

ISABEL DURÁN MUÑOZ

[iduran@uco.es](mailto:iduran@uco.es)

EVA LUCÍA JIMÉNEZ-  
NAVARRO

[lucia.jimenez@uco.es](mailto:lucia.jimenez@uco.es)

Universidad de Córdoba,  
Spain

# ADVENCOR, the corpus of adventure tourism: Design, Compilation and Management

# CONTENTS

1. Introduction
2. Objectives
3. ADVENCOR Corpus
  1. Design
  2. Compilation
  3. Management
4. Conclusions and further research
5. References



I

# INTRODUCTION



# I. INTRODUCTION: ADVENTURE TOURISM

- ❑ Since the 70s, but... current innovative ways to attract potential tourists
- ❑ Alternative tourism (Fennell, 2003)
- ❑ Distinctive characteristics:
  - Opposite to mass tourism (e.g., sun and beach)
  - Natural environment
  - Unfamiliar places
  - Elements of risk at different grades
  - Physical challenge

---

2

OBJECTIVES

---

## 2. OBJECTIVES

1. To present the design and compilation of the ADVENCOR corpus
2. To explore corpus management techniques with different tools and purposes

---

3

ADVENCOR  
CORPUS

---

## 3.1. CORPUS DESIGN

External criteria (Biber, 1993; Sinclair, 2003):

- 1) time of publication
- 2) geography
- 3) text type
- 4) text length
- 5) level of specialization

Type of compilation

- Pros and cons of automatic compilation
- 'New Corpus' function in *Sketch Engine* (<https://www.sketchengine.eu/>)



## 3.2. CORPUS COMPILATION: STEPS

- ❑ Automatic corpus compilation with *Sketch Engine*

### CREATE CORPUS

ADVENCOR

1. CREATE CORPUS > 2. ADD TEXTS > 3. COMPILE

Build your own private corpus from texts on the web or from your own documents.

Name Advencor

Corpus type  Single language corpus  
 Multilingual corpus

Language English

Description promotional texts about adventure tourism

Storage used: 1,118,903 of 2,500,000 words (44%)

Available features

- part of speech
- N-grams
- Concordance
- Keywords
- wordsketch
- lemma
- Frequency distribution
- GDEX
- Thesaurus
- Term extraction
- Wordlist

# STEP I. SELECTION OF SEEDS

adventure sport	adventure tour	adventure tourism
adventure activity	kayak	parasailing
adventure activities	cycling	parachuting
trekking	mountain biking	skydiving
hiking	riding	caving
canyoning	potholing	paragliding
mountaineering	climbing	hang gliding
kayaking	bungee jumping	dogsledding
hike	speleology	canoeing
trek	rappel	adventure
canyon	rafting	zip lining

## STEP II. SELECTION OF TUPLES

- ❑ Automatic creation of “tuples” → random combination of three different seeds (e.g., “trekking kayak adventure tourism”)

# STEP III. SELECTION OF WEBSITES

1. Automatic generation of a list of potentially relevant webpages
2. Manual revision to refine the searches and to guarantee successful results
3. De-selection of irrelevant URLs according to the project purpose:
  1. *Wikipedia, Amazon, social networks (e.g. Facebook, Pinterest), YouTube, Scribd, eBay, etc.*
  2. Websites not originally written in English
  3. Websites not published by public or private institutions (including travel companies or agencies), such as articles, blogs, etc.
4. After manual revision, 30% of the URLs were discarded

## 3.2. CORPUS COMPILATION: DATA

<b>Number of tokens</b>	1,045,755
<b>Number of types</b>	49,988
<b>Type of corpus</b>	Specialized/electronic
<b>Size</b>	Small
<b>Mode</b>	Written
<b>Language</b>	English
<b>Domain / Subdomain</b>	Tourism / Adventure tourism
<b>Genre</b>	Promotional
<b>Text length</b>	Complete texts
<b>Purpose</b>	Terminological and phraseological analyses
<b>Communicative situation</b>	Semi-specialized
<b>Publication date</b>	Recent
<b>Source of texts</b>	Websites
<b>Publishers</b>	English-speaking public or private institutions

---

3.3.

CORPUS  
MANAGEMENT

# 3.3.1. CORPUS MANAGEMENT: *TERMOSTAT WEB 3.0 (I)*

## TermoStat Web 3.0

### New corpus

File  advencor.txt

Language  ▾

Extraction

single-word terms

multi-word terms (no

Categories (0) ▾

- adjectives
- adverbs
- nouns
- verbs

All corpora submitted to TermoStat MUST be in **PLAIN TEXT** (no PDF, no Microsoft Word) format.  
Please convert DOC files using Word and PDF using Adobe Acrobat or some other tool.

<http://termostat.ling.umontreal.ca/>

- Free access
- Five languages: English, Portuguese, French, Italian and Spanish
- A source for specialized terminology extraction → single terms and multi-lexical terms

# 3.3.1. CORPUS MANAGEMENT: TERMOSTAT WEB 3.0 (II)

Results				
<span>List of terms</span> <span>Cloud</span> <span>Stat</span> <span>Structuration</span> <span>Bigrams</span>				
Candidate (grouping variant)	Frequency (Specificity)	Score +	Variants	Pattern
<b>adventure</b>	2974	129.02	<i>adventure</i> <i>adventures</i>	Common_Noun
<b>trip</b>	2368	103.89	<i>trip</i> <i>trips</i>	Common_Noun
<b>activity</b>	3305	102.67	<i>activity</i> <i>activities</i>	Common_Noun
<b>tour</b>	2096	94.95	<i>tour</i> <i>tours</i>	Common_Noun
<b>mountain</b>	1794	92.05	<i>mountain</i> <i>mountains</i>	Common_Noun
<b>trail</b>	1498	90.96	<i>trail</i> <i>trails</i>	Common_Noun
<b>experience</b>	2194	78.65	<i>experience</i> <i>experiences</i>	Common_Noun
<b>guide</b>	1297	77.2	<i>guide</i> <i>guides</i>	Common_Noun
<b>kayak</b>	795	68.49	<i>kayak</i> <i>kayaks</i>	Common_Noun
<b>jump</b>	940	67.16	<i>jump</i> <i>jumps</i>	Common_Noun
<b>hike</b>	776	65.28	<i>hike</i> <i>hikes</i>	Common_Noun
<b>trek</b>	707	63.64	<i>trek</i> <i>treks</i>	Common_Noun

- Durán Muñoz, I. (2019). Adjectives and their keyness. A corpus-based analysis in English tourism. *Corpora*, 14(3). 351–378.
- 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.



## 3.3.2. CORPUS MANAGEMENT: *SKETCH ENGINE* (I)

**DASHBOARD** Advencor

**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
- Parallel Concordance**  
Translation search
- Wordlist**  
Frequency list
- N-grams**  
Multiword expressions (MWEs)
- Keywords**  
Terminology extraction
- Trends**  
Diachronic analysis, neologisms
- Text type analysis**  
Statistics of the whole corpus
- OneClick Dictionary**  
Automatic dictionary drafting

<https://www.sketchengine.eu/>

- Free institutional access
- Over 90 languages
- A tool for corpus building, management and exploration
- 11 functions

# 3.3.2. CORPUS MANAGEMENT: SKETCH ENGINE (II)

WORD SKETCH Advencor

climb as verb 1,980x

Get more space

modifiers of "climb"

<b>steeply</b>	5	10.16	...
climbs steeply			
<b>gradually</b>	4	9.64	...
Climb mostly gradually			
<b>up</b>	9	9.58	...
Climb up			
<b>easily</b>	7	9.44	...
capable of easily climbing leaning adjustable ladders			
<b>uphill</b>	3	9.43	...
climbs uphill			
<b>easy</b>	3	9.41	...
easy climbs			
<b>first</b>	5	9.13	...
world , successfully first climbed in 1952 by			
<b>directly</b>	3	8.98	...
climb directly the			
<b>then</b>	8	8.52	...
then climb			
<b>usually</b>	4	8.37	...
We usually climb in the Langdale			
<b>down</b>	3	8.28	...
down climbing			
<b>all</b>	4	8.24	...
all climbed into the			

objects of "climb"

<b>wall</b>	40	10.34	...
climbing wall			
<b>route</b>	35	9.83	...
climbing routes			
<b>technique</b>	23	9.54	...
climbing techniques and			
<b>mountain</b>	24	9.52	...
climb the mountain			
<b>area</b>	29	9.48	...
climbing areas			
<b>experience</b>	37	9.41	...
climbing experience			
<b>gear</b>	19	9.15	...
climbing gear			
<b>guide</b>	20	9.05	...
climbing guides			
<b>Ferrata</b>	15	9	...
hrs Climbing the Via Ferrata in			
<b>harness</b>	16	8.98	...
climbing harness			
<b>shoe</b>	16	8.97	...
climbing shoes			
<b>peak</b>	16	8.92	...
climbing peak			

subjects of "climb"

<b>rock</b>	204	13	...
rock climbing			
<b>ice</b>	69	11.76	...
ice climbing			
<b>sport</b>	13	9.22	...
sport climbs			
<b>hrs</b>	10	9.15	...
hrs Climbing the Via Ferrata			
<b>rope</b>	10	9.02	...
rope climbing			
<b>mountain</b>	11	8.97	...
mountain climbing			
<b>pitch</b>	8	8.81	...
pitch climbs			
<b>peak</b>	8	8.71	...
peak climbing			
<b>steep</b>	7	8.65	...
touring " with steep climbs and descents			
<b>Rock</b>	7	8.62	...
Rock climbing			
<b>beginner</b>	6	8.4	...
Our guided half-day beginner climbing and rappelling program			
<b>Peak</b>	6	8.38	...
Peak climbing in Nepal			

- Jiménez Navarro, E. L. (2020). *Treatment and representation of verb collocations in the specialized language of adventure tourism*. Doctoral dissertation. Universidad de Córdoba, Spain.

- 'Keywords' function
- 'Word Sketch' function
- 'Concordance' function

4

# CONCLUSIONS AND FURTHER RESEARCH

# 4. CONCLUSIONS

- ❑ Importance of corpus design and compilation steps
- ❑ Re-use of specialized corpus for different purposes:
  - Terminological analysis of adjectives and motion verbs
  - Phraseological analysis of collocations of motion verbs
  - Development of the specialized dictionary *DicoAdventure* (in progress)
- ❑ ADVENCOR in another language: Spanish

## 4. FURTHER RESEARCH WITH ADVENCOR

1. Analysis of other word types, such as nouns and adverbs
2. Study and classification of other types of verbs
3. Extraction of the collocations of other terms (e.g., nouns) and types of verbs
4. Contrastive analysis of different features with Spanish
5. ...

---

5

REFERENCES

---

# 5. REFERENCES

- ❑ Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8(4), 243–257.
- ❑ *DicoAdventure*. An online dictionary of adventure tourism. (2021). Montréal, Canada: OLST.
- ❑ Durán Muñoz, I. (2019). Adjectives and their keyness. A corpus-based analysis in English tourism. *Corpora*, 14(3), 351–378.
- ❑ 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.
- ❑ Fennell, D. (2003). *Ecotourism. Second edition*. London/New York, England/NY: Routledge.
- ❑ Jiménez-Navarro, E. L. (2020). *Treatment and representation of verb collocations in the specialized language of adventure tourism*. Doctoral dissertation. Universidad de Córdoba, Spain.
- ❑ Sinclair, J. (2003). 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.
- ❑ *Sketch Engine* [Computer software]. Brno, Czech Republic: Lexical Computing Limited.
- ❑ *TermoStat Web 3.0* [Computer software]. Montréal, Canada: Patrick Drouin.



**THANK YOU  
FOR YOUR ATTENTION**



ISABEL DURÁN MUÑOZ

[iduran@uco.es](mailto:iduran@uco.es)

EVA LUCÍA JIMÉNEZ-  
NAVARRO

[lucia.jimenez@uco.es](mailto:lucia.jimenez@uco.es)

Universidad de Córdoba,  
Spain

# ADVENCOR, the corpus of adventure tourism: Design, Compilation and Management