A CORPUS-DRIVEN ANALYSIS OF ADJECTIVE/ NOUN COLLOCATIONS IN TRAVEL JOURNALISM IN ENGLISH, ITALIAN AND POLISH[1]

David Finbar Brett
dbrett@uniss.it
Università degli studi di Sassari, Italy

Barbara Loranc-Paszylk
bloranc@ath.bielsko.pl
University of Bielsko-Biala, Poland

Antonio Pinna
dedalo@uniss.it
Università degli studi di Sassari, Italy

Abstract

This paper describes the compilation and subsequent analysis of a comparable corpus of travel journalism in three languages (English, Italian, and Polish). By means of a corpus-driven methodology, our study focuses on adjective/noun pairings, extracting a list of statistically significant collocations for each language and observing differences and similarities with those of the other two. Social Networks Analysis tools are used to highlight the most productive collocates. Finally, collocations concerning selected themes are analysed across the three corpora, highlighting how this approach may provide valuable input to the production of reference materials for translators.

Abstract in italiano

L'articolo descrive la compilazione e la successiva analisi di un corpus comparabile plurilingue (inglese, italiano e polacco) di testi giornalistici di viaggio. Nel focalizzarsi sulle collocazioni aggettivo/sostantivo, lo studio affianca la linguistica dei corpora all'analisi delle reti sociali. Mediante la prima, l'esame di liste di collocazioni statisticamente significative per ciascuna lingua ha evidenziato differenze e somiglianze tra i tre corpora, mentre la seconda ha individuato i collocati più produttivi. Infine, l'analisi delle collocazioni relative ad alcuni temi specifici mostra come questo approccio offra un input utile per la produzione di materiale di riferimento per i traduttori.

Parole chiave: Collocazioni; Giornalismo di viaggio; Inglese; Italiano; Polacco.

1. Introduction

Tourism has been widely recognized as a global economic force that contributes significantly to the shaping of contemporary society. In 2018 the sector reached the 1.4 billion mark in terms of international tourist arrivals, while its export earnings grew to 1.7 trillion US dollars (World Tourism Organization, 2019). The travel journalism sector has undergone a similar growth, and its role as a key player in shaping destination images and convincingly conveying them to mass audiences has recently garnered considerable attention on the part of media scholars (e.g. Hanusch 2010; Hanusch & Fürsich 2014a; Pirolli 2019). However, attention to travel reportage paid by linguists so far has been scarce and limited in scope to one or two languages (e.g. Brett 2018; Brett & Pinna 2015; Canals & Liverani 2010; Pinna 2018).

Hence much work remains to be done as regards the language of travel journalism in general, and as Taylor and Marchi (2018: 9) note when discussing under-researched content, "we might also consider languages which are under-researched, often due to lack of resources at the level of both corpora and expertise. Similarly, we might think about the relative lack of studies on multilingual corpora." Our contribution to linguistic research on this genre consists of the compilation and subsequent analysis of a multilingual corpus comprising three languages (English, Italian and Polish) belonging to different sociocultural contexts.

By means of a corpus-driven methodology, our study investigates adjective/noun collocations, a phenomenon that has not yet been investigated in travel reportages and one to which little attention has been paid in the language of tourism.[2] Given their functions of describing and evaluating specific referents, adjectives play a prominent role in constructing destination images and thus provide a vital contribution to the purpose of the genre (Durán-Muñoz 2019: 354). Our decision to study adjective/noun collocations constitutes an attempt to identify recurrent associations of specific descriptions/evaluation and referring expressions by focussing on the following points:

1. the differences and similarities in the frequencies of adjective/noun collocations
2. the differences and similarities in what the most frequent adjective/noun collocations denote
3. connectivity, i.e. the most productive collocates in the three languages, whether these are adjectives or nouns, and whether they are general terms, or closely connected to the subject at hand
4. syntactic variability: in the case of Italian and Polish, adjectives may be placed before and after nouns. This raises the question of whether there are more collocations in one order than in another, and whether there are any collocations that can be found in both orders.

The analysis will then proceed to focus on selected themes that emerge in the results, comparing the related collocations across the three corpora. In this way our study aims to make a contribution to translators and practitioners in travel journalism by highlighting useful information regarding certain collocations in their specific contexts of use and notable cultural differences between the three corpora. The importance of real-world examples in assisting the translator's decision-making process has oft been noted, for instance: "Contextual information is extremely valuable because it shows how the word behaves in a specific communicative setting and also exemplifies how a collocation is used in real language" (Castro & Faber 2014: 232).

---

2. While collocation is a notoriously difficult term to define (Gries, 2013), the sense in which it is used in the current work is "the tendency of two words to -occur, or as the tendency of one word to attract another" (Hunston, 2002: 68).

1.1. Travel writing and travel reportage

Travel writing constitutes a supra-generic category that includes a wide variety of different (sub)genres, from travel books and tourist guidebooks to maps and itineraries, all sharing a fundamental interest in travel (Witosz 2007). Thompson (2011: 26) maintains that travel writing can only be broadly defined as a constellation of different types of texts sharing some combination of common attributes, the central feature of which is the first-person, non-fictional narrative of travel. Among these, texts belonging to the genre of travel reportage are characterized as factual accounts of travellers' experiences, typically describing and commenting on their trips, usually produced by professional journalists and published in dedicated newspaper sections and magazines, although nowadays travel accounts are increasingly written by amateurs and posted on personal blogs on the Internet.

Scholars of media communication and journalism studies have noted the increasing academic attention paid to this genre (e.g. Fürsich & Kavoori 2001; Hanusch 2010; Hanusch & Fürsich 2014a). In particular, Hanusch and Fürsich (2014b: 5) underline the effects of the expansion of global tourism on the media industry, one that has triggered growing interest in travel-related journalism worldwide and provided an expanding market for travel advertising. In this respect, travel journalism plays a role in the globalized economy by promoting a cosmopolitan identity for the affluent classes worldwide and contributing to the construction of tourist destination images. For Hanusch and Fürsich (2014b: 10), as a type of lifestyle journalism, travel reportage is differentiated from hard news by its commercial orientation, in that it "primarily addresses its audience as consumers, providing them with factual information and advice, often in entertaining ways, about goods and services they can use in their daily lives" (Hanusch 2013: 4). Information, guidance, and entertainment are therefore identified as the main objectives of the genre.

The provision of factual information points to a critical difference between the practices of professional journalism, i.e. the reporting of factual accounts, and travel writing, which allows the inclusion of fictional elements. However, this clear cut distinction is questioned by Thompson (2011: 30) who maintains that "the apparent truthfulness and factuality of a travelogue

is always to some degree a rhetorical effect; and we must remember also that any form of travel text is always a constructed, crafted artefact". Moreover, travel writing has seen various famous authors straddle the divide between travel journalism and literature, such as the British Lawrence Osborne, the Americans Bill Bryson and Paul Theroux, the Italians Tiziano Terzani and Guido Piovene, and the Poles Ryszard Kapuściński and Jacek Hugo-Bader.

Travel reportage may be more protean than the academic taxonomies and definitions would like it to be and various cultural traditions position it along a cline between literary fiction and journalism, as is the case of Polish travel reportage, for instance, a genre that emerged in the Polish literary tradition in the late 19th century (Moroz 2015). Traditionally, travel reportage texts found in Polish newspapers were authored by journalists who focused on reporting their travel experiences (Rajter 2004). The genre evolved in the 20th century from linear, retrospective narratives into polyphonic travels, characterised by a centralised position of narrative persona and an increased use of creative fiction techniques (Moroz 2015).

The proximity between literature and journalism in the Italian travel reportage tradition is summarized by Massimo Bontempelli's (1938: 82) aphorism "one can be a journalist without being a writer, but to be a writer one has to be a journalist." As a matter of fact, contributors to Italian travel reportage have included some of the best writers of the 20th century, for whom travels abroad constituted not only mere visits to foreign destinations, but also pastures new for their imagination, promises of intellectual freedom and personal renewal (De Luca & Scarpa 2012: 812). For others, especially after World War II, travel reportages on their Italian tours allowed the combination of social representation and a search for identity in a world in rapid socioeconomic transformation (Lombardinilo 2016: 76).

A recent survey of travel writing using British and American travelogues is offered by Thompson (2011), who explores how the genre has managed to report the world, reveal the narrating self, and represent the other. In the field of journalism studies, Hanusch and Fürsich (2014a) edited a collection of essays that is not limited to travel journalism in the West, but also takes into consideration India (Raman & Choudary 2014) and China (Bao 2014). Finally, Pirolli (2019) explores important aspects of the practice of travel reporting in the digital age.

1.2. Linguistic studies of tourism discourse and travel reportage

In his seminal work *The Language of Tourism*, Dann (1996) quotes and/or analyzes various linguistic examples taken from travelogues. Moreover, the inclusion of travel reportages in the family of texts involved in the language of tourism is vouchsafed for not only by their topic, but also by their commercial orientation (Hanusch & Fürsich 2014b: 10).

There has been some interest in the linguistic analysis of travel writing in Polish, with studies exploring various elements of tourist discourse in thematic areas such as: urban spaces (Duda 2015), geographical regions (Żarski 2013) and individual countries (Graf 2018). Some studies utilise corpora, such as Kudełko (2016), whose analysis of Polish travel texts and guidebooks about Spain published between 1910-2010 shows how stereotypes of Spanish culture and axiological recommendations of famous places were reflected in these texts. Other studies investigating tourism discourse in Polish guidebooks demonstrate a number of ways in which it has been infused with value-laden information (Podkidacz 2004). These include ekphrastic descriptions of buildings and artworks to attribute positive connotations to history and art (Stanisławek 2013), high frequency of positive evaluative adjectives, the superlatives, metaphorically rich noun phrases and highly formulaic predicate forms to promote persuasive communication and cultural stereotyping (Zarski 2013).

Italian academics' interest in the study of tourism discourse is especially evident in the field of foreign languages, e.g. Calvi (2000) and Gotti (2006). Translation from/into Italian of tourism texts has been studied by Nigro (2006), Margarito *et al.* (2011) and Baumann (2018). However, attention to travel reportage has been scarce and limited to its production in other languages (e.g. Canals & Liverani, 2010, Pinna 2018). In relation to the methodology used here, Brett (2018) employs Social Network Analysis to study the phenomenon of connectivity, extracting networks of collocates from a 1-million-word corpus of travel reports from *The Guardian*, while Brett and Pinna (2015) apply the Part-of-Speech-gram technique to study the inflected superlative adjectives in a 450,000-word corpus of travelogues from the BBC website and not only demonstrate that inflected superlatives

are characteristic of the language of travel writing, but also that they are typically used in a small series of highly frequent constructions with limited lexical variation.

2. Materials and methods

2.1. Corpus compilation

The analysis illustrated in this paper necessitated the compilation of three comparable corpora of travel journalism for the three languages discussed: English, Italian and Polish. An attempt was made to select articles from newspapers of a comparable standing in the three respective speech communities. The authors had already compiled a collection of articles from the 'Travel' section of the British broadsheet *The Guardian* called the Guardian Travel Corpus (GTC). This consisted of a total of 1204 articles, amounting to one million tokens. These articles appeared in the online version of the newspaper (https://www.guardian.co.uk) over a period from 2006-2011. When compiling comparable corpora in Italian and Polish, the choice fell on *La Repubblica* (https://www.repubblica.it/) and *Gazeta* (https://www.gazeta.pl/), respectively, both of which are considered to be quality publications, aimed at an educated middle-class readership. Just as *The Guardian* has a 'Travel' section, *La Repubblica* has a section entitled 'Viaggi' and *Gazeta* has one called 'Podróże'.

The GTC was compiled semi-automatically in the following way:

1. The pages of the archive of the travel section were downloaded using gnu wget, (http://www.gnu.org/software/wget/). This is a free software package for retrieving files using http and other widely-used Internet protocols. In the Window OS it can be used with ms-dos to loop through incrementing addresses (e.g. https://www.theguardian.com/uk/travel?p=1, https://www.theguardian.com/uk/travel?p=2, etc.), retrieving and saving each destination file.

2. Tailor-made perl scripts developed by the authors were used to scan the html of each page of the archive for links to articles. These links were then saved, but only if they met certain criteria, e.g. the link had to contain the word "travel", while "picture", "audio", "gallery" and "video" were filtered out. This process was enacted to make sure

    that only samples of travel articles were included in the corpus, as opposed to articles from other sections of the newspaper, or pages presenting multimedia, which cannot be considered to be examples of travel journalism stricto sensu.

3.   The result of step 2 is a list of URLs. This was then fed to gnu wget, which proceeded to download the html at each URL.

4.   The html of each file was then analysed using another perl script compiled by the authors so as to identify the start and the end of the article proper, and hence eliminate all the 'boilerplate' (advertisements, links to other articles and all other extraneous content). Metadata about data of publication, author and keywords for each article were also collected.

5.   The cleaned html was then converted to the txt format using another perl script.

The same procedure was followed for the compilation of the Italian and Polish corpora, with one main difference: while *Gazeta* (like *The Guardian*) allows the reader to browse through all the articles it has ever published by way of a centralised archive (via the URL https://podroze.gazeta.pl/podroze/0,0.html?str=1; one may progress through the archive simply by augmenting the value of str), *La Repubblica* allows access only to a maximum of twelve pages (amounting to approximately the 120 most recent articles). Hence an alternative strategy was necessary in order to gather a large enough sample to allow direct comparison with the corpora in the other languages. This strategy involved searching for archive pages with two variables: the page number and a tag. An example of this is: https://www.repubblica.it/viaggi/ricerca?tags=Irlanda&p=1

    In this case the compiler proposes a tag and the script conducts repeated attempts to download the URL with incrementing values of p. At some point the URL will lead to a non-existent page and the script is aborted.

    Two sets of tags were provided: the names of all the European countries and, considering the vast amount of internal tourism, those of all the Italian regions. While this strategy did allow the compilation of a corpus for Italian travel journalism of similar dimensions to that for English and Polish, it

is important to note that a certain amount of Euro-centric, and especially Italo-centric, bias has been introduced.

The procedure described above resulted in three 1M-word comparable corpora of travel journalism in English, Italian and Polish. Some variability was noted in the composition of the corpora: the English, Italian and Polish sections were composed of 1204, 725, and 1084 articles. Hence, the Italian section contained articles that were on average longer (1379 tokens), than those of the English (830 tokens) and Polish (922 tokens).

2.2. Annotation for part-of-speech and extraction of collocations

The texts were annotated for Part-of-Speech (PoS) using Tree Tagger,[3] a tool which not only attributes a PoS tag to each token in the text, but also provides its lemma. The parameters used for the tagging process were downloaded and installed separately.[4]

Thereafter, the lemmas constituted the focus of the work. When dealing with adjective/noun pairs in English, there are essentially two variants, that with the singular and plural form of the noun, hence the conversion to lemmas merged just two pairs of word forms into one (e.g. short break, short breaks > SHORT BREAK). Similarly, in order to calculate the total frequency in the corpus (necessary for the test for strength of collocation), using lemmas made no difference to the adjective count, while that of the nouns was the sum of the singular and plural forms. For the other two languages analysed, the impact of using lemmas was far higher. Italian adjectives usually have four forms, masc. sing, masc. plur., fem. sing., and fem. plur. Some have even more, e.g. BELLO > *bello*, *bel*, *belli*, *bei*, *begli*, *bella*, *belle*.

In Polish, both nouns and adjectives decline in case, number and gender. All singular nouns are either masculine, feminine or neuter. Further to that, among masculine nouns there is another differentiation between animate

---

3. http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/
4. The English tagset is described at https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/Penn-Treebank-Tagset.pdf; the Italian tagset developed by Prof. Achim Stein, University of Stuttgart, is described at https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/italian-tagset.txt; and the Polish tagset trained on the Polish National Corpus, developed by Adam Przepiórkowski, is described at http://nkjp.pl/poliqarp/help/ense2.html.

and inanimate nouns. In the plural form, nouns and adjectives distinguish only between personal and non-personal gender. Despite the fact that there are seven cases, endings often overlap, for example, masculine animate adjectives have an identical form for the accusative and genitive case. Consequently, the adjective DOBRY 'good' has the following forms: in the singular: *dobry, dobrego, dobremu, dobrym, dobre, dobra, dobrej, dobrą*, and in the plural: *dobre, dobrych, dobrym, dobrymi, dobrzy*, arriving at 11 different forms in total. (Zagórska-Brooks 1975). Hence, working with lemma pairs, rather than word forms, becomes essential when searching for collocates and compiling wordlists of single items, in order to avoid inputting data to the statistical test that are essentially meaningless.[5]

The collocate pairs were extracted using perl scripts written by the authors. Initially a wordlist for the single lemmas was created. Thereafter, all the immediately adjacent lemma pairs tagged as adjective + noun were extracted. These data were then collapsed into a list composed of type and frequency.[6] The script then took each lemma pair, recorded its frequency, as well as that of each element's frequency on the single lemma wordlist. These data, along with the total number of tokens formed the input for a statistical test to identify pairs whose tendency to co-occur is above that of chance. The statistical procedure adopted was that of mutual information, the cut-off value for significance was 3 and the minimum frequency for collocations was 5.

---

5. It is important to note that the results reported below are combinations of lemmas, and not word forms. Therefore, all adjectives are in the masculine, regardless of the gender of the modified noun. For example, the Italian collocation *luna piena* 'full moon' is reported as LUNA PIENO.

6. The Polish list had to undergo considerable post-processing to eliminate a) proper nouns, e.g. NOWA ZELANDIA; comprising 4.91% of the original list of collocates b) determiners, including several types of pronouns, such as: possessives (e.g. NASZ KONTYNENT 'our continent'), negatives (e.g. ŻADEN PROBLEM 'no problem'), indefinite particles (e.g. NIEKTÓRY DOM 'a house'), as well as predeterminers (e.g. TAKI OBIEKT 'such a facility'), numerals, e.g. 10 ROK 'tenth year' and dates, e.g. 14 LUTY 'February 14' - deletions of determiners reduced the original wordlist by 30% c) duplicated words, e.g. BARWA BARWA, (2.88%) d) other erroneous inclusions which comprised 0.8% of the original wordlist.

As indicated above, the initial focus was on the prototypical order for combining adjectives and nouns in the three languages, i.e. ADJ+NOUN for English and Polish, and NOUN+ADJ for Italian. However, examination of the concordance lines for Italian and Polish suggested time and time again that the presence of the inverted order variant was not negligible. In fact, on repeating the procedure for the inverted order variant form (ADJ+NOUN for Italian and NOUN+ADJ for Polish) the numbers were indeed substantial (see Section 3.1), and furthermore the vast majority of collocations displayed a preference for one order or the other.

2.3. *Networks of collocates*

Brezina et al (2015:146) describe the importance of connectivity as a property of collocations, beside those of distance, frequency, exclusivity, directionality, dispersion and type-token distribution among collocates. Bearing in mind the oft-quoted statement of Firth's (1957: 11) "You shall know a word by the company it keeps", this study made use of Gephi (https://gephi.org/), a tool developed in the field of Social Networks Analysis (henceforth SNA). This tool, on importing appropriately formatted data concerning collocations (see Brett 2018), allows the production of detailed graphs highlighting the hub collocates (those which are particularly productive in the formation of collocations) and those which, to the contrary, collocate with only one other word. In SNA terminology, each item is called a node; connections between nodes are called edges; and the number of edges a given node has is called its degree. Just like directionality is a property of collocation (Gries 2013), edges can be directional, or undirected. For the purposes of the present analysis, directionality was not calculated for the dataset, though this may well be incorporated in future studies.

The decisions regarding the formatting of the graphs are the following:

1. Nodes are colour-coded for Part-of-Speech (green for adjectives, red for nouns)
2. Node size reflects frequency of the node word in the corpus
3. Edge size reflects the frequency of the collocation

3. Results and discussion

3.1. *Types and tokens*

Considerable variation was observed in the number of collocations that the three corpora yielded: English provided 1050 types, with a sum of 11512 tokens, Italian, 765 types, with a sum of 8102 tokens and Polish, 993 types, amounting to 10330 tokens. Therefore, at least as far as concerns the ADJ+NOUN (or NOUN+ADJ for Italian) structure, it would appear that English is the language in which there is greatest formulaicity in travel journalism, and Italian the language which resorts to it the least. Polish would appear to lie somewhere between the two. However, factoring in the variant syntactic pattern (ADJ+NOUN for Italian, and NOUN+ADJ for Polish) overturned these results. Details can be found in Tables 1 and 2. In light of these new data English and Italian appear to be extremely similar, and it is Polish that emerges as the most formulaic, both in terms of types and tokens.

	English	Italian	Polish
Types (ADJ+NOUN)	1050	365	993
Types (NOUN+ADJ)		765	431
Total	1050	1130	1424

Table 1. Statistically significant collocate pairs in the three corpora: types.

	English	Italian	Polish
Tokens (ADJ+NOUN)	11512	3404	10330
Tokens (NOUN+ADJ)		8102	5076
Total	11512	11506	15406

Table 2. Statistically significant collocate pairs in the three corpora: tokens.

The pattern is somewhat different if we take into consideration the number of collocations with a frequency greater than or equal to fifty. The types and their frequencies are listed in Table 3. English has 20, Italian has 12 and Polish has 22.

English	LAST YEAR (200), FIRST TIME (116), NEXT DAY (101), YOUNG PEOPLE (99), FULL BOARD (90), MORE INFORMATION (85), NEXT YEAR (79), FURTHER INFORMATION (76), NEXT MORNING (73), NATIONAL PARK (70), GOOD PLACE (70), FEW DAY (60), LOCAL PEOPLE (60), GOOD WAY (58), LAST WEEK (57), WEST COAST (55), FEW YEAR (55), CHIEF EXECUTIVE (53), LIVE MUSIC (53), OPEN AIR (52)
Italian (NOUN+ADJ)	CENTRO STORICO (397), PARCO NAZIONALE (123), PARCO NATURALE (86), RISERVA NATURALE (68), PISTA CICLABILE (61), MUSEO ARCHEOLOGICO (59), METRO QUADRATO (55), GUERRA MONDIALE (53), MACCHIA MEDITERRANEO (52)
Italian (ADJ+NOUN)	GRAN PARTE (106), ULTIMO ANNO (68), GRANDE PARTE (60)
Polish (ADJ+NOUN)	STARY MIASTO (185), DUŻY ATRAKCJA (83), PIĘKNY WIDOK (77), DUŻY MIASTO (73), ŚWIATOWY DZIEDZICTWO (71), PRAWY STRONA (70), LEWY STRONA (67), WYSOKI SZCZYT (64), JEDEN DZIEŃ (57), DOBRY MIEJSCE (53), ŚREDNI TEMPERATURA (50)
Polish (NOUN+ADJ)	PARK NARODOWY (244), WOJNA ŚWIATOWY (120), LINIA LOTNICZY (107), MATKA BOSKI (85), INFORMACJA TURYSTYCZNY (80), BILET LOTNICZY (71), STRONA INTERNETOWY (70), MUR OBRONNY (66), ATRAKCJA TURYSTYCZNY (59), OŚRODEK NARCIARSKI (54)

Table 3. Types and frequencies of the statistically significant ADJ+NOUN (or NOUN+ADJ) collocations found in the three corpora with frequency ≥ 50.

3.2. Syntactic variation

As has been noted above, two of the languages under analysis, Italian and Polish, allow variability in the position of the adjective with respect to the noun. Italian generally prefers to place the adjective after the noun (Serianni 1989), Polish before (Zagórska-Brooks 1975). This study suggests that, at least with regards to the text type taken into consideration, adjective/noun collocations in Italian and Polish occur in the canonical form twice as often

as in the variant form. This proportion remains the same regardless of whether types or tokens are taken into consideration.

A further point of interest is whether collocations display exclusive preferences for a particular order, or whether there are collocations that are statistically significant in both the canonical and the variant forms, and if so, what the proportions involved are. The data present a very clear picture: adjective/noun collocations display a distinct preference for a particular order: just 39 types (2.7% of the total number of types in each form), corresponding to 427 tokens (2.8% of total number of tokens in each form) appeared on both lists in Polish. This separation was even more extreme in the Italian data, as the statistically significant collocate pairs in both the canonical and the variant form consisted of only 11 types (1.0% of total types), corresponding to 87 tokens (0.8% of total tokens). Even when the collocations appeared on both lists, a tendency to occur in one form or the other was still observed in the majority of cases. Figure 1 presents this phenomenon for Italian. It is interesting to note that the collocations that were present in both forms displayed a preference for the variant form, i.e. ADJ+NOUN. For example, the frequency of ANTICO BORGO is 20, whereas that of BORGO ANTICO is 10. Therefore, it displays a preference for the variant form, and hence is plotted one third along the axis spanning a range from -1 to 1. The collocate pairs in the centre of the graph (close to 0) display little or no preference for one form or the other. For obvious reasons of legibility, the collocate pairs that are significant only in one form are not plotted, but if they were, they would all be aligned at -1 or 1 on the x-axis.

The topic of connectivity is discussed in detail in the next section. However, it may be fitting at this point to observe that some collocates are particularly productive in the variant syntactic form. These are all adjectives, and have to do with size (GRANDE, PICCOLO, LUNGO), age (NUOVO, VECCHIO, ANTICO) and positive evaluation (BELLO, SPLENDIDO, SPETTACOLARE).
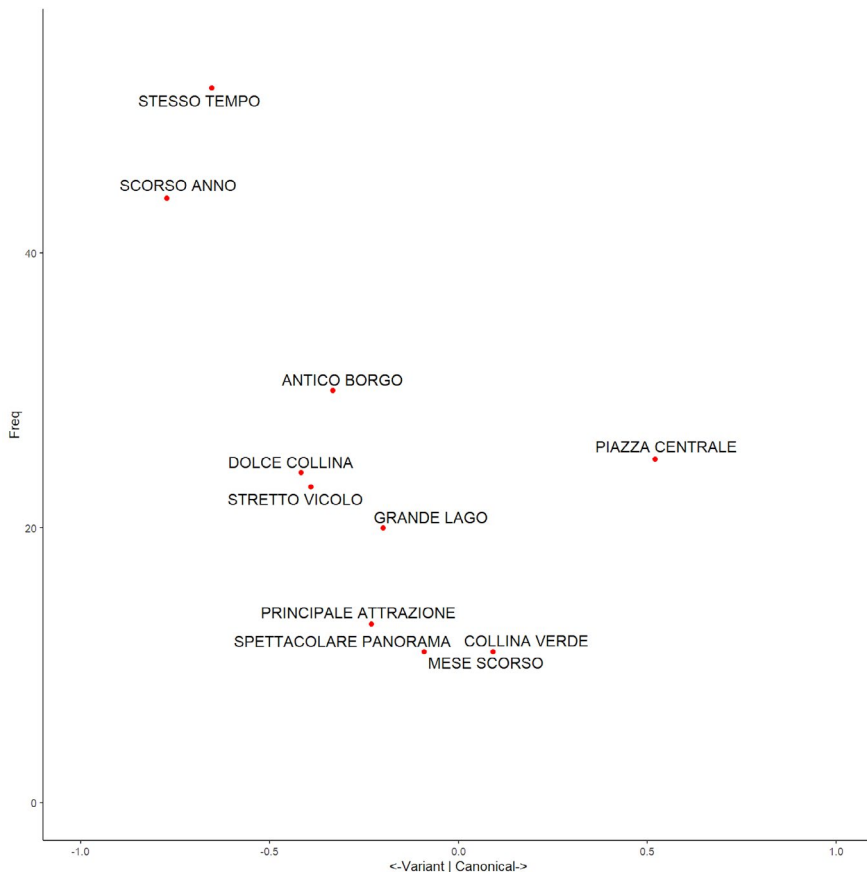
Figure 1. Adjective/noun collocations in the Italian corpus that are statistically significant in both the canonical and the variant order. The x-axis represents the tendency to occur in the canonical (right) and the variant (left) order.

With respect to the Polish data, taken as a whole the collocate pairs that were statistically significant in both orders did not appear to display a preference for a particular order (Fig. 2). On the level of the individual pair, some preferred the canonical (ADJ+NOUN) form, (e.g. CZERWONY SZLAK, WOLNY CZAS), other preferred the variant form (e.g. ATRAKCJA TURYSTYCZNY, TRASA NARCIARSKI, WODA MINERALNY, ŻYCIE NOCNY), while still

more occurred equally in both forms (e.g. TURYSTA INDYWIDUALNY, WODA MORSKI).

The most productive collocates in the variant syntactic form in Polish are, similar to Italian, all adjectives, however they are more specific to the subject matter: TURYSTYCZNY, NARCIARSKI, MIEJSKI. At this point, it is important to note that in Polish slight differences of meaning can be conveyed through noun adjective order. Therefore, if the collocate pair takes on the NOUN+ADJ order, the adjective classifies the noun based on its intrinsic quality. The example of ATRAKCJA TURYSTYCZNY can illustrate this tendency. The canonical form, TURYSTYCZNY ATRAKCJA refers to an entertainment which only potentially can be attractive to tourists, as it is primarily used for other purposes (e.g. *korzystanie z miejskiej kolejki, która sama w sobie może stanowić turystyczną atrakcję* 'taking the urban railway may in itself be a tourist attraction'). On the other hand, the variant form, ATRAKCJA TURYSTYCZNY denotes an entertainment primarily meant for tourists (*na terenie jeziora znajduje się kilka atrakcji turystycznych, dla których warto na kilka godzin podnieść się z łóżka* 'near the lake are a few tourist attractions worth getting up from bed for a few hours') and it tends to be used with a preceding adjective in the superlative form (*Ale to w końcu jedna z najbardziej znanych atrakcji turystycznych świata* 'After all, it is one of the most famous tourist attractions in the world').

With respect to collocate pairs formed with MIEJSKI and NARCIARSKI, the following tendency can be observed: when in the variant, NOUN+ADJ order, the collocations tend to be used in the plural form (*Z placu odjeżdżają autobusy miejskie;* 'City buses depart from this square'; *Wogezy nie słyną ani z narciarskich tras, ani z zapierających dech panoram,* 'The Vosges are not famous for their ski trails, nor for breathtaking views'). The canonical form, on the other hand, especially in the case of MIEJSKI AUTOBUS, is almost exclusively used in the singular form. It is also important to note that, if the collocate pair follows the NOUN+ADJ order, it tends to be pre-modified by another adjective, which does not occur often in the canonical form (*Są tu trzy dość trudne trasy narciarskie,* 'there are three quite difficult ski trails').
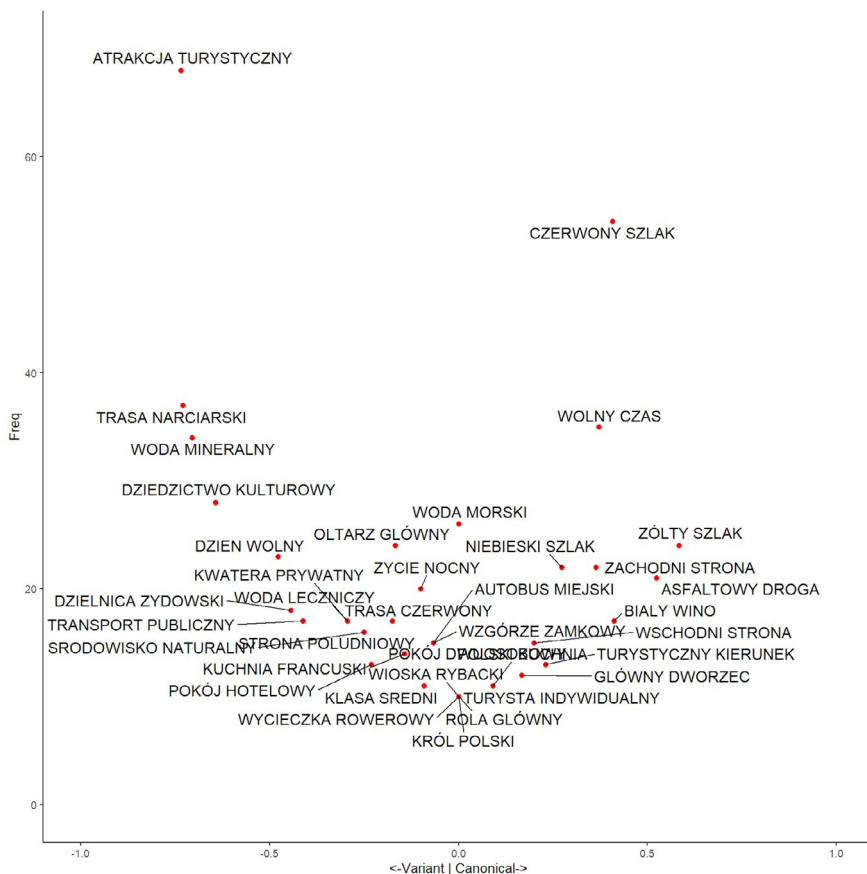
Figure 2. Adjective/noun collocations in the Polish corpus that are statistically significant in both the canonical and the variant order. The x-axis represents the tendency to occur in the canonical (right) and the variant (left) order.

In conclusion, these results may be of interest to translation practitioners and researchers as, when translating from one language to another, it is not sufficient solely to have great familiarity with the collocations in both languages (Baker 2018:53; Taylor 1998:26), but it is also necessary to be aware of the syntactic preferences that these may display in languages that allow such flexibility.

3.3. Connectivity

The lists of collocates were imported to Gephi to facilitate the identification of hub collocates, i.e., those nodes that are most productive in the creation of collocations. The degree of each node (i.e., the number of connections that it has to other nodes) is calculated by running the average degree test. This test provides an indication of the overall connectedness of the network. The average degree results for the English, Italian and Polish data were 2.565, 2.045 and 2.630, respectively. However, these results concerned only the canonical syntactic form. When the results were integrated with the data regarding the variant form, the average degree results for Italian and Polish were updated to 2.503 and 2.711, respectively.

Therefore, the overall connectedness figures are quite similar across the three corpora. Had substantial differences emerged, for example with the data for one language having a particularly low value, it would suggest that the dataset in question was composed of a greater proportion of isolates (i.e., pairs of words that collocate only with each other), as opposed to hubs.

Naturally, such isolate pairs can be found in all three datasets. They generally tend to be either technical terms, such as TIDAL BORE, INLAND WATERWAY and RENEWABLE ENERGY or low-frequency collocations such as BEATEN TRACK, HIDDEN GEM and PLAIN SAILING, to provide examples from the English dataset. A similar pattern was found in the other two languages. The isolates constituting technical terms in Italian included: SCARTAMENTO RIDOTTO 'narrow gauge', SET CINEMATOGRAFICO 'film set', and RITO PROPIZIATORIO 'propitiatory rite', whereas the low-frequency collocations included PIEDE NUDO 'bare foot', MANTO NEVOSO 'snow cover' and LUNA PIENO 'full moon'. The isolates in the variant syntactic form data were on the whole of the latter type, examples include CONTINUO EVOLUZIONE 'continuous development', TARDO POMERIGGIO 'late afternoon' and LARGO ANTICIPO 'well in advance'.

A few representative examples of technical terms found in Polish include OBFITY OPAD 'heavy precipitation', PRZEDNIA SZYBA 'windscreen', and SWOBODNY PRZEPŁYW 'free flow' and low-frequency collocations can be exemplified by the following: PRZYJEMNY CHŁÓD 'pleasant chill',

SZKLANA KOPUŁA 'glass dome', and AKTYWNY WYPOCZYNEK 'active recreation'.

Greater differences are to be seen when observing the nodes that are most connected: there are two aspects that are striking. Firstly, here the Italian corpus appears to be the outlier, with less than half the number of collocates that act as hubs, in comparison to English and Polish. A second point to be made is that the lemmas that do appear to be particularly productive in the formation of collocations are by and large specifically connected with the subject at hand in the Italian corpus, these include terms relating to culture (e.g. STORICO, MEDIEVALE, CULTURALE 'historical, medieval, cultural') and places (e.g. CENTRO, LOCALE, CAPITALE 'centre/town, local, capital'). The hub collocates found for English are mostly lemmas that could pertain to any domain. In fact, all of the noun lemmas appear within the top 100 most frequent nouns in the BNC, with the exception of BEACH. Similarly, the adjective collocates are all on the list of the top 100 most frequent adjectives in the BNC, with the exception of GAY and NEXT.[7] Interestingly, Polish appears to lie in the middle of these two extremes also with respect to the nature of its hub collocates. While quite a few are general, all-purpose words (e.g. DOBRY, DUŻY, MIEJSCE, INNY, CZĘŚĆ 'good, big, a place, the other, a part'), many more relate to the specific subject matters dealt with in travel journalism (e.g. MIASTO, WODA, DROGA, ATRAKCJA, BRZEG 'a town, water, a road, an attraction, a river bank').

3.4. Themes

The analysis will now focus on a number of themes that emerge as being recurrent in the collocations extracted from the corpora in the three languages. Differences and similarities will be highlighted, while one caveat must always be borne in mind: the current analysis is an observation solely of adjective/noun collocations. The absence of a given collocation in a particular corpus, corresponding to collocations in one or two of the other corpora, does not necessarily mean that this entity or concept is not widely dealt with or discussed in that corpus. Its absence from the list of statistically

---

7. BNC wordlists available at http://www.kilgarriff.co.uk/bnc-readme.html#raw

significant collocations could be due to the fact that it is expressed with a different pattern (e.g. a compound noun, or even a sole noun). Alternatively, it could be expressed by way of a number of adjective/noun expressions that are not frequent enough, or do not display a strong enough attraction, to reach statistical significance.

3.4.1. Theme 1: Human settlements

By far the collocation with the highest frequency across the three corpora is the Italian CENTRO STORICO (397). As noted above, CENTRO constitutes a hub, with 11 collocates. This is partially due to the polysemy of the word itself, and the different senses of the word can be observed in the list of collocates:

1. the core of a larger entity (STORICO, CITTADINO). The first would translate into English as 'old town' (i.e. the historical nucleus of a town/city), the second 'town/city centre'.
2. a medium/large settlement (ABITATO, URBANO, MEDIOVALE). The first two would translate simply as 'town' or 'city', the last as 'medieval town'.
3. a place of great importance (ARTISTICO, CULTURALE, SPIRITUALE). These would translate as 'artistic/cultural/spiritual centre'.
4. a place with the facilities for a specific activity (COMMERCIALE, TERMALE, BALNEARE). These would translate as 'shopping centre/ mall, spa resort, seaside resort'.

STORICO is also a hub collocate, contributing to the formation of no fewer than 21 pairings. Amongst these we find QUARTIERE STORICO (5), NUCLEO STORICO (5), both of which have very similar meanings to CENTRO STORICO. Another collocate of STORICO is BORGO (6), which in turn collocates with ANTICO (10), hence providing two collocations with practically the same meaning, 'old/historic village', hinging on the noun BORGO. Of interest is the fact that the latter collocation is present also in the variant order list. In fact, ANTICO BORGO (20) is actually twice as frequent as the collocation in the canonical form.

There is another near synonym of STORICO that appears in the list of collocates: VECCHIO. This is present in a sole collocate pair, one that to all intents and purposes would again appear to have exactly the same meaning as CENTRO STORICO, CITTA' VECCHIO (8). However, an examination of the concordance lines reveals that it is used almost exclusively in non-Italian contexts.
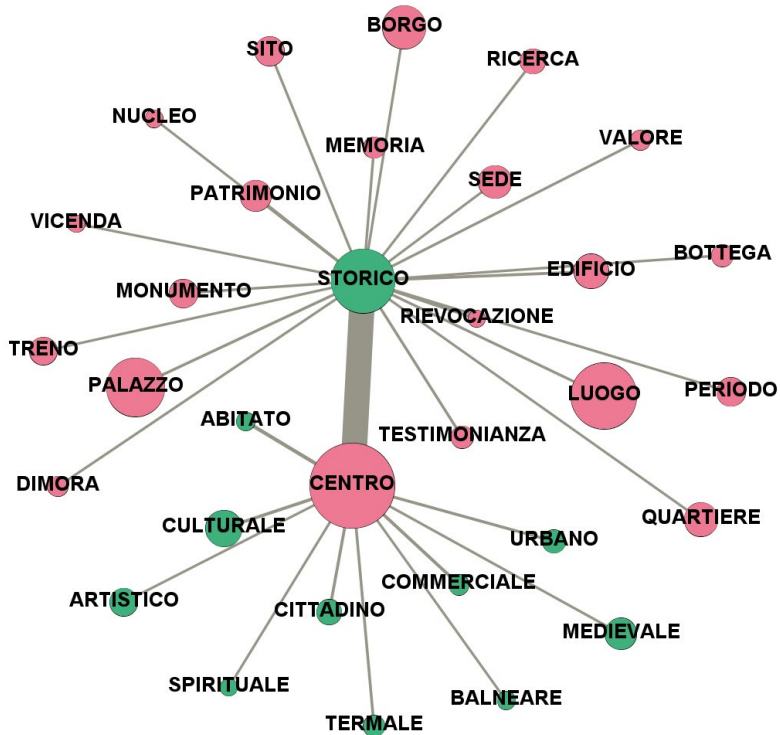

Figure 3. Network graph showing the collocates of CENTRO and STORICO

In Polish, MIASTO 'town' has 11 collocates, and that with the highest frequency is with STARY 'old'.

The following senses of the word MIASTO can be illustrated by looking at its collocates:

1. The oldest, most picturesque part of a town where most of the historic sites are located, would be represented by the collocation STARY MIASTO. This would translate into English as 'old town'. Other collocations that express the concept of STARY MIASTO are: STARY CENTRUM, HISTORYCZNY CENTRUM, and ZABYTKOWY CENTRUM - a closer look at the concordance lines shows that all three collocations are used predominantly in non-Polish contexts (*znajdują się oczywiście w centrum miejscowości, przy starym mieście Główna plaża Lloret de Mar,* 'they are obviously located in the town centre, near the Old Town, where the main beach in Lloret de Mar can be found'). It is important to note, however, that the same concept as STARY MIASTO, CENTRO STORICO and OLD TOWN can be expressed using a single noun, STARÓWKA (a derivative of STARY), of which there are 91 occurrences in the corpus.
2. A large area inhabited by a number of inhabitants where facilities are located and services are provided (DUŻE MIASTO) This would translate into English as 'city, big town';
3. A place the town or city in which a (famous) person used to live (RODZINNE MIASTO). This would translate into English as 'hometown';
4. A tourist destination, a foreign place worth visiting (EUROPEJSKI, WŁOSKI). This would translate into English simply as 'town' or 'city';
5. A place of great (religious or spiritual) importance, similar to Italian: ANTYCZNY, STAROŻYTNY, ŚWIĘTY. This sense is also expressed by the collocation WAŻNY CENTRUM. English equivalents would be 'ancient city', 'religious centre', 'important centre'.

The concept of TOWN CENTRE is expressed in Polish through the collocation ŚCISŁY CENTRUM that denotes the most central area of the town, in which all the major sites are located. An examination of concordance lines reveals that this collocation is used when referring to conveniently located places, especially accommodation.

There are two near synonyms of STARY that can be found in the list of collocations: DAWNY and ZABYTKOWY. While STARY is a more generic adjective that denotes old age and tends to form collocations with inanimate

nouns that refer to buildings (RATUSZ, 'townhall') or constructions erected as a unified community (STARY CMENTARZ', 'old cemetery'), as well as animate nouns (STARY DRZEWO, 'old tree'), its synonym ZABYTKOWY forms pairings only with inanimate nouns that refer to buildings or human settlements. DAWNY, on the other hand, collocates with both inanimate (DAWNY DZIELNICA, 'old district') and animate nouns (DAWNY MIESZKAŃCY, 'former inhabitants'), as well as nouns denoting abstract concepts (DAWNY CZAS, 'old time', DAWNY ŚWIETNOŚĆ, 'past glory') and carries the meaning of past state/activity that is no longer part of the present.
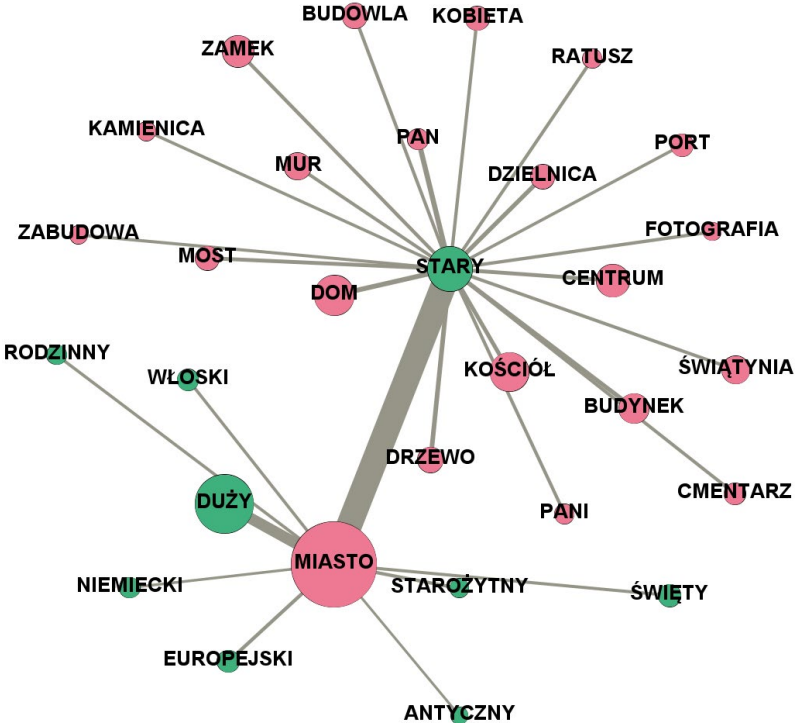

Figure 4. Network graph showing the collocates of STARY and MIASTO

The presence in the English corpus of equivalents of CENTRO STORICO/ STARY MIASTO is rather underwhelming. The only option for expressing this would appear to be OLD TOWN (27). Similarly, collocations referring to the age of human settlements are limited to MEDIEVAL TOWN (11) and HISTORIC TOWN (8).

3.4.2. Theme 2: Destination Appeal

Another theme of interest is that of evaluation, specifically the notion of attraction. Tourism, based as it is on personal and group preferences, which in turn are influenced by fads and fashions, is a particularly fickle and unpredictable market, prone to mass vagaries and whims. It is perhaps no mere coincidence that places to be visited are imbued with animation and construed as being active sentient beings that attract tourists, in the same way that humans and animals attract potential mates. This metaphor is persistent across the three corpora, as statistically significant collocations have been found featuring ATTRACTION, ATTRAZIONE and ATRAKCJA. In Italian the noun ATTRAZIONE collocates with three adjective lemmas TURISTICO (13), NATURALE (6) and PRINCIPALE (5). The tokens of the most frequent collocation, that with TURISTICO, are more or less in equal measure singular and plural. In some instances, one may detect a slightly negative semantic prosody, as if denoting places/sites etc. that are very much *on* the beaten path, (e.g. *Ma la vera Brac è molto di più di queste attrazioni turistiche e per chi vuole scoprirla davvero* 'But there is much more to Brac than these tourist attractions, and for those who really want to discover it'). In other cases, the prosody is decidedly positive (e.g. *Non altrettanto scontato è invece il fatto che siano considerate attrazioni turistiche a pieno titolo, di quelle che, per intenderci, valgono una deviazione, se non il viaggio* 'The fact is not so obvious, however, that they are considered fully-fledged tourist attractions, those which, to make things clear, are worth taking a detour for, if not the whole trip').

The tokens of the collocation with PRINCIPALE are almost all plural, and it is of interest to note that three out of the five form part of an identical string: *Una delle attrazioni principali* 'One of the main attractions'. Similarly, the tokens of ATTRAZIONE NATURALE are all plural, and three out of

the six instances are *una delle attrazioni naturali* 'One of the natural attractions'. Here the semantic prosody is markedly positive, and the superlative is present in the co-text in four instances, one example being *tra le attrazioni naturali più affascinanti e spettacolari del nostro Paese* 'Amongst the most fascinating and spectacular natural attractions in our country'.

The noun ATRAKCJA is particularly productive in Polish forming pairs with 9 adjectives, of which two, such as WIELKI and DUŻY are close synonyms. The most frequent collocation is formed with the adjective DUŻY (83) 'large/big attraction'. The tokens of the collocation do not reveal a clear preference for either singular or plural, but almost all use the superlative form of the adjective, therefore conveying positive semantic prosody. In contrast to this, all the tokens of the collocate pair WIELKI ATRAKCJA form part of a three-item string, WIELKA ATRAKCJA TURYSTYCZNA, and, interestingly, only occur in the singular.

The tokens of the collocation with GŁÓWNY (40), 'main attraction', bear some resemblance to the tokens of Italian PRINCIPALE, namely, almost all occur in the plural, and if they are in the singular, they tend to form part of an identical string: *Jedną z głównych atrakcji*, 'One of the main attractions'. The instances of two collocations, DODATKOWY ATRAKCJA (24), 'additional attraction', and CIEKAWY ATRAKCJA (11), 'interesting attraction', generally follow the same pattern. The tokens of these collocations occur both in the singular and the plural. In the case of the latter, they tend to be preceded by quantifiers of amount, for example [*w programie bardzo dużo ciekawych atrakcji z lokalnej flory i fauny,* 'there are lots of interesting attractions related to local flora and fauna in the program']; suggesting a large number of additional or interesting attractions. A similar sense to the abovementioned is conveyed through another pair of collocates, LICZNY ATRAKCJA (7), which exclusively occurs in the plural form and translates into English as 'numerous attractions'. Interestingly, these tokens of DODATKOWY ATRAKCJA and CIEKAWY ATRAKCJA, which are not preceded by the quantifiers of amount, tend to occur at the very beginning of the sentence signalling a novel aspect of the information being conveyed, (e.g. *Ciekawą atrakcją jest też stojący tuż obok kolumny św Trójcy miejski ratusz,* 'An interesting tourist attraction is the townhall located next to the column of the holy Trinity').

The instances of the collocation WAŻNA ATRAKCJA (6), 'important attraction', follows a pattern similar to DUŻA ATRAKCJA, i.e. the tokens of the collocation occur both in the singular and in the plural form, and interestingly they all use the superlative form of the adjective (e.g. *Do najważniejszych atrakcji regionu należą: spływ Dunajcem na drewnianych tratwach,* 'the Dunajec river rafting ride is the most important tourist attraction of the region').

The tokens of the collocation NOWY ATRAKCJA (6) are almost exclusively in the singular form, and similar to English, they are used with respect to theme or aqua parks, and there is also an occasional instance of the superlative NAJNOWSZY, 'the newest', referring to attractions for children (e.g. *także dla rodzin podróżujących z dziećmi. Najnowszą atrakcją dla tych ostatnich ,* 'also for the families travelling with kids. The newest attraction for the latter').

Polish also avails of the adjective ATRAKCYJNY which derives from ATRAKCJA. In our dataset it collocates with the following three nouns: MIEJSCE (6) 'place', OFERTA (6) 'offer' and CENA (5) 'price'. In the first case, all the tokens of ATRAKCYJNY MIEJSCE are in the plural and denote a place which tourists would find appealing and worth visiting (e.g. *posiadłości z atrakcyjnymi miejscami odpoczynku dla turystów,* 'estates offering attractive places for tourists to relax'). On the other hand, the two other collocates: ATRAKCYJNY CENA and ATRAKCYJNY OFERTA are used in the sense of there being a bargain, i.e. affordable and not overpriced. These would translate into English as 'reasonable price' and 'good offer', respectively, (e.g. *To była naprawdę wyjątkowo atrakcyjna oferta, tzw . last minute,* 'It was really a good offer, so-called last minute'; *Ich zaletą prócz atrakcyjnej ceny jest znakomita lokalizacja,* 'Their advantage is, apart from a reasonable price, a perfect location').

The collocates of ATTRACTION in the English corpus all have the function of evaluating the reasons for visiting a particular locality and/or event. MAIN (10) would appear at first sight to be a direct counterpart of the Italian PRINCIPALE. However, only half of the instances are in the plural, and none concern *one of the main attractions*. Most of the examples are equative constructions with the collocation being either before or after the copula verb, e.g. *but the main attraction is the food*; *The main attractions are the people and*

the stunning scenery plains with volcanic hills. NEW (8) is again sometimes singular, sometimes plural, with two instances of the superlative NEWEST, implying there are also other 'new' attractions. The pattern is almost invariably 'the new(est) attraction(s) at + PROPER NOUN'. It is of interest to note that the proper noun in question is always that of a theme park (e.g. Alton Towers), a zoo or an aquarium (e.g. Amazon World/Blue Reef), or an educational museum (e.g. National Space Centre). Therefore, NEW would appear to collocate with one of the senses of ATTRACTION, that denoting a sight or an activity which is exciting and novel, particularly appealing to children.

One glaring absence in the English list is a counterpart of ATTRAZIONE TURISTICO and TURYSTYCZNY ATRAKCJA: an equivalent does exist in the English corpus, but to the contrary of the other two languages, it does not adhere to the ADJ+NOUN pattern, illustrating the caveat mentioned at the start of the section. The collocation in question is TOURIST ATTRACTION, a compound noun, that is correctly annotated in the corpus as being NOUN + NOUN. Remarkably, its frequency, 11, is almost identical to that of its Italian and Polish counterparts. En passant, it was noted above that concepts requiring ADJ+NOUN in one language could be expressed as single words in another. One strong candidate as an equivalent of ATTRAZIONE PRINCIPALE and DUŻY/GŁÓWNY ATRAKCJA would be HIGHLIGHT. There are 99 instances of this in the Guardian Travel Corpus, with the singular and plural forms in roughly equal proportions. In comparison, in the BNC, the two forms sum to a frequency of 930.[8]

4. Final remarks

The aim of the current paper is twofold. Firstly, it constitutes an attempt to demonstrate how corpus linguistics techniques can contribute to a better understanding of collocation across languages. Secondly, it aims to explore differences and similarities in collocation patterns in different languages, especially with the aim of garnering information useful to translators. With regards to the first objective, Baker (2018: 56) remarks:

---

8. Applying the chi-squared test, the presence of HIGHLIGHT in the Guardian Travel Corpus is statistically significant at $p \leq 0.001$, with a result of 92.79.

Every word in a language can be said to have a range of items with which it is compatible, to a greater or lesser degree. Range here refers to the set of collocates, that is other words, which are typically associated with the word in question. Some words have a much broader collocational range than others.

While this observation is extremely insightful, it was destined to remain anecdotal, until the phenomenon of connectivity became a topic of interest in corpus linguistics, and methodologies started to appear that allowed the objective measurement of the feature (see Brezina *et al.* 2015). This study demonstrates how tools developed outside the field of corpus linguistics can be harnessed to highlight the presence of hubs, what we can consider super-collocates, those terms that are particularly productive in the creation of collocate pairs. These are no other than the terms with a "much broader collocational range than others" referred to above. Similarly, isolate pairs, words that only collocate with each other can be identified with ease.

In terms of the second point, an overall observation of the data would suggest that the language of Italian travel journalism is slightly less formulaic than that of English and Polish, at least with regard to adjective/noun pairs.

During the analysis an interesting point emerged which would seem to constitute an exception to the view that the "most important difference between grammatical and lexical choices, as far as translation is concerned, is that grammatical choices are largely obligatory while lexical choices are largely optional" (Baker 2018: 96).

In Italian and Polish, certain adjective/noun collocations, admittedly limited in number, were seen to be statistically significant in both the canonical order and in a variant, marked, order. Hence, what is essentially a grammatical choice, allows an option. Since this phenomenon concerned only certain lemmas, almost exclusively adjectives, it would appear to constitute an important intersection between grammar and lexis. For example, while ANTICO BORGO and BORGO ANTICO are both attested, at such frequencies to be both statistically significant collocations, the first, a variant order form, is twice as frequent. This type of knowledge is of use to the translator who is translating a travel journalism text into Italian and aims to reproduce similar lexical/grammatical choices to those enacted by a travel journalist writing in the target language.

In fact, Castro & Faber (2014: 205), in a paper describing the most representative English and Spanish collocation dictionaries for general language with the aim of evaluating how useful they may be for translators, observe that:

There is a general consensus among translators that phraseological information in lexicographic resources is crucial, especially in the final production of the target language text. In this phase, the translator may need grammatical and syntactic information related to terms, including collocations in the target language.

It is hard to imagine how objective data concerning the lexical/syntactic behaviour of collocations may be gleaned without recourse to the corpus-driven methodologies described in this paper, especially where special domains of human endeavour are concerned.

References

BAKER, Mona. (2018) *In Other Words. A Coursebook on Translation* (3rd ed.), London: Routledge.

BAO, Jiannu. (2014) "Going with the flow: Chinese travel journalism in transition." In: Hanusch, Folker & Elfriede Fürsich (eds.) 2014. *Travel Journalism. Exploring Production, Impact and Culture*. Basingstoke: Palgrave Macmillan, pp.134-151.

BAUMANN, Tania. (2018) *Reiseführer – Sprach- und Kulturmittlung im Tourismus / Le guide turistiche – mediazione linguistica e culturale in ambito turistico*. Bern: Peter Lang.

BONTEMPELLI, Massimo. (1938) *L'avventura novecentista. Selva polemica (1926-1938)*. Firenze: Vallecchi.

BRETT, David. (2018) "Social Network Analysis and the Analysis of Collocations in the Language of Travel Journalism." In Baumann, Tania (ed.) 2018. *Reiseführer – Sprach- und Kulturmittlung im Tourismus / Le guide turistiche – mediazione linguistica e culturale in ambito turistico*. Bern: Peter Lang, pp.183-207.

BRETT, David & Pinna, Antonio. (2015) "Patterns, fixedness and variability: using PoS-grams to find phraseologies in the language of travel journalism." *Procedia - Social and Behavioral Sciences* 198, pp. 52 – 57.

Brezina, Vaclav; McEnery, Tony & Wattam, Stephen. (2015) "Collocations in context: A new perspective on collocation networks." *International Journal of Corpus Linguistics* 20:2, pp. 139-173.

Calvi, Maria Vittoria. (2000) *Il linguaggio spagnolo del turismo.* Viareggio: Baroni.

Canals, Jordi & Elena Liverani (eds.) (2010) *Viaggiare con la parola.* Milano: Franco Angeli.

Castro, Miriam Buendía, & Pamela Faber. (2014) "Collocation dictionaries: a comparative analysis." *MonTI. Monografías de Traducción e Interpretación* 6, pp. 203-235.

Dann, Graham. (1996) *The Language of Tourism. A Sociolinguistic Perspective.* Wallingford: CAB International.

De Luca, Bernardo & Domenico Scarpa. (2012) "Gli scrittori in viaggio". In Scarpa, Domenico (ed.) 2012. *Atlante della Letteratura Italiana. Dal Romanticismo a oggi.* Vol. III. Torino: Einaudi, pp. 812-821.

Duda, Beata. (2015) *Dyskursywne i tekstowe reprezentacje współczesnej przestrzeni miejskiej* [Representations of contemporary urban spaces through discourse and text]. Katowice: Uniwersytet Śląski. Unpublished doctoral dissertation.

Durán-Muñoz, Isabel. (2019) "Adjectives and their keyness: a corpus-based analysis of tourism discourse in English." *Corpora* 14: 3, pp. 351–378.

Firth, John R. (1957) *Papers in Linguistics 1934–1951.* London: Oxford University Press.

Fürsich, Elfriede & Anandam P. Kavoori (2001) "Mapping a critical framework for the study of travel journalism." *International Journal of Cultural Studies* 4:2, pp. 149-171.

Gries, Stefan Th. (2013) "50-something years of work on collocations: what is or should be next …" *International Journal of Corpus Linguistics* 181, pp. 137–165.

Gotti, Maurizio. (2006) "The language of tourism as specialized discourse." In: Palusci, Oriana & Sabrina Francesconi (eds.) 2006. *Translating Tourism. Linguistic/Cultural Representations.* Trento: Editrice Università degli studi di Trento, pp.15-34.

Graf, Paweł. (2018)" Teoria i praktyka dyskursu turystycznego (na przykładzie Słowacji i Rzymu wpisanych w przewodniki)." [Theory and Praxis of the Tourist Discourse (on the Example of Slovakia and Rome as Seen in Guidebooks)]. In: Rejter, Artur; Ewa Biłas-Pleszak; Joanna Przyklenk & Katarzyna Sujkowska-Sobisz (eds.) 2018. *Wędrówka, podróż, migracja w*

języku i kulturze. Katowice: Wydawnictwo Uniwersytetu Śląskiego, pp. 433-448.

HANUSCH, Folker. (2010) "The dimensions of travel journalism: Exploring new fields for journalism research beyond the news." *Journalism Studie*s 11:1, pp. 68-82.

HANUSCH, Folker. (2013) "Broadening the focus: the case of lifestyle journalism as a field of scholarly enquiry." In: Hanusch, Folker (ed.) 2013. *Lifestyle Journalism*. Abingdon: Routledge, pp.1-10.

HANUSCH, Folker & Fürsich, Elfriede (eds.) (2014a) *Travel Journalism. Exploring Production, Impact and Culture*. Basingstoke: Palgrave Macmillan.

HANUSCH, Folker & Fürsich, Elfriede. (2014b) "On the relevance of travel journalism: An introduction." In: Hanusch, Folker & Fürsich, Elfriede (eds.) 2014. *Travel Journalism. Exploring Production, Impact and Culture*. Basingstoke: Palgrave Macmillan, pp.1-17.

HUNSTON, Susan. (2002) *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

JAWORSKA, Sylvia. (2013) "The quest for the 'local' and 'authentic': Corpus-based explorations into the discursive constructions of tourist destinations in British and German commercial travel advertising." In: Höhmann, Doris (ed.) 2013. *Tourismuskommunikation. Im Spannungsfeld von Sprach- und Kulturkontakt*. Frankfurt am Main: Peter Lang, pp. 75-100.

KUDEŁKO, Joanna. (2016) *Językowo-kulturowa rekreacja wizerunku Hiszpanii w polskich tekstach turystycznych w XX wieku* [Linguistic and cultural re-creation of an image of Spain in Polish tourist texts published in the 20th century]. Lublin: Katolicki Uniwersytet Lubelski Jana Pawła II. Unpublished doctoral dissertation.

LOMBARDINILO, Andrea. (2016) "Metafisica quotidiana. Viaggio e società nel giornalismo culturale di Flaiano e Manganelli." *Prospettiva persona* 95-96, pp. 76-81.

MANCA, Elena. (2008) "From phraseology to culture. Qualifying adjectives in the language of tourism." *International Journal of Corpus Linguistics* 13:3, pp. 368–385.

MARCHI, Anna & Taylor, Charlotte (2018) "Introduction: Partiality and reflexivity." In: Taylor, Charlotte & Anna Marchi (eds.) 2018. *Corpus approaches to discourse: A critical review*. London: Routledge, pp. 1-15.

Margarito, Mariagrazia; Marie Hediard & Nadine Celotti. (2011) *La comunicazione turistica - Lingue culture istituzioni a confronto / La communication touristique - Langues cultures institutions en face-à-face.* Torino: Libreria Cortina.

Moroz, Grzegorz. (2015) "Ryszard Kapuściński: between Polish and Anglophone travel writing." *Studies in Travel Writing* 19:2, pp. 169-183.

Nigro, Maria Giovanna. (2006) *Il linguaggio specialistico del turismo. Aspetti storici, teorici e traduttivi.* Roma: Aracne.

Pinna, Antonio. (2018) "Affect in the language of travel journalism". In: Baumann, Tania (ed.) 2018. *Reiseführer – Sprach- und Kulturmittlung im Tourismus / Le guide turistiche – mediazione linguistica e culturale in ambito turistico.* Bern: Peter Lang, pp. 151-182.

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

Podkidacz, Waldemar. (2004) "Aksjologizacja obrazu świata w tekstach przewodników turystycznych [Using axiology to depict the world image in the guidebooks]." *Poradnik Językowy* 7, pp.45-58.

Rajter, Artur. (2004) "Wzorzec tekstowy reportażu podróżniczego w aspekcie ewolucji gatunku mowy: próba syntezy." [The textual model of travel reportage in the context of the evolution of speech genres: An attempt at synthethis]. In: Malinowska, Elżbieta & Dariusz Rott (eds.) 2004. *Wokół reportażu podróżniczego*, tom 2 [Of travel reportage, vol. 2]. Katowice: Wydawnictwo Uniwersytetu Śląskiego, pp. 8–17.

Raman, Usha & Choudary, Divya. (2014) "Have travelled, will write: User-generated content and new travel journalism." In: Hanusch, Folker & Elfriede Fürsich (eds.) 2014. *Travel Journalism. Exploring Production, Impact and Culture.* Basingstoke: Palgrave Macmillan, pp. 116-133.

Serianni, Luca. (1989) *Grammatica italiana.* Torino: UTET Libreria.

Stanisławek, Joanna. (2013) "Ekfraza w tekstach użytkowych na przykładzie ofert turystycznych [Ekphrasis in the texts of tourist offers] ." *Rozprawy Komisji Językowej* 59, pp. 253-262.

Taylor, Charlotte & Marchi, Anna. (eds.) (2018) *Corpus approaches to discourse: A critical review.* London: Routledge.

Taylor, Christopher. (1998) *Language to Language: A Practical and Theoretical Guide for Italian/English Translators.* Cambridge: Cambridge University Press.

Thompson, Carl. (2011) *Travel Writing.* London: Routledge.

Witosz, Bożena. (2007) "Gatunki podróżnicze w typologicznym ujęciu genologii lingwistycznej." [Travel genres in the typology of linguistic geneaology]. In: Rott, Dariusz (ed.) 2007. *Wokół reportażu podróżniczego* [Of travel reportage]. Katowice: Wydawnictwo Uniwersytetu Śląskiego, pp. 11–29.

World Tourism Organization UNWTO (2019) *International Tourism Highlights 2019*. Online version: <https://www.e-unwto.org/doi/pdf/10.18111/9789284421152> (last accessed 30th May 2020)

Zagorska Brooks, Maria. (1975) *Polish reference grammar.* The Hague: Mouton.

Żarski, Waldemar. (2013) "Językowo-kulturowy obraz Śląska i Ślązaków w przewodnikach turystycznych." [Linguistic and cultural portraying of Silesia and the Silesians in tourist guides]. In: Ursel, Marian & Olga Taranek-Wolańska (eds.) 2013. *Śląskie pogranicza kultur, t. 2,*[Silesian borders of culture, vol.2] Wrocław: Oficyna Wydawnicza ATUT, pp.159-184.

BIONOTE / BIONOTA

David Finbar Brett is a full-time researcher at the University of Sassari and has been working in Italy for over 25 years: initially in the field of teaching English as a foreign language, and more recently in that of research in the sector of English language and translation. His main research interests include corpus linguistics, e-learning and foreign language learning, and computer assisted pronunciation training. He has given numerous presentations on these topics in international conferences and has held workshops on CALL, corpus linguistics and EFL materials development in Italy, France, Slovenia, Spain, Poland and Cyprus.

Barbara Loranc-Paszylk holds a PhD in Applied Linguistics from University of Silesia, Poland. She works as assistant professor at University of Bielsko-Biała, Poland. Her research interests focus on exploring various linguistic aspects of telecollaboration as well as innovative uses of digital resources in foreign language teaching and learning. She has published in international journals and edited volumes in the field of second language acquisition.

Antonio Pinna has an MPhil in Corpus Linguistics from the University of Birmingham (UK). He works as associate professor of English Language at the University of Sassari (Italy) where he teaches Pragmatics, (Critical)

Discourse Analysis, and English for Tourism Studies at both undergraduate and postgraduate level. His research interests include U.S. Presidential discourse, News discourse, and applications of Corpus Linguistics to various discourse types.

David Finbar Brett è ricercatore a tempo indeterminato presso l'Università di Sassari e lavora in Italia da oltre 25 anni: inizialmente nel campo dell'insegnamento dell'inglese come lingua straniera, successivamente nella ricerca nel settore scientifico Lingua Inglese e Traduzione. I suoi interessi di ricerca principali includono la linguistica dei corpora, l'e-learning e l'apprendimento delle lingue straniere, gli strumenti digitali per l'apprendimento della pronuncia. Ha presentato le sue ricerche su questi argomenti in numerosi congressi internazionali e ha tenuto workshop sul CALL, la linguistica dei corpora e lo sviluppo dei materiali in Italia, Francia, Slovenia, Spagna, Polonia e Cipro.

Barbara Loranc-Paszylk ha conseguito un dottorato di ricerca in Linguistica Applicata presso l'Università della Slesia, Polonia. Attualmente è Professore Associato all'Università di Bielsko-Biała, Polonia. I suoi interessi di ricerca riguardano l'esplorazione di vari aspetti linguistici della tele-collaborazione oltre agli utilizzi innovativi delle risorse digitali per l'insegnamento e l'apprendimento delle lingue straniere. Ha pubblicato su riviste internazionali e ha curato diversi volumi nel campo dell'acquisizione della seconda lingua.

Antonio Pinna ha un MPhil in Linguistica dei Corpora presso l'Università di Birmingham (UK). Attualmente è professore associato di Lingua Inglese all'Università di Sassari (Italia) dove insegna Pragmatica, Analisi Critica del Discorso e Lingua Inglese per il Turismo nei corsi di studio triennali e magistrali. Tra i suoi interessi di ricerca ci sono il discorso presidenziale statunitense, il linguaggio della stampa britannica e le applicazioni della linguistica dei corpora a varie tipologie di discorso.