UNIVERZITA PALACKÉHO V OLOMOUCI

Pedagogická fakulta v Olomouci

BAKALÁŘSKÁ PRÁCE

Olomouc 2018 Lukáš Váňa

UNIVERZITA PALACKÉHO V OLOMOUCI

Pedagogická fakulta v Olomouci

Ústav cizích jazyků

BAKALÁŘSKÁ PRÁCE

Lukáš Váňa

Využití jazykových korpusů ve výuce anglického jazyka

Using of language corpora in English language teaching

Olomouc 2018

Školitel bakalářské práce

Doc. PhDr. Václav Řeřicha, CSc.

Prohlášení

Čestně prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně, za použití uvedené literatury a pramenů.

V Olomouci dne 21. 6. 2018 ……

vlastnoruční podpis

Abstract

The bachelor's thesis deals with the usage of language corpora in teaching English. In the theoretical part, corpus linguistics as one of the modern linguistic branches and language corpora, their development, types and construction are introduced. The issues of corpora in English lessons are also presented. In the practical part, working sheets for teachers are created.

Key words: Corpus, corpus linguistics, worksheets, teaching

Abstrakt

Bakalářská práce se zabývá využitím jazykových korpusů ve výuce anglického jazyka. V teoretické části je představena korpusová lingvistika, jakožto jedno z moderních odvětví jazykovědy, a jazykové korpusy, jejich vývoj, druhy a výstavba. Je také představena problematika korpusů ve výuce angličtiny. V praktické části jsou vytvořeny pracovní listy pro učitele.

Klíčová slova: Korpus, korpusová lingvistika, pracovní listy, výuka

CONTENT

INTRODUCTION

Today's word is surrounded by information and communication technologies, developing by mile steps, and the languages are no exception. Modern technologies in language teaching are commonly used at any type of schools today. Classrooms are equipped by computers, overhead projectors, interactive boards etc. Teachers can arrange computer classrooms (when the lesson is not directly held there) for teaching English to make it more interesting. There are also various educational programmes, e-learnings etc. to make the lessons innovative.

During my university studies in Opava I first encountered corpus linguistics – a discipline which development relates to computers and information technologies. I quickly realized that this is the way I wanted to set out. A connection of modern technologies and my liking for English (especially language teaching methodology) led to a deeper interest how to implement corpus-based activities in the English language teaching.

The aim of the bachelor's thesis is to propose and create self-made corpus-based worksheets for teachers.

I am sure that many teachers have no idea what corpus linguistics actually means, that is why it is necessary to reach this aim gradually.

First of all, it is necessary to outline the issue of this linguistic branch. The theoretical part, which forms about half of the thesis, serves as a small introduction to corpus linguistics and its implementation in the English language teaching. For this purpose, various books and internet websites of this branch were read up. Origin and conception of corpus linguistics is presented first. Corpora as the main unit of corpus linguistics are then defined and divided. The most important ones are introduced. The usage of corpora in the English language teaching is indicated finally.

The practical part, which forms the second part of the thesis, proposes corpus-based worksheets to be used in the English lessons at grammar schools. These worksheets were made on an individual basis. There were four principles according to which the worksheets were made. The first one – each worksheet represents a specific topic, typical for English. These topics correspond somehow with the ones commonly used for the English leaving exam. It is not possible to cover all the topics that is why I chose only some of them. The second principle – the worksheets have the same types of exercises but different assignments.

The third principle – the right answers are quoted at the end of the thesis as an appendix, not in the worksheets themselves. The fourth and the last principle – any of the worksheets should not serve as a written exam and therefore there are no points written with single tasks so that teachers could choose how to evaluate them. Computer or notebook with the internet connection is necessary to be able to fill in the worksheets. The materials are not focused on certain grammatical features, the tasks correspond somehow with the given topic.

The thesis should only bring the basic outline of language corpora. This issue is much broader, and its detailed explanation could confuse people who are going to encounter this linguistic branch for the first time. For more information, people can search various publications either discussing general questions of corpus linguistics or specific topics of this branch.

I expect the thesis to increase an interest in modern teaching methods of languages among teachers. Of course, it is necessary to use traditional teaching methods sometimes, but it is still harder to engage students' attention, that is why modern ways of teaching English are a good option how to attract them. I believe some teachers are possibly going to be interested in them more and they could search further how to make their lessons much more interesting.

1. ORIGIN AND CONCEPTION OF CORPUS LINGUISTICS

1.1 Origin of corpus linguistics

Bennett (2010, p. 2) briefly describes the origins and development of corpus linguistics:

> "The principles of corpus linguistics have been around for almost a century. Lexicographers, or dictionary makers, have been collecting examples of language in use to help accurately define words since at least the late 19th century. Before computers, these examples of language were essentially collected on small slips of paper and organized in pigeon holes. The advent of computers led to the creation of what we consider to be modern-day corpora. The first computer-based corpus, the Brown corpus, was created in 1961 and comprised about 1 million words. Today, generalized corpora are hundreds of millions of words in size, and corpus linguistics is making outstanding contributions to the fields of second language research and teaching."

One of the most interesting ideas is that corpus linguistics is not a new discipline because the origins are almost one hundred years old. In fact, we have it mostly connected with a rapid growth of information and communication technologies in recent years, which enables to store big amount of data electronically, but its origin is older than any computer has ever been made. It means that corpus linguistics did not develop from computers or any other technologies, these devices only enabled to extend the branch of linguistics which had been known before.

1.2 Conception of corpus linguistics

There are many authors describing the conception of corpus linguistics nowadays.

Bennett (2010, p. 2) states that corpus linguistics approaches the study of language in use through corpora, which is a large, principled collection of naturally occurring examples of language stored electronically. According to Bennett (2010, p. 2), corpus linguistics serves to answer two fundamental research questions:

a) "What particular patterns are associated with lexical or grammatical features?"

b) "How do these patterns differ within varieties and registers?"

As Bennett (2010, p. 2, 3) further states, it is important not to understand only what corpus linguistics is, but also what corpus linguistics is not.

a) "Corpus linguistics is not able to provide negative evidence."

Bennett (2010, p. 3) explains that a corpus can't tell us what's possible or correct or not possible or incorrect in language; it can only tell us what is or is not present in the corpus. Many instructors should therefore believe that if a corpus does not present a particular manner to express a certain idea, then the corpus is not faulty because the manner is perhaps not very common in the register represented by the corpus.

b) "Corpus linguistics is not able to explain why."

Bennett (2010 p. 3) states that it can't tell us why something is the way it is, only tell us what is. To find out why, we, as users of language, use our intuition.

c) "Corpus linguistics is not able to provide all possible language at one time."

According to Bennett (2010, p. 3), corpus cannot be a representative of all language, no matter how planned, principled or large it is. This means that even in a corpus containing one billion words, all instances of use of a language may not be present.

On the other hand, Nadja Nesselhauf (2011, p. 2), an associate professor in Heidelberg university, describes corpus linguistics as the analysis of naturally occurring language based on computerized corpora, which is performed with the help of the computer, i.e. with specialised software, and considers the frequency of the phenomena investigated. It is evident that Nesselhauf takes into account a contemporary approach of corpus linguistics based on information technologies.

Becoming aware of corpus linguistics, we may be wondering what "corpus" (plural "corpora") in linguistics means. This term is explained in the next chapter.

2. LANGUAGE CORPUS

2.1 Definition of the word "corpus"

To understand corpus linguistics more, it is necessary to explain the term "*corpus*".

According to Nadja Nesselhauf (2011, p. 2), *"A corpus can be defined as a systematic collection of naturally occurring texts (of both written and spoken language)."*

Nesselhauf (2011, p. 2) explains that *"systematic"* means that the structure and contents of the corpus follows certain principles and that information on the exact composition of the corpus is available to the researcher. She also states that although corpus can refer to any systematic text collection, it is often used today only to refer to systematic text collections that have been computerized.

Definition of corpus also presents Bennett (2010, p. 12) who states that "*A corpus is a principled collection of authentic texts stored electronically that can be used to discover information about language that may not have been noticed through intuition alone."*

It is evident from both the definitions that corpus is a collection of texts organized systematically. These texts are organized according to clues so that we can easily search them. According to the types of texts forming a certain corpus we distinguish several types of language corpora, some of them are mentioned further in the thesis.

A bit different approach brings Aston (1997, p. 4) who does not restrict only to one definition of "corpus" but he renders two of them taken from the Oxford English Dictionary particularly referring to language.

a) Corpus is *"a body or collection of writings or the like; the whole body of literature on any subject"*.

Aston (1997, p. 4) explains this definition on the example of the "Shakespearean corpus", meaning the entire collection of texts by Shakespeare.

b) Corpus is *"the body of written or spoken material upon which a linguistic analysis is based"*.

According to Aston (1997, p. 4) this is the sense of the word which the phrase "corpus linguistics" comes from.

If we think about both the definitions we find out the they can overlap because any author's work can be subject of a linguistic analysis.

2.2 The beginnings and periods of electronic corpus

Meyer (2004, p. 1) mentions that the beginnings of electronic corpus were not easy. Even though the creators of the first machine-readable English corpus W. Nelson Francis and Henry Kučera are nowadays regarded as pioneers and visionaries in the corpus linguistics community, the Brown corpus was not warmly excepted by many members of the linguistics community at that time. The reason was that generative grammar dominated linguistics and there was little tolerance for approaches to linguistic study that did not adhere to the rules of this branch of linguistics.

"*Generative grammar is a precisely formulated set of rules whose output is all (and only) the sentences of a language - i.e., of the language that it generates."* (Encyclopaedia Britannica, 1998)

Francis (1992, p. 28) speaks about a leading generative grammarian of the time characterizing the creation of the Brown Corpus as *"a useless and foolhardy enterprise"* because *"the only legitimate source of grammatical knowledge"* about a language was the intuitions of the native speaker, which could not be obtained from a corpus.

According to Meyer (2004, p. 1) some linguists still believe this, but many linguists of all persuasions are now far more open to the idea of using linguistic corpora for both descriptive and theoretical studies of language. Many corpus linguists are actively engaged in language theory, and many generative grammarians have shown an increasing concern for the data upon which their theories are based, even though data collection remains a marginal concern in modern generative theory.

The relationship between generative grammar and corpus linguistics is far more complicated and is not explained in detail as it is not the issue of the thesis.

Čermák (2011, p. 13, 14) says that there have been three periods of of language corpora development:

a) 1960s – 1980s

This period is characterized by gaining the experience how to build first one million-corpora.

Čermák (2011, p. 12) explains that the beginnings were plain and slow and were motivated by a desire to better understand a language. First corpora arose in the English-speaking area.

b) 1980s – 2000

According to Čermák (2011, p. 12) it is the period of increasing the size of corpora to approximately twenty million words due to gaining data by first scanners. It was later (in the 1990s) replaced by texts from electronic typesetting and the size increased in some cases to hundred million of words.

c) 2000 and more

Čermák (2011, p. 12) states that new texts which have never been printed have arrived in corpora (e.g. from the internet), the size has increased to billions of words recently.

2.3 Types of corpora

There are many types of corpora mentioned in various publications. It is important to realize what you are going to look for to know what kind of corpus you should consult.

Bennett (2010, p. 13) states that there are approximately eight types of corpora – generalized, specialized, learner, pedagogic, historical, parallel, comparable and monitor and that the first four mentioned types of corpora are the most useful to be used in the language teaching. These four corpora are explained more in details.

a) Generalized corpora

Bennett (2010, s. 13) explains that this is the broadest type of corpus, often very large, containing more than 100 million words and variety of language. This type of corpus contains written texts such as newspapers, magazine articles or works of fiction and nonfiction and spoken transcript such as informal conversations, government proceedings or business meetings. British National Corpus (BNC) and Corpus of Contemporary American English

(COCA) are the examples of generalized corpora. This type of corpus should be consulted if generalizations about language as a whole are to be drawn.

b) Specialized corpora

Bennett (2010, p. 13) states that *"A specialized corpus contains texts of a certain type and aims to be representative of the language of this type."* According to Bennett (2010, p. 13) these corpora are often created to answer very specific questions. The examples can be CHILDES corpus, containing language used by children, or a medical corpus containing language used by nurses and hospital staff.

c) Learner corpora

Bennett (2010, p. 14) explains that this is a kind of specialized corpus consisting of written texts or/and spoken transcripts of language used by students who are acquiring the language. One of the best-known learner corpus is the International Corpus of Learner English containing essays written by English language learners with 14 different native languages.

d) Pedagogic corpora

According to Bennett (2010, p. 14) *"A pedagogic corpus is a corpus that contains language used in classroom settings."* He further states that it can include e.g. academic textbooks or transcripts of classroom interaction and can be used to ensure students are learning useful language, or to examine teacher-student dynamics.

2.4 Questions we can answer with corpora

According to Bennett (2010, p. 4) corpus linguistics looks to see what patterns are associated with lexical and grammatical features. Corpora searching provides answers to questions like these:

a) What are the most frequent words and phrases in English?

b) What are the differences between spoken and written English?

c) What tenses do people use most frequently?

d) What prepositions follow particular verbs?

e) *How do people use words like can, may, and might?*

f) *Which words are used in more formal situations and which are used in more informal ones?*

g) *How often do people use idiomatic expressions?*

h) *How many words must a learner know to participate in everyday conversation?*

i) *How many different words do native speakers generally use in conversation?*

Bennett (2010, p. 4) explains that for the most part, these questions don't look particularly revolutionary. We already know the answers to many of them. We can open up almost any grammar, vocabulary, conversation, or writing textbook and find the answers.

Bennett (2010, p. 5) further states that corpora do not contain the same number of words, so we cannot use a simple frequency count to see in which corpus a word is more common. He shows the issue on the example of the word *very* which occurs in the spoken portion of the COCA 195,000 times and in the written portion of the COCA 198,000 times. We might conclude from looking only at the simple frequency count that *very* is used only slightly more in written language. However, as the written portion of the COCA is much larger than the spoken one, we can only get an accurate comparison by calculating how many times *very* occurs per million words. This is called the normed count.

2.5 A small view into language corpora creation

Development of any corpora is a long-lasting process which contains planning a corpus creation, collecting, computerizing, annotating, and analysing data. In the thesis I focus on the first step, planning a corpus creation. This step includes some major issues. Meyer (2004, p. 30) mentions several of them:

a) *The overall length of a corpus*

Meyer (2004, p. 32) states that earlier corpora were relatively short mainly because of logistical difficulties. This process has been eased with the invention of optical scanners. Nevertheless, the technology has not progressed to the point where it can highly expedite the collection of speech. That is why, for example, 90 % of the BNC is writing and only 10 % speech.

Meyer (2004, p. 32-33) clarifies that when determining a length of a corpus, it is first important to compare the resources that will be available to create it with the amount of time it will take to collect texts for inclusion, computerizing, annotating, and parsing. The next step is to consider the length of a corpus to permit the kinds of studies one envisions for it.

Meyer (2004, p. 30) proposes that *"in general, the lengthier the corpus, the better"*.

b) The types of genres to include in a corpus

Meyer (2004, p. 34-38) explains this issue on the comparison of the BNC and the International Corpus of English (ICE). These corpora are different in the composition but represent similar genres of spoken and written English. There is an important methodological question: why these genres and not the others? This question can be answered according to types of corpora and the particular kind of studies which we can carry out on them. As both the mentioned corpora are multi-purpose corpora, intended to be used for many different purposes, they contain a broad range of genres.

According to Meyer (2004, p. 34-38), on the other hand, general corpora do not always contain a full representation of a genre, therefore special-purpose corpora are being developed such as the Michigan Corpus of Academic Spoken English.

c) The length of individual text samples to be included in a corpus

Meyer (2004, p. 38-40) presents that corpora vary in terms of the length of the individual text samples that they contain. Most corpora tend to contain text fragments rather than complete texts as there are numerous logistical obstacles that make the inclusion of complete texts in corpora nearly impossible (e.g. books which are quite lengthy). Nevertheless, it is possible to take excerpts that themselves form a coherent unit such as in the American component of ICE and many other earlier corpora as well.

d) Determining the number of texts and range of speakers and writers to include in a corpus

Meyer (2004, p. 44-45) demonstrates several factors that need to be considered prior to collecting texts for inclusion in a corpus. I chose four of them:

- Lengthier corpora are better than shorter one but more important than the sheer length of a corpus is the range of genres included within it.

- The range of genres to be included is determined by whether it will be a multi-purpose corpus or a special-purpose one.
- It is more practical to include text fragments rather than complete texts.
- The more variation exists in a genre, the more samples of a genre are needed.

e) The time-frame for selecting speakers and texts

According to Meyer (2004, p. 45-46) most corpora contain samples of speech or writing that have been written or recorded within a specific time-frame.

Meyer (2004, p. 45-46) further develops this issue on the example of synchronic and diachronic corpora. In creating a synchronic corpus, the time-frame should be narrow enough to provide an accurate view of contemporary English undisturbed by language change. However, linguists disagree about whether purely synchronic studies are even possible: new words come into the language every day, indicating that language change is a constant process. With diachronic corpora which are used to study historical periods of English, the time-frame for texts is quite easy to determine, since the various historical periods of English are fairly well defined. However, it is important not only to cover predetermined historical periods of English but also to think through how significant events can be best covered in the particular corpus.

f) Sampling native vs. non-native speakers of English

According to Meyer (2004, p. 46-48) it is necessary to distinguish whether a language in particular country is first or second (additional). For example, corpus of American English would be best formed by speech or writing of native speakers of American English as United States is a country where English is a first language. On the other hand, corpus of Nigerian English should not be formed by native speakers of English residing in Nigeria, as it is a variety of English spoken primarily a second (additional) language.

Meyer further clarifies (2004, p. 46-48) that there were specific criteria established for selecting the individuals whose writing or speech will be included in the corpus. These criteria are, for example:

- How many years an individual has used English
- In what contexts they have used English

- How much education in English they have had

g) Controlling for sociolinguistic variables

According to Meyer (2004, p. 48-53) there are sociolinguistic variables that will need to be considered before selecting the speakers and writers whose texts are supposed to be included in a corpus. Some of them apply to the collection of both spoken and written texts; others are more particular to spoken texts.

These variables are:

a. General balance

b. Age

c. Level of education

d. Dialect variation

e. Social contexts and social relationships

It is clear, that planning a corpus creation is a long-lasting effort with many steps which need to be observed. As I have already mentioned it is only the beginning of the whole process the research of which would require much more space and thus is not mentioned here in detail.

2.6 A short description of the British National Corpus and the Czech National Corpus

2.6.1 British National Corpus

The British National Corpus (BNC) is a 100-million-word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English from the later part of the 20th century, both spoken and written. The latest edition is the *BNC XML Edition*, released in 2007.

The written part of the BNC (90%) includes, for example, extracts from regional and national newspapers, academic books and popular fiction or school and university essays. The spoken part (10%) consists of orthographic transcriptions of unscripted informal conversations (recorded by carefully selected volunteers) and spoken language collected in

different contexts, ranging from formal business or government meetings to radio shows and phone-ins.

The BNC project was carried out and is managed by the BNC Consortium**,** an industrial/academic consortium led by Oxford University Press. Work on building the corpus began in 1991 and was completed in 1994. Making the BNC was a joint effort of a large number of participants; organisations and individuals. No new texts have been added after the completion of the project, but the corpus was slightly revised prior to the release of the second edition *BNC World* (2001) and the third edition *BNC XML Edition* (2007).

The XML Edition of the BNC contains 4049 texts. In total, it comprises just under 100 million orthographic words.[1]

The BNC enables two possibilities to access. The first one is simple search from the website http://www.natcorp.ox.ac.uk/, where you do not log in, the second one is more advanced search available at https://corpus.byu.edu/bnc/ where you must sign up and then log in every time you want to enter the corpus. The second website enables to research various sections (academic, newspapers etc.) and the results are also recounted in millions which makes it possible to compare with the occurrence of the same word in other corpora.

2.6.2 Czech National Corpus

According to Čermák (2004, p. 154) Czech National Corpus (CNC) is a project which products map and monitor different forms of language to make language data and particular tools for their usage accessible. The project is academic, non-commercial and opened to all interested people. CNC is being developed by the Institute of Czech National Corpus at the Charles University, led by František Čermák, and was established in 1994. The Institute of Czech National Corpus is responsible for building and further development of CNC and creating its own methodology. CNC is the main name for several entities and segments developed from electronic texts of a different specialization and range.

The access to the corpus is available via https://www.korpus.cz/. It is possible to use a simple search, or you can sign up and log in to unlock all the functions.

[1] British National Corpus, [cit.14.3.2018]. Dostupné z URL: http://www.natcorp.ox.ac.uk/

2.7 Small information about other English language corpora

Apart from the British National Corpus there are many other English language corpora of various types, varieties etc. A choice of some of them is listed in the following table:

Corpus	Variety/ies	Time (span)	Number of words	Text type(s)
Large general corpora				
COCA	AmE	1990-2011	425 million	various spoken and written text types
American National Corpus	AmE	PDE	20 million	written and spoken
Comparable corpora of written British and American English				
The Lancaster-Oslo/Bergen Corpus (LOB)	BrE	1961	1 million	2000-word samples from different written text types
Brown	AmE	1961	1 million	same as LOB
Corpora of other varieties				
Wellington Corpus of written New Zealand English	NZE	1986-1990	1 million	same as LOB
The Australian Corpus of English	AuE	1986	1 million	same as LOB
Other regional corpora – International corpus of English (ICE)				
ICE-New Zealand	NZE	1989-1994	1 million	different spoken and written genres
ICE-Singapore	SingE	PDE	1 million	different spoken and written genres
ICE-India	IndE	PDE	1 million	different spoken and written genres
ICE-Philippines	PhiE	PDE	1 million	different spoken and written genres
ICE-Canada	CanE	PDE	1 million	different spoken and written genres

ICE-Hong Kong	HKE	PDE	1 million	different spoken and written genres
ICE-Ireland	IrE	PDE	1 million	different spoken and written genres
Spoken corpora				
Corpus of London Teenage Language	BrE	1991	500,000	spontaneous spoken language of London teenagers
The Michigan Corpus of Academic Spoken English	AmE	1997-2001	1,8 million	academic speech (native and non-native)
Historical corpus				
Helsinki Corpus (diachronic part)	BrE	750-700	1,5 million	different text types
Learner corpus				
International Corpus of Learner English	learner English	1990s	2,5 million	argumentative essays written by learners with different languages
Corpus of child language				
Polytechnic of Wales Corpus	BrE	1978-1984	65,000	speech of children between 6 and 12 (spontaneous and interviews

Table 1 A choice of English language corpora, my own product according to data by Nesselhauf (2011)

3. CORPORA IN ENHLIGH LESSONS

In previous chapters there was theoretical basis for corpora and corpus linguistics outlined. As I am aware of the fact that many teachers do not know much (if anything at all) about this issue it was necessary to break into it first. Definitions, history, types of corpora etc. were presented. This chapter is dedicated to the ways of preparing corpus-designed activities.

Many teachers use dictionaries or course books in their language teaching every day. They know how to work with them, they can connect their usage with e.g. electronic devices, but they are not aware of the fact that these are actually corpus-based materials. Nevertheless, self-made published materials incorporating language corpora in the English lessons are not very common. This thesis is supposed to bring several worksheets to be used when teaching English.

3.1 Corpus use in the classroom today

Garcia (2012, p. 37) explains that even though the electronic corpora have become more and more accessible to amateurs over the past decade, the direct use of corpora still seems to be far away from becoming constant part of language teaching. According to the only survey of using corpora in schools in 2004, the biggest problem is that almost 80% of secondary school teachers (who were interviewed) have never heard of corpora.

According to Garcia (2012, p. 38) it is necessary to held workshops, talks and conferences for teachers (rather than for corpus linguists). Teachers at schools must be aware of the fact that it is not necessary to install any complicated software or pay for any access. Even though there are various introductions or tutorials, not many teachers seem to have found about corpora for themselves as it is more practical for them to use traditional language materials.

3.2 A framework for creating corpus-designed activities

Corpus-designed activities enable to make a lesson much more interesting. One of the ways of integrating these activities into the lessons is to prepare them on our own. We can find these materials on the internet (if we have a good luck, as they are not very common), but

they do not have to be suitable for our particular purpose. It is then better to create them by ourselves.

Bennett (2010, p. 18-21) describes seven steps of creating corpus-designed activities. These steps are:

a) Ask a research question

The "research question" for a corpus-designed activity could be for example: What is the difference between *through* and *between*?

b) Determine the register on which your students are focused

According to Bennett (2010, p. 11) "*register is defined as situation of use*". If we create a corpus-based activity, it is necessary to know which register is relevant for students. If they should practise informal conversation, a corpus of academic papers will not be helpful.

c) Select a corpus appropriate for the register

Bennett (2010, p. 11) states that the most important thing is if the corpus contains authentic language for real-life communication.

d) Utilize a concordancing programme for quantitative analysis

Bennett (2010, p. 11) describes that a concordancing programme must be used to access the language stored in the corpus. We have to be sure to understand all the search function.

e) Engage in qualitative analysis

According to Bennett (2010, p. 11) qualitative analysis as the last step of corpus approach will answer the question *Why?* For qualitative analysis, we take the quantitative information given by the corpus concordancing programme and determine its significance.

f) Create exercises for students

Bennett (2010, p. 11) presents that a central element of corpus-based activities is traditional fill-in-the-blanks and gap-fill activities.

g) Engage student in a whole-language activity

Bennett (2010, p. 11) indicates that it is important to realize that the steps of the process may not take place specifically in this order. We may start either with a research question, a corpus, a register, or even a whole language activity.

Bennett (2010, p. 20) describes that these steps can be applied mostly for an intermediate to high-intermediate level of English. We can modify corpus-designed activities for low language level in these steps:

a) Ask simple research question

Bennett (2010, p. 20) presents an example of simple activity for low language level which could be a wrong usage of the verb *to have* instead of *to be* to tell how old someone is.

b) Find your concordance line

According to Bennett (2010, p. 21) it is better to investigate in materials around us such as newspapers, magazines, textbooks than to utilize a whole corpus of authentic English.

c) Adapt lines for students' level

Bennett (2010, p. 21) suggests that the materials for corpus-designed activities could be modified for sentence structure and vocabulary but the main features must stay intact. He gives a following example:

"The two disciplines do not appear on the surface to have very much in common. Historically, though, anthropologists and epidemiologists have worked together for a very long period of time."

"The study of culture and the study of people's health do not seem similar, though they have been studied together for a very long time."

d) Present students with fewer lines

Bennett (2010, p. 21) states that it is necessary for students to have 20-50 lines, 10 lines may be enough for lower-level students.

e) Encourage group or whole class work

According to Bennett (2010, p. 21) another great way to complete a corpus-designed activity is a whole class work without being too challenging for the students' level.

3.3 Possibilities in searching the British National Corpus

British National Corpus, which must be used when filling in the worksheets, has two accesses as already mentioned above in the thesis. Possibilities in searching both websites are now explained more in details with an example as a screenshot.

a) A simple search of the BNC

The website http://www.natcorp.ox.ac.uk/corpus/index.xml enables to access the British National Corpus (BNC) for a simple search without a necessity to be signed and logged in. This simple access does not allow to search various sections (e.g. academic, newspapers). You can sort out the search results according to relevance (e.g. date, title) or you can pick the results from a particular year only. There is also an "advanced" search which enables to sort out the results according to content or location. At the bottom of the page there are *"related searches"* for the particular word such as types of the word (e.g. room – seminar room) or filetype. The number of searches is not recounted in millions.

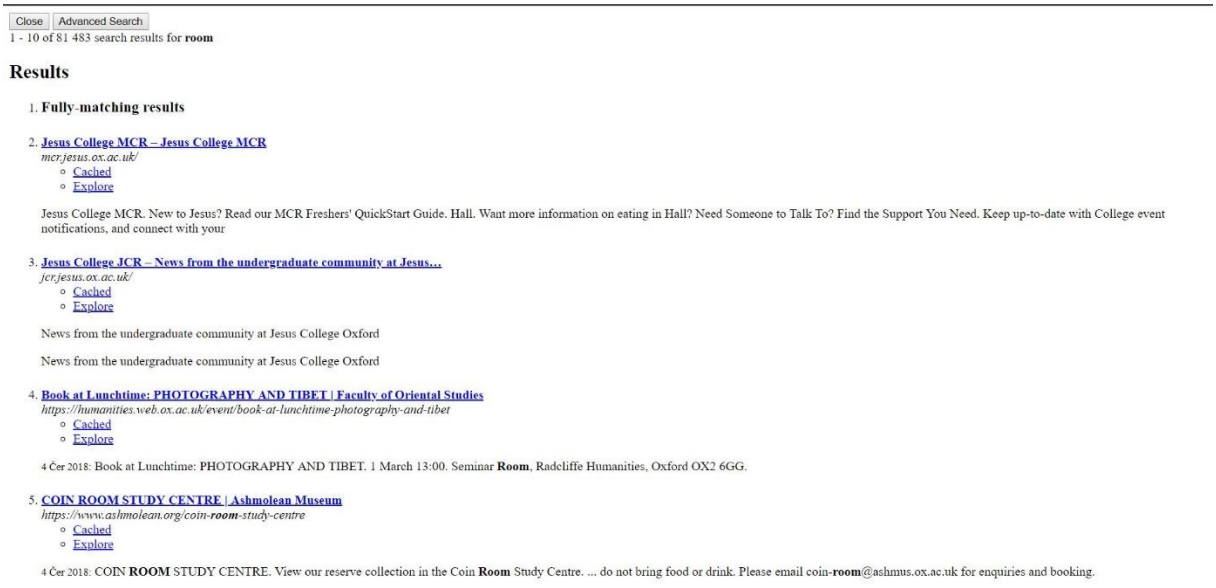


Figure 1An example of a simple search, adopted from http://www.natcorp.ox.ac.uk/corpus/index.xml (2018)

b) A more advanced search of the BNC

The second website https://corpus.byu.edu/bnc/ is for more advanced search where you must sign up and log in first to get the access there. It is more practical to create a fictitious user name when you want to work with this corpus in your lessons. It would be in fact rather impractical and time consuming when all your pupils or students should create their own account during lesson unless they do it at home what I do not suppose. The full search enables to get a chart (either vertical or horizontal – you can choose) sorting the relevance of a particular word in all sections. The results are recounted in millions which enables to compare them with those of the same word at any other corpora where the results are recounted as well. You can click the sections to find out more information. Apart from the words themselves you can also find collocations in the corpus. The example could be "*room – living"* which indicates either a living room, a part of a house or a flat, or living as a present participle (e.g. I was living in a room …). There is also a possibility to compare two or more words together.

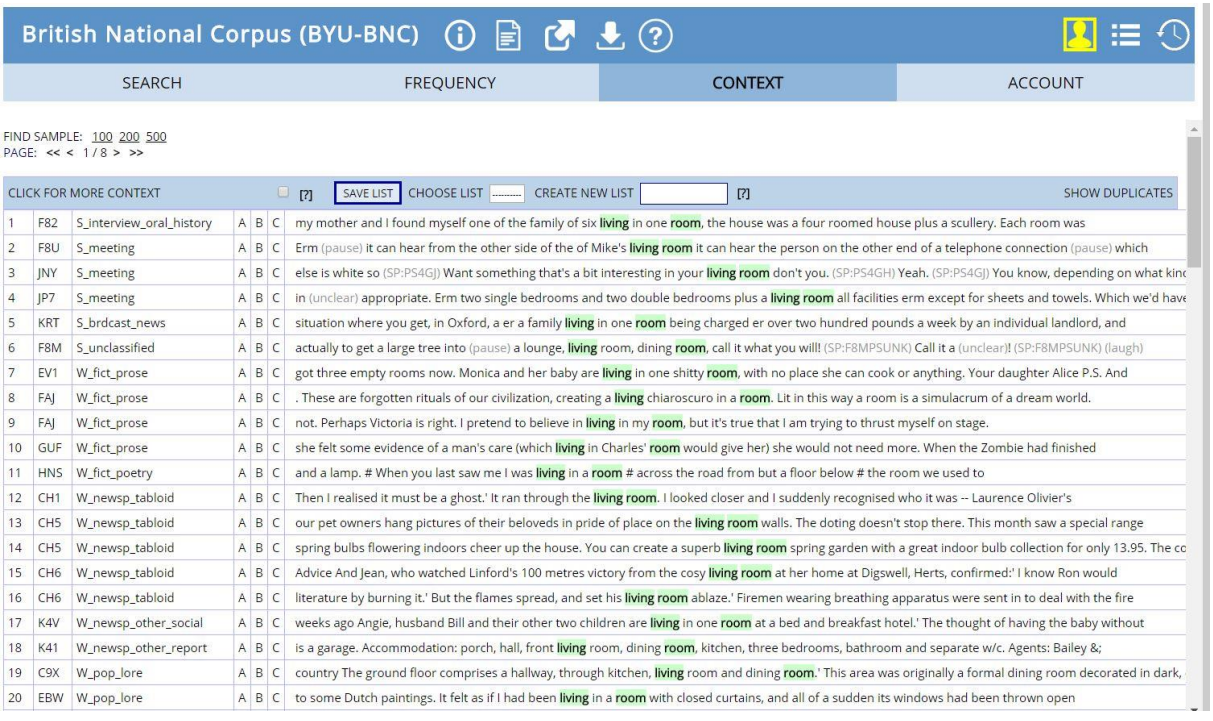


Figure 2 An example of a collocation, adopted from https://corpus.byu.edu/bnc/ (2018)

4. CORPUS-DESIGNED MATERIALS FOR ENGLISH

This chapter brings self-made corpus-designed materials to be used in the English lessons. Their structure and the way of creation was described in the introduction. It is not appropriate for pupils to use these materials without any explanations if they have not worked with them so far. The materials are written from the position of a teacher. There are (according to me) some practical steps which should be observed before working with these materials (not necessarily in the particular order). I created a list of these steps:

a) Once you decide to use these materials in your lesson make sure that a computer classroom is available as some of the tasks are to be fulfilled via the internet. The computer classroom can be either free or you can change it if the class who is supposed to have a lesson there do not need it at that time.

b) Be sure that all your pupils can sit in front of their own computer. It is not suitable if two or more pupils should sit together.

c) Create a same fictitious user name and password for all your pupils before. It is inappropriate to do it in the lesson.

d) Once you are in the classroom, explain a work with the British National Corpus at https://corpus.byu.edu/bnc/ before pupils are supposed to do the tasks. Use an overhead projector to see the websites, show your students how to log in, how to search in different sections etc. Try to be short and objective.

e) Your fictious user name and password should be visible for your pupils through the whole lessons (e.g. at the blackboard).

f) Some of the task are supposed to be done with the whole class and they do not need a computer. Try to lead these activities (e.g. start and finish of these activities). As the computer classr1oom do not have to provide enough space you can do these activities at the corridor but not disturbing lessons in neighbouring classrooms.

g) Explain to your pupils that any of the corpus-based activities is by no means a written exam. They should not be afraid of it. You can give extra good mark for it to encourage your pupils in learning English.

h) Be sure that you really know what to do. It is nothing worse than an unprepared teacher. Try to help your student, be ready to answer questions etc.

Corpus-designed activities – Worksheet 1

Topic: Family

Ice-breaking game

a) Pupils come up with as many adjectives describing family members as they know. You are going to write these adjectives down on the blackboard.

b) Pupils then try to find opposite words or synonyms of these adjectives, if there are any.

Whole class speaking activity

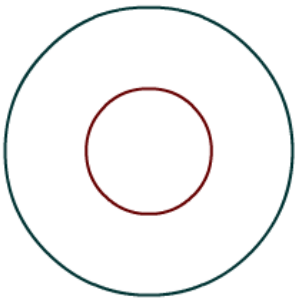
All pupils and the teacher make two circles of themselves (inner and outer ones) as shown in the picture (standing face to face).

The people from the inner circle speak about their family to those in the outer one for 30 seconds exactly (the teacher measures the time).

Then the inner circle moves left so that the people in it could speak to another person in the outer circle, having the same assignments.

When the people from the inner circle comes back to their original position, the people from the outer one start to speak.

You can see that pupils do not probably say the same information about their families every time as they have no time to prepare the speech.



Cloze activity

Pupils fill in the gaps with family members.

My brother's name is Andrew. He has 9 years old son Paul, who's my _____.

I have two grandmothers – Lucy and Katy. I am their _____.

Peter is going to marry my sister Laura soon. He will be my new _____.

Charles' wife suddenly passed away last month. He is a _____ now.

Tina has become a _____. My cousin Henry has just made a proposal to her.

Working with the internet corpora

Pupils fill in the chart with the right answers from the previous task and find out the number of collocations.

Family members		**The number of collocations**
Brother		
Grandmother		
Sister		
Wife		
Cousin		

Pupils find out the frequency of the following words recounted per millions.

Mother Mum Father Dad

In what section are informal expressions (mum and dad) covered the most?

Writing and reading

Pupils make a short piece of writing (approximately 10 sentences) describing their idea of a future family (wife/husband, children, ideal house or flat etc.).

They will read their stories in pairs.

Corpus-designed activities – Worksheet 2

Topic: Housing

Ice-breaking game

Pupils make four groups of approximately the same number of people.

In their groups, they will take a big piece of paper and make a mind map, big and colourful, for the topic "House" to present it in front of their classmates.

Whole class speaking activity

Pupils go around the classroom and ask different classmates various questions concerning their housing.

You can prepare your own question as well.

Encourage them to ask you as well.

Multiple choice

Pupils circle a correct answer.

Are you going to **do**/**make** your bed now, Anabel?

Could you please **come**/**go** over to my house? I need to ask you something.

I must go home. I think I **have**/**get** no other choice.

Will you be so kind and **tell**/**give** me an advice?

I was not successful, but I think I will **make**/**do** an attempt again.

Working with the internet corpora

Pupils translate these expressions into English and find their frequency recounted per millions. They have to find an example of a sentence where the collocation is used in a different meaning than the Czech translation (words are separately).

Obývací pokoj

Jídelna

Koupelna

Pupils write down a frequency of these words in different sections of recounted per millions. Forget about a different meaning of some of them.

Word	**Section**	**Recounted per millions**
Doors	Academic	
Stairs	Newspapers	
Roof	Magazines	
Garage	Spoken	
Bed	Fiction	

Writing and reading

Pupils use the following words (not necessarily in the given order) to make a fictious story of approximately 10 sentences. They can use the words more times.

They read the story in front of the whole class and vote for the others to choose the most interesting story.

House

Small

Children (child)

Brave

Happen

Corpus-based activities – Worksheet 3

Topic: My favourite

Ice-breaking game

When we say "my favourite personality" we usually think of a singer or an actor, that we like a lot. But it can also be your favourite animal etc.

Pupils make pairs and write down 3 favourite animals each. They brainstorm positive adjectives for each animal (for example: dog – friendly).

Whole class speaking activity

Pupils write a famous person's name on a piece of paper.

You collect the papers to make sure they have chosen people who everyone knows.

They go one by one and you put their pieces of paper with a famous person's name on their back. Don't let their classmates see who their person is.

Now all of them mingle and look at each other's names. They ask questions about themselves. (e.g. Am I alive or dead? Am I from this country? Am I a boy or a girl?) They are not allowed to ask their name. They must guess who they are.

You can add rules to this activity (e.g. they can only ask three questions).

Error activity

Pupils find a mistake in each of the sentences, underline it and write the right expression.

My friend Peter is only 1,6 m big.

I had never had a chance to see Beyoncé live.

Unfortunately, my favourite actor's new film were postponed.

Michael Jackson, which I admire, is truly a legend.

The teachers of English and music are my favourite once.

Working with the internet corpora

Pupils find a word which these synonyms collocate the most with. If there is a punctuation, render the next expression.

Word 1	**The most frequent collocation**	**Word 2**	**The most frequent collocation**
Small		Little	
Famous		Popular	
Big		Large	
Warm		Hot	
Beautiful		Gorgeous	

Famous singers

a) Pupils find the frequency of these singers. Sometimes there is no frequency for the singer. They have to guess why.

b) Are the searches of Sia connected with the famous singer? If not, what are they?

Famous person	**Frequency**
Michael Jackson	
Rihanna	
Beyoncé	
Freddie Mercury	
Sia	

Role-play writing

Pupils write a short story of approximately 15 sentences following this plot.

Kelly and Mary are twins. Although they look the same, they are totally different in their behaviour and do not get on well with each other. They fell in love with the same man – a rich businessman Peter.

Corpus-based activities – Worksheet 4

Topic: Jobs and hobbies

Ice-breaking game

Pupils make pairs and take five cards of professions from you so that their partners cannot see them.

Their task is to use pantomime to explain the profession to their partners who have to guess what profession it is.

Whole class speaking activity

Pupils try to discover as many characteristics as they have in common with their classmates (hobbies, appearance …).

They go around the class and ask their classmates.

The one who finds the most is the winner.

You can give them a limit of, for example, 5 things.

Contextual analysis

Pupils explain why *have to* or *must* is used in any of these sentences.

I cannot go out. I **have to** learn as I have bad marks.

I **must** stop spoking.

You **must not** eat here. It is forbidden.

The book has already been found. We **do not have to** look for it anymore.

You **must** see the film.

Working with the internet corpora

Pupils compare these equivalents of men's and women's professions (in general, no section). Which words do these ones collocate the most? They have to render the first position only, if there is a punctuation, then the next one.

Word 1	**The most frequent colocation**	**Word 2**	**The most frequent colocation**
Actor		Actress	
Monk		Nun	
Waiter		Waitress	
Usher		Usherette	

Pupils render 5 activities they like to do in their free time (one word each, no compound words) and find their frequency in spoken section.

Activities they like to do	**Frequency in spoken section**

Writing and reading

Pupils use these three sentences to make a fictious story of approximately 10 sentences. They must not change the sentences given.

My father is a teacher.

He likes travelling.

We often go cycling.

Corpus-based activities – Worksheet 5

Topic: Travelling and holiday

Ice-breaking game

Pupils make four groups of approximately the same number of people.

They try to list as many countries with their capital cities as they know in two minutes.

The most successful group wins.

Whole class speaking activity

Pupils make four groups of approximately the same number of people.

Each one receives a bag with seven different objects in it.

At a signal, each group opens its bag, takes out its objects and invents an oral story incorporating all the objects found in the bag.

Cloze activity

Practising a vocabulary. Pupils fill in the gaps with the means of transport.

"Mind the gap" is a well-known announcement at London _____.

When you want to travel to Australia, the fastest way is to use a _____.

Small _____ are usually found on lakes and rivers.

Škoda is a type of _____ produced in Mladá Boleslav.

A type of bus used for package tours is called a _____.

Working with the internet corpora

Pupils find expressions of the world class given which collocate with the expression in the first column the most.

Expression	Word class	The most frequent colocation
Book	Noun	
Fast	Verb	
Travel	Number	
Hide	Adjective	
Long	Adverb	

There are expressions typical for any country. Pupils find a frequency of those in the following table, if there is any (collocations)?

Country	Expression	Frequency
Italy	Wine	
Netherlands	Mill	
USA	Wheat	
Croatia	Sea	
Japan	Car	

Writing and reading

Pupils finish a fictious story of approximately 10 sentences beginning with "I was sitting in a bus when…". They should make it as funny as they can, but they must not use swear words.

Corpus-based activities – Worksheet 6

Topic: Sports

Ice-breaking game

One of the pupils says a word connected to sports. A person next to him/her repeats the same word and tells a new one.

The game continues in the same way.

Once he/she forgets to say any word, he/she is left out.

Whole class speaking activity

Three pupils sit in a line in front of the class. They are sport experts.

The rest of the class then ask them questions and they answer them. Each expert uses one word at a time only.

Example:	Do you think that Petra Kvitová will win this year's Wimbledon?
Expert 1	No
Expert 2	Serena
Expert 3	Williams
Expert 1	will
Expert 2	win
Expert 3	it

Cloze activity

Practising a vocabulary. Pupils fill in the gaps with a verb.

My sister _____ gymnastics. She is really good in it.

Tennis player _____ by taking steroids and received a punishment.

When Lionel Messi _____ the crowd went wild.

Russia _____ the ice hockey match against Canada 3:0.

The ball was _____ by the goalkeeper in a very last second.

Working with the internet corpora

Pupils choose five different sport branches and find their frequency in newspapers and magazines recounted per millions.

Sport branch	**Newspapers**	**Magazines**

Pupils compare the frequency of the following expressions connected with sports in two different sections recounted per millions. Forget about a different meaning of some of them.

Expression	**Section 1**		**Section 2**	
Ball	Academic		Fiction	
High	Newspapers		Magazines	
Goal	Spoken		Newspapers	
Throw	Non-academic		Academic	
Jogging	Spoken		Magazines	

Writing and reading in pairs

Pupils make pairs and write down a chat conversation between a sportsman (one of them) and an interviewer (the other one). It is supposed to be a formal conversation between them (approximately 15 phrases or sentences each).

Corpus-based activities – Worksheet 7

Topic: Food and dishes

Ice-breaking game

Pupils make four groups of approximately the same number of people and write down as many types of food as they know in two minutes.

Each group read their ideas after that. Which group wrote the most types of food?

Whole class speaking activity

Bring several pictures of dishes and cut them into pieces according to the number of pupils in the class.

Each of the pupils receives a piece of a picture (a puzzle) so that nobody can see it.

They task is to go around the class and ask their classmates what they have on their puzzle to put them together.

The first complete picture wins.

Cloze activity

Practising a vocabulary. Pupils fill in the gaps with the right expression.

How many _____ of bread are, on average, in a _____ of bread.

My mother ate a _____ of chocolate yesterday.

Tom's parents like a _____ of black tea for breakfast.

Could I have a _____ of water? I am thirsty.

Working with the internet corpora

The types of fruit finishing with -berry usually make problems among pupils. Pupils find their frequency recounted per millions and try to realize what kind of fruit it is.

Word	**Recounted per millions**
Strawberry	
Raspberry	
Blueberry	
Blackberry	
Gooseberry	

Pupils choose their five favourite types of food and find their frequency recounted per millions in spoken section. Forget about a different meaning of some words.

Type of food	**Recounted per millions**

Writing and reading in groups

Pupils make four groups of approximately the same number of people and plan a party. You can give them possibilities such as a Christmas party, a birthday party, a wedding party, an end-of-school party etc.

Each group then make a shopping list for the selected celebration, including food, decorations, invitation cards, gift bags, etc.

When the groups are ready, they share their ideas with the others.

CONCLUSION

The bachelor's thesis dealt with the issue of language corpora and their usage in English language teaching. In the theoretical part there were basic principles of corpus linguistics and corpora concerning development, types, structure etc. presented. It shows itself that this branch of linguistics is still not very well-known, especially for people who are not professionals in terms of English linguistics. The theoretical part was supposed to serve as an introduction to this issue with awareness that some of possible readers could take a deeper interest in it. The practical part brought several corpus-designed worksheets to use these materials further in the English lessons to make them more interesting. Teachers can make their own corpus-designed worksheets either for low-level or higher-level students, following several steps as indicated in the thesis.

In the thesis there were several works of different authors (either Czech or foreign), dealing with the issues of corpus linguistics and language corpora, quoted. Each of the authors deals with a bit different section of this branch, only a few of them overlap in some topics. The most important one for me was Bennett and his "Using Corpora in the Language Learning Classroom". This work brought the most substantial piece of knowledge to me and was used in the thesis more times. According to me, the book is very-well written with sections which are good to understand and helped to create a good basis of the thesis.

I believe that teaching English will not stagnate with simple "swotting up" the vocabulary and grammar rules what is a bit horrifying for many pupils and students. Hopefully, it will bring much more pleasure of learning this beautiful language which is undoubtedly a world language spoken by hundred millions of people for different purposes every day. It needs time to improve this situation even though there are lots of good teachers who can incorporate new techniques of learning English to their lessons. One of the ways is that teachers should gain awareness of new strategies in teaching English, not only corpus-based activities, as early as they find their first job vacancy, this means at universities. However, it is not only about the studies, it is also about the teacher's willingness to free from usual methods and bring something new into their lessons. And I hope it is gradually approaching.

REFERENCES

Bibliography:

BENNETT, Gena R. *Using Corpora in the Language Learning Classroom: Corpus linguistics for teachers*. Ann Arbor, MI: University of Michigan Press, 2010. ISBN 0472033859.

ČERMÁK, František. Korpusy včera, dnes a zítra. In: ČERMÁK, F., KUČERA, K., PETKEVIČ, V., ed. *Studie z korpusové lingvistiky 2, Korpusová lingvistika, výzkum a výstavba korpusů*. Praha: Lidové Noviny, 2006. ISBN 978-80-7422-115-6.

ČERMÁK, František a Věra SCHMIEDTOVÁ. Český národní korpus – základní charakteristika a širší souvislosti. *Národní knihovna: Knihovnická revue*. Národní knihovna České republiky, 2004, **15**(3), 16. ISSN 0862-7487.

FRANCIS, Winthrop Nelson. Language corpora B.C. In: SVARTVIK, Jan. *Directions in corpus linguistics: proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991*. New York: Mouton de Gruyter, 1992. ISBN 3-11-012826-8.

FRANKENBERG-GARCIA, Ana. Integrating corpora with everyday language teaching. In: THOMAS, James a Alex BOULTON. *Input, Process and Product: Developments in Teaching and Language Corpora*. Brno: Masaryk University Press, 2012. ISBN 978-80-210-5896-5.

GUY ASTON a LOU BURNARD. *The BNC handbook: exploring the British National Corpus with SARA*. Edinburgh: Edinburgh University Press, 1998. ISBN 07-486-1054-5.

MEYER, Charles F. *English corpus linguistics: an introduction*. New York: Cambridge University Press, 2002. ISBN 0-521-00490-x.

Online sources:

Generative grammar. *Encyclopaedia Britannica* [online]. [cit. 2018-06-04]. Dostupné z: https://www.britannica.com/topic/generative-grammar

NESSELHAUF, Nadja. Corpus linguistics: a practical introduction. In: *Nadja Nesselhauf* [online]. September 2011 [cit. 2018-02-25]. Dostupné z: http://www.as.uni-heidelberg.de/personen/Nesselhauf/files/Corpus%20Linguistics%20Practical%20Introduction.pdf

British National Corpus [online]. [cit. 2018-03-01]. Dostupné z: http://www.natcorp.ox.ac.uk/corpus/index.xml

British National Corpus [online]. [cit. 2018-03-01]. Dostupné z: https://corpus.byu.edu/bnc/

List of figures

List of tables

List of abbreviations

APPENDIXES

Appendix A – Right answers in the worksheet 1 (My family)

Cloze Activity

My brother's name is Andrew. He has 9 years old son Paul, who's my **nephew**.

I have two grandmothers – Lucy and Katy. I am their **grandson**.

Peter is going to marry my sister Laura soon. He will be my new **brother-in-law**.

Charles' wife suddenly passed away last month. He is a **widower** now.

Tina has become a **fiancée**. My cousin Henry has just made a proposal to her.

Working with the internet corpora

Family members		**The number of collocations**
Brother	Nephew	7
Grandmother	Grandson	1
Sister	Brother-in-law	16
Wife	Bride	16
Cousin	Fiancée	0

mother	241.23	
mum	79.88	spoken
father	225.08	
dad	64.71	spoken

Appendix B – Right answers in the worksheet 2 (Housing)

Multiple choice

Are you going to **make** your bed now, Anabel?

Could you please **come** over to my house? I need to ask you something.

I must go home. I think I **have** no other choice.

Will you be so kind and **give** me an advice?

I was not successful, but I think I will **make** an attempt again.

Working with the internet corpora

living room	7.10
dining room	8.17
bathroom	23.08

Word	**Section**	**Recounted per millions**
Doors	Academic	8.54
Stairs	Newspapers	10.41
Roof	Magazines	32.36
Garage	Spoken	45.16
Bed	Fiction	476.32

Appendix C – Right answers in the worksheet 3 (My favourite)

Error activity

My friend Peter is only 1,6 m **big**.	**tall**
I **had** never had a chance to see Beyoncé live.	**have**
Unfortunately, my favourite actor's new film **were** postponed.	**was**
Michael Jackson, **which** I admire, is truly a legend.	**who**
The teachers of English and music are my favourite **once**.	**ones**

Working with the internet corpora

Word 1	**The most frequent collocation**	**Word 2**	**The most frequent collocation**
Small	Relatively	Little	Bit
Famous	Most	Popular	Most
Big	Big	Large	Number
Warm	Keep	Hot	Water
Beautiful	Most	Gorgeous	Oh

Famous person	**Frequency**
Michael Jackson	194
Rihanna	0
Beyoncé	44
Freddie Mercury	0
Sia	3

Sia is an abbreviation, an Italian expression and a scientific term

Appendix D – Right answers in the worksheet 4 (Jobs, hobbies)

Contextual analysis

I cannot go out. I *have to* learn as I have bad marks. obligation from the outside

I *must* stop spoking. I want

You *must not* eat here. It is forbidden. nesmíš

The book has already been found. We *do not have to* look for it anymore. nemusíme

You *must* see the film. it is worth seeing

Working with the internet corpora

Word 1	**The most frequent colocation**	**Word 2**	**The most frequent colocation**
Actor	An	Actress	Best
Monk	Buddhist	Nun	Priest
Waiter	Head	Waitress	Cocktail
Usher	Hall	Usherette	An

Appendix E – Right answers in the worksheet 5 (Travelling, holiday)

Cloze Activity

"Mind the gap" is a well-known announcement at London **underground**.

When you want to travel to Australia, the fastest way is to use a **plane**.

Small **boats** are usually found on lakes and rivers.

Škoda is a type of **car** produced in Mladá Boleslav.

A type of bus used for package tours is called a **coach**.

Working with the internet corpora

Expression	Word class	The most frequent colocation
Book	Noun	Book
Fast	Verb	Going
Travel	Number	Two
Hide	Adjective	Unable
Long	Adverb	So

Country	Expression	Frequency
Italy	Wine	5
Netherlands	Mill	0
USA	Wheat	1
Croatia	Sea	0
Japan	Car	23

Appendix F – Right answers in the worksheet 6 (Sports)

Cloze Activity

My sister **does** gymnastics. She is really good in it.

Tennis player **cheated** by taking steroids and received a punishment.

When Lionel Messi **scored** the crowd went wild.

Russia **won** the ice hockey match against Canada 3:0.

The ball was **caught** by the goalkeeper in a very last second.

Working with the internet corpora

Expression	**Section 1**		**Section 2**	
Ball	Academic	13.04	Fiction	55.38
High	Newspapers	448.91	Magazines	441.98
Goal	Spoken	44.06	Newspapers	263.22
Throw	Non-academic	22.49	Academic	11.61
Jogging	Spoken	3.21	Magazines	2.89

Appendix G – Right answers in the worksheet 7 (Food, dishes)

Cloze Activity

How many **slices** of bread are, on average, in a **loaf** of bread.

My mother ate a **bar** of chocolate yesterday.

Tom's parents like a **cup** of black tea for breakfast.

Could I have a **glass** of water? I am thirsty.

Working with the internet corpora

Word	**Recounted per millions**
Strawberry	3,71
Raspberry	1,63
Blueberry	0,11
Blackberry	1,44
Gooseberry	0,62