + noun and one noun + verb collocations. As for the *logDice* scores, without a doubt they are significantly higher in the specialized corpus and the difference is of 4.56 points on an average. In fact, there are several collocations whose *logDice* in the ADVENCOR corpus doubles (or almost doubles) their *logDice* in the 'enTenTen13,' for instance, *gradually descend* (11.58 vs. 5.83), *descend (a) gully* (10.97 vs. 5.24), *gently descend/descend gently* (10.41 vs. 5.27), *descend down* (8.18 vs. 4.2). Moreover, the collocation *descend*

back was not identified in the reference corpus. These results indicate that the collocations selected are highly specialized in the domain of adventure tourism.

The collocation showing the highest *logDice* score in the ADVENCOR corpus is *gradually descend* (11.58 points), which is, incidentally, the only collocation found in contexts representing both types of motion, so it is included in *descend*$_1$ and *descend*$_2$ in *DicoAdventure*. In total, fictive motion is represented by a higher number of collocations (i.e., five), since *descend (a) gully, descend steeply, slowly descend/descend slowly* and *(a) trail descend* must be added to *gradually descend*. On the other hand, four collocations express actual motion: the abovementioned *gradually descend* and *gently descend/descend gently*, *descend down* and *descend back*.

With respect to the meaning of the verb, *descend*-v describes vertical motion, more specifically, downward direction of motion, which involves a subject moving to a lower position. It can be used transitively and be accompanied by the path that is descended, for example, *Then we will **descend** the river, enjoying the untouched nature, several slides, jumps (optional up to 8mt), caves,* or intransitively and can show the source or the destination of the movement, for instance, *Experience the Samaria Gorge and **descend** from the Omalos Plateau; Once we started **descending** to Winteregg, it started to drizzle a bit.*

descend$_1$

descend$_1$ encompasses collocations whose contexts describe an animate entity performing a real movement. It is defined as follows: "A TOURIST moves in a downward DIRECTION along a PATH from SOURCE to DESTINATION." The collocations illustrating this meaning are *gradually descend, gently descend/descend gently, descend down* and *descend back*:

Collocation	Type	Explanation	Examples
gradually descend	adverb + verb	~ in a specific MANNER	(69) We then descended back through the clouds and drifted on slowly, covering at total of 8 miles, **gradually descending** with still more breathtaking views below us. (70) We **gradually descend** to Cheplung village from where we get a glimpse of Mt Khumbila, a sacred mountain which has never been.

Collocation	Type	Explanation	Examples
gently descend / descend gently	adverb + verb / verb + adverb	~ in a specific MANNER	(71) You will **gently descend** the 21 switchbacks, where you will enjoy bi-coastal views of Maui's central valley. (72) The instructor will pull the ripcord and you will both **descend gently** to the ground sharing a canopy engineered for two people.
descend down	verb + adverb	~ in a downward DIRECTION	(73) After breakfast **descend down** for 9 kms to the gate where you get your transport back to Nairobi, making a lunch stop in Nanyuki. (74) Your canopy opens and you **descend** safely **down** to earth.
descend back	verb + adverb	~ in a specific DIRECTION	(75) We then **descended back** through the clouds and drifted on slowly, covering at total of 8 miles, gradually descending with still more breathtaking views below us. (76) To **descend back** to the base, you need to rappel down the rock face using a rope to help you slide down safely.

As it is shown in Table 28, four collocations represent actual motion of *descend*-v. They belong to only one lexical type, that is, adverb + verb/verb + adverb, where the collocate provides information of two kinds about the verb: (1) *gradually* and *gently* reveal the MANNER in which the motion is developed, that is, at a slow speed and in small stages (*gradually*) and in a careful and non-violent way (*gently*), and (2) information about the DIRECTION of the movement is provided by *down* (i.e., downward) and *back* (i.e., toward the source of motion or a previous point). Except for example (71), all the uses of the verb are intransitive.

As for the span between the elements of these combinations, the vast majority of contexts show the elements of the collocations right next to each other, with the exception of (74): *descend safely down*, where the intervening word indicates the manner of the motion, that is, in a way that is safe. The span +1 is slightly more common than -1, since it can be observed in four contexts, namely, (72), (73), (75) and (76), whereas the span -1 is used in three examples, that is, (69), (70) and (71). It means that, while the adverb *gradually* precedes the verb in both contexts within a span of -1 (no co-occurrences of *descend gradually* were retrieved from the ADVENCOR corpus), *gently* occurs both before

(span: -1) and after (span: +1) the verb. On the other hand, *down* and *back* always follow the verb.

descend₂

descend₂ displays fictive motion events, that is, an inanimate entity is not actually moving, and is defined as follows: "A trail, track, etc. (PATH) has a downward course (DIRECTION) that departs from a SOURCE through a PLACE to DESTINATION." The collocations illustrating this meaning are *gradually descend, descend (a) gully, descend steeply, slowly descend/descend slowly* and *(a) trail descend:*

Collocation	Type	Explanation	Examples
gradually descend	adverb + verb	~ in a specific MANNER	(77) The trail again **gradually descends** until Phakding. (78) After walking along the rim, our hike **gradually descends** 1100 feet, averaging about 320 feet of loss per mile.
descend (a) gully	verb + noun	~ a specific PLACE	(79) After **descending** the **gully**, the route follows a traversing path to more vertical walls, where a rising path with chiselled stairs cuts across the rock. (80) From this pass, you can optionally go up to the summit of Monte Paterno, but this isn't required to complete the route, which **descends** a **gully** below and somewhat to the right.
descend steeply	verb + adverb	~ in a specific MANNER	(81) After **descending steeply**, the trail goes through rice terraces into a side canyon, when we cross the Khaksewa Khola (River) on a long suspension bridge (1540 meter). (82) The water stream has hollowed out a deep ravine which is **descending steeply** toward the Storo plain.
slowly descend / descend slowly	adverb + verb / verb + adverb	~ at a specific SPEED	(83) The path crosses open areas with nice views and wildflowers and then **slowly descends** into the forest. (84) Follow the trail. First, it's **descending slowly**, then a steep

Collocation	Type	Explanation	Examples
			serpentine finally takes you down to the river.
(a) trail descend	noun + verb	A PATH ~s	(85) Follow the **trail**. First, it's **descending** slowly, then a steep serpentine finally takes you down to the river.[220]
			(86) The **trail** again gradually **descends** until Phakding.[221]

The collocations containing *descend*-v and representing fictive motion are five. As in *descend*$_1$, the most frequent lexical type is adverb + verb/verb + adverb (three types), followed by verb + noun (one type) and noun + verb (one type). When the verb is modified by an adverb, the latter expresses the MANNER in which the motion is made, that is, *gradually* (at a slow speed and in small stages) and *steeply* (in a difficult way), and the SPEED at which the motion takes place: *slowly*. Then, the verb + noun collocation, *descend (a) gully*, includes the PLACE where the descent takes place. The last collocation selected is *(a) trail descend*, which unveils the inanimate entity (PATH) that seems to be performing the motion. With regards to the span between the verb and its collocates, it is shorter than in previous entries, since it is -1/+1 in 80% of the examples and +2 in 20% of them (more specifically, in the verb + noun collocation, in whose contexts the noun is introduced by the determiners *the* (79) and *a* (80)).

Regarding the subject participating in the fictive motion events, it is included in the collocation *(a) trail descend* (within a span of -3 in both examples). In the rest of the collocations, it needs to be identified in the contexts. For instance, *trail* is also the subject in (77), (81) and (84) (in fact, the context (84) is the same as (85), being the latter used to exemplify the collocation *(a) trail descend*). Other subjects are *route* (two tokens), *hike* (one token), *ravine* (one token) and *path* (one token), which were searched for in Sketch Engine to know whether they were also identified as nouns collocating with *descend*-v in the subject position. As a result, we discovered that *(a) hike descend* is regarded both as a strong and specialized collocation, given that its *logDice* score is of 9.45 points and it was not extracted from the 'enTenTen13' when doing the same search. Nevertheless, it was not included in our selected collocations because, despite fulfilling the criterion related to the *logDice* score (more than 7 points), it has only one token in the ADVENCOR

[220] Examples (84) and (85) show the same context, but the focus is on different collocations.
[221] Examples (77) and (86) show the same context, but the focus is on different collocations.

corpus, therefore, it does not satisfy the criterion related to the frequency of the word combination (at least two tokens). On the other hand, as the rest of the subjects, that is, *route*, *ravine* and *path*, were not identified through the Word Sketch function, they were manually checked in the Concordance function, which produced the following results:

1) *(a) route descend*: four tokens of *route* + *descend* within a span of -5 were identified in the ADVENCOR corpus, with a *logDice* score of 7.03 points. However, only one of them was considered a true collocation, which coincides with (80) above: *From this pass, you can optionally go up to the summit of Monte Paterno, but this isn't required to complete the* ***route****, which* ***descends*** *a gully below and somewhat to the right*. In order to know the real *logDice* score of this collocation, the parameters were changed and the span was reduced to -2 (to focus only on this concrete example). After that, the result was a *logDice* of 5.03 points, two points lower than in the previous search which included three false collocations. On the other hand, the same type of search was carried out but, in this case, setting a span of +4 so that example (79) could be retrieved, and the results were three tokens where only one represented a true collocation, that is, example (79). The *logDice* obtained was 7.31 points. Similar searches were made in the reference corpus and the results were 236 tokens with a *logDice* of 2.70 points when the span was set in -2 words and 292 tokens with a *logDice* of 3.01 points when the span was set in +4 words. The differences in the points obtained in both corpora reveal that *(a) route descend* can still be regarded as a specialized collocation. Furthermore, we think that it must be encoded in $descend_2$, for it fulfills both criteria established in this dissertation for the collocation extraction (at least in one of the searches), namely, *logDice* score and frequency;

2) *(a) ravine descend*: the only context identified for the combination *ravine* + *descend* is the one in (82): *The water stream has hollowed out a deep* ***ravine*** *which is* ***descending*** *steeply toward the Storo plain*. This collocation has a *logDice* score of 7.39 points, which means that it fulfils one of the criteria established in this work when extracting collocations. In addition to this, it is important to realize that, although the word sketch of *descend*-v in the 'enTenTen13' did not extract *ravine* as a collocate, the Concordance function did. Thus, the results showed six tokens with a *logDice* of 1.76 points, but

only two were true collocations. Consequently, we can firmly declare that *(a) ravine descend* is a specialized collocation too; and

3) *(a) path descend*: two co-occurrences of *path* + *descend* were detected in the specialized corpus within two different spans: -5 and -3, and with a *logDice* of 7.21 points. Both contexts were manually checked and only one true collocation was identified, which differs from (83): *The **path** to Tserang **descends** below the red and grey cliffs across the valley to small river*. The reason for not having detected (83) is probably because, as it has been aforesaid, a span of -5/+5 is the maximum allowed in the Concordance function, but the span in (83) is -12. In order to know the degree of specialization of *(a) path descend*, the collocation was searched in the same way in the 'enTenTen13' and the *logDice* obtained was of 3.72 points, almost half the *logDice* in the ADVENCOR corpus. Accordingly, we can claim once again that it is a specialized collocation in the SL of adventure tourism. Moreover, as it satisfies the two criteria applied in the collocation extraction in this work (i.e., *logDice* and frequency), we think that it must be included in the entry for *descend*$_2$ in *DicoAdventure*.

As for the span between the verb and the subjects undertaking the action, the only exception is (83) (span: -12), but in the rest of the examples the subject is located close to the verb, more specifically, within a span of -2 (examples (78) and (80)), -3 (examples (77), (82) and (84)), +3 (example (81)) and +4 (example (79)).

To sum up, eight collocations of *descend*-v have been analyzed in this work. Actual motion events are represented by four collocations: *gradually descend, gently descend/descend gently, descend down* and *descend back*, while fictive motion events are represented by five collocations: *gradually descend, descend (a) gully, descend steeply, slowly descend/descend slowly* and *(a) trail descend*. It means that *gradually descend* is included in *descend*$_1$ and *descend*$_2$.

The most frequent lexical collocation containing *descend*-v is the combination of the verb plus an adverb (six types, 75%). In fact, real motion is only described with this type of combination, in which the collocate reveals the MANNER (*gradually* and *gently*) and the DIRECTION (*down* and *back*) of the movement. By the same token, the adverb + verb/verb + adverb collocation is the most common type in *descend*$_2$, but, in this group,

there are two collocates describing the MANNER (*gradually* and *steeply*) and one collocate specifying the SPEED (*slowly*) of the movement. In addition, this entry also includes a verb + noun and a noun + verb collocation, where the collocate is the PLACE that is descended (*gully*), on the one hand, and the PATH that seems to be descending (*trail*), on the other. An element in common is that the span between the parts of the collocations in context is rather short in both entries. To clarify, it is -1/+1 in all the adverb + verb/verb + adverb collocations (with the only exception of example (74), *descend safely down*), +2 in the verb + noun collocation and -3 in the noun + verb collocation.

Last but not least, we must say that four more noun + verb collocations were discovered when analyzing the inanimate subjects of the contexts in *descend*$_2$ (either through the Word Sketch or the Concordance function), namely: *(a) hike descend, (a) route descend, (a) ravine descend* and *(a) path descend*. Their *logDice* scores in the ADVENCOR corpus were compared to their *logDice* scores in the reference corpus and the following results were achieved: 9.45 versus 0 points, 5.03 versus 2.70 points, 7.39 versus 1.76 points, 7.21 versus 3.72 points, respectively. Therefore, it was concluded that these collocations are both strong associations of words and specialized collocations in the SL of adventure tourism. What is more, *(a) route descend* and *(a) path descend* deserve to be included in *descend*$_2$ because they fulfil the *logDice* (7.31 and 7.21 points, respectively) and frequency (two tokens each) criteria established for collocation extraction in this dissertation.

5.4.5. *head*-v

The results obtained for *head*-v in the Word Sketch function of Sketch Engine were modifiers and subjects of the verb. The collocations selected together with their *logDice* score and frequency in both corpora, specialized and reference, are shown in descending order in Table 29:

Table 29

logDice *Score and Frequency of the Collocations Containing* head-*v*

Collocation	***logDice* in ADVENCOR**	**Frequency in ADVENCOR**	***logDice* in 'enTenTen13'**	**Frequency in 'enTenTen13'**
head north	11.35	8	9.9	19,362
head back	10.03	20	8.53	97,084
head up	9.86	28	5.46	4,696
(a) path head	9.75	2	5.07	605
head directly	9.5	3	5.78	3,301
head down	9.33	5	7.75	15,567
(a) trail head	8.8	5	6.35	1,524

The collocations selected for *head*-v were seven and belong to two different lexical types: verb + adverb (five types) and noun + verb (two types). While the former is found in actual (*head north, head back, head up* and *head down*) and fictive (*head north* and *head directly*) motion events, the latter expresses only fictive motion events (*(a) path head* and *(a) trail head*). As it can be seen, one collocation participates in both types of motion: *head north*, so its contexts are examined in the two entries for *head*-v available in *DicoAdventure*.

In fact, it is *head north* the strongest collocation from the list, with a *logDice* score of 11.35 points. However, it is not the most frequent, since it has eight tokens and occupies the third position behind *head up* (28) and *head back* (20), showing that a higher frequency does not imply a greater strength of association. Regarding their degree of specialization, the *logDice* scores in the ADVENCOR corpus are higher than those in the reference corpus, with a difference of 2.82 points on an average. This fact leads us to readily admit that the collocations containing *head*-v are more typical of the specialized corpus, therefore, they must be seen as specialized collocations.

Turning to the meaning of this motion verb, it describes direction oriented toward a destination and is always used intransitively, for instance, *After two hours of fun and adrenaline, we will* ***head*** *to the Balsa River and enjoy two hours of whitewater action on a beautiful river; Once you're all kitted out, we will* ***head*** *into the trees to begin our tour.*

head₁

head₁ presents collocations where motion is performed by an animate subject and is defined in the following terms: "A TOURIST follows a PATH from SOURCE to DESTINATION in a particular DIRECTION." The collocations illustrating this meaning are *head north, head back, head up* and *head down:*

Collocation	Type	Explanation	Examples
head north	verb + adverb	~ in a specific DIRECTION	(87) Many of them can be found in the 6-million-acre Adirondack Park – the largest protected wilderness area east of the Mississippi – but nature is pretty much everywhere as soon as you **head north** and west of the city. (88) Stroll sandy beaches and hike along the rugged coastline, take photos from rock platforms, and **head north** to see forests, green gullies, pristine lakes and gushing waterfalls.
head back	verb + adverb	~ in a specific DIRECTION	(89) The trail back is pretty easy to follow, but if you're directionally challenged like me, make sure you find some other hikers in your group to **head back** with. (90) Time is taken to enjoy the winter wonderland as you make your away through the scenic forest trails before **heading back**.
head up	verb + adverb	~ in an upward DIRECTION	(91) The idea is that you move (and unlock) one carabiner at a time as you **head up** to the rocky cliffs of the mountainside, placing it above the next fixed bolt (so that you always have one line on the mountain, securing you, at all times). (92) From Cairns you will **head up** into the rainforest to the Barron Falls Power Station Bridge.
head down	verb + adverb	~ in a downward DIRECTION	(93) We then spent a little while preparing the campsite before **heading down** to have a look at the Pwll-Cym Sump from the inside.

Collocation	Type	Explanation	Examples
			(94) After forging deep into the cave, still following the Rao Thuong River and clambering up to the lookout point with its views over the beach campsite, we **headed down** to the tents where we dropped our gear.

In total, four collocations containing *head*-v represent a motion event where there is an animate entity having a displacement in space. They all have the syntactic structure verb + adverb and the span used is +1 in all the examples. The collocate always refers to the DIRECTION followed by the movement, which can be, for instance, toward a cardinal point: *north*, toward the source of motion or a previous point: *back*, upward: *up*, or downward: *down*. Furthermore, a closer look into the concordance lines extracted from the ADVENCOR corpus revealed that the combination of *head*-v and a cardinal point is more common than one could expect at first:

1) *head southeast*: the word sketch of the verb in the ADVENCOR corpus regarded this collocation as a strong association in the light of its high *logDice* score: 11.09 points, which doubles the score achieved in the reference corpus: 5.1 points. The reason for not having been extracted in our search was that it did not satisfy the frequency criterion, as it only has one token: *Turn right again and **head southeast** through dense pines and knee-high grass along part of the Continental Divide Trail*;

2) *head east*: this collocation only got one token in the word sketch of the verb and it achieved a *logDice* score of 8.76 points, but we had a feeling that more contexts could be found. Thus, the Concordance function was used and it produced eight results with a *logDice* score of 8.64 points (this discovery suggests that the Word Sketch function of Sketch Engine is not absolutely perfect and there can be a margin of error). From these results, only two could be regarded as real examples of the collocation in the domain of adventure tourism, namely: (1) *At this stage the trek is comfortable and quite spectacular, we **head east** through a lake to arrive at a privileged viewpoint: the huge Qooroq glacier, one of the busiest in all of Greenland with over 200,000 tons of ice a day flowing in to the sea*, and (2) *We depart from our Nong Khiaw office at 8:30 am or 1:00 pm and **head** 10 Km **east** to the starting*

point, where you will receive instruction about how to zipline and safety advice. The same type of search in the reference corpus resulted in 1,551 tokens with a *logDice* of 5.11 points, accordingly, we can state that *head east* is a specialized collocation; and

3) *head south*: it was retrieved in the word sketch of *head*-v with a *logDice* score of 9.78 points and three tokens, but only one of them dropped an explicit reference to adventure tourism: *From the Fonda Savio hut we **head south** into a stone filled basin and climb to the base of the Cima del Cadin de NE*. These figures were matched against the reference corpus and the results achieved were 10.22 points of *logDice* score and 22,299 tokens. Therefore, the higher *logDice* score in the 'enTenTen13' shows that the combination *head + south* is also frequent and a strong association in the GL.

In view of these results, it can be stated that the combination of *head*-v with a cardinal point is common, specialized and strong (also proven in the reference corpus). For this reason, all the collocations containing *head* + cardinal point are collected in the same row in *head$_1$*, which implies that any cardinal point may follow the verb when it represents actual motion events.

head$_2$

head$_2$ includes collocations whose contexts describe an inanimate entity performing motion but it does not actually move, thus, it is defined as "A PATH has a particular course that goes from SOURCE to DESTINATION following a particular DIRECTION." The collocations illustrating this meaning are *head north, (a) path head, head directly* and *(a) trail head:*

Collocation	Type	Explanation	Examples
head north	verb + adverb	~ in a specific DIRECTION	(95) There is a steep trail **heading north** up the gully that you can take to approach the routes on the SW side of the tower. (96) Just follow the road **heading north** and after two kilometers it will join main road from Mestia to Ushguli at Ughviri pass.

Collocation	Type	Explanation	Examples
(a) path head	noun + verb	A PATH ~s	(97) Once there, find a small **path heading** northeast. (98) Leave the village by a **path heading** by the river to the east, to the foot of Adishi glacier.
head directly	verb + adverb	~ in a specific MANNER	(99) From this creek bed the trail turns sharply to the west and **heads directly** toward the shoreline of Lewis Lake. (100) The trail from here continues to the newly built "5 seasons" camp/restaurant. From here, it mostly traverses the eastern side of the valley, crosses the river (no bridge, you will have to ford it) and **heads directly** toward striking green/yellow tent.
(a) trail head	noun + verb	A PATH ~s	(101) There is a steep **trail heading** north up the gully that you can take to approach the routes on the SW side of the tower.[222] (102) From this creek bed the **trail** turns sharply to the west and **heads** directly toward the shoreline of Lewis Lake.[223]

The number of collocations containing *head*-v which depict fictive motion is four. Two types of lexical collocations are found: verb + adverb (50%) and noun + verb (50%). In the former group, the collocate alludes to the DIRECTION followed by the movement: *north*, and the MANNER in which it is made: *directly*. In the latter group, the collocate describes the PATH that seems to make the movement: *path* and *trail*. In all the instances, the span between the collocate and the verb is +1 or -1, except for example (102), in which the subject occurs in a span of -7 from the verb, since they are separated by another clause.

As we have seen in *head*$_1$, the combinations of *head*-v and a cardinal point are common in actual motion events. By the same token, it also happens in fictive motion events. For instance, although *head north* was the only collocation of this type that satisfied the criteria established in the Word Sketch function for collocation extraction, the Concordance function produced the following results:

[222] Examples (95) and (101) show the same context, but the focus is on different collocations.
[223] Examples (99) and (102) show the same context, but the focus is on different collocations.

1) *head east*: one example representing fictive motion was retrieved: *Soon after you reach the top of Nevada Falls, the Merced Lake Trail and John Muir Trail* ***head east*** *up the canyon*. The *logDice* score of the collocation in the ADVENCOR corpus was 8.64 points, whereas it was 4.48 points in the reference corpus; as a result, it can be stated that it is a specialized collocation;

2) *head west*: one co-occurrence exemplifying fictive motion was found: *The Rocky Mountain Line* ***heads west*** *from Centennial City across the central Rockies*. The *logDice* score of the collocation in the ADVENCOR corpus was 7.21 points, whereas it was 4.60 points in the reference corpus, thus, *head west* can be regarded as a specialized collocation; and

3) *head northeast*: the only example of this collocation in the ADVENCOR corpus, which was identified in the word sketch of the verb with a *logDice* of 8.89 points, represents fictive motion: *Once there, find a small path* ***heading northeast*** (it coincides with example (97) above). The degree of specialization of the collocation is quite high, given that its *logDice* in the 'enTenTen13' was of 5.72 points.

As a result, given the high number of co-occurrences of *head*-v with a cardinal point and their high degree of specialization, all the collocations containing *head* + cardinal point are gathered in the same row in $head_2$, which means that this verb can be followed by any cardinal point in fictive motion events.

As for the subjects of the verb + adverb collocations, it is *trail* in 75% of the examples, which is identified as a nominal collocate in *(a) trail head*, included in $head_2$. The remaining subject, *road*, was also provided in the word sketch of the verb and the collocation *(a) road head* got 9.56 points of *logDice* score. However, only one example illustrating the domain of adventure tourism was found (example (96)). The same search was conducted in the 'enTenTen13' and this collocation obtained 6.42 points (and had 2,165 tokens), consequently, *(a) road head* can be considered a specialized collocation. On the other hand, *(mountain) line* is the subject in the example representing *head west*, but it was not extracted as a collocate of *head*-v either through the Word Sketch or the Concordance function from the ADVENCOR corpus.

With respect to the span between the subjects and the verb in verb + adverb collocations, it is -1 in the examples of *head* + cardinal point, while it is much wider in

the examples of *head directly*. More specifically, it is -7 in (99) and -21 in (100), although the context of the latter must be analyzed more in-depth. To explain, the personal pronoun referring to the real subject, *it*, is found within a span of -21, but the real subject, *trail*, is located 34 positions to the left of the base: *The trail from here continues to the newly built "5 seasons" camp/restaurant. From here, it mostly traverses the eastern side of the valley, crosses the river (no bridge, you will have to ford it) and* ***heads directly*** *toward striking green/yellow tent*. This example demonstrates that the elements of a collocation can occur in different sentences (cf. §2.2.1.1.).

Summing up, seven collocations were extracted for *head*-v expressing the following meanings: (1) actual motion: *head north, head back, head up* and *head down*, and (2) fictive motion: *head north, (a) path head, head directly* and *(a) trail head.* Thus, *head north* is included in $head_1$ and $head_2$. The most frequent lexical type is verb + adverb (71.43%), followed by noun + verb (28.57%). Actual motion events are represented only by verb + adverb collocations, whereas in fictive motion events half of them are verb + adverb and the other half are noun + verb collocations. The adverbs usually show the DIRECTION followed by the movement: *north, back, up* and *down*, although they can also allude to the MANNER in which it is developed: *directly*. The nominal collocates describe the inanimate entity which seems to be participating in the motion event: *path* and *trail*. The span between the collocates and the verb is +1 in verb + adverb collocations and normally -1 in noun + verb collocations, except for the context (102), where it is -7.

It must be remembered that a deeper study on the co-occurrences of *head*-v with cardinal points reported three more collocations in $head_1$: *head southeast, head east* and *head south*, and three more collocations in $head_2$: *head east*, *head west* and *head northeast*; they are all collected in the same row along with *head north* in each of the entries in *DicoAdventure*. Additionally, we discovered the noun + verb collocation *(a) road head* in the word sketch of the verb. The *logDice* scores obtained by these collocations matched against the reference corpus concluded that they can be regarded as specialized collocations.

Finally, we must highlight that the collocation *(a) trip head* was also extracted from the ADVENCOR corpus and fulfilled both of the criteria established, the minimum frequency and the minimum *logDice* score. Nevertheless, it was excluded from this dissertation because there was a scarce number of contexts that allowed us to create a new

entry in *DicoAdventure*, for it represents a new meaning, namely: an inanimate entity (TRAVEL) really performing motion.

After having carefully selected the collocations of the motion verbs and analyzed the type of motion represented, they were encoded in *DicoAdventure*, which will be explained in the last section of this chapter.

5.5. Implementation of collocations in *DicoAdventure*

The implementation of specialized collocations of motion verbs in *DicoAdventure*, the dictionary of adventure tourism (cf. §4.2.), was the primary objective of this dissertation. This resource is freely accessible at http://olst.ling.umontreal.ca/dicoadventure/ and is hosted by the OLST at the University of Montreal (Canada). As it was mentioned in Section 4.5., the 'Collocations' tab in each of the entries which produced collocations in this work would include the following items of information organized into four columns:

1) First, the 'Collocation' column showing the collocations themselves organized according to their *logDice* score in the ADVENCOR corpus (i.e., the strongest collocations would appear first);
2) Second, the 'Type' column identifying the lexical type of the collocation;
3) Third, the 'Explanation' column providing an explanation of the collocation in terms of the semantic role of the collocates; and
4) Fourth, the 'Examples' column illustrating the collocations in context by means of annotated examples selectively extracted from the specialized corpus; it would be accessible via the 'Ex' button included in the 'Explanation' column.

Information about the language used to build the dictionary is not within the scope of this work, thus, we will not minutely examine this field. Nevertheless, we can still supply brief details and mention that the type of language used is eXtensible Markup Language (XML), the one greatly used to encode information by lexicographers. In fact, it is said to be the emerging standard for data presentation and exchange on the World Wide Web (Bray, Paoli, & Sperberg-McQueen, 1998, as cited in Ide, Bonhomme, & Romary, 2000, p. 1). In an XML database, each part of the data structure is given an XML

tag, which can record information of any nature, such as the example represented in Figure 73 (Kiefer & Van Sterkenburg, 2003, p. 361):

Figure 73

Example of XML Tags

<LEMMA/>clerk<LEMMA\>
<WORDTYPE/>noun<WORDTYPE\>
<LEMMA MEANING NUMBER/>1<LEMMA MEANING NUMBER\>
<LEMMA MEANING/>employee charged with writing<LEMMA MEANING\>
<LEMMA MEANING NUMBER/>2<LEMMA MEANING NUMBER\>
<LEMMA MEANING/>particular low rank in the official hierarchy<LEMMA MEANING>

In the case of *DicoAdventure*, the XML files are stored in a server called eXist[224] and they can be edited by using the Oxygen XML Editor.[225] We used the Oxygen 20.0 version and access was gained thanks to the helpful guidance provided by Benoît Robichaud, researcher at the OLST. The following text presents an example of XML code used in the dictionary, which corresponds to the 'Collocations' tab of the entry for *drive*$_1$:[226]

```
<famille nom="Collocation">
 <lien-lexical>
    <explication-ra>~ a VEHICLE_WITHOUT_ENGINE</explication-ra>
    <explication-tt/>
    <type>verb + noun</type>
    <lien identificateur="drive (a) team" numero-acception="">drive (a)
team</lien>
 </lien-lexical>
 <lien-lexical>
    <explication-ra>~ a VEHICLE_WITHOUT_ENGINE</explication-ra>
    <explication-tt/>
    <type>verb + noun</type>
    <lien identificateur="drive (a) sled" numero-acception="">drive (a)
sled</lien>
 </lien-lexical>
 <lien-lexical>
    <explication-ra>A TOURIST ~s</explication-ra>
    <explication-tt/>
    <type>noun + verb</type>
    <lien identificateur="(a) participant drive" numero-acception="">(a)
```

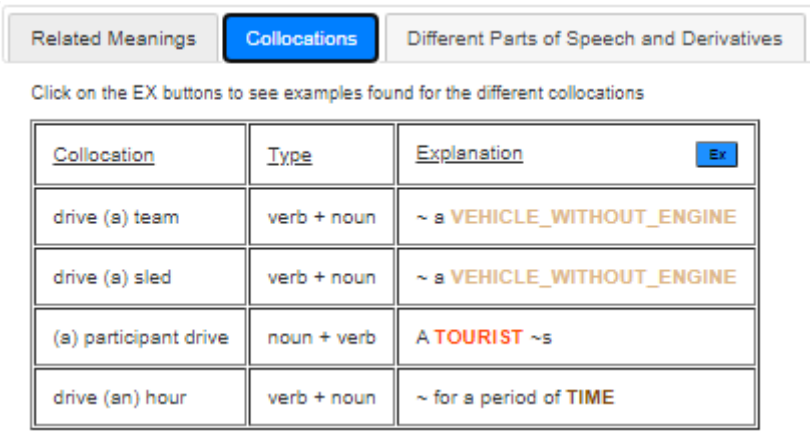
[224] See Siegel & Retter (2015) for a careful description of eXist.
[225] https://www.oxygenxml.com/about_us.html (Last accessed: 25/10/2020).
[226] Different colors are used to represent distinct elements.

participant drive</lien>
</lien-lexical>
<lien-lexical>
<explication-ra>~ for a period of TIME</explication-ra>
<explication-tt/>
<type>verb + noun</type>
<lien identificateur="drive (an) hour" numero-acception="">drive (an) hour</lien>
</lien-lexical>
</famille>

As it can be observed, every XML file represents a hierarchy of nested objects and contains an opening tag and a closing tag. The former begins with a left angle bracket ("<"), is followed by an element name and finishes with a right angle bracket (">"), for example, "<famille nom="Collocation">." The latter begins with the same type of bracket preceded by a slash ("</"), includes exactly the same name as the opening tag and closes with a right angle bracket too, for example, "</famille>." The information contained in every XML file consists of *attributes* and *titles*. Attributes are the elements which are never presented in a browser, such as "nom=," whereas titles are the elements that will be displayed. For instance, "famille," which is the name of the tag opening and closing the XML file, defines the 'Collocations' tab. Then, the information about each of the collocations included in the entry is represented by a tag called "lien-lexical;" as we can see, four files use this tag, which means that four collocations were selected for the verb *drive*$_1$ (each of these files makes a row in the table). Moreover, each of these files includes three items of information on every collocation, introduced by distinct tags: (1) every collocation of the verb uses the "lien" tag, for example, "<lien identificateur="(a) participant drive" numero-acception="">(a) participant drive</lien>," which corresponds to the first column shown in the 'Collocations' tab in the dictionary; (2) the lexical type of the collocation uses the "type" tag, such as "<type>noun + verb</type>," which is the second column in the 'Collocations' section; and (3) the explanation of each collocation uses the "explication-ra" tag, in which the semantic roles of the collocates are specified (i.e., the capital letters), for example, "<explication-ra>A TOURIST ~s</explication-ra>;" this is the third column of the table. Figure 74 illustrates the appearance of the XML file in a web browser:

Figure 74

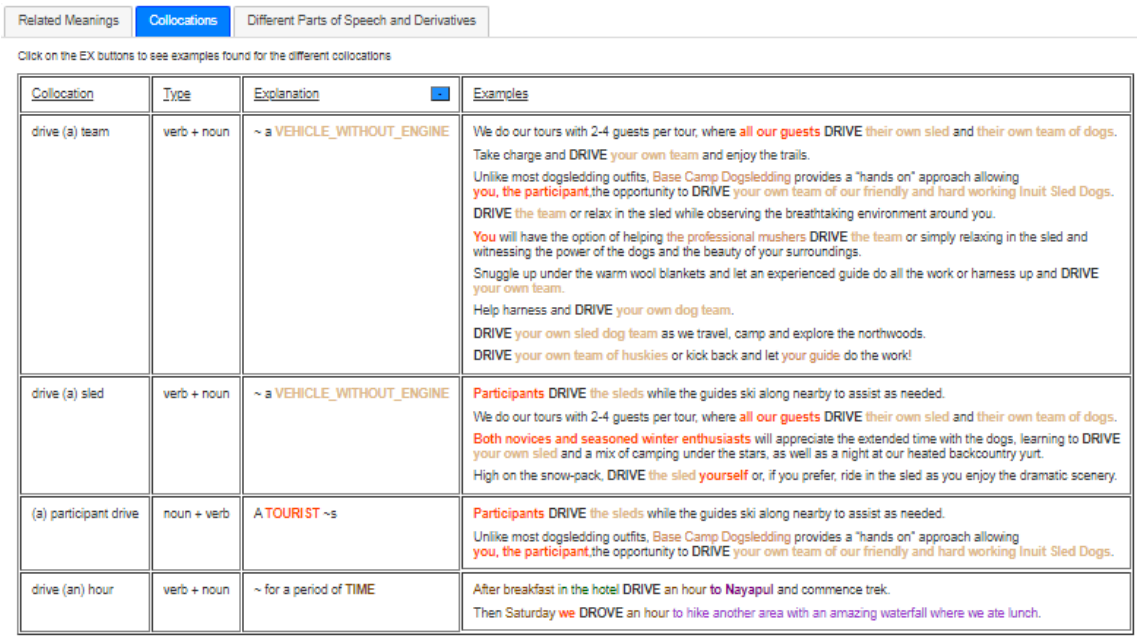
'Collocations' Tab in the Entry for drive$_1$ *in* DicoAdventure *('Examples' Column Unexpanded)*



As we can directly observe, the 'Collocations' tab of *DicoAdventure* uses the template suggested in Section 4.5. In this case, Figure 74 displays a table illustrating the information of the four collocations selected for *drive$_1$* in an accessible and clear way. Thus, the three distinct columns match the XML files previously described: first, the collocations are shown; second, their lexical types are specified; and third, an explanation of the collocations including the semantic roles of the collocates (which appeared in capital letters in the XML file), which are represented in colors (cf. §4.2.). Additionally, the 'Explanation' column offers access to the 'Examples' column, that is, the annotated examples of the collocations in context extracted from the ADVENCOR corpus, through the 'Ex' button. The annotation of the examples depends on the semantic roles played by the linguistic elements in the sentences and use a different color for every semantic role (as it can be seen in the explanations of the collocations). For instance, light brown is used for VEHICLE_WITHOUT_ENGINE, red for TOURIST, purple for DESTINATION and lilac for PURPOSE, among others. Figure 75 shows the result of the annotations in the examples of the collocations encoded in the entry for *drive$_1$*:

Figure 75

'Collocations' Tab in the Entry for drive$_1$ *in* DicoAdventure *('Examples' Column Expanded)*



Related Meanings | Collocations | Different Parts of Speech and Derivatives

Click on the EX buttons to see examples found for the different collocations

Collocation	Type	Explanation	Examples
drive (a) team	verb + noun	~ a VEHICLE_WITHOUT_ENGINE	We do our tours with 2-4 guests per tour, where all our guests DRIVE their own sled and their own team of dogs. Take charge and DRIVE your own team and enjoy the trails. Unlike most dogsledding outfits, Base Camp Dogsledding provides a "hands on" approach allowing you, the participant,the opportunity to DRIVE your own team of our friendly and hard working Inuit Sled Dogs. DRIVE the team or relax in the sled while observing the breathtaking environment around you. You will have the option of helping the professional mushers DRIVE the team or simply relaxing in the sled and witnessing the power of the dogs and the beauty of your surroundings. Snuggle up under the warm wool blankets and let an experienced guide do all the work or harness up and DRIVE your own team. Help harness and DRIVE your own dog team. DRIVE your own sled dog team as we travel, camp and explore the northwoods. DRIVE your own team of huskies or kick back and let your guide do the work!
drive (a) sled	verb + noun	~ a VEHICLE_WITHOUT_ENGINE	Participants DRIVE the sleds while the guides ski along nearby to assist as needed. We do our tours with 2-4 guests per tour, where all our guests DRIVE their own sled and their own team of dogs. Both novices and seasoned winter enthusiasts will appreciate the extended time with the dogs, learning to DRIVE your own sled and a mix of camping under the stars, as well as a night at our heated backcountry yurt. High on the snow-pack, DRIVE the sled yourself or, if you prefer, ride in the sled as you enjoy the dramatic scenery.
(a) participant drive	noun + verb	A TOURIST ~s	Participants DRIVE the sleds while the guides ski along nearby to assist as needed. Unlike most dogsledding outfits, Base Camp Dogsledding provides a "hands on" approach allowing you, the participant,the opportunity to DRIVE your own team of our friendly and hard working Inuit Sled Dogs.
drive (an) hour	verb + noun	~ for a period of TIME	After breakfast in the hotel DRIVE an hour to Nayapul and commence trek. Then Saturday we DROVE an hour to hike another area with an amazing waterfall where we ate lunch.

All the collocations selected in this work went through the same implementation process and they are all encoded in *DicoAdventure* in the same way, which was the main objective of this research. They were also considered for the analysis performed in the following chapter.

CHAPTER 6. ANALYSIS

[...] it is important to base one's analysis of language on real data –actual instances of speech or writing– rather than on data that are contrived or "made-up."

(Meyer, 2002, p. xiii)

Chapter 6 will analyze the extensive data derived from the methodological steps explained in Chapter 4 and presented in Chapter 5 according to four aspects: first, the keyness and collocation production of the motion verbs (§6.1.); second, the strength and frequency of the collocations selected (§6.2.); third, the lexical types of these collocations (§6.3.); and fourth, the meaning of the verbal bases and their co-occurrents in terms of the Frame Semantics theory (§6.4.). To this end, this overall analysis will cover all the motion verbs[227] which produced collocations representing any type of motion, that is, those examined in Chapter 5[228] and those included in the Appendix.[229]

6.1. Verbal keyness and collocation production

Step one of our methodology (cf. §4.4.1.) extracted 978 potential candidate verbs for the bases of the collocations that would be implemented in *DicoAdventure*. After manual close inspection, 118 motion verbs related to the field of adventure tourism were selected (cf. Table 22). Next, step two allowed us to extract collocations containing these verbs by applying two criteria: (1) a minimum *logDice* score of 7 points, and (2) a minimum occurrence of two tokens. At this point, less than half of the verbs (55, cf. Table 23) produced collocations. However, step three, according to which collocations were

[227] In total, 30 verbs and 36 meanings will be considered, given that six of the bases (viz., *ascend*-v, *climb*-v, *cross*-v, *descend*-v, *head*-v and *ride*-v) produced two different meanings. It must be remembered that this is an ongoing project, so *DicoAdventure* may include more data in the future.
[228] *ascend*-v, *climb*-v, *cross*-v, *descend*-v and *head*-v.
[229] *abseil*-v, *arrive*-v, *drive*-v, *enter*-v, *exit*-v, *fall*-v, *float*-v, *fly*-v, *glide*-v, *hike*-v, *jump*-v, *land*-v, *navigate*-v, *paddle*-v, *reach*-v, *ride*-v, *scale*-v, *scramble*-v, *skydive*-v, *slide*-v, *swim*-v, *traverse*-v, *trek*-v, *venture*-v and *walk*-v.

classified in terms of whether they depicted actual or fictive motion, revealed that only 30 items (i.e., 25.48% of the motion verbs selected in step one, cf. Table 24) produced collocations that met all eligibility criteria, that is, these collocations were identified in at least two contexts representing the same type of meaning and made an explicit reference to the domain of adventure tourism. Albeit all the 30 verbs were more recurrent in the ADVENCOR corpus than in the reference corpus, that is, the 'enTenTen13,' their keyness scores were quite different, as some of them are more typical of the SL under analysis than others.[230] It is visually represented in Figure 76, which shows a term cloud where verbs displaying a higher keyness score[231] look bigger:

Figure 76

Term Cloud of the Motion Verbs Whose Collocations Met all Eligibility Criteria

As it can be observed, *skydive*-v is the verb with the highest keyness score (279.650), that is, there is a higher difference with respect to the rest of the verbs if its frequency in the ADVENCOR corpus is compared to its frequency in the 'enTenTen13.' Other verbs with high keyness scores are *trek*-v (142.670), *abseil*-v (131.580), *glide*-v (62.140), *hike*-v (57.490), *paddle*-v (53.330) and *climb*-v (42.610), most of which allude to an adventure activity that is land-based, such as trekking, abseiling, hiking and climbing. These results

[230] Generally, a higher value (e.g., 100, 200) implies a higher frequency, whereas a lower value (e.g., 1, 0.1) implies a low frequency (more rare verbs).
[231] The keyness scores of these verbs are shown in Table 22 (cf. §5.1.).

show that these verbs are more typical of the SL of adventure tourism than others, for example, *arrive*-v, *drive*-v, *enter*-v or *walk*-v. Nevertheless, it does not imply a higher degree of collocation production. To put it differently, only one collocation was extracted for *skydive*-v, *abseil*-v and *glide*-v, two collocations for *trek*-v and three for *hike*-v. On the other hand, *paddle*-v produced eight collocations and *climb*-v 15. Table 30 displays the number of collocations that each of the selected bases produced in the ADVENCOR corpus in descending order:

Table 30

Number of Collocations Produced by the Motion Verbs Whose Collocations Met all Eligibility Criteria

	Verb	Number of collocations produced
1.	*climb*-v	15
2.	*reach*-v	14
3.	*fly*-v	10
4.	*cross*-v	9
5.	*descend*-v	8
6.	*paddle*-v	8
7.	*head*-v	7
8.	*ride*-v	7
9.	*ascend*-v	5
10.	*drive*-v	4
11.	*navigate*-v	4
12.	*walk*-v	4
13.	*hike*-v	3
14.	*scale*-v	3
15.	*float*-v	2
16.	*slide*-v	2
17.	*traverse*-v	2
18.	*trek*-v	2
19.	*abseil*-v	1
20.	*arrive*-v	1
21.	*enter*-v	1
22.	*exit*-v	1

	Verb	Number of collocations produced
23.	*fall*-v	1
24.	*glide*-v	1
25.	*jump*-v	1
26.	*land*-v	1
27.	*scramble*-v	1
28.	*skydive*-v	1
29.	*swim*-v	1
30.	*venture*-v	1

In total, 121 collocations were selected. As we can see, the most productive verbs in terms of collocations are *climb*-v, *reach*-v, *fly*-v, *cross*-v, *descend*-v, *paddle*-v, *head*-v, *ride*-v and *ascend*-v, which are the bases of five or more collocations. However, if we focus on the collocations which represent actual motion,[232] the most productive verbs are *reach*$_1$-v, *climb*$_1$-v, *fly*$_1$-v, *paddle*$_1$-v and *ride*$_1$-v. That is, *cross*$_1$-v, *descend*$_1$-v, *head*$_1$-v and *ascend*$_1$-v are dropped from the list because they produced a high number of collocations representing fictive motion. Table 31 illustrates the verbs along with the number of collocations describing actual motion in descending order:

Table 31

Number of Collocations Representing Actual Motion in the ADVENCOR Corpus

	Verb	Number of collocations representing actual motion
1.	*reach*$_1$-v	14
2.	*climb*$_1$-v	10
3.	*fly*$_1$-v	10
4.	*paddle*$_1$-v	8
5.	*ride*$_1$-v	5
6.	*cross*$_1$-v	4
7.	*descend*$_1$-v	4
8.	*drive*$_1$-v	4

[232] Step three of our methodology classified the collocations extracted into two groups: (1) representation of actual motion, and (2) representation of fictive motion. It was possible thanks to the concordance lines that showed the collocations in context extracted from the specialized corpus.

	Verb	Number of collocations representing actual motion
9.	*head*$_1$-v	4
10.	*navigate*$_1$-v	4
11.	*walk*$_1$-v	4
12.	*hike*$_1$-v	3
13.	*scale*$_1$-v	3
14.	*float*$_1$-v	2
15.	*ride*$_2$-v[233]	2
16.	*slide*$_1$-v	2
17.	*traverse*$_1$-v	2
18.	*trek*$_1$-v	2
19.	*abseil*$_1$-v	1
20.	*arrive*$_1$-v	1
21.	*ascend*$_1$-v	1
22.	*enter*$_1$-v	1
23.	*exit*$_1$-v	1
24.	*fall*$_1$-v	1
25.	*glide*$_1$-v	1
26.	*jump*$_1$-v	1
27.	*land*$_1$-v	1
28.	*scramble*$_1$-v	1
29.	*skydive*$_1$-v	1
30.	*swim*$_1$-v	1
31.	*venture*$_1$-v	1

In total, 100 collocations display actual motion events. On the other hand, Table 32 shows the verbs along with the number of collocations depicting fictive motion; this number amounts to 28:

[233]*Ride*-v has two different meanings describing actual motion, so entry *ride*$_1$ in *DicoAdventure* includes collocations where a tourist controls and guides the movements of a vehicle, for example, *It is a sport where one **rides bicycles***, and entry *ride*$_2$ includes collocations where a tourist travels on a vehicle as a passenger, for instance, *Hikers can **ride** the ski **lift**.*

Table 32

Number of Collocations Representing Fictive Motion in the ADVENCOR Corpus

	Verb	Number of collocations representing fictive motion
1.	*cross*$_2$-v	8
2.	*climb*$_2$-v	7
3.	*descend*$_2$-v	5
4.	*ascend*$_2$-v	4
5.	*head*$_2$-v	4

Given that the verbs *ascend*-v, *climb*-v, *cross*-v, *descend*-v and *head*-v produced collocations conveying two different meanings, that is, actual and fictive motion, they can be considered the richest verbs in terms of meaning in this study (along with *ride*-v, whose collocations express two types of actual motion). Table 33 summarizes the data obtained from these verbs, which were analyzed in Section 5.4.:

Table 33

Verbs Which Produced Collocations Representing Actual and Fictive Motion in the ADVENCOR Corpus

	Verb	Total number of collocations	Number of collocations representing only actual motion	Number of collocations representing only fictive motion	Number of collocations representing actual and fictive motion
1.	*climb*-v	15	8	5	2
2.	*cross*-v	9	1	5	3
3.	*descend*-v	8	3	4	1
4.	*head*-v	7	3	3	1
5.	*ascend*-v	5	1	4	0

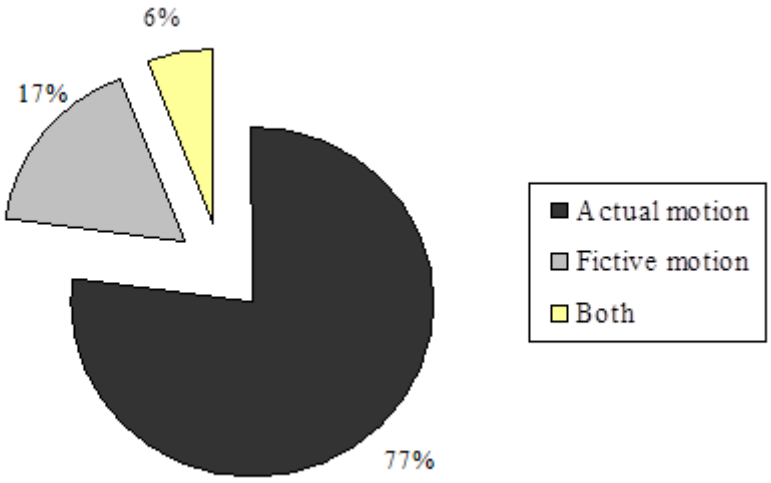
As Table 33 indicates, *climb*-v is the verb which produced more collocations in total (15) and is also the verb with more collocations depicting actual motion (10). However, fictive motion events are mainly described by collocations containing the base *cross*$_2$-v (eight in total). In short, the total number of collocations produced by the five verbs representing

each type of motion, actual and fictive, is: (1) *ascend*$_1$-v: 1, *ascend*$_2$-v: 4; (2) *climb*$_1$-v: 10, *climb*$_2$-v: 7; (3) *cross*$_1$-v: 4, *cross*$_2$-v: 8; (4) *descend*$_1$-v: 4, *descend*$_2$-v: 5; and (5) *head*$_1$-v: 4, *head*$_2$-v: 4.

To sum up, the number of collocations that satisfied all eligibility criteria established in this work was 121, from which 128 distinct meanings emerged. In other words, 93 collocations occurred only in actual motion events, 21 described only fictive motion events and seven were found in both types of motion events. It goes without saying that actual motion is more recurrent in the specialized corpus than fictive motion. These results are shown in Figure 77:

Figure 77

Motion Representation by Collocations in the ADVENCOR Corpus



After having examined the verbal keyness score of the candidate verbs which produced collocations meeting all eligibility criteria and the collocation production of the 36 bases in terms of the type of motion represented, the following section will evaluate the strength of association between the elements of the collocations retrieved for each of the motion verbs.

6.2. Collocation strength and frequency

As it was explained in Section 5.4., the tables illustrating the collocations selected included four items of information: (1) their *logDice* score in the specialized corpus, (2)

their frequency in the specialized corpus, (3) their *logDice* score in the reference corpus, and (4) their frequency in the reference corpus.

After considering the results obtained from the 36 verbal bases,[234] a fundamental aspect is that in more than 95% of the cases the *logDice* score of the collocations is higher in the ADVENCOR corpus than in the 'enTenTen13,' which means that they can be regarded as specialized collocations. In fact, the only exceptions are: *climb (a) mountain* (9.52 vs. 10.09), *climb (a) tree* (7.09 vs. 8.85), *fall back* (7.35 vs. 8.08), *reach (a) peak* (8.04 vs. 8.05), *ride (a) bike* (10.6 vs. 11.36) and *walk (one) mile* (8.87 vs. 9.04). Nevertheless, we think that these minor differences between both scores must not question the relevance of these collocations in the SL of adventure tourism,[235] for three major reasons: (1) they refer to actions closely linked to this domain; (2) the elements found in each of the contexts retrieved also pertain to this field, such as example (17): *Guests are attached to a safety cable system as they* ***climb*** *the* ***mountain****, enjoying the views of the canyon and Ogden on their way up* or example (202): *Via ferratas are just awesome because you really don't have to be a mountaineer, nor a rock-climber to* ***reach*** *wild* ***peaks*** *with jawdropping views*; and (3) they fulfilled all the eligibility criteria. Thus, they were without doubt included in this dissertation.

On the other hand, in Section 4.4.2. we explained that the *logDice* score is an association measure that tells how strong an association between two words is. Additionally, it is related to typicality, which means that collocations displaying higher scores are less predictable and are worthy of learning. In order to know which verbs produced stronger collocations in this research, Table 34 lists the average *logDice* score of the collocations retrieved for the 30 different verbs in the specialized corpus in descending order, along with the number of collocations that they produced:

[234] Ten bases were covered in Section 5.4. and 26 bases are gathered in the Appendix.
[235] Their high scores in the reference corpus imply that they are also common in the GL.

Table 34

*Collocation Strength (*logDice*) of the Verbs in the* ADVENCOR *Corpus*

	Verb	Average *logDice* in ADVENCOR	Number of collocations produced
1.	*exit*-v	12.02	1
2.	*skydive*-v	11.96	1
3.	*glide*-v	11.14	1
4.	*arrive*-v	11.04	1
5.	*land*-v	10.43	1
6.	*abseil*-v	10.38	1
7.	*hike*-v	10.34	3
8.	*jump*-v	10.2	1
9.	*ascend*-v	9.95	5
10.	*enter*-v	9.89	1
11.	*descend*-v	9.88	8
12.	*head*-v	9.8	7
13.	*scale*-v	9.76	3
14.	*float*-v	9.72	2
15.	*drive*-v	9.62	4
16.	*cross*-v	9.61	9
17.	*ride*-v	9.51	7
18.	*fly*-v	9.44	10
19.	*paddle*-v	9.35	8
20.	*slide*-v	9.11	2
21.	*navigate*-v	8.97	4
22.	*climb*-v	8.91	15
23.	*traverse*-v	8.79	2
24.	*reach*-v	8.77	14
25.	*walk*-v	8.5	4
26.	*venture*-v	8.44	1
27.	*trek*-v	8.1	2
28.	*swim*-v	7.69	1
29.	*fall*-v	7.35	1
30.	*scramble*-v	7.18	1

As it can be observed, the bases which produced stronger collocations in the specialized corpus are *exit*-v, *skydive*-v, *glide*-v, *arrive*-v, *land*-v, *abseil*-v, *hike*-v and *jump*-v, for their average *logDice* score was ≥ 10 points. Two features shared are: first, these verbs only describe actual motion, and second, they were not highly productive, as all of them produced only one collocation, except for *hike*-v (three collocations). The most productive verbs are located by the middle and the bottom of the table, such as *cross*-v (nine collocations, row 16), *fly*-v (10 collocations, row 18), *climb*-v (15 collocations, row 22) and *reach*-v (14 collocations, row 24). On the other hand, Table 35 lists the average *logDice* score of the collocations in the reference corpus in descending order:

Table 35

*Collocation Strength (*logDice*) of the Verbs in the 'enTenTen13'*

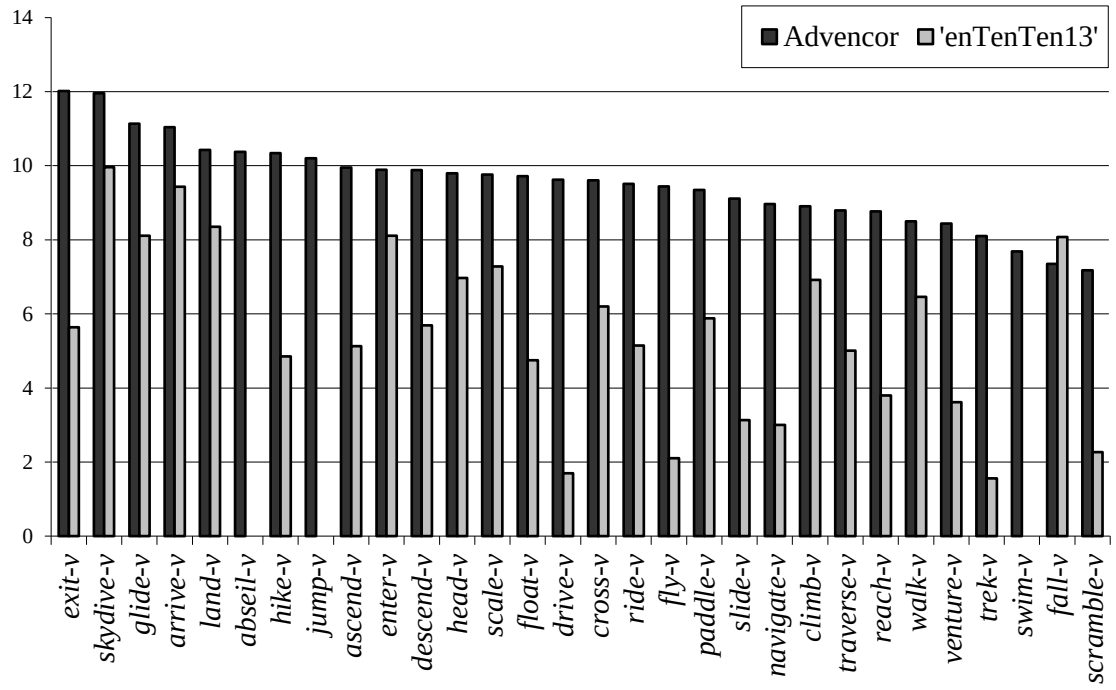
	Verb	**Average *logDice* in 'enTenTen13'**	**Number of collocations produced**
1.	*skydive*-v	9.96	1
2.	*arrive*-v	9.43	1
3.	*land*-v	8.35	1
4.	*enter*-v	8.11	1
5.	*glide*-v	8.11	1
6.	*fall*-v	8.08	1
7.	*scale*-v	7.28	3
8.	*head*-v	6.97	7
9.	*climb*-v	6.92	15
10.	*walk*-v	6.46	4
11.	*cross*-v	6.2	9
12.	*paddle*-v	5.88	8
13.	*descend*-v	5.69	8
14.	*exit*-v	5.64	1
15.	*ride*-v	5.15	7
16.	*ascend*-v	5.13	5
17.	*traverse*-v	5.01	2
18.	*hike*-v	4.85	3
19.	*float*-v	4.75	2
20.	*reach*-v	3.8	14
21.	*venture*-v	3.62	1

	Verb	Average *logDice* in 'enTenTen13'	Number of collocations produced
22.	*slide*-v	3.13	2
23.	*navigate*-v	3	4
24.	*scramble*-v	2.27	1
25.	*fly*-v	2.1	10
26.	*drive*-v	1.7	4
27.	*trek*-v	1.56	2
28.	*abseil*-v	0	1
29.	*jump*-v	0	1
30.	*swim*-v	0	1

In this case, stronger collocations were found in the following verbs: *skydive*-v, *arrive*-v, *land*-v, *enter*-v, *glide*-v, *fall*-v and *scale*-v, since the resulting average *logDice* score was ≥ 7 points. These verbs have in common the characteristics mentioned above, that is, their meanings only represent actual motion and they were not highly productive in terms of collocations (again, all of them produced only one collocation, except for one verb: *scale*-v, which produced three). Unlike in Table 34, the most productive verbs are scattered around the table, for instance, *climb*-v (15 collocations, row 9), *cross*-v (nine collocations, row 11), *reach*-v (14 collocations, row 20) and *fly*-v (10 collocations, row 25). Figure 78 helps to visualize the differences in terms of collocation strength of the verbs in both corpora:

Figure 78

*Collocation Strength (*logDice*) of the Verbs in the ADVENCOR Corpus and the 'enTenTen13'*



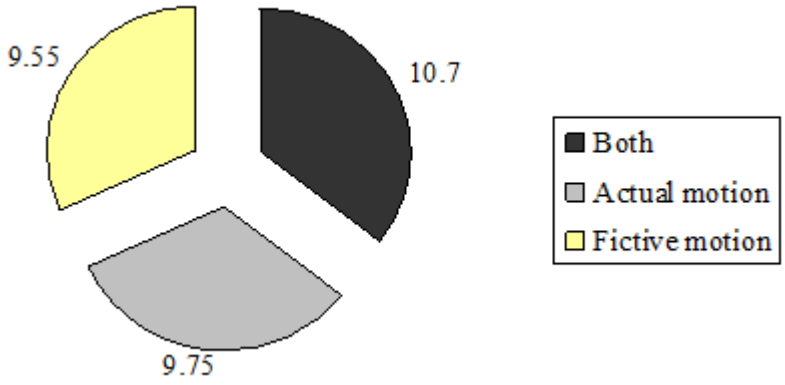
As Figure 78 illustrates, the average *logDice* scores in the 'enTenTen13' are lower than those in the ADVENCOR corpus in all cases, with the exception of *fall*-v, where the only collocation produced, that is, *fall back*, achieved a score of 8.08 in the reference corpus and 7.35 points in the specialized corpus. In fact, the average score in the ADVENCOR corpus is 9.53 points, whereas it is 4.98 points in the 'enTenTen13.' To put it differently, the former almost doubles the latter, which means that the collocations extracted from the specialized corpus are more typical there and must, therefore, be considered specialized collocations.

On the other hand, in terms of the type of motion depicted, the elements of those collocations describing both actual and fictive motion events, for example, *climb up, cross (a) bridge, gradually descend, head north*, seem to be more strongly associated than any others, as the average score achieved when considering the four verbs which produced this type of collocations, namely, *climb*-v, *cross*-v, *descend*-v and *head*-v, was 10.7 points. After that, the average *logDice* score achieved by the 30 verbs whose collocations depicted actual motion was 9.75 points, closely followed by the result

obtained from the five verbs which produced collocations representing fictive motion (i.e., *cross*$_2$-v, *climb*$_2$-v, *descend*$_2$-v, *ascend*$_2$-v and *head*$_2$-v): 9.55 points.[236] Figure 79 visually depicts these results:

Figure 79

*Collocation Strength (*logDice*) in Terms of the Type of Motion Represented in the* ADVENCOR *Corpus*



Finally, one conclusion that can be drawn from these findings is that collocation production and collocation strength are not interrelated, for they do not increase at the same pace. Thus, the collocations produced by a highly productive verb, such as *climb*-v (15 collocations), might have a medium *logDice* score (8.91 in the specialized corpus and 6.92 in the reference corpus), whereas the collocations produced by a rather unproductive verb, such as *skydive*-v (one collocation), might have a very high *logDice* score (11.96 in the specialized corpus and 9.96 in the reference corpus).[237]

Moving on to the collocation frequency in this dissertation, the verbs which produced more tokens in the ADVENCOR corpus were *climb*-v, *hike*-v, *reach*-v, *head*-v, *cross*-v and *fly*-v, since over 60 contexts were identified containing these bases. These results are shown in Table 36:

[236] It must be clarified that these results were obtained from the ADVENCOR corpus, given that all the concordance lines retrieved from the corpus, unlike in the 'enTenTen13,' were manually checked so as to make sure of the type of motion described in each of them.
[237] Incidentally, *skydive*-v was the verb which obtained the highest keyness score in the verbal terminology extraction performed in this work.

Table 36

Number of Tokens of the Collocations in the ADVENCOR Corpus

	Verb	Number of tokens of the collocations in ADVENCOR	Number of collocations produced
1.	*climb*-v	199	15
2.	*hike*-v	148	3
3.	*reach*-v	105	14
4.	*head*-v	71	7
5.	*cross*-v	70	9
6.	*fly*-v	61	10
7.	*drive*-v	49	4
8.	*ride*-v	37	7
9.	*descend*-v	32	8
10.	*paddle*-v	30	8
11.	*ascend*-v	19	5
12.	*walk*-v	18	4
13.	*scale*-v	12	3
14.	*navigate*-v	11	4
15.	*slide*-v	10	2
16.	*trek*-v	9	2
17.	*exit*-v	9	1
18.	*skydive*-v	9	1
19.	*venture*-v	9	1
20.	*enter*-v	7	1
21.	*float*-v	6	2
22.	*traverse*-v	5	2
23.	*arrive*-v	5	1
24.	*jump*-v	5	1
25.	*land*-v	4	1
26.	*scramble*-v	4	1
27.	*fall*-v	3	1
28.	*swim*-v	3	1
29.	*abseil*-v	2	1
30.	*glide*-v	2	1

One may expect that the verbs which produced a higher number of collocations also presented more tokens in total, such as *climb*-v (15 types, 199 tokens), *reach*-v (14 types, 105 tokens) and *fly*-v (10 types, 61 tokens). However, the cases of *hike*-v (three types, 148 tokens) and *head*-v (seven types, 71 tokens) must be highlighted, given that they were not highly productive in terms of collocations, but the collocations retrieved had a frequent recurrence in the corpus. Thus, we can state that collocation production and collocation frequency are not interrelated, for one verb may produce a substantial amount of rather infrequent collocations or a tiny amount of extremely frequent collocations. On the other hand, the reccurrence of collocations in the reference corpus is counted in thousands of tokens for almost every verb, as Table 37 displays:

Table 37

Number of Tokens of the Collocations in the 'enTenTen13'

	Verb	Number of tokens of the collocations in 'enTenTen13'	Number of collocations produced
1.	*reach*-v	167,592	14
2.	*head*-v	142,139	7
3.	*ride*-v	108,841	7
4.	*climb*-v	98,742	15
5.	*fall*-v	76,625	1
6.	*cross*-v	65,736	9
7.	*walk*-v	37,456	4
8.	*hike*-v	37,108	3
9.	*enter*-v	33,776	1
10.	*arrive*-v	17,485	1
11.	*drive*-v	11,492	4
12.	*fly*-v	7,261	10
13.	*descend*-v	6,416	8
14.	*paddle*-v	5,696	8
15.	*scale*-v	5,594	3
16.	*navigate*-v	5,495	4
17.	*slide*-v	4,789	2
18.	*land*-v	4,249	1
19.	*float*-v	3,711	2

	Verb	Number of tokens of the collocations in 'enTenTen13'	Number of collocations produced
20.	*scramble*-v	2,663	1
21.	*ascend*-v	1,740	5
22.	*trek*-v	1,590	2
23.	*glide*-v	1,009	1
24.	*venture*-v	655	1
25.	*exit*-v	440	1
26.	*traverse*-v	375	2
27.	*skydive*-v	75	1
28.	*abseil*-v	0	1
29.	*jump*-v	0	1
30.	*swim*-v	0	1

As it can be seen, the verbs which produced more tokens in the 'enTenTen13' were *reach*-v, *head*-v, *ride*-v, *climb*-v, *fall*-v and *cross*-v (more than 60,000). This is almost the same top six as in the specialized corpus, although *ride*-v and *fall*-v are new. In fact, *fall*-v must be emphasized, as it only produced one collocation but it recurred very frequently in the corpus (76,625 times).

On the other hand, in principle it may be stated that the collocation frequency in the reference corpus outnumbers that in the ADVENCOR corpus, as the former is measured in thousands and the latter is mainly measured in tens. Nevertheless, if the normalized frequency per million[238] is considered, it must be pointed out that the collocations selected would occur more frequently in the ADVENCOR corpus than in the 'enTenTen13.' Table 38 illustrates these differences (the verbs are arranged in descending order according to their normalized frequency per million in the ADVENCOR corpus):

[238] It must be remembered that this score allows linguists to compare frequencies in corpora of different sizes and is calculated by dividing the overall frequency of the collocations in the corpus by the number of words contained in the corpus and, then, the result is multiplied by 1,000,000.

Table 38

Normalized Frequency per Million of the Collocations in the ADVENCOR Corpus and the 'enTenTen13'

	Verb	Normalized frequency per million of the collocations in ADVENCOR	Normalized frequency per million of the collocations in 'enTenTen13'
1.	*climb*-v	168.75	5.01
2.	*hike*-v	125.50	1.71
3.	*reach*-v	89.04	8.51
4.	*head*-v	60.2	7.22
5.	*cross*-v	59.36	3.33
6.	*fly*-v	51.72	0.36
7.	*drive*-v	41.55	0.58
8.	*ride*-v	31.37	5.52
9.	*descend*-v	27.13	0.32
10.	*paddle*-v	25.44	0.28
11.	*ascend*-v	16.11	0.08
12.	*walk*-v	15.26	1.9
13.	*scale*-v	10.17	0.28
14.	*navigate*-v	9.32	0.27
15.	*slide*-v	8.48	0.24
16.	*trek*-v	7.63	0.08
17.	*exit*-v	7.63	0.02
18.	*skydive*-v	7.63	0.003
19.	*venture*-v	7.63	0.03
20.	*enter*-v	5.93	1.71
21.	*float*-v	5.08	0.18
22.	*traverse*-v	4.24	0.01
23.	*arrive*-v	4.24	0.88
24.	*jump*-v	4.24	0
25.	*land*-v	3.39	0.21
26.	*scramble*-v	3.39	0.13
27.	*fall*-v	2.54	3.89
28.	*swim*-v	2.54	0
29.	*abseil*-v	1.69	0
30.	*glide*-v	1.69	0.05

As we can see, in all cases the normalized frequency per million of the collocations in the ADVENCOR corpus is much higher than that in the 'enTenTen13.'[239] Therefore, if blocks of one thousand million words were considered, the collocations would occur more repeatedly in the specialized corpus than in the reference corpus. In other words, they would be more typical of the SL of adventure tourism, which allows us to openly admit that we are dealing with specialized collocations of this SL and must be implemented in *DicoAdventure*.

With respect to the four verbs which produced collocations whose contexts unveiled the description of both types of motion, namely, *climb*-v, *cross*-v, *descend*-v and *head*-v, Table 39 shows the collocations of this type retrieved for each base along with the number of tokens representing each type of motion:

Table 39

Number of Tokens of the Collocations Representing Actual and Fictive Motion in the ADVENCOR Corpus

	Verb	Collocation representing actual and fictive motion	Number of tokens of the collocation representing actual motion	Number of tokens of the collocation representing fictive motion
1.	*climb*$_{1,2}$-v	*climb up*	27	5
		climb (a) face	5	2
2.	*cross*$_{1,2}$-v	*cross (a) bridge*	18	2
		cross (a) river	10	12
		cross (a) stream	7	3
3.	*descend*$_{1,2}$-v	*gradually descend*	2	5
4.	*head*$_{1,2}$-v	*head north*	5	3

As it can be observed, contexts depicting actual motion were more frequent in collocations that represented both types of motion event, with the exceptions of *cross (a) river* (10 vs. 12) and *gradually descend* (2 vs. 5). In principle, there seems to be no apparent explanation for that, since these collocations do not share any features, for they

[239] There is only one exception: *fall*-v (2.54 vs. 3.89).

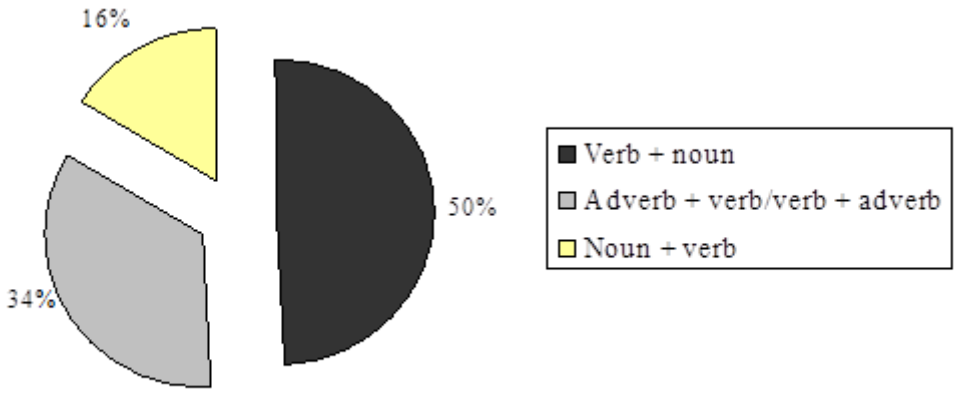
contain a distinct verbal base and they belong to different lexical types.[240] An analysis of the lexical types of the collocations selected in this dissertation will be provided in the next section.

6.3. Lexical types of the collocations selected

Section 4.4.2. explained that this work was focused on lexical collocations, that is, those consisting of two content words, such as Le2-Verb + noun, Le3-Noun + verb, Le6-Adverb + verb/verb + adverb and Le7-Verb + adjective (cf. Table 12). Then, Section 5.2. revealed that no verb + adjective collocations were extracted for the verbal bases selected in step one of our methodology. As a consequence, the analysis carried out in this section will deal with the frequency and collocation strength of the remaining lexical types, that is, verb + noun, noun + verb and adverb + verb/verb + adverb collocations. Figure 80 displays the frequency of each type in the ADVENCOR corpus:

Figure 80

Frequency of the Lexical Types of the Collocations Selected in the ADVENCOR Corpus



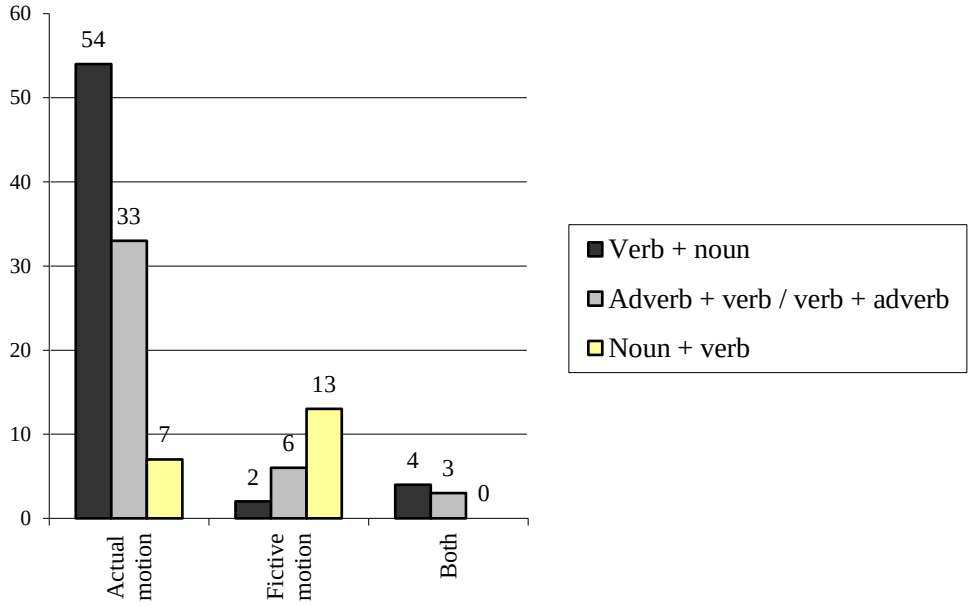
As it can be observed, the collocations containing the verbal base and a nominal collocate in the object position, such as *drive (a) team, navigate (a) waterfall, reach (a) peak, ride (a) wave*, occupy half of the diagram (60 types in total). After that, the combinations where the verb is modified by an adverb, occurring either before or after the base, for

[240] A key point is that the collocations representing actual and fictive motion events containing the verbs *cross*$_{1,2}$-v, *descend*$_{1,2}$-v *and head*$_{1,2}$-v were the strongest collocations produced by the verb involved, as it was observed in Sections 5.4.3., 5.4.4. and 5.4.5., respectively.

example, *arrive early, gently paddle, tandem skydive, trek down*, are present in 34% of the total (42 types). Finally, the least recurrent type of collocation is that where the base is preceded by a noun in the subject position, for example, *(a) trail ascend, (a) glider fly, (a) path head, (a) guide walk*, since it makes 16% of the total (20 types). In terms of the type of motion described by the collocations, the results are very much alike, as we can see in Figure 81:

Figure 81

Frequency of the Lexical Types of the Collocations Selected in Terms of the Type of Motion Represented in the ADVENCOR Corpus



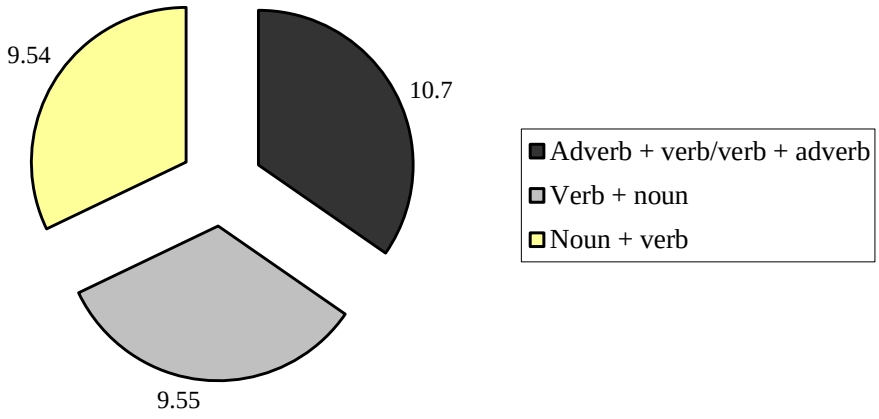
As Figure 81 shows, actual motion events and those contexts where both types of motion, actual and fictive, were discovered obtained similar results, given that verb + noun collocations are the most recurrent type (57.45% and 57.14%, respectively), followed by adverb + verb/verb + adverb collocations (35.1% and 42.86%, respectively). Regarding noun + verb collocations, they only occupy 7.45% of the actual motion events and have no representation in those events representing both types of motion. An explanation for this is that the noun of these combinations designates the subject of the verb, which is an animate entity in actual motion events (i.e., a person) and an inanimate entity in fictive motion events (e.g., a track, a road), therefore, they never coincide. Nevertheless, noun + verb collocations is the most common type found in fictive motion events (61.9%), which

means that the subject appearing in these lexical collocations usually refers to an inanimate entity, such as *(a) climb ascend* or *(a) trek cross*. After that, adverb + verb/verb + adverb collocations also occupy the second place in this type of motion events (28.57%) and, finally, there are only two types of verb + noun collocations (9.53%), namely, *ascend (a) slope* and *descend (a) gully*.

Moving on to the strength of the lexical types in the ADVENCOR corpus, the most recurrent one, that is, verb + noun, is not the one achieving the highest score, for it was found in adverb + verb/verb + adverb collocations. On the other hand, the noun + verb type was the one obtaining the lowest *logDice* score, although there is an extremely slight difference with respect to the strength of verb + noun collocations, as it is illustrated in Figure 82:

Figure 82

*Strength (*logDice*) of the Lexical Types of the Collocations Selected in the ADVENCOR Corpus*

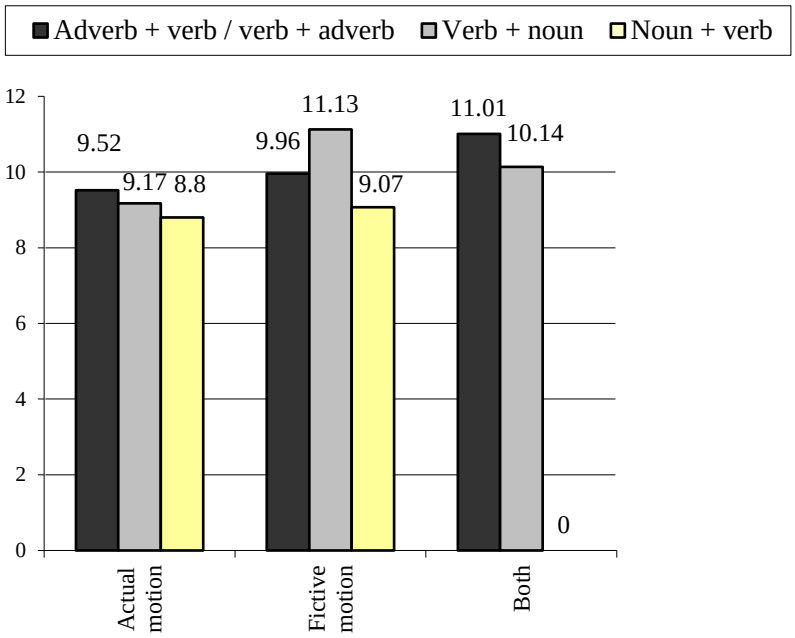


If we consider the type of motion represented by the abovementioned lexical types of collocations, Figure 83 indicates that adverb + verb/verb + adverb collocations are also the strongest type of the collocations describing only actual motion events as well as in those describing both types of motion. Similarly, this type is followed by verb + noun and noun + verb collocations, although the latter is not presented in collocations depicting both types of motion, for the sound reason mentioned above. On the contrary, the strongest collocations in fictive motion events are of the type verb + noun (despite the fact that the most recurrent type is noun + verb), with a difference of almost 1.5 points

with respect to adverb + verb/verb + adverb collocations and of more than 2 points with respect to verb + noun collocations:

Figure 83

*Strength (*logDice*) of the Lexical Types of the Collocations Selected in Terms of the Type of Motion Described in the ADVENCOR Corpus*



In some cases, the type of word accompanying the base of the collocation is motivated by the meaning of the verb, such as a noun in the object position designating a vehicle (e.g., *fly (a) parachute*) or an adverb indicating the direction of the movement (e.g., *paddle back*). This is the topic explored in the next section.

6.4. Semantic analysis of the motion verbs and their co-occurrents

As it was explained in Section 1.5.1.1., the theory of Frame Semantics by Fillmore (1977, 1982, 1985) posits that words are not directly related to each other, but they belong to semantic frames which provide speakers with the conceptual base to determine the concept that a word designates. More specifically, the elements contained in these frames are entities involved in a specific situation, therefore, they help to interpret words. The application of Frame Semantics is FN, the frame-based database which includes hundreds of frames evoked by common vocabulary in English. The information collected in this

database was used as a source of reference to perform the semantic analysis of the bases of the collocations implemented in *DicoAdventure*, as it provides a solid basis to approach the meaning of the verbs and to explore the occurrence of their collocates[241] so as to better understand how knowledge is organized in the field of adventure tourism (cf. Durán-Muñoz, in press; Durán-Muñoz & L'Homme, 2020). Nevertheless, a key point is that this information could not cater for all needs identified in the study of the SL of adventure tourism, given that it is a resource covering GL, therefore, tailored frames were created in order to provide a more detailed analysis that satisfies the requirements set in this SL.[242]

First of all, it must be said that a total of 12 semantic frames were discovered,[243] which are shown in Table 40 in descending order according to the number of motion verbs evoking them:

Table 40

Semantic Frames Evoked by the Bases of the Collocations

	Semantic frame	Base
1.	OPERATE_VEHICLE	*drive*$_1$-v
		float$_1$-v
		fly$_1$-v
		glide$_1$-v
		navigate$_1$-v
		paddle$_1$-v
		ride$_1$-v
2.	SELF_MOTION	*head*$_1$-v
		hike$_1$-v
		scramble$_1$-v
		swim$_1$-v
		trek$_1$-v
		venture$_1$-v
		walk$_1$-v

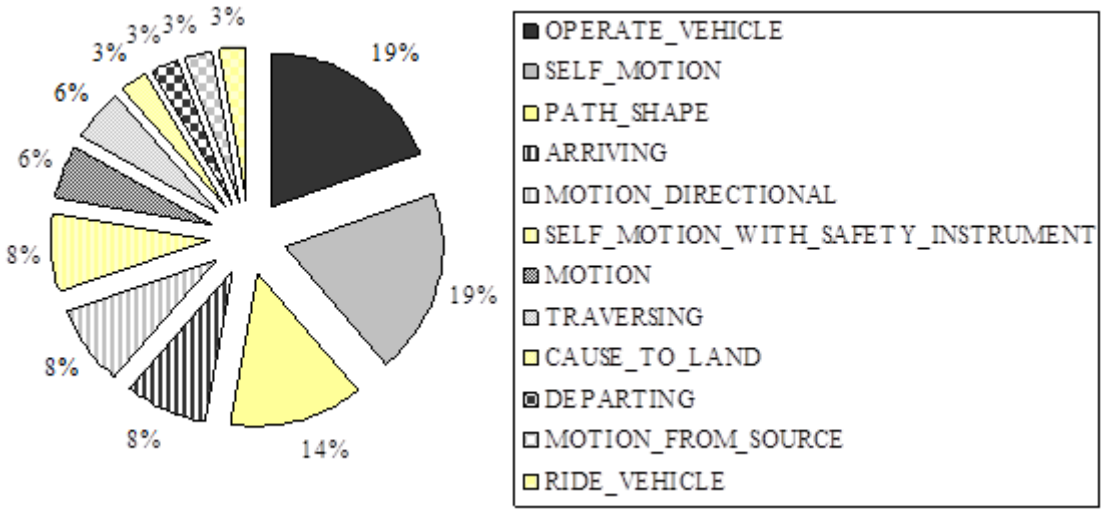
[241] According to Martin (2008, p. 56), collocates zoom in on a specific aspect of the frame triggered by the base.
[242] It must be emphasized that it was discovered in research carried out as part of an ongoing project which was set up before writing the current dissertation. The analysis presented here is a contribution to the dictionary and is based on the information supplied by previous findings.
[243] More semantic frames may be discovered in the future.

	Semantic frame	Base
3.	PATH_SHAPE	*ascend*$_{2}$-v
		climb$_{2}$-v
		cross$_{2}$-v
		descend$_{2}$-v
		head$_{2}$-v
4.	ARRIVING	*arrive*$_{1}$-v
		enter$_{1}$-v
		reach$_{1}$-v
5.	MOTION_DIRECTIONAL	*ascend*$_{1}$-v
		descend$_{1}$-v
		fall$_{1}$-v
6.	SELF_MOTION_WITH_SAFETY_INSTRUMENT	*abseil*$_{1}$-v
		climb$_{1}$-v
		scale$_{1}$-v
7.	MOTION	*skydive*$_{1}$-v
		slide$_{1}$-v
8.	TRAVERSING	*cross*$_{1}$-v
		traverse$_{1}$-v
9.	CAUSE_TO_LAND	*land*$_{1}$-v
10.	DEPARTING	*exit*$_{1}$-v
11.	MOTION_FROM_SOURCE	*jump*$_{1}$-v
12.	RIDE_VEHICLE	*ride*$_{2}$-v

As we can observe, the semantic frames evoked by a higher number of bases are OPERATE_VEHICLE (seven types) and SELF_MOTION (seven types). Then, we find PATH_SHAPE (five types), followed by ARRIVING (three types), MOTION_DIRECTIONAL (three types), SELF_MOTION_WITH_SAFETY_INSTRUMENT (three types), MOTION (two types) and TRAVERSING (two types). Finally, the CAUSE_TO_LAND, DEPARTING, MOTION_FROM_SOURCE and RIDE_VEHICLE frames are evoked by one base. Figure 84 represents the frequency of the semantic frames in the ADVENCOR corpus:

Figure 84

Frequency of the Semantic Frames in the ADVENCOR Corpus



As Figure 84 reveals, the most recurrent semantic frames in the specialized corpus are OPERATE_VEHICLE (19%) and SELF_MOTION (19%). First, OPERATE_VEHICLE was evoked by seven verbal bases, namely: *drive*$_1$-v, *float*$_1$-v, *fly*$_1$-v, *glide*$_1$-v, *navigate*$_1$-v, *paddle*$_1$-v and *ride*$_1$-v, which produced 34 collocations in total, as some of the most productive verbs are included in this frame, that is, *fly*$_1$-v (10 collocations), *paddle*$_1$-v (eight collocations) and *ride*$_1$-v (five collocations). The bases *navigate*$_1$-v and *paddle*$_1$-v have very close meanings, since the former may be considered a hypernym of the latter, and they share the same argument structure too (this argument structure is also shared by *ride*$_1$-v), but we will see that they do not share any collocates. In the same vein, *float*$_1$-v, *fly*$_1$-v and *glide*$_1$-v have quite similar argument structures, in fact, *float*$_1$-v and *glide*$_1$-v might be regarded as hyponyms of the other, but they do not have any collocates in common.

This frame involves a vehicle and someone who controls it, that is, the person participating in the activity, who can be a TOURIST or a RESPONSIBLE;[244] accordingly, all the verbs describe actual motion. The vehicle may be a VEHICLE_WITH_ENGINE or a VEHICLE_WITHOUT_ENGINE, and the collocates of the verbs may describe the PLACE or the PATH in which the motion takes place or the DIRECTION followed by the vehicle, among others. Sometimes, the verbs evoking this frame unveil the part of nature where the

[244] The semantic roles (which can also be referred to as FEs; Atkins & Rundell, 2008, p. 145) identified in the motion verbs implemented in *DicoAdventure* were presented in Section 4.2. (cf. Table 17).

activity is undertaken, for example, land, air or water. The semantic roles expressed in the collocations belonging to the OPERATE_VEHICLE frame are summarized in Table 41:

Table 41

Semantic Roles Represented by the Collocates of the Verbs Evoking the OPERATE_VEHICLE Frame

	Semantic role	Collocation
1.	VEHICLE_WITHOUT_ENGINE	*drive (a) team*
		drive (a) sled
		fly (a) parachute
		fly (a) paraglider
		paddle (a) canoe
		paddle (a) kayak
		paddle (a) raft
		ride (a) bike
		ride (a) bicycle
		ride (a) zipline
		ride (a) raft
2.	PATH	*navigate rapids*
		navigate (a) section
		navigate water
		navigate (a) waterfall
		paddle (a) river
		ride (a) wave
3.	TOURIST/RESPONSIBLE	*(a) participant drive*
		(a) pilot fly
		(a) glider fly
		(an) instructor fly
		(a) kayaker paddle
4.	MANNER	*fly solo*
		fly tandem
		silently glide/glide silently
		paddle hard
		gently paddle
5.	DIRECTION	*float down*

	Semantic role	Collocation
		fly backwards
		paddle back
6.	PLACE	*fly (a) site*
7.	SPEED	*slowly float*
8.	TIME	*drive (an) hour*
9.	VEHICLE_WITH_ENGINE	*fly (a) paramotor*

As Table 41 shows, nine are the semantic roles of the OPERATE_VEHICLE frame described by the collocates. First of all, it must be remembered that arguments are distinguished from circumstantials (cf. §4.2.). The former are included in the argument structure of the verb, whereas the latter are optional. More specifically, Durán-Muñoz and L'Homme (2020, p. 45) define them in the following way: "arguments [...] are core components of the meanings of motion verbs in adventure tourism, and circumstantials [...] add peripheral information but are optional as far as characterizing the meaning is concerned." In this case, five are the arguments identified, namely, VEHICLE_WITHOUT_ENGINE, PATH, TOURIST/RESPONSIBLE, PLACE and VEHICLE_WITH_ENGINE.

First, 11 verb + noun collocations make a specific reference to the VEHICLE_WITHOUT_ENGINE that is controlled: *drive (a) team, drive (a) sled, fly (a) parachute, fly (a) paraglider, paddle (a) canoe, paddle (a) kayak, paddle (a) raft, ride (a) bike, ride (a) bicycle, ride (a) zipline* and *ride (a) raft*, and one of them specifies the VEHICLE_WITH_ENGINE that is controlled: *fly (a) paramotor*. In most of the cases, the verb unveils the part of nature in which the action takes place, such as *fly*$_1$-v and *paddle*$_1$-v, although the vehicles designated by the collocations also tell whether they are used on land (e.g., *sled, bike* and *bicycle*), in the air (e.g., *parachute, paramotor, paraglider* and *zipline*) or in the water (e.g., *canoe, kayak* and *raft*). It is important to realize that *raft* co-occurs with two different verbs: *paddle*$_1$-v and *ride*$_1$-v, which have a similar argument structure and whose features in the ADVENCOR corpus are alike: their *logDice* scores are 8.62 and 8.43, respectively, and both of them had two tokens in the corpus.

Second, verb + noun collocations can also include the PATH that is followed in the motion (six collocations): *navigate rapids, navigate (a) section, navigate water, navigate (a) waterfall, paddle (a) river* and *ride (a) wave*. In this case, all the collocates represent a PATH located in the water and 66.67% of the collocations contain *navigate*$_1$-v. One collocation contains *paddle*$_1$-v, although its collocate, *river*, does not co-occur with

navigate$_1$-v, despite being two closely semantically connected verbs. Last but not least, *ride*$_1$-v also collocates with a noun related to water, *wave*, although as aforementioned it also collocates with nouns related to land (e.g., *bike*). On the other hand, the PLACE where the motion occurs described in *fly (a) site* refers to a place located on land, as it is reflected in the contexts (127) *DO NOT attempt to **fly** this **site** for the first time without a local guide!* and (128) *We spend the following days **flying** the **sites** around Lake Hovsgol at the foot of eastern Sayan Mountain* in the Appendix.

Third, nominal collocates in the subject position designate the TOURIST or RESPONSIBLE related to the motion (five collocations): *(a) participant drive, (a) pilot fly, (a) glider fly, (an) instructor fly* and *(a) kayaker paddle*. More specifically, four of the collocates refer to the tourist participating in an adventure activity, that is, *participant, pilot, glider* and *kayaker*, whereas *instructor* clearly defines a person responsible for the activity.

Regarding the circumstantials found in the OPERATE_VEHICLE frame, they add information on different aspects of the verbs:

1) First, five collocations include the MANNER in which the motion is carried out: *fly solo, fly tandem, silently glide/glide silently, paddle hard* and *gently paddle*. The collocates *solo* and *tandem* allude to the company that a person undertaking the activity has; *silently* tells that no sound is made when gliding; finally, the activity of paddling can require considerable physical effort, that is, it is *hard*, or can be made carefully, with slow movements and no sudden changes, that is, *gently*;

2) Second, the DIRECTION taken in the motion event is represented by three collocations: *float down, fly backwards* and *paddle back*, and two directions are shown: toward a lower position: *down*, or toward the source of motion or a previous point: *backwards* and *back*;

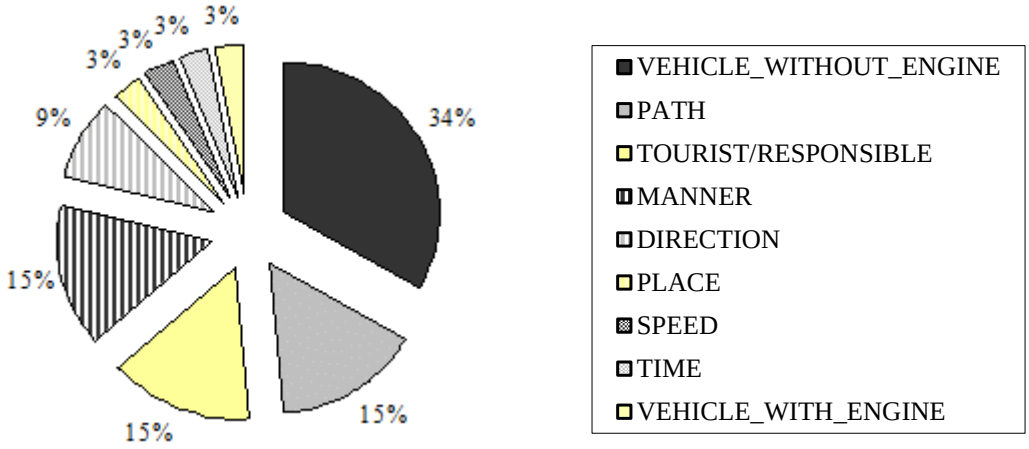
3) Third, the specific SPEED at which the displacement occurs is included in *slowly float*, which means that it does not happen fast; and

4) Fourth, the period of TIME during which the motion develops is described in *drive (an) hour*.

Figure 85 shows the frequency of the distinct semantic roles in the collocations belonging to the OPERATE_VEHICLE frame:

Figure 85

Frequency of the Semantic Roles Represented by the Collocates of the Verbs Evoking the OPERATE_VEHICLE Frame



As it can be seen, VEHICLE_WITHOUT_ENGINE is the most recurrent semantic role in the collocations (34%), since it forms the core of the meaning of the OPERATE_VEHICLE frame. Then, PATH, TOURIST/RESPONSIBLE and MANNER are represented by 15% of the collocates, and DIRECTION by 9% of them. Finally, PLACE, SPEED, TIME and VEHICLE_WITH_ENGINE are depicted by 3% of the total. Such a variety of semantic roles may be related to the lack of repetition of the collocates. Additionally, it may be explained by the fact that the verbs describe a variety of activities that are undertaken in different scenarios, such as on land (e.g., *drive*$_1$-v), in the air (e.g., *fly*$_1$-v) and in the water (e.g., *navigate*$_1$-v, *paddle*$_1$-v). On the other hand, the collocates of *ride*$_1$-v imply that this verb can be used in land-based (*ride (a) bike*, *ride (a) bicycle*), air-based (*ride (a) zipline*) and water-based (*ride (a) raft*) activities.

Second, the SELF_MOTION frame is also evoked by seven bases (19%), namely: *head*$_1$-v, *hike*$_1$-v, *scramble*$_1$-v, *swim*$_1$-v, *trek*$_1$-v, *venture*$_1$-v and *walk*$_1$-v, which produced 16 collocations in total. The bases *head*$_1$-v, *hike*$_1$-v, *trek*$_1$-v and *walk*$_1$-v are intimately related, as it will be shown. This frame entails a living being, who can be a TOURIST or a RESPONSIBLE, following a specific DIRECTION along a PATH. In addition, different items

of information on the motion may be given, such as the DISTANCE covered. All the verbs evoking this frame represent actual motion, since it "prototypically involves individuals moving under their own power by means of their bodies."[245] Table 42 represents the semantic roles identified in the collocates of these verbs:

Table 42

Semantic Roles Represented by the Collocates of the Verbs Evoking the SELF_MOTION Frame

	Semantic role	Collocation
1.	DIRECTION	*head north*
		head back
		head up
		head down
		hike down
		hike back
		scramble up
		trek down
		trek up
		walk down
2.	DISTANCE	*walk (one) mile*
3.	MANNER	*swim (one's) way*
4.	PATH	*hike (a) trail*
5.	RESPONSIBLE	*(a) guide walk*
6.	SOURCE	*venture out*
7.	TIME	*walk (one) hour*

As it can be observed, seven semantic roles are evoked by the collocates accompanying the verbal bases in the SELF_MOTION frame, four of which are arguments (i.e., DIRECTION, PATH, RESPONSIBLE, SOURCE) and three of which are circumstantials (i.e., DISTANCE, MANNER, TIME).

First, DIRECTION is expressed by ten verb + adverb collocations and it can be: (1) toward a lower position: *head down, hike down, trek down* and *walk down*; (2) toward a higher position: *head north, head up, scramble up* and *trek up*; and (3) toward the source

[245] https://framenet.icsi.berkeley.edu/fndrupal/frameIndex (Last accessed: 25/10/2020).

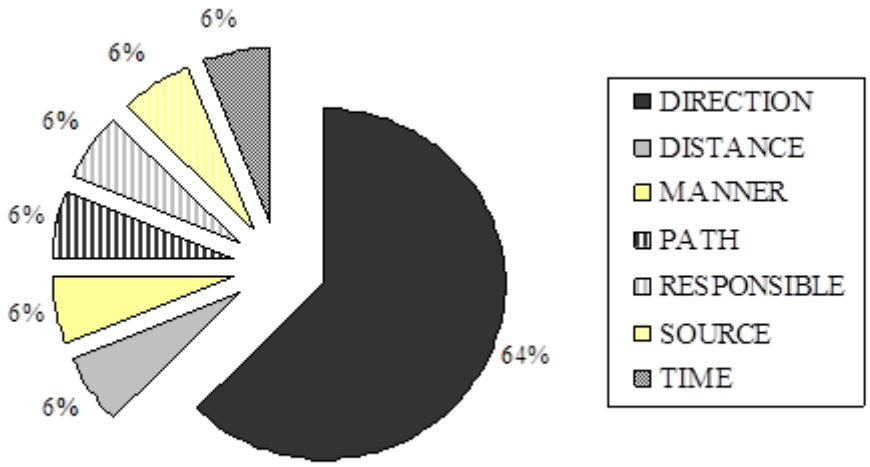
of motion or a previous point: *head back* and *hike back*. Therefore, the collocate *down* co-occurs with four different verbs, *head*$_1$-v, *hike*$_1$-v, *trek*$_1$-v and *walk*$_1$-v, being the strongest collocation *hike down* (9.7 points of *logDice* score), closely followed by *walk down* (9.37 points), *head down* (9.33 points) and a bit further by *trek down* (8.24 points). A key point is that these verbs share the same argument structure, in fact, *head*$_1$-v, *hike*$_1$-v and *trek*$_1$-v can be regarded as hyponyms of *walk*$_1$-v. Moreover, *hike*$_1$-v and *trek*$_1$-v are sometimes used interchangeably, in spite of the fact that they imply distinct durations and levels of difficulty of the activity; more specifically, *trek*$_1$-v involves a longer walk that requires a bigger effort. These divergences in meaning can explain the differences in the collocational preferences of these four verbs, for they only share the collocate *down* (albeit *up* is shared by *head*$_1$-v and *trek*$_1$-v).

Regarding the rest of the semantic roles, they are represented by different lexical collocations and by only one collocation each. For instance, the verb + noun collocation *walk (a) mile* suggests the DISTANCE covered in the displacement. Then, the nominal collocate in *swim (one's) way* alludes to the MANNER of the motion, which is further clarified in context (238) in the Appendix: *using whichever tactic best suits the next obstacle.* A verb + noun collocation is also used to reveal the PATH followed in the motion in *hike (a) trail*. Next, the nominal collocate in the subject position in *(a) guide walk* unveils the RESPONSIBLE for the activity. After that, in *venture out* the adverb implies a definite starting point of the motion, that is, the SOURCE. Finally, the nominal collocate in the object position in *walk (one) hour* expresses the period of TIME for which the action takes place.

Figure 86 illustrates the frequency of the semantic roles represented by the collocations evoking the SELF_MOTION frame:

Figure 86

Frequency of the Semantic Roles Represented by the Collocates of the Verbs Evoking the SELF_MOTION *Frame*



As we can see in Figure 86, DIRECTION occupies more than half (64%) of the representation of semantic roles in the SELF_MOTION frame. After that, all the other semantic roles, that is, DISTANCE, MANNER, PATH, RESPONSIBLE, SOURCE and TIME, are represented by 6% of the collocates. The high frequency of DIRECTION may be due to the fact that, normally, verbs evoking this frame involve individuals who are in control of their bodies while making the movement, therefore, it is not surprising that the phraseology of these verbs reveals the direction followed.

Third, the PATH_SHAPE frame is evoked by five bases (14%), namely: *ascend*$_2$-v, *climb*$_2$-v, *cross*$_2$-v, *descend*$_2$-v and *head*$_2$-v, which produced 28 collocations in total. These verbs have one feature in common: they describe fictive motion of a stationary PATH. It occurs through a PLACE/PATH and the collocates accompanying these verbs may describe the DIRECTION (in some cases, it could be inferred from the meaning of the verb, e.g., *ascend*$_2$-v), MANNER or SPEED of the motion, among others, as it is shown in Table 43:

Table 43

Semantic Roles Represented by the Collocates of the Verbs Evoking the PATH_SHAPE Frame

	Semantic role	Collocation
1.	PATH (subject position)	*(a) climb ascend*
		(a) trail ascend
		(a) route ascend
		(a) pitch climb
		(a) trail climb
		(a) bridge cross
		(a) road cross
		(a) trail cross
		(a) trek cross
		(a) route cross
		(a) trail descend
		(a) path head
		(a) trail head
2.	MANNER	*climb steeply*
		gradually climb
		gradually descend
		descend steeply
		head directly
3.	DIRECTION	*climb up*
		climb uphill
		head north
4.	PATH (object position)	*cross (a) bridge*
		cross (a) river
		cross (a) stream
5.	PLACE	*ascend (a) slope*
		climb (a) face
		descend (a) gully
6.	SPEED	*slowly descend/descend slowly*

As Table 43 illustrates, the collocations depicting fictive motion events represent six semantic roles of the PATH_SHAPE frame (the arguments are PATH –both in the subject and object positions–, PLACE and DIRECTION). The most recurrent one is PATH in the

subject position, which designates the stationary entity that seems to undergo displacement. In total, 13 noun + verb collocations belong in this group, namely: *(a) climb ascend, (a) trail ascend, (a) route ascend, (a) pitch climb, (a) trail climb, (a) bridge cross, (a) road cross, (a) trail cross, (a) trek cross, (a) route cross, (a) trail descend, (a) path head* and *(a) trail head*. Without a doubt, *trail* stands out for its occurrence with all the verbs evoking the frame, although it must be highlighted that the strongest associations are *(a) trail ascend* and *(a) trail descend* (9.48 points of *logDice* score each), that is, with the two verbs from which the DIRECTION of motion can be inferred. Then, *(a) trail head* is the third strongest collocation (8.8 points), followed by *(a) trail cross* (8.39 points) and, finally, *(a) trail climb* (7.76 points). Additionally, *route* collocates with *ascend*$_2$-v and *cross*$_2$-v, and also at this point it must be remembered that the searches carried out through the Concordance function of Sketch Engine in Section 5.4.2. and Section 5.4.4. allowed us to know that *(a) route climb* and *(a) route descend* were also collocations worthy of inclusion in *DicoAdventure* (they achieved 7.03 and 7.31 points and occurred in three and two different contexts, respectively). By the same token, the collocation *(a) path descend* also satisfied the eligibility criteria (7.21 points, two tokens) (cf. §5.4.4.), therefore, it was also selected for its implementation in the dictionary. This way, *path* collocates with *descend*$_2$-v and *head*$_2$-v, but no other nominal collocates co-occur with more than one verb in the subject position. It may be surprising, given that some nouns are semantically related, such as *trail*, *trek* and *path*. In fact, all the nominal collocates, except for *bridge*, designate a type of track made on a surface (*bridge* also names a kind of track but is supported by a structure).

Second, MANNER is represented by five adverb + verb/verb + adverb collocations: *climb steeply, gradually climb, gradually descend, descend steeply* and *head directly*, and it describes a movement made with no difficulties: *steeply*, in small stages and at a low speed: *gradually*, or with no stops: *directly*. From the collocations containing the collocate *steeply*, the latter, *descend steeply*, is a bit stronger than the former, *climb steeply* (10.82 vs. 10.16 points). Similarly, the collocation including *gradually* is stronger with the verb *descend* than with the verb *climb* (11.58 vs. 9.64 points).

Thirdly, DIRECTION is depicted in *climb up, climb uphill* and *head north*, all of which emphasize the upward direction of motion. However, there is a slight difference. While the combinations of *climb*$_2$-v may be reasonably inferred given the intrinsic features of the verb (*climb* usually designates upward motion, albeit we showed in Section

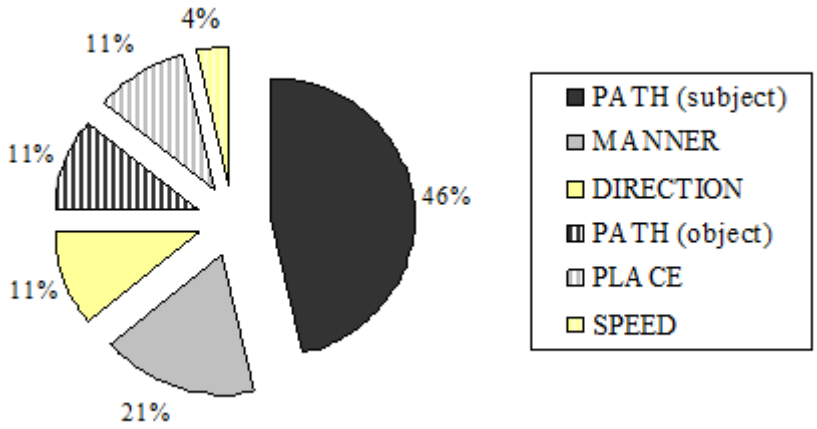
5.4.2. that downward direction can also be specified), *head north* implies upward direction by means of the cardinal point *north*. In fact, Section 5.4.5. unveiled a high co-occurrence of *head*-v with cardinal points representing actual and fictive motion events.

Then, the semantic roles PATH (in the object position) and PLACE are represented by three collocations each. On the one hand, PATH (in the object position) is identified in *cross (a) bridge*, *cross (a) river* and *cross (a) stream* and the collocates can be classified into two groups: (1) natural structures located in the water: *river* and *stream*, and (2) built structures: *bridge*. On the other hand, PLACE is present in the collocations *ascend (a) slope*, *climb (a) face* and *descend (a) gully*. In this case, all the collocates describe a natural structure located on land. Finally, the SPEED of the motion is mentioned in *slowly descend/descend slowly*, which indicates that the descent is not fast.

Figure 87 illustrates the frequency of the semantic roles found in the PATH_SHAPE frame:

Figure 87

Frequency of the Semantic Roles Represented by the Collocates of the Verbs Evoking the PATH_SHAPE Frame



As Figure 87 depicts, PATH represented as the inanimate subject performing the motion occupies almost half of the collocates (46%). It is interesting because in the frames previously analyzed, OPERATE_VEHICLE and SELF_MOTION, both of them representing actual motion, collocations containing the doer of the action were infrequent. One sound reason may be that verbs expressing actual motion usually go in the imperative form (e.g., example (232): *Embrace the fear and tandem skydive*) or use the subject *you* (e.g.,

example (251): *You will walk 5 miles to the bridge*). For this reason, we can state that noun + verb collocations are more common in fictive motion events. Secondly, MANNER is identified in 21% of the collocations. Then, DIRECTION, PATH (in the object position) and PLACE obtain the same representation in the ADVENCOR corpus: 11%, and, finally, the SPEED of the motion is mentioned in just 4%.

Next, ARRIVING (8%) is ranked fourth and is evoked by three bases, namely: *arrive*$_1$-v, *enter*$_1$-v and *reach*$_1$-v, which produced 16 collocations in total. More specifically, this high number of collocations is due to the substantial level of collocation production of *reach*$_1$-v (14 types). The three verbs share the same argument structure, in which a TOURIST or a RESPONSIBLE and a DESTINATION are involved, therefore, they depict actual motion events. The collocates can express other semantic roles, as it is illustrated in Table 44:

Table 44

Semantic Roles Represented by the Collocates of the Verbs Evoking the ARRIVING Frame

	Semantic role	Collocation
1.	DESTINATION	*enter (a) world*
		reach (the) summit
		reach (the) top
		reach (a) destination
		reach (a) point
		reach (the) bottom
		reach (a) bridge
		reach (a) spot
		reach (a) canyon
		reach (a) peak
		reach (the) base
		reach (the) base camp
		reach (a) mountain
2.	DISTANCE	*reach (an) altitude*
3.	TIME	*arrive early*
4.	TOURIST	*(a) participant reach*

As Table 44 shows, the number of semantic roles evoked by the collocations containing *arrive*$_1$-v, *enter*$_1$-v and *reach*$_1$-v are four. It goes without saying that a higher percentage of collocations refers to one of the arguments of the verbs, that is, DESTINATION. It is described by 13 verb + noun collocations, in which the nominal collocate specifies the type of destination, which can be:

1) The highest point of something: *reach (the) summit, reach (the) top* and *reach (a) peak*;
2) The lowest part of something: *reach (the) bottom* and *reach (the) base*;
3) A built structure: *reach (a) bridge* and *reach (the) base camp*;
4) A natural structure: *reach (a) canyon* and *reach (a) mountain*; and
5) An unspecified destination whose details can only be known by checking the contexts: *enter (a) world* (e.g., *the world of ice*, example (116) in the Appendix), *reach (a) destination* (e.g., *reaching the bungee destination*, example (185)), *reach (a) point* (e.g., *reach a point where trekking high in the trees is a truly memorable experience*, example (188)) and *reach (a) spot* (e.g., *reached a dive spot*, example (197)).

Regarding their collocation strength in terms of *logDice* score, the strongest associations of words belong in the first group (9.99 points on an average), which may be due to the fact that reaching a high point of a natural structure, for example, the peak of a mountain, is the main objective in a wide range of adventure activities, providing individuals with greater satisfaction than reaching other points. It is followed by the words designating an unspecified destination (9.11 points) and then collocations in group two are stronger than those in group three (8.52 vs. 8.34 points). Finally, group four, that is, those collocations containing a verb and a natural structure, represents the weakest collocations (7.92 points).

On the other hand, the second argument is TOURIST, mentioned in the noun + verb collocation *(a) participant reach*. In this case, the nominal collocate does not specify the type of participant in the adventure activity, as we could see in previous examples of collocations, such as *(a) glider fly* or *(a) kayaker paddle*.

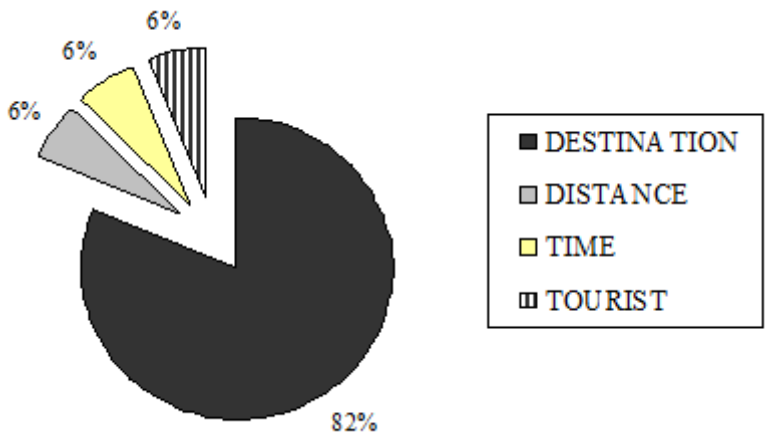
Finally, the collocations *reach (an) altitude* and *arrive early* allude to the circumstantials DISTANCE and TIME. They are collocations of different bases and different

lexical types. Thus, the former is a verb + noun collocation where the object informs of the amount of space covered by the displacement. More specifically, to know this detail, the context needs to be checked, such as *reach a sufficient altitude* (example (193)) and *reach altitudes of several thousand feet* (example (194)). The latter is a verb + adverb collocation where the modifier implies that the movement should take place before something else happens (i.e., *early*).

The frequency of the semantic roles revealed by the collocates of the verbs evoking the ARRIVING frame is shown in Figure 88:

Figure 88

Frequency of the Semantic Roles Represented by the Collocates of the Verbs Evoking the ARRIVING Frame



As it has been seen, the collocations mentioning the DESTINATION of the motion massively outnumbers the collocations alluding to DISTANCE, TIME and TOURIST (82% vs. 6%). It may be due to the fact that DESTINATION forms the core of the meaning of the ARRIVING frame.

Next, the MOTION_DIRECTIONAL frame is ranked fifth and is evoked by three bases (8%), namely: *ascend*$_1$-v, *descend*$_1$-v and *fall*$_1$-v, which produced six collocations in total. In this frame, an individual moves in a certain DIRECTION, which is often determined by gravity or other natural forces and inherently expressed by the verbs. Thus, *ascend*$_1$-v implies an upward direction and *descend*$_1$-v and *fall*$_1$-v imply a downward direction. The individual moving is a living entity, such as a TOURIST or a RESPONSIBLE, so the type of

event described entails actual motion. The semantic roles expressed by the collocates of *ascend*$_1$-v, *descend*$_1$-v and *fall*$_1$-v are shown in Table 45:

Table 45

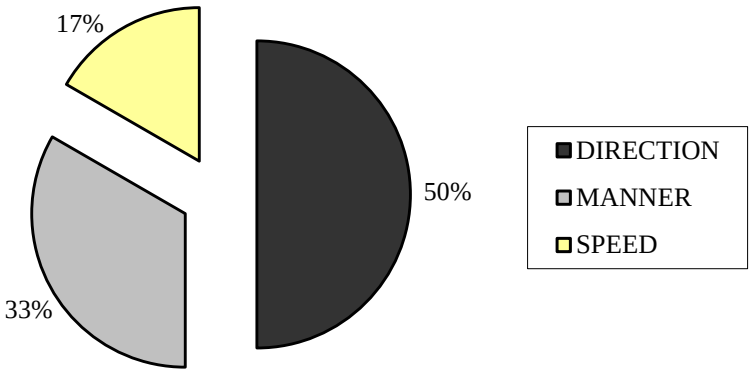
Semantic Roles Represented by the Collocates of the Verbs Evoking the MOTION_DIRECTIONAL *Frame*

	Semantic role	Collocation
1.	DIRECTION	*descend down*
		descend back
		fall back
2.	MANNER	*gradually descend*
		gently descend/descend gently
3.	SPEED	*ascend slowly*

As it can be observed, three are the semantic roles expressed by the adverb + verb/verb + adverb collocations evoking the MOTION_DIRECTIONAL frame. First, despite the fact that the DIRECTION of the motion is implied in the three bases, this is the most recurrent semantic role (in fact, this is an argument of the verb). On the one hand, the adverbial collocate in *descend down* emphasizes the downward direction suggested by the verb. On the other hand, the collocate *back* occurs in *descend back* and *fall back* and refers to a displacement toward the source of motion or a previous point, which is performed in a downward direction (inferred from both verbs). Both collocations are equally strong (7.37 vs. 7.35 points). Second, two collocations describe the MANNER of the descent: *gradually descend* (in small stages and slowly) and *gently descend/descend gently* (with slow movements and no sudden changes). Third, the SPEED at which the movement is made is mentioned in *ascend slowly* (i.e., not fast). The frequency of the distinct semantic roles in this frame is shown in Figure 89:

Figure 89

Frequency of the Semantic Roles Represented by the Collocates of the Verbs Evoking the MOTION_DIRECTIONAL *Frame*



As we can see, despite knowing that the very bases evoking the MOTION_DIRECTIONAL frame specify the DIRECTION of motion, the collocations evoking this argument (50%) outnumber the collocations describing the MANNER (33%) and the SPEED (17%) of the displacement.

The following frame in terms of frequency in the ADVENCOR corpus is SELF_MOTION_WITH_SAFETY_INSTRUMENT (8%), also evoked by three bases, namely: *abseil*$_1$-v, *climb*$_1$-v and *scale*$_1$-v, which produced a total of 14 collocations. The argument structures of these verbs are very similar, albeit an essential difference can be stressed. While *climb*$_1$-v describes motion which can be in an upward or downward direction, it is always downward in *abseil*$_1$-v and upward in *scale*$_1$-v. An interesting point is that this frame is reminiscent of the SELF_MOTION frame, in which a living entity, that is, a TOURIST or a RESPONSIBLE, follows a specific DIRECTION through a PATH. Nevertheless, in this case a SAFETY_INSTRUMENT is required so as to guarantee the individual's safety. Table 46 illustrates the semantic roles identified in the collocations belonging to the SELF_MOTION_WITH_SAFETY_INSTRUMENT frame:

Table 46

Semantic Roles Represented by the Collocates of the Verbs Evoking the SELF_MOTION_WITH_SAFETY_INSTRUMENT *Frame*

	Semantic role	Collocation
1.	PATH	*climb (a) wall*
		climb (a) route
		climb (a) mountain
		climb (a) face
		climb (a) waterfall
		climb (a) cliff
		climb (a) tree
		scale height
		scale (a) wall
2.	DIRECTION	*climb up*
		climb down
3.	PLACE	*climb (a) peak*
		scale (a) peak
4.	SPEED	*slowly abseil/abseil slowly*

As it can be observed, the semantic roles evoked by the collocations of the verbs in the SELF_MOTION_WITH_SAFETY_INSTRUMENT frame are four; two of them are arguments (i.e., PATH and DIRECTION) and two are circumstantials (i.e., PLACE and SPEED). First, PATH is much more recurrent than the rest, as it is described in nine verb + noun collocations, which can be classified into three groups according to the reference of the collocate:

1) A natural structure found on land: *climb (a) wall, climb (a) route, climb (a) mountain, climb (a) face, climb (a) cliff, climb (a) tree* and *scale (a) wall*;

2) A natural structure found in the water: *climb (a) waterfall*; and

3) An unspecified path: *scale height*.

On an average, stronger collocations are found in group three (11.54 points), followed by those in group one (8.75 points) and, finally, those in group two (7.57 points). Two significant aspects that must be mentioned are the following. First, *climb (a) route* does not describe a concrete natural structure like in the rest of the collocations in the first group, but *route* designates the PATH followed when climbing a mountain, so it activates

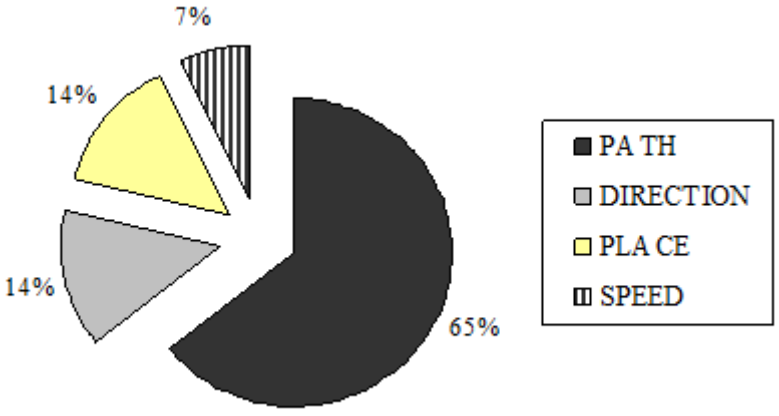
the PART FOR WHOLE metonymy. Second, *scale height* describes a non-tangible PATH, given that the contexts allow us to know that it is located in the air: *scaling new heights over the skies* (example (223)) and *scale new heights in the sky* (example (224)).

After PATH, the DIRECTION of motion is described in two verb + adverb collocations: *climb up* and *climb down*, therefore, it can be upward or downward. PLACE is also evoked by two verb + noun collocations: *climb (a) peak* and *scale (a) peak*. More specifically, the shared nominal collocate alludes to the top of a place, so the displacement is carried upward. In other words, as it was mentioned in Section 5.4.2., the PART FOR WHOLE metonymy is activated, since it highlights the highest point of a mountain, so the motion refers to the whole process of climbing it. Last but not least, the SPEED at which the movement occurs is represented in the only collocation produced by *abseil*$_1$-v: *slowly abseil/abseil slowly*, it being in a way that is not fast.

As abovementioned, PATH is the most frequent semantic role of all the ones described in the collocations evoking the SELF_MOTION_WITH_SAFETY_INSTRUMENT frame, as Figure 90 illustrates:

Figure 90

Frequency of the Semantic Roles Represented by the Collocates of the Verbs Evoking the SELF_MOTION_WITH_SAFETY_INSTRUMENT Frame



As we can see, PATH is represented by more than half (65%) of the collocations evoking the SELF_MOTION_WITH_SAFETY_INSTRUMENT frame. Then, DIRECTION occupies 14% of the representation as well as PLACE, and SPEED is the least recurrent one (7%). Albeit we said that this frame could remind us of the SELF_MOTION frame, it must be emphasized

that the results for each of them are not similar. Thus, PATH is the most frequent semantic role in the former (65%), whereas it is DIRECTION in the latter (64%). Then, DIRECTION is evoked by 14% of the collocates in the former and PATH is evoked by 6% of the collocates in the latter.

Next, the frame ranked seventh is MOTION, which is evoked by two verbs (6%), namely: *skydive*$_1$-v and *slide*$_1$-v, which produced a total of three collocations. This frame describes actual motion events, for it entails a living being, that is, a TOURIST or a RESPONSIBLE, moving from one point (SOURCE) to another (DESTINATION) through a PATH. Another semantic role which can appear in the frame is MANNER, as it is displayed in Table 47:

Table 47

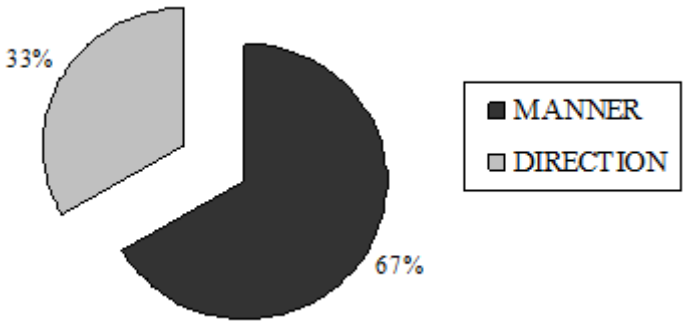
Semantic Roles Represented by the Collocates of the Verbs Evoking the MOTION Frame

	Semantic role	Collocation
1.	MANNER	*tandem skydive*
		slide (one's) way
2.	DIRECTION	*slide down*

Table 47 shows the two semantic roles evoked by the collocates co-occurring with *skydive*$_1$-v and *slide*$_1$-v. First, MANNER, which is a circumstantial, is depicted by *tandem skydive* and *slide (one's) way*. The former is an adverb + verb collocation and alludes to the company of a person who is practicing the adventure activity. More specifically, it means that there are two people skydiving, normally a RESPONSIBLE and a TOURIST (the first one taking care of the second one). The latter refers to the personal style in which the activity is undertaken (contexts (235) and (236) add no more details about that). Second, the DIRECTION (argument) of motion is unveiled by the verb + adverb collocation *slide down*, that is, downward. Figure 91 illustrates the frequency of these two semantic roles in the MOTION frame:

Figure 91

Frequency of the Semantic Roles Represented by the Collocates of the Verbs Evoking the MOTION *Frame*



As it can be observed, the representation of MANNER in the MOTION frame in the ADVENCOR corpus doubles that of DIRECTION (67% vs. 33%), in spite of the fact that the latter is more central to the meaning of the frame.

Next, the TRAVERSING frame is also evoked by two verbal bases (6%), namely: *cross*$_1$-v and *traverse*$_1$-v, which produced six collocations in total. Both verbs have the same argument structure and can be considered synonyms. The frame involves a human being (a TOURIST or a RESPONSIBLE) who moves along a PATH which goes from a SOURCE to a DESTINATION. All the contexts represent actual motion events and PATH, one of the arguments of the verbs, seems to be the most important semantic role of the frame, as it is shown in Table 48:

Table 48

Semantic Roles Represented by the Collocates of the Verbs Evoking the TRAVERSING *Frame*

	Semantic role	Collocation
1.	PATH	*cross (a) bridge*
		cross (a) river
		cross (a) stream
		cross (a) meadow
		traverse (a) canyon
		traverse (a) river

As we can see, 100% of the collocations contain the verb plus a nominal collocate in the object position and describe the PATH followed by *cross*$_{1}$-v and *traverse*$_{1}$-v. The collocates can be classified according to the type of path represented: (1) a natural structure in the water: *cross (a) river, cross (a) stream* and *traverse (a) river*; (2) a natural structure on land: *cross (a) meadow* and *traverse (a) canyon*; and (3) a built structure: *cross (a) bridge*. *River* is the only collocate which occurs with both bases, albeit the association with *cross*$_{1}$-v is much stronger than the association with *traverse*$_{1}$-v (11 vs. 8.24 points). On an average, the collocation in group three is stronger than those in group one and in group two (11.13 vs. 9.93 and 8.88, respectively).

The following frame, CAUSE_TO_LAND, is ranked ninth and is evoked by *land*$_{1}$-v (3%), which produced the collocation *land safely*. This frame represents actual motion events, given that an individual, either a TOURIST or a RESPONSIBLE, controls a VEHICLE_WITH_ENGINE or a VEHICLE_WITHOUT_ENGINE, which usually does not operate on land (e.g., a paramotor, a glider), and maneuvers it into land. The adverb modifying the verb in the only collocation found in this frame describes the MANNER (circumstantial) in which the motion is performed, that is, in a way that is not likely to cause damage, injury or harm.

Next, DEPARTING is also evoked by one verb: *exit*$_{1}$-v (3%), which produced one collocation: *exit (an) aircraft*. The meaning of this frame has two main semantic roles, a TOURIST or a RESPONSIBLE who carries out the motion and the SOURCE from which he/she moves away. Accordingly, the nominal collocate in *exit (an) aircraft* specifies the starting point of the movement (i.e., the SOURCE, which is an argument of the verb). Despite being only one collocation, it must be remembered that it is the strongest collocation retrieved from the ADVENCOR corpus (12.02 points) and it has a high number of tokens (nine).

The penultimate frame in terms of representation in the specialized corpus is MOTION_FROM_SOURCE, evoked by *jump*$_{1}$-v (3%), which produced the collocation *jump solo*. This frame is characterized because a living being, that is, a TOURIST or a RESPONSIBLE, leaves a SOURCE in order to reach a DESTINATION, therefore, the events represented describe actual motion. In this case, the adverb expresses the MANNER (circumstantial) in which the activity is performed, that is, with no company.

Last but not least, the RIDE_VEHICLE frame is evoked by the verbal base *ride*$_{2}$-v (3%), which produced two collocations. In this frame, a TOURIST or a RESPONSIBLE is

located in a VEHICLE_WITH_ENGINE or a VEHICLE_WITHOUT_ENGINE and is moved while he/she has no power to intervene. The MANNER in which the motion is developed can be specified, as it is shown in Table 49:

Table 49

Semantic Roles Represented by the Collocates of the Verbs Evoking the RIDE_VEHICLE Frame

	Semantic role	Collocation
1.	MANNER	*ride tandem*
2.	VEHICLE_WITH_ENGINE	*ride (a) lift*

As it can be observed, two semantic roles are evoked by the collocates of *ride*$_2$-v. First, the circumstantial MANNER of motion refers to the company of a responsible when undertaking the activity: *ride tandem*. Second, *ride (a) lift* specifies the argument VEHICLE_WITH_ENGINE, in which the motion takes place. Regarding their strength, the former is a bit stronger than the latter (9.34 vs. 9.11 points). Their representation in the specialized corpus is the same, albeit VEHICLE_WITH_ENGINE is more central to the meaning of the frame.

The analysis in terms of Frame Semantics previously performed reveals the motivation of collocates generated by the meanings of the verbs evoking each frame, for they represent the semantic roles of the elements in the frame. Table 50 lists the 121 collocations produced by the motion verbs selected in this work according to the semantic role represented:[246]

Table 50

Semantic Roles Represented by the Collocates of the Verbs in the ADVENCOR Corpus

	Semantic role	Collocation
1.	PATH (object position)	*climb (a) wall* *climb (a) route* *climb (a) mountain* *climb (a) waterfall*

[246] First, they are alphabetically arranged in terms of the base, and second, when the verb is the same, they are organized according to the *logDice* score achieved in the specialized corpus (in descending order).

	Semantic role	Collocation
		climb (a) cliff
		climb (a) tree
		cross (a) bridge
		cross (a) river
		cross (a) stream
		cross (a) meadow
		hike (a) trail
		navigate rapids
		navigate (a) section
		navigate water
		navigate (a) waterfall
		paddle (a) river
		ride (a) wave
		scale height
		scale (a) wall
		traverse (a) canyon
		traverse (a) river
2.	DIRECTION	*climb up*
		climb uphill
		climb down
		descend down
		descend back
		fall back
		float down
		fly backwards
		head north
		head back
		head up
		head down
		hike down
		hike back
		paddle back
		scramble up
		slide down
		trek down
		trek up

	Semantic role	Collocation
		walk down
3.	MANNER	*climb steeply*
		gradually climb
		descend steeply
		gradually descend
		gently descend/descend gently
		fly solo
		fly tandem
		silently glide/glide silently
		head directly
		jump solo
		land safely
		paddle hard
		gently paddle
		ride tandem
		tandem skydive
		slide (one's) way
		swim (one's) way
4.	DESTINATION	*enter (a) world*
		reach (the) summit
		reach (the) top
		reach (a) destination
		reach (a) point
		reach (the) bottom
		reach (a) bridge
		reach (a) spot
		reach (a) canyon
		reach (a) peak
		reach (the) base
		reach (the) base camp
		reach (a) mountain
5.	PATH (subject position)	*(a) climb ascend*
		(a) trail ascend
		(a) route ascend
		(a) pitch climb
		(a) trail climb

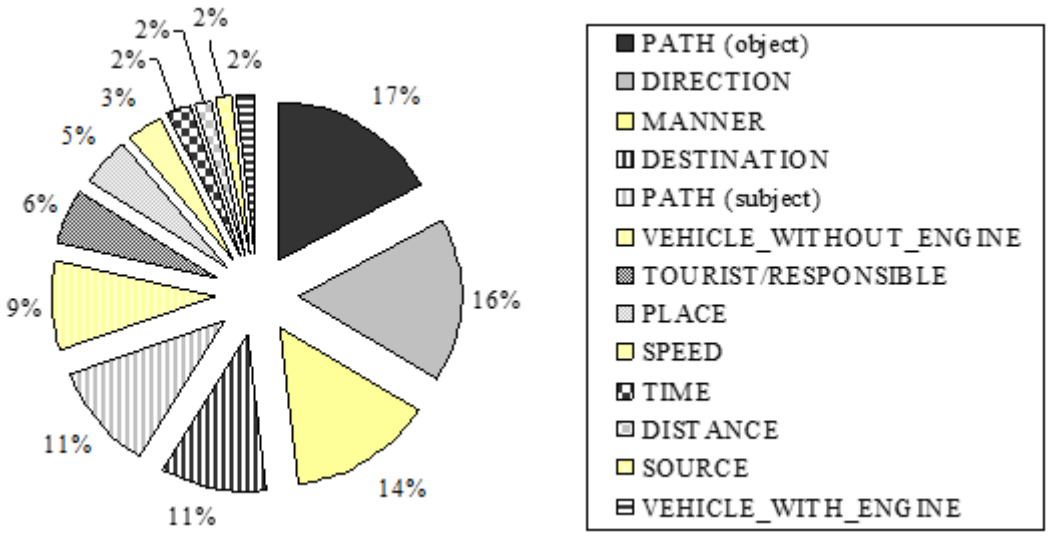
	Semantic role	Collocation
		(a) bridge cross
		(a) road cross
		(a) trail cross
		(a) trek cross
		(a) route cross
		(a) trail descend
		(a) path head
		(a) trail head
6.	VEHICLE_WITHOUT_ENGINE	*drive (a) team*
		drive (a) sled
		fly (a) parachute
		fly (a) paraglider
		paddle (a) canoe
		paddle (a) kayak
		paddle (a) raft
		ride (a) bike
		ride (a) bicycle
		ride (a) zipline
		ride (a) raft
7.	TOURIST/RESPONSIBLE	*(a) participant drive*
		(a) pilot fly
		(a) glider fly
		(an) instructor fly
		(a) kayaker paddle
		(a) participant reach
		(a) guide walk
8.	PLACE	*ascend (a) slope*
		climb (a) peak
		climb (a) face
		descend (a) gully
		fly (a) site
		scale (a) peak
9.	SPEED	*slowly abseil/abseil slowly*
		ascend slowly
		slowly descend/descend slowly
		slowly float

	Semantic role	Collocation
10.	TIME	*arrive early*
		drive (an) hour
		walk (one) hour
11.	DISTANCE	*reach (an) altitude*
		walk (one) mile
12.	SOURCE	*exit (an) aircraft*
		venture out
13.	VEHICLE_WITH_ENGINE	*fly (a) paramotor*
		ride (a) lift

As it can be observed, 13 semantic roles were identified in the collocations extracted from the specialized corpus. PATH is the most recurrent one and refers to the path along which the adventure activity is developed (21 collocations; in this case, the verbs are accompanied by a nominal collocate in the object position) and to the stationary path that apparently is performing the movement (13 collocations; here, the nominal collocate occurs in the subject position and belongs to word combinations depicting fictive motion events). Second, the DIRECTION of the action is displayed in 20 verb + adverb collocations and is closely followed by the MANNER in which the motion occurs (17 collocations), in which verbs are also modified by an adverb, either occurring before or after the base, or by a noun occurring after the base. After that, the end point of the adventure activity, that is, the DESTINATION, is shown in 13 verb + noun collocations, immediately followed by the VEHICLE_WITHOUT_ENGINE employed in a specific activity (11 collocations), specified in nominal collocates in the object position. The rest of the semantic roles are illustrated by less than 10 collocations each, namely: TOURIST/RESPONSIBLE (seven noun + verb collocations), which describes the subject in actual motion events; the PLACE where the activity takes place (six verb + noun collocations); the SPEED at which the action occurs (four adverb + verb/verb + adverb collocations); the period of TIME during which the action is developed (three collocations of two different types: verb + adverb and verb + noun); the DISTANCE covered in the motion (two verb + noun collocations); the starting point of the adventure activity, that is, the SOURCE (two collocations, which are also of two types, verb + noun and verb + adverb); and, finally, the VEHICLE_WITH_ENGINE used (two verb + noun collocations). Figure 92 summarizes the frequency of the semantic roles in the specialized corpus:

Figure 92

Frequency of the Semantic Roles in the ADVENCOR Corpus



Turning to the repetition of the collocates, it must be highlighted in the following semantic roles:

1) In DIRECTION, there are several groups of verbs that co-occur with the same collocate, such as *climb/descend/float/head/hike/slide/trek/walk* + *down*, *descend/fall/head/hike/paddle* + *back*, *climb/head/scramble/trek* + *up* (as aforementioned, some combinations with the same collocate are explained because the verbs are semantically related, e.g., *head*$_1$-v, *hike*$_1$-v, *trek*$_1$-v and *walk*$_1$-v);

2) In SPEED, all the displacements are undergone in a way that is not fast, reflected in the collocate *slowly*: *slowly abseil/abseil slowly, ascend slowly, slowly descend/descend slowly* and *slowly float*;

3) When the PATH is in the subject position, that is, it represents the inanimate entity performing fictive motion, *trail* is the most common collocate, as it co-occurs with all the verbs conveying this meaning, namely: *ascend*$_2$-v, *climb*$_2$-v, *cross*$_2$-v, *descend*$_2$-v and *head*$_2$-v;

4) When the PATH is in the object position, *river* is the one collocating with a higher number of verbs, namely: *cross*$_1$-v, *paddle*$_1$-v and *traverse*$_1$-v (it must

be remembered that the first one and the last one share the same argument structure and can be considered synonyms); and

5) In MANNER, the most recurrent collocate is *tandem*, which co-occurs with *fly*$_1$-v, *ride*$_2$-v and *skydive*$_1$-v.

The rest of the collocates occurred ≤ 2 times in the ADVENCOR corpus. With respect to whether they tend to represent arguments or circumstantials, there is a general tendency toward the representation of arguments (i.e., core components of the meanings of motion verbs), albeit some circumstantials are also highly typical of the SL of adventure tourism. For instance, MANNER is described in 66.67% of the frames (i.e., OPERATE_VEHICLE, SELF_MOTION, PATH_SHAPE, MOTION_DIRECTIONAL, MOTION, CAUSE_TO_LAND, MOTION_FROM_SOURCE and RIDE_VEHICLE) and SPEED is represented in 33.33% of the frames (i.e., OPERATE_VEHICLE, PATH_SHAPE, MOTION_DIRECTIONAL and SELF_MOTION_WITH_SAFETY_INSTRUMENT). Finally, TIME occurs in 25% of the frames (i.e., OPERATE_VEHICLE, SELF_MOTION and ARRIVING). These results, which help to better understand how motion verbs behave in context and to confirm their argument structures (cf. Durán-Muñoz & L'Homme, 2020), are summarized in Table 51:

Table 51

Arguments and Circumstantials Represented by the Collocations in the Semantic Frames Evoked

	Semantic frame	Arguments	Circumstantials
1.	OPERATE_VEHICLE	VEHICLE_WITHOUT_ENGINE	MANNER
		PLACE	DIRECTION
		PATH	SPEED
		TOURIST/RESPONSIBLE	TIME
		VEHICLE_WITH_ENGINE	
2.	SELF_MOTION	DIRECTION	DISTANCE
		PATH	MANNER
		RESPONSIBLE	TIME
		SOURCE	
3.	PATH_SHAPE	PATH (subject)	MANNER
		PLACE	SPEED
		PATH (object)	

	Semantic frame	Arguments	Circumstantials
		DIRECTION	
4.	ARRIVING	DESTINATION	DISTANCE
		TOURIST	TIME
5.	MOTION_DIRECTIONAL	DIRECTION	MANNER
			SPEED
6.	SELF_MOTION_WITH_SAFETY_INSTRUMENT	PATH	PLACE
		DIRECTION	SPEED
7.	MOTION	DIRECTION	MANNER
8.	TRAVERSING	PATH	
9.	CAUSE_TO_LAND		MANNER
10.	DEPARTING	SOURCE	
11.	MOTION_FROM_SOURCE		MANNER
12.	RIDE_VEHICLE	VEHICLE_WITH_ENGINE	MANNER

Moving on to the strength of the collocates retrieved, it could be said that, roughly speaking, all of them are strong, for they collocate with a reduced number of motion verbs. Nevertheless, *down* and *back*, showing the DIRECTION of motion, and *trail*, specifying the PATH which seems to perform the movement, may be regarded as weak collocates, as they can combine with a wider range of verbs. Be that as it may, the stringent criteria used for the selection of collocations (i.e., (1) a minimum *logDice* score of ≥ 7 points, (2) a minimum recurrence of two times in the specialized corpus, (3) description of the same meaning, and (4) an explicit reference to the field of adventure tourism) make them worthy of implementation in *DicoAdventure*, which is the primary objective established in this dissertation.

Finally, regarding the span that separates the verbal bases from their collocates, it must be highlighted that, in general, this is shorter in adverb + verb/verb + adverb collocations (i.e., -1/+1), for example, *as you **slowly float** back to earth* (example (121)), *you must **paddle hard** for most of the trip* (example (166)). Then, when a nominal collocate specifies the subject of the motion, in most of the cases it occurs right next to the base (i.e., -1), such as ***Participants drive** the sleds* (example (111)), *The **trails ascend** step by step till Ulleri* (example (8)); however, they can also co-occur in wider spans, for example, *a sea **kayaker** who **paddles** in busy waters* (example (171)), *although **participants** never **reach** the same heights* (example (196)), and even in different clauses, such as *Then the **trail** continues down a hot steep path, **crosses** the suspension bridge*

(example (63)), *From this creek bed the **trail** turns sharply to the west and **heads** directly* (example (102)). Last but not least, spans are much more varied in verb + noun collocations, since they can occur (almost) immediately in the text, such as *and **ride** the **waves** into the shore* (example (214)), *while **scaling peaks** throughout New Zealand's Southern Alps* (example (228)), or in broader spans separated by words which add nuances of meaning to the collocate, for example, *we **enter** a different **world*** (example (115)), ***navigate** the pristine swirling **waters*** (example (162)), or specify the reference of the collocate, for instance, ***Paddle** the majestic Colorado **River*** (example (179)), *we **reached** the route's pinch **point*** (example (187)). To conclude, no more inescapable conclusions can be drawn to this respect, as spans entirely depend on contexts, which can enormously vary from one text to another.

CHAPTER 7. CONCLUSIONS

The primary objective set in this doctoral dissertation was the implementation of collocations in the entries for motion verbs in the specialized concept-based dictionary *DicoAdventure*. In order to attain it, six specific objectives were defined; three of them were attained in Part I and three of them were attained in Part II.

The first aim was to explore the theories which influence specialized lexicography, the discipline which deals with the creation of terminological resources with the purpose of facilitating communication in specialized situations. It was achieved in Chapter 1, which first explained the view of SLs as functional subsystems belonging in the GL, in which they are located at the same level. It means that the SL of adventure tourism must be regarded as an SL which occupies a position as important as the positions of other SLs, such as the environment, law or science.

After that, Lexicology and Traditional Terminology were compared, and it led to the investigation of four modern theories of Terminology that endeavored to overcome the drawbacks found in the Traditional Terminology. For instance, Socioterminology highlighted the role of SLs in specialized communication, thus, the contexts where they appeared should be considered. Similarly, the CTT also emphasized the role of terminology in the spread of specialized knowledge and worked on the assumption that it allows the creation of products which help to such task. In the third place, the STT roundly criticized the concept of standardization postulated by the GTT, given that it constricted language and was incompatible with language evolution. Finally, the FBT shared several premises with the previous theories, but it also incorporated other psychological and linguistic models which allowed the organization of specialized knowledge. The conclusion drawn from this study was that modern Terminology was closer to Lexicology. Moreover, the applied branches of these two were explored, and we explained that the concept of specialized lexicography emerged with the rapprochement between Lexicography and (traditional) Terminography.

The last part of the chapter addressed three applied linguistic branches which have enhanced specialized lexicography, namely: (1) Cognitive Linguistics (and Frame Semantics), which helps to better understand the organization of language and its phraseology; (2) Corpus Linguistics, which offers an efficient method of analyzing

authentic language examples so as to know how it is used; and (3) Computational Linguistics, which counts on applications that make it possible the analysis of natural language.

The second aim pursued in this work was to define the concept of collocation and, more specifically, specialized collocation, and was attained in Chapter 2. Accordingly, we first explored the origins and development of Phraseology and emphasized the role played by a group of Russian scholars who aimed to categorize word combinations according to the semantic opacity and the degree of fixedness of their parts. Consequently, collocations could be placed along a continuum between idiomatic expressions and free word combinations. This assumption has influenced more recent phraseological taxonomies proposed in the twenty-first century. Nevertheless, these criteria were not enough to frame the concept of collocation. Then, a statistical approach gradually developed in the 1970s and 1980s was highlighted, and it is the one more widely used in combination with the view presented by the Russian linguists (i.e., a mixed or hybrid approach is commonly adopted when it comes to collocations), given that it offers a method of identification and extraction of word combinations.

Furthermore, the functional and formal features of collocations were examined, as well as the types that can be found. We also compared these word associations to others, such as idiomatic expressions, compounds or phrasal verbs, and, finally, defined the concept. Thus, a collocation is considered a frequent co-occurrence of two words which hold a syntactic relationship. It is perceived as a unit of language whose elements enjoy a different status, given that one of them is chosen first and evokes the accompanying item. Regarding its meaning, semantic transparency may vary from one collocation to another. With respect to the concept of specialized collocation, this definition also applies, but other features must be recognized, such as the integration of terms, their denotative function and their use by a specialized community, for they carry specialized knowledge (for this reason, they must be considered in terminological products). For instance, several works showed the special phraseology detected in the SL of tourism. Similarly, it goes without saying that these features also characterize the specialized collocations analyzed in this research.

The third goal established here was to discover the items of information that could guarantee an effective representation of collocations in a specialized dictionary, which

was reached in Chapter 3. With this objective in mind, the contents of lexicographical and terminological resources were addressed, and we found out that, in both cases, word combinations are worthy of implementation. Accordingly, a comprehensive analysis of the phraseological information described in three types of resources was conducted: (1) five important English general dictionaries, (2) five English collocation dictionaries, and (3) five terminological resources.

First, the analysis of the English general dictionaries unveiled a similar treatment of collocations in their entry for *bottle*. However, a special treatment by the OALD, the CDE and the MED must be highlighted, as they contain specific references to these word combinations. Second, the analysis of the English collocation dictionaries suggested a progressive improvement from the ECD to the MCD. For instance, while the former did not explicitly address collocations, the only negative aspect discussed in the latter was the lack of nominal collocates in verb entries. Third, the analysis of the terminological products revealed the non-representation of verb collocations.

The assessment of the analyses carried out allowed the proposal of a set of guidelines for a comprehensive specialized dictionary and collocational description in its entries. Therefore, the four items that a terminographer should consider are: (1) the types of collocations to be encoded; (2) the method used to retrieve, select and organize them; (3) the place they would take in the macrostructure and/or microstructure of the dictionary; and (4) the pieces of information that would accompany the entry. Moreover, this assessment led to the conclusion that it would be entirely appropriate that the following items should be included in number (4): (1) the lexical type of the collocation, (2) an explanation of the collocation, (3) corpus annotated sentences, and (4) equivalents in other languages (in the case of bilingual and multilingual dictionaries).

In the fourth place, we aimed to extract the verbal bases of the collocations to be implemented in *DicoAdventure* from a specialized corpus. The methodology employed was explained in Chapter 4 and the results were presented in Chapter 5. The Keywords function of Sketch Engine was used and the extraction of candidate motion verbs was performed according to their keyness score, which compared the frequency of the terms in the specialized corpus and the reference corpus (the ‘enTenTen13’) in order to find the ones that were specific to the former. No threshold was set at this point, albeit a minimum frequency of two tokens was decided.

The software automatically extracted 978 items, which were manually checked. After that, 860 units were discarded for several reasons, such as their incorrect tagging as a verb (which proved that automatic taggers are not 100% accurate), their lack of relation to the domain of adventure tourism or their meaning different from motion representation. Thus, the resulting list totaled 118 verbs (12% of the total).

Objective number five centered on the extraction of collocations produced by the 118 units previously mentioned. Again, the extraction of these word associations was explained in Chapter 4 and the results were shown in Chapter 5. In this step, the Word Sketch function of Sketch Engine was relied on and the 'enTenTen13' was the reference corpus used. The focus was on lexical collocations, that is, verbs combining with nouns, adverbs and adjectives. At this point, two criteria were established: (1) a *logDice* score of ≥ 7 points so that the collocations could be considered strong and worthy of inclusion in the specialized dictionary, and (2) a frequency of ≥ 2 tokens in the ADVENCOR corpus. Despite the fact that the Concordance function of Sketch Engine also extracts collocations, we decided not to trust it because it is less sophisticated and does not apply any linguistic criteria (although it proved useful to identify other collocations that had not been extracted through the Word Sketch function, as it was explained in Section 5.3.).

The results revealed that no verb + adjective combinations were retrieved from the specialized corpus. Additionally, it was discovered that not all the candidate verbs were successful, for less than half of them (46.61%, 55 verbs in total) produced collocations after applying the extraction criteria. Added to that, we realized that *logDice* scores and frequencies of collocations were not interrelated, for collocations assigned the highest scores were not necessarily the most recurrent ones.

Next, the last objective set in this dissertation was the classification of the collocations selected into two categories according to the meaning expressed, that is, actual motion or fictive motion. The Concordance function of Sketch Engine made it possible, since it provides the user with examples of the collocations in context directly taken from the corpus. The criteria established at this point were two: first, the collocation should occur in at least two contexts conveying the same meaning, and second, the contexts should contain a specific reference to the domain of adventure tourism. The application of these criteria revealed that only 54.54% of the motion verbs selected (30

items) produced collocations (it means 25.42% of the first list of candidate verbs extracted).

The results showed that some of the collocations extracted for five verbs (viz., *ascend*-v, *climb*-v, *cross*-v, *descend*-v and *head*-v) were identified in contexts depicting active and fictive motion, which were fully analyzed in Section 5.4., thus, they have two entries in the dictionary. In addition, the 'Collocations' option available in the Concordance function of Sketch Engine was sometimes used so as to identify new collocations which had not been previously extracted and which met all eligibility criteria. As it could be observed, the Sketch Engine software served all the specific purposes of this study. Other tools were evaluated (i.e., TermoStat Web 3.0 and WordSmith Tools 7.0), but they were not as effective as Sketch Engine to fulfill the needs of this research.

Once all the collocations had been rigorously selected, they were implemented in *DicoAdventure*. To this end, a template based on the assessment of the phraseological representation of collocations in lexicographical and terminological resources investigated for the attainment of objective three was designed. Section 4.5. explained the place that the 'Collocations' tab would occupy in the microstructure of the dictionary and proposed a model for collocational representation based on a table containing three columns: (1) the collocation itself, (2) its lexical type, and (3) an explanation using the semantic roles of the verbs' argument structure evoked by the collocates. Moreover, the third column would provide access to corpus annotated sentences illustrating the collocation in context. To our mind, this representation is adequate and comprehensible to all types of user, from professionals to laypersons.

Moving on to the data retrieved in this research, a detailed study was provided in Chapter 6. Some of the most interesting findings are summarized in the following lines:

1) The motion verbs which seemed to be more typical of the SL of adventure tourism did not present a higher degree of collocation production, such as *skydive*-v, *abseil*-v or *glide*-v;

2) The richest verbs in terms of meaning were *ascend*-v, *climb*-v, *cross*-v, *descend*-v, *head*-v and *ride*-v, for they conveyed at least two types of meaning, which are represented in different entries in *DicoAdventure*;

3) In total, 121 collocations met all eligibility criteria and 128 distinct meanings were identified; it means that seven collocations (viz., *climb up, climb (a)*

face, cross (a) bridge, cross (a) river, cross (a) stream, gradually descend and *head north*) represented two different types of meaning, more specifically, actual motion and fictive motion (incidentally, their *logDice* scores were higher than others, which means that their elements seem to be more strongly associated);

4) Except for six collocations (viz., *climb (a) mountain, climb (a) tree, fall back, reach (a) peak, ride (a) bike* and *walk (one) mile*), all of them achieved a higher *logDice* score in the specialized corpus, which indicated that they must be considered strong word associations and must be implemented in the specialized dictionary (the other six must also be included because all the elements of the contexts pertained to the SL of adventure tourism and they fulfilled all eligibility criteria);

5) Collocation production and collocation strength are not interrelated, as they do not increase at the same pace; in other words, highly productive verbs may produce moderately strong word associations, such as *climb*-v and *skydive*-v;

6) By the same token, collocation production and collocation frequency are not interrelated, for one verb may produce a small amount of collocations which are extremely recurrent in the text, for example, *hike*-v;

7) When the normalized frequency per million of the collocations selected was calculated in both corpora, it was discovered that they would occur much more frequently in the SL of adventure tourism than in the GL, which leads to the conclusion that they are specialized and worthy of inclusion in a terminological product;

8) The frame-based analysis of the motion verbs and the collocations performed in this research helped to better understand the organization of knowledge in the domain of adventure tourism and proved the motivation of collocates generated by the meanings of the verbs evoking each frame;

9) Verbs sharing the same argument structures and being semantically similar did not always share the same collocates (e.g., *navigate*$_1$-v and *paddle*$_1$-v), which may be explained due to the slight divergences identified in their meanings;

10) Noun + verb collocations were normally identified in fictive motion events (e.g., *(a) climb ascend, (a) trail descend*), given that verbs expressing actual motion usually appeared in the imperative form or used the subject *you*;

11) The PART FOR WHOLE metonymy was identified in some collocations (e.g., *climb (a) route, scale (a) peak*), which means that they must be interpreted figuratively;

12) Collocations highlighting the highest point of a natural element, such as *reach (the) summit, reach (the) top* or *reach (a) peak*, were considered strong associations of words given their high *logDice* scores, which may be due to the fact that reaching the highest point of a mountain is the main objective in a wide variety of adventure activities, and also the experience which provides adventurers with greater satisfaction;

13) Collocates expressing DIRECTION, SPEED, PATH and MANNER were the most repeated ones, for example, *down, slowly, trail, tandem*; and

14) Although there was a general tendency of the collocates to represent the arguments of the verbs (i.e., the core semantic roles), several circumstantials typical of the SL of adventure tourism were also frequent, such as MANNER, SPEED and TIME.

Despite acknowledging that these conclusions prove broadly satisfactory and admitting that this research has confirmed the hypothesis formulated, that is, the need for the implementation of verb collocations in terminological products, we must recognize some limitations:

1) Phraseological information encoded in verb entries in English general dictionaries could not be examined, given that phraseology does not usually appear in this type of entries. However, our analysis of the representation of phraseological units in verb entries included in English collocation dictionaries and multilingual terminological resources helped us to propose a (successful) model for the implementation of collocations in the entries for motion verbs in *DicoAdventure*, the primary objective set in this dissertation;

2) The thresholds established for the collocation extraction (i.e., *logDice* score $\geq$ 7 points and frequency $\geq$ 2 tokens) may have been tighter (e.g., *logDice*

score ≥ 10 points and frequency ≥ 5 tokens). Nevertheless, we think that the strict criteria established in the subsequent collocation classification (i.e., the existence of at least two contexts conveying the same meaning and making a specific reference to the domain of adventure tourism) helped to make a careful selection of the collocations worthy of implementation in the specialized dictionary; and

3) In some cases, the Word Sketch function of Sketch Engine was not enough for the identification of collocations, so the Collocations option included in the Concordance function was used. Two plausible reasons may be: (1) the Word Sketch function applies a strict linguistic criterion, and (2) it focuses on words co-occurring within the same sentence. Thus, an incorrect tagging of the words in the corpus or collocations whose elements co-occurred in distinct sentences may have skewed the final results.

Finally, and knowing that *DicoAdventure* is an ongoing project, we present possible lines of further research:

1) Grammatical collocations of the motion verbs analyzed (i.e., their government patterns) may be examined and incorporated into their entries;

2) The implementation of verbs which do not express motion and other word types, such as nouns, adjectives and adverbs, along with their collocations, may be achieved;

3) Less stringent criteria may be applied for selecting collocations so that *DicoAdventure* can host a greater range of these word associations;

4) Other languages may be explored so that the dictionary can be bilingual (e.g., English–Spanish) or even multilingual (e.g., English–Spanish–French); and

5) Collocation equivalents in other languages may be provided, given that in many cases they cannot be translated literally.

Overall, our aim is to offer a valuable resource which gathers the terminology and frequent combinations typical of the SL of adventure tourism. The result would be a quality product which offers real help to any type of user, such as translators, tourist guides, teachers or students. Additionally, the user-friendly interface of the database and

the suitable arrangement of the information implemented in the entries may inspire other terminological products dealing with different SLs.

CHAPTER 8. REFERENCES

Aarts, J., & Meijs, W. (Eds.). (1984). *Corpus linguistics: Recent developments in the use of computer corpora in English language research*. Amsterdam, The Netherlands: Rodopi.

Addison, G. (1999). Adventure tourism and ecotourism. In J. C. Miles & S. Priest (Eds.), *Adventure programming* (pp. 415–430). State College, PA: Venture Publishing.

Adolphs, S., & Knight, D. (2010). Building a spoken corpus: What are the basics? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 38–52). London/New York, England/NY: Routledge.

Ahmad, K., & Rogers, M. (2001). The analysis of text corpora for the creation of advanced terminology databases. In S. E. Wright & G. Budin (Eds.), *Handbook of terminology management. Volume II* (pp. 725–760). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Aisenstadt, E. (1979). Collocability restrictions in dictionaries. *ITL International Journal of Applied Linguistics, 45/46*, 71–74. doi:10.1075/itl.45-46.10ais

Alberts, M. (2001). Lexicography versus Terminography. *Lexikos, 11*, 71–84. doi:10.5788/11-0-840

Alcaraz Varó, E. (2000). *El inglés profesional y académico*. Madrid, Spain: Alianza.

Alcaraz Varó, E., Hugues, B., Campos Pardillos, M. A., Pina Medina, V. M., & Alesón Carbonell, M. A. (2006). *Diccionario de términos de turismo y de ocio. Inglés-Español, Spanish-English*. Barcelona, Spain: Ariel.

Aldhubayi, L., & Alyahya, M. (2014). Automated Arabic antonym extraction using a corpus analysis tool. *Journal of Theoretical and Applied Information Technology, 70*(3), 422–433. Retrieved from http://www.jatit.org/volumes/Vol70No3/7Vol70No3.pdf

Alipour, M., Robichaud, B., & L'Homme, M.-C. (2015). Towards an electronic specialized dictionary for learners. In I. Kosem, M. Jakubíček, J. Kallas, & S. Krek (Eds.), *Electronic lexicography in the 21st century: Linking lexical data in the*

digital age (pp. 51–69). Ljubljana/Brighton, Slovenia/England: Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd.

Alonso Campos, A., & Torner Castells, S. (2010). Adjectives and collocations in specialized texts: Lexicographical implications. In A. Dykstra & T. Schoonheim (Eds.), *Proceedings of the 14th EURALEX International Congress* (pp. 872–881). Leeuwarden, The Netherlands: Fryske Akademy – Afûk. Retrieved from https://euralex.org/publications/adjectives-and-collocations-in-specialized-texts-lexicographical-implications/

Alonso Campos, A., Williams, G., & DeCesaris, J. (2017). Studying lexical meaning in context. From collocation to collocational networks and resonance. In S. Torner Castells & E. Bernal (Eds.), *Collocations and other lexical combinations in Spanish. Theoretical, lexicographical and applied perspectives* (pp. 41–74). London/New York, England/NY: Routledge.

Amosova, N. N. (1963). *Osnovy anglijskoj frazeologii*. Leningrad, Russia: Izdatelstvo Leningradskogo Universyteta.

Anderson, W. J. (2006). *The phraseology of administrative French. A corpus-based study*. Amsterdam/New York, The Netherlands/NY: Rodopi.

AntConc [Computer software]. Tokyo, Japan: Laurence Anthony.

Arnold, I. V. (1973). *The English word*. Moscow, Russia: Vyssaja Skola.

Atkins, B. T., & Rundell, M. (2008). *The Oxford guide to practical lexicography*. Oxford, England: Oxford University Press.

Atkins, S., Clear, J., & Ostler, N. (1992). Corpus design criteria. *Literary and Linguistic Computing, 7*(1), 1–16. doi:10.1093/llc/7.1.1

Baer, P. de, Kerremans, K., & Termmerman, R. (2006). Bridging communication gaps between legal experts in multilingual Europe: Discussion of a tool for exploring terminological and legal knowledge resources. In E. Corino, C. Marello, & C. Onesti (Eds.), *Proceedings of the 12th EURALEX International Congress* (pp. 813–818). Alessandria, Italy: Edizioni dell'Orso. Retrieved from https://euralex.org/publications/bridging-communication-gaps-between-legal-experts-in-multilingual-europe-discussion-of-a-tool-for-exploring-terminological-and-legal-knowledge-resources/

Baer, P. de, Meersman, R., & Temmerman, R. (2009). Termontography and DOGMA for Knowledge Engineering within PROLIX. In R. Meersman, P. Herrero, & R. Dillon (Eds.), *On the move to meaningful Internet systems: OTM 2009 Workshops* (pp. 534–543). Berlin, Germany: Springer.

Bajčić, M. (2017). *New insights into the semantics of legal concepts and the legal dictionary*. Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Baker, P. (2006). *Using corpora in discourse analysis*. London/NewYork, England/NY: Continuum.

Baldwin, T., & Kim, S. N. (2010). Multiword expressions. In N. Indurkhya & F. J. Damerau (Eds.), *Handbook of Natural Language Processing. Second edition* (pp. 267–292). Boca Raton, FL: Taylor & Francis Group.

Bally, C. (1909). *Traité de stylistique française. Volume I*. Heidelberg, Germany: Carl. Winters Universitätsbuchhandlung.

Bally, C. (1919). *Traité de stylistique française. Volume II*. Heidelberg, Germany: Carl. Winters Universitätsbuchhandlung.

Bally, C. (1951). *Traité de stylistique française. Volume I. Second edition*. Paris, France: Librairie C. Klincksieck

Barnbrook, G., Mason, O., & Krishnamurthy, R. (2013). *Collocation. Applications and implications.* Basingstoke, England: Palgrave Macmillan.

Barsalou, L. W. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes, 18*(5–6), 513–562. doi:10.1080/01690960344000026

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617–645. doi:10.1146/annurev.psych.59.103006.093639

Bartsch, S. (2004). *Structural and functional properties of collocations in English: A corpus study of lexical and pragmatic constraints on lexical co-occurrence*. Tübingen, Germany: Gunter Narr Verlag Tübingen.

Bauer, L. (1983). *English word-formation*. Cambridge, England: Cambridge University Press.

Baynat Monreal, M. E. (2017). 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. Retrieved from http://cedille.webs.ull.es/13/03baynat.pdf

Beaugrande, R. de. (1987). Special purpose language and linguistic theory. *UNESCO ALSED-LSP Newsletter, 10/2*(25), 2–10.

Benson, M., Benson, E., & Ilson, R. (1986a). *Lexicographic description of English*. Amsterdam, The Netherlands: John Benjamins Publishing Company.

Benson, M., Benson, E., & Ilson, R. (1986b). *The BBI Combinatory Dictionary of English: A guide to word combinations.* Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Benson, M., Benson, E., & Ilson, R. (1997). *The BBI Combinatory Dictionary of English word combinations. Revised edition*. Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Benson, M., Benson, E., & Ilson, R. (2009). *The BBI Combinatory Dictionary of English: Your guide to collocations and grammar. Third edition*. Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Bergenholtz, H. (1996). Grundfragen der Fachlexikographie. In M. Gellerstam et al. (Eds.), *Proceedings of the 7th EURALEX International Congress* (pp. 731–758). Göteborg, Sweden: Göteborg University. Retrieved from https://euralex.org/publications/grundfragen-der-fachlexikographie/

Bergenholtz, H. (1995). Wodurch unterscheidet sich Fachlexikographie von Terminographie? *Lexicographica, International Annual for Lexicography, 11*, 37–46. doi:10.1515/9783110244151.50

Bergenholtz, H. (1998). Das schlaue Buch. Vermittlung von Informationen für textbezogene und textunabhängige Fragestellungen. In A. Zettersten, V. H. Pedersen, & J. E. Mogensen (Eds.), *Symposium on Lexicography VIII* (pp. 93–110). Tübingen, Germany: Niemeyer.

Bergenholtz, H., & Kaufmann, U. (1997). Terminography and lexicography. A critical survey of dictionaries from a single specialized field. *Hermes, Journal of Linguistics, 10*(18), 91–125. doi:10.7146/hjlcb.v10i18.25413

Bergenholtz, H., & Nielsen, S. (2002). Terms in the language of culture-dependent LSP dictionaries. *Lexicographica, International Annual for Lexicography, 18*, 5–18. doi:10.1515/9783484604476.5

Bergenholtz, H., & Tarp, S. (1995). *Manual of specialized lexicography. The preparation of specialized dictionaries*. Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Bergenholtz, H., & Tarp, S. (2002). Die moderne lexikographische Funktionslehre. Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchsgegenstände verstehen. *Lexicographica, International Annual for Lexicography, 22*, 145–155.

Bergenholtz, H., & Tarp, S. (2003). Two opposing theories: On H. E. Wiegand's recent discovery of lexicographic functions. *Hermes, Journal of Linguistics, 31*, 171–196. doi:10.7146/hjlcb.v16i31.25743

Bergenholtz, H., & Tarp, S. (2010). LSP lexicography or terminography? The lexicographer's point of view. In P. A. Fuertes-Olivera (Ed.), *Specialized dictionaries for learners* (pp. 27–37). Berlin/New York, Germany/NY: Walter de Gruyter.

Bernardini, S., Baroni, M., & Evert, S. (2006). A WaCky introduction. In M. Baroni & S. Bernardini (Eds.), *WaCky! Working papers on the Web as Corpus* (pp. 9–40). Bologna, Italy: Gedit Edizioni.

Bevilacqua, C. R. (2004). *Unidades fraseológicas especializadas eventivas: Descripción y reglas de formación en el ámbito de la energía solar* (Doctoral thesis, Universitat Pompeu Fabra, Barcelona, Spain). Retrieved from http://hdl.handle.net/10803/7515

Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing, 8*(4), 243–257. doi:10.1093/llc/8.4.243

Biber, D. (1995). *Dimensions of register variaton. A cross-linguistic comparison*. Cambridge, England: Cambridge University Press.

Biber, D., & Barbieri, F. (2006). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*, 263–286. doi:10.1016/j.esp.2006.08.003

Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgård & S. Oksefjell (Eds.), *Out of corpora: Studies in honour of Stig Johansson* (pp. 181–189). Amsterdam, The Netherlands: Rodopi.

Biber, D., & Reppen, R. (2015). Introduction. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 1–8). Cambridge, England: Cambridge University Press.

Biel, Ł., Biernacka, A., & Jopek-Bosiacka, A. (2018). Collocations of terms in EU competition law: A corpus analysis of EU English collocations. In S. Marino, Ł. Biel, M. Bajčić, & V. Sosoni (Eds.), *Language and Law. The role of language and translation in EU competition law* (pp. 249–274). Cham, Switzerland: Springer.

Bieliaieva, O., Lysanets, Y. V., Znamenska, I., Rozhenko, I., & Nikolaieva, N. (2017). Terminological collocations in medical Latin and English: A comparative study. *Wiadomosci Lekarskie, 70*(1), 139–143.

Birkner, V. (2015). Advantages and disadvantages of employing corpus evidence in sociolinguistic studies. *The Teacher Magazine, 2*(126), 11–17. Retrieved from https://www.academia.edu/11326263/Advantages_and_Disadvantages_of_employing_corpus_evidence_in_sociolinguistic_studies

Boers, F., & Lindstromberg, S. (Eds.). (2008). *Cognitive linguistic approaches to teaching vocabulary and phraseology.* Berlin/New York, Germany/NY: Mouton de Gruyter.

Bogaards, P. (1996). Dictionaries for learners of English. *International Journal of Lexicography, 9*(4), 277–320. doi:10.1093/ijl/9.4.277

Bolshakov, I. A., & Gelbukh, A. (2004). *Computational linguistics. Models, resources, applications*. Mexico DF, Mexico: Instituto Politécnico Nacional.

Borja Albi, A. (2000). *El texto jurídico inglés y su traducción al español*. Barcelona, Spain: Ariel.

Boulanger, J. C. (1991). Une lecture socioculturelle de la Terminologie. *Cahiers de Linguistique Sociale, 18*, 13–30.

Boulanger, J. C. (1995). Présentation: Images et parcours de la socioterminologie. *Meta, Journal des Traducteurs, 40*(2), 194–205. doi:10.7202/002117ar

Bowker, L., & Pearson, J. (2002). *Working with specialized language. A practical guide to using corpora*. London/New York, England/NY: Routledge.

Bravo Gozalo, J. M., & Fernández Nistal, P. (1998). La lingüística del corpus, las nuevas tecnologías de la información y los estudios de traducción en la década de 1990. In P. Fernández Nistal & J. M. Bravo Gonzalo (Eds.), *La traducción: Orientaciones lingüísticas y culturales* (pp. 205–257). Valladolid, Spain: Universidad de Valladolid.

Bray, T., Paoli, J., & Sperberg-McQueen, C. M. (Eds.). (1998). *Extensible Markup Language (XML). Version 1.0. W3C Recommendation*. Retrieved from http://www.w3.org:TR/1998/REC-xml-19980210.

Brezina, V. (2018). *Statistics in corpus linguistics. A practical guide*. Cambridge, England: Cambridge University Press.

Buckley, R. (2010). *Adventure tourism management*. Oxford/Burlington, England/MA: Elsevier.

Buendía Castro, M. (2012). Verb dynamics. *Terminology, 18*(2), 149–166. doi:10.1075/term.18.2.01bue

Buendía Castro, M. (2013). Phraseology in specialized language and its representation in environmental knowledge resources (Doctoral thesis, Universidad de Granada, Granada, Spain). Retrieved from https://hera.ugr.es/tesisugr/21876927.pdf

Buendía Castro, M., & Faber Benítez, P. (2014). Collocation dictionaries: A comparative analysis. *MonTI, 6*, 203–235. doi:10.6035/MonTI.2014.6.7

Buendía Castro, M., & Faber Benítez, P. (2017). Verb collocations for natural disasters: A contrastive study of Mexican and peninsular Spanish. *Sendebar, 28*, 189–208. Retrieved from https://revistaseug.ugr.es/public/sendebar/sendebar28_completo.pdf

Buendía Castro, M., & López Rodríguez, C. I. (2013). The Web for Corpus and the Web as Corpus in translator training. *New Voices in Translation Studies, 10*, 54–71. Retrieved from https://www.academia.edu/21619077/The_Web_for_Corpus_and_the_Web_as_Corpus_in_Translator_Training

Buendía Castro, M., Montero Martínez, S., & Faber Benítez, P. (2014). Verb collocations and phraseology in English. *Yearbook of Phraseology, 5*, 57–94. doi:10.1515/phras-2014-0004

Burger, H. (1998). *Phraseologie. Eine Einführung am Beispiel des Deutschen*. Berlin, Germany: Erich Schmidt.

Cabré Castellví, M. T. (1993). *La terminología. Teoría, metodología, aplicaciones*. Barcelona, Spain: Antártida/Empúries.

Cabré Castellví, M. T. (1999). *Terminology. Theory, methods and applications*. Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Cabré Castellví, M. T. (2000). Elements for a theory of terminology: Towards an alternative paradigm. *Terminology, 6*(1), 35–57. doi:10.1075/term.6.1.03cab

Cabré Castellví, M. T. (2002). Textos especializados y unidades de conocimiento: Metodología y tipologización (I). In J. García Palacios & M. T. Fuentes Morán (Eds.), *Texto, terminología y traducción* (pp. 15–36). Salamanca, Spain: Ediciones Almar.

Cabré Castellví, M. T. (2003). Theories of terminology. Their description, prescription and explanation. *Terminology, 9*(2), 163–199. doi:10.1075/term.9.2.03cab

Cabré Castellví, M. T. (2005). La Terminología, una disciplina en evolución: Pasado, presente y algunos elementos de futuro. *Debate terminológico, 1*. Retrieved from https://www.researchgate.net/publication/237821583_La_Terminologia_una_disciplina_en_evolucion_pasado_presente_y_algunos_elementos_de_futuro

Cabré Castellví, M. T., Estopà Bagot, R., & Lorente Casafont, M. (1996, November). *Terminología y fraseología*. Paper presented at the V Simposio de Terminología Iberoamericana, Ciudad de México (Mexico).

Calvi, M. V. (2010). Los géneros discursivos en la lengua del turismo: Una propuesta de clasificación. *Ibérica, 13*, 9–32. Retrieved from https://www.redalyc.org/pdf/2870/287024099002.pdf

Calvi, M. V., & Bonomi, M. (2008). El lenguaje del turismo: De los textos especializados a la Comunidad del viajero. In C. Navarro, R. M. Rodríguez Abella, & F. Dalle

Pezze (Eds.), *La comunicación especializada: Terminología y traducción* (pp. 181–202). Bern, Switzerland: Peter Lang.

Cambridge Dictionary of English. (2020). Cambridge, England: Cambridge University Press.

Campo, Á. (2012). *The reception of Eugen Wüster's work and the development of Terminology* (Doctoral thesis, Université de Montréal, Montréal, Canada). Retrieved from https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/9198/Campo_Angela_2012_these.pdf

Cao, D., & Deignan, A. (2019). Using an online collocation dictionary to support learners' L2 writing. In C. Wright, L. Harvey, & J. Simpson (Eds.), *Voices and practices in applied linguistics. Diversifying a discipline* (pp. 233–249). York, England: White Rose University Press.

Cappelli, G. (2006). *Sun, sea, sex and the unspoilt countryside. How the English language makes tourists out of readers*. Pari, Italy: Pari Publishing.

Carter, R. (1998). *Vocabulary*. London, England: Routledge.

Carter, R., & McCarthy, M. J. (2001). Size isn't everything: Spoken English, corpus and the classroom. *TESOL Quarterly, 35*(2), 337–340. doi:10.2307/3587654

Casademont, A. J. (2014). On the elements activating the transmission of specialized knowledge in verbs. *Terminology, 20*(1), 92–116. doi:10.1075/term.20.1.05joa

Cervero, M. J., & Pichardo Castro, F. (2000). *Aprender y enseñar vocabulario*. Madrid, Spain: Edelsa.

Chien, L.-F., & Chen, C.-L. (2001). Incremental extraction of domain-specific terms from online text resources. In C. Jacquemin & M.-C. L'Homme (Eds.), *Recent advances in computational linguistics* (pp. 89–109). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Choueka, Y. (1988). Looking for needles in a haystack or locating interesting collocational expressions in large textual databases. In C. Fluhr & D. Walker (Eds.), *Proceedings of RIAO '88: User-oriented content-based text and image*

handling (pp. 609–623). Paris, France: Le Center de Hautes Études Internationales D'Informatique Documentaire.

Chrupała, G. (2014, January). *Language-agnostic processing of microblog data with text embeddings*. Poster session presented at the 24th Meeting of Computational Linguistics, Leiden (The Netherlands).

Cinková, S., & Pomikálek, J. (2006). LEMPAS: A make-do lemmatizer for the Swedish PAROLE-corpus. *The Prague Bulletin of Mathematical Linguistics, 86,* 47–54. Retrieved from http://ufal.mff.cuni.cz/pbml/86/cinkova-pomikalek.pdf

Clancy, B., & O'Keeffe, A. (2015). Pragmatics. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 235–251). Cambridge, England: Cambridge University Press.

Clear, J. (1993). From Firth principles. Computational tools for the study of collocation. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 271−292). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Clear, J. (1994). I can't see the sense in a large corpus. In F. Kiefer, G. Kiss, & J. Pajzs (Eds.), *Papers in computational lexicography, COMPLEX '94* (pp. 33–48). Budapest, Hungary: Research Institute for Linguistics, Hungarian Academy of Sciences.

Cloke, P., & Perkins, H. C. (1998). Cracking the canyon with the awesome foursome: Representations of adventure tourism in New Zealand. *Environment and Planning D: Society and Space, 16*, 185−218. Retrieved from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.465.3212&rep=rep1&type=pdf

Cohen, H., & Lefebvre, C. (2005). Bridging the category divide. In H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (pp. 1–15). Amsterdam, The Netherlands: Elsevier.

Collins English Dictionary. (2020). Glasgow, Scotland: HarperCollins Publishers.

ConcGram 1.0. A phraseological search engine [Computer software]. Hong Kong, China: Chris Greaves.

Condamines, A., & Rebeyrolle, J. (2001). Searching for and identifying conceptual relationships via a corpus-based approach to a Terminological Knowledge Base (CTKB): Method and Results. In C. Jacquemin & M.-C. L'Homme (Eds.), *Recent advances in computational linguistics* (pp. 127–148). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Corbacho Sánchez, A. (2005). Los conceptos de *Tour(ismus)* y *Reise* en los fraseologismos alemanes y su traducción al español. *Pragmalingüística, 13,* 65–75. doi:10.25267/Pragmalinguistica.2005.i13.04

Corbacho Sánchez, A. (2017). Colocaciones verbo-nominales de la correspondencia hotelera en alemán: Incidencia en la enseñanza y traducción al español. *Encuentro, 26,* 94–112. Retrieved from https://ebuah.uah.es/dspace/handle/10017/35079

Corpas Pastor, G. (1996). *Manual de fraseología española*. Madrid, Spain: Gredos.

Corpas Pastor, G. (2008). *Investigar con corpus en traducción: Los retos de un nuevo paradigma*. Frankfurt am Main, Germany: Peter Lang.

Corpas Pastor, G. (Ed.). (2016). *Computerized and corpus-based approaches to phraseology: Monolingual and multilingual perspectives*. Geneva, Switzerland: Editions Tradulex.

Corpas Pastor, G., & Colson, J.-P. (Eds.). (2020). *Computational phraseology*. Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Corpas Pastor, G., & Mitkov, R. (Eds.). (2019). *Lecture Notes in Artificial Intelligence. Volume 11755. Computational and corpus-based phraseology*. Cham, Switzerland: Springer.

Corpas Pastor, G., & Seghiri Domínguez, M. (2010). Size matters: A quantitative approach to corpus representativeness. In R. Rabadán, M. Fernández López, & T. Guzmán González (Eds.), *Lengua, traducción, recepción en honor de Julio César Santoyo* (pp. 111–145). León, Spain: Universidad de León, Área de Publicaciones.

Coulmas, F. (Ed.). (1981). *Conversational routine. Explorations in standardized communication situations and prepatterned speech*. The Hague, The Netherlands: Mouton de Gruyter.

Coupland, N., & Jaworski, A. (2001). Discourse. In P. Cobley (Ed.), *The Routledge Companion to semiotics and linguistics* (pp. 134–148). London, England: Routledge.

Cowie, A. P. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics, 2*(3), 223–235. doi:10.1093/applin/II.3.223

Cowie, A. P. (1988). Stable and creative aspects of vocabulary use. In R. Carter & M. McCarthy (Eds.), *Vocabulary and language teaching* (pp. 126−139). London, England: Longman.

Cowie, A. P. (1998a). Past achievements and current trends. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 1–20). Oxford, England: Clarendon Press.

Cowie, A. P. (1998b). Phraseological dictionaries: Some east-west comparisons. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 209–228). Oxford, England: Clarendon Press.

Cowie, A. P. (2001). Speech formulae in English: Problems of analysis and dictionary treatment. In G. van der Meer (Ed.), *Making sense: From lexeme to discourse. In honor of Werner Abraham at the occasion of his retirement* (pp. 1–12). Groningen, The Netherlands: Center for language and Cognition.

Croft, W., & Cruse, D. A. (2004). *Cognitive Linguistics*. Cambridge, England: Cambridge University Press.

Crowther, J., Dignen, S., & Lea, D. (2002). *Oxford collocations dictionary for students of English.* Oxford, England: Oxford University Press.

Cruse, A. (1986). *Lexical semantics*. Cambridge, England: Cambridge University Press.

Cuenca, M. J., & Hilferty, J. (1999). *Introducción a la lingüística cognitiva*. Barcelona, Spain: Ariel.

Culpeper, J., & Demmen, J. (2015). Keywords. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 90–105). Cambridge, England: Cambridge University Press.

Curado Fuentes, A. (2001). Lexical behaviour in academic and technical corpora: Implications for ESP development. *Language Learning & Technology, 5*(3), 106–

129. Retrieved from https://www.researchgate.net/publication/255595552_Lexical_behaviour_in_academic_and_technical_corpora_Implications_for_ESP_development

Daille, B. (1994). *Approche mixte pour l'extraction automatique de terminologie: Statistiques lexicales et filtres linguistiques* (Unpublished doctoral dissertation). Université Paris 7, Paris, France.

Daille, B. (2001). Qualitative terminology extraction: Identifying relational adjectives. In C. Jacquemin & M.-C. L'Homme (Eds.), *Recent advances in computational linguistics* (pp. 149–166). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Dann, G. (1996). *The language of tourism. A sociolinguistic perspective*. Wallingford, England: CAB International.

Davies, M. (2011). Synchronic and diachronic uses of corpora. In V. Viana, S. Zyngier, & G. Barnbrook (Eds.), *Perspectives on corpus linguistics* (pp. 63–80). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Davies, M. (2015). Corpora: An introduction. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 11–31). Cambridge, England: Cambridge University Press.

Dereza, O. (2019). Lemmatisation for under-resourced languages with sequence-to-sequence learning: A case of Early Irish. *EPiC Series in Language and Linguistics, 4*, 113–124. doi:10.29007/CXTL

DicoAdventure. An online dictionary of adventure tourism. (2020). Montréal, Canada: OLST.

DiCoEnviro. Le dictionnaire fondamental de l'environnement. (2020). Montréal, Canada: OLST.

DiCoInfo. Le dictionnaire fondamental de l'informatique et de l'Internet. (2020). Montréal, Canada: OLST.

Donalies, E. (2004). *Grammatik des Deutschen im Europäischen Vergleich. Kombinatorische Begriffsbildung: Substantivkomposition*. Mannheim, Germany: Institut für Deutsche Sprache.

Drouin, P. (2003). Term extraction using non-technical corpora as a point of leverage. *Terminology, 9*(1), 99–115. doi:10.1075/term.9.1.06dro

Durán-Muñoz, I. (2011). *El trabajo ontoterminográfico aplicado a la traducción de textos del turismo de aventura (español-inglés-alemán): Fases de preparación, elaboración y edición* (Doctoral thesis, Universidad de Málaga, Málaga, Spain). Retrieved from https://riuma.uma.es/xmlui/bitstream/handle/10630/4847/TD%20de%20Isabel%20Dur%c3%a1n%20Mu%c3%b1oz.pdf?sequence=1

Durán-Muñoz, I. (2012). *La ontoterminografía aplicada a la traducción. Propuesta metodológica para la elaboración de recursos terminológicos dirigidos a traductores*. Frankfurt am Main, Germany: Peter Lang.

Durán-Muñoz, I. (2013). El turismo de aventura en textos promocionales y su terminología. In M. A. Candel Mora (Ed.), *El lenguaje del turismo y la publicidad: Terminología y traducción* (pp. 251–277). Barcelona, Spain: Wolters Kluwer.

Durán-Muñoz, I. (2016). Producing frame-based definitions: A case study. *Terminology, 22*(2), 224–250. doi:10.1075/term.22.2.04mun

Durán-Muñoz, I. (2019). Adjectives and their keyness: A corpus-based analysis of tourism discourse in English. *Corpora, 14*(3), 351–378. doi:10.3366/cor.2019.0178

Durán-Muñoz, I. (in press). *DicoAdventure* y la terminología del turismo de aventura: Propuesta de diccionario en línea. In T. Barceló Martínez & F. García Luque (Eds.), (in press [pp.391–413]). Valencia, Spain: Tirant Lo Blanch.

Durán-Muñoz, I., & L'Homme, M.-C. (2020). Diving into English motion verbs from a lexico-semantic approach. A corpus-based analysis of adventure tourism. *Terminology, 26*(1), 33–59. doi:10.1075/term.00041.dur

EcoLexicon. (2020). Granada, Spain: LexiCon Research Group.

El Emary, I. M. M., & Atwan, J. (2005). Designing and building an automatic information retrieval system for handling the Arabic data. *American Journal of Applied Sciences, 2*(11), 1520–1525. doi:10.3844/ajassp.2005.1520.1525

Evans, V., Bergen, B. K., & Zinken, J. (Eds.). (2007). *The cognitive linguistics reader*. London/Oakville, England/Canada: Equinox.

Evert, S. (2005). *The statistics of word cooccurrences. Word pairs and collocations* (Doctoral thesis, Universität Stuttgart, Stuttgart, Germany). Retrieved from https://elib.uni-stuttgart.de/bitstream/11682/2573/1/Evert2005phd.pdf

Evert, S. (2009). Corpora and collocations. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics. An international handbook. Volume II* (pp. 1212–1248). Berlin/New York, Germany/NY: Walter de Gruyter.

Faber Benítez, P. (2011). The dynamics of specialized knowledge representation: Simulational reconstruction or the perception–action interface. *Terminology, 17*(1), 9–29. doi:10.1075/term.17.1.02fab

Faber Benítez, P. (2015). Frames as a framework for terminology. In H. J. Kockaert & F. Steurs (Eds.), *Handbook of terminology. Volume I* (pp. 14–33). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company. doi:0.1075/hot.1.02fra1

Faber Benítez, P., & Cabezas García, M. (2019). Specialized knowledge representation: From terms to frames. *Research in Language, 17*(2), 197–211. doi:10.2478/rela-2019-0012

Faber Benítez, P., & León Araúz, P. (2010). Dinamismo conceptual en las bases de conocimiento terminológico: El caso de EcoLexicon. *Íkala, Revista de Lenguaje y Cultura, 15*(25), 75–100. Retrieved from https://www.redalyc.org/pdf/2550/255020319004.pdf

Faber Benítez, P., & León-Aráuz, P. (2019). Frame-based terminology applied to military science: Transforming a glossary into a knowledge resource. *Lexicography, 6*, 105–131. doi:10.1007/s40607-019-00060-y

Faber Benítez, P., & López-Rodríguez, C. I. (2012). Terminology and specialized language. In P. Faber Benítez (Ed.), *A cognitive linguistics view of terminology*

and specialized language (pp. 9–31). Berlin/Boston, Germany/MA: De Gruyter Mouton.

Faber Benítez, P., & Mairal Usón, R. (1999). *Constructing a lexicon of English verbs*. Berlin/NewYork, Germany/NY: Mouton de Gruyter.

Faber Benítez, P., Márquez Linares, C., & Vega Expósito, M. (2005). Framing terminology: A process-oriented approach. *Meta, Journal des Traducteurs, 50*(4), 1–11. doi:10.7202/019916ar

Faber Benítez, P., et al. (2006). Process-oriented terminology management in the domain of Coastal Engineering. *Terminology, 12*(2), 189–213. doi:10.1075/term.12.2.03fab

Fairclough, N. (1992). *Discourse and social change*. Cambridge, England: Polity Press.

Fano, R. M. (1961). *Transmission of information: A statistical theory of communications*. New York, NY: MIT Press.

Fasold, R. (1990). *The sociolinguistics of language. Introduction to sociolinguistics. Volume II.* Oxford, England: Basil Blackwell.

Fauconnier, G. (1994). *Mental spaces: Aspects of meaning construction in natural language*. Cambridge, England: Cambridge University Press.

Fauconnier, G., & Turner, M. (1998). Conceptual integration networks. *Cognitive Science, 22*(2), 133–187. doi:10.1016/S0364-0213(99)80038-X

Fauconnier, G., & Turner, M. (2002). *The way we think: Conceptual blending and the mind's hidden complexities*. New York, NY: Basic Books

Fauconnier, G., & Turner, M. (2003). Conceptual blending, form and meaning. *Recherches en Communication, 19*, 57–86. doi:10.14428/rec.v19i19.48413

Federici, E. (2018). *Translation theory and practice. Cultural differences in tourism and advertising*. Naples, Italy: Paolo Loffredo, Iniziative Editoriali.

Fennell, D. (2003). *Ecotourism. Second edition*. London/New Tork, England/NY: Routledge.

Fernández Fernández, J. (2011). *Representación del conocimiento terminológico difundido en la web. Estudio del dominio del cambio climático* (Doctoral thesis,

Universidad de Granada, Granada, Spain). Retrieved from https://hera.ugr.es/tesisugr/20745539.pdf

Fillmore, C. J. (1976). Frame semantics and the nature of language. *Annals of the New York Academy of Sciences, 280*(1), 20–32. doi:10.1111/j.1749-6632.1976.tb25467.x

Fillmore, C. J. (1977). Scenes and frames semantics. In A. Zampolli (Ed.), *Linguistic structures processing* (pp. 55−88). Amsterdam/New York/Oxford, The Netherlands/NY/England: North-Holland Publishing Company.

Fillmore, C. J. (1982). Frame semantics. In Linguistic Society of Korea (Ed.), *Linguistics in the morning calm* (pp. 111−137). Seoul, South Korea: Hanshin Publishing Company.

Fillmore, C. J. (1985). Frames and the semantics of understanding. *Quaderni di Semantica, 6*(2), 222−254. Retrieved from http://www.icsi.berkeley.edu/pubs/ai/framesand85.pdf

Fillmore, C. J. (1992). "Corpus linguistics" or "Computer-aided armchair linguistics." In J. Svartvik (Ed.), *Directions in corpus linguistics* (pp. 35–60). Berlin/New York, Germany/NY: Mouton de Gruyter.

Fillmore, C. J. (2006). Frame semantics. In D. Geeraerts (Ed.), *Cognitive linguistics: Basic readings* (pp. 373–400). Berlin/New York, Germany/NY: Mouton de Gruyter.

Fillmore, C. J., & Atkins, B. T. (1992). Toward a frame-based lexicon: The semantics of RISK and its neighbors. In A. Lehrer & E. Kittay (Eds.), *Frames, fields and contrast: New essays in semantics and lexical organization* (pp. 75–102). Hillsdale, NJ: Lawrence Erlbaum Associates.

Firth, J. R. (1957). *Papers in Linguistics, 1934-1951*. London, England: Oxford University Press.

Firth, J. R. (1962). A synopsis of linguistic theory, 1930-1955. In J. R. Firth et al. (Eds.), *Studies in linguistic analysis* (pp. 1–32). Oxford, England: Basil Blackwell.

Fletcher, W. H. (2004). Making the web more useful as a source for linguistic corpora. In U. Connor & T. A. Upton (Eds.), *Applied corpus linguistics. A multidimensional perspective* (pp. 191–205). Amsterdam/New York, The Netherlands/NY: Rodopi.

Fletcher, W. H. (2007). Concordancing the web: Promise and problems, tools and techniques. In C. Mair, C. F. Meyer, & N. Oostdijk (Eds.), *Corpus linguistics and the web* (pp. 25–45). Amsterdam/New York, The Netherlands/NY: Rodopi.

Fletcher, W. H. (2013). Corpus analysis of the World Wide Web. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–9). Hoboken, NJ: Wiley-Blackwell.

Flowerdew, L. (2004). The argument for using English specialized corpora to understand academic and professional language. In U. Connor & T. A. Upton (Eds.), *Discourse in the professions. Perspectives from corpus linguistics* (pp. 11–33). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Fontenelle, F. (1992). Collocation acquisition from a corpus or from a dictionary: A comparison. In H. Tommola, K. Varantola, T. Salmi-Tolonen, & J. Schopp (Eds.), *Proceedings of the 5th EURALEX International Congress* (pp. 221–228). Tampere, Finland: Tampereen YIiopisto. Retrieved from https://euralex.org/publications/collocation-acquisition-from-a-corpus-or-from-a-dictionary-a-comparison/

Foucault, M. (1972). *The archaeology of knowledge*. London, England: Tavistock.

FrameNet. (2020). Berkeley, CA: International Computer Science Institute.

Frankenberg-García, A. (2018). Investigating the collocations available to EAP writers. *Journal of English for Academic Purposes, 35*, 93–104. doi:10.1016/j.jeap.2018.07.003

Fuertes-Olivera, P. A. (2011). Michael Rundell. Macmillan collocations dictionary for learners of English. *Language Value, 3*(1), 153–161. doi:10.6035/LanguageV.2011.3.7

Fuertes-Olivera, P. A., & Tarp, S. (2014). *Theory and practice of specialised online dictionaries. Lexicography versus Terminography*. Berlin/Boston, Germany/MA: Walter de Gruyter.

Fuster-Márquez, M. (2014). Lexical bundles and phrase frames in the language of hotel websites. *English Text Construction, 7*(1), 84–121. doi:10.1075/etc.7.1.04fus

Fuster-Márquez, M. (2017). The discourse of US hotel websites: Variation through the interruptibility of lexical bundles. In M. Gotti, S. Maci, & M. Sala (Eds.*), Ways of seeing, ways of being: Representing the voices of tourism* (pp. 401–420). Bern/Berlin/Brussels/Frankfurt am Main/New York/Oxford/Wien, Switzerland/Germany/Belgium/Germany/NY/England/Austria: Peter Lang.

Fuster-Márquez, M., & Pennock-Speck, B. (2015). Target frames in British hotel websites. *International Journal of English Studies, 15*(1), 51–69. doi:10.6018/ijes/2015/1/213231

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning, 67*(S1), 155–179. doi:10.1111/lang.12225

Gabrielatos, C., & Marchi, A. (2012, September). *Keyness: Appropriate metrics and practical issues*. Paper presented at the CADS International Conference, Bologna (Italy).

Gambier, Y. (1991). Travail et vocabulaire spécialisés: Prolégomènes à une socio-terminologie. *Meta, Journal des Traducteurs, 36*(1), 8–15. Retrieved from https://id.erudit.org/iderudit/002795ar

Gambier, Y. (1993). Vers une histoire sociale de la terminologie. *Le Langage et L'homme, 28*(4), 233–246.

Gaudin, F. (1993). *Pour una socioterminologie: Des problèmes sémantiques aux pratiques institutionnelles*. Rouen, France: Publications de l'Université de Rouen.

Gaudin, F. (2003). *Socioterminologie: Une approche sociolinguistique de la terminologie*. Brussels, Belgium: De Boeck/Duculot.

Geeraerts, D. (2006). Introduction. A rough guide to Cognitive Linguistics. In D. Geeraerts (Ed.), *Cognitive linguistics: Basic readings* (pp. 1–28). Berlin/New York, Germany/NY: Mouton de Gruyter.

Gledhill, C. J. (2000). *Collocations in science writing*. Tübingen, Germany: Gunter Narr Berlag Tübingen.

Glynn, D., & Fischer, K. (Eds.). (2010). *Quantitative methods in cognitive semantics: Corpus-driven approaches*. Berlin/New York, Germany/NY: Walter de Gruyter.

Golberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago, IL: University of Chicago.

Goodwin, C. (1981). *Conversational organization: The interaction between speaker and hearer*. New York, NY: Academic Press.

Gotti, M. (2006). The language of tourism as specialized discourse. In O. Palusci & S. Francesconi (Eds.), *Translating tourism linguistic/cultural representations* (pp. 15–34). Trento, Italy: Editrice Università degli Studi di Trento.

Granger, S. (2012). Introduction: Electronic lexicography – from challenge to opportunity. In S. Granger & M. Paquot (Eds.), *Electronic phraseology* (pp. 1–11). Oxford, England: Oxford University Press.

Granger, S., & Paquot, M. (2008). Disentangling the phraseological web. In S. Granger & F. Meunier (Eds.), *Phraseology. An interdisciplinary perspective* (pp. 27–49). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Gray, B., & Biber, D. (2015). Phraseology. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 125–145). Cambridge, England: Cambridge University Press.

Greaves, C. (2009). *ConcGram 1.0: A phraseological search engine*. Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Greenbaum, S. (1974). Some verb-intensifier collocations in American and British English. *American Speech, 49*(1/2), 79−89. doi:10.2307/3087920

Gries, S. T. (2010). Useful statistics for corpus linguistics. In A. Sánchez & M. Almela (Eds.), *A mosaic of corpus linguistics: Selected approaches* (pp. 269−291). Frankfurt am Main, Germany: Peter Lang.

Gries, S. T. (2011). Methodological and interdisciplinary stance in Corpus Linguistics. In V. Viana, S. Zyngier, & G. Barnbrook (Eds.), *Perspectives on corpus linguistics*

(pp. 81–98). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Grieve, J. (2015). Dialect variation. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 363–380). Cambridge, England: Cambridge University Press.

Grishman, R. (1986). *Computational linguistics. An introduction*. Cambridge, England: Cambridge University Press.

Gupta, S. (2018). Methods and materials in a thesis. In S. C. Parija & V. Kate (Eds.), *Thesis writing for Master's and Ph.D. program* (pp. 141−151). Singapore, Singapore: Springer.

Gyllstad, H. (2007). *Testing English collocations: Developing receptive tests for use with advanced Swedish learners* (Doctoral thesis, Lund University, Lund, Sweden). Retrieved from https://lup.lub.lu.se/search/ws/files/5893676/2172422.pdf

Halliday, M. A. K. (1966). Lexis as a linguistic level. In C. Bazell, J. Catford, M. Halliday, & R. Robins (Eds.), *In memory of J. R. Firth* (pp. 148−162). London, England: Longman.

Halliday, M. A. K. (1978). *Language as social semiotic*. London, England: Edward Arnold.

Halliday, M. A. K. (1994). *An introduction to functional grammar. Second edition*. London, England: Edward Arnold.

Halliday, M. A. K. (1996). On grammar and grammatics. In R. Hasan, C. Cloran, & D. G. Butt (Eds.), *Functional descriptions: Theory in practice* (pp. 1−38). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Halliday, M. A. K. (2002). *On grammar*. London, England: Continuum.

Halliday, M. A. K. (2004). Lexicology. In M. A. K. Halliday, W. Teubert, C. Yallop, & A. Čermáková (Eds.), *Lexicology and corpus linguistics. An introduction* (pp. 1–122). London/New York, England/NY: Continuum.

Handl, S. (2009). Towards collocational webs for presenting collocations in learners' dictionaries. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in*

another language. Multiple interpretations (pp. 69–85). Hampshire, England: Palgrave Macmillan.

Hanks, P. (2012). Corpus evidence and electronic lexicography. In S. Granger & M. Paquot (Eds.), *Electronic lexicography* (pp. 57–82). Oxford, England: Oxford University Press.

HASK: Collocation Databases. (2020). Łódź, Poland: PELCRA Group.

Hausmann, F. J. (1979). Un dictionnaire des collocations est-il possible? *Travaux de Linguistique et de Litterature, XVII*(17), 187−195.

Hausmann, F. J. (1989). *Wörterbücher: Ein internationales Handbuch zur Lexikographie / Dictionaries: An international encyclopedia of lexicography / Dictionnaires: Encyclopédie internationale de lexicographie. Volume I.* Berlin/New York, Germany/NY: Walter de Gruyter.

Heid, U. (1994). On ways words work together – Topics in lexical combinatorics. In W. Martin, W. Meijs, M. Moerland, E. ten Pas, P. van Sterkenburg, & P. Vossen (Eds.), *Proceedings of the 6th EURALEX International Congress* (pp. 226–257). Amsterdam, The Netherlands: Euralex. Retrieved from https://euralex.org/publications/on-ways-words-work-together-topics-in-lexical-combinatorics/

Heid, U. (2008). Computational phraseology. An overview. In S. Granger & F. Meunier (Eds.), *Phraseology. An interdisciplinary perspective* (pp. 337–360). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Heid, U. (2009). Corpus linguistics and lexicography. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics. An international handbook. Volume I* (pp. 131–153). Berlin/New York, Germany/NY: Walter de Gruyter.

Heid, U., & Freibott, G. (1991). Collocations dans une base de données terminologique et lexicale. *Meta, Journal des Traducteurs, 36*(1), 77–91. doi:10.7202/002882AR

Heid, U., & Gouws, R. H. (2006). A model for a multifunctional dictionary of collocations. In E. Corino, C. Marello, & C. Onesti (Eds.), *Proceedings of the 12th EURALEX International Congress* (pp. 979–988). Alessandria, Italy: Edizioni

dell'Orso. Retrieved from https://euralex.org/publications/a-model-for-a-multifunctional-dictionary-of-collocations/

Hill, J. (2000). Revising priorities: From grammatical failure to collocational success. In M. Lewis (Ed.), *Teaching collocation: Further developments in the lexical approach* (pp. 47–69). Hove, England: Language Teaching Publications.

Hobbs, J. R., & Bear, J. (1994). Two principles of parse preference. In A. Zampolli, N. Calzolari, & M. Palmer (Eds.), *Current issues in computational linguistics: In honour of Don Walker* (pp. 503–512). Pisa, Italy: Springer-Science+Business Media Dordrecht.

Hoffmann, L. (1979). Towards a theory of LSP. Elements of a methodology of LSP analysis. *Fachsprache, 1*, 1–2, 12–17.

Howarth, P. A. (1996). *Phraseology in English academic writing. Some Implications for language learning and dictionary making*. Tübingen, Germany: Niemeyer.

Howarth, P. A. (1998). The phraseology of learners' academic writing. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 161–186). Oxford, England: Clarendon Press.

Hudson, R. (1988). The linguistic foundations for lexical research and dictionary design. *International Journal of Lexicography, 1*(4), 287–312. doi:10.1093/ijl/1.4.287

Hunston, S. (2009). Collection strategies and design decisions. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics. An international handbook. Volume I* (pp. 154–168). Berlin/New York, Germany/NY: Walter de Gruyter.

Hunston, S., & Francis, G. (2000). *Pattern grammar. A corpus-driven approach to the lexical grammar of English*. Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

IATE. Interactive Terminology for Europe (2020). Luxembourg, Grand Duchy of Luxembourg: Translation Centre for the Bodies of the European Union.

Ibarretxe-Antuñano, I. (2004). What is cognitive linguistics? A new framework for the study of Basque. *Cahiers, 10*(2), 3–30. Retrieved from http://www.afls.net/cahiers/cover10.2.php

Ide, N., Bonhomme, P., & Romary, L. (2000). XCES: An XML-based encoding standard for linguistic corpora. *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC'00)* (pp. 1–6). Retrieved from https://www.aclweb.org/anthology/L00-1127/

Jackson, H. (1988). *Words and their meaning*. London, England: Longman.

Jackson, H., & Amvela, E. Z. (2007). *Words, meaning and vocabulary. An introduction to modern English lexicology. Second edition*. London/New York, England/NY: Continuum.

Jakobson, R. (1960). Linguistics and poetics. In T. A. Sebeok (Ed.), *Style in language* (pp. 350–377). Cambridge, MA: The Massachusetts Institute of Technology.

Jenkins, J. J. (1970). The 1952 Minnesota word association norms. In L. Postman & G. Keppel (Eds.), *Norms of word associations* (pp. 1–38). New York, NY: Academic Press.

Jiménez-Navarro, E. L. (2019). Nominal collocations in scientific English: A frame-semantic approach. In G. Corpas Pastor & R. Mitkov (Eds.), *Lecture Notes in Artificial Intelligence. Volume 11755. Computational and corpus-based phraseology* (pp. 187–199). Cham, Switzerland: Springer.

Johansson, S. (2009). Some aspects of the development of corpus linguistics in the 1970s and 1980s. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics. An international handbook. Volume I* (pp. 33–53). Berlin/New York, Germany/NY: Walter de Gruyter.

Juilland, A. G., Brodin, D. R., & Davidovitch, C. (1971). *Frequency dictionary of French words.* The Hague, The Netherlands: Mouton de Gruyter.

Karlsson, F. (2009). Early generative linguistics and empirical methodology. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics. An international handbook. Volume I* (pp. 14–32). Berlin/New York, Germany/NY: Walter de Gruyter.

Kay, M. (2003). Introduction. In R. Mitkov (Ed.), *The Oxford handbook of computational linguistics* (pp. xvii–xx). Oxford, England: Oxford University Press.

Kay, M. (2005). A life of language. *Computational Linguistics, 31*(4): 425–438. doi:10.1162/089120105775299159

Keller, E. (1979). Gambits: Conversational strategy signals. *Journal of Pragmatics, 3*(3-4), 219−238. doi:10.1016/0378-2166(79)90032-8

Kenning, M.-M. (2010). What are parallel and comparable corpora and how can we use them? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 487–500). London/New York, England/NY: Routledge.

Kestemont, M., De Pauw, G., Van Nie, R., & Daelemans, V. (2017). Lemmatization for variation-rich languages using deep learning. *Digital Scholarship in the Humanities, 32*(4), 797–815. doi:10.1093/llc/fqw034

Kheirzadeha, S., & Marandi, S. S. (2014). Concordancing as a tool in learning collocations: The case of Iranian EFL learners. *Procedia – Social and Behavioral Sciences, 98*, 940–949. doi: 10.1016/j.sbspro.2014.03.503

Kiefer, F., & Van Sterkenburg, P. (2003). Design and production of monolingual dictionaries. In. P. van Sterkenburg (Ed.), *A practical guide to lexicography* (pp. 350–365). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Kilgarriff, A. (2009). Simple maths for keywords. In M. Mahlberg, V. González-Díaz, & C. Smith (Eds.), *Proceedings of the Corpus Linguistics Conference 2009 (CL2009)* (pp. 1–6). Liverpool, England: University of Liverpool.

Kilgarriff, A., Kovář, V., Krek, S., Srdanovic, I., & Tiberius, C. (2010). A quantitative evaluation of word sketches. In A. Dykstra & T. Schoonheim (Eds.), *Proceedings of the 14th EURALEX International Congress* (pp. 372–379). Leeuwarden, The Netherlands: Fryske Akademy – Afûk. Retrieved from https://euralex.org/publications/a-quantitative-evaluation-of-word-sketches/

Kilgarriff, A., & Rundell, M. (2002). Lexical profiling software and its lexicographic applications: A case study. In A. Braasch & C. Povlsen (Eds), *Proceedings of the 10th EURALEX International Congress* (pp. 807–818). Copenhagen, Denmark: Center for Sprogteknologi. Retrieved from https://euralex.org/publications/lexical-profiling-software-and-its-lexicographic-applications-a-case-study/

Kilgarriff, A., Rychlý, P., Smrž, P., & Tugwell, D. (2004). The Sketch Engine. In G. Williams & S. Vessier (Eds.), *Proceedings of the 11th EURALEX International*

Congress (pp. 105–115). Lorient, France: Université de Bretagne-Sud, Faculté des lettres et des sciences humaines. Retrieved from https://euralex.org/publications/the-sketch-engine/

Kilgarriff, A., et al. (2014). The Sketch Engine: Ten years on. *Lexicography, 1*(1), 7–36. doi:10.1007/s40607-014-0009-9

Kjellmer, G. (1984). Some thoughts on collocational distinctiveness. In J. Aarts & W. Meijs (Eds.), *Corpus linguistics* (pp. 163−171). Amsterdam, The Netherlands: Rodopi.

Kjellmer, G. (1987). Aspects of English collocations. In W. Meijs (Ed.), *Corpus linguistics and beyond. Proceedings of the 7th International Conference on English Language Research on Computerized Corpora* (pp. 133–140). Amsterdam, The Netherlands: Rodopi.

Kjellmer, G. (1994). *A dictionary of English collocations: Based on the Brown Corpus*. Oxford, England: Clarendon Press.

Klappenbach, R. (1968). Probleme der Phraseologie. *Wissenschaftliche Zeitschrift der Karl-Marx-Universität, 17*(2), 221–227.

Kocourek, R. (1982). *La langue française de la technique et de la science*. Wiesbaden, Germany: Brandstetter.

Koike, K. (2001). *Colocaciones léxicas en el español actual: Estudio formal y léxico-semántico.* Alcalá de Henares, Spain: Universidad de Alcalá de Henares.

Koprowski, M. (2005). Investigating the usefulness of lexical phrases in contemporary coursebooks. *ELT Journal, 59*(4), 322–332. doi:10.1093/elt/cci061

Kučera, H. (1992). The odd couple: The linguist and the software engineer. The struggle for high quality computerized language aids. In J. Svartvik (Ed.), *Directions in corpus linguistics* (pp. 401–420). Berlin/New York, Germany/NY: Mouton de Gruyter.

Lakoff, G. (1987). *Women, fire and dangerous things. What categories reveal about the mind*. Chicago, IL: The University of Chicago Press.

Lakoff, G., & Johnson, M. (1980). Metaphors we live by. Chicago, IL: The University of Chicago Press.

Langacker, R. (1987). *Foundations of cognitive grammar. Volume I. Theoretical prerequisites*. Stanford, CA: Stanford University Press.

Langacker, R. (1991). *Foundations of cognitive grammar. Volume II. Descriptive application*. Stanford, CA: Stanford University Press.

Langlotz, A. (2001, April). *Cognitive principles of idiom variation*. Paper presented at the 4th International Symposium on Phraseology, Rome (Italy).

Lavid, J. (2005). *Lenguaje y nuevas tecnologías. Nuevas perspectivas, métodos y herramientas para el lingüista del siglo XXI*. Madrid, Spain: Cátedra.

Lavid, J. (2007). To the memory of John Sinclair, professor of modern English language. *Estudios Ingleses de la Universidad Complutense*, *15*, 9−12. Retrieved from https://dialnet.unirioja.es/servlet/articulo?codigo=2522905

Lea, D. (2007). Making a collocations dictionary. *Zeitschrift für Anglistik und Amerikanistik, 55*(3), 261−271. doi:10.1515/zaa.2007.55.3.261

Lea, J. (1988). *Tourism and development in the third world*. London/New York, England/NY: Routledge.

Lee, D. Y. W. (2010). What corpora are available? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 107–121). London/New York, England/NY: Routledge.

Leech, G. (1991). The state of the art in corpus linguistics. In K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics. Studies in honour of Jan Svartvik* (pp. 8−29). New York/London, NY/England: Routledge.

Leech, G. (1993). Corpus annotation schemes. *Literary and Linguistic Computing, 8*(4), 275–281. doi:10.1093/llc/8.4.275

Leech, G. (2011). Principles and applications of corpus linguistics. In V. Viana, S. Zyngier, & G. Barnbrook (Eds.), *Perspectives on corpus linguistics* (pp. 155−170). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Leech, G. (2015). Descriptive grammar. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 146–160). Cambridge, England: Cambridge University Press.

Lehmberg, T., & Wörner, K. (2009). Annotation standards. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics. An international handbook. Volume I* (pp. 484–153). Berlin/New York, Germany/NY: Walter de Gruyter.

León Araúz, P. (2009). *Representación multidimensional del conocimiento especializado: El uso de marcos desde la macroestructura hasta la microestructura* (Doctoral thesis, Universidad de Granada, Granada, Spain). Retrieved from http://lexicon.ugr.es/pub/leonphd

Lew, R., Frankenberg-García, A., Rees, G. P., Roberts, J. C., & Sharma, N. (2018). ColloCaid: A real-time tool to help academic writers with English collocations. In J. Čibej, V. Gorjanc, I. Kosem, & S. Krek (Eds.), *Proceedings of the 18th EURALEX International Congress* (pp. 247–254). Ljubljana, Slovenia: Ljubljana University Press. Retrieved from https://euralex.org/publications/collocaid-a-real-time-tool-to-help-academic-writers-with-english-collocations/

Lewis, M. (1993). *The lexical approach: The state of ELT and the way forward*. Hove, England: Language Teaching Publications.

Lewis, M. (Ed.). (2000). *Teaching collocation: Further developments in the lexical approach*. Hove, England: Language Teaching Publications.

Lindquist, H. (2009). *Corpus linguistics and the description of English*. Edinburgh, Scotland: Edinburgh University Press.

Lipka, L. (1974). Probleme der Analyse englischer Idioms aus struktureller und generativer Sicht. *Linguistik und Didaktik, 20*, 274–285.

Longman Dictionary of Contemporary English. (2020). London, England: Pearson.

López Rodríguez, C. I. (2007). Understanding scientific communication through the extraction of the conceptual and rhetorical information codified by verbs. *Terminology, 13*(1), 61–84. doi:10.1075/term.13.1.04lop

Lorente Casafont, M. (2002). Terminología y fraseología especializada: Del léxico a la sintaxis. In G. Guerrero Ramos & M. F. Pérez Lagos (Eds.), *Panorama actual de la terminología* (pp. 159–179). Granada, Spain: Comares.

Lorente Casafont, M. (2007). Les unitats lèxiques verbals dels textos especialitzats. Redefinició d'una proposta de classificació. In M. Lorente Casafont, R. Estopà

Bagot, J. Freixa, J. Martí, & C. Tebé (Eds.), *Estudis de lingüística y de lingüística aplicada en honor de M. Teresa Cabré Castellví* (pp. 365–380). Barcelona, Spain: IULA.

Lorente Casafont, M., Martínez-Salom, A., Santamaría, I., & Vargas Sierra, C. (2017). Specialized collocations in specialized dictionaries. In S. Torner Castells & E. Bernal (Eds.), *Collocations and other lexical combinations in Spanish. Theoretical, lexicographical and applied perspectives* (pp. 200–222). London/New York, England/NY: Routledge.

Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 157–176). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

L'Homme, M.-C. (1998). Le statut du verbe en langue de spécialité et sa description lexicographique. *Cahiers de Lexicographie, 73*(2), 61–84. doi:10.15122/ISBN.978-2-8124-4325-1.P.0065

L'Homme, M.-C. (2003). Capturing the lexical structure in special subject fields with verbs and verbal derivatives. A model for specialized lexicography. *International Journal of Lexicography, 16*(4), 403–422. doi:10.1093/ijl/16.4.403

L'Homme, M.-C. (2004). A lexico-semantic approach to the structuring of terminology. *Proceedings of CompuTerm 2004: 3rd International Workshop on Computational Terminology*, 7–13. Retrieved from https://www.aclweb.org/anthology/W04-1801

L'Homme, M.-C. (2009a). A methodology for describing collocations in a specialized dictionary. In S. Nielsen & S. Tarp (Eds.), *Lexicography in the 21st century: In honor of Henning Bergenholtz* (pp. 237–256). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

L'Homme, M.-C. (2009b). *Dictionary compiled by the ECLECTIK team – Meaning-Text Linguistics Observatory (OLST) – University of Montreal*. Retrieved from http://olst.ling.umontreal.ca/dicoinfo/DiCoInfo%20manual_Eng.pdf

L’Homme, M.-C. (2010). Designing terminological dictionaries for learners based on lexical semantics: The representation of *actants*. In P. A. Fuertes-Olivera (Ed.), *Specialized dictionaries for learners* (pp. 141–153). Berlin/New York, Germany/NY: Walter de Gruyter.

L’Homme, M.-C. (2012). Le verbe terminologique: Un portrait de travaux récents. *SHS Web of Conferences, 1*(2261-2424), 93–107. doi:10.1051/shsconf/20120100340

L’Homme, M.-C. (2014). Dictionaries and the digital revolution: A focus on users and lexical databases. *International Journal of Lexicography, 27*(4), 331–340. doi:10.1093/ijl/ecu023

L’Homme, M.-C. (2017). Combinatoire spécialisée: Trois perspectives et des enseignements pour la terminologie. *TTR, 30*(1-2), 215–241. doi:10.7202/1060025ar

L’Homme, M.-C. (2020). *Lexical semantics for terminology*. Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

L’Homme, M.-C., & Bertrand, C. (2000). Specialized lexical combinations: Should they be described as collocations or in terms of selectional restrictions? In U. Heid, S. Evert, E. Lehmann, & C. Rohrer (Eds.), *Proceedings of the 9th EURALEX International Congress* (pp. 497–506). Sttutgart, Germany: Euralex. Retrieved from https://euralex.org/publications/specialized-lexical-combinations-should-they-be-described-as-collocations-or-in-terms-of-selectional-restrictions/

L’Homme, M.-C., & Laneville, M.-E. (2009). *Le dictionnaire fondamental de l’environnement – Dictionnaire élaboré par l’équipe ÉCLECTIK – Observatoire de Linguistique Sens-Texte (OLST).* Retrieved from http://olst.ling.umontreal.ca/dicoenviro/manuel-DiCoEnviro.pdf

L’Homme, M.-C., & Leroyer, P. (2009). Combining the semantics of collocations with situation-driven search paths in specialized dictionaries. *Terminology, 15*(2), 258–283. doi:10.1075/term.15.2.05lho

L’Homme, M.-C., Robichaud, B., & Subirats Rüggeberg, C. (2014). Discovering frames in specialized domains. In N. Calzolari et al. (Eds.), *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC'14)* (pp. 1364–1371). Reykjavik, Iceland: European Language Resources Association

(ELRA). Retrieved from http://www.lrec-conf.org/proceedings/lrec2014/index.html

Maci, S. M. (2018). An introduction to English tourism discourse. *Sociolinguistica, 32*(1), 25–42. doi:10.1515/soci-2018-0004

Macmillan English Dictionary. (2020). London, England: Macmillan Education Limited.

Manca, E. (2008). From phraseology to culture: Qualifying adjectives in the language of tourism. *International Journal of Corpus Linguistics, 13*(3), 368–385. doi:10.1075/ijcl.13.3.07man

Mănescu, O. M. (2020). The complexity of the language of tourism. *British and American Studies, 26*, 221–229. Retrieved from https://www.ceeol.com/search/article-detail?id=871104

Mart, C. T. (2012). How to teach phrasal verbs. *English Language Teaching, 5*(6), 114–118. doi:10.5539/elt.v5n6p114

Martin, W. (2008). A unified approach to semantic frames and collocational patterns. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective* (pp. 51–65). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Maynard, D., & Ananiadou, S. (2001). Term extraction using a similarity-based approach. In C. Jacquemin & M.-C. L'Homme (Eds.), *Recent advances in computational linguistics* (pp. 261–278). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

McCarthy, M., & O'Keeffe, A. (2010). Historical perspective. What are corpora and how have they evolved? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 3–13). London/New York, England/NY: Routledge.

McEnery, T. (2003). Corpus linguistics. In R. Mitkov (Ed.), *The Oxford handbook of computational linguistics* (pp. 448–463). Oxford, England: Oxford University Press.

McEnery, T., & Wilson, A. (2001). *Corpus linguistics. An introduction*. Edinburgh, Scotland: Edinburgh University Press.

McGee, I. (2012). Collocation dictionaries as inductive learning resources in data-driven learning – An analysis and evaluation. *International Journal of Lexicography, 25*(3), 319–361. doi:10.1093/ijl/ecr040

McIntosh, C., Francis, B., & Poole, R. (2009). *Oxford collocations dictionary for students of English*. Oxford, England: Oxford University Press.

Medlik, S. (2003). *Dictionary of travel, tourism and hospitality. Third edition*. London/New York, England/NY: Routledge.

Mehler, A. (2009). Large text networks as an object of corpus linguistic studies. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics. An international handbook. Volume I* (pp. 328–382). Berlin/New York, Germany/NY: Walter de Gruyter

Mel'čuk, I. A. (1981). Meaning-text models: A recent trend in Soviet linguistics. *Annual Review of Anthropology, 10*, 27−62. doi:10.1146/annurev.an.10.100181.000331

Mel'čuk, I. A. (1988). *Dependency syntax: Theory and practice*. Albany, NY: SUNY Press.

Mel'čuk, I. A. (1996). Lexical functions: A tool for the description of lexical relations in a lexicon. In L. Wanner (Ed.), *Lexical functions in lexicography and natural language processing* (pp. 37–102). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Mel'čuk, I. A. (1998). Collocations and lexical functions. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 23–53). Oxford, England: Clarendon Press.

Mel'čuk, I. A. (2006). Explanatory combinatorial dictionary. In G. Sica (Ed.), *Open problems in linguistics and lexicography* (pp. 225–355). Milan, Italy: Polimetrica.

Mel'čuk, I. A. (2012). Phraseology in the language, in the dictionary, and in the computer. *Yearbook of Phraseology, 3*(1), 31–56. doi:10.1515/phras-2012-0003

Mel'čuk, I. A., & Polguère, A. (1987). A formal lexicon in the Meaning-Text Theory (or how to do lexica with words). *Computational Linguistics, 13*(3-4), 261−275. Retrieved from https://www.aclweb.org/anthology/J87-3006.pdf

Meunier, F., & Granger, S. (Eds.). (2008). *Phraseology in foreign language learning and teaching*. Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Meyer, C. F. (2002). *English corpus linguistics. An introduction*. Cambridge, England: Cambridge University Press.

Meyer, C. F. (2009). Pre-electronic corpora. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics. An international handbook. Volume I* (pp. 1–14). Berlin/New York, Germany/NY: Walter de Gruyter.

Meyer, I., & Mackintosh, K. (2000). When terms move into our everyday lives: An overview of de-terminologization. *Terminology, 6*(1), 111–138. doi:10.1075/term.6.1.07mey

Miguel Aparicio, E. de. (Ed.). (2009). *Panorama de la lexicología*. Barcelona, Spain: Ariel.

Milićević, J. (2006). A short guide to the Meaning-Text linguistic theory. *Journal of Koralex, 8*, 187–233. Retrieved from http://olst.ling.umontreal.ca/pdf/IntroMTTJM.pdf

Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 211–277). New York, NY: McGraw-Hill Book.

Mitchell, T. F. (1971). Linguistics 'goings-on': Collocations and other lexical matters arising on the syntactic record. *Archivum Linguisticum, 2*, 35−69.

Mitchell, T. F. (1975). *Principles of Firthian linguistics*. London, England: Longman.

Mitkov, R. (Ed.). (2017). *Lecture Notes in Artificial Intelligence. Volume 10596. Computational and corpus-based phraseology*. Cham, Switzerland: Springer.

Mitkov, R., Monti, J., Corpas Pastor, G., & Seretan, V. (Eds.). (2018). *Multiword units in machine translation and translation technology*. Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Mittmann, B. (1999). The treatment of collocations in OALD5, LDOCE3, COBUILD2 and CIDE. In T. Herbst & K. Popp (Eds.), *The perfect learners' dictionary* (pp. 101–111). Tübingen, Germany: Max Niemeyer Verlag.

Moon, R. (2008). Dictionaries and collocation. In S. Granger & F. Meunier (Eds.), *Phraseology. An interdisciplinary perspective* (pp. 313–336). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Moreno Jaén, M. (2009). *Recopilación, desarrollo pedagógico y evaluación de un banco de colocaciones frecuentes de la lengua inglesa a través de la lingüística de corpus y computacional* (Doctoral thesis, Universidad de Granada, Granada, Spain). Retrieved from digibug.ugr.es/handle/10481/2311

Muller, T. E., & Cleaver, M. (2000). Targeting the CANZUS baby boomer explorer and adventurer segments. *Journal of Vacation Marketing, 6*(2), 154–169. doi:10.1177/135676670000600206

Multilingual Dictionary of Tourism (2014). Valencia, Spain: COMETVAL Group.

Murison-Bowie, S. (1993). *MicroConcord manual. An introduction to the practices and principles of concordancing in language teaching*. Oxford, England: Oxford University Press.

Murphy, G. L., & Medin, D. L. (1985). The role theories in conceptual coherence. *Psychological Review, 92*(3), 289–316. doi:10.1037/0033-295X.92.3.289

Nelson, M. (2010). Building a written corpus. What are the basics? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 53–65). London/New York, England/NY: Routledge.

Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Nigro, M. G. (2006). The language of tourism as LSP? A corpus-based study of the discourse of guidebooks. In H. Picht (Ed.), *Modern approaches to terminological theories and applications* (pp. 187–197). Bern/Berlin/Brussels/Frankfurt am Main/New York/Oxford/Wien, Switzerland/Germany/Belgium/Germany/NY/England/Austria: Peter Lang.

Nivre, J. (2009). Treebanks. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics. An international handbook. Volume I* (pp. 225–241). Berlin/New York, Germany/NY: Walter de Gruyter.

Nord, C. (2006). Translating a purposeful activity: A prospective approach. *TEFLIN Journal, 17*(2), 131–143. Retrieved from https://core.ac.uk/download/pdf/233167343.pdf

Nuray, R., & Can, F. (2006). Automatic ranking of information retrieval systems using data fusion. *Information Processing & Management, 42*(3), 595–614. doi:10.1016/j.ipm.2005.03.023

Nurmukhamedov, U. (2011). Macmillan collocations dictionary for learners of English. *ELT Journal, 65*(1), 96–99. doi:10.1093/elt/ccq083

Olivera Betrán, J., & Olivera Betrán, A. (1995). Propuesta de una clasificación taxonómica de las actividades físicas de aventura en la naturaleza. Marco conceptual y análisis de los criterios elegidos. *Apuntes: Educación Física y Deportes, 41*, 108–123. Retrieved from https://hemeroteca.revista-apunts.com/apunts/articulos/41/es/041_108-123_es.pdf

Omazić, M. (2005). Cognitive linguistic theories in phraseology. *Jezikoslovlje, 6*(1), 37–56. Retrieved from https://hrcak.srce.hr/index.php?show=clanak&id_clanak_jezik=48921

Orăsan, C., Ha, L. A., Evans, R., Hasler, L., & Mitkov, R. (2007). Corpora for computational linguistics. *Ilha do Desterro. A Journal of English Language, Literatures in English and Cultural Studies, 52*, 65–101. doi:10.5007/%25x

Orliac, B. (2008). Extracting specialized collocations using lexical functions. In S. Granger & F. Meunier (Eds.), *Phraseology. An interdisciplinary perspective* (pp. 377–390). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Oxford Advanced Learner's Dictionary. (2020). Oxford, England: Oxford University Press.

O'Halloran, K. (2010). How to use corpus linguistics in the study of media discourse. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 563–577). London, England: Routledge.

Page, S. (2007). *Tourism management. Managing for change. Second edition*. Oxford/Burlington, England/MA: Elsevier.

Palmer, H. E. (1933). *Second interim report on English collocations*. Tokyo, Japan: Kaitakusha.

Paquot, M. (2015). Lexicography and phraseology. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 460–477). Cambridge, England: Cambridge University Press.

Patiño García, P. (2014). Towards a definition of specialized collocation. In G. Quiroz Herrera & P. Patiño García (Eds.), *LSP in Colombia: Advances and challenges* (pp. 119–133). Bern/Berlin/Brussels/Frankfurt am Main/New York/Oxford/Wien, Switzerland/Germany/Belgium/Germany/NY/England/Austria: Peter Lang.

Pauw, G. de, & Schryver, G.-M. de. (2008). Improving the computational morphological analysis of a Swahili corpus for lexicographic purposes. *Lexikos, 18*, 303–318. doi:10.5788/18-0-488

Pavel, S. (1993). Vers une méthode de recherche phraséologique en langue de spécialité. *L'Actualité Terminologique, 26*(2), 9–13.

Pearce, D. (2001, June). *Synonymy in collocation extraction*. Paper presented at the 2nd Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL-2001), Pittsburgh, PA (United States).

Pearson, J. (1998). *Terms in context*. Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Peñalver Torres, M. T. (2004). El turismo activo como alternativa y complemento al modelo turístico en la Región de Murcia. *Cuadernos de Turismo, 14*, 179–215. Retrieved from https://digitum.um.es/digitum/handle/10201/13029

Pérez Hernández, M. C. (2002). Explotación de los córpora textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento. *Estudios de Lingüística del Español (ELiEs), 18*. Retrieved from http://elies.rediris.es/elies18/

Peruzzo, K. (2014). Term extraction and management based on event templates. An empirical study on an EU corpus. *Terminology, 20*(2), 151–170. doi:10.1075/term.20.2.02per

Pęzik, P. (2014). Graph-based analysis of collocational profiles. In V. Jesenšek & P. Grzybek (Eds.), *Phraseologie im Wörterbuch und Korpus* (pp. 227–243). Maribor/Bielsko‑Biała/Budapest/Kansas/Prague, Slovenia/Poland/Hungary/MO/Czech Republic: Filozofska fakuteta.

Pęzik, P. (2018). *Facets of prefabrication. Perspectives on modelling and detecting phraseological units*. Łódź, Poland: Łódź University Press.

Philip, G. (2011). *Colouring meaning. Collocation and connotation in figurative language.* Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Piccioni, S., & Pontrandolfo, G. (2019). La construcción del espacio turístico a través de la fraseología metafórica. *Linguistik Online, 94*(1/19), 137–153. doi:10.13092/lo.94.5436

Picht, H. (1987). Terms and their LSP environment – LSP phraseology. *Meta, Journal des Traducteurs, 32*(2), 149–155. Retrieved from https://www.academia.edu/20914152/Terms_and_their_LSP_Environment_LSP_Phraseology

Picht, H., & Draskau, J. (1985). *Terminology: An introduction*. Guildford, England: The University of Surrey.

Pimentel, J. (2014). Using frame semantics to build a bilingual lexical resource on legal terminology. In H. J. Kockaert & F. Steurs (Eds.), *Handbook of terminology. Volume I* (pp. 427–450). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Plisson, J., Lavrac, N., & Mladenic, D. (2004). A rule-based approach to word lemmatization. *Proceedings of the 7th International Multi-Conference Information Society IS 2004, Volume I,* 83–86. Retrieved from https://ailab.ijs.si/dunja/SiKDD2004/Papers/Pillson-Lematization.pdf

Pritchard, A., & Jaworski, A. (2005). Introduction. Discourse, communication and tourism dialogues. In A. Jaworski & A. Pritchard (Eds.), *Discourse, communication and tourism* (pp. 1–16). Clevedon/Buffalo/Toronto, England/NY/Canada: Channel View Publications.

Pustejovsky, J. (1995). *The generative lexicon*. Cambridge, MA: MIT Press.

Raţă, G., & Petroman, I. (2012). English borrowings in the Romanian of tourism (sites of travel agencies). In G. Raţă, I. Petroman, & C. Petroman (Eds.), *The English of tourism* (pp. 49–54). Newcastle upon Tyne, England: Cambridge Scholars Publishing.

Raţă, G., Petroman, C., & Petroman, I. (2012). English borrowings in the Romanian of agritourism (Internet sites). In G. Raţă, I. Petroman, & C. Petroman (Eds.), *The English of tourism* (pp. 55–58). Newcastle upon Tyne, England: Cambridge Scholars Publishing.

Rayson, P. (2001, May). *WMATRIX: A web-based corpus processing environment*. Software demonstration presented at ICAME 2001 Conference, Louvain-la-Neuve (Belgium).

Rayson, P. (2015). Computational tools and methods for corpus compilation. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 11–31). Cambridge, England: Cambridge University Press.

Reily Collins, V. (2008). *Tourism Society's Dictionary for the tourism industry*. Wallingford, England: CAB International.

Reimerink, A., & León Arauz, P. (2017). Predicate-argument analysis to build a phraseology module and to increase conceptual relation expressiveness. In R. Mitkov (Ed.), *Lecture Notes in Artificial Intelligence. Volume 10596. Computational and corpus-based phraseology* (pp. 176–190). Cham, Switzerland: Springer.

Reppen, R. (2010). Building a corpus: What are the key considerations? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 31–37). London/New York, England/NY: Routledge.

Rey, A. (1976). Néologisme: Un pseudo-concept? *Cahiers de Lexicologie 28*(1), 3–17. doi:10.15122/isbn.978-2-8124-4281-0.p.0005

Rogerson, C. M. (2007). The challenges of developing adventure tourism in South Africa. *Africa Insights, 37*(2), 228–244. doi:10.4314/ai.v37i2.22506

Rojas García, J., & Faber Benítez, P. (2019). Extraction of terms for the construction of semantic frames for named bays. *Argentinian Journal of Applied Linguistics, 7*(1), 27–57. Retrieved from http://lexicon.ugr.es/pdf/Garcia-Rojas8c4_2019.pdf

Romaine, S. (2009). Corpus linguistics and sociolinguistics. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics. An international handbook. Volume I* (pp. 96–111). Berlin/New York, Germany/NY: Walter de Gruyter.

Römer, U. (2009). Corpora and language teaching. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics. An international handbook. Volume I* (pp. 112–131). Berlin/New York, Germany/NY: Walter de Gruyter.

Romero Aguilera, L. (2015). *El tratamiento de las colocaciones en la lexicografía monolingüe general del español (siglos XVII-XXI)* (Doctoral thesis, Universitat d'Alacant, Alicante, Spain). Retrieved from http://www.cervantesvirtual.com/obra/el-tratamiento-de-las-colocaciones-en-lexicografia-monolingue-general-del-espanol-siglos-xvii-xxi/

Rondeau, G. (1983). *Introduction à la terminologie*. Quebec, Canada: Gaëtan Morin.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111–144). New York, NY: Academic Press.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*(3), 192–233. doi:10.1037/0096-3445.104.3.192

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rühlemann, C. (2010). What can a corpus tell us about pragmatics? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 288–301). London, England: Routledge.

Ruiz-Garrido, M. F., & Saorín-Iborra, A. M. (2013). Language for tourism. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–6). Hoboken, NJ: Wiley-Blackwell.

Rundell, M. (2010). *Macmillan collocations dictionary for learners of English*. Oxford, England: Macmillan Publishers Ltd.

Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., Baker, C. F., & Scheffczyk, J. (2016). *FrameNet II: Extended theory and practice*. Retrieved from https://framenet.icsi.berkeley.edu/fndrupal/the_book

Rychlý, P. (2008). A lexicographer-friendly association score. In P. Sojka & A. Horák (Eds.), *Proceedings of Recent Advances in Slavonic Natural Language Processing (RASLAN 2008)* (pp. 6–9). Brno, Czech Republic Masaryk University. Retrieved from https://www.fi.muni.cz/usr/sojka/download/raslan2008/13.pdf

Sag, I. A., Baldwin, T., Bond, F., Copestake, A., & Flickinger, D. (2002). Multiword Expressions: A pain in the neck for NLP. In A. F. Gelbukh (Ed.), *Lecture Notes in Computer Science. Volume 2276. Computational Linguistics and Intelligent Text Processing* (pp. 1–15). Cham, Switzerland: Springer.

Sager, J. C. (1990). *A practical course in terminology processing*. Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Sager, J. C., Dungworth, D., & McDonald, P. F. (1980). *English special languages. Principles and practice in science and technology*. Wiesbaden, Germany: Brandstetter.

Sanz Vicente, M. L. (2011). *Análisis contrastivo de la terminología de la teledetección. La traducción de compuestos sintagmáticos nominales del inglés al español* (Unpublished doctoral dissertation). Universidad de Salamanca, Salamanca, Spain.

Scalise, S., & Vogel, I. (2010). Why compounding? In S. Scalise & I. Vogel (Eds.), *Cross-disciplinary issues in compounding* (pp. 1–18). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Schank, R., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding. An inquiry into human knowledge structures*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schneider, F. (1998). *Studien zur kontextuellen Fachlexikographie. Das deutsch-französische Wörterbuch der Rechnungslegung*. Tübingen, Germany: Niemeyer Verlag.

Schryver, G.-M. de. (2002). Web for/as Corpus: A perspective for the African languages. *Nordic Journal of African Studies, 11*(2), 266–282. Retrieved from http://www.njas.helsinki.fi/pdf-files/vol11num2/schryver.pdf

Scott, M. (2001). Comparing corpora and identifying key words, collocations, and frequency distributions through the WordSmith Tools suite of computer programs. In M. Ghadessy, A. Henry, & R. L. Roseberry (Eds.), *Small corpus studies and ELT: Theory and practice* (pp. 47–67). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Scott, M. (2011). The technological aspect of corpus Linguistics. In V. Viana, S. Zyngier, & G. Barnbrook (Eds.), *Perspectives on corpus linguistics* (pp. 213–219). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Scott, M. (2016). *WordSmith Tools version 7*. Stroud, England: Lexical Analysis Software.

Seretan, V. (2011). *Syntax-based collocation extraction*. Dordrecht/Heidelberg/London/New York, The Netherlands/Germany/NY/England: Springer.

Siegel, E., & Retter, A. (2015). *eXist. A NoSQL document database and application platform*. Sebastopol, CA: O'Reilly.

Siepmann, D. (2005). Collocation, colligation and encoding dictionaries. Part II: Lexicographical aspects. *International Journal of Lexicography, 19*(1), 1–39. doi:10.1093/ijl/eci051

Sinclair, J. (Ed.). (1987). *Looking up. An account of the COBUILD project in lexical computing*. London/Glasgow, England/Scotland: Collins.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford, England: Oxford University Press.

Sinclair, J. (1998). The lexical item. In E. Weigand (Ed.), *Contrastive lexical semantics* (pp. 1−24). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Sinclair, J. (2000). Lexical grammar. *Naujoji Metodologija*, *24*, 191–203. Retrieved from http://donelaitis.vdu.lt/publikacijos/sinclair.pdf

Sinclair, J. (2003a). Corpora for lexicography. In. P. van Sterkenburg (Ed.), *A practical guide to lexicography* (pp. 167–178). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Sinclair, J. (2003b). Corpora processing. In. P. van Sterkenburg (Ed.), *A practical guide to lexicography* (pp. 179–193). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Sinclair, J. (2004). The search for units of meaning. In J. Sinclair (Ed.), *Trust the text: Language, corpus and discourse* (pp. 24– 48). London/New York, England/NY: Routledge. (Reprinted from *Textus, 9*(1), 75–106, 1996).

Sketch Engine [Computer software]. Brno, Czech Republic: Lexical Computing Limited.

Smadja, F. A. (1993). Retrieving collocations from text: Xtract. *Computational Linguistics, 19*(1), 143−177. Retrieved from https://www.aclweb.org/anthology/J93-1007.pdf

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge/London, MA/England: Harvard University Press.

Snyder, J., & Stonehouse, B. (Eds.). (2007). *Prospects for polar tourism*. Wallingford, England: CAB International.

Sorokin, A. A., & Shavrina, T. (2016). *Automatic spelling correction for Russian social media texts.* Paper presented at the International Conference on Computational Linguistics and Intellectual Technologies "Dialogue," Moscow (Russia).

Stefanowitsch, A., & Gries, S. T. (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics, 8*(2), 209−243. doi:10.1075/ijcl.8.2.03ste

Stefanowitsch, A., & Gries, S. T. (2004). Extending collostructional analysis. A corpus-based perspective on 'alternations.' *International Journal of Corpus Linguistics, 9*(1), 97−129. doi:10.1075/ijcl.9.1.06gri

Stefanowitsch, A., & Gries, S. T. (Eds.). (2006). *Corpus-based approaches to metaphor and metonymy*. Berlin/New York, Germany/NY: Mouton de Gruyter.

Stefanowitsch, A., & Gries, S. T. (2008). Channel and constructional meaning: A collostructional case study. In G. Kristiansen & R. Dirven (Eds.), *Cognitive sociolinguistics: Language variation, cultural models, social systems* (pp. 129−152). Berlin, Germany: Mouton De Gruyter.

Štefic, L., Mravak-Stipetić, M., & Borić, V. (2010). Collocations in the language of dental medicine: Examples in oral medicine. *Acta Stomatologica Croatica, 44*(3), 176−187. Retrieved from https://hrcak.srce.hr/58754

Strong, G., & Dyer, P. (2009). *Access accents: Received Pronunciation (RP). An accent training resource for actors*. London, England: Bloomsbury Publishing PLC.

Stubbs, M. (1983). *Discourse analysis: The sociolinguistic analysis of natural language*. Chicago, IL: The University of Chicago Press.

Sung, H. H. (1996). *Definition of adventure travel: Conceptual framework for empirical application* (Unpublished master's thesis). Purdue University, West Lafayette, IN, United States.

Sung, H. H., Morrison, A. M., & O'Leary, J. T. (1996). Definition of adventure travel: Conceptual framework for empirical application for the providers' perspective. *Asia Pacific Journal of Tourism Research, 1*(2), 47–67. doi:10.1080/10941669708721975

Svartvik, J. (Ed.). (1990). *The London corpus of spoken English: Description and research*. Lund, Sweden: Lund University Press.

Swarbrooke, J., Beard, C., Leckie, S., & Pomfret, G. (2003). *Adventure tourism. The new frontier*. Oxford/Burlington, England/MA: Butterworth-Heinemann.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science, 12*(1), 49–100. doi:10.1207/s15516709cog1201_2

Tarp, S. (1992). *Prolegomena til teknisk ordbog* (Unpublished doctoral dissertation). Aarhus University School of Business and Social Sciences, Aarhus and Herning, Denmark.

Tarp, S. (1994). Funktionen in Fachwörterbüchern. In B. Schaeder & H. Bergenholtz (Eds.), *Fachlexikographie. Fachwissen und seine Repräsentation in Wörterbüchern* (pp. 229–246). Tübingen, Germany: Gunter Narr Verlag.

Tarp, S. (1995). Wörterbuchfunktionen: Utopische und realistische Vorschläge für die bilinguale Lexikographie. In H. E. Wiegand & P. Virtanen (Eds.), *Studien zur zweisprachigen Lexikographie mit Deutsch II* (pp. 17–51). Hildesheim/New York, Germany/NY: Georg Olms Verlag.

Tarp, S. (1998). Leksikografien på egne ben. Fordelingsstrukturer og byggedele i et brugerorienteret perspektiv. *Hermes, Journal of Linguistics, 21*, 121–137. doi:10.7146/hjlcb.v11i21.25479

Tarp, S. (2000). Theoretical challenges to practical specialised lexicography. *Lexikos, 10*, 189–208. doi:10.5788/10-0-894

Tarp, S. (2001). Lexicography and the linguistic concepts of homonymy and polysemy. *Lexicographica, International Annual for Lexicography, 17*, 22–39. doi:10.1515/9783110244212.22

Tarp, S. (2002). Translation dictionaries and bilingual dictionaries. Two different concepts. *Journal of Translation Studies, 7*, 59–84.

Tarp, S. (2009). Reflections on lexicographical user research. *Lexicos, 19*, 275–296. doi:10.5788/19-0-440

Tebé, C. (2003). Terminología: Historia y organización. *Curso de postgrado online de Introducción a la terminología*. Barcelona, Spain: Institut Universitary de Lingüística Aplicada, Universitat Pompeu Fabra.

Temmerman, R. (1997). Questioning the univocity ideal. The difference between socio-cognitive Terminology and traditional Terminology. *Hermes, Journal of Language and Communication in Business, 10*(18), 51–90. doi:10.7146/hjlcb.v10i18.25412

Temmerman, R. (1998). Why traditional terminology theory impedes a realistic description of categories and terms in the life sciences. *Terminology, 5*(1), 77–92. doi:10.1075/term.5.1.07tem

Temmerman, R. (2000). *Towards new ways of terminology description. The sociocognitive approach.* Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Temmerman, R., & Kerremans, K. (2003). Termontography: Ontology building and the sociocognitive approach to terminology description. In J. Mírovský, A. Kotěšovcová, & E. Hajičová (Eds.), *Proceedings of XVII International Congress of Linguists* (pp. 1–10). Prague, Czech Republic: Matfyzpress. Retrieved from https://www.academia.edu/851013/Termontography_Ontology_building_and_the_sociocognitive_approach_to_terminology_description

TermoStat Web 3.0 [Computer software]. Montréal, Canada: Patrick Drouin.

Teubert, W. (2004). Language and corpus linguistics. In M. A. K. Halliday, W. Teubert, C. Yallop, & A. Čermáková (Eds.), *Lexicology and corpus linguistics. An introduction* (pp. 73–112). London/New York, England/NY: Continuum.

Teubert, W., & Čermáková, A. (2004). Directions in corpus linguistics. In M. A. K. Halliday, W. Teubert, C. Yallop, & A. Čermáková (Eds.), *Lexicology and corpus linguistics. An introduction* (pp. 113–165). London/New York, England/NY: Continuum.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work.* Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Tognini-Bonelli, E. (2002). Corpora and LSP: Issues and implications. In C. T. Torsello, K. Ackerley, & E. Castello (Eds.), *Corpora for university language teachers* (pp. 31–47). Bern/Berlin/Brussels/Frankfurt am Main/New York/Oxford/Wien, Switzerland/Germany/Belgium/Germany/NY/England/Austria: Peter Lang.

Tognini-Bonelli, E. (2010). Theoretical overview of the evolution of corpus linguistics. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 14–27). London/New York, England/NY: Routledge.

Torner Castells, S., & Bernal, E. (2017). Collocations in learner's dictionaries. In S. Torner Castells & E. Bernal (Eds.), *Collocations and other lexical combinations in Spanish. Theoretical, lexicographical and applied perspectives* (pp. 157–172). London/New York, England/NY: Routledge.

Trono, A. (2015). Politics, policy and the practice of religious tourism. In R. Raj & K. Griffin (Eds.), *Religious tourism and pilgrimage management. An international perspective. Second edition* (pp. 16–36). Wallingford, England: CAB International.

Ungerer, F., & Schmid, H.-J. (1996). *An introduction to cognitive linguistics. Second edition.* Harlow, England: Pearson.

Ureña Gómez Moreno, J. M., & Buendía Castro, M. (2017). Semantic and conceptual aspects of volcano verb collocates within the natural disaster domain: A frame-based terminology approach. In M. Grygiel (Ed.), *Cognitive approaches to specialist languages* (pp. 330–350). Newcastle upon Tyne, England: Cambridge Scholars Publishing.

Van der Wouden, T. (1994). *Negative contexts* (Doctoral thesis, Rijksuniversiteit Groningen, Groningen, The Netherlands). Retrieved from https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.408.2910&rep=rep1&type=pdf

Van der Wouden, T. (1997). *Negative contexts. Collocation, polarity and multiple negation*. London/New York, England/NY: Routledge.

Van Leeuwen, T. (1993). Genre and field in critical discourse analysis: A synopsis. *Discourse & Society, 4*(2), 193–223. doi:10.1177/0957926593004002004

Van Sterkenburg, P. (2003). 'The' dictionary: Definition and history, In. P. van Sterkenburg (Ed.), *A practical guide to lexicography* (pp. 1–17). Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Varantola, K. (1986). Special language and general language: Linguistic and didactic aspects. *UNESCO ALSED–LSP Newsletters, 9*(2), 10–19.

Vargas Sierra, C. (2006). Diseño de un corpus especializado con fines terminográficos: El Corpus de la Piedra Natural. *Debate Terminológico, 2*(7), 1–20. Retrieved from http://rua.ua.es/dspace/handle/10045/9426

Vargas Sierra, C. (2010). Combinatoria terminológica y diccionarios especializados para traductores. In M. Ibáñez Rodríguez (Ed.), *Lenguas de especialidad y terminología* (pp. 18–49). Granada, Spain: Comares.

Vargas Sierra, C. (2012a). Caracterización contrastiva (inglés-español) de la combinatoria léxica especializada: Estudio basado en un corpus especializado. In E. Mogorrón Huerta & S. Mejri (Eds.), *Lenguas especializadas, fijación y traducción* (pp. 173–196). Alicante, Spain: Universitat d'Alacant.

Vargas Sierra, C. (2012b). La tecnología de corpus en el contexto profesional y académico de la traducción y la terminología: Panorama actual, recursos y perspectivas. In M. A. Candel Mora & E. Ortega Arjonilla (Eds.), *Tecnología, traducción y cultura* (pp. 67–99). Valencia, Spain: Tirant Humanidades.

Verlinde, S., & Binon, J. (1994). The dictionnaire contextual du français économique: A production-oriented dictionary of business French. In W. Martin, W. Meijs, M. Moerland, E. ten Pas, P. van Sterkenburg, & P. Vossen (Eds.), *Proceedings of the 6th EURALEX International Congress* (pp. 523–529). Amsterdam, The Netherlands: Euralex. Retrieved from https://euralex.org/publications/the-dictionnaire-contextuel-du-francais-economique-a-production-oriented-dictionary-of-business-french/

Verlinde, S., Binon, J., & Van Dyck, J. (2000). *Dictionnaire contextuel du français économique – Tome A: L'entreprise*. Antwerpen/Apeldoorn, The Netherlands: Garant.

Verlinde, S., Folon, J., Binon, J., & Van Dyck, J. (2000). *Dictionnaire contextuel du français économique – Tome B: Le commerce*. Antwerpen/Apeldoorn, The Netherlands: Garant.

Verlinde, S., Folon, J., Binon, J., & Van Dyck, J. (2003a). *Dictionnaire contextuel du français économique – Tome C: Les finances*. Antwerpen/Apeldoorn, The Netherlands: Garant.

Verlinde, S., Folon, J., Binon, J., & Van Dyck, J. (2003b). *Dictionnaire contextuel du français économique – Tome D: L'emploi*. Antwerpen/Apeldoorn, The Netherlands: Garant.

Viana, V., Zyngier, S., & Barnbrook, G. (Eds.). (2011). *Perspectives on corpus linguistics*. Amsterdam/Philadelphia, The Netherlands/PA: John Benjamins Publishing Company.

Vicente-López, E., Campos, L. M. de, Fernández-Luna, J. M., Huete, J. F., Tagua-Jiménez, A., & Tur-Vigil, C. (2015). An automatic methodology to evaluate personalized information retrieval systems. *User Modeling and User-Adapted Interaction, 25*(1), 1–37. doi:10.1007/s11257-014-9148-9

Villayandre Llamazares, M. (2008). Lingüística con corpus (I). *Estudios Humanísticos. Filología, 30*, 329–349. Retrieved from http://revpubli.unileon.es/index.php/EEHHFilologia/article/view/2847/2024

Vinogradov, V. V. (1947). *Izbrannye Trudy. Leksikologija i leksikografija*. Moscow, Russia: Nauka.

Walker, C. (2009). The treatment of collocation by learners' dictionaries, collocational dictionaries and dictionaries of business English. *International Journal of Lexicography, 22*(3), 281–299. doi:10.1093/ijl/ecp016

Wang, Y. (2018). "Language" and "discourse:" Two perspectives on linguistic philosophy. *Semiotica, 224*, 295–312.

Weinreich, U. (1969). Problems in the analysis of idioms. In J. Puhvel (Ed.), *Substance and structure of language* (pp. 23–81). Berkeley/Los Angeles, CA: University of California Press.

Wichmann, A. (2009). Speech corpora and spoken corpora. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics. An international handbook. Volume I* (pp. 187–207). Berlin/New York, Germany/NY: Walter de Gruyter.

Wiegand, H. E. (1995). Der kulturelle beitrag der lexicographie zur umgestaltung osteuropas. *Lexicographica, International Annual for Lexicography, 11*, 210–218.

Willis, D. (1990). *The lexical syllabus: A new approach to language teaching*. London, England: Collins.

WMATRIX (Version 4) [Computer software]. Lancaster, England: Lancaster University.

Wolter, B. (2001). Comparing the L1 and L2 mental lexicon. *Studies in Second Language Acquisition, 23*, 41–69. doi:10.1017/S0272263101001024

WordSmith Tools (Version 8) [Computer software]. Stroud, England: Lexical Analysis Software.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge, England: Cambridge University Press.

Wüster, E. (1931). *Internationale Sprachnormung in der Technik, besonders in der Elektrotechnik: Die nationale Sprachnormung und ihre Verallgemeinerung*. Berlin, Germany: VDI Verlag.

Wüster, E. (1968). *The machine tool. An interlingual dictionary of basic concepts*. London, England: Technical Press.

Xiao, R. (2015). Collocation. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 106−124). Cambridge, England: Cambridge University Press.

Yang, H. (1986). A new technique for identifying scientific and technical terms and describing science texts. *Literary and Linguistic Computing, 1*(2), 93−103. doi:10.1093/llc/1.2.93

Zgusta, L. (1971). *Manual of lexicography*. The Hague, The Netherlands: Mouton de Gruyter.

Zinsmeister, H., Hinrichs, E., Kübler, S., & Witt, A. (2009). Linguistically annotated corpora: Quality assurance, reusability and sustainability. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics. An international handbook. Volume I* (pp. 759–776). Berlin/New York, Germany/NY: Walter de Gruyter.

Žolkovskij, A. K., & Mel'čuk, I. A. (1965). O vozmožnom metode i instrumentax semantičeskogo sinteza. *Naučno-Texničeskaja Informacija, 5*, 23–28.

Žolkovskij, A. K., & Mel'čuk, I. A. (1966). O sisteme semantičeskogo sinteza. I. Stroenie slovarja. *Naučno-Texničeskaja Informacija, 11*, 48–55.

Žolkovskij, A. K., & Mel'čuk, I. A. (1967). O semantičeskom sinteze. *Problemy Kibernetiki, 19*, 177–238.

Zurick, D. N. (1992). Adventure travel and sustainable tourism in the peripheral economy of Nepal. *Annals of the Association of American Geographers, 82*(4), 608–628. doi:10.1111/j.1467-8306.1992.tb01720.x

APPENDIX

1. *abseil*-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
slowly abseil / abseil slowly	10.38	2	--	--

abseil$_1$

Collocation	Type	Explanation	Examples
slowly abseil / abseil slowly	adverb + verb / verb + adverb	~ at a specific SPEED	(103) Our experienced instructors will also be there to guide you as you **slowly abseil** down the wall. (104) **Abseil slowly** alongside your instructor who, although on a separate rope, provides information, reassurance... and a backup safety tether.

2. *arrive*-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
arrive early	11.04	5	9.43	17,485

arrive$_1$

Collocation	Type	Explanation	Examples
arrive early	verb + adverb	~ at a specific TIME	(105) Our airfield planning permission allows us to fly from 9am to 8pm (or dark) on a daily basis and it's always advisable to **arrive early** as we operate on a first come first served system. (106) 'Start early and **arrive early**' is the cardinal rule of trekking 6.

3. *drive*-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
drive (a) team	11.85	34	--	--
drive (a) sled	10.9	11	--	--
(a) participant drive	8.07	2	--	--
drive (an) hour	7.66	2	6.9	11,792

***drive*$_1$**

Collocation	Type	Explanation	Examples
drive (a) team	verb + noun	~ a VEHICLE_WITHOUT_ENGINE	(107) **Drive** the **team** or relax in the sled while observing the breathtaking environment around you. (108) You will have the option of helping the professional mushers **drive** the **team** or simply relaxing in the sled and witnessing the power of the dogs and the beauty of your surroundings.
drive (a) sled	verb + noun	~ a VEHICLE_WITHOUT_ENGINE	(109) Both novices and seasoned winter enthusiasts will appreciate the extended time with the dogs, learning to **drive** your own **sled** and a mix of camping under the stars, as well as a night at our heated backcountry yurt. (110) High on the snow-pack, **drive** the **sled** yourself or, if you prefer, ride in the sled as you enjoy the dramatic scenery.
(a) participant drive	noun + verb	A TOURIST ~s	(111) **Participants drive** the sleds while the guides ski along nearby to assist as needed. (112) Unlike most dogsledding outfits, Base Camp Dogsledding provides a "hands on" approach allowing you, the **participant**, the opportunity to **drive** your own team of

Collocation	Type	Explanation	Examples
			our friendly and hard-working Inuit Sled Dogs.
drive (an) hour	verb + noun	~ for a period of TIME	(113) After breakfast in the hotel **drive** an **hour** to Nayapul and commence trek. (114) Then Saturday we **drove** an **hour** to hike another area with an amazing waterfall where we ate lunch.

4. *enter*-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
enter (a) world	9.89	7	8.11	33,776

***enter*$_1$**

Collocation	Type	Explanation	Examples
enter (a) world	verb + noun	~ a specific DESTINATION	(115) Waving our good-byes, we **enter** a different **world** of a grander scale dominated by jungle, soaring limestone rock faces, and flourishing ferns. (116) While you **enter** the **world** of ice – ice rocks, dripping stalactites and of course the slippery snow stand there to greet you.

5. *exit*-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
exit (an) aircraft	12.02	9	5.64	440

exit₁

Collocation	Type	Explanation	Examples
exit (an) aircraft	verb + noun	~ a SOURCE	(117) As you **exit** the **aircraft** at approx. 3000ft, your parachute will be automatically deployed by a Static Line, securely attached to the aircraft, for added safety. (118) Weather permitting, you will **exit** the **aircraft** from around 10,000 feet, experiencing 30 seconds of freefall, which is quite a rush!

6. *fall₁*-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
fall back	7.35	3	8.08	76,625

fall₁

Collocation	Type	Explanation	Examples
fall back	verb + adverb	~ in a specific DIRECTION	(119) Wingsuit Flying is one of the oldest adrenaline pumping sports that can topple and twist you in mid-air while you are **falling back** at a top speed to mother Earth. (120) When the jumper jumps off the structure for free-falling, hovering above the ground and the rebounds, the cord stretches and the jumper flies upward again, then **falls back** down.

7. *float*-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
slowly float	9.95	2	4.89	890
float down	9.49	4	4.62	2,821

***float*$_1$**

Collocation	Type	Explanation	Examples
slowly float	adverb + verb	~ at a specific SPEED	(121) Just enjoy breath-taking views of Cape Town and Table Mountain, as you **slowly float** back to earth. (122) At just the right moment, you pull the chord and your descent is stopped as your parachute unfurls and you **slowly float** down to solid ground.
float down	verb + adverb	~ in a downward DIRECTION	(123) The parachute ride is exhilarating as you **float down** back to earth. (124) At just the right moment, you pull the chord and your descent is stopped as your parachute unfurls and you slowly **float down** to solid ground.

8. *fly*-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
(a) pilot fly	10.71	13	8.33	5,588
fly (a) site	10.43	19	--	--
fly solo	9.62	2	5.98	537
(a) glider fly	9.61	4	--	--
fly backwards	9.61	2	6.75	1,136
fly tandem	9.23	5	--	--
fly (a) parachute	9.15	6	--	--
fly (a) paramotor	9.12	4	--	--
fly (a) paraglider	9.01	4	--	--

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
(an) instructor fly	7.92	2	--	--

***fly*$_1$**

Collocation	Type	Explanation	Examples
(a) pilot fly	noun + verb	A TOURIST ~s	(125) Trim speed is defined as the speed a glider is flying when the **pilot** is **flying** hands off. (126) Parascending **pilots fly** canopies that are towed into the air by a Land Rover or winch before gliding back down to land.
fly (a) site	verb + noun	~ over a specific PLACE	(127) DO NOT attempt to **fly** this **site** for the first time without a local guide! (128) We spend the following days **flying** the **sites** around Lake Hovsgol at the foot of eastern Sayan Mountain.
fly solo	verb + adverb	~ in a specific MANNER	(129) **Fly solo** or soar across the skyline in pairs! (130) Similarly, in paragliding, you can either book into a course where you will learn the ropes in order to be able to **fly solo**, or you can book onto a tandem flight where an experienced paraglider takes control.
(a) glider fly	noun + verb	A TOURIST ~s	(131) The glide ratio has improved which allows **gliders** to **fly** for extended time periods. (132) The **glider flies** in a very "bird like" way that emulates the soaring flight of eagles and hawks.
fly backwards	verb + adverb	~ in a specific DIRECTION	(133) On this popular zip, you can **fly backwards**, spin upside down or even just enjoy the race! (134) Our specialized safety system allows a true freestyle zipping experience, you can

Collocation	Type	Explanation	Examples
			fly backwards, forwards, anyway you like, including the DO THE DANGLETM experience, flying upside down above the canyon!
fly tandem	verb + adverb	~ in a specific MANNER	(135) After your morning session you will **fly tandem** with an instructor by your side to an altitude of either 1,500 feet or 3,000 feet above beautiful Lookout Valley. (136) You can **fly tandem** or solo, but we recommend tandem because the experience is so special you will want to share every moment.
fly (a) parachute	verb + noun	~ a VEHICLE_WITHOUT_ENGINE	(137) Learning to skydive in British Columbia is a life-changing adventure, be it the thrill of freefall or learning to **fly** your own **parachute**. (138) The silence is amazing, the view sensational and during the 5-minute descent you will be taught how to **fly** the **parachute** to a gentle landing.
fly (a) paramotor	verb + noun	~ a VEHICLE_WITH_ENGINE	(139) Learning to **fly** a **paramotor** before making your initial flights under power, you'll first need to learn to fly the wing. (140) No CAA licence is required to **fly** a **paramotor**, but you still have to know and obey the rules and regulations applying to UK Airspace.
fly (a) paraglider	verb + noun	~ a VEHICLE_WITHOUT_ENGINE	(141) Camps Paragliding is a recreational extreme sport that involves **flying paragliders**, which are foot-launched, free-flying, lightweight glider aircrafts that lack a rigid primary structure. (142) When folded up to fit in a backpack and the weight of scale, some of the very few

Collocation	Type	Explanation	Examples
			that have attracted mountaineers and mountain tops are those arriving by **flying** a **paraglider**.
(an) instructor fly	noun + verb	A RESPONSIBLE ~S	(143) Once the jumper deploys the parachute, the **instructor flies** away and manages his/her own canopies. (144) If you'd like, you can assist the **instructor flying** the parachute and together you will fly back to the ground for a soft landing.

9. *glide*-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
silently glide / glide silently	11.14	2	8.11	1,009

***glide*$_1$**

Collocation	Type	Explanation	Examples
silently glide / glide silently	adverb + verb / verb + adverb	~ in a specific MANNER	(145) The paraglider has no motor and uses the rising air to **silently glide** above the ground, resulting in a mind-blowing experience and up to 20 minutes of flying time. (146) Immerse your senses in a world of natural beauty, **gliding silently** through sweeping mountain views of rural Jamaica from nearly a mile of ziplines through a tropical forest.

10. *hike*-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
hike (a) trail	12.04	131	11.25	35,977
hike down	9.7	5	3.31	1,131
hike back	9.28	12	--	--

***hike*$_1$**

Collocation	Type	Explanation	Examples
hike (a) trail	verb + noun	~ a specific PATH	(147) I have **hiked** this **trail** in fall and spring, and it gets hot so make sure you carry lots of water and food, as there is no food available on trail, though Bungee America does provide water near the bridge if you are hiking with the group. (148) There are scenic roads and **trails** to **hike** in the region.
hike down	verb + adverb	~ in a downward DIRECTION	(149) Transportation to this remote ecosystem is included in the tour; then trekkers will **hike down** through a coffee plantation to enter Quebrada Rosa, a beautiful and unspoiled creek. (150) For a breathtaking bird's eye view of the town, **hike down** toward Mouriq or climb up to the Mellal Gorges.
hike back	verb + adverb	~ in a specific DIRECTION	(151) Only suggestion would be to not let individuals **hike back** alone without a guide. (152) After refueling and refreshing, we'll **hike back** through canyon country to our trailhead.

11. *jump*-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
jump solo	10.2	5	--	--

***jump*$_1$**

Collocation	Type	Explanation	Examples
jump solo	verb + adverb	~ in a specific MANNER	(153) You **jump solo** from a height of approximately 3,500 feet and your parachute deploys automatically. (154) You'll **jump** either **solo** or in tandem attached to an instructor who controls all aspects of the jump.

12. *land*-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
land safely	10.43	4	8.35	4,249

***land*$_1$**

Collocation	Type	Explanation	Examples
land safely	verb + adverb	~ in a specific MANNER	(155) After coming out to a certain height, the skydiver is required to pull out his parachute to **land safely** on the ground. (156) Once you come down to specific altitude, pull out your parachute to **land safely** on the ground.

13. *navigate*-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
navigate rapids	9.19	2	5.56	252
navigate (a) section	9.03	3	--	--
navigate water	8.86	4	6.45	5,243
navigate (a) waterfall	8.82	2	--	--

***navigate*$_1$**

Collocation	Type	Explanation	Examples
navigate rapids	verb + noun	~ a specific PATH	(157) My favorite experience was **navigating** class IV and V **rapids** while carefully avoiding capsizing into its highly dangerous currents. (158) Feel your adrenaline pumping as you **navigate** powerful **rapids** and white water.
navigate (a) section	verb + noun	~ a specific PATH	(159) If you were **navigating** this **section** by boat, I imagine it would require a little extra effort from both your arms and back. (160) That's because there's a via Ferrata built into the route, helping hikers **navigate** waterfalls and vertical **sections** of rock.
navigate water	verb + noun	~ a specific PATH	(161) If this is your first time in a canoe, you'll have no problem **navigating** the calm **waters** of the Youghiogheny – it's waiting for you! (162) Alongside your highly qualified guide, **navigate** the pristine swirling **waters** and boulders that create a thrilling rafting obstacle course that suits all levels of experience.
navigate (a) waterfall	verb + noun	~ a specific PATH	(163) That's because there's a via Ferrata built into the route, helping hikers **navigate waterfalls** and vertical sections of rock. (164) The next several hours will be full of adrenalin as you swim through rapids, jump off cliffs, scramble up rocks and **navigate** huge **waterfalls**, for sure lots of fun.

14. *paddle*-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
paddle hard	10.19	2	1.72	185
gently paddle	10.07	2	2.4	57
paddle (a) canoe	10.05	7	9.99	1,868
(a) kayaker paddle	10.04	2	8.28	153
paddle (a) kayak	9.72	7	9.42	1,147
paddle (a) raft	8.62	2	6.25	130
paddle back	8.39	6	2.17	1,125
paddle (a) river	7.75	2	6.87	1,031

***paddle*$_1$**

Collocation	Type	Explanation	Examples
paddle hard	verb + adverb	~ in a specific MANNER	(165) We all keep **paddling hard**. (166) When Sport Rafting you must **paddle hard** for most of the trip.
gently paddle	adverb + verb	~ in a specific MANNER	(167) All you need to do is sit back and **gently paddle** your way down stream as your knowledgeable guide explains all about the local area and the flora and fauna in this stunning part of Wisconsin. (168) Carve your surfboard into iconic waves or **gently paddle** along our beautiful coastline in a canoe built for two.
paddle (a) canoe	verb + noun	~ a VEHICLE_WITHOUT_ENGINE	(169) Enjoy the stunning scenery and natural wilderness of this special area as you **paddle** your **canoe** and attempt to stay dry! (170) **Paddle** your **canoe** across the peaceful waters of Derwentwater or down a river in the Lake District.

Collocation	Type	Explanation	Examples
(a) kayaker paddle	noun + verb	A TOURIST ~s	(171) It could be an option for a sea **kayaker** who **paddles** in busy waters where there is lots of shipping. (172) The **kayakers** also **paddle** into Krabi River opposite town, and out to the Koh Hong island group.
paddle (a) kayak	verb + noun	~ a VEHICLE_WITHOUT_ENGINE	(173) CLASSIII-IV experience is a MUST, whether you are **paddling** your own **kayak** or whitewater canoe, or whether you choose to paddle one from our fleet (no additional charge). (174) Not only will you be able to **paddle** a **kayak** up one of Kauai's gentle and meandering rivers but since these rivers cut right through the dense and luscious jungles your Kauai kayak tour will deliver you on one of its scenic rivers into the very heart of Kauai.
paddle (a) raft	verb + noun	~ a VEHICLE_WITHOUT_ENGINE	(175) Build your own raft, using barrels, planks and rope, then navigate a course around obstacles or **paddle** your **raft** to an island for a picnic! (176) Navigate a course around obstacles or **paddle** your **raft** to an island for a picnic.
paddle back	verb + adverb	~ in a specific DIRECTION	(177) They'll drive you up the river, prepare reliable equipment for you to use, and let you **paddle back** down the gentle waters. (178) Same as start, you **paddle back** to the beginning.
paddle (a) river	verb + noun	~ a specific PATH	(179) **Paddle** the majestic Colorado **River** with our expert guides.

Collocation	Type	Explanation	Examples
			(180) You can experience unforgettable outdoor activities, scale heights, **paddle** the **river** and explore underground.

15. *reach*-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
reach (the) summit	11.13	27	6.9	8,957
reach (the) top	10.82	21	8.08	22,859
reach (a) destination	9.19	9	8.31	25,996
reach (a) point	9.18	9	8.84	73,804
reach (the) bottom	9.01	6	6.69	8,073
reach (a) bridge	8.77	6	--	--
reach (an) altitude	8.46	4	--	--
(a) participant reach	8.24	2	--	--
reach (a) spot	8.21	4	--	--
reach (a) canyon	8.2	4	--	--
reach (a) peak	8.04	4	8.05	20,449
reach (the) base	8.04	3	6.38	7,454
reach (the) base camp	7.92	3	--	--
reach (a) mountain	7.65	3	--	--

***reach*$_1$**

Collocation	Type	Explanation	Examples
reach (the) summit	verb + noun	~ a specific DESTINATION	(181) This all sounds quite intense and quite frankly when you're several metres up from the ground it can be just that, but there is nothing like the sensation of **reaching** the **summit**.

Collocation	Type	Explanation	Examples
			(182) There is nothing like the sensation of **reaching** the **summit**.
reach (the) top	verb + noun	~ a specific DESTINATION	(183) As you **reach** the **top**, you are rewarded by the magnificent views, then you descend to Dzongla at noon. (184) Get ready for stunning mountain views, big skies and a huge sense of achievement when you **reach** the **top**.
reach (a) destination	verb + noun	~ a specific DESTINATION	(185) I loved the hike that we had upon **reaching** the bungee **destination** and how we were hip-deep in water at some times. (186) During your peak climbing trip, you will be accompanied by expert climbing guides and support teams who can help you **reach** your **destination**.
reach (a) point	verb + noun	~ a specific DESTINATION	(187) Ascending a steep glacier, we **reached** the route's pinch **point**, where Shackleton pushed his uncanny luck to the max. (188) These high and long suspension bridges will play with your head at first, but as you learn to trust the equipment, you'll **reach** a **point** where trekking high in the trees is a truly memorable experience!
reach (the) bottom	verb + noun	~ a specific DESTINATION	(189) Follow that back towards Neates Glen until you reach Pilcher Trail, which exits on the left hand side just before you **reach** the **bottom** of Grand Canyon. (190) Despite divers going as deep as 120 meters, no one has **reached** the **bottom** yet, leaving lots to be discovered.
reach (a) bridge	verb + noun	~ a specific DESTINATION	(191) Once you **reach** the **bridge**, you'll relax for 15 minutes or so and then you'll

Collocation	Type	Explanation	Examples
			go through "jump school" for about 30 minutes. (192) Luckily, we ran into other hikers that helped us **reach** the **bridge.**
reach (an) altitude	verb + noun	~ a specific DISTANCE	(193) As you glide and **reach** a sufficient **altitude,** pull open your parachute and land feet first on the ground. (194) It started as gliding down hills on low performance kites, but now pilots can stay airborne for hours, **reach altitudes** of several thousand feet and reach speeds of over one hundred kilometres per hour.
(a) participant reach	noun + verb	A TOURIST ~s	(195) Once **participants reach** the top of the rock, they will be exposed to stunning views before making their way to the 100 ft. rappel. (196) Snowkiting is another former of high-flying adventure, although **participants** never **reach** the same heights as paragliders.
reach (a) spot	verb + noun	~ a specific DESTINATION	(197) Then we made our way up across a slippery cliff and **reached** a dive **spot**. (198) You ride along in a boat to the voice of a tour guide until you **reach** a **spot** where whales frequently appear.
reach (a) canyon	verb + noun	~ a specific DESTINATION	(199) Park at the Powerlines just past Katoomba Airfield, walk down into Katoomba Creek and follow the creek for a few hours to **reach** the **canyon**. (200) Once in Katoomba Creek simply follow it until you **reach** the **canyon** – this should take around 2.5 hours.
reach (a) peak	verb + noun	~ the top of a DESTINATION	(201) Once **reaching** the **peaks** of 4,000 meters / 14,000 feet you will be freefalling into one of the most beautiful skydive locations on earth.

Collocation	Type	Explanation	Examples
			(202) Via ferratas are just awesome because you really don't have to be a mountaineer, nor a rock-climber to **reach** wild **peaks** with jawdropping views.
reach (the) base	verb + noun	~ a specific DESTINATION	(203) A short way beyond the Refuge Locatelli, we **reach** the **base** of the via ferrata on the north side of the Torre di Toblin.
			(204) The fun and thrill of heli-skiing begins with the drop at high altitudes and continues till you **reach** the skiing **base**!
reach (the) base camp	verb + noun	~ a specific DESTINATION	(205) The walk passes over rocky dunes and moraine and streams, till you **reach** the **base camp**.
			(206) As we **reach** the **base camp**, we enjoy the incredible melt patterns and admire the gently rounded ice towers on the upper part of the Khumbu glacier.
reach (a) mountain	verb + noun	~ a specific DESTINATION	(207) A 6000m peak climbed in alpine style from a high camp may not be the same challenge as climbing an 8000m mountain where you spend weeks camped at its base while you establish higher camps, but you can combine the two activities – trekking and mountaineering – into a single expedition, since often you must trek for days to **reach** the **mountain** in the first place.
			(208) Trekking to Muktinath allows you to **reach** the high **mountains** of the Himalayas within a short time frame and without ascending and descending to extreme altitude.

16. *ride*$_1$-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
ride (a) bike	10.6	13	11.36	70,858
ride (a) bicycle	10.12	6	10.05	20,011
ride (a) wave	9.75	5	9.41	17,218
ride (a) zipline	9.26	4	--	--
ride (a) raft	8.43	2	--	--

***ride*$_1$**

Collocation	Type	Explanation	Examples
ride (a) bike	verb + noun	~ a VEHICLE_WITHOUT_ENGINE	(209) To make this jump, you just have to be able to **ride** a **bike** along the 30-meter long ramp. (210) No experience required but must be able to **ride** a **bike** comfortably.
ride (a) bicycle	verb + noun	~ a VEHICLE_WITHOUT_ENGINE	(211) It is a sport where one **rides bicycles** over rough territory, using specially designed mountain bikes which are able to withstand the vagaries of going over bumpy and uneven pathways. (212) You are going to **ride** your **bicycle** on different high-height streets, including the Khardungla pass, the world's most elevated motorable street, 5606 m over the ocean level.
ride (a) wave	verb + noun	~ a specific PATH	(213) You actually **ride** those **waves** and leap as soon as a rising wave approaches. (214) Paddle past the most easterly point of the Australian mainland, see the Julian Rocks Marine Reserve, and **ride** the **waves** into the shore.
ride (a) zipline	verb + noun	~s a VEHICLE_WITHOUT_ENGINE	(215) Guides have the final authority as to determining guests' ability to **ride** the **zipline**.

Collocation	Type	Explanation	Examples
			(216) Once at the waterfall those there to **ride** the **zipline** are given a safety harness, helmet and gloves.
ride (a) raft	verb + noun	~ a VEHICLE_WITHOUT_ENGINE	(217) Suited up with a helmet and flotation vest, you will paddle and **ride** your inflatable **raft** down the grade 2-3 rapids, splashing down huge drops and through raging white water to reach the calm water at the bottom of the gorge. (218) You will be given the option to **ride** the **raft** or walk the short portage trail at every significant rapid.

17. *ride*$_2$-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
ride tandem	9.34	4	--	--
ride (a) lift	9.11	3	5.28	754

***ride*$_2$**

Collocation	Type	Explanation	Examples
ride tandem	verb + adverb	~ in a specific MANNER	(219) If your child is under 70lbs, they qualify as Little Zippers and get to **ride tandem** with one of our awesome fun guides. (220) If a child can't comprehend the zip line, they can **ride tandem** with a guide.
ride (a) lift	verb + noun	~ a VEHICLE_WITH_ENGINE	(221) Hikers can **ride** the ski **lift**, which is operational on weekends during summer, to the 7800-foot level to the Top Of The Notch Restaurant and, on a clear day, a view of Catalina Island.

Collocation	Type	Explanation	Examples
			(222) If you're short on time or stamina, cut a few miles off the trek by **riding** the **lift** up to 7,800 feet.

18. *scale*-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
scale height	11.54	8	8.07	2,320
scale (a) wall	9.24	2	7.09	2,595
scale (a) peak	8.5	2	6.7	679

***scale*$_1$**

Collocation	Type	Explanation	Examples
scale height	verb + noun	~ a specific PATH	(223) If you keep your eyes open, you would certainly find the extreme adventure and thrill of **scaling** new **heights** over the skies in a place near you. (224) If you are passionate enough to **scale** new **heights** in the sky, then skydiving can be your newfound interests or rather call it love!
scale (a) wall	verb + noun	~ a specific PATH	(225) All sessions will be geared to your level, so whether you have never even stepped into a pair of climbing shoes before, or have already **scaled** a couple of **walls** in your time, you can learn lots from your qualified guides who are all highly experienced and qualified climbers and mountaineers. (226) When climbing Grand Canyon and **scaling** the sheer **walls** of challenging mountains, it's always best to get a guide for the safest and most exciting adventures.

Collocation	Type	Explanation	Examples
scale (a) peak	verb + noun	~ up to the top of a PLACE	(227) Nothing better sums up the outdoors than the centuries-old human endeavor to **scale** mighty **peaks**. (228) Developed by our team of pre-eminent guides the ACC is ideal for those looking to learn the latest and most relevant alpine climbing techniques while **scaling peaks** throughout New Zealand's Southern Alps.

19. *scramble*-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
scramble up	7.18	4	2.27	2,663

***scramble*$_1$**

Collocation	Type	Explanation	Examples
scramble up	verb + adverb	~ in an upward DIRECTION	(229) **Scramble up** to the ledge and start from there, traversing left under the curtain. (230) These have been developed now into amazing playgrounds where you cross cable bridges, zip along cables, **scramble up** netting and climb steep ladders and rock walls, always attached to a safety cable for protection.

20. *skydive*-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
tandem skydive	11.96	9	9.96	75

skydive₁

Collocation	Type	Explanation	Examples
tandem skydive	adverb + verb	~ in a specific MANNER	(231) Embrace the excitement and soak in the views as you **tandem skydive** or go solo – a true heart-stopping adventure! (232) Embrace the fear and **tandem skydive**, then whitewater raft down the Shotover River.

21. *slide*-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
slide down	10.06	6	6.27	4,789
slide (one's) way	8.17	4	--	--

slide₁

Collocation	Type	Explanation	Examples
slide down	verb + adverb	~ in a downward DIRECTION	(233) To descend back to the base, you need to rappel down the rock face using a rope to help you **slide down** safely. (234) Guides won't hesitate to take pictures of you while you **slide down**.
slide (one's) way	verb + noun	~ in a specific MANNER	(235) The Aerial Trekking Course includes over 30 games in the trees that test agility and stamina and require you and the fam to climb, scramble, swing, **slide** your **way** from station to station. (236) Discover the incredible natural beauty of the Blue Mountains as you abseil down waterfalls and swim, jump, and **slide** your **way** along rocky canyon descents.

22. *swim*-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
swim (one's) way	7.69	3	--	--

***swim*$_1$**

Collocation	Type	Explanation	Examples
swim (one's) way	verb + noun	~ in a specific MANNER	(237) Starting at the top of the canyon, you will abseil, zip-line, cliff-jump, rock-slide and **swim** your **way** down from one pool to the next, using whichever tactic best suits the next obstacle. (238) Some areas of the river are shallow, sometimes waist deep, but most of the time it was deep that you needed to **swim** your **way** across through it.

23. *traverse*-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
traverse (a) canyon	9.34	3	5.08	84
traverse (a) river	8.24	2	4.95	291

***traverse*$_1$**

Collocation	Type	Explanation	Examples
traverse (a) canyon	verb + noun	~ a specific PATH	(239) Cross off an ultimate bucket list adventure, **traversing** the **canyon** from rim to rim. (240) On days we are not on tour we spend our time training, **traversing** our **canyons**, and kayaking our rivers.
traverse (a) river	verb + noun	~ a specific PATH	(241) Rafting is not available there, you have to **traverse** the

Collocation	Type	Explanation	Examples
			river manually by doing a lot of swimming and trekking. (242) Use the paddle to **traverse** the sparkling **river**.

24. *trek*-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
trek down	8.24	2	2.15	506
trek up	7.97	7	0.97	1,084

***trek*$_1$**

Collocation	Type	Explanation	Examples
trek down	verb + adverb	~ in a downward DIRECTION	(243) After breakfast in Gorakshep, we **trek down** to Dhugla, then continue staying high above the valley floor in Imja valley. (244) You will **trek down** to lodge at Gokyo on the same day.
trek up	verb + adverb	~ in an upward DIRECTION	(245) One morning we decided to **trek up** along the hills surrounding the lake. (246) Once our pilots, Toni and Tobias with Airtime Paragliding, arrived, it was time to **trek up** to the take off point.

25. *venture*-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
venture out	8.44	9	3.62	655

***venture*$_1$**

Collocation	Type	Explanation	Examples
venture out	verb + adverb	~ away from a SOURCE	(247) Those more experienced can **venture out** in the sea on their own. (248) In order to avoid injury, seek advice from local guides, clubs, or associations before **venturing out**.

26. *walk*-v

Collocation	*logDice* in ADVENCOR	Frequency in ADVENCOR	*logDice* in 'enTenTen13'	Frequency in 'enTenTen13'
walk down	9.37	5	7.27	13,026
walk (one) mile	8.87	6	9.04	21,495
walk (one) hour	8.72	5	5.13	2,282
(a) guide walk	7.07	2	4.41	653

***walk*$_1$**

Collocation	Type	Explanation	Examples
walk down	verb + adverb	~ in a downward DIRECTION	(249) Park at the Powerlines just past Katoomba Airfield, **walk down** into Katoomba Creek and follow the creek for a few hours to reach the canyon. (250) We **walked down** off the mountain 20 meters apart, silently lost in our own thoughts.
walk (one) mile	verb + noun	~ a specific DISTANCE	(251) You will **walk** 5 **miles** to the bridge at a brisk pace. (252) In order to participate in this adventure, you must be able to **walk** 2 **miles** and kayak 60-90 minutes.
walk (one) hour	verb + noun	~ for a period of TIME	(253) It was a great group who all enjoyed each the company and shared experience of having to **walk** 7 **hours** a day and get acclimatised before reaching the

Collocation	Type	Explanation	Examples
			Annapurna Base Camp at over 4000 metres. (254) Try the sled run on Breitenstein mountain from the village of Fischbachau; while there is no chairlift here to whisk you to the top (you **walk** an **hour** up a mountain road, dragging your sled behind you), there is a little wooden hut at the summit laced with icicles that serves cake and coffee to fortify you after the hike and before your descent.
(a) guide walk	noun + verb	A RESPONSIBLE ~s	(255) Our **guides walk** with you, provide interpretation and love to show you our beautiful home. (256) The anticipation and excitement builds as your group and **guides walk** through the vast archway of the Devils Coachhouse and into the McKeowns valley to one of the most beautiful of Jenolan's caves.