

UNIVERZITA PALACKÉHO V OLOMOUCI
PEDAGOGICKÁ FAKULTA
Ústav cizích jazyků

Diplomová práce

Bc. Vojtěch Přecechtěl

**English Vocabulary Frequency and Its
Use at Lower Secondary Schools**

Olomouc 2016

Supervisor

Doc. PhDr. Václav Řeřicha, CSc.

Prohlášení

Prohlašuji, že jsem závěrečnou práci vypracoval samostatně a všechny použité prameny a literaturu uvádím v seznamu použitých zdrojů a literatury.

V Olomouci 20. 4. 2016

.....

vlastnoruční podpis

Acknowledgement

I would like to thank my supervisor Doc. PhDr. Václav Řeřicha, CSc. for his support and valuable comments on the content, style and presentation of my final project.

Abstract

Vocabulary frequency is a topic that has raised interest of many researchers and teachers over the last seventy years. It appears to be a reasonable and objective criterion for vocabulary selection. This thesis deals with the possible impact of the vocabulary frequency aspect on English foreign language (EFL) teaching at lower secondary schools. Vocabulary frequency lists provide support for vocabulary level analysis and teachers can use the related tools in order to classify the language level of a text, determine the text difficulties and create new learning materials based on this objective method. The subjects of the research in this thesis are wordlists of six English textbooks for lower secondary school learners. The aim of the research is to point out the frequency aspect presence amongst other vocabulary aspects and discuss its contribution to the EFL teaching. The results show the distribution of the six wordlists in the frequency levels of two reference vocabulary frequency lists. The *NGSL* covers 60 – 82 % of the six textbook wordlists. This proportion of the wordlists stands for the high frequency vocabulary. The research proves that the high-frequency vocabulary is present and covers the majority of items in the six textbook wordlists.

Contents

INTRODUCTION	8
1 INTRODUCTION TO VOCABULARY FREQUENCY.....	9
1.1 VOCABULARY	9
1.2 PASSIVE AND ACTIVE VOCABULARY.....	11
1.3 VOCABULARY SIZE	12
1.4 VOCABULARY FREQUENCY.....	13
1.4.1 <i>High-frequency vocabulary</i>	13
1.4.2 <i>Mid-frequency vocabulary</i>	14
1.4.3 <i>Low-frequency vocabulary</i>	15
1.4.4 <i>Vocabulary coverage</i>	15
1.5 ENGLISH CORPORA.....	16
1.5.1 <i>Types of corpora</i>	17
1.5.2 <i>Data processing</i>	18
1.6 FREQUENCY LISTS	19
1.6.1 <i>General Service List</i>	20
1.6.2 <i>Academic Word List</i>	23
1.6.3 <i>New General Service List</i>	24
1.6.4 <i>New Academic Wordlist</i>	26
1.6.5 <i>GSL, New-GSL and NGSL comparison</i>	26
1.6.6 <i>BNC-COCA</i>	27
1.6.7 <i>Vocabulary profilers</i>	28
1.7 VOCABULARY OF A CZECH LEARNER OF ENGLISH	30
2 EFL VOCABULARY TEACHING	31
2.1 FRAMEWORKS	31
2.2 TEACHING AND LEARNING STRATEGIES.....	33
2.2.1 <i>Classroom English</i>	35
2.2.2 <i>CLIL</i>	35
2.2.3 <i>Vocabulary selection</i>	36
2.3 MATERIALS BASED ON WORD FREQUENCY	36
2.3.1 <i>Dictionaries</i>	37
2.3.2 <i>Textbooks</i>	38
2.3.3 <i>Wordlists</i>	39
2.3.4 <i>Graded readers</i>	40
2.3.5 <i>Software</i>	41
2.3.6 <i>Games</i>	42

3	RESEARCH.....	45
3.1	METHODOLOGY.....	45
3.1.1	<i>Textbooks selection.....</i>	45
3.1.2	<i>Description of the wordlists.....</i>	46
3.1.3	<i>Data processing.....</i>	48
3.1.4	<i>Reference list selection.....</i>	48
3.1.5	<i>Vocabulary profiler tool selection.....</i>	49
3.2	RESULTS.....	50
3.2.1	<i>Project 4 the Third Edition.....</i>	50
3.2.2	<i>Hello Kids! Angličtina pro 7. ročník ZŠ.....</i>	52
3.2.3	<i>More! L3.....</i>	54
3.2.4	<i>English Plus 2.....</i>	56
3.2.5	<i>Messages 3.....</i>	58
3.2.6	<i>Angličtina 7 Way to Win.....</i>	60
3.3	RESEARCH SUMMARY.....	62
3.4	METHODOLOGY DISCUSSION.....	64
3.5	RESEARCH CONCLUSION.....	65
4	CONCLUSION.....	67
	LIST OF ABBREVIATIONS.....	69
	LIST OF SOURCES.....	70
	BIBLIOGRAPHY.....	70
	ONLINE AND OTHER SOURCES.....	71
	APPENDICES.....	76
	APPENDIX 1 – WORD FAMILY LEVELS.....	77
	APPENDIX 2 - NGSL FREQUENCY DIVISION OVERVIEW.....	78
4.1.1	<i>Appendix 2a.....</i>	78
4.1.2	<i>Appendix 2b.....</i>	79
4.1.3	<i>Appendix 2c.....</i>	80
	APPENDIX 3 - IMAGES.....	81
	APPENDIX 4 - VOCABULARY DISTRIBUTION DETAILS.....	83
	<i>Appendix 4.1a - Project 4 the Third Edition (Hutchinson, 2009) (NGSL).....</i>	83
	<i>Appendix 4.1b - Project 4 the Third Edition (Hutchinson, 2009) (BNC-COCA).....</i>	85
	<i>Appendix 4.2a - Hello Kids! Angličtina pro 7. ročník ZŠ (Zahálková, 2010) (NGSL).....</i>	88
	<i>Appendix 4.2b - Hello Kids! Angličtina pro 7. ročník ZŠ (Zahálková, 2010) (BNC-COCA).....</i>	90
	<i>Appendix 4.3a - More! L3 (Puchta et al., 2008) (NGSL).....</i>	93
	<i>Appendix 4.3b - More! L3 (Puchta et al., 2008) (BNC-COCA).....</i>	96

<i>Appendix 4.4a - English Plus 2 (Hardy-Gould & Mellersh, 2011) (NGSL)</i>	99
<i>Appendix 4.4b - English Plus 2 (Hardy-Gould & Mellersh, 2011) (BNC-COCA)</i>	102
<i>Appendix 4.5a - Messages 3 (Goodey et al., 2006) (NGSL)</i>	105
<i>Appendix 4.5b - Messages 3 (Goodey et al., 2006) (BNC-COCA)</i>	107
<i>Appendix 4.6a - Angličtina 7 Way to Win (Betáková & Dvořáková, 2006) (NGSL)</i>	109
<i>Appendix 4.6b - Angličtina 7 Way to Win (Betáková & Dvořáková, 2006) (BNC-COCA)</i>	111
ANNOTATION	113

Introduction

I have chosen this topic due to my interest in English foreign language (EFL) teaching. I find the vocabulary to be a very important part of any language. I attended a lecture about Oxford Learner's Dictionary at the English department of the Faculty of Education at Palacky University few years ago (2014). The lesson I learnt was that if I need to communicate in English, I can learn new words with frequency in mind. With around the first 2,000 word families, I can say about 80 % of what I would like.

Later on, I got an opportunity to write my thesis about the vocabulary frequency and I simply took it. The whole work aims at lower secondary school EFL teaching because I have some experience with it and plan on teaching the lower secondary learners.

The main role of my thesis is to cover this question:

“Does the vocabulary frequency aspect impact English teaching and learning?”.

During the work pre-research, I came up with following hypotheses that I also used to set the direction of this work:

- 1) *“English core language can be based on the vocabulary frequency aspect.”*
- 2) *“English textbooks used among Czech learners reflect the core language.”*
- 3) *“The EFL learners should learn at first the core vocabulary of the target language.”*
- 4) *“Through comparing the Czech and English core language, the EFL materials can be adjusted for Czech learners of English.”*

I would like to summarize vocabulary teaching and learning methods as it is essential for the practical use at lower secondary schools. I believe it is essential to gather information related to lower secondary schools English language vocabulary teaching.

Another goal is to create a methodology for research of English textbook wordlists and check frequency of the vocabulary. The results should show how the authors of the textbooks worked and coped with the vocabulary frequency aspect for teaching.

1 Introduction to vocabulary frequency

Lexicology is a study of different aspects of vocabulary. For pedagogical purposes, I focused on the studies dealing with the vocabulary frequency in a relation to English language learning and teaching. In this section, you will find information about main features related to vocabulary frequency researches.

1.1 Vocabulary

Vocabulary is dealt throughout my entire thesis. There is no doubt about its importance for EFL teaching.

Knowledge of a language (Smith et al., 2016, p. 57): *“is divided into our knowledge of vocabulary on the one hand, and our knowledge of how to combine that vocabulary into sentences on the other.”*

“Vocabulary is central to English language teaching because without sufficient vocabulary students cannot understand others or express their own ideas.” (Lessard-Clouston, 2013, p. 2)

In the work *Pygmalion*, George Bernard Shaw says that words are important - If one cannot say what they means, they will never mean what they says and one should always mean what they says. (Lessard-Clouston, 2013, p. 2)

Vocabulary exists as a relative term, which can carry many different meanings. In this section, I would like to mention some definitions that may be useful for this work. *“...vocabulary is the total number of words that are needed to communicate ideas and express the speakers' meaning.”* (Alqahtani, 2015, p. 25)

“Vocabulary can be defined as the words of a language, including single items and phrases or chunks of several words which convey a particular meaning, the way individual words do. Vocabulary addresses single lexical items—words with specific meaning(s)—but it also includes lexical phrases or chunks.” (Lessard-Clouston, 2013, p. 2)

When teachers want their learners to learn new words, they should know what to focus on. A word has aspects that teachers need to be aware of. One of word aspects is its form, which involves pronunciation, spelling and other word parts (such as prefix, root, and suffix). Another word aspect is meaning, it says what the concept behind the word is, what people imagine and what associations they make when they encounter the word or expression. The final aspect is its use, the use depends on the grammar structures, collocations and constrains on use. At all of the three aspects, receptive and productive dimension can be taken look at. Learning a word involves all together 18 types of lexical knowledge, as seen in the *Table 1*. (Lessard-Clouston, 2013, p. 3)

Aspect	Component	Receptive knowledge	Productive knowledge
Form	spoken	What does the word sound like?	How is the word pronounced?
	written	What does the word look like?	How is the word written and spelled?
	word parts	What parts are recognizable in this word?	What word parts are needed to express the meaning?
Meaning	form and meaning	What meaning does this word form signal?	What word form can be used to express this meaning?
	concepts and referents	What is included in this concept?	What items can the concept refer to?
	associations	What other words does this make people think of?	What other words could people use instead of this one?
Use	grammatical functions	In what patterns does the word occur?	In what patterns must people use this word?
	collocations	What words or types of words occur with this one?	What words or types of words must people use with this one?
	Constraints on use (register, frequency)	Where, when, and how often would people expect to meet this word?	Where, when, and how often can people use this word?

Table 1 - What is involved in Knowing a Word (Lessard-Clouston, 2013, p. 3)

I find this quotation by Saramago corresponding with my idea of vocabulary:

“Human vocabulary is still not capable, and probably never will be of knowing, recognizing, and communicating everything that can be humanly experienced and felt.”

Smith et al. write in their book *Chomsky: Ideas and Ideals* (2016, p. 58) about so-called “*mental lexicon*”, which is a knowledge of the vocabulary that individual has stored in his or her head. This may be in conflict with knowledge of the community of scholars or information stored in books.

It is clear that the concepts of words differ for each language. In the Czech Republic, the word for mushrooms “*houby*” will surely provide different associations than the English word. Therefore, we could speak about individual understanding of the concept, connotation, denotation and the word’s use. From my perspective, comparing the first and foreign language concepts is supportive for learning.

According to Smith et al. (2016, p. 58), vocabulary rather consists of lexical items than individual words. For instance, all of these items can be regarded as vocabulary:

- single words “*question*”;
- expressions “*get up*”;
- idioms like “*to bury the hatchet*”;
- clichés and collocations such as “*it’s a hard life*” or “*good morning*”;
- or phrases, for example “*What’s your name?*”.

I am aware of other important aspects of vocabulary such as grammatical categories, which are important for the meaning, context and word formation.

1.2 Passive and active vocabulary

In methodology, there is a need to deal with the vocabulary on many levels. Frequency of vocabulary generally does not need to mean anything for a learner but if we speak of frequency of learners’ vocabulary, it has a great importance. Person’s frequency of the vocabulary use can tell us what the learner remembers and uses actively, this is often called “*productive vocabulary*”. What the learner would not use, is unsure about and though has only a passive knowledge, is called “*receptive vocabulary*”. Vocabulary comprehension depends on the learner’s knowledge of the vocabulary aspects. (Lebedová, 2011, pp. 7-8)

At first, learners get some passive knowledge of a word. They get in touch with it again and learn about its different meanings and use. Once they know these aspects, they can try to turn the passive knowledge into active by using the word in different contexts without any assistance. (Lebedová, 2011, p. 8) Of course, the more words covering the text we know the better we can get ease of reception and enjoy that text. (Schmitts, 2014, p. 495)

Active vocabulary of a language develops and even native speakers may from time to time encounter a new word they would like to look up in a dictionary. (Lessard-Clouston, 2013, p. 6)

1.3 Vocabulary size

English language vocabulary size is often argued and differs according to what is defined as a word, as described in the previous section. Oxford University Press (OUP, 2016a) writes that it is impossible to count all words in a language. Some of the researchers attempted to estimate the total number of words in English. Google in cooperation with the Global Language Monitor came up with almost one million entries based on a research of words used in media. (McCarten, 2013, p. 1)

In my opinion, the vocabulary size of each individual person differs in many ways. Its range and size depends on how much the person makes effort, learning strategy, time, materials and so on. It is possible to estimate an approximate vocabulary size. There are vocabulary size tests available, which I reviewed very briefly. Most of the tests estimate your vocabulary size based on frequency so one does not have to spend as much time and effort. The selection of sets of words, that are tested, is made with your previous answers in mind. The applications calculate the probability of your future answers and provide you with set of words you may not know. (testyourvocab.com & Ghent University, 2016)

In numbers, it usually asks about 100 different word meanings before it estimates the total number of the words a person knows. In the methodology, I learnt that homographs and other confusing words are excluded from the set of words you are tested on. (testyourvocab.com & Ghent University, 2016)

1.4 Vocabulary frequency

As we could learn from the first subchapter, the frequency is one of the vocabulary aspects. In my own words, I would explain frequency as information about how frequent the word is in a specified range of texts.

Vocabulary frequency can be sorted in many ways, for pedagogical purpose I find useful the three-frequency band division into **high**, **mid** and **low frequency vocabulary** by the Schmitts (2014) who define the vocabulary range for each band based on coverage in various texts. The Schmitts (2014, pp. 485-486) set up the boundaries dependently on the vocabulary use and coverage. The 3,000 word families (boundary between high and mid-frequency) are enough for everyday purpose – conversation with the coverage of 95 %. The **mid-frequency vocabulary** (up to 9,000 word families) covers 95 % of wide range of authentic texts, is essential for academic studies in English. (Schmitts, 2014, pp. 494-497) The **low-frequency vocabulary** contains words that you would find in a text very rarely. Of course, in real situations, this strict division may fail and the boundaries of the bands usually blend when one acquires and learns new vocabulary. (Schmitts, 2014, p. 495)

1.4.1 High-frequency vocabulary

Core vocabulary is the part of high frequency vocabulary, which appears in any text. Cvrček (2011, p. 1) established his research on the concept of Basic English [Ogden, 1930 & Crystal, 1997] which was a small subset of English vocabulary and grammar with a highest possible text coverage as it proved to play crucial role in determining the core elements.

There is a core of language which appears in every text - among four corpora (*LOB*¹, *BNC*², *BE06*³ and *EnTenTen12*⁴), there is a stable vocabulary core of 2,122 word families, which stands for 70.7 % of the texts overall. (Březina, 2013, p. 1) The term core vocabulary matches with the high-frequency vocabulary in many aspects, from my point of view.

¹ Refers to Lancaster-Oslo/Bergen Corpus [1961]

² Refers to British National Corpus (1980s - 1993)

³ Refers to The British English 2006 corpus (BE06). Compiled by Paul Baker.

⁴ Refers to English TenTen web corpus (2012) (Lexical computing, 2016)

High-frequency vocabulary usually appears in any text. It does not depend on the text genre or style so much, contains a lot of function words (words with high coverage), but also the most used words carrying the information. The first 1,000 word families in *COCA* corpus are made up of function words by 43 %. This is also a reason why they have so much coverage. (Schmitts, 2014, p. 488)

The Schmitts (2014, p. 486) based the frequency division a recent study by Nation [2006], who researched the BNC. They argue that the core vocabulary should include the most frequent 3,000 word families. Their arguments for this come from occurrence of the 3,000 level words in the Brown's written English corpus and Cobb's corpus of novels by Jack London. Beyond the 3,000 level, the learning opportunities from reading begin to taper off quickly. (Schmitts, 2014, p. 489)

Considering English listening and conversation, the 95% coverage appears to be enough for comprehension - that is the range between the 2,000 and 3,000 word families. (Schmitts, 2014, p. 490) McCarten (2007, p. 1) comes up with the number of 1,800 word families for 80% spoken corpus coverage.

I understand that learning high-frequency vocabulary is the breaking point of the language learning. After this imaginary border of the 3,000 word families, new possibilities of learning are open. At this level the learner should be able to learn from the context, they can use monolingual dictionaries where word meaning is explained in the context, he can slowly start reading authentic texts and study on his own just the unknown words as they make less than 5 % of the text. (Schmitts, 2014, p. 490)

1.4.2 Mid-frequency vocabulary

The mid-frequency vocabulary covers the range between the high and low frequency vocabulary. To some extent, the academic and technical vocabulary can be classified as the mid-frequency range. (Schmitts, 2014, p. 493)

The achievement of 98% text coverage is recommended for enjoyable listening and watching movies. This is largely dependent on mastering the words in the mid-frequency range. For horror, drama and crime genre, it makes around 5,000 word families to achieve that coverage while for war and animation 9-10,000 word families. (Schmitts, 2014, p. 495)

The mid-frequency vocabulary is essential when operating in English across a range of different topics and situations, for instance to cope with the university studies in English. In a class of high-intermediate students, only 2 % of the speech were beyond the high-frequency bands in Horst's analysis of 32 hours classroom discourse. Mid-frequency words are also used for defining other words – technical words are often explained in the texts. (Schmitts, 2014, p. 495)

1.4.3 Low-frequency vocabulary

The low-frequency vocabulary is usually considered beyond the 10,000 frequency level. The Schmitts (2014, p. 484-485) think that the threshold should be decreased to the 9,000 frequency level because 8,000 – 9,000 word families are sufficient for a learner to be able to read a wide range of authentic novels or newspapers without assistance. The 9,000 level stands for 98% coverage in texts of this type. However, reader's comprehension of text is not always complete. *“For example if reader's vocabulary covers 95-98 % of a text, he is likely to comprehend 60-68 % of that text.”* (Schmitts, 2014, pp. 489-490)

“Low frequency vocabulary occurs so infrequently that it is not worth spending classroom time on. [Nation]” It is better to teach the vocabulary learning strategies so that the learners learn these rare words on their own. (Schmitts, 2014, p. 485)

1.4.4 Vocabulary coverage

I understand vocabulary coverage as an effect of the frequency aspect. I would say the coverage is in direct proportion to the frequency aspect. The more frequent the vocabulary item is the greater coverage it gets. I already mentioned the function words, such as the verb *“to be”* and its inflections or the definite article *“the”*, belong to the high frequency words and therefore get great coverage.

The frequency is an important vocabulary aspect. The approaches to work with frequency differ. I believe the most suitable attitude is to use a methodology that reflects the practical situation when teaching. From the reading, I conclude the high-frequency band is suitable for teaching the lower secondary learners.

1.5 English Corpora

In this chapter, I would like to present some information for better understanding of corpora. Corpus data can tell us many things about the vocabulary – which words and expressions are common and rare, differences in speaking and writing, what vocabulary the people use in different situations, the words that often collocate or underline the words that are used to organize and manage a text. (McCarten, 2013, p. 3)

Corpus is a collection of texts stored in a database. The data is usually used for lexicology researches. “*Corpus research is based on hypothesis that corpora data collections reflect the language truly and precisely.*” (Cvrček, 2011, p. 1)

“Quantitative analysis of a corpus can give us statistical ideas while qualitative helps us observe the words’ use in contexts. Corpus itself cannot replace a teacher as it is raw data in need to be considered more but it can supply the teacher and learners with hints on what is important to be taught.” (McCarten, 2013, p. 3)

The choice of corpus for a research is not easy. There are many different corpora with focus on the text purpose, style and genre. Very important information about corpus is the target language, first language, language level, size and nature of the text database. (CECL, 2016)

Examples of two well-known general English monolingual corpora and brief description:

Corpus	Size	Composition
British National Corpus (BNC)⁵	100 million words	Written text (90 %) Spoken text (10 %)
Corpus of Contemporary American English (COCA)⁶	520 million words	spoken, fiction, magazine, newspaper, academic (20 % each)

Table 2 - Two well-known general English monolingual corpora

⁵ Source: DAVIES, M. (2004-) *BYU-BNC. (Based on the British National Corpus from Oxford University Press)*. Available online at <http://corpus.byu.edu/bnc/>.

⁶ Source: DAVIES, M. (2008-) *The Corpus of Contemporary American English: 520 million words, 1990-present*. Available online at <http://corpus.byu.edu/coca/>.

1.5.1 Types of corpora

There are different purposes of corpora. A corpus can be of a reference nature, it is set up once and not modified so the researchers can rely on it in the long term, or non-reference nature as it develops through the time. Apart from the spoken and written language register, it can be focused on a particular genre, author or text and literary style. Furthermore, it can be defined as a composition of various kinds of text. (Rundell, 1998, pp. 320-321)

There are various kinds of corpora depending on the text type and style according to Keddie [2016]:

- *Books*
- *Magazines*
- *Newspapers*
- *Emails*
- *Television*
- *Radio*
- *Conversations*

Nowadays, there also web corpora such as *enTenTen* (2012) or *czTenTen* (2012) composed just from website texts. (Lexical Computing, 2016)

For linguistic and lexical research purposes, there are also corpora of other kinds:

- *Samples of written US English*
- *Samples of spoken British English*
- *Business correspondence*
- *Legal contracts*
- *Old English*
- *Children's speech*
- *Learner's*
- *Native speaker's*

Nature of a text can differ in many aspects and except for communicative and informative function, it can be amusing, philosophic, academic, political, technical or different purpose text. Considering the nature, the texts will have different features.

Considering education, I would like to describe some of the kinds mentioned above. **Native speaker's** corpus can be the reference point for EFL teachers, as it should reflect the real language use. It can also show development of the language or different use through the time. (Keddie, 2016) **Learner's** corpus may be a good choice for investigation of differences between the first and the target language, as the research by Roca-Varela (2013, p. 561) dealing with Spanish English learner's corpus shows.

(Keddie, 2016) **Spoken language** corpus is a database of records and their transcriptions. In spoken language, you can usually see what you need to learn in order to take a part in a conversation. The repetitive nature of spoken language results in many functional words, omissions and short phrases. **Written language** corpora contain texts of literary works, newspapers or articles. They mostly involve longer sentences and more complex grammatical structures. Therefore, these texts are usually more demanding on comprehension. (Keddie, 2016)

1.5.2 Data processing

Rundell (1998, p. 320) explains that the data for lexicographic research is usually gained from corpora. Corpora are processed into individual items based on specified criteria and patterns that have significant impact on the research possibilities. There is a need to not only count the occurrences of the words but also recognize their class and grammatical categories.

According to Keddie (2016), information technologies brought an improvement for corpora analysis and synthesis. Computer software can be a good time-saver when processing corpora data into a vocabulary frequency list. Nevertheless, there are areas or language situations the software cannot deal with and needs a human assistance. (Rundell, 1998, p. 337)

In the commonly used corpora, the data is usually available online for a limited number of queries a day. For instance, the online tools at <http://www.wordandphrase.info> by Davies (2004-), which access the *COCA* and the *BNC* data, are good enough to analyse a word or phrase, however, with higher number of words in the queries, the search progress tends to be slow or stops working.

A selection of a text from an age specific corpus can help teachers prepare a teaching material. The corpus data carry a frequency information of the word distribution among the corpus texts. In a corpus, we can search for any word, phrase or collocation and use the corpus contexts to illustrate various meanings of the vocabulary item in the classes. There is free corpus data available on the internet, but it demands further investigation in terms of its use.

1.6 Frequency lists

In this section, I would like to focus on the history and the present of English language frequency wordlists. The beginnings must have been very complicated as there was no automatization and all data was human processed. (Krishnamurthy, 2008, p. 238) Furthermore, I would like to introduce the most known and used wordlists and compare them with their current alternatives. We are aware that there are other than the wordlists mentioned in this section.

Wordlists based on different corpora can serve for vocabulary studies. When setting up a wordlist, clear rules should be applied and described in the methodology. In case of following the identical methodology, the research results should be comparable. (Gilner, 2011, p. 79-80)

Each text in a corpus has its individual purpose and aspects such as style, genre, vocabulary range and size. In terms of vocabulary frequency, the wider variety of texts and the higher total number of words, the more the results based on the corpus can be generalized. (Krishnamurthy, 2008, p. 232)

Creation process of a frequency list seems to be simple – the data from the source corpora are turned into a wordlist based on the vocabulary frequency and coverage. Notably, here must be clear criteria applied on the word selection process. The wordlist maker needs to decide questions like:

- *What do we consider to be a word?*
- *How to deal with the words in the same word family?*
- *Should we include proper names, numbers or other common words?*

Some words are often excluded from the lists for pedagogical and other reasons. For instance, these are the proper nouns, days of a week and months of a year, numbers, abbreviations and single letters. (Browne & Culligan & Phillips, 2013a) There is no need to include them, as learners should learn them all to be able to use the language for their needs. Monday is at the same level of importance as Wednesday, January as February and so on. (Browne, 2014, p. 43)

1.6.1 General Service List

A General Service List of English Words (GSL) by Michael West (1953) is a set of approximately 2,000 headwords representing their word families. (Gilner, 2011, p. 70) Bauman [1995] and Gilner (2011, p. 66) agreed that the list does not contain exactly the most common 2,000 words, although frequency was one of the main factors when making the word selection. The list is sorted alphabetically and contains lemmas including some derivations, frequency rank and meanings. Note that some derivations are included as new items.

The GSL was based on a two and half million word corpus⁷ that was collected under a grant from the Rockefeller Foundation in 1938. (Gilner, 2011, p. 69)

“This list has had a wide influence for many years, serving as the basis for graded readers as well as other material. Texts based on the GSL are still on sale, but the list itself is out of print.” (Bauman, [1995])

Gilner (2011, p. 65-66) introduces that approximately 2,000 words accounts for 70 - 95 % of any text, analyses based on various corpora keep returning the same lexical set and this set of stable words can be called a core vocabulary. The *GSL* is the best researched from all frequency range based lists and it only differs in around 10% word coverage in comparison with more recent lists.

⁷ Later, the corpus was enhanced to five million words.

1.6.1.1 Characteristics of the GSL

Gilner (2011, p. 70) summarizes that contemporary researchers identify with following characteristics of the *GSL* as the most relevant: “*Frequency, universality (words used in all countries), utility (words used to talk about a wide range of topics), and usefulness (words that can be used to describe or define other words).*” [Carter, 1998; Carter & McCarthy, 1988; Nation, 1990]

Many versions of the West’s *GSL* are available on the internet. Some of them are sorted alphabetically as the original one and the meanings are not mostly included. In addition, the frequency information is not available in all of them. An example entry of the original West’s *GSL*:

Lemma + POS	Freq. rank	Sense #	Senses
POOR, adj.	1096e	-1	(having little money) Rich and poor Poor people’s children

Table 3 - An example entry from West’s (1953) “*A General Service List of English Words.*” (Gilner, 2011, p. 70)

1.6.1.2 Critiques of the GSL

Engels [1968] considered the *GSL* range, and said that last 2,000 words cannot be called the *GSL* because there was small occurrence of these words in texts. Gilner (2011, p. 71) commented the research that Engel’s methodology failed due to lack of words he used for the research. Richards [1974] found the *GSL* being dated especially due to new words associated with technology. Gilner and Morales [2008] stated that the *GSL* is non-expandable as the word selection criteria are not clear. (Gilner, 2011, p. 72)

Nation [2004] also researched the *GSL* and applied it on a modern corpus. “*His conclusion is that given the differences in distribution, coverage, and content - possibly a result of the age of the GSL, Nation speculates—a replacement might merit consideration although it is not entirely clear how it should be formulated. Of note, subsequent refinement of Nation’s BNC list has lessened the differences in distribution, coverage, and content with the GSL.*”

Gilner (2011, p. 76) summarizes the critique with the conclusion that although the speculations about the criteria of the *GSL* word selection are justified and the *GSL* does not exactly equate to the most frequent words, modern studies showed the *GSL* words are among the most frequent in the English language.

1.6.1.3 History of the GSL

I would like to mention some parts from history of the *GSL* because it appears to be essential for understanding the studies on all English wordlists. In the vocabulary frequency studies, each English reference wordlist is compared to the *GSL*.

Interim Report on Vocabulary Selection [Faucett et al., 1936] is an annotated vocabulary list of about 2,000 words. It was the result of two conferences held in the mid-1930s. *Interim Report* was sponsored by the Carnegie Corporation with the purpose of examining “*the part played by wordlists in the teaching of English as a foreign language*”. (Gilner, 2011, p. 66)

There were different linguists involved with different ideas on vocabulary selection. Thorndike’s *The Teacher’s Word Book* [1921] and *A Teacher’s Word Book of 20,000 Words* [1931] are among the first publications in English related to the vocabulary frequency. “*Thorndike [1931] presented frequency information for the 10,000 most frequently occurring words in a corpus of 4.5 million running words from 41 sources including the Bible, children’s story books, textbooks, trade manuals, and periodicals.*” (Gilner, 2011, p. 67)

Another early important wordlist was by Horn [1926] who reported inflections and derivations separately in his wordlist of the 10,000 most frequent words published in *A Basic Writing Vocabulary*, Thorndike [1932] did not count work with inflections and derivations separately. (Gilner, 2011, p. 67)

Maki and Faucett [1932] collaborated on combination of the most extensive and credited frequency lists of the time – Horn’s and Thorndike’s, based on objective criteria. Their work was published as “*A Study of English-Word Values Statistically Determined from the Latest Extensive Word Counts*”. (Gilner, 2011, p. 68)

Palmer [1931] (a director of Institute for Research in English Teaching in Tokyo) set a 3,000 headwords word-list based on Thorndike's. It contained the headwords and their most common derivatives and compounds. West [1927] isolated about 1,800 words based on experience and intuition in the work *New Methods Readers Series* [1927]. (Gilner, 2011, p. 68)

During the first conference, *The Use of English as a World Language* held 1934 in New York the criteria for word selection were discussed. It involved consideration up to 5,000 words, voting for the doubtful words, review and evaluation. It brought together researchers Faucett, Palmer, Thorndike and West who framed, classified and itemized the wordlist. They used two attitudes for the word selection. The first selection was based on the real frequency, which is objective. (Gilner, 2011, p. 68)

“The objective selection was primarily based on the Faucett and Maki (1932) frequency list while the subjective criteria was embodied in Palmer's principles, on the one hand, and in Palmer's word-list and West's definition vocabulary, on the other.” (Gilner, 2011, p. 68)

Secondly, they discussed each word subjectively and decided about its inclusion and exclusion. *“They used subjective criteria for inclusion/exclusion of words as consisting of: structural value, universality, subject range, definition words, word-building potential, and style.”* [Faucett et al., 1936] (Gilner, 2011, p. 69)

(Gilner, 2011, p. 69) The second *Carnegie conference* took place in London in 1935. They reviewed the wordlist, stated that the list is rather tentative and emphasized a desire for feedback and the aim was to evaluate.

In 1939, arrangements for revision of the list were made and West was invited to carry out the work. The WWII postponed the process but in 1953 *A General Service List of English Words* was finally published. (Gilner, 2011, p. 69)

1.6.2 Academic Word List

Academic Word List (AWL) was developed by Coxhead in 1998 as her MA thesis at the School of Linguistics and Applied Language Studies at Victoria University of Wellington, New Zealand.

The list contains 570 word families, which were selected according to principles to complement the *GSL*. The list does not include words that are in the most frequent 2,000 English words included in the *GSL*. It completes the *GSL* in terms of academic vocabulary and provides the frequent academic vocabulary especially useful for the learners that would like to study at university. (Coxhead, 2000, p. 213)

1.6.3 New General Service List

Recently (2013), there have been two attempts to establish a new general service list. One of them is the *new-GSL* by Březina and Gablasová. Secondly, the *NGSL* by Browne, Culligan and Phillips, which is now available in the version *NGSL 1.01* (2014).

1.6.3.1 new-GSL

“The new-GSL is a list of ~2,500 common English vocabulary based on four language corpora of the total size of over 12 billion running words. It can be used for both teaching and research purposes.” (Březina, 2014)

As the authors say on their websites, their wordlist can be accessed online and used for pedagogical and research purposes. They comment on and describe their wordlist creation in a video. In addition, they provide a set of tools that can be used to analyse a text. (Březina, 2014) The *new-GSL* takes into account frequency, dispersion, distribution (stability) across language corpora. The wordlist covers between 80.1 and 81.7 % of the text in the source corpora. Each item in the wordlist is a headword and its inflectional variants, only the most frequent lemmas are included. Altogether, it makes 2494 items (2116 base part + 378 current vocabulary). The wordlist’s size was reduced by almost 40 % with similar coverage. (Březina & Gablasová, 2013, p. 3-5)

The corpora *LOB*, *BNC*, *BE06* and *EnTenTen12* were used for its creation and it is based primarily on British variety of English. The *new-GSL* can be expanded because the objective criteria of word selection were applied. The *new-GSL* is also available in *Applied Linguistics Online*. (Březina & Gablasová, 2013, p. 18-19)

1.6.3.2 NGSL

Another from the recent general service lists is the *NGSL*. It tries to replace the *GSL* with better text coverage. It is based on a 273 million word subselection of the *Cambridge English Corpus (CEC)* containing two billion words. According to the authors, the *NGSL* provides the coverage of 92.34% with 2,818 lemmas compared to the *GSL* with the coverage of 84.24 % (3,623 lemmas) in the *CEC*. The 1.01 version of the *NGSL* often gets around 92% coverage for most general English texts. (Browne et al., 2013a)

The *NGSL* is available in a version with lemmas and statistics up to the frequency of about 31,000 lemmas. In addition, there is a version with definitions in English and Japanese. (Browne et al., 2013a) To illustrate, I list the first ten items of the *NGSL*, there is always a headword, frequency rate and the inflectional forms on the list.

the	1
be	2 am, m, is, isn, s, are, re, aren, ain, been, was, wasn, were, weren, being, bein, beings
and	3 and s
of	4
to	5
a	6 an
in	7 ins
have	8 haven, ve, has, hasn, had, hadn, d, having, hafta
it	9 its
you	10 y, ye, your, yours

Table 4 – The first ten items of the *NGSL* and their inflectional forms (Browne et al., 2013a)

As it is said on the wordlist’s portal (Browne, 2014): “*the most important words for second language learners of English*”. Big advantage of the *NGSL* is that the list, although it is quite recent, found its application in many free learning, teaching, analytical and editing tools and EFL textbooks. (Browne et al., 2013a)

1.6.4 New Academic Wordlist

New Academic Wordlist (NAWL) by Browne, Culligan and Phillips (2013b) was created to complete the *NGSL* with the most frequent academic vocabulary. The criteria of word selection were identical with the *NGSL*. “*The NAWL is a list of 963 words derived from an academic corpus containing about 288 million words. It is available in alphabetical order, with inflected forms, and with Standard Frequency Indices.*” (Browne et al., 2013b)

LEMMAS (INFLECTED FORMS)

abdominal, abdominals

absorb, absorbs, absorbed, absorbing, absorbings

accelerate, accelerates, accelerated, accelerating, acceleratings

Table 5 – Three example items of the NAWL with inflected forms (Browne et al., 2013b)

1.6.5 GSL, New-GSL and *NGSL* comparison

From my point of view, all the lists mentioned above can be used for research. The *GSL* contains some archaic words, which are not common anymore in the recent English language. Therefore, it is advisable to choose from some of its successors as the *new-GSL* or the *NGSL*.

Subjective aspects were applied while setting up the *GSL*. Later on there was a need to work with the service list so the *new-GSL* authors took the downsides, tried to deal with them and created another general service list. Headwords, word forms or families are considered differently than in the *GSL*. The items in the *new-GSL* are lemmas and they include only their inflectional forms. The *new-GSL* should be expandable as it was created with special attention to following the criteria of word selection. This could be a good starting point for further researches. (Březina & Gablasová, 2013, p. 19)

Another general service is the *NGSL* completed with the *NAWL*, they are a major update of the *GSL* and the *AWL* with a helpful set of tools, which are provided to work with a text. The *NGSL* was made to be as similar with the *GSL* as possible. So it does not contain the academic vocabulary, the *NAWL* was created to complement the *NGSL* with academic vocabulary, it is the same relation as between the *GSL* and the *AWL*. The *new-GSL* may carry out similar results as it has more than 80% coverage in the source corpora, the authors also provide a good set of tools.

However, these wordlist with exceptions (for example the *NGSL*) only provide frequency information for the high-frequency words, more precisely up to their total headwords or lemmas count.

The criteria for vocabulary selection into wordlists were subject of many researches, Bauer and Nation in their publication *Word Families* [1993] presented a table of possibilities how word families can be considered. The *GSL* is out of the table as it presents some of the derivations as headwords and some not at the same time. The *new-GSL* uses the *Level 2* criteria as can be seen in the Table 14⁸. The *NGSL* has its own criteria, it is inspired by the *Level 2* criteria but has some extra – for example the British and American spelling count for only one lexeme, also some items were excluded and some will be included again with another major update. (Browne, 2014, pp. 39-40)

1.6.6 BNC-COCA

The *British National Corpus* (BNC) and the *Corpus of Contemporary American English* (COCA) provided data for the BNC-COCA list [2010], which provides 25 k-levels⁹. Paul Nation created the BNC-COCA list in cooperation with Mark Davies by integration the two corpora into one. (Nation, 2012, pp. 1-2)

The data was added into *Lextutor*¹⁰ in 2013 so it is suitable for online text analysis and it shows each single word in the analysed text and its coverage in 25 k-levels which is very helpful for an overview. (Cobb, 2015)

⁸ For details, see the Table 14 in Appendix 1.

⁹ The 25 k-levels correspond with 25,000 frequency levels.

¹⁰ <http://www.lexutor.ca/> (Cobb, 2014)

1.6.7 Vocabulary profilers

Vocabulary profiler is an application that compares an entered text with a frequency wordlist, it usually highlights and sorts items according to their frequency. It also shows the coverage information for each item. This can help with **text level selection** for learners of English. “*Vocabulary profilers break texts down by word frequencies in the language at large, as opposed to in the text itself. Most of them divide the words of texts into either first or second thousand levels, academic words, and the remainder or off-list, or the BNC based 20 levels plus off-list. VP is used for many research and teaching purposes (like matching text to learner via Levels Test).*” (Cobb, 2015)

Vocabulary profilers are useful when creating any language material. The vocabulary frequency level can be checked, it helps to estimate the language level of the whole text. Then the less common items can be taught in context. From my experience, teachers usually do this text level analysis spontaneously and then ask students in the lesson about unknown words. Sometimes the words tend to be only translated into learners’ first language because it is fast, instead of explaining them in other words and different contexts or practicing them.

The vocabulary profiler *Compleat Web VP!* with a set of other tools by Cobb (2015) is available for free use on his website <http://www.lex tutor.ca/>, these tools are the most powerful, have the most options. The possibility to choose from the *GSL-AWL*, *NGSL-NAWL*, *BNC*, *BNC-COCA* and *French-v5* frequency wordlists makes it unique. The output is a set of tables and colour scales for each frequency level of the selected wordlist.

The most useful vocabulary profiler tool for teachers is VocabKitchen’s *CEFR Vocabulary Profiler* (Garner, 2016). It classifies all the words in a text into the *CEFR* language A1 - C2 levels with very detailed information and coloured overview. It highlights the words according to the level so it is very helpful when deciding the level of a text. There is *Image 1* in *Appendix 3* for better illustration.

Another example of vocabulary profiling tool is *English Vocabulary Profile* (CUP, 2015), which covers all the six levels according to the *CEFR* for both British and American English. It helps decide what to teach on a particular level and what is out of the bounds of the level of the learner.

I find this tool very useful for **vocabulary selection**. Especially because it displays the CEFR levels to which the different meanings of the word are related as shown in the *Image 2 in Appendix 3*. It did not work as well as expected with phrases, but the search was successful in case of typing just a single word.

CUP (2015) also published their *English Grammar Profile* tool. It has very similar interface to the profiler mentioned above (see *Image 3 in Appendix 3*) but offers unique possibility of classifying grammar into the *CEFR* language levels, which can be also helpful to know.

Range (Nation, 2016) is another piece of software that is available for vocabulary profiling. It is provided with *GSL and AWL* or *BNC* word frequency lists but it is not any problem to add another frequency list. It is available to download on Nation's webpage. *AntWordProfiler* (Anthony, 2014) can produce very similar results but has even more options. Both of the applications can serve much better when processing more data than the online tools. This application is also freely available at author's website. Anthony also offers other tools for working with corpora and texts.

Many other lexicographers use commercial software called *Sketch Engine* (Lexical Computing, 2016), it is far more powerful, it has access to wide spectrum of corpora and researchers can compose their corpus by an easy selection from the corpora available.

The wordlists are intended on research but also on teaching and many authors find them useful for creating teaching materials. For instance, Březina (2013, p. 1) says that a learner need to acquire the form and the variety of meanings of a given lexical item, lists can aid teachers and students of EFL with the selection of appropriate materials.

1.7 Vocabulary of a Czech learner of English

Roca-Varela (2013, p. 559) researched the words “*actual, career, pretend*” from the high-frequency vocabulary band at Spanish learners of English. These words were selected intentionally, because in Spanish there are words seemingly similar but they differ in meaning so called false friends. She focused on accuracy of these words’ use. She took *LIDSEI* data, which is an oral production database of advanced Spanish learners of English, and compared the words’ use with their actual use. She found out that non-native speakers of English might face serious problems with accuracy and deviant use of these words. The Spanish learners’ use differed from the native speakers’ use of the words. She concluded that although these words are high frequency, the advanced learners could not use them properly. She came up with a question whether the word-frequency helps learners to better acquisition of the words or not.

When searching sources for my work I noticed that studies on differences of the Czech and the English language are often done in the translation-grammar approach.

There is no English language corpus of Czech learners of English so far (2016). For Czech teachers of English it would be great to have one because then they could point out the most problematic vocabulary and word structures for the Czech learners, redesign the materials, which would be very convenient. There are some English language issues related to this subject discussed on the portal for Czech learners of English <http://www.helpforenglish.cz>. (Vít, 2010-2014)

From my experience, for Czech learners it is quite difficult to understand English modality and to pronounce the phonemes not existing in the Czech language. Some pupils often have problems due to the different word order in the two languages.

I believe the learners at lower secondary schools should already know the *International Phonetic Alphabet* (IPA) to transcribe the words’ pronunciation. I meet many learners, who were not taught or are not used to use the *IPA*, especially at lower secondary schools. During my teaching practice, it showed that learners need to be reminded to record the new vocabulary items and checked if they do so.

2 EFL vocabulary teaching

In the beginning of this chapter, I would like to note that each learner has different learning abilities and personal preferences depending on his personality. Therefore, every learner has different learning strategies and preferences. The teaching style should correspond with the learners' learning style to make the process effective. If the styles does not suit at all, the learners may be frustrated and demotivated. Teaching methods carry an obvious impact on the learning process. (El-Hmoudova, 2015, p. 60-61) It is up to the teacher to diversify the teaching process for the learners' needs. In practice it is always a time management question. Information in this section should present possible ways of vocabulary teaching and learning with use of the vocabulary frequency aspect.

Schmitt & McCarthy (1997, p. 8) claim that up to 20,000 word families, the native speakers add around 1,000 word families a year to their vocabulary size. Learning vocabulary is very individual. For non-native speakers there is an initial gap, but the rate of vocabulary growth can be the same. Significant growth is visible in the second language environment [Goulden, Nation & Read, 1990].

Schmitt and McCarthy (1997, p. 10-11) are convinced that with 2,000 words, a learner knows 80 % of the words in a text and approximately two words on a line are unknown. At least 95% coverage is needed for successful guessing of unknown words and reasonable comprehension, much larger vocabulary stands for the remaining 5 %.

2.1 Frameworks

Frameworks for education specify the standards. I would like to focus on the frameworks used in the Czech Republic. Generally, the frameworks say what the learner should be able to do at each level of the language learning. These concepts are usually accepted as some kind of milestones in the learner's progress, it mirrors in the learning activities, vocabulary and the teaching approach.

Common European Framework of Reference for Languages (CEFR) is one of the frameworks and therefore it clearly says the learner's abilities and competences at a certain level of the language. Teachers can follow the provided scales and use them for learner's assessment.

It focuses on measuring the four language skills – reading, speaking, writing and listening. There are six language levels from the starter (A1) to proficient stage of a language (C2). (CEFR, 2016) The *CEFR* includes statements about the vocabulary range of a learner at each of six levels, A1 to C2 as in the Table 6:

C2	Has a good command of a very broad lexical repertoire including idiomatic expressions and colloquialisms; shows awareness of connotative levels of meaning.
C1	Has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions; little obvious searching for expressions or avoidance strategies. Good command of idiomatic expressions and colloquialisms.
B2	Has a good range of vocabulary for matters connected to his/her field and most general topics. Can vary formulation to avoid frequent repetition, but lexical gaps can still cause hesitation and circumlocutions.
B1	Has a sufficient vocabulary to express him/herself with some circumlocutions on most topics pertinent to his/her everyday life such as family, hobbies and interests, work, travel, and current events.
A2	Has sufficient vocabulary to conduct routine, everyday transactions involving familiar situations and topics. Has sufficient vocabulary for the expression of basic communicative needs. Has sufficient vocabulary for coping with simple survival needs.
A1	Has a basic vocabulary repertoire of isolated words and phrases related to particular concrete situations.

Table 6 - Vocabulary range of a learner statements (CEFR, 2001, p. 112)

In the Czech Republic, the *Framework Education Programme for Basic Education “Rámcový vzdělávací program pro základní vzdělávání” (RVP ZV, 2013)* follows the *CEFR* language standards. It does not say how many words a learner needs for a certain level of the language but it describes the abilities and competences of a learner at each level. (RVP ZV, 2013) This information helps schools set up their own *School Education Programme – “Školní vzdělávací program” (ŠVP)* which is more specific in terms of the school and what the learners learn. For EFL teaching the study programme and the lesson plans can be based on an English textbook. *Oxford University Press* (OUP, 2016c) presents on their Czech website the examples of *ŠVP* for their published books. Nevertheless, the other authors also provide the schools with this material. Either it is on the publisher’s website or it is included in the teacher’s book.

According to the *CEFR* (2001), the B1 language level is called “*threshold*”, and it clearly corresponds with the Schmitts’ high frequency level (2014), which is the first 3,000 word families. They say that beyond this level a learner has new opportunities for learning from various context by himself, for instance by watching films and reading. For the A2 level learners, the knowledge of the words up to the 2,000 frequency level should be sufficient.

2.2 Teaching and learning strategies

Jařab (2016) and some of his friends learnt *Oxford English Dictionary* by heart when trying to learn English. During his informal meeting with public, he commented on this fact laughing that it is not the way languages are possible to be learnt. In this section, I would like to point out strategies that may work.

According to Alharbi (2015, p. 505), there are different strategies for EFL vocabulary learning, I will name and describe at least some of them. **Cognitive strategy** is focused on reading and writing skills, the strategy aims at vocabulary relations, for example synonyms. The tasks of the learners connected with learning a new word would be checking if there are synonyms, using a semantic map and experience different contexts of the new word. All of these steps should lead to the expansion of vocabulary size, easier learning, remembering and development of connection for word relations and usage. **Meta-cognitive & social strategies** are often used by the applications of frequency wordlists. They are related to learning definitions with contexts. The learners would guess the meaning from the context, check the words’ context in dictionaries, check the meaning and then apply it through the four language skills. All of that should help the proper use of new words and apply the previous knowledge to guess the meaning in the future. **The social strategy** itself is based on communication, listening and pronunciation in conversation and listening. That should enhance the learners’ speaking and listening skills. Another strategy that is often used is the **bookmark strategy** (meta-cognitive), which makes the learners set up their personal vocabulary notebook. Regular revision and checking new words is essential for this strategy. It supports learners’ memory and improves long-term memory practices. The last strategy I would like to name is **memory strategy**. There are some techniques for learning vocabulary, learners can write the new words (helps to remember), picture or visualize them, create some cards and use other mnemonics. The vocabulary comprehension increases using these strategies.

According to Shen (2003, p. 192), the common vocabulary learning strategies can be divided into contextual and decontextual. The contextual strategies are the activities related to the four language skills: *listening, speaking, reading and writing*. *Wordlists, flash cards and conventional use of dictionaries* belong amongst the decontextual strategies. However, there are also strategies that are in between and connect the two extremes: “*word grouping, word concept association, aural and visual imagery, keyword, physical response, physical sensation and semantic mapping.*”

It is recommended for the learners to acquire the high frequency words (3,000) as fast as possible. For initial learning, there are not many different options than using wordlists and cards, it is fast and so it is good for idle moments in the class. (Schmitt & McCarthy, 1997, p. 11-12) Lessard-Clouston (2013, p. 5) advises teachers to begin with learner’s strong vocabulary dimensions, learning vocabulary should be individual, another point is that each teacher usually focuses on some of the dimensions more.

Roca-Varela (2013, p. 560-561) encourages EFL teachers to teach lexical items through illustrative examples, which make their use and meaning clear. Each language may have different denotation for lookalike words. Students should be taught differences between languages in order not to make these mistakes. Furthermore, the EFL learners should be led to study high-frequency words more properly. While learning new words or their usage, they should pay attention to their form, meaning and linguistic features.

Alber (2014) uses for her teaching very similar vocabulary division as we can see at the *NGSL*. There are three 1,000 levels according to the frequency. When she prepares a material or recommends a book for reading, she divides the vocabulary into the frequency bands, creates charts with the words and asks learners to check if they understand the words or not. Afterwards, the learners are prepared to focus just on the text comprehension and mostly they enjoy the activity.

The learning can be done from context by reading or listening, then the learning is cumulative, it is relatively slow but learning from context gets more important with growing level of the learner’s language. The process is expected to continue even outside the classes. (Schmitt & McCarthy, 1997, p. 11-12) In the vocabulary learning process, it is important to achieve a balance between meaning, fluency and form focused activities. (Schmitt & McCarthy, 1997, p. 11-12)

Learners need multiple exposures to a word before they fully understand that word and can apply it. From the time of the first exposure, the vocabulary should be revised, more context provided and applied. Learning from the context that is appropriate to the age group of the learners and their level of language is much more effective than just filling in definitions or copying wordlists with words that they will never use. *“This is why every classroom should have a killer classroom library stocked full of high-interest, age appropriate books.”* (Alber, 2014)

It is also important to deal with the number of the items that the learners are able to learn during one lesson. For lower language levels, learning around eight items per a lesson should be manageable and for the higher level learners it would be about twelve items. (Gairns and Redman, 1986, p. 66) I think this attitude is present in current English textbooks. I have experienced as a learner that there has always been just about ten new items a lesson to learn.

2.2.1 Classroom English

In the time of reading the Schmitts’ article (2014), I was teaching at my second teaching practice and a part of it was a classroom English project. It was about the use of phrases for commands and work in the classroom. The Schmitts also write that teachers tend to use the high frequency vocabulary when teaching. Then I thought that the classroom English can be a very good example for real use of vocabulary frequency aspect and at the same time, it is one of the teaching strategies. The classroom language repeats, is easy to demonstrate and besides understand and the learners’ comprehension is visible immediately.

2.2.2 CLIL

Content and Language Integrated Learning (CLIL), according to Klimova (2012, p. 572), is an ordinary learning of the subjects that are taught in the target language instead of the learners’ first language. It can motivate the learners as they see the real use and the context of the target language and help them acquire new vocabulary. As for the vocabulary frequency, the teachers can provide such materials that would contain the common words in areas of the subject. Additionally, it can highlight the difficult vocabulary that can be focused on. I am aware the use and frequency will be in conflict, so I would consider frequency as a complementary and supportive aspect.

2.2.3 Vocabulary selection

It is not effective for learning to supply learners with all the different meanings and vocabulary features first time they encounter the word. Preferably, a teacher should select what to teach first. Word spelling, pronunciation and basic meaning is essential. In another lessons, teacher can add more information, phrases and meanings according to the learners' language level. (McCarten, 2007, p. 18)

There should always be a compromise between learners' and teacher's possibilities. The demands on learning involve the language level, difficulty, atmosphere, interests, needs, motivation and so on. Vocabulary has so many dimensions¹¹ that they cannot be covered at once. A teacher can teach opposites, creating words by word formation and use any context to show the meaning and the word concepts. According to Šmajdlerová (2010, p. 19) the selection of vocabulary can be done based on these important criteria – *“frequency, coverage, need and level.”*

2.3 Materials based on word frequency

In this section, I would like to describe materials, which make use of the vocabulary frequency. I would like to provide some examples with a brief description of each material and suggestions for working with it. In my opinion, materials on vocabulary should try to cover all the important aspects of vocabulary, spelling, pronunciation and its use in context. It should be the teacher's choice to come up with a way to teach the items as already mentioned.

Information about word frequency can help us work with the teaching materials. Frequency lists are useful to help us decide about what to teach and in what order. The rare expressions can be taught later on, while we can focus on teaching the vocabulary items used more often. To illustrate, when teaching a large vocabulary set (colours, hobbies, types of music, cloths, health, etc.), it can be beneficial to teach the items the learners would use at first. Frequency information use in corpus-informed materials can be almost invisible but can bring an ease and some kind of a starting point when deciding about a material. (McCarten, 2007, p. 4)

¹¹ For more details see the Table 1 in the 1.1 Vocabulary chapter of this thesis (p. 10)

This topic is discussed even in the article *A reassessment of frequency and vocabulary size in L2 vocabulary teaching* by the Schmitts (2014), which mentions the importance of reaching the first 3,000 frequency level.

2.3.1 Dictionaries

Rundell (1998, pp. 315-317) argues that **monolingual dictionaries** (MLDs) nowadays are a good support for language learners. There are MLDs meant to be used mainly by native speakers, but from 1920s, they evolved and there are dictionaries for the purpose of the learners – **learner dictionaries** (LDs). LDs should aim to cover just a selected subset of the lexicon. LDs usually omit words that are highly technical or rare. Descriptions of words should be so clear that no further searching is needed.

Application corpus data to the dictionary making process prove the frequency use for learning a language. Rundell (1998, p. 337) points out the limits of automatization in terms of analysis, discovery and description of meaning.

The lists of defining vocabulary are created in order to define the words in a learner's dictionary. These words are usually the high frequency vocabulary. (Schmitts, 2014, p. 491)

MLDs found their use in electronic book readers, for example Amazon brought the possibility to read e-books with an immediate assistance of a reader's choice dictionary, for example *Oxford Dictionary of English* or *New Oxford American Dictionary*. (Amazon, 2016)

For the learners of English generally, I would recommend using one of the learner's dictionaries by *Oxford*, *Cambridge* or *Macmillan*. For the lower secondary school learners, it should be sufficient to use the dictionaries containing “*essential*” in their name, it covers the vocabulary of target level A1-A2, in other words elementary and pre-intermediate. (*englishbooks.cz*, 2016a)

However, there is the Czech edition of *Oxford Student's Dictionary*, which is available and has the Czech translation included so it is more than advisable to use this version as there is no other learner's dictionary like this. The publication's name is *Oxford Studijní slovník + CD-ROM - česká edice* and it was published in 2010 (OUP, 2010).

As it says, the target level is between B1 and C1, the vocabulary selection is based on the frequency list *Oxford 3000™ keywords*. It provides words with British and American pronunciation, easy explanations and for each meaning at least one sample sentence.

2.3.2 Textbooks

Okamoto (2015, pp. 2-3) writes about methods and criteria of selecting vocabulary for EFL textbooks. Vocabulary can be selected by making use of the corpus analysis, and relying on native speakers' judgements. Reliability in vocabulary selection is a topic for discussion. Native speakers feel that words beyond 11,000 word level are extremely rare. They seem to make reasonable judgements based on the frequency of their actual word use up to the 7,000 or 6,000 frequency level. Okamoto claims that native speakers should be regarded as complementary "*yardsticks*" for selecting vocabulary to teach under temporally restricted conditions. (Okamoto, 2015, p. 6)

Textbooks are also a subject of many researches. For example, an analysis of vocabulary from the bestselling *New Headway Upper-intermediate* showed that high frequency vocabulary provided 95.5% coverage of the textbook (44,877 running words), 66.4 % of the remaining 1,005 word families occurred only once, only 12.1 % occurred five and more times. (Schmitts, 2014, p. 499) This shows a small amount of words recycled though the text so it is up to the teacher to use the vocabulary for instructions - classroom English, or supply the learners with other materials to revise the vocabulary, OUP (2016b) offers such extra materials in the workbooks and online, each publisher may have different approach to these materials. There are more examples of textbooks in the research part of the thesis.

Among Czech learners of English at the lower secondary schools, for example, these textbook series are used (SEVT, 2016):

- *Project*
- *More!*
- *English Plus*
- *Messages*
- *Angličtina 6-9 Way to Win*

2.3.3 Wordlists

This section is related to the wordlist as a textbook part or a learner's support, note that the wordlist here has mainly learning purposes and it does not cover all the vocabulary dimensions as the *Table 1*¹² shows.

OUP (2016b) freely offers wordlists, which are parts of their published textbooks. They can be usually found at the end of the book or a unit, therefore the vocabulary list concerns specific topics according to the unit's content. I provide examples and details about textbook wordlists in the research.

Other examples of wordlists can be the general service lists that serve for research and word selection for education. Many publishers dealing with the language materials like *Oxford*, *Cambridge* or *Macmillan* try to sort out the corpus data into wordlists that would be helpful for teaching.

Candel (2013) describes OUP's wordlist of 3,000 most common words called *Oxford 3000*TM [2012]. It is a list of frequency ranked words made for learning. It contains descriptions, meanings and the *CEFR* language level for each word.

According to Candel (2013), the list is very helpful for his learners, he provides them with the list in electronic form in the beginning of the school year and they fill in the definitions, the word family information and use the words in sentences. The electronic form of the wordlist with empty columns for students can be found on the internet in a version for PCs and tablets with operating system *Windows* and *iOS*.

¹² The Table 1 is presented on page 10 in this thesis

2.3.4 Graded readers

Graded readers are books that provide simplified adaptations of original stories based on vocabulary frequency. This can be helpful especially in the beginning to help develop the vocabulary fast. The main benefit is learning meaningfully from the context that can be entertaining for the learner. On the other hand, these books sometimes do not cover pronunciation, which is a necessary part of vocabulary. This may later be an issue for the learners.

Most of the graded reader series finish at around 3,000 word family level, reaching this level is an important stage for EFL learners. Oxford series have vocabulary level between 200 and 5,000 headwords. Although the books are based on some limited vocabulary, this does not mean that only the high frequency words are used. There are different counts of headwords for different levels of English, this is a matter for each publisher. To illustrate, let's say the first 1,000 frequency level would stand for three quarters of a text, however, the rest would be covered by words from between 1,000 and 9,000 headwords frequency range. (Schmitts, 2014, p. 491, 500)

In bookshops, we can find the books from publishers such as *Macmillan Readers*, *Penguin Readers* and *Collin English Readers*. (englishbooks.cz, 2016b) There may be also bilingual versions. There are graded readers freely available on Paul Nation's website always in several level versions (Nation, 2016). Nation (2013, p. 18) advises teachers to provide elementary and intermediate learners with graded readers books and spend one quarter of the course time on the extensive reading of the appropriate level books.

I really recommend the BBC Learning English portal (BBC, 2016) with materials for EFL learning and teaching sorted according to the language level. There can be found materials for **reading** – simplified news or popular science articles. They also provide listening exercises and what I like the most from what is available – a **drama**, these are simplified novels turned into short acts that can be used for a **roleplay**.

2.3.5 Software

Educational software focused on vocabulary is usually available for many different platforms. Many applications aimed at mobile devices such as tablets or smartphones are accessible online or on the PC at the same time. Textbook content related software is often provided in the textbook and another is present at the publisher's website.

I reviewed some of the recent applications not tied to a textbook that are available for mobile devices and apply the vocabulary frequency aspect.

*Memrise*¹³ offers learning based on frequency. The main purpose of the application is learning words by matching them with the right meaning or pronunciation (audio) and it is supported by images. Learning can be scheduled and reminded in notifications.

The users can choose from various courses, some of them are made from the wordlists that are available, for example the *NGSL* or textbook wordlists (*New Headway*, *New English File*). Furthermore, there are other than the English language options.

I really appreciate that the learning is done in the target language and the large offer of courses. I would recommend this application to the learners who want to revise their vocabulary and add some new items.

*Duolingo*¹⁴ aims at the people who like to use their first language when learning another language. Apart from vocabulary, it teaches the grammar. The grammar-translation approach is used together with listening. The application asks the learner to translate even whole sentences.

It is a complete online course for the languages, which are available. There are no other options other than choosing the first and the target language. At the first launch, the application provides a test to adjust to the learner's language level. It also supports scheduled learning and provides a reminder.

¹³ *Memrise*. [software] Memrise Ltd, 2016. Available at URL: <http://www.memrise.com>

¹⁴ *Duolingo*. [software] 2016. Available at URL: <http://www.duolingo.com>

I liked the fact the vocabulary and grammar is provided in topics. I would let the learners try the application and decide whether they like the way of language learning or not.

*Quizlet*¹⁵ provides vocabulary learning based on matching definitions with the word forms according to the meaning, audio is included. It is quite innovative as it encourages teamwork in a game for the whole class, where the teams compete.

I recommend this application to the teachers, the advantage over the other applications is that they can create study sets of words for their students. Considering the game, it can be another fun activity in the class.

Another useful application is *Wordle* (Feinberg, 2014), it can transfer any text to image cloud of words where the word font size will be based on number of occurrences of each word in the text. It is very useful for introduction of any topic, the brainstorming or summary activities.

2.3.6 Games

Games are a perfect way of learning new vocabulary. Frequency may help in selection of vocabulary. *CEFR* proposes the communicative teaching approach, which supports game activities in the classroom. Vocabulary can be taught through such activities in combination with any of the language skills.

“Some of the frequency information is fun to know and can be used in guessing game activities in class.” (McCarten, 2007, p. 4) For example, students can guess the most used phrase for saying hello, talking about weather or clothes.

On the portal <http://www.freerice.com>, one can donate an amount of rice to the people in need when matching a meaning with a word form correctly. (Lessard-Clouston, 2013, p. 6) The game is based on a frequency wordlist similarly to the applications mentioned above but does not offer so many options.

¹⁵ *Quizlet*. [software] 2016. Available at URL: <https://quizlet.com>

Guessing vocabulary frequency can be a game that tests your frequency awareness. Macmillan Dictionary (2016) published such game on their web pages. The game is called *Red Words*. The rules are clear, it displays a word and one has to guess the word frequency on a scale of one to three stars. Three stars stand for the most common 2,500 English words, two stars represent the next 2,500 most common words and the only star refers to the words in between 5,000 and 7,500 frequency bands. *Macmillan Dictionary* use this three stars system in their dictionaries to demonstrate the frequency aspect of the words to the learners. (Macmillan Publishers, 2016)

Another online game based on frequency and a set of questions is *QGame*. The rules of the game are that one specifies an animal before the game starts. *QGame* asks a set of 15 questions on the animal properties and one selects the answers. After answering these questions, *QGame* estimates the animal that one had in mind. The game is focused on practising different animal names. (Přichystal, 2013)

McCarten (2007, p. 18) suggests teaching vocabulary in relation to grammar patterns in games (*I like/ I can't stand game*). It is advisable to focus each game on a particular word meaning and grammatical structure.

Many game-like activities can be done in a lesson using an interactive board. This can save time for preparation of the activity, because there are many materials available on the internet ready to be used. The classic games take use of the frequency aspect.

For example, in the well-known game *Hangman*, which is about learners' guessing the right word spelling according to some clues (usually left empty spaces, general descriptions), the teacher can use the word frequency rate for better motivation at the beginning or at the end of the activity. He or she can ask the learners to guess how frequent the word is. The vocabulary frequency can be also used just to complete the information or prepare a reasonable set of vocabulary.

In the same way a modification of the game *Who am I?* can be played. In this game, the learners' ask yes or no questions in order to guess what or who the person is. The modification could be implemented in a way that the teacher would use the frequency of the word as a guessing clue to motivate the learners. Another modification could be just completing the information about the subject of the guessing.

A very popular game activity amongst my students is *AZ Quiz*, which is a knowledge competition show on a Czech television channel. The game ends by connecting at least three edges of the playing field (usually in a shape of a triangle) which is divided into many other fields labelled by the initial letters of words a teacher prepares. The competitors ask the teacher for a question based on the initial letters. The teacher is the moderator of the competition who uses a prepared set of words and their definitions, reads the definition according to the initial letter of the competitors' choice. If the learners answer correctly, they get the field, otherwise they just lose their turn. The game is fun to be played also in teams. It should not be a problem to motivate three teams of four students to compete. The frequency aspect can be applied in the competition word selection. The game can focus on a particular language level and topic.

Word dominoes is another beloved game of my students. The teacher prepares a set of objects, words or picture cards and asks the learners to play domino. It means the students will match two cards for a reason that they should explain. Again, the game vocabulary can be adjusted to the learners' needs based on the frequency.

I could mention many other games including some of them with use of the physical movements or creative skills. The attitude to the vocabulary frequency aspect would always be very similar.

3 Research

In this chapter, I would like to cover the question on how to deal with the English vocabulary frequency data. My research is based on considering various wordlists included in textbooks or workbooks for use in the lower secondary classes and their vocabulary frequency analysis with help of a reference general service list.

The aim of this research is to find out whether high frequency vocabulary items are present in the wordlists in several English textbooks used among Czech lower secondary learners of English and to determine to what extent.

3.1 Methodology

The premise of this research is that the wordlists in English textbooks represent the vocabulary that needs to be learnt.

The methodology needs to deal with the selection of the following items:

- English textbooks
- a reference wordlist
- a vocabulary profiler tool

3.1.1 Textbooks selection

For this research purposes, I considered the target language level and use of the English textbooks amongst Czech learners as the criteria to select the books containing wordlists.

The first criterion is that the level of language should correspond with the lower secondary school language level according to *RVP ZV* (2013) and *CEFR* (2001) as discussed in the theory. This means that books of target level A1-B1 can be selected. Secondly, the textbook series should be used among Czech learners of English. I choose to base the textbooks series selection on the list of bestselling English textbook series in Czech e-shop *SEVT*. The bestselling textbook series include following: “*Project Third Edition*” by Hutchinson (2009), “*Hello Kids!*” by Zahálková (2010), “*More!*” by Puchta et al. (2008), “*English Plus*” by Hardy-Gould & Mellersh (2011), “*Messages*” by Goodey et al. (2006), and “*Angličtina Way to Win*” by Betáková & Dvořáková (2006). Except for “*Hello Kids!*” series, which consists of books even for the primary school learners, these series cover the lower secondary schools. (SEVT, 2016)

I also compared the textbook series selection with my own experience with the textbooks that I used during my teaching practices and I believe the selection reflects the use of textbooks at several lower secondary schools in Olomouc region. Due to the thesis aims, I made a selection of books with the target language level A2 (pre-intermediate), which reflects the target language of the 9th grade learners according to the Czech school system, which is needed for the high school studies. In addition, the textbooks were approved for use at Czech elementary schools (MŠMT, 2013). I present the selected books containing the wordlists for research below:

- *Project 4 the Third Edition Workbook (Czech Version)* (Hutchinson, 2009).
- *Hello Kids! Angličtina pro 7. ročník základní školy - pracovní sešit* (Zahálková, 2010)
- *More! Level 3 Student's Book with interactive CD-ROM* (Puchta et al., 2008)
- *English Plus 2 Workbook with MultiROM (Czech Edition)* (Hardy-Gould & Mellersh, 2011)
- *Messages 3 Student's Book* (Goodey et al., 2006)
- *Angličtina 7 Way to Win* (Betáková & Dvořáková, 2006)

3.1.2 Description of the wordlists

Wordlists with vocabulary were available in all the textbook series I selected. Each of the selected textbook wordlists is briefly described in this section. The textbook usually presents the vocabulary in the student's book in the context of articles, listening and other exercises for better learning experience and then the new vocabulary is listed in the wordlist. In workbooks, the authors focus on practising the new grammar and vocabulary.

3.1.2.1 Project 4 the Third Edition (Hutchinson, 2009)

The textbook series aims at the pupils of the sixth to ninth grade, there are five books, the first book can be used for fifth or sixth grade and the last for ninth grade learners. The wordlists in *Project the Third Edition* textbook series are part of their workbooks usually printed at the end. The vocabulary is divided into sections according to the textbook units and unit names are used as the section titles. The wordlists include single, two or more words expressions, one version of the IPA transcription and usually only one of the meanings in Czech language is listed.

3.1.2.2 *Hello Kids!* (Zahálková, 2010)

Czech series of English textbooks *Hello Kids* by Zahálková provide a wordlist in the workbook sorted into sections according to the textbook units. In the student's book, the wordlist is also available sorted alphabetically with notes about the units. The wordlist includes its pronunciation and Czech translation for each item. The pronunciation is not according to the *IPA*, for example, the pronunciation of the word "bench" is transcribed as /benč/, which should be /bentʃ/ according to the *IPA*.

3.1.2.3 *More! Level 3* (Puchta et al., 2008)

The textbook series includes textbooks of four different language levels so it nicely fits the Czech school system lower secondary grades. In *More! L3*, the textbook wordlist contains vocabulary items sorted into sections according to the units in the student's book, the unit title is on the list, one version of the *IPA* transcription and translation into the Czech language is present.

3.1.2.4 *English Plus 2* (Hardy-Gould & Mellersh, 2011)

The textbook series consists of four books for the 6 - 9th grade. The wordlist in *English Plus 2* contains items sorted according to the units. For items related to the same topic, there are pictures demonstrating the concept of the item, moreover, there is the English word equivalent for each image. The *IPA* pronunciation is present on the list. The *IPA* symbols are explained using simple words at the beginning of the list. There is no Czech translation of the items.

3.1.2.5 *Messages 3* (Goodey et al., 2006)

The textbook series composes of four different level textbooks. Each level has a student's book with the classroom activities, a workbook for further grammar and vocabulary practice and a teacher's book with methodology. The wordlist is a part of the student's book. It is divided into sections according to the topics that are covered in the modules. The units that are part of the modules contain grammar. The wordlist is set up from words, phrases and collocations. Each item provides pronunciation and Czech translation. Many of the items are explained in additional phrases.

3.1.2.6 *Angličtina 7 Way to Win* (Betáková & Dvořáková, 2006)

The textbook series provide full support for Czech learners of English. The vocabulary is presented in the articles in the student's book and then in the lists at the end of each lesson. The wordlist contains a lot of phrases, the vocabulary meaning is shown in the context of the articles, the *IPA* transcription of the pronunciation and the Czech translation. I have taught 34 lessons using this textbook series and I can conclude that the textbook fits the learner's needs in terms of vocabulary.

3.1.3 Data processing

I transferred all the English data from the wordlists into an Excel table file (Appendix 5), which can be easily analysed. The wordlists contained not only single words but also phrases, collocations, definitions and phonetic transcription. Note that definitions and pronunciation related items are excluded from this research. See the *Table 7* for the items count of each wordlist.

English Plus 2 wordlist (Hardy-Gould & Mellersh, 2011)	809 items
Project 4 the Third Edition wordlist (Hutchinson, 2009)	678 items
More! L3 wordlist (Puchta et al., 2008)	653 items
Hello Kids! Angličtina pro 7. ročník ZŠ wordlist (Zahálková, 2010)	529 items
Messages 3 wordlist (Goodey et al., 2006)	407 items
Angličtina 7 Way to Win wordlist (Betáková & Dvořáková, 2006)	369 items

Table 7 – Textbook wordlists sorted according to the number of items

3.1.4 Reference list selection

Because the focus of this work is the A2 level, the data in the general service lists should be sufficient for the research and correspond with what the learners learn. For the actual analysis, I considered using one out of the three general service lists that are described in the theory part of my work. Finally, I decided to work with *NGSL and NAWL* because my pre-research showed that it had the best coverage among the general service lists. I consider it suitable for pedagogical purposes as the academic vocabulary results can be seen separately and it is used widely with a very good support of various tools. This will show the results up to the 2,500 frequency level.

Then I would like to compare these results with the *BNC-COCA wordlists* because they can classify the words into the levels up to the 25-k level (25,000 frequency level) and that would illustrate the vocabulary coverage beyond the high frequency vocabulary level.

3.1.5 Vocabulary profiler tool selection

I decided the textbook wordlist data should be put into a vocabulary profiler tool for the analysis. Note that most of the vocabulary profiler tools are made primarily to analyse texts although they can also analyse the words and phrases separately.

As for the vocabulary profiler tool, there are many options to choose from (see the section 1.6.7). These textbook wordlists do not contain so much data so I can also select the online tools. For the *NGSL* and *NAWL* and *BNC-COCA* frequency wordlists I selected *Compleat Web VP!* (Cobb, 2014) which is available online. It outputs the needed data in a reliable way, it has been developed over many years and provides options appropriate for this research. These wordlists contain only single words so while using the *NGSL* and *NAWL* we will get output in tokens and using the *BNC-COCA* option will list word families. Comparison will be made based on tokens percentage in each *NGSL* frequency level.

Firstly, the vocabulary profiler analyses a wordlist, it will classify the items into tokens according to the reference wordlist criteria. It will provide an overview in a table with the information about the token coverage by the reference lists and at the same time detailed overview of vocabulary division among the frequency bands of the reference wordlist. The overview table data will be used to demonstrate the coverage of each frequency level, based on which I will create figures. All the data will be added to the thesis in the appendices (Appendix 2 and 4).

Note that all of the vocabulary profiler tools output information about the token coverage and distribution into the individual frequency levels. In this case, the occurrence numbers of individual tokens, as seen in the appendices, are irrelevant because the main aim is to check the frequency levels of the items in the textbook wordlists.

Some of the words will be out of the reference lists frequency range because they were either excluded from the vocabulary selection when setting up the *NGSL*, *NAWL* and *BNC-COCA wordlist* or the particular textbook wordlist provides words not covered in the frequency reference wordlists selected for this research.

3.2 Results

In this section, I would like to show the results from the research for each book wordlist separately. I provide two figures showing all of the results for a quick overview¹⁶. For each of the wordlists full data from the research are available in appendices. To demonstrate the *NGSL* levels in the research, I created for each wordlist a table with ten example items for each of the *NGSL* category. In appendices, complete information is available for each wordlist. The data is based on the output of *Compleat Web VP* (Cobb, 2014). In the following figures, the *NGSL_1* is the first 1,000 lemmas, *NGSL_2* second 1,000 lemmas and the *NGSL_3* stands for the following 801 lemmas according to the *NGSL*. The *NAWL* band contains the frequent academic vocabulary.

3.2.1 Project 4 the Third Edition¹⁷

The wordlist belonging to the textbook *Project 4 Third Edition* provides the learners with 678 items, which were classified by the vocabulary profiler as 903 tokens. From the total number of tokens, 45 % tokens match the first *NGSL* frequency level, 18 % match the second level and 8 % of the tokens fall into the third *NGSL* level. Almost **71 % of the tokens** are presented by the high-frequency words in the *NGSL*. About 5 % tokens stand for academic words (*NAWL*) and 24 % items are out of the frequency boundaries of the *NGSL*.¹⁸

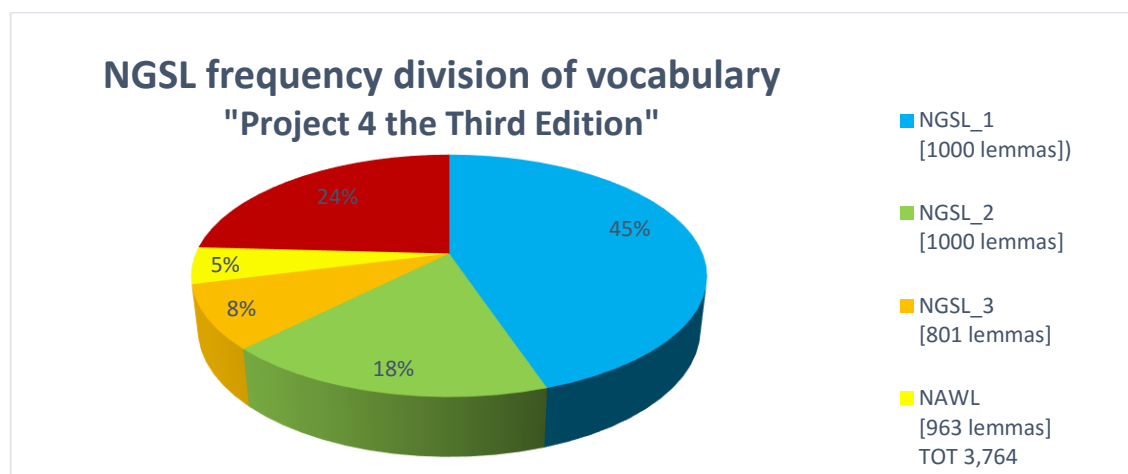


Figure 1 – The *NGSL* frequency level division of *Project 4 the Third Edition* wordlist

¹⁶ Complete research data is available in Appendix 5 (CD) and in Appendix 2 (tables) and 4

¹⁷ All the section refers to Hutchinson (2009), data is the output of *Compleat Web VP* (Cobb, 2014)

¹⁸ For more details see Table 15 in Appendix 2a and raw data in Appendix 4.1

While the *NGSL* shows 218 tokens to be out of the list, the *BNC-COCA wordlists* (Figure 2) show that most items were classified into the first 3,000 frequency level and the rest counts as just a minor number of the word families.

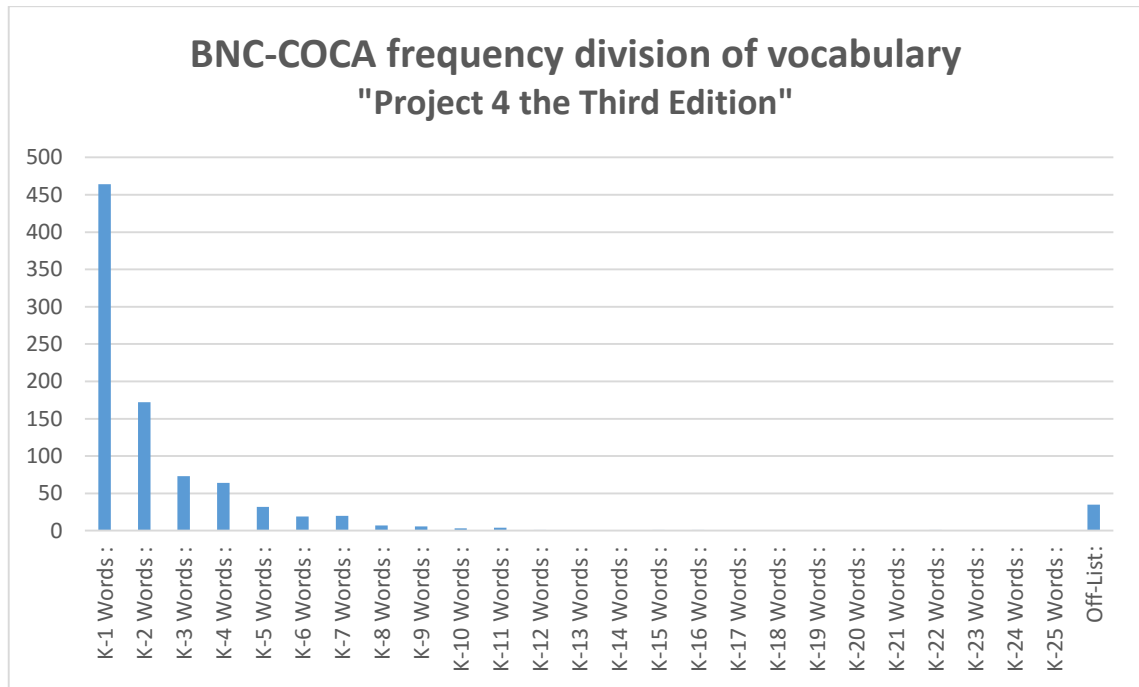


Figure 2 - The *BNC-COCA* frequency level division of *Project 4 the Third Edition* wordlist

I appreciate that most of the wordlist items (almost 71 %) fit the *NGSL* levels so it proves that the vocabulary contains the high-frequency words. In addition, I considered the number of items out of the *NGSL* level boundary, the words like *appliance*, *bandage*, *carbohydrate*, *folder* or *strenuous* are so specific that they simply cannot fit these levels.

Freq. Level	Example items
NGSL_1	able, be, foreign, have, low, pull, pressure, read, tell, worry
NGSL_2	abroad, afford, army, blow, calm, criminal, excite, excuse, hide, smell
NGSL_3	adjust, dig, electric, pollution, leather, loose, magic, stomach, trap, unite
NAWL	absorb, calcium, cattle, goods, lifetime, liver, molecule, oxygen, ray, tribe
Off-List:	alloy, cabbage, cotton, donation, dive, dump, fame, tablet, typhoon, wool

Table 8 - Ten example items from *Project 4 the Third Edition* for each *NGSL* level

The *BNC-COCA wordlists* better point out the low-frequency words. The words that did not reach the k-25 level were for example *CPU, desktop, download, fundraising* or *tugboat*. In my opinion, these words can be understood well, but are classified this way because of their frequency. It is interesting that for almost all of them the word formation process was compounding.

3.2.2 Hello Kids! Angličtina pro 7. ročník ZŠ¹⁹

The wordlist of 529 items was classified as 624 tokens. About 36 % of the items belong to the first, 15 % to the second and around 9 % to the third NGSL frequency level. Almost 5 % of the tokens corresponds with the NAWL and a big percentage of tokens – about 35 % are out of all the levels. About **60% of the tokens** on the wordlist is of the high-frequency vocabulary.²⁰

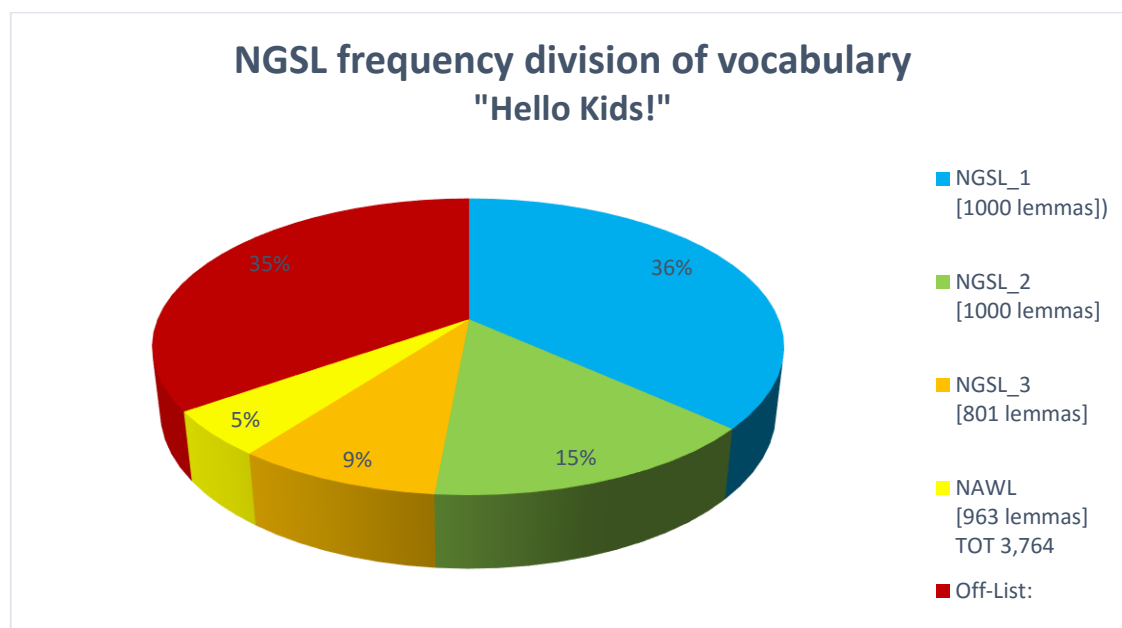


Figure 3 - The NGSL frequency level division of Hello Kids! wordlist

The data shows that most of the vocabulary that the learners should learn is from high-frequency bands²¹. The words out of the *NGSL* boundaries are, for example, *beech, blister, diagonally* and *platypus*, I do not consider these words so much useful for everyday conversation.

¹⁹ All the section refers to Zahálková (2010), data is the output of *Compleat Web VP* (Cobb, 2014)

²⁰ For more details see Table 16 in Appendix 2a and raw data in Appendix 4.2

²¹ Up to the 3,000 frequency level

Freq. Level	Example items
NGSL_1	age, change, do, make, land, leave, instead, money, run, surprise
NGSL_2	accident, battle, desk, enemy, heat, settle, smell, speed, ticket, unfortunately
NGSL_3	altogether, ceremony, counter, illegal, muscle, prisoner, pub, slave, van, web
NAWL	accent, arrow, cattle, liver, lung, nasty, oxygen, painful, recupe, tribe
Off-List:	avenue, Roman, clan, disabled, Welsh, pollute, sealion, skull, vomit, yawn

Table 9 - Ten example items from *Hello Kids! Angličtina pro 7. Ročník ZŠ* for each *NGSL* level

On the other hand, there are proper words like *Roman* or *Welsh* that were not included while setting up the *NGSL* because of the vocabulary selection rules of the *NGSL* although these words are useful for the learners to learn. Many one topic related words appear in the third frequency *NGSL* and the *NAWL* level.

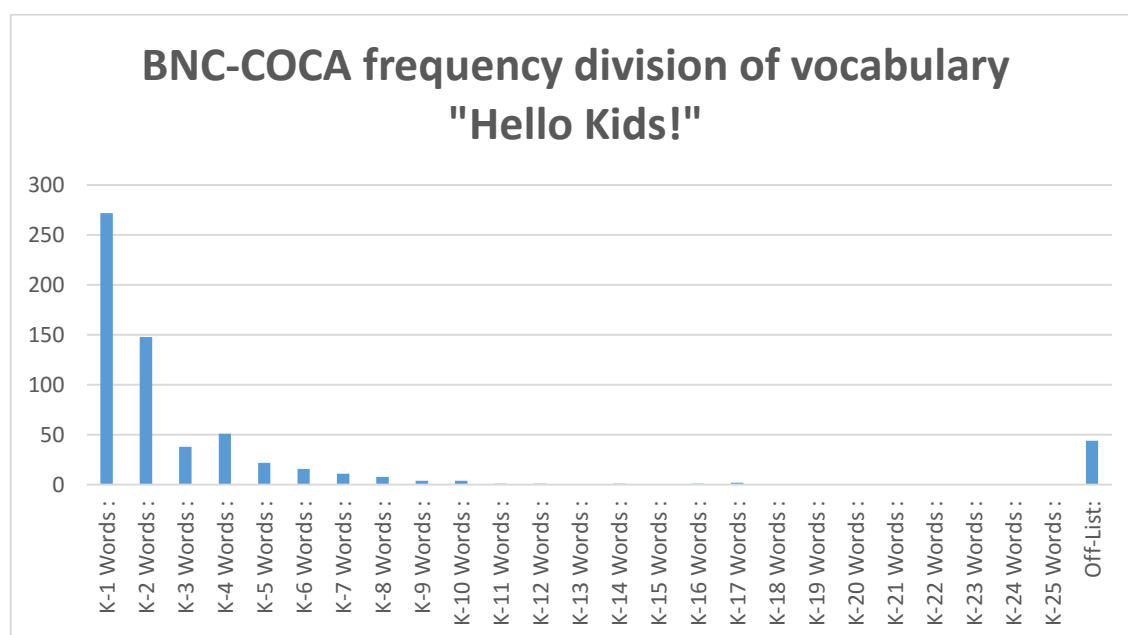


Figure 4 - The *BNC-COCA* frequency level division of *Hello Kids! Angličtina pro 7. ročník ZŠ* wordlist

According to the *BNC-COCA* lists, most of the items belong to the first 4,000 frequency level. It confirms that the high occurrence of high-frequency words on the wordlist.

3.2.3 More! L3²²

The wordlist belonging to *More! L3* textbook contains 653 items and they were classified as 756 tokens. About 31 % of the tokens are for the first, 21 % for the second and 11 % for the third NGSL frequency level, which stands for **nearly 64 % of the tokens**. Almost 31 % of the tokens stand for the off-list items and about 5 % for the academic vocabulary.²³

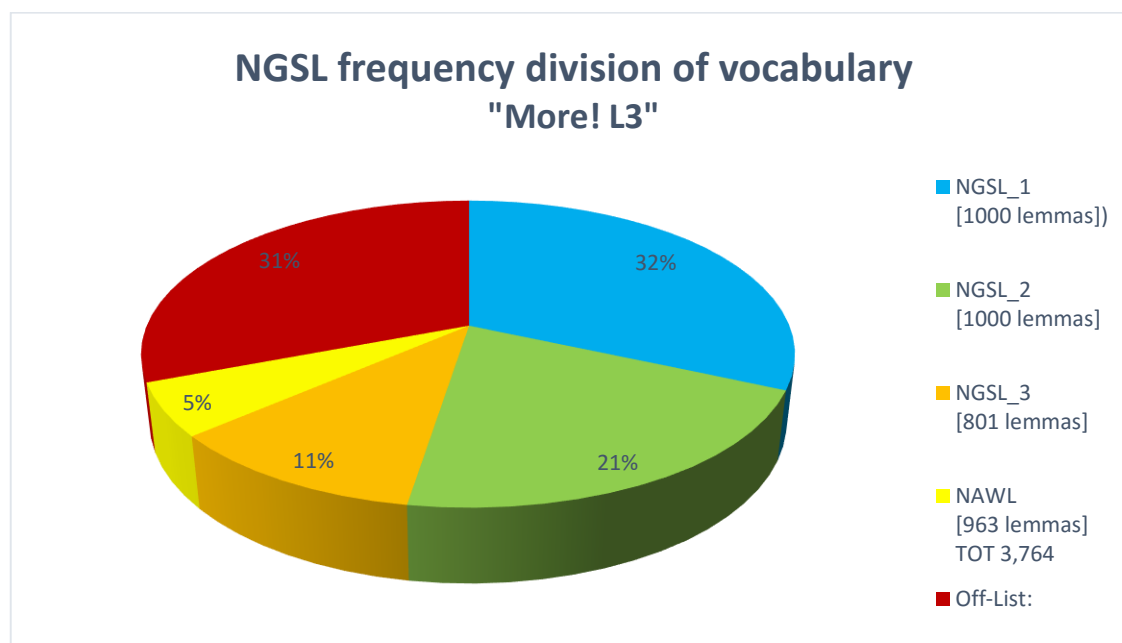


Figure 5 - The NGSL frequency level division of *More! L3* wordlist

Freq. Level	Example items
NGSL_1	add, after, build, cover, deal, event, experience, light, with, you
NGSL_2	abroad, advertisement, definitely, investigate, load, narrow, noise, twice, waste
NGSL_3	admire, apologize, cough, invent, illegal, hero, possession, refugee, trap, violent
NAWL	architect, breakdown, cone, ecology, insect, invade, mechanic, solar, treaty, tribe
Off-List:	aftershock, alibi, buggy, cd, carving, mudslide, palmtop, spymaster, torn, usb

Table 10 - Ten example items from *More! L3* for each NGSL level

²² All the section refers to Puchta et al. (2008), data is the output of *Compleat Web VP* (Cobb, 2014)

²³ For more details see Table 17 in Appendix 2b and raw data in Appendix 4.3

I think the classification is reliable and for some reason, there are such words present in the textbook. There are words connected to grammar, classroom topic or countries that are needed to learn. On the other hand, the words such as *spymaster* or *palmtop* have a limited use and are of a very low frequency.

The data from the *BNC-COCA wordlists* provides an overview and tell us that a significant number of items is classified into the first four frequency levels. See the *Figures 5 and 6* for an overview.

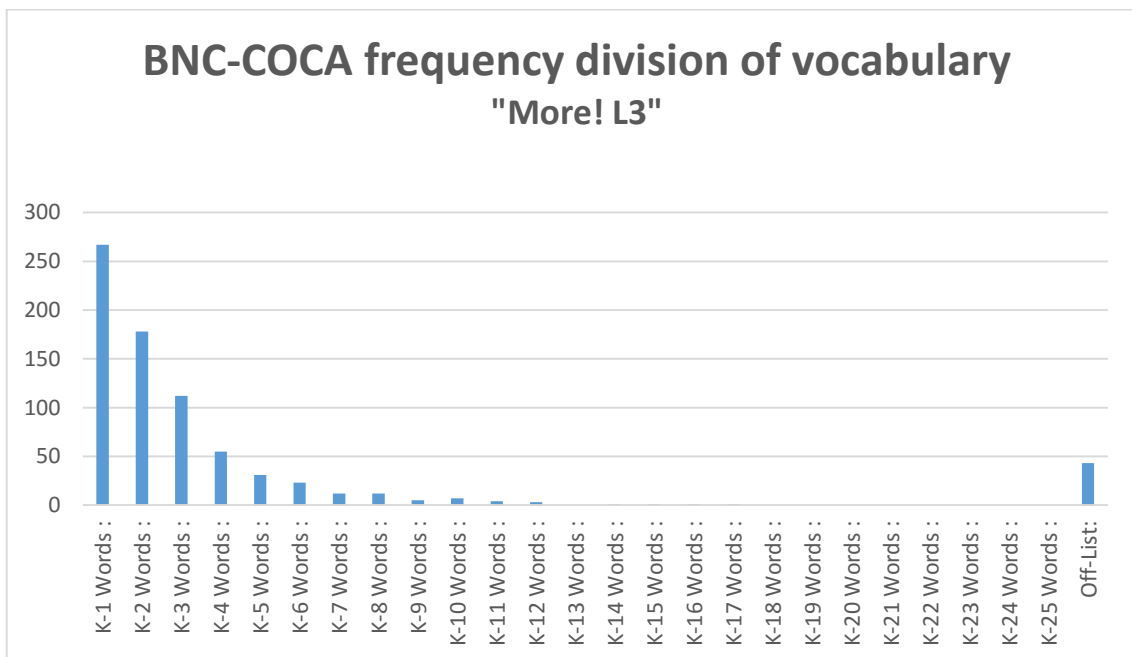


Figure 6 - The BNC-COCA frequency level division of More! L3 wordlist

The wordlist proved to contain a great deal of vocabulary that is not high frequency. The use and topics won over the vocabulary frequency aspect in many cases of the wordlist vocabulary selection. From my point of view, the authors intended to provide the learners with appropriate topics of their interest. The learners will have to cope with some uncommon words. Nevertheless, it is also a way to go, I would be interesting to know whether the text provided in the textbook is an authentic English text or not.

3.2.4 English Plus 2²⁴

English Plus 2 workbook is composed of 809 items being the largest from the analysed wordlists. It is present by 46 % tokens in the first, by 18 % in the second and 9 % in the third *NGSL* frequency level. Altogether, it makes **about 73 % that stands for the high-frequency vocabulary**. The number of 234 tokens (23 %) is classified out of the *NGSL* frequency levels. As for the *NAWL*, it is present in 4 % of the tokens.²⁵

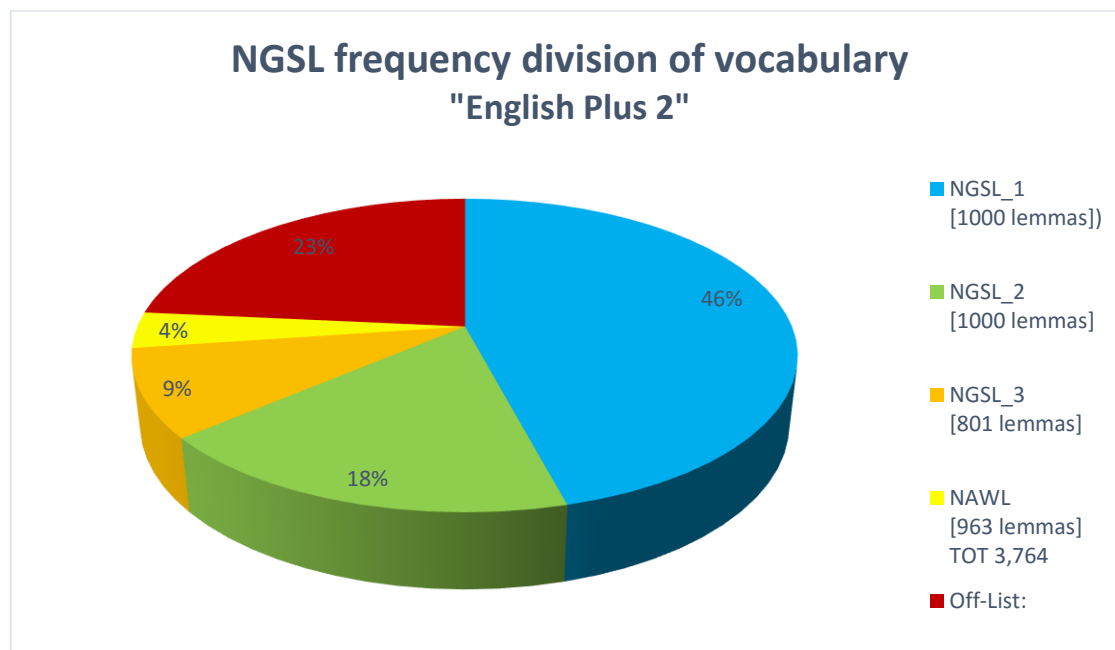


Figure 7 - The *NGSL* frequency level division of *English Plus 2* wordlist

It is a surprise for me that this wordlist contains so many tokens that are covered by the first *NGSL* frequency level. I can explain this fact only by the presence of whole phrases on the list.

²⁴ All the section refers to Hardy-Gould et al. (2011), data is the output of *Compleat Web VP* (Cobb, 2014)

²⁵ For more details see Table 18 in Appendix 2b and raw data in Appendix 4.4

Freq. Level	Example items
NGSL_1	a, about, agree, century, food, history, imagine, man, save, traditional
NGSL_2	actor, atmosphere, bet, careful, enemy, journey, label, photo, youth, strange
NGSL_3	aggressive, ceremony, joke, nose, sponsor, tournament, nose, tooth, twin, weird
NAWL	approximate, diary, goat, cheer, chess, memorize, noisy, poster, ridiculous, solar
Off-List:	aspirin, disabled, exotic, internet, kayak, karate, microwave, millenium, surfer, trophy

Table 11 - Ten example items from English Plus 2 for each NGSL level

In the graphs, we can see that most of the tokens belong to the first three frequency bands. In the off-list items there are, for example *microwave*, *millennium* and *trophy* what was surprising for me was that *internet* was also one of them. Overall, the high frequency vocabulary is involved, but there are of course some exceptions.

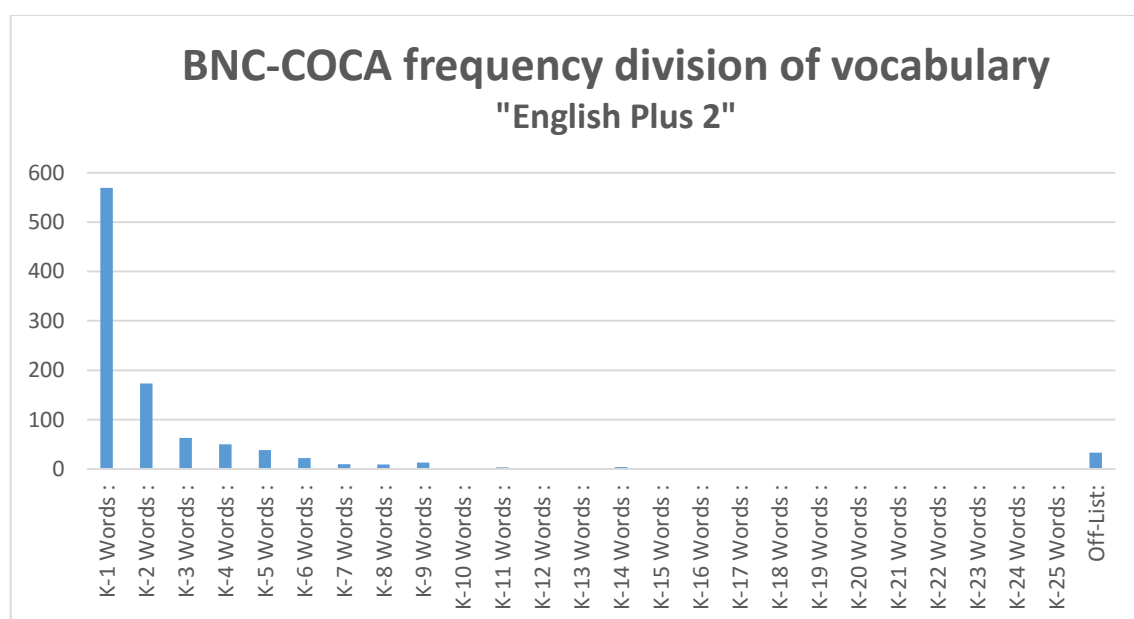


Figure 8 - The BNC-COCA frequency level division of English Plus 2 wordlist

When I compare the *BNC-COCA list* and the *NGSL coverage*, it only confirms the fact that the three first levels, which stand for the high-frequency vocabulary are an essential part of the wordlist vocabulary.

3.2.5 Messages 3²⁶

There are 407 items in the *Messages 3* wordlist, which classified as 836 tokens. Most of the tokens (65 %) are in the first *NGSL* frequency level. The other levels have very low coverage of tokens – 10 % for the second and 7 % for the third *NGSL* frequency level, though the *NGSL* itself covers **82% of all the tokens**. Academic vocabulary is just 2% and the off-list tokens carry the percentage of about 16 % out of the total number of tokens.²⁷

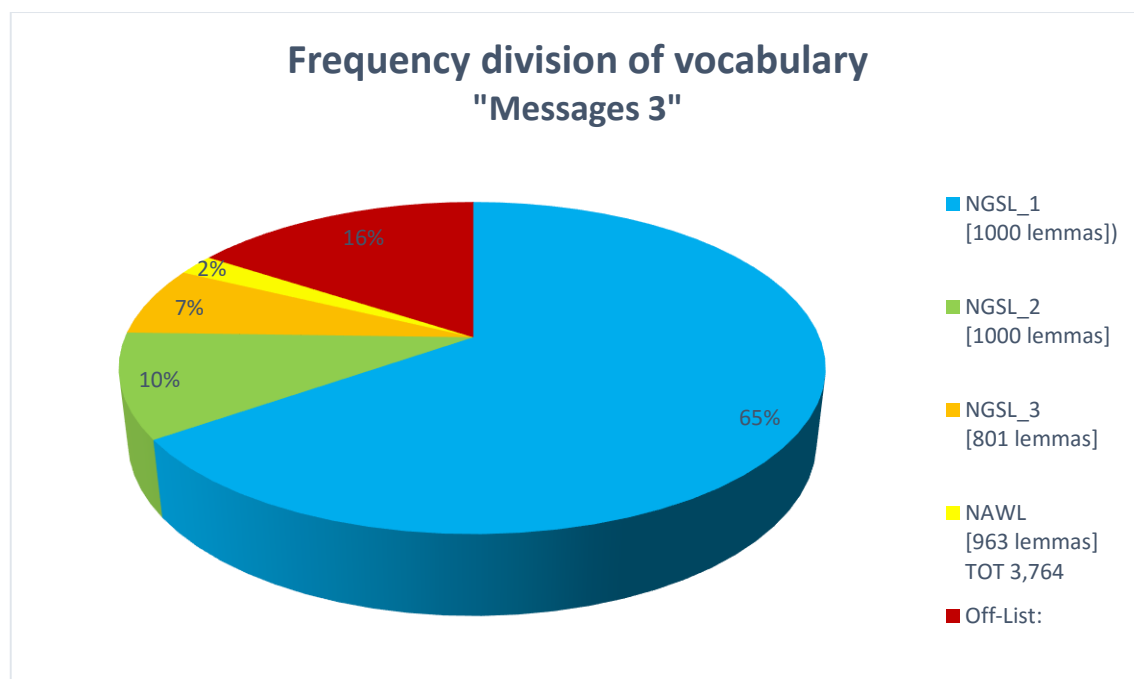


Figure 9 - The *NGSL* frequency level division of *Messages 3* wordlist

As we can see in the figures, a major number of tokens is covered by the *NGSL*. High-frequency vocabulary is clearly present in the textbook's wordlist. The first frequency counts so many tokens because there are many phrases on the list. I assume the vocabulary frequency aspect influenced the vocabulary selection of the list.

²⁶ All the section refers to Goodey et al. (2006), data is the output of *Compleat Web VP* (Cobb, 2014)

²⁷ For more details see Table 19 in Appendix 2c and raw data in Appendix 4.5

Freq. Level	Example items
NGSL_1	about, argue, attack, character, fine, give, heart, number, relationship, that
NGSL_2	abroad, afraid, award, friendly, occasion, sick, reply, theater, traffic, wash
NGSL_3	admire, awful, castle, custom, knife, salt, silver, stomach, survival, throat
NAWL	arrow, bracket, click, comma, injection, invert, nicely, plug, realistic, robot
Off-List:	ache, ambitious, funeral, homesick, Japanese, Poland, skateboarding, tracksuit, trainers, zorbing

Table 12 - Ten example items from Messages 3 for each NGSL level

Even the *BNC-COCA wordlists* confirm these data. The frequency levels from K-6 up to K-25 have covered no tokens. The *Messages 3* wordlist must have been composed with the frequency aspect in mind otherwise I believe it could not be so notably focused on the first frequency levels.

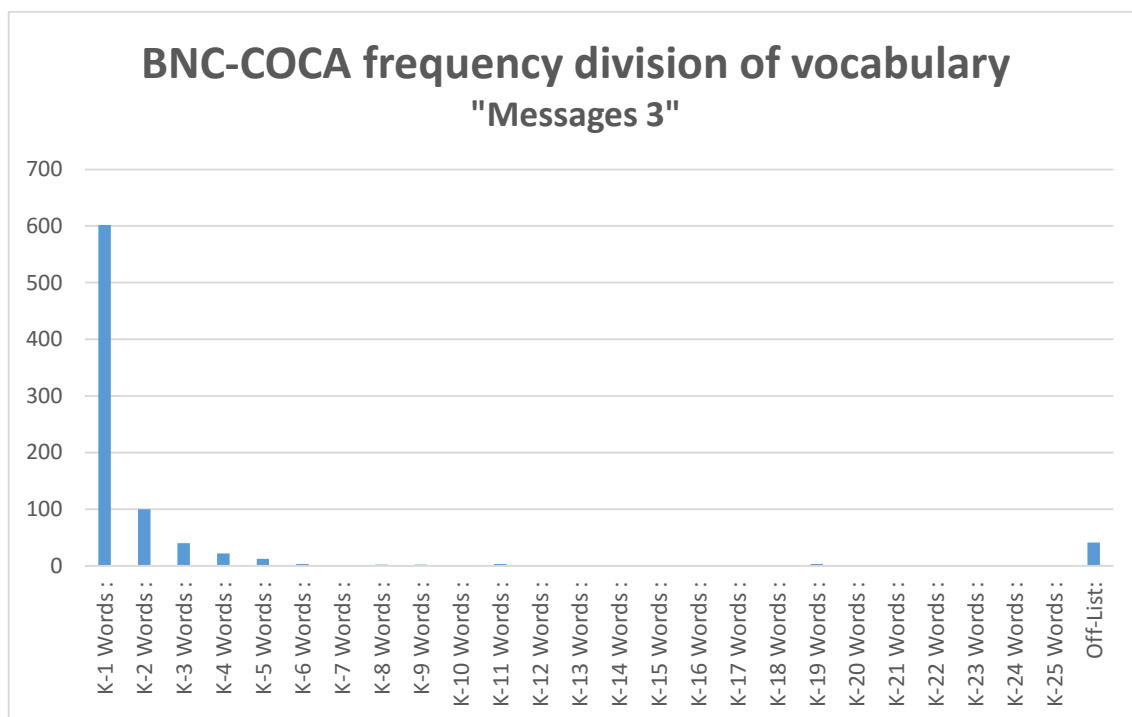


Figure 10 - The BNC-COCA frequency level division of Messages L3 wordlist

I will explain the great distribution of words in the first *NGSL* and *BNC-COCA* frequency level by the presence of phrases in the wordlist. It also explains why there were 407 items considered as 836 tokens. We can also notice a very low number of the tokens beyond the first two frequency bands and no token beyond the K-5 *BNC-COCA* frequency level.

The off-list items also count as just a little number. Plenty of these tokens are related to sports (*zorb*, *swimsuit*, *tracksuit*, *snowboarding*, *bungee*, *canoeing*) and others are related to nationality and travelling (*adventurous*, *American*, *Argentina*, *Britain*) the proper nouns belong to this category automatically because of the *NGSL* selection criteria. For example, the off-list item “*spear*” has a limited meaning and belongs to the mid-frequency vocabulary.

3.2.6 Angličtina 7 Way to Win²⁸

The textbook wordlist is composed of 361 items and were transferred into 535 tokens by the vocabulary profiler. The majority of the tokens is covered by the *NGSL* – 54 % tokens are in the first, 20 % in the second and 8 % in the third frequency band - that makes **82% coverage** in these high frequency levels. A small portion of these tokens is for the academic wordlist, it makes about 2 %. Only 16 % of the total number of tokens are out of the *NGSL*.²⁹

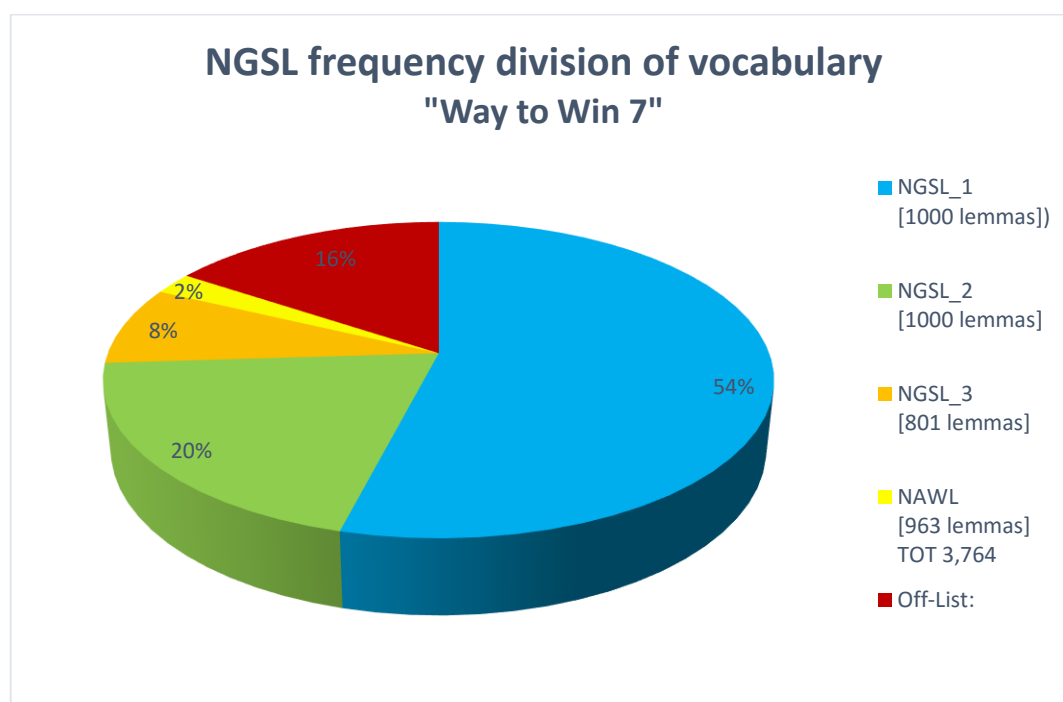


Figure 11 - The *NGSL* frequency level division of *Angličtina 7 Way to Win* wordlist

²⁸ All section refers to Betáková & Dvořáková (2006), data is the output of *Compleat Web VP* (Cobb, 2014)

²⁹ For more details see Table 20 in Appendix 2c and raw data in Appendix 4.6

The great distribution of words in the first *NGSL* and *BNC-COCA* frequency level is caused by the phrases present in the wordlist. The *NGSL* covers 82 % of this vocabulary, which is very close to its coverage of a general English text.

Freq. Level	Example items
NGSL_1	about, all, almost, choose, end, environment, inside, think, science, water
NGSL_2	angry, bike, burn, destroy, engine, glass, lesson, lock, plate, winner
NGSL_3	adventure, bath, fortune, giant, pen, pollution, rescue, sink, web, whisper
NAWL	acid, biology, cheat, diagram, harvest, horizon, poster, quiz, shuttle, solar
Off-List:	anorak, clockwise, dice, dock, Indian, hurricane, pavement, recycle, skates, tuna

Table 13 - Ten example items from *Angličtina 7 Way to Win* for each *NGSL* level

The frequency division shows high-frequency wordlist focus in both the *NGSL* and the *BNC-COCA* wordlists. Two first frequencies stand out in both of the graphs. I believe, it is caused by the large presence of phrases and collocations among the wordlist items. Other frequency levels do not play any significant role. This is also valid for the off-list items.

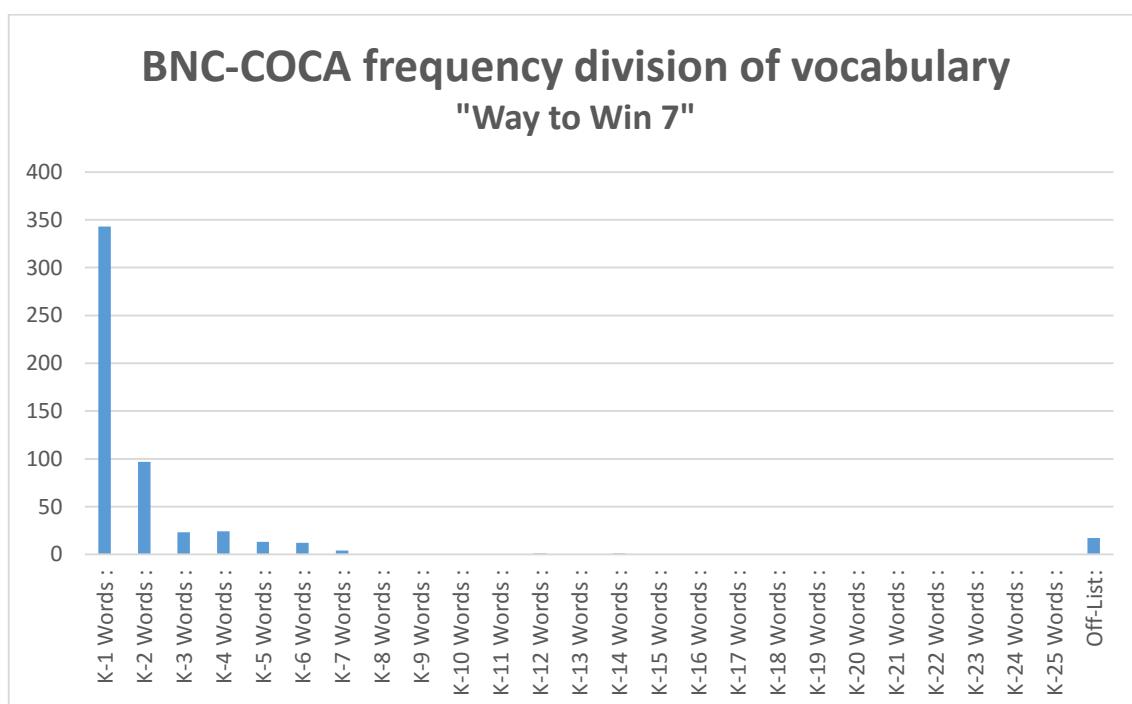


Figure 12 - The *BNC-COCA* frequency level division of *Angličtina 7 Way to Win* wordlist

In this case, I did not find any proper nouns that would be excluded from the wordlist selection except for “*Indian*”. I understand the off-list items are the words that are not common enough to appear on the list (for example *lifeboat*, *lifejacket* and *spacesuit*). For a textbook made by Czech authors, it did pretty well, I did not expect so high focus on the first two frequency levels.

3.3 Research summary

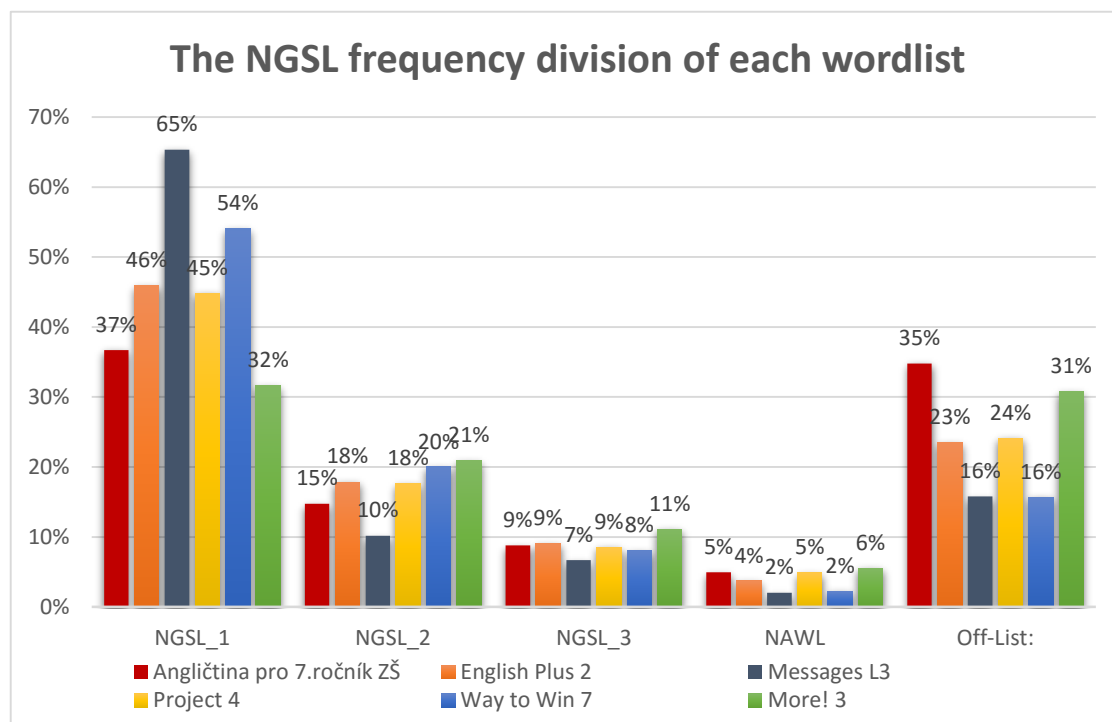


Figure 13 – Comparison of the NGSL frequency division of each wordlist

As we can see in the *Figure 12*, a major number of tokens is covered by the *NGSL*. High-frequency vocabulary is clearly present in each textbook wordlist. From the vocabulary present in the wordlists 60 - 82 % of the tokens is covered by the *NGSL* frequency levels. If the first frequency level accounts notably more tokens, there are usually present phrases on the wordlist. These phrases contain the core vocabulary, which raises the coverage percentage. I assume the vocabulary frequency aspect influenced the vocabulary selection of the vocabulary for the lists, the authors must have been thinking what to include and what is useless. The use and variety of meanings also increase the frequency rate at the words so the frequency aspect can be taken as an effect of the vocabulary need and use.

When we compare the wordlists at each *NGSL* level, we can see very few differences in the third *NGSL* and *NAWL* frequency level. The least difference is in the third *NGSL* frequency level. What makes the difference is the first *NGSL* frequency level and the off-list items.

I would like to point out some facts. It showed that the *Messages L3* wordlist uses the vocabulary frequency aspect more than the rest of the wordlists. I was surprised when I saw this comparison. *Angličtina 7 Way to Win* wordlist has got so many tokens in the first frequency levels even if it is a Czech textbook.

Both *Project 4 the Third Edition* and *English Plus 2* were published by Oxford University Press, therefore, these wordlists have very close percentage of tokens in all the levels. The off-list items can also stand for the high-frequency words but were excluded from the *NGSL* selection because they are proper nouns. However, there is a great deal of words that are not used frequently.

The first two *NGSL* frequency levels contained a lot of prepositions the verbs of general use – “*to be, do, go, have*” and others. This was expected and mentioned in the theory part of the thesis.

Generally, the off-list items were made up by words concentrating on a specific area. I noticed that there were a lot of words related to sports, illnesses and criminality. Plenty of these tokens are related to sports (*zorbining, swimsuit, tracksuit, snowboarding, bungee, canoeing*) and are the area of interest of the young learners so the authors used the situation to encourage the learners to provide words they may be interested in.

Many other words are related to nationality and travelling (*adventurous, American, Argentina, Britain*). No wonder, travelling is an evergreen amongst the topics and it is dealt with in any language book, the main purpose of a language is to communicate and travelling is the greatest opportunity and also one of the reasons of learning a foreign language. The proper nouns belong to the off-list words automatically because of the *NGSL* word selection criteria.

3.4 Methodology discussion

The selection of the textbook wordlists was reasonable, selected target language level agrees with the theory and I do not find any reason to argue the selection criteria.

The vocabulary profiler *Compleat Web VP!* (Cobb, 2014) showed to perform as expected. On the one hand, the results tell us general information about the tokens' coverage by the *NGLS* and *BNC-COCA wordlist*. On the other hand, the detailed information on tokens' distribution in the *NGSL* and *BNC-COCA* frequency levels is also available.

Selection of the *NGSL* and *BNC-COCA wordlists* for reference was also a good step. I was sure I did not make any mistake because I knew I could check and compare the two independent results with each other and I did.

I consider my research methodology valid, as the results are adequate and more or less confirm the theory and my expectations.

3.5 Research conclusion

As can be seen from the results, the high-frequency vocabulary is present amongst the wordlists. I assume the target level of A2 should be covered by the *NGSL* because the threshold level according to CEFR is the B1 target level. This is the point where the new possibilities of learning are open. According to the Schmitts (2014, p. 492), this threshold could correspond with the knowledge of the first 3,000 word families. There must have been other reasons to include uncommon words, and I would like to lead a discussion about these areas below.

Note that we would not like the textbooks to be put together strictly based on vocabulary frequency. We understand that there are other aspects, which have a big impact on the vocabulary learning process. At the same time, we see the advantages of vocabulary frequency during examination and assessment and believe that the teachers should be aware of the word frequency at least for this purpose. In addition, the frequency aspect is useful for analysing the language level of the learning materials and can be used as a reference point for a text level analysis and possible adjustment of the text.

Frequency is not a coincidental vocabulary aspect, it shows the importance of grammar structures – the determiners or prepositions in the English language, and therefore, it is advisable to use these relations in practice.

I would like to make use of the information from previous parts of the work and the research data as well. It appears there are some aspects other than the vocabulary frequency, which play a role in the textbook wordlist word selection when composing a textbook and other lesson focused materials. I would like to mention some of them.

Appropriateness should be considered first when preparing a lesson. Each lesson should be prepared for the age group of the learners. For each age, there are specific needs and the learning materials can be adjusted to the appropriate level. The choice of the materials is essential for the learner's interest and motivation. It is very effective to teach the vocabulary that the learners use in their first language.

Motivation in the lesson plays a big part for the learners in terms of their attitude to the subject they are taught and their involvement. Motivation initiates an interest. Therefore, the vocabulary selection for specific topics of interest should be preferred to the frequency aspect. Compromises should be made whilst choosing a topic, the frequency of the items can be a part of the vocabulary selection. In some situations, the vocabulary frequency can motivate the learners to learn.

There is an example of the list *Oxford 3,000TM* [OUP, 2010] use with the learners as mentioned in the section with materials. They simply fill in the list, watch and follow their progress. This means they have clear aims, which will be supportive. (Candel, 2013)

Every teacher needs to know and do some **time management**. The vocabulary is learnt best in context, so the learner can get the meaning, concept of the word and at the same time the pronunciation and use. The timing of the lesson should allow for a revision at the end of every lesson. In addition, the concentration and attention change during the lesson, therefore, it is better to alternate the lesson plan activities. The younger the learners are, the more often the activities should change.

Another big topic of teachers is the lesson **volume**. It is logical the learners cannot learn too many words at once. Human memory is able to remember a limited number of items. According to the learners' age, lesson time and vocabulary difficulty, approximately 10 items per a lesson can be taught. It is not possible to learn all the vocabulary dimensions at once. The teachers need to keep in mind that if the vocabulary is presented just once, the learning is very ineffective. It needs to be recycled.

In practice, this is very hard to do and further research could be made in terms of vocabulary recycling in English language textbooks. The teachers can show the learners the learning possibilities and give more options of vocabulary learning through all the language skills.

The aim of this research has been reached and I can say that the vocabulary present in the wordlists in the textbooks of English used among Czech learners of English is the high frequency vocabulary in large part. The core vocabulary was present in the wordlists even if there were no sentences and grammatical structures, this proves it is essential for the learners and the composition of textbook wordlists is reasonable.

4 Conclusion

I think I reached my thesis aims although many new questions appeared during writing the piece of work. I found out that focusing on a particular topic in a language (the vocabulary frequency aspect in this case) is not as easy as I expected. The topic is closely interrelated to vocabulary and corpus language studies. I had to sort out so much information just to figure out how the frequency aspect works and still, when I continue reading, I find new interesting pieces of information. I wish to deepen my knowledge of vocabulary frequency use in EFL teaching further.

My first hypothesis has been confirmed because the theory shows the English core language corresponds with the high frequency words in the general service lists. Furthermore, the theory and the research parts of the work agree that there is a core language vocabulary. Even among the wordlists that were the subject of my research, there is at least about 60 % of the tokens, which belongs to the high-frequency vocabulary. This confirms my second hypothesis. It would be interesting to investigate this area further, carry out text level analysis of the entire English textbooks and compare it with my research results.

For the third hypothesis, I cannot state that it is necessary to learn the core vocabulary of the English language at first, although it is beneficial for the learners and the authors of English textbooks already consider and more or less apply the vocabulary frequency aspect when composing a language textbook. Nevertheless, it cannot be easily checked if they do so intentionally. If they provided only the vocabulary of general authentic texts, the *NGSL* could cover between around 90 % of the vocabulary³⁰ especially of the function words.

My hypothesis number four “*Through comparing Czech and English core language, EFL materials can be adjusted for Czech learners of English.*” stays unanswered because it could be a topic of another piece of work and a subject of further research. However, the results could be very interesting for both vocabulary and grammar teaching purposes.

³⁰ These numbers are my deduction based on the information about the *NGSL*'s general text coverage, which is usually around 90%.

The main aim of the thesis was to cover the question: “*Does vocabulary frequency impact English teaching and learning?*”. I found out that the vocabulary frequency is an important aspect of vocabulary and it is present in any text whether we like or not. It proved in both the theory and research part of the thesis that the high-frequency vocabulary is essential for language learning and teaching. Therefore, the question should rather be: “*How much should the teachers pay attention to vocabulary frequency aspect when teaching English?*”. Another research could be done in the field of the vocabulary frequency in the lessons of English.

Lastly, I would like to say that teaching is so complex activity that there is no time to focus on using all the vocabulary dimensions in the lessons. The benefits of vocabulary frequency aspect are valuable especially, when estimating the language level of a text, preparing a language material and concerning learners’ work or performance assessment.

List of abbreviations

AWL – Academic Word List

BNC – British National Corpus

CEC – Cambridge English Corpus

CEFR – Common European Framework of Reference for Languages

CLIL – Content and Language Integrated Learning

COCA – Corpus of Contemporary American English

CUP – Cambridge University Press

EFL – English as a Foreign English

ELT – English Language Teaching

ESL – English as a Second Language

GSL – A General Service List of English Words (*West, 1953*)

L1 – Mother tongue / First language

L2 – Second language

LD – Learner's dictionary

MLD – Monolingual dictionary

NAWL – New Academic Word List

new-GSL – New General Service List (*Březina & Gablasová, 2013*)

NGSL – New General Service List (*Browne et al., 2013*)

OUP – Oxford University Press

RVP ZV – Rámcový vzdělávací program pro základní vzdělávání

ŠVP – Školní vzdělávací program / School Education Programme

VP – Vocabulary profiler

List of sources

Bibliography

BETÁKOVÁ, L. & DVOŘÁKOVÁ, K., 2006. *Angličtina 7 Way to Win*. Fraus. ISBN 80-7238-389-2 EAN 9788072383894 (104p).

CEFR, 2001. *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Council of Europe, Cambridge University Press.

ČERMÁK, F. & KŘEN, M. *Frekvenční slovník češtiny*. Nakladatelství Lidové noviny, Praha 2004. ISBN 80-7106-676-1.

GAIRNS, R. & REDMAN, S., 1986. *Working with Words (A guide to teaching and learning vocabulary)*. Cambridge: Cambridge University Press. ISBN 0 521 31709 6.

GOODEY, D. & GOODEY, N. & CRAVEN, M., 2006. *Messages 3 Student's Book*. Cambridge University Press. ISBN13 9780521614337 (144p).

HARDY-GOULD, J. & MELLERSH, K., 2011. *English Plus 2 Workbook with MultiROM (Czech Edition)*. Oxford University Press. ISBN: 978-0-19-47901-5, EAN: 9780194749015 (112p).

HUTCHINSON, T., 2009. *Project 4 the Third Edition Workbook (Czech Version)*. Oxford University Press, 2009. ISBN: 978-0-947636-1-5 EAN: 9780194763615 (79p).

LESSARD-CLOUSTON, M., 2013. *Teaching Vocabulary*. TESOL International Association. ISBN – 9781931185974 (pp. 1-46).

MCCARTEN, J., 2007. *Teaching Vocabulary*. Cambridge: Cambridge University Press. ISBN-13 978-0-521-94325-3.

NATION, P., 2013. *What Should Every EFL Teacher Know?* Compass Publishing (March 1, 2013) ISBN-13: 978-1599662664 (pp. 1-238).

OUP, 2010. *Oxford Studijní slovník + CD-ROM - česká edice výkladový slovník angličtiny s českým překladem*. Oxford University Press. ISBN: 9780194306553.

PRŮCHA, J. & WALTEROVÁ, E. & MAREŠ, J., 2003. *Pedagogický slovník*. 4. Aktualizované vydání. Praha: Portál. ISBN 80-7178-772-8.

PUCHTA, H. et al., 2008. *More! Level 3 Student's Book with interactive CD-ROM*. Cambridge University Press. ISBN: 9780521713078 EAN: 9780521713078.

RUNDELL, M., 1998. *Recent trends in English pedagogical lexicography*. Oxford: Oxford University Press, *International Journal of Lexicography*, Vol 11 No. 4.[cit. 2015-04-14] (pp. 315-342).

RVP ZV, 2013. *Rámcový vzdělávací program pro základní vzdělávání*. Praha: Ministerstvo školství, mládeže a tělovýchovy.

SMITH, N. & ALLOTT, N. & ALLOT, N. 2016. *Chomsky: Ideas and Ideals*. Cambridge University Press, 3rd Edition. ISBN-9781107082144 (pp. 474).

WILLIS, D., 1990. *Lexical syllabus*. [e-book] London: Collins ELT. The University of Birmingham. [cit. 2016-01-20] Dostupné také z: <<http://www.birmingham.ac.uk/schools/edacs/departments/englishlanguage/research/research/sources/lexical-syllabus.aspx>>.

ZAHÁLKOVÁ, M., 2010. *Hello Kids! Angličtina pro 7. ročník základní školy - pracovní sešit*. SPN, 2010. ISBN: 9788072354856 (80p).

Online and other sources

ALBER, R., 2014. *Doing It Differently: Tips for Teaching Vocabulary*. [online] [cit. 2016-02-25] Dostupné z: <<http://www.edutopia.org/blog/vocabulary-instruction-teaching-tips-rebecca-alber>>.

ALHARBI, M., A., 2015. *Building Vocabulary for Language Learning: Approach for ESL Learners to Study New Vocabulary*. University of Memphis (USA) © Journal of International Students Volume 5, Issue 4 (2015), ISSN: 2162-3104 Print/ ISSN: 2166-3750 Dostupné z <<http://jistudents.org/>> (pp. 501-511).

ALQAHTANI, M., 2015. *The importance of vocabulary in language learning and how to be taught*. International Journal of Teaching and Education, 2015, 3.3. (pp. 21-34).

Amazon, 2016. *Reference dictionaries thesauruses eBooks Kindle..* [online search] [cit. 2016-02-22] Dostupné z: <<http://www.amazon.com/Reference-Dictionaries-Thesauruses-eBooks-Kindle/b?ie=UTF8&node=2224239011>>.

ANTHONY, L., 2014. *AntWordProfiler (Version 1.4.1)* [Computer Software]. Tokyo, Japan: Waseda University. [cit. 2016-01-05] Dostupné z: <<http://www.laurenceanthony.net>>.

BAUMAN, J., [1995]. *About the General Service List*. [online] Enterprise Training Group. [cit. 2015-06-22]. <<http://jbauman.com/aboutgsl.html>>.

BBC, 2016. *Learning English*. [online] [cit. 2016-02-23] Dostupné z: <<http://www.bbc.co.uk/learningenglish/english/>>.

BŘEZINA, V. & GABLASOVÁ, D., 2013. *Is There a Core General Vocabulary? Introducing the New General Service List*. [online] Applied Linguistics, 2013-08-25. [cit. 2015-04-05] Dostupné z: <<http://applied.oxfordjournals.org/content/early/2013/08/25/applin.amt018.abstract>> (pp. 1-22).

BŘEZINA, V., 2014. *New-GSL portal*. [online] [cit. 2016-02-22] Dostupné z: <<http://corpora.lancs.ac.uk/vocab/index.php>>.

BROWNE, C., 2014. *The New General Service List Version 1.01: Getting Better All the Time*. [online] Korea TESOL Journal Vol. 11, No. 1, 2014. [cit. 2015-06-22] Dostupné z: <<http://www.newgeneralservicelist.org/s/KOTESOL2014-344f.pdf>>.

BROWNE, C., CULLIGAN, B. & PHILLIPS, J., 2013a. *The New General Service List*. [online] [cit. 2016-01-22] Dostupné z: <<http://www.newgeneralservicelist.org>>.

- BROWNE, C., CULLIGAN, B. & PHILLIPS, J., 2013b. *The New Academic Wordlist*. [online] [cit. 2016-01-22] Dostupné z: <<http://www.newacademicwordlist.org>>.
- CANDEL, B., 2013. *Who is the Oxford 3,000™ actually for?* Oxford University Press. [online] [cit. 2016-03-24] Dostupné z: <<http://oupeltglobalblog.com/2013/09/23/who-is-the-oxford-3000-actually-for/>>.
- CECL, 2016. *Learner corpora around the world*. [online] [cit. 2016-02-10] Dostupné z: <<http://www.uclouvain.be/en-cecl-lcworld.html>>.
- CINKE, J., 2015. *Učebnice anglického jazyka pro 2. stupeň základní školy - analýza a srovnání* [online]. Hradec Králové, 2015 [cit. 2016-01-23]. Diplomová práce. Univerzita Hradec Králové, Pedagogická fakulta. Vedoucí práce Mgr. Olga Vraštilová, M.A., Ph.D. Dostupné z: <<http://theses.cz/id/w19mqo/>>.
- COBB, T., 2015. *VocabProfile Home*. [online] [cit. 2016-02-20] Updated 2015 Nov 30 Dostupné z: <<http://www.lexutor.ca/vp/>>.
- COXHEAD, A., 2000. *A New Academic Wordlist*. TESOL Quarterly, 2000, 34(2). (pp. 213-238).
- CUP, 2015. *English Profile*. [online] [cit. 2016-01-20] Dostupné z: <<http://www.englishprofile.org/>>.
- CVRČEK, V., 2011. *How Large is the Core Language?* [online] Corpus Linguistics, Birmingham. [cit. 2015-04-05] Dostupné z: <<http://www.birmingham.ac.uk/Documents/college-artslaw/corpus/conference-archives/2011/Paper-145.pdf>> (pp. 1-23).
- DAVIES, M., 2004-. *BYU-BNC. (Based on the British National Corpus from Oxford University Press)*. [online] [cit. 2016-03-23] Dostupné z: <<http://corpus.byu.edu/bnc/>>.
- DAVIES, M., 2008-. *The Corpus of Contemporary American English: 520 million words, 1990-present*. [online] [cit. 2016-02-23] Dostupné z: <<http://corpus.byu.edu/coca/>>.
- Duolingo*. [software] 2016. [cit. 2016-02-23] Dostupné z: <<http://www.duolingo.com>>.
- EL-HMOUDOVA, D. & MILKOVA, E., 2015. *Variations and Frequencies in Learning Styles in a Group of Czech English as Foreign Language Learners*. [online] Procedia - Social and Behavioral Sciences, Volume 182, 13 May 2015, ISSN 1877-0428 [cit. 2016-01-20] Dostupné z: ScienceDirect (pp. 60-66).
- englishbooks.cz, 2016a. *Essential dictionary*. [online search] [cit. 2016-03-12] <<http://www.englishbooks.cz/vyhledavani.html?q=essential+dictionary>>.
- englishbooks.cz, 2016b. *Zjednodušená četba*. [online] [cit. 2016-02-26] Dostupné z URL: <<http://www.englishbooks.cz/zjednodusena-cetba/>>.
- FEINBERG, J. 2014. *Wordle*. [online] [cit. 2016-02-20] Dostupné z: <<http://www.wordle.net>>.

- GARNER, J. (2016) *CEFR Vocabulary Profiler*. [online] VocabKitchen, 2016. [cit. 2016-03-20] Dostupné z: <<http://vocabkitchen.com/profiler/cefr>>.
- Ghent University, [2016]. *Test of vocabulary size* [online] [cit. 2016-02-20] Dostupné z: <<http://vocabulary.ugent.be/>>.
- GILNER, L., 2011. *A primer on the General Service List*. Reading in a Foreign Language, April 2011, Volume 23, No. 1. ISSN 1539-0578 (pp. 65-83).
- IMP Dictionary and thesaurus*. [online] [cit. 2016-01-20] Dostupné z: <<https://www.learnthat.org/dictionary/>>.
- JARĀB, J., 2016. *Neoficiální setkání UP host s prof. Josefem Jarábem*. [účast 2016-02-16] Olomouc. [cit. 2016-02-25] Dostupné z: <<https://www.mixcloud.com/UPAIROlomouc14/up-host-s-prof-josefem-ja%C5%99abem/>> [0:43:10 – 0:50:00].
- KEDDIE, J., 2016. *Introduction to corpora*. [online] [cit. 2016-02-25] Dostupné z: <<http://www.onestopenenglish.com/skills/vocabulary/corpora/introduction-to-corpora/155104.article>>.
- KLIMOVA, B. F., 2012. *CLIL and the teaching of foreign languages*. Procedia-Social and Behavioral Sciences, 2012, 47: 572-576.
- KRISHNAMURTHY, R., 2008. *Corpus-driven lexicography*. Aston University, Birmingham, UK. International Journal of Lexicography, Vol. 21 No. 3. Advance access publication 5 July 2008 Oxford University Press. doi:10.1093/ijl/ecn028 (pp.231-242).
- LEBEDOVÁ, M., 2011. *Practising vocabulary through game-based activities* [online]. Brno. [cit. 2016-01-20] Bakalářská práce. Masarykova univerzita, Pedagogická fakulta. Vedoucí práce doc. Mgr. Světlana Hanušová, Ph.D. Dostupné z: <<http://theses.cz/id/yhkfd5>>.
- Lexical Computing, 2016. *Text Corpora*. [Online list] [cit. 2016-04-12] Dostupné z: <<https://www.sketchengine.co.uk/corpora/>>.
- LIPKA, L., 1992. *An Outline of English Lexicology: Lexical Structure, Word Semantics, and Word-Formation*. Max Niemeyer Verlag Tübingen, Second Edition. ISBN 3-484-41003-5 ISSN 0178-7861 (pp.218).
- Macmillan Publishers, 2016. *Language Games*. [online] [cit. 2016-02-23] Dostupné z: <<http://www.macmillandictionary.com/language-games/>>.
- Memrise*. [software] Memrise Ltd, 2016. [cit. 2016-02-23] Dostupné z: <<http://www.memrise.com>>.
- MŠMT, 2013. *Schvalovací doložky učebnic*. [online] [cit. 2016-01-20] Dostupné z: <<http://www.msmt.cz/vzdelavani/skolstvi-v-cr/schvalovaci-dolozky-ucebnic-2013>>
- NATION, P., [2016]. *Nation's webpage*. [online] [cit. 2016-02-20] Dostupné z: <<http://www.victoria.ac.nz/lals/about/staff/paul-nation>>

- NATION, P., 2012. *The BNC/COCA word family lists*. [online] [cit. 2016-02-20] Dostupné z: <http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Information-on-the-BNC_COCA-word-family-lists.pdf>
- OKAMOTO, M., 2015. *Is corpus word frequency a good yardstick for selecting words to teach? Threshold levels for vocabulary selection*. [online] System, Volume 51, July 2015, ISSN 0346-251X [cit. 2016-01-05] Dostupné z: <<http://www.sciencedirect.com/science/article/pii/S0346251X1500038X>> (pp. 1-10)
- OUP, 2016a. *How many words are there in the English language?* [online] Oxford Dictionaries. Oxford University Press. [cit. 2016-01-22]. Dostupné z: <<http://www.oxforddictionaries.com/words/how-many-words-are-there-in-the-english-language>>.
- OUP, 2016b. *Slovníčky*. [online]. [cit. 2016-01-22]. Dostupné na WWW: <https://elt.oup.com/general_content/cz/kestazeni/slovnicky?cc=cz&selLanguage=cs>
- OUP, 2016c. *RVP*. [online]. [cit. 2016-01-22]. Dostupné na <https://elt.oup.com/general_content/cz/kestazeni/rvp?cc=cz&selLanguage=cs>
- PŘICHYSTAL, Š., 2013. *QGame*. [online software] [cit. 2016-02-22] Brno: Masaryk University Brno in the Czech Republic. Dostupné z: <<https://nlp.fi.muni.cz/projekty/qgame/play-question-game/>>.
- Quizlet*. [software] 2016. [cit. 2016-02-23] Dostupné z: <<https://quizlet.com/>>.
- ROCA-VARELA, M. L., 2013. *High-frequency English Words in Spoken Learner Language: Actual, career and pretend as a Case in Point*. [online] Procedia-Social and Behavioral Sciences, 2013, 95. [cit. 2016-01-20] (pp 557-562) Dostupné z: *ScienceDirect*
- SCHMITT, N. & MCCARTHY, M., 1997. *Vocabulary: Description, acquisition and pedagogy*. Cambridge: Cambridge University Press.
- SCHMITT, N. & SCHMITT, D., 2014. *A reassessment of frequency and vocabulary size in L2 vocabulary teaching*. Cambridge: Cambridge University Press, Language Teaching, 2014, 47.04 doi:10.1017/S0261444812000018. (pp. 484-503)
- SEVT, 2016. *Angličtina – učebnice ZŠ - nejprodávanejší*. [online search] [cit. 2016-03-20] <<http://www.sevt.cz/obchod/knihkupectvi/cizi-jazyk/anglictina/ucebnice-zs/?order=bestsellers&71=Ucebnice>>.
- SHEN, Wei-Wei., 2003. *Current trends of vocabulary teaching and learning strategies for EFL settings*. Feng chia journal of Humanities and social sciences, 2003, 7. (pp. 187-224).
- ŠMAJDLEROVÁ, S., 2010. *Výuka a učení se anglické slovní zásoby*. [online]. Olomouc. [cit. 2016-02-21]. Diplomová práce. Univerzita Palackého v Olomouci, Pedagogická fakulta. Vedoucí práce Simon Gill Dostupné z: <<http://theses.cz/id/upkas2/>>.

testyourvocab.com, 2016. *Vocabulary size test*. [online] [cit. 2016-02-20] Dostupné z: <<http://testyourvocab.com/>>.

VÍT, M., 2010 – 2014. *Czenglish alert*. [online articles] [cit. 2016-02-15] Dostupné z: <<http://www.helpforenglish.cz/gramatika/ruzne/czenglish-alert>>.

Vocabulary Builder. [online] [cit. 2016-01-20] Dostupné z: <http://www.online-utility.org/english/vocabulary_builder.jsp>

WebLingua, 2014. *Text Inspector*. [online software] [cit. 2016-02-23] Dostupné z: <<http://textinspector.com/workflow>>.

WEST, M., 1953. *General service list of English words*. [online] [cit. 2015-06-22]. Dostupné z: <<http://www.academia.edu/4791005/Gsl-the-general-service-list-by-michael-west-1953>>

Word Family Framework. [2016]. [online] [cit. 2016-01-20] Dostupné z: <<https://www.learnenglish.org.uk/wff/index.html>>

Appendices

Appendix 1

Table 14 - Word family levels according to Bauer and Nation [1993] (Nation, 2012).. 77

Appendix 2a

Table 15 - Project 4 the Third Edition NGSL frequency division details 78

Table 16 - Hello Kids! Angličtina pro 7. ročník ZŠ NGSL frequency division details . 78

Appendix 2b

Table 17 - More! L3 NGSL frequency division details 79

Table 18 - English Plus 2 NGSL frequency division details 79

Appendix 2c

Table 19 – Messages 3 NGSL frequency division details 80

Table 20 – Angličtina 7 Way to Win NGSL frequency division details 80

Appendix 3 - Images

Image 1 - Screenshot of CEFR Vocabulary Profiler (Garner, 2016)..... 81

Image 2 - Screenshot of English Vocabulary Profile (CUP, 2015) 81

Image 3 - Screenshot of English Grammar Profile (CUP, 2015) 82

Appendix 4 – Vocabulary distribution details of each wordlist

Appendix 4.1 - Project 4 the Third Edition (Hutchinson, 2009)

Appendix 4.2 - Hello Kids! Angličtina pro 7. ročník ZŠ (Zahálková, 2010)

Appendix 4.3 - More! L3 (Puchta et al., 2008)

Appendix 4.4 - English Plus 2 (Hardy-Gould & Mellersh, 2011) (NGSL)

Appendix 4.5 - Messages 3 (Goodey et al., 2006)

Appendix 4.6 - Angličtina 7 Way to Win (Betáková & Dvořáková, 2006)

Appendix 5 - CD

Appendix 1 – Word family levels

Table 15 – Word family levels according to Bauer and Nation [1993] (Nation, 2012)

Level 1	A different form is a different word. Capitalization is ignored.
Level 2	Regularly inflected words are part of the same family. The inflectional categories are -plural; third person singular present tense; past tense; past participle; -ing; comparative; superlative; possessive.
Level 3	-able, -er, -ish, -less, -ly, -ness, -th, -y, non-, un-, all with restricted uses
Level 4	-al, -ation, -ess, -ful, -ism, -ist, -ity, -ize, -ment, -ous, in-, all with restricted uses
Level 5	-age (leakage), -al (arrival), -ally (idiotically), -an (American), -ance (clearance), -ant (consultant), -ary (revolutionary), -atory (confirmatory), -dom (kingdom; officialdom), -eer (black marketeer), -en (wooden), -en (widen), -ence (emergence), -ent (absorbent), -ery (bakery; trickery), -ese (Japanese; officialese), -esque (picturesque), -ette (usherette; roomette), -hood (childhood), -i (Israeli), -ian (phonetician; Johnsonian), -ite (Paisleyite; also chemical meaning), -let (coverlet), -ling (duckling), -ly (leisurely), -most (topmost), -ory (contradictory), -ship (studentship), -ward (homeward), -ways (crossways), -wise (endwise; discussion-wise), anti- (anti-inflation), ante- (anteroom), arch- (archbishop), bi- (biplane), circum- (circumnavigate), counter- (counter-attack), en- (encage; enslave), ex- (ex-president), fore- (forename), hyper- (hyperactive), inter- (inter-African, interweave), mid- (mid-week), mis- (misfit), neo- (neo-colonialism), post- (post-date), pro- (pro-British), semi- (semi-automatic), sub- (subclassify; subterranean), un- (untie; unburden)
Level 6	-able, -ee, -ic, -ify, -ion, -ist, -ition, -ive, -th, -y, pre-, re-

Table 14 - Word family levels according to Bauer and Nation [1993] (Nation, 2012)

Appendix 2 - NGSL frequency division overview

4.1.1 Appendix 2a

Tables 16, 17 - Textbook wordlist research NGSL frequency division overview

FREQ. LEVEL	LEMMAS (%)	TYPES (%)	TOKENS (%)	CUMUL. TOKEN %
NGSL_1 [1000 LEMMAS])	242 (49.29)	256 (35.61)	405 (44.85)	44.85
NGSL_2 [1000 LEMMAS]	136 (27.70)	138 (19.19)	159 (17.61)	62.46
NGSL_3 [801 LEMMAS]	72 (14.66)	72 (10.01)	77 (8.53)	70.99
NAWL [963 LEMMAS] TOT 3,764	41 (8.35)	41 (5.70)	44 (4.87)	75.86
OFF-LIST:	??	211 (29.35)	218 (24.14)	100.00
TOTAL (UNROUNDED)	491+?	719 (100)	903 (100)	≈100.00

Table 15 - Project 4 the Third Edition (Hutchinson, 2009) NGSL frequency division details

Freq. Level	Lemmas (%)	Types (%)	Tokens (%)	Cumul. token %
NGSL_1 [1000 lemmas])	185 (52.41)	190 (32.82)	229 (36.70)	36.70
NGSL_2 [1000 lemmas]	85 (24.08)	88 (15.20)	92 (14.74)	51.44
NGSL_3 [801 lemmas]	52 (14.73)	54 (9.33)	55 (8.81)	60.25
NAWL [963 lemmas] TOT 3,764	31 (8.78)	31 (5.35)	31 (4.97)	65.22
Off-List:	??	215 (37.13)	217 (34.78)	100.00
Total (unrounded)	353+?	579 (100)	624 (100)	≈100.00

Table 16 - Hello Kids! Angličtina pro 7. ročník ZŠ (Zahálková, 2010) NGSL frequency division details

4.1.2 Appendix 2b

Tables 18, 19 - Textbook wordlist research NGSL frequency division details

Freq. Level	Lemmas (%)	Types (%)	Tokens (%)	Cumul. token %
NGSL_1 [1000 lemmas]	197 (41.39)	200 (28.05)	239 (31.61)	31.61
NGSL_2 [1000 lemmas]	154 (32.35)	155 (21.74)	158 (20.90)	52.51
NGSL_3 [801 lemmas]	83 (17.44)	83 (11.64)	84 (11.11)	63.62
NAWL [963 lemmas]	42 (8.82)	42 (5.89)	42 (5.56)	69.18
TOT 3,764				
Off-List:	??	232 (32.54)	233 (30.82)	100.00
Total (unrounded)	476+?	713 (100)	756 (100)	≈100.00

Table 17 - *More! Level 3 (Puchta et al., 2008) NGSL frequency division details*

Freq. Level	Lemmas (%)	Types (%)	Tokens (%)	Cumul. token %
NGSL_1 [1000 lemmas]	292 (52.14)	317 (38.94)	457 (45.88)	45.88
NGSL_2 [1000 lemmas]	149 (26.61)	158 (19.41)	178 (17.87)	63.75
NGSL_3 [801 lemmas]	83 (14.82)	85 (10.44)	90 (9.04)	72.79
NAWL [963 lemmas]	36 (6.43)	36 (4.42)	37 (3.71)	76.50
TOT 3,764				
Off-List:	??	217 (26.66)	234 (23.49)	99.99
Total (unrounded)	560+?	814 (100)	996 (100)	≈100.00

Table 18 - *English Plus 2 (Hardy-Gould & Mellersh, 2011) NGSL frequency division details*

4.1.3 Appendix 2c

Tables 20, 21 - Textbook wordlist research NGSL frequency division details

Freq. Level	Lemmas (%)	Types (%)	Tokens (%)	Cumul. token %
NGSL_1 [1000 lemmas]	240 (62.99)	261 (48.69)	546 (65.31)	65.31
NGSL_2 [1000 lemmas]	75 (19.69)	77 (14.37)	85 (10.17)	75.48
NGSL_3 [801 lemmas]	51 (13.39)	52 (9.70)	56 (6.70)	82.18
NAWL [963 lemmas] TOT 3,764	15 (3.94)	16 (2.99)	17 (2.03)	84.21
Off-List:	??	129 (24.07)	132 (15.79)	100.00
Total (unrounded)	381+?	536 (100)	836 (100)	≈100.00

Table 19 – Messages 3 (Goodey et al., 2006) NGSL frequency division details


Freq. Level	Lemmas (%)	Types (%)	Tokens (%)	Cumul. token %
NGSL_1 [1000 lemmas]	151 (58.53)	158 (48.47)	289 (54.02)	54.02
NGSL_2 [1000 lemmas]	69 (26.74)	69 (21.17)	107 (20.00)	74.02
NGSL_3 [801 lemmas]	28 (10.85)	28 (8.59)	43 (8.04)	82.06
NAWL [963 lemmas] TOT 3,764	10 (3.88)	10 (3.07)	12 (2.24)	84.30
Off-List:	??	61 (18.71)	84 (15.70)	100.00
Total (unrounded)	258+?	326 (100)	535 (100)	≈100.00

Table 20 – Angličtina 7 Way to Win (Betáková & Dvořáková, 2006) NGSL frequency division details

Appendix 3 - Images

CEFR Vocabulary Profiler

Clear Show: **A1** **A2** **B1** **B2** **C1** **C2** **DICTIONARY**
ENABLED
Double Click on any word.

 | Stroop Mode:

Word count: 320

I can't recall when, but it was for the writing assessment.

And one of the concerns at that point is what happens if students have more time?

And so in that assessment, it was a comparison of 20 minutes and 15 minutes to see if in fact the additional time made a difference.

I'm mentioning this here not because it's writing, but people might have been concerned about how much time was given to the students to read and then write the extended passage.

And they found that with the additional time, it made almost no difference.

I can't recall when, but it was for the writing assessment.

And one of the concerns at that point is what happens if students have more time?

And so in that assessment, it was a comparison of 20 minutes and 15 minutes to see if in fact the additional time made a difference.

I'm mentioning this here not because it's writing, but people might have been concerned about how much time was given to the students to read and then write the extended passage.

And they found that with the additional time, it made almost no difference.

Image 1 - Screenshot of CEFR Vocabulary Profiler (Garner, 2016)

English Vocabulary Profile

British English American English


Choose level:

- A1
- A1-A2
- A1-B1
- A1-B2
- A1-C1
- A1-C2
- A2 only
- B1 only
- B2 only
- C1 only
- C2 only

Browse A-Z
OR
Enter a word or phrase

Hide culturally sensitive words **Search**

English Profile Home
About English Vocabulary Profile
Word of the Week
Help
Feedback

able  /'eɪ.bəl/

Word family:
Nouns: ability, disability, inability
Verbs: enable
Adjectives: able, disabled, unable

► **ADJECTIVE**

be able to do sth
A2 to have the ability to do something or the possibility of doing something
Dictionary examples:
He'll be able to help you.
Will she be able to do the work?
I'm sorry that I wasn't able to phone you yesterday.
It's so wonderful being able to see the sea from my window.

Learner example:
I will be able to go to your house next Friday.

CLEVER
C2 clever or good at doing something
Dictionary example:
She's a very able student.

Learner example:
They contend that this will lead to better grades, better results in key stage test[s], and a more able work force.

Image 2 - Screenshot of English Vocabulary Profile (CUP, 2015)

English Grammar Profile Online

Search... Level A1 A2 B1 B2 C1 C2

Results 1 - 20 of 1239 Sort by: SuperCategory ▾ Ascending ▾ 20 ▾

SuperCategory	SubCategory	Level	Can-do statement	Example	Details
ADJECTIVES	combining	A1	FORM: COMBINING TWO ADJECTIVES WITH 'AND' Can use 'and' to join a limited range of common adjectives.	<input type="button" value="Example"/>	<input type="button" value="Details"/>
ADJECTIVES	combining	A2	FORM: COMBINING TWO ADJECTIVES WITH 'BUT' Can use 'but' to join a limited range of common adjectives, after 'be'.	<input type="button" value="Example"/>	<input type="button" value="Details"/>
ADJECTIVES	combining	B1	FORM: BEFORE THE NOUN Can use a comma to combine two adjectives used before the noun, following the usual order of adjective types.	<input type="button" value="Example"/>	<input type="button" value="Details"/>
ADJECTIVES	combining	B1	FORM: COMBINING COMPARATIVE ADJECTIVES WITH 'AND' Can use 'and' to join a limited range of comparative adjectives ► adjectives: comparatives	<input type="button" value="Example"/>	<input type="button" value="Details"/>
ADJECTIVES	combining	B1	FORM: COMBINING MORE THAN TWO ADJECTIVES Can use commas and 'and' to join more than two adjectives, after 'be'.	<input type="button" value="Example"/>	<input type="button" value="Details"/>

Image 3 - Screenshot of English Grammar Profile (CUP, 2015)

Appendix 4 - Vocabulary distribution details

Appendix 4.1a - Project 4 the Third Edition (Hutchinson, 2009) (NGSL)

Lemmas List [↑]

Lemma [number of tokens]

NGSL-1 [lems 242 : types 256 : tokens 405]

able_[2] accord_[1] after_[1] age_[2] agency_[1] anyway_[1] attack_[1] average_[1] avoid_[1]
away_[2] back_[2] balance_[1] bar_[1] be_[13] become_[1] begin_[1] bit_[1] break_[2]
bring_[1] build_[1] business_[1] buy_[1] campaign_[1] can_[1] capital_[1] carry_[1] case_[1]
catch_[2] cell_[1] chance_[1] change_[2] check_[3] claim_[1] come_[1] common_[1]
community_[1] computer_[3] contain_[1] cost_[1] could_[3] cross_[1] cut_[3] deal_[1]
difficult_[1] difficulty_[1] discover_[1] discuss_[1] discussion_[1] do_[7] door_[1] down_[5]
draw_[1] drink_[1] drive_[5] drop_[1] eat_[1] economic_[1] environment_[1] event_[1]
exercise_[1] expect_[1] experience_[1] eye_[1] fall_[4] far_[1] feel_[4] fight_[1] find_[3]
fire_[2] fit_[1] fly_[1] food_[1] for_[2] foreign_[1] form_[1] full_[2] game_[1] generation_[1]
get_[5] give_[4] go_[4] good_[3] grow_[1] hand_[1] hard_[1] have_[6] head_[1] hear_[2]
heart_[1] high_[1] hit_[1] hold_[2] how_[2] human_[1] i_[5] idea_[1] imagine_[1] improve_[1]
in_[5] individual_[1] industry_[1] influence_[1] it_[9] keep_[2] key_[1] know_[1] learn_[1]
leave_[1] letter_[1] lie_[1] like_[1] likely_[1] long_[1] look_[3] lose_[1] low_[1] magazine_[1]
make_[1] man_[1] mark_[3] material_[2] mean_[1] memory_[1] mention_[1] might_[1]
mind_[2] modern_[2] money_[2] more_[1] name_[1] natural_[1] nature_[1] never_[1] next_[1]
no_[1] not_[4] of_[5] off_[3] offer_[1] oil_[1] on_[7] open_[1] out_[7] over_[3] paper_[1]
particularly_[1] pay_[1] piece_[1] please_[1] point_[1] policy_[1] pound_[1] power_[1]
prefer_[1] pressure_[1] produce_[1] promise_[1] protect_[1] pull_[2] put_[2] question_[1]
quite_[1] raise_[1] read_[1] really_[1] recognize_[1] red_[2] release_[2] replace_[1] right_[1]
ring_[1] rise_[1] room_[1] round_[2] run_[3] save_[1] say_[1] school_[1] screen_[1] see_[1]
sell_[1] send_[1] set_[1] share_[2] short_[1] should_[2] sing_[1] sit_[1] size_[1] so_[2]
something_[2] sorry_[1] speak_[1] spend_[1] sport_[1] stand_[1] star_[2] station_[1] stay_[1]
stop_[1] store_[1] style_[1] success_[1] successful_[1] supply_[1] suppose_[2] surprise_[1]
take_[3] tax_[1] teach_[1] teacher_[1] team_[1] technology_[1] tell_[1] that_[3] the_[3]
think_[1] through_[3] throw_[1] time_[1] to_[9] travel_[1] treatment_[1] trip_[1] try_[1]
understand_[1] up_[5] use_[2] walk_[1] water_[1] wear_[2] well_[3] will_[1] win_[1]
window_[1] with_[4] wonder_[1] work_[1] worry_[1] would_[1] write_[1] you_[4]

NGSL-2 [lems 136 : types 138 : tokens 159]

abroad_[1] actor_[1] ad_[1] afford_[1] amaze_[1] appreciate_[1] army_[1] atmosphere_[1]
attitude_[1] battle_[2] blame_[1] blood_[1] blow_[1] bone_[1] bottom_[1] branch_[1] burn_[1]
calm_[1] cancer_[1] classic_[1] climate_[1] climb_[2] comfortable_[1] complex_[1]
connect_[1] connection_[1] cool_[1] criminal_[1] defeat_[1] destroy_[1] diet_[1]
disappoint_[1] disk_[1] double_[1] enemy_[1] engine_[1] equipment_[2] escape_[2] excite_[1]
excuse_[1] expensive_[1] extreme_[1] famous_[1] farmer_[1] fat_[2] feed_[1] finger_[1]
flat_[2] freedom_[1] freeze_[1] fuel_[1] gap_[1] gas_[1] glad_[1] glass_[1] global_[2] gold_[2]
guard_[1] habit_[1] hang_[1] hide_[2] hunt_[1] hurt_[1] ice_[2] ignore_[1] import_[1]
incident_[1] industrial_[1] joint_[1] jump_[2] king_[1] knock_[1] launch_[1] leg_[1] lift_[1]
luck_[2] lucky_[1] metal_[1] monitor_[1] mouse_[2] neighborhood_[1] novel_[2] online_[1]
opposite_[1] ordinary_[1] originally_[1] path_[1] peace_[1] plain_[1] plastic_[1] pocket_[2]
pop_[1] prison_[1] protein_[1] refuse_[1] reporter_[1] reserve_[1] ride_[1] safety_[1]
scientist_[1] settle_[1] shake_[1] sharp_[1] shock_[1] shoot_[4] shut_[1] sick_[1] skin_[2]
slip_[1] smell_[1] soft_[1] software_[1] spell_[1] steal_[2] stick_[2] stone_[2] storm_[1]
straight_[1] survey_[1] survive_[1] swim_[1] tear_[1] tool_[1] tour_[1] union_[1] unusual_[1]
user_[1] volunteer_[2] wake_[1] warm_[1] waste_[1] weak_[1] weapon_[1] weather_[1]
wood_[1] yard_[1]

NGSL-3 [lems 72 : types 72 : tokens 77]

adjust_[1] adjustment_[1] bend_[1] bite_[3] blind_[1] brilliant_[1] bury_[1] cable_[1] cap_[1] chest_[1] coal_[1] crew_[1] currency_[1] designer_[1] dig_[1] disagree_[1] elderly_[1] electric_[1] electricity_[1] embarrass_[1] era_[1] excitement_[1] false_[1] fiction_[2] flag_[1] gate_[1] gentle_[1] gradually_[1] immigrant_[1] infection_[1] iron_[1] jacket_[1] knee_[1] leather_[1] lip_[1] loose_[1] magic_[1] menu_[2] nervous_[1] noun_[1] pollution_[1] portion_[1] portrait_[1] potato_[1] pretend_[1] rescue_[1] revise_[1] rival_[1] scream_[1] shade_[1] sheep_[1] shore_[1] silver_[1] sink_[1] steel_[2] stomach_[1] sugar_[1] sweep_[1] tight_[1] tongue_[1] tournament_[1] trap_[1] trick_[1] twin_[1] unite_[1] upset_[1] verb_[1] virus_[1] vital_[1] wooden_[1] wound_[1] zone_[1]

NAWL [lems 41 : types 41 : tokens 44]

absorb_[1] aluminum_[1] arrow_[2] bleed_[1] calcium_[1] cattle_[1] cheat_[1] click_[1] comma_[1] diagnosis_[1] emperor_[1] fossil_[1] goods_[1] injection_[1] invade_[1] invasion_[1] legend_[1] lifetime_[1] liver_[1] lung_[1] mineral_[2] molecule_[1] non_[1] oxygen_[1] painful_[1] potassium_[1] quotation_[1] ray_[1] rotation_[1] sniff_[1] soluble_[2] sophisticate_[1] stripe_[1] swell_[1] sword_[1] synthetic_[1] thumb_[1] treaty_[1] tribe_[1] tropical_[1] vitamin_[1]

OFFLIST: [?: types 211 : tokens 218]

adjective_[1] alloy_[1] ambition_[1] ankle_[1] annoyed_[1] anthem_[1] apostrophe_[1] appliance_[1] archaeologist_[1] archer_[1] armour_[1] axe_[1] baggy_[1] bakelite_[1] bandage_[1] bark_[1] bc_[1] bow_[1] breaker_[1] brigade_[1] bronze_[1] burglary_[1] butterflies_[1] cabbage_[1] calf_[1] camcorder_[1] capsule_[1] carbohydrate_[1] careless_[1] celebrity_[1] cgi_[1] charcoal_[1] climber_[1] cloak_[1] cloth_[1] comic_[1] congratulations_[1] conquer_[1] contestant_[1] conumber_[1] cookery_[1] copper_[1] corn_[1] corrupt_[1] cotton_[1] cpu_[1] crocodile_[1] crooked_[1] crossroads_[1] crown_[1] cruel_[1] crust_[1] cub_[1] cursor_[1] cyclone_[1] deforestation_[1] den_[1] denim_[1] desktop_[1] disguise_[2] dive_[1] donation_[1] doughnut_[1] download_[1] drought_[1] dump_[1] escalator_[1] euro_[1] evacuate_[1] explorer_[1] extinct_[1] eyesight_[1] fake_[1] fame_[1] fashionable_[1] fi_[1] firewall_[1] flared_[1] folder_[1] forehead_[1] fundraising_[1] fur_[1] gamble_[1] generator_[1] gig_[1] greenhouse_[1] handcuff_[1] handcuffs_[1] handicapped_[1] hardware_[1] headphones_[1] heel_[1] heeled_[1] helmet_[1] hibernate_[1] hunter_[1] hurricane_[1] icon_[1] internet_[1] invader_[1] itchy_[1] jelly_[1] jellyfish_[1] junk_[1] kangaroo_[1] keyboard_[1] kidnap_[1] kingdom_[1] knight_[1] koala_[1] laptop_[1] litter_[1] magnesium_[1] mat_[1] medieval_[1] melt_[2] merry_[1] miner_[1] monument_[1] notebook_[1] nylon_[1] optimist_[1] ore_[1] orphan_[1] orphanage_[1] outback_[1] outlaw_[1] palace_[1] palm_[1] password_[1] patent_[1] penguin_[1] pessimist_[1] phosphorus_[1] photographer_[1] pity_[1] plaster_[1] platypus_[1] poisonous_[1] polluted_[1] polyester_[1] polystyrene_[1] posture_[1] premiership_[1] prescription_[1] presenter_[1] prosperity_[1] publicity_[1] pvc_[1] queen_[1] quit_[1] raffle_[1] rainforest_[1] realist_[1] reassure_[1] recyclable_[1] recycle_[1] rivet_[1] rower_[1] rubber_[1] saltwater_[1] sardine_[1] saucer_[1] seatbelt_[1] settler_[1] shark_[1] sheriff_[1] shield_[1] silk_[1] skyscraper_[1] sleeved_[2] sling_[1] sore_[2] souvenir_[1] spectator_[1] spinach_[1] standby_[1] steam_[1] steep_[1] strenuous_[1] stunt_[3] suburb_[1] suck_[1] sunbathe_[1] sunglasses_[1] sunscreen_[1] sunshine_[1] tablet_[1] tailor_[1] tanker_[2] temple_[1] thigh_[1] tin_[1] tugboat_[1] tv_[1] typhoon_[1] underwear_[1] vandalized_[1] vegetarian_[1] waterproof_[1] webcam_[1] website_[1] whale_[1] wholesaler_[1] wi_[1] wool_[1] woolen_[1] workmen_[1] wow_[1] wrist_[1] zinc_[1]

Appendix 4.1b - Project 4 the Third Edition (Hutchinson, 2009) (BNC-COCA)

BNC-COCA-1,000 Families: [fams 278 : types 304 : tokens 464]

able_[2] act_[1] advertise_[1] afford_[1] after_[1] age_[2] amaze_[1] any_[1] away_[2]
back_[2] bar_[1] be_[13] become_[1] begin_[1] better_[1] bit_[1] blood_[1] blow_[1] bone_[1]
bottom_[1] break_[3] bring_[1] build_[1] burn_[1] business_[1] buy_[1] can_[1] care_[1]
carry_[1] case_[1] catch_[2] chance_[1] change_[2] check_[3] climb_[3] come_[1] comfort_[1]
computer_[3] cook_[1] cool_[1] cost_[1] could_[3] cross_[1] cut_[3] deal_[1] difficult_[2]
dig_[1] discover_[1] do_[7] door_[1] double_[1] down_[5] draw_[1] drink_[1] drive_[5]
drop_[1] eat_[1] engine_[1] excite_[2] excuse_[1] expect_[1] expensive_[1] experience_[1]
eye_[1] fall_[4] far_[1] farm_[1] fat_[2] feed_[1] feel_[4] fight_[1] find_[3] finger_[1] fire_[2]
fit_[1] flat_[2] fly_[1] food_[1] for_[2] form_[1] free_[1] freeze_[1] full_[2] game_[1] gas_[1]
gentle_[1] get_[5] give_[4] glad_[1] glass_[1] go_[4] gold_[2] good_[2] grow_[1] hand_[1]
hang_[1] hard_[1] have_[6] head_[1] hear_[2] heart_[1] hide_[2] high_[1] hit_[1] hold_[2]
how_[2] human_[1] hunt_[2] hurt_[1] i_[5] ice_[2] idea_[1] imagine_[1] in_[5] internet_[1]
it_[9] jump_[2] keep_[2] key_[1] king_[2] knock_[1] know_[1] learn_[1] leave_[1] leg_[1]
letter_[1] lie_[1] lift_[1] like_[1] lip_[1] long_[1] look_[3] lose_[1] low_[1] luck_[3] make_[1]
man_[1] mark_[3] mean_[1] mention_[1] might_[1] mind_[2] money_[2] more_[1] name_[1]
nature_[2] neighbour_[1] never_[1] next_[1] no_[1] not_[4] of_[5] off_[3] offer_[1] oil_[1]
on_[7] open_[1] out_[7] over_[3] pain_[1] paper_[1] particular_[1] pay_[1] photograph_[1]
piece_[1] please_[1] point_[1] pop_[1] power_[1] present_[1] prison_[1] promise_[1]
protect_[1] pull_[2] put_[2] queen_[1] question_[1] quite_[1] raise_[1] read_[1] real_[1]
really_[1] red_[2] report_[1] ride_[1] right_[1] ring_[1] rise_[1] room_[1] round_[2] run_[3]
safe_[1] save_[1] say_[1] school_[1] science_[1] see_[1] sell_[1] send_[1] set_[1] settle_[2]
shake_[1] share_[2] shoot_[4] short_[1] should_[2] shut_[1] sick_[1] sing_[1] sit_[1] size_[1]
skin_[2] slip_[1] smell_[1] so_[2] soft_[1] some_[2] sorry_[1] speak_[1] spend_[1] sport_[1]
stand_[1] star_[2] station_[1] stay_[1] steal_[2] stick_[2] stone_[2] stop_[1] store_[1]
straight_[1] sun_[1] suppose_[2] surprise_[1] swim_[1] take_[3] tax_[1] teach_[2] team_[1]
tear_[1] television_[1] tell_[1] that_[3] the_[3] think_[1] through_[3] throw_[1] tight_[1]
time_[1] to_[9] travel_[1] treat_[1] trip_[1] try_[1] understand_[1] up_[5] use_[3] usual_[1]
wake_[1] walk_[1] warm_[1] waste_[1] water_[1] wear_[2] weather_[1] well_[3] will_[1]
win_[1] window_[1] with_[4] wonder_[1] wood_[2] work_[1] worry_[1] would_[1] write_[1]
yard_[1] you_[4]

BNC-COCA-2,000 Families: [fams 152 : types 159 : tokens 172]

according_[1] annoy_[1] appreciate_[1] army_[1] assure_[1] atmosphere_[1] attack_[1]
attitude_[1] average_[1] avoid_[1] balance_[1] bark_[1] battle_[2] bend_[1] bite_[3] blame_[1]
bleed_[1] blind_[1] bow_[1] branch_[1] brilliant_[1] bury_[1] cable_[1] calm_[1] cap_[1]
capital_[1] cheat_[1] chest_[1] claim_[1] classic_[1] cloth_[1] coal_[1] common_[1]
community_[1] connect_[2] contain_[1] cotton_[1] criminal_[1] crown_[1] cruel_[1] design_[1]
destroy_[1] diet_[1] disappoint_[1] discuss_[2] dive_[1] dump_[1] economy_[1] elder_[1]
electric_[2] embarrass_[1] enemy_[1] environment_[1] equipment_[2] escape_[2] event_[1]
exercise_[1] extreme_[1] famous_[1] fashion_[1] flag_[1] foreign_[1] fur_[1] gate_[1]
generation_[1] guard_[1] habit_[1] ignore_[1] improve_[1] individual_[1] industry_[2]
influence_[1] iron_[1] jacket_[1] knee_[1] likely_[1] loose_[1] magazine_[1] magic_[1]
material_[2] melt_[2] memory_[1] merry_[1] metal_[1] modern_[2] mouse_[2] nervous_[1]
non_[1] opposite_[1] ordinary_[1] original_[1] path_[1] peace_[1] pity_[1] plain_[1] plastic_[1]
pocket_[2] poison_[1] policy_[1] pollute_[2] potato_[1] prefer_[1] pressure_[1] pretend_[1]
produce_[1] quit_[1] quote_[1] ray_[1] recognize_[1] refuse_[1] release_[2] replace_[1]
reserve_[1] row_[1] scream_[1] screen_[1] shade_[1] sharp_[1] sheep_[1] shock_[1] shore_[1]
silver_[1] sink_[1] sore_[2] spell_[1] steam_[1] steel_[2] stomach_[1] storm_[1] style_[1]
success_[2] suck_[1] sugar_[1] supply_[1] survive_[1] sweep_[1] sword_[1] technology_[1]
tin_[1] tongue_[1] tool_[1] tour_[1] trap_[1] trick_[1] twin_[1] union_[1] unite_[1] upset_[1]
weak_[1] weapon_[1] wool_[2] wound_[1]

BNC-COCA-3,000 Families: [fams 66 : types 69 : tokens 73]

abroad_[1] absorb_[1] adjust_[2] agency_[1] campaign_[1] cancer_[1] cattle_[1] cell_[1]
climate_[1] complex_[1] contest_[1] corrupt_[1] crew_[1] currency_[1] defeat_[1] disagree_[1]
disc_[1] donate_[1] era_[1] explore_[1] false_[1] fiction_[2] fuel_[1] gap_[1] global_[2]
goods_[1] gradual_[1] heel_[2] immigrant_[1] import_[1] incident_[1] infect_[1] inject_[1]
invasion_[1] joint_[1] launch_[1] leather_[1] legend_[1] menu_[2] miner_[1] molecule_[1]
monitor_[1] novel_[2] palace_[1] palm_[1] portion_[1] portrait_[1] prosper_[1] protein_[1]
rescue_[1] revise_[1] rival_[1] shield_[1] silk_[1] software_[1] sophisticated_[1] suburb_[1]
survey_[1] swell_[1] tournament_[1] treaty_[1] tribe_[1] virus_[1] vital_[1] volunteer_[2]
zone_[1]

BNC-COCA-4,000 Families: [fams 58 : types 60 : tokens 64]

aluminium_[1] ambition_[1] ankle_[1] armour_[1] arrow_[2] axe_[1] bronze_[1] butterfly_[1]
celebrity_[1] click_[1] comic_[1] congratulate_[1] copper_[1] corn_[1] crust_[1] diagnosis_[1]
disguise_[2] emperor_[1] evacuate_[1] fake_[1] fame_[1] forehead_[1] fossil_[1] gamble_[1]
greenhouse_[1] handicap_[1] hardware_[1] helmet_[1] hurricane_[1] invade_[2] keyboard_[1]
kidnap_[1] liver_[1] lung_[1] mat_[1] medieval_[1] mineral_[2] monument_[1] optimist_[1]
oxygen_[1] patent_[1] premier_[1] prescription_[1] publicity_[1] recycle_[2] rotate_[1]
rubber_[1] sheriff_[1] sleeve_[2] steep_[1] stripe_[1] temple_[1] thigh_[1] thumb_[1]
tropics_[1] vitamin_[1] whale_[1] wrist_[1]

BNC-COCA-5,000 Families: [fams 31 : types 32 : tokens 32]

appliance_[1] brigade_[1] burgle_[1] calf_[1] carbohydrate_[1] cloak_[1] conquer_[1]
crook_[1] cub_[1] drought_[1] extinct_[1] flare_[1] generator_[1] gig_[1] icon_[1] junk_[1]
knight_[1] litter_[1] orphan_[2] plaster_[1] posture_[1] pounding_[1] shark_[1] sniff_[1]
spectator_[1] synthetic_[1] tablet_[1] tailor_[1] vegetarian_[1] verb_[1] wholesale_[1]

BNC-COCA-6,000 Families: [fams 18 : types 18 : tokens 19]

adjective_[1] bandage_[1] cabbage_[1] calcium_[1] capsule_[1] charcoal_[1] den_[1] euro_[1]
folder_[1] itch_[1] jelly_[1] nylon_[1] ore_[1] outlaw_[1] sling_[1] souvenir_[1] tanker_[2]
vandal_[1]

BNC-COCA-7,000 Families: [fams 16 : types 17 : tokens 20]

alloy_[1] anthem_[1] archer_[1] crocodile_[1] denim_[1] handcuff_[2] noun_[1] password_[1]
potassium_[1] rivet_[1] saucer_[1] soluble_[2] spinach_[1] strenuous_[1] stunt_[3] zinc_[1]

BNC-COCA-8,000 Families: [fams 7 : types 7 : tokens 7]

baggy_[1] camcorder_[1] deforest_[1] doughnut_[1] magnesium_[1] penguin_[1] polyester_[1]

BNC-COCA-9,000 Families: [fams 6 : types 6 : tokens 6]

comma_[1] cursor_[1] kangaroo_[1] phosphorus_[1] raffle_[1] sardine_[1]

BNC-COCA-10,000 Families: [fams 3 : types 3 : tokens 3]

cyclone_[1] escalator_[1] hibernate_[1]

BNC-COCA-11,000 Families: [fams 4 : types 4 : tokens 4]

apostrophe_[1] outback_[1] pessimist_[1] polystyrene_[1]

BNC-COCA-12,000 Families: [fams 1 : types 1 : tokens 1]

typhoon_[1]

BNC-COCA-15,000 Families: [fams 1 : types 1 : tokens 1]

koala_[1]

BNC-COCA-16,000 Families: [fams 1 : types 1 : tokens 1]

platypus_[1]

BNC-COCA-22,000 Families: [fams 1 : types 1 : tokens 1]

bakelite_[1]

OFFLIST: [?: types 35 : tokens 35]

archaeologist_[1] bc_[1] cgi_[1] conumber_[1] cpu_[1] crossroads_[1] desktop_[1]
download_[1] eyesight_[1] fi_[1] firewall_[1] fundraising_[1] headphones_[1] jellyfish_[1]
laptop_[1] lifetime_[1] notebook_[1] online_[1] pvc_[1] rainforest_[1] saltwater_[1]
seatbelt_[1] skyscraper_[1] standby_[1] sunbathe_[1] sunglasses_[1] sunscreen_[1] tugboat_[1]
underwear_[1] waterproof_[1] webcam_[1] website_[1] wi_[1] workmen_[1] wow_[1]

Appendix 4.2a - Hello Kids! Angličtina pro 7. ročník ZŠ (Zahálková, 2010) (NGSL)

Lemmas List [↑]

Lemma [number of tokens]

NGSL-1 [lems 185 : types 190 : tokens 229]

adult_[1] age_[3] air_[1] almost_[1] attack_[1] balance_[1] bank_[1] base_[1] be_[1]
become_[1] bill_[1] bit_[1] black_[1] board_[1] break_[1] bring_[1] build_[1] buy_[1] by_[1]
can_[1] car_[1] care_[2] carry_[1] certain_[1] change_[1] church_[1] close_[1] come_[1]
community_[1] competition_[1] control_[2] cover_[1] cross_[1] cut_[1] damage_[1]
depend_[1] do_[1] draw_[1] drink_[1] drive_[2] eat_[1] establish_[1] event_[1] ever_[1]
experience_[1] eye_[1] fact_[1] fall_[1] feel_[2] fight_[1] fill_[2] financial_[1] find_[1] fly_[1]
foot_[1] for_[1] forget_[1] form_[2] free_[1] get_[2] give_[1] go_[1] ground_[1] have_[1]
he_[1] hear_[1] high_[1] hit_[1] hold_[1] human_[1] in_[2] include_[1] instead_[1] just_[1]
keep_[2] kill_[1] know_[2] knowledge_[1] land_[1] language_[1] leader_[1] leave_[1] life_[3]
light_[2] live_[1] lose_[1] lot_[1] low_[1] make_[1] marry_[1] measure_[1] medium_[1]
meet_[1] memory_[1] money_[1] move_[1] movement_[1] natural_[1] necessary_[1] need_[1]
nor_[1] of_[6] officer_[1] oil_[1] on_[3] operation_[1] order_[1] own_[1] pass_[1] pattern_[1]
pick_[1] place_[1] plant_[2] population_[1] protect_[1] public_[1] pull_[1] put_[1] reach_[2]
read_[1] real_[1] receive_[1] recognize_[1] reduce_[1] remain_[1] replace_[1] respect_[1]
return_[1] right_[2] ring_[1] room_[1] rule_[1] run_[1] sale_[1] save_[1] say_[1] school_[1]
see_[1] sell_[1] send_[1] sense_[1] set_[1] side_[1] sign_[1] similar_[1] simple_[1] sing_[1]
single_[1] sit_[1] sleep_[1] sort_[1] sound_[1] space_[1] speak_[1] spend_[1] stand_[1] star_[1]
surprise_[1] system_[1] table_[1] take_[2] teach_[1] tell_[1] the_[6] think_[1] through_[1]
throw_[1] to_[2] top_[1] toward_[1] trade_[1] train_[1] tree_[9] understand_[1] up_[2]
useful_[1] voice_[2] vote_[1] wait_[1] war_[2] wear_[1] well_[1] win_[1] world_[2] write_[1]

NGSL-2 [lems 85 : types 88 : tokens 92]

abroad_[1] accident_[1] aid_[1] alive_[1] battle_[1] beat_[1] beauty_[1] belong_[1] blood_[1]
bottom_[1] brain_[1] breathe_[1] burn_[1] calm_[1] cash_[1] chain_[1] chief_[1] civil_[1]
climate_[1] crash_[2] danger_[1] declare_[1] desk_[1] destroy_[1] egg_[1] enemy_[1]
equal_[1] escape_[1] explore_[1] extremely_[1] fair_[1] fashion_[1] freeze_[1] gas_[1]
guard_[1] hang_[2] heat_[2] hide_[1] honor_[1] hurt_[1] lay_[1] medicine_[1] metal_[1]
milk_[1] mostly_[1] neighborhood_[1] neither_[1] noise_[1] origin_[1] path_[1] powerful_[1]
properly_[1] religious_[1] ride_[1] row_[1] settle_[1] shake_[1] sharp_[1] ship_[1] shoot_[1]
smell_[1] speed_[1] spell_[2] spirit_[1] steal_[1] stick_[2] stone_[1] strange_[1] swim_[1]
taste_[1] tear_[1] thick_[1] ticket_[3] tiny_[1] tool_[1] traffic_[1] transport_[1] typical_[1]
unfortunately_[1] vehicle_[1] waste_[1] wave_[1] weapon_[1] wed_[1] wild_[1]

NGSL-3 [lems 52 : types 54 : tokens 55]

altogether_[1] barrier_[1] blind_[1] ceremony_[1] chest_[1] cough_[1] counter_[1] craft_[1]
cream_[1] curious_[1] dust_[1] electricity_[1] entertainment_[1] entrance_[1] exam_[1]
forever_[1] found_[1] grain_[1] hook_[1] illegal_[1] immigrant_[1] iron_[2] lend_[1] lip_[1]
liquid_[1] loud_[1] muscle_[1] nerve_[1] port_[1] pot_[1] prisoner_[1] province_[1] pub_[1]
pump_[1] roof_[1] ruin_[1] sail_[3] seal_[1] sensitive_[1] servant_[1] shell_[1] slave_[1]
stir_[1] stomach_[1] sweep_[1] tale_[1] tap_[1] tent_[1] tongue_[1] universe_[1] van_[1]
web_[1]

NAWL [lems 31 : types 31 : tokens 31]

accent_[1] antibiotic_[1] arrow_[1] bleed_[1] cattle_[1] cheat_[1] civilization_[1] fever_[1]
freely_[1] kidney_[1] legend_[1] lifestyle_[1] liver_[1] lung_[1] mechanic_[1] nasty_[1]
organ_[1] oxygen_[1] painful_[1] pest_[1] physically_[1] powder_[1] recipe_[1] rope_[1]
slavery_[1] solar_[1] swell_[1] sword_[1] toxic_[1] tribe_[1] tropical_[1]

OFFLIST: [?: types 215 : tokens 217]

almond_[1] amusing_[1] ankle_[1] apricot_[1] avenue_[1] bagpipes_[1] bark_[1] barley_[1]
battery_[1] bay_[1] beast_[1] beaver_[1] beech_[1] belly_[1] bench_[1] bishop_[1] bison_[1]
blink_[1] blister_[1] blueberry_[1] boar_[1] bow_[1] bracelet_[1] brake_[1] brewery_[1]
british_[1] bronze_[1] buffalo_[1] bush_[1] cacao_[1] calais_[1] captain_[1] carriage_[1]
carving_[1] cave_[1] celt_[1] celtic_[1] christianity_[1] clan_[1] cloth_[1] conqueror_[1]
coral_[1] cotton_[1] crawl_[1] croatia_[1] crossroads_[1] cuisine_[1] deer_[2] dessert_[1]
diagonally_[1] diamond_[1] disabled_[1] dough_[1] dragon_[1] eagle_[1] earrings_[1]
earthquake_[1] egyptian_[1] elk_[1] equator_[1] explorer_[1] eyelid_[1] faint_[1] fame_[1]
feast_[1] fiddles_[1] fingerprints_[1] fireball_[1] flame_[1] flavour_[1] forehead_[1]
forester_[1] fortunately_[1] fox_[1] frown_[1] fur_[1] gaelic_[1] galaxy_[1] glitter_[1]
goodies_[1] graze_[1] greedy_[1] gum_[1] handicapped_[1] hare_[1] hazelnut_[1]
hedgehog_[1] homesick_[1] ii_[1] ingredients_[1] inland_[1] irish_[1] isles_[1] jar_[1]
jewellery_[1] jews_[1] jungle_[1] ketchup_[1] knight_[1] lamp_[1] leprechaun_[1] lid_[1]
lifeboat_[1] litter_[1] lizard_[1] lobster_[1] lonely_[1] mammal_[1] maple_[1] margarine_[1]
mayo_[1] miracle_[1] monument_[1] moose_[1] mushroom_[1] mustard_[1] nail_[1]
nickname_[1] nursery_[2] oak_[1] oats_[1] octopus_[1] onion_[1] orbit_[1] overnight_[1]
passport_[1] pasta_[1] paw_[1] piggy_[1] pin_[1] pinch_[1] pine_[1] pirate_[1] platypus_[1]
pod_[1] pollute_[1] pottery_[1] pouch_[1] prawn_[1] prehistory_[1] priest_[1] rainforest_[1]
rash_[1] raspberry_[1] rattle_[1] recycle_[1] reef_[1] relaxation_[1] reservations_[1] reuse_[1]
rib_[1] riches_[1] rocky_[1] roman_[1] roommate_[1] salmon_[1] saver_[1] scots_[1]
sealion_[1] settler_[1] shave_[1] shipwreck_[1] shorten_[1] sightseeing_[1] silk_[1] skull_[1]
skyscraper_[1] sleeper_[1] snail_[1] soapy_[1] spear_[1] spill_[1] spruce_[1] squirrel_[1]
starfish_[1] sting_[1] storehouse_[1] subtitles_[1] suck_[1] sunburnt_[1] sunrise_[1] sunset_[1]
swallow_[1] thatched_[1] thief_[1] thieves_[1] tobacco_[1] toffee_[1] tonsillitis_[1] trader_[1]
trunk_[1] tunnel_[1] tyre_[1] unfair_[1] untouched_[1] viking_[1] vomit_[1] voyage_[1]
warrior_[1] warship_[1] waterfall_[1] waterproof_[1] weed_[1] weightlifting_[1] wellies_[1]
welsh_[1] whale_[1] wheelchair_[1] whipped_[1] whistle_[1] woodcutter_[1] wool_[1]
yawn_[1] yellowish_[1] zebra_[1]

Appendix 4.2b - Hello Kids! Angličtina pro 7. ročník ZŠ (Zahálková, 2010) (BNC-COCA)

Families List [↑]

Family [number of tokens]

BNC-COCA-1,000 Families: [fams 214 : types 232 : tokens 272]

age_[3] air_[1] almost_[1] bank_[1] base_[1] be_[1] beat_[1] beauty_[1] become_[1] bill_[1] bit_[1] black_[1] blood_[1] board_[1] bottom_[1] break_[1] bring_[1] build_[1] burn_[1] bush_[1] buy_[1] by_[1] can_[1] car_[1] care_[2] carry_[1] certain_[1] change_[1] church_[1] close_[1] come_[1] control_[2] cover_[1] cross_[1] cut_[1] danger_[1] depend_[1] do_[1] draw_[1] drink_[1] drive_[2] eat_[1] egg_[1] ever_[1] experience_[1] eye_[1] fact_[1] fair_[2] fall_[1] feel_[2] fight_[1] fill_[2] find_[2] fly_[1] foot_[1] for_[1] forest_[1] forget_[1] form_[2] fortunate_[2] free_[2] freeze_[1] gas_[1] get_[2] give_[1] go_[1] good_[1] ground_[1] hang_[2] have_[1] he_[1] hear_[1] heat_[2] hide_[1] high_[1] history_[1] hit_[1] hold_[1] honour_[1] human_[1] hurt_[1] in_[2] instead_[1] just_[1] keep_[2] kill_[1] know_[2] land_[1] lay_[1] lead_[1] leave_[1] life_[3] light_[2] lip_[1] live_[1] lose_[1] lot_[1] loud_[1] low_[1] make_[1] marry_[1] meet_[1] milk_[1] money_[1] most_[1] move_[2] nature_[1] necessary_[1] need_[1] neighbour_[1] noise_[1] of_[6] officer_[1] oil_[1] on_[3] order_[1] own_[1] pain_[1] pass_[1] pick_[1] place_[1] plant_[2] pot_[2] power_[1] prison_[1] proper_[1] protect_[1] public_[1] pull_[1] put_[1] reach_[2] read_[1] real_[1] return_[1] rich_[1] ride_[1] right_[1] rights_[1] ring_[1] rock_[1] room_[1] rule_[1] run_[1] sail_[3] save_[2] say_[1] school_[1] see_[1] sell_[1] send_[1] sense_[1] serve_[1] set_[1] settle_[2] shake_[1] ship_[1] shoot_[1] short_[1] side_[1] sign_[1] simple_[1] sing_[1] single_[1] sit_[1] sleep_[2] smell_[1] sort_[1] sound_[1] space_[1] speak_[1] spend_[1] stand_[1] star_[1] steal_[1] stick_[2] stone_[1] strange_[1] surprise_[1] swim_[1] system_[1] table_[1] take_[2] taste_[1] teach_[1] tear_[1] tell_[1] the_[6] thick_[1] think_[1] through_[1] throw_[1] to_[2] top_[1] touch_[1] toward_[1] train_[1] tree_[9] understand_[1] up_[2] use_[2] van_[1] voice_[2] wait_[1] war_[2] waste_[1] wave_[1] wear_[1] wed_[1] well_[1] wild_[1] win_[1] world_[2] write_[1] yellow_[1]

BNC-COCA-2,000 Families: [fams 140 : types 144 : tokens 148]

accent_[1] accident_[1] adult_[1] aid_[1] alive_[1] altogether_[1] amuse_[1] attack_[1] balance_[1] bark_[1] battery_[1] battle_[1] bay_[1] belong_[1] bleed_[1] blind_[1] bow_[1] brain_[1] brake_[1] breathe_[1] calm_[1] captain_[1] cash_[1] chain_[1] cheat_[1] chest_[1] chief_[1] cloth_[1] community_[1] competition_[1] cotton_[1] cough_[1] counter_[1] crash_[2] crawl_[1] cream_[1] curious_[1] damage_[1] desk_[1] destroy_[1] dragon_[1] dust_[1] electric_[1] enemy_[1] entertain_[1] equal_[1] escape_[1] establish_[1] event_[1] exam_[1] extreme_[1] faint_[1] fashion_[1] finance_[1] flame_[1] fox_[1] fur_[1] guard_[1] hook_[1] include_[1] iron_[2] knowledge_[1] lamp_[1] language_[1] legal_[1] lend_[1] lid_[1] lone_[1] measure_[1] medicine_[1] memory_[1] metal_[1] muscle_[1] nail_[1] nasty_[1] neither_[1] nerve_[1] nor_[1] oak_[1] onion_[1] operate_[1] path_[1] pattern_[1] physical_[1] pig_[1] pin_[1] pine_[1] pollute_[1] population_[1] port_[1] pub_[1] pump_[1] receive_[1] recipe_[1] recognize_[1] reduce_[1] relax_[1] remain_[1] replace_[1] reserve_[1] respect_[1] roof_[1] rope_[1] row_[1] ruin_[1] sale_[1] seal_[1] sharp_[1] shave_[1] shell_[1] similar_[1] slave_[2] speed_[1] spell_[2] spirit_[1] stir_[1] stomach_[1] suck_[1] swallow_[1] sweep_[1] sword_[1] tale_[1] tap_[1] tent_[1] thief_[2] ticket_[3] tiny_[1] tongue_[1] tool_[1] trade_[2] traffic_[1] trunk_[1] typical_[1] vehicle_[1] vote_[1] weapon_[1] weed_[1] whip_[1] whistle_[1] wool_[1]

BNC-COCA-3,000 Families: [fams 37 : types 38 : tokens 38]

abroad_[1] barrier_[1] bench_[1] bishop_[1] carve_[1] cattle_[1] ceremony_[1] civil_[1]
civilise_[1] climate_[1] craft_[1] declare_[1] disabled_[1] entrance_[1] explore_[2] flavour_[1]
grain_[1] immigrant_[1] ingredient_[1] legend_[1] liquid_[1] mechanic_[1] medium_[1]
organ_[1] origin_[1] powder_[1] priest_[1] province_[1] religious_[1] sensitive_[1] silk_[1]
spill_[1] swell_[1] transport_[1] tribe_[1] tunnel_[1] universe_[1]

BNC-COCA-4,000 Families: [fams 49 : types 49 : tokens 51]

ankle_[1] arrow_[1] avenue_[1] beast_[1] belly_[1] blink_[1] brew_[1] bronze_[1] carriage_[1]
cave_[1] deer_[2] diamond_[1] eagle_[1] fame_[1] fever_[1] forehead_[1] frown_[1]
galaxy_[1] graze_[1] greed_[1] handicap_[1] jar_[1] jewellery_[1] jungle_[1] liver_[1] lung_[1]
mammal_[1] miracle_[1] monument_[1] mushroom_[1] nursery_[2] orbit_[1] overnight_[1]
oxygen_[1] pinch_[1] rattle_[1] recycle_[1] rib_[1] salmon_[1] skull_[1] soap_[1] solar_[1]
sting_[1] tobacco_[1] toxic_[1] tropics_[1] warrior_[1] webs_[1] whale_[1]

BNC-COCA-5,000 Families: [fams 22 : types 22 : tokens 22]

antibiotic_[1] clan_[1] conquer_[1] coral_[1] dessert_[1] earthquake_[1] feast_[1] fiddle_[1]
glitter_[1] gum_[1] inland_[1] kidney_[1] knight_[1] litter_[1] nickname_[1] passport_[1]
pasta_[1] pest_[1] pirate_[1] reef_[1] spear_[1] voyage_[1]

BNC-COCA-6,000 Families: [fams 16 : types 16 : tokens 16]

almond_[1] blister_[1] cuisine_[1] diagonal_[1] dough_[1] hare_[1] isle_[1] lizard_[1]
lobster_[1] maple_[1] mustard_[1] paw_[1] rash_[1] squirrel_[1] vomit_[1] yawn_[1]

BNC-COCA-7,000 Families: [fams 11 : types 11 : tokens 11]

apricot_[1] barley_[1] bracelet_[1] buffalo_[1] equator_[1] margarine_[1] oats_[1] pod_[1]
pouch_[1] raspberry_[1] snail_[1]

BNC-COCA-8,000 Families: [fams 8 : types 8 : tokens 8]

beaver_[1] beech_[1] boar_[1] hazel_[1] spruce_[1] subtitle_[1] thatch_[1] tyre_[1]

BNC-COCA-9,000 Families: [fams 4 : types 4 : tokens 4]

moose_[1] octopus_[1] prawn_[1] zebra_[1]

BNC-COCA-10,000 Families: [fams 4 : types 4 : tokens 4]

bison_[1] elk_[1] hedgehog_[1] ketchup_[1]

BNC-COCA-11,000 Families: [fams 1 : types 1 : tokens 1]

toffee_[1]

BNC-COCA-12,000 Families: [fams 1 : types 1 : tokens 1]

leprechaun_[1]

BNC-COCA-14,000 Families: [fams 1 : types 1 : tokens 1]

wellie_[1]

BNC-COCA-16,000 Families: [fams 1 : types 1 : tokens 1]

platypus_[1]

BNC-COCA-17,000 Families: [fams 2 : types 2 : tokens 2]

cacao_[1] tonsillitis_[1]

OFFLIST: [?: types 44 : tokens 44]

bagpipes_[1] blueberry_[1] british_[1] calais_[1] celt_[1] celtic_[1] christianity_[1] croatia_[1]
crossroads_[1] earrings_[1] egyptian_[1] eyelid_[1] fingerprints_[1] fireball_[1] forever_[1]
gaelic_[1] homesick_[1] ii_[1] irish_[1] jews_[1] lifeboat_[1] lifestyle_[1] mayo_[1]
rainforest_[1] roman_[1] roommate_[1] scots_[1] sealion_[1] shipwreck_[1] sightseeing_[1]
skyscraper_[1] starfish_[1] storehouse_[1] sunburnt_[1] sunrise_[1] sunset_[1] viking_[1]
warship_[1] waterfall_[1] waterproof_[1] weightlifting_[1] welsh_[1] wheelchair_[1]
woodcutter_[1]

Appendix 4.3a - More! L3 (Puchta et al., 2008) (NGSL)

Lemmas List

Lemma [number of tokens]

NGSL-1 [lems 197 : types 200 : tokens 239]

across_[1] action_[1] add_[1] advantage_[1] affect_[1] after_[1] age_[1] agreement_[1] aim_[1]
allow_[1] and_[1] argument_[1] art_[1] article_[1] author_[1] average_[1] away_[1] back_[1]
be_[4] bear_[2] believe_[1] bill_[1] blue_[1] break_[1] build_[2] business_[1] by_[1] career_[1]
case_[1] cell_[1] century_[1] character_[1] charge_[1] choice_[1] clear_[1] come_[2]
community_[1] competition_[1] completely_[1] control_[1] country_[1] couple_[1] court_[1]
cover_[1] current_[1] deal_[1] decide_[1] design_[1] determine_[1] discover_[1] do_[1]
doubt_[1] down_[2] drop_[1] earn_[1] easy_[1] economic_[1] economy_[1] effect_[1] end_[1]
energy_[2] event_[1] ever_[1] exactly_[1] experience_[1] face_[1] fail_[1] fall_[1] field_[2]
film_[1] final_[1] fly_[1] for_[1] force_[1] foreign_[1] game_[1] generation_[1] get_[1]
give_[1] ground_[1] grow_[2] half_[1] hand_[1] identify_[1] in_[3] increase_[1]
international_[1] into_[4] just_[1] keep_[1] kid_[1] kill_[1] law_[1] lead_[1] leader_[1] let_[1]
lie_[1] light_[1] listen_[1] local_[1] look_[1] machine_[1] make_[3] mark_[1] market_[1]
marry_[1] material_[1] mean_[1] mind_[1] mother_[1] movie_[1] must_[1] national_[1]
object_[1] of_[3] off_[1] oil_[1] on_[2] open_[1] opinion_[1] opportunity_[1] out_[5] own_[1]
paper_[1] park_[2] performance_[1] plan_[1] plant_[1] play_[1] player_[2] police_[1]
population_[1] positive_[1] power_[2] production_[1] protect_[1] provide_[1] put_[1] reach_[1]
recent_[1] relationship_[1] release_[1] resource_[1] result_[1] review_[1] right_[1] round_[1]
rule_[1] run_[1] seat_[1] sell_[1] sense_[1] separate_[1] series_[1] serve_[1] set_[1] share_[1]
show_[1] sign_[2] similar_[1] since_[1] single_[1] skill_[1] so_[1] society_[1] source_[1]
special_[1] stage_[1] stand_[1] stuff_[1] successful_[1] support_[1] take_[2] the_[1] throw_[1]
to_[2] total_[1] touch_[1] traditional_[1] train_[1] trouble_[1] turn_[2] understand_[1] unit_[1]
up_[10] various_[1] war_[2] water_[2] way_[1] weight_[1] well_[1] wish_[1] with_[1]
worry_[1] year_[1] yet_[1] you_[1]

NGSL-2 [lems 154 : types 155 : tokens 158]

abroad_[1] advertisement_[1] afford_[1] afraid_[1] alternative_[1] ancient_[1]
approximately_[1] arrest_[1] artist_[1] background_[1] battle_[1] bet_[1] bind_[1] blood_[1]
blow_[1] bone_[1] border_[1] bottle_[1] breathe_[1] burn_[1] camera_[1] chart_[1] coach_[1]
collapse_[1] column_[1] commit_[1] complaint_[1] concert_[1] confidence_[1] contrast_[1]
council_[1] crime_[1] cup_[1] defeat_[1] defense_[1] definitely_[1] democracy_[1]
demonstrate_[1] diet_[2] disappoint_[1] earth_[1] equipment_[1] escape_[1] exhibition_[1]
factory_[1] flat_[1] flight_[1] flow_[1] folk_[1] fresh_[1] gap_[1] glass_[1] global_[1]
grade_[1] gun_[1] heat_[1] heavy_[1] helpful_[1] hide_[1] hunt_[1] illness_[1] independent_[1]
investigate_[1] island_[1] justice_[1] label_[1] landscape_[1] lawyer_[1] leadership_[1]
literature_[1] load_[1] location_[1] lucky_[1] mass_[1] massive_[1] metal_[1] mobile_[1]
multiple_[1] narrow_[1] native_[1] neighbor_[1] neither_[1] noise_[1] nuclear_[1] nurse_[1]
package_[1] pain_[1] panel_[1] permanent_[1] pilot_[1] plot_[1] politician_[1] producer_[1]
protest_[2] proud_[1] reduction_[1] refuse_[1] regulation_[1] remind_[1] remove_[1] repair_[1]
responsible_[1] ride_[1] route_[1] rush_[1] salary_[1] scale_[1] scheme_[1] scientist_[1]
self_[1] seriously_[1] settlement_[1] shake_[1] shoot_[1] shot_[1] slide_[1] smoke_[1]
snow_[1] solve_[1] soul_[1] speaker_[2] spot_[1] spread_[1] stick_[1] stone_[1] strength_[1]
studio_[1] sum_[1] survive_[1] switch_[1] talent_[1] telephone_[1] tennis_[1] theater_[1]
thick_[1] tiny_[1] tip_[1] tire_[1] toy_[1] traffic_[2] twice_[1] typical_[1] unusual_[1]
violence_[1] volume_[1] volunteer_[1] warn_[1] wash_[1] waste_[1] weak_[1] wind_[1]
wonderful_[1] wood_[1] writer_[1]

NGSL-3 [lems 83 : types 83 : tokens 84]

admire_[1] adventure_[1] album_[1] apologize_[1] behave_[1] bowl_[1] charm_[1] cheese_[1]
cigarette_[1] classical_[1] coin_[1] controversial_[1] cooperation_[1] cough_[1] crazy_[1]
cream_[1] creative_[1] currency_[1] depress_[1] desert_[2] digital_[1] disaster_[1] electric_[1]
entrance_[1] evil_[1] extract_[1] fancy_[1] fence_[1] festival_[1] flexible_[1] flood_[1]
fold_[1] found_[1] gallery_[1] gaze_[1] gently_[1] harbor_[1] hero_[1] illegal_[1]
immigrant_[1] invent_[1] justify_[1] lazy_[1] march_[1] mirror_[1] mixture_[1] moon_[1]
musician_[1] ocean_[1] overcome_[1] panic_[1] parallel_[1] peer_[1] pipe_[1] port_[1]
possession_[1] printer_[1] prize_[1] proof_[1] pump_[1] refugee_[1] resort_[1] revise_[1]
roof_[1] scream_[1] seal_[1] shine_[1] smart_[1] specialist_[1] strict_[1] strip_[1] survival_[1]
sweep_[1] symbol_[1] tower_[1] trap_[1] tube_[1] tune_[1] violent_[1] wander_[1] web_[1]
whisper_[1] wooden_[1]

NAWL [lems 42 : types 42 : tokens 42]

architect_[1] assignment_[1] blank_[1] breakdown_[1] cone_[1] consumption_[1] container_[1]
continent_[1] coordinate_[1] delta_[1] diameter_[1] dilemma_[1] diverse_[1] dye_[1]
dynamic_[1] ecology_[1] fever_[1] hip_[1] horizontal_[1] identification_[1] immune_[1]
incredible_[1] infect_[1] influential_[1] insect_[1] invade_[1] jazz_[1] mechanic_[1]
multiply_[1] planner_[1] pronounce_[1] questionnaire_[1] reactor_[1] republic_[1]
ridiculous_[1] sensation_[1] sneeze_[1] solar_[1] sophisticate_[1] subtract_[1] treaty_[1]
tribe_[1]

OFFLIST: [?: types 232 : tokens 233]

aftershock_[1] aliby_[1] amateur_[1] amuse_[1] animated_[1] apartheid_[1] appliance_[1]
aquarium_[1] archaeologist_[1] ash_[1] ashore_[1] atlas_[1] automated_[1] avalanche_[1]
babysitting_[1] bagel_[1] baker_[1] balcony_[1] bamboo_[1] baseball_[1] bean_[1]
beheaded_[1] berry_[1] bestselling_[1] blackout_[1] blond_[1] brick_[1] brighten_[1]
buggy_[1] bulb_[1] bump_[1] businessman_[1] buzz_[1] cache_[1] canteen_[1] caption_[1]
carve_[1] carving_[1] cassette_[1] castaway_[1] casual_[1] catastrophe_[1] cd_[1] chilly_[1]
circular_[1] clarinet_[1] clockwise_[1] cocoa_[1] coconut_[1] comic_[1] compilation_[1]
compressed_[1] console_[1] copper_[1] cosmetics_[1] crane_[1] crush_[1] deadly_[1] deer_[1]
den_[1] desktop_[1] diagonal_[1] digit_[1] dimwit_[1] dinosaur_[1] dirt_[1] disco_[1]
dishwasher_[1] dockyards_[1] doorbell_[1] doorkey_[1] drought_[1] drummer_[1]
dynamite_[1] earrings_[1] earthquake_[1] eco_[1] electrician_[1] elegant_[1] energetic_[1]
envious_[1] epic_[1] erupt_[1] evacuate_[1] expedition_[1] explosion_[1] fantasy_[1] flare_[1]
flu_[1] fox_[1] french_[1] fries_[1] geocaching_[1] glacier_[1] gps_[1] grizzly_[1] handful_[1]
handkerchief_[1] handprint_[1] handset_[1] headmaster_[1] headphones_[1] helmet_[1]
heroine_[1] hibernate_[1] hop_[1] housework_[1] hurricane_[1] hydroelectric_[1] idol_[1]
igloo_[1] indie_[1] influenza_[1] ingredient_[1] ink_[1] insulation_[1] intelligent_[1]
intonation_[1] keyboard_[1] lama_[1] lan_[1] lava_[1] leaflet_[1] lentil_[1] leopard_[1]
litter_[1] lonely_[1] lousy_[1] lout_[1] magnifying_[1] magpie_[1] marine_[1] marsh_[1]
megastore_[1] melt_[1] metallic_[1] miracle_[1] modem_[1] monarchy_[1] monument_[1]
mosquito_[1] motorway_[1] mpnumber_[1] mud_[1] mudslide_[1] mummy_[1] muscular_[1]
nightmare_[1] notebook_[1] olive_[1] outbreak_[1] palmtop_[1] pandemic_[1] parade_[1]
paradise_[1] parliament_[1] passionate_[1] pavement_[1] peaceful_[1] peach_[1] petition_[1]
petrol_[1] pity_[1] poison_[1] poker_[1] pond_[1] popcorn_[1] postpone_[1] prehistoric_[1]
programmer_[1] punk_[1] raft_[1] ranger_[1] rap_[1] recycle_[1] renewable_[1] resin_[1]
reuse_[1] rhino_[1] rom_[1] rubbish_[1] rude_[1] safari_[1] saxophone_[1] scruffy_[1]
sheepdog_[1] shrink_[1] slogan_[1] spike_[1] sporty_[1] spymaster_[1] starving_[1]
storyline_[1] stove_[1] straightaway_[1] stunt_[1] suitcase_[1] sulk_[1] superstition_[1]
superstitious_[1] swarm_[1] sympathy_[1] temple_[1] terror_[1] tide_[1] toiletries_[1] torn_[2]
trailer_[1] trance_[1] treasure_[1] treehouse_[1] tremor_[1] truffle_[1] tsunami_[1] tunnel_[1]

turbine_[1] tv_[1] usb_[1] vet_[1] villain_[1] vocals_[1] volcano_[1] warden_[1] warrior_[1]
waterfall_[1] webpage_[1] website_[1] wetland_[1] wilderness_[1] windscreen_[1]
windsurfing_[1] wipers_[1]

Appendix 4.3b - More! L3 (Puchta et al., 2008) (BNC-COCA)

Families List [↑] Family [number of tokens]

BNC-COCA-1,000 Families: [fams 215 : types 228 : tokens 267]

across_[1] act_[1] add_[1] advertise_[1] afford_[1] afraid_[1] after_[1] age_[1] agree_[1]
allow_[1] and_[1] art_[2] away_[1] back_[1] be_[4] bear_[2] believe_[1] bet_[1] bill_[1]
blood_[1] blow_[1] blue_[1] bone_[1] bottle_[1] break_[1] bright_[1] build_[2] burn_[1]
business_[1] by_[1] case_[1] charge_[1] choice_[1] clear_[1] come_[2] complete_[1]
control_[1] country_[1] couple_[1] court_[1] cover_[1] crazy_[1] crime_[1] cup_[1] dead_[1]
deal_[1] decide_[1] definite_[1] dirty_[1] discover_[1] do_[1] doubt_[1] down_[2] drop_[1]
earth_[1] easy_[1] end_[1] ever_[1] exact_[1] experience_[1] face_[1] fall_[1] field_[2]
film_[1] final_[1] find_[1] flat_[1] fly_[1] for_[1] force_[1] fresh_[1] game_[1] gentle_[1]
get_[1] give_[1] glass_[1] ground_[1] grow_[2] gun_[1] half_[1] hand_[2] heat_[1] heavy_[1]
help_[1] hide_[1] hunt_[1] in_[3] into_[4] island_[1] just_[1] keep_[1] kill_[1] law_[1] lazy_[1]
lead_[3] let_[1] lie_[1] light_[1] listen_[1] load_[1] local_[1] look_[1] luck_[1] machine_[1]
make_[3] mark_[1] market_[1] marry_[1] mean_[1] mind_[1] mother_[1] movie_[1] mum_[1]
music_[1] must_[1] nation_[1] neighbour_[1] new_[1] noise_[1] nurse_[1] of_[3] off_[1]
oil_[1] on_[2] open_[1] out_[5] own_[1] pack_[1] pain_[1] paper_[1] park_[2] plan_[2]
plant_[1] play_[3] police_[1] power_[2] programme_[1] protect_[1] put_[1] reach_[1]
recent_[1] relate_[1] responsible_[1] ride_[1] rights_[1] round_[1] rubbish_[1] rule_[1] run_[1]
science_[1] seat_[1] self_[1] sell_[1] sense_[1] serious_[1] serve_[1] set_[1] settle_[1]
shake_[1] share_[1] shoot_[2] show_[1] sign_[2] since_[1] single_[1] smoke_[1] snow_[1]
so_[1] speak_[2] special_[2] sport_[1] spot_[1] stage_[1] stand_[1] stick_[1] stone_[1] stuff_[1]
support_[1] take_[2] tear_[2] telephone_[1] television_[1] the_[1] thick_[1] throw_[1] tire_[1]
to_[2] total_[1] touch_[1] train_[1] trouble_[1] turn_[2] two_[1] understand_[1] up_[10]
use_[1] usual_[1] war_[2] wash_[1] waste_[1] water_[2] way_[1] web_[1] weight_[1] well_[1]
wind_[1] wish_[1] with_[1] wonder_[1] wood_[2] worry_[1] write_[1] year_[1] yet_[1]
you_[1]

BNC-COCA-2,000 Families: [fams 167 : types 173 : tokens 178]

admire_[1] advantage_[1] adventure_[1] affect_[1] amuse_[1] argue_[1] arrest_[1] article_[1]
average_[1] background_[1] bake_[1] battle_[1] bean_[1] bind_[1] blank_[1] blonde_[1]
bowl_[1] breathe_[1] brick_[1] bump_[1] camera_[1] career_[1] casual_[1] century_[1]
character_[1] charm_[1] cheese_[1] cigarette_[1] circle_[1] classic_[1] coach_[1] commit_[1]
community_[1] competition_[1] contain_[1] cough_[1] council_[1] cream_[1] create_[1]
current_[1] defence_[1] depress_[1] desert_[2] design_[1] determine_[1] diet_[2]
disappoint_[1] drum_[1] earn_[1] economy_[2] effect_[1] electric_[2] energy_[3]
equipment_[1] escape_[1] event_[1] evil_[1] fail_[1] fancy_[1] fence_[1] flight_[1] flood_[1]
flow_[1] fold_[1] folk_[1] foreign_[1] fox_[1] fry_[1] generation_[1] grade_[1] hero_[1]
identify_[2] ill_[1] increase_[1] incredible_[1] influence_[1] investigate_[1] justice_[1]
lawyer_[1] legal_[1] locate_[1] lone_[1] march_[1] mass_[1] massive_[1] material_[1] melt_[1]
metal_[2] mirror_[1] moon_[1] mud_[1] narrow_[1] native_[1] neither_[1] object_[1]
opinion_[1] opportunity_[1] panic_[1] peace_[1] perform_[1] pipe_[1] pity_[1] poison_[1]
politics_[1] population_[1] port_[1] positive_[1] possess_[1] print_[1] produce_[1] product_[1]
pronounce_[1] protest_[2] proud_[1] provide_[1] pump_[1] react_[1] reduce_[1] refuse_[1]
release_[1] remind_[1] remove_[1] repair_[1] result_[1] ridiculous_[1] roof_[1] rude_[1]
rush_[1] salary_[1] scale_[1] scream_[1] seal_[1] separate_[1] series_[1] shine_[1] similar_[1]
skill_[1] slide_[1] smart_[1] society_[1] soul_[1] spread_[1] starve_[1] strength_[1] strip_[1]
success_[1] survive_[2] sweep_[1] switch_[1] theatre_[1] tide_[1] tiny_[1] tip_[1] tower_[1]
toy_[1] tradition_[1] traffic_[2] trap_[1] tune_[1] typical_[1] unit_[1] various_[1] violent_[1]
wander_[1] warn_[1] weak_[1] wipe_[1]

BNC-COCA-3,000 Families: [fams 111 : types 112 : tokens 112]

abroad_[1] aim_[1] album_[1] alternative_[1] ancient_[1] apology_[1] approximate_[1]
archaeology_[1] architect_[1] assign_[1] author_[1] behave_[1] border_[1] carve_[2] cell_[1]
chart_[1] chill_[1] coin_[1] collapse_[1] column_[1] complaint_[1] concert_[1] confidence_[1]
consumption_[1] continent_[1] contrast_[1] controversy_[1] cooperate_[1] coordinate_[1]
crush_[1] currency_[1] defeat_[1] democracy_[1] demonstrate_[1] digital_[1] disaster_[1]
diverse_[1] entrance_[1] exhibit_[1] explosion_[1] extract_[1] factory_[1] fantasy_[1]
festival_[1] flexible_[1] gallery_[1] gap_[1] gaze_[1] global_[1] harbor_[1] hip_[1]
immigrant_[1] immune_[1] independent_[1] infect_[1] ingredient_[1] international_[1]
invent_[1] justify_[1] label_[1] landscape_[1] literature_[1] marine_[1] mechanic_[1]
mixture_[1] mobile_[1] multiple_[1] nuclear_[1] ocean_[1] overcome_[1] panel_[1]
parallel_[1] parliament_[1] passion_[1] pave_[1] peer_[1] permanent_[1] petrol_[1] pilot_[1]
plot_[1] prize_[1] proof_[1] refuge_[1] regulate_[1] republic_[1] resort_[1] resource_[1]
review_[1] revise_[1] route_[1] scheme_[1] solve_[1] sophisticated_[1] source_[1] strict_[1]
studio_[1] sum_[1] symbol_[1] sympathy_[1] talent_[1] tennis_[1] terror_[1] treasure_[1]
treaty_[1] tribe_[1] tube_[1] tunnel_[1] violence_[1] volume_[1] volunteer_[1] whisper_[1]

BNC-COCA-4,000 Families: [fams 55 : types 55 : tokens 55]

amateur_[1] animate_[1] ash_[1] baseball_[1] bulb_[1] buzz_[1] comic_[1] compile_[1]
compress_[1] console_[1] copper_[1] deer_[1] diameter_[1] dilemma_[1] dynamic_[1]
elegant_[1] erupt_[1] evacuate_[1] expedition_[1] fever_[1] helmet_[1] hop_[1] horizontal_[1]
hurricane_[1] ink_[1] insect_[1] insulate_[1] intelligent_[1] invade_[1] jazz_[1] keyboard_[1]
miracle_[1] monarch_[1] monument_[1] multiply_[1] nightmare_[1] olive_[1] parade_[1]
petition_[1] poke_[1] pond_[1] postpone_[1] questionnaire_[1] ranger_[1] rap_[1] recycle_[1]
sensation_[1] shrink_[1] solar_[1] spike_[1] temple_[1] veterinarian_[1] vocal_[1] volcano_[1]
warrior_[1]

BNC-COCA-5,000 Families: [fams 30 : types 31 : tokens 31]

appliance_[1] aquarium_[1] automate_[1] balcony_[1] cassette_[1] catastrophe_[1] cone_[1]
cosmetic_[1] crane_[1] delta_[1] dinosaur_[1] drought_[1] dye_[1] earthquake_[1] ecology_[1]
epic_[1] flare_[1] handkerchief_[1] influenza_[2] leaflet_[1] litter_[1] marsh_[1] modem_[1]
mosquito_[1] paradise_[1] peach_[1] slogan_[1] stove_[1] trailer_[1] wilderness_[1]

BNC-COCA-6,000 Families: [fams 23 : types 23 : tokens 23]

apartheid_[1] ashore_[1] berry_[1] caption_[1] cocoa_[1] coconut_[1] den_[1] diagonal_[1]
digit_[1] disco_[1] glacier_[1] heroine_[1] idol_[1] kidding_[1] lava_[1] magnify_[1]
muscular_[1] punk_[1] raft_[1] swarm_[1] turbine_[1] villain_[1] warden_[1]

BNC-COCA-7,000 Families: [fams 12 : types 12 : tokens 12]

atlas_[1] avalanche_[1] bamboo_[1] canteen_[1] clockwise_[1] intone_[1] lousy_[1]
prehistoric_[1] resin_[1] saxophone_[1] stunt_[1] subtract_[1]

BNC-COCA-8,000 Families: [fams 12 : types 12 : tokens 12]

buggy_[1] cache_[1] clarinet_[1] leopard_[1] rhino_[1] safari_[1] scruff_[1] sneeze_[1]
sulk_[1] superstition_[1] trance_[1] tremor_[1]

BNC-COCA-9,000 Families: [fams 5 : types 5 : tokens 5]

dynamite_[1] envious_[1] indie_[1] superstitious_[1] tsunami_[1]

BNC-COCA-10,000 Families: [fams 7 : types 7 : tokens 7]

behead_[1] grizzly_[1] handset_[1] hibernate_[1] hydroelectric_[1] lentil_[1] truffle_[1]

BNC-COCA-11,000 Families: [fams 4 : types 4 : tokens 4]

lout_[1] rom_[1] straightaway_[1] toiletry_[1]

BNC-COCA-12,000 Families: [fams 3 : types 3 : tokens 3]

bagel_[1] magpie_[1] pandemic_[1]

BNC-COCA-14,000 Families: [fams 1 : types 1 : tokens 1]

igloo_[1]

BNC-COCA-15,000 Families: [fams 1 : types 1 : tokens 1]

palmtop_[1]

BNC-COCA-16,000 Families: [fams 1 : types 1 : tokens 1]

lama_[1]

BNC-COCA-17,000 Families: [fams 1 : types 1 : tokens 1]

megastore_[1]

OFFLIST: [?: types 43 : tokens 43]

aftershock_[1] alibi_[1] babysitting_[1] bestselling_[1] blackout_[1] breakdown_[1]
businessman_[1] castaway_[1] cd_[1] desktop_[1] dimwit_[1] dishwasher_[1] dockyards_[1]
doorbell_[1] doorway_[1] earrings_[1] eco_[1] french_[1] geocaching_[1] gps_[1] handprint_[1]
headmaster_[1] headphones_[1] housework_[1] lan_[1] motorway_[1] mpnumber_[1]
mudslide_[1] notebook_[1] outbreak_[1] popcorn_[1] sheepdog_[1] spymaster_[1]
storyline_[1] suitcase_[1] treehouse_[1] usb_[1] waterfall_[1] webpage_[1] website_[1]
wetland_[1] windscreen_[1] windsurfing_[1]

Appendix 4.4a - English Plus 2 (Hardy-Gould & Mellersh, 2011) (NGSL)

Lemmas List [↑]

Lemma [number of tokens]

NGSL-1 [lems 292 : types 317 : tokens 457]

a_[2] about_[6] accept_[1] across_[1] act_[1] after_[1] afternoon_[1] against_[1] agree_[1] allow_[1] always_[1] and_[1] another_[1] argue_[1] arm_[1] around_[1] arrive_[1] as_[2] at_[2] attack_[1] average_[1] baby_[1] back_[1] bad_[1] base_[1] basic_[1] be_[2] bear_[2] beautiful_[1] become_[2] bed_[2] behind_[1] between_[1] board_[1] book_[2] break_[3] brother_[1] bus_[2] buy_[1] call_[1] car_[2] card_[3] century_[1] child_[2] class_[1] clean_[3] clear_[1] clothes_[2] common_[1] company_[2] competition_[2] computer_[1] contact_[1] cost_[1] country_[1] couple_[1] court_[1] cross_[2] cut_[1] dance_[2] dark_[1] daughter_[1] day_[1] death_[1] decision_[2] describe_[1] difficult_[1] do_[7] dog_[2] down_[2] draw_[2] dream_[1] drive_[1] easy_[1] encourage_[1] energy_[1] evening_[1] ever_[1] everybody_[1] exercise_[1] exist_[1] expect_[1] fall_[3] family_[1] far_[1] fast_[1] father_[1] fear_[1] few_[2] finally_[1] find_[1] first_[2] fish_[2] floor_[1] fly_[2] food_[1] foot_[1] for_[1] forget_[1] forward_[1] friend_[3] from_[2] front_[2] game_[1] get_[2] go_[7] goal_[1] good_[1] group_[1] grow_[2] guess_[1] half_[1] hand_[1] happen_[1] happy_[1] hard_[1] have_[2] head_[1] hear_[1] help_[1] history_[1] holiday_[1] home_[4] hospital_[1] hour_[1] house_[4] how_[1] however_[1] human_[1] husband_[1] imagine_[1] immediately_[1] in_[3] interest_[2] into_[1] it_[1] job_[1] keep_[1] key_[3] kid_[1] kind_[1] land_[1] last_[2] learn_[1] leave_[2] light_[3] listen_[1] live_[1] look_[2] lose_[2] machine_[1] magazine_[2] make_[4] man_[1] manager_[1] marry_[1] match_[1] mean_[1] meet_[2] memory_[4] middle_[1] minute_[1] modern_[1] money_[1] month_[1] morning_[1] mother_[1] move_[2] music_[2] near_[1] need_[1] never_[1] next_[1] nice_[1] note_[1] number_[2] of_[7] off_[2] officer_[1] often_[1] on_[6] organize_[1] out_[2] outside_[1] over_[1] paint_[1] parent_[1] party_[1] pass_[1] patient_[1] phone_[2] picture_[1] plan_[1] play_[6] player_[3] police_[1] positive_[1] practice_[1] president_[1] price_[1] professional_[1] program_[2] public_[1] quarter_[1] race_[3] read_[1] real_[1] recognize_[1] record_[1] remember_[1] right_[1] ring_[1] risk_[1] road_[3] rock_[1] room_[3] rule_[1] run_[2] save_[1] school_[2] science_[1] score_[1] sea_[1] season_[1] seat_[1] second_[1] see_[1] send_[1] series_[1] serious_[1] shop_[2] short_[2] show_[1] side_[1] similar_[1] sing_[1] sister_[1] site_[1] situation_[1] sleep_[1] sometimes_[1] son_[1] soon_[1] space_[1] sport_[2] stage_[1] stand_[1] start_[2] stop_[1] story_[1] street_[1] study_[1] support_[1] suppose_[2] table_[3] take_[5] teacher_[1] team_[1] the_[14] then_[1] thing_[1] though_[1] through_[1] throw_[1] to_[8] top_[2] touch_[1] toward_[1] traditional_[1] train_[3] travel_[1] turn_[1] under_[2] university_[2] up_[5] usually_[1] view_[1] visit_[1] walk_[3] watch_[2] week_[1] well_[1] what_[2] wife_[1] win_[2] window_[1] worker_[1] worry_[1] write_[2] wrong_[1] year_[1] you_[6]

NGSL-2 [lems 149 : types 158 : tokens 178]

actor_[1] adopt_[1] advertise_[1] aid_[1] amaze_[1] angry_[1] atmosphere_[1] attractive_[1] bag_[1] band_[1] beach_[1] bet_[1] bike_[2] bird_[1] block_[1] boat_[1] bone_[1] bore_[1] bother_[1] bottom_[1] brand_[1] breakfast_[1] burn_[2] busy_[1] camera_[1] careful_[2] celebrate_[1] chair_[1] cheap_[1] climb_[2] coach_[1] coat_[1] collect_[1] compete_[2] concert_[1] construction_[1] cook_[2] copy_[1] corner_[1] crash_[1] crowd_[1] cry_[1] cycle_[3] danger_[1] dangerous_[1] decade_[1] definitely_[1] desk_[1] dinner_[1] dozen_[1] drama_[1] driver_[1] egg_[1] enemy_[1] equipment_[1] essential_[2] fan_[1] farm_[1] finger_[1] flat_[2] folk_[1] football_[2] forest_[1] formal_[1] friendly_[1] funny_[1] glass_[2] graduate_[1] hardly_[1] heavy_[1] helpful_[1] hide_[1] horse_[1] hunt_[1] ice_[1] ideal_[1] injury_[1] joke_[1] journalist_[1] journey_[1] jump_[2] kick_[1] king_[1] label_[1] lake_[1] left_[1] leg_[1] library_[1] lucky_[1] lunch_[1] mathematics_[1] meal_[1] medical_[1] metal_[1] mobile_[1] moral_[1] mountain_[1] musical_[2] neck_[1] negative_[1] neither_[1] net_[1] nobody_[1] novel_[1] owner_[1] pack_[1] partner_[1] photo_[1] pilot_[1] plane_[1] plastic_[1] practical_[1] rain_[1] rare_[1] repeat_[1] reporter_[1] rich_[1] ride_[2] river_[1]

row_2] safe_1] shirt_1] shoe_2] shoulder_1] shout_1] slow_1] smoke_1] spell_1]
storm_1] straight_2] strange_1] stress_1] swim_5] talent_1] tennis_2] ticket_3] tie_1]
topic_2] tour_1] toy_1] traffic_2] transport_1] wake_1] wash_3] wave_1] wild_1]
winner_1] writer_1] youth_1]

NGSL-3 [lems 83 : types 85 : tokens 90]

aggressive_1] apartment_1] aunt_1] bath_1] billion_1] bite_1] boot_1] bowl_1]
brush_1] button_1] cap_1] ceremony_1] champion_1] chest_1] coal_1] competitor_1]
compose_1] cousin_1] creative_1] dare_2] desert_1] designer_1] destruction_1]
disagree_1] entertainment_1] exam_3] exposure_1] faithfully_1] forever_1] formula_2]
fortune_1] frighten_1] furniture_1] grandmother_1] hero_1] hundred_1] hurry_1]
illegal_1] injure_1] invent_2] iron_1] keen_1] knee_1] laboratory_1] mad_1] million_1]
mirror_1] nervous_1] nose_1] ocean_1] opera_1] orange_1] pig_1] pitch_1] planet_1]
poem_1] possession_1] reckon_1] rescue_1] sail_1] scare_1] secondary_1] shower_1]
singer_1] sink_1] ski_3] specialist_1] sponsor_1] stair_1] strict_1] stupid_1]
supporter_1] teenager_1] tent_1] therapy_1] thousand_1] tooth_1] tournament_1]
trick_1] twin_1] upset_1] valley_1] weird_1]

NAWL [lems 36 : types 36 : tokens 37]

approximate_1] artistic_1] athletic_1] barrel_1] cheer_1] chemistry_1] chess_2]
cinema_1] composer_1] conference_1] diary_1] goat_1] homework_1] informal_1]
insect_1] kilometer_1] liter_1] memorize_1] metaphor_1] millimeter_1] monkey_1]
neat_1] noisy_1] non_1] photographic_1] pole_1] poster_1] questionnaire_1]
ridiculous_1] rope_1] semi_1] sensible_1] snake_1] solar_1] stadium_1] syllable_1]

OFFLIST: [?: types 217 : tokens 234]

ambitious_1] ambulance_2] ankle_1] arachnophobia_1] armchair_1] arrogant_1]
aspirin_2] backpack_1] balcony_1] ballet_2] banana_1] bandage_2] basketball_1]
blog_1] bookcase_1] born_1] brave_1] bruise_1] builder_1] bungalow_1] butterfly_1]
camel_1] captain_1] cathedral_1] celebrity_1] claustrophobia_1] coaster_1] colourful_1]
convertible_1] cottage_1] crisp_1] crossroads_1] crutches_1] cupboard_1] curry_1]
cute_1] dancer_1] daredevil_1] dentist_1] detached_2] detective_1] detector_1]
disabled_1] disorganized_1] dive_1] diver_1] diving_2] documentary_1] dolphin_1]
domesticated_1] drawers_1] elbow_1] elephant_1] english_1] enthusiastic_1] exotic_1]
finalist_1] firefighter_1] fluently_1] fond_1] french_1] frog_1] generous_1]
geography_1] granddaughter_1] grandfather_1] grandparent_1] grandson_1]
gymnasium_1] handball_1] helicopter_1] helmet_2] heroine_1] houseboat_1] iceland_1]
ict_1] id_1] impatient_1] instructor_1] intelligent_1] internet_1] inventor_1] ireland_1]
japanese_1] jeans_1] jewellery_1] judo_1] junction_1] jungle_1] karate_1] kart_1]
kayak_1] korea_1] lamp_1] laptop_1] lonely_1] lorry_1] loser_1] manga_1] mansion_1]
marathon_1] memorable_1] microwave_1] millennium_1] modest_1] monastery_1]
monk_1] moody_1] moped_1] motocross_1] motorbike_1] mpnumber_1] narrator_1]
naughty_1] nephew_1] niece_1] nil_1] nought_1] octopus_1] outgoing_1] painter_1]
parachute_1] paramedic_2] parrot_1] pavement_1] pe_1] peaceful_1] pedestrian_2]
phobia_1] phobic_1] piranha_1] plaster_2] polio_1] prince_1] prodigy_1]
programmer_1] prohibit_1] prosperity_1] pullover_1] purse_1] pyramid_1] queen_1]
rally_1] recite_1] referee_1] rhyme_1] robes_1] roller_1] rubbish_1] rugby_1] russia_1]
sandals_1] scary_1] scooter_1] scorpion_1] scuba_1] sentimental_1] shark_1] shy_1]
skateboard_2] skates_1] sling_1] snack_1] snorkel_1] snowboard_1] socks_2] sofa_1]
spectacular_1] spicy_1] spider_1] sprain_1] stately_1] sth_1] stretcher_1] stunt_2]
sunglasses_1] superstition_1] superstitious_1] supper_1] surf_1] surfboard_2] surfer_1]
swimmer_1] teddy_1] terraced_1] terrified_1] thunderstorm_1] tidy_1] tightrope_1]
toe_1] tracksuit_2] trainer_1] trainers_2] tram_1] trendy_1] triskaidekaphobia_1]

trophy_[1] tunnel_[1] turkey_[1] tv_[2] umbrella_[1] unambitious_[1] underground_[1]
underwater_[1] unfriendly_[1] unhelpful_[1] unimaginative_[1] unlucky_[1] unpleasant_[1]
vacuuming_[1] verse_[1] volcano_[1] wallet_[1] wardrobe_[1] wetsuit_[1] whale_[1] wrist_[1]

Appendix 4.4b - English Plus 2 (Hardy-Gould & Mellersh, 2011) (BNC-COCA)

Families List [↑]

Family [number of tokens]

BNC-COCA-1,000 Families: [fams 362 : types 424 : tokens 569]

a_[2] about_[6] accept_[1] across_[1] act_[2] advertise_[1] after_[1] afternoon_[1] against_[1] agree_[1] allow_[1] always_[1] amaze_[1] and_[1] angry_[1] another_[1] arm_[1] around_[1] arrive_[1] art_[1] as_[2] at_[2] aunt_[1] baby_[1] back_[1] bad_[1] bag_[1] base_[1] basic_[1] bath_[1] be_[2] beach_[1] bear_[2] beauty_[1] become_[2] bed_[2] behind_[1] bet_[1] between_[1] billion_[1] bird_[1] board_[1] boat_[1] bone_[1] book_[2] boring_[1] born_[1] bother_[1] bottom_[1] break_[3] breakfast_[1] brother_[1] build_[1] burn_[2] bus_[2] busy_[1] buy_[1] call_[1] car_[2] card_[3] care_[2] chair_[1] cheap_[1] child_[2] class_[1] clean_[3] clear_[1] climb_[2] clothes_[2] coat_[1] collect_[1] colour_[1] company_[2] computer_[1] cook_[2] corner_[1] cost_[1] country_[1] couple_[1] court_[1] cross_[2] cry_[1] cut_[1] dance_[3] danger_[2] dark_[1] daughter_[1] day_[1] death_[1] definite_[1] difficult_[1] dinner_[1] do_[7] dog_[2] down_[2] draw_[2] dream_[1] drive_[2] easy_[1] egg_[1] evening_[1] ever_[1] every_[1] expect_[1] fall_[3] family_[1] far_[1] farm_[1] fast_[1] father_[1] fear_[1] few_[2] final_[2] find_[1] finger_[1] first_[2] fish_[2] flat_[1] floor_[1] fly_[2] food_[1] foot_[1] football_[2] for_[1] forest_[1] forget_[1] forward_[1] friend_[5] fright_[1] from_[2] front_[2] fun_[1] game_[1] get_[2] glass_[2] go_[7] good_[1] grandfather_[5] group_[1] grow_[2] guess_[1] half_[1] hand_[1] happen_[1] happy_[1] hard_[1] hardly_[1] have_[2] head_[1] hear_[1] heavy_[1] help_[3] hide_[1] history_[1] holiday_[1] home_[4] horse_[1] hospital_[1] hour_[1] house_[4] how_[1] however_[1] human_[1] hundred_[1] hunt_[1] hurry_[1] husband_[1] ice_[1] imagine_[2] in_[3] interest_[2] internet_[1] into_[1] it_[1] job_[1] joke_[1] jump_[2] keep_[1] key_[3] kick_[1] kind_[1] king_[1] lake_[1] land_[1] last_[2] learn_[1] leave_[2] left_[1] leg_[1] light_[3] listen_[1] live_[1] look_[2] lose_[3] luck_[2] lunch_[1] machine_[1] mad_[1] make_[4] man_[1] manage_[1] marry_[1] meal_[1] mean_[1] meet_[2] middle_[1] million_[1] minute_[1] money_[1] month_[1] morning_[1] mother_[1] mountain_[1] move_[2] music_[4] naughty_[1] near_[1] neat_[1] neck_[1] need_[1] never_[1] next_[1] nice_[1] nobody_[1] noise_[1] nose_[1] note_[1] number_[2] of_[7] off_[2] officer_[1] often_[1] on_[6] orange_[1] out_[3] over_[1] owned_[1] pack_[1] paint_[2] parent_[1] party_[1] pass_[1] photograph_[2] picture_[1] plan_[1] play_[9] police_[1] price_[1] prince_[1] programme_[3] public_[1] quarter_[1] queen_[1] race_[3] rain_[1] read_[1] real_[1] record_[1] remember_[1] report_[1] rich_[1] ride_[2] right_[1] ring_[1] river_[1] road_[3] rock_[1] roll_[1] room_[3] rubbish_[1] rule_[1] run_[2] safe_[1] sail_[1] save_[1] scare_[2] school_[2] science_[1] sea_[1] seat_[1] second_[2] see_[1] send_[1] serious_[1] shirt_[1] shoe_[2] shop_[2] shoulder_[1] shout_[1] show_[1] shy_[1] side_[1] sing_[2] sister_[1] situation_[1] sleep_[1] slow_[1] smoke_[1] some_[1] son_[1] soon_[1] space_[1] special_[1] sport_[2] stage_[1] stand_[1] start_[2] stop_[1] story_[1] straight_[2] strange_[1] street_[1] study_[1] stupid_[1] support_[2] suppose_[2] swim_[6] table_[3] take_[5] teach_[1] team_[1] telephone_[2] television_[2] the_[14] then_[1] thing_[1] though_[1] thousand_[1] through_[1] throw_[1] tie_[1] to_[8] tooth_[1] top_[2] touch_[1] toward_[1] train_[6] travel_[1] turn_[1] under_[2] up_[5] usual_[1] view_[1] visit_[1] wake_[1] walk_[3] wash_[3] watch_[2] wave_[1] week_[1] well_[1] what_[2] wife_[1] wild_[1] win_[3] window_[1] work_[1] worry_[1] write_[3] wrong_[1] year_[1] you_[6]

BNC-COCA-2,000 Families: [fams 148 : types 156 : tokens 173]

aid_[1] apartment_[1] argue_[1] atmosphere_[1] attack_[1] attract_[1] average_[1] banana_[1]
band_[1] bike_[2] bite_[1] block_[1] boot_[1] bowl_[1] brand_[1] brave_[1] brush_[1]
button_[1] camera_[1] cap_[1] captain_[1] century_[1] champion_[1] cheer_[1] chest_[1]
coach_[1] coal_[1] common_[1] competition_[2] contact_[1] copy_[1] cottage_[1] cousin_[1]
crash_[1] create_[1] crisp_[1] crowd_[1] dare_[2] decision_[2] describe_[1] desert_[1]
design_[1] desk_[1] detect_[2] dive_[4] dozen_[1] drama_[1] drawer_[1] elephant_[1]
encourage_[1] enemy_[1] energy_[1] entertain_[1] equipment_[1] exam_[3] exercise_[1]
exist_[1] expose_[1] faith_[1] fan_[1] folk_[1] fortune_[1] frog_[1] furniture_[1] goal_[1]
hero_[1] immediate_[1] injure_[2] instruct_[1] jeans_[1] journey_[1] keen_[1] kilometre_[1]
knee_[1] laboratory_[1] lamp_[1] legal_[1] library_[1] lone_[1] magazine_[2] match_[1]
mathematics_[1] medical_[1] memory_[5] metal_[1] microwave_[1] mirror_[1] modern_[1]
monkey_[1] mood_[1] neither_[1] nervous_[1] non_[1] organize_[1] partner_[1] patient_[1]
peace_[1] pig_[1] pitch_[1] plane_[1] planet_[1] plastic_[1] pleasant_[1] poem_[1] pole_[1]
positive_[1] possess_[1] practical_[1] practise_[1] president_[1] profession_[1] rare_[1]
reckon_[1] recognize_[1] repeat_[1] ridiculous_[1] risk_[1] rope_[1] row_[2] score_[1]
season_[1] series_[1] shower_[1] similar_[1] sink_[1] site_[1] ski_[3] snake_[1] sock_[2]
spell_[1] stairs_[1] storm_[1] stress_[1] teenage_[1] tent_[1] ticket_[3] toe_[1] topic_[2]
tour_[1] toy_[1] tradition_[1] traffic_[2] trick_[1] twin_[1] university_[2] upset_[1] valley_[1]
weird_[1]

BNC-COCA-3,000 Families: [fams 53 : types 57 : tokens 63]

adopt_[1] aggressive_[1] approximate_[1] athlete_[1] celebrate_[1] ceremony_[1] compete_[3]
compose_[2] concert_[1] confer_[1] construct_[1] convert_[1] cycle_[3] decade_[1]
destruction_[1] disabled_[1] disagree_[1] enthusiastic_[1] essential_[2] formal_[1] formula_[2]
generous_[1] geography_[1] graduate_[1] ideal_[1] invent_[3] journalist_[1] label_[1]
mobile_[1] modest_[1] moral_[1] narrate_[1] negative_[1] net_[1] novel_[1] ocean_[1]
opera_[1] pave_[1] pilot_[1] prohibit_[1] prosper_[1] rally_[1] rescue_[1] sensible_[1]
sponsor_[1] strict_[1] talent_[1] tennis_[2] therapy_[1] tournament_[1] transport_[1] tunnel_[1]
youth_[1]

BNC-COCA-4,000 Families: [fams 45 : types 46 : tokens 50]

ambitious_[2] ambulance_[2] ankle_[1] arrogant_[1] ballet_[2] barrel_[1] bruise_[1]
butterfly_[1] cathedral_[1] celebrity_[1] chemistry_[1] cinema_[1] detach_[2] diary_[1]
documentary_[1] dolphin_[1] elbow_[1] exotic_[1] fond_[1] goat_[1] helicopter_[1] helmet_[2]
impatient_[1] informal_[1] insect_[1] intelligent_[1] jewellery_[1] jungle_[1] metaphor_[1]
nought_[1] poster_[1] questionnaire_[1] sentiment_[1] solar_[1] spectacular_[1] spice_[1]
stadium_[1] supper_[1] terrace_[1] terrify_[1] trophy_[1] verse_[1] volcano_[1] whale_[1]
wrist_[1]

BNC-COCA-5,000 Families: [fams 34 : types 35 : tokens 38]

balcony_[1] basketball_[1] cupboard_[1] dentist_[1] flats_[1] gymnasium_[1] id_[1]
junction_[1] mansion_[1] marathon_[1] memorable_[1] millennium_[1] monastery_[1]
monk_[1] nephew_[1] pedestrian_[2] plaster_[2] purse_[1] pyramid_[1] recite_[1] referee_[1]
robe_[1] shark_[1] shorts_[2] skate_[1] snack_[1] sofa_[1] spider_[1] surf_[2] tidy_[1]
umbrella_[1] vacuum_[1] wallet_[1] wardrobe_[1]

BNC-COCA-6,000 Families: [fams 20 : types 20 : tokens 22]

armchair_[1] bandage_[2] bungalow_[1] camel_[1] chess_[2] curry_[1] cute_[1] firefight_[1]
fluent_[1] heroine_[1] ironed_[1] kidding_[1] litre_[1] niece_[1] parachute_[1] parrot_[1]
rhyme_[1] rugby_[1] sling_[1] syllable_[1]

BNC-COCA-7,000 Families: [fams 8 : types 8 : tokens 10]

aspirin_[2] millimetre_[1] sandal_[1] stately_[1] stunt_[2] teddy_[1] tram_[1] trendy_[1]

BNC-COCA-8,000 Families: [fams 9 : types 9 : tokens 9]

crutch_[1] domesticate_[1] karate_[1] kayak_[1] lorry_[1] nil_[1] scoot_[1] stretcher_[1]
superstition_[1]

BNC-COCA-9,000 Families: [fams 11 : types 12 : tokens 13]

claustrophobia_[1] disorganized_[1] octopus_[1] paramedic_[2] phobia_[2] polio_[1]
prodigy_[1] scorpion_[1] semi_[1] sprain_[1] superstitious_[1]

BNC-COCA-10,000 Families: [fams 2 : types 2 : tokens 2]

coaster_[1] scuba_[1]

BNC-COCA-11,000 Families: [fams 3 : types 3 : tokens 3]

judo_[1] mope_[1] pullover_[1]

BNC-COCA-13,000 Families: [fams 2 : types 2 : tokens 2]

daredevil_[1] snorkel_[1]

BNC-COCA-14,000 Families: [fams 3 : types 3 : tokens 4]

blog_[1] piranha_[1] tracksuit_[2]

BNC-COCA-15,000 Families: [fams 2 : types 2 : tokens 2]

kart_[1] wetsuit_[1]

BNC-COCA-16,000 Families: [fams 1 : types 1 : tokens 1]

motocross_[1]

BNC-COCA-18,000 Families: [fams 1 : types 1 : tokens 1]

manga_[1]

BNC-COCA-24,000 Families: [fams 1 : types 1 : tokens 1]

arachnophobia_[1]

OFFLIST: [?: types 31 : tokens 33]

backpack_[1] bookcase_[1] crossroads_[1] english_[1] forever_[1] french_[1] handball_[1]
homework_[1] houseboat_[1] iceland_[1] ict_[1] ireland_[1] japanese_[1] korea_[1] laptop_[1]
motorbike_[1] mpnumber_[1] outgoing_[1] pe_[1] russia_[1] skateboard_[2] snowboard_[1]
sth_[1] sunglasses_[1] surfboard_[2] thunderstorm_[1] tightrope_[1] triskaidekophobia_[1]
turkey_[1] underground_[1] underwater_[1]

Appendix 4.5a - Messages 3 (Goodey et al., 2006) (NGSL)

Lemmas List [↑]

Lemma [number of tokens]

NGSL-1 [lems 240 : types 261 : tokens 546]

a_[1] about_[4] action_[1] after_[1] again_[1] age_[1] all_[2] almost_[1] and_[4] argue_[1]
art_[1] as_[1] ask_[1] at_[2] attack_[1] back_[1] bad_[1] be_[13] bed_[1] before_[3] begin_[1]
big_[1] blue_[1] box_[1] break_[1] bus_[1] business_[1] but_[1] call_[1] campaign_[1] can_[2]
capital_[1] car_[1] care_[1] cent_[1] century_[1] certainly_[1] character_[1] clearly_[1]
clock_[1] close_[1] competition_[1] computer_[2] could_[3] country_[1] couple_[1] course_[1]
create_[1] culture_[1] cut_[1] day_[2] department_[1] develop_[1] do_[15] dress_[1] each_[1]
early_[1] easily_[1] easy_[1] enter_[1] fall_[1] fast_[1] father_[2] feature_[1] feel_[2] few_[1]
film_[1] fine_[2] first_[2] fish_[1] fit_[1] fly_[1] for_[3] friend_[1] full_[2] get_[11] give_[1]
go_[3] goal_[1] good_[3] great_[2] half_[1] hand_[1] hard_[1] have_[9] he_[6] heart_[1]
high_[2] hit_[2] holiday_[1] home_[3] hour_[1] how_[4] i_[5] idea_[1] if_[1] image_[1]
imagine_[1] important_[1] improve_[1] in_[6] interest_[2] international_[1] into_[1] it_[6]
keep_[1] kind_[1] land_[1] language_[1] large_[1] last_[1] late_[1] lead_[1] leave_[1] let_[1]
letter_[2] like_[7] link_[1] listen_[1] live_[1] long_[1] look_[4] love_[1] machine_[1] main_[2]
make_[1] mark_[2] matter_[1] mean_[2] medium_[1] meet_[2] message_[1] mind_[1] miss_[1]
more_[1] most_[1] mr_[1] much_[1] nearly_[1] nice_[2] no_[1] normal_[1] not_[8] number_[1]
of_[5] off_[2] okay_[1] on_[6] one_[3] order_[1] other_[1] out_[2] over_[1] page_[3] park_[2]
pass_[1] pay_[1] per_[1] perform_[1] performance_[1] phone_[2] plant_[1] please_[5]
point_[1] popular_[1] problem_[1] produce_[2] protect_[1] put_[1] quarter_[2] question_[1]
quickly_[1] ready_[1] really_[1] record_[1] red_[1] relationship_[1] release_[1] right_[3]
ring_[1] room_[2] say_[2] school_[1] several_[1] shall_[2] shop_[2] short_[1] size_[2]
small_[1] so_[2] some_[1] someone_[3] sorry_[1] speak_[2] spend_[1] star_[1] start_[1]
stay_[1] stop_[2] store_[2] successful_[1] suddenly_[2] sure_[1] take_[5] talk_[1] tell_[1]
than_[1] thank_[4] that_[14] the_[9] then_[2] they_[1] think_[4] this_[2] time_[3] to_[89]
top_[1] touch_[1] trip_[1] try_[2] turn_[2] up_[1] very_[1] visit_[1] want_[1] we_[3] week_[1]
welcome_[1] well_[3] what_[5] when_[1] why_[1] will_[1] with_[4] work_[1] worry_[2]
wrong_[1] year_[1] yesterday_[1] you_[13]

NGSL-2 [lems 75 : types 77 : tokens 85]

abroad_[1] achievement_[1] adopt_[1] afraid_[1] aid_[1] amaze_[1] approximately_[1]
award_[1] boat_[1] branch_[1] breakfast_[1] carefully_[1] cheap_[1] chip_[4] climb_[1]
communicate_[1] connect_[2] crash_[1] dear_[1] definitely_[1] destroy_[1] engineer_[1]
exhibition_[1] expensive_[1] expert_[1] fault_[2] friendly_[1] gas_[1] glass_[2] gold_[1] hi_[1]
hurt_[1] ice_[1] illness_[1] independent_[1] instruction_[1] journey_[1] jump_[3] kiss_[1]
leg_[1] length_[1] lift_[1] lunch_[1] metal_[1] mobile_[1] museum_[1] occasion_[1]
ordinary_[1] pack_[1] plastic_[1] plate_[1] pop_[1] predict_[1] reply_[2] row_[1] scene_[1]
shake_[2] shout_[1] sick_[1] slowly_[1] software_[1] strange_[1] survive_[1] swim_[1]
terrible_[1] theater_[1] theme_[1] traffic_[1] unusual_[1] volunteer_[1] wash_[1] wed_[1]
wild_[1] wonderful_[1] wood_[1]

NGSL-3 [lems 51 : types 52 : tokens 56]

admire_[1] alarm_[1] album_[1] apologize_[1] awful_[1] badly_[1] billion_[1] boot_[1]
castle_[2] confident_[1] crazy_[1] cream_[1] custom_[1] exhaust_[1] fantastic_[2] five_[1]
friday_[1] gallery_[1] grass_[1] greet_[1] hello_[1] honest_[1] hundred_[1] knife_[2]
laboratory_[1] lazy_[1] leather_[1] menu_[1] million_[1] nineteen_[1] ocean_[1] potato_[1]
quietly_[1] sail_[2] salt_[2] secondary_[1] seven_[1] shelter_[1] silver_[1] singer_[1] sink_[1]
stomach_[1] stupid_[1] successfully_[1] survival_[1] thousand_[1] three_[1] throat_[1] two_[1]
upset_[1] web_[1]

NAWL [lems 15 : types 16 : tokens 17]

arrow_[1] bracket_[1] cinema_[2] clever_[1] click_[1] comma_[2] injection_[1] invert_[1]
nicely_[1] pardon_[1] plug_[1] prediction_[1] realistic_[1] robot_[1] stadium_[1]

OFFLIST: [?: types 129 : tokens 132]

ache_[1] adj_[1] adventurous_[1] ambitious_[1] ambulance_[1] american_[1] angrily_[1]
animated_[1] annoy_[1] annoyed_[1] anxiously_[1] apostrophe_[1] aquarium_[1] argentina_[1]
argentinian_[1] australia_[1] australian_[1] backache_[1] biography_[1] biscuits_[1] bow_[1]
britain_[1] british_[2] bungee_[1] burp_[1] cafeteria_[1] calmly_[1] canada_[1] canadian_[1]
canoeing_[1] cartoon_[1] cathedral_[1] classmate_[1] closet_[1] cookies_[1] cotton_[1]
crisps_[1] dessert_[1] dining_[1] diving_[1] earache_[1] elevator_[1] endangered_[1]
english_[2] exclamation_[1] faint_[1] fame_[1] fork_[1] france_[1] french_[1] fries_[1]
funeral_[1] generous_[1] goggles_[1] greece_[1] greek_[1] happily_[1] hatchet_[1]
homesick_[1] irritated_[1] italian_[1] italy_[1] jam_[1] japan_[1] japanese_[1] jones_[1]
jumper_[1] lick_[1] loudly_[1] luckily_[1] medal_[1] mexican_[1] mexico_[1] moody_[1]
mosque_[1] nod_[1] noisily_[1] palace_[1] pants_[1] pavement_[1] pepper_[1] petrol_[1]
plaise_[1] poland_[1] polish_[1] polite_[1] politely_[1] pollute_[1] rainforest_[1] rebel_[1]
sadly_[1] script_[1] scuba_[1] serviette_[2] shy_[1] sidewalk_[1] skateboarding_[1]
sneakers_[1] snowboarding_[1] socks_[1] sore_[1] spain_[1] spanish_[1] spear_[1] spoon_[1]
starter_[1] starving_[1] supergroup_[1] surfing_[1] swimsuit_[1] temple_[1] tidy_[1]
toothache_[1] tracksuit_[1] trainers_[1] trousers_[1] trunks_[1] unfriendly_[1] unkind_[1]
unplug_[1] unpopular_[1] untidy_[1] usa_[1] vacation_[1] wardrobe_[1] wetsuit_[1]
windsor_[1] worldwide_[1] zorbing_[1]

Appendix 4.5b - Messages 3 (Goodey et al., 2006) (BNC-COCA)

Families List [↑]

Family [number of tokens]

BNC-COCA-1,000 Families: [fams 278 : types 311 : tokens 602]

a_[1] about_[4] act_[1] afraid_[1] after_[1] again_[1] age_[1] all_[2] almost_[1] amaze_[1]
and_[4] angry_[1] art_[1] as_[1] ask_[1] at_[2] awful_[1] back_[1] bad_[2] be_[13] bed_[1]
before_[3] begin_[1] big_[1] billion_[1] blue_[1] boat_[1] box_[1] break_[1] breakfast_[1]
bus_[1] business_[1] but_[1] call_[1] can_[2] car_[1] care_[2] certain_[1] cheap_[1] chip_[4]
clear_[1] climb_[1] clock_[1] close_[1] computer_[2] could_[3] country_[1] couple_[1]
course_[1] crazy_[1] cut_[1] danger_[1] day_[2] dear_[1] definite_[1] do_[15] dress_[1]
each_[1] early_[1] easy_[2] enter_[1] expensive_[1] fall_[1] fast_[1] father_[2] feel_[2] few_[1]
film_[1] fine_[2] first_[2] fish_[1] fit_[1] five_[1] fly_[1] for_[3] friday_[1] friend_[3] full_[2]
gas_[1] get_[11] give_[1] glass_[2] go_[3] gold_[1] good_[3] grass_[1] great_[2] half_[1]
hand_[1] happy_[1] hard_[1] have_[9] he_[6] heart_[1] hello_[2] high_[2] hit_[2] holiday_[1]
home_[3] honest_[1] hour_[1] how_[4] hundred_[1] hurt_[1] i_[5] ice_[1] idea_[1] if_[1]
imagine_[1] important_[1] in_[6] interest_[2] into_[1] it_[6] jump_[4] keep_[1] kind_[1]
kiss_[1] land_[1] large_[1] last_[1] late_[1] lazy_[1] lead_[1] leave_[1] leg_[1] let_[1]
letter_[2] lift_[1] like_[7] listen_[1] live_[1] long_[1] look_[4] loud_[1] love_[1] luck_[1]
lunch_[1] machine_[1] main_[2] make_[1] mark_[2] matter_[1] mean_[2] meet_[2] million_[1]
mind_[1] miss_[1] mister_[1] more_[1] most_[1] much_[1] near_[1] nice_[3] nine_[1] no_[1]
noise_[1] normal_[1] not_[8] number_[1] of_[5] off_[2] ok_[1] on_[6] one_[3] order_[1]
other_[1] out_[2] over_[1] pack_[1] page_[3] pardon_[1] park_[2] pass_[1] pay_[1] plant_[1]
please_[5] point_[1] pop_[1] problem_[1] protect_[1] put_[1] quarter_[2] question_[1]
quick_[1] quiet_[1] ready_[1] real_[1] really_[1] record_[1] red_[1] relate_[1] reply_[2]
right_[3] ring_[1] room_[2] sad_[1] sail_[2] say_[2] school_[1] second_[1] seven_[1]
several_[1] shake_[2] shall_[2] shop_[2] shout_[1] shy_[1] sick_[1] sing_[1] size_[2] slow_[1]
small_[1] so_[2] some_[4] sorry_[1] speak_[2] spend_[1] star_[1] start_[2] stay_[1] stop_[2]
store_[2] strange_[1] stupid_[1] sudden_[2] sure_[1] swim_[1] take_[5] talk_[1] telephone_[2]
tell_[1] terrible_[1] than_[1] thank_[4] that_[14] the_[9] then_[2] they_[1] think_[4] this_[2]
thousand_[1] three_[1] throat_[1] time_[3] to_[89] top_[1] touch_[1] train_[1] trip_[1] try_[2]
turn_[2] two_[1] up_[1] usual_[1] very_[1] visit_[1] want_[1] wash_[1] we_[3] web_[1]
wed_[1] week_[1] well_[3] what_[5] when_[1] why_[1] wild_[1] will_[1] with_[4] wonder_[1]
wood_[1] work_[1] worry_[2] wrong_[1] year_[1] yesterday_[1] you_[13]

BNC-COCA-2,000 Families: [fams 86 : types 94 : tokens 100]

admire_[1] adventure_[1] aid_[1] alarm_[1] annoy_[2] anxious_[1] argue_[1] attack_[1]
biscuit_[1] boot_[1] bow_[1] branch_[1] calm_[1] canoe_[1] capital_[1] castle_[2] cent_[1]
century_[1] character_[1] clever_[1] competition_[1] connect_[2] cotton_[1] crash_[1]
cream_[1] create_[1] crisp_[1] culture_[1] department_[1] destroy_[1] develop_[1] dine_[1]
dive_[1] engineer_[1] exhaust_[1] faint_[1] fantastic_[2] fault_[2] feature_[1] fry_[1] goal_[1]
ill_[1] image_[1] improve_[1] instruct_[1] jam_[1] journey_[1] kindly_[1] knife_[2]
laboratory_[1] language_[1] length_[1] message_[1] metal_[1] mood_[1] occasion_[1]
ordinary_[1] per_[1] perform_[2] plastic_[1] plate_[1] plug_[2] polish_[1] polite_[2]
pollute_[1] popular_[2] potato_[1] produce_[2] release_[1] row_[1] salt_[2] scene_[1]
shelter_[1] silver_[1] sink_[1] sock_[1] sore_[1] starve_[1] stomach_[1] success_[2]
survive_[2] theatre_[1] traffic_[1] trunk_[1] upset_[1] welcome_[1]

BNC-COCA-3,000 Families: [fams 39 : types 40 : tokens 40]

abroad_[1] ache_[1] achieve_[1] adopt_[1] album_[1] apology_[1] approximate_[1] award_[1]
campaign_[1] communicate_[1] confident_[1] custom_[1] elevate_[1] exhibit_[1] expert_[1]
funeral_[1] gallery_[1] generous_[1] greet_[1] independent_[1] inject_[1] international_[1]
leather_[1] link_[1] medium_[1] menu_[1] mobile_[1] museum_[1] nod_[1] ocean_[1]

palace_[1] pave_[1] pepper_[1] petrol_[1] predict_[2] rebel_[1] software_[1] theme_[1]
volunteer_[1]

BNC-COCA-4,000 Families: [fams 21 : types 21 : tokens 22]

ambitious_[1] ambulance_[1] animate_[1] arrow_[1] biography_[1] cartoon_[1] cathedral_[1]
cinema_[2] click_[1] exclaim_[1] fame_[1] fork_[1] irritate_[1] lick_[1] medal_[1] robot_[1]
script_[1] spoon_[1] stadium_[1] temple_[1] trousers_[1]

BNC-COCA-5,000 Families: [fams 11 : types 12 : tokens 12]

aquarium_[1] bracket_[1] dessert_[1] pants_[1] shorts_[1] sneak_[1] spear_[1] surf_[1] tidy_[2]
vacation_[1] wardrobe_[1]

BNC-COCA-6,000 Families: [fams 3 : types 3 : tokens 3]

closet_[1] invert_[1] mosque_[1]

BNC-COCA-7,000 Families: [fams 1 : types 1 : tokens 1]

cookie_[1]

BNC-COCA-8,000 Families: [fams 2 : types 2 : tokens 2]

cafeteria_[1] goggle_[1]

BNC-COCA-9,000 Families: [fams 1 : types 2 : tokens 2]

comma_[2]

BNC-COCA-10,000 Families: [fams 1 : types 1 : tokens 1]

scuba_[1]

BNC-COCA-11,000 Families: [fams 3 : types 3 : tokens 3]

apostrophe_[1] burp_[1] hatchet_[1]

BNC-COCA-13,000 Families: [fams 1 : types 1 : tokens 1]

bungee_[1]

BNC-COCA-14,000 Families: [fams 1 : types 1 : tokens 1]

tracksuit_[1]

BNC-COCA-15,000 Families: [fams 1 : types 1 : tokens 1]

wetsuit_[1]

BNC-COCA-16,000 Families: [fams 1 : types 1 : tokens 1]

plaiice_[1]

BNC-COCA-19,000 Families: [fams 2 : types 2 : tokens 3]

serviette_[2] supergroup_[1]

OFFLIST: [?: types 40 : tokens 41]

adj_[1] american_[1] argentina_[1] argentinian_[1] australia_[1] australian_[1] backache_[1]
britain_[1] british_[2] canada_[1] canadian_[1] classmate_[1] earache_[1] english_[2]
france_[1] french_[1] greece_[1] greek_[1] homesick_[1] italian_[1] italy_[1] japan_[1]
japanese_[1] jones_[1] mexican_[1] mexico_[1] poland_[1] rainforest_[1] sidewalk_[1]
skateboarding_[1] snowboarding_[1] spain_[1] spanish_[1] swimsuit_[1] toothache_[1] usa_[1]
windsor_[1] worldwide_[1] zorbing_[1]

Appendix 4.6a - Angličtina 7 Way to Win (Betáková & Dvořáková, 2006) (NGSL)

Lemmas List [↑]

Lemma [number of tokens]

NGSL-1 [lems 151 : types 158 : tokens 289]

about_[1] after_[1] against_[2] all_[1] almost_[2] area_[2] ask_[1] at_[3] away_[2] back_[1]
bad_[1] be_[3] break_[2] campaign_[1] can_[2] car_[1] cause_[2] check_[2] choice_[2]
choose_[2] come_[1] competition_[2] control_[1] cost_[2] course_[1] die_[1] do_[5] dog_[2]
down_[4] dress_[2] drink_[2] drop_[1] each_[2] either_[1] end_[2] energy_[3] environment_[2]
ever_[2] fall_[4] favorite_[2] feel_[2] fire_[2] fish_[1] for_[3] force_[3] forget_[1] front_[1]
future_[2] go_[1] great_[1] grow_[1] hard_[2] have_[2] he_[1] help_[1] how_[1] i_[2] idea_[1]
if_[1] important_[2] include_[2] inside_[1] into_[1] it_[5] kill_[1] last_[1] let_[1] letter_[1]
lie_[2] like_[2] list_[2] look_[1] love_[1] machine_[2] make_[1] match_[1] mean_[1]
message_[1] model_[2] move_[2] music_[2] never_[2] newspaper_[2] next_[2] night_[1]
not_[4] of_[2] off_[2] officer_[2] on_[5] once_[1] out_[3] outside_[1] page_[1] part_[1]
party_[4] phone_[1] player_[1] point_[1] police_[4] power_[2] put_[2] race_[2] reach_[1]
really_[2] record_[3] report_[2] science_[1] screen_[2] service_[3] she_[2] shop_[2] site_[1]
somebody_[1] something_[3] sorry_[1] sort_[1] space_[2] station_[5] study_[1] suddenly_[2]
sure_[2] term_[2] that_[2] the_[7] think_[1] throw_[2] time_[2] to_[8] too_[2] top_[1]
touch_[1] toward_[1] track_[1] train_[2] turn_[4] understand_[1] up_[1] walk_[2] want_[1]
water_[2] we_[1] what_[1] who_[1] why_[1] win_[2] with_[1] worry_[4] would_[3] write_[3]
you_[6]

NGSL-2 [lems 69 : types 69 : tokens 107]

accident_[2] alternative_[2] angry_[2] arrange_[2] assistant_[2] bike_[1] blow_[1] bone_[2]
burn_[1] careful_[2] celebrate_[1] coach_[2] coast_[3] cool_[1] count_[3] crash_[2] cry_[1]
cycle_[1] dangerous_[2] destroy_[1] electronic_[1] emergency_[2] engine_[1]
environmental_[1] factory_[2] famous_[1] friendly_[1] glass_[2] hate_[2] hurt_[2] injury_[2]
instrument_[1] invite_[1] leg_[1] lesson_[1] library_[2] load_[2] lock_[1] lovely_[1] meal_[2]
metal_[2] musical_[1] narrow_[1] neighbor_[1] net_[1] noise_[2] pain_[1] plastic_[2] plate_[2]
presentation_[2] print_[2] quick_[2] radio_[1] rain_[1] remark_[1] ride_[1] sad_[2] safe_[2]
safety_[2] shake_[1] shock_[1] shout_[2] stick_[1] studio_[2] volunteer_[1] waste_[2] wild_[2]
wind_[2] winner_[1]

NGSL-3 [lems 28 : types 28 : tokens 43]

adventure_[2] bath_[1] brilliant_[3] cake_[1] chicken_[1] dust_[2] exhaust_[2] fancy_[2]
fantastic_[1] festival_[2] flash_[1] float_[1] fortune_[1] giant_[1] harbor_[1] nervous_[2]
pen_[1] pipe_[2] pollution_[1] potato_[2] rescue_[1] sail_[2] seal_[4] shower_[1] sink_[1]
van_[1] web_[2] whisper_[1]

NAWL [lems 10 : types 10 : tokens 12]

acid_[1] biology_[1] cheat_[1] diagram_[1] harvest_[1] horizon_[1] poster_[1] quiz_[2]
shuttle_[1] solar_[2]

OFFLIST: [?: types 61 : tokens 84]

ambulance_[2] anorak_[1] anticlockwise_[1] asleep_[2] bonfire_[1] brigade_[2] browse_[1]
butter_[2] captain_[1] cd_[1] clockwise_[1] dice_[1] disabled_[1] dock_[2] dolphin_[1]
embarrassed_[1] ferry_[2] fetch_[2] fireworks_[2] fisherman_[1] fishermen_[1] flare_[1]
grapes_[2] helmet_[1] hurricane_[1] indian_[1] inline_[1] internet_[1] jewellery_[1]
librarian_[2] lifeboat_[2] lifejacket_[2] mayday_[1] organizer_[2] overtake_[1] pal_[1]
parade_[1] pavement_[2] peanut_[2] penfriend_[1] pollute_[1] raft_[2] rattle_[1] recycle_[1]
recycling_[1] riddle_[1] rocket_[1] rubbish_[2] sailor_[2] sanctuary_[4] sausage_[2]
schoolmate_[2] settler_[1] sincerely_[1] skates_[1] spacesuit_[2] teller_[1] thanksgiving_[1]
tragedy_[1] tuna_[1]

Appendix 4.6b - Angličtina 7 Way to Win (Betáková & Dvořáková, 2006) (BNC-COCA)

Families List [↑]

Family [number of tokens]

BNC-COCA-1,000 Families: [fams 182 : types 195 : tokens 343]

about_[1] after_[1] against_[2] all_[1] almost_[2] angry_[2] area_[2] arrange_[2] ask_[1] at_[3] away_[2] back_[1] bad_[1] bath_[1] be_[3] blow_[1] bone_[2] break_[2] burn_[1] cake_[1] can_[2] car_[1] care_[2] cause_[2] check_[2] chicken_[1] choice_[2] choose_[2] come_[1] control_[1] cool_[1] cost_[2] count_[3] course_[1] cry_[1] danger_[2] die_[1] do_[5] dog_[2] down_[4] dress_[2] drink_[2] drop_[1] each_[2] either_[1] end_[2] engine_[1] ever_[2] fall_[4] favourite_[2] feel_[2] fire_[2] fish_[1] for_[3] force_[3] forget_[1] friend_[1] front_[1] glass_[2] go_[1] great_[1] grow_[1] hard_[2] hate_[2] have_[2] he_[1] help_[1] how_[1] hurt_[2] i_[2] idea_[1] if_[1] important_[2] inside_[1] internet_[1] into_[1] it_[5] kill_[1] last_[1] leg_[1] let_[1] letter_[1] lie_[2] like_[2] list_[2] load_[2] lock_[1] look_[1] love_[1] lovely_[1] machine_[2] make_[1] meal_[2] mean_[1] move_[2] music_[3] neighbour_[1] never_[2] next_[2] night_[1] noise_[2] not_[4] of_[2] off_[2] officer_[2] on_[5] once_[1] out_[4] page_[1] pain_[1] part_[1] party_[4] play_[1] point_[1] police_[4] power_[2] present_[2] put_[2] quick_[2] race_[2] radio_[1] rain_[1] reach_[1] really_[2] record_[3] report_[2] ride_[1] rubbish_[2] sad_[2] safe_[4] sail_[4] science_[1] service_[3] settle_[1] shake_[1] she_[2] shop_[2] shout_[2] some_[4] sorry_[1] sort_[1] space_[2] station_[5] stick_[1] study_[1] sudden_[2] sure_[2] telephone_[1] term_[2] that_[2] the_[7] think_[1] throw_[2] time_[2] to_[8] too_[2] top_[1] touch_[1] toward_[1] track_[1] train_[2] turn_[4] understand_[1] up_[1] van_[1] walk_[2] want_[1] waste_[2] water_[2] we_[1] web_[2] what_[1] who_[1] why_[1] wild_[2] win_[3] wind_[2] with_[1] worry_[4] would_[3] write_[3] you_[6]

BNC-COCA-2,000 Families: [fams 56 : types 58 : tokens 97]

accident_[2] adventure_[2] asleep_[2] assist_[2] bike_[1] brilliant_[3] butter_[2] captain_[1] cheat_[1] coach_[2] coast_[3] competition_[2] crash_[2] destroy_[1] dust_[2] energy_[3] environment_[3] exhaust_[2] famous_[1] fancy_[2] fantastic_[1] fetch_[2] flash_[1] float_[1] fortune_[1] future_[2] giant_[1] include_[2] injure_[2] instrument_[1] invite_[1] lesson_[1] library_[4] match_[1] message_[1] metal_[2] model_[2] narrow_[1] nervous_[2] newspaper_[2] organize_[2] pen_[1] pipe_[2] plastic_[2] plate_[2] pollute_[1] potato_[2] print_[2] remark_[1] sausage_[2] screen_[2] seal_[4] shock_[1] shower_[1] sink_[1] site_[1]

BNC-COCA-3,000 Families: [fams 17 : types 17 : tokens 23]

alternative_[2] campaign_[1] celebrate_[1] cycle_[1] disabled_[1] electronic_[1] emergency_[2] factory_[2] festival_[2] harbor_[1] net_[1] pave_[2] rescue_[1] studio_[2] tragedy_[1] volunteer_[1] whisper_[1]

BNC-COCA-4,000 Families: [fams 19 : types 20 : tokens 24]

acid_[1] ambulance_[2] biology_[1] diagram_[1] dock_[2] dolphin_[1] ferry_[2] harvest_[1] helmet_[1] horizon_[1] hurricane_[1] jewellery_[1] parade_[1] poster_[1] rattle_[1] recycle_[2] rocket_[1] sincere_[1] solar_[2]

BNC-COCA-5,000 Families: [fams 8 : types 8 : tokens 13]

brigade_[2] dice_[1] flare_[1] grape_[2] pal_[1] sanctuary_[4] shuttle_[1] skate_[1]

BNC-COCA-6,000 Families: [fams 8 : types 8 : tokens 12]

browse_[1] firework_[2] overtake_[1] peanut_[2] quiz_[2] raft_[2] riddle_[1] tuna_[1]

BNC-COCA-7,000 Families: [fams 3 : types 4 : tokens 4]

bonfire_[1] clockwise_[2] teller_[1]

BNC-COCA-12,000 Families: [fams 1 : types 1 : tokens 1]

anorak_[1]

BNC-COCA-14,000 Families: [fams 1 : types 1 : tokens 1]

mayday_[1]

OFFLIST: [?: types 14 : tokens 17]

cd_[1] embarrassed_[1] fisherman_[1] fishermen_[1] indian_[1] inline_[1] lifeboat_[2]
lifejacket_[2] penfriend_[1] pollute_[1] schoolmate_[2] spacesuit_[2] thanksgiving_[1]

Annotation

Jméno a příjmení:	Vojtěch Přecechtěl
Katedra:	Ústav cizích jazyků
Vedoucí práce:	Doc. PhDr. Václav Řeřicha, CSc.
Rok obhajoby:	2016

Název práce:	Frekvence anglické slovní zásoby a její využití na 2. stupni ZŠ
Název v angličtině:	English Vocabulary Frequency and Its Use at Lower Secondary Schools
Anotace práce:	Diplomová práce pojednává o možném vlivu frekvence slovní zásoby na výuku angličtiny jako cizího jazyka a slouží jako úvod do této problematiky. Frekvenční slovníky pro obecné potřeby poskytují oporu při analýze textu a s nimi spojené nástroje mohou být používány k posouzení jazykové úrovně textu, k nalezení problémových částí textu a tvoření nových výukových materiálů, které by využívaly této objektivní metody. Předmětem výzkumu této práce jsou slovníčky obsažené v učebnicích angličtiny. V práci jsou uvedeny informace o použité metodologii a srovnání šesti slovníčků z učebnic anglického jazyka, které se používají pro výuku na druhém stupni základních škol. Cílem tohoto výzkumu je poukázat na frekvenci slovní zásoby jako jednoho z mnoha aspektů slovní zásoby a nastínit jeho význam ve výuce anglického jazyka.
Klíčová slova:	slovní zásoba, frekvence slovní zásoby, korpus, frekvenční seznamy slov, analýza textu, druhý stupeň základní školy, výuka, výuka angličtiny jako cizího jazyka, výuka jazyka, RVP ZV, výukové materiály a aplikace
Anotace v angličtině:	Purpose of this thesis is to discuss the possible influence of the vocabulary frequency aspect on EFL teaching. The general service lists provide support for text level analysis and teachers can use the related tools in order to classify the language level of texts, find out the text difficulties and create new learning materials based on this objective method. The subject of the research in this piece of work is an English textbook wordlist. The article shows methodology and compares data of six textbook wordlists' analyses used amongst the Czech lower secondary school learners of the

	English language. The aim of the research is to point out the frequency aspect presence amongst other vocabulary aspects and discuss its meaning for EFL teaching.
Klíčová slova v angličtině:	vocabulary, vocabulary frequency, corpus, general service list, text analysis, lower secondary schools, education, EFL, language learning, CEFR, learning materials and applications
Přílohy vázané v práci:	Appendix 1 – Word family levels Appendix 2 – Images Appendix 3 – NGSL frequency division overview Appendix 4 – Vocabulary distribution details CD
Rozsah práce:	66 stran
Jazyk práce:	Angličtina

Resumé

Magisterská diplomová práce pojednává o možném vlivu frekvence slovní zásoby na výuku angličtiny jako cizího jazyka se zaměřením na žáky druhého stupně základních škol. V první polovině teoretické části práce jsou prezentovány poznatky a znalosti potřebné k uvedení do problematiky frekvence slov, zabývá se tématy, jako je slovní zásoba, korpus, frekvenční slovníky a analýza slovní zásoby. V navazující teoretické části práce se prohlubuje zaměření na cílovou skupinu žáků, jak úroveň jazyka, tak uvedením vybraných učebních strategií a materiálů zaměřených na získávání slovní zásoby. Poslední část práce se věnuje zejména výzkumu přítomnosti a vlivu frekvence slovní zásoby na její výuku na druhém stupni základních škol. Předmětem zkoumání jsou slovníčky z učebnic angličtiny, které se používají na českých školách. Je sledována distribuce položek slovníčků v jednotlivých frekvenčních úrovních. Výstupem výzkumu je vzájemné srovnání slovníčků a také jejich porovnání s poznatky teoretické části práce, založené na aspektu frekvence slovní zásoby.