

# Collocations in blogs

---

**Stanić, Monika**

**Undergraduate thesis / Završni rad**

**2023**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Pula / Sveučilište Jurja Dobrile u Puli**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:137:766944>

*Rights / Prava:* [In copyright](#)/[Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2024-11-20**



*Repository / Repozitorij:*

[Digital Repository Juraj Dobrila University of Pula](#)



Sveučilište Jurja Dobrile u Puli  
Filozofski fakultet

Monika Stanić  
**COLLOCATIONS IN BLOGS**  
Završni rad

Pula, rujan 2023.

Sveučilište Jurja Dobrile u Puli  
Filozofski fakultet

Monika Stanić

## **COLLOCATIONS IN BLOGS**

Završni rad

**JMBAG:** 0303097482, redoviti student

**Studijski smjer:** preddiplomski studij Engleski jezik i književnost, Japanski jezik i kultura

**Predmet:** Morfosintaksa engleskog jezika

**Znanstveno područje:** Humanističke znanosti

**Znanstveno polje:** Frazeologija

**Znanstvena grana:** Anglistika

**Mentor:** Katja Dobrić-Basaneže

Pula, rujan 2023.



## IZJAVA O AKADEMSKOJ ČESTITOSTI

Ja, dolje potpisana \_\_\_Monika Stanić\_\_\_, kandidatkinja za prvostupnicu Engleskog jezika i književnosti i Japanskog jezika i kulture izjavljujem da je ovaj Završni rad rezultat isključivo mogega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na objavljenu literaturu kao što to pokazuju korištene bilješke i bibliografija. Izjavljujem da nijedan dio Završnog rada nije napisan na nedopušteni način, odnosno da nije prepisan iz kojega necitiranog rada te da nijedan dio rada ne krši bilo čija autorska prava. Izjavljujem, također, da nijedan dio rada nije iskorišten za koji drugi rad pri bilo kojoj drugoj visokoškolskoj, znanstvenoj ili radnoj ustanovi.

Student

---

U Puli, \_\_\_\_\_, \_\_\_\_\_ godine



## IZJAVA O KORIŠTENJU AUTORSKOGA DJELA

Ja, \_\_\_Monika Stanić\_\_\_, dajem odobrenje Sveučilištu Jurja Dobrile u Puli, nositelju prava korištenja, da moj završni rad pod nazivom „\_\_\_Collocations in blogs\_“ upotrijebi da tako navedeno autorsko djelo objavi u javnoj internetskoj bazi Sveučilišne knjižnice Sveučilišta Jurja Dobrile u Puli te preslika u javnu internetsku bazu završnih radova Nacionalne i sveučilišne knjižnice (stavljanje na raspolaganje javnosti), sve u skladu sa Zakonom o autorskom pravu i drugim srodnim pravima i dobrom akademskom praksom, a radi promicanja otvorenoga, slobodnoga pristupa znanstvenim informacijama. Za korištenje autorskog djela na gore navedeni način ne potražujem naknadu.

Potpis

---

U Puli, \_\_\_\_\_, \_\_\_\_\_ godine

## Contents:

<b>1. INTRODUCTION</b> .....	1
<b>2. BLOGS</b> .....	2
<b>3. PHRASEOLOGY</b> .....	4
<b>3.1. Typologies of phraseological units</b> .....	5
<b>3.2. Collocations</b> .....	9
<b>4. CORPUS LINGUISTICS</b> .....	13
<b>4.1. What is a linguistic corpus?</b> .....	13
<b>5. CORPUS AND METHODOLOGY</b> .....	16
<b>6. RESULTS</b> .....	20
<b>6.1. Blogs 1</b> .....	20
<b>6.2. Blogs 2</b> .....	21
<b>6.3. Blogs 3</b> .....	24
<b>7. ANALYSIS AND DISCUSSION</b> .....	26
<b>8. CONCLUSION</b> .....	28
<b>BIBLIOGRAPHY</b> .....	29
<b>APPENDIX</b> .....	32
<b>ABSTRACT</b> .....	34
<b>SAŽETAK</b> .....	35

## 1. INTRODUCTION

Unlike other linguistic disciplines, phraseology is relatively new and therefore much less researched. That also applies to the study of collocations, which are an essential part of language competence, given the fact that using them correctly makes you sound like a native speaker (Patekar & Košuta, 2022:9-10). Naturally, each register possesses its own specific terminology and phraseology. For instance, in the language of medicine there are collocations like *terminal illness*, *to prescribe drugs*, *to respond well to treatment*,<sup>1</sup>etc. When it comes to blogs, however, there is no blog-specific terminology, which would bring about a subject-field phraseology that bloggers would use. After all, their goal is to persuade readers to read their posts. For this reason, we can claim that blog posts are characterized by informal language that addresses their target audience. Since they can include a variety of topics, each blog web page necessarily also uses collocations centering around key terms covering different topics (e.g. culinary or beauty blogs).

This thesis will provide an insight into the most frequently used collocations in the selected blogs and examine what they can tell us about blogs themselves and the type of language that they use.

The thesis is divided into eight parts. After the introduction, it discusses the history of blogs, that is, how and why they became popular. In the third part, the thesis provides an overview of phraseology as a discipline, including the typologies of phraseological units. In other words, it examines how different authors refer to similar concepts using different terminology. Additionally, this section also explores the etymology, usage, and common types of collocations. Next section includes a definition of both corpus linguistics and a linguistic corpus. The methodology section describes the corpus as well as the entire research process, from data collection to data analysis, including the tools that were used. Finally, the thesis provides an analysis of the data and some concluding remarks.

---

<sup>1</sup> Online OXFORD Collocation Dictionary (n.d.), Retrieved August 20, 2023, from <https://www.freecollocation.com/>

## 2. 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 will 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. The term "weblog" was first shortened to "blog" by the programmer Peter Merholz in 1999 and the short term is nowadays more used than the original one.<sup>2</sup>

During the 1990s, when blogs first emerged, they were primarily online diaries. We can thus claim that a weblog is an online platform where individuals or groups can write and publish various types of content. They can be the work of 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 talk about. Blog content creators also refer to the blog environment as the blogosphere. (*What is a blog?*, 2022)

Blogs come in various forms, depending on the preferences of the creators. Some people prefer writing them in the form of online diaries, while others prefer incorporating more visuals. As media outlets have become a part of our lives, companies have developed corporate blogs, which serve as strategic tools in their marketing because you can get immediate feedback through comments on your blog posts. According to Jeremy Wright (Wright, 2005;2), "basing your business decisions on actual customer feedback and market intelligence is the smartest business move you could ever make."

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. Reasons may vary depending on whether it is a personal or a business blog, but they often include making a name for yourself or building yourself a reputation, increasing web traffic or blog visits and, having a connection with your readers, etc. Finally, for companies blogging serves the purpose

---

<sup>2</sup> WDD STAFF. (2011). *A Brief History of Blogging*. Retrieved September 11, 2023, from <https://www.webdesignerdepot.com/2011/03/a-brief-history-of-blogging/>



of content marketing as well as ensuring loyalty for their customers. (*What is a blog?*, 2022)

Since bloggers often lack an educational background in journalism, the quality of a blog is typically assessed based on the benefits it offers to the 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." (Cho, 2012:6)

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

### 3. PHRASEOLOGY

Phraseology is “a subfield of lexicology dealing with the study of word combinations rather than single words” (Granger & Paquot, 2008:3). Phraseology studies idioms, proverbs, collocations, and other types of phraseological units that play a vital role in everyday language and contribute to effective communication. It is important to note that it hasn't been long since phraseology had its acknowledgement among Western linguists, which it received very early from the linguists in the former Soviet Union and Eastern Europe in general (Cowie, 1998:1).

Phraseology not only plays a crucial role in language learning, but it is also essential for the true understanding of language. This is because phraseological units often have specific meanings when used together, which may differ from the meanings of their individual constituents. Additionally, these expressions include a cultural element, and understanding is vital for effective communication, particularly with individuals from different cultural backgrounds. Therefore, it is important for linguists and learners of both the first and second languages to study phraseology thoroughly.

The study of related disciplines, including lexicology, lexicography, discourse analysis, corpus linguistics, cognitive linguistics, psycholinguistics, stylistics, etc., has shown that phraseology and phraseological units are frequently used in both speech and writing. Omazić (Omazić, 2015:17) suggests that modern linguistics started to study language as a mental phenomenon and that it focuses on studying language in use rather than its ideal state. That shifted the interest of linguistic scholars towards the previously forgotten and neglected registers.

It is evident from the above paragraphs that phraseological units have multiple functions. For example, if we look at them from the psycholinguistic perspective, they are something that we have learned and that we use in our everyday communication. Therefore, while we are talking, we access our brain and pull them from our memory. In this way, we can see how language acquisition depends on learning the already ready-made formulas, that is, phraseological units (Omazić, 2015:28).

Igor Mel'čuk emphasizes that collocations form a central part of the entire linguistic system, and acknowledges that "there is...no universally accepted formal definition of collocations" (Mel'čuk, 1998:24). However, Mel'čuk (Mel'čuk, 1998:24) also states that collocations are "a subclass of what are known as set phrases" and set phrases are "fixed (frozen) phrases, word combinations, idioms, etc."

We believe that collocations are a useful tool for blogging because they can improve the effectiveness of writing as well as make blogs more engaging for their audience. One way to use collocations in your blog is to consider the topic of your post and identify common collocations that are associated with it. For example, if you are writing about relationships, some of the collocations one might include are *red flags* and *non-verbal communication*. Similarly, a post about healthy lifestyle, could use collocations like *balanced diet* or *full-body workout*. Knowing your target audience is thus important, since one needs to use collocations they are familiar with in order for those readers to identify with the blog.

### **3.1. Typologies of phraseological units**

Since phraseology is a relatively young and complex field of study, setting clear boundaries and adopting a unified typology of phraseological units is challenging. Consequently, numerous researchers propose their own categorizations, leading to multiple terms for a similar concept, which can be confusing for people who get acquainted with the discipline. For example, 'word combinations', 'multi-word expressions', 'collocations', and 'phraseological units' refer to a similar concept but different researchers use different terms.

Cowie (Cowie, 1988), for instance, uses the term 'word combinations' and divides them into:

- 1) Composites, which include:
  - a) Restricted collocations (e.g. *heavy sleeper*)
  - b) Figurative idioms (e.g. *fishy*)
  - c) Pure idioms (e.g. *a piece of cake*)
- 2) Formulae, which are further subdivided into:
  - a) Routine formulae (e.g. *good morning* or *see you later*)
  - b) Speech formulae (e.g. *if anything*)

According to Cowie (Cowie 1998: 4 cited in Granger & Paquot 2008:36) composites are units “which function syntactically at or below the level of a sentence” while formulae are units “which function pragmatically as autonomous utterances” (Granger and Paquot 2008: 36). *Heavy sleeper* is an example of a restricted collocation, which means that the meaning of the individual word is restricted to that two-word phrase. Here, adjective *heavy* does not mean that the person sleeping is heavy in weight, but rather that they sleep deeply. An example of a figurative idiom is *something is fishy*, where *fishy* refers to something suspicious or weird. Lastly, the example of a pure idiom example would best be illustrated with the unit *a piece of cake*, whose meaning is ‘easy/without problems’.

Routine formulae are phrases that frequently appear in certain types of recurring social settings (e.g. *good morning* or *see you later*). Speech formulae are phrases or expressions that are usually in the form of a sentence and they convey speaker's attitude and make the conversation flow more easily. An example of a speech formula is *if anything*, which means ‘on the contrary’.

Mel'čuk's typology (Mel'čuk, 1998) is very similar to Cowie's and he distinguishes between:

- 1) pragmatic phrasemes
  - a) pragmatemes (e.g. *watch your step*)
- 2) semantic phrasemes, which are further divided into:
  - a) semi-phrasemes or collocations (e.g. *to pay a visit*)
  - b) quasi-phrasemes or quasi-idioms (e.g. *barbed wire*)
  - c) full phrasemes or idioms (e.g. *kill two birds with one stone*)

His semantic phrasemes are very similar to Cowie's composites and pragmatic phrasemes are very close to Cowie's formulae.

Gläser (Gläser, 1994/95), on the other hand, makes a distinction between ‘sentence-like’ and ‘word-like’ phraseological units. Her typology consists of ‘nominations’ and ‘propositions’, but there is also a transition area, which includes:

- 1) Nominations, i.e. word-like phraseological units that are further divided into:
  - a) idioms (e.g. *green thumb*)

- b) non-idioms that consist of:
- terminological word groups (e.g. *atomic warfare*)
  - onymic entities (e.g. *the Rocky Mountains*)
  - clichés (e.g. *happily ever after*)
  - paraphrasal verbs (e.g. *to wink at*)
  - operator or function words (e.g. *by my side*)

2) Propositions, i.e. sentence-like phraseological units that consist of:

- a) commonplaces (e.g. *The pen is mightier than the sword.*)
- b) routine formulae (e.g. *How do you do?*)
- b) proverbs (e.g. *Actions speak louder than words.*)
- c) commandments and maxims (e.g. *know thyself*)
- d) slogans (e.g. *just do it*)
- e) quotations and winged forms (e.g. *Love looks not with the eyes, but with the mind, and therefore is winged Cupid painted blind*)

It must be pointed out that there are some phraseological units that can be classified either as 'nominations' or 'propositions'. These include:

- a) irreversible binominals (e.g. *dead or alive*)
- b) stereotyped comparisons or similes (e.g. *as light as a feather*)
- c) proverbial sayings (e.g. *All that glitters is not gold.*)
- d) fragments of proverbs (e.g. *spilled milk*)
- e) literary allusions and fragments of quotations (e.g. *Stay gold, Ponyboy*)

Rosamund Moon's typology (Moon, 1998), on the other hand, consists of three major groups of phraseological units that include 'anomalous collocations', 'formulae' and 'metaphor'. Anomalous collocations deal with lexico-grammar problems and are further subdivided into:

- 1) ill-formed collocations (e.g. *gigantic breakfast*)
- 2) cranberry collocations (e.g. *break the ice*)
- 3) defective collocations (e.g. *make a favor*)
- 4) phraseological collocations (e.g. *give a hand*)

Formulae deal with problems of pragmatics and we differentiate between:

- 1) simple formulae (e.g. *you know*)
- 2) sayings (e.g. *A little bird told me*)
- 3) proverbs (e.g. *A picture is worth a thousand words*)
- 4) similes (e.g. *as white as snow*)

Metaphors deal with semantic problems and includes:

- 1) transparent metaphors (e.g. *a sea of troubles*)
- 2) semi-transparent metaphors (e.g. *time is a thief*)
- 3) opaque metaphors (e.g. *kick the bucket*)

Burger's typology (Burger, 1998), on the other hand, is founded on the way phraseological units work in discourse. He divides phraseological units into:

- 1) referential units:
  - a) propositional
  - b) nominative:
    - collocations
    - partial idioms (e.g. *raining buckets*)
    - idioms
- 2) structural units
- 3) communicative units:
  - a) routine formulae

Granger and Paquot's typology (Granger & Paquot, 2008) resembles Burger's, given the fact that it is based on its function in discourse. They distinguish between:

- 1) referential phrasemes:
  - a) idioms
  - b) irreversible bi- and trinomials (e.g. *alive and kicking*)
  - c) similes
  - d) compounds (e.g. *pickpocket*)
  - e) phrasal verbs (e.g. *give in*)

- f) grammatical collocations (e.g. *pay attention*)
  - g) (lexical) collocations (e.g. *child care*)
- 2) textual phrasemes:
- a) complex prepositions (e.g. *in spite of*)
  - b) complex conjunctions (e.g. *even though*)
  - c) linking adverbials (e.g. *on the other hand*)
  - d) textual sentence stems (e.g. *to illustrate this*)
- 3) communicative phrasemes:
- a) speech act formulae (e.g. *Would you mind...?*)
  - b) attitudinal formulae (e.g. *thank you*)
  - c) proverbs and proverb fragments (e.g. *Better late...*)
  - d) slogans (e.g. *Think Different.*)
  - e) commonplaces (e.g. *Don't judge a book by its cover.*)
  - f) idiomatic sentences
  - g) quotations

In simple words, referential phrasemes refer to real-life objects while textual phrasemes are used for linking the content together so it would make sense. Communicative phrasemes, on the other hand, serve as participants in the conversation.

### **3.2. Collocations**

According to G. Barnbrook (Barnbrook et al., 2013:3) the word collocation was originally introduced into English in the sixteenth century and its usage has undergone significant changes over time. It stems from the Latin word *collocatio* (*collocare*) which means set or place together (Harper, n.d.).

G. Barnbrook (Barnbrook et al., 2013:3) suggests that the word is typically used in three ways:

- 1) to describe the way words naturally group together in texts,
- 2) to describe the analysis tool used to investigate this grouping and to analyze its significance and implications, and
- 3) to describe a process of language production in which pre-fabricated language chunks are used to construct utterances.

The fact that we can find collocations in dictionary entries and other texts is in itself evidence of its usage over the centuries. This demonstrates that many lexicographers from different periods recognized the importance of collocations (Barnbrook et al., 2013:4).

Concordances can also provide insight on the use of collocation. Concordance is “a collection of the occurrences of a word-form, each in its own textual environment” (Sinclair, 1991:32). In other words, it is a list or an index. At first, concordances were used as indices for the Bible, and eventually, they were applied on texts that were thought to be important enough to need assistance for their analysis (Barnbrook et al., 2013:4).

According to G. Barnbrook (Barnbrook et al., 2013:4), in the 20th century there was a hyperproduction of dictionaries for non-native English speakers and even before that in the 19th century there were style guides which offered advice about language use. With the publication of these dictionaries, the lexicographers' perspective on collocations has shifted from the dependance on information it can provide about word meanings to the acceptance that it is an essential part of language acquisition. The evolution of linguistic theory during the 20th century was influenced by this attitude change, which ultimately resulted in the recognition of collocation as “an underlying principle of language production and interpretation” (Barnbrook et al., 2013:4).

According to John Sinclair (Sinclair, 1991:109), there are two models we can use in order to interpret the meaning of texts, as relying on only one is not enough. There is the open-choice principle and the idiom principle. To better understand them, both principles will be explained at this point. Additionally, the focus will be on the one that is applied to collocations.

In simple words, the open-choice principle shows us how a text came into existence on the basis of multiple possible choices. Due to these multiple possible choices the only way to somehow control that vast number of open possibilities is, according to John Sinclair (Sinclair 1991:109), the principle of “grammaticalness”. This principle is also called the ‘slot-and-filler’ model, where texts represent empty slots that need to be filled and practically any word can occur within those slots.



It is obvious that words do not appear in texts at random and that the open-choice principle does not impose strong enough restrictions on consecutive selections. According to the idiom principle, on the other hand, language users have access to a vast number of semi-preconstructed phrases that represent single choices, despite appearing to be regarded as independent segments. It is seen as inferior because it does not fit the open-choice model. Moon explains this as follows: “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” (Moon 1998:29).

The main difference between the idiom principle and the open-choice principle is the inability, or limitations, related to the verbal interaction in certain expressions, which means 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” (Erman & Warren, 2000:32).

Collocations operate on the idiom principle, as words seem to occur together in groups or pairs. According to John Sinclair (Sinclair, 1998:113), there is a “dependency of meaning” which “correlates with the operation of the idiom principle to make fewer and bigger choices”. Collocations support this claim, since they represent words that often co-occur together. Subsequently, those words that appear together frequently because they have a specific meaning when combined is what makes them significant as a pair.

There are several types of collocations defined on the basis of word classes of their constituents. Some of the most common types are:

- 1) Noun + noun
  - *board game, desk job, gold mine, head teacher, etc.*
- 2) Noun + verb
  - *lions roar, dogs bark, bomb went off, etc.*
- 3) Verb + noun
  - *pay attention, break the law, make money, save money, etc.*
- 4) Adjective + noun
  - *great pride, empty words, free speech, fatal mistake, etc.*

5) Adverb + adjective

- *bitterly cold, highly controversial, blissfully ignorant, etc.*

6) Adverb + verb

- *deeply rooted, badly damage, strongly deny, etc.*

7) Verb + expression with preposition

- *burst into giggles, ask in advance, care about etc.*

It must be pointed out that corpus linguistics has a different understanding of collocations than traditional phraseology. In corpus linguistics, collocations are defined as “a series of words or terms that co-occur more often than would be expected by chance”.<sup>3</sup> One example from this research would be *to keep a reader*, which a corpus tool lists as a potential collocation, but every linguist knows that the verb *keep* requires another obligatory complement here (e.g. *to keep a reader interested*).

---

<sup>3</sup> Online OXFORD Collocation Dictionary (n.d.), Retrieved September 11, 2023, from <https://www.freecollocation.com/>

## **4. CORPUS LINGUISTICS**

Despite being quite unpopular in the 1960s and 1970s, corpus linguistics is nowadays one of the most widely used methodologies in linguistics. Initially, it was mainly used in English linguistics, particularly in the study of grammar. According to McEnery and Wilson (McEnery & Wilson, 2001:1) corpus linguistics is “the study of language based on examples of real-life language use”. While this definition is correct, it is also very broad, which means that it covers a lot of methodological approaches that do not belong to corpus linguistics. Some authors like Cheng and Meyer describe corpus linguistics as “the compilation and analysis of corpora” (Cheng, 2012:6; Meyer, 2002:11). That description alludes that a corpus is a compilation and a recording of ‘real life language use’. (Stefanowitsch, 2020:20)

When approaching corpus linguistics as a language learner, one may wonder whether corpus linguistics is a methodology or a branch of linguistics. According to McEnery and Wilson (McEnery & Wilson, 2001:2) corpus linguistics is not a branch of linguistics in the same sense as pragmatics and syntax, which focus on explaining specific aspects of language use. Instead, corpus linguistics is more of a methodology than an area of linguistics that aims to describe or explain some certain aspects of language. It is a methodology that can be applied to almost all areas of linguistics, but it does not explain or represent an area of linguistics itself. As corpus linguistics is “not defined by the object of study...the object of corpus linguistics is not the study of corpora. It is rather the study of language through corpora” (Johansson, 1995:19 cited in Biel, 2009:2), or “a methodology with a theoretical status” (McEnery et al., 2006:7-8 cited in Biel, 2009:2).

### **4.1. What is a linguistic corpus?**

The definition of the word “corpus” can vary slightly depending on the academic field. In general, it refers to a collection of texts. For example, in literary studies, this collection of texts might include all author's works (e.g. all William Blake's poems). It can also include

works belonging to specific genre or time, such as all novels from the 19<sup>th</sup> century. In theology, it can refer to a particular translation of the Bible.

According to Anatol Stefanowitsch (Stefanowitsch, 2020:22-23), in corpus linguistics, the term is used differently and it refers to an electronic collection of texts, hence, examples of language use that meet a specific criterion. A corpus must thus represent authentic examples of language use, be representative of the language or language diversity which is being researched and, last but not least, it must be of considerable size. It is also important to add that these electronic corpora are mostly, but not always, annotated which means that they add metadata to the text<sup>4</sup> (e.g. indicating a word's word class).

Authentic language in spoken form, according to Anatol Stefanowitsch (Stefanowitsch 2020:23), is a "language produced for the purpose of communication, not for linguistic analysis or even with the knowledge that it might be used for such a purpose." In the case of written language, it is easier to meet the criterion of authenticity because the writer does not know that his or her writing will be taken as a sample for linguistic research.

The second criterion a reliable corpus must meet is its representativeness. A representative sample is a portion of a population that, in terms of the distribution of the phenomenon under research, is equal to the population as a whole. The distribution of linguistic phenomena, hence, grammatical structures (nouns, verbs, etc.) in a corpus which is a sample of language use, would therefore need to match the distribution in the language as a whole for it to be considered representative of that language.

Corpus size, on the other hand, is said to contribute to the second criterion. From a logical point of view, there needs to be a correlation between sample size and representativeness. For example, if we include every possible variation of a language into our corpus, it would inevitably be representative, and if the sample size were to be reduced, the representativeness would not immediately disappear, but it would decrease pretty quickly. (Stefanowitsch, 2020:34) According to Anatol Stefanowitsch (Stefanowitsch, 2020:37), if we take out just 1% of all texts written in a given language,

---

<sup>4</sup> Sketch Engine. (n.d.). *Adding structures and metadata*. Retrieved September 13, 2023, from <https://www.sketchengine.eu/guide/annotating-corpus-text/?highlight=annotation>

then perhaps entire language variations are already lost. Stefanowitsch also suggests that “current corpora that at least make an honest attempt at diversity currently range from one million to about half a billion” (Stefanowitsch, 2020:38). However, according to Bowker and Pearson (Bowker & Pearson, 2002:48) “corpora intended for LSP (Language for specific purposes) can be smaller than those used for LGP (Language for general purposes) studies”.

It is important to note that corpus can be both written and spoken. Texts in written corpora can be taken from scanned newspapers, books, works of literature, etc. while spoken corpora includes transcripts of spoken language such as everyday conversations, recorded phone calls, radio broadcasts, TV shows, etc. (McCarthy, 2004:1). Further distinctions include monolingual corpora, which include texts in only one chosen language, and multilingual corpora, which include texts in two or more languages. A parallel corpus, on the other hand, would include two or more monolingual corpora that are each other’s translations (e.g. EUR-Lex). A comparable corpus in turn contains two or more monolingual corpora that are not translations, but include the same domain (*Sketch Engine*, 2023).

## 5. CORPUS AND METHODOLOGY

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 (110,308 words), concerned with books and writing and including topics such as book reviews and writing tips.
- 2) Blogs 2 (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 (112,215 words), including personal blogs and topics ranging from personal experiences, travels, and finances to tips about better cooking, simple lifestyle, etc.

The first two are extremely popular topics among bloggers, while the third topic covers a wide range of subjects. The purpose of selecting these topics was to explore potential differences in collocations between them. Texts stem from the time period between 2011 and 2023, with most of the texts being from recent years.

It must be pointed out that corpus-based investigations of blogs conducted thus far mostly focus on register and discourse analysis. One such study was conducted by Yang Wu (2015), who examined register variation in corporate blogs. He found that corporate blogs are a complex genre because of 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 each other (e.g. industries companies belong to, context of situation), among others (Yang, 2015:269).

Furthermore, it seems that research into collocations in blogs are also scarce. A somewhat related study includes a research conducted by Ursula Lutzky and Andrew Kehoe (Lutzky & Kehoe, 2016) that included both blog posts and readers' comments on those posts. Lutzky and Kehoe conducted a research on swearing in computer mediated communication. They used a corpus linguistic approach and produced an initial list of swearwords and then they extracted their collocates. Their research, however, differs

from the research conducted for the purpose of this thesis, since it only focuses on the collocates of swearwords.

The corpus collected for the purpose of this study includes 43 blogs, whose posts were used and compiled into a Word document and categorized according to their three chosen topics (1. Books and writing, 2. Health and fitness, 3. Personal blogs). Sketch Engine was used as the primary tool for this research. The Word documents were uploaded into Sketch Engine, which created three separate language corpora. After creating corpora, the data analysis began by extracting nouns from all corpora using a wordlist, and then selecting the most frequent nouns for further investigation. Wordsketches of the ten most frequent nouns in each corpus were then investigated and analyzed.

Firstly, the wordlists of nouns from these corpora were examined. For the purpose of this research, we only focused on the first 10 nouns due to their high frequency (see Table 1). In Blogs 3, there is one exception, given the fact that the 10<sup>th</sup> most frequent noun was, the word *something*, which we decided to exclude, and include the 11th most frequent noun.

Noun	Frequency ? ↓	Frequency Per Million ? ↓	Noun	Frequency ? ↓	Frequency Per Million ? ↓
1 character	1,145	8,879.07 ...	26 backstory	123	953.82 ...
2 story	1,032	8,002.79 ...	27 event	121	938.31 ...
3 book	426	3,303.48 ...	28 mystery	120	930.56 ...
4 reader	307	2,380.68 ...	29 post	112	868.52 ...
5 plot	304	2,357.41 ...	30 moment	107	829.75 ...
6 way	259	2,008.45 ...	31 truth	106	821.99 ...
7 scene	226	1,752.55 ...	32 goal	106	821.99 ...
8 arc	205	1,589.70 ...	33 lie	102	790.97 ...
9 world	204	1,581.95 ...	34 action	97	752.20 ...
10 time	204	1,581.95 ...	35 chapter	96	744.45 ...
11 life	197	1,527.66 ...	36 technique	91	705.67 ...
12 thing	188	1,457.87 ...	37 fact	85	659.14 ...
13 writer	186	1,442.36 ...	38 word	85	659.14 ...
14 example	177	1,372.57 ...	39 fantasy	84	651.39 ...
15 point	166	1,287.27 ...	40 perspective	84	651.39 ...
16 fiction	162	1,256.25 ...	41 experience	83	643.64 ...

**Image 1.** Wordlist nouns of Blogs 1

These nouns serve as our “nodes” and we investigated their collocates (see Table 1). Sinclair (Sinclair, 1991:115) describes a “node” as the word in a collocation that is being

examined and the term “collocates” as any word that “occur in the specified environment of a node.”

<u>Blogs 1</u>	<u>Blogs 2</u>	<u>Blogs 3</u>
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)
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)

**Table 1.** Most frequent nouns used as nodes in Blogs 1, Blogs 2 and Blogs 3

Given the topics of our subcorpora, some nodes were expected (e.g. *character*, *story*, *reader*, *book*, etc.). What was not expected, however, was that there was a high frequency of nouns involving the time dimension in Blogs 3.

Collocates of the selected nouns were then analysed by means of the Word Sketch function, which “processes the word’s collocates and other words in its surroundings. It can be used as a one-page summary of the word’s grammatical and collocational behavior.” (*Sketch Engine*, 2023). It is for this reason that this function is perfect for investigating collocations. It produces many grammatical relations (“modifiers of X”, “verbs” with X as subject, etc.) which can serve as a typology of collocations based on the word classes of their constituents (see Image 2).



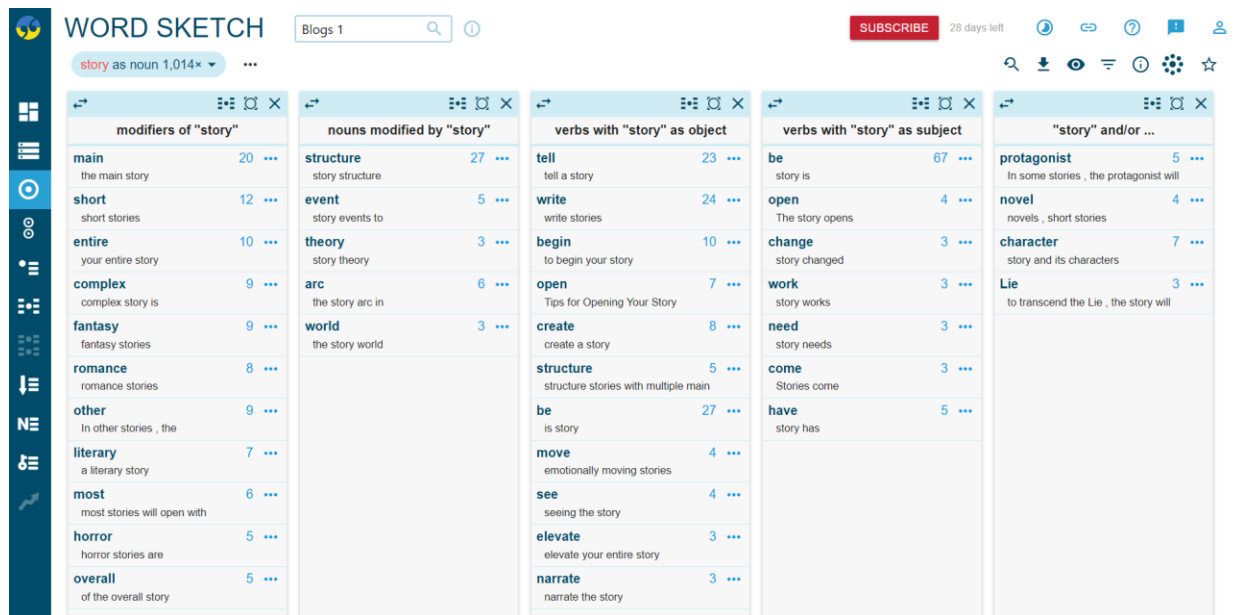


Image 2: Word sketch of the noun story

Since Sketch Engine lists many grammatical relations, we decided to focus only on “modifiers of X” and “verbs with X as object”, because they provided most information on the collocability of nodes. A separate table was then composed for each subcorpus, which helped us analyze the results (see Tables 2, 3 and 4).

## 6. RESULTS

Results of the analysis suggest that all three subcorpora include collocations that bloggers use to attract more views and readings. The following subsections will elaborate on the results in more specific detail.

### 6.1. Blogs 1

At the end of the data collection, Blogs 1 had a total of 111, 308 words. Blog posts concerning book reviews and summaries as well as writing tips that focus on creating an emotionally packed scene to overcoming writer's block were used.

Node	Collocations	
	modifiers of X	verbs with X as object
Character	<i>main character, female character, cardboard character, interesting character</i>	<i>to introduce a character, to create a character, to develop a character</i>
Story	<i>main story, short story, complex story, fantasy story, romance story, literary story</i>	<i>to tell a story, to write a story, to begin a story, create a story, to narrate a story</i>
Book	<i>crime book, popular book, favorite book, insightful book, nonfiction book</i>	<i>to read a book, to write a book, to sell a book, to review a book, to end a book, to devote a book</i>
Reader	<i>potential readers, modern readers, beta readers</i>	<i>to hook a reader, to grab a reader, to satisfy a reader, to help a reader</i>
Plot	<i>main plot, contrived plot</i>	<i>to turn the plot, to drive the plot, to complicate the plot, to advance the plot</i>

Way	<i>best way, certain way, only way, different ways, correct way</i>	<i>to feel a certain way, to adopt a particular way, to work this way</i>
Scene	<i>opening scene, emotional scene, redemption scene, exciting scene, action scene, love scene</i>	<i>to write a scene, to choose a scene, to execute a scene, to rewrite a scene</i>
Arc	<i>character arc, archetypal arc, story arc, hero arc</i>	<i>to explore an arc, to create an arc, to write an arc, to follow an arc</i>
World	<i>normal world, new world, fantasy world, adventure world, external world, story world</i>	<i>to see the world, to alter the world, to depict the world, to improve the world</i>
Time	<i>long time, real time, medieval times, more time</i>	<i>to spend time, to take the time to, to dedicate the time, to give time</i>

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

Results show that the most frequent nouns such as *character* and *story* display more collocations than the noun *time*, which does not occur as frequently. It is also interesting to note that even though most of the collocations refer to the fictitious world inside the book, there are some that depict the outside world (e.g. *external world, normal world*). Results also show that some of these collocations are used as writing tips (e.g. *to create an arc, to write a scene, to begin a story, etc.*)

## 6.2. Blogs 2

Blogs 2 consists of a total of 110, 642 words. This subcorpus includes the following topics: different strength trainings, foods to avoid before workouts, starting a sport, dehydration, coping mechanisms for stress, sleep inertia and allergies, among others. Table 3 below shows the most frequent collocations from the analyzed grammatical structures.

Node	Collocations	
	modifiers of X	verbs with X as object
Body	<i>upper body, entire body, human body, lower body, whole body</i>	<i>to move the body, to force the body to, to help the body, to give the body, to affect the body</i>
Muscle	<i>core muscles, strong muscles, glute muscles, calf muscles</i>	<i>to build muscle, to gain muscle, to strengthen the muscles, to target a muscle, to rebuild the muscle, to grow muscles</i>
Weight	<i>heavy weights, body weight, light weight, scale weight, moderate weight</i>	<i>to lose weight, lift weights, to gain weight, to add weight, to keep the weight, to maintain a weight</i>
Time	<i>more time, long time, enough time, difficult times</i>	<i>to spend time, to take the time to, to save time, to create more time, to need more time</i>
Exercise	<i>best exercise, moderate-intensity exercise, aerobic exercise, compound exercise, breathing exercise, bodyweight exercise, cardiovascular exercise</i>	<i>to perform an exercise, to do an exercise, to follow an exercise, to add an exercise, to make an exercise</i>
People	<i>most people, active people, many people, different people</i>	<i>to help people, to interrupt people, to describe people, to know people</i>

Day	<i>rest day, single day, busy day, long day, cardio day</i>	<i>to start the day, to tackle the day, to fill the day</i>
Food	<i>healthy food, fast food, certain food, spicy food, nutrient-dense food, fried food, plant-based food</i>	<i>to eat food, to process food, to avoid food, to choose food, to digest food, to include food, to consume food</i>
Calorie	<i>total calories, maintenance calories, many calories, enough calories</i>	<i>to burn calories, to eat calories, to consume calories, to contain calories, to count calories, to cut calories, to add calories</i>
Training	<i>strength training, resistance training, interval training, barbell training, bodyweight training, Pilates training</i>	<i>to start training, to guide the training, to advance the training, to demand a training, to handle a training</i>

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

Results show that most collocations refer to the body and its muscles as well as to food and the calories which subsequently impact weight. It is interesting to note that collocations in Table 3 are to a certain extent interrelated because most of these concepts are connected while discussing one’s physical well-being. Some of those interrelated collocations are:

- 1) Muscle and Exercise- *Strength training* is essential for *building muscles*.
- 2) Weight and Exercise- *Cardiovascular exercises* are effective if you want *to lose weight*.
- 3) Calories and Weight- Consuming way too *many calories* can result in *gaining weight*.

### 6.3. Blogs 3

Blogs 3 consists of 112 215 words and is concerned more with personal blogs, in which bloggers describe their travels and finance solutions as well as how they developed certain habits or how they deal with certain situations. Therefore, the texts collected for this subcorpus are more subjective than the ones in the other 2 subcorpora, as evidenced by the collocations displayed in Table 4 below.

Node	Collocations	
	modifiers of X	verbs with X as object
Time	<i>screen time, more time, enough time, hard time, long time</i>	<i>to spend time, to take the time, to have time, to waste time</i>
Day	<i>normal day, recovery days, work days, long days</i>	<i>to start the day, to write about your day</i>
People	<i>most people, black people, toxic people</i>	<i>to let people know, to reward people, to notice people, to help people</i>
Thing	<i>important thing, same thing, many things, small things, great things, strange things</i>	<i>to do things, to try things, to get things done, to crush things, to choose a thing, to make things</i>
Way	<i>great way, easy way, best way, hard way, different way, simple way</i>	<i>to find a way, to feel some way, to learn the hard way</i>
Year	<i>last year, gap year, teenage years, past year, childbearing years</i>	<i>to spend a year, to end the year, to reach a year</i>
Life	<i>full life, daily life, real life, glam life, chaotic life, balanced life</i>	<i>to live life, to control life, to slow down life, to pause a life, to impact a life</i>

Kid	<i>thrifty kid, underage kid, young kid</i>	<i>to teach a kid, to empower a kid, to involve a kid, to provide a kid with, to help a kid, to want a kid, to get a kid</i>
Habit	<i>monthly habits, common habits, daily habits, important habits, walking habits</i>	<i>to track a habit, to build a habit, to improve a habit, to break a habit</i>
Money	<i>enough money, more money, extra money, unlimited money, lunch money</i>	<i>to save money, to earn money, to make money, to steal money</i>

**Table 4:** Collocations of 10 most frequent nouns in Blogs 3

Results show that collocations from the table above all refer to various aspects of daily life, relationships, routines, and experiences. The following examples show how these collocations are interrelated in our everyday language:

- 1) Time and Day- You have *enough time to write about your day*.
- 2) Year and Life- *Last year, I had to slow down my life*.
- 3) Habit and Money- You can *save money* if you have good *monthly* spending *habits*.

## 7. ANALYSIS AND DISCUSSION

In order to analyze the results, we compared collocations from the three tables to see whether they have some common denominator and whether they can reveal something about the topic the blogs focus on. 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), and *main plot* (193.87), among others. These collocations reflect the importance of creating a well-developed character arc and knowing how to effectively tell a story, as well as the significance of both the main character and the main plot in a book.

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. Although this was somewhat expected, most collocations in Blogs 2 refer to body parts. Subsequently, the emphasis is on strength training and building muscles, which suggests that nowadays this is one of the most popular ways of exercise. Similarly, collocations referring to cardio training/exercise are also quite frequent, suggesting that these exercises are the second most popular type of workouts.

In Blogs 3, on the other hand, the most frequent collocations are *most people* (159.54), *to spend time* (98.76), *to save money* (98.76), *best way* (98.76), *next day* (129.15), etc. There is an emphasis on spending time, whether it be in learning new hobbies or skills, or simply spending time with family and friends. In personal blogs with various themes we can see the importance of the time dimension (e.g. *to spend time*, *work days*, *last year*, *teenage years*, *to end the year*, etc.). As there were blog posts about finance and advice on finance as well, we can see here that *to save money* was also one of the most frequent collocations.

Collocations in our subcorpora can say a lot about the nouns that we used as our nodes. For example, collocates of the noun *reader* (e.g. *potential readers*, *modern readers*, *to hook a reader*, *to satisfy a reader*, *to help a reader*, etc.) emphasize the importance of attracting and satisfying readers. This also clearly shows that bloggers want readers to focus on specific themes while reading a blog post, which can especially be seen through collocations in Blogs 1 and Blogs 2. Since Blogs 3 focuses on personal experiences, the



collocations extracted from this subcorpus can also be found in Blogs 1 and Blogs 2 (e.g. *enough/hard time, normal/long day, best/hard way* and *most people*).

Results also suggest that there is some contrast between the collocations. For instance, if we look at the noun *world* and its collocations *normal world, new world, fantasy world, adventure world, external world, story world, to see the world, to alter the world, to depict the world, to improve the world*, we can see that there is a distinction between the ‘real or external world’ and the ‘fantasy or made-up world’.

Many of these analyzed collocations indeed represent popular keywords that online guides for bloggers use to estimate potential popularity or “likes” if you enter one of these collocations as a search word. One of the guides is the Keyword Difficulty Checker<sup>5</sup> which was tested on collocations from Blogs 2. Thus, for instance, the search word *bodyweight exercise* suggests that it would be very hard to rank in top 10 search results among bloggers with this keyword (on a scale of 1 to 100, this keyword had a difficulty estimated at 74). There are many bloggers, who advise new bloggers to use two- or three-word combinations to grab the reader’s attention, but it is also worth mentioning that blogging is quite competitive, therefore, it is difficult to rank high even with ‘popular’ keywords.

Collocations listed in Tables 2, 3 and 4 demonstrate the potential for readers to learn from blog posts. If we look at the collocations used, we can tell that blog posts are generally easy to understand, but they can also be used as a learning resource for language learners. Blogs 3, for instance, includes collocation *to make money*, which some learners can interpret wrong (e.g. in Croatian the direct translation would be *napraviti novac*, which doesn’t refer to earning money, but literally to the process of creating (paper) money).

Finally, nouns depicting time are something that all three subcorpora have in common (e.g. *time, year, and day* occur as nodes in all three subcorpora) and they very often include metaphorical collocations (e.g. *to spend a year, to waste time* or *to tackle the day*).

---

<sup>5</sup> Ahrefs. (n.d). Keyword Difficulty Checker. Retrieved September 12, 2023, from <https://ahrefs.com/keyword-difficulty>

## 8. CONCLUSION

The aim of this thesis was to study collocations that occur in blogs and examine what they can tell us about the nature of blogs as well as point to the differences between subject fields blogs focus on. While each of the analysed subcorpora focuses on one topic, there are collocations involving the time dimension that occur in all three. The typicality of collocations occurring in each corpus can also be confirmed by consulting the online guides for bloggers, since the analysed collocations very often occur as keywords in those tools.

One conclusion that can be drawn from this thesis is that collocations are inevitable in language use and by using them we can communicate more naturally and fluently. This thesis represents a modest attempt of the corpus-based analysis of collocations in blogs and an insight into the real-life use of phraseology.

Future research, however, might include a bigger corpus focusing on one subject field only, which could provide more reliable results regarding both the types of collocations and the genre.

This research can be beneficial for both bloggers and students. Bloggers can use the findings of this study and recognize the importance of collocations in their writing. By selecting appropriate and attention-grabbing collocations, bloggers can effectively capture the interest of their readers and enhance the overall quality of their content.

For students, this research offers educational value because it highlights the significance of collocations in language learning and use. By reading interesting blog posts and personal experiences, students can improve their English language proficiency. They can observe how collocations are used in real-life and gain a deeper understanding of how words combine and interact to convey specific meanings and ideas.

## BIBLIOGRAPHY

1. Ahrefs. (n.d). Keyword Difficulty Checker. Retrieved September 12, 2023, from <https://ahrefs.com/keyword-difficulty>
2. Barnbrook, G., Mason O., Krishnamurthy R. (2013). *Collocation Applications and Implications*. London, UK: Palgrave Macmillan.
3. Biel, L. (2009). *Corpus-Based Studies of Legal Language for Translation Purposes. Methodological and practical potential*. In *Reconceptualizing LSP. Online proceedings of the XVII European LSP symposium*, 1-15.
4. Burger, H. (1998). *Phraseologie. Eine Einführung am Beispiel des Deutschen*. Berlin: ErichSchmid
5. Cheng, Winnie. 2012. *Exploring corpus linguistics: language in action*. London; New York, NY: Routledge.
6. Cho, J.D. (2012). *Blog, Inc : blogging for passion, profit, and to create community*. San Francisco, CA: Chronicle Books.
7. Cowie, A. P. (1998). *Phraseology: Theory, Analysis, and Applications*. Oxford: Clarendon Press.
8. Cowie, A. P. (1988). Stable and creative aspects of vocabulary use. In R. Carter, & M. J., McCarthy (Eds.), *Vocabulary and Language Teaching* (1<sup>st</sup> ed., pp. 126-137). London: Longman.
9. Erman, B., & Warren, B. (2000). *The idiom principle and the open choice principle*. Berlin, Germany: De Gruyter.
10. Gläser, R. (1994/95). Relations between Phraseology and Terminology with Special Reference to English. *ALFA* 7(8). 41-60.
11. Granger, S., Paquot, M., (2008). Disentangling the phraseological web. In S., Granger & F., Meunier (Eds): *Phraseology: An Interdisciplinary Perspective* (pp. 27-50). Amsterdam, Netherlands: John Benjamins.
12. Harper, D. (n.d.). *Etymology of collocation*. Online Etymology Dictionary. <https://www.etymonline.com/word/collocation>
13. Johansson, S. (1995). Mens sana in corpore sano: On the Role of Corpora in Linguistic Research. *The European English Messenger*, 4(2), 19-25.

14. Lutzky, U., & Kehoe, A. (2016). Your blog is (the) shit. *International Journal of Corpus Linguistics IJCL*, 21(2), 165-191. DOI: 10.1075/ijcl.21.2.02lut
15. McCarthy, M. (2004). *Touchstone: From Corpus to Course Book*. Cambridge: Cambridge University Press
16. McEnery, T., Wilson, A. (2001). *Corpus Linguistics*. Edinburgh University Press.
17. McEnery, T., Xiao, R., Tono, Y. (2006). *Corpus-Based Language Studies. An Advanced Resource Book*. London/New York: Routledge
18. Meyer, Charles F. 2002. *English corpus linguistics: an introduction* (Studies in English language). Cambridge, UK ; New York: Cambridge University Press.
19. Mel'čuk, I. (1998). Collocations and Lexical Functions. In A. P., Cowie (Ed.), *Phraseology: Theory, Analysis, and Applications* (1<sup>st</sup> ed., pp. 23-55). Oxford/New York: Oxford University Press.
20. Moon, R. (1998). *Fixed expressions and idioms in English: A corpus-based approach*. Oxford: Clarendon Press.
21. Moon, R. (1998). Frequencies and Forms of Phrasal Lexemes in English. In A. P., Cowie (Ed.), *Phraseology: Theory, Analysis, and Applications* (1<sup>st</sup> ed., pp. 79-101). Oxford/New York: Oxford University Press.
22. Omazić, M. (2015). *Phraseology through the looking glass*. Osijek, Croatia: Josip Juraj Strossmayer University of Osijek, Faculty of Humanities and Social Sciences.
23. Online OXFORD Collocation Dictionary (n.d.), Retrieved August 20, 2023, from <https://www.freecollocation.com/>
24. Patekar, J., Košuta, N. (2022). Croatian EFL learners' collocational competence: Congruent and incongruent collocations. *Training, language and culture*, 6(3), 9-31. <https://doi.org/10.22363/2521-442X-2022-6-3-9-31>
25. Sinclair, J. (1991). *Corpus, Concordance and Collocation*. UK: Oxford University Press.
26. Sketch Engine. (n.d.). <https://www.sketchengine.eu/>
27. Stefanowitsch, A. (2020). *Corpus linguistics, A guide to the methodology*. Berlin, Germany: Language Science Press.
28. WDD STAFF. (2011, March 14). *A Brief History of Blogging*. Webdesigner Depot. <https://www.webdesignerdepot.com/2011/03/a-brief-history-of-blogging/>

29. *What is a blog?: Content Marketing Glossary* (2022, November 15). Textbroker.  
<https://www.textbroker.com/blogging>
30. *Word sketch - collocations and word combinations* (2023, May 09). Sketch Engine.  
<https://www.sketchengine.eu/guide/word-sketch-collocations-and-word-combinations/>
31. Wright, J. (2005). *Blog Marketing: The Revolutionary New Way to Increase Sales, Build Your Brand, and Get Exceptional Results*. New York: McGraw Hill.

## APPENDIX

List of blogs used for the design of the corpus:

1. Book nerdection <https://booknerdection.com/>
2. Born fitness <https://www.bornfitness.com/>
3. Breaking muscle <https://breakingmuscle.com/>
4. Brewing writer <https://brewingwriter.com/>
5. Chris Guillebeau <https://chrisguillebeau.com/>
6. Cookie and Kate <https://cookieandkate.com/>
7. David Lebovitz <https://www.davidlebovitz.com/>
8. Design for mankind <https://designformankind.com/>
9. Elise Blaha <https://eliseblaha.typepad.com/>
10. Everywhereist <https://www.everywhereist.com/>
11. Fit bottomed girls <https://fitbottomedgirls.com/>
12. Fitnessista <https://fitnessista.com/>
13. Gala darling <https://galadarling.com/>
14. Girls gone strong <https://www.girlsgonestrong.com/>
15. Global us home <https://www.fitbit.com/global/us/home>
16. Goins writer <https://goinswriter.com/>
17. Healthline <https://www.healthline.com/>
18. Helping writers become authors <https://www.helpingwritersbecomeauthors.com/>
19. <https://achs.edu/>
20. James clear <https://jamesclear.com/>
21. Jane Friedman <https://janefriedman.com/>
22. Jerry Jenkins <https://jerryjenkins.com/>
23. Kids lit book café <https://kidslitbookcafe.com/>
24. Live strong <https://www.livestrong.com/>
25. Love that max <https://www.lovethatmax.com/>
26. Marks daily apple <https://www.marksdailyapple.com/>
27. Miss thrifty <https://www.miss-thrifty.co.uk/>
28. Motherhood and mayhem <https://motherhoodandmayhem.online/>

29. Mr money mustache <https://www.mrmoneymustache.com/>
30. My fitness pal <https://blog.myfitnesspal.com/>
31. Nerd fitness <https://www.nerdfitness.com/>
32. Noob gains <https://www.noobgains.com/>
33. Okdani <https://www.okdani.com/>
34. Runners world <https://www.runnersworld.com/>
35. Techs avvy mama <https://techsavvymama.com/>
36. The blog a broad <https://theblogabroad.com/>
37. The creative penn <https://www.thecreativepenn.com/>
38. The write life <https://thewritelife.com/>
39. Thoughts on papyrus <https://thoughtsonpapyrus.com/>
40. Wowlit <https://wowlit.org/>
41. Write to done <https://writetodone.com/>
42. Yoga journal <https://www.yogajournal.com/>
43. Zen habits <https://zenhabits.net/>

## **ABSTRACT**

Phraseology is important for effective communication as well as for understanding the cultural aspects of languages. There are a lot of languages for specific purposes (LSPs) that possess their own terminologies and phraseologies (e.g. the language of medicine, legal language, etc.). This research, however, does not focus on such specialized languages. Instead, it examines collocations in blogs to provide insight into the phraseology of an informal piece of writing. The aim of this research was to study collocations that occur in blogs and examine what they can tell us about the nature of blogs as well as to point out the differences between the subject fields blogs focus on. This thesis adopts a corpus-based approach by focusing on collocations in an English monolingual corpus consisting of three subcorpora. Results suggest that the most frequent collocations were topic-related ones and that they do not offer a general picture of the collocability shared between blogs. They do, however, show how bloggers use collocations to address their target audience and grab their attention. This research can thus be beneficial for both bloggers, who can improve their writing and students who want to improve their English language proficiency.

**Keywords:** blogs, collocations, corpus-based research, phraseology



## SAŽETAK

Frazeologija je važna za učinkovitu komunikaciju i razumijevanje kulturnih aspekata nekog jezika. Danas postoje mnogi specijalizirani jezici (LSP), koji posjeduju vlastitu terminologiju i frazeologiju (npr. jezik medicine, pravni jezik). Ovo se istraživanje ne usredotočuje na takve specijalizirane jezike. Umjesto toga, bavi se istraživanjem kolokacija u blogovima kako bi pružilo uvid u frazeologiju neformalne prirode. Stoga je cilj ovog istraživanja bio proučiti kolokacije koje se pojavljuju u blogovima i ispitati što nam one mogu reći o samim blogovima te ukazati na razliku između tematskih područja na koje se blogovi usredotočuju. Ovaj rad služi se korpusnim pristupom u istraživanju kolokacija u jednojezičnom engleskom korpusu koji se sastoji od 3 podkorpusa. Rezultati pokazuju da su najčešće kolokacije bile tematske i da one ne nude opću sliku o kolokacijama u blogovima općenito. Međutim, upućuju na to da blogeri koriste kolokacije kako bi privukli pažnju ciljane publike. Stoga ovo istraživanje može biti korisno i za blogere koji mogu poboljšati svoje pisanje i za studente koji žele poboljšati svoje znanje engleskog jezika.

***Ključne riječi:*** blogovi, frazeologija, kolokacije, korpusno istraživanje