

Univerzita Karlova

Filozofická fakulta

Ústav bohemistických studií

Bakalářská práce

Alexandra Maslova

Kolokace lemmat „země“ a „nebe“ v ČNK

Collocations of the lemmas „země“ and „nebe“ in the Czech National Corpus

Praha 2017

Vedoucí práce: PaedDr. Ilona Starý Kořánová

Prohlášení:

Prohlašuji, že jsem bakalářskou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze, dne 4.8.2017

.....

Alexandra Maslova

Ráda bych poděkovala vedoucí práce PaedDr. Iloně Starý Kořánové za trpělivost a cenné rady a své rodině a přátelům za podporu.

Abstrakt

Tato bakalářská práce se zabývá kolokacemi lemmat “nebe” a “země” v Českém národním korpusu. Práce je rozdělená do dvou částí: teoretické a praktické. Teoretická část se zaměřuje na vysvětlení pojmů základních lexikálních jednotek, lemmatu a kolokace. Kromě toho zahrnuje stručný popis frazeologie a idiomatiky a seznámení s Českým národním korpusem. Praktická část obsahuje materiály získané při hledání ve slovnících a v korpusu, jejich srovnávací analýzu a podrobnější popis vyhledaných v korpusu dat.

Klíčová slova: ČNK, kolokace, lemma, frazém

Summary

In this bachelor thesis we are examining lemmas “nebe” and “země” in the Czech National Corpus. The thesis consists of two parts: theoretical and practical. The theoretical part is focused on explaining the terms of basic lexical units, lemmas and collocations. Beyond that, it contains a brief description of phraseology and idiomatics and introduction to the Czech National Corpus. The practical part contains the materials obtained in the search in dictionaries and in the corpus, its comparative analysis and more detailed description of data founded in the corpus.

Keywords: ČNK, collocation, lemma, phrase

OBSAH:

ÚVOD.....	9
TEORETICKÁ ČÁST.....	10
1. Lexikální jednotky v systému a jejich pojetí.....	10
1.1. Slovo, lexém a pojmenování.....	10
1.2. Lexém: podrobnější pojetí.....	11
2. Lemma a lemmatizace.....	13
2.1. Pojetí lemmatu.....	13
2.2. Vztah lemmatu a lexému.....	13
2.3. Pojetí lemmatizace.....	14
2.4. Problémy lemmatizace.....	14
2.5. Proces lemmatizace.....	15
3. Kolokace.....	16
3.1. Obecné pojetí kolokace.....	16
3.2. Kolokační paradigma, kolokabilita, kolokát.....	17
3.3. Typy kolokace.....	17
3.4. Frekvence a asociační míry.....	18
4. Frazém a idiom.....	19
4.1. Idiomatika a frazeologie: jazykové disciplíny.....	19
4.2. Definice frazému a idiomu.....	20

4.3. Formálně-strukturní klasifikace frazémů a idiomů podle F. Čermáka.....	21
4.3.1. Kolokační frazémy.....	21
4.3.2. Propoziční frazémy.....	22
4.3.3. Lexikální frazémy.....	22
5. Korpusová lingvistika a jazykový korpus.....	24
5.1. Co je to korpus?	24
5.2. Typy korpusů.....	25
5.3. Český národní korpus.....	26
PRAKTICKÁ ČÁST.....	27
1. Výzkum 1.: slovníky versus korpus	29
1.1. Slovník české frazeologie a idiomatiky.....	29
1.2. Postup práce při výzkumu 1.	29
1.3. Lemma „nebe“ ve slovnících.....	30
1.4. Lemma „země“ ve slovnících.....	39
1.5. Výsledky výzkumu 1.	52
1.6. Lidová rčení J. Zaorálka.....	53
2. Výzkum 2: lemmata „nebe“ a „země“ v ČNK.....	56
2.1. Postup práce při hledání v ČNK.....	56
2.2.1. Lemma „nebe“ v ČNK.....	57

2.2.2. Analýza získaného materiálu.....	62
2.3.1. Lemma „země“ v ČNK.....	63
2.3.2. Analýza získaného materiálu.....	68
ZÁVĚR.....	70
Seznam použité literatury.....	72
Seznam použitých slovníků.....	73
Seznam použitých internetových zdrojů.....	73

Seznam zkratk:

SČFI1	ČERMÁK, František, Jiří HRONEK a Jaroslav MACHAČ. Slovník české frazeologie a idiomatiky. 1, Přirovnání. Praha: Leda, 2009. ISBN 978-80-7335-215-8.
SČFI2	ČERMÁK, František, Jiří HRONEK a Jaroslav MACHAČ. Slovník české frazeologie a idiomatiky. 2, Výrazy neslovesné. Praha: Leda, 2009. ISBN 978-80-7335-215-8.
SČFI3	ČERMÁK, František, Jiří HRONEK a Jaroslav MACHAČ. Slovník české frazeologie a idiomatiky. 3, Výrazy slovesné. Praha: Leda, 2009. ISBN 978-80-7335-215-8.
SČFI4	HRONEK, Jiří, Jaroslav MACHAČ a František ČERMÁK. Slovník české frazeologie a idiomatiky. 4, Výrazy větné. Praha: Leda, 2009. ISBN 978-80-7335-215-8.
SČFI5	ČERMÁK, František a Jan HOLUB. Slovník české frazeologie a idiomatiky. 5, Onomaziologický slovník. Voznice: Leda, 2016. ISBN 978-80-7335-440-4.
LR	ZAORÁLEK, Jaroslav. Lidová rčení. Vyd. 3., reprint 1. vyd. z r. 1947. Praha: Aurora, 1996. ISBN 80-85974-10-X.

apod.	a podobně
atd.	a tak dále
ČNK	Český národní korpus
jaz.	jazykový
IF	Idiomatika a frazeologie
n.	nebo
náb.	náboženský
např.	například

resp.	respektive
srov.	srovnat
popř.	popřípadě
tj.	to jest, to je
tzn.	to znamená
tzv.	takzvaný
vs.	versus

ÚVOD

Frazeologie a idiomatika je docela mladá větev lingvistiky, která vznikla ve 20. století a stále se velmi aktivně rozvíje. Frazémy a idiomy jsou nedílnou součástí jazyka a mění se spolu s vývojem společnosti. Velice důležitou událostí pro rozvoj české frazeologie a idiomatiky jako vědecké disciplíny bylo založení Českého národního korpusu na konci 20. století, který umožňuje přístup k rozsáhlým jazykovým datům.

Táto práce se zabývá zkoumáním kolokací lemmat „nebe“ a „země“. Cílem dané práce je zjistit, zda se všechny frazémy a idiomy s lexémy „nebe“ a „země“, zachycené ve slovnících, vyskytují v ČNK či ne. Dalším cílem je třídění materiálu, získaného v korpusu, respektive vyhledávání nových frazémů, dosud nezaznamenaných slovníky.

Práce je rozdělená na dvě části – teoretickou a praktickou.

Předmětem zájmu teoretické části je vymezení takových základních pojmů, jako slovo, lexém, lemma, lemmatizace, kolokace, frazém a dalších. Kromě toho daná část obsahuje stručný popis frazeologie a idiomatiky jako vědecké disciplíny a přehled klasifikace frazémů a idiomů podle F. Čermáka. Poslední kapitola je věnována seznámení s jazykovými korpusy včetně Českého národního korpusu.

Praktická část obsahuje dva výzkumy. První výzkum je zaměřen na srovnávací analýzu dat, nalezených ve slovnících a materiálu, získaného v korpusu. Druhý výzkum je věnován přímo zpracování, třídění a rozboru frazémů a idiomů, vyhledaných v korpusu. Nalezený materiál je uspořádán do tabulek pro lepší přehlednost.

Údaje k analýze byly získány z Českého národního korpusu řady SYN, konkrétně synv4, a dále také ze slovníku, a to především ze *Slovníku české frazeologie a idiomatiky* (F. Čermák, J. Hronek) a ze slovníku *Lidová rčení* (J. Zaorálek).

Táto práce může být přínosná pro studenty bohemistiky, kteří chtějí rozšířit slovní zásobu a prozkoumat frazeologii a idiomatiku jako lingvistickou disciplínu. Rovněž může posloužit k odstranění chyb v korpusu, nalezených v procesu vyhledávání informace.

TEORETICKÁ ČÁST

1. Lexikální jednotky v systému a jejich pojetí

1.1. Slovo, lexém a pojmenování

V knize F. Čermáka *Lexikon a sémantika* (2010, s. 17-26) v kapitole *Slovo, lexém a pojmenování* se František Čermák zaměřuje na problém vymezení pojmu základní (minimální, tradiční) lexikální jednotky, která má svoji funkci. F. Čermák nepochybuje, že za takovou jednotku v moderní lexikologii považujeme lexém, ale dodává, že na něj „však navazují, resp. mu předcházejí další dva termíny, pojmenování (nominace) a slovo.“¹

Na začátku se F. Čermák zaměřuje na vysvětlení pojmu *slovo*, které bylo dlouho chápáno jako základová jednotka jazyka: „Tradiční a starou jednotkou je **slovo**, obvykle chápané intuitivně jen jako jediná diskrétní forma, tj. celek psaný dohromady (grafické slovo). V lingvistice patří vůbec k nejstarším pojmům s dlouhou a proměnlivou historií.“²

Další definicí, kterou autor uvádí, je termín lexém: „**Lexém** je systémová (tj. ustálená, langueová) formálně samostatná, avšak funkčně a abstraktně chápaná znaková jednotka, která je formálně jednoduchá nebo složená (víceslovná), tj. tvořená jediným tvarem nebo úhrnem všech příslušných (flektivních) tvarů patřících funkčně k sobě.“³ Podle Čermáka se v češtině implikuje škála „od (a) jednoslovného lexému flektivně neměnného (*ale, že, relé<...>*) přes dva typy funkčních grup tvarů, (b) jednoslovných (*otec, otce, otci, otců*, tj. všech 10 flektivních tvarů lexému *otec<...>*), a (c) grup jen textově víceslovných (*byl bych šel* atd., lexém *jít, <...>*), až po grupy vždy, tj. i systémově, víceslovné, a tedy i dva typy kombinací systémově víceslovných, a to (d) víceslovné termíny (*orgány činné v trestním řízení, kyselina siřičitá*) a (e) frazémy (*vzít do zaječích, <...>*) včetně všech jejich flektivních tvarů (tj. (d) i (e)).“⁴

Se zřetelem k základní jednotce, tj. *lexému*, vymezuje pak autor i další dvě, *slovo* a *pojmenování*. „Slovo je jediný konkrétní a textový (parolový) tvar lexému, popř. i sám lexém, jde-li o jednoslovný a formálně neměnný lexém.“⁵ Musíme zde vysvětlit, že obzvlášť u flexivních jazyků

¹ČERMÁK, F. (2010, s. 17)

²ČERMÁK, F. (2010, s. 17)

³ČERMÁK, F. (2010, s. 17)

⁴ČERMÁK, F. (2010, s. 17-18)

⁵ČERMÁK, F. (2010, s. 18)

(na rozdíl od jazyků bez flexe) dochází k tomu, že přirozená jazyková jednotka může být “rozbita” (např. *stydím* je čistě mechanicky chápáno jako jeden tvar, zatímco *se* jako druhý). Uvedený příklad slouží jako potvrzení toho, že parolové hledisko má své nevýhody a nemůže být vždy funkčním.

Nakonec se Čermák dostává k vysvětlení termínu *pojmenování* (nominace), a to přes popis vztahu lexému a slova:

„U vztahu lexém a slovo v dnešním pojetí jde tedy o souvztažnost langue-parole, kdy se první, tj. jednotka systémová, realizuje v textu v *tvar*, jednotku druhou. Avšak lexém vnímaný procesuálně, resp. z onomaziologického hlediska, kdy se vnímá ne jeho realizace, ale tvoření (ať už skutečné či jen ve smyslu přechodu od významu k formě), se často nazývá jinak. Protože v tomto odlišném pohledu mluvčí svou výchozí myšlenku, obsah, představu (psychologické východisko) lexikálně pojmenovává, označuje slovem či slovy, tj. dává jí formu (jazykový rezultat), nazývá se takováto jednotka případně také **pojmenování (nominace)**.“⁶

Nominace je formálně shodná s některou z podob lexému uvedených výš, co se ale týče jejich uplatnění v praxi, jde tady o dva různé procesy. Častějším případem je myšlenková volba nějakého z ustálených, již existujících lexémů, a to buď jednoslovného nebo víceslovného (např. až spadneš z nebe místo nikdy). Druhou situací je skombinování lexémů do víceslovného označení tak podrobného a dlouhého, jak ho bude mluvčí potřebovat, v takovém případě mluvíme o pojmenování (např. jedno neustálené pojmenování v 8 slovech *tamten malý krásný domeček tamhle vlevo toho parku*).

1.2.Lexém: podrobnější pojetí

Podívejme se podrobněji na pojem lexém, který považujeme za základní lexikální jednotku.

V *Encyklopedickém slovníku češtiny* (J. Bachmannová, 2002) najdeme toto pojetí lexému a těsně souvisejících s ním odvozenin:

„Abstraktní formálně významová jednotka lexikonu nadřazená všem svým konkrétním manifestacím v řeči, u ohebných slov zahrnující celá paradigma (*miluji, miloval jsem, budu*

⁶ČERMÁK, F. (2010, s. 18)

milovat...; dívka, dívky, dívkou...). Konkrétní manifestace lexému bývá označována jako **lex**, z hlediska variantnosti jako **alolex**; též typ a token. Víceznačný/ polysémní lexém nazývají někteří lingvisté **hyperlexém**, jeho jednotlivé monosémní lexémy lexie (tak zejména Filipec v Člex, 1985). Většinou synonymně s termínem se dříve užíval termín **lexikální jednotka**, synonymně s termínem lexie termín **základní lexikální jednotka**. Lexémy mohou být jednoslovné i víceslovné (víceslovností není míněná členitost gramatických tvarů typu *miloval jsem*). Víceslovnými lexémy jsou ustálené pravidelné i nepravidelné kolokace, tj. víceslovné termíny a frazémy-idiomy (*např. kysličník uhličitý, srovnávací jazykověda; máslo na hlavě, vyrazit trumf z ruky*).⁷

V své publikaci *Korpusová lingvistika: první vstup* se M. Šulc přiklání k této definici lexému:

„Lexém=lexikální jednotka=jednotka slovníku, slovní zásoby. Uvažujeme-li o slově jako o „lexému“, nezajímá nás jeden jeho konkrétní souhrn gramatických kategorií, vytažený z textu (kontextu), ale je pro nás abstraktní jednotkou lexikonu, a jako takový zahrnuje celé morfologické paradigma (všechny flektivní tvary). Základní slovníkový (reprezentativní) tvar slova/lexikální jednotky.“⁸

Můžeme se tady všimnout toho, že M. Šulc ve svém díle zaměřeném přímo na korpusovou lingvistiku bere pojmy lexém a lexikální jednotka jako dva rovnocenné. Budeme se v této bakalářské práci držet daného hlediska.

⁷BACHMANNOVÁ, J. (2002, s. 242)

⁸ŠULC, M. (1999, s. 77)

2. Lemma a lemmatizace

2.1. Pojetí lemmatu

Přestože považujeme lexém za základní lexikální jednotku, v dané práci se zabýváme výzkumem dat získaných v korpusu, proto bychom potřebovali obrátit svou pozornost na další termín, termín **lemma**.

Encyklopedický slovník češtiny od J. Bachmannové uvádí toto pojetí lemmatu:

„Jednotlivý slovníkový vstup. Reprezentativní podoba lexému, charakterizovaná jistou gramatickou formou (u jmen nejčastěji nom. sg., u sloves v č. inf.), jistým pravopisem a jistou výslovností, popř. rozrůzněná dubletou nebo variantností. V klasických slovnících uvádí lemma heslový odstavec (v tomto případě se užívá spíše termínu **heslové slovo**). Termín lemma je běžný zejména v korpusové lexikografii. Tzv. **lemmatizací** se přiřazuje lemma textovým tvarům lexémů (slovům).“⁹

M. Šulc charakterizuje lemma jako „slovní tvar, který je dohodou vybrán (pro potřeby zejména slovníkářství) jako reprezentující celé paradigma. Stojí v hlavičce slovníkového hesla.“¹⁰

Lemma představuje reprezentativní slovníkovou podobu hesla, která se přiděluje každé formě v korpusu během automatického zpracování jazyka v procesu lemmatizace.

Obecně platí, že

- lemmatem každého českého podstatného jména je forma jeho **nom. sg.** (např. tvary *nebem, nebí, nebích* mají lemma **nebe**)
- u přídavných jmen je to **nom. sg. mask.** (např. tvary *dobrého, dobrou, dobrých* mají lemma **dobrý**)
- u sloves je to **inf.** (tvary *dělal, dělám, děláme* mají lemma **dělat**).

2.2. Vztah lemmatu a lexému

Z hlediska určitých aspektů jsou lemma a lexém analogické, liší se ovšem tím, že lemma je vždycky jednotka jednoslovná a „<...>“ ve většině korpusů koresponduje s výskytem slovního

⁹BACHMANNOVÁ, J. (2002, s. 242)

¹⁰ŠULC, M. (1999, s. 77)

tvaru v poměru 1:1. V každém korpusu je tedy počet lemmat menší, než počet slovních tvarů (např. ve stomilionovém korpusu SYN2010 najdeme 1,7 mil. různých slovních tvarů, ale jenom 786 tis. různých lemmat).¹¹

2.3. Pojetí lemmatizace

Proces **lemmatizace** je součástí automatické morfologické anotace (anotace je „proces, při němž se ručně či automaticky připojují interpretační lingvistické, strukturní údaje a/nebo metatextové údaje k textovým datům korpusu.“¹²) Zásada lemmatizace spočívá v tom, že se každému slovnímu tvaru přiřazuje lemma v korpusu.

„Lemmatizace je typicky součástí procesu desambiguace (zjednoznačnění) slovních tvarů v textu na základě kontextu. Lemmatizace je jednoduchá a nekontextová (patří-li lemmatizovaný slovní tvar k paradigmatu jediného lexému, např. slovesnému tvaru *vytvoříme* bude přiřazeno lemma *vytvořit* jakožto reprezentativní podoba příslušného slovesného lexému bez ohledu na kontext).“¹³

2.4. Problémy lemmatizace

Automatická lemmatizace s sebou přináší i jisté komplikace. Lemmatizovaný slovní tvar může být například homonymní (to znamená, že patří k paradigmatům víc než jednoho lexému), a v takovém případě se na základě kontextu má rozhodnout, které lemma se přiřadí tomu či jinému tvaru. Při automatické lemmatizaci, stejně jako při každém automatickém procesu, může dojít k chybě, proto by měl uživatel korpusu dávat pozor na získané výsledky (případně je kontrolovat).

Na webové stránce <https://wiki.korpus.cz/doku.php/pojmy:lemma> (slovníček z korpusové lingvistiky-lemma) jsou zaznamenané následující problémy spojené s lemmatizací: „Velkým lingvistickým i počítačovým problémem je lemmatizace víceslovných spojení. Jiným neřešeným problémem při automatické lemmatizaci je lemmatizace všech tvarů pod jediné lemma i tam, kde to není patřičné: např. zdvořilá prosba o dovození projít *Dovolíte?* se neodráží v žádném

¹¹ dostupné na: <<https://cs.wiktionary.org/wiki/lemma>>

¹² dostupné na: <<https://wiki.korpus.cz/doku.php/pojmy:anotace>>

¹³ dostupné na: <<https://wiki.korpus.cz/doku.php/pojmy:lemma>>

z registrovaných významů slova *dovolit*, protože není součástí výlučně morfologické lemmatizace. Podobné je to i u frazémů, kde nelze tvar *holičkách* (frazému *nechat na holičkách*) lemmatizovat jako *holičky* (tvar, který navíc vůbec neexistuje).¹⁴

2.5. Proces lemmatizace

Automatická lemmatizace se realizuje pomocí počítačového programu zvaného lemmatizátor, „<...>“ který bývá součástí morfologického taggeru <...>.¹⁵ Smysl lemmatizace spočívá jednak v identifikaci náležitého lexému u homonymních slovních tvarů v daném kontextu, jednak v tom, aby uživatel korpusu měl možnost pracovat nejen se slovními tvary, ale i s lemmaty reprezentujícími příslušné lexémy a jejich paradigmata (a to v podstatě usnadňuje práci s korpusem).

¹⁴dostupné na: <<https://wiki.korpus.cz/doku.php/pojmy:lemma>>

¹⁵dostupné na: <<https://wiki.korpus.cz/doku.php/pojmy:lemma>>

3. Kolokace

3.1. Obecné pojetí kolokace

Pojem kolokace (z lat. *collocare* = spolu-umístit)¹⁶ vznikl ve 30. letech 20. století, zavedl ho do jazykovědy J. R. Firth (1957), do češtiny se daný termín dostal díky F. Čermákovi (1982).

Termín *kolokace* v rámci korpusové lingvistiky znamená „smysluplné, ustálené, syntagmatické spojení dvou (nebo víc) slovních tvarů (někdy celých lexémů) v blízkém kontextu.“¹⁷ Dříve se pro označení daného jevu používal termín *slovní spojení*.

Michal Šulc uvádí dané pojetí kolokace ve výběrovém slovníčku pojmů z lingvistiky ve svém díle *Korpusová lingvistika: první vstup*:

„Kolokace je zvláště obvyklý/častý (a tedy do určité míry předpověditelný) souvýskyt (=společný výskyt) samostatných lexikálních jednotek. Je realizací syntagmatického vztahu uvnitř lexikonu a k jeho vzniku většinou lze nalézt odpovídající interpretaci.“

V díle *Kolokace* F. Čermák vymezuje kolokace takto:

„Kolokace je smysluplné spojení slov/ lexémů, resp. lexikální syntagma, zvl. v podobě víceslovných pojmenování, jehož vznik je podmíněn jejich vzájemnou kolokabilitou, a tedy i kompatibilitou.“¹⁸

V téže práci F. Čermák poznamenává, že „od kolokace je třeba odlišit bigram, trigram, tetragram atd., tj. dvou-, troj-, čtyř – či vícečlennou kombinaci, resp. postupné a po sobě jdoucí kombinace dvou, tři, čtyř atd. všech sousedních slov, kterou z celého textu, resp. korpusu mechanicky, postupně a poměrně jednoduše vyčleňuje některý z mnoha už dostupných programů <...> Výsledkem jsou tedy např. všechny dvojice (trojice atd.) sousedních slov. To znamená, že v takto mechanickém a obvykle velkém úhrnu kombinací je i mnoho kombinací nesmyslných,

¹⁶ČERMÁK, F., ŠULC, M. (2006, s. 10)

¹⁷dostupné na: <<https://wiki.korpus.cz/doku.php/pojmy:kolokace>>

¹⁸ ČERMÁK, F., ŠULC, M. (2006, s. 10)

náhodných apod. Teprve z úhrnu bigramů lze pak získávat, obvykle poměrně pracně, smysluplné kolokace.¹⁹

3.2. Kolokační paradigma, kolokabilita, kolokát

„Pro slovo, které se k výchozímu slovu druží jako komponent dané kolokace, se razí termín *kolokát*.²⁰ Pojmem *kolokační paradigma* označujeme seznam kolokátů jednoho slova, tedy potenciálních syntagmat lexému. Kolokabilitou slova se pak někdy nazývá jeho „kombinatorická schopnost, resp. syntagmatický potenciál.“²¹

F. Čermák v svém díle *Frazeologie a idiomatika česká a obecná* popisuje kolokabilitu takto:

„Kolokabilita je obecně individuální formální a významová spojitelnost, schopnost kombinace každého jazykového prvku s jiným nebo jinými, vytvářená jeho jedním nebo více kolokačními paradigmaty a (u pravidelných kombinací) podmíněná jeho kompatibilitou s nimi; spolu s valencí patří k hlavním složkám syntagmatickosti každého jazykového prvku. Individuální realizaci kolokability vzniká konkrétní kolokace.“²²

3.3. Typy kolokace

Na základě údajů z *Encyklopedického slovníku češtiny* (J. Bachmannová, J. Balhar, R. Blatná, 2002) můžeme vyčlenit několik typů kolokace:

(1a) **systémové pravidelné kolokace**, to znamená víceslovné termíny, např. *kyselina sírová*; (2a) **systémové nepravidelné kolokace**, to jest frazémy a idiomy, např. *růžové brýle*. Dané rozdělení je založeno na distinkci systém-text (ustálenost-neustálenost)²³. Kromě toho můžeme rozlišovat **textovou pravidelnou kolokaci**, tj. běžnou (např. *letní dovolená*), a **textovou nepravidelnou kolokaci**, pro kterou je charakteristické užívání individuálních autorských metafor (např. *třeskutě vtipný*).

¹⁹ČERMÁK, F., ŠULC, M. (2006, s. 11)

²⁰ČERMÁK, F., ŠULC, M. (2006, s. 11)

²¹ČERMÁK, F., ŠULC, M. (2006, s. 11)

²²ČERMÁK, F. (2007, s. 29)

²³BACHMANNŮV, J. (2002, s. 218)

3.4. Frekvence a asociační míry

Důležitým pojmem, spojeným s kolokací, je frekvence. Termín *frekvence* označuje v korpusové lingvistice počet výskytů jevu v korpusu, „četnost formy nebo jevu. Uvádí se buď v **absolutních hodnotách**, např. lemma *pes* se ve 100 milionovém korpusu SYN2010 vyskytuje 17 701krát, nebo jako údaj **relativní**, např. lemma *pes* se v SYN2010 vyskytuje (po zohlednění rozdílného počtu slov a pozic v korpusu) 145krát na milion slov (používané zkratky jsou **ipm** = instances per million nebo **ppm** = parts per million).“²⁴

Absolutní frekvence vyžaduje pro zhodnocení prostého počtu výskytů slov v korpusech další upřesnění, zatímco relativní frekvence ukazuje četnost jevu sama o sobě a dává možnost porovnat mezi sebou korpusy nebo texty různé velikosti.

Dalším termínem těsně spojeným s kolokací je asociační míra. Pojmem *asociační míra* (z angl. Association measures) označujeme matematické vzorce (postupy), které používáme pro vyhledání kolokací v korpusu. K nejstarším měřám, kterých se užívá nejvíc, patří MI-score (index vzájemné informace) a t-score (t-test).

„MI-score, míra vzájemné informace, měří sílu asociace mezi dvěma slovy, u níž jde v zásadě o podíl pravděpodobnosti výskytu těchto dvou slov spolu a výskytu každého z těchto slov nezávisle. <...> T-test, resp. t-score, je test signifikantnosti založený na rozdílu mezi pozorovaným a předpokládaným výsledkem. Také se mu říká míra kontrastu, protože je lepší mimo jiné pro frekventovaná a přehlížená slova, pro funkční slova apod.“²⁵

²⁴ dostupné na: <<https://wiki.korpus.cz/doku.php/pojmy:frekvence>>

²⁵ ČERMÁK, F., ŠULC, M. (2006, s. 13)

4. Frazém a idiom

4.1. Idiomaticita a frazeologie: jazykové disciplíny

Idiomaticita a frazeologie (dále IF) je poměrně mladou lingvistickou disciplínou, která se zabývá studiem a popisem frazémů a idiomů.

V *Encyklopedickém slovníku češtiny* od Bachmannové se IF popisuje takto:

„Frazeologie a idiomaticita je oblast frazémů a idiomů všech úrovní; <...> Nejde o zvláštní jaz. rovinu, ale o oblast ustálených anomálních kombinací, která je komplementární k oblasti pravidelných jaz. kombinací jednotek všeho druhu a na všech relativních rovinách.“²⁶

Velmi těžce se dá určit přesná hranice mezi frazeologií a idiomaticitou, protože jako větve lingvistiky je IF těsně spojená s dalšími jazykovědnými disciplínami a svým způsobem na ně navazuje nebo z nich vyplývá. Dané téma popisuje F. Čermák v své publikaci *Frazeologie a idiomaticita česká a obecná* (2007), autor dokonce věnoval této otázce celou kapitolu, podíváme se ni podrobněji.

Na tvorbě frazémů a idiomů se dřív účastnila jak syntax, tak i morfologie. Je jisté, že frazémy existují už na úrovni kombinace morfému, v **morfologii**, a to potvrzují například frazémy „uvnitř“ hranic slova: *měkkosrdcatý* vs. *mít měkké srdce*, *svéhlavý* vs. *mít svou hlavu*.²⁷ V obou případech mezi výrazy není zásadní rozdíl, i když to starší teorie neuznávala a prezentovala je jako dvě různé věci.²⁸

Vztah **syntaxe** a frazeologie (eventuálně jejich restrikce) je velice zajímavý jazykový jev. Tak například synchronně v pravidelném jazyce neexistuje *ležet něčím*, ale v IF ano, nebo například nemůžeme tvořit kombinace substantivum + adverbium (podle syntaktických pravidel), a naproti tomu existuje frazém *místo nahoře*.²⁹

Frazeologie se propojuje s **fonologií**, a to zvláště kvůli tomu, že frazémy větné mají svou specifickou intonaci, srov. *To zrovna!*, vyslovují se s jedinou specifickou intonací.³⁰ IF je pak i

²⁶ BACHMANNOVÁ, J. (2002, s. 142)

²⁷ ČERMÁK, F. (2007, s. 14)

²⁸ ČERMÁK, F. (2007, s. 14)

²⁹ ČERMÁK, F. (2007, s. 14)

³⁰ ČERMÁK, F. (2007, s. 14)

významným prostředkem jazykové **stylistiky**, jde tady především o frazémy mluvené a nespisovné; podobnou roli zde hraje z více hledisek i **rétorika**.

Vzhledem k tomu autor poznamenává:

„I fakt většinové příslušnosti české IF do mluveného (a nespisovného) jazyka ukazuje mimochodem na omezenost a nefunkčnost teorie spisovnosti. Vzato striktně by pak spisovná čeština neměla mít frazeologii (téměř) žádnou.“³¹

K dalším jazykovým disciplínám, které mají vztah k frazeologii, můžeme přiřadit **dialektologii**, historicky pak málo studovanou **etymologii**, volně i lingvistickou **sémiotiku**.

Na konci kapitoly F. Čermák dochází k následujícímu závěru:

„Dodejme ještě, že **mezijazykově**, zvláště v oblasti srovnání a učení jazyků, se frazeologie a idiomatika spíše intuitivně odjakživa chápaly jako cosi zvlášť obtížného a nesnadno přenositelného z jednoho jazyka do druhého. <...> Frazémy a idiomy jsou jazykové universale, jejich vznik a distribuce jsou však velmi specifické.“³²

V další kapitole *IF a nejjazyk: jiné obory a disciplíny* autor mluví o vztahu frazeologie a nelingvistických věd. Čermák tu označuje, že IF je ovlivněná či nějak navazuje na **historii**, **antropologii**, **literaturu**, **hudbu**, **malířství**, **psychologii** atd.

4.2. Definice frazému a idiomu

Frazému a idiomu tradičně rozumíme jako ustálenému a reprodukovatelnému spojení slov, jehož význam je buď zčásti, nebo zcela neodvoditelný z významu jeho komponentů. Toto vymezení však nevyhovuje, protože nepokrývá všechny typy ani všechny roviny.³³ V podstatě platí, že pokud analyzujeme kombinatorický útvar z hlediska *formálních* rysů, mluvíme o **frazému**, zatímco analýza z hlediska relativních *sémantických* rysů opravňuje užití pojmu **idiom**.

V *Encyklopedickém slovníku češtiny* najdeme tuto definici:

³¹ ČERMÁK, F. (2007, s. 14)

³² ČERMÁK, F. (2007, s. 15)

³³ BACHMANNOVÁ, J. (2002, s. 139)

„<...> je to taková jedinečná ustálená kombinace minimálně dvou prvků, z nichž některý (popř. žádný) nefunguje stejným způsobem v žádné jiné kombinaci n. více kombinacích, resp. vyskytuje se v takové funkci pouze ve výrazu jediném, popř. několika málo.“³⁴

F. Čermák však (v díle *Frazeologie a idiomatika česká a obecná*, v němž je zahrnutá skoro stejná definice) poukazuje na to, že výše uvedené pojetí nevyhovuje všem typům frazémů, a proto nemůže být univerzální. Podle Čermáka shromáždění všech vlastností frazému a idiomu do jednoznačné konkrétní definice v podstatě skoro není možné. Pokud bychom zdůraznili to, že všechny vlastnosti jsou jen projevy vlastností jazykového systému a struktury³⁵, mohli bychom pak vyčlenit shrnutý a zobecněný pojem, podle kterého frazém a idiom je „<...> takové nemodelové a ustálené syntagma prvků, z nichž (aspoň) jeden je z hlediska druhého členem extrémně omezeného a zavřeného paradigmatu (formálně a většinou i sémanticky).“³⁶

4.3. Formálně-strukturní klasifikace frazémů a idiomů podle F. Čermáka

Hledisek, podle kterých můžeme členit frazémy a idiomy, je hodně, v dnešní lingvistice však nejsou všechna dokonalá: pro svou přílišnou obecnost, pro nedostatek spolehlivých pramenů a informací, pro náročnost a neujasněnost kritérií atd. V dané bakalářské práci se zaměříme na klasifikaci formálně-strukturní, kterou upřednostňuje F. Čermák. Formálně-strukturní klasifikace je založená na evidenci vstupních komponentů a rozeznává frazémy *kolokační*, *propoziční* a *lexikální*.

4.3.1. Kolokační frazémy

Kolokační frazém je „frazém s funkcí kolokace (nevětné kombinace lexikálních tvarů) a s komponenty z roviny jednoslovných lexémů.“³⁷ Patří sem frazémy „tradiční“, které známe intuitivně. Komponenty kolokačních frazémů a idiomů jsou slova, která netvoří sama o sobě promluvu; díky různosti struktur a počtu výrazů je tato třída nejvýraznější oblastí IF.³⁸ Podle Čermáka se frazémy této třídy rozkládají do tří velkých subtypů: 16 základních binárních kombinací a struktur autosémantických komponentů (např. *verbum-substantivum*, *verbum-*

³⁴ BACHMANNOVÁ, J. (2002, s. 139-140)

³⁵ ČERMÁK, F. (2007, s. 15)

³⁶ ČERMÁK, F. (2007, s. 32)

³⁷ BACHMANNOVÁ, J. (2002, s. 141)

³⁸ ČERMÁK, F. (2007, s. 44)

adverbium, srov. *číst někomu levity, dopadnout blbě*), 7 základních binárních kombinací a struktur autosémantických komponentů (např. partikule-konjunkce, srov. *ještě že*) a 8 kombinací a struktur smíšených z komponentů obou subtypů (např. prepozice-substantivum, srov. *do omrzení*).³⁹

4.3.2. Propoziční frazémy

Další skupinu představují **propoziční frazémy**. Na základě množství zúčastněných mluvčích a promluv rozlišujeme frazémy a idiomy **propoziční a polypropoziční monosubjektové**, proti kterým stojí frazémy **polypropoziční intersubjektové**.

„Frazémy a idiomy propoziční, které jsou formálně již hotovou propozicí (promluvou), jsou vytvářeny kombinací komponentů z roviny lexikální nebo kolokační, zatímco frazémy a idiomy polypropoziční jsou vytvářeny z komponentů z roviny propoziční (přítomnost aspoň dvou propozic). Oba typy mají jediného společného mluvčího.“⁴⁰

Pragmatická funkce propozičních frazému může mít pět základních podob, na což navazuje následné členění:

- Fakultativní třída frazémů (F) slouží k vyjádření běžné informace různého druhu;
- Voluntativní (direktivní) třída frazémů (V) slouží k vyvolání nějakého účinku u posluchače, např. *Neřeš to!*
- Expresivní třída frazémů (Ex) slouží k vyjádření postoje mluvčího k situaci, např. *Klobouk dolů!*
- Emocionální třída frazémů (Em) slouží k sdělení své emocionální reakce vůči posluchači, např. *To je ale kus!*
- Deklarativní třída frazémů (D) slouží k vyjádření nastolení nějaké situace, např. *Hej rup!*

4.3.3. Lexikální frazémy

Do poslední skupiny v Čermákově klasifikaci zařazujeme **lexikální frazémy**, které „patří svými komponenty do roviny morfémové (slovotvorní a flektivní) a mají jedinou syntetickou,

³⁹ BACHMANNOVÁ, J. (2002, s. 141)

⁴⁰ ČERMÁK, F. (2007, s. 53)

nediskrétní formu.⁴¹ Kompletní popis dané nejspornější oblasti frazeologie zatím nebyl vytvořen, i když lexikální frazémy jsou stejně anomální jako frazémy jiných typů. Funkčně jsou frazémy této skupiny jednoslovným lexémem, jako příklad můžeme uvést kompozitum *každopádně* nebo prefixální deriváty slovesa *být* - *nabýt*, *přibýt* atd.

⁴¹ ČERMÁK, F. (2007, s. 60)

5. Korpusová lingvistika a jazykový korpus

“Korpusová lingvistika je odvětví lingvistiky, jehož rozvoj je spjat s příchodem počítačů a informačních technologií. Tato disciplína zkoumá jazyk pomocí elektronických jazykových korpusů a zabývá se i výstavbou těchto korpusů, jejich zpracováním a příslušnou metodologií.”⁴²

Obecný význam korpusové lingvistiky „spočívá ve snaze prostřednictvím korpusu lépe poznat funkce a struktury přirozeného jazyka.”⁴³

I když je korpusová lingvistika poměrně mladou disciplínou, je to velice perspektivní a rychle se rozvíjející odvětví aplikované jazykovědy. Jelikož se daná bakalářská práce zaměřuje na výzkum dat získaných v ČNK, uvedeme tady základní informaci o korpusu obecně, o jeho typech a o Českém národním korpusu.

5.1. Co je to korpus?

Slovo *korpus* má latinský původ (*corpus*, *-oris* = 1. tělo, těleso hmota 2. tělo, postava 3. uspořádané těleso, celek, kmen, soubor, sbor) a přišlo do češtiny z angličtiny (*corpus*, pl. *corpora* nebo *corpuses*), kde znamená jednak sbírku všech psaných textů jednoho určitého druhu nebo jednoho člověka, jednak sbírku informací či materiálů, určených ke studiu.⁴⁴

V *Encyklopedickém slovníku češtiny* (2002) lze najít následující definici a popis smyslu jazykového korpusu:

„Jazykový korpus je vnitřně strukturovaný a unifikovaný soubor zpravidla elektronicky (počítačově) uložených jazykových dat, a to většinou v textové podobě. Skladba dat jazykového korpusu (textů) je koncipována s ohledem na určitý cíl (např. vytvořit textovou a lexikální základnu pro tvorbu slovníků odrážející slovní zásobu daného období), vůči němuž je pak považován za více méně reprezentativní. Smysl jazykového korpusu tkví v tom, že textová data se v něm nacházejí ve svém přirozeném kontextu, a pokud je jazykový korpus dostatečně rozsáhlý (v řádu milionů slovních výskytů),

⁴²dostupné na: <https://cs.wikipedia.org/wiki/Korpusová_lingvistika>

⁴³BACHMANNOVÁ, J. (2002, s. 248)

⁴⁴ŠULC, M. (1999, s. 9)

zachycuje především základní a typická užití jazykových jednotek a jejich skutečnou frekvenční distribuci.“

Michal Šulc v své práci *Korpusová lingvistika: první vstup* uvádí další pojetí korpusu:

„Korpus je rozsáhlý soubor elektronických textů, cíleně shromážděný jako referenční zdroj pro vědecké studium jazyka a pro zpracování užitečných jazykových nástrojů, který je v jednotném formátu, je lingvisticky označovaný a který lze z hlediska skladby považovat za jistým způsobem vyvážený.“

Korpus může sloužit jako zdroj materiálu nejenom lingvistům, „ale i širokému spektru zájemců z jiných oborů k mnohostrannému poznání jazyka, zákonitostí lidského myšlení a kultury“.⁴⁵

5.2. Typy korpusů

Korpusové projekty lze dělit podle různých kritérií, následující přehled, popsany M. Šulcem v publikaci *Korpusová lingvistika: první vstup*, je soustředěn na kritéria úzce související s jazykem.

Podle množství obsažených jazyků můžeme korpusy dělit na **korpusy jednoho jazyka** a **paralelní korpusy**.⁴⁶ Korpusem jednoho jazyka (jednojazyčným) je například Český národní korpus. Paralelní korpus (parallel corpora) obsahuje stejné texty ve dvou nebo více jazycích (jde většinou o originál a překlady).

Podle časového záběru dělíme korpusy na **diachronní** a **synchronní**. Diachronní korpusy (diachronic corpora) obsahují jazykový materiál z delšího časového období a slouží právě k poznání jazyka v jeho vývoje; synchronní korpusy (synchronic corpora) se zabývají texty z daného, zpravidla kratšího období a zkoumají jazyk v daném období bez zřetele k jeho vývoji.

Některé synchronní korpusy byly sestaveny za účelem blíže nespecifikovaného lingvistického výzkumu.⁴⁷ Nedávají přednost žádné jazykové rovině, oblasti lingvistiky či stylové příslušnosti. Říká se jim **všeobecné** korpusy (general corpora). Všeobecné korpusy slouží většinou ke studiu

⁴⁵ŠULC, M. (1999, s. 11.)

⁴⁶ŠULC, M. (1999, s.12.)

⁴⁷ŠULC, M. (1999, s.12.)

slovní zásoby, gramatiky nebo diskurzu; bývají také někde v literatuře citovány jako **základní** (core corpora), pokud byly použity v komparativních studiích.

Většina korpusů byla a je sestavována ke konkrétnímu výzkumu.⁴⁸ Cíl projektu je v daném případě určen předem a samozřejmě ovlivňuje tvorbu korpusu, kterému se říká **specializovaný** (specialized corpora). Na takových korpusech lze zkoumat např. rozvoj jazyka u dětí, nesouhlas v konverzaci, direktivní jazykové techniky učitele, regionální či dialektové variety, nestandardní/neformální jazyk, chyby studentů určitého jazyka ap.⁴⁹

Dále rozlišujeme korpusy **mluveného jazyka** (spoken corpora) a **psaného jazyka** (written corpora). Korpusy mluveného jazyka jsou představeny ve značně menším počtu než korpusy jazyka psaného, a to je kvůli mnohostranné náročnosti a nákladnosti sběru materiálu (nahrávky mluvy jsou speciálně zakódované, zpravidla v souladu s mezinárodní fonetickou abecedou IPA⁵⁰).

Rozlišujeme také korpusy obecné a nářeční, dále „z hlediska interpretace a uložení jazykové korpusy surových textů, tj. v nezpracované podobě, n. naopak korpus foneticky, morfologicky, syntakticky, sémanticky i jinak interpretovaný, resp. označkový, anotovaný“.⁵¹

5.3. Český národní korpus

Český národní korpus (dále ČNK) je projekt, který byl založen v roce 1994 s cílem vytvářet a zpřístupňovat „rozsáhlá jazyková data pro výuku a výzkum ve formě elektronických korpusů“.⁵² V dnešní době umožňuje přístup k víc než třem miliardám slov (3,6 mld. slov v českých jednojazyčných korpusech a 1,5 mld. slov v cizojazyčných paralelních korpusech).

V Českém národním korpusu se užívá dotazovací jazyk CQL, který slouží k podrobnějšímu vymezení dotazu pomocí regulérních výrazů.

⁴⁸ŠULC, M. (1999, s.12.)

⁴⁹ŠULC, M. (1999, s.12.)

⁵⁰BACHMANNOVÁ, J. (2002, s. 235)

⁵¹BACHMANNOVÁ, J. (2002, s. 235)

⁵²dostupné na: <<https://wiki.korpus.cz/doku.php/cnk:uvod>>

„Projekt ČNK provozují zejména pracovníci dvou ústavů FF UK: Ústavu Českého národního korpusu a Ústavu teoretické a počítačové lingvistiky. Na sběru materiálů a koordinaci dílčích činností se ovšem kromě toho podílejí i více než dvě stovky externistů z celé ČR.“⁵³

ČNK je dostupný pro veřejnost, je bezplatný. Na úvodní webové stránce korpusu www.korpus.cz lze najít odkaz na Wiki stránku, kde jsou představeny manuály k aplikacím ČNK, přehled korpusů dostupných v rámci ČNK, krátký kurz práce s korpusem, slovníček základních pojmů z korpusové lingvistiky atd.

⁵³dostupné na: <<https://wiki.korpus.cz/doku.php/cnk:uvod>>

PRAKTICKÁ ČÁST

Údaje k analýze byly získány ze slovníků a Českého národního korpusu. Bylo použito celkem 5 slovníků: První až čtvrtý díl SČFI (*Slovník české frazeologie a idiomatiky. 1, Přirovnání* (F. Čermák, J. Hronek), *Slovník české frazeologie a idiomatiky. 2, Výrazy neslovesné* (F. Čermák, J. Hronek, J. Machač), *Slovník české frazeologie a idiomatiky. 3, Výrazy slovesné* (F. Čermák, J. Hronek, J. Machač), *Slovník české frazeologie a idiomatiky. 4, Výrazy větné* (F. Čermák, J. Hronek, J. Machač)) a dílo *Lidová rčení* od J. Zaorálka.

Informace v ČNK byly vyhledávány v psaných korpusech řady SYN, konkrétně **synv4**.

Práce se následně zabývá několika výzkumy. Zaprvé se zaměříme na materiál zachycený ve slovnících a provedeme analýzu toho, zda se všechny nalezené frazémy a idiomy vyskytují v **synv4**. Výrazy ze slovníku, které se v korpusu v rámci kolokačních lemmat neobjeví, dohledáme manuálně (výzkum 1).

Zadruhé se podíváme na materiály získané v korpusu a provedeme jejich stručnou analýzu, třídění do skupin a zaměříme se na nové frazémy či významy frazému, dosud nezaznamenaný slovníky (výzkum 2).

ZDROJ	Zkratka
ČERMÁK, František, Jiří HRONEK a Jaroslav MACHAČ. <i>Slovník české frazeologie a idiomatiky. 1, Přirovnání</i> . Praha: Leda, 2009. ISBN 978-80-7335-215-8.	SČFI1
ČERMÁK, František, Jiří HRONEK a Jaroslav MACHAČ. <i>Slovník české frazeologie a idiomatiky. 2, Výrazy neslovesné</i> . Praha: Leda, 2009. ISBN 978-80-7335-215-8.	SČFI2
ČERMÁK, František, Jiří HRONEK a Jaroslav MACHAČ. <i>Slovník české frazeologie a idiomatiky. 3, Výrazy slovesné</i> . Praha: Leda, 2009. ISBN 978-80-7335-215-8.	SČFI3

HRONEK, Jiří, Jaroslav MACHAČ a František ČERMÁK. <i>Slovník české frazeologie a idiomatiky. 4, Výrazy větné</i> . Praha: Leda, 2009. ISBN 978-80-7335-215-8.	SČFI4
ZAORÁLEK, Jaroslav. <i>Lidová rčení</i> . Vyd. 3., reprint 1. vyd. z r. 1947. Praha: Aurora, 1996. ISBN 80-85974-10-X.	LR

1. Výzkum 1.: slovníky versus korpus

1.1. Slovník české frazeologie a idiomatiky

Slovník české frazeologie a idiomatiky z edice Františka Čermáka je jedním ze dvou základních zdrojů informace pro praktickou část této bakalářské práce, proto se krátce podíváme na historie jeho vzniku.

Slovník české frazeologie a idiomatiky (dále také SČFI) od prof. PhDr. Františka Čermáka, DrSc. a kolektiva spolupracovníků vyšel poprvé v nakladatelství Academia v roce 1983 (1. díl *Přirovnání* v roce 1983, 2. díl *Výrazy neslovesné* v roce 1988 a 3. díl ve dvou svazcích *Výrazy slovesné* v roce 1994).⁵⁴ Druhé vydání slovníku (vyšlo roku 2009) je nutně doplněno a modernizováno, nové tu jsou především *Výrazy větné* ve 4. díle, který vychází poprvé.⁵⁵ Celkově slovník obsahuje 16790 hesel všech typů a „<...> je největším známým komplexním a relativně úplným popisem současné frazeologie nějakého jazyka vůbec.“⁵⁶

1.2. Postup práce při výzkumu 1.

Frazémy a idiomy obsahující lemmata „nebe“ a „země“ byly nejprve vyhledány ve slovnících a abecedně uspořádány do dvou tabulek: Tabulka 1. obsahuje frazémy a idiomy s jazykovou jednotkou „nebe“, Tabulka 2. s jazykovou jednotkou „země“.

Zde je nutné poznamenat, že 1. díl SČFI obsahuje přirovnání, 2. díl výrazy neslovesné, 3. díl výrazy slovesné a 4. díl výrazy větné; v tabulkách máme zachovány všechny nalezené frazémy a chápeme je jako samostatné jazykové jednotky, i když mají stejný základ, např.

⁵⁴ ČERMÁK, F., HRONEK, J., MACHAČ, J., SČFI1 (2009, s. 7)

⁵⁵ ČERMÁK, F., HRONEK, J., MACHAČ, J., SČFI1 (2009, s. 7)

⁵⁶ ČERMÁK, F., HOLUB, J., SČFI5 (2016, s. 8)

- Přirovnání: *jako by (mu) to spadlo z nebe,*
- Neslovesný výraz: *spadlý z nebe,*
- Slovesný výraz: *jako když to spadne z nebe,*
- Větný výraz: *spadnout z nebe.*

V prvním sloupci vlevo je uveden zdroj informace (tady konkrétní díl slovníku), druhý sloupec obsahuje samotný frazém, třetí sloupec pak jeho význam. Sloupec vpravo zahrnuje údaje o tom, zda se daný frazém vyskytuje v ČNK či ne. Zde je důležité dodat, že pod výskytem v korpusu rozumíme přítomnost lemmatu v rámci přesně **kolokačního lemmatu** (col_lemma). Podrobnější postup hledání v korpusu je popsán dále v kapitole 2.1. Poznámka „ANO“ znamená, že lze daný frazém najít v korpusu přesně v takové podobě, jak je zachycen ve slovníku, poznámka „NE“ znamená, že se tento frazém vůbec nevyskytuje v korpusu jako kolokační lemma, zatímco poznámka „ANO/NE“ ukazuje na to, že se vybraný frazém objevuje v ČNK, ale v trochu jiné podobě, než ve slovníku.

1.3.Lemma „nebe“ ve slovnících

Nebe je podstatné jméno středního rodu, slovo má své kořene v praslovanštině (nebo).

V internetové jazykové příručce (<http://prirucka.ujc.cas.cz>) je uvedeno několik definic jazykové jednotky „nebe“ s odkazem na *Slovník spisovné češtiny pro školu a veřejnost* (2014):

„1. obloha: jasné nebe; spát pod širým nebem venku, bez přístřeší

2. (v náb.) sídlo boha, bohů a místo posmrtné blaženosti: Otče náš, jenž jsi na nebesích (modlitba)

3. nebesa stříška z látky (nad postelí): postel s nebesy; nést nad knězem nebesa, baldachýn.“⁵⁷

⁵⁷ dostupné na: <<http://prirucka.ujc.cas.cz/?slovo=nebe>>

Tabulka 1

Tabulka 1 představuje seznam všech frazémů s lemmatem „nebe“ nalezených v SČFI.

Zdroj	Frazém	Význam	Výskyt daného frazému v ČNK
SČFI1 (Přirovnání)	(být) modrý j. nebe	Být mile modrý, (popř.:) a vzbuzovat tak naději.	ANO
SČFI1	je tu j. v nebi	Je tady velmi krásně a milé, klid a bezpečí, popř. pohodlí; cítím se tu šťastný, bezpečný a uvolněný.	ANO
SČFI1	(je to) j. by (mu) to spadlo z nebe / j. když to spadne z nebe	Překvapivě, náhle, neočekávaně a většinou bezpracně něco získat, najít, dostat apod.; je to vítané (i když to není nezbytně zapotřebí).	ANO
SČFI1	(je tam) schodů / (to jsou) schody j. do nebe	Je tam velmi mnoho schodů; jsou to neobyčejně dlouhé schody s vysokými stupni.	ANO
SČFI1	mít se / žít si / cítit se / být j. v (sedmém) nebi	<ol style="list-style-type: none"> 1. Vést příjemný, blahobytný život bez materiálních starostí, (často:) zásluhou druhého. 2. Být, cítit se velmi šťastný, mít opojnou radost. 	ANO
SČFI1	to je (j.) nebe a dudy	Jsou to dvě věci zcela rozdílné, je v tom velký rozdíl.	ANO

SČFI1	objevil se j. když spadne z nebe / j. by z nebe spadl / j. z nebe spadlý	Ukázal se, objevil se, přišel náhle, zcela nečekaně, popř. nevysvětlitelně.	ANO
SČFI1	být / koukat j. z nebe spadlý n. j. by spadl z nebe / j. když spadne z nebe	Být, tvářit se velmi překvapeně, popř. zaraženě, nechápavě n. zmateně.	ANO
SČFI2 (Výrazy neslovesné)	nebe a dudy	Dvě věci (při srovnání vlastností, povahy, úrovně) zcela rozdílné, které nelze směšovat ani připodobňovat.	ANO
SČFI2	nebe na zemi n. království boží na zemi	Šťastný, blahobytný, spokojený, zcela bezstarostný život.	ANO
SČFI2	mezi nebem a zemí	<ol style="list-style-type: none"> 1. Ve výši nad zemským povrchem, vysoko ve vzduchu. 2. (Jevy, věci) působící dojmem tajemnosti pro svou nepochopitelnost (v ironickém smyslu též o věcech, jevech běžných, ale překvapivých.) 	ANO
SČFI2	spadlý z nebe n. z nebe spadlý	Takový (jev), který nastal, k němuž došlo zcela neočekávaně, překvapivě, popř. (o úspěšném výsledku) bez vynaložení nějakých úsilí; takový (člověk, např. návštěva), který se objevil, přišel zcela nenadále, nečekaně.	ANO

SČFI2	pod širým nebem	Ve volné přírodě, venku, bez ochrany proti nepohodě (nikoli v domě, pod střechou apod., jak je v daném případě obvyklejší).	ANO
SČFI2	do nebe volající	(Čin, akt někoho se nežádoucím způsobem, nepříznivě dotýkající, omyl apod.) tak nápadně, evidentně nesprávný, že zasluhuje odsouzení, trest (zprav. jen zesiluje význam příslušného výrazu).	ANO
SČFI3 (Výrazy slovesné)	Dobry / dobre nebe!	To je těžko uvěřitelné překvapení; jsem tím silně udivený.	NE
SČFI3	To je (jako) nebe a dudy!	To jsou zcela rozdílné věci, nelze je směřovat ani připodobňovat, mezi nimi je zásadní rozdíl, přičemž toto je mnohem lepší.	ANO
SČFI3	Nebe je modré / modry jako chrpa.	Nebe je krásné a příjemně modré.	NE
SČFI3	Snes / snesl by jí modry / modré z nebe.	Má ji moc rád a ochotně a se sebezapřením by pro ni udělal všechno (reálné) n. přinesl i velkou oběť.	ANO
SČFI3	Nebe je (modry / modré) jako šmolka. n. Obloha je modrá jako šmolka.	Venku je krásně a vyvolává to dobrou náladu; obloha je příjemně modrá a je hezké počasí.	ANO

SČFI3	Nebe je (zatažený) jako pytel.	Obloha je silně, temně zamračená a bude zřejmě pršet, je škaredě.	NE
SČFI3	Žádný / žádnéj učený z nebe nespád / nespád.	To se taky naučíš, jen se nevzdávej; vytrvej, příště se ti to už třeba povede; každý se to musel někdy naučit.	NE
SČFI3	Žádný / žádnéj učený z nebe nespád, ale hlupáky / blbce jako by shazovali /vyhazoval.	Trvale hloupých lidí je všude nepříjemně a nenápadně moc; hlupáci mezi lidmi jakoby převažují a ztrpčují život.	NE
SČFI3	Nebe je jako vymetený / vysmejčený. n. Obloha je jako vymetená / vysmejčená.	Obloha je zela bezmračná, jasná a je n. bude tam dobré počasí.	ANO
SČFI3	Když já vám povídám, že je nebe na zemi, pravdu mám, věřte mi.	Život je pěkný, daří se mi a já mám z něj radost, měj proto z něj radost i ty.	ANO/NE
SČFI3	Nebe a země pomínou, ale slova má nepomínou.	(Ty, moje) myšlenky jsou velmi závažné a mají trvalou platnost.	ANO/NE
SČFI3	Je to (tu) jako v nebi.	Je tu krajně hezky, klid a pohoda a cítím se tu šťastný a spokojený.	ANO
SČFI3	Tebe mi seslalo samo nebe! n. Samo nebe (nám) ho poslalo / seslalo! n. Tebe nám poslal sám osud.	Jsem moc rád, že přicházíš, protože jedině n. zrovna ty mi, nám můžeš pomoci.	ANO

SČFI3	(Je to) jako by (mu) to spadlo z nebe / jako když to spadne z nebe	Je to překvapivě tady a je to vítané n. vhodné.	ANO
SČFI3	To volá do nebe. n. To je až do nebe volající. n. ... (až) to volá do nebe.	To je nespravedlivé a drzé chování; to je ostudné, nemorální a provokativní jednání.	ANO
SČFI4 (Výrazy větné)	být nebe a dudy	Být naprosto rozdílný.	ANO
SČFI4	být / bejt / cítit se v sedmém / devátém / desátém nebi n. bejt v sedmým / devátým / desátým nebi	Být krajně šťastný a mít velkou radost, euforii.	ANO
SČFI4	mít v někom své nebe	Vidět v někom svůj ideál a být do něj zamilovaný.	ANO
SČFI4	mít nebe na zemi	Mít se výborně a žít klidně a šťastně, bez potíží a problémů.	ANO
SČFI4	přijít / dostat se do nebe	Být spasený (před peklem), být odměněn stálou blažeností.	ANO
SČFI4	růst do nebe	Být nápadně, neúměrně vysoký a růst stále dál.	ANO
SČFI4	slíbit / slibovat někomu modré / modrý z nebe	Slíbit, že uděláš cokoliv	ANO

SČFI4	snést někomu modré / modrý z nebe	Milovat někoho silně a obětavě a být ochoten (pro jeho potěšení) mu splnit všechna jeho přání.	ANO
SČFI4	spadnout z nebe	Objevit se zcela náhle, znenadání, popř. bez důvodů.	ANO
SČFI4	vidět / mít v někom / něčem své / svý nebe	Vidět v někom svůj ideál a být do něj zamilovaný.	ANO
SČFI4	volat do nebe n. být do nebe volající	Být krajně, nápadně nespravedlivý, popř. ostudný n. provokativní a vyžadovat nápravu, zákrok.	ANO
SČFI4	vychvalovat / vynášet / velebit někoho / něco až do (samého / samýho) nebe / nebes / oblak	Chválit někoho, něco velmi vysoce a všestranně, popř. nekriticky a přehnaně; projevovat pro někoho, něco své velké nadšení.	ANO

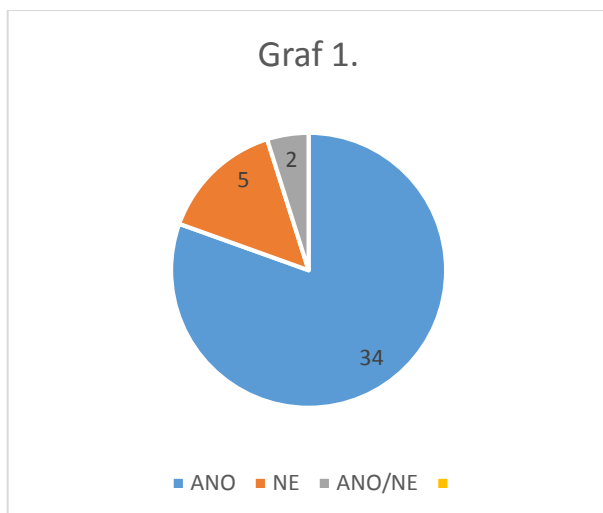
Na základě údajů z Tabulky 1 lze zjistit, že frazémy s poznámkami ANO, NE, ANO/NE jsou v daném poměru:

ANO – 34 výskytů z 41 (82,92%),

NE – 5 výskytů z 41 (12,19%),

ANO/NE – 2 výskyty z 41 (4,87%).

Grafické zobrazení dané proporce ilustruje **Graf 1.**; frazémy, které se objevují v SČFI a nejsou nalezeny v korpusu jako kolokační lemmata (tj. s poznámkou NE nebo ANO/NE) jsou dále zachyceny zvlášť v **Tabulce 1.1** pro jejich snadnější popis a zpracování.



Tabulka 1.1

Za účelem komplexního popisu daných frazémů provedeme jejich manuální vyhledávání v korpusu, například dotaz pro druhý frazém vypadá takto:

[lemma="nebe"] [lemma="být"] [lemma="modrý"] [lemma="jako"] [lemma="chrpa"],

(v korpusu synv4, dotazovací jazyk CQL). V pravém sloupci tabulky 1.1 jsou zahrnuty výsledky hledání, eventuálně i počet výskytů frazémů. Pro snadnější přehled používáme také poznámky ANO a NE.

Zdroj	Frazém	Význam	Výskyt daného frazému v ČNK (v rámci kolokačního lemmatu)	Výsledky manuálního vyhledávání v ČNK
SČFI3	Dobrý / dobré nebe!	To je těžko uvěřitelné překvapení; jsem tím silně udivený.	NE	ANO 29 výskytů

SČFI3	Nebe je modré / modrý jako chrpa.	Nebe je krásné a příjemně modré.	NE	NE
SČFI3	Nebe je (zatažený) jako pytel.	Obloha je silně, temně zamračená a bude zřejmě pršet, je škaredě-	NE	NE
SČFI3	Žádný / žádnéj učení z nebe nespád / nespád.	To se taky naučíš, jen se nevzdávej; vytrvej, příště se ti to už třeba povede; každý se to musel někdy naučit.	NE	ANO 13 výskytů
SČFI3	Žádný / žádnéj učení z nebe nespád, ale hlupáky / blbce jako by shazovali /vyhazoval.	Trvale hloupých lidí je všude nepříjemně a nenápadně moc; hlupáci mezi lidmi jakoby převažují a ztrpčují život.	NE	NE
SČFI3	Když já vám povídám, že je nebe na zemi, pravdu mám, věřte mi.	Život je pěkný, daří se mi a já mám z něj radost, měj proto z něj radost i ty.	ANO/NE	ANO 1 výskyt
SČFI3	Nebe a země pominou, ale	(Ty, moje) myšlenky jsou velmi závažné a	ANO/NE	ANO 7 výskytů

	slova má nepominou.	mají trvalou platnost.		
--	--------------------------------	---------------------------	--	--

Výsledky provedeného výzkumu a závěr jsou představené v kapitole 1.5. spolu s výsledky lemmatu „země“.

1.4.Lemma „země“ ve slovnících

„Země“ (také „zem“) je podstatné jméno ženského rodu. V internetové jazykové příručce (<http://prirucka.ujc.cas.cz>) je uvedeno několik definic jazykové jednotky „země“ s odkazem na *Slovník spisovné češtiny pro školu a veřejnost* (2014):

1. planeta sluneční soustavy (třetí co do vzdálenosti od Slunce) obývaná lidmi, zeměkoule, tento svět,
2. pevnina, souš,
3. území tvořící polit., správní celek, stát, kraj,
4. zemský povrch, půda, po které chodíme; (v místnostech) podlaha,
5. půda, hlína. ⁵⁸

Při hledání lemmatu „země“ v korpusu budeme rozlišovat velikost prvního písmene, zkoumáme pouze jazykovou jednotku „země“, nikoli „Země“, která je názvem planety sluneční soustavy, tzn. vyloučíme první význam.

⁵⁸dostupné na: <<http://prirucka.ujc.cas.cz/?slovo=zem%C4%9B>>

Tabulka 2

Tabulka 2 představuje seznam všech frazémů s lemmatem „země“ nalezených v SČFI.

Zdroj	Frazém	Význam	Výskyt daného frazému v ČNK
SČFI1 (Přirovnání)	být černý j. zem	Být silně zašpiněný, nemytý, nepraný n. dlouho nečištěný; popř. být odporně tmavý a černý.	NE
SČFI1	byl/ stál j. do země vrostlý / j. by vrostl do země	Leknutím, překvapením náhle nevěděl, co má dělat; byl tak ohromen a zaražen, že se chvíli nemohl ani pohnout; byl šokován.	ANO
SČFI1	objevil se j. by ze země vyrostl	Objevil se někde náhle, zcela nevysvětlitelně, překvapivě; přišel náhle, tiše a zcela nepozorovaně.	ANO
SČFI1	utíká j. by pod ním země hořela	Běží pryč nápadně rychle, divoce, panicky a ze všech sil; vystrašeně utíká rychle pryč.	NE
SČFI1	zmizel / ztratil se j. by se pod ním / po něm zem slehla /slehne n. j. by se do země propadl n. j. když se do země propadne	Zmizel náhle, zcela nevysvětlitelně a beze stopy.	ANO

SČFI1	zůstal j. když ho do země vrazí	Leknutím, překvapením náhle nevěděl, co má dělat; byl tak ohromen a zaražen, že se chvíli nemohl ani pohnout; byl šokován.	NE
SČFI2 (Výrazy neslovesné)	země (tisíce) jezer	Finsko, ekvivalent finského názvu země – Suomi.	ANO
SČFI2	země kalicha	Husitské Čechy.	ANO
SČFI2	země mlékem a strdím oplývající	Úrodný kraj, území poskytující v naprosto dostatečné míře vše potřebné k živobytí.	ANO
SČFI2	země neomezených možností	Spojené státy americké jako stát, kde je možno individuálně vyniknout, získat bohatství, postavení ap.	ANO
SČFI2	země nikoho	Pás území mezi frontami válčících stran vojensky neobsazený.	ANO
SČFI2	země pruhů a hvězd	Označení Spojených států amerických (podle státní vlajky).	ANO
SČFI2	země slunce	Japonsko.	ANO
SČFI2	země jitřní svěžesti	Korea.	NE
SČFI2	rodná země / zem	Stát, krajina, kde se člověk narodil a s níž je proto spojen pevnými pouty.	ANO

SČFI2	suchá země	Souš, popř. pevnina na rozdíl od moře.	ANO
SČFI2	Svatá země	Palestina (zvl. v historickém kontextu).	ANO
SČFI2	zaslíbená země	1. Palestina (jako biblické a historické území v Přední Asii) 2. Místo, prostředí vytouženého štěstí, blahobytu, stav poskytující člověku plné uspokojení.	ANO
SČFI3 (Výrazy slovesné)	Toto je země mlékem a strdím oplývající.	To je země, území pro nás, vás ideální, krásné a přátelské, plné přírodního bohatství. A. Jirásek (1851–1930), <i>Staré pověsti české</i> (1894).	ANO
SČFI3	Aj, zde leží zem ta (před okem mým slzy ronícím, někdy kolébka, nyní národu mého rakev).	Tady je ta má ztracená vlast, kterou oplakávám. (J. Kollár (1793–1852), <i>Slavy dcera</i> (1824), předzpěv.)	ANO
SČFI3	Dost v zem železa na dobré meče.	Nebojte se, vzchopte se a bojujte, zbraní je dostatek! (J. Neruda (1834–1891), báseň <i>Jen dál!</i> , <i>Zpěvy páteční</i> (1896).)	NE
SČFI3	Už je pod zemí.	Být (už) mrtvý a pohřbený.	NE

SČFI3	Dostaneš takovou, že chytneš druhou vo zem/ zed'!	Já tě potrestám!; zbiju ti za trest!; dám ti velkou facku!	ANO/NE
SČFI3	Už ho kreje / kryje (černá) zem.	Už je mrtvý.	NE
SČFI3	Nezasluhuje / nezaslouží si, aby ho zem nesla / že ho zem nosí.	Je to velmi špatný, zlý člověk; je to darebák a zločinec (a neměl by vůbec žít).	NE
SČFI3	Zmizel, jako by se po něm / pod ním zem / země slehla. n. Slehla se pod ním zem.	Zmizel, odešel zcela nevysvětlitelně a nepozorovaně a nelze ho najít.	ANO
SČFI3	Zůstal, jako by k zemi přirost / přirostl. n. Stál jako do země vrostlej / vrostlý.	Náhle se zarazil, zcela znehybněl a nemohl se pohnout z místa.	ANO
SČFI3	Česká země, matka chleba.	České země, Čechy jsou pro Čechy nejen mateřskou zemí, ale dávají jim i obživu.	NE
SČFI3	A to je ta krásná země, země česká, domov můj.	Česká země, Čechy jsou má krásná milovaná vlast. (J. K. Tyl (1808–1856), <i>Fidlovačka</i> (1834).)	ANO
SČFI3	Budiž mu země lehká! n. Necht' je mu země lehká.	Loučím se se zemřelým naposled a přeju mu po smrti klid.	ANO

SČFI3	Jiná země, jiný / jinej mrav.	Každá země má specifické zvyky a normy života, a ty je třeba (i návštěvníky) dodržovat a respektovat.	ANO
SČFI3	Čí země, toho náboženství	Vítěz, vítězný stát vnucuje svou ideologii, náboženství, výklad podmaněným.	NE
SČFI3	Země nepotřebuje nové kontinenty, ale nové lidi.	Společnost potřebuje především nově myslící a tvůrčí lidi, a ne jen senzační objevy nových, ale nerozvinutých území.	ANO
SČFI3	Ať se do země propadnu (jestli ne-)! 	Opravdu mi věř, je to tak; já si nevymýšlím, je to pravda.	ANO/NE
SČFI3	Copak se propadnul / propadl do země? n. Snad / přece se nepropadnul do země!	Vůbec ho nemůžu najít, a to mě zlobí, někde tu přece musí být.	ANO/NE
SČFI3	(Zmizel) jako když se do země / vody propadne. n. Jako by se do země propadnul.	Zmizel, odešel náhle, a zcela nevysvětlitelně, záhadně; najednou se to ztratilo n. zmizelo (je to asi někde zastrčené) a nikde to tu není možné najít.	ANO
SČFI3	Nejradši bych se propad / propadnul do země.	Velmi se stydím, jsem se styděl a cítil jsem svou chybu (tak, že bych radši odešel, kdybych mohl).	ANO/NE

SČFI3	Už rostu / zarůstám do země.	Já už nerostu (a nepotřebuju proto pořád větší šaty), spíš se zmenšuju (a blížím se smrti).	ANO
SČFI3	Utíká, jako by za ním země hořela.	Utíká, běží velmi rychle (popř. a zbrkle, neelegantně, nedůstojně).	NE
SČFI3	Sotva je ho u země vidět.	Je velmi malý a nízkého vzrůstu; je velmi drobné postavy.	NE
SČFI3	Radší / raději bych se viděl pod zemí, (než)...	To rozhodně nemůžu a nejsem ochotný připustit a budu tomu osobně všemožně bránit; nemůžu s tím zásadně souhlasit n. to udělat (popř. a postavím se proti i za cenu vlastního života).	NE
SČFI3	Objevil se, jako by ze země vyrost / vyrostl. n. jako by ze země vyrost.	Objevil se, vynořil se, vyvstal (přede mnou) náhle a nepozorovaně.	ANO/NE
SČFI4 (Výrazy větné)	dívat se / hledět do země	Mít sklopenou hlavu, dívat se dolů a projevovat tak svůj stud n. mlčenlivě rozpaky, popř. zatvrzelé odmítání.	ANO
SČFI4	sotva se země dotýkat	Pohybovat se lehce, plavně a vzletně; téměř se vznášet; chodit lehce a elegantně.	NE
SČFI4	hnít v zemi	Být (už) mrtvý a pohřbený.	ANO

SČFI4	jít k zemi	Prudce upadnout n. se zřítit k zemi.	ANO
SČFI4	klanět se (někomu) až k zemi	Projevovat někomu neobvykle výraznou n. přehnanou a obv. i poníženou úctu, popř. vděčnost.	ANO
SČFI4	ležet / spát na hole zemi	Ležet n. spát primitivně, nedůstojně a nepohodlně.	ANO
SČFI4	odlepit se od země	Vzlétnout, vznést se, zvednout se do vzduchu.	ANO
SČFI4	poroučet se k zemi / na zem	1. Omdlít a upadnout. 2. (S)padnout, skácet se.	ANO
SČFI4	poslat někoho k zemi / na zem	Srazit někoho na zem.	ANO
SČFI4	praštit sebou na zem	1. Omdlít a bezvládně upadnout. 2. Prudce a těžce upadnout. 3. Prudce si lehnout, padnout na zem.	ANO
SČFI4	propadnout se do země	Náhle, nevysvětlitelně a beze stopy se (někam) ztratit, popř. zamotat, zamíchat, schovat (a nebýt vidět).	ANO
SČFI4	přitisknout / tisknout někoho k zemi	Utlačovat někoho a omezovat ho v životních právech, volnosti, pohybu n. projevu.	NE

SČFI4	růst/ zarůstat do země	Stárnout a ochabovat a nemít už vyhlídky na dlouhý život a rozkvět; blížit se konci života.	ANO
SČFI4	sebrat se ze země	Zvednout se, vstát.	NE
SČFI4	skropit zem krví	Bojovat za to místo urputně a obětavě, popř. hrdinsky a ztratit, nechat tam mnoho životů a krvi.	NE
SČFI4	srazit někoho k zemi / na zem	Povalit, srazit někoho prudce na zem.	ANO
SČFI4	srovnat něco se zemí	Rozbít, rozbořit něco úplně do základu, nenechat stát nic.	ANO
SČFI4	stát na pevné zemi	Být (si) zcela jistý a realistický; mít pevné důvody a cítit reálnou jistotu; postupovat realisticky.	ANO
SČFI4	vydupat něco ze země	Vytvořit překvapivě, téměř zázračně něco z téměř ničeho n. si to vynutit; uskutečnit něco navzdory krajním potížím, nedostatku apod.	ANO
SČFI4	žít při zemi	<ol style="list-style-type: none"> 1. Žít ustrašeně, nenápadně a bez vzletu; žít primitivně. 2. Žít skromně a reálně, bez velkých nároků. 	ANO

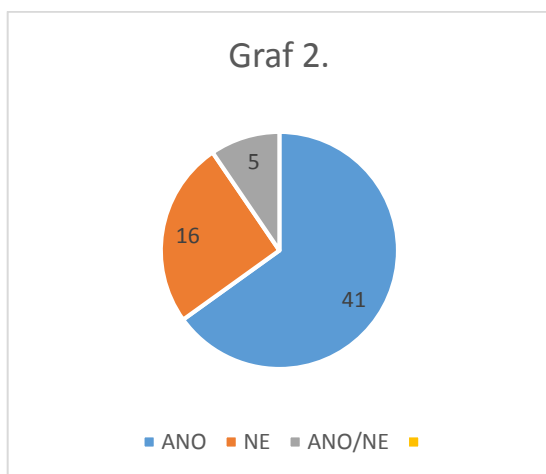
Na základě údajů z tabulky 2 lze zjistit, že frazémy s poznámkami ANO, NE, ANO/NE jsou tady v daném poměru:

ANO – 41 výskytů z 62 (66,12%),

NE – 16 výskytů z 62 (25,80%),

ANO/NE – 5 výskytů z 62 (8,06%).

Grafické zobrazení dané proporce ilustruje **Graf 2.**; frazémy, které se objevují v SČFI a nejsou nalezeny v korpusu jako kolokační lemmata (tj. s poznámkou NE nebo ANO/NE) jsou dále zachyceny zvlášť v **Tabulce 2.2** pro jejich snadnější popis a zpracování.



Tabulka 2.2

Tabulka 2.2 má stejnou strukturu jako tabulka 1.1, uvedené frazémy byly též vyhledány v ČNK manuálně.

Zdroj	Frazém	Význam	Výskyt daného frazému v ČNK (v rámci kolokačního lemmatu)	Výsledky manuálního vyhledávání v ČNK

SČFI1	být černý j. zem	Být silně zašpiněný, nemytý, nepraný n. dlouho nečištěný; popř. být odporně tmavý a černý.	NE	NE
SČFI1	zůstal j. když ho do země vrazí	Leknutím, překvapením náhle nevěděl, co má dělat; byl tak ohromen a zaražen, že se chvíli nemohl ani pohnout; byl šokován.	NE	ANO 14 výskytů
SČFI2	země jitřní svěžestí	Korea.	NE	ANO 31 výskytů
SČFI3	Dost v zem železa na dobré meče.	Nebojte se, vzchopte se a bojujte, zbraní je dostatek! (J. Neruda (1834–1891), báseň <i>Jen dál!</i> , <i>Zpěvy páteční</i> (1896).)	NE	ANO 1 výskyt
SČFI3	Už je pod zemí.	Být (už) mrtvý a pohřbený.	NE	ANO 14 výskytů
SČFI3	Dostaneš takovou, že chytneš druhou vo zem/ zed'!	Já tě potrestám!; zbiju ti za trest!; dám ti velkou facku!	ANO/NE	NE

SČFI3	Už ho kreje / kryje (černá) zem.	Už je mrtvý.	NE	ANO 1 výskyt
SČFI3	Nezasluhuje / nezaslouží si, aby ho zem nesla / že ho zem nosí.	Je to velmi špatný, zlý člověk; je to darebák a zločinec (a neměl by vůbec žít).	NE	NE
SČFI3	Česká země, matka chleba.	České země, Čechy jsou pro Čechy nejen mateřskou zemí, ale dávají jim i obživu.	NE	ANO 1 výskyt
SČFI3	Čí země, toho náboženství	Vítěz, vítězný stát vnucuje svou ideologii, náboženství, výklad podmaněným.	NE	ANO 4 výskyty (2 výskyty: čí panství, toho náboženství, 11 výskytů čí/koho vlada, toho náboženství)
SČFI3	At' se do země propadnu (jestli ne-)!	Opravdu mi věř, je to tak; já si nevymýšlím, je to pravda.	ANO/NE	ANO 3 výskyty

SČFI3	Copak se propadnul / propadl do země? n. Snad / přece se nepropadnul do země!	Vůbec ho nemůžu najít, a to mě zlobí, někde tu přece musí být.	ANO/NE	ANO 1 výskyt
SČFI3	Nejradši bych se propad / propadnul do země.	Velmi se stydím, jsem se styděl a cítil jsem svou chybu (tak, že bych radši odešel, kdybych mohl).	ANO/NE	NE
SČFI3	Utíká, jako by za ním země hořela.	Utíká, běží velmi rychle (popř. a zbrkle, neelegantně, nedůstojně).	NE	NE
SČFI3	Sotva je ho u země vidět.	Je velmi malý a nízkého vzrůstu; je velmi drobné postavy.	NE	NE
SČFI3	Radši / raději bych se viděl pod zemí, (než)...	To rozhodně nemůžu a nejsem ochotný připustit a budu tomu osobně všemožně bránit; nemůžu s tím zásadně souhlasit n. to udělat (popř. a postavím se proti i za cenu vlastního života).	NE	ANO 1 výskyt

SČFI3	Objevil se, jako by ze země vyrost / vyrostl. n. jako by ze země vyrost.	Objevil se, vynořil se, vyvstal (přede mnou) náhle a nepozorovaně.	ANO/NE	NE
SČFI4	sotva se země dotýkat	Pohybovat se lehce, plavně a vzletně; téměř se vznášet; chodit lehce a elegantně.	NE	NE
SČFI4	přitisknout / tisknout někoho k zemi	Utlačovat někoho a omezovat ho v životních právech, volnosti, pohybu n. projevu.	NE	ANO 33 výskytů
SČFI4	sebrat se ze země	Zvednout se, vstát.	NE	ANO 6 výskytů
SČFI4	skropit zem krví	Bojovat za to místo urputně a obětavě, popř. hrdinsky a ztratit, nechat tam mnoho životů a krvi.	NE	ANO 1 výskyt

1.5.Výsledek výzkumu 1

Tabulka 1.1 se skládá z 7 položek, 4 ze 7 výrazů se objevily v korpusu v procesu podrobnějšího manuálního vyhledávání. Tabulka 2.2 zahrnuje 21 výrazů, 13 z nich bylo nalezeno v ČNK při manuálním hledání. Na základě této informace dospíváme ke dvěma závěrům:

- 1) I přesto, že 2. vydání Slovníku české frazeologie a idiomatiky od Františka Čermáka vyšlo v roce 2009 a je tím pádem daný slovník poměrně moderní, může obsahovat i výrazy, které jsou velice málo frekventované (např. mají pouze 1 výskyt v ČNK) nebo se skoro vůbec neužívají (podle údajů z ČNK).
- 2) Velká část výrazů ze slovníku, které se neobjevily v ČNK jako kolokační lemmata, byly pak dohledány manuálně pomocí zadání jiného, více specifikovaného, dotazu. Daný jev v své řadě znamená, že se při lemmatizaci jednotlivých frazémů mohou objevit drobné chyby a nedostatky (např. výraz „země jitřní svěžesti“ má 31 výskytů v korpusu, ale není označen jako kolokační lemma, i když fakticky je kolokačním lemmatem).

1.6. Lidová rčení J. Zaorálka

Jaroslav Zaorálek (1896–1947) byl především známý český překladatel, uměl víc než 5 jazyků, avšak „ovládl češtinu ve všech stylových rovinách a historických vrstvách, zapisoval si i lidová rčení, která vydal tiskem.“⁵⁹

Jeho dílo *Lidová rčení* (dále také LR) vyšlo roku 1947 a během několika desetiletí sloužilo jako hlavní sbírka českého frazeologického materiálu. SČFI F. Čermáka, který vznikl od 60. let, svým způsobem doplnil Zaorálkovo dílo.⁶⁰ V předmluvě k prvnímu vydání SČFI z roku 1981 F. Čermák napsal, že výrazné procento výrazů ze Zaorálkova slovníku je již neznámé a neužívané.⁶¹

Jazyk je velice proměnlivý systém, proto lze předpokládat, že většina výrazů z LR, 70 let staré publikace, se v korpusu už nenajdou.

V Tabulce 3 jsou zahrnuty frazémy s lexémem „nebe“ (frazém „země“ se v daném slovníku nevyskytuje). V sloupci vpravo je představen počet nalezených výsledků v korpusu. Číslo vlevo před lomítkem označuje výsledky hledání, které odpovídají významu frazému ze Zaorálkova

⁵⁹ dostupné na: <https://cs.wikipedia.org/wiki/Jaroslav_Zaor%C3%A1lek>

⁶⁰ ČERMÁK, F., HRONEK, J., MACHAČ, J., SČFI1 (2009, s. 11)

⁶¹ ČERMÁK, F., HRONEK, J., MACHAČ, J., SČFI1 (2009, s. 11)

slovníku, číslo vpravo za lomítkem je celkový počet výsledků (včetně jazykových jednotek, které mají stejnou strukturu, ale jiný význam).

Tabulka 3

Zdroj	Frazém	Význam	Výskyt v ČNK
LR	ještě se nebe hýbe	Ještě je čas, ještě není všemu konec, ještě se to může stát.	0/10
LR	nebe se otvírá	Blýská se.	2/9
LR	pije, až se nebe točí	Velmi, až do opilosti.	-
LR	otevřelo se mu nebe	Měl veliké štěstí.	-
LR	být na kraji samého nebe	Býti velmi šťasten.	1/1
LR	chce se prstem nebe dotýkati	Je pyšný, domýšlivý.	3/3
LR	rád by ho za řadry (na rukou) do nebe donesl	Má ho velmi rád.	-
LR	střílet do nebe	Hloupě mluvit.	1/7
LR	šel do nebe	Umřel.	přes 100 z 1792
LR	to volá do nebe	To je hrozné, neslýchané apod.	-
LR	do nebe (n. pod nebe) – něco –dělat (např. křičet, chválit)	Velmi mnoho, nesmírně, bouřlivě to dělat.	křičet do nebe-12, chválit do nebe-3 atd.
LR	do nebe nesmí a v peklo ho (ještě) nechtějí	Nemůže umřít, dlouho (n. těžce) umírá.	1/1
LR	bydlet u samého nebe	Velmi vysoko bydlet.	-

LR	všecko sežene, i kdyby to měl u nebe vzít	Stůj, co stůj to opatří.	-
LR	až spadneš z nebe	Nikdy.	-
LR	kouká na nás z nebe	Je nebožtík.	9/9
LR	klel, že se to k nebi psalo	Velmi klel.	-
LR	mít – někde – hanácké nebe	Míti tam blahobyt, mítí se tam velmi dobře.	-
LR	vrátí na potkání v nebi	Nikdy nevrátí.	-

Výsledek hledání:

Na základě dat získaných v korpusu můžeme konstatovat, že většina výrazů z LR je archaických. 10 z 19 frazémů není ani v korpusu zachyceno, v podstatě pouze 2 z 19 výrazů se ještě používá běžně, všechny ostatní jenom velice málo. I přesto, že obsah Lidových rčení většinou neodpovídá stavu dnešního jazyka, resp. neodráží soudobou jazykovou realitu, lze používat této slovník za účelem zkoumání vývoje jazyka.

2. Výzkum 2: lemmata „nebe“ a „země“ v ČNK

Ve výzkumu 2 se zaměříme na analýzu dat získaných z korpusu **synv4**.

Korpus SYN verze 4 zahrnuje korpusy SYN2000, SYN2005, SYN2006PUB, SYN2009PUB, SYN2010PUB a SYN2015, a kromě toho také dosud nezveřejněnou publicistiku převážně z let 2010–2014 v ročních objemech přesahujících 200 mil. slov.⁶² Korpus **synv4** obsahuje celkem 3,6 mld. slov.

2.1. Postup práce při hledání v ČNK

Na příkladu lemmatu „nebe“ popíšeme postup práce s korpusem v daném výzkumu. Při hledání v korpusu zvolíme typ dotazu „**Lemma**“ a uvedeme v poli „**Dotaz**“ lemma „nebe“. Ve vybraném korpusu se objeví **168 010** výskytů daného lemmatu. Dále v řádku menu nahoře vybereme „**Frekvence**“ a potom „**Vlastní...**“. V objeveném se okénku zvolíme možnost **col_lemma [collocation_lemma]** v poli „**Atribut**“ a vybereme volbu „**Vytvořit frekvenční seznam**“. Na otevřené stránce nalezneme seznam frekvenční distribuce, údaje z něj jsou představeny v **Tabulce 4**.

První sloupec tabulky z levé strany obsahuje frazém. Číslo před frazémem označuje jeho pořadí v seznamu frekvenční distribuce. Tabulka je sestavena tak, že vedle konkrétního výrazu ve stejném řádku uvádíme i další výrazy, které mají stejný základ nebo stejný význam (např. hned pod větným výrazem číslo tři „spadnout z nebe“ je uveden další výraz „padat z nebe“ atd.). Druhý sloupec obsahuje informaci o počtu výskytů daného kolokačního lemmatu, třetí sloupec představuje relativní hodnotu počtu výskytů na milion výsledků (pokud je daný index menší než 0,01 nezobrazuje se v korpusu automaticky a v tabulce používáme znak „-“).

Ve čtvrtém sloupci uvádíme, zda je daný výraz zachycen ve SČFI, používáme tady stejné poznámky ANO a NE jako ve Výzkumu 1. Pokud se ve vyhledaných výsledcích objeví chyba, např. výraz neobsahující lemma „nebe“ je mylně přiřazen do daného seznamu, označíme tento jev poznámkou „chyba“ a frazém označíme pro lepší odlišení šedou barvou. Pokud v seznamu frekvenční distribuce lemmatu „nebe“ nalezneme frazém či skupinu frazémů, jejichž základem je jiné slovo než „nebe“, uvedeme to slovo též v pravém sloupci.

Stejným způsobem sestavíme i **Tabulku 5**, zahrnující seznam frekvenčních distribucí lemmatu „země“ v kapitole 2.3.1. Hned pod tabulkami se nachází analýza údajů z něj.

⁶² dostupné na: < <https://wiki.korpus.cz/doku.php/cnk:syn:verze4> >

2.2.1. Lemma „nebe“ v ČNK

Tabulka 4

Frazém	Počet výskytů	ipm	Vyskytuje se daný výraz ve SČFI?
1. celkový počet výskytů 168 010, 110 896 výskytů není označeno jako kolokační lemmata (col_lemma)			
2. pod širým nebem	34 491	7,93	ANO
3. spadnout z nebe	3 017	0,69	ANO
6. padat z nebe	1 841	0,42	NE
24. spadlý z nebe	224	0,05	ANO
30. z nebe spadlý	109	0,03	ANO
47. spadnout s nebe	44	0,01	NE
60. padat s nebe	12	-	NE
67. spadlý s nebe	5	-	NE
4. mezi nebem a zemí	2 886	0,66	NE
5. do nebe volající	2 102	0,48	ANO
26. volat do nebe	158	0,04	ANO
7. jako blesk z čistého nebe	1 495	0,34	blesk, hrom
9. blesk z čistého nebe	1 135	0,26	
10. přijít jako blesk z čistého nebe	1 056	0,24	
19. být jako blesk z čistého nebe	328	0,08	
27. zasáhnout jako blesk z čistého nebe	131	0,03	
29. zapůsobit jako blesk z čistého nebe	112	0,03	

31. udeřit jako blesk z čistého nebe	102	0,02	
39. přicházet jako blesk z čistého nebe	59	0,01	
42. působit jako blesk z čistého nebe	58	0,01	
41. objevit se jako blesk z čistého nebe	58	0,01	
48. jako blesk z nebe	43	0,01	
64. jako hrom z čistého nebe	9	-	
8. být v sedmém nebi	1 387	0,32	ANO
38. cítit se v sedmém nebi	70	0,02	ANO
73. mít se jako v sedmém nebi	2	-	ANO
74. žít jako v sedmém nebi	2	-	ANO
79. být v devátém nebi	1	-	ANO
83. být v desátém nebi	1	-	ANO
11. nebe na zemi	978	0,22	ANO
55. mít nebe na zemi	18	-	ANO
12. dostat se do nebe	502	0,12	ANO
13. snést modrý z nebe	421	0,1	ANO
23. slibovat modré z nebe	236	0,05	ANO
33. snášet modrý z nebe	90	0,02	ANO
54. přinést modrý z nebe	19	-	NE
66. slíbit modré z nebe	6	-	ANO
14. růst do nebe	405	0,09	ANO

15. vynášet do nebe	372	0,09	ANO
43. vychvalovat do nebe	56	0,01	ANO
57. chválit do nebe	16	-	ANO
70. velebení do nebe	5	-	ANO
16. nebe a dudy	370	0,09	ANO
17. být nebe a dudy	369	0,08	ANO
20. být jako nebe a dudy	277	0,06	ANO
18. přijít do nebe	344	0,08	ANO
21. žádný strom neroste do nebe	255	0,06	strom
22. rána z čistého nebe	241	0,06	rána
32. přijít jako rána z čistého nebe	91	0,02	
35. jako rána z čistého nebe	77	0,02	
25. dar z nebe	213	0,05	dar
28. bože na nebi	114	0,03	NE
34. obrátit oči k nebi	89	0,02	oči
36. obracet oči k nebi	77	0,02	
37. cítit se jako v nebi	71	0,02	ANO
40. být jako v nebi	58	0,01	ANO
65. žít jako v nebi	8	-	ANO
72. mít se jako v nebi	3	-	ANO
44. samo nebe nám něco seslalo	56	0,01	ANO
46. samo nebe něco seslalo	45	0,01	ANO
45. dej mu pánbůh nebe	54	0,01	NE

49. na počátku stvořil Bůh nebe a zemi	36	0,01	NE
50. mana z nebe	33	0,01	NE
82. biblická mana z nebe	1	-	NE
51. modrý jako nebe	28	0,01	ANO
59. být modrý jako nebe	14	-	ANO
52. věčné nebe	25	0,01	NE
62. dát věčné nebe	10	-	NE
53. nebe je jako vymetené	21	-	ANO
56. jako hvězd na nebi	17	-	NE
58. mít své nebe	15	-	ANO
78. vidět své nebe	1	-	ANO
61. nebe jako šmolka	10	-	ANO
77. nebe je jako šmolka	1	-	ANO
63. pohnout nebem i zemí	10	-	NE
69. na nebi není ani mráčku	5	-	NE
68. V jako blesk z čistého nebe	5	-	Chyba
71. psí hlas do nebe nedoletí	3	-	NE
80. psí hlas do nebe nedojde	1	-	NE
75. schodů jako do nebe	2	-	ANO
76. N_jako hrom	1	-	Chyba
81. co se něčeho týká	1	-	Chyba

2.2.2. Analýza získaného materiálu

- V procesu značkování korpusu může dojít k chybě. V tabulce uvedené výše obsahuje 3 z 83 položek (3,61%) chybu (68. V jako blesk z čistého nebe, 76. N_jako hrom, 81. co se něčeho týká). Dané mylné výrazy jsou však pouze mechanickou chybou při značkování, nikoli projevem nesprávného užití jazyka.
- Některé frazémy ve slovníku nenajdeme u lexému „nebe“, ale u jiných lexémů:
strom: žádný strom neroste do nebe - 255 výskytů,
rána: rána z čistého nebe – 241 výskyt, přijít jako rána z čistého nebe – 91 výskyt,
jako rána z čistého nebe - 77 výskytů,
dar: dar z nebe – 213 výskytů,
bůh: bože na nebi – 114 výskytů,
oči: obrátit oči k nebi – 89 výskytů, obracet oči k nebi – 77 výskytů.
Dané frazémy mají velkou míru frekvence.
- Velmi zajímavou skupinu představují frazémy se slovem „blesk“: nesou přesně určený význam, který ale může být vyjádřen jednou z alternativ, jichž je velký počet (12 variant: přijít /být / zapůsobit jako blesk z čistého nebe atd.).
- Zvláštní skupinu tvoří výrazy s náboženskou tematikou:
dej mu pánbůh nebe – 54 výskyty,
na počátku stvořil Bůh nebe a zemi – 36 výskytů,
mana z nebe – 33 výskyty, biblická mana z nebe – 1 výskyt,
věčné nebe – 25 výskytů, dat věčné nebe – 10 výskytů.
Dané frazémy nejsou zaznamenány v SČFI, mají specifický a úzce určený význam. Výraz „na počátku stvořil Bůh nebe a zemi“ je přímo citace z Bible, v korpusu se ale vyskytuje v publicistice, i když zachovává svůj význam.
- Na základě korpusových dat lze zjistit, která z variant téhož frazému se používá nejvíc:
být v sedmém nebi – 1387 výskytů,
cítit se v sedmém nebi – 70 výskytů,
mít se jako v sedmém nebi – 2 výskyty,
žít jako v sedmém nebi – 2 výskyty,
být v devátém nebi – 1 výskyt,
být v desátém nebi – 1.

Můžeme konstatovat, že předním frazémem z dané skupiny je frazém „být v sedmém nebi“, zatímco frazémy „být v devátém nebi“ a „být v desátém nebi“, rovněž zachycené v SČFI, se v korpusu objevily pouze jednou. Navíc lze dodat, že slovní spojení „sedmé nebe“ se hlavně používá se slovesem „být“.

- Na základě korpusových dat lze rovněž sledovat historický vývoj jazyka. Podívejme se na jednu z nejfrekventovanějších skupin frazémů:

spadnout z nebe 3017 X spadnout s nebe 44

padat z nebe 1841 X padat s nebe 12

spadlý z nebe 109 X spadlý s nebe 5.

V dnešní době se používá předložka „z“ s genetivem, zatímco dřív převažovala předložka „s“, která se v soudobém jazyce skoro nevyskytuje s druhým pádem.

- Velká zásluha ČNK spočívá v tom, že tady lze vyhledávat a nalézt **nové jazykové spojení a konstrukce**, ještě nezaznamenané slovníky. Příkladem z tabulky 4. slouží dané výrazy.

Výraz „psí hlas do nebe nedoletí“ má 3 výskyty v korpusu: – 3 výskyty:

- „Mohou úřadům pěkně zatopit . S druhými naopak může zatočit úřad - je jich nezajímavá hrstka , v jednom městě , kvůli jedné ulici , na jednom náměstí . **Jejich psí hlas do nebe nedoletí** . A vrchol všeho - představte si jednoho demonstranta , an kvůli oprávněnému požadavku zvýšení svého platu zatarasí ulici . Toho strážník rovnou sbalí. Jeho svobodomyšlný křik o spravedlivém požadavku vyvolá ve všech dojem...“⁶³
- „ I když dobře vím , že **psí hlas do nebe nedoletí** , rozhodl jsem se pro klid svého svědomí touto sebevražednou formou upozornit na trvale se zhoršující stav životního prostředí lidí , bydlících u silnic , zejména 1 . třídy , procházejících obcemi , “ píše ve svém dopise referent...“⁶⁴
- „ Lavičky v parku - jak zabránit mládeži , aby si místo na sedačku sedala na opěradla a mnohdy špinavými botami sedačky znečišťovala ? Tak to je můj dodatečný příspěvek k anketě . I když - **psí hlas do nebe nedoletí**.“⁶⁵

Výraz „psí hlas do nebe nedojde“ se vyskytuje v korpusu jednou:

⁶³ dostupné na:

<https://kontext.korpus.cz/view?ctxattrs=word&attr_vmode=visible&pagesize=40&refs=%3Ddoc.title&q=~AFMq8BF2ti&viewmode=kwic&attrs=word%2Ccol_lemma&corpname=omezeni%2Fsyn_v4&structs=p&attr_allpos=kw>

⁶⁴ dostupné na:

<https://kontext.korpus.cz/view?ctxattrs=word&attr_vmode=visible&pagesize=40&refs=%3Ddoc.title&q=~AFMq8BF2ti&viewmode=kwic&attrs=word%2Ccol_lemma&corpname=omezeni%2Fsyn_v4&structs=p&attr_allpos=kw>

⁶⁵ dostupné na:

<https://kontext.korpus.cz/view?ctxattrs=word&attr_vmode=visible&pagesize=40&refs=%3Ddoc.title&q=~AFMq8BF2ti&viewmode=kwic&attrs=word%2Ccol_lemma&corpname=omezeni%2Fsyn_v4&structs=p&attr_allpos=kw>

- „Jejich autorem je brněnský loutkoherec a boledický rodák Radim Koráb . " Chodím se psem na procházky a ten mě nutí stále se zastavovat a koukat kolem sebe . Tak vznikají mé práce , " řekl ke kolekci devíti barevných snímků s přírodními náměty nazvané **Psí hlas do nebe nedojde** . Sedmatřicetiletý autor , jenž pochází ze slavné ochotnické rodiny Korábů v Boleradicích , se pokaždé nechá inspirovat jarem , různými strukturami a veškerými přírodními úkazy a krásami .⁶⁶

Podle kontextu poznáme, že se tyto frazémy používají pro označení marné snahy, a to se většinou týče činů provedených v společenství nebo v politickém prostředí. Výjimkou je poslední ukázka textu, podle které se dá těžce poznat, zda byl použit výraz „psí hlas do nebe nedojde“ v přeneseném významu či ne. 3 z 4 výrazů byly nalezeny v novinách Deníky Moravia, vzhledem k tomu lze předpokládat, že se daný výraz používá především na Moravě.

2.3.1. Lemma „země“ v ČNK

Tabulka 5, seznam frekvenčních distribucí lemmatu „země“, byla sestavena podle stejného postupu, popsaného v kapitole 2.1.

Tabulka 5

Frazém	Počet výskytů	ipm	Vyskytuje se daný výraz ve SČFI?
1. celkový počet výskytů 2 645 140, 2 575 745 výskytů není označeno jako kolokační lemma (col_lemma)			
2. rodná země	9 709	2,23	ANO
3. srovnat se zemí	8 298	1,91	ANO
4. srazit k zemi	4 545	1,05	ANO
6. srazit na zem	4 261	0,98	ANO
5. jít k zemi	4 537	1,04	ANO

⁶⁶ dostupné na:

<https://kontext.korpus.cz/view?ctxattrs=word&attr_vmode=visible&pagesize=40&refs=%3Ddoc.title&q=~cR785Hn5pD&viewmode=kwic&attrs=word%2Ccol_lemma&corpname=omezeni%2Fsyn_v4&structs=p&attr_allpos=kw>

7. po někom se zem slehla	3 949	0,91	ANO
32. jako by se zem slehla	233	0,05	ANO
54. pod někým se zem slehla	42	0,01	ANO
8. svatá země	3522	0,81	ANO
9. poslat k zemi	3 235	0,74	ANO
26. poslat na zem	521	0,12	ANO
10. ráj na zemi	3 115	0,72	ráj, nebe
18. nebe na zemi	980	0,23	
45. mít ráj na zemi	84	0,02	
66. mít nebe na zemi	18	-	
11. mezi nebem a zemí	2886	0,66	NE
12. držet se při zemi	2561	0,59	NE
13. peklo na zemi	2420	0,56	peklo
56. připravit peklo na zemi	42	0,01	
60. mít peklo na zemi	39	0,01	
14. země nikoho	1777	0,41	ANO
15. zaslíbená země	1433	0,33	ANO
16. matka země	1175	0,27	matka
24. matička země	552	0,13	
87. políbit matičku zemi	2	-	
17. poroučet se k zemi	1150	0,26	ANO
44. poroučet se na zem	88	0,02	ANO
19. zadupat do země	904	0,21	NE

22. vydupat ze země	748	0,17	NE
20. propadnout se do země	797	0,18	ANO
90. jako když se do země propadne	1	-	ANO
21. odlepit se od země	793	0,18	ANO
23. dívat se do země	575	0,13	ANO
28. hledět do země	309	0,07	NE
25. země neomezených možností	530	0,12	ANO
27. suchá země	423	0,1	ANO
29. země tisíce jezer	270	0,06	ANO
30. sůl země	255	0,06	sůl
35. být solí země	148	0,03	
31. zmizet z povrchu země	239	0,05	povrch
48. smést z povrchu země	74	0,02	
50. vymazat z povrchu země	56	0,01	
33. stát oběma nohama na zemi	211	0,05	NE
36. stát na pevné zemi	145	0,03	ANO
42. být na pevné zemi	95	0,02	NE
43. stát oběma nohama pevně na zemi	94	0,02	NE
34. budiž mu země lehká	167	0,04	ANO
58. budiž někomu země lehká	40	0,01	ANO
64. ať je mu země lehká	25	0,01	NE
77. nechť je někomu země lehká	6	-	ANO

37. proletáři všech zemí,spojte se	134	0,03	NE
38. jiná země, jiný mrav	118	0,03	ANO
39. dary země	116	0,03	dar
40. země slunce	102	0,02	slunce
41. praštit sebou na zem	101	0,02	NE
46. království boží na zemi	83	0,02	NE
47. spát na holé zemi	81	0,02	ANO
57. ležet na holé zemi	40	0,01	ANO
49. klanět se až k zemi	59	0,01	ANO
51. syn země	48	0,01	
80. synové a dcery země	3	-	syn
52. a to je ta krásná země, země česká, domov můj	46	0,01	ANO
53. země jezer	43	0,01	NE
55. dědic české země	42	0,01	dědič
59. růst do země	39	0,01	ANO
74. zarůstat do země	8	-	ANO
61. na počátku stvořil Bůh nebe a zemi	36	0,01	NE
62. země pruhů a hvězd	36	0,01	ANO
63. země mlékem a strdím oplývající	28	0,01	ANO
69. země mléko a strdí	15	-	NE
65. hnít v zemi	24	0,01	ANO
67. žít při zemi	15	-	ANO

68. jako by ze země vyrost	15	-	ANO
70. cti otce svého i matku svou, abys dlouho žil a dobře se ti vedlo na zemi	10	-	NE
71. pohnout nebem i zemí	10	-	NE
72. vzít druhou o zem	8	-	NE NE TAK
73. náměstek Kristův na zemi	8	-	NE
81. stát jako do země vrostlý	3	-	ANO
75. vrátit se z oblaků na zem	7	-	NE
76. zlatý prut země české	7	-	NE
78. N jako hrom	4	-	Chyba
79. propadnout se hanbou do země	4	-	ANO (voda)
86. propadnout se studem do země	3	-	
91. moct se propadnout hanbou do země	1	-	
82. mít očistec na zemi	3	-	NE
83. země kalicha	3	-	ANO
84. země nepotřebuje nové kontinenty, ale nové lidi	3	-	ANO
85. ach a jen země je má	3	-	NE
88. nezůstávat za někým pozadu	2	-	Chyba
89. co se něčeho týká	2	-	Chyba
92. někdo to s tebou myslí dobře	1	-	Chyba
93. někdo k pohledání	1	-	Chyba

94. Aj, zde leží zem ta před okem mým slzy ronícím, někdy kolébka, nyní národu mého rakev	1	-	ANO
---	---	---	-----

2.3.2. Analýza získaného materiálu

- V předchozí kapitole (analýza tabulky 4) bylo naznačeno, že při anotaci jazykových jednotek v korpusu může dojít k drobným chybám či překlepům a tím pádem se v seznamu frekvenční distribuce lemmatu můžou objevit i frazémy, které by do toho seznamu neměly patřit. 5 z 94 (**5,31 %**) položek v seznamu frekvenční distribuce lemmatu „země“ je chybných. Tyto frazémy (78. N jako hrom, 88. nezůstávat za někým pozadu, 89. co se něčeho týká, 92. někdo to s tebou myslí dobře, 93. někdo k pohledání) vydělíme stejně šedou barvou pro lepší viditelnost a uvedeme v pravém sloupci poznámku „chyba“.

- Některé frazémy nenalezeme ve slovníku u lexému „země“, ale u jiných lexémů: Dané frazémy mají poměrně velký počet výskytů:

peklo: peklo na zemi – 2420 výskytů, připravit peklo na zemi – 42 výskyty, mít peklo na zemi – 39 výskytů,

sůl: sůl země – 255 výskytů, být solí země – 148 výskytů,

povrch: zmizet z povrchu země – 239 výskytů, smést z povrchu země – 74 výskytů,

vymazat z povrchu země – 56 výskytů,

dar: dary země – 116 výskytů,

slunce: země slunce – 102 výskyty,

syn: syn země – 48 výskytů, synové a dcery země – 3 výskyty,

dědic: dědic české země – 42 výskyty.

- Na základě korpusových dat lze například zjistit, se kterou předložkou se nejčastěji užívá sloveso v konkrétním frazému:

srazit k zemi 4545 X srazit na zem 4261,

poslat k zemi 3235 X poslat na zem 521,

poroučet se k zemi 1150 X poroučet se na zem 88.

Číslo vpravo od výrazů označují počet výskytů frazémů v korpusu. Frazémy „srazit k zemi“ a „srazit na zem“ se objevují v ČNK skoro ve stejném počtu, zatímco u sloves v dalších dvojicích je o hodně frekventovanější předložka „k“ (a tím pádem valence slovesa s dativem), než předložka „na“ s akuzativem. Daný příklad ukazuje na to, že korpus může být nezbytným zdrojem informace pro zkoumání valence sloves.

- Korpus umožňuje sledování rozvoje frazému od základní k rozšířené variantě, např.:

stát oběma nohama na zemi (211 výskytů) + stát na pevné zemi (145 výskytů) = stát oběma nohama pevně na zemi (94 výskyty).

Proměna výrazu se však může projevat nejenom v jeho rozvinutí, ale i v jeho zkracování, např. frazém „země mléko a strdí“, který má 15 výskytů v korpusu, vznikl z frazému „země mlékem a strdím oplývající“, který se vyskytuje v korpusu 28krát. Daný výraz je znám od Aloise Jiráska (Staré pověsti české), ale zajímavý je ten fakt, že se dané výrazy objevují zpravidla v publicistice (81,4 % výskytů), a to v různých novinách z 21. století, nikoli v historických odborných textech nebo próze. Daný jev můžeme chápat tak, že se lze pomocí korpusových dat všimnout tendence „oživení“ ustálených výrazů či citací a přiřazení jim nového kontextu a tím pádem i aktualizace jejich významu.

- Další možnost, kterou nám poskytuje korpus, je srovnávací analýza dat z hlediska jejich aktuálnosti v soudobém jazyce. Probereme to u následujícího příkladu:

budiž mu země lehká – 167 výskytů, budiž někomu země lehká – 40 výskytů,
ať je mu země lehká – 25 výskytů,
nechť je někomu země lehká – 6 výskytů.

Částice budiž – ať – nechť se objevují v korpusu u daného výrazu v poměru 86,97 % – 10,5 % – 2,52 %.

Pokud se podíváme do **korpusu SyD**, kde můžeme porovnat užití těchto částic v diachronním pohledu, tyto částice se tady budou objevovat v poměru 5,99 % – 72,56 % – 21,28 %.

Přestože částice „ať“ je celkově ze tří výše uvedených nejfrekventovanější, u výrazů budiž/ať/nechť převažuje částice „budiž“.

Popsaný paradox může být dán historickým kontextem, eventuálně ustáleností spojení, které se nemění s časem, i když jedna z jeho soustav je rozhodně zastaralá a velice málo frekventovaná.

ZÁVĚR

Daná bakalářská práce měla za cíl zkoumat kolokací lemmat „nebe“ a „země“. Materiál k analýze byl nalezen ve slovnících a v korpusech řady synv4. Praktická část práce reprezentuje dva výzkumy.

První výzkum se zaměřuje na srovnávací analýzu dat získaných ve slovnících a materiálu, nalezeného v korpusu. Frazémy a idiomy obsahující lexémy „nebe“ a „země“ byly vyhledány v prvním až čtvrtém dílech *Slovníku české frazeologie a idiomatiky* Františka Čermáka a zařazeny do dvou přehledných tabulek: Tabulka 1 obsahuje všechny výrazy s jazykovou jednotkou „nebe“ a Tabulka 2 zahrnuje frazémy a idiomy s lexémem „země“. Kromě samotného výrazu a jeho významu, mají této tabulky zvláštní sloupec s poznámkami o tom, zda se konkrétní položka ze slovníku objevuje v korpusu v rámci seznamu zkoumaných kolokačních lemmat. Bylo zjištěno, že převážná většina frazémů ze slovníku se vyskytují v korpusu, u lexému „nebe“ 82,92 %, u lexému „země“ 66 %. Ostatní výrazy byly dodatečně manuálně vyhledány v korpusu a zařazeny do Tabulky 1.1 a Tabulky 2.2. Bylo zjištěno, že přibližně polovina z těchto výrazů se vyskytuje v ČNK, zatímco zbývající frazémy se v korpusu ani neobjevily. Daná součást výzkumu svědčí o tom, že:

- Ne všechny jednotky zařazené do slovníku se běžně používají v živém jazyce,
- I přesto, že je korpus obrovský mimořádný strukturovaný systém, mohou se v něm objevit drobné nedostatky a chyby při kódování jednotlivých lemmat.

Zvláštní podkapitola v dané práci je věnována publikaci *Lidová rčení* Jaroslava Zaorálka z roku 1947. 19 frazémů s lexémem „nebe“ nalezené v daném slovníku jsou zařazeny do Tabulky 3. V ČNK se objevují pouze 10 z 19 frazémů, přičemž jenom 2 z 10 se běžně používají v dnešním jazyce, ostatní se vyskytují velice málo. Výsledky výzkumu svědčí o tom, že jazyk je rychle se vyvíjející a složitý systém, ve kterém se neustále vznikají nové formy, zatímco zastaralé formy mizí.

Druhý výzkum se zaměřuje přímo na analýzu korpusových dat. Upravené seznamy frekvenční distribuce lemmat „nebe“ a „země“ jsou představeny v Tabulce 4 a Tabulce 5. Kromě samotného výrazu zahrnuje tabulka údaje o počtu jeho výskytu v ČNK, relativní hodnotu počtu výskytů na

milion výsledků a poznámky s odkazem na to, zda se daný frazém vyskytuje v SČFI. Na základě získaných v korpusu údajů lze provádět různé zkoumání, např.:

- Lze zjistit, která z variant téhož frazému je nejfrekventovanější,
- Lze sledovat historický vývoj jazyka,
- Lze zkoumat rozvoj frazému od základní k rozšířené variantě,
- Lze zkoumat valenci sloves atd.

ČNK umožňuje nejen pohled na historický vývoj jazyka, ale i na realitu soudobé češtiny. Největší zásluha Českého národního korpusu spočívá v tom, že lze v něm nalézt nové jazykové konstrukce, dosud nezaznamenaný slovníky (např. výraz „psí hlas do nebe nedoletí“).

Frazeologie a idiomatika představuje oblast jazykovědy, kterou je potřeba neustále zkoumat i dále vzhledem k její rychlému vývoji. Můžou k tomu bez pochyby přispívat jazykové korpusy, které nejlépe odráží stav jazyka. Pevně doufám, že daná bakalářská práce může posloužit jako zdroj inspirace pro studenty bohemistiky a zájemce o češtinu a probudit snahu k hlubšímu poznání jazyka.

Seznam použité literatury:

ČECHOVÁ, Marie. *Čeština - řeč a jazyk* [online]. 2. přepracované vydání. Praha: ISV nakladatelství, 2000 [cit. 2017-06-29]. ISBN 80-85866-57-9. Dostupné z:

<http://alephuk.cuni.cz/CKIS-28.html>

ČECHOVÁ, Marie. *Řeč o řeči*. Praha: Academia, 2012. Lingvistika. ISBN 978-80-200-2069-7.

ČERMÁK, František. *Idiomatika a frazeologie češtiny*. Praha: Univerzita Karlova, 1982.

ČERMÁK, František a Renata BLATNÁ. *Jak využívat Český národní korpus*. Praha: Nakladatelství Lidové noviny, 2005. ISBN 80-7106-736-9.

ČERMÁK, František a Michal ŠULC. *Kolokace*. Praha: Nakladatelství Lidové noviny, 2006.

Studie z korpusové lingvistiky. ISBN 80-7106-863-2. Dostupné také z:

http://toc.nkp.cz/NKC/200704/contents/nkc20071706565_1.pdf

ČERMÁK, František. *Frazeologie a idiomatika: česká a obecná = Czech and general phraseology*. Praha: Karolinum, 2007. ISBN 978-80-246-1371-0.

ČERMÁK, František. *Lexikon a sémantika*. Praha: NLN, Nakladatelství Lidové noviny, 2010. ISBN 978-80-7422-020-3.

FILIPEC, Josef a František ČERMÁK. *Česká lexikologie*. Praha: Academia, 1985. Studie a práce lingvistické.

STĚPANOVÁ, Ludmila. *Historie a etymologie českých rčení: bibliografie pramenů*. Praha: Karolinum, 1998. ISBN 80-7184-669-4.

ŠULC, Michal. *Korpusová lingvistika: první vstup*. Praha: Karolinum, 1999. ISBN 80-7184-847-

Seznam použitých slovníků:

BACHMANNOVÁ, Jarmila, Jan BALHAR, Renata BLATNÁ, et al. *Encyklopedický slovník češtiny*. Praha: Nakladatelství Lidové noviny, 2002. ISBN 80-7106-484-X.

BEČKA, Josef Václav. *Slovník synonym a frazeologismů*. 3. vyd. upravené a dopl. Praha: Novinář, 1982. Knihovnička novináře.

ČERMÁK, František, Jiří HRONEK a Jaroslav MACHAČ. *Slovník české frazeologie a idiomatiky. 1, Přirovnání*. Praha: Leda, 2009. ISBN 978-80-7335-215-8.

ČERMÁK, František, Jiří HRONEK a Jaroslav MACHAČ. *Slovník české frazeologie a idiomatiky. 2, Výrazy neslovesné*. Praha: Leda, 2009. ISBN 978-80-7335-215-8.

ČERMÁK, František, Jiří HRONEK a Jaroslav MACHAČ. *Slovník české frazeologie a idiomatiky. 3, Výrazy slovesné*. Praha: Leda, 2009. ISBN 978-80-7335-215-8.

HRONEK, Jiří, Jaroslav MACHAČ a František ČERMÁK. *Slovník české frazeologie a idiomatiky. 4, Výrazy větné*. Praha: Leda, 2009. ISBN 978-80-7335-215-8.

ČERMÁK, František a Jan HOLUB. *Slovník české frazeologie a idiomatiky. 5, Onomaziologický slovník*. Voznice: Leda, 2016. ISBN 978-80-7335-440-4.

ČERVENÁ, Vlasta. *Slovník spisovné češtiny pro školu a veřejnost: s dodatkem Ministerstva školství, mládeže a tělovýchovy České republiky*. Vyd. 4., dotisk. Praha: Academia, 2014. ISBN 978-80-200-1080-3.

ZAORÁLEK, Jaroslav. *Lidová rčení*. Vyd. 3., reprint 1. vyd. z r. 1947. Praha: Aurora, 1996. ISBN 80-85974-10-X.

Seznam použitých internetových zdrojů:

<http://wiki.korpus.cz/doku.php>

<http://korpus.cz/>

<http://prirucka.ujc.cas.cz>