ISABEL DURÁN MUÑOZ

iduran@uco.es

EVA LUCÍA JIMÉNEZ-NAVARRO

lucia.jimenez@uco.es

Universidad de Córdoba, Spain

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

#### CONTENTS

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

#### 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

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

3

### ADVENCOR CORPUS

#### 3.1. CORPUS DESIGN

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

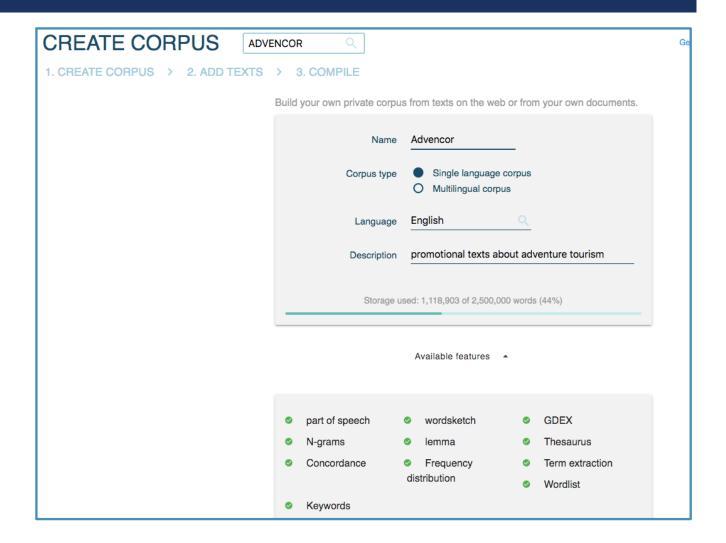
- I) time of publication
- 2) geography
- 3) text type
- 4) text length
- 5) level of specialization

#### 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



#### 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

Number of tokens	1,045,755		
Number of types	49,988		
Type of corpus	Specialized/electronic		
Size	Small		
Mode	Written		
Language	English		
Domain / Subdomain	Tourism / 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		

3.3.

### CORPUS MANAGEMENT

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

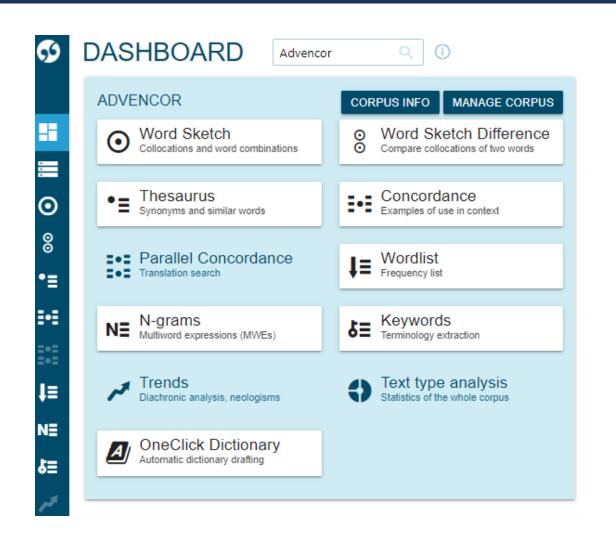


## 3.3.1. CORPUS MANAGEMENT: TERMOSTAT WEB 3.0 (II)

Results							
List of terms Cloud Stat Structuration Bigrams Score							
Candidate (grouping variant)	Frequency		Variants	Pattern			
adventure	2974	129.02	adventure adventures	Common_Noun			
trip	2368	103.89	trip trips	Common_Noun			
activity	3305	102.67	activity activities	Common_Noun			
tour	2096	94.95	tour tours	Common_Noun			
mountain	1794	92.05	mountain mountains	Common_Noun			
trail	1498	90.96	trail trails	Common_Noun			
experience	2194	78.65	experience experiences	Common_Noun			
guide	1297	77.2	guide guides	Common_Noun			
kayak	795	68.49	kayak kayaks	Common_Noun			
jump	940	67.16	jump jumps	Common_Noun			
hike	776	65.28	hike hikes	Common_Noun			
trek	707	63.64	trek treks	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 lexicosemantic approach. A corpusbased analysis of adventure tourism. *Terminology*, 26(1). 33–59.

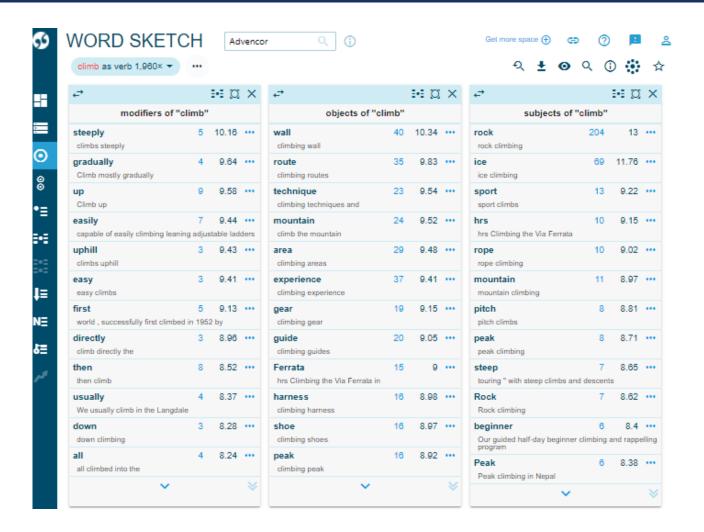
#### 3.3.2. CORPUS MANAGEMENT: SKETCH ENGINE (I)



https://www.sketchengine.eu/

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

#### 3.3.2. CORPUS MANAGEMENT: SKETCH ENGINE (II)



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 Tork, 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

EVA LUCÍA JIMÉNEZ-NAVARRO

lucia.jimenez@uco.es

Universidad de Córdoba, Spain

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