

David Finbar Brett

University of Sassari
dbrett@uniss.it
ORCID: 0000-0003-4285-0604

Antonio Pinna

University of Sassari
dedalo@uniss.it
ORCID: 0000-0001-6608-2246

Barbara Loranc

University of Bielsko-Biala
bloranc@ubb.edu.pl
ORCID: 0000-0003-2495-3482

**Collocates Pairs and Collocations in Travel Journalism
in English, Italian and Polish****1. Introduction**

The considerable growth of the tourism industry in the past decades has been strongly influenced by the importance given to leisure travel in affluent societies. This has led to an increased production of travel-related publications to cater to the expanding global audience's interests in travel activities and experiences. Pirolli¹ points out that prospective travellers now heavily rely on these media outlets for information and advice when making decisions about their leisure trips. Travel journalism thus plays a vital role in connecting the tourism industry's services with an enthusiastic global audience, shaping destination images, and inspiring potential tourists. As a type of lifestyle journalism, travel articles have a

¹ Bryan Pirolli, *Travel Journalism: Informing Tourists in the Digital Age* (London: Routledge, 2019):3.

distinct commercial and service-oriented nature². Their primary purpose is to provide practical and valuable information about travel destinations and services. Pirolli³ highlights the significance of contextualising firsthand travel accounts to generate expectations among potential tourists. In this sense, an author's representation of the destination contributes to both soliciting her/his readers' desire to live out the experiences narrated in the report and possibly framing her/his readers' interpretation of the future holiday. Information, persuasion and guidance are thus interwoven in travel journalism output and render it an interesting site of research of contemporary market-oriented communication.

The language of tourism can demonstrate a linguistic and socio-cultural perspective of a given culture⁴. Because the perceptions of reality and its attractiveness may differ from one culture to another, these variations can be observed not only in the choice of words used, but also in cultural orientations. Therefore to translate the meaning successfully into the target language, it is not sufficient to identify translation equivalents at the linguistic level, as it is important to consider cultural orientations as well. For example, Manca⁵ analysed the differences between the British and the Italian cultures in relation to the way holiday offers at farmhouses, hotels and campsites were promoted. She showed that, while Italian descriptions utilised abstract nouns and metaphorical descriptions to appeal to the customers' five senses, the English ones relied on factual and concrete descriptions. She concluded that a literal translation would turn out to be inadequate and fail to reach the intended audience, as the texts would ignore the target-culture stylistic conventions and aesthetic traditions.

Although travel journalism has attracted some recent academic interest in the areas of media and journalism studies,⁶ the genre of travel reportage

² See: Folker Hanusch and Elfriede Fürsich, eds., *Travel Journalism: Exploring Production, Impact, and Culture* (Basingstoke: Palgrave Macmillan, 2014); Bryan Pirolli, *Travel Journalism: Informing Tourists in the Digital Age* (London: Routledge, 2019); Ben Cocking, *Travel Journalism and Travel Media: Identities, Places, and Imaginings* (London: Palgrave, 2020).

³ Pirolli, *Travel Journalism: Informing Tourists in the Digital Age*.

⁴ David Katan, *Translating Cultures: An Introduction for Translators, Interpreters, and Mediators* (New York: Routledge, 2014).

⁵ Emanuela Manca, "The Language of Tourism in English and Italian: Investigating the Concept of Nature between Culture and Usage," *ESP Across Cultures* 1 (2004): 53–65.

⁶ Folker Hanusch and Elfriede Fürsich, eds., *Travel Journalism: Exploring Production, Impact, and Culture*; Bryan Pirolli, *Travel Journalism: Informing Tourists in the Digital Age*; Ben Cocking, *Travel Journalism and Travel Media: Identities, Places, and Imagining*.

is still under-researched from a linguistic point of view.⁷ While various semiotic techniques contribute to the effectiveness of these publications from a communicative standpoint, analysing their written content and form helps us to understand important linguistic strategies commonly employed to portray leisure travel destinations with the concurring aims to capture readers' attention, generate interest, and ultimately convert them into consumers of specific tourist services. As a contribution to covering this gap in research and, at the same time, extending the investigation of travel journalism output to different socio-cultural contexts of production and consumption, we focus on collocate pairs and collocations in travel reports in three European languages, English, Italian and Polish.

The study of collocation is a useful way of uncovering specific recurrent semantic associations that are at work within a certain socio-cultural context, be it a community of practice restricted to the journalistic profession or larger, and possibly fuzzier, social groups based on shared interests (e.g. readers of travel accounts and amateurs reporting on their trips), language and/or culture. In this paper our goal is the identification of corresponding collocations in English, Italian and Polish, which may contribute to the recognition of common discursive practices within this profession in the West, but also highlight some variation in accordance with assumed differences in sensitivity and taste among the target audiences.

2. Collocate pairs and collocations

The earliest attested use of *collocation* with a meaning akin to that used in contemporary linguistics dates to Palmer's celebrated *Second Interim Report*

⁷Recent corpus-assisted linguistic studies have been carried out by the authors of the present paper, namely: David Brett and Antonio Pinna, "Patterns, Fixedness, and Variability: Using PoS-grams to Find Phraseologies in the Language of Travel Journalism" *Procedia. Social and Behavioral Sciences* 198 (2015): 52–57; David Brett, "Social Network Analysis and the Analysis of Collocations in the Language of Travel Journalism," in *Reiseführer – Sprach- und Kulturmittlung im Tourismus*, ed. Tania Baumann (Bern: Peter Lang, 2018), 183–205; David Brett and Antonio Pinna, "Using Sem-grams to Study Metaphors in Travel Journalism," in *Stranijeziciturizam. Foreign Languages and Tourism*, ed. Vlado Sušac, Tomislav Krpan, and Ivana Lozo (Zadar: Morepress, 2020), 13–29; David Brett, Barbara Loranc-Paszyk, and Antonio Pinna, "A Corpus-Driven Analysis of Adjective/Noun Collocations in Travel Journalism in English, Italian, and Polish," *MonTI* 13 (2021): 114–147; Antonio Pinna, "Evaluative Phraseologies in Travel Journalism: A Look at The Guardian Travel Section," *Annals of the University of Craiova. Series Philology-English* 21, no. 1–2 (2020): 226–242; Antonio Pinna and David Brett, "Phraseologies Realizing Synthetic Personalization in Travel Articles," in *Le Parole del Turismo*, ed. Lorenzo Devilla and Marta GaliñanesGallen (Alessandria: Ed. dell'Orso, 2022), 161–176.

on *English Collocations* in 1933.⁸ However, after almost a century of work in the field the term still does not have a clear, unanimously agreed upon meaning. To the contrary, Gries⁹ suggests that “the notion of ‘collocation’ is probably best characterized as a radial category whose different senses are related to each other and grouped around one or more somewhat central senses, but whose senses can also be related to each other only rather indirectly”. Similarly, McEnery and Hardie report “a great multitude of different definitions” for the term.¹⁰ Evert¹¹ makes the distinction between empirical and theoretical concepts when discussing the meaning of collocation. The former regards quantifiable co-occurrences of lexical items that can be observed in natural language. The latter are “idiosyncratic multiword expressions, defined by linguistic tests and speaker intuitions”. While in the future AI may assist in the identification of such idiosyncratic patterning, at present they cannot be elicited by automatic processes. In the following discussion co-occurrences that are identified by means of a statistical test for association will be referred to as collocate pairs. The term *collocation* will only be used when one or both of the words is “lexically determined and has a modified or bleached meaning”.¹² Hence, the results of the automatic procedure described in the methodology section are not to be considered collocations, but rather potential collocations, and only subsequent analysis by the research team will be able to ascertain whether or not they are indeed so. The results and discussion of the present study will initially focus on collocate pairs and their semantic groupings. Subsequently, *stricto sensu* collocations will be illustrated, also making reference to their occurrences in the texts.

Hence, the research questions of this study are the following:

- 1) What are the most widely dispersed statistically significant collocate pairs in travel journalism in English, Italian and Polish?

⁸ Harold E. Palmer, “Second interim report on English collocations,” in the *Tenth Annual Conference of English Teachers under the Auspices of the Institute for Research in English Teaching* (Tokyo: Institute for Research in English Teaching, 1933).

⁹ Stefan Th. Gries, “Analysing Dispersion,” in *A Practical Handbook of Corpus Linguistics* (Cham: Springer International Publishing, 2021), 138.

¹⁰ Tony McEnery and Andrew Hardie. *Corpus linguistics: Method, theory and practice*. (Cambridge: Cambridge University Press, 2011), 123.

¹¹ Stefan Evert, “Corpora and Collocations,” in *Corpus Linguistics: An International Handbook*, ed. Anke Lüdeling et al. (Berlin: De Gruyter, 2009).

¹² Stefan Evert, “Corpora and Collocations”, 1214.

- 2) Can fully-fledged collocations be identified amongst the collocate pairs identified in RQ1?

3. Methodology

In order to investigate the research questions formulated in the previous section, we used three comparable corpora of travel journalism in English, Italian and Polish, which had been compiled for a previous study.¹³ For English, articles published over a period of seven years were taken from the ‘Travel’ section of *The Guardian*.¹⁴ The Italian section was composed of texts from the ‘Viaggi’ [Travels] section of *La Repubblica*¹⁵ and the Polish section was composed of articles from the ‘Podróże’ [Travels] section of *Gazeta*.¹⁶ The three newspapers are of a comparable standing in each culture, as they are all considered to be quality publications, catering to a middle-class readership, with medium-high education. The articles were saved using a semi-automatic procedure described in detail in the aforementioned work by the authors. The resulting corpus consisted of three 1M-word sections. Some variability was noted in the composition of these sub-corpora: the English, Italian and Polish sections were composed of 1204, 725, and 1084 articles, respectively. Hence, the Italian section contained articles that were on average longer (1379 tokens), than those of the English (830 tokens) and Polish (922 tokens).

The sub-corpora were then annotated for part of speech (PoS) and lemma with the TreeTagger tool¹⁷. Collocate pairs (4L 4R) featuring lexical lemmas (adjectives, common nouns and lexical verbs)¹⁸ were then extracted from the three corpora by way of a tailor-made Perl script, featuring the application of the mutual information (MI) test. Only the collocate pairs that scored 3 or more on the MI test were considered statistically significant. The procedure adopted is analogous to that of the

¹³ David Brett, Barbara Loranc-Paszylk, and Antonio Pinna, “A Corpus-Driven Analysis of Adjective/Noun Collocations in Travel Journalism in English, Italian, and Polish”.

¹⁴ <http://www.theguardian.com>

¹⁵ <https://www.repubblica.it/>

¹⁶ <https://www.gazeta.pl/>

¹⁷ <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

¹⁸ Each unit tested for co-occurrence was composed of a lemma followed by underscore+PoS tag (e.g. town_NN). Note that the tags were modified so that all nouns were normalised into singulars, and all adjectives and verbs into base forms.

Concgram program,¹⁹ i.e. collocations are extracted in a “grammatically blind” manner,²⁰ rather than undergoing syntactic constraints. Given that dispersion is a stronger and more unique predictor of word naming and lexical choice than token frequency,²¹ a dispersion filter was then applied to the results (occurrence in at least 20 texts) to avoid taking into consideration collocate pairs present in high concentrations in only a small number of texts. The utility of such a filter can be illustrated with an example from the data: there are 41 instances of the statistically significant collocate pair AVERE_VER+DIRITTO_NOM [HAVE+RIGHT]. Other pairs with the same frequency are LIVELLO_NOM+MARE_NOM [SEA+LEVEL], CULTURA_NOM+CAPITALE_NOM [CULTURE+CAPITAL], PISTA_NOM+SCI_NOM [SKI+SLOPE]. However, the former is present only in 7 texts, and furthermore, 34 of its 41 instances are all from a single text presenting the constitution of the Republic of Užupis in Vilnius. The latter group of collocate pairs instead feature in a number of texts ranging from 28 to 33. Therefore, placing a dispersion filter avoids giving undue coverage to features that are certainly frequent, but only in very limited segments of the corpus. The results were subsequently analysed and classified by semantic field. The lists of collocate pairs for each language were initially tagged for semantic field in an independent manner. The lists of categories were then compared, and a single set of categories was agreed on for the three languages.

4. Results and discussion

Before proceeding with a qualitative analysis of the results, some quantitative data will be presented. After applying the dispersion filter of presence in at least 20 different texts, 251, 318 and 309 collocate pairs remained in the English, Italian and Polish results, respectively.

It is of interest to note that reducing the dispersion threshold to presence in at least 10 texts would have increased the results to be analysed

¹⁹ Christopher Greaves, *ConcGram 1.0* (Amsterdam: Benjamins, 2009).

²⁰ Adriano Ferraresi, Silvia Bernardini, Giuseppe Picci, and Marco Baroni, “Web Corpora for Bilingual Lexicography: A Pilot Study of English/French Collocation Extraction and Translation,” in *Using Corpora in Contrastive and Translation Studies*, ed. Richard Xiao (Newcastle upon Tyne: Cambridge Scholars Publishing, 2010), 338.

²¹ Gries, Stefan Th. “Analysing Dispersion.” In *A Practical Handbook of Corpus Linguistics*, eds Magali Paquot and Stefan Th. Gries (Cham: Springer International Publishing, 2021), 99–118.

roughly fourfold, as the collocate pairs would in that case number 996, 1036 and 1366, for English, Italian and Polish, respectively. Another point that needs to be made concerns the differences in composition of the subsections of the corpus. Given that there were considerably more texts in the English (1204) than in the Italian section (725), it was therefore “easier” for the English collocate pairs to meet the criteria. Nevertheless, as reported above, there were 67 (27%) more collocate pairs in the Italian section, suggesting that it is somewhat more formulaic.

Collocate pairs and their thematic categories

The categories identified in the data were the following ACCOMMODATION, FOOD, NATURE, SPORT, CULTURE+HISTORY, PEOPLE, PLACE, TRANSPORT, LOGISTICS, PRICE, ACTIVITY, EVALUATION, ROUTE, BUSINESS, and TIME. A substantial number of collocations could not be placed into any particular category and hence they were added to an OTHER category. The following analysis focuses just on the first five of the above list, partly because they constituted the most compact and easily identifiable groupings. It is important to highlight that the following discussion focuses on collocate pairs composed of lemmas (and PoS tags), not on word forms, so adjectives are presented in their base (i.e. masculine) form in Italian and Polish, irrespective of the gender of the noun constituting the other component of the pair.

4.1. NATURE

The NATURE category consists of collocate pairs to do with the natural environment, including landscape features, vegetation and weather.

| | |
|---------|--|
| English | PARK_NN+NATIONAL_JJ (70), AIR_NN+OPEN_JJ (52), RESERVE_NN+NATURE_NN (41), MOUNTAIN_NN+HIGH_JJ (35), BEACH_NN+SANDY_JJ (34), SET_VV+SUN_NN (28), WATER_NN+CLEAR_JJ (28), ISLAND_NN+SMALL_JJ (24), SAND_NN+WHITE_JJ (27), BEACH_NN+WHITE_JJ (20), PEAK_NN+HIGH_JJ (21), FOREST_NN+PINE_NN (19) |
|---------|--|

| | |
|---------|--|
| Italian | <p>CRISTALLINO_ADJ+ACQUA_NOM (31), LAGO_NOM+ACQUA_NOM (23), MARE_NOM+AFFACCIARE_VER (33), NATURALE_ADJ+AMBIENTE_NOM (24), NATURALE_ADJ+AREA_NOM (25), PROTEGGERE_VER+AREA_NOM (27), PROTETTO_ADJ+AREA_NOM (35), APERTO_ADJ+ARIA_NOM (32), NATURALE_ADJ+BELLEZZA_NOM (30), MONTUOSO_ADJ+CATENA_NOM (21), NATURA_NOM+CONTATTO_NOM (21), FAUNA_NOM+FLORA_NOM (21), FIUME_NOM+FOCE_NOM (25), NATURA_NOM+IMMERGERE_VER (40), PARCO_NOM+INTERNO_NOM (32), MEDITERRANEO_ADJ+MACCHIA_NOM (53), CRISTALLINO_ADJ+MARE_NOM (31), INCONTAMINATO_ADJ+NATURA_NOM (36), SELVAGGIO_ADJ+NATURA_NOM (23), MOZZAFIATO_ADJ+PANORAMA_NOM (26), NATURALE_ADJ+PARCO_NOM (92), NAZIONALE_ADJ+PARCO_NOM (138), REGIONALE_ADJ+PARCO_NOM (22), MARE_NOM+PICCO_NOM (31), NATURALE_ADJ+RISERVA_NOM (69), FIUME_NOM+RIVA_NOM (31), MARE_NOM+RIVA_NOM (25), BIANCO_ADJ+SABBIA_NOM (35), ROCCIA_NOM+SCAVARE_VER (29), ACQUA_NOM+SPECCHIO_NOM (24), BIANCO_ADJ+SPIAGGIA_NOM (26), MARE_NOM+VISTA_NOM (26)</p> |
| Polish | <p>PANORAMA_NN+WSPANIAŁY_ADJ (21), NARODOWY_ADJ+PARK_NN (248), SZCZYT_NN+WYSOKI_ADJ (73), ZIEMIA_NN+TRZĘSIENIE_NN (59), WODA_NN+CIEPŁY_ADJ (46), GÓRA_NN+WYSOKI_ADJ (42), PLAŻA_NN+PIĘKNY_ADJ (40), WIDOKOWY_ADJ+PUNKT_NN (36), SŁOŃCE_NN+ZACHÓD_NN (40), JEZIORO_NN+BRZEG_NN (33), WODA_NN+CZYSTY_ADJ (32), POWIETRZE_NN+ŚWIEŻY_ADJ (30), RZĘKA_NN+DOLINA_NN (31), SŁOŃCE_NN+WŚCHÓD_NN (33), NIEBO_NN+GOŁY_ADJ (36), WIDOK_NN+ROZCIĄGAĆ_VER (27), KRAJOBRAZOWY_ADJ+PARK_NN (36), RZĘKA_NN+BRZEG_NN (29), WIDOK_NN+ROZTACZAĆ_VER (25), PTAK_NN+GATUNEK_NN (25), PRZYRODA_NN+REZERWAT_NN (25), MINERALNY_ADJ+WODA_NN (29), GÓRSKI_ADJ+PASMO_NN (22), JEZIORO_NN+LAS_NN</p> |

| | |
|--------|--|
| Polish | NN (21), LAS_NN+SOSNOWY_ADJ (23), TEMPERATURA_NN+WYSOKI_ADJ (26), POGODA_NN+DOBRY_ADJ (37), STOPIEŃ_NN+TEMPERATURA_NN (35), WIATR_NN+SILNY_ADJ (27), TEMPERATURA_NN+ŚREDNI_ADJ (58), WYNOSIĆ_VER+TEMPERATURA_NN (45) |
|--------|--|

Table 1. Collocate pairs in the NATURE category in each of the three languages with relative frequencies in brackets.

In this category, the investigation of the English corpus yielded the smallest number of collocate pairs in comparison with the other two corpora. They are all concerned with identifying and describing water-related or land-related attractions (e.g. WATER CLEAR, BEACH SANDY, BEACH WHITE, PARK NATIONAL, MOUNTAIN HIGH, FOREST PINE). The Italian corpus returned collocate pairs that can be grouped into two main classes defined on the basis of their textual function: the first identifies and describes an attraction that is water- or land-related (e.g. ACQUA CRISTALLINO [WATER CRYSTAL-CLEAR], FIUME FOCE [RIVER MOUTH], MARE RIVA [SEA SHORE]; MONTUOSO CATENA [MOUNTAIN RANGE], MEDITERRANEO MACCHIA [MEDITERRANEAN MAQUIS], NATURALE PARCO [NATURAL PARK]), while the second provides a positive evaluation of a destination (e.g. MOZZAFIATO PANORAMA [BREATH-TAKING PANORAMA], NATURALE BELLEZZA [NATURAL BEAUTY], INCONTAMINATO NATURA [UNSPOILED NATURE]). The Polish corpus shows a greater variety of thematic classes, apart from the classes shared with Italian, such as: various water and land-related attractions (e.g. WODA CZYSTY [WATER CLEAN], RZEKA BRZEG [RIVER BANK], LAS SOSNOWY [FOREST PINE], GÓRA WYSOKI [MOUNTAIN HIGH]) and a positive evaluation of the places (PANORAMA WSPANIAŁA [SKYLINE MARVELOUS], PLAŻA PIĘKNA [BEACH BEAUTIFUL]), there are also collocate pairs related to weather and climate (e.g. POGODA DOBRA [WEATHER GOOD], TEMPERATURA WYSOKA [TEMPERATURE HIGH], WODA CIEPŁA [WATER WARM], WIATR SILNY [WIND STRONG]).

4.2. SPORT

| | |
|---------|--|
| English | POOL_NN+SWIMMING_NN (69), BIKING_NN+MOUNTAIN_NN (42), COURSE_NN+GOLF_NN (33), RINK_NN+ICE_NN (45), WATER_NN+WHITE_JJ (46), BOAT_NN+SMALL_JJ (27), DIVING_NN+SCUBA_NN (25), MOUNTAIN_NN+CLIMB_VV (26), RESORT_NN+SKI_JJ (27), BIKE_NN+MOUNTAIN_NN (22), CLIMB_VV+ROCK_NN (26) |
| Italian | SCIISTICO_ADJ+COMPRESORIO_NOM (29), RISALITA_NOM+IMPIANTO_NOM (44), PISTA_NOM+KM_NOM (39), BIKE_NOM+MOUNTAIN_VER (35), CICLABILE_ADJ+PISTA_NOM (65), SCI_NOM+PISTA_NOM (41), FONDO_NOM+SCI_NOM (45), INVERNALE_ADJ+SPORT_NOM (21) |
| Polish | ROWEROWY_ADJ+ŚCIEŻKA_NN (30), SPORT_NN+UPRAWIAĆ_VER (34), NARCIARSKI_ADJ+TRASA_NN (39), NARTA_NN+JAZDA_NN (28), WODNY_ADJ+SPORT_NN (30), ROWEROWY_ADJ+TRASA_NN (25), TURYSTYCZNY_ADJ+SZLAK_NN (31), NARCIARSKI_ADJ+STACJA_NN (27), SPRZĘT_NN+WYPOŻYCZALNIA_NN (23), NARCIARSKI_ADJ+OŚRODEK_NN (57) |

Table 2. Collocate pairs in the SPORT category in each of the three languages with relative frequencies in brackets.

In this category, the English corpus returned collocate pairs dealing with a larger variety of open air sports activities than that found in the other two corpora, from winter sports (RESORT SKI, RINK ICE) to cycling (BIKING MOUNTAIN, BIKE MOUNTAIN) to water-related activities (POOL SWIMMING, DIVING SCUBA), mountain climbing (MOUNTAIN CLIMB, CLIMB ROCK) and golf (COURSE GOLF). The Italian corpus yielded collocate pairs dealing with two main sports activities, i.e. winter sports (e.g. INVERNALE SPORT [WINTER SPORT], SCI PISTA [SKI SLOPE], RISALITA IMPIANTO [LIFT SKI]), and cycling (BIKE MOUNTAIN, CICLABILE PISTA [BIKE LANE]). In a similar vein, the collocate pairs found in the Polish corpus can be classified into the winter sports group (e.g. NARCIARSKI TRASA [SKI ROUTE]; NARCIARSKI OŚRODEK [SKI CENTRE]), the cycling one (e.g. ROWEROWY ŚCIEŻKA

[CYCLOPATH], ROWEROWA TRASA [BIKE ROUTE]) and there is just one collocate pair related to the water sport category (e.g. WODNY SPORT [WATER SPORT]).

4.3 ACCOMMODATION

| | |
|---------|---|
| English | ACCOMMODATION_NN+INCLUDE_VV (126), NIGHT_NN+SPEND_VV (53), NIGHT_NN+OFFER_VV (77), ROOM_NN+DOUBLE_JJ (49), ACCOMMODATION_NN+NIGHT_NN (42), STAY_VV+PLACE_NN (35), HOTEL_NN+BOUTIQUE_NN (38), HOTEL_NN+STAR_NN (40), ACCOMMODATION_NN+BOARD_NN (39), ROOM_NN+DINING_NN (47), ROOM_NN+HOTEL_NN (38), CATERING_NN+SELF_JJ (35), CATERING_NN+SELF_NN (41), SHARING_NN+BASE_VV (45), STAY_NN+NIGHT_NN (35), HOUSE_NN+COUNTRY_NN (33) ACCOMMODATION_NN+TRANSFER_NN (29), NIGHT_NN+PERSON_NN (33), INCLUDE_VV+SHARING_NN (28), ACCOMMODATION_NN+WEEK_NN (27), CABIN_NN+LOG_NN (25) |
| Italian | RICETTIVO_ADJ+STRUTTURA_NOM (28) |
| Polish | TANI_ADJ+HOTEL_NN (48), HOTEL_NN+LUKSUSOWY_ADJ (42), HOTEL_NN+NOCLEG_NN (29), NOCLEGOWY_ADJ+MIEJSCE_NN (26), ŁAZIENKA_NN+POKÓJ_NN (23), |

Table 3. Collocate pairs in the ACCOMMODATION category in each of the three languages with relative frequencies in brackets.

In this category, the largest number of collocate pairs (17) is found in the English corpus that comprises thematic areas providing information on the price of accommodation (e.g. ACCOMMODATION WEEK, NIGHT PERSON, STAY NIGHT), the services included (e.g. ACCOMMODATION INCLUDE, ACCOMMODATION BOARD, ACCOMMODATION TRANSFER), the type of accommodation (e.g. CABIN LOG, HOUSE COUNTRY, HOTEL STAR) or a more detailed description of the accommodation (e.g. SELF-CATERING, ROOM DINING, ROOM HOTEL). The Italian corpus provided only one collocate pair (RICETTIVO STRUTTURA[ACCOMMODATION FACILITY]), which as a

superordinate concept functions as a cover term that subsumes various types of accommodation. The Polish corpus, on the other hand, yielded collocate pairs dealing with either the description of the place offering accommodation (e.g. NOCLEG HOTEL [ACCOMMODATION HOTEL], ŁAZIENKA POKÓJ [BATHROOM ROOM]) or the indication of its price (e.g. TANI HOTEL [CHEAP HOTEL], HOTEL LUKSUSOWY [HOTEL LUXURIOUS]).

4.4. CULTURE and HISTORY

| | |
|---------|---|
| English | MUSIC_NN+LIVE_JJ (54), FESTIVAL_NN+MUSIC_NN (53), CENTRE_NN+VISITOR_NN (31), WAR_NN+WORLD_NN (38), WAR_NN+SECOND_JJ (21), WORLD_NN+SECOND_JJ (21) |
| Italian | PERA_NOM+AMMIRARE_VER (22), ARCHEOLOGICO_ADJ+AREA_NOM (27), CONTEMPORANEO_ADJ+ARTE_NOM (104), CULTURA_NOM+ARTE_NOM (33), MODERNO_ADJ+ARTE_NOM (39), LOCALE_ADJ+ARTIGIANATO_NOM (24), INTERNAZIONALE_ADJ+ARTISTA_NOM (20), CULTURALE_ADJ+ARTISTICO_ADJ (24), CULTURALE_ADJ+BENE_NOM (36) CULTURA_NOM+CAPITALE_NOM (41), CULTURALE_ADJ+CENTRO_NOM (39), GOTICO_ADJ+CHIESA_NOM (21), PATRIMONIO_NOM+DICHIARARE_VER (21), CULTURA_NOM+EUROPEO_ADJ (28), CULTURALE_ADJ+EVENTO_NOM (21), ARTE_NOM+GALLERIA_NOM (62), MUSEO_NOM+INTERESSANTE_ADJ (24), DEDICARE_VER+MOSTRA_NOM (27), APRIRE_VER+MUSEO_NOM (31), ARCHEOLOGICO_ADJ+MUSEO_NOM (64), ARTE_NOM+MUSEO_NOM (73), ARTE ARTO_NOM+MUSEO_NOM (34), CIVICO_ADJ+MUSEO_NOM (36), CONTEMPORANEO_ADJ+MUSEO_NOM (23), DEDICARE_VER+MUSEO_NOM (48), NAZIONALE_ADJ+MUSEO_NOM (59), CLASSICO_ADJ+MUSICA_NOM (25), ARTE_NOM+OPERA_NOM (95), ARTISTA_NOM+OPERA_NOM (33), MOSTRA_NOM+OSPITARE_VER (29), MUSEO_NOM+OSPITARE_VER (53), OCHIESA_NOM+PALAZZO_NOM (26), COMUNALE_ADJ+PALAZZO_NOM (24), |

| | |
|---------|---|
| Italian | <p>ARCHEOLOGICO_ADJ+PARCO_NOM (32), ARTISTICO_ADJ+PATRIMONIO_NOM (24), CULTURALE_ADJ+PATRIMONIO_NOM (38), MONDIALE_ADJ+PATRIMONIO_NOM (44), UMANITÀ_NOM+PATRIMONIO_NOM (47), UMANITÀ_NOM+PATRIMONIO_NOM (25), MUSEO_NOM+SEDE_NOM (42), ESPOSITIVO_ADJ+SPAZIO_NOM (30), GOTICO_ADJ+STILE_NOM (30), ARTE_NOM+STORIA_NOM (37), CULTURA_NOM+STORIA_NOM (23), ARTISTICO_ADJ+STORICO_ADJ (21), LOCALE_ADJ+TRADIZIONE_NOM (21), POPOLARE_ADJ+TRADIZIONE_NOM (22), MUSEO_NOM+VISITA_NOM (33), MUSEO_NOM+VISITARE_VER (64), SESSANTA_ADJ+ANNO_NOM (35), MEDIEVALE_ADJ+ANTICO_ADJ (20), ROMANO_ADJ+ANTICO_ADJ (33), TRADIZIONE_NOM+ANTICO_ADJ (35), NATALE_ADJ+CASA_NOM (37), SECOLO_NOM+COSTRUIRE_VER (21), ROMANO_ADJ+EPOCA_NOM (37), SECOLO_NOM+FINE_NOM (38), MONDIALE_ADJ+GUERRA_NOM (58), MONDIALE_ADJ+PRIMO_ADJ (29), STORIA_NOM+RICCO_ADJ (32), SECOLO_NOM+RISALIRE_VER (68), GUERRA_NOM+SECONDO_ADJ (44), MONDIALE_ADJ+SECONDO_ADJ (49), ARCHEOLOGICO_ADJ+SITO_NOM (33)</p> |
| Polish | <p>KOŚCIÓŁ_NN+GOTYCKI_ADJ (39), KOŚCIÓŁ_NN+BAROKOWY_ADJ (35), KLASZTOR_NN+KOŚCIÓŁ_NN (26), SZTUKA_NN+DZIEŁO_NN (47), SZTUKA_NN+MUZEUM_NN (48), MUZEUM_NN+MIEŚCIĆ_VER (36), NARODOWY_ADJ+MUZEUM_NN (25), FESTIWAL_NN+ODBYWAĆ_VER (23), SZTUKA_NN+GALERIA_NN (26), ZABYTEK_NN+WAŻNY_ADJ (24), KONCERT_NN+ODBYWAĆ_VER (24), ZABYTEK_NN+CENNY_ADJ (23), KULTUROWY_ADJ+DZIEDZICTWO_NN (23), WSPÓŁCZESNY_ADJ+SZTUKA_NN (22), ŚWIATOWY_ADJ+WOJNA_NN (120), WOJNA_NN+II_ADJ (80), OBRONNY_ADJ+MUR_NN (69), WOJNA_NN+CZAS_NN (68), WOJNA_NN+I_ADJ (36), WIEK_NN+POŁOWA_NN (51), POMNIK_NN+STAĆ_VER (23), MUR_NN+OTACZAĆ_VER (21), ZAMEK_NN+ŚREDNIOWIECZNY_ADJ (22), ZAMEK_NN+RUINA_NN (40)</p> |

Table 4. Collocate pairs in the CULTURE and HISTORY in each of the three languages with relative frequencies in brackets.

In this category, the English corpus provided a handful of collocate pairs identifying an attraction (FESTIVAL MUSIC, MUSIC LIVE, CENTRE VISITOR) or denoting a historical event (WAR SECOND, WAR WORLD, WORLD SECOND). The largest number of collocate pairs (62) was found in the Italian corpus. These can be grouped into three main thematic areas, dealing with the functions of identifying an attraction (e.g. ARTE MUSEO [ART MUSEUM], GOTICO CHIESA [GOTHIC CHURCH]), defining the age of the attraction (e.g. ROMANO EPOCA [ROMAN EPOCH], SECOLORISALIRE [CENTURYDATE]), or evaluating the destination (e.g. MUSEO INTERESSANTE [MUSEUM INTERESTING], STORIA RICCO [HISTORY RICH]). The first is the largest with 41 collocate pairs which form networks of hyponymic and meronymic relationships. For example, most of the attractions (e.g. ARCHEOLOGICO MUSEO [ARCHAEOLOGICAL MUSEUM], CIVICO MUSEO [CITY MUSEUM], ARTE GALLERIA [ART GALLERY], CHIESA PALAZZO [CHURCH PALACE], NATALE CASA [BIRTH HOUSE], ARCHEOLOGICO PARCO [ARCHAEOLOGICAL PARK], ARCHEOLOGICO SITO [ARCHAEOLOGICAL SITE]) are part of heritage (ARTISTICO PATRIMONIO [ARTISTIC HERITAGE], CULTURALE PATRIMONIO [CULTURAL HERITAGE], UMANITA' PATRIMONIO [HUMANITY HERITAGE], MONDIALE PATRIMONIO [WORLD HERITAGE]). At the same time, these attractions are cultural assets (CULTURALE BENE [CULTURAL ASSET]) and, frequently, exhibition areas (ESPOSITIVO SPAZIO [EXHIBITION AREA]) where works of art of renowned artists (ARTE OPERA [ART WORK], ARTISTA OPERA [ARTIST WORK], INTERNAZIONALE ARTISTA [INTERNATIONAL ARTIST]) are exhibited (DEDICARE MOSTRA [DEDICATE EXHIBITION], MOSTRA OSPITARE [EXHIBITION HOST]), sometimes on special occasions (CULTURALE EVENTO [CULTURAL EVENT]). The second group contains 11 collocate pairs and focuses on the period a certain attraction dates to (e.g. SECOLO COSTRUIRE [CENTURY BUILD], SECOLO FINE [CENTURY END], ROMANO ANTICO [ROMAN ANCIENT]) with the greatest emphasis placed on recent history (e.g. MONDIALE PRIMO [WORLD FIRST], GUERRA SECONDO [WAR SECOND], SESSANTA ANNO [SIXTY YEAR]). Positive assessment of a destination by the ten collocate pairs of the third group may be straightforward

(MUSEO INTERESSANTE [MUSEUM INTERESTING]), expressed in the co-text of the collocate pair (e.g. ARTE CULTURA [ART CULTURE], ARTE STORIA [ART HISTORY], ARTISTICO CULTURALE [ARTISTIC CULTURAL])²², or implicit, for example by strongly advising a visit to a certain museum (e.g. MUSEO VISITA [MUSEUM VISIT], MUSEO VISITARE [MUSEUM VISIT]).²³

The number of collocate pairs (26) in the Polish corpus was halfway between those of the English and Italian corpora. These pairs can be classified into three thematic areas. Similarly to Italian, in the first group, these focus on identifying an attraction (e.g. MUZEUM NARODOWE [MUSEUM NATIONAL], GALERIA SZTUKI [GALLERY ART], KOŚCIÓŁ KLASZTOR [CHURCH MONASTERY], OBRONNY MUR [DEFENCE WALL], ZAMEK RUINA [CASTLE RUIN]) or indicating the age of the attraction (WIEK POŁOWA [CENTURY HALF], WOJNAŚWIATOWY [WAR WORLD]). The second thematic area includes collocate pairs which focus on the characteristic style certain attractions incorporate that can assign them to a specific aesthetic trend (e.g. WSPÓŁCZESNY SZTUKA [CONTEMPORARY ART], KOŚCIÓŁ GOTYCKI [CHURCH GOTHIC], KOŚCIÓŁ BAROKOWY [CHURCH BAROQUE], ZAMEK ŚREDNIOWIECZNY [CASTLE MEDIEVAL]). Interestingly, the most frequent (74) collocate pairs refer to attractions that have been erected for predominantly religious purposes. Finally, the third group of collocate pairs includes a small number of examples referring to positive assessment of an attraction (e.g. ZABYTEK WAŻNY [MONUMENT IMPORTANT], ZABYTEK CENNY [MONUMENT VALUABLE]).

²² These collocate pairs are part of longer lists that provide an overview of what a destination has to offer and are associated with a general positive evaluation. For example, “piacevole meta per un weekend tra arte, cultura, natura e buona cucina” [pleasant destination for a weekend of art, culture, nature and good food], “un mix affascinante di arte, storia e gastronomia” [a fascinating mix of art, history and gastronomy], “un territorio ricco di eccellenze artistiche, culturali e gastronomiche” [an area rich in artistic, cultural and gastronomic excellence].

²³ Circa 50% of these collocate pairs are explicit suggestions for tourists to visit a specific museum. For example, “museo da visitare” [the museum is a must-see], “va visitato” [it must be visited], “è d’obbligo la visita” [a visit is imperative], “non può mancare una visita” [a visit can’t be missed].

4.5. FOOD

| | |
|---------|--|
| English | BOARD_NN+FULL_JJ (89), MEAL_NN+INCLUDE_VV (67), BREAKFAST_NN+INCLUDE_VV (67), BOARD_NN+INCLUDE_VV (58), CREAM_NN+ICE_NN (53), FULL_JJ+INCLUDE_VV (49), BOARD_NN+NIGHT_NN (51), RESTAURANT_NN+BAR_NN (38), BOARD_NN+HALF_NN (33), DRINK_NN+FOOD_NN (38), MEAL_NN+ACCOMMODATION_NN (35), PRODUCE_NN+LOCAL_JJ (34), CHIP_NN+FISH_NN (35), FOOD_NN+ACCOMMODATION_NN (22), ACCOMMODATION_NN+FULL_JJ (22), FARM_NN+ORGANIC_JJ (23), TEA_NN+AFTERNOON_NN (23), WINE_NN+MULL_VV (33) |
| Italian | RISTORANTE_NOM+BAR_NOM (23), TIPICO_ADJ+PIATTO_NOM (32), COLAZIONE_NOM+PRIMO_ADJ (29), LOCALE_ADJ+PRODOTTO_NOM (34), TIPICO_ADJ+PRODOTTO_NOM (52) |
| Polish | TANI_ADJ+ZJEŚĆ_VER (42), RESTAURACJA_NN+HOTEL_NN (40), BAR_NN+RESTAURACJA_NN (38), MORZE_NN+OWOC_NN (44), RESTAURACJA_NN+DOBRY_ADJ (26), GĘBA_NN+NIEBO_NN (23), WODA_NN+PIĆ_VER (22) |

Table 5. Collocate pairs in the FOOD category in each of the three languages with relative frequencies in brackets.

In this category, the English corpus shows the largest number (19) of collocate pairs in comparison with the other two corpora. The most comprehensive thematic area is the one dealing with the services included in a specific holiday offer (e.g. BOARD FULL, MEAL INCLUDE, BREAKFAST INCLUDE), which ends up overlapping with the similar group found in the ACCOMMODATION category (see Section 4.3), while the other areas focus on the genuineness of ingredients (FARM ORGANIC, PRODUCE LOCAL), identify places where one can eat or drink (e.g. RESTAURANT BAR), or list types of food/drink that are served (e.g. CREAM ICE, WINE MULL, CHIP FISH). Perhaps unexpectedly, the Italian corpus yielded the smallest number of collocate pairs (5), three of which highlight the authentic character of the dishes and their ingredients (TIPICO PIATTO [TYPICAL DISH], TIPICO PRODOTTO [TYPICAL PRODUCT], LOCALE PRODOTTO [LOCAL PRODUCT]). The Polish

corpus provided collocate pairs either related to description of the place offering food and the name of the dish (e.g. RESTAURACJA HOTEL [RESTAURANT HOTEL], BAR RESTAURACJA [BAR RESTAURANT], MORZE OWOC [SEA FRUIT]), or positive evaluation of the food served (e.g. RESTAURACJA DOBRY [RESTAURANT GOOD], GĘBA NIEBO [MOUTH HEAVEN]).

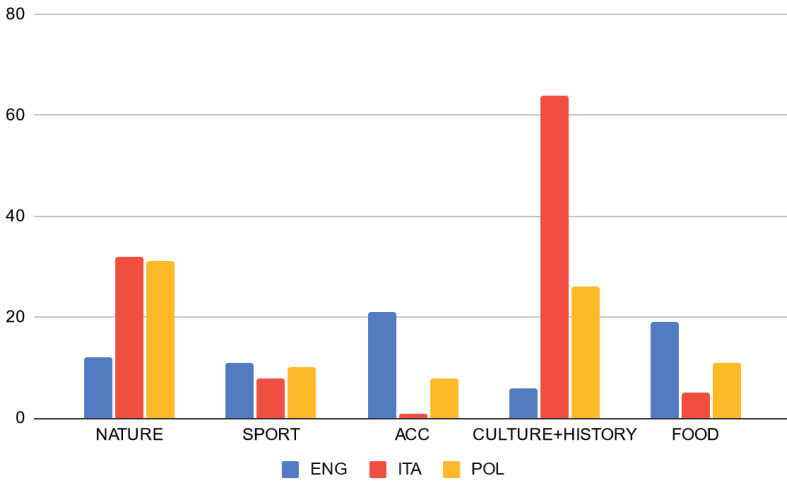


Figure 1. The sum of the collocation types in the five categories in English, Italian and Polish.

Finally, some general considerations can be made about the proportions of these categories in the three sub-corpora. While the number of collocate pairs in the SPORT category is similar in all of the languages, English has the highest numbers in the ACCOMMODATION and FOOD sections, and the lowest in CULTURE + HISTORY. Conversely, Italian has the most in CULTURE + HISTORY and the fewest in ACCOMMODATION and FOOD. Polish is more or less in the middle in all three categories. This may suggest that English travel journalism focuses more on practical matters associated with travelling, whereas Italian concentrates on attractions, particularly heritage sites. Italian and Polish both have almost three times as many collocations connected with NATURE in comparison to English,

which again can be interpreted as an enhanced interest in depicting visually attractive realities for Italian and Polish readers.

4.6. Fully-fledged collocations

Analysis of the collocate pairs led to the identification of a certain number of what could be considered fully-fledged collocations, i.e. combinations that not only show statistically significant co-occurrence, but that are also lexically determined and feature some kind of semantic shift. The following list is merely exemplative and the matter will be returned to in greater depth in future research.

- 1) WHITE_JJ WATER_NN, indicating an outdoor activity (typically rafting) that takes place where waters are rough. For instance: “The white water moments were spectacular: in one stretch of rapids, a team mate flipped out of the raft, landed in the water ahead of us, then disappeared as the dinghy ran him over.”
- 2) BOUTIQUE_NN HOTEL_NN, indicating a small upmarket, possibly themed hotel. For instance: “New entry Mr & Mrs Smith brings a bit of glamour to the list with its mouth watering portfolio of sexy boutique hotels”.
- 3) HALF_NN BOARD_NN (34 instances) and FULL_JJ BOARD_NN (90 instances), indicating that one and two main meals are provided with accommodation. For instance: “Apartments £91 per adult per night half board, £23 per child, including activities”. While the Italian equivalents PENSIONE COMPLETA[FULL BOARD] and MEZZA PENSIONE [HALF BOARD] are attested in the relevant subcorpus, they do not pass the dispersion filter of ≥ 20 texts, as they appear in only 6 and 17 texts, respectively. Polish, on the other hand, has straightforward equivalents of the collocations, that is PEŁNE WYŻYWIENIE [FULL BOARD] and CZĘŚCIOWE WYŻYWIENIE [HALF-BOARD].
- 4) In 14 out of 21 instances, APRIRE_VER+MUSEO_NOM is included in the fixed expression “museo a cielo aperto” [an open air museum] (literally, museum in the open sky). The place in question is not a museum as such, but the implication is that there is such a wealth of heritage that it could be. Almost half of these instances are preceded by the adjective pair “vero e proprio” which functions as a marker of precision, thus pre-empting the possible assumption on the part of the

reader that this description of the destination is an exaggeration. The expression is always accompanied by positive evaluation, for example, “alla scoperta di uno dei gioielli naturali che distingue questo territorio: il museo a cielo aperto di Arte Sella” [discover one of the natural jewels that distinguishes this territory: the Arte Sella open air museum].

- 5) The theme of the great outdoors is shared by all three sections of the corpus. There were 52 instances of AIR_NN+OPEN_JJ in English, usually associated with places of social gathering and their typical activities (e.g. restaurants, theatres, parties), such as: “gales have forced the last minute cancellation of its open air street party, the main event which attracts 100,000 partygoers”. There were 32 instances of APERTO_ADJ+ARIA_NOM in the Italian section, usually employed to characterise a holiday experience or activity involving individual physical movement (e.g. hiking or cycling), and associated with positive evaluation, for example “il luogo ideale per chi ama le attività montane all’aria aperta”. There were 30 instances of the Polish POWIETRZE_NN+ŚWIEŻY_ADJ (30) in most cases it is used to describe spending time outdoors as a part of a healthy lifestyle, for example “chętniej wybieramy aktywny wypoczynek na świeżym powietrzu zamiast leniwego siedzenia w domu”. Similarly to English, the collocation is also used to describe a social event or eating out organised outside the building because of warm weather. The Polish corpus also included 36 instances of the collocate pair NIEBO_NN+GOŁY_ADJ which are all realisations of the collocation *pod gołym niebem* [UNDER THE NAKED SKY] meaning “outdoors, not inside a building”. It generally refers to cultural or social events taking place in the open air, for example, “Jak co roku wiele osób spędzi Sylwestra pod gołym niebem na Krupówkach”.
- 6) All 21 instances of the collocate pair GĘBA_NN+NIEBO_NN form a collocation *niebo w gębie* [HEAVEN IN THE MOUTH] which is used to describe food that tastes heavenly. It reflects a humorous but very positive take on evaluating food. In fact, associating the word *niebo* [HEAVEN] with the colloquial word for mouth, *gęba*, similarly to the English *gob*, makes the expression almost an oxymoron. Interestingly, in all instances it collocates with the verb *feel* which underlines the sensual delight of tasting great food, for example, “można tu poczuć niebo w gębie, próbując miejscowych potraw”.

5. Conclusions

This brief study of collocate pairs in travel journalism in English, Italian and Polish, allows us to make some tentative conclusions about the different angles taken by journalists in these three languages. On the basis of the results obtained, the most marked difference between the texts in these three languages concerns the preponderance of collocate pairs connected with CULTURE in the Italian section, followed by the large quantity of pairs connected with NATURE in the Italian and Polish sections in comparison to that in the English section. English, on the other hand, features a relatively large number of pairs in the ACCOMMODATION field. This resonates with the conclusions made by Manca (2004) who found Italian descriptions to be rich in abstract nouns and metaphorical descriptions, whereas the English ones were more factual and concrete.

One finding that is to say the least surprising is the low number of collocate pairs in the Italian section dealing with FOOD. Given that Italy is widely considered to be one of the world's undisputed culinary capitals, the fact that there are so few pairs in comparison to English and Polish may suggest that the methodology may be in need of some fine-tuning. The first factor to be addressed is that of the very high dispersion filter of presence in at least 20 texts. In fact, if the threshold is lowered to presence in at least 10 texts, the number of collocate pairs rises from 7 to 53, amongst which we see the appearance of dishes that can only be found in some regions (TARTUFO_NOM BIANCO_ADJ [WHITE TRUFFLE], PESCE_NOM FRESCO_ADJ [FRESH FISH], FRUTTO_NOM MARE_NOM [SEAFOOD]), hyponyms of the PIATTO_NOM+TIPICO_ADJ that made it past the higher threshold, being present in 32 texts. This is to be contrasted with the English FISH_NN+CHIP_NN, a dish that can be found more or less everywhere in Britain, as well as in other localities. Hence, the low number of Italian collocate pairs in the FOOD category may not necessarily be a telltale sign of flawed methodology, but it may instead reflect a culture in which the culinary tradition is very heterogeneous, as every region, even town, has its own *piatto tipico* and this results in relatively few collocate pairs in comparison to cultures with more homogenous traditions.

Caution must certainly be exercised when interpreting the presence of collocate pairs concerning a particular theme as indicative of greater

treatment of that topic in a particular language, as there may be collocation gaps, in which a two- or three-word collocation in one language is expressed with a single word in another, e.g. vested interest (English) > *interesse* (Italian);²⁴ *avvenuta ricezione* (Italian) > receipt (English). The fondness of the English language for compounds also complicates the matter somewhat, as the vagaries in their orthographic representation (space, hyphen, no space) lead to concrete problems in what is considered to be a “token” in corpus linguistics. For example, the Italian *impianto di risalita* and the Polish *wyciąg narciarski* translate into English as *skilift*. Nevertheless, the preponderance of data such as those concerning the CULTURE section in Italian would suggest that there is greater focus on these matters in Italian as opposed to English and Polish travel journalism, a matter to be explored in greater depth in future research using qualitative and quantitative means.

The insights gained from the current study are of use to professional and trainee translators working with the languages studied. Collocations are known to be one of the more challenging aspects of translational action as they “appear arbitrarily to juxtapose nouns with verb-nouns”,²⁵ and the results of the current study provides a list of recurrent word combinations in the three languages that can be consulted to verify whether a particular combination is frequently used in travel journalism in the language in question.

Finally, a clear caveat is where the articles were sourced in the three languages. While it can be argued that the texts were produced by a large number of different authors, nevertheless, they were all harvested from a single newspaper in each language. Future work will focus on extending the corpora to include texts from other newspapers, enhancing the representativeness of the study and allowing us to generalise with more certainty about the results.

²⁴ Christopher Taylor, *Language to Language. A practical and theoretical guide for Italian/English translators* (Cambridge: Cambridge University Press, 1998), 55.

²⁵ Peter Newmark, *A Textbook of Translation* (New York: Prentice Hall, 2003), 146.

David Finbar Brett, Antonio Pinna, Barbara Loranc

Abstract

Using a corpus-driven approach, this paper will address an important aspect of formulaicity in travel journalism, an under-researched genre that has hitherto received scarce attention on the part of academics in general and linguists in particular. In order to compile three 1-million-word comparable corpora, travel articles were downloaded from the online versions of the *Guardian*, *La Repubblica*, and *Gazeta*, to constitute representative samples of the genre in English, Italian and Polish. The texts were annotated for part-of-speech and lemma using TreeTagger. Collocate pairs were then extracted from the annotated data using tailor-made Perl scripts. Subsequently a dispersion filter of presence in at least 20 different texts was applied.

The resulting collocate pairs were placed into thematic groups, including: ACCOMMODATION, FOOD, NATURE, SPORT, CULTURE+HISTORY, PEOPLE, PLACE, TRANSPORT, LOGISTICS, PRICE, ACTIVITY, EVALUATION, ROUTE, BUSINESS and TIME. The first five categories were subjected to further analysis, and it emerged that there were notable differences in their relative proportions between the three languages. The results show a number of collocations that could be identified amongst the collocate pairs in the three languages.

Keywords: collocation, travel journalism, English, Italian, Polish

Słowa kluczowe: kolokacja, dziennikarstwo podróżnicze, angielski, włoski, polski