

**UNIVERZITA KARLOVA**

**Filozofická fakulta**

**Ústav anglického jazyka a didaktiky**

**Filologie – anglický jazyk**

**Autoreferát disertační práce**

**Mgr. Jaroslav Emmer**

*Anglická přirovnání v korpusových datech*

**English Idioms of Comparison in Corpus Data**

Vedoucí práce:            prof. PhDr. Aleš Klégr

2023

## 1 Introduction

This work deals with idioms of comparison. Similes are often viewed as figures of speech belonging to the literary domain. However, speakers also use them in everyday communication because they have a range of functions and effects, making them valuable tools for various purposes. They have standardised forms but also novel variants and other transformations. This perhaps makes them appear less frequent, but they typically stand out when used.

The aim of this work is to collect a representative sample of adjectival similes in present-day English, describe it, look at the representation of adjectival similes in English dictionaries and compare the English sample with a Czech one. At the same time, the study aims to describe the methodology of searching for adjectival similes in corpora and its associated pitfalls.

For several reasons, purely linguistic studies researching adjectival similes are relatively scarce. Firstly, they are traditionally associated with literature. Secondly, they are sometimes overshadowed by metaphors and dismissed as nothing but explicit variants. Thirdly, adjectival similes are multi-word units and require larger datasets to provide enough instances for an investigation.

The research conducted in this work uses corpus linguistics as a methodology for data mining. Adjectival similes were extracted from four heterogeneous corpora and subjected to a thorough analysis. The formal frame suggests that mining simile from corpora should be a simple task. However, the formal frame is not exclusive to adjectival similes. Consequently, it is necessary to formulate precise criteria delimiting what is considered an adjectival simile.

## 2 Similes

Since comparing is vital for understanding the world, it is also reflected in language. As a result, probably all languages have a plethora of grammatical and lexical means used for comparing that their speakers recognise as established and conventional. One traditionally recognised category of set expressions serving this purpose is called idioms of comparison, or even more commonly, *similes*. Similes can generally be described as idiomatic units of varying form and meaning whose purpose is to reinforce the feature ascribed to the *target*, the subject of comparison, by comparing it to a prototypical bearer of such a feature – the *source*. The shared feature is either explicit or implicit and may display a varying degree of prominence.

## **2.1 The simile interpretation of metaphor**

Many scholars, starting with Aristotle, argued that metaphors are just elliptical similes and tried to support this view with various arguments. Many simile theories suggest that metaphors are nothing but implicit similes. The validity of these theories is questionable, but they help form the basis of our understanding of the undeniable difference between similes and metaphors. Barnden (2009: 81) argues that “non-paralleled items are often crucial to the metaphorical effect.” He demonstrates with examples that, contrary to popular belief, not every linguistic metaphor requires mapping to be understood and attributes the lack of attention paid to non-parallelism in theories of metaphor to vague descriptions of how context is used in the interpretation of metaphors. As Glucksberg and Haught (2006: 361) put it, “metaphors and similes are understood in their own right: the metaphor as a categorisation assertion, the simile as an assertion of similitude”.

## **2.2 Literal comparison vs simile**

In this work, similes are considered (idiomatic) phraseological units of varying degrees of conventionalisation, both syntactic and semantic. Consequently, the established part-of-speech sequences allow us to harvest most similes from corpora with a predesigned query. However, these syntactic frames are not exclusive to similes, and the query often yields many irrelevant strings. Therefore, it is imperative to establish a clear definition of similes as opposed to literal comparisons of two (or more) entities.

Čermák (2007: 386) describes the difference between ‘non-established comparisons’, whose purpose is to simply express a perceived similarity of two entities at a specific moment, and ‘established comparisons’ (similes), which are phraseological units stored in our mental lexicon as single items. It follows from the distinction that non-established comparisons are readily generated by speakers of the language with a free selection of the components, while similes are acquired as whole phraseological units.

Although some similes are literal in meaning, their respective component slots remain limited in terms of lexical variability, which makes them syntagmatically idiomatic. Similes are ready-made phraseological units and, as Čermák (2007: 393) points out, attempts at decomposition and analysis of individual components and their role seldom provide any beneficial outcome. Conversely, literal comparisons are only restricted by the syntactic frame and allow for the free paradigmatic selection of the lexical elements and the analysis of their constituents.

### 2.3 Explicit vs implicit similes

The default simile structure comprises five semantic components, each with its own specifics and function.

<b>comparandum</b> <b>(target)</b>	<b>relator</b>	<b>tertium</b> <b>comparationis</b> <b>(ground)</b>	<b>comparator</b>	<b>comparatum</b> <b>(source)</b>
<i>Robin</i>	<i>was</i>	<i>high</i>	<i>as</i>	<i>a kite</i>
<i>Lenny</i>	<i>cried</i>	<i>(-)</i>	<i>like</i>	<i>a baby</i>

Figure 1. The semantic components of similes.

Naturally, not all similes contain all five components explicitly. Nevertheless, two prototypical forms illustrated in Figure 1 serve as the base forms for all structural variants. These simile forms present a target (*Robin* and *Lenny*) and a source (*kite* and *baby*), the formally mandatory elements of the simile. However, *Robin was high as a kite* also contains an explicit ground *high*, making the simile ‘explicit’ (Čermák, 2007: 401). The explicitness means there is no free ground interpretation, as the simile states the ground. This is not the case in *Lenny cried like a baby*; therefore, such similes are ‘implicit’ (ibid.). In implicit similes, the ground is either implied by the relator or is not required for the interpretation. Frequently, explicit similes are called adjectival and implicit similes are called verbal. This is based on the central components’ part of speech.

### 2.4 Standard similes

It is equally important to establish what is considered a standard simile. Standard similes are conventionalised, recurrent expressions that are not limited to specific situations. They are acquired by speakers of the language and relisted in dictionaries for the purpose of description and preservation (hence standard). The recurrence makes them possible to be traced in corpora and justifies the attention, as they can be used in communication to convey established and readily recognised meanings.

Conversely, ‘nonce simile’ is a term for one-off expressions that are typically used in poetry to trigger specific aesthetic effects. These are primarily the subject of literary analysis and are of little interest to linguistic enquiry for two reasons. Firstly, they are often purposeless (or even ineffective) for standard communication, as they are created for a particular occasion. Secondly,

as the source slot is frequently extensive, they are difficult to track down, making their collection arduous.

## 2.5 The two main functions of similes

The two main functions of similes are intensification and humour. Moon (2008: 7) describes adjectival similes as “pragmatically emphasizing the degree of the quality denoted by the adjective.” The intensification function is undisputed, but the question remains as to how much of the comparison between the target and the source remains active in the interpretation of these similes.

As for humour, Veale (2013: 4) notes that “the simile form continues to be a frequent, varied and ingeniously crafted vehicle for expressing humorous intent.” This does not pertain to nonce similes only. Norrick (1986: 48-51) discusses humour in similes and elaborates on the phenomena contributing to particular similes being perceived as humorous, namely incongruence of the source, irony, punning and euphemisms.

## 3 Adjectival similes

The formalised frame ‘(as) ADJECTIVE as NOUN PHRASE’ is considered the prototypical form, as it is represented in most standard adjectival similes. The initial *as* is a non-obligatory element and its omission is common in spoken English. For a CQL query construction, it is best to disregard the initial *as*. The central component is the adjective, both semantically and syntactically. Therefore, its presence is the primary determining factor in diagnosing adjectival similes. The comparator *as* is obligatory and occurs in all standard similes. The last element is the noun phrase, which, in standard similes, is typically limited to a head noun and a possible determiner.

### 3.1 Adjectival simile types

According to Moon (2008: 12), there are three main approaches to classifying an independent simile. The ‘conceptual approach’ allows us to cluster many synonymous similes into a single conceptual simile, such as (AS) HARD/TOUGH AS ROCK/NAILS/OLD BOOTS. The ‘dictionary approach’ simplifies the clustering. Either the noun or adjective become the fixed element with the other slot listing possible transformations, for instance, (AS) WHITE AS A SHEET/A GHOST/DEATH/MARBLE/MILK/PAPER/SNOW or (AS) PALE/WHITE AS A GHOST. The ‘type approach’ (or corpus approach) disregards clustering completely. Strings such as *(as) sharp as a razor* and *(as) sharp as a tack* are treated as two separate entities.

All three approaches are used in this work but for different purposes. The conceptual and the dictionary approaches are used in the description of the most frequent similes, while the type approach is used in quantitative analyses.

### **3.2 Formal properties**

Several prosodic features recur in adjectival similes. As Moon (2008: 5) points out, “many similes are alliterative, assonantal, or both.” In this context, ‘many’ represents about 20% of all the standard similes. This makes them easier for speakers to remember and reproduce. Arguably, striking prosodic features might be one of the major factors motivating the existence of some similes, such as (AS) FIT AS A FIDDLE or (AS) DRUNK AS A SKUNK.

Regarding other formal properties, both the ground and the source tend to be single-word phrases. The lexical elements in similes are mostly of Germanic origin, which is reflected in the number of syllables. This further reinforces the idea that similes are not used as precise semantic descriptions. Their functions are rather to emphasise, entertain or compare the target entity to something generally familiar.

## **4 Methodology**

Corpus evidence is imperative for this study. The corpora used for data acquisition are the spoken section of the British National Corpus 2014 (Spoken BNC2014; 11,422,617 tokens), the British National Corpus (BNC; 112,102,325), Corpus of Contemporary American English (COCA; 1,001,610,938 tokens), and the English Web 2015 (EW15; 13,190,556,334 tokens). The Spoken BNC2014 exclusively contains spoken data, the BNC and COCA combine written and spoken data (with spoken data amounting to approximately one-tenth of the overall sample size), and the EW15 comprises only written data. For this reason, any comparisons must be carefully considered, as data diversity can and will distort the results. This work primarily does not compare particular periods, dialects, genres, and age or gender demographics; it focuses on the frequency and tagging of adjectival similes in corpora.

The process of constructing a query is unique for each corpus and the respective queries underwent (sometimes significant) refinement before deemed satisfactory. Table 1 presents the overview of the queries used for simile mining.

Corpus	Query
<b>Spoken BNC2014</b>	[pos="JJ.* RR.*"&lemma!=".* y almost bad far fine just like much only quite soon twice well"] [lemma="as"] [pos="AT.* MC"]? [pos="JJ.*"]? [pos="N..* JJ.*"&lemma!="--.* fuck hell other possible shit well"]
<b>BNC</b>	[pos="AJ.* AV.*"&hw!=".* y about almost available bad far fine just least like much off on only out quite so soon together twice up well"] [hw="as"] [pos="AT.* CRD"]? [pos="AJ.*"]? [pos="N..* AJ.*"&hw!="fuck hell likely other possible shit well"]
<b>COCA</b>	[j*] as [n*] [j*] as [a*] [n*] [j*] as [a*] * [n*] [j*] as [mc*] * [n*]
<b>EW15</b>	[tag="JJ.*"&word!=".* y about almost available bad early English far fine interesting just many much nice off on only out possible quite same such so soon together twice up well"] [word="as"] [tag="DT.* CD.*"]? [tag="N..*"&word!="fuck hell likely other possible shit well"]

Table 1. The queries used for simile extraction in the four corpora.

#### 4.1 Research questions and hypotheses

The purpose of this research is to map adjectival similes in corpora. This includes their frequencies, tagging, and meanings and functions. The general task is complemented by several other goals. The process of directing the research started with the inductive approach by formulating the research questions. Then, using the deductive approach and drawing on general observations, the expected answers to the research questions were stated as hypotheses. The hypotheses are formulated as statements of tendencies rather than universalities, which is a standard approach in corpus linguistics (Stefanowitsch, 2020: 68).

1) **Research question:** *How effectively do dictionaries utilise corpus evidence regarding adjectival similes?*

**Hypothesis:** *Higher corpus frequencies of adjectival similes tend to result in their better dictionary coverage.*

**Result:** Dictionaries lag behind corpus data considerably. Many frequent adjectival similes are not covered by dictionaries.

2) **Research question:** *Are adjectival similes easy to extract from corpora using general queries?*

**Hypothesis:** *The recall-precision ratio of adjectival similes is effective when using CQL pattern-queries.*

**Result:** The CQL pattern-queries have limited effectiveness. Since the pattern is not restricted to adjectival similes, the precision is generally suboptimal. As for the recall, several similes are not returned when using generalised queries due to tagging inconsistencies of individual elements.

- 3) **Research question:** *Do prominent prosodic features make the adjectival simile more frequent?*

**Hypothesis:** *Prosodically marked adjectival similes do not occur more frequently than prosodically unmarked ones.*

**Result:** This hypothesis proved true. Prosodic markedness does not appear to be a significant factor regarding the simile's frequency.

- 4) **Research question:** *How frequent are animal sources in adjectival similes?*

**Hypothesis:** *Adjectival similes with 'animal' sources constitute a large group within the dataset.*

**Result:** This hypothesis proved true. No other semantic category reaches the frequency of 'animal' similes, unless very general.

- 5) **Research question:** *Is there a significant lexical overlap between English and Czech adjectival similes?*

**Hypothesis:** *Frequent English adjectival similes have lexical equivalents in Czech.*

**Result:** This is only partly true. While we can observe a significant lexical overlap in many cases, several other examples illustrate the uniqueness of the simile stock in each language.

## 5 A quantitative survey of findings

	SBNC14	BNC	COCA	EW15
Corpus size (number of tokens)	11,422,617	112,102,325	1,001,610,938	13,190,556,334
Original cut-off	not applied	4	8	40
Absolute frequency of simile tokens (retrieved)	96	788	7194	38251
Absolute frequency of simile types (retrieved)	56	115	189	218
Token/type ratio (tokens per simile)	1.714	6.852	38.063	175.763
Relative frequency of simile tokens (i.p.m.)	8.404	7.029	7.182	2.900

Table 2. Statistical overviews for the researched corpora.



Table 2 presents statistical overviews for each corpus. A ‘token’ represents any adjectival simile occurrence, and a ‘type’ is a formally unique adjectival simile, regardless of its frequency. Therefore, five instances of *deaf as a post* and three instances of *high as a kite* equal eight tokens and two types. The token-type ratio illustrates an average number of tokens per simile type. The relative frequency represents the number of similes per million words within the corpus.

Another important aspect of this study is the treatment of adjectival simile types. This work considers *sharp as a razor* and *sharp as razors* two different simile types. This is because neither singular nor plural sources can be treated as the general default for adjectival similes. While some similes appear with both singular and plural sources, others are restricted to one form. For example, (AS) HARD AS NAILS, (AS) HARMLESS AS DOVES or (AS) WIDE AS SAUCERS do not have a singular source variant attested in the dataset. Similarly, many similes are attested in the dataset only with a singular source, such as (AS) CLEAR AS A BELL, (AS) HIGH AS A KITE or (AS) QUIET AS A TOMB. The ability to count the source may be attributed to some similes, but it does not work for many others. As this is directly related to the semantic content of each simile, its preferred targets, and often also the context, it is best to treat all plural-source similes as unique types.

### **5.1 Cut-offs and assessing the significance**

The larger corpora limit the number of returns, often due to server load. Consequently, an imposed cut-off must be accepted when using generalised queries. The imposed cut-off can be lowered by specifying the query, but this means sacrificing the recall of formally marginal simile strings, such as *(as) thick as two short planks* or *(as) poor as a church mouse*. Ideally, no cut-off should be used for simile investigation, as it is a relatively infrequent multi-word unit compared to other constructions. However, the sheer amount of data often makes not applying any cut-off impossible.

A cut-off is a convenience tool used to avoid the necessity of browsing through long lists of data. Additionally, it is often used to prevent less frequent units from negatively affecting statistical calculations. However, the massive drawback is the omission of many less common multi-word units that are nevertheless firmly established in the lexicon. As a result, the cut-offs in this work are only used when imposed by the technical limitations or size of the corpus.

### **5.2 The representative sample of adjectival similes**

The essential part of this research was mining adjectival similes from the selected corpora and creating a list of all the unique types within each corpus. These lists were put together, and

any duplicates were removed. After a thorough revision based on the criteria outlined in the methodology section of the work, the final list comprises 309 adjectival simile types collected from all four corpora (Appendix 1).

## 6 Frequency-based core of adjectival similes

The 60 most frequent adjectival similes from the whole sample are described in more detail. The analysis of each simile presented in this section is descriptive, with quantitative data for illustration. An example from one of the corpora is always provided for each simile. The structure and content in the example texts from the corpora are unedited – any typos or spelling mistakes are preserved. Each simile is interpreted in terms of meaning and its typical interaction with frequent targets. Where relevant or otherwise significant, etymological comments are included to explain the meaning or function of the simile. Semantically and structurally related similes and comments concerning prominent formal properties are included. Here is an example of the descriptive analysis, illustrated by the most frequent in the dataset.

### (AS) CLEAR AS DAY

*However, many of the other common traits were **clear as day**: petty theft, vandalism and setting things on fire, inability to feel remorse, excessive lying, defiance of authority, lack of empathy.*

(EW15; 256958)

The most frequent simile in the dataset is (AS) CLEAR AS DAY. The meaning of this simile is ‘easy to see or understand’ or ‘obvious’. A common target is *memories* and *laws* but also a visual experience.

The simile (AS) PLAIN AS DAY appears to be semantically and functionally identical, but dictionaries do not list it as a variant of (AS) CLEAR AS DAY. The data suggest that (AS) CLEAR AS DAY is more common in American English. Moreover, it is the most frequent simile in both the COCA ( $n=273$ ; 0.273 i.p.m.) and the EW15 ( $n=1341$ ; 0.101 i.p.m.). Additionally, it is the only simile in the EW15 to reach the relative frequency of 0.10 instances per million words.

Another simile with the meaning ‘easy to see or understand’ is (AS) CLEAR AS CRYSTAL. However, this simile is not used with the meaning ‘obvious’. (AS) CLEAR AS DAY also has a humorous counterpart (AS) CLEAR AS MUD, which is used as irony with the meaning ‘not easy to understand’ or ‘not obvious’.

Lastly, the simile (AS) BRIGHT AS DAY shares the same source, but its meaning differs. The source in (AS) CLEAR AS DAY alludes to clarity in comprehension, whereas in (AS) BRIGHT AS DAY, it refers to light.

## 7 Frequency-based core of adjectival similes in dictionaries

The 60 most frequent adjectival similes are searched in seven sources to check the simile coverage. The sources are *English Idioms and How to Use Them*, *Cambridge Dictionary*, *Oxford Advanced Learner's Dictionary*, *Collins online dictionary*, *Merriam-Webster online dictionary*, *The Free Dictionary*, and the website *Simile Stack*. The data are presented in tables along with commentaries (see Table 3 as an example).

Simile	EI	Cam	Ox	Col	MW	TFD	SiS
CLEAR AS DAY	x	✓	✓	✓	✓	✓	x
WHITE AS SNOW	x	x	x	✓	x	✓	✓
TOUGH AS NAILS	x	✓	✓	✓	✓	✓	✓
EASY AS PIE	✓	✓	✓	✓	✓	✓	✓
HARD AS A ROCK	x	x	x	x	✓	✓	x
GOOD AS GOLD	x	✓	✓	✓	✓	✓	✓
LIGHT AS A FEATHER	✓	✓	x	✓	✓	✓	✓
COLD AS ICE	✓	✓	x	✓	x	✓	✓
PLAIN AS DAY	x	x	✓	✓	✓	✓	✓
OLD AS TIME	x	x	x	x	x	✓	x
SMOOTH AS SILK	x	✓	x	x	✓	✓	✓
AMERICAN AS APPLE PIE	x	✓	✓	✓	✓	✓	x

Table 3. The top 12 simile coverage in the sources.

The similes are classified assigned ‘coverage tiers’ (ranging from Tier 1 to Tier 5) according to their source coverage. Figure 2 represents the coverage summary.

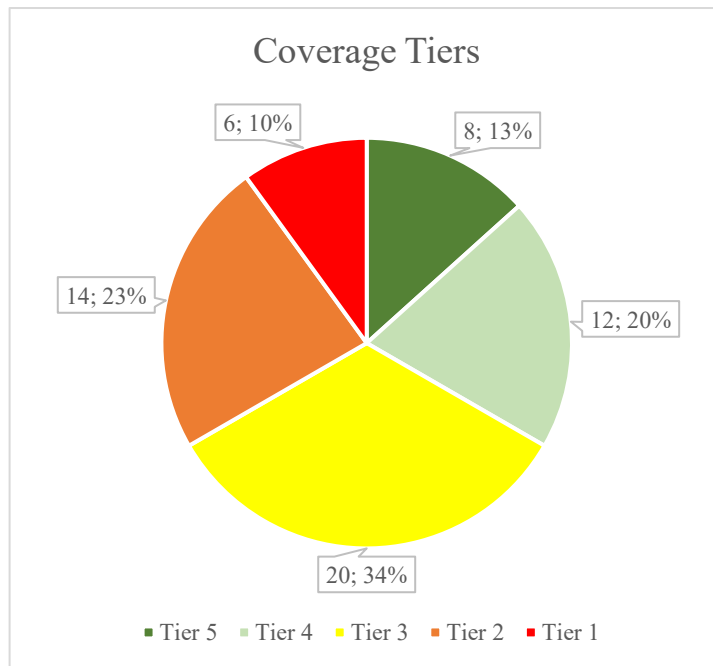


Figure 2. An overview of the adjectival simile Tiers regarding dictionary coverage.

Table 4 provides an overview of the overall dictionary coverage of adjectival similes.

Dictionary	Coverage
English Idioms	28.33%
Cambridge	48.33%
Oxford	36.67%
Collins	58.33%
Merriam-Webster	53.33%
The Free Dictionary	86.67%
Simile Stack	71.67%

Table 4. Overall dictionary coverage of adjectival similes.

## 8 Formal analysis of adjectival similes

Firstly, similes are classified according to the number of syllables and stress-timed rhythm. This classification is general and applies to all adjectival similes. Most standard similes are dimetric (two feet) or trimetric (three feet) – the few extensive similes in the dataset are marginal examples. The prevalence of dimetric similes is largely due to two specific patterns: ‘SuS’ ((AS) CLEAR AS DAY) and ‘SuuS’ ((AS) LIGHT AS A FEATHER).<sup>1</sup>

Secondly, similes are classified according to prosodic markedness. Prosodically marked similes contain alliteration, assonance, or rhythm. Overall, 20.06% of all the similes in the

<sup>1</sup> S = stressed syllable; u = unstressed syllable.

dataset are marked. 38 types contain an alliterating consonant (e.g. (AS) CLEAR AS CRYSTAL), 21 types are assonantal (e.g. (AS) EASY AS ABC), and only five types rhyme (e.g. (AS) THICK AS A BRICK).

## **9 Content analysis of adjectival similes**

Content classifications are approached from two perspectives, based on the constituting element brought to focus.

The ground-centred classification considers the ground a head element of an adjectival simile. The categories are based on clustering adjectival similes sharing a generalised semantic feature expressed by the ground, such as ‘colour’. The five general categories are ‘colour or light’ (e.g. (AS) WHITE AS SNOW), ‘character or behaviour’ (e.g. (AS) DAFT AS A BRUSH), ‘appearance or physical description’ (e.g. (AS) CLEAN AS A WHISTLE), ‘sensation or emotion’ (e.g. (AS) HIGH AS A KITE), and ‘material quality or texture’ (e.g. (AS) HARD AS A STONE).

The source-centred classification is less fragmented than the ground-centred one, suggesting that the source slot is more limiting regarding its potential representatives. This aligns with the premise that the source is usually a familiar entity. The three categories are ‘animals’ (e.g. (AS) BLIND AS A BAT), ‘natural entities or material’ (e.g. (AS) QUICK AS LIGHTNING), and ‘human-made objects’ (e.g. (AS) WARM AS TOAST). The latter two can be further subdivided into more specific categories but this potentially makes the categorisation too fragmented.

## **10 Comparison of corpus-extracted English and Czech adjectival similes**

The comparison of the English sample and the Czech list of similes by no means exhaustive and represents just a brief survey of similarities and differences observed in English and Czech. The list of Czech similes (Appendix 2) was kindly made available by M. Hnátková and V. Petkevič. It was compiled for the purposes of the Czech electronic database of MWEs, LEMUR, using data from Čermák et al. (1983–2009), FRANTALEX (a list of phrasemes and collocations compiled by M. Hnátková) and two Czech SYN corpora (Křen et al. 2015, 2019). After revision, the list comprises 886 adjectival simile types, which is considerably more than the list of English similes (309 types).

One general observation can be made regarding the transformational variability in the source slot in Czech similes. Many grounds in the list appear with a great number of sources. This variability does not appear in the English similes collected from the researched corpora. However, the list of Czech similes is a combined product of corpus data (the SYN corpora) and

dictionary data. It is important to note that dictionaries typically inherit existing word lists and update them by adding new items. Many items become obsolete over time and may no longer be attested in present-day corpora, but they are rarely removed from the dictionaries. Consequently, the list of Czech similes naturally contains more types as it represents the typical ‘dictionary bloating’.

Table 5 presents an overview of ground-centred categories along with the type frequencies for the English and Czech simile types in this research. The brackets show relative frequencies of simile types per one hundred (t.p.h.) to illustrate how frequent the individual categories are within the samples.

Category	English	Czech
<b>Colour or light</b>	30 (9.709)	95 (10.722)
<b>Character or behaviour</b>	63 (20.388)	197 (22.235)
<b>Appearance or physical description</b>	43 (13.916)	246 (27.765)
<b>Sensation or emotion</b>	45 (14.563)	102 (11.512)
<b>Material quality or texture</b>	35 (11.327)	109 (12.302)
<b>Intoxication</b>	3 (0.971)	28 (3.160)

*Table 5. English and Czech ground-centred category distribution.*

The comparison of animal sources suggests that ‘animal’ adjectival similes are more prevalent in Czech. The English list yields 54 instances (17.476 t.p.h.) of similes with an animal source, whereas the Czech list contains 264 unique types (29.797 t.p.h.). Additionally, due to the larger size of the dataset, some Czech animal sources appear with more unique grounds than the English ones. However, this trend should not be overstated, as many animal sources still remain exclusive to single similes.

## 11 Discussion of findings

The hypotheses in this work and their testing were intended as complementary goals to the overall survey of adjectival similes in the selected corpora. The primary focus was to map adjectival similes in corpora, their annotation, and scrutinise the mining process. Using the data mined from four English corpora, I collected a reliable sample of adjectival similes objectively reflecting the current state of their use in English, which was the primary goal of this thesis. By imposing a frequency limit to this representative sample, I then compiled and described what could be termed a ‘similes minimum’ (by analogy to the so-called paremiological minimum), in other words, a set of similes that can be regarded as the essential minimal knowledge of

similes displayed by native speakers which learners of English should at least know, or ideally, be able to use.

This work is by no means exhaustive, and many aspects of adjectival similes require further research. For instance, the common conception that similes are primarily aesthetic literary devices requires extensive data analysis. The similes extracted in this work suggest that adjectival similes are frequently used in spoken English, and some appear almost exclusively colloquial ((AS) COMMON AS MUCK or (AS) HARD AS A ROCK).

Another aspect is the exclusiveness of adjectival similes regarding regional variants. This work occasionally comments on regional preferences of certain similes, but it offers no complex account of regional variation. The dataset hints at many similes being popular in specific English-speaking countries. However, more data would need to be analysed to obtain conclusive evidence.

Adjectival simile transformations are likewise an interesting area, especially compound adjectival similes (CAS). Many adjectival similes have CAS forms, whose frequencies are much higher than those of standard simile forms. CAS demonstrate the conceptual similes' ability to adapt syntactically. The empirical evidence does not suggest that standard similes are cognitively prominent lexical units. This is where an investigation of CAS could help us explain what helps conceptual similes survive in the language. Moon (2008: 34) argues that *as*-similes "likely (...) belong to receptive vocabularies" and considers them "negligible in terms of frequency" in corpora. Although standard adjectival similes undisputedly are infrequent compared to other single- and multi-word units, their CAS transformations offer a significantly different picture.

## 12 Conclusions

This work's primary purpose was the investigation of adjectival similes in corpus data. Four different corpora were chosen for the simile extraction to ensure data heterogeneity, allowing the collection of similes from various types of discourse. This offered insight into how various corpora handle the annotation of multi-word units and how restrictive the corpora are both technically and functionally.

The process of data mining was a tasking endeavour, but it revealed a lot about adjectival similes, their various alternative forms, and the individual corpora. The precision of general CQL queries is overwhelmingly insufficient, and individually searching every simile is extremely time-consuming and limits the data we can obtain. The formal frame appears frequent in English, albeit only marginally used by adjectival similes.

Standard dictionaries often contain a limited simile set, which typically does not reflect the empirical evidence. Using specialised idiom or simile dictionaries does not help either. Firstly, they are exceedingly rare since online dictionaries typically incorporate idioms entries or whole sections. Secondly, dictionaries such as Sommer's *Similes Dictionary* (2013) contain many creative similes that are not attested in corpus data and are often clearly author-specific expressions. These dictionaries contribute to the literary image of similes and distort their overall perception.

Naturally, several limitations ought to be mentioned. Any corpus-based research is limited by the chosen corpora. The data never represent a complete picture and only provide a fraction of the language reality. Although conclusions based on corpus investigations cannot be viewed as ultimate truths, they can be indicative of general tendencies in the language with great reliability.

Furthermore, corpus-based studies are restricted by our theoretical preconceptions. This can perfectly be demonstrated by adjectival similes. On many occasions throughout this research, I considered including various transformations to complement the standardised *as*-forms in the overall frequencies. This was partly due to the standard similes being relatively infrequent but also because, fundamentally, the transformations represent the comparison concept equally. Ultimately, the statistics only include the standard forms due to space constraints. However, the transformations are believed to belong in the adjectival simile research.

To conclude, this work sheds light on many aspects of adjectival similes. Conversely, it also gives rise to many questions. The goals of this enquiry were achieved, but more research is necessary to further our understanding of adjectival similes and their various forms.

## References

- Barnden, J. (2009). Metaphor and Context: A Perspective from Artificial Intelligence. In Musolff, A., Zinken, J. (eds) *Metaphor and Discourse*, pp. 79-94. Palgrave Macmillan, London. [https://doi.org/10.1057/9780230594647\\_6](https://doi.org/10.1057/9780230594647_6)
- Čermák, F. (2007). *Frazeologie a idiomatika česká a obecná / Czech and General Phraseology*. Praha: Univerzita Karlova v Praze – Nakladatelství Karolinum.
- Glucksberg, S., Haught, C. (2006). On the Relation Between Metaphor and Simile: When Comparison Fails. In *Mind & Language*, Vol. 21 No. 3, pp. 360–378.
- Moon, R. (2008). Conventionalized *as*-similes in English: a problem case. *International Journal of Corpus Linguistics*, 13(1), pp. 3-37. <https://doi.org/10.1075/ijel.13.1.03moo>
- Norricks, N. R. (1986). Stock Similes. *Journal of Literary Semantics XV/1*, pp. 39-52.
- Stefanowitsch, A. (2020). *Corpus linguistics: A guide to the methodology*. Berlin: Language Science Press.



Veale, T. (2013). Humorous similes. In Ford, T. (Ed.), *Humor: International Journal of Humor Research*, 26(1), pp. 3-22.

## Corpora

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

Jakubiček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). The TenTen corpus family. *7th International Corpus Linguistics Conference CL*, pp. 125-127.

Křen, M., Cvrček, V., Čapka, T., Čermáková, A., Hnátková, M., Chlumská, L., Jelínek, T., Kovářiková, D., Petkevič, V., Procházka, P., Skoumalová, H., Škrabal, M., Truneček, P., Vondříčka, P. Et Zasina, A. (2015). *SYN2015: reprezentativní korpus psané češtiny*. Praha: Ústav Českého národního korpusu FF UK. Available at: <http://www.korpus.cz>

Křen, M., Cvrček, V., Čapka, T., Čermáková, A., Hnátková, M., Chlumská, L., Jelínek, T., Kovářiková, D., Petkevič, V., Procházka, P., Skoumalová, H., Škrabal, M., Truneček, P., Vondříčka, P. et Zasina, A. (2019). *Korpus SYN, verze 8 z 12. 12. 2019*. Praha: Ústav Českého národního korpusu FF UK. Available at: <http://www.korpus.cz>

Love, R., Dembry, C., Hardie, A., Brezina, V. and McEnery, T. (2017). The Spoken BNC2014: designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics*, 22(3): pp. 319-344. DOI: 10.1075/ijcl.22.3.02lov

The British National Corpus (BNC) – Hoffmann, S., & Evert, S. (1996). BNCweb (CQP-edition). Accessed from <https://cqpweb.lancs.ac.uk/> (last accessed on June 25, 2023).

## Sources

Cambridge dictionary. Cambridge University Press. Accessed from <https://dictionary.cambridge.org/> (Last accessed on June 25, 2023)

Collins dictionary (online). Collins. Accessed from <https://www.collinsdictionary.com/> (last accessed on June 25, 2023)

Merriam-Webster dictionary. Accessed from <https://www.merriam-webster.com> (Last accessed on June 25, 2023)

Oxford Learner's Dictionaries. Oxford University Press. Accessed from <https://www.oxfordlearnersdictionaries.com/> (Last accessed on June 25, 2023)

Seidl, J., McMordie, W. (1988). *English idioms and How to Use Them* (5th ed.). Oxford University Press.

Simile Stack. Accessed from <https://www.similestack.com/> (Last accessed on June 25, 2023)

Sommer, E. (2013). *Similes Dictionary* (2nd ed.). Visible Ink Press.

The Free Dictionary. Farlex, Inc. Accessed from <https://www.thefreedictionary.com/> (Last accessed on June 25, 2023)

## APPENDIX 1 – THE ENGLISH SIMILE SAMPLE (309 items)

American as apple pie	brave as hope
big as a barn	bright as a button
big as a basketball	bright as day
big as a field	bright as the Sun
big as a football	busy as a bee
big as a football field	busy as bees
big as a football stadium	clean as a whistle
big as a house	clear as a bell
big as a mountain	clear as crystal
big as a whale	clear as day
big as houses	clear as daylight
big as life	clear as glass
big as marbles	clear as mud
big as saucers	clear as the sun
big as Texas	clear as water
big as the sky	cold as ice
black as a crow	cold as stone
black as a raven	common as dirt
black as a raven's wing	common as muck
black as coal	cool as a cucumber
black as ink	cool as ice
black as midnight	crazy as a loon
black as night	crooked as a dog's hind leg
black as pitch	cute as a bug
black as soot	cute as a bug's ear
black as the ace of spades	cute as a button
black as the night	daft as a brush
black as thunder	dappy as pig shit
blind as a bat	dark as night
blue as the sky	dark as the night
bold as a lion	dark as the night sky
bold as brass	dead as a dodo
brave as a lion	dead as a doornail

dead as mutton	good as her word
deaf as a post	good as his word
deaf as posts	good as sex
different as day and night	green as grass
different as chalk and cheese	guilty as sin
different as chalk from cheese	happy as a clam
different as night and day	happy as a lark
drunk as a lord	happy as a pig
drunk as a skunk	happy as a pig in muck
dry as a bone	happy as a pig in mud
dry as a desert	happy as a pig in shit
dry as dust	happy as clams
dull as dishwater	happy as Larry
dumb as a box of hair	hard as a rock
dumb as a box of rocks	hard as concrete
dumb as a post	hard as diamond
dumb as a rock	hard as diamonds
dumb as bricks	hard as granite
dumb as dirt	hard as iron
dumb as rocks	hard as nails
easy as ABC	hard as rock
easy as pie	hard as rocks
fast as lightning	hard as steel
fast as the wind	hard as stone
fat as a pig	harmless as doves
fit as a fiddle	healthy as a horse
flat as a board	heavy as lead
flat as a pancake	high as a house
free as a bird	high as a kite
free as the wind	high as an elephant's eye
fresh as a daisy	high as the sky
gentle as a lamb	hot as an oven
gentle as doves	hot as balls
good as gold	hot as fire

hot as hell	old as the world
hot as the sun	old as time
hungry as a wolf	old as war
cheap as chips	pale as a ghost
innocent as doves	pale as death
keen as mustard	pale as the moon
large as life	pissed as a fart
light as a feather	pissed as a nook
light as air	plain as a pikestaff
lonely as a cloud	plain as day
mad as a hatter	plain as the nose on (one's) face
mad as a hornet	pleased as punch
mad as a march hare	poor as a church mouse
miserable as sin	poor as dirt
naked as a jaybird	poor as church mice
naked as the day (one) came to this world	precious as gold
naked as the day (one) was born	pretty as a picture
natural as breathing	proud as a peacock
neat as a pin	proud as punch
nervous as a cat	pure as gold
nice as pie	pure as snow
old as America	pure as the driven snow
old as civilization	quick as a flash
old as dirt	quick as a rat
old as history	quick as a wink
old as humanity	quick as lightning
old as mankind	quiet as a grave
old as Methuselah	quiet as a church
old as the Bible	quiet as a church mouse
old as the earth	quiet as a mouse
old as the hills	quiet as a tomb
old as the history of sth	quiet as the grave
old as the republic	rare as hen's teeth
old as the universe	red as a beet

red as a beetroot	smooth as velvet
red as a tomato	snug as a bug
red as blood	soft as a baby's
regular as clockwork	soft as a baby's behind
right as rain	soft as a baby's bottom
rich as Croesus	soft as a baby's butt
rough as arseholes	soft as butter
safe as houses	soft as cotton
serious as a heart attack	soft as silk
sharp as a knife	soft as velvet
sharp as a razor	solid as a rock
sharp as a tack	solid as rock
sharp as knives	steady as a rock
sharp as razors	stiff as a log
sick as a dog	stiff as a board
sick as a parrot	stiff as a poker
sick as your secrets	still as a statue
silent as the grave	still as a stone
simple as ABC	straight as a die
simple as pie	straight as a ramrod
skinny as a rail	straight as an arrow
skinny as a rake	strong as a bear
slippery as an eel	strong as a bull
slow as molasses	strong as a horse
smart as a whip	strong as a lion
smooth as a baby	strong as an oak
smooth as a baby's ass	strong as an ox
smooth as a baby's behind	strong as death
smooth as a baby's bottom	strong as iron
smooth as a baby's bum	strong as steel
smooth as a baby's butt	stubborn as a mule
smooth as butter	sure as night
smooth as glass	sure as the sun
smooth as silk	sweet as candy

sweet as honey

white as snow

sweet as pie

white as the snow

sweet as sugar

wide as saucers

sweet as wine

wise as a serpent

tall as a house

wise as serpents

tall as a tree

thick as a brick

thick as pig shit

thick as pigshit

thick as thieves

thick as two short planks

thin as a rail

thin as a rake

thin as paper

tight as a drum

tough as nails

tough as old boots

transparent as glass

true as steel

ugly as sin

unique as a fingerprint

warm as toast

weak as a kitten

weak as water

white as a ghost

white as a sheet

white as alabaster

white as bone

white as death

white as chalk

white as marble

white as milk

white as paper

white as sheets

## APPENDIX 2 – THE CZECH SIMILE LIST (886 items)

bázlivý jako králík	blbý jako štandlík
bázlivý jako zajíc	blbý jako tágo
bdělý jako Argus	blbý jako troky
bezbranný jako dítě	blbý jako vantrok
bezmocný jako dítě	blbý jako vantroky
bezradný jako dítě	bledý jako alabastr
bílý jako alabastr	bledý jako duch
bílý jako křída	bledý jako křída
bílý jako lilie	bledý jako měsíc
bílý jako mléko	bledý jako mramor
bílý jako mramor	bledý jako mrtvola
bílý jako mrtvola	bledý jako papír
bílý jako padlý sníh	bledý jako sedma
bílý jako papír	bledý jako slonová kost
bílý jako sejra	bledý jako smrt
bílý jako slonová kost	bledý jako stěna
bílý jako smrt	bledý jako tvaroh
bílý jako sníh	bledý jako vosk
bílý jako stěna	bledý jako z vosku
bílý jako tvaroh	bledý jako zeď
bílý jako vápno	bledý jako zelenka
bílý jako z alabastru	bledý jako zelinka
bílý jako zeď	bohatý jako Argentinec
blbý jako bedna kytu	bohatý jako koza rohatá
blbý jako bota	bohatý jako Krésus
blbý jako kopyto	bohatý jako Rothschild
blbý jako motyka	bojovný jako kohout
blbý jako pařez	bradatý jako Kristus
blbý jako patník	bujný jako hříbě
blbý jako poleno	bystrý jako sokol
blbý jako pučtok	černý jako antracit
blbý jako rádio	černý jako ašant
blbý jako škopek	černý jako bota

černý jako cikán	čilý jako pometlo
černý jako cikánka	čilý jako rtuť
černý jako čert	čilý jako ryba
černý jako ďábel	čilý jako rybička
černý jako eben	čilý jako veverka
černý jako havran	čiperný jako veverka
černý jako havraní křídla	čistý jako alabastr
černý jako hřích	čistý jako klíčka
černý jako inkoust	čistý jako křišťál
černý jako kolomaz	čistý jako lilie
černý jako mouřenín	čistý jako lilium
černý jako noc	čistý jako padlý sníh
černý jako saze	čistý jako panic
černý jako smola	čistý jako sklo
černý jako smůla	čistý jako slovo boží
černý jako uhel	čistý jako studánka
černý jako vrána	čistý jako z alabastru
červený jako cihla	čistý jako z bavlnky
červený jako jablíčko	čistý jako z cukru
červený jako jahoda	čistý jako ze škatulky
červený jako kohout	čistý jako zrcadlo
červený jako králík	děravý jako cedník
červený jako malina	děravý jako ementál
červený jako panenka	děravý jako řešeto
červený jako paprika	děravý jako síto
červený jako prase	děravý jako sýr
červený jako rak	divoký jako ďábel
červený jako růže	dlouhý jako chmelová tyčka
červený jako růžička	dlouhý jako jitrnice
červený jako vlčí mák	dlouhý jako Lovosice
čilá jako koroptev	dlouhý jako skládací metr
čilý jako čečetka	dlouhý jako slonbidlo
čilý jako koroptvička	dlouhý jako štangle
čilý jako mník	dlouhý jako tasemnice



dlouhý jako tyčka	hladký jako mramor
dlouhý jako týden	hladký jako mýdlo
dobrý jako anděl	hladký jako oblázek
dotěrný jako moucha	hladký jako olej
dotěrný jako štěnice	hladký jako samet
drsný jako smirkový papír	hladký jako sklo
drsný jako struhadlo	hladký jako stůl
drzý jako opice	hladký jako úhoř
drzý jako stádo opic	hladký jako zrcadlo
drzý jako štěnice	hladový jako herec
dutý jako bambus	hladový jako lev
falešný jako Jidáš	hladový jako pes
falešný jako kočka	hladový jako vlk
falešný jako pětník	hloupý jako bota
fousatý jako Krakonoš	hloupý jako dlabaný troky
fousatý jako Kristus	hloupý jako dřevo
hbitý jako jelen	hloupý jako husa
hbitý jako ještěrka	hloupý jako kopyto
hbitý jako lasice	hloupý jako osel
hbitý jako lasička	hloupý jako pařez
hbitý jako ohař	hloupý jako patník
hebký jako hedvábí	hloupý jako tágo
hebký jako samet	hloupý jako tele
hezký jako anděl	hloupý jako troky
hezký jako andělíček	hluchý jako pařez
hezký jako andílek	hluchý jako poleno
hezký jako obrázek	hluchý jako špalek
hezký jako panenka	hluchý jako tetřev
hladký jako alabastr	hnědý jako hovno
hladký jako dětská prdelka	hodný jako anděl
hladký jako had	holý jako koleno
hladký jako hedvábí	horký jako oheň
hladký jako led	horký jako turecká fangle
hladký jako mejdlo	hořký jako blín

hořký jako pelyněk	chladný jako mramor
hořký jako utrejch	chladný jako ocel
hořký jako zeměžluč	chladný jako psí čumák
hořký jako žluč	chladný jako smrt
hravý jako kotě	chladný jako sníh
hrubý jako smirkový papír	chlupatý jako Ezau
hubatý jako čert	chlupatý jako medvěd
hubený jako bič	chlupatý jako opice
hubený jako bidlo	chudý jako kostelní myš
hubený jako drožkářská kobyła	chudý jako myš
hubený jako chroust	chytrý jako advokát
hubený jako chrt	chytrý jako čert
hubený jako komár	chytrý jako četník
hubený jako koza	chytrý jako liška
hubený jako louč	chytrý jako opice
hubený jako luňák	chytrý jako opička
hubený jako lunt	chytrý jako policajt
hubený jako nit	chytrý jako rádio
hubený jako nitka	chytrý jako stádo opic
hubený jako nudle	jasný jako Brno
hubený jako pes	jasný jako den
hubený jako sirka	jasný jako facka
hubený jako šindel	jasný jako křišťál
hubený jako špejle	jasný jako pár facek
hubený jako tříška	jednoduchý jako facka
hubený jako tyčka	jednoduchý jako násobilka
hubený jako žízala	jednoduchý jako pár facek
huňatý jako medvěd	jemný jako hedvábí
hustý jako hrachová polévka	jemný jako pavučinka
hustý jako mlha	jemný jako peří
hustý jako smetana	jemný jako peříčko
chladný jako hrobka	klidný jako beránek
chladný jako kámen	klidný jako smrt
chladný jako led	kluzký jako mejdlo

kluzký jako mýdlo	lehký jako peří
kluzký jako úhoř	lehký jako peříčko
krásná jako bohyně	lehký jako pírkó
krásná jako víla	lehoučkový jako pápěří
krásný jako Adónis	lehounkový jako pápěří
krásný jako anděl	lehounkový jako peří
krásný jako andělíček	lehounkový jako peříčko
krásný jako andílek	lehounkový jako pírkó
krásný jako antický bůh	lesklý jako drahokam
krásný jako Apollón	lesklý jako směla
krásný jako bůh	lesklý jako zrcadlo
krásný jako madona	levý jako šavle
krásný jako obrázek	levý jako ta šavle
krásný jako růže	líný jako kanec
krásný jako sen	líný jako kapr
krásný jako Venuše	líný jako prase
krásný jako z pohádky	líný jako svině
krásný jako zrcadlo	líný jako štěnice
krotký jako beránek	líný jako valach
krutý jako Nero	líný jako vepř
křehký jako pápěří	líný jako veš
křivý jako když kanec chčije	lstivý jako had
křivý jako když pes chčije	lstivý jako liška
křivý jako turecká šavle	lstivý jako Odysseus
kudrnatý jako beránek	malý jako cvoček
kudrnatý jako ovce	malý jako cvrček
kyselý jako ocet	malý jako klíčka
kyselý jako šťovík	malý jako náprstek
lakomý jako čert	malý jako špendlíková hlavička
lakomý jako chrt	mazaný jako advokát
ledový jako rampouch	mazaný jako čert
lehký jako facka	mazaný jako liška
lehký jako pápěrka	mazlivý jako kotě
lehký jako pár facek	měkký jako hedvábí

měkký jako houba	mrštný jako lasička
měkký jako koberec	mrštný jako opice
měkký jako kobereček	mrštný jako úhoř
měkký jako máslo	mrštný jako veverka
měkký jako mech	mrzutý jako dudy
měkký jako samet	mrzutý jako staré dudy
měkký jako tvaroh	nabalený jako cibule
měkký jako vosk	nabalený jako pumpa
milý jako trn v patě	nacucaný jako houba
mírná jako holubice	nacucaný jako žok
mírná jako holubička	nadýchaný jako obláček
mírný jako beránek	nadýchaný jako peříčko
mírný jako jehně	nadýchaný jako peřinka
mírný jako ovce	nahá jako Eva
mírný jako ovečka	nahý jako Adam
mlsný jako čert	nahý jako ašant
mlsný jako kocour	nalitý jako dělo
mlsný jako kočka	nalitý jako dobytek
mlsný jako koza	nalitý jako pupen
modrý jako nebe	nalitý jako slíva
modrý jako noc	nalitý jako snop
modrý jako pomněnky	nalitý jako sud
modrý jako šmolka	nalitý jako žok
mokrý jako hastrman	naložený jako mezek
mokrý jako hnůj	naložený jako soumar
mokrý jako houba	namačkání jako herynci
mokrý jako myš	namačkání jako sardinky
moudrý jako Šalamoun	namačkání jako slanečci
moudrý jako Šalomoun	napařený jako Dán
mrštný jako had	napínavý jako kšandy
mrštný jako hádě	napitý jako houba
mrštný jako ještěrka	napjatý jako luk
mrštný jako kočka	napjatý jako provazy
mrštný jako lasice	napjatý jako struna

napjatý jako strunky	opilý jako štěně
napjatý jako špagát	opilý jako zvíře
napjatý jako tětíva	opilý jako žok
napnutý jako kšandy	opuštěný jako hruška v poli
napnutý jako provazy	opuštěný jako hruška v širém poli
napnutý jako struna	opuštěný jako pes
napnutý jako špagát	ospalý jako kotě
napružený jako péro	ospalý jako sysel
nasraný jako brigadýr	ospalý jako štěně
nasraný jako kanonýr	ostrý jako břitva
nasraný jako kýbl	ostrý jako dýka
natažený jako struna	ostrý jako jed
navlečený jako cibule	ostrý jako jehla
navlečený jako na severní točnu	ostrý jako meč
němý jako kapr	ostrý jako nůž
němý jako ryba	ostrý jako šídlo
němý jako rybička	ostrý jako žiletky
neoblomný jako skála	ostříhaný jako trestanec
neomylný jako papež	ošklivá jako čarodějnice
nešťastný jako šafářův dvoreček	ošklivá jako ropucha
neústupný jako pařež	ošklivá jako stará čarodějnice
nevinný jako anděl	ošklivý jako ďábel
nevinný jako andílek	ošklivý jako noc
nevinný jako lilie	ošklivý jako opice
nevinný jako lilium	oškubaný jako lípa
oblečený jako hadroš	oteklý jako bakule
oblečený jako na severní točnu	otrávený jako šváb
ohavný jako noc	otrhaný jako cikán
ohyzdný jako noc	otrhaný jako lípa
opálený jako ašant	ožralý jako čuně
opálený jako Indián	ožralý jako Dán
opilý jako čuně	ožralý jako dělo
opilý jako Dán	ožralý jako prase
opilý jako mraky	ožralý jako slíva

ožralý jako svině	pracovitý jako mraveneček
ožralý jako štěně	pracovitý jako pejsek
pádný jako pěst	pracovitý jako včela
paličatý jako mezek	pracovitý jako včelička
paličatý jako osel	pracovitý jako včelka
pěkný jako panáček	prohnutý jako luk
pevný jako ocel	proměnlivý jako počasí
pevný jako skála	prostrhý jaký řešeto
pevný jako židovská víra	prostrhý jaký síto
pevný jako žula	protivný jaký činže
pichlavý jaký ježek	provrtaný jaký řešeto
pilný jaký mravenec	pružný jaký proutek
pilný jaký mraveneček	pružný jaký vrbový proutek
pilný jaký včela	přelétavý jaký motýl
pilný jaký včelička	přesný jaký hodinky
pilný jaký včelka	přesný jaký hodiny
pitomý jaký ovce	přesný jaký švýcarské hodinky
pitomý jaký pučtok	pyšný jaký páv
pitomý jaký tágo	pyšný jaký pávice
pitomý jaký vantrok	rezavý jaký liška
pitomý jaký vantroky	rovný jaký když kanec chčije
plachý jaký laň	rovný jaký když střelí
plachý jaký laňka	rovný jaký mlat
plachý jaký srna	rovný jaký pravítko
platný jaký pes v kostele	rovný jaký stůl
plavý jaký lvíce	rovný jaký svíce
plochý jaký stůl	rovný jaký šíp
podobný jaký vejce vejci	rovný jaký šňůra
pomalý jaký šnek	rovný jaký turecká šavle
pomalý jaký želva	rozbitý jaký turecká fangle
poslušný jaký pejsek	rozcuchaná jaký čarodějnice
potrhlý jaký švec	rozcuchaná jaký stará čarodějnice
pověřivý jaký stará baba	rozpálený jaký cihlička
pověřivý jaký stará bába	rozpálený jaký dvířka od kamen

rozpálený jako kamna	silný jako dub
rozpálený jako oheň	silný jako golem
rozpálený jako pec	silný jako Herkules
rozpálený jako trajfus	silný jako hrom
rozpálený jako žehlička	silný jako kobyla
rozstřílený jako cedník	silný jako kůň
roztažený jako žába	silný jako lev
roztrhaný jako turecká fangle	silný jako lvi
rozvrzaný jako staré schody	silný jako medvěd
rozzuřený jako bejk	silný jako noha
rozzuřený jako býk	silný jako Samson
rudý jako krocán	silný jako smrt
rudý jako mák	silný jako tur
rudý jako pivoňka	silný jako z oceli
rudý jako rak	skromná jako fialka
růžový jako prasátko	skromný jako fialinka
růžový jako selátko	skromný jako chudobka
rychlý jako blesk	slabý jako čaj
rychlý jako ďábel	slabý jako dech
rychlý jako chrt	slabý jako duch
rychlý jako jelen	slabý jako hnilička
rychlý jako ohař	slabý jako komár
rychlý jako raketa	slabý jako moucha
rychlý jako sokol	slabý jako pápěrka
rychlý jako střela	slabý jako pápěří
rychlý jako šíp	slabý jako třtina
rychlý jako šipka	slabý jako z kozy duch
rychlý jako vítr	sladký jako cecek
scvrklý jako křížala	sladký jako cukr
sdílný jako led	sladký jako cumel
shrbený jako stařec	sladký jako med
silný jako bejk	sladký jako vánek
silný jako Bivoj	slepý jako kotě
silný jako býk	slepý jako krtek

slepý jako netopýr	svobodný jako Amerika
slepý jako patrona	šedivý jako popel
slizký jako had	šeredný jako ďábel
smrdutý jako zkažený zub	šeredný jako noc
smutný jako želva	šikovný jako hrom do police
snědý jako cikán	šikovný jako motovidlo
snědý jako cikánka	široký jako almara
sprostý jako dlaždič	škaredá jako ropucha
sprostý jako erární deka	škaredý jako noc
sprostý jako kanálník	škaredý jako opice
starý jako Abrahám	špičatý jako jehla
starý jako lidstvo	špičatý jako šídlo
starý jako lidstvo samo	špinavý jako bota
starý jako Metuzalém	špinavý jako cikán
starý jako svět	špinavý jako kanec
stejný jako každý druhý	špinavý jako onuce
stejný jako každý jiný	špinavý jako prase
stísněný jako klíčka	špinavý jako špek
studený jako had	šťastný jako blecha
studený jako kámen	šťastný jako děčko
studený jako kus ledu	šťastný jako malé dítě
studený jako led	štíhlá v pase jako vosa
studený jako mramor	štíhlý jako bříza
studený jako psí čumák	štíhlý jako gazela
studený jako psí nos	štíhlý jako jedle
studený jako rampouch	štíhlý jako laň
studený jako ryba	štíhlý jako laňka
suchý jako drn	štíhlý jako proutek
suchý jako louč	štíhlý jako svíce
suchý jako šindel	štíhlý jako svíčka
suchý jako trn	švorcový jako eso pikový
suchý jako troud	švorcový jako pikové eso
svěží jako broskev	temný jako noc
svěží jako orosená broskev	tenký jako hůlky



tenký jako lupínek	tlustý jako koule
tenký jako nit	tlustý jako měch
tenký jako nitka	tlustý jako noha
tenký jako nudle	tlustý jako prase
tenký jako papír	tlustý jako šandlík
tenký jako pavučinka	tlustý jako valach
tenký jako sirka	tlustý jako vepř
tenký jako struna	tmavý jako noc
tenký jako strunka	trpělivý jako beránek
tenký jako tříška	trpělivý jako Job
tenký jako vlas	trpělivý jako ovečka
tenký jako vlásek	tuhý jako houžev
teplý jako chcanky	tuhý jako podrážka
teplý jako kafe	tupý jako motyka
teplý jako polívka	tupý jako pařez
těžký jako cent	tupý jako poleno
těžký jako hrom	tupý jako prdel
těžký jako kámen	tvárný jako vosk
těžký jako kráva	tvrdohlavý jako beran
těžký jako olovo	tvrdohlavý jako kozel
těžký jako prase	tvrdohlavý jako mezek
tichý jako myš	tvrdohlavý jako osel
tichý jako myška	tvrdý jako beton
tichý jako oukropeček	tvrdý jako brok
tichý jako pěna	tvrdý jako hřebíky
tichý jako smrt	tvrdý jako kámen
tlustý jako bagoun	tvrdý jako kost
tlustý jako basa	tvrdý jako křemen
tlustý jako bečka	tvrdý jako mlat
tlustý jako bejk	tvrdý jako oblázek
tlustý jako buřt	tvrdý jako ocel
tlustý jako býk	tvrdý jako podrážka
tlustý jako čuně	tvrdý jako roh
tlustý jako hroch	tvrdý jako skála

tvrdý jako suk	velký jako kolo od vozu
tvrdý jako z křemene	velký jako kráva
tvrdý jako z oceli	velký jako medvěd
tvrdý jako železo	velký jako nákladník
tvrdý jako žula	velký jako pěst
učenlivý jako opice	velký jako pětník
ukecaný jako stará bába	velký jako skřivánek
úlisný jako had	velký jako tele
umazaný jako uhlíř	velký jako vejce
umíněný jako beran	věrný jako pes
umíněný jako mezek	volný jako motýl
umouněný jako cikáně	volný jako pták
urostlý jako bříza	vyčůraný jako mraky
urostlý jako topol	vychrtlý jako lusk
úskočný jako had	vychrtlý jako pes
uštvaný jako kůň	vyjukaný jako poledne
utahaný jako alík	vylekaný jako králík
utahaný jako čokl	vylekaný jako zajíc
utahaný jako kočka	vymáčkaný jako citrón
utahaný jako kotě	vymóděný jako hadroš
utahaný jako kůň	vypasený jako čuně
utahaný jako pes	vypasený jako louka
utahaný jako štěně	vypasený jako prase
útlý jako proutek	vypasený jako vepř
uvázaný jako pes u boudy	vyplašený jako králík
úzký jako jitrnice	vyplašený jako poledne
veliký jako pěst	vyplašený jako zajíc
veliký jako slon	vyprahlý jako poušť
veliký jako stodola	vyprahlý jako Sahara
velký jako almara	vyprahlý jako troudu
velký jako Brno	vyschlý jako došek
velký jako holubí vejce	vyschlý jako louč
velký jako hrachy	vyschlý jako treska
velký jako kobyla	vyschlý jako trn

vyschlý jako troud	zelený jako sedma
vysoký jako slonbidlo	zkroucený jako paragraf
vysoký jako věž	zkroucený jako preclík
vystrojený jako hadstroš	zlitý jako Dán
vysušený jako treska	zlitý jako prase
vyzáblý jako pes	zlý jako pes
vyžraný jako prase	zmačkaný jako cumel
vyžraný jako zub	zmačkaný jako onuce
vzácný jako šafrán	zmalovaný jako Indián
vzteklý jako čert	zmatený jako Goro před Tokiem
vzteklý jako ďábel	zmrzlý jako drozd
vzteklý jako křeček	zmrzlý jako hovno
vzteklý jako pes	zmrzlý jako kost
záhadný jako smrt	zmrzlý jako preclík
zahrabaný jako krtek	zmrzlý jako rampouch
zalezlý jako krtek	známý jako falešný pětník
zamilovaný jako student	zpitý jako dobytek
zamračený jako noc	zpocení jako myš
zarostlý jako Ezau	zrádný jako had
zarostlý jako Tarzan	zrudlý jako krocan
zavřený jako v kleci	zrzavý jako liška
zazobaný jako sysel	ztuhlý jako hovno
zdravý jako buk	ztuhlý jako rampouch
zdravý jako dub	ztuhlý jako sloup
zdravý jako hřib	ztuhlý jako socha
zdravý jako křen	zvědavý jako Eva
zdravý jako lípa	zvědavý jako opice
zdravý jako pařízek	zvědavý jako opička
zdravý jako ryba	zvědavý jako stará baba
zdravý jako rybička	zvědavý jako stará bába
zdravý jako rys	zvědavý jako straka
zdravý jako řípa	žárlivý jako Othello
zdravý jako tuřín	živý jako rtuť
zelený jako brčál	žízlivý jako poušť

žlutý jako citrón

žlutý jako kanárek

žlutý jako šafrán

žravý jako kobyłka

### **Publications (WoS indexed)**

- Dvořáková, K., Emmer, J., Janktová, R., Klementová, K. (2023). The influence of remote learning environment and use of technology on university students' behavioural engagement in contingency online learning. *Tuning Journal for Higher Education*, 10(2), 271–300. <https://doi.org/10.18543/tjhe.2327>
- Dvořáková, K., Emmer, J., Janktová, R., Klementová, K. (2021). From F2F to ERT: University students' perception of remote learning during the first COVID-19 lockdown. *Journal on Efficiency and Responsibility in Education and Science*, 14(2), 89–100. <https://doi.org/10.7160/eriesj.2021.140203>
- Emmer, J. (2020). Boring as hell: a corpus study of intensifying post-modification of predicative adjectives in the ADJ as NOUN frame. *Linguistica Pragensia*, 2020/2(30), 127–137. DOI: 18059635.2020.2.2

### **Active Conference Participation**

- 2023 Hradec Králové Anglophone Conference 2023  
paper: Understanding words with multiple senses. Polysemy and the implications for EFL vocabulary teaching and learning.
- 2019 Hradec Králové Anglophone Conference 2019  
paper: On the Existence of Linking Semivowels
- 2018 14th ESSE Conference, Masaryk University, Brno, Czech Republic  
paper: Boring as hell: Intensifying postmodification of predicative adjectives of the 'ADJ as NOUN' type in the Spoken BNC2014
- 2018 Hradec Králové Anglophone Conference 2018  
paper: Institutionalized '(as) ADJ as NG' similes in the Spoken BNC2014

### **Pedagogical Activities**

- 2016-present Lecturer of English  
Subjects taught: Lexicology, Phonetics and Phonology, Pragmatics, Stylistics  
Department of English Studies, Faculty of Education  
University of South Bohemia in České Budějovice