

DEPARTMENT OF ENGLISH LANGUAGE TEACHING

## **Collocations in Czech pop culture**

Seminar work

Název předmětu: Lexikologie a frazeologie CJP/6PLEX

Akademický rok: 2023/2024

Obor: Anglický jazyk se zaměřením na vzdělávání

Vyučující: Mgr. Gabriela Entlová, Ph.D.

2.ročník / 2.semestr

Jméno studenta: Prokop

Golláš

číslo studenta:

D22129

I state that this submitted thesis is my original work and that I elaborated it myself. I cited all references and other sources that I used to work up the thesis. Those references and other sources are given in the list of references at the end.

V Bohumíně, 3.4.2023

Prokop Gollán

# CONTENTS

<b>CONTENTS .....</b>	<b>3</b>
<b>INTRODUCTION .....</b>	<b>4</b>
<b>1 THEORETICAL PART .....</b>	<b>4</b>
1.1 Collocations.....	4
1.1.1 The definition of collocations .....	4
1.1.2 Types of collocations .....	5
1.1.3 Examples of collocations in Czech and English.....	6
<b>2 PRACTICAL PART .....</b>	<b>7</b>
2.1 Articles .....	7
2.2 Corpora .....	8
2.3 Collocations of lexical units.....	8
<b>CONCLUSION .....</b>	<b>13</b>
<b>LITERATURE.....</b>	<b>14</b>

# INTRODUCTION

In this semestral project, I will be writing about a lexical topic – collocations. More precisely collocations that are used in articles about Czech pop culture and the music industry. This work will mainly focus on topic-specific words and collocations and compare their usage and frequency per million words with British National Corpus (BNC) and Corpus of Contemporary American English (COCA). The comparison of frequency per million words will be processed by using Sketch Engine (an online tool to create corpuses) to make my own corpus using topic-related articles and manually analysing the three mentioned corpuses and the chosen collocations between them.

The seminar work will contain two main parts. The theoretical part will describe collocations and its definition and types of collocations, and the practical part will focus on the description of chosen articles and corpuses and will contain graphs and tables of the analysed collocations to better visualize the word frequency per million words.

I chose this topic because I find the way collocations make one word have a completely different meaning or multiple meanings interesting. The topics were chosen because I am interested in the Czech music industry, and I will use the information from this seminar paper in my bachelor thesis.

## 1 THEORETICAL PART

As I already mentioned in the introduction, in this part I will focus on the definition of collocations, an explanation of what collocations are using multiple examples, and list of different types of collocations.

### 1.1 Collocations

#### 1.1.1 The definition of collocations

According to Kvetko (2009), “the term *collocation* is most frequently used in the following meanings:

- A **tendency** to (the likelihood of) co-occurrence, i.e. the way in which some words regularly combine (collocate) with other words (natural to native

speakers), e.g.: *commit* collates with *crime*, *blond* with *hair* (not a \*blond car);

- A **combination** of words that regularly occur together, combines in a predictable way, has a relatively restricted combinatory range and transparent meaning: *submit an application/a proposal, flock of sheep, interested in.*” (p. 101)

According to Peprník (1992), “A collocation is an established combination of words with a single meaning so that they form a lexical unit. This characteristic distinguishes it from a free association, with which it has the same syntactic structure. The stability of a collocation is demonstrated not only by its consistent meaning but also by the interchangeability of its individual components with synonyms. For example, the adjective in collocation water wheel cannot be interchanged by another, whereas in the free association broken wheel the adjective can be replaced with damaged, destroyed, ruined, etc.” (my own translation, original in Czech language on p. 49)

### 1.1.2 Types of collocations

As in many lexical topics, multiple linguists have different opinions and different views on various definitions, classifications, and types. Collocations are no exception. I have found diverse opinions on types of collocations through the internet and linguistic books. The most normalized and used types of collocations are **grammatical** and **lexical** collocations. Collocation is classed as **grammatical** if the main (dominant) word is followed by a grammatical word: *look forward to*, *consist of*. Collocation classed as **lexical** must consist of at least two equal components: *catch a cold*, *heavy wind*. (Kvetko, 2009)

Collocations occur in almost every language regularly as Kvetko (2009) explains “regardless of different views and classifications, languages differ (independently of grammatical structures) in their word combinations. It means that different language communities may combine different words, or stress different features of the extralinguistic reality” (p. 102)

### 1.1.3 Examples of collocations in Czech and English

Collocations are widely used both in Czech and English language and I will list a few collocations and the translation in each language for a better understanding of collocations.

#### **Czech collocations:**

- Dělat rozhodnutí – to make decision (literal translation do decision)
- Dělat si starosti – to worry (literal translation do worry)
- Dát pozor – to pay attention (literal translation give attention)

#### **English collocations:**

- Clever trick – well executed/thought out trick
- Cold war - an ongoing political rivalry between the United States and the Soviet Union and their respective allies that developed after World War II.
- Bitter cold - extremely and unpleasantly cold

Someone might find collocations and idioms confusing so here is a list of a few **English idioms** to clear up any confusion:

- Hit the books – to start studying
- Kick the bucket – to die
- Break the ice – to initiate a conversation

As the examples show, there is clear pattern in which can be seen that collocations have a natural similarity in literal meaning and true (lexical) meaning while idioms differ completely in literal translation and lexical meaning.

## 2 PRACTICAL PART

As I already mentioned in the introduction, in the practical part I will work with an online tool called Sketch Engine which is a multi-purpose tool focusing on corpuses and the word within the corpuses. I made a corpus of three different articles about music, pop culture, and the Czech music industry. This corpus will be comparing specific collocations and frequency of word usage per million words. This frequency will be also used to analyse the specific collocations in my ‘Czech pop culture’ related corpus and BNC plus COCA. The main goal of this analysis is to determine whether some word combinations (collocations) are used more in the corpus I made and British National Corpus and Corpus of Contemporary American English.

The practical part will include a summary of each article I chose to make the Czech pop culture corpus, describe the mentioned corpuses, and the graphs and charts of collocations.

In the end, I will also make a conclusion on this topic where I will describe the findings and observations, I made during working on this seminar work.

### 2.1 Articles

For the articles, I chose these three articles that combine into the ‘**Czech pop culture**’ corpus which makes the base and most important part of this seminar work. These are the three articles with a brief description and the date they were published:

1. **Musically mapped: Czech popular music as a second ‘world sound’** by C. Michael Elavsky was published in a major international journal *European Journal of Cultural Studies* in 2011. This article is about the development of Czech popular music in the global perspective.
2. **The music mainstream in communism revisited: a corpus analysis of Czechoslovak pop lyrics (1962–1991)** by J. Blüml was published in *Journal of Contemporary Central and Eastern Europe* in 2022. Based on the name of this article, this article is a corpus analysis of Czechoslovak pop lyrics from 1962 to 1991. This corpus analysis was made by Sketch Engine.
3. **Cultural memory, underground and mainstream culture reminiscences in the Czech Republic** by Z. Skořepová was published in *Ethnomusicology Forum* in 2019. This article interviews two contemporary artists, the underground

singer Dáša Vokatá and the popular actor Oldřich Kaiser. This article looks into the underground scene in late socialism.

## 2.2 Corporuses

1. **Czech pop culture** is a corpus made by me using the great online tool Sketch Engine. I uploaded the three mentioned articles: Musically mapped: Czech popular music as a second ‘world sound’ by C. Michael Elavsky, The music mainstream in communism revisited: a corpus analysis of Czechoslovak pop lyrics (1962-1991) by J. Blüml, Cultural memory, underground and mainstream culture reminiscences in the Czech Republic by Z. Skořepová. The corpus contains 32,068 words.
2. **British national corpus (BNC)** formerly known as the ‘BYU Corpora’ was originally created by Oxford University press. It contains 100 million words of text from a wide range (spoken, magazines, newspaper, academic, etc.)
3. **Corpus of Contemporary American English (COCA)** is composed of more than one billion words in 458,179 texts of multiple genres (TV and Movies subtitles, spoken, fiction, magazines, etc.)

## 2.3 Collocations of lexical units

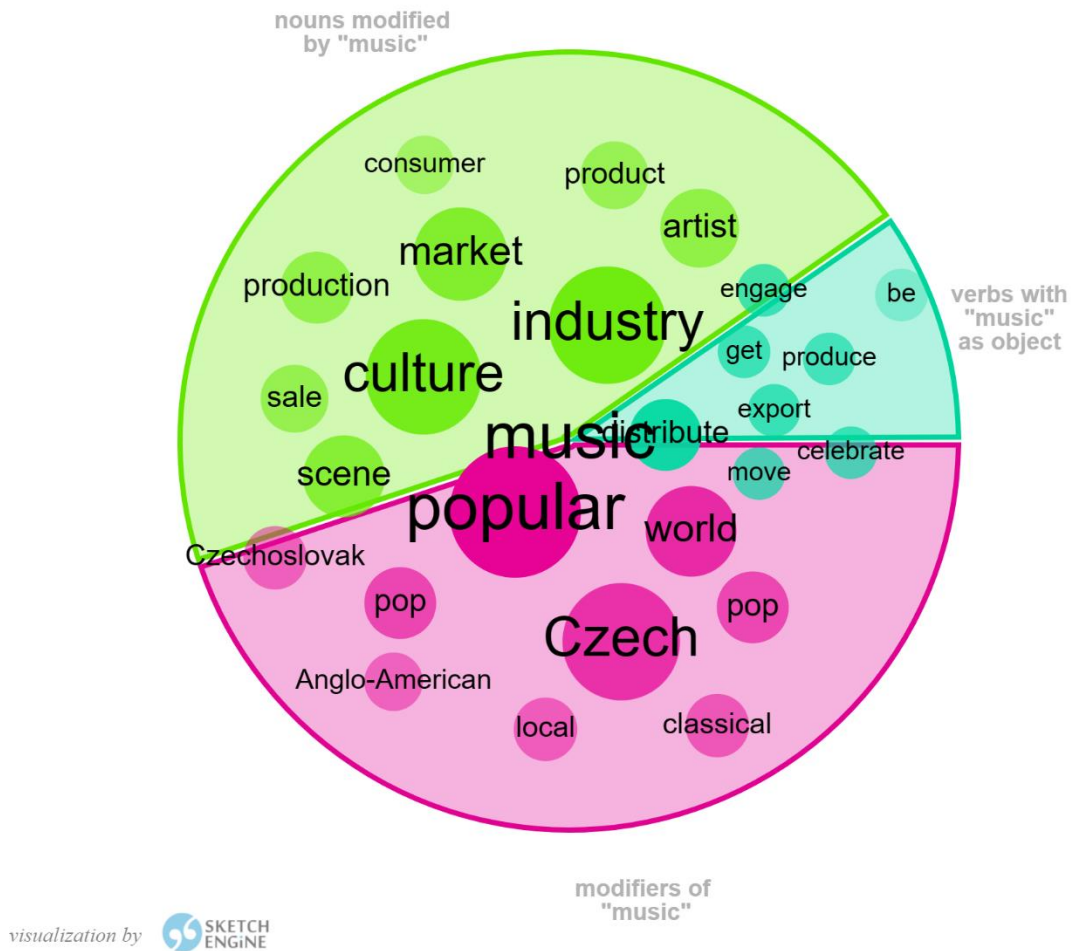
In the last part of my seminar paper, I analysed four words that relate to the topic ‘Czech pop culture’ and are commonly used with modifiers – making collocations. I tried to choose the most fitting words and while browsing Sketch Engine I found the words ‘music’, ‘song’, ‘artist’, and ‘industry’ were used regularly and had a decent number of interesting modifiers.

The tables below the generated graphs show the modifier and its frequency per million words in my corpus, BNC, and COCA. The words are listed by the highest to the lowest frequency per million words in my corpus.

The frequency per million words was calculated by dividing the number each word was found in the corpus by the number of all the words in the corpus and multiplying it by one million.



### 2.3.1 Collocations of ‘music’

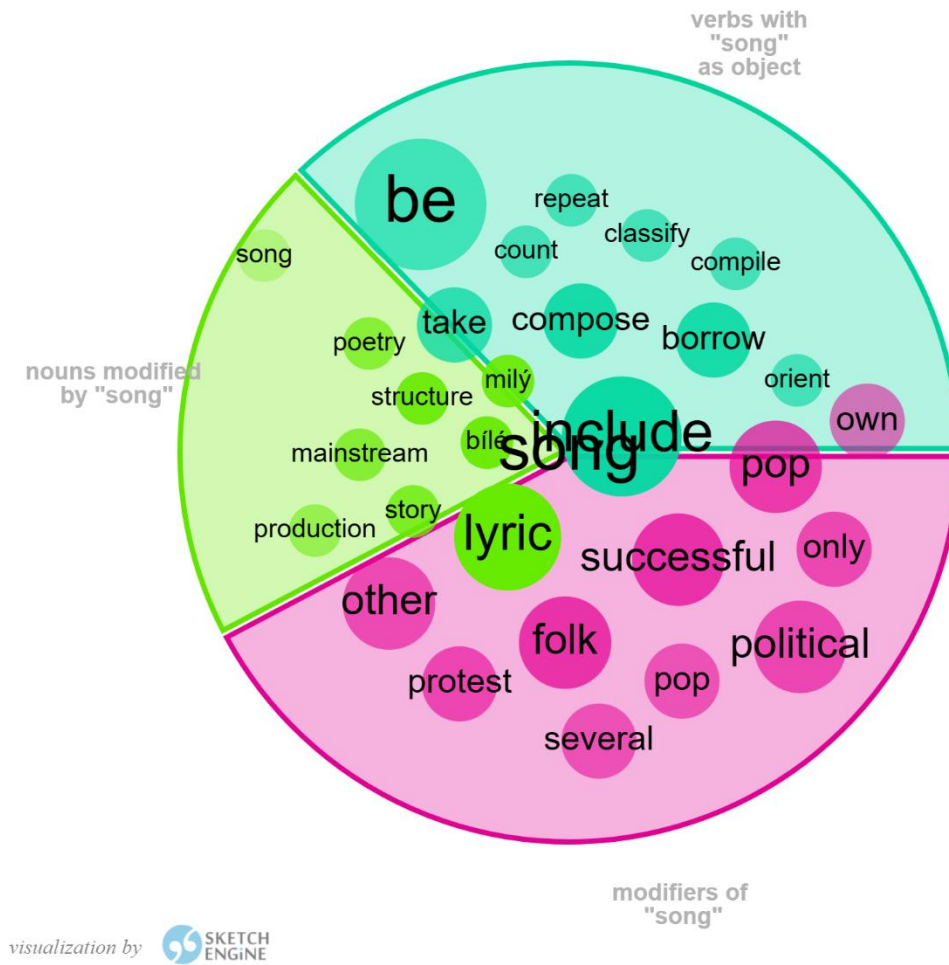


Collocations of ‘music’	Frequency pmw in my corpus	Frequency pmw in BNC	Frequency pmw in COCA
popular	997,88	2,43	2,28
Czech	748,41	0,01	0,01
industry	748,41	1,20	2,27
culture	717,23	0,35	0,81
market	405,39	0,11	0,20

pmw = per million words

This graph investigates the collocations of the word ‘music’. As I expected, the most dominant variations of collocations include the modifiers ‘popular’ and ‘Czech’. This dominance is due to my corpus related to Czech pop culture. I was not surprised by the fact that the collocation ‘popular music’ is most frequently used in all three corpora as it is a globally used term. If I skip the modifier ‘Czech’ as it is too specific, the least frequent collocation is with ‘music’ and ‘market’.

### 2.3.2 Collocations of ‘song’



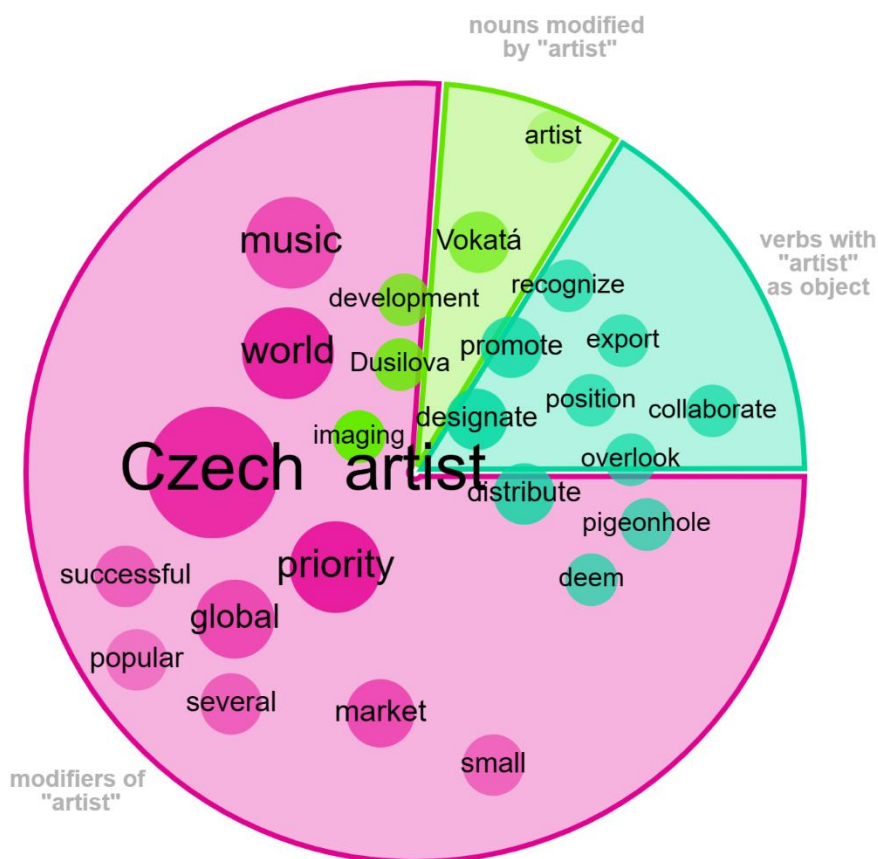
visualization by  SKETCH ENGINE

Collocations of ‘song’	Frequency pmw in my corpus	Frequency pmw in BNC	Frequency pmw in COCA
include	155,92	0,05	0,07
lyric	124,73	0,10	1,03
successful	93,56	0,03	0,03
pop	93,56	0,36	0,41
political	93,56	0,03	0,05

pmw = per million words

This graph focuses on the collocations of ‘song’. The most frequently used collocation in my corpus contains the verb ‘include’ with ‘song’ as an object (e.g. including songs, including pop songs). I did not count the verb ‘be’ as it is not used as a collocation in any of the contexts in my corpus. What I was the most surprised about was the word ‘lyric’ being a noun modified by ‘song’. The table shows that in COCA this collocation is much more frequent than in BNC.

### 2.3.3 Collocations of ‘artist’



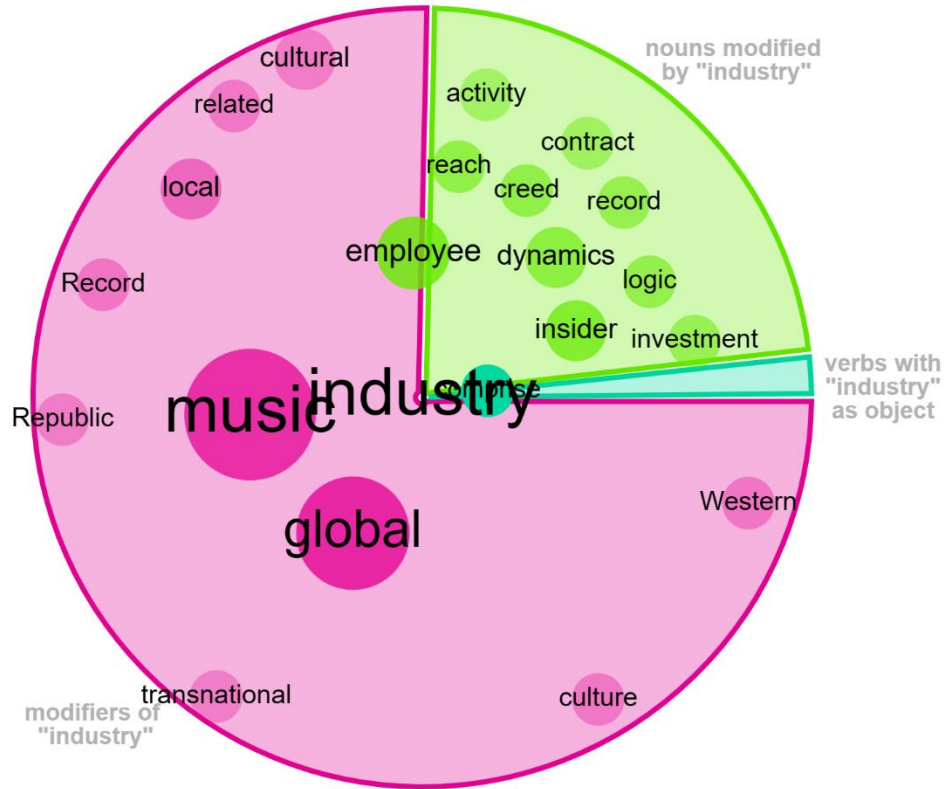
visualization by  SKETCH ENGINE

Collocations of ‘artist’	Frequency pmw in my corpus	Frequency pmw in BNC	Frequency pmw in COCA
priority	249,47	0	0,00 (2 in the whole corpus)
world	249,47	0,11	0,21
global	155,92	0	0,02
market	93,56	0,01	0,06
popular	62,37	0,10	0,11

pmw = per million words

The most dominant version of collocations in this table were collocations with modifiers of ‘artist’. The collocation ‘world artist’ is quite understandable but it shares the same frequency per million words as ‘priority artist’. I find this statistic odd because it has not been found in BNC at all and in COCA it is found only twice in the whole one billion word corpus.

### 2.3.4 Collocations of ‘industry’



visualization by  SKETCH ENGINE

Collocations of ‘industry’	Frequency pmw in my corpus	Frequency pmw in BNC	Frequency pmw in COCA
global	498,94	0,16	0,43
emloyee	124,73	0,03	0,02
insider	62,37	0,04	0,08
dynamics	62,37	0,02	0,04
local	62,37	1.67	0,46

pmw = per million words

For the last analysis, I chose collocations including ‘industry’ because I thought the results might be interesting. There were two dominant types of collocations, modifiers of ‘industry’ and nouns modified by ‘industry’. In my corpus, the collocation with the highest frequency per million words was ‘global industry’. I was surprised by the fact that the collocation ‘local industry’ has a higher frequency per million words in BNC than COCA while having much fewer words in the corpus.

## CONCLUSION

This seminar work was about a lexical topic collocations. I have summarized the definition and types of collocations as well as some useful examples and the difference between collocations used in the Czech and English languages. I have worked with multiple sources, mainly the great literature by P. Kvetko, J. Peprník, and P. Kolář, and explained the topic collocations using the literature by these authors.

In the second part of this seminar paper, I have chosen three different articles with my topic of choice 'Czech pop culture'. I have made a corpus including these three articles using the tool Sketch Engine and then I analysed each word using Word Sketch which is included in Sketch Engine. These analysed words were then compared by frequency per million words in the corpus I made using Sketch Engine, British National Corpus, and Corpus of Contemporary American English.

Some results were as expected and some results were interesting and odd. While working on this seminar paper I have realized that the English language is so unique and interesting. I have learned how to use Sketch Engine and how to search for specific words in both BNC and COCA. I will put this knowledge to good use in my continuing study of the English language and in my forthcoming bachelor thesis 'Anglicisms in the Czech and Slovak music industry.'

## LITERATURE

- Kvetko, P. (2009). *English Lexicology in Theory and Practice* (2nd Revised edition). Univerzita Sv. Cyrila a Metoda v Trnave.
- Peprník, J. (1992). *Anglická lexikologie*. Rektorát Univerzity Palackého v Olomouci.
- Kolář, P. (2006). *A guide to English Lexicon*. Slezská univerzita v Opavě.
- Elavsky, C. M. (2011). Musically mapped: Czech popular music as a second ‘world sound’. *European Journal of Cultural Studies*, 14(1), 3-24. <https://doi.org/10.1177/1367549410370074>
- Blüml, J. (2022). The music mainstream in communism revisited: a corpus analysis of Czechoslovak pop lyrics (1962–1991). *Journal of Contemporary Central and Eastern Europe*, 30(2), 147–165. <https://doi.org/10.1080/25739638.2022.2089384>
- Skořepová, Z. (2019). Cultural memory, underground and mainstream culture reminiscences in the Czech Republic. *Ethnomusicology Forum*, 28(3), 362–383. <https://doi.org/10.1080/17411912.2020.1755332>

## ONLINE TOOLS

- <https://www.sketchengine.eu/>
- <https://www.grammarly.com/>
- <https://www.english-corpora.org/coca/>
- <https://www.english-corpora.org/bnc/>