Jihočeská univerzita v Českých Budějovicích

Pedagogická fakulta

Katedra anglistiky

Diplomová práce

Measuring authenticity of dialogues: English textbooks and sitcom scripts

Zkoumání míry autentičnosti dialogů v učebnicích anglického jazyka a scénářích situačních komedií

Vypracovala: Bc. Alžběta Hálová

Vedoucí práce: Mgr. Jaroslav Emmer

České Budějovice 2023

Prohlášení

Prohlašuji, že svoji diplomovou práci jsem vypracovala samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své diplomové práce, a to v nezkrácené podobě, elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

V Českých Budějovicích, dne

....

Alžběta Hálová

Anotace

Tato diplomová práce zjišťuje míru autentičnosti modelových dialogů v učebnicích anglického jazyka a scénářů situačních komedií.

Jejím cílem je prokázat, zda scénáristé situačních komedií předkládají autentičtější dialogy než autoři učebnic anglického jazyka.

Teoretická část je věnována procesu komunikace, jejím modelům a kanálům. Dále je ve stručnosti popsán pragmatický aspekt sociální interakce a je vyzdvižena klíčová role kontextu při procesu komunikace. Pozornost je také věnovaná kolokacím a žánru situační komedie.

V praktické části je představena analýza nasbíraných dialogů ze čtyř situačních komedií a učebnic angličtiny pro pokročilé uživatele (úroveň C1-C2). Jsou zvoleny dva sitcomy od britských tvůrců a dva od amerických, které jsou srovnány s britskými a americkými učebnicemi angličtiny (CAE a TOEFL). Výzkum se soustředí na nejčastěji se vyskytující podstatná jména a jejich kolokace. Výsledky výzkumu jsou následně také porovnány s národními korpusy anglické a americké mluvené angličtiny.

Výsledek diplomové práce ukazuje, že situační komedie i učebnice pro pokročilé uživatele angličtiny zahrnují autentické dialogy.

Abstract

This diploma thesis measures the authenticity of model dialogues in English textbooks for advanced learners and conversations occurring in situation comedies.

The theoretical part focuses on the communication process, its models and channels. Subsequently, context is presented as well as its importance in speech. Attention is then paid to collocations as well as to classification of the genre of situation comedy.

The practical part presents the research of sitcom and textbooks dialogues. The data were collected from four sitcoms (two of British and American origin) and CAE and TOEFL textbooks. The analysis concentrates on the most frequent nouns and their collocations. The findings are then compared, and the most frequent expressions are studied with the help of national corpora of British and American spoken English.

The result of this thesis is that both textbook and sitcom dialogues include authentic material.

Table of contents

List of abbreviations

BNC	British National Corpus
CAE	Certificate of Advanced English
HIMYM	How I Met Your Mother
INBT	The Inbetweeners
OANC	Open American National Corpus
TBBT	The Big Bang Theory
TOEFL	Test of English as a Foreign Language

1. Introduction

While in the past, people used to learn languages through memorizing lists of words and grammatical rules from textbooks or repeating tape recordings, these days, there are many modern approaches and methods to English learning and teaching. With the development of information technologies and electronic devices, the methods and resources for teaching and learning English, or languages in general, also grow and so is their quality.

The learners very often use books, music, PC games or films to practise and improve their English. What also became very popular in the last twenty years are sitcoms. Nowadays, there is not a teenager who does not watch sitcoms on TV or their computer.

The learning potential of sitcoms is great, because they often introduce common people set in common locations, and usually contain simple sentences with phrases that tend to repeat a lot. But most importantly, they make people laugh and it is fun to watch them. And so, it is a great combination of both entertainment and learning.

In the last ten years, quite a considerable amount of research has been made on the topic of sitcom or tv shows in general, mainly its usage in teaching. The authors examined their ability to improve English language skills, primarily listening and speaking. Some even point out that sitcoms can be also used as a tool for practising grammar structures.

What they see as the main reason for using sitcoms in English class is that they have a large positive impact on learners' motivation as well as a friendly learning environment and atmosphere among pupils.

Another advantage of the usage of these audio-visual aids is that the learners are more likely to remember a new phrase occurring in a humorous background than in a textbook model dialogue.

Some authors also express their opinion about the authenticity of sitcoms dialogues. As these tv series are intended mainly for native speakers of English, it indicates that the scripts include real-life speech.

The purpose of this thesis is to expand the existing research on phrases and expressions occurring in situation comedies and to examine their actual usage in everyday life.

This thesis aims to measure the authenticity of model dialogues occurring in advanced grammar textbooks in comparison with sitcom dialogues. The hypothesis is that sitcom scripts include more authentic dialogues than those in grammar textbooks.

The theoretical part deals firstly with the process of communication, its models, and channels. Subsequently, it is described how people communicate in different contexts. Collocations are then defined and classified. Different approaches to collocations are presented along with available research methods used for studying collocations. Then the sitcom genre is characterised and analysed in terms of language.

In the practical part, CAE and TOEFL textbooks are analysed as well as four situation comedies, two of American and British origin. How I Met Your Mother and The Big Bang Theory are chosen from the American setting. These sitcoms are picked because they are very popular among Czech learners of English, especially teenagers. They also occur on most of the streaming platforms available in the Czech Republic. From the British setting, Chewing Gum and The Inbetweeners are analysed.

At first, #Lancsbox is presented as the software used for this analysis along with its main functions. After that, the research itself is carried out and findings are gathered. The object of the research is often occurring collocations in model dialogues of grammar textbooks and conversations in sitcom scripts. Subsequently, the results from textbooks and sitcoms are compared together. The authenticity of the results is then measured with the use of the national corpora.

The result of this thesis is the evaluation of the authenticity of model dialogues appearing in British and American advanced grammar textbooks.

2. Theoretical part

2.1. Social communication and its variety

Communication is the most basic competence of every living being, from bacterium, plants or animals to humankind. The form of communication is unique in every possible way, and every individual is special in both producing and receiving new information. Nevertheless, the intention is always the same – to share some new information. Each organism has a natural tendency to relate to another living creature. It greatly corresponds with the meaning of the word communication itself, which means "to share" or "to be in a relation with".[1]

Animals can create a message based only on a non-verbal way. Human beings' uniqueness lies in their competence to convey a message consisting of both of non-verbal and verbal signals. People communicate primarily through visual and auditory signals. Humans, as well as animals, need to communicate in order to survive. But maybe it is more about the need for social contact, which is crucial for survival and leads people to communicate.

2.2. Communication process and models of communication

For communication to happen, at least two participants must present – a sender and a receiver. The sender has to come up with a meaning, information which is then transformed and encoded into a message. After that, a communication channel is chosen, a medium transmitting the data, and the idea is sent to the receiver. A process of decoding occurs after the listener receives and accepts the signal. The receiver translates the message and interprets it. The last part of a conversation is when the listener provides feedback, responding to the delivered news. This is probably the most crucial step of the communication process because the communicators check whether they understood each other and whether the receiver interpreted the information in the way the sender intended it. In other words, if the exchange fulfilled its purpose.

What is also present during the exchange is noise. It can be defined as anything distracting the listener's attention from the message he is receiving. Noise is any accidental

---

[1] Paul Cobley, "Communication: Definitions and Concepts", (2008-06-05), https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781405186407.wbiecc071.

changes to the transmitted signal.[2] Noise is inevitable, and due to it, communicators can misinterpret or misunderstand the transferred information. Wood[3] lists four kinds of noise:

- Physiological including hunger, a headache, fatigue and other physiological symptoms.
- Physical, which is the effect of the environment such as sounds of others around or lighting.
- Psychological refers to the state of mind of the participants.
- Semantic noise, words that are not understood, for instance, technical language or abbreviations.

It would be impossible to describe every model of communication, as every author has its own approach to message exchange. Some of these models are more complex than others, each focusing on different signal transmission parts.

The first model of communication was designed by mathematicians Shannon and Weaver, who were the founders of the linear model. They viewed communication only as a one-way process focusing mainly on the part of encoding and decoding. This mathematical theory was considered the most influential and functioned as an example to other authors.

[2] Daniel Chandler, A Dictionary of Media and Communication (Aberystwyth University: Oxford University Press, 2020), 726

[3] Julia T Wood, Interpersonal Communication: Everyday Encounters (The University of North Carolina at Chapel Hill: Wadsworth Publishing, 2012), 15
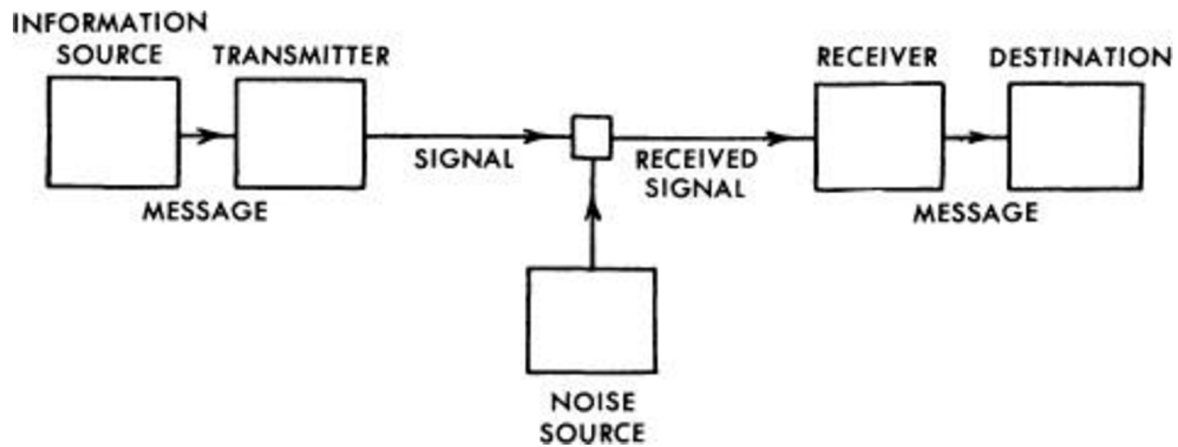
Fig. 1 - The Shannon-Weaver linear model[4]

A similar graphical scheme was called the Lasswell Formula. Lasswell used a strategy of asking several basic questions to understand communication.
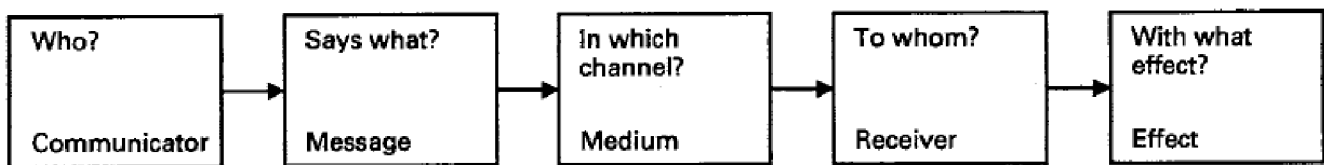
Fig. 2 - The Lasswell Formula[5]

In the linear models, the roles of the participants were fixed. Furthermore, these models lacked feedback and did not picture real-life conversations. This was solved when transactional and interactional models were designed.

---

[4] Claude Shannon, The Mathematical Theory of Communication (Chicago: University of Illinois Press, 1963), 33

[5] Sven Windahl, Denis McQuail, Communication Models for the Study of Mass Communications (Abingdon: Routledge, 2013), 13

The linear models can be perceived as communication to the masses. On the other hand, the interactional approach fully acknowledges the feedback between the communicators. It is a transmission that goes both ways.
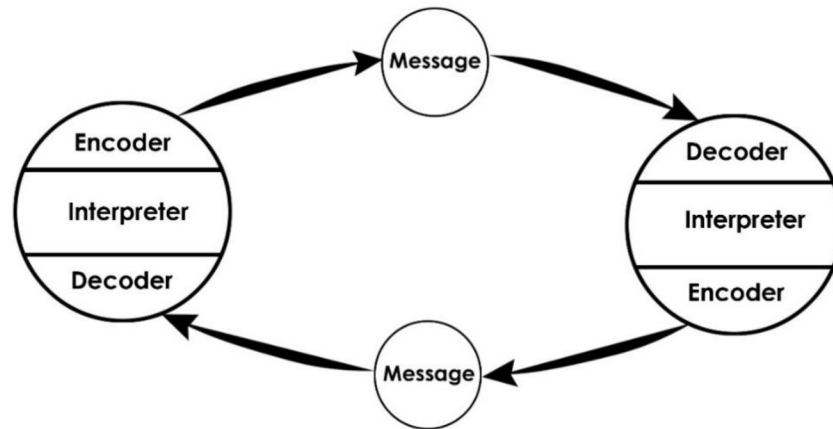
Fig. 3 - The Osgood and Schramm interactional model[6]

2.3. Types of communication

While communicating, two central signal systems are used – verbal and non-verbal. The verbal system includes mainly verbal signals transferred by air and received by hearing. The non-verbal or extra-linguistic system represents all other channels of sharing meaning. These are gestures, facial expressions and tactility. In a real conversation, both of the main systems work simultaneously.[7]

2.3.1. Theory of non-verbal communication

The non-verbal conversation includes every transmission of information that does not include any speech element.[8]

---

[6] Sven Windahl, Denis McQuail, Communication Models for the Study of Mass Communications (Abingdon: Routledge, 2013), 19

[7] Joseph A. DeVito, Základy mezilidské komunikace (Praha: Grada Publishing a.s., 2001), 99

[8] Zbyněk Vybíral, Psychologie komunikace (Praha: PORTÁL, s. r. o., 2013), 81

Floyd defines nonverbal communication as "behaviours and characteristics that convey meaning without the use of words (...) ". He adds that in speaking, non-lexical behaviour and physical appearance generally prevail over the speech itself. Furthermore, it can also reveal more information than oral transmission.[9]

In a real-life conversation, speaking non-verbally often comes naturally, and it often completes the verbal part of any utterance. It is very often a subconscious process that cannot be controlled, so it can be said that it is pretty easy to master this type of communication. When imagining a situation where one's intention is to lie, one often keeps a blank expression because it makes it more believable. Mimics, along with other nonverbal behaviour, usually catch one in a lie. The thing is that a non-lexical message is more believable. In fact, it is usually believed over the verbal one. The reason may be because, as mentioned earlier, nonverbal behaviour is more of a subconscious process that is difficult to control. Alternatively, it is just easier to believe in something possible to perceive visually than by hearing.[10]

Functions of nonverbal communication

According to Floyd, people use nonverbal codes for two main reasons:

- to invite and maintain a conversation,
- to express emotions[11].

Channels of nonverbal communication

Křivohlavý, Argyle, Floyd and other authors agree on these channels of nonverbal communication:

- gesticulation, head movement and other body movements
- body posture and position
- facial expressions or facial displays

---

[9] Kory Floyd, Interpersonal Communication (USA: McGraw-Hill Education, 2011), 178-179
[10] Kory Floyd, Interpersonal Communication (USA: McGraw-Hill Education, 2011), 182
[11] Kory Floyd, Interpersonal Communication (USA: McGraw-Hill Education, 2011), 184,246

- eye contact and other eye behaviour, for example pupil size

- choice of distance

- use of smell

- physical contact and touch behaviour

- vocal behaviour such as intonation, volume, rate, pronunciation, articulation and others

- physical appearance including clothing and other accessories

- visual features and objects[12]

2.3.2. Theory of verbal communication

Verbal (from verbalis) means consisting of words. Verbal communication is the act of using words (either written or spoken) to share information. People use symbol systems for communication – languages.

Intrapersonal communication

Through intrapersonal communication, we talk to, get to know and judge ourselves.[13] Janoušek highlights several aspects of intrapersonal communication: conscious, intentional or unintentional, audible or inaudible, involuntary and spontaneous.[14]

There is no need to go into detail about how inner speech functions as, according to a recent study, 96% of adults say they have self-talk, and 25% have audible internal dialogue. Talking to oneself can have positive effects on mental health.[15]

---

[12] Jaroslav Křivohlavý, Jak si navzájem lépe porozumíme (Praha: Svoboda, 1988); Michael Argyle, Bodily Communication (New York: International Universities Press, Inc, 1975); Kory Floyd, Interpersonal Communication (USA: McGraw-Hill Education, 2011), 190

[13] Joseph A. DeVito, Základy mezilidské komunikace (Praha: Grada Publishing a.s., 2001), 15

[14] Jaromír Janoušek, "Intrapersonální komunikace a vnitřní řeč," Československá psychologie 45, no. 6 (2001): 481-492, http://cspsych.psu.cas.cz/result.php?id=270

[15] WebMD Editorial Contributors, "Why Do People Talk to Themselves?" WebMD, (2021-06-28), https://www.webmd.com/balance/why-people-talk-to-themselves

Interpersonal communication

DeVito describes interpersonal communication as a process of sending and receiving information. He adds that communication is constantly disturbed by noise.[16]

Via interpersonal communication, we affect others and interact with them. We get to know them and let them know us. The purpose is to start, maintain, end, corrupt a relationship or make it right.[17]

Communication principles

According to DeVito, there are three principles of interpersonal communication which have to apply:

- inevitability
- irreversibility
- unrepeatability

He claims that communicating is inevitable because we often send signals to others unknowingly and unintentionally. Therefore, we send a message to others even in cases we do not wish to. For the communication to be complete, the other side has to receive our signal – without it, the principle of inevitability would not apply. As for the second principle, when we say something, we cannot "unsay" it, we cannot take it back. The best we can do is to weaken the effects of our actions – to say sorry or try to rephrase our statement.[18]

Functions of interpersonal communication

Every communication has its function and purpose, which affect the receiver and give the communication a meaning.

DeVito claims that every person communicates with a particular intention and that some motivation drives them. He feels that the intention of communication is to convey a message. He categorises five main goals of conversation:

---

[16] Joseph A. DeVito, Základy mezilidské komunikace (Praha: Grada Publishing a.s., 2001), 20
[17] Joseph A. DeVito, Základy mezilidské komunikace (Praha: Grada Publishing a.s., 2001), 15
[18] Joseph A. DeVito, Základy mezilidské komunikace (Praha: Grada Publishing a.s., 2001), 33

- to learn – to get knowledge about others, about the world and himself,

- to connect – to start relationships with others, to interact with others,

- to help – to listen to others and offer them solutions,

- to influence – to strengthen or change approaches of others or their behaviour,

- to play – to enjoy a particular moment[19].

Forms of interpersonal communication

Interpersonal communication takes various forms according to one's social background and role. It is logical and natural to speak differently with family members and friends, colleagues and superiors, or strangers. Here are some examples of forms of communication between persons, according to Vybíral:

- intimate communication of a couple in private

- private communication of a couple in public

- a communication controlled from one side and that has divided roles (an interview, an interrogation)

- informal communication in a small group (family and relatives)

- formal, polite communication in a small group (a business meeting)

- communication in business agreements and contracts[20]

2.4. Pragmatic aspect of communication

No matter how wide one's knowledge is about communication, its types, or forms, one has to also consider the pragmatic aspect of a conversation. It is not enough to analyse what people say but what they mean by their utterances.

In the previous chapter, communication was described as a process of encoding and decoding of information which is transferred through a communication channel.

---

[19] Joseph A. DeVito, Základy mezilidské komunikace (Praha: Grada Publishing a.s., 2001), 33
[20] Zbyněk Vybíral, Psychologie komunikace (Praha: PORTÁL, s. r. o., 2013), 39

But communication includes more than just a transformation of signal into meaning. It is much more complex and includes processes beyond what is verbally stated. It is about what is not said out loud but is still understood from the given situation.

There is a big gap between the meaning of the utterance and the true intentions of it. Here is an example of a situation, that happened to me during my teaching.

The lesson was just starting, and I had earlier opened a door when no one was in the classroom but me to let some fresh air inside. The door leads straight outside. I was about to start the lesson, when one of my students said, looking at the open door: "Aren't we gonna be cold this way?" While apologizing and replying that I just wanted some fresh air I got up and closed the door right away.

The true intention of the sentence "Aren't we gonna be cold this way?" was actually "Do you think we can close the door?".

If I only considered the semantic meaning of the sentence, I would probably respond something like "Yes, we probably are." and stayed sitting. But given the situation and that I was also feeling the cold, I reacted the way I did.

While semantics studies abstract meanings of words, pragmatics deals with what people mean by their utterances, what the speaker says, and how the listener interprets it. Yule[21] describes pragmatics as "study of speaker meaning". The main concern of pragmatics is therefore the intention of utterances.

2.4.1. Context and its types

The Cambridge dictionary describes context broadly as "the influences and events related to a particular event or situation".[22] When talking about specific types of contexts, the definition becomes more concrete. Huang distinguishes three types of contexts: [23]

- <u>Physical context</u>

Physical context refers to the physical setting of the utterance. To understand the utterance properly, the knowledge of the physical environment is needed.

---

[21] G. Yule, H. Widdowson, Pragmatics: Introduction to Language Study Series (Oxford: Oxford University Press, 1996), 3

[22] Online Cambridge dictionary. (2023-12-04). https://dictionary.cambridge.org/

[23] Y. Huang, Pragmatics (Oxford: Oxford University Press, 2014), 34

Example:

She is not the author of the painting, she is.

This sentence is accompanied by pointing at two different women. Without the knowledge of the physical setting, this utterance would not be understood.

- General knowledge context

It is the person's awareness of the world that helps to deduce the meaning of these utterances.

Example:

I visited Rome last year. The Colosseum was beautiful.
I visited Brussels. The Colosseum was beautiful.

The second sentence is pragmatically abnormal, because with the general knowledge of the world, it is clear that there is no Colosseum in Brussels.

- Linguistic context

During the exchange, very often misinterpretations occur, and the speakers have to clarify the intentions of their speech. People create the ideas in their head and when they transfer them into speech, they sound different. Either the speakers do not think about the formulation of their ideas long enough before saying them out loud, or they are just not able to use the right words to describe their thoughts. Sometimes they do not use enough words, or they leave out those essential ones. Context makes the conversation easier to comprehend.

Linguistic context can be briefly described as a conversational setting.[24] It is the set of conditions surrounding a conversation influencing it and making it understandable. But this

---

[24] P. Schumacher, R. Finkbeiner, What is a Context?: Linguistic Approaches and Challenges (Amsterdam: John Benjamins Publishing Company, 2012), 10

does not mean that context is only forming conversation, on the contrary, context and conversation are interrelated, and they are in balance with one another.

People often use the phrase "out of context", for example: "That is not what I meant. You took my words completely out of context!" This happens when the receiver takes the message not as a whole but without its surrounding text. The change of the intended meaning can be quite significant. A typical example is a journalist who interviews some celebrity and writes half of the answers out of context just to make the article more exciting for the readers.

Context also prevents semantic and syntactic ambiguities. Here are some examples of ambiguous sentences:

After work, Jenny went to the bank.

Either the speaker went to the institution which deals with money or went near the river.

Children get sad about nothing.

Either they get sad about every little thing, or they never get sad.

When adding some context to these sentences, they are no longer ambiguous.

- After work, Jenny went to the bank. She was really overworked, but as soon as she sat down and listened to the sound of the river, she felt better.
- After work, Jenny went to the bank. She was surprised to find out that her financial advisor retired.
- Children get sad about nothing. They are always laughing and excited about everything.
- Children get sad about nothing. All it takes is to take away their favourite toy.

2.5. Lexical and grammatical words

In the English language there are open and closed word classes. Open word class includes a lot of words as it accepts new members. Here belong all content or lexical words. That includes nouns, verbs, adverbs, adjectives and interjections. This word class is permanently changing as new words often enter or leave the language. Content words (also called contentives) have descriptive content. [25] They carry the meaning of the text and without them it would not be comprehensible.

Closed word class is a much smaller category and there are all function or grammatical words. Function words are prepositions, pronouns, auxiliaries, conjunctions and articles. A function word, also called a functor does not have descriptive meaning but has primarily grammatical function. It conveys information about grammatical categories such as number, person, tense, et cetera.[26] Function words represent a significant role in English grammar. These words are not changing over time. [27] They create connections between the items of the sentence and hold them together. [28]

Radford presents a simple test on how to determine whether a word is a functor or a contentive. If a word has a corresponding antonym, then it is a contentive, in the opposite case it is a functional word (old/young, give/take, comedy/tragedy; because/-, be/-, the/-).[29]

Halliday says that there is no line dividing grammar words from content words.[30] By that he means that these two groups are interrelated, and in a sentence, they are both essential as they complement each other.

---

[25] Andrew Radford, English Syntax: An Introduction (Cambridge: Cambridge University Press, 2004), 34
[26] Andrew Radford, English Syntax: An Introduction (Cambridge: Cambridge University Press, 2004), 339
[27] Randolph Quirk, Sidney Greenbaum, A Comprehensive Grammar of the English Language (London: Longman, 1989), 68
[28] Ronald Carter, Michael McCarthy, Cambridge Grammar of English: A Comprehensive Guide (Cambridge: Cambridge University Press, 2006), 778
[29] Andrew Radford, Syntactic Theory and the Structure of English: A Minimalist Approach (Cambridge: Cambridge University Press, 1997), 45
[30] M.A.K. Halliday, Colin Yallop, Lexicology: A Short Introduction (London: Continuum, 2007), 4

Lexical and grammatical words and their distribution in texts

It is not possible to generally say which of these dominate the sentence and which specific words of these groups will occur there. The choice of lexical words in a text is closely related to the topic and context of the text. On the contrary, grammatical words are less dependent on the topic and context.

Later in the theoretical part lexical words in sitcoms and textbooks will be examined more closely.

2.6. Collocations

Collocations belong to the domain of the field of phraseology that studies fixed multi-word units, such as phrasal verbs or idioms.

There is not just one accepted definition of collocation, and many linguists share the opinion that it is a highly controversial phenomenon in linguistics. As Čermák points out, collocations, their recognition and usage are nowadays one of the most central subjects of corpus linguistics.[31]

In the literature devoted to collocations the terms *co-occurrence, tendency, close relation* or *combination* often repeat.

According to the Cambridge dictionary, collocations are „*a combination of words formed when two or more words are often used together in a way that sounds correct.* "[32] Granger and Meunier define a collocation as „*a word group consisting of two conceptual elements (…).*" They develop this definition and clarify that collocations are combinations of multiple words (usually two), in a semantic and syntactic relation and depend on each other.[33] For Fairclough collocations mean commonly and regularly used patterns of words.[34] A more complex clarification comes from Cruse, who states that a collocation means a series of lexemes that usually co-occur. He adds that these combinations are semantically transparent or non-

---

[31] F. Čermák, M. Šulc, Kolokace (Praha: NLN, 2006), 9
[32] Cambridge dictionary. (2022-12-16). https://dictionary.cambridge.org/
[33] S. Granger, F. Meunier, Phraseology: An interdisciplinary perspective (Amsterdam: John Benjamins Publishing Company, 2008), 56
[34] N. Fairclough, Analysing Discourse: Textual Analysis for Social Research (London: Routledge, 2003), 213

idiomatic; in other words, each lexical item included in a collocation has its own meaning.[35] Moreover, as Cruse also illustrates, in different collocates, the components also have their unique meanings (hard time/ hard rock/hard man/hard drive).[36]

In this respect, they quite differ from idioms, which are treated as complex semantic units. When figuring out the meaning of an idiom, a definition that each part of the idiom carries is of no importance as well as of no use (for example, out of your mind – to behave in a crazy or stupid way; hit the road – to begin a journey).

The first significant figure in collocation research is Palmer, who examines collocations according to their syntactic patterns.

Nevertheless, Firth is the one who is considered the father of collocations. He claims that "You shall know the word by the company it keeps".[37] He proposes that the meaning of a word is not just based on the word alone (in isolation), but also on the specific combinations it forms with other words.

Firth's theory is based on the belief that it is necessary to analyse collocations in terms of their context. This idea had a big impact on the future research of collocations as it contributed to the development of corpus linguistics.

Sinclair, Hunston, Greenbaum and other neo-Firthian linguists continues to examine collocations statistically via corpus data. Not only that it made it easier to analyse collocations, but this new approach also helped to better understand them.

Neo-Firthian linguists criticizes the intuitive approach to collocations as they found it insufficient and unreliable, and considered corpus data essential for their study.[38]

---

[35] D. A. Cruse, Lexical Semantics (Cambridge Textbooks in Linguistics) (Cambridge: Cambridge University Press, 1986), 40

[36] D. A. Cruse, A Glossary of Semantics and Pragmatics (Glossaries in Linguistics) (Edinburgh: Edinburgh University Press, 2006), 27

[37] Hans Lindquist, Corpus Linguistics and the Description of English (Edinburgh: EUP, 2009), 71

[38] Douglas Biber, Randi Reppen, The Cambridge Handbook of English Corpus Linguistics (Cambridge: Cambridge University Press, 2015), 108

2.6.1. Collocations in corpus linguistics

A corpus is a computer-based collection of texts of spoken and written language.[39] Corpus linguistics developed in the early 1960s, with the invention of computers.[40] It is relatively a new method of linguistic research in comparison with other approaches.

During the 1980s and 1990s, large English corpora were created including the British National Corpus and Corpus of Contemporary American English.[41]

Using corpora as a tool for the examination of collocations presents several advantages. The corpus works with a wide amount of information, and it includes authentic data from diverse sources like articles, books and real-life conversations.

For its large capacity, the corpus provides rich and real-life context of collocations which enables to analyse them and study their grammatical and lexical patterns.

The corpus also includes collocation frequencies which help to distinguish strong and weak collocations.

Every collocation consists of a node, a central lexeme of a collocation, and is surrounded by collocates, which are dependent on the node. While searching for a certain collocation in the corpus, it is necessary to select a search span – it determines the number of tokens (words) the collocation should have. In the corpus the maximum span available is 5 (-5 to +5 span).[42] It means that the corpus finds five words preceding or following the node. The higher the span is selected, the less significant the findings are.[43]

2.6.2. Measure of collocation

Halliday determines the *measure of collocation* representing the probability of occurrence of a lexeme according to the presence of a particular word.[44] Crystal uses the term

[39] Ronald Carter, Michael McCarthy, Cambridge Grammar of English: A Comprehensive Guide (Cambridge: CUP, 2006), 10

[40] Geoffrey Sampson, Diana McCarthy, Corpus Linguistics: Readings in a Widening Discipline (London: Continuum, 2005), 2

[41] David Crystal, A Dictionary of Linguistics and Phonetics (Hoboken: Wiley-Blackwell, 2008), 117

[42] Tony McEnery, Andrew Hardie, Corpus Linguistics: Method, Theory and Practice (Cambridge: Cambridge University Press, 2011), 124

[43] Hans Lindquist, Corpus Linguistics and the Description of English (Edinburgh: EUP, 2009), 73

[44] M. Halliday, Ch. Matthiessen, Halliday's Introduction to Functional Grammar (London: Routledge, 2014), 59

collocability or *collocational range* when talking about the ability of items to collocate and combine with one another.[45] Lyons sees collocational range rather as the set of contexts where collocations can be found.[46]

2.6.3. Classification of collocations

McCarthy and O'Dell distinguish three types of collocations – *strong, fixed* and *weak. Strong collocations* include words in very close relation, and they hardly ever collocate with other words.

The form of *fixed collocations* cannot be changed at all. No item of a fixed collocation can be replaced with another. McCarthy and O'Dell call them idioms. Cruse uses the term *bound collocations* and explains that they usually do not separate and behave as idioms, such as *helter-skelter,* meaning confused or chaotic.

Weak collocations comprise lexical units that combine with many other words, such as a *great deal, great thanks, a great idea or great depression.*[47][48]

Hill and Lewis list different types of collocations:

- adjective + noun (high school, social media, long time)
- verb + noun (make sense, pay attention, save money)
- noun + verb (time spend, years pass)
- noun + noun (health care, customer service, credit card)
- adverb + adjective (most important, very good, more likely)
- verb + adverb (know how, read more, learn how)[49]

---

[45] David Crystal, A Dictionary of Linguistics and Phonetics (Hoboken: Wiley-Blackwell, 2008), 86

[46] John Lyons, Linguistic Semantics: An Introduction (Cambridge: Cambridge University Press, 1995), 62

[47] F. O'Dell, M. McCarthy, English Collocations in Use Advanced (Cambridge: Cambridge University Press, 2017), 8

[48] D. A. Cruse, Lexical Semantics (Cambridge Textbooks in Linguistics) (Cambridge: Cambridge University Press, 1986), 41

[49] J. Hill, M. Lewis, LTP Dictionary of Selected Collocations (Boston: Cengage Learning, Inc, 1997), 6

2.6.4. Grammatical x lexical collocations

Granger and Meunier distinguish between lexical and grammatical collocations. Lexical collocations consist of two lexical words (other people, high quality, good idea, take care...), whereas grammatical collocations are formed from a lexical term and a functional word, as in part of, lot of, back to, responsible for or better than.[50]

As Lipka highlights, for a non-native speaker of English, it can be quite challenging to learn and use the suitable fixed collocation, given the fact that the items involved are sometimes semantically and morphologically very similar (big/great/large/huge; more easily/much easier).[51] According to Čermák, a native speaker uses collocations subconsciously, but is often in doubt when asked to explain why they used a certain collocation.[52]

2.6.5. Predicting collocations

It is impossible to predict collocations no matter to which extent one's awareness of the world is. In some cases, certain words collocate together without any reasonable or logical explanation (for example, coffee with milk is white coffee even though it does not even have white colour, but more beige).[53]

By studying collocations via corpus and examining their patterns, it can make it easier to predict them.

2.6.6. Importance of collocations

Collocations play an important role in the English language. They help to distinguish polysemous words and prevent ambiguity.

I **feel bad** that I did not help you. (guilty)

I am really **bad at sports**. (not skilful)

---

[50] S. Granger, F. Meunier, Phraseology: An interdisciplinary perspective (Amsterdam: John Benjamins Publishing Company, 2008), 33
[51] Leonhard Lipka, An Outline of English Lexicology (Berlin: De Gruyter, 1992), 167
[52] F. Čermák, M. Šulc, Kolokace (Praha: NLN, 2006), 9
[53] David Crystal, The Cambridge Encyclopedia of the English Language (Cambridge: Cambridge University Press, 2019), 162

The **fruit** in the fridge has **gone bad**. (spoiled, rotten)

She has a runny nose and a **bad cough**. (severe, intensive)

Every movie has a **bad guy**. (evil character, antihero)

The example shows that the adjective *bad* carries several meanings. The words that collocate with it are the ones that distinguish its particular meaning. The interpretation of the word also depends on the context in which the word *bad* occurs.

Using collocations in speech correctly shows that the person is an advanced user of English. It makes the utterance sound more natural and native-like. [54] [55]

Corpus-based research shows that the majority of texts include a considerable number of collocations.[56] This indicates that native speakers apply them regularly in their speech. They use them naturally in comparison with learners of English who acquire them only after learning them by heart.

---

[54] Douglas Biber, Randi Reppen, The Cambridge Handbook of English Corpus Linguistics (Cambridge: Cambridge University Press, 2015), 115

[55] Michael McCarthy, Felicity O'Dell, English Collocations In Use (Cambridge: Cambridge University Press, 2005), 6

[56] DĄBROWSKA, Ewa. Words that go together: Measuring individual differences in native speakers' knowledge of collocations. *The Mental Lexicon*. John Benjamins Publishing Company, 2014, **9**(3), 401-418.

2.7. Situation comedy

2.7.1. History

Sitcom, short for situation comedy is a television series genre of comedy involving several characters, usually a small number of them. They appear in a few constant and repeating settings. The characters are usually closely related, they are groups of best friends or relatives, usually a whole family stars in one sitcom.

These days, there are a great many situation comedies available on plenty of streaming services and their popularity is getting bigger and bigger every day. For example, Netflix, which is considered one of the most used video streaming services in the world, has more than 200 million paid memberships worldwide, and in the Czech Republic, there were almost half a million people in 2021 paying for this streaming platform.[57]

Sitcom as it is known today, transmitted by television, appeared in the 1950s. However, situation comedy was known even before that. It started as "radio comedy" in the 1930s. Unsurprisingly, this genre originated in the United States. In the Czech Republic the first radio sitcom did not appear until the 70s.

It involved a comedian telling jokes and stories in a microphone, which were enriched by a band performance. Later on, these radio programs evolved into variety shows and funny sketches were added. The most famous ones were *The Jack Benny Program* and *Amos 'n' Andy.*[58]

During these years, a specific radio situation comedy format developed. It had a form of a narrative. It had all features of a sitcom. The characters were often married couples. One of them, *Vic and Sade*, husband and wife, also had a son. These short radio sketches were full of humorous dialogues and were mainly intended for housewives.

With the invention of television in the US during the 1950s, situation comedy finally found its place on the screen. Radio sitcoms were slowly losing their popularity after that, and television sitcoms started to dominate.

---

[57] Statista Research Department, " Netflix subscription count in Czechia 2021," Statista, (2022-07-07), https://www.statista.com/statistics/1285620/netflix-subscriptions-czechia/
[58] Skretvedt, R., Sterling, Christopher H.,"Radio," *Encyclopedia Britannica*, https://wwwritannica.com/topic/radio.

This was the post-war era and people were really in need of some entertainment and distraction. At that time, it was all about settling down, buying a house, getting married and starting a family. And television had a big impact on it too, especially sitcom programs. They portrayed a perfect and stable family financially secure. This prototypical family included a working father, a stay-at-home mother and several children. One of the most popular television sitcoms was "I love Lucy". It followed a home life of an ordinary married couple. It was a huge hit, and it had a big influence on many families, who later copied the characters' lifestyle.[59]

Throughout the years, the topics of sitcoms have naturally changed regarding social and political background, but the principle of the genre stayed the same.

2.7.2. Sitcoms and language

The authors of situation comedies usually write about real-life issues. Sometimes they use their own lives and experience as inspiration. By adding some humorous context to those everyday situations, it creates a funny plot. The writers try to make the content as much entertaining as they can for the viewers. The episodes are sometimes filmed in front of a real audience, whose laughter is then kept in the background. Alternatively, a laugh track is inserted in the show to encourage people to laugh.

Sitcoms are sometimes a combination of comedy and satire or parody. They usually criticize politics, different cultures and religions. Here are some illustrations from different sitcoms.

In How I Met Your Mother they use satire when discussing the Canadian nation. As one of the characters is Canadian, there are cases when the rest of the group is mocking Canadian customs and stereotypes.

In The Big Bang Theory appears religious satire on Jewish and Indian culture. The sitcom also satirically portrays the differences between geeks and other people.

In The Inbetweeners the aim of the writers is to satirically describe the coming of age of the main characters and what it is like to be a British teenager.

---

[59] Mary M. Dalton, The Sitcom Reader: America Viewed And Skewed (Albany: State University of New York Press, 2005), 49,87,90

The sitcom scripts portray relationships between friends, family members, or romantic relationships. The protagonists usually lead ordinary lives and deal with everyday life issues like money problems, career, health, and all kinds of relationships, which are probably the most important thing in their lives. The lives of the characters evolve slowly throughout the seasons. There are not any sudden changes in the main storyline.

Sitcoms have their characteristic structure. One episode usually lasts approximately 20 minutes. It consists of short scenes and usually includes a simple plot twist which is resolved by the end of the episode. The protagonists usually exchange short and simple sentences.

Since sitcoms imitate real lives of people, the characters' dialogues are very similar to everyday speech as well. This also affects the choice of language in terms of grammar, syntax and lexis.

Grammatical structures occurring in sitcom dialogues are usually simple and easy to comprehend. They include simple tenses, and the speakers usually avoid using advanced grammatical structures. [60]

The language used in sitcoms is informal. Informal English involves freely structured sentences and colloquial expressions. Those are expressions used mainly in spoken English in casual and familiar situations.[61] They include contractions, slang, neologisms, phrasal verbs, idioms and collocations, which are the main concern of this thesis.[62]

Slang refers to informal words or phrases typical for a particular social group.[63] In sitcoms often appears British or American slang.

In British sitcoms typically occur swear words. There are not many profanities in American sitcoms as they have more strict rules when it comes to mainstream broadcasts.

The essential part of sitcom is humour. The scripts include a lot of jokes and humorous comments, which very often repeat throughout the series. This is also connected with the term catchphrase. Those are expressions repeatedly used by a certain person.[64] They usually become clichés after some time or gradually disappear.

---

[60] Sun Quan, The Applicability in Selection of American sitcoms as Materials for Teaching English listening and speaking, 2016.
[61] Online Collins dictionary. (2023-06-08). https://www.collinsdictionary.com/
[62] David Crystal, A Dictionary of Linguistics and Phonetics (Hoboken: Wiley-Blackwell, 2008), 195
[63] Gregory James, Dictionary of Lexicography (Oxfordshire: Routledge, 1998), 127
[64] Merriam-Webster Dictionary. (2023-06-08). https://www.merriam-webster.com/

Another term for catchphrase is a running gag, a joke repeated multiple times by the same person.[65] A sitcom known for a high frequency of catchphrases is How I Met Your Mother. Here are some examples of catchphrases from the show:

- "What up!"
- "Awesome!"
- "Haaaave you met Ted?"
- "Suit up!"
- "Legen-wait for it-dary!"

Another example of a catchphrase is „Bazinga" from The Big Bang Theory. This phrase is usually said after a joke to emphasize it.

A key element of humour is irony and sarcasm. In every sitcom, there is usually one character constantly reacting ironically and sarcastically to the people around them. Typical examples are Chandler Bing from Friends, Sheldon Cooper from The Big Bang Theory, or Charlie Harper from Two and a Half Men.

The vocabulary of a sitcom is closely related to its topics. Because sitcoms mostly include various relationships, it is also the central topic of this genre. The characters usually discuss together their friendships, romantic relationships, partnerships, or family relations.

One advantage of sitcoms in terms of lexis is that they react more flexibly to new trends, and because of this they often include current topics. It is also reflected in the occurring vocabulary.

The vocabulary is also influenced by the physical settings of the show. The characters often visit public places like cafés, bars, restaurants, gyms, hospitals, offices, stores, schools, airports, et cetera. These different settings create various contexts and therefore influence the lexical words and phrases occurring in the episodes.

For instance, in How I Met Your Mother or The Big Bang Theory the characters often meet at a bar or a restaurant. There can be heard some beverage and food vocabulary as well as various phrases:

- "What can I get you?"

---

[65] Merriam-Webster Dictionary. (2023-06-11). https://www.merriam-webster.com/

- "I'll have…."
- "Cheers!"
- "Does anybody what anything else?"
- "Does anybody need coffee?"
- "Can I get some coffee, please?"
- "Are you ready to order?"

On the shows, the characters often go to the hospital. There can be heard medical vocabulary like *the emergency room, illness, disease, injury, accident, dislocated, diabetes, migraines, measle*s, et cetera.

When visiting a shopping centre, there appear clothing vocabulary *(department, second floor, top floor, on sale, dressing room*, et cetera) as well as phrases like

- "This one suits you."
- "You should try this on."
- "I will pick out something. "

At the airport, there are words like *boarding pass, departure, first class, economy class,* and others. The people use phrases like

- "Enjoy your flight."
- "Welcome aboard."
- "That flight has been cancelled/delayed."
- "Attention, passengers, we are ready for boarding."

The speech fluency and speed in sitcoms is similar to a regular performance of a native speaker. A common rate of speech of a person from an English-speaking country is from 150 to 170 words per minute. After running a test of speed on the characters from the sitcom

Friends, it showed that their speaking rate was 200 wpm.[66] This shows that even though they are only actors repeating their lines, they try to express themselves naturally.

In sitcoms, one can hear various accents, as there often occur actors of different nationalities and cultures. Especially the last couple of years a big emphasis has been put on the cultural diversity among the casts of tv shows and movies. Apart from American and British English, which are the most frequent ones, there are sometimes Australian, Indian, or African accents.

Another feature of sitcom is exaggerating reactions to situations. This sometimes leads to unnatural or incorrect intonation.

Even though the sitcoms dialogues reflect real-life situations and deal with everyday issues, there is the question of to which extent they are "real". The scripts are made-up by writers who have one main goal – to amuse the audience. They do not intend to portray realistic events. Given the fact that the length of an episode is twenty minutes,

2.7.3. Sitcoms in teaching

Sitcoms can be used for teaching the most common grammatical structures like present and past tenses, irregular or modal verbs.

By exposing to colloquial expressions in sitcom dialogues, it can help to acquire new everyday phrases idioms and collocations. They also include many phrasal verbs, which are essential in the English language, but for non-native speakers of English, it is often difficult to learn them. Through sitcoms, one can introduce various vocabulary, depending on their settings.

While using sitcoms for learning English, the content is presented in multiple means – apart from the auditory presentation, where the pupils listen to the characters' dialogues, there are also visual aids projected on the screen as well as subtitles either in the original language or students' mother tongue. While watching the episode, the students not only hear the speakers like in a regular listening exercise, but they also see how they communicate non-verbally. This may greatly increase the students' understanding of the viewed episode because

---

[66] Sun Quan, The Applicability in Selection of American sitcoms as Materials for Teaching English listening and speaking, 2016.

even if they do not fully comprehend the verbal part of the dialogue, the characters' gestures and facial expressions make it much easier.

2.8. CAE and TOEFL textbooks and language

In comparison with situation comedies, textbooks include a much broader scale of topics. They mainly include college-related themes, for example lectures, seminars, seminar papers and consultations.

Furthermore, the authors of textbooks do not have to follow any storylines, so there are almost no limitations of the content. Consequently, they include other topics like travelling, nature, sports, culture, food, art, fashion, sports, politics and many more.

They present various settings and provide rich contexts, such as university premises, lecture halls, professors' and admission offices, libraries, laboratories, et cetera.

For their variety of topics, contexts and physical settings they differ from the sitcoms, which are more limited in that sense. The reason for this diversity is that the authors aim to expose the learners to a wide range of vocabulary, grammar structures and phrases. After all, the main purpose of textbooks is to educate the users.

The main aim of sitcoms is to entertain the audience and humour plays an essential role in them. Textbooks, on the other hand, usually do not prioritize the appearance of humour. Similarly, they rarely include irony, sarcasm and swear words, because they are designed mainly as a learning tool, and they present clear and literal content.

Even though they include current topics, they are not as modern as sitcoms, because they lack the newest trends concerning, for example new technologies, recent cultural and political events, et cetera. It is because sitcom episodes usually come out every week, while textbooks take a lot more time to edit and publish.

The language of textbooks is far more formal than sitcoms. Even though there also occur colloquial expressions, they are not that common. Here are some examples of colloquialisms from the CAE and TOEFL textbooks:

- Whadda'ya mean?
- So, I'm outta luck?
- See ya!

- What's up?

The formality of textbooks is caused by the context of the dialogues. The participants of the dialogues are usually professors and pupils. The exchanges in these academic settings are very polite and impersonal. The speakers address each other by last names usually including titles. They express themselves in more complex sentences and discuss mainly academic issues like essays, presentations, curriculum, exams, et cetera. They avoid talking about any private matters such as personal relationships, money, or health problems.

There also occur conversations between classmates, which are less formal and include more colloquial expressions. Apart from school, they share their work experiences, memories from travelling, sports, or free time activities.

3. Practical part

The practical part aims to measure the authenticity of sitcom dialogues and conversations occurring in English grammar textbooks. These two sources are then compared together, and it is determined to which extent textbook dialogues are authentic. The hypothesis is that sitcom dialogues are more authentic than the conversations in textbooks. However, as mentioned previously, situation comedies are composed of dialogues that someone still makes up, and their purpose is mainly to amuse others. Consequently, they cannot be considered real-life conversations.

3.1. Methodology

I have chosen four different sitcoms (two from the British scene and two from the American) - *The Inbetweeners* and *Chewing Gum* from the British setting and *The Big Bang Theory* and *How I Met Your Mother* from the American. These four sitcoms were chosen, because they are considered the most famous ones among this genre, especially from the American setting. It is also easy to access them as they are available on many streaming services. It was not difficult to choose from the American sitcoms, because this genre is very popular in this given culture and there are plenty of them. On the contrary, the British authors create more comedy/drama tv shows and there are not many sitcoms as such.

After the selection of resources, I have downloaded subtitles of chosen sitcom episodes. This part was the easiest one because they are freely accessible. More problematic and necessary part was then to make some changes in the subtitle transcripts and delete some redundant text, which was a little bit time-consuming, as it was a lot of content.

As the next step, I have collected dialogues from listening transcripts appearing in textbooks. Here, it was crucial to deal with similar levels of English. The authors of sitcoms are native speakers of English, so the level of the language used in their scripts is proficient (level C2). To work with, for example B1 textbooks would not be effective, because they deal with much easier grammar, vocabulary, et cetera, and there would not be enough (if any) data to analyse.

Consequently, I have chosen to work with textbooks for advanced users of English (C1-C2 level) to balance the difference of language levels. After going through several of these textbooks I have realized that they do not contain any model dialogues in comparison with textbooks of lower levels. The authors reasonably presume that the learners are already experienced enough in the language.

I have therefore decided to work with CAE and TOEFL Practice tests textbooks because they include transcripts of dialogues of their listening exercises.

My goal was to create corpora consisting at least of 100 000 words, so the final results would be as conclusive as possible. One episode of mentioned sitcoms has approximately 5 000 words. I have used subtitles of 5 episodes of each sitcom, so each corpus had 25 000 words. I have then collected scripted dialogues from textbooks with the same number of words (25 000 each). The total number of words I have worked with was 125 000.

Once I have gathered all the needed data, I have uploaded it to the software #Lancsbox and created six corpora. Then I have conducted quantitative research.

3.2. #Lancsbox

#Lancsbox is used for corpus analysis designed at Lancaster University by Vaclav Brezina. The function of this software is to analyse uploaded data, sort it, display it visually and arrange it according to its frequency. This software makes it possible to work with an existing corpus or to upload a new one. It has six main tools: KWIC, GraphColl, Whelk, Words, Ngrams and Text.

KWIC is a tool that finds a chosen keyword (node) in context. *GraphColl's* function is to find a list of collocations appearing in the corpus. The results are also displayed in the form of a graph, where the keyword is in the middle and its collocations around it. *Whelk* presents the frequencies of the node in different corpus files. *Words* generates all the words of the corpus according to their frequency and is also able to compare two corpora with one another. It also shows the part of speech of each word in the list. *Ngrams* searches for sequences of words (minimum two words) and presents their frequencies. *Text* shows the whole context in which the keyword was used.

3.3. Research

During my research, I have used only the tools *Words* and *GraphColl*. Once I have understood how these tools function and how to work with them, it was quite simple to conduct the research and I have enjoyed doing it. I have examined both lexical and grammatical collocations occurring in the chosen sitcoms and textbooks. The study was divided into three main stages described below.

3.3.1. Words

In the first stage of my research, I have been using the tool *Words* to create a list of ten most frequent nouns occurring in each sitcom and textbook of my corpus. To do so, I have selected the type of searching to *Lemma* to limit the findings only to their basic forms. Subsequently, I have set a filter **_N*, so only nouns would appear in the list. As a final step, I have gone through the list and skipped all proper nouns and got the final list.

This way, after I have tested all six corpora I have uploaded, I have got 60 most frequent nouns occurring in the chosen sitcoms and English textbooks.
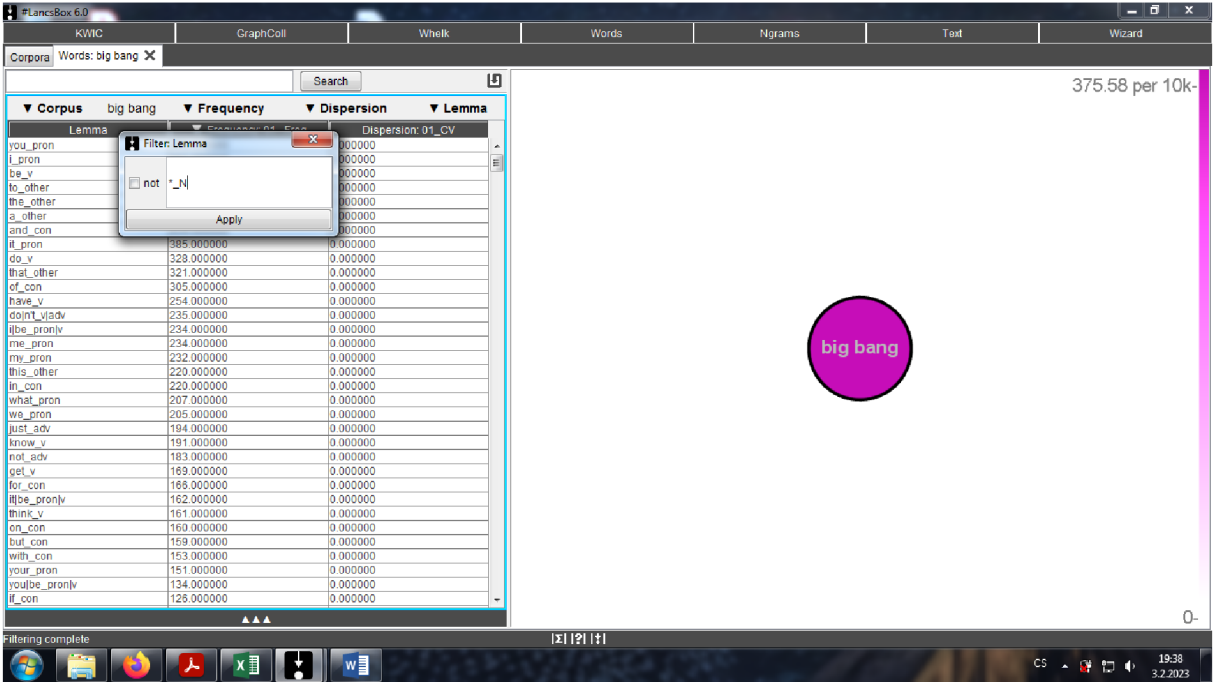
Fig. 4 – Working with the tool Words *in #Lancsbox*

TBBT		Frequency
1.	time	62
2.	guy	49
3.	thing	48
4.	friend	39
5.	kid	33
6.	way	31
7.	fun	30
8.	woman	26
9.	year	24
10.	life	24
TOTAL		366

HIMYM		Frequency
11.	time	61
12.	room	52
13.	thing	50
14.	mom	39
15.	year	36
16.	guy	36
17.	gift	35
18.	man	33
19.	way	33
20.	girl	29
TOTAL		404

Chewing gum		Frequency
31.	man	63
32.	thing	55
33.	time	55
34.	mum	42
35.	sex	37
36.	friend	37
37.	guy	36
38.	people	35
39.	girl	32
40.	life	27
TOTAL		419

TOEFL		Frequency
41.	class	78
42.	time	63
43.	people	61
44.	thing	60
45.	way	42
46.	week	40
47.	computer	35
48.	problem	33
49.	kind	33
50.	paper	31
TOTAL		476

	CAE	Frequency
51.	people	147
52.	thing	86
53.	time	61
54.	year	59
55.	way	55
56.	day	48
57.	world	32
58.	shop	24
59.	family	24
60.	book	23
TOTAL		559

Fig. 5 - 60 most frequent nouns in sitcoms and textbooks

Below there are graphical representations of the occurring words in the analysed sitcoms and textbooks.
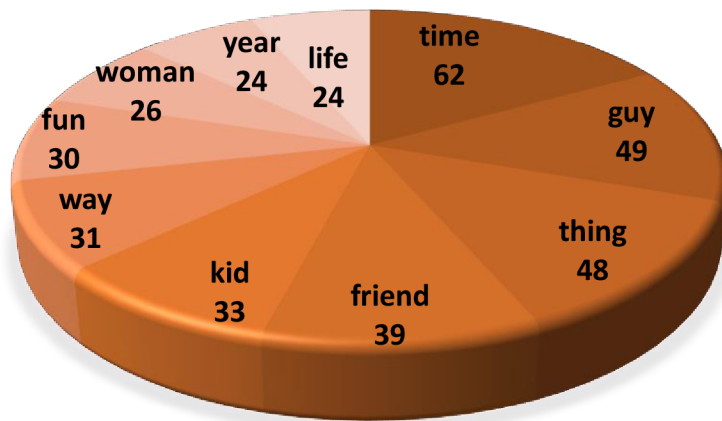
Fig. 6 – Top 10 occurring words in The Big Bang Theory
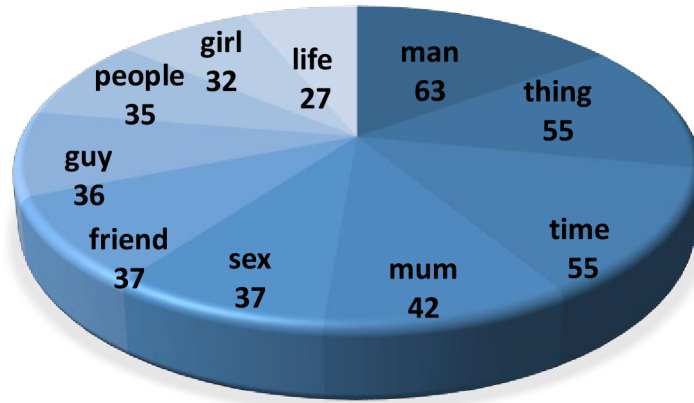
Fig. 7 – Top 10 occurring words in Chewing Gum
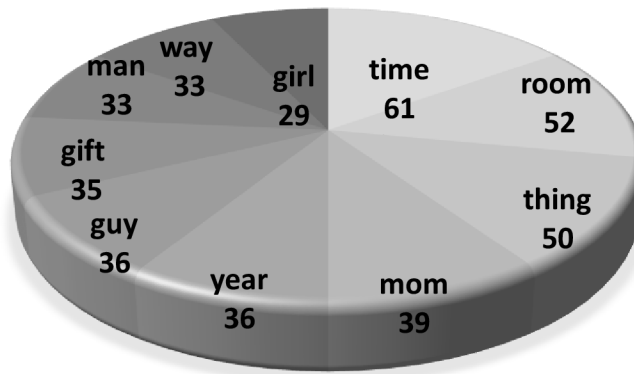
Fig. 8 – Top 10 occurring words in How I Met Your Mother
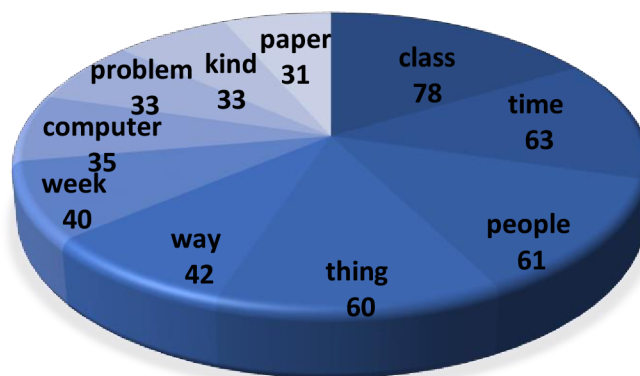
Fig. 9 – Top 10 occurring words in TOEFL textbooks

Fig. 10 – Top 10 occurring words in Inbetweeners
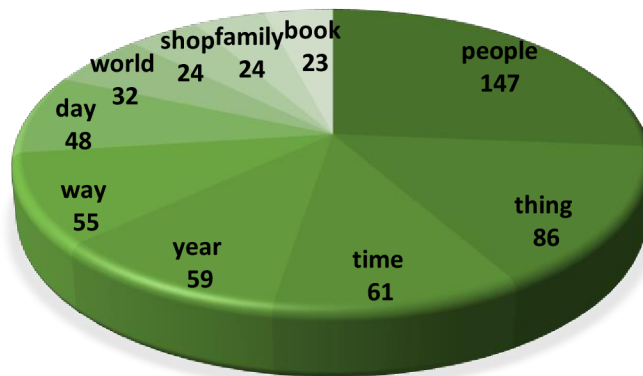
Fig. 11 – Top 10 occurring words in CAE textbooks
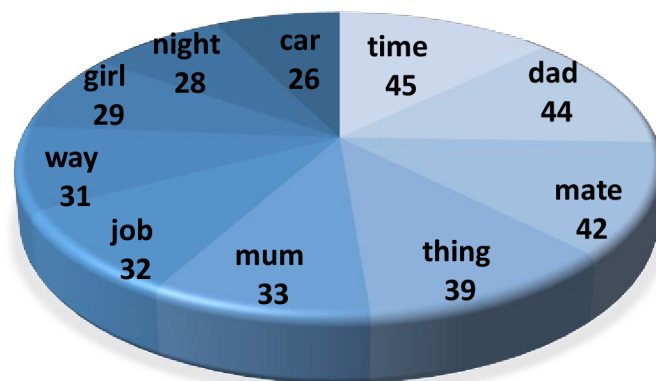
Fig. 12 – Top 10 occurring words in Inbetweeners - censored version

Looking at the pie charts above, we can see that several words repeat themselves. The word *time* is present in every of the examined corpus, in half of the cases it is listed as the most frequent word. Another word that appears in every pie chart is *thing.* The third most common word is *way*.

Comparison of findings

After comparing the results of the American sitcoms, half of the results (five out of ten lemmas) were the same: *time, guy, thing, way* and *year.* In the case of the British sitcoms, four out of ten words were the same: *time, thing, mum* and *girl.*
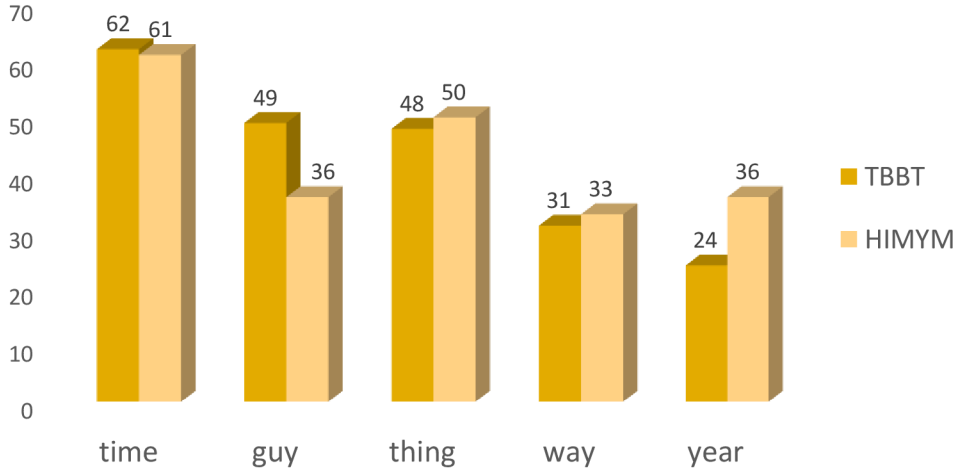
Fig. 13 – Comparison of frequent words in The Big Bang Theory and How I Met Your Mother
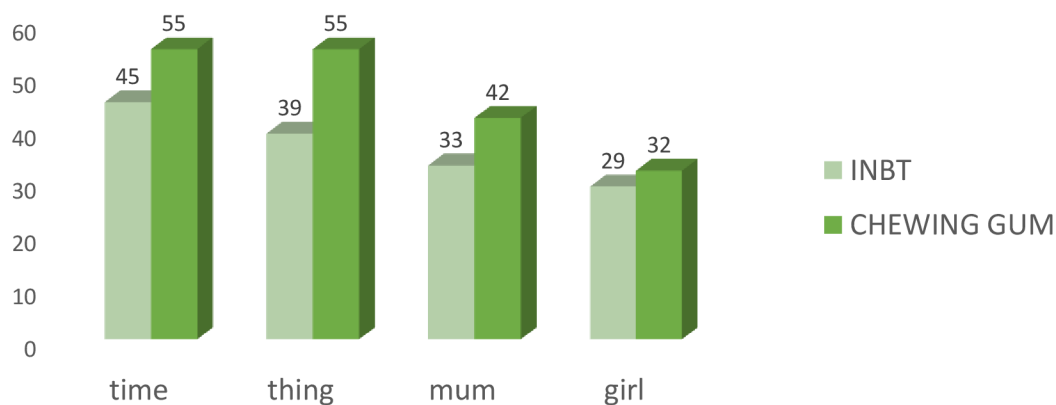
Fig. 14 – Comparison of frequent words in Inbetweeners and Chewing Gum

The graphs show that the number of occurrences of words in sitcoms was in most cases the same with minimal deviation.

I was then comparing the sitcoms with the textbooks, which was the main focus of the practical part. As for the TOEFL textbooks, The Big Bang Theory and How I Met Your Mother, *time, thing* and *way* were the words that appeared in all three of them.
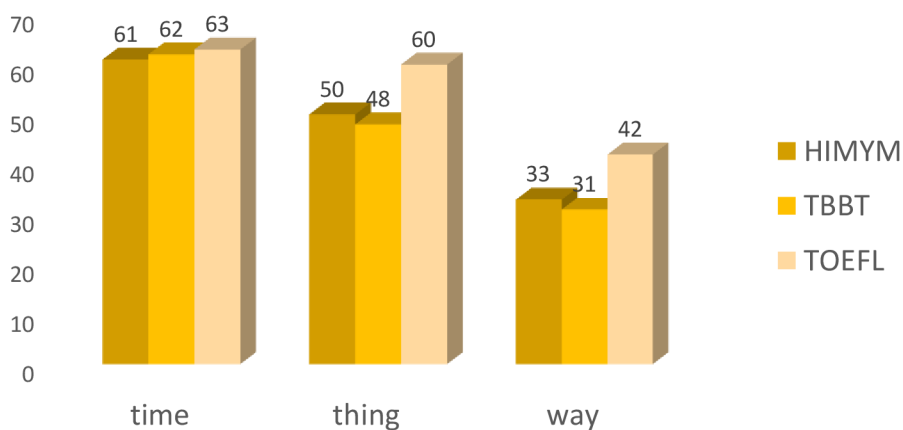
Fig. 15 – Comparison of frequent words in How I Met Your Mother, The Big Bang Theory and TOEFL textbooks

In British sitcoms and textbooks, the values were more varied, especially in British sitcoms and textbooks. In the case of CAE textbooks and The Inbetweeners, the most frequent words were *time, thing* and *way,* while in the case of CAE and Chewing Gum it was *thing, time*

and *people*. The occurrences in CAE textbooks were distinctly higher in comparison with the sitcoms. In the case of *people* in figure 17, the difference was particularly big. Perhaps the reason for this is that the CAE textbooks are written more neutrally. In Chewing Gum, the characters address each other more informally, and therefore there rather appear expressions like "guys, buddies, mates", et cetera.
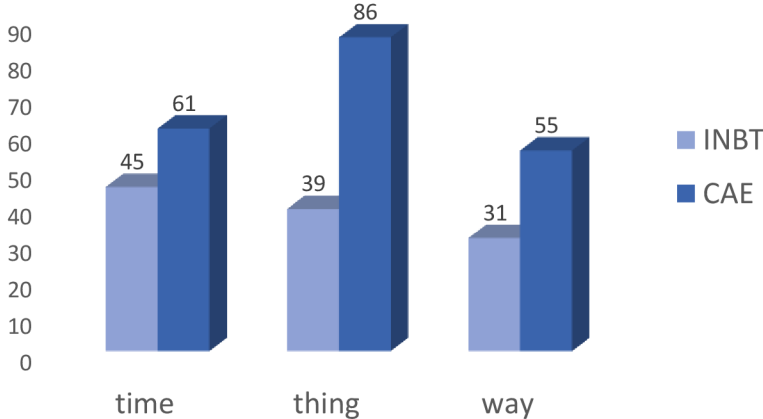
Fig. 16 – Comparison of frequent words in Inbetweeners and CAE textbooks
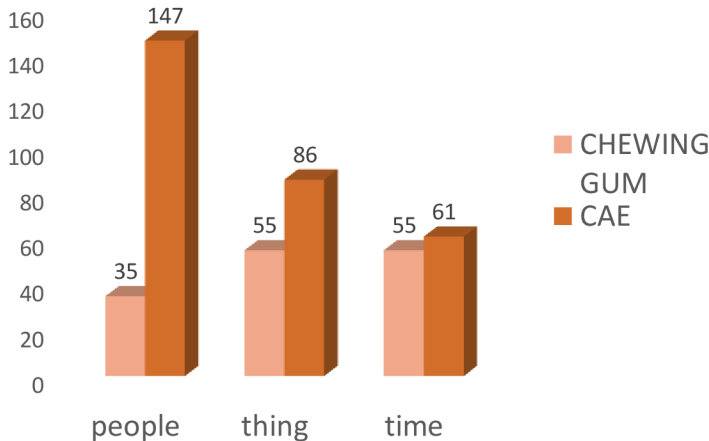
Fig. 17 – Comparison of frequent words in Chewing Gum and CAE textbooks

3.3.2. GraphColl

In the second stage of my research, I have worked with the tool *GraphColl*. Again, I have selected the type to *Lemma* and the same filter. Then I have switched the Statistic value in Threshold to 2.0. This helped me to decrease the number of findings, so the graph would not be too full. Finally, I have set the Span from -5 to 5.

By setting the span to its maximum I was able to get more findings. However, it also meant that some collocating words came from different sentences and thus some results showed more contextual collocations than phraseological.

I have used the findings from the tables above and looked up their most frequent collocations (again avoiding proper nouns).

This way, I have been able to find the noun + noun collocations. I have been also interested in the noun + adjective and noun + verb collocations. So, I have selected two other filters to find those as well. (*_ADJ; *_V).
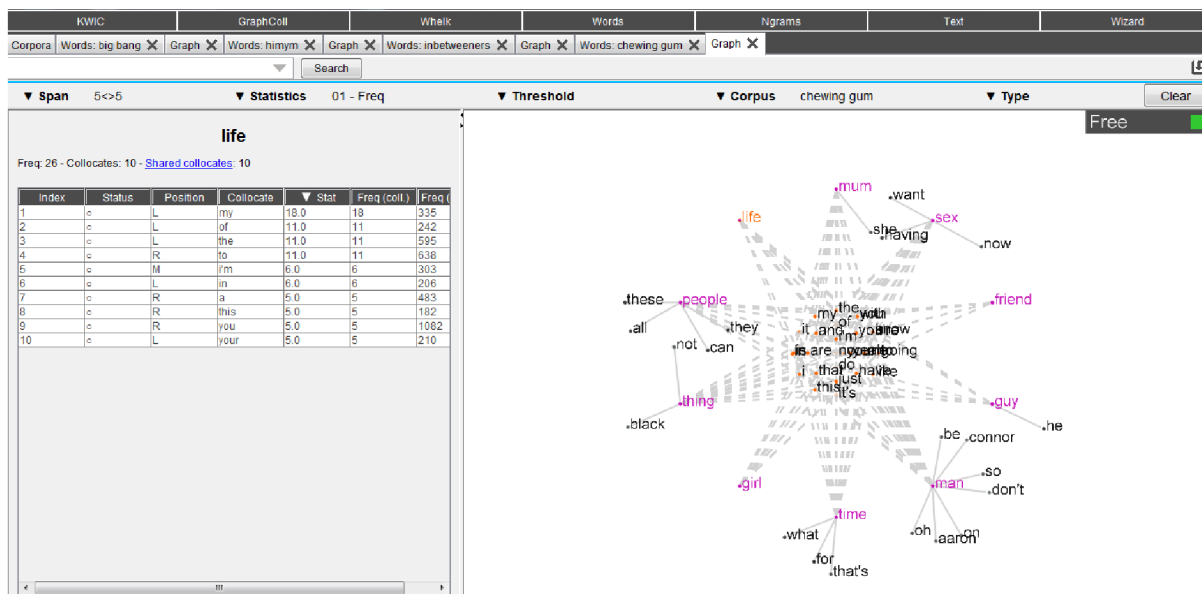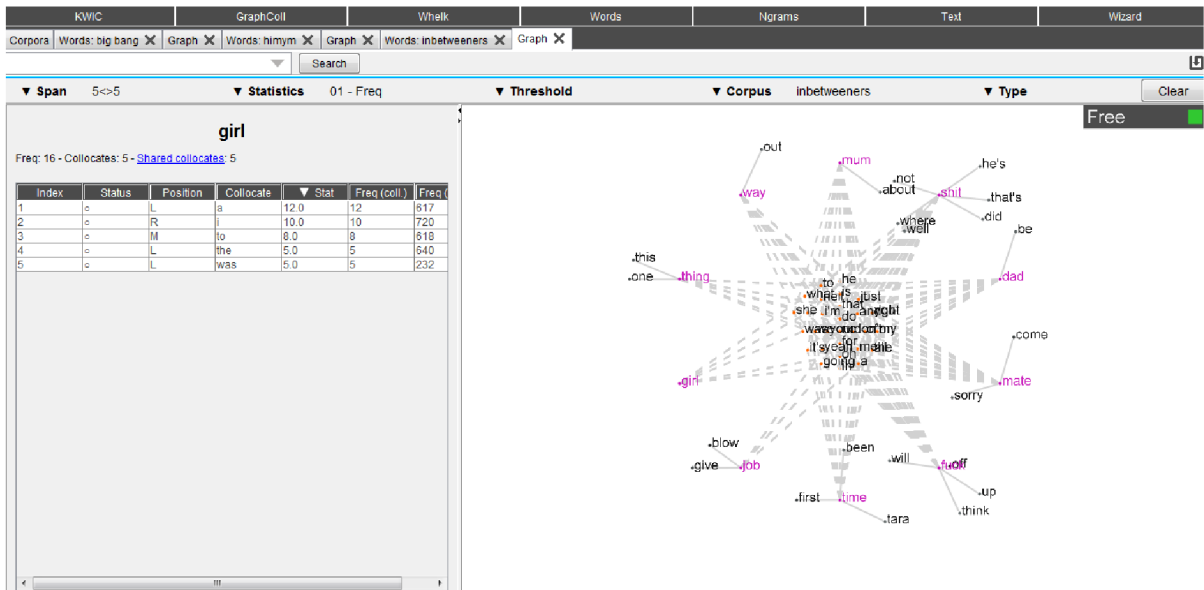
Fig. 18 – Working with the tool GraphColl *in #Lancsbox to find collocations*

The findings were diverse and broad, and each sitcom and textbook offered different collocations. I have limited the list again to ten most frequent collocations (see Appendix).

The results were influenced by the topics of the sitcoms and textbooks as the following example demonstrates.

In the case of The Big Bang Theory (its last season), the characters were either married or engaged, and so they dealt with relationship issues, family and children stuff. The results showed that there collocated words such as *marry, kid, guy, woman*, et cetera.

The whole last season of How I Met Your Mother centred around a wedding of two of the protagonists. It took place in a hotel and all the characters are there. The collocations that appeared in the dialogues were therefore associated with wedding and hotel, for example *wedding gift, ring, anniversary, room service* and others.

In The Inbetweeners, there was a group of boys attending comprehensive school who enjoyed their teenage years partying as well as experiencing their first relationships. When looking at the censored version, there occurred collocations with *girl, pub, night, car, fun, kiss, smoke*, et cetera.

The main character of Chewing Gum came from a strict Christian family and was trying to live a normal adolescent life and experience all that comes with it. As the characters were religious, there were a lot of collocations related to religion – *Jesus, God, lord*, et cetera.

As stated earlier, the CAE and TOEFL textbooks are primarily school-related, but they cover many different topics as well. The results found there were, therefore, more diverse presenting collocations with *people, class, college, dorm, book, world, travel, money, research, nature, family* and many more.

Frequent collocations

This part is dedicated to the examination and comparison of sitcoms with textbooks in terms of the occurrence of the same collocations.

In the case of N + N collocations, there were not many matches. Moreover, the results were not relevant as the same collocations occurred only in the textbooks or only in the sitcoms. In the case of ADJ + N collocations, the co-occurrences were a bit higher. As for V + N collocations, the most frequent verbs collocating in the corpora were *be, have* and *do*, which is not surprising as these are the primary auxiliary verbs.

Collocation	Co-occurrence					
	TBBT	HIMYM	TOEFL	INBT	CHEWING GUM	CAE
people + thing	x	x	3	x	x	4
thing + kind	x	x	2	x	x	2
mom (mum) + dad	x	x	x	4	4	x
first + thing	3	x	3	x	x	x
first + time	5	x	4	9	x	5
good + friend	14	x	x	x	3	x
good + thing	x	9	5	x	x	4
good + time	2	3	2	x	3	2
good + way	x	2	x	x	x	2
last + time	2	3	x	x	3	2
long + time	8	x	x	x	2	x
next + time	3	4	4	x	x	x
okay + time	x	x	x	2	4	x
other + people	x	x	6	x	2	12
same + thing	x	x	3	x	2	x
whole + thing	3	x	x	2	x	3
more + people	x	x	6	x	x	8
other + thing	x	x	9	x	x	2
other + way	x	x	2	x	x	2
be + time	22	20	16	16	13	24
be + thing	10	16	21	6	19	28
be + way	11	11	12	7	10	15

Fig. 19 – Table of frequent collocations and their co-occurrences

The following graphs illustrate the most significant collocations of my research - collocations with the highest number of occurrences, and collocations repeating in most of the analysed sitcoms and textbooks.

Figure 20 shows three most repeated collocations in the research. At first glance, it is clear that the collocation *first + time* dominates the graph. It occurs the most on the show The Inbetweeners. There it is mostly related to having sex for the first time, since the characters are all teenagers, and mainly talk about their relationships.

Most repeated collocations
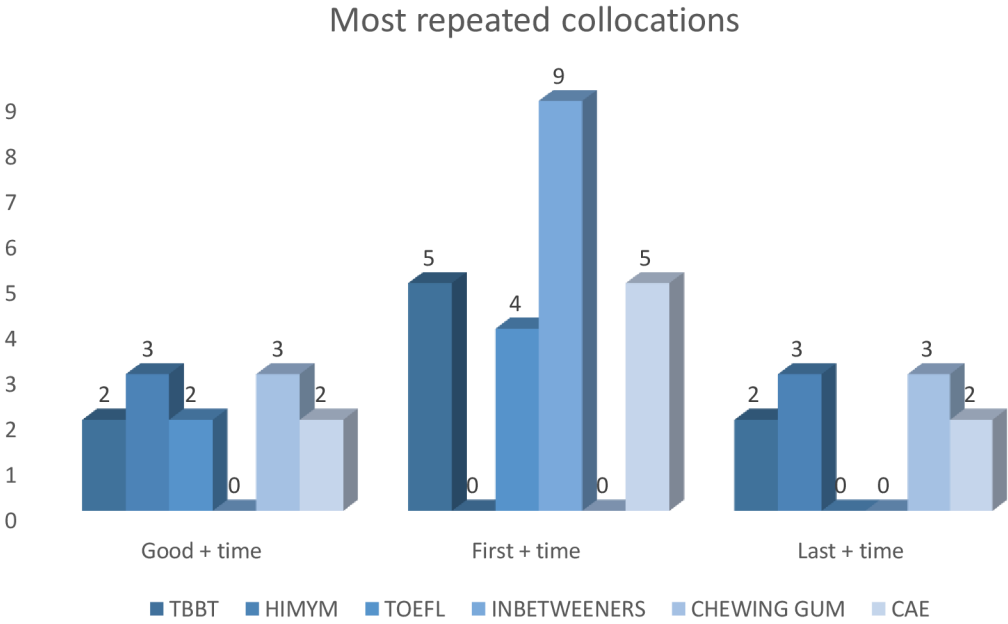
Fig. 20 – The most repeated collocations in sitcoms and textbooks

Figure 21 illustrates the collocations with the highest frequencies. Here, the results were surprising. It showed that the occurrence of a certain collocation is sometimes individual and even two very thematically similar sitcoms as TBBT and HIMYM show very different results (as in *good + friend*).
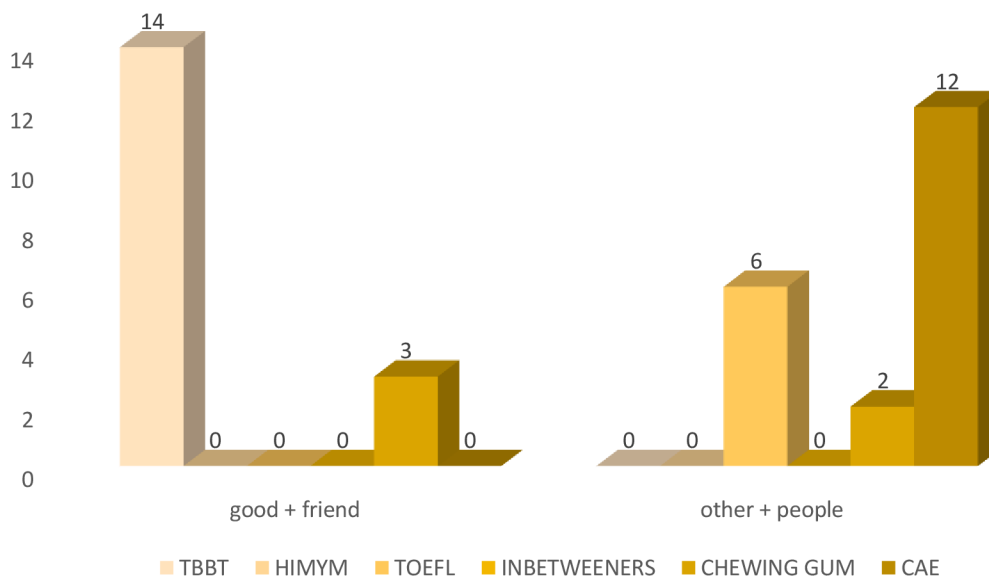
Fig. 21 – Collocations with the highest frequencies

Figure 22 depicts frequent collocates of the verb *to be*. Here prevailed the collocation *be + thing*. The analysed dialogues included collocations like "the thing is…, this is the kind of thing…, the most important thing is…", which are all common collocations in English.
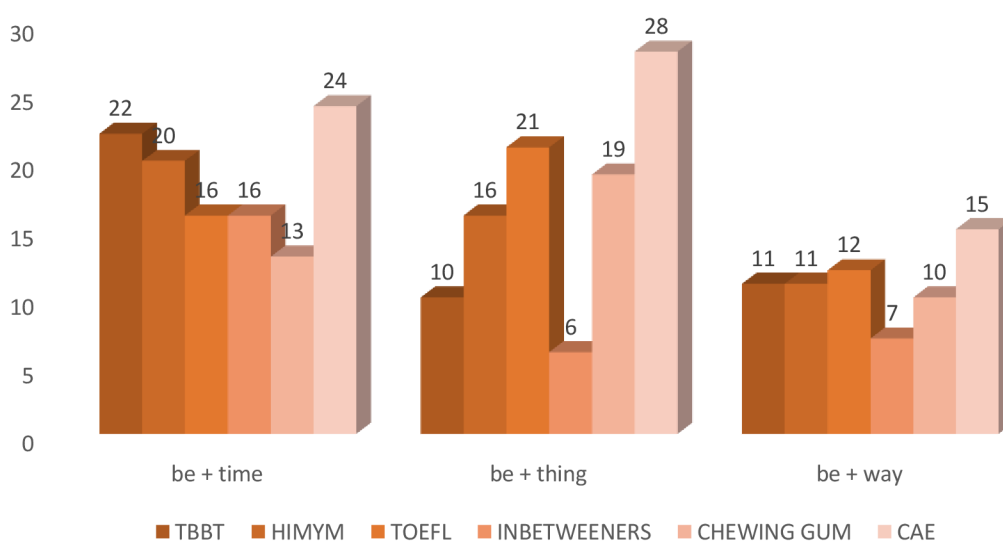
Fig. 22 – Table of frequent collocations and their co-occurrences

51

The use of the national corpora

In the final part of my research, I have worked with one British and one American corpus. My original plan was to use the British National Corpus and The Corpus of Contemporary American English. Due to some difficulties, I have chosen a different American corpus of a smaller extent instead. So, my final choice of corpora was the BNC and The Open American National Corpus (OANC). This corpus I have uploaded to #Lancsbox and once again worked with the function GraphColl.

Collocation	Co-occurrence		Per mil.	
	OANC	BNC	OANC	BNC
people + thing	263	135	82,2	13,6
thing + kind	500	280	156,3	28,3
mum (mom) + dad	67	277	20,9	27,8
first + thing	129	424	40,3	42,8
first + time	264	691	82,5	69,8
good + friend	51	24	15,9	2,4
good + thing	330	301	103,1	30,4
good + time	222	238	69,4	24
good + way	144	178	45	17,9
last + time	205	831	64,1	83,9
long + time	527	939	164,7	94,8
next + time	102	467	31,9	47,2
okay + time	45	4	14,1	0,4
other + people	464	1106	145	111,7
same + thing	405	552	126,6	55,8
whole + thing	256	287	80	28,9
more + people	397	593	124,1	59,9
other + thing	645	554	201,6	55,9
other + way	199	574	62,2	57,9
be + time	275	839	85,9	84,7
be + thing	444	480	138,8	48,5
be + way	272	565	85	57,1

Fig. 23 – Table of frequent collocations and their co-occurrences in the national corpora

I dealt only with the spoken sections of these corpora, where the BNC contains 9,963,663 words and OANC 3,217,772 words. As these two corpora differ quite a lot in size, the tables below include the occurrences of the analysed collocations as well as their appearance per million words. This way, the results would be more informative.

I aimed to provide a broader perspective on the usage of the findings of my research in everyday spoken English.

2.3.4. Commentary and discussion of results

Here, I am going to comment on collocations with the highest co-occurrences in both my and the national corpora, which were already graphically portrayed above.

It is clear, that to compare my corpus and the national corpora is for their different sizes meaningless. The main intention is therefore to compare only the results of my corpus. The BNC and OANC data are to give a general idea of the usage.

In the case of American sitcoms and textbooks, there were several collocations worth mentioning. Even though *good + friend* occurred in TBBT in my corpus most frequently, there was no appearance of it in TOEFL. It was the same with *long + time*, which also repeated quite a lot in TBBT and in TOEFL not whatsoever. On the other hand, collocations *next + time* and *good + time*, co-occurred similarly both in the sitcoms and the textbooks.

As for the British setting, the most frequent collocations were *other + people* and *more + people*, but appearing only in CAE, while in the sitcoms their appearance was either rare or non-existent. But for example, when looking at the collocations with time (first + time, good + time, last + time), they co-occurred with similar frequency.

To summarize, collocations with *time* (*good + time, last + time, next + time*) repeated the most in the list.

While thinking about the reason why these particular collocations were the most frequent ones, I have concluded that given the fact that the characters are good friends and spend a lot of time together, it is only natural that they talk about recent events or discuss plans. It is therefore clear why collocations with *time* were so frequent in the dialogues.

After going through the co-occurrences from the national corpora, I have noticed that when the frequencies were high in the national corpora, the co-occurrences in my corpus also tended to be higher, especially in the case of the TOEFL and CAE textbooks. It showed that the authors of textbooks are use corpus data while designing their textbooks to make them more authentic. This was a great discovery for me, even though it partly disproved my hypothesis. However, I was pleasantly surprised that these grammar textbooks include more authentic material than I have originally thought. Textbook transcripts were also more in accordance with the BNC and OANC than sitcom dialogues with the national corpora.

Both textbook and sitcom dialogues contained a lot of real-life expressions. However, sitcom conversations presented a smaller diversity of collocations than the textbook dialogues. It corresponds with the characterisation of sitcoms and textbooks already

mentioned earlier - that sitcoms are limited to the content as they follow one main storyline, whereas textbooks present various independent topics.

3.4. Obstacles during research

I was facing some difficulties while doing my research. When creating my own corpus, it was necessary to make some corrections for #Lancsbox to process the textbook transcripts and sitcom subtitles. This meant mainly to remove any symbols which were not regular letters of the alphabet.

Then, when I have been working with the corpus, I had to omit any proper nouns appearing in the list. Also, I have skipped any catchphrases occurring in the corpus (for example, "Bazinga!" in TBBT, "Legendary!" in HIMYM, et cetera).

Another thing was the presence of swear words in the sitcoms, which distorted the results, and led me to create censored versions.

The most problematic task was to work with national corpora, especially the COCA. For its size (127,396,916 words), it was sometimes impossible to find certain collocations, because the system reported errors, as in the case of "people", "other" or "time". I have therefore decided to use another corpus.

The word **other** occurs **1,454,276** times in the corpus.

In COCA, you can usually find the collocates for high-frequency words like this, as long as:

1) you search by "lemma" (capitalize the "dictionary form" of the word(s), e.g. **DECIDE** instead of **decides**)

2) you leave the "span" set to 4 words left and 4 words right, and

3) you don't limit by section (e.g. by time period, genre, or dialect, depending on the corpus).

4) you don't choose to see the frequency by section (i.e. you need to de-select the box to the left of "Sections" in the search form).

You might also re-do the search with one of the following:

1. Reverse the WORD and COLLOCATES fields, by putting the least frequent word in the [WORD] field, or

2. Do the search as a string search, e.g. **ADJ health**, instead of **health + ADJ** collocates, or **VERB the money** instead of **money + NOUN** collocates.

Fig. 24 – Error reporting while working with the COCA

4. Conclusion

This thesis aimed to measure the authenticity of model dialogues in grammar textbooks of advanced English.

The theoretical part dealt with the communication process, mainly with its pragmatic aspect. The practical part focused on the analysis of collected dialogues and evaluating whether they consist of genuine speech. The main area of interest were the most frequent nouns and words they collocate with (noun + noun, adjective + noun and verb + noun collocations).

The hypothesis was that sitcom dialogues are more authentic than textbook dialogues. The results showed several things. Firstly, the collocations with time co-occurred the most frequently. Secondly, the textbook transcripts contain wider diversity of collocations.

To measure the authenticity, the findings were run through the British National Corpus and the Open American National Corpus. It was discovered that textbook transcripts correspond more often (in comparison with the sitcom dialogues) with the national corpora.

An interesting thing was that in some instances, a high-frequency sitcom co-occurrence (as in good + friend in TBBT) meant a rare frequency in the national corpora.

The most surprising discovery of the research was the measured authenticity of TOEFL and CAE textbooks transcripts, which was higher than expected. Initially, the author assumed that these English textbooks are composed of phrases and vocabulary that are rarely used by native speakers. The research showed the opposite - that these model dialogues include a lot of everyday collocations and so they can be considered natural and spontaneous.

The result of the practical part was that even though the values from the created and the national corpora differed sometimes, the numbers showed that both the grammar textbooks and sitcoms dialogues included authentic material.

While conducting this research, only advanced textbooks were analysed. It would be interesting to expand this study and examine upper-intermediate or intermediate English textbooks used for secondary and high school education. This way, the teachers would be able to choose genuine materials for their pupils more easily.

This thesis proposed the possibility of teaching phrases and vocabulary with the use of sitcoms. As they can develop both productive and receptive skills, it is possible to use them

for various tasks, for example as a presentation of new grammatical structures, phrases and vocabulary, as a listening quiz, or as a revision of already discussed topic.

The research showed that they contain some authentic collocations. Nevertheless, further research on phraseological units would be useful, for instance, the occurrence of phrasal verbs.

The last point is about the occurrence of swear words in some of the examined data. When using sitcoms as a learning material, especially for younger students, one should also have in mind that some of them (especially those of British origin) are not censored.

5. Resumé

Téma této diplomové práce zní *Zkoumání míry autentičnosti dialogů v učebnicích anglického jazyka a scénářích situačních komedií.* Klade si za cíl zjistit míru autentičnosti vzorových dialogů uvedených v učebnicích anglického jazyka pro pokročilé využívaných českými studenty angličtiny. Tato práce zkoumá také scénáře nejznámějších situačních komedií americké a britské produkce. Hypotéza zní, že scénáristé situačních komedií předkládají reálnější dialogy než autoři učebnic angličtiny.

Teoretická část je zaměřena na komunikaci jakožto nepostradatelnou součást společnosti a života lidí. Pozornost je věnována procesu dorozumívání, nejznámějším modelům komunikace a komunikačním kanálům. Dále jsou zmíněny typy komunikace a jejich odlišnosti. Následně autorka hlouběji rozebírá průběh komunikace a poukazuje na její pragmatickou rovinu. Zabývá se kontextem a jeho významem v komunikaci. Pozornost je poté věnovaná definování a klasifikaci kolokací. Dále je popsána korpusová metoda výzkumu kolokací a její přínos. Na závěr teoretické části je charakterizován žánr situační komedie, jeho vznik, historický vývoj, a scénáře sitcomů jsou následně rozebrány z hlediska jazyka.

V praktické části jsou nejprve ze čtyř zvolených situačních komedií (dvou britské a americké produkce) nasbíraná potřebná data. Je využito veřejně dostupných titulků v původním znění. Stejným způsobem jsou nashromážděna data z učebnic britské a americké angličtiny pro pokročilé uživatele (úroveň C1-C2). Je pracováno s učebnicemi pro přípravu na CAE a TOEFL zkoušku.

Dále je představen korpusový software #Lancsbox a jeho základní funkce, s jehož pomocí je proveden výzkum. Nashromážděná data jsou do softwaru nahrána a je vytvořen seznam nejčastějších podstatných jmen a jejich frekvence, přičemž jsou vynechána všechna vlastní jména. Kvůli výskytu vulgarismů v některých seriálech jsou uváděny také cenzurované verze. Výsledky jsou následně mezi sebou porovnány.

V druhé fázi výzkumu jsou ke zjištěným podstatným jménům přiřazeny podstatná, přídavná jména a slovesa, která s nimi nejčastěji kolokovala. Data jsou mezi sebou opět porovnána a je vytvořen seznam nejčastějších kolokací a jejich výskytů v sitcomech a učebnicích angličtiny.

Poslední fází praktické části je samotné měření autentičnosti modelových dialogů. Naměřené hodnoty z učebnic a sitcomů jsou srovnávány s národními korpusy britské a

americké angličtiny. Původním záměrem autorky bylo využít Britský národní korpus (BNC) a Korpus současné americké angličtiny (COCA). Kvůli jeho velkému rozsahu bylo ale s americkým korpusem obtížné pracovat, a proto byl zvolen menší korpus (OANC). Během práce s národními korpusy byly využity pouze části obsahující mluvený jazyk.

Ačkoliv autorka předpokládala, že tvůrci učebnic angličtiny předkládají umělé (nepřirozené) dialogy, výsledky výzkumu ukazují, že obsahují každodenní fráze i slovní zásobu.

Výsledkem práce je poukázání na fakt, že situační komedie i učebnice pro pokročilé uživatele angličtiny obsahují autentický materiál.

6. Bibliography

ARGYLE, Michael. *Bodily Communication*. New York: International Universities Press, 1975.

BIBER, Douglas. *The Cambridge Handbook of English Corpus Linguistics.* Cambridge: Cambridge University Press, 2015.

CARTER, Ronald. *Cambridge Grammar of English*. Cambridge: Cambridge University Press, 2006

ČERMÁK, F., M. ŠULC. *Kolokace*. Praha: NLN, 2006.

CHANDLER, Daniel. *A Dictionary of Media and Communication*. Aberystwyth University: Oxford University Press, 2020.

CRUSE, D. A. *A Glossary of Semantics and Pragmatics (Glossaries in Linguistics)*. Edinburgh: Edinburgh University Press, 2006.

CRUSE, D. A. *Lexical Semantics (Cambridge Textbooks in Linguistics)*. Cambridge: Cambridge University Press, 1986.

CRYSTAL, David. *A Dictionary of Linguistics and Phonetics*. Hoboken: Wiley-Blackwell, 2008.

CRYSTAL, David. *The Cambridge Encyclopedia of the English Language*. Cambridge: Cambridge University Press, 2019.

DALTON, Mary M. *The Sitcom Reader: America Viewed And Skewed*. Albany: State University of New York Press, 2005.

DEVITO, Joseph A. *Základy mezilidské komunikace*. Praha: Grada Publishing, 2001.

FAIRCLOUGH, N. *Analysing Discourse: Textual Analysis for Social Research*. London: Routledge, 2003.

FLOYD, Kory. *Interpersonal Communication*. USA: McGraw-Hill Education, 2011.

GRANGER, S., F. MEUNIER. *Phraseology: An interdisciplinary perspective*. Amsterdam: John Benjamins Publishing Company, 2008.

HALLIDAY, M., CH. MATTHIESSEN. *Halliday's Introduction to Functional Grammar*. London: Routledge, 2014.

HALLIDAY, Michael. *Lexicology: A Short Introduction*. London: Continuum, 2007.

HILL, J., M. LEWIS. *LTP Dictionary of Selected Collocations*. Boston: Cengage Learning, 1997.

HUANG, Y. *Pragmatics*. Oxford: Oxford University Press, 2014.

JAMES, Gregory. *Dictionary of Lexicography*. Oxfordshire: Routledge, 1998.

KŘIVOHLAVÝ, Jaroslav. *Jak si navzájem lépe porozumíme*. Praha: Svoboda, 1988.

LINDQUIST, Hans. *Corpus Linguistics and the Description of English*. Edinburgh: EUP, 2009

LIPKA, Leonhard. *An Outline of English Lexicology*. Berlin: De Gruyter, 1992.

LYONS, John. *Linguistic Semantics: An Introduction*. Cambridge: Cambridge University Press, 1995.

MCCARTHY, M. *English Collocations in Use Advanced*. Cambridge: Cambridge University Press, 2017.

MCCARTHY, Michael. *English Collocations In Use*. Cambridge: Cambridge University Press, 2005.

MCENERY, Tony. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press, 2011

O'DELL, Felicity. *English Collocations in Use Advanced*. Cambridge: Cambridge University Press, 2017.

QUIRK, Randolph. *A Comprehensive Grammar of the English Language.* London: Longman, 1989

RADFORD, Andrew. *English Syntax: An Introduction*. Cambridge: Cambridge University Press, 2004.

RADFORD, Andrew. *Syntactic Theory and the Structure of English: A Minimalist Approach.* Cambridge: Cambridge University Press, 1997.

SAMPSON, Geoffrey. *Corpus Linguistics: Readings in a Widening Discipline*. London: Continuum, 2005.

SCHUMACHER, P. a R. FINKBEINER. *What is a Context?: Linguistic Approaches and Challenges*. Amsterdam: John Benjamins Publishing Company, 2012.

SHANNON, Claude. *The Mathematical Theory of Communication*. Chicago: University of Illinois Press, 1963.

VYBÍRAL, Zbyněk. *Psychologie komunikace*. Praha: PORTÁL, 2013.

WINDAHL, Sven, Denis MCQUAIL. *Communication Models for the Study of Mass Communications*. Abingdon: Routledge, 2013.

WOOD, Julia T. *Interpersonal Communication : Everyday Encounters*. The University of North Carolina at Chapel Hill: Wadsworth Publishing, 2012.

YULE, G., H. WIDDOWSON. *Pragmatics: Introduction to Language Study Series*. Oxford: Oxford University Press, 1996.

Articles

COBLEY, Paul. *Communication: Definitions and Concepts* [online]. 2008-06-05 [cit. 2023-04-12]. Available at: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781405186407.wbiecc071.

DĄBROWSKA, Ewa. Words that go together: Measuring individual differences in native speakers' knowledge of collocations. *The Mental Lexicon*. John Benjamins Publishing Company, 2014, **9**(3), 401-418.

JANOUŠEK, Jaromír. Intrapersonální komunikace a vnitřní řeč. *Československá psychologie* [online]. 2001, **45**(6), 481-492 [cit. 2023-04-12]. Available at: http://cspsych.psu.cas.cz/result.php?id=270

QUAN, Sun. The Applicability in Selection of American sitcoms as Materials for Teaching English listening and speaking, 2016.

Statista Research Department. Netflix subscription count in Czechia 2021. *Statista* [online]. 2022-07-07 [cit. 2023-04-12]. Available at: https://www.statista.com/statistics/1285620/netflix-subscriptions-czechia/

WebMD Editorial Contributors. Why Do People Talk to Themselves?. *WebMD* [online]. 2021-06-28 [cit. 2023-04-12]. Available at: https://www.webmd.com/balance/why-people-talk-to-themselves

Other online sources

Merriam-Webster Dictionary [online]. [cit. 2023-06-08]. Available at: https://www.merriam-webster.com/

Cambridge dictionary [online]. [cit. 2023-04-12]. Available at: https://dictionary.cambridge.org/

Online Collins dictionary [online]. [cit. 2023-06-08]. Available at: https://www.collinsdictionary.com/

Skretvedt, R., Sterling, Christopher H., Radio. *Encyclopedia Britannica* [online]. [cit. 2023-04-12]. Available at: https://wwwritannica.com/topic/radio.

Corpora

Davies, Mark. (2008-) *The Corpus of Contemporary American English (COCA)*. Available online at https://www.english-corpora.org/coca/.

BNC Consortium, *The British National Corpus, XML Edition*, 2007, Oxford Text Archive, URL: https://www.english-corpora.org/bnc/

Brezina, V., Weill-Tessier, P., & McEnery, A. (2021). #LancsBox v. 6.x. [software package]. Available at http://corpora.lancs.ac.uk/lancsbox/

7. Appendix

Collocating words

The Big Bang Theory

N + N

time	thing	friend	fun
charity	fun	world	night
nickel	guy		number
place			thing
pumpkin			
sex			
week			

ADJ + N

time	guy	thing	friend	way	fun	year
long	good	first	good	normal	much	first
first	sweet	important	new	bad	more	
good		whole		instructive	good	
next				weird		
last						

N + V

time	guy	thing	friend	kid	way	fun	woman	year
be	be	be	be	be	be	have	be	can
can	know	have	grow	make	go	make	marry	
see	do	say	have	do	want	pass	meet	
do	think	tell	give		eat	think	know	
believe		can	know		flirt			
build		know	trust		know			
find		see			laugh			
get					make			
go					mean			
happen					try			

How I Met Your Mother

N + N

time	room	thing	mom	year	guy	gift	man	girl
friend	service	cousin	egg	anniversary	miracle	wedding	game	coffee
gym	ghost	idea	ring	today	wedding	anniversary	poker	list
lawyer	bar		daughter	celibate		friend	time	mouth-breather
man	key		jellyfish	year			year	
time	night		photo					
	door							
	hooker							
	lock							

N + ADJ

time	room	thing	year	guy	gift	man	way	girl
next	cold	good	full	bad	innocent	good	good	little
good	crappy	weird	last	good		more		wrong
last	easy			great				
other	more			other				
perfect								
possible								
single								

N + V

time	room	thing	mom	year	guy	gift	man	way	girl
be	be	be	be	be	be	get	be	be	get
have	get	do	get	know	die	be	say	know	be
see	order	happen	do	take	meet	give	fold	have	give
know	do	can	make	think		do	get	do	do
call	say	let	have			arrive	have	work	arrive
think	call	look	know			know	play	can	know
feel	see	put	ring			say	start	feel	say
go		say	terrify			send		tell	send
happen		take	want						
take		tell	bang						

The Inbetweeners

N + N

time	dad	shit	fuck	mum	job
arm	god	car	dad	dad	blow
god	mum	fuck	game	year	hand
	job	hole	god	day	
		game	man	fuck	
		house	mum	shit	
		light	shit	standard	
		music	way		
		pub			

N + ADJ

time	dad	mate	shit	fuck	thing	mum	job	girl
first	last	cool	fat	mental	big	bad	good	fit
great		good	good	old	whole	first		
more		grim	old	right		right		
ok			back	serious				
sleepy			own					
tall								

N + V

time	dad	mate	shit	fuck	thing	mum	job	way	girl
be	be	be	be	do	do	be	be	be	be
do	go	come	do	be	be	know	give	do	have
go	get	do	go	go	get	do	get	go	fancy
think	have	get	get	fuck	have	see	go	look	fuck
get	leave	can	give	get	know	buy	have	come	might
have	use	go	have	have	smoke	find	see	get	want
spend	call	have	think	say	work	go	stick	have	
drive	come	say	look	think		have		jump	
take	die	see	try	can		tell		know	
want	do	look	can	look		count		say	

The Inbetweeners – censored version

N+ N

time	dad	mate	mum	night	car
arm	god	mate	dad	car	god
god	mum		year	fun	night
	job		day	time	
			standard		

N + ADJ

time	dad	mate	thing	mum	job	girl	night	car
first	last	cool	big	bad	good	fit	good	least
great		good	whole	first			last	
more		grim		right				
sleepy								
tall								

N + V

time	dad	mate	thing	mum	job	way	girl	night	car
be	be	be	do	be	be	be	be	be	be
do	go	come	be	know	give	do	have	get	go
go	get	do	get	do	get	go	fancy	have	have
think	have	get	have	see	go	look	want	do	get
get	leave	can	know	buy	have	come		go	know
have	use	go	smoke	find	see	get		kiss	lock
spend	call	have	work	go	stick	have		think	matter
drive	come	say		have		jump			
take	die	see		tell		know			
want	do	look		count		say			

Chewing Gum

N + N

man	thing	time	mum	sex	guy	people	girl	life
guy	guy	boyfriend	joy	bum	man	fuck	fuck	day
butcher	thing	eye	string	centre		penguin		lord
jesus			dad	dick		sex		
waste			flyer	people				
chill			sister	sin				
fuck			tongue					
god			work					
limit								

N + ADJ

man	thing	time	mum	sex	friend	guy	people	girl
gay	black	ok	great	only	good	other	black	beautiful
	cool	same	right	dickcentric	more	little	buff	good
	only	good	sick		ok	right	other	
	same	last				white		
		long						

N + V

man	thing	time	mum	sex	friend	guy	people	girl	life
be	be	be	be	have	be	be	be	be	be
know	do	go	say	be	know	go	do	do	live
say	get	know	go	want	see	know	can	have	get
have	know	do	do	do		remind	find	look	go
get	say	have	live	go			hurt	talk	become
come	can	see	tell	can			know		do
go	feel	take	thank	get			look		feel
want	fuck	tell	come	like			come		reign
do	make	come	get	say			have		say
find	mime	get	have	suck			hear		try

Chewing Gum – censored version

N + N

man	thing	time	mum	sex	guy	people	life
guy	guy	boyfriend	joy	bum	man	penguin	day
butcher		eye	string	centre		sex	lord
jesus			dad	people			
waste			flyer	sin			
chill			sister				
god			tongue				
limit			work				

N + ADJ

man	thing	time	mum	sex	friend	guy	people	girl
gay	black	ok	great	only	good	other	black	beautiful
	cool	same	right		more	little	buff	good
	only	good	sick		ok	right	other	
	same	last						
		long						

N + V

man	thing	time	mum	sex	friend	guy	people	girl	life
be	be	be	be	have	be	be	be	be	be
know	do	go	say	be	know	go	do	do	live
say	get	know	go	want	see	know	can	have	get
have	know	do	do	do		remind	find	look	go
get	say	have	live	go			hurt	talk	become
come	can	see	tell	can			know		do
go	feel	take	thank	get			look		feel
want	make	tell	come	like			come		reign
do	mime	come	get	say			have		say
find	see	get	have	think			hear		try

N + N

people	thing	time	way	day	world	shop	family	book
stage	hour	stage	mum	hour	people	sheep	song	child
time	kind	year	thing	country	sense	success	child	club
year	people	thing		end	way			manuscript
stuff	presentation	track		paper				
thing				people				
track				programme				

N + ADJ

people	thing	time	year	way	day	world	shop	book
other	bad	first	last	own	interesting	other	loud	good
many	good	good	first	same	few	own		more
different	whole	last	public	good		good		
good	great			long		great		
most	long			many		modern		
same	other			old		natural		
easy				other				
more				real				
own								
bad								

N + V

people	thing	time	year	way	day	world	shop	family	book
be	be	be	be	be	be	be	be	be	be
have	do	have	spend	get	buy	go	go	have	do
get	get	do	do	have	come	have	can	can	write
can	have	go	have	see	give	involve	use	get	have
do	can	come	know	can	go	travel	have	sing	get
make	decide	get	make	do	say		make	want	
think	find	know		feel	look				
go	know	think		go	make				
know	mention	download		learn	spend				
live	must	give		live	work				

TOEFL

N + N

class	time	people	thing	way	week	computer	kind	problem	paper
college	day	group	kind	degree	class	lab	poem	ocean	term
dorm	hour	kind		fluid	point		behaviour		currency
program	people	thing		nature			hinge		money
review	rest	area		opinion			people		research
week		atmosphere					thing		
		difference							
		dog							
		idea							
		point							
		pressure							

N + ADJ

class	time	people	thing	way	week	computer	kind	problem	paper
sure	first	other	other	cloudy	next	more	hard	good	good
	good	more	first	other	last	okay	light	last	long
	long	good	good		good		old	same	short
	much	much	same				sure	sure	valuable
	next								well-structed
	available								
	full								
	most								
	ok								

N + V

class	time	people	thing	way	week	computer	kind	problem	paper
be	be	be	be	be	have	be	be	be	be
take	can	have	do	get	be	have	do	have	have
have	spend	do	know	can	come	can	make	solve	see
get	have	know	mean	do	do	run	believe	can	do
think	know	mean		have	know	do	get	come	write
do	do	get		look	say	freeze	have		compare
know	take	go		pay		mean	take		get
need	need	say		talk		see			try
want	read	walk		breathe		use			use
ask	seem	come		go		work			