

Saggio scientifico originale – Izvorni znanstveni rad – Original scientific paper

UDK <040.774.6BLOG:002>:811.111'322.2

<040.774.6BLOG:070.446>:811.111'322.2

<040.774.6BLOG:82-94>:811.111'322.2

DOI: 10.32728/studpol/2024.13.01.01

COLLOCATIONS IN BLOGS: A CORPUS-BASED INQUIRY

Monika Stanić
student at the Juraj Dobrila University of Pula
mostanic@unipu.hr

Katja Dobrić Basanež
Juraj Dobrila University of Pula
kdobri@unipu.hr
ORCID: 0000-0002-0261-6015

ABSTRACT

This paper presents a corpus-based analysis of collocations in blogs. It analyses collocations in three small subcorpora, focusing on books and writing, health and fitness, and personal experiences. The analysis includes a selection of the ten most frequent nouns from the wordlists produced by Sketch Engine^[1] and the investigation of their word sketches. The collocations included in the analysis are of the “modifiers of x” and “verbs with x as object” structure. The results suggest that blogs are a snapshot of a domain-specific phraseology, thus representing a valuable resource both for bloggers and language learners.

[1] Kilgariff, A., Baisa, V., Bušta, J., Jakubiček, M., Kovář, V., Michelfeit, J., Riehly, P. Suchomel, Vít, *The Sketch Engine: ten years on, Lexicography*, 1-30, 2014.

Key words: blogs, collocations, corpora, specialized language

INTRODUCTION

Collocations are an essential part of language competence, given the fact that using them correctly makes you sound like a native speaker^[2]. Awareness of this arose in the 20th century with the significant increase in the publication of dictionaries for non-native English speakers. The lexicographers' perspective on collocations then shifted from the information they could provide about word meanings to the acceptance of their being an essential part of language acquisition^[3]. The evolution of linguistic theory during the 20th century was influenced by this change in attitude, which ultimately resulted in the recognition of collocation as "an underlying principle of language production and interpretation"^[4]. As a matter of fact, it is obvious that words do not appear in texts at random and that language users have access to a vast number of semi-preconstructed phrases. Consequently, "a single choice in one slot may be made which dictates which elements will fill the next slot or slots, and prevents the exercise of free choice"^[5]. Two principles are at play here, namely, the open-choice and the idiom principle^[6]. Collocations operate on the latter one, as words seem to occur together in groups or pairs and there is a "dependency of meaning" which "correlates with the operation of the idiom principle to make fewer and bigger choices"^[7]. In other words, there are limitations related to the interaction between the constituents of collocations, suggesting that "at least one member of the prefab cannot be replaced by a synonymous item without causing a change of meaning or function and/or idiomaticity"^[8].

[2] Patekar, J., Košuta, N., *Croatian EFL learners' collocational competence: Congruent and incongruent collocations, Training, language and culture*, 6(3), 9-31, 2022, cit. pp. 9-10

[3] Barnbrook, G., Mason O., Krishnamurthy R., *Collocation Applications and Implications*, Palgrave Macmillan, London, 2013, cit. p. 4

[4] Ibidem.

[5] Moon, R. *Fixed expressions and idioms in English: A corpus-based approach*, Clarendon Press, Oxford, 1998, cit. p. 29

[6] Sinclair, J., *Corpus, Concordance and Collocation*. Oxford University Press, UK, 1991.

[7] Ibidem, cit. p. 113

[8] Erman, B., Warren, B., *The idiom principle and the open choice principle*, De Gruyter, Berlin, 2000, cit. p. 32

Despite the fact that to this date there have been numerous studies addressing this lexical phenomenon, “there is...no universally accepted formal definition of collocations”^[9]. This may be attributed to the fact that collocations are studied from various perspectives, most of which can be grouped into two main approaches, namely, the frequency-based approach and the traditional phraseological one. According to the first approach, also known as the Firthian^[10], a collocation is regarded as a combination of words, the occurrence of which is greater than chance^[11]. The second approach also takes frequency of occurrence into account, yet adopts a dual perspective by combining frequency with the semantic aspect, enabling researchers to differentiate between free word combinations and the more restricted ones to which collocations also belong. For the purpose of this paper, our understanding of collocations will take a combined approach, and include mostly binary units^[12] occurring in the corpora at least twice, but also involving a higher degree of semantic cohesion than exhibited by free word combinations. The paper focuses on collocations in blogs, a type of discourse involving informal language. Since blogs can include a variety of topics, each blog web page is hypothesized to use collocations covering those topics (e.g., culinary or beauty blogs), and, in this sense, includes instances of a certain specialized domain. It must be noted, however, that in language for specific purposes (LSP) a word combination recognized by a corpus software as a potential collocation might also represent a multi-word term. A case in point is the combination *terminal illness* that can be regarded either as a more specific term of the noun *illness*, or a term-embedding collocation^[13] (Biel 2014). We thus expect that the corpus will reveal both instances of multi-word terms and collocations, but, since it is not the purpose of this paper to dissect these differences, we will regard all units as collocations. The goal of the paper is, after all, to gain insight

[9] Mel’čuk, I., *Collocations and Lexical Functions*. In Cowie, A.P. (ed.), *Phraseology: Theory, Analysis, and Applications*, Oxford University Press, Oxford, 23-55, 1998, cit. p. 24

[10] Firth, J.R., *Modes of Meaning*. In John Rupert Firth (ed), *Papers in Linguistics 1934–1951*, Oxford University Press, London, 190–215, 1957, cit. p. 10

[11] Biber, D., Finegan, E., Johansson, S., Conrad, S., Leech, G., *Longman Grammar of Spoken and Written English*, Pearson, Edinburgh, 1999.

[12] Some collocations occur reveal a triple structure (e.g., *to feel a certain way*).

[13] Biel, Ł., *Lost in the Eurofog: The Textual Fit of EU Translated Law*, Peter Lang, Frankfurt am Main et al., 2014.

into the presence of specialized collocations in blogs both in the form of noun and verb phrases. In the first part of the paper, we will address the historical development of blogs, that is, how and why they became popular. We will then address the concept of collocation and describe our corpus and the methodological procedure, from data collection to data analysis. The results and analysis section will outline the most salient collocation examples and discuss their relevance for the discourse of blogs.

1. BLOGS

Nowadays, everyone is familiar with a blog, but few know how it came about. When Justin Hall, a student at Swarthmore College, started writing his personal web page in 1994, nobody thought that this would mark the birth of blogs. Jorn Barger, a blogger for Robot Wisdom, first used the term *weblog* to describe his method of “logging the web”, while browsing the internet in 1997. Subsequently, the term *weblog* was shortened to *blog* by the programmer Peter Merholz in 1999 and this short form is nowadays more common than the original one.^[14]

During the 1990s, when blogs first emerged, they were primarily used as online diaries. A weblog is, after all, an online platform where individuals or groups can write and publish various types of content. The content can be created by a single person or a group of people who work together to write either the same type of content or they individually settle on a topic that they would like to write about. Blog content creators also refer to the blog environment as the blogosphere^[15]. According to the Pew Research Centre^[16], since 2006 blogging has decreased in popularity among teenagers and young adults, and although this might imply that they no longer write blogs, they do “blog-like things”^[17] on other social networking sites.

[14] WDD STAFF. (2011). *A Brief History of Blogging*. Retrieved from <https://www.webdesignerdepot.com/2011/03/a-brief-history-of-blogging/> (Last access: 18th May 2024)

[15] *What is a blog?: Content Marketing Glossary*. Retrieved from <https://www.textbroker.com/blogging> (Last access: 20th June 2024)

[16] Social Media and Young Adults. <https://www.pewresearch.org/internet/2010/02/03/social-media-and-young-adults/> (Last access: 12th November 2024)

[17] Blogging ‘Peaks’, But Reports of its Death are Exaggerated. Retrieved from <https://www.wired.com/2010/12/long-live-blogging/> (Last access: 12th November 2024)

Blogs can take various forms, depending on the preferences of their creators. Some people prefer writing them in the form of online diaries, while others prefer incorporating visuals, such as pictures. As media outlets have become an indispensable part of our lives, companies have developed corporate blogs, which serve as strategic tools in their marketing because one can get immediate feedback through comments on blog posts.

The majority of blogs focus on specific topics such as travel, fashion, or cooking. Blogs are written for a variety of reasons, including the joy of writing, the exchange of knowledge and information, as well as the exploration of new trends. Furthermore, bloggers often intend to make a name for themselves, increase their web traffic or blog visits, and cherish connection with their readers. Finally, for companies, blogging serves the purpose of content marketing as well as ensuring loyalty to their customers^[18].

Since bloggers often lack an educational background in journalism or literature, the quality of a blog is typically assessed based on the benefits it offers its readers. Blogs are often written in the first person to add a personal touch, as they frequently present original ideas and personal experiences. This is also one of the elements highlighted by the successful blogger Cho Joy, who states that “...while everyone wants to know the secret to blogging success, the common denominator is to have a genuine and passionate voice that readers can connect with”^[19].

Today, it is crucial to regularly update your blog with new content. Additionally, a blog serves as a tool for effective interaction with its readers. This can be achieved by observing the number of times readers share blog posts on social media or by enabling comments from readers on specific posts. In this sense, a blog more resembles a conversation than a one-sided post or monologue.

Corpus-based investigations of blogs mostly focus on register and discourse analysis, thus producing a gap in the research on collocations in blogs. One such study^[20] on register variation in corporate blogs found

[18] *What is a blog?: Content Marketing Glossary*. Retrieved from <https://www.textbroker.com/blogging> (Last access: 20th June 2024)

[19] Cho, J.D., *Blog, Inc : blogging for passion, profit, and to create community*. Chronicle Books, San Francisco, 2012, cit. p. 6

[20] Yang, Wu, *A Corpus-based Register Analysis of Corporate Blogs – Text Types and Linguistic*

that corporate blogs are a complex genre due to their business setting (e.g., most of them are used for marketing and sales) and that there are aspects of register variation that are inseparable from one another (e.g., industries companies belong to, the context of situation). As a result, most blogs focusing on one specific topic or domain will necessarily include their own subject-field terminology and phraseology despite the fact that bloggers are not field experts. Our corpus only includes blog posts without readers' comments, unlike the study conducted by Lutzky and Kehoe^[21] that included both blog posts and readers' comments. Their study adopts a corpus linguistic approach; it departs from a list of swearwords and investigates their collocates. Our paper, on the other hand, studies the collocates of the most frequent keywords from the three subcorpora, thus revealing the subject-field phraseology of blogs and possibly contributing to its potential applicability in blogs to attract more readers.

2. CORPUS AND METHODOLOGY

As pointed out elsewhere in this paper, this research focuses on the analysis of collocations in blogs from a corpus-based perspective. The corpus examined for the purpose of this research includes three subcorpora:

1. Blogs 1 (consisting of 111,308 words), concerned with books and writing and including topics such as book reviews and writing tips.
2. Blogs 2 (consisting of 110,642 words), concerned with health and fitness and including topics about stress, various health problems and illnesses, workouts and eating healthy.
3. Blogs 3 (consisting of 112,215 words), including personal blogs and topics ranging from personal experiences, travel, and finance to tips about better cooking, simple lifestyles, etc.

The corpus collected for the purpose of this study includes 43 blogs, the posts of which we used and compiled into separate Word documents and categorized according to the chosen topics (1. Books and writing, 2. Health

Features, University of Liverpool, doctoral dissertation.

[21] Lutzky, U., Kehoe, A., *Your blog is (the) shit*, *International Journal of Corpus Linguistics IJCL*, 21(2), 165-191, 2016. DOI: 10.1075/ijcl.21.2.02lut

and fitness, 3. Personal blogs). We then uploaded the Word documents into Sketch Engine, which created three separate language corpora. After creating corpora, we extracted nouns by means of the Wordlist function, and selected the most frequent nouns for further investigation. We first examined the wordlists (see Image 1) and then decided to focus on the first 10 nouns based on their high frequency. In Blogs 3, there is one exception, given the fact that the 10th most frequent noun was the word *something*, which we decided to exclude, and include the 11th most frequent noun.

Image 1. A wordlist of nouns in Blogs 1

Noun	Frequency	Noun	Frequency	Noun	Frequency	Noun	Frequency
1 character	1,145	14 example	177	27 event	121	40 perspective	84
2 story	1,032	15 point	166	28 mystery	120	41 experience	83
3 book	426	16 fiction	162	29 post	112	42 audience	83
4 reader	307	17 people	154	30 moment	107	43 other	82
5 plot	304	18 theme	149	31 truth	106	44 romance	81
6 way	250	19 protagonist	148	32 goal	106	45 structure	80
7 scene	226	20 conflict	142	33 lie	102	46 choice	80
8 arc	205	21 author	140	34 action	97	47 blog	79
9 world	204	22 genre	137	35 chapter	96	48 reason	77
10 time	204	23 something	135	36 technique	91	49 part	75
11 life	197	24 type	129	37 fact	85	50 relationship	75
12 thing	188	25 novel	127	38 word	85		
13 writer	188	26 backstory	123	39 fantasy	84		

These nouns served as our *nodes* (see Table 1), that is, words in a collocation that are here examined^[22].

Table 1. The investigated nodes and their relative frequency in the corpora

Blogs 1	Blogs 2	Blogs 3
character (8,879.07)	body (3,447.11)	time (2,856.54)
story (8,002.79)	muscle (2,642.79)	day (2,210.78)
book (3,303.48)	weight (2,351.70)	people (2,187.98)
reader (2,380.68)	time (2,313.40)	thing (1,846.11)
plot (2,357.41)	exercise (2,244.45)	way (1,754.95)

[22] Sinclair, J., *Corpus, Concordance and Collocation*. Oxford University Press, UK, 1991, cit. p. 115

way (2,008.45)	people (2,137.21)	year (1,565.02)
scene (1,752.55)	day (1,976.35)	life (1,534.63)
arc (1,589.70)	food (1,907.40)	kid (995.23)
world (1,581.95)	calorie (1,899.74)	habit (881.27)
time (1,581.95)	training (1,899.74)	money (850.88)

Given the topics of our subcorpora, some nodes are more specialized than others and we can claim they constitute the core of certain specialized domains (e.g., *character, story, reader, book*, etc.). The high frequency of nouns denoting the time dimension in blogs is, on the other hand, more typical of general language, but since blogs report on what people do in their spare time, their high frequency is also unsurprising.

Our next step included investigating collocates through the Word Sketch function, which “can be used as a one-page summary of the word’s grammatical and collocational behavior”^[23]. In other words, word sketches of a query noun consist of many grammatical relations, depending on the syntactic function of the node in question (see Image 2).

Image 2. Word sketch of the noun story

The screenshot displays the 'WORD SKETCH' interface for the noun 'story'. It features a search bar at the top with 'Blogs 1' and a 'SUBSCRIBE' button. The main area is divided into several columns, each representing a different grammatical relation:

- modifiers of "story"**: main (20), short (12), entire (10), complex (9), fantasy (9), romance (8), other (9), literary (7), most (6), horror (5), overall (5).
- nouns modified by "story"**: structure (27), event (5), theory (3), arc (6), world (3).
- verbs with "story" as object**: tell (23), write (24), begin (10), open (7), create (8), structure (5), be (27), move (4), see (4), elevate (3), narrate (3).
- verbs with "story" as subject**: be (67), open (4), change (3), work (3), need (3), come (3), have (5).
- "story" and/or ...**: protagonist (5), novel (4), character (7), Lie (3).

[23] Sketch Engine. Retrieved from <https://www.sketchengine.eu/> (Last access: 20th May 2024)

For the purpose of our study, we selected two grammatical relations, namely, “modifiers of X” and “verbs with X as object”, assuming they would provide the most information on the collocability of nodes. A separate table with the most frequent collocates was composed for each subcorpus, which helped us analyze the results (see Tables 2, 3 and 4 in the next section). Finally, we checked whether collocations from our tables occur in enTenTen21 in order to confirm their typicality.

3. ANALYSIS AND RESULTS

The results analysed in this section include the most frequent collocates of the selected nouns categorized either as Adjective + Noun, Noun + Noun (“modifiers of x”) or Verb + Noun (“verbs with x as object”) collocations. It must be pointed out, however, that Sketch Engine sometimes lists potential collocates under the wrong grammatical relation. By way of example, the unit *moving stories* occurs in Sketch Engine as a V+N collocation due to the fact that the software falsely recognizes the syntactic function of *moving*. Since the gerund form of the verb here expresses the pre-modifying function, we listed it under combinations consisting of Adjective+Noun.

Blogs 1

Blogs 1 consist of a total of 111,308 words and include book reviews, summaries, as well as writing tips that range from creating an emotionally loaded scene to overcoming writer’s block. Both the nouns and their collocates reflect these topics (see Table 2).

Table 2. Collocations of the 10 most frequent nouns in Blogs 1

Node	Collocations	
	Adjective / Noun + Noun	Verb + Noun
character	<i>main, female, cardboard, better, interesting</i>	<i>to introduce, to create, to develop, to show, to follow</i>
Story	<i>main, short, complex, fantasy, romance, moving</i>	<i>to tell, to write, to begin, to open, to create</i>

Book	<i>crime, popular, favorite, insightful, nonfiction book</i>	<i>to read, to write, to sell, to review, to end</i>
reader	<i>potential, beta, most, modern</i>	<i>to hook, to grab, to keep, to satisfy, to help</i>
Plot	<i>main, contrived, entire</i>	<i>to turn, to drive, to complicate, to advance, to have</i>
Way	<i>best, certain, only, different, correct</i>	<i>to feel a (certain) way, to adopt a (particular) way, to work (this) way, to underestimate, to detail</i>
Scene	<i>opening, emotional, redemption, exciting, action, love</i>	<i>to write, to break, to choose, to execute, to rewrite</i>
Arc	<i>character, archetypal, story, hero, life</i>	<i>to explore, to create, to write, to follow</i>
World	<i>normal, new, fantasy, adventure, external</i>	<i>to see, to alter, to depict, to improve, to save</i>
time	<i>long, real, medieval (times), more, last</i>	<i>to spend, to take, to allot, to dedicate, to give</i>

In Blogs 1, the most frequent collocations are *character arc* (690.16), *tell a story* (178.36), *read a book* (147.34), *main character* (449.77), *main plot* (193.87), *cardboard character* (131.83), etc. The typicality of these collocations is confirmed by the data from enTenTen21, where most of these collocations occur in the top 10 candidates of the queried nodes. They reflect the importance of creating a well-developed character arc and narrating the story effectively and stress the significance of both the main character and the main plot in a book. Other collocations from the table point to the most popular genres (e.g., *romance stories* and *crime books*) as well as to the importance of female characters in the book. Most collocations refer to the fictitious world inside the book, but there are some that depict the reality (e.g., *external world*, *normal world*). The contrast is also evident in the V+N collocations, where the collocate *to alter* can only refer to the world depicted in the book, whereas *to see* can refer to both the fictitious and the real world. Therefore, we might assume that bloggers do

not write about the content of a book for the purpose of the review *per se*, but also relate this content to real life, making readers more involved, as evidenced by the following concordances:

Thus, to make sales every day, you need readers to see your books every day.

And it can slip into a common trap of “reporting” – as if the character is retelling a story from a remove, rather than as if readers are living with them directly.

What is more, readers are treated as active participants in the process of publication, providing the author feedback prior to publication (e.g., *beta readers*). It also seems that the books blogs refer to are relatively new and aimed at attracting potential readers. This is achieved metaphorically through *hooking* or *grabbing*, reducing readers to animals (fish) or objects, given that the collocate *hook* induces the meaning of seizure as if by hook^[24], and *grab* the meaning of taking or seizing with a sudden motion^[25]. Furthermore, these blogs serve as some sort of instructions or writing tips, as evidenced by some collocations (e.g., *to create an arc, to write a scene, to begin a story, etc.*). Finally, the collocates listed in Table 2 suggest that domain-specific terms which include the nouns *character* and *story* are more frequent than general nouns (e.g., *time*). While in Blogs 3 the noun *time* occurs only in singular, Blogs 1 reveal instances of plural usage, where the collocation metonymically represents a specific historical period (e.g., *medieval times*).

4. BLOGS 2

Blogs 2 consists of a total of 110,642 words. This subcorpus includes topics ranging from different strength trainings, foods to avoid before workouts, starting a sport, dehydration, coping mechanisms for stress, sleep inertia, allergies, etc. Table 3 below lists the most frequent collocations from the analyzed grammatical relations.

[24] <https://www.merriam-webster.com/dictionary/hook>. (Last access: 11th November 2024)

[25] <https://www.merriam-webster.com/dictionary/grab>. (Last access: 11th November 2024)

Table 3. Collocations of the 10 most frequent nouns in Blogs 2

Node	Collocations	
	Adjective / Noun + Noun	Verb + Noun
Body	<i>upper, entire, human, lower, whole</i>	<i>to move, to force, to help, to give, to affect</i>
muscle	<i>push, core, tight, strong, glute</i>	<i>to build, to gain, to strengthen, to target, to rebuild</i>
weight	<i>heavy, body, light, scale, moderate</i>	<i>to lose, to lift, to gain, to add, to keep, to maintain</i>
Time	<i>more, long, enough, difficult, recovery</i>	<i>to spend, to take, to save, to create, to need</i>
exercise	<i>moderate-intensity, aerobic, compound, breathing, bodyweight, cardiovascular</i>	<i>to perform, to do, to follow, to add, to make</i>
People	<i>most, active, many, other, different</i>	<i>to help, to interrupt, to describe, to know</i>
Day	<i>rest, single, busy, long, cardio</i>	<i>to start, to tackle, to rehydrate, to fill, to skip</i>
food	<i>healthy, fast, spicy, nutrient-dense, fried</i>	<i>to eat, to process, to avoid, to choose, to digest</i>
calorie	<i>total, maintenance, many, enough, total</i>	<i>to burn, to eat, to consume, to contain, to count, to cut</i>
training	<i>strength, resistance, interval, barbell, bodyweight</i>	<i>to start, to guide, to advance, to demand, to handle</i>

The most frequent collocations in Blogs 2 are *to build muscles* (283.43), *to lose weight* (352.37), *strength training* (459.62), *to burn calories* (145.54), *body weight* (130.22), etc. The collocations occur in enTenTen21 with high frequencies, again confirming their typicality. Although this was somewhat expected, most collocations in Blogs 2 refer to different body parts. Subsequently, the emphasis is on strength training and building muscles, which suggests that nowadays this is one of the most popular forms of exercise. Similarly, collocations referring to cardio training are also quite frequent, suggesting that these exercises are the second most

popular type of workout. Avoiding certain types of food (e.g., *spicy food*, *dried food*), paying attention to the number of calories we consume and aspiring to maintain moderate weight is also accentuated, suggesting that being healthy must include different aspects of care, as confirmed by the following concordances:

Strength training is essential for building muscles.

Cardiovascular exercises are effective if you want to lose weight.

Consuming way too many calories can result in gaining weight.

The time dimension also seems to be relevant here, but it very often includes metonymical reference to either the body or the exercise (e.g., *skip your cardio day*, *rehydrate your day*, etc.)

5. BLOGS 3

Blogs 3 are comprised of 112,215 words and they include personal blogs, in which bloggers write about places they travel to and their financial solutions as well as how they developed certain habits or how they deal with certain situations. We can thus assume that the texts collected for this subcorpus are more subjective than the ones in the other 2 subcorpora, as witnessed by the collocations from Table 4 below.

Table 4. Collocations of the 10 most frequent nouns in Blogs 3

Node	Collocations	
	Adjective / Noun + Noun	Verb + Noun
time	<i>screen, more, enough, hard, long</i>	<i>to spend, to take, to have, to waste, to track</i>
Day	<i>normal, recovery, work, long, dreaded</i>	<i>to start, to write (about), to dread, to dedicate, to miss</i>
People	<i>most, black, toxic, sh*tty, few</i>	<i>to let, to reward, to kill, to notice, to help</i>
Thing	<i>important, same, many, small, great</i>	<i>to do, to try, to crush (things), to choose, to make (things)</i>

Way	<i>great, easy, best, hard, different</i>	<i>to find, to finesse, to claw, to weave, to pave</i>
Year	<i>last, gap, teenage, past, childbearing</i>	<i>to spend, to end, to reach, to celebrate, to carry</i>
Life	<i>full, daily, real, glam, chaotic</i>	<i>to live, to control, to slow down, to pause, to impact</i>
Kid	<i>thrifty, underage, young, white, little</i>	<i>to teach, to empower, to involve, to provide (with), to want</i>
Habit	<i>monthly, common, daily, important, walking</i>	<i>to track, to build, to improve, to scale, to complete</i>
money	<i>enough, more, extra, unlimited, lunch</i>	<i>to save, to earn, to make, to pay, to steal</i>

The most frequent collocations here include *most people* (159.54), *to spend time* (98.76), *to save money* (98.76), *best way* (98.76), *next day* (129.15), etc. Again, most of these collocations occur in the top 10 collocations of the same nodes in enTenTen21. This subcorpus places a lot of emphasis on spending time, whether it be on learning new hobbies or skills, or simply spending time with family and friends. The time dimension is overrepresented here, as evidenced by both the nodes (e.g., *time, day, year*) and the collocates occurring with the nodes not expressing temporality (e.g., *monthly habits*). Furthermore, time can refer metonymically to the specific portion of one's life (e.g., *work days, teenage years*) or as a financial resource (e.g., *to spend time*), as supported by the following concordances:

Saying yes costs you time in the future.

No is a form of time credit.

Having money also seems to be an important part of one's personal life since collocations stress the importance of having enough or extra money, which is mostly achieved by saving it. The concept of life itself is either portrayed as a container which is full if one is satisfied with it (e.g., *full life*), or a motion we can interrupt (e.g., *to slow down / pause a life*). Finally, since Blogs 3 refer to personal experiences in general, it is not

surprising that the collocations found in this subcorpus also occur in Blogs 1 and Blogs 2 (e.g., *enough/hard time*, *normal/long day*, *best/hard way and most people*).

The comparison of results from all subcorpora suggests that blogs focusing on specific topics use specialized collocations centering around the key terms one might expect to occur in a given domain. Blogs might thus represent a small-scale specialized discourse. Furthermore, some collocations are of a highly appellative character, leading us to believe that blogs, due to their intention to attract more views and reads, could represent a subtype of operative texts^[26]. Along these lines, bloggers act as field experts and producers of texts from which their readers can benefit both in terms of content and the typical linguistic expressions occurring in a certain specialized language. That said, the collocations listed in Table 2, 3 and 4 suggest that blogs might represent a useful learning resource for language learners; by reading blogs learners can observe how collocations are used in real-life, thus gaining a deeper understanding of how words combine to convey specific meanings and ideas.

6. CONCLUDING REMARKS

The aim of this paper was to study collocations that occur in different types of blogs and examine what they can tell us about the nature of blogs. The results of the study suggest that bloggers use domain-specific collocations, which is confirmed by their high frequency in the reference corpus. It seems that bloggers design content, which as regards phraseology, resembles a certain specialized language. Accordingly, both bloggers and students can profit from the content of blogs, though they are written by non-experts. Bloggers can use the findings of this study and recognize that by selecting domain-specific collocations, they can effectively capture the interest of their readers and enhance the overall quality of their texts. For students, this research offers educational value because it highlights the significance of collocations in language learning and use. Such corpus-

[26] Reiss, K., *Type, kind and individuality of text: decision-making in translation*. In Venuti, Lawrence (ed). *The Translation Studies Reader*. London: Routledge, 168-179, 1981 / 2004.

based investigations of blogs can thus represent a useful resource for second-language acquisition classrooms, given that they, on one hand, report on popular topics, and, on the other, include typical collocations in those domains.

Although this paper represents a modest attempt at the corpus-based analysis of collocations in blogs, it offers insight into the real-life use of specialized phraseology, suggesting that it penetrates into general language. Future research, however, might profit from a bigger corpus focusing on one subject field only, which could provide more reliable results regarding types of collocations. Additionally, such corpus-based investigation can be combined with discourse analysis to reveal to what degree the frequency of certain linguistic expressions contributes to the success of rhetorical strategies.

REFERENCES

- Barnbrook, G., Mason O., Krishnamurthy R., *Collocation Applications and Implications*, Palgrave Macmillan, London, 2013.
- Biber, D., Finegan, E., Johansson, S., Conrad, S., Leech, G., *Longman Grammar of Spoken and Written English*, Pearson, Edinburgh, 1999.
- Biel, Ł., *Lost in the Eurofog: The Textual Fit of EU Translated Law*, Peter Lang, Frankfurt am Main et al., 2014.
- Cho, J.D., *Blog, Inc: blogging for passion, profit, and to create community*. Chronicle Books, San Francisco, 2012.
- Erman, B., Warren, B., *The idiom principle and the open choice principle*, De Gruyter, Berlin, 2000.
- Firth, J.R., *Modes of Meaning*. In John Rupert Firth (ed), *Papers in Linguistics 1934–1951*, Oxford University Press, London, 190–215, 1957.
- Kilgariff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Richly, P. Suchomel, Vít, *The Sketch Engine: ten years on, Lexicography*, 1-30, 2014.
- Lutzky, U., Kehoe, A., *Your blog is (the) shit*, *International Journal of Corpus Linguistics IJCL*, 21(2), 165-191, 2016. DOI: 10.1075/ijcl.21.2.02lut
- Mel'čuk, I., *Collocations and Lexical Functions*. In Cowie, A.P. (ed.), *Phraseology: Theory, Analysis, and Applications*, Oxford University Press, Oxford, 23-55, 1998.
- Moon, R. *Fixed expressions and idioms in English: A corpus-based approach*, Clarendon Press, Oxford, 1998.
- Patekar, J., Košuta, N., *Croatian EFL learners' collocational competence: Congruent and incongruent collocations, Training, language and culture*, 6(3), 9-31, 2022. <https://doi.org/10.22363/2521-442X-2022-6-3-9-31>
- Reiss, K., *Type, kind and individuality of text: decision-making in translation*. In Venuti, Lawrence (ed). *The Translation Studies Reader*. London: Routledge, 168-179, 1981 / 2004.

Sinclair, J., *Corpus, Concordance and Collocation*. Oxford University Press, UK, 1991.

Sketch Engine. Retrieved from <https://www.sketchengine.eu/> (Last access: 20th May 2024)

Yang, Wu, *A Corpus-based Register Analysis of Corporate Blogs – Text Types and Linguistic Features*, University of Liverpool, doctoral dissertation.

WDD STAFF. (2011). *A Brief History of Blogging*. Retrieved from <https://www.webdesignerdepot.com/2011/03/a-brief-history-of-blogging/> (Last access: 18th May 2024)

What is a blog?: Content Marketing Glossary. Retrieved from <https://www.textbroker.com/blogging> (Last access: 20th June 2024)

COLLOCAZIONI NEI BLOG: UN'INDAGINE BASATA SUL CORPUS

RIASSUNTO

Questo articolo presenta un'analisi basata sul corpus delle collocazioni nei blog, un tipo di linguaggio che coinvolge il linguaggio informale. Nella prima parte dell'articolo si affronta lo sviluppo storico dei blog, ovvero come e perché sono diventati popolari. Segue poi la presentazione del concetto di collocazione e la descrizione del corpus. Si passa poi alla metodologia di ricerca, dalla raccolta all'analisi dei dati. La parte dedicata alla discussione dei risultati delinea gli esempi di collocazione più salienti e ne discute la rilevanza per il linguaggio dei blog, mentre le osservazioni conclusive suggeriscono le potenziali implicazioni dei risultati e le linee di ricerca future. Gli studi sui blog basati sul corpus si concentrano per lo più sull'analisi del registro e del discorso (in ingl. *register and discourse analysis*), producendo così una lacuna nella ricerca sulle collocazioni nei blog. Uno di questi studi (Yang 2015) sulla variazione di registro nei blog aziendali ha rilevato che i blog aziendali sono un genere complesso a causa del loro contesto commerciale (ad esempio, la maggior parte di essi sono utilizzati per il marketing e le vendite) e che ci sono aspetti della variazione di registro che sono inseparabili l'uno dall'altro (ad esempio, le industrie a cui appartengono le aziende, il contesto della situazione). Il nostro corpus comprende solo post di blog senza commenti dei lettori, a differenza dello studio condotto da Lutzky e Kehoe (2016) che ha incluso sia i post di blog sia i commenti dei lettori su tali post. Il loro studio adotta un approccio basato sulla linguistica dei corpora; parte da un elenco di parolacce e analizza i loro collocati. Il nostro lavoro, invece, studia le collocazioni delle parole chiave più frequenti, rivelando così la fraseologia del campo tematico dei blog e indicando la sua potenziale applicabilità nei blog per attirare più lettori.

Il corpus progettato ai fini del presente lavoro comprende tre piccoli sottocorpora: Blog 1, su libri e scrittura, Blog 2, su salute e fitness, e Blog 3, su esperienze personali. L'analisi comprende dieci sostantivi risultati i

più frequenti nelle liste di parole prodotte da Sketch Engine^[27] e l'analisi dei loro *wordsketches*. Le collocazioni incluse nell'analisi sono della struttura “modificatori di x” e “verbi con x come oggetto”. La loro tipicità è confermata dalle interrogazioni del corpus enTenTen21.

Tutti e tre i sottocorpora rivelano che i blogger utilizzano correttamente le collocazioni specializzate, il che è confermato dalla loro alta frequenza nel corpus di riferimento. I risultati suggeriscono che i blog incorporano la fraseologia tipica del dominio preso in esame, rappresentando così un piccolo campione nel linguaggio analizzato. Inoltre, alcune collocazioni hanno un carattere fortemente appellativo, il che ci porta a credere che i blog, per il loro intento di attirare più visualizzazioni e letture, possano rappresentare un sottotipo di testi operativi.

Sebbene questo lavoro rappresenti un modesto tentativo di analisi basato sul corpus delle collocazioni nei blog, offre una visione dell'uso reale della fraseologia specializzata, suggerendo che essa penetra nel linguaggio generale.

Parole chiave: blog, collocazioni, corpora, fraseologia specializzata

[27] Kilgariff, A., Baisa, V., Bušta, J., Jakubiček, M., Kovář, V., Michelfeit, J., Riehlý, P. Suchomel, Vít, *The Sketch Engine: ten years on, Lexicography*, 1-30, 2014.