FAKULTA
PŘÍRODOVĚDNĚ-HUMANITNÍ
A PEDAGOGICKÁ TUL

Bakalářská práce

A Corpus-Based Comparison of the Nouns "Problem" and "Issue"

Studijní program:	B0114A300068 Anglický jazyk se zaměřením na vzdělávání
Studijní obory:	Anglický jazyk se zaměřením na vzdělávání Základy společenských věd se zaměřením na vzdělávání
Autor práce:	**Barbora Hypšová**
Vedoucí práce:	Mgr. Jaromír Haupt, PhD. Katedra anglického jazyka

Liberec 2023

FAKULTA
PŘÍRODOVĚDNĚ-HUMANITNÍ
A PEDAGOGICKÁ TUL

Zadání bakalářské práce

A Corpus-Based Comparison of the Nouns "Problem" and "Issue"

Jméno a příjmení:	**Barbora Hypšová**
Osobní číslo:	P20000644
Studijní program:	B0114A300068 Anglický jazyk se zaměřením na vzdělávání
Specializace:	Anglický jazyk se zaměřením na vzdělávání Základy společenských věd se zaměřením na vzdělávání
Zadávající katedra:	Katedra anglického jazyka
Akademický rok:	2021/2022

Zásady pro vypracování:

Cílem práce je porovnat použití anglických substantiv *problem* a *issue* z hlediska významu a použití.
Metody:
1. Studium literatury z oblasti korpusové lingvistiky.
2. Zjištění, jakým způsobem zkoumaná substantiva popisují slovníky a jiné zdroje.
3. Volba vhodných korpusů a korpusová analýza zkoumaných slov s využitím nástroje Sketchengine.
4. Vyhodnocení výsledků a jejich konfrontace se slovníky a dalšími zdroji.

Rozsah grafických prací:
Rozsah pracovní zprávy:
Forma zpracování práce: tištěná/elektronická
Jazyk práce: Angličtina

Seznam odborné literatury:

Hoey, Michael. 2005. *Lexical priming: a new theory of words and language*. London: Routledge.
Jarunwaraphan, Boonrak, and Prima Mallikamas. 2020. 'A Corpus-Based Study of English Synonyms: Chance and Opportunity'. *REFLections* 27 (2): 218–45.
Jirananthiporn, Somjit. 2018. 'Is This Problem Giving You Trouble? A Corpus-Based Examination of the Differences between the Nouns Problem and Trouble'. *Thoughts*, no. 2 (December): 1–25.
Meyer, Charles F. 2002. *English corpus linguistics: an introduction*. Cambridge: Cambridge University Press.
Stewart, Dominic. 2009. *Semantic Prosody: A Critical Evaluation*. 1st edition. New York: Routledge.

Vedoucí práce: Mgr. Jaromír Haupt, PhD.
Katedra anglického jazyka

Datum zadání práce: 24. června 2022
Předpokládaný termín odevzdání: 14. července 2023

L.S.

prof. RNDr. Jan Picek, CSc.
děkan

Mgr. Zénó Vernyik, Ph.D.
vedoucí katedry

V Liberci dne 24. června 2022

Prohlášení

Prohlašuji, že svou bakalářskou práci jsem vypracovala samostatně jako původní dílo s použitím uvedené literatury a na základě konzultací s vedoucím mé bakalářské práce a konzultantem.

Jsem si vědoma toho, že na mou bakalářskou práci se plně vztahuje zákon č. 121/2000 Sb., o právu autorském, zejména § 60 – školní dílo.

Beru na vědomí, že Technická univerzita v Liberci nezasahuje do mých autorských práv užitím mé bakalářské práce pro vnitřní potřebu Technické univerzity v Liberci.

Užiji-li bakalářskou práci nebo poskytnu-li licenci k jejímu využití, jsem si vědoma povinnosti informovat o této skutečnosti Technickou univerzitu v Liberci; v tomto případě má Technická univerzita v Liberci právo ode mne požadovat úhradu nákladů, které vynaložila na vytvoření díla, až do jejich skutečné výše.

Současně čestně prohlašuji, že text elektronické podoby práce vložený do IS/STAG se shoduje s textem tištěné podoby práce.

Beru na vědomí, že má bakalářská práce bude zveřejněna Technickou univerzitou v Liberci v souladu s § 47b zákona č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších předpisů.

Jsem si vědoma následků, které podle zákona o vysokých školách mohou vyplývat z porušení tohoto prohlášení.

24. dubna 2023 Barbora Hypšová

Acknowledgements

I would like to express my sincere gratitude to Mgr. Jaromír Haupt, Ph.D., my thesis supervisor, for his guidance, insightful feedback, and support throughout the process of writing this thesis. His extensive knowledge and expertise in the field of linguistics were invaluable in shaping my research and improving the quality of this work. Additionally, his many suggestive comments and valuable insights helped me to complete this thesis.

Furthermore, I am also grateful to my family, for their positive comments and instant support while writing. Their encouragement and support have been invaluable in keeping me motivated and focused.

Anotace

Tato bakalářská práce si klade za cíl zanalyzovat problematický fenomén synonymie, a to na pozadí korpusového porovnání substantiv *issue* a *problem*. Práce je rozdělena do dvou hlavních částí, z nichž teoretická část představuje úvod do lexikální sémantiky a korpusové lingvistiky. Praktická část následně uvádí použitou metodologii, zdroje dat, analýzu slovníkových a dalších definic dvou zkoumaných substantiv a taktéž korpusovou analýzu provedenou s pomocí nástroje Sketch Engine. Korpusová analýza zkoumá frekvenci použití a kolokační vzorce obou zkoumaných substantiv v korpusu English Web 2020 se záměrem získat informace o jejich použití a významu. Autor si u obou substantiv *issue* a *problem* klade za cíl určit stupeň synonymie a na základě dat získaných korpusovou analýzou a jejím srovnáním s dalšími lexikografickými zdroji identifikovat, jaké rozdíly a podobnosti mezi těmito dvěma slovy existují.

Klíčová slova

lexikální sémantika, synonymie, korpusová lingvistika, korpus, kolokace, Sketch Engine, issue, problem

Annotation

This Bachelor's thesis aims to investigate the complex phenomenon of synonymy through a corpus-based comparison of the nouns *issue* and *problem*. The thesis is structured into two main parts: the theoretical part, which provides an introduction to lexical semantics and corpus linguistics, and the practical part, which outlines the methodology and sources of data, analyses the definitions of the target nouns provided by monolingual dictionaries and other internet sources, and conducts a corpus analysis using the Sketch Engine search software. The corpus analysis examines the frequency of use and collocational patterns of the two target nouns in the English Web 2020 corpus, providing valuable insights into their usage and meaning. The author aims to determine the degree to which the nouns *issue* and *problem* are synonymous, and identify the differences and overlaps between them, based on the findings from the corpus analysis and comparison with various lexicographical sources.

Keywords

lexical semantics, synonymy, corpus linguistics, corpus, collocation, Sketch Engine, issue, problem

List of Contents

List of Tables

List of Abbreviations

AAD	Ask Any Difference
CaD	Cambridge Dictionary
CoD	Collins Dictionary
D	Dictionary
DB	Difference Between
LE	Learning English
MWD	Merriam-Webster Dictionary
P	Peedia

Introduction

It is said that it is the language that distinguishes human beings from other living creations. As Trask (1999) states: "Without language, we could hardly have created the human world we know. Our development of everything from music to warfare could never have come about in the absence of language" (1). However, even the language lives with us and constantly changes within its use. When there is the need to create a new word for something currently discussed in society, the language users develop a new term. Moreover, as Čermák (2010) points out, inventing a new term also leads to developing its synonyms to prevent the repetition of this newly invented word (267). Even though synonyms are not essential for everyday communication, they are commonly used (263).

Despite being a common concept, synonymy is a complex phenomenon that goes beyond its simple definition, "the sameness of meaning" (Cruse 2001, 8761). As words can have multiple meanings, and their usage may differ depending on the context and the speaker, it might be challenging to discern the nuances between synonyms. However, in the past decades, there has been a significant development in the field of corpus linguistics, which enabled researchers to analyse the use of naturally occurring language and observe patterns which could be otherwise overlooked (Hunston 2022, 1). Corpus linguistics tools enable analysis and observation of the use of words in a language, providing insights into their utilisation and contextual nuances.

The thesis aims to conduct a corpus-based comparison of the use and the meanings of the nouns *issue* and *problem*, which are generally considered synonymous. The findings will be confronted with various lexicographical sources to decide whether these two words are synonymous and, if so, to what extent they overlap and differ. The definitions of *issue* and

problem will be analysed and mutually compared through online monolingual dictionaries and other web sources to distinguish their meanings. In the main part of the thesis, a study of a language in use in corpus English Web 2020 will be conducted with the help of the Sketch Engine tool. The data achieved by using tools of word sketch and concordance will provide a more detailed perspective on using the two target words concerning their collocational behaviour.

The bachelor thesis is divided into two main parts: a theoretical one and a practical one. The former provides a theoretical background for the practical work with the nouns *issue* and *problem*. The theoretical part's first chapter offers an introduction to lexical semantics, the types of meaning and lexical relations. The second chapter provides a theoretical background of corpus linguistics. The practical part starts with the outline of the methodology and the sources of data. The next chapter studies the definitions provided by monolingual dictionaries and other internet sources. Chapter 5 uses corpus analysis to answer research questions, examining the frequency of use and collocational patterns of the two target nouns. The last chapter presents the results of the corpus analysis and summarises the findings acquired about the nouns *issue* and *problem*.

1 Lexical Semantics

Linguistics is an umbrella term for many other studies, such as phonology, syntax and semantics, which Cruse (2001) lists as the most important ones (8758). They all study a language from different perspectives: phonology studies the sounds and sound systems of language, syntax analyses the grammar, and semantics studies the meanings of words. For the purpose of the object of the present thesis, the last-mentioned linguistics study is the area discussed in more detail in this part. As mentioned, lexical semantics is a subfield of linguistics concerned with studying the meaning of words and how words combine to form the meaning of sentences. Murphy (2010) notes that the term *lexical* refers to the lexicon, which is "a collection of meaningful linguistic expressions from which more complex linguistic expressions are built" (3). Such meaningful lexical expressions are often words, but not always. For this reason, as Murphy explains, lexical semantics is actually the study of lexeme meaning since "not all words are lexemes and not all lexemes are words" (6).

A word is a unit of language that has a meaning in context and can have multiple forms, depending on its use in a sentence (such as *run* and *running*). A lexeme, on the other hand, is the underlying semantic representation of a word that can be inflected or changed in form to create different word-forms (such as the lexeme *run* that creates the word forms *runs* and *running*). A lexeme represents the core meaning of a word, which can be shared by its different forms. Lyons (1996) distinguishes two sets of expressions used in a language: lexemes, the vocabulary units of a language, and lexically composite expressions, which are constructed out of lexemes "by means of the grammatical (i.e., syntactic and morphological) rules of the language" (51). However, for the purpose of this thesis, the terms "word" and "lexeme" will be used interchangeably because "most word-expressions, in all languages that have words, are lexically simple" (51).

Lexical semantics is a crucial component of language as it provides a systematic way to describe and understand the meaning of words, their relationships to each other, and how they contribute to the overall meaning of a sentence. The object of its study makes it clear that it is related to several scientific fields, including once mentioned linguistics, furthermore to psychology, philosophy and artificial intelligence (Osherson 1995, 311). It is a field whose interdisciplinary nature highlights the importance of studying the meaning of words, how they contribute to the overall meaning of sentences, and the broad impact that this knowledge can have on our understanding of the world.

1.1 Meaning

In lexical semantics, meaning is understood as the concept and information that speakers of a language associate with a word and how these concepts are used in context to convey intended meanings. As Lyons (1996) points out by providing several examples of the verb *mean* in four different sentences, the meaning of spoken or written language utterances depends on the context in which they are used (4). Murphy (2010) states that "a single thing or word can "be meaningful" in many different ways" (29). The language expressions perceived as a part of an utterance can be understood in a much larger variety of ways than first occur to us when we hear or see them out of context (Lyons 1996, 4). Nevertheless, Murphy (2010) notes a crucial point at this stage: "Where we are provided with less contextual information, we rely on interference to fill in the gaps in our mental picture of the situation" (31). For this purpose, which is to avoid misunderstandings, it is essential to distinguish and describe different types of meanings.

1.1.1 Types of Meaning

In lexical semantics, there are several different types of meaning that words can have. According to Cruse (2001), there are two types of meaning: the "inherent meaning" and the "meaning-in-context" (8758). The former, also known as lexical or dictionary meaning, refers to the intrinsic or essential nature of something. Cruse states it as the primary concern of lexical semantics and provides us with an example of the inherent meaning of a *boy*, which he describes as *an immature human being* (8758). In contrast, the latter is the meaning of a word in a specific context and is the main concern of pragmatics. Meaning in context takes into account the way words are used in a specific context, like in the example that Cruse gives: "The child runs to his mother." From this, it can be concluded that the *child* is a boy as "his" gives us the information (8758).

Besides this distinction, Murphy (2010) lists and describes other dimensions of meaning, such as denotative (or conceptual or cognitive), connotative, and social meaning (32). Denotative meaning is the literal one that can be found in dictionaries and defines what the word refers to (32). For example, the denotative meaning of the word *dog* is a domesticated animal that is often kept as a pet. On the other hand, the connotative meaning, or simply connotation, is a semantic association that a word has (33). This is the emotional or cultural meaning that a word can carry beyond its primary definition. Connotative meaning is subjective and varies between individuals and cultures. For example, the word *home* can have a positive connotation for some people, while others may associate it with negative emotions. The last term, social meaning, deals with the social aspect that can be drawn from the expression one uses (33). An example of such meaning Murphy (2010) shows on the use of the word *howdy* (meaning hello), from which one can make an assumption about where the speaker is from or

a piece of information about the speaker's social status or attitude towards the person he or she speaks with. (34)

Leech (1985) proposes even more types of meaning to provide a comprehensive framework for understanding the complex nature of linguistic meaning. According to Leech, there are seven types of meaning: conceptual meaning, connotative meaning, social meaning, affective meaning, reflected meaning, collocative meaning and thematic meaning (9).

As evident, some of the terms overlap with the types of meaning that Murphy (2010) describes, namely the denotative, connotative and social meanings. Similarly to Murphy, Leech also assigns priority to the conceptual meaning, whose study can provide us with a "configuration of abstract symbols which is its 'semantic representation', and which shows exactly what we need to know if we are to distinguish that meaning from all other possible sentence meanings in the language" (11). In other words, this type of meaning refers to the cognitive or mental representation of an object, action, or idea and encompasses a word's basic or literal definition, e.g., the word *woman* is defined as human + male + adult (10).

On the other hand, the connotative meaning is "communicative value an expression has by virtue of what it refers to, over and above its purely conceptual content" (12). It encompasses the additional meanings and implications that a word can carry as a result of its cultural or historical context, e.g., *woman* can be associated with the word *compassionate* (12). Social meaning is the last of the three meanings that both Murphy (2010) and Leech (1985) distinguish. Social meaning is the information that language reveals about the social context in which it is used. The dimensions of social meaning may vary depending on dialect, time, province, status, modality and singularity (Murphy 2010, 14).

The fourth type of meaning that Leech describes is affective meaning, which is often conveyed through the conceptual or connotative content of the words used. In this type of

meaning, factors such as the tone of voice or politeness play an important role (16). The reflected meaning is the "meaning that arises in cases of multiple conceptual meaning, when a sense of a word forms part of our response to another sense" (16). Except these, two more types of meaning are noted. Firstly, it is the collocative meaning which "consists of the associations a word acquires on account of the meanings of words which tends to occur in its environment" (17). Leech's thematic meaning refers to choosing between different grammatical constructions. For example, a speaker might use passive or active voice to emphasise different aspects (19).

1.2 Lexical Relations

Language is a system in which its particular items are connected. And with meanings, it is the same. Even the meanings are interconnected (Lyons 1996, 5), and the relations between them Cruse (2001) describes as the subject of sense or lexical relations, which are divided into two parts – paradigmatic and syntagmatic (8761). Both paradigmatic and syntagmatic relationships are essential in determining the meaning of words and sentences in a language. Hence, to determine the meanings of two target words, *issue* and *problem*, the lexical relations have to be explained.

Murphy (2010) states a similar distinction as Cruse (2001) and describes syntagmatic relations as relations between words associated with one another, like *dogs* and *bark* (108). He points out that these words usually belong to different parts of speech, as evident in the example with *dogs* (noun) and *bark* (verb). In brief, syntagmatic relationships involve the combination of words to create meaningful sentences. The other type of paradigmatic relations Murphy (2010) describes as the ones that form a paradigm, a set of examples showing a pattern, and refers to them as paradigmatic relations (109). He provides an example with a colour paradigm

in which the words *red* and *white* belong and points out that the words in this type of relationship usually belong to the same word class and are also substitutable for each other (109).

On the other hand, Cruse (1986) says that paradigmatic relations are the ones which "represent systems of choices a speaker faces when encoding his message" (86). Cruse's definition highlights the importance of paradigmatic relations in the structure and organization of language, as they allow speakers to use different words to express similar meanings and provide a basis for the creation of new words and expressions. Moreover, Cruse (1986) states that the syntagmatic aspects serve discourse cohesion, adding necessary informational redundancy to the message and, at the same time controlling the semantic contribution of individual utterance elements through disambiguation or by signalling alternative interpretation strategies (86). Cruse (2001) provides various examples and explanations of paradigmatic relations that have figured prominently in discussions. Namely, these are:

- synonymy = the sameness of meaning (brave and heroic)
- hyponymy = the relation of inclusion between a hyponym and a superordinate (*truck* and *vehicle*)
- meronymy = the relation of inclusion between an item and its parts (*finger* and *hand*)
- incompatibility = the relation between sister hyponyms of the same superordinate (*dog, cat* and *animal*)
- oppositeness (antonyms) = incompatibility between two-member sets (8761).

From the list above, synonymy is the main concern to complete the aim of the thesis, which is to distinguish the use and meanings of the two synonymous nouns, *issue* and *problem*. However, before examining synonyms, discussing the diachronic perspective of meaning is necessary. As mentioned in the introduction, language is flexible and changes in ways that influence the meanings of words. Through these changes, language aims to maintain its

effectiveness in everyday communication. An example might be a situation in which words get new senses, which is nowadays a common phenomenon in the area of personal computers, where terms such as *mouse* are used (Cruse 2001, 8762). It is an example of polysemy which Cruse (2001) describes as a situation when a word has more than one meaning, and the meaning is determined by its context (8760). However, the diachronic changes are gradual and do not affect only the polysemy of words but also the synonymy. As Čermák (2010) states, with reference to Breál, from a diachronic perspective, synonyms tend to shift their meaning and gradually separate each other (268).

1.2.1 Synonymy

As apparent from the Lexical Relations part, synonymy is one of a couple of semantic relations between the senses of words (Murphy 2010, 108). In the English language, it is quite a frequent phenomenon, which is, to a certain extent, a result of the historical development of the English language, whose vocabulary range has been influenced by many other languages such as French, Latin or Greek (Palmer 1976, 59). The term "synonym" comes from the Greek roots *syn*, which means "alike", and *onym*, which means "name" (Murphy 2010, 110). Its definition that synonyms are words that mean the same (110) reflects the Greek terms. However, perfect synonymy, a situation when words have the same meaning and can be substituted without a change, exists only in rare cases. As Crystal (2002) points out, "there may be no lexemes which have exactly the same meaning" (164).

Murphy (2010) notes the substitutability test to determine whether two words are synonymous. If the two words can be used in the same sentence without changing their meaning, we can refer to them as synonyms, as he shows in the two sentences with the words *person* and *human* (110):

A *person* is standing beside me.

A *human* is standing beside me.

However, when comparing the words *man* and *person*, it can be found that these are not synonyms because one can only substitute the other by changing the meaning. Whereas *person* can represent any human being, e.g., woman, girl, boy, *man* does not, and that is why the two words are not synonyms (110). For this reason, Crystal (2002) describes a list of nuances which separate each lexeme or a context in which one lexeme may appear but the other cannot (164). The differences are as follows: a dialect one that may exist in American and British English (e.g., *autumn* and *fall*), a stylistic one, referring to the (in)formality of words (e.g., *insane* and *loony*), a collocational one which reflects the well-formed sequence of words going together (e.g., *rancid* and *rotten*) and an emotional or connotational one dealing with politeness (e.g., *youth* and *youngster*). Crystal (2002) states that even though perfect synonymy is rare and the differences are present, for most practical purposes, they can be ignored (164).

Regarding the differences mentioned above, Cruse (2000) distinguishes three degrees of synonymy, which are absolute synonymy, near-synonymy and propositional synonymy (156).

As Čermák (2010) points out, acquiring knowledge of a language that developed throughout history allowed us to revise the linguistic attitudes toward synonyms, especially the idea of absolute synonymy (264). The increasing popularity and development of technology during the 20th century enabled linguists the detailed language study through corpora. Simultaneously, it was discovered that no words might have precisely the same meaning (Crystal 2002, 164); therefore, referring to words as "perfect synonyms" is not suitable.

A. Absolute synonymy

Absolute synonymy refers to a situation where two words have exactly the same meaning and can be used interchangeably in all contexts without changing the meaning of a sentence. For absolute synonyms, it is necessary to share the same collocational range, which refers to the set of words that typically occur together in a language (Lyons 1996, 62). Absolute synonyms are extremely rare (Lyons 1996, 61), as words often have subtle differences in connotation or usage. At this stage, Lyons (1996) lists three conditions that the two supposedly absolute synonyms must satisfy:

I. all their meanings are identical;

II. they are synonymous in all contexts;

III. they are semantically equivalent (i.e., their meaning or meanings are identical) on all dimensions of meaning, descriptive and non-descriptive) (61).

The differences are apparent in the examples of *big* and *large*, often regarded as synonymous, which Lyons (1996, 61) and Cruse (2000, 157) provide. Lyons (1996) states that the two terms are not absolute synonyms because "*big* has at least one meaning which it does not share with *large*"62). Moreover, as evident from the examples of two seemingly absolute synonyms *big* and *large* that Cruse (2000) gives, one lexical term tends to be more normal ("+"), whereas the other is less normal ("−") (157).

He's a *big* baby, isn't he? (+)

He's a *large* baby, isn't he? (−)

In the context of absolute synonyms, Murphy (2010) notes that the two words might not share all of their senses, but if they share at least one sense that means the same as one of the other word's senses, we can talk about sense synonyms (111). However, even though English

has many words whose meaning is really close, it is still rare to find a pair which would be perfectly synonymous even for just one of their senses (111). Nevertheless, some examples that would suit the definition of absolute synonyms can often be found in technical names for things like plants, animals or chemicals (111).

B. Near-synonymy

Compared to absolute synonymy, a situation in which a word's senses overlap is much more common in English. Lyons (1996) explains this category of synonymy as a situation when the two expressions have similar meanings but are not identical (60). Cruse (2000) characterises four aspects in which near-synonyms may differ:

I. adjacent position on scale of "degree": *fog:mist*, *laugh:chuckle*…
II. certain adverbial specialization of verbs: *amble:stroll*, *chuckle:giggle*…
III. aspectual distinction: *calm:placid* (state vs disposition)
IV. difference of prototype centre: *brave* (prototypically physical):*courageous* (prototypically involves intellectual and moral factors) (160).

An example of near-synonyms is a pair of *large* and *big*, presented above when disproving that they are absolute synonyms. Even though the two words have similar meanings, they also have slightly different nuances, connotations, or shades of meaning. Whereas both *large* and *big* describe size, *big* may have a connotation of being more substantial or impressive. Saeed (2009) states that synonyms may also differ as a result of distributional constraints, which may be regional, formal and attitudinal (66). He provides examples of words referring to the police, where the regional differences contain terms such as the *guards*, used in Irish English, or the *old Bill*, used in British English. On the other hand, terms such as *police officer* tend to be more formal than others, and the speaker's attitude also plays a role since it adds a negative (*fuzz*) or a neutral connotation (*cop*) (66). Saeed (2009) also points out that taboo words,

expressions considered socially unacceptable or inappropriate, are also a significant source of synonymy (65).

C. Propositional synonymy

Propositional synonymy refers to the phenomenon in which two different sentences have the same meaning regarding their truth conditions. Cruse (2000, 158) defines it as a situation in which two lexical items may be swapped without affecting their truth-conditional features. Cruse's distinction between propositional synonymy and other forms of semantic equivalence is important in linguistic semantics, as it highlights the role of form and logic in determining the truth conditions of sentences. By focusing on the relationship between the predicates in sentences, Cruse provides a new perspective on the nature of meaning and the role of logic in language. An example is the propositional synonymy between the nouns *fiddle* and *violin*, whose use in a sentence depends on the speaker's characteristics (158). While a professional player would prefer to use the term *fiddle*, Crystal notes that someone not involved in violinist culture would use the term *violin*. Nevertheless, both terms have the same truth conditions.

2 Corpus Linguistics

Corpus linguistics is a branch of linguistics that McEnery and Wilson (1996) describe as “the study of language based on examples of real-life language use” (1). Studying real-life language use and analysing large amounts of text data can identify word use patterns in different contexts and potentially help with the problematic phenomenon of synonymy. Corpus linguistics is a relatively new branch of linguistics whose development accelerated in the second half of the 20th century with the use of computers and has grown especially in the 2000s (Lindquist 2009, 1). The word *corpus* has a Latin origin and refers to *body,* from which it may

be concluded that *corpus* is “a body of language or more specifically, a (usually) very large collection of naturally occurring language, stored as computer files” (Baker 2010, 6).

From the language the corpus stores, which can be spoken, written or signed, many can be found to acquire information about frequencies, co-occurrence and meanings (Hunston 2022, 1). Hence, corpus linguistics has been particularly useful in analysing language in many linguistics areas, such as in lexicography, language teaching, translation, stylistics, grammar, gender studies, forensic linguistics or computational linguistics (Tognini-Bonelli 2001, 1). As corpus linguistics has proven to be helpful in analysing language, Hunston (2022) points out that new concepts were developed, such as the theory of Units of Meaning by Sinclair (2004) or lexical priming by Hoey (2005) (4). All of this provides valuable insights into the subtle nuances of meaning between apparently similar words and clarifies their appropriate use in different contexts.

Researchers' work with corpora for many years depended on analysing printed texts (McEnery and Wilson 2001, 31). Lindquist (2009) notes regarding the data analysis that “many of the earliest descriptions were quite biased, based on the compiler's own way of speaking and writing and on their own personal views on what constituted the correct use of the English language” (2). However, it changed when computers appeared and enabled machine-readable analysis, which lowered the probability of errors and accelerated the time spent searching through the corpus (McEnery and Hardie 2012, 2). The first electronic collection of English text is the Brown Corpus, which included 500 samples of 2,000 words from various text genres (Lindquist 2009, 3). Nowadays, as the vast majority of collections of texts are stored on digital media, the term *corpus* is always used to refer to electronic corpora (3). The ability to analyse and calculate huge databases and collections of texts is generally considered to be the greatest advantage of computers over humans (25).

McEnery and Hardie (2012) note that to read whichever corpus rapidly and reliably, one can exploit tools such as concordances to look at a word in context (2). Concordance is an example of a qualitative approach in corpus linguistics, which is often used to investigate how language is used in particular contexts (Hunston 2022, 47). Usually, the KWIC (Key Word in Context) format showing the searched word in the centre of each line is used in the context of concordance (Sinclair 1991, 32). As Hunston (2022) remarks, the qualitative approach usually follows the quantitative approach, with whom most corpus studies begin and which, on the other hand, seeks to obtain a frequency list of words (47).

Nevertheless, there are other examples of quantitative approaches in corpus linguistics, such as studying collocations or keywords. A frequency list is used to identify how often a word form (referred to as type) occurs in a given corpus, while it also provides the number of occurrences of each word form (84). There are two types of frequency: firstly, absolute or raw frequency, which simply counts the number of times a particular word occurs in a corpus without taking into consideration the size of a corpus and secondly, normalised or relative frequency, which enables the comparison of different size corpora as it divides the word frequency by the size of given corpora (85). Another way to compare two corpora is to use keywords tool, which shows "how much more frequent a word is in one corpus than another" (88). Finally, it is the collocation to which a separate paragraph is dedicated.

2.1 Collocation

In corpus linguistics, collocation refers to the tendency of certain words to co-occur with other words in a language. For the purpose of the present thesis, it is necessary to describe collocation because it may serve as one of a list of nuances that show the subtle differences

between the use of the two near-synonymous nouns *issue* and *problem*, as discussed in the "synonymy" subchapter.

There is a great multitude of different definitions of collocation, which may emphasise different aspects, such as the native speakers' or researchers' intuition (McEnery and Hardie 2012, 123). One of the first definitions of collocation was created by H. E. Palmer, claiming that "a collocation is a succession of two or more words that must be learnt as an integral whole and not pieced together from its component parts" (1933, title page, as cited in Lindquist 2009, 71). On the other hand, one of the most popular definitions is the one by J. R. Firth (1957), who noted, "you shall know a word by a company it keeps" (11, as cited in Hunston 2022, 93). In this definition, he emphasised, "how the meaning of individual words is influenced by other words that it frequently occurs together with" (Lindquist 2009, 72). Later in the 1960s, Sinclair and Halliday defined collocation as "the more-frequent-than-average co-occurrence of two lexical items within five words of text" (Krishnamurthy in Sinclair et al. 2004: xiii, as cited in Lindquist 2009, 73).

Based on the definitions above, Lindquist (2009) distinguishes between "window collocations" and "adjacent collocations" (73). The former refers to words that appear together within usually four or five words to the left and to the right "but which do not necessarily stand in a direct grammatical relationship with it" (73). When interpreting the collocation data, one has to take into account the used statistical measure (78). Such a measure can be mutual information (MI) score which shows the collocational strength between the keyword and its co-occurrence (76). Another statistical measure is also logDice, which shows how typical or strong a collocation is (Word sketch - collocations and word combinations).

The latter term, adjacent collocations, refers to words that appear immediately together with no other words in between (Lindquist 2009, 78). Adjacent collocations show frequent

choices of speakers or writers which are influenced by the words they have just uttered or written (78). Both types of collocations can affect the way we analyse and understand language use. While window collocations may be more helpful in analysing larger patterns of language use, adjacent collocations may be more beneficial for identifying specific phrases or expressions, which can help teachers accept more than just one variant of language in use (87).

2.1.1 Terms Used in Corpus Linguistics

However, besides collocation, there are other terms in corpus linguistics to explain. In the Corpus Analysis part, the terms such as node and collocate are used. The former is a word or phrase that is the focus of analysis, while the latter term frequently appears in proximity to the node (Sketch Engine). Together, a node and a collocate form a collocation. Later, when describing a corpus, three key terms, token, type and lemma, are used. Token refers to the total number of words in a corpus, and a type to the number of unique words (Baker, Hardie and McEnery 2006, 162). The last term, lemma, is described as the base or canonical form of a word (104), which may appear in various inflected forms. Moreover, it is essential to distinguish and understand the concepts of colligation, semantic prosody, semantic preference and semantic association. As evident, these terms are really similar, but in fact, each refers to a slightly different phenomenon built upon a collocation analysis (McEnery and Hardie 2012, 132).

Colligation is described by Sinclair (2004) as "the co-occurrence of grammatical phenomena" (142). In other words, colligation refers to the grammatical patterns that words tend to occur in. For example, the verb *make* often collocates with the noun *decision* (e.g., *make a decision*).

Semantic prosody (or discourse prosody) refers to the positive or negative connotations that a word acquires based on its frequent co-occurrence with other words and can only be

discovered by analysis of a concordance (McEnery and Hardie 2012, 136). Sinclair (1991) provides an example of the word *happen*, which is usually associated with unpleasant, bad things, such as accidents (112, as cited in Hoey 2005, 22). In simply terms, semantic prosody "describes the speaker's communicative purpose" (Lindquist 2009, 57).

Semantic preference is similar to semantic prosody as both are "abstractions across multiple, different collocations" (McEnery and Hardie 2012, 137), and hence, sometimes, these terms overlap and lead to confusion. However, as McEnery and Hardie (2012) point out, the difference is that "semantic preference may be in favour of any definable semantic field", whereas "semantic prosody is always either for positive evaluation or for negative evaluation semantic preference" (137). Moreover, semantic prosody is said to be more suggestive compared to semantic preference that does not reveal attitudes (Baker 2010, 132).

Semantic association is described by Hoey (2005) as a phenomenon that "exists when a word or word sequence is associated in the mind of a language user with a semantic set or class, some members of which are also collocates for that user" (24). However, when Hoey (2005) uses the term semantic association, he actually refers to semantic prosody (23).

2.2 Types of Corpora

Even though the term *corpus* has already been described in the introductory paragraph on corpus linguistics, concerning the practical part of the thesis, in which English Web 2020 will be used to analyse the use of the nouns *issue* and *problem*, types of corpora should be briefly described. Lindquist (2009) divides different types of corpora into the following categories: spoken, written, general, specialised, historical, parallel and multilingual corpora and also provides other similar sources of information such as dictionaries, text archives, newspaper CD-ROMs and web, which may serve as a source of linguistics investigation (23).

Spoken corpora are collections of the spoken language of interviews or (telephone) conversations, which have to be transcribed, and as a result, their creation is expensive (11).

Written corpora are made up of written texts and can be used for studying language use in written contexts. To investigate general patterns of language use, the general corpora containing texts from a wide range of sources and genres are useful (18). When focusing on a specific domain, such as the academic or sports journalism language, specialised corpora is often used (18). Historical (or diachronic) corpora contain texts from past times and can be used to study language change and evolution over time (19). Parallel and multilingual corpora contain texts in more languages. Both can be useful for translation studies and comparative linguistic (20).

On the other hand, Hunston (2022) lists some other types of corpora, such as the synchronic, allowing studying language in a specific period (23), or diachronic corpora, tracing the development of a language over time (24). An interesting one is the monitor corpus, which is never finished and continues to grow as new data are added (26). She also mentions two corpora with applications to language learning and teaching: a learner corpus and a pedagogic corpus (26). As obvious, there is a wide range of different types of corpora, each designed for different purposes (43). Hence, when choosing a corpus to work with, one has to consider the purpose of the survey. Hunston (2022) provides a list of some commonly used corpora of English, of which the first three are:

- The British National Corpus (BNC): a 100-million-word corpus of written and spoken British English compiled in the 1990s,
- BNC-2014: a 100-million-word corpus of written and spoken British English compiled in the 2010s,

- The Brown corpus, a 1-million-word corpus of written American English, compiled in the 1970s. (44)

The list continues with other corpora, such as the COCA (Corpus of Contemporary American English), CoHA (Corpus of Historical American English) or the English Web corpora, which are "large corpora drawn from the internet" (45). A detailed description of one of them, namely the English Web 2020, will follow in the next chapter as it is the chosen corpus for comparing the two target nouns *issue* and *problem*.

3 Methodology and Data

To distinguish the use of the two nouns *issue* and *problem* in the English language, the English Web 2020 corpus was analysed. This corpus is accessible only through Sketch Engine software, which provides a great variety of tools for data analysis. Therefore, the chosen corpora, the software and its tools will be described in the following paragraphs. However, as the analysis of the two nous will develop further, some research questions will occur. Hence, the methodology will be described for each research question separately.

3.1 English Web 2020

English Web 2020 corpus is one of many corpora in the English Web containing data collected from the internet. In recent years, web corpora have become increasingly popular, given the vast amount of data that can be gathered from online sources and used for further linguistic investigation (Lindquist 2009, 22). The corpora with such data are referred to as the TenTen Corpus Family, which is an abbreviation for the target corpus size of 10+ billion words per language (Sketch Engine). The English Web 2020 corpus, accessible from Sketch Engine, is one of the largest web-based corpora of contemporary English. It comprises over 43 billion

of tokens and approximately 36 billion words of English texts from websites, blogs, and other online sources of states where English is the official language.

Sketch Engine's website states that part of the corpus is annotated by genre and topic. Out of the whole corpus size, genre annotation covers 2.2 billion tokens (5.2%), and topic classification comprehends 17.8 billion tokens (41.4%) (Sketch Engine). By genre, referring to writing style, the English Web 2020 contains data from news, discussions, blogs and legal. By topic, the data is divided into categories: arts, business, games, health, home, recreation, reference, science, sport, society and technology (Sketch Engine). Its size, diversity of genres and representativeness make it an ideal resource for linguistic research on a wide range of topics, including the aim of the present thesis. With the help of the search tool Sketch Engine, allowing access to the data of English Web 2020, a more detailed insight into the use of the nouns *issue* and *problem* is gained.

3.2 Sketch Engine Search Software

Sketch Engine is a corpus manager and a software for corpus analysis enabling researchers to explore and analyse large collections of language data. The software contains hundreds of ready-made corpora in more than 90 languages (Hurston 2022, 10). Except for analysing the corpora, Sketch Engine enables users to build their own corpora. Moreover, as language study has proven useful for language learning, the Sketch Engine version for English Language Learners called SKELL was designed (Kilgarriff et al., 2014, 66).

To analyse the English Web corpora, the following tools can be used: word sketch, thesaurus, keywords, word lists, n-grams, concordance and text type analysis (Sketch Engine). Word sketch tool shows "word's grammatical and collocational behaviour" (Kilgarriff et al., 2014, 9). The thesaurus provides a list of synonyms and similar words; the keywords tool

enables corpora comparison (30), and the word lists tool shows parts of speech and organises them by frequency (Sketch Engine). The n-grams tool lists the frequency of multi-word units; concordance provides examples of a word in context, and text type analysis is used for statistical analysis (Sketch Engine). Regarding the aim of the thesis, the word sketch tool and concordance are the most relevant ones to investigate the use of the two target nouns. However, before analysing the data in Sketch Engine's English Web 2020 corpus, the definitions and use of the two nouns occurring on the internet are introduced.

4 Analysis of the definitions in dictionaries and other sources

In an attempt to determine whether and to what extent the two nouns *issue* and *problem* are synonymous and how language users use them, their definitions provided by dictionaries and other sources will be examined and will serve as a starting point for further research in the English Web 2020 corpus. The monolingual dictionaries are Cambridge Dictionary (CaD), Merriam-Webster Dictionary (MWD), Collins Dictionary (CoD) and Dictionary (D). All the definitions provided by dictionaries and other sources were retrieved online. Firstly, the dictionary definitions of both target words are analysed separately for each noun. Secondly, other sources' definitions are described and mutually compared with the dictionaries' findings to determine whether the meanings ascribed to *issue* and *problem* vary and, if so, to what extent.

4.1 Dictionary Definitions

As the dictionaries claim, the word *problem* can be used as a noun or as an adjective. Nevertheless, as the thesis aims to investigate the use of *issue* and *problem* as nouns, the main concern is dedicated to the dictionary definitions of *problem* as a noun. The noun *problem* is defined similarly across the selected dictionaries as a situation that requires a solution as it is explicitly tied to difficulty or challenge. However, as the dictionaries provide more than one definition of a problem, the sequence of the analysis of their definitions is conducted regarding their ordering in each dictionary, from the primary definitions to the rest.

Regarding the primary definitions, CaD defines *problem* as "a situation, person, or thing that needs attention and needs to be dealt with or solved" (n.d.). Such *problems* can be *financial* or *health*. On the other hand, there are two definitions of *problem* in CoD, distinguishing between British (BrE) and American English (AmE). In BrE, it is described as "any thing, matter, person, etc., that is difficult to deal with, solve, or overcome", whereas in AmE as "a question proposed for solution or consideration" (n.d.). Similarly, MWD defines it as "a question raised for inquiry, consideration, or solution" (n.d.). Finally, D defines *problem* as "any question or matter involving doubt, uncertainty, or difficulty" (n.d.). Clearly, there are some similarities in the primary definitions of CaD and Cod BrE as well as between CoD's AmE, MWD's and D's primary definitions referring to questions.

Generally, the definitions refer to *problem* to be set for a solution as it is said to be "a source of perplexity, distress or vexation" (MWD n.d.). Consequently, *problem* explains the difficulty in understanding something or accepting it, such as in "I have a problem with your saying that" (MWD n.d.). *Problem* can also refer to a question, puzzle, or proposition that requires inquiry, consideration, or solution, often in mathematics or physics. In most cases, the

dictionary entries describe *problem* as an A1-level word. However, when referring to a (mathematical) puzzle requiring logical thought to solve it, *problem* is stated as an A2-level word (CoD n.d.). Such an example is *mathematical problem*. Except for the definitions, each dictionary provides a list of synonyms through the thesaurus tool. The CaD, CoD, and D, list the word *difficulty* as the most suitable synonym for the noun *problem*. However, MWD and D list *issue* to be one of the most relevant words to use instead of *problem*.

The noun *issue* can have multiple meanings depending on the context in which it can be used as a noun or a verb. Regarding the aim of the present thesis, only the noun's definitions are analysed. In the primary definition provided by CaD, the word *problem* is used to describe the meaning of *issue*, which is "a subject or problem that people are thinking and talking about" (n.d.). For example, *environmental* or *ethical issues* are listed in this sense. Similarly, MWD describes it as "a vital or unsettled matter", such as *economic issues* (n.d.). MWD describe the meaning of *issue* by using the words *concern* and *problem* (n.d.). Cod provides a definition claiming that "issue is an important subject that people are arguing about or discussing" (n.d.). On the other hand, the top definition in D is different, stating that *issue* is "the act of sending out or putting forth" (n.d.). Words such as *promulgation* and *distribution* are used to explain the definition.

In CaD and CoD the word *issue,* described by the primary definitions above, is a B1-level word (n.d.). However, the definition by D demonstrated that *issue* can also refer to publishing a set of copies of newspapers or magazines, and in this sense, the dictionaries describe it as a B2-level word (n.d.). Such an example used in context is: "There's an article on motorbikes in the latest/next issue" (CaD n.d.). Another meaning of *issue* is described as a dispute between two or more parties (MWD n.d.). Usually, when expressing a strong disagreement, the noun *issue* creates a collocation with the verb *take*, as obvious in the example

in CoD: “I took issue with him over his interpretation of the instructions” (n.d.) Also, the noun *issue* can be used to refer to the most important point in a situation or discussion (CoD n.d.) or to point out that “an unsettled matter is ready for a decision” (MWD n.d.). To conclude, *issue* refers to a topic or problem that people are discussing, disagreeing about, or trying to resolve.

Interestingly, both nouns *issue* and *problem* can be used in collocation with the verb *have*. To *have a problem with something/someone* means to be annoyed or offended by something (CaD n.d.), whereas *to have an issue with somebody/something* is described as “to dislike or disapprove of someone or something and let it upset or worry you” (CaD n.d.). Such examples are: “She can smoke - I don't have a problem with that” and “I'm leaving now. Does anyone have an issue with that?” (CaD n.d.).

4.2 Other Sources

The term "other sources” refers to websites providing an explanation of the nouns *issue* and *problem*. Even though their explanations may differ from the ones provided by dictionaries, they can serve as a valuable source of information for later analysis. However, as these other sources are not academic, they can be unreliable, and the observations they present can serve as the source of research questions. Despite the fact that there is a great deal of websites on which this phenomenon is described, only four of the ones with information about the author and date of publication were chosen to be studied. The websites’ names are: Learning English (LE), Difference Between (DB), Pediaa (P) and Ask Any Difference (AAD). Each of them is examined individually, and at the end of the chapter, the definitions of other sources and dictionaries are summarised.

According to the article in LE, *problem* is a situation that involves some difficulty to overcome or a question to solve it, while an *issue* is a more formal word which may occasionally

mean the same as *problem*; however, it can also refer to the point of a debate which many people are concerned about, such as *environmental issue of plastic pollution* (Robbins 2020). Furthermore, she notes that there are many words similar to *problem*, namely *trouble* or *challenge*.

Nedha (2011) notes that even though the two words are often used interchangeably as synonyms, they are in fact not. As the major difference she sees the fact that *problem* is used with the intention to be solved, whereas *issue* emphasises controversy which is not present in *problem*. She summarises the differences into four categories: controversy, scope, character and harm. There is no controversy in *problem*, compared to *issue*, which is significant for its controversy. Regarding the scope, *problem* affects the organisation or institution as a whole, whereas *issue* involves one or a few individuals. *Problem* seems to be more personal compared to *issue* which is organisational in character. And finally, it is the harm, which may be present in *issue*, but not in *problem* (Nedha 2011).

Hasa (2016) points out that despite the fact that both, *issue* and *problem*, may be used interchangeably in some cases, for instance when referring to challenging situations or matters, there are some differences between them. The main one is that *problem* is "a harmful and unwelcome matter or situation that needs to be dealt with", whereas *issue* is "an important topic or problem for debate or discussion" (Hasa 2016). Hasa provides a list of areas in which the two nouns differ. Except for their definitions (see above), it is also controversy, solution and usage. If *issue* is used, a debate or controversy can be expected. On the other hand, *problem* does not specifically indicate debate nor controversy. She points out that *issue* collocates with the verb *deal with*, whereas *problem* often appear with the verb *solve*. Regarding the usage, *issue* tends to be used less in comparison with *problem* which occurs more commonly (Hasa 2016).

Smith (2023) explains that *issue* and *problem* both refer to challenging situations and require attention to be solved. Still, she provides a comparison table of the two nouns in which the differences are categorised according to definition, usage, solutions, impacts and outcomes. The first difference emerges when *issue* and *problem* are defined: while *issue* is a small matter dealt with by discussion and not affecting people's lives, *problem* refers to a more significant and negative situation requiring an instant solution. According to Smith, *issue* is used more than *problem* as people try to solve the situation by discussion sooner than it becomes a problem. To solve an *issue*, one has to open discussion or debate, whereas when solving a *problem*, an instant and well-thought-out decision should be made. From the point of impacts, *problem* is claimed to have a negative effect compared to the neutral impact of *issue*. The outcomes of *issue* are not life-altering, in contrast to *problem* which is described as life-altering phenomenon, evoking bad conditions between people (Smith 2023).

To conclude, the data provided by the dictionaries and other sources show that the definitions and explanations of the nouns *issue* and *problem* vary. Generally, the dictionaries and the other sources agree that the two target nouns have really close meanings, however, they cannot be used interchangeably as some differences between them exist. Usually in the definitions, *problem* refers to some difficulty that should be solved, whereas *issue* may evoke debate or controversy in one of its senses. However, there are many examples in which the inconsistency of definitions occurs. For instance, the differences between the information about the verb collocates of *issue* and *problem*, the different list of synonyms for the two nouns, or in the data in categories of usage, solution and impact. Hence, in order to describe the use of the

nouns *issue* and *problem*, the aim of the analysis in the English Web 2020 is to provide answers to the following questions:

1. Which of the two nouns, *issue* and *problem,* is preferred and in which genre and topic do these two target nouns appear?
2. What are the collocational patterns of *issue* and *problem,* and is *issue* related to the controversy*,* whereas *problem* to difficulty?

5 Corpus Analysis

The corpus analysis of the use of the nouns *issue* and *problem* is undertaken in English Web 2020, a corpus accessed through Sketch Engine software. The corpus tools word sketch and concordance were used to retrieve information about the frequency of use and to analyse the collocational behaviour of the two target nouns. The chapter is divided into two main sections. Firstly, the frequency of use of *issue* and *problem* is discussed, including the text type analysis by genre and topic. Secondly, the collocates of *issue* and *problem* are examined, providing insights into the collocational patterns. The methodology of each analysis is described separately to explain the data acquisition in detail.

5.1 Frequency of Use of *Issue* and *Problem*

Table 1 provides basic statistics of the use of the nouns *issue* and *problem* on English language webs in a diachronic perspective in the years 2020, 2013 and 2008. In addition to the main corpus English Web 2020, two other corpora are discussed to acquire more data about the frequency of use of the two target nouns across the past 15 years. The concordance tool was used to provide the pieces of information about the corpus size, absolute frequency and relative frequency of the use of the lemma nouns *issue* and *problem*. As evident from the table, the size

of the English web corpora differs across the time, hence the relative frequency is listed (see the "Hits per Million Tokens") to be able to compare the use of the target nouns across different-size corpora. On the other hand, the absolute frequency refers to the number of hits in a given corpus, i.e., how many times a node occurs there, which is displayed in the "Hits" column.

Table 1: Frequency of lemma nouns *issue* and *problem* in the English Webs corpora

Corpus	**Corpus size (tokens)**	**Lemma (noun)**	**Hits**	**Hits per Million Tokens**
English Web 2020	43,125,207,462	*issue*	15,695,211	363.95
		problem	15,713,604	364.37
English Web 2013	22,728,686,012	*issue*	7,929,859	348.89
		problem	9,811,990	431.7
English Web 2008	3,268,798,627	*issue*	1,415,686	433.09
		problem	1,428,775	437.09

As described and explained in the Other Sources subchapter, some consider the noun *problem* to be more frequent than *issue*, whereas others claim the exact opposite. However, the data in Table 1 showed that the noun *problem* seems slightly more common than the noun *issue* in all examined corpora. Converted to relative frequency, both lemmas of the target nouns had an approximately similar frequency of use. However, the most significant difference between the use of the nouns occurred in the English Web 2013, where the difference was over 82 hits per million tokens compared to the difference of 4 hits per million tokens in the English Web 2008 and 0.42 hits per million tokens in the English Web 2020.

Nevertheless, this underlying statistic shows a general preference for using the lemma nouns *issue* and *problem*. It is important to note that the frequency of use can vary significantly depending on the specific genres or topics being examined. Therefore, further analysis is needed

to determine the contextual factors that influence the frequency of use of *issue* and *problem* in the English Web 2020 corpus.

As mentioned in the subchapter English Web 2020, the target corpus contains over 36 billion words. To ensure that a huge amount of data is easily studied and analysed, annotation is used. Leech (2004) describes it as "the practice of adding interpretative linguistic information to a corpus". There are many different types of annotation, for instance, the part-of-speech (POS); however, as visible in Table 2 and Table 3, the topic and genre annotations are to be discussed as the target corpus English Web 2020 is partially annotated by these two categories. The topic classification and genre annotation were accessed through the concordance text analysis tool providing statistics calculated from the first 10 million hits of the node (Sketch Engine). Similarly, as in Table 1, two types of frequency are listed to show the raw data in the target corpus and ensure the comparison of the data across different-size corpora. To provide examples of the two nodes used in a context, the Good Dictionary Examples (GDEX) tool is used. It is a system for evaluating sentences to determine their appropriateness for use as illustrative examples in dictionaries or as effective pedagogical aids (Sketch Engine).

5.1.1 Analysis of Topic Types

The Sketch Engine distinguishes 12 different topic categories displayed in Table 2. The number of hits of the two lemma nouns *issue* and *problem* differs in each of the 12 topics. Nevertheless, some differences are subtle compared to others, in which the number of hits of one node is twice as big as the number of hits of the second one. The data show that both lemmas are the most frequently used in reference, society and technology. However, the ranking and the number of hits in these three categories are different for each lemma. The order of the top three topics by frequency is as follows:

issue	1. reference	2. society	3. technology
problem	1. technology	2. society	3. reference.

Table 2: Frequency of lemma nouns *issue* and *problem* according to different topic

Absolute frequency (AF) / Relative text type frequency per million tokens (RF)												
Lemma/ Topic	**arts**		**business**		**games**		**health**		**home**		**recreation**	
	AF	RF	AF	RF	AF	RF	AF	RF	AF	RF	AF	RF
issue	281,255	6.52	310,617	7.2	58,493	1.36	175,032	4.06	94,171	2.18	52,502	1.22
problem	254,068	5.89	271,079	6.29	68,378	1.59	376,316	8.73	206,752	4.79	96,660	2.24
Absolute frequency (AF) / Relative text type frequency (RF)												
Lemma/ Topic	**reference**		**science**		**sports**		**society**		**technology**		**news**	
	AF	RF	AF	RF	AF	RF	AF	RF	AF	RF	AF	RF
issue	796,663	18.47	202,479	4.7	119,463	2.77	555,980	12.89	499,852	11.59	152,380	3.53
problem	425,989	9.88	249,882	5.79	145,605	3.38	514,768	11.94	794,531	18.42	129,539	3,00

The sequence of the remaining nine topics varies depending on the node, which is visible in Table 2. Interestingly, the frequency of two of the three most frequent topics listed above significantly differs between the two nodes. These topics are reference and technology. Whereas *issue* seems to be more than twice as common as *problem* in reference topic, *problem* tends to be used more frequently in technology topic. The sentences below are the GDEX examples of use of the two nodes regarding the topics of reference (1) and technology (2):

(1) Due to these issues and difficulties with deployment, many carriers were not able to or delayed acquisition of these updated capabilities.

(2) Wow! everyone is having the same problem: unspecified error.

The data acquired in the target corpus also show that there are three more areas in which the frequency differences are the most significant. Namely in the topics of health, home and

recreation. In all of these categories, the lemma noun *problem* is double times as frequent as the lemma noun *issue*. The GDEX examples of the lemma noun *problem* used in context in the the topics og health (3), home (4) and recreation (5) include:

(3) Health problems - depressive tendencies and a worsening eye problem - led to his suicide on 10 August 1970.

(4) Deer are beautiful - but they can cause a lot of problems in your backyard.

(5) Proper thread handling is required to avoid problems while sewing.

The seven remaining categories of topics can be divided according to a little predominance of one of the nodes. The sentences below demonstrate the use of the nodes in the categories regarding their predominance in a topic. In four of the 12 topics, the lemma noun *issue* is slightly more frequent than the lemma noun *problem*. These categories contain the topics of arts (6), business (7), society (8) and news (9), which are all examples of social studies areas. On the other hand, in science (10), sports (11) and games (12), the node *problem* seems to be a bit more common than issue.

(6) As QI features several images during each episode there are copyright issues.

(7) If comfortable, employees can even speak with clients or email about non-complex issues.

(8) A wide variety of social issues were covered.

(9) On Wednesday, Iran's top naval commander expounded on the issue of control to the nation's English-language Press TV.

(10) The problem of the detection and mapping of a stochastic gravitational wave background (SGWB), either of cosmological or astrophysical origin…

(11) Skilled wingers aren't a big problem for this club right now.

(12) I just had a problem with a lot of the game pieces.

5.1.2 Occurrence Across Different Styles

In Sketch Engine, the corpora can be annotated by four different genres categories referring to writing styles, which are blog, discussion, legal and news. Table 3 provides an insight into each category regarding the absolute and relative frequency of the two lemma nouns *issue* and *problem*. Interestingly, the absolute frequency of the target noun lemmas reveals that both lemmas are, in most cases, used primarily in the context of news, secondly in blog, thirdly in discussion and finally in legal. The data show that except for the category of blog, whose frequency is approximately the same for both nodes, the three other categories display more significant contrast between the use of the two nodes.

Table 3: Frequency of lemma nouns *issue* and *problem* according to different genre

Absolute frequency (AF) / Relative text type frequency per million tokens (RF)								
Lemma/genre	blog		discussion		legal		news	
	AF	RF	AF	RF	AF	RF	AF	RF
issue	84,373	1.96	47,735	1.11	45,484	1.05	342,882	7.95
problem	99,572	2.31	88,109	2.04	5,968	0.14	188,964	4.38

There are two genre categories, in which the lemma noun *issue* is comparatively more common. The frequency of the node *issue* is several times higher in the category of legal and news. The most distinct difference occurs in the genre of legal, where the lemma noun *issue* is used 7.5 times more frequently than the lemma noun *problem*. In the news category, the frequency of *issue* exceeds 1.8 times the use of *problem*. Such GDEX examples of the node *issue* used in legal (13) and news category (14) include:

(13) Findings are made by a trial court to dispose of the issues raised in the pleadings.

(14) “While we often disagreed on many issues, I always found Tim to be honorable and a man of integrity,” Triantafilou said.

The categories of blog and discussion are the ones in which the lemma noun *problem* is, by its frequency of use, more common than *issue*. However, the distinction between their frequencies is not as significant in the blog and discussion genre as in the two other categories analysed above. In the discussion genre, *problem* exceeds the use of *issue* over 1,8 times. In the blog category, the relative frequency of use of both lemma nouns is similar, differing by the value of 0,35. The two sentences below illustrate the node *problem* in the blog category (15), and discussion category (16):

(15) The Louvre is actually not a good place to look at art and if moving the Mona Lisa to a dedicated gallery elsewhere can help solve that problem, they should do it.

(16) The problem is that too many people have too many different opinions. I think this would be true of all of the areas you mention.

To conclude, the data provided by Sketch Engine concordance text analysis tool displayed the information about the frequency of use of the nodes *issue* and *problem* across the 12 different topics and four different genres. Generally, *problem* was used comparatively more than *issue* in most cases in both of these categories. Out of 12 different topics, *problem* was used in seven categories more often than *issue*. In the genre category, in which four categories were distinguished, *problem* appeared more frequently in two groups. The higher frequency of use of problem corresponds to the information acquired on the website Pedia and simultaneously disproves the statement by Smith (2023) that the noun issue tends to be more common.

In the categories of topic, both lemma nouns *issue* and *problem* were most frequently used in reference, technology and health. Surprisingly, these three topics and the topic of

recreation were the categories in which the major differences between the frequencies of the two lemmas occurred. On the other hand, subtle differences were found in the topics of arts, business, games, science, sports, society and news. Except for the genres of games, science and sports, *issue* was slightly more frequent than *problem*.

The genre analysis of the frequency of use in four different categories revealed that both lemma nouns *issue* and *problem* tend to be most commonly used in the news genre and the least in the legal genre. In both of these, *issue* was more frequent than *problem*, especially in the legal category, where the difference was the most significant. In the genres of blog and discussion, *problem* was used comparatively more than *issue*.

5.2 Collocational Patterns of *Issue* and *Problem*

To distinguish the subtle nuances between the meanings of the two target nouns *issue* and *problem,* and discover to what extent these two nouns are synonymous, their collocational patterns will be analysed. The data in the English Web 2020 corpus was accessed through Sketch Engine software through the word sketch tool, which provides a detailed insight into the collocates in close proximity to the target nouns. As the comparison between the two lemma nouns via collocation is made, the word sketch difference tool will be used too. The word sketch difference tool effectively compares the word sketches of the two nodes and highlights the differences in their collocates (Sketch Engine). Moreover, each node has its own colour (green or red) generated by Sketch Engine, whose shade “indicates the strength of the collocation” (Sketch Engine). The nodes' collocates are divided into many categories in which they are stored regarding the typicality score. In addition, the concordance is easily accessible through a direct link displayed in each category next to the individual collocates.

As noted, and briefly described in the Collocation subchapter, there are more statistical measurements to use when analysing the collocational patterns of words; however, the word sketch sorts its data by the logDice score. Rychlý (2008) states that the maximum value of the logDice score is 14; nevertheless, the value usually won't exceed 10. He also points out that if two scores are compared, plus one point equals twice as often collocation and plus seven points means approximately 100 times more frequent collocation (9). The higher the score, the stronger the collocation between the two words is, meaning that they tend to co-occur together and not that often with other collocates. On the other hand, the score is lower, the more frequently the collocate co-occurs with other words, creating weak collocation (Word sketch - collocations and word combinations). If the frequency of a particular term or collocation equals 0, it will be assumed that the logDice score is 0, as in such a huge amount of data, there are no such examples in use.

The analysis of the collocational patterns in word sketch and word sketch difference will mainly focus on several selected categories whose data will be confronted with the information provided by the websites listed in the Other sources subchapter.

5.2.1 Verb Collocates

There are several categories of the verb collocates, however, the main ones are the following: *verbs with "issue/problem" as object* and *verbs with "issue/problem" as subject*. Both of the categories will be further analysed and described regarding the data accessed through the tools of word sketch and word sketch difference. The tables below do not provide a full list of verb collocates but only a selection of them. The full list of verb collocates can be found in the Appendix 1 and 2 at the end of this thesis.

Table 4: Word sketch: verbs with "*issue*" and "*problem*" as objects

verbs with "*issue*" as object			verbs with "*problem*" as object		
verb collocate	**frequency**	**logDice score**	**verb collocate**	**frequency**	**logDice score**
address	455,002	10.7	solve	738,960	11.4
resolve	173,651	9.8	cause	298,150	9.5
discuss	185,918	9.4	fix	166,941	9.1
raise	196,268	9.3	address	179,416	9.1
relate	158,815	9.2	resolve	95,466	8.6

Table 5: Word sketch: verbs with "*issue*" and "*problem*" as subjects

verbs with "*issue*" as subject			verbs with "*problem*" as subject		
verb collocate	**frequency**	**logDice score**	**verb collocate**	**frequency**	**logDice score**
face	71,633	9.1	solve	192,904	10.9
relate	58,054	9.0	arise	61,775	9.0
affect	61,923	9.0	face	43,991	8.2
arise	48,845	9.0	occur	35,431	7.8
surround	50,229	8.8	lie	25,951	7.6

Tables 4 and 5 display the top five verb collocates of the nodes *issue* and *problem* in the English Web 2020 in categories where the nodes function as objects and subjects. The word sketch tool in Sketch Engine was used to retrieve the data. As evident from both tables, three verbs co-occur with both nodes: *address*, *resolve* and *arise*. The verbs *address* and *resolve* are among the top five most frequent verb collocates where both nodes function as objects. On the other hand, the verb *arise* is listed as one of the most frequently used verbs where the nodes function as subjects. However, when the logDice score is considered, it can be seen that the node *issue* is bound to these three mutual verbs collocates more than the node *problem*. Even

though the verb *arise* is less frequently used with the node *issue*, its typicality score is the same as for *problem* whose frequency is higher.

Although the logDice score usually won't exceed 10, there are three examples of the verb collocates, in tables 4 and 5, with a higher typicality score. Such a high number signifies that these verbs typically co-occur with the nodes and not that frequently with other words. For the node *issue* there is one verb collocate to be frequently found in its close proximity, and that is the verb *address*. A GDEX example of such collocation with *issue* as object + *address* in concordance is:

(1) Nor does it address issues of community autonomy.

For the node *problem*, *solve* is a typical verb collocate in whose company the node serves either as an object or a subject. The collocation of the noun *problem* as an object and the verb *solve* is really strong regarding the logDice score showing the number over 11. Similarly, the verb *solve* frequently occurs with *problem* when the node functions as a subject. The two sentences below illustrate the use of *solve* + *problem* as an object and as a subject, respectively.

(2) He knows how to help my kid to understand and solve the problem.

(3) Fix that, put the cost curve on a downward trajectory, and all the other problems solve themselves.

In the category where the nodes serve as objects, there are two verb collocates with similar typicality scores that are worth discussing. Whereas for *issue* collocating with the verb *discuss*, the node *problem* is likely to appear in co-occurrence with the verb *cause*. Both collocations exceed the value of 9.4; hence, the collocates and their nodes are often to be seen with each other. The following sentences provide two examples of such verb collocates of both nodes functioning as objects.

(4) A small group of students would gather to discuss the issues surrounding educated women and their lives.

(5) Contact with dog dirt can pass on toxocariasis which causes serious eye problems including blindness.

Although the rest of the verb collocates displayed in Table 4 and Table 5 are also quite characteristic for the nodes, regarding the logDice score, they won't be exclusively commented. Generally, as it has been said several times, the higher the score, the more typical and stronger the collocation is. For the purpose of the present thesis, the verb collocates of the two nodes are better to compare using the word sketch difference tool as displayed in Table 6.

Table 6: Word sketch difference: verbs with "*issue/problem*" as object

verbs with "*issue/problem*" as object					
Selected verb collocates		frequency		logDice score	
		issue	problem	issue	problem
debate	*issue/problem*	21,352	618	6.9	1.4
discuss		185,918	39,256	9.4	6.9
report		27,768	36,508	6.7	6.8
fix		110,999	166,941	8.8	9.1
cause		55,198	298,150	7.2	9.5
solve		79,292	738,960	8.5	11.4

Even though the Sketch Engine enables its users to download a visualisation of the searched word and its collocates, its readability is not suitable and therefore the data was transformed to the form of Table 6. The table shows six selected verbs collocating with the nodes *issue* and *problem* as objects. The selection of the verbs in Table 6 was made regarding the frequency and typicality of the collocation expressed through the colour and its shade. The

colour demonstrates the preference of the collocate to co-occur with the node and the shade shows the strength of the collocation (Word sketch difference - compare words via collocations). The green colour signifies the preference for the node *issue*, whereas the red colour represents the preference for the node *problem*. As visible, for each node, two strong and frequent verb collocates were chosen. The two not coloured collocates represent verbs without preference for either of the nodes.

The most significant difference between the nodes occurs in the collocation with the verb *debate*. The collocation of *issue* + *debate* is 35 times more frequent that the same collocation with the node *problem*. On the other hand, the node *problem* is most likely to be accompanied by the verb collocate *solve* that is not that frequent for *issue*. Hence, as the data show, *issue* tends to be *debated* whereas *problem* is rather *solved*. To illustrate such context, the sentence below shows the GDEX collocations of *issue* + *debate* (6). The collocation of problem + solve is illustrated in Example 2 in the Verb collocates subchapter.

(6) The health effects of abortion on women are not easily separated from the hotly debated social issues that surround the practice of abortion.

The verb *discuss* is the third most frequently used verb collocate of *issue* functioning as object. Even though that the verbs *discuss* and *debate* are really close in meanings, the data show that the collocation with *issue* + *discuss* is stronger and more typical, exceeding the score 9.4. Nevertheless, the score difference between the same verb collocate and the node *problem* is not that obvious as with the verb *debate* as explained in the previous paragraph. Proportionally, a similar score difference can be seen between the verb collocate *cause* and the node *problem*. The typicality score is, in this case, 9.5 compared to 7.2 for the same verb collocate and the node *issue*. Examples 4 and 5 (see above) demonstrate the use of the verb collocates *discuss* and *cause* with the nodes *issue* and *problem* in context.

Besides the verb collocates typical of either of the nodes, some verbs seem neutral since the collocation between them and a node has approximately the same typicality score. From the complete list of such verb collocates, only two were chosen and displayed in Table 6. While the verb collocate *report* is slightly more typical of the node *issue*, the second neutral verb *fix* is a little more frequent with *problem*. Yet, the difference between the logDice scores of these two collocations is negligible. Since their typicality score is considered, it can be seen that the verb *fix* is more typical of both of the nodes than the verb *report*, whose logDice score ranges the value of 6.7– 6.8. Examples 7 and 8 demonstrate the collocation between the two nodes and the verb collocate *report*, whereas the other two illustrate the collocation of *fix + issue* and *fix + problem*.

(7) Please report web-related issues to EMAIL Copyright 2010 Zyntax Consulting BV, All Rights Reserved.

(8) Please report any problems via email: EMAIL.

(9) We have fixed some application issues.

(10) Developers now have access to free and inexpensive developer tools that can help find and fix many common application problems.

In conclusion, Tables 4, 5 and 6 displayed the verb collocates of the nodes *issue* and *problem*. Each table and its data were commented on and described regarding the studied phenomenon. The analysis revealed that the verb collocates of *issue* and *problem* differ in both studied categories where the nodes function as objects and subjects. Simultaneously, in both categories, some similarities concerning the verb collocates were also observed. The verb collocates of the nodes were firstly analysed separately through the word sketch and then compared via the word sketch difference tool in Sketch Engine. The comparison enabled

observation of the differences in the use of the nodes and their verb collocates and detected the node's preference to (not to) combine with certain verbs.

The verbs *address* and *solve* must be listed to illustrate the most typical verb collocates of the two nouns *issue* and *problem* functioning as objects. Both collocations have a logDice score higher than 10, which proves that such collocation is strong and typical. As the collocation *address* + *issue* reaches the score of 10.7 and the collocation *solve* + *problem* gets at the value of 11.4, it is highly probable to notice the use of the node and its typical verb collocate close to each other. At the level of analysis of the collocations where the nodes functioned as subjects, the verb *arise* had the same logDice score for both nodes and was included in the top five most frequent verb collocates of the target words.

The data in the word sketch difference showed that the most significant difference, concerning the verb collocates, occurs with the verb *debate*, whose logDice score was with *issue* almost five times higher compared to *problem*. The comparison also revealed that even though the verb *solve* tends to occur frequently with the node *problem*, the collocation with the node *issue* is also quite common. There were also two verb collocates, *discuss* and *cause*, which were 1.3 times more typical of one of the nodes. Whereas the verb *discuss* tends to join with the node *issue*, the verb *cause* seems to be more common for *problem*. However, the verbs *report* and *fix* illustrated that some verbs tend to appear near both nodes without a significant difference concerning the typicality score.

Discussion

In order to offer a comprehensive overview, the current study will now juxtapose the verb collocate findings derived from the English Web 2020 with the information from the Other sources subchapter.

In substance, all of the sources discussed in the Other sources subchapter agreed that *issue* requires a discussion or debate. Hence, it can be concluded that to achieve such act, one has to *debate* or *discuss* about an *issue*. Such verbs, like *discuss* or *debate*, as described above, were among the most frequently used verb collocates of the node *issue*. In addition, Hasa (2016) also pointed out that the typical verb collocate for the node *issue* is the phrasal verb *deal with*.

As the investigation of the phrasal verbs was not included in the verb collocates analysis, the tendency of the verb *deal with* to appear in collocation with *issue* will be briefly analysed now by using the concordance tool. The simple query type of the phrasal verb *deal with* appears in the English Web 2020 more than four million times. When the +1 range of the collocates is displayed, showing the first right word next to the node, the first listed and most frequent item is truly the word *issues*. There were 22,557 co-occurrences of the collocate *issues* within the selected range. However, the value of 5.90 of the logDice score signifies that the phrasal verb also collocates with other words and not exclusively with the noun *issue*. Hence, the statement by Hasa (2016) is true, but it is important to note that the phrasal verb *deal with* is not one of the most commonly used verb collocates of the node *issue*.

When the other sources described the verb collocates, they pointed out that the node *problem* is commonly used with the verb *solve*. Hasa (2016) states that the word *solve* is among the words mostly used in relation to *problem*. As the analysis of the verb collocates explained, the logDice score of the collocation *solve + problem* is unusually high, emphasising that these two words tend to occur near and with each other exclusively. Therefore, it can be stated that Robbins (2020), Nedha (2011) and Hasa (2016) were right.

Moreover, when speaking about the differences existing between the two nouns, the other sources were not in agreement about *what* the *issue* or *problem* is. Nedha (2011) emphasised the variety of opinions between people and the instant debate about *issue* and

pointed out that *homosexuality* is an example of *issue*. On the other hand, in relation to *poverty* she used the word *problem*. Hasa (2016) provided more examples and stated that *poverty* or *abortion* are *issues*, whereas *racism* or *sexism* are *problems*. Nevertheless, when giving examples in use, she mentioned *racism* with *issue* as well as with *problem*. Hence, with the help of word sketch difference, the analysis of the category where the nodes serve as subject complements after the linking verb *to be* (X is a "problem/issue") was conducted. The data showed that people prefer to use the words *poverty* and *racism* more with *problem*, whereas the word *abortion* with *issue*. The word *homosexuality* was not included in this list. Regarding the preference and the most significant differences between the two nodes, the word *reform* was used for *issue* and the word *hell* for *problem*.

5.2.2 Modifiers of the Target Nouns

Another category to analyse to distinguish the subtle differences between the target nouns *issue* and *problem* is the one of their modifiers. Hence, firstly the modifiers of each node will be listed and described, and secondly, the node's modifiers will be mutually compared via the word sketch difference.

Table 7: Word sketch: Top 5 modifiers of *issue* and *problem*

Top 5 modifiers of *issue* and *problem*					
issue			*problem*		
modifier	**frequency**	**logDice**	**modifier**	**frequency**	**logDice**
health	230,968	8.9	health	250,239	9.1
important	162,306	8.4	serious	171,107	9.1
key	135,072	8.3	big	213,400	8.6
environmental	104,963	8.2	major	134,862	8.2
major	113,450	7.8	real	134,725	8.2

Table 7 displays the top five modifiers of the lemma nouns *issue* and *problem* accessed through the word sketch tool. The modifiers are listed vertically according to the strength of the collocation with the node. Generally, it can be stated that the strength of the collocates of the node *problem* seems to be higher compared to the node *issue*, whose collocational scores are at the same positions lower. At the same time, the top five collocations with *problem* are, according to frequency, more common, which reflects the foregoing finding that *problem* is used more than *issue.*

As can be observed from the table above, the strongest collocate of *issue* and *problem* is the word *health*, creating the collocations *health* + *issue* and *health* + *problem*. There is not that significant difference between the two collocations regarding the logDice score, however, the frequency differs by more than 20 000 hits. The GDEX examples below illustrate the use of the strongest collocates in terms of modifiers of the nodes *issue* and *problem*, respectively.

(1) The pet's obesity would naturally result in other health issues.

(2) Overweight/obesity is a major health problem.

Another collocational similarity occurs with the modifier *major*, which is among the top five collocates that both nodes share. Again, for *problem,* the logDice score, as well as the frequency, is higher in comparison to *issue.* Examples 3 and 4 provide an illustration of such collocation with the collocate *major* and each node.

(3) Even though homesickness is completely normal, it's still a major issue for a great deal of people when they're abroad.

(4) Major problem was lack of capacity in health care workers on this issue.

According to the information presented in Table 7, there are three other modifiers that are different for each of the nodes. The noun *issue* is most commonly modified by *important, key* and *environmental*, whereas the noun *problem* often collocates with the words *serious, big* and *real*. However, according to the website Thesaurus, some of the three different modifiers of *issue* and *problem*, listed in Table 7, are considered synonymous. The synonym pairs include *important* and *serious* (Synonyms of important), *key* and *major* (Synonyms of key), and *major* and *big* (Synonyms of major). This support the claim that the nouns *issue* and *problem* are similar in meaning as even some of their typical collocates, in terms of modifiers, are mutual synonyms. The following examples illustrate the use of the modifiers *important* and *key* of the node *issue and* the modifiers *serious* and *big* of the node *problem*.

(5) Social reform and home rule for Ireland were important domestic issues after 1900.

(6) During the preliminary design phase, one key issue to be addressed was the effect of extreme wind loads on the structure.

(7) GERD can lead to more serious health problems over time.

(8) It has been a big problem down through history and continues to be a problem.

The modifiers *environmental* and *real*, whose logDice score equals 8.2, were not listed as synonyms of neither of the modifiers displayed in Table 7. The two GDEX examples below provide examples of the use of *environmental* and *real* with their typical nodes.

(9) Public Lab is a community and non-profit democratizing science to address environmental issues that affect people.

(10) Bad news: This disease can be a real health problem and pain in the neck making it very difficult to cope up with daily activities of life.

Table 8: Word sketch difference: modifiers of "*issue/problem*"

Modifiers of "issue/problem"					
Selected modifier		frequency		logDice score	
		issue	problem	issue	problem
contentious	*issue/problem*	21,677	0	6.2	—
controversial		28,992	414	6.6	0.8
technical		57,309	45,759	7.3	7.2
particular		35,345	32,952	6.4	6.5
optimization		382	15,404	0.4	6.1
math		0	15,641	—	6.1

For the same reasons explained when interpreting the data in Table 6, the data about the modifiers compared via word sketch difference was transformed to the form of Table 8. The table shows six selected modifiers of the nodes *issue* and *problem*. The selection of the modifiers was made regarding the frequency and typicality of the collocation expressed through the colour and its shade. The colour demonstrates the preference of the collocate to co-occur with the node and the shade shows the strength of the collocation (Sketch Engine). The green colour signifies the preference for the node *issue*, whereas the red colour represents the

preference for the node *problem*. For each node, two strong and frequent modifiers are displayed. The two not coloured collocates represent modifiers without preference for either of the nodes.

The first interesting finding about the difference in the use of modifiers occurs with the words *contentious*, used exclusively with *issue,* and *math*, which collocates only with *problem*. Rychlý (2008) states that when the logDice value is 0, as in examples with contentious and *math*, it means "there is less than 1 co-occurrence of XY per 16,000 X or 16,000 Y" (9). The Examples 11 and 12 show a GDEX sentences in which the two nodes and their collocates *contentious* and *math* are used.

(11) The withdrawal of U.S. military forces from Iraq was a contentious issue in the United States for much of the 2000s.

(12) Please complete the math problem to prove you are not a robot: 1 + 2 =

Moreover, according to Thesaurus.com, the adjective *contentious* is the synonym for the adjective *controversial*, listed as the second entry from the top of Table 8. In this case, the collocational difference between the two nodes and their collocates is still significant; however, the frequency of *controversial + problem* is higher than 0 in contrast to the collocation *contentious + problem* whose frequency was 0. Nevertheless, the collocation *controversial + issue* is definitely much more frequent. A similar difference between the two nodes occurs with the collocate *optimization* which is more frequently used with the node *problem* than with *issue*. The nodes *issue* and *problem* are used in the following sentences to exemplify their frequent use with the modifiers *controversial* and *optimization*.

(13) After six years of discussing controversial issues at BCC, I can moderate conversations and make everyone feel welcome to speak, even when they disagree with the majority.

(14) Most optimization problems found in the real world cannot be solved using analytical methods.

The middle part of Table 8, in which the two modifiers without preference for either of the nodes are displayed, represents some of the shared modifiers that both nouns *issue* and *problem* co-occur with. Their collocational logDice scores differ by one-tenth in both these cases, However, the frequency of *technical/particular* + *issue* is slightly more frequent than that of *technical/particular* + *problem*. When the two collocates are compared regarding the logDice score, it is evident that the collocations with *technical* are stronger than the collocations with *particular.* As both modifiers are relatively frequent to use with both target nouns, their collocations are illustrated in the sentences below.

(15) If you experience any technical issues, please contact Servicedesk (EMAIL).

(16) Customers experiencing technical problems with their PCs are generally instructed to contact the manufacturer for assistance.

(17) We know that affordable housing is a particular issue in rural areas and this has a real impact on those communities.

(18) Heroin use among high school students is a particular problem.

To further explore the differences between the two nouns *issue* and *problem*, an analysis of their modifiers was conducted. Firstly, the top five modifiers of each node were identified and compared using the word sketch tool and the results were transformed to the form of Table 7. Secondly, the word sketch difference analysis of six selected modifiers was conducted and displayed in Table 8.

The results in Table 7 showed that *problem* has stronger collocates compared to *issue*, whose collocational logDice scores with the top five modifiers were lower. Additionally, the top five collocations with *problem* were, regarding the frequency, higher compared to *issue*,

which supports the foregoing finding that *problem* is used more than *issue*. The two nodes share the strongest collocate, the word *health*, and the modifier *major*, which also appeared in the top five collocates of *issue* and *problem*. In both cases, *problem* had a higher frequency and logDice score. It was revealed that there are three modifiers that are unique for each node. The noun *issue* is commonly modified by *important, key* and *environmental*, whereas *problem* often collocates with *serious, big* and *real.* However, it should be noted that, according to Thesaurus.com, some of these modifiers are considered synonymous, such as *important* and *serious*, *key* and *major*, and *major* and *big*. This supports the notion that *issue* and *problem* are similar in meaning as they share mutual synonyms among their typical collocates. Similarly, Table 8 revealed significant differences between the use of modifiers of the key nouns. Such difference is that the modifier *contentious* is used exclusively with *issue*, while the modifier *math* collocates only with *problem*. On the other hand, both nouns collocate with the modifiers *technical* and *particular*, having similar frequencies and logDice scores.

Discussion

Moving forward, the present study aims to supplement the findings outlined in the previous section by engaging with the information obtained from the various websites listed in the Other sources subchapter.

In general, the other sources described *issue* as a matter of controversy and debate, while *problem* was generally referred to as something rather negative, causing trouble, harm or challenge. The analysis of the modifiers of the two nouns conducted via word sketch and word sketch difference tools revealed the top five collocates and the main differences in using their modifiers which will be now confronted with the data provided by other sources.

From the data, summarised above, it can be concluded that *issue* is linked with controversy as the word sketch difference exposed that the modifiers *contentious* and *controversial* are the ones in whose use the two target nouns differ the most. Moreover, the further investigation revealed that these modifiers are synonyms. The collocations containing the noun *issue* were not only several times more frequent than the ones with *problem*, but also their logDice scores were significantly higher than the ones containing the second node. Hence, *issue* is, according to the research, definitely more commonly associated with matters that people consider *controversial*. Such a *controversial* example might be *environmental issue(s)*, which illustrates the use of one of the five strongest modifiers of *issue* in terms of logDice score. *Environmental issue* is an example used for example by Robbins (2020), when explaining the difference between *issue, problem* and *matter.*

On the other hand, when Nedha (2011) mentions *controversy* to further explain the use of *issue*, she points to homosexuality viewed as *social issue*.. The data diplayed in the full list of modifiers showed that the word *issue* tends to be modified by *social* more than the noun *problem*. Moreover, when the top differences are analysed, there can be found modifiers such as *bond, justice, right, regulatory, policy, compliance* that prefer to collocate with *issue* than with *problem.* Such collocations not only evoke the impact of *issue* on more than just one individual but also indicate its use in legal or business environments. Hence, when Robbins (2020) stated that issue is in comparison with *problem* more formal, her argument was valid.

As introduced, the noun *problem* was commonly described by the other sources as something difficult, challenging or having a negative effect on people's lives. Even though the words *difficult* and *challenging* are included in the list of words modifying the noun *problem*, they are not, according to frequency and logDice score, among its most typical co-occurrences. Nevertheless, these two modifiers are more likely to be used with *problem* than with the noun

issue. Although the modifiers *difficult* and *challenging* were not displayed in Table 7 or 8, they can be found in the Appendix 3 where the full list of modifiers is included. However, the research revealed that *problem* is often modified by words like *serious, big* or *major*, which are adjectives referring to the scope of its impact.

Moreover, when the full list of modifiers accessed through the word sketch difference is further analysed, it can be seen that *problem* is often modified by words related to *health problems*, which may potentially have negative effect on people's lives. Among other modifiers, whose use differs between the two target nouns the most, are words like *vision, heart, alcohol, respiratory, drug, sleep, behavior, skin* or *chronic*. Therefore, it can be stated that when Smith (2023) assumed that *problem* has in comparison with *issue* negative life-altering impacts on people's lives, she was right. People on the English webs usually use the noun *problem* in such contexts. However, the variety of different *health problems* indicates the personal character of *problem*, and hence, Nedha's (2011) statement about *problem* being more personal was true.

6 Conclusion

The present study aimed to conduct a corpus-based comparison of the use and meanings of the nouns *issue* and *problem*, with the purpose of determining the extent to which these two words are synonymous and whether they overlap or differ in meaning. Firstly, the dictionaries were analysed and the data they provided were then compared with the information found on websites discussing the differences between *issue* and *problem*.

The analysis of four online monolingual dictionaries revealed that both target nouns *issue* and *problem* may function as nouns, however, *issue* can also be used as a verb, whereas *problem* as an adjective. Hence, the analysis was narrowed down to their noun definitions. As

emerged from the analysis, the noun *issue* has two different level meanings. Its B1-level meaning refers to a subject or disagreement that is debated, whereas the B2-level meaning refers to the act of publishing a set of copies. Similarly, the noun *problem* has two different level meanings. If *problem* refers to something that needs attention to solve it, it is described as a beginner, A1-level word. On the other hand, if it refers to some logical, mathematical question, it is an A2-level word. Hence, the noun *problem* seems to be part of the basic vocabulary knowledge, whereas *issue* is slightly more advanced.

The dictionary definitions of *issue* and *problem* were confronted with four non-academic web sources and the points of the inconsistency of data about the use and collocational patterns served as a source of research questions to be answered via the English Web 2020 analysis. The aim of the first research question was to reveal which of the target nouns is preferred and in which genre and topic do they appear. The aim of the second analysis was to study the collocational patterns of *issue* and *problem* to distinguish whether the former is related to controversy, whereas the latter to difficulty.

To answer the former research question, the analysis of the target corpus via concordance tool was conducted. The results of the data analysis indicated that the noun *problem* is used more than the noun *issue* not only in the English Web 2020 but also in two other corpora that were included to illustrate the diachronic preference. This finding corresponds to the fact that *problem* belongs to the basic vocabulary knowledge, and therefore, more language users may know it and potentially use it.

To provide a comprehensive overview of the use of *issue* and *problem* in the English Web 2020 corpus, their occurrence across different genres and topics was further studied. The data showed that *problem* was used more frequently than *issue* across most topics and genres. Both nouns were commonly used in the topics of technology, health and recreation. At the same

time, in these topics, the most significant differences occurred. Other topics showed slight differences and in most of them, *issue* was used more. Interestingly, as the use of *issue* was in the legal and news genres several times more frequent than for *problem*, and simultaneously, *problem* predominated in blog and discussion genres, it can be stated that *issue* seems to be used in more formal contexts and therefore more formal.

The study of the collocational patterns was the most extensive part of the present thesis and aimed to focus on selected categories such as the verb collocates and the modifiers of *issue* and *problem*. The verb collocates and modifiers were firstly analysed separately for each node and then mutually compared to observe the existing differences and similarities. The study was undertaken via Sketch Engine's word sketch and word sketch difference tools.

The data suggests that there are some verbs that are more typical of one node than the other. Notably, the verbs *address* and *solve* were found to be the most typical collocates for *issue* and *problem*, respectively, when the nodes functioned as objects. On the other hand, when serving as subjects, *issue* typically collocated with the verb *face* and *problem* with the verb *solve*. Both nodes, functioning as subjects, collocate with the same typicality score with the verb *arise*. However, the mutual comparison of the verb collocates, revealing the most significant differences existing between the two words, showed that *discuss* and *solve* were the verbs whose usage differed the most between the two nouns. On the other hand, some verb collocates, such as *report* or *fix*, did not prefer either of the nodes and collocated with both with the same typicality.

The analysis of the modifiers indicated that *problem* has stronger collocates and higher frequencies of use than *issue*. Both nouns shared, among their top five strongest collocates, the modifiers *health* and *major*. However, *issue* was commonly modified by *important*, *key* and *environment*, whereas *problem* often collocated with *serious, big* and *real*. Interestingly, further

study revealed that a few of the top five modifiers were, according to Thesaurus.com, mutual synonyms. When the target nouns *issue* and *problem* were compared via word sketch difference, some similarities in their modifiers emerged. For instance, in the words *technical* or *particular*, which collocated with both nouns with the same typicality. The study found that the greatest disparities involved the modifiers *contentious*, which only collocated with *issue*, and *math*, which was a typical collocate of *problem*.

In conclusion, this study aimed to compare the use and meanings of the nouns *issue* and *problem* and determine if they are synonymous and overlap or differ in meaning. As the two lexemes do not share all their lexical meanings, verb collocates and modifiers and therefore do not meet the three conditions of absolute synonymy, introduced in the theoretical part, they may be classified as near-synonyms. However, only selected categories of collocates were analysed and studied; therefore, further research would be needed to provide a more comprehensive understanding of the use and meanings of *issue* and *problem*. The analysis could be extended to a broader range of collocates and contexts and confront the present findings with data provided by another English corpus.

References

Baker, Paul. 2010. *Sociolinguistics and Corpus Linguistics*. Edinburgh University Press.

Baker, Paul, Andrew Hardie, and Tony McEnery. 2006. *A Glossary of Corpus Linguistics*. Edinburgh University Press.

'Cambridge Dictionary'. n.d. Cambridge Dictionary. Accessed 10 April 2023. https://dictionary.cambridge.org/.

Čermák, František. 2010. *Lexikon a sémantika*. Nakladatelství Lidové noviny.

'Collins Dictionary'. n.d. Collins Dictionary. Accessed 10 April 2023. https://www.collinsdictionary.com/.

Cruse, D. A. 1986. *Lexical Semantics*. Cambridge University Press.

Cruse, D. A. 2000. *Meaning in Language: An Introduction to Semantics and Pragmatics*. Oxford University Press.

Cruse, D. A. 2001. 'Lexical Semantics'. In *International Encyclopedia of the Social & Behavioral Sciences*, edited by Neil J. Smelser and Paul B. Baltes, 8758–64. Oxford: Pergamon. https://doi.org/10.1016/B0-08-043076-7/02990-9.

Crystal, David. 1994. *The Cambridge Encyclopedia of the English Language*. Cambridge [England], New York: Cambridge University Press.

'Dictionary'. n.d. Dictionary. Accessed 10 April 2023. https://www.dictionary.com/.

Hasa. 2016. 'Difference Between Issue and Problem | Comparison of Definition, Meaning and Usage'. Pediaa. 2016. https://pediaa.com/difference-between-issue-and-problem/.

Hoey, Michael. 2005. *Lexical Priming: A New Theory of Words and Language*. Psychology Press.

Hunston, Susan. 2022. *Corpora in Applied Linguistics*. 2nd edition. Cambridge: Cambridge University Press.

Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, and Vít Suchomel. 2014. 'The Sketch Engine: Ten Years On'. *Lexicography* 1 (1): 7–36. https://doi.org/10.1007/s40607-014-0009-9.

Leech, Geoffrey. 1985. *Semantics: The Study of Meaning*. Second edition. Penguin.

Leech, Geoffrey. 2004. 'Developing Linguistic Corpora: A Guide to Good Practice'. Ahds - Literature, Languages and Linguistics. 2004. https://users.ox.ac.uk/~martinw/dlc/chapter2.htm.

Lindquist, Hans. 2009. *Corpus Linguistics and the Description of English*. Edinburgh University Press Ltd.

Lyons, John. 1996. *Linguistic Semantics: An Introduction*. Cambridge Approaches to Linguistics. Cambridge University Press.

McEnery, Tony, and Andrew Hardie. 2012. *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press.

McEnery, Tony, and Andrew Wilson. 2001. *Corpus Linguistics: An Introduction*. Edinburgh University Press.

'Merriam-Webster Dictionary'. n.d. Merriam-Webster Dictionary. Accessed 10 April 2023. https://www.merriam-webster.com/.

Murphy, M. Lynne. 2010. *Lexical Meaning*. Cambridge Textbooks in Linguistics. Cambridge University Press.

Nedha. 2011. 'Difference Between Problem and Issue'. Difference Between. 25 February 2011. https://www.differencebetween.com/difference-between-problem-and-issue/.

Osherson, Daniel N., Lila R. Gleitman, Mark Liberman, and Edward E. Smith. 1995. *An Invitation to Cognitive Science: Language*. MIT Press.

Palmer, F. R. 1976. *Semantics: A New Outline*. Cambridge University Press.

Robbins, Jill. 2020. 'Problem, Issue or Matter?' Learning English. 24 July 2020. https://learningenglish.voanews.com/a/problem-issue-or-matter-/5512446.html.

Rychlý, Pavel. 2008. 'A Lexicographer-Friendly Association Score', 6–9.

Saeed, John I. 2009. *Semantics*. 3rd edition. Introducing Linguistics.

Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford University Press.

Sinclair, John. 2004. *Trust the Text: Language, Corpus and Discourse*. Taylor & Francis.

Sketch Engine. n.d. Accessed 12 March 2023. https://www.sketchengine.eu/.

Smith, Emma. 2023. 'Difference Between Issue and Problem'. Ask Any Difference. 2023. https://askanydifference.com/difference-between-issue-and-problem/?utm_content=cmp-true.

Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Vol. 6. Studies in Corpus Linguistics. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Thesaurus. n.d. Thesaurus. Accessed 10 April 2023. https://www.thesaurus.com/.

Trask, R. L. 1999. *Language: The Basics (2nd Ed.)*. 2nd edition. Taylor and Francis.

List of Appendices

Appendices

Appendix 1: Word sketch: verbs with *issue* and *problem* as objects

verbs with *issue* as object		
Collocate	**Freq**	**Score**
address	455 002	10.7
resolve	173 651	9.8
discuss	185 918	9.4
raise	196 268	9.3
relate	158 815	9.2
fix	110 999	8.8
solve	79 292	8.5
tackle	62 889	8.3
identify	68 994	7.8
face	55 004	7.6
explore	50 782	7.6
regard	44 239	7.3
highlight	33 003	7.2
cause	55 198	7.2
examine	33 942	7.2
consider	50 229	7.1
understand	39 462	7.1
have	453 303	7.1
involve	43 938	7.0
cover	41 406	7.0

verbs with *problem* as object		
Collocate	**Freq**	**Score**
solve	738 960	11.4
cause	298 150	9.5
fix	166 941	9.1
address	179 416	9.1
resolve	95 466	8.6
have	1 256 033	8.5
face	109 985	8.4
pose	58 623	7.9
tackle	57 445	7.9
experience	68 131	7.8
encounter	50 860	7.7
avoid	61 804	7.7
associate	57 656	7.5
overcome	44 643	7.5
identify	62 548	7.5
be	1 602 703	7.4
correct	35 939	7.2
relate	46 974	7.2
create	98 355	7.2
present	50 853	7.1

verbs with *issue* as object		
Collocate	**Freq**	**Score**
become	64 817	7.0
associate	33 483	7.0
be	1 148 556	6.9
handle	27 833	6.9
debate	21 352	6.9
experience	29 323	6.9
avoid	27 993	6.8
press	21 920	6.8
report	27 768	6.7
underlie	20 873	6.7
emerge	20 795	6.7
remain	23 285	6.5
present	26 677	6.4
investigate	18 019	6.4
decide	15 752	6.4
settle	15 090	6.3
concern	18 760	6.3
encounter	15 032	6.3
bring	28 621	6.1
publish	20 845	6.1
overcome	13 688	6.1
ignore	14 202	6.1
study	16 235	6.1

verbs with *problem* as object		
Collocate	**Freq**	**Score**
discuss	39 256	6.9
report	36 508	6.8
become	59 621	6.7
prevent	31 885	6.6
exacerbate	21 045	6.5
understand	31 000	6.5
underlie	23 087	6.5
eliminate	23 184	6.5
diagnose	20 756	6.5
grow	29 466	6.5
compound	18 406	6.3
think	31 056	6.3
see	73 454	6.3
alleviate	16 761	6.2
treat	21 116	6.2
handle	19 754	6.1
approach	16 440	6.0
remain	19 198	6.0
suffer	16 665	6.0
highlight	16 626	6.0
find	48 068	5.9
describe	18 857	5.8
explain	15 291	5.8

verbs with *issue* as object		
Collocate	**Freq**	**Score**
correct	12 590	6.0
confront	12 379	6.0
clarify	11 885	6.0
take	72 089	5.9
approach	11 958	5.8
mention	12 912	5.8
think	18 435	5.7
explain	11 350	5.6
know	23 310	5.6
frame	8 665	5.5
manage	12 944	5.5
follow	26 157	5.4
analyze	9 094	5.4
review	9 891	5.4
research	8 192	5.4
see	38 177	5.4
force	9 770	5.4
confuse	7 627	5.4
pose	7 372	5.2
complicate	6 720	5.2
escalate	6 529	5.2
put	14 242	5.2
treat	8 739	5.2

verbs with *problem* as object		
Collocate	**Freq**	**Score**
consider	21 624	5.7
remedy	11 090	5.6
develop	24 037	5.6
get	47 693	5.6
ignore	12 302	5.6
indicate	12 932	5.6
detect	12 118	5.5
study	13 506	5.5
notice	11 926	5.5
cure	10 679	5.5
reduce	19 244	5.5
define	13 481	5.4
recognize	11 700	5.4
combat	9 729	5.4
know	21 723	5.3
raise	14 254	5.3
mention	11 162	5.3
discover	10 877	5.3
investigate	10 186	5.3
rectify	8 672	5.3
analyze	9 671	5.2
confront	8 767	5.2
mitigate	8 339	5.2

verbs with *issue* as object		
Collocate	**Freq**	**Score**
prevent	9 877	5.2
include	39 331	5.1
find	25 524	5.1
create	21 390	5.1
surround	7 602	5.0
determine	8 767	5.0
outline	6 452	5.0
describe	9 423	5.0
revisit	5 635	5.0
deal	5 659	4.9
define	7 985	4.9
read	10 269	4.9
eliminate	6 378	4.9
mitigate	5 342	4.9
discover	6 659	4.8
push	6 211	4.8
notice	5 795	4.8
link	6 031	4.8
diagnose	5 152	4.8
say	20 886	4.8
combat	5 111	4.8
cite	5 409	4.7
leave	12 221	4.7

verbs with *problem* as object		
Collocate	**Freq**	**Score**
illustrate	8 394	5.0
lie	7 468	5.0
reveal	8 448	4.9
acknowledge	7 233	4.9
breathe	6 780	4.9
attack	7 459	4.9
recur	6 594	4.8
examine	7 546	4.8
involve	10 304	4.7
bring	12 187	4.7
manage	8 683	4.7
perceive	6 295	4.7
say	20 971	4.7
regard	8 211	4.7
anticipate	5 890	4.6
increase	11 958	4.6
show	12 459	4.6
minimize	5 933	4.6
teethe	5 363	4.6
worsen	5 384	4.6
introduce	7 328	4.6
exist	6 331	4.6
press	5 819	4.5

verbs with *issue* as object		
Collocate	**Freq**	**Score**
get	23 557	4.6
introduce	6 487	4.6
connect	6 025	4.6
communicate	4 507	4.6
view	5 394	4.5
alleviate	4 158	4.5
detect	4 817	4.5
list	5 253	4.5
devote	4 185	4.4
rectify	3 843	4.4
pursue	4 598	4.4

verbs with *problem* as object		
Collocate	**Freq**	**Score**
sort	5 333	4.5
give	19 604	4.5
aggravate	5 083	4.5
spot	5 047	4.4
learn	7 382	4.4
explore	6 484	4.4
repair	4 861	4.4
follow	13 284	4.4
share	7 421	4.3
note	5 237	4.3
believe	5 688	4.3

Appendix 2: Word sketch: verbs with *issue* and *problem* as subjects

verbs with *issue* as subject		
Collocate	**Freq**	**Score**
face	71 633	9.1
relate	58 054	9.0
affect	61 923	9.0
arise	48 845	9.0
surround	50 229	8.8
concern	35 953	8.5
regard	37 319	8.0
involve	28 437	7.6
pertain	14 192	7.5
impact	12 508	7.3
raise	14 212	7.0
include	65 291	6.9
range	11 322	6.9
confront	7 367	6.7
remain	16 195	6.4
cause	10 444	6.1
plague	4 532	6.0
feature	9 501	6.0
occur	8 455	5.9
be	937 207	5.9
become	17 891	5.8
stem	4 061	5.7

verbs with *problem* as subject		
Collocate	**Freq**	**Score**
solve	192 904	10.9
arise	61 775	9.0
face	43 991	8.2
occur	35 431	7.8
lie	25 951	7.6
exist	20 650	7.2
persist	12 543	7.1
affect	19 041	7.1
involve	18 274	6.8
stem	9 912	6.7
get	37 422	6.7
be	1 448 610	6.5
plague	8 169	6.5
cause	15 437	6.5
seem	17 633	6.3
remain	15 231	6.2
relate	9 511	6.2
concern	8 013	6.0
start	16 017	6.0
confront	6 038	6.0
find	15 899	6.0
encounter	5 878	6.0

verbs with *issue* as subject		
Collocate	**Freq**	**Score**
come	24 903	5.7
emerge	4 764	5.7
require	9 238	5.7
seem	9 699	5.6
contain	8 487	5.6
need	12 145	5.6
exist	5 190	5.5
prevent	3 576	5.5
appear	7 809	5.4
go	18 589	5.3
get	13 102	5.3
deal	3 120	5.2
result	3 566	5.2
present	4 775	5.2
persist	2 645	5.2
address	3 310	5.2
have	140 114	5.1
discuss	3 204	5.1
bring	4 980	5.0
lie	3 531	5.0
please	2 969	5.0
focus	3 549	5.0
happen	3 800	5.0

verbs with *problem* as subject		
Collocate	**Freq**	**Score**
require	11 845	5.9
result	6 475	5.8
become	18 297	5.8
use	26 148	5.8
happen	7 295	5.7
regard	8 651	5.7
come	24 936	5.7
go	23 881	5.6
appear	9 761	5.6
emerge	4 989	5.5
include	25 262	5.5
keep	6 309	5.4
begin	11 343	5.4
disappear	3 842	5.2
access	3 609	5.2
do	35 923	5.2
continue	7 511	5.1
develop	5 349	5.1
beset	2 776	5.0
range	3 487	4.9
try	5 595	4.9
have	119 105	4.9
deal	2 887	4.8

verbs with *issue* as subject		
Collocate	**Freq**	**Score**
underlie	2 129	4.9
dominate	2 391	4.8
continue	5 651	4.8
highlight	2 314	4.7
lead	5 246	4.7
do	24 559	4.7
revolve	1 838	4.7
drive	2 823	4.7
cover	3 321	4.6
follow	5 184	4.5
keep	2 899	4.5
divide	1 649	4.5
like	2 383	4.4
surface	1 515	4.4
report	3 206	4.3
start	4 372	4.3
apply	2 071	4.2
use	8 615	4.2
influence	1 683	4.2
make	9 543	4.2
contribute	1 815	4.2
force	1 538	4.2
find	4 171	4.2

verbs with *problem* as subject		
Collocate	**Freq**	**Score**
prevent	2 826	4.8
please	2 948	4.7
lead	5 791	4.7
need	6 913	4.7
understand	2 849	4.6
surround	3 233	4.6
grow	3 905	4.5
set	3 864	4.5
install	2 208	4.5
force	2 256	4.5
run	4 678	4.4
connect	2 369	4.4
follow	5 299	4.4
afflict	1 889	4.4
make	10 577	4.3
surface	1 814	4.3
worsen	1 743	4.3
pay	2 542	4.3
tend	2 383	4.2
present	2 701	4.2
increase	2 596	4.2
turn	2 857	4.1
pose	1 752	4.1

verbs with *issue* as subject		
Collocate	**Freq**	**Score**
tend	1 938	4.2
encounter	1 347	4.2
mean	2 779	4.1
challenge	1 427	4.1
identify	1 705	4.1
hit	1 889	4.1
pend	1 260	4.1
begin	4 194	4.0
note	1 847	4.0
track	1 329	4.0
threaten	1 377	4.0
take	6 538	4.0
deserve	1 263	4.0
head	1 404	4.0
create	2 559	4.0
gain	1 416	4.0
fall	2 201	3.9
explore	1 386	3.9
pose	1 238	3.9
shape	1 161	3.9
depend	1 617	3.9
receive	2 761	3.9
leave	2 542	3.9

verbs with *problem* as subject		
Collocate	**Freq**	**Score**
identify	2 072	4.1
mean	3 080	4.1
create	3 113	4.1
report	3 055	4.1
put	2 527	4.0
let	1 902	4.0
read	1 986	4.0
describe	2 425	4.0
depend	1 833	3.9
manifest	1 294	3.8
take	6 103	3.8
originate	1 323	3.8
call	3 557	3.8
impact	1 372	3.8
like	1 804	3.8
log	1 266	3.8
see	2 872	3.8
reach	1 867	3.8
hit	1 762	3.8
crop	1 194	3.8
raise	1 684	3.7
occure	1 178	3.7
stay	1 488	3.7

verbs with *issue* as subject		
Collocate	**Freq**	**Score**
play	2 890	3.9
reflect	1 351	3.8
crop	962	3.8
put	1 813	3.8
matter	997	3.8
effect	947	3.8
hamper	938	3.7
see	2 526	3.7
limit	1 078	3.7

verbs with *problem* as subject		
Collocate	**Freq**	**Score**
sleep	1 250	3.7
figure	1 237	3.7
bring	2 316	3.7
apply	1 696	3.7
delay	1 195	3.7
escalate	1 183	3.7
downloading	1 149	3.7
pertain	1 235	3.7
resolve	1 163	3.7

Appendix 3: Word sketch: Modifiers of *issue* and *problem*

modifiers of *issue*		
Collocate	**Freq**	**Score**
health	230 968	8.9
important	162 306	8.4
key	135 072	8.3
environmental	104 963	8.2
major	113 450	7.8
legal	88 488	7.8
social	114 269	7.7
serious	73 567	7.7
safety	69 141	7.6
security	77 097	7.6
special	97 065	7.6
critical	67 925	7.6
policy	65 455	7.6
mental	67 292	7.6
complex	65 136	7.6
big	112 412	7.5
related	56 578	7.4
ethical	48 174	7.3
technical	57 309	7.3
other	268 013	7.3
real	76 069	7.2
current	74 276	7.2

modifiers of *problem*		
Collocate	**Freq**	**Score**
health	250 239	9.1
serious	171 107	9.1
big	213 400	8.6
major	134 862	8.2
real	134 725	8.2
common	89 326	7.9
only	104 404	7.8
complex	62 731	7.8
mental	62 091	7.7
potential	62 328	7.6
same	147 002	7.6
main	72 559	7.3
similar	50 611	7.2
technical	45 759	7.2
environmental	45 443	7.2
social	64 033	7.0
difficult	30 334	6.9
heart	30 621	6.8
significant	41 147	6.8
fundamental	29 330	6.8
severe	29 326	6.8
financial	43 702	6.8

modifiers of *issue*		
Collocate	**Freq**	**Score**
political	63 558	7.1
right	42 375	7.0
specific	56 535	7.0
main	55 626	6.8
same	86 631	6.7
such	62 743	6.7
controversial	28 992	6.6
various	47 403	6.5
global	41 156	6.5
potential	34 657	6.5
many	98 101	6.5
economic	36 380	6.4
particular	35 345	6.4
significant	35 607	6.4
performance	27 820	6.4
common	35 339	6.3
similar	31 762	6.3
minor	25 436	6.3
contemporary	25 230	6.3
fundamental	24 195	6.3
contentious	21 677	6.3
gender	23 308	6.2
pressing	21 344	6.2

modifiers of *problem*		
Collocate	**Freq**	**Score**
huge	36 452	6.8
medical	40 586	6.8
specific	40 563	6.7
many	104 711	6.6
other	162 695	6.6
economic	34 757	6.6
particular	32 952	6.5
skin	21 144	6.4
such	43 281	6.3
related	22 163	6.3
drug	22 361	6.3
psychological	18 300	6.2
behavioral	16 757	6.2
practical	19 544	6.1
minor	18 636	6.1
real-world	16 115	6.1
respiratory	16 188	6.1
chronic	17 308	6.1
optimization	15 404	6.1
structural	16 689	6.1
math	15 641	6.1
legal	22 901	6.0
security	21 511	6.0

modifiers of *issue*		
Collocate	**Freq**	**Score**
management	28 593	6.2
privacy	22 332	6.2
several	45 590	6.2
only	38 619	6.2
public	43 353	6.2
difficult	22 829	6.2
moral	22 272	6.1
medical	29 603	6.1
quality	25 879	6.1
sensitive	20 600	6.1
justice	19 955	6.1
late	34 652	6.0
bond	18 530	6.0
first	73 329	6.0
regulatory	19 213	5.9
central	22 538	5.9
relevant	19 741	5.9
human	29 685	5.8
compliance	16 542	5.8
water	24 317	5.8
personal	25 393	5.8
back	17 765	5.8
next	31 225	5.7

modifiers of *problem*		
Collocate	**Freq**	**Score**
global	24 748	6.0
more	58 321	6.0
behavior	14 088	5.9
emotional	15 600	5.9
sleep	13 963	5.9
basic	19 018	5.9
few	37 482	5.8
ongoing	15 220	5.8
own	38 311	5.8
possible	17 253	5.8
current	25 515	5.8
vision	12 508	5.7
important	22 457	5.7
physical	17 783	5.7
performance	14 485	5.7
mechanical	12 921	5.7
obvious	12 643	5.7
business	22 253	5.7
memory	12 898	5.6
challenging	11 780	5.6
debt	12 049	5.6
various	22 484	5.6
alcohol	11 665	5.6

modifiers of *issue*		
Collocate	**Freq**	**Score**
financial	23 774	5.7
development	19 373	5.7
core	16 444	5.7
crucial	15 858	5.7
unresolved	14 986	5.7
few	37 531	5.7
national	23 873	5.7
whole	23 941	5.7
certain	23 057	5.7
outstanding	16 277	5.7
broad	17 171	5.6
cultural	17 899	5.6
tax	18 838	5.6
recent	21 921	5.6
practical	15 989	5.6
ongoing	15 173	5.5
local	30 048	5.5
compatibility	12 857	5.5
international	20 686	5.5
emotional	14 266	5.5
constitutional	13 249	5.5
immigration	12 915	5.4
life	15 472	5.4

modifiers of *problem*		
Collocate	**Freq**	**Score**
world	15 187	5.6
personal	19 233	5.6
water	17 994	5.6
systemic	10 843	5.5
digestive	10 679	5.5
eye	11 660	5.5
several	25 814	5.5
political	19 080	5.5
hard	13 846	5.5
back	11 779	5.4
pressing	9 881	5.4
design	13 067	5.4
engineering	11 063	5.4
quality	13 641	5.4
kidney	9 841	5.4
communication	11 691	5.4
long-term	11 447	5.4
pollution	9 619	5.4
traffic	10 540	5.4
word	10 515	5.3
persistent	9 371	5.3
safety	11 332	5.3
drinking	9 251	5.3

modifiers of *issue*		
Collocate	**Freq**	**Score**
single	19 659	5.4
governance	12 136	5.4
sustainability	12 021	5.4
different	29 091	5.3
family	17 082	5.3
huge	15 202	5.3
second	23 586	5.3
climate	14 031	5.3
complicated	11 277	5.3

modifiers of *problem*		
Collocate	**Freq**	**Score**
numerous	11 373	5.3
little	19 517	5.3
key	14 087	5.3
further	13 367	5.3
immediate	10 281	5.3
mathematical	9 207	5.3
relationship	9 062	5.3
abuse	9 017	5.3
dental	9 473	5.3