

JIHOČESKÁ UNIVERZITA V ČESKÝCH BUDĚJOVICÍCH
FILOZOFICKÁ FAKULTA
ÚSTAV ROMANISTIKY

DIPLOMOVÁ PRÁCE

KAPESNÍ ČESKO-FRANCOUZSKÉ SLOVNÍKY:
VÝBĚR HESEL VS. FREKVENCE LEMMAT V KORPUSU

Vedoucí práce: doc. PhDr. Jan Radimský, Ph.D.

Autor práce: Tereza Antošová

Studijní obor: Francouzský jazyk

Ročník: 3.

2016

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

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

V Českých Budějovicích, 6. dubna 2016

.....
Tereza Antošová

Poděkování

Ráda bych touto cestou poděkovala doc. PhDr. Janu Radimskému, Ph.D. za jeho vstřícný přístup, cenné rady a odborné vedení při zpracování této práce.

OBSAH

Anotace.....	6
Annotation.....	7
Úvod.	8
I. Teoretická část	
1. Slovník a lexikografie.....	10
2. Slovníky a jejich typologie.....	11
2.1. Překladové slovníky.....	19
2.2. Makrostruktura slovníku.....	22
2.2.1. Úvod a přílohy.....	22
2.2.2. Slovníková stať.....	23
2.2.2.1. Řazení slov ve slovnících.....	23
2.2.2.2. Výběr hesel do slovníku.....	26
2.3. Mikrostruktura slovníku.....	29
2.3.1. Typy hesel.....	30
2.3.2. Výstavba hesla.....	31
2.3.3. Struktura a složky hesla.....	32
2.3.3.1. Lemma neboli slovníkový vstup.....	33
2.3.3.2. Gramatická charakteristika.....	34
2.3.3.3. Ekvivalenty.....	34
2.3.3.4. Valence.....	38
2.3.3.5. Exemplifikace.....	38
2.3.3.6. Víceslovné lexémy.....	40
2.3.3.7. Technický odkaz ve slovníkovém hesle.....	41
2.4. Homonymie a polysémie.....	41
2.5. Metajazyk ve slovníku.....	43
2.6. Slovníkářovo desatero.....	44
3. Korpus.....	46
3.1. Typologie korpusů.....	49
3.2. Český národní korpus.....	50
3.2.1. Korpus SYN 2005.....	51
3.3. Výhody a nevýhody neboli možnosti a meze korpusu.....	52
3.4. Další druhy materiálů pro lingvistické účely.....	55

3.5. Lexikografické využití korpusu.....	57
4. Frekvenční slovník češtiny.....	61
<i>II. Empirická část</i>	
1. Úvod.....	63
2. Materiál.....	64
2.1. Slovníky.....	64
2.1.1. FRAGMENT.....	66
2.1.2. FIN.....	68
2.1.3. OLOMOUC.....	70
2.1.4. LINGEA.....	71
2.1.5. LEDA.....	72
2.2. Korpus SYN2005.....	73
3. Metoda.....	74
3.1. Srovnání jednotlivých slovníků	75
3.2. Frekvence vs. slovníková hesla.....	76
Závěr.....	85
Resumé.....	87
Seznam použité literatury.....	89

Anotace

Práce si klade za cíl zhodnotit vybrané překladové slovníky, a sice současné česko-francouzské kapesní slovníky. Hodnocena bude celková výstavba slovníku, ale především výběr hesel, a to jednak z hlediska frekvence na základě zvoleného korpusu, jednak bude brán ohled i na potenciální uživatele. Účelem je tedy zhodnotit, do jaké míry tyto slovníky odpovídají potřebám předpokládaných adresátů. Kromě studie je zařazena i teoretická část, která se zabývá typologií slovníků, jsou zde vysvětleny základní pojmy z lexikografie, řeší se zde otázky související s výstavbou slovníků. Součástí jsou i kapitoly pojednávající o korpusech a frekvenčních slovnících. Teoretické poznatky z první části jsou potom využity při analýze slovníků v empirické části a zohledněny při celkovém hodnocení slovníků.

klíčová slova:

slovník, korpus, frekvence, heslo, lexikografie

Annotation

The aim of this dissertation is to critically evaluate the set of chosen dictionaries, concretely the set of five Czech-French contemporary pocket dictionaries. The evaluation concerns not only the overall composition but mainly the choice of respective entries in order to determine to which extent they satisfy supposed users and meet their requirements. The data are analyzed from the point of view of frequency based on the chosen corpus as well as from the point of view of potential user's needs. Besides the study, the present dissertation also includes the theoretical part which presents typology of dictionaries and the key terms from the field of lexicography are explained. This part also focuses on questions concerning dictionary composition. Furthermore the chapters about corpuses and frequency dictionaries are included. The theoretical observations from the first part are taken into consideration in the second part as well as in the overall evaluation of results.

key words

dictionary, corpus, frequency, entry, lexicography

Úvod

Slovníky, resp. jejich dřívější formy, člověku sloužily již od starověku a v naší společnosti hrají významnou roli i v dnešní době. Snad každý někdy hledal ve slovníku, ať už se jednalo o slovník výkladový nebo překladový. Slovníky nám totiž pomáhají zdokonalovat jak jazyk cizí, tak i jazyk mateřský. Jejich kvalita a kvantita odráží úroveň naší společnosti a zanechávají tak důležitý kulturní odkaz.

Právě slovníky představují hlavní téma této práce, která se soustřeďuje na konkrétní druh dvojjazyčných slovníků, a sice na současné česko-francouzské kapesní slovníky. Práce si klade za cíl zhodnotit výběr hesel v těchto malých překladových slovnících, a to jak z hlediska korpusové frekvence, tak z hlediska potřeb potenciálních uživatelů slovníku. Budeme se tedy zabývat otázkami, které řeší lexikograf při výběru hesel a výstavbě slovníku a zároveň se budeme snažit zjistit, do jaké míry odráží vybrané slovníky korpusovou frekvenci, frekvenční seznamy a nakolik vybrané slovníky vyhovují požadavkům stanovených jejich autory.

Práce je rozdělena na tři hlavní části: 1) úvod, 2) hlavní stať, která se dělí na část teoretickou a empirickou, 3) závěr.

Část teoretická se zabývá typologií slovníků, zvláštní pozornost je pak věnována slovníkům překladovým. Dále se tato část zaměřuje na otázky spojené s výstavbou slovníku, jsou zde vysvětleny základní pojmy související s touto problematikou. Pojednává se zde tedy o makrostruktuře a mikrostruktuře slovníku a jejich jednotlivých prvcích. Součástí teoretické části jsou i kapitoly o korpusech a frekvenčních slovnících.

Vzhledem k tomu, že hlavním tématem této práce jsou kapesní slovníky, druhá, empirická, část, je věnována analýze vybraného souboru slovníků. Nejprve jsou tedy zařazeny kapitoly uvádějící potřebný materiál – slovníky a korpus. Je představena pětice vybraných slovníků, u nichž jsme se snažili zhodnotit kladné a záporné aspekty, a to na základě teoretických poznatků z první části. Další kapitola je věnována korpusu, který jsme během našeho výzkumu použili, a sice korpus SYN2005. Dále je pak popsána použitá metoda a následuje představení samotných výsledků rešerše. Závěrem jsou zhodnoceny výsledky vyplývající ze studie a shrnuty důležité poznatky.

Dále bychom rádi zmínili, že zdrojem inspirace pro druhou část byla studie M. Šulce *Kapesní slovníky a frekvence lemmat v korpusu*, o níž se naše rešerše opírá. M. Šulc prováděl obdobný výzkum rovněž na souboru pěti kapesních slovníků, nicméně česko-anglických. Dalším rozdílem je použitý korpus - Šulc konfrontoval hesla s frekvenčním seznamem korpusu FSC2000, který je variantou korpusu SYN2000, zatímco v naší rešerši bylo porovnání provedeno na korpusu novějším, a to SYN2005. Tento korpus jsme zvolili proto, že slovníky z našeho souboru vznikaly a vyšly právě v tomto období.

Část II.

Teoretická část

1. Slovník a lexikografie

Vzhledem k tomu, že se naše práce týká problematiky slovníku a jeho výstavby, budeme se zabývat různými aspekty z oblasti lexikografie. Proto si na úvod zmíníme několik poznatků týkající se tohoto oboru aplikované lingvistiky.

Lexikografie neboli česky slovníkářství, je věda, která se, jak už její český název napovídá, zabývá teorií a praxí vytváření slovníků. Tuto disciplínu lze rozdělit na teoretickou, neboli metalexikografii, která zkoumá historii, strukturu, či metodologii slovníků a odlišnosti jejich použití, popř. i typologii slovníků, a na praktickou lexikografii, která se zabývá vytvářením slovníků.¹ Lexikografie tak v praxi aplikuje teoretické poznatky lexikologie, tedy disciplíny zabývající se lexikálními jednotkami určitého jazyka a jejich užíváním. Ačkoli se jedná o dva úzce spjaté obory, vztah mezi lexikologií a lexikografií není tak zřejmý, jak se může zdát. Někdy se totiž uvádí, že lexikologie se dělí do několika subdisciplín, mezi nimiž i lexikografie, popř. se na lexikografii nahlíží jako na aplikovanou lexikologii, jak dokládá i F. Čermák: „Lexikografická věda bývá často charakterizována jako aplikovaný úsek nauky o slovní zásobě, tedy jako aplikovaná část lexikologie.“² Pokud bychom přijali toto hledisko, měl by zde existovat vztah nadřazenosti (lexikologie) a podřazenosti (lexikografie), nicméně vzájemný vztah těchto dvou disciplín je složitější. Zaprvé bychom mohli zmínit fakt, že „lexikografie ve své primitivní podobě se objevila mnohem dříve než lexikologie coby teoretická disciplína studující slovní zásobu. Prvopočátky pokusů o popis smyslu některých znaků a snahy o zaznamenání objektů, které jsou těmito znaky popisovány, můžeme klást již do období dávného starověku.“³ Zadruhé uvedme, že lexikografie při tvorbě slovníků vychází z lexikologické teorie, naopak lexikologie při svých analýzách čerpá z poznatků při zpracování slovníků. Oba tyto důležité obory aplikované lingvistiky mají však jedno společné, a tím je slovní zásoba. Rozdíly mezi lexikologií a lexikografií plynou z toho, jak ke slovní zásobě přistupují.

¹ Brabcová, 2012 : 11

² Čermák, Blatná, 1995 : 16

³ Kapusta, 2008 : 9

2. Slovníky a jejich typologie

Na úvod si vypůjčme tvrzení českého lexikografa a národního buditele Josefa Jungmanna, za jehož časů, tedy v období národního obrození, nastal největší rozvoj slovníkářství v českých zemích. Jungmann v předmluvě ke svému dílu *Slovník česko-německý* definuje slovník následovně:

„Slovník národního jazyka náleží mezi první potřeby vzdělaného člověka.“⁴

Můžeme si klást otázku, zda od dob národního obrození slovník neztratil svou důležitost a i v dnešní moderní době náleží mezi první potřeby vzdělaného člověka, nicméně stále platí, že pojmem slovník rozumíme „souborné zpracování slovní zásoby jazyka (nebo její určité složky) v jednotlivých heslech, nejčastěji abecedně řazených, s připojeným výkladem nebo překladem resp. knižní soubor těchto hesel“⁵

Slovník je tedy základem jazykového systému, je „depozitářem všech jazykových jednotek, depozitářem velmi složitě a přitom funkčně organizovaným.“⁶ F. Čermák definuje slovník takto: „Slovník je příručka mapující lexikálně-sémantické korelace a všechny další systémové vztahy mezi dvěma jazyky, s nimi pevně spjaté, která má komplexní paradigmaticko-syntagmatickou povahu.“⁷ Slovník bývá někdy také nazýván jako lexikon, ten bychom však spíše definovali jako úhrn slov (lexémů) daného jazyka.

Čermák uvádí různé nominativní aspekty slovníku. Dle Čermáka „podle různých přístupů vyvstávají i různé stránky slovníku. V přehledu jde především o tyto:⁸

1. složení: *apelativa* (jména obecná) a *propria* (jména vlastní)

Pozn.: Propria studuje subdisciplína **onomastika**, resp. zvl. **antroponomastika** (názvy lidí) a **toponomastika** (názvy míst)

2. procesuálně: **onomaziologie** (od významu formě, přiřazování), **sémaziologie** (od formy k významu, analýza, dešifrace)
3. teoreticky a deskriptivně: **lexikologie** (teorie), **lexikografie** (aplikace)

⁴ Jungmann, J.: Slovník česko-německý. www.kramerius.nkp.cz [online].

⁵ Slovník. www.priLucka.ujc.cas.cz [online].

⁶ Čermák, Blatná 1995 : 95

⁷ Tamtéž : 231

⁸ Čermák, 2011 : 192

4. šíří uplatnění: obecná a nespécifická část slovníku a **terminologie** (založená na exaktní definici denotátu různých profesí)
5. časově: **synchronní** slovník a **diachronní** slovník(y) jednotlivých stavů jazyka a **etymologie** (původ slov)
6. kombinatoricky: *pravidelné* kombinace (terminologické, náhodné a jiné) a *anomální* kombinace, studované **frazeologií a idomatičkou**
7. funkčně (v pražské tradici): **funkční onomatologie** (nauka o pojmenování, resp. nominacích, zde jen neustálených), orientovaná na přiřazování lexému (formy) výchozímu významu
8. znakově: *indexy* (zvl. pronomina a propria), *ikony* (onomatopoeia) a *symbols* (ostatní)

V Čermákově pojetí by měl slovník mít základní vybavení dvojího typu:⁹

- Měl by zařazovat lexém do systému souvislostí, především tříd, kam ten který lexém patří
- Měl by naznačovat valenci a základní syntagmatiku, spojitelnost, pro každé heslo, resp. heslové slovo i ekvivalent

Tyto zásady často nejsou respektovány především v malých kapesních slovnících, k čemuž se F. Čermák vyjadřuje následovně: „Je to věc zcela zásadní a základní a naopak to, co běžná praxe vydává za slovníky, je jen hromada pracných a iritujících polotovarů. Toto platí zvl. o malých a kapesních slovnících. Shrnuto jinak, znamená to, že slovník by se neměl dělat podle zásady slovíčko za slovíčko (kromě slovníků terminologických, např. z matematiky, fyziky apod.) a protože domácímu slovu odpovídá jednoznačně cizí slovo jen málokdy, je třeba řídit zásadou, že heslu odpovídá jakožto komplex informací taky ‚heslo‘ (resp. heslová stať, jak se vlastně nazývá úhrn ekvivalentů a dalšího).“¹⁰

Slovníky se tradičně vyskytují nejčastěji v knižní podobě, k dostání jsou však i digitální slovníky na CD či na internetu. V poslední době se objevují i slovníky dostupné v rámci mobilních aplikací. Mezi nejvýznamnější česká nakladatelství slovníků patří Fraus, Leda, Lingea, Fin či TZone.¹¹

⁹ Čermák, 1989 : 3

¹⁰ Tamtéž

¹¹ Slovník-cz.eu [online]

Za zmínku rovněž stojí, že některé jazyky, jako je např. angličtina či francouzština rozlišují dvě slova, korespondující s českým slovem slovník. I když je tato problematika komplexnější, uveďme si alespoň zjednodušeně, v čem spočívá základní rozdíl. Jednak existuje slovo *dictionary*, které odkazuje na knižní, popř. elektronickou, formu slovníku obsahující slova a jejich významy. Jednak máme také slovo *vocabulary*, pod kterým si lze představit seznam slovní zásoby určitého jazyka či osoby, popř. seznam slov užitých v nějaké publikaci, článku, apod. Tyto dva různé pojmy lze přeložit jako *slovník (dictionary)* a *slovní zásoba (vocabulary)*.

Čermák se domnívá, že povaha slovníku je dána jeho cílem, typem uživatele a potřebou, kterou má slovník plnit. Takto lze rozhodnout, zda půjde např. o slovník studijní, anebo jen o komentovaný seznam slov. Mezi těmito póly samozřejmě existuje řada stupňů a podob. Stejnými ohledy je dán i rozsah slovníku.¹² Každý typ slovníku má tak své nároky teoretické, lingvistické i mimo-lingvistické, vyžaduje jistou metodu a metodiku, jakož i realizaci různých pracovních etap, ať už se jedná o jednojazyčné slovníky výkladové, např. z lékařského či technického prostředí, nebo slovník spisovné češtiny, do kterého lze nahlédnout, řešíme-li nejasnosti týkající se mateřského jazyka. Při studiu cizího jazyka zase uplatníme slovníky dvojjazyčné neboli překladové. Slovníky nám tak pomáhají zdokonalovat jak jazyk mateřský, tak cizí.¹³ Obecně lze říci, že existují různá kritéria a náhledy, podle kterých lze slovníky dělit.

Slovníky můžeme dělit podle rozsahu. Autoři F. Čermák a R. Blatná rozlišují v publikaci *Manuál lexikografie* slovníky dle rozsahu takto¹⁴:

- A) kapesní slovník - obsahuje 10 – 15 tisíc hesel
- B) střední slovník - obsahuje 35 – 50 tisíc hesel
- C) velký slovník - obvykle 80 tisíc a více hesel, se snahou být vyčerpávající.

¹² Čermák, 1989 : 1

¹³ Saksová, V. 2014 : 36

¹⁴ Čermák, Blatná, 1995 : 233

Je ovšem nutno brát v potaz, že kritéria rozsahu se mohou dle jednotlivých zdrojů lišit. Např. J. Šabršula v knize *Základy francouzské lexikologie* rozlišuje slovníky na základě velikosti následovně¹⁵:

A) minislovníky, mezi které dále řadíme tzv.

- a) vokabuláře (dvojjazyčný slovník v podobě přílohy k vydání textu cizího autora)
- b) glosáře (seznam slov s jejich vysvětlením)

B) malý slovník má většinou 10 – 15 tisíc hesel

C) střední slovník obsahuje 35 – 50 tisíc hesel

D) velký slovník, tzv. Thesaurus, jehož snahou je pokrýt co možná nejvíce informací

Malý slovník, má rozsah přibližně 10-15 tisíc hesel, bývá také označován jako kapesní. Obvykle zastupuje základní slovní zásobu. Jsou zde obsažena podstatná jména, přídavná jména a slovesa, zařazeny mohou být i nejužívanější spojky, předložky, částice a citoslovce. Jeho formát je obvykle A6 nebo menší.¹⁶ Některá nakladatelství vyčleňují v rámci malých slovníků různé další podtypy. Kromě kapesních slovníků se tak můžeme setkat např. i s označením šikovný; praktický; cestovní; nejen na cesty; kolibří; minislovník; středně malý; velký kapesní.

Střední slovník obsahuje zpravidla mezi 30-50 tisíci hesly. Obvykle obsahuje jak aktivní, tak pasivní slovní zásobu, jakož i základy gramatiky. Jsou zastoupeny předložky, spojky, částice i citoslovce. Rovněž může být zařazen i přehled nepravidelných sloves. Formát slovníku bývá A5.¹⁷

Počet hesel velkého slovníku se pohybuje od 80 tisíc a výše. Vzhledem ke svému rozsahu může obsahovat i několik svazků. Ve slovníku bývá zahrnuta jak aktivní, tak pasivní slovní zásoba, všechny slovní druhy a gramatické nepravidelnosti. Velké slovníky zpravidla obsahují také frazeologii, idiomy či přísloví. Většina hesel obsahuje informaci

¹⁵ Šabršula, 1983 : 247

¹⁶ Brabcová, 2012 : 16

¹⁷ Tamtéž

o valenci a jejich použití. Někdy mohou být součástí slovníku i zeměpisné údaje, či seznamy vlastních jmen.¹⁸

Dále lze slovníky dělit podle počtu jazyků na slovníky:

- A) Výkladové (jednojazyčné)
- B) Překladové (dvou a vícejazyčné)

Výkladový slovník místo překladu slov nabízí jejich výklad, resp. vysvětlení, co dané slovo znamená. Výklad je napsán ve stejném jazyce jako popisované slovo, slovník tudíž označujeme jako jednojazyčný, neboli monolingvní. Výkladový slovník tedy předpokládá určitou znalost jazyka, a proto není vhodný pro úplné začátečníky daného jazyka.¹⁹ Výkladové slovníky lze dále dělit podle různých kritérií a oblastí zaměření.

Překladové slovníky nejčastěji obsahují dvě části. Na základě toho rozlišujeme slovníky jednosměrné a obousměrné. V jednosměrném slovníku se obvykle překládá pouze z cizího jazyka do vlastního, popř. z vlastního do cizího, obousměrný slovník pak poskytuje uživateli informace v obou jazycích. I tyto slovníky mohou být specializované a omezovat se jen na odborné termíny z určité oblasti (např. terminologie z oblasti práva, chemie či lékařství).²⁰

U překladových slovníků lze dále ještě rozlišit, zda se jedná o slovník aktivní nebo pasivní. Aktivní slovník se zaměřuje hlavně na jazyk cílový, v němž by měla být vyjádřena informace gramatická, sémantická, stylistická, a to tak, aby byla srozumitelná pro uživatele, který není rodilým mluvčím. Naproti tomu pasivní slovník často obsahuje pouhý seznam přeložených hesel a spoléhá na uživatelovy jazykové znalosti a schopnosti, které mu umožní tyto překlady rozlišit. Takto ušetřený prostor je pak využit ke zvýšení počtu hesel a významů. Překladovým slovníkům se budeme blíže věnovat v následující kapitole.²¹

„Vícejazyčné slovníky bývají většinou jen velmi základní pomůckou pro překlad z a do více cílových jazyků, jejich značnou nevýhodou je omezený prostor pro výklad

¹⁸ Brabcová, 2012 : 16

¹⁹ Co je slovník? www.ajslovicka.cz [online]

²⁰ Pfliegerová, 2013 : 27

²¹ Brabcová : 15

jednotlivých hesel. Všeobecné vícejazyčné slovníky byly velmi oblíbené hlavně v minulosti, dnes se spíše objevují vícejazyčné slovníky konkrétně tematicky zaměřené (např. kulinářské, medicínské, apod.) nebo v podobě jazykových příruček pro turisty a cestovatele, kterým poslouží jako základní nástroj pro porozumění nejběžnějším výrazům.“²²

Majja V. Lazová se problematice typologie překladových slovníků věnuje ve svém článku *Obecné a dílčí problémy při práci na dvojjazyčných slovnících různých typů*.²³ Podle autorky je možné slovníky řadit podle toho, pro jaký specifický okruh uživatelů je publikace určena. Na základě toho kritéria je možno rozlišit nejen strukturu, ale i samotný obsah slovníků. Vymezit lze pět základních kategorií:

1. Slovníky pro žáky základních a středních škol, pro osoby začínající studovat cizí jazyk
-slovník obsahuje do 12.000 hesel
2. Slovníky pro studenty technických, nejazykových vysokých škol
-slovník obsahuje do 25 000 hesel
3. Slovníky pro studenty jazykových vysokých škol
-slovník obsahuje do 35 000 hesel
4. Slovníky pro široký okruh uživatelů, kteří jsou schopní číst uměleckou literaturu a tisk v cizích jazycích
-slovník obsahuje do 50 000 hesel
5. Slovníky pro odbornou část veřejnosti (pro vědecké pracovníky, překladatele)
-slovník obsahuje 100 000 a více hesel

²² Pfliegerová, 2013 : 25

²³ Lazová, 1974: 133-138 in Pfliegerová, 2013

Jednu z možných typologií podle příručky *Manuál lexikografie*²⁴ zachycuje následující tabulka. Znázorňuje schéma, které třídí typy slovníků od nejvíce obecných ke konkrétním podle jejich základních rysů:

Rys	Typ slovníku
1) (Obsah) obecně	A) Encyklopedické
	B) Jazykové (viz 2)
2) (Čas) jazykové	A) Synchronní (viz 3)
	A) Diachronní a. historické b. etymologické
3) (Jazyky) synchronní	A) Jednojazyčné (viz 4)
	B) Vícejazyčné/ překladové a. Dvojazyčné b. vícejazyčné
4) (Cíl) jednojazyčné	A) Normativní (obvykle školní)
	B) Deskriptivní/ vědecké (viz 5)
5) (Uspořádání) deskriptivní	A) Alfabetické (viz 6)
	B) Nealfabetické a. Retrográdní b. Frekvenční c. věcné/ tezaurové
6) (Oblast) alfabetické	A) Dílčí a. Individuální i. autora ii. díla b. geograficky stratifikované i. dialektické c. sociálně stratifikované d. argotické aj. e. oborově věcné i. terminologické ii. slangové f. oborově jazykové i. onomaziologické ii. frazeologické iii. onomastické
	B) celonárodní (viz 7)
7) (Východisko)	A) onomaziologické a. Tezaurus/ věcný b. Synonymické c. Hyponymické aj.
	B) Sémaziologické a. Obecné-dílčí b. Ortografické-ortoepické
8) (Médium uložení)	A) knižní
	B) elektronické

²⁴ Čermák, Blatná, 1995 : 11-12

Je nutné podotknout, že některé kategorie se vzájemně prolínají, jiné se naopak mohou ještě dále větvit.

Velmi specifické členění slovníků představil stoupenec pražské školy, český rusista ruského původu Leontij Kopeckij²⁵, který uvádí, že podle adresáta slovníku je možné vyčlenit okruh příruček pro čtenáře uměleckých textů a publicistiky, běžných tiskovin, čtenáře z oblasti humanitních věd, dále pak slovníky pro technické a přírodovědné obory, ale také pro slovníky z oblasti archeologie, psychologie, apod. Jak ale ve své práci namítá Pfliegerová, „tato typologizace je však poněkud zavádějící a nepřesná, jelikož nelze obsáhnout všechny možné oblasti zájmů potencionálních adresátů, se kterými by autor slovníku mohl počítat již při jeho koncipování. Z tohoto důvodu je daleko přínosnější členění slovníků uvedené výše.“²⁶

Z výše uvedeného vyplývá, že „třídění a klasifikace slovníků je velmi obsáhlá a nelze vymezit jedno kritérium, které by bylo závazné pro všechny. Jednotlivé druhy slovníků v sobě mohou zahrnovat více vymezených rysů a zaměření, na druhé straně mohou být některé slovníky natolik specifické, že je složité je do nějaké kategorie zařadit. Dalším faktem pak je, že nejrůznějších druhů slovníků je tolik, že by bylo možné jejich schéma dělit na nespočet menších a menších kategorií, což by však bylo značně nepřehledné.“²⁷ Snažili jsme se proto uvést reprezentativní vzorek možných typologií dělení slovníků.

Pro shrnutí lze říci, že „slovníky bývají určeny kolektivnímu uživateli, tj. jednotlivcům různých jazykových znalostí, předpokladů, zájmů, nelingvistům, mluvčím téhož jazyka i cizincům.“²⁸ „Dobře definovaný cíl, a tedy zřetel k typu uživatele a reálné potřebě tak předurčuje slovníkový rezultat co do parametrů. Typologie slovníků je tedy dána potřebám jednotlivých uživatelů (které mohou být různorodé). Rozhoduje tu tedy poptávka, kde rovnováha mezi snahou vyhovět takto požadavkům trhu a nutností zachovat alespoň minimální teoreticko-odbornou úroveň může být dost delikátní. Bývá smutně porušována zvláště u malých komerčních slovníků. Tvůrci slovníku se někdy snaží vytvořit slovník o co širším záběru pro co největší počet uživatelů. Takový slovník však může být vágní, a tudíž hrozí, že povede k nejistotě a nebude dobře sloužit nikomu.

²⁵ Kopeckij, 1974 : 38-39 in Pfliegerová, 2013

²⁶ Pfliegerová, 2013 : 26

²⁷ Pfliegerová, 2013 : 22

²⁸ Čermák, Blatná, 1995 : 18

Proto by měl správným stanovením cíle, funkce, povahy, rozsahu a typu slovníku být dán soubor východisek k vlastní tvorbě slovníku, která by však nikdy neměla být pouhou kompilací z předchozích publikací.²⁹ Ve druhé, empirické, části této práce se budeme zabývat právě dvojjazyčnými kapesními slovníky, jejich výstavbou a analýzou slovníkových hesel. Tento průzkum si klade za cíl zjistit, jak autoři při výstavbě těchto slovníků postupují, co zohledňují a komu slovníky primárně adresují. S tím souvisí otázka cílového uživatele. Obvykle je jím uživatel, který nemá s daným jazykem mnoho zkušeností, tj. začátečník. Tím může být turista či student. Z toho vyplývá, že kapesní slovníky se sice zaměřují na určitou skupinu uživatelů, zároveň se ale snaží vytvořit slovník pro co nejširší okruh uživatelů. Výběr hesel by tedy měl být korespondovat s požadavky těchto uživatelů. Měla by být zařazena hesla nejfrekventovanější. Toto téma je hlavním předmětem druhé části práce, kde budeme blíže zkoumat, která kritéria ovlivňují výběr hesel, ale i celou výstavbu slovníku a nakolik se od sebe vybrané kapesní slovníky v tomto ohledu liší.

2.1. Překladové slovníky

Jelikož druhá část této práce budeme věnována vybraným česko-francouzským kapesním slovníkům, které se řadí mezi slovníky dvojjazyčné neboli překladové, bude následující podkapitola věnována právě tomuto druhu slovníků.

Jak již bylo naznačeno, překladový (zvl. dvoujazyčný, bilingvní) slovník je takový, který obsahuje hesla ve výchozím jazyce a definuje je překladem do jazyka cílového. Podnětem k jejich vzniku je potřeba dorozumět se v cizí řeči. Takový slovník tedy slouží k usnadnění kontaktu mezi komunitami s různými jazyky, je prostředkem ke komunikaci mezi těmito komunitami. Tento druh slovníku je užíván k překladům nejrůznějšího druhu, napomáhá k osvojení jazyka a rozšiřování slovní zásoby. „Předmětem překladového slovníku je poskytnout adekvátní překlad. To znamená nalézt takové lexikální jednotky cílového jazyka, které by po použití v daném kontextu zachovaly co nejpřesněji veškeré informace o daném slovu, a splnily tak podmínky funkční ekvivalence překladu.“³⁰ Zjednodušeně by se dalo říci, že překladový slovník je dílo, jehož náplní jsou překlady slov z jazyka A do jazyka B. F. Čermák ve své stati

²⁹ Čermák, Blatná, 1995 : 91

³⁰ Faltová, 2007 : 10

Překladová lexikologie charakterizuje překladový slovník jako „hledání možností a prostředků vzájemné substituce, tj. specifičtěji hledání takových lexikálních jednotek cílového jazyka (ekvivalentů), které po vložení do daného kontextu poskytnou hladký překlad resp. překlad adekvátní.“³¹

F. Čermák uvádí některá úskalí při tvorbě překladových slovníků. Tvorba překladových, přesněji dvoujazyčných slovníků, je prakticko-teoretická činnost, trpící řadou problémů. Obecně je třeba si připomenout především tyto³²:

1. Cizí jazyk, do kterého se výchozí jazyk slovníkově převádí, by měl pokrývat rodilý mluvčí, tj. cizinec a naopak: ani ten nejlepší překladatel se nemůže rovnat tomu, kdo jazyk zná od narození.
2. Slovník je umění možného: v každém slovníku jsou nějaké nedokonalosti (popř. chyby) a je mylné se domnívat, že lze vytvořit slovník bez chyb (pojetí, výklad chyby ostatně záleží na teoretickém východisku, a je tedy dost relativní) je třeba je ovšem eliminovat na minimum a slovníkář se proto musí umět vyrovnat se svou omezenou znalostí
3. Slovník nelze chápat jako seznam slov, ke kterým hledáme jiná slova v dalším seznamu, ale jako komplexní a strukturovaný soubor hesel ke kterým hledáme ekvivalentní hesla v druhém souboru. Místo o slovníku by tedy bylo výstižnější mluvit o ‚hesláři‘.

Jak Čermák dále tvrdí „je nedobrou tradicí překladových slovníků, že svou materiálovou základnu přebírají ze slovníků výkladových. Pro překladový slovník je třeba mít k dispozici počítačové korpusy obou jazyků (ideálně paralelní). Oproti tištěným textům je v nich totiž obsaženo několikanásobně větší množství materiálu v obou jazycích, což umožňuje vyhledat ekvivalenty v různých kontextech a tím snáze najít nejvhodnější heslo.“³³

³¹ Čermák, Blatná, 1995 : 233

³² Čermák, Blatná 1995 : 231

³³ Čermák, Blatná 1995 : 231

Cílem překladového slovníku je postihnout slovní zásobu cizího jazyka takovým způsobem, aby uživatel pochopil význam každé lexikální jednotky. Dle potřeb, které překladové slovníky nejčastěji plní, je Čermák shrnuje do tří okruhů³⁴:

- 1) slovníky pro ne odbornou veřejnost – příručky slouží např. při cestování nebo obchodě, zahrnují běžnou, každodenní komunikaci, uživatelé mohou být turisté, studenti, překladatelé a tlumočníci
- 2) slovníky pro odborné překladatele – obsahují odbornou terminologii
- 3) slovníky pro studenty (zvl. univerzitní) a lingvisty - studium daného jazyka

Tomuto rozlišení odpovídají i nejčastější typy slovníků, tj. obecné, studijní a terminologické.

Dle Čermáka lze překladové slovníky rozdělit podle následujících kritérií³⁵:

- A) Podle počtu – slovníky jednojazyčné, dvojjazyčné či vícejazyčné
- B) Podle nositele jazyka
 - a. národní jazyk
 - i. spisovný jazyk
 - ii. slang
 - iii. argot
 - b. idiolekt
- C) Podle rozsahu komunikace – slovník obecný nebo s omezenou oblastí platnosti, kterou může být např. hantýrka
- D) Podle oboru jazyka –
 - a. slovník jazyka obecného
 - b. slovník týkající se určité jazykové nauky, např. frazeologie
- E) Podle oboru věcného
 - a. Slovník specializovaný, např. terminologický
 - b. Slovník nespécializovaný
- F) Podle času
 - a. Slovník diachronní, který sleduje vývoj a proměny jazyka v průběhu času

³⁴ Čermák, Blatná 1995 : 232

³⁵ Čermák, Blatná 1995 : 235

- b. Slovník synchronní, který zkoumá současný jazyk
- G) Podle zaměření slovníku např. slovník obecný či studijní

2.2. Makrostruktura slovníku

Makrostruktura neboli organizace slovníku je dle Čermáka smysluplné a jednoznačné uspořádání a provázání hesel do celku. Toto uspořádání hesel (tj. heslových slov, popř. částí slov a spojení slov) tvoří stať slovníku. Soubor heslových slov bývá někdy označován jako nomenklatura. V širším pojetí se k makrostruktuře může řadit i úvodní a závěrečná část slovníku. Tyto složky tvoří makrostrukturu slovníku.³⁶

Má-li být slovník relativně uzavřeným a samostatným referenčním médiem, má kromě hlavní části, tj. slovníkové stati, obsahovat také:

- A) návod, jak slovník používat
- B) informaci o organizaci hesel a struktuře slovníku
- C) vysvětlení použitých zkratk
- D) vysvětlení metajazyka
- E) stručný přehled potřebné gramatiky

Dodejme ještě, že tyto údaje se mohou objevovat buď v úvodu, nebo na konci, popř. přímo na deskách slovníku.

Pro uživatele slovníku je makrostruktura velmi důležitá, neboť udává celkovou organizaci, a tím i přehlednost slovníku, která je zásadní pro rychlou orientaci a přesné vyhledávání hesel. Slovník s dobře zpracovanou mikrostrukturou je přehledný a umožňuje nám tak rychle a spolehlivě nacházet.

2.2.1. Úvod a přílohy

Úvodní částí slovníku rozumíme titulní strany, předmluvu, kde bývá specifikován účel publikace, dále základní informace o obsahu a struktuře slovníku, jakož i uživatelé,

³⁶ Čermák, Blatná 1995 : 236

kterým je slovník primárně určen. V předmluvě také bývá představen autor či autorský kolektiv, lektori apod. V úvodu by též měly být zmíněny všeobecné informace o slovníku a použitých pramenech. Dále by v úvodní části měla být popsána struktura hesel a podány instrukce pro používání slovníku, tj. vysvětlení použitých symbolů a zkratek. V některých slovnících je uveden i rejstřík. Další složkou úvodu může být gramatická část s tabulkami se systémem sloves, popř. i další slovní druhy. Tato gramatická část bývá často umístěna až ve slovníkové části, jako předěl mezi dvěma oddíly, v některých případech můžeme tuto část nalézt i na konci slovníku. Jako další bývá v úvodu obsažena abeceda cílového jazyka, jakož i poznámky k výslovnosti, které zpravidla bývají doprovázené příklady slov.

Přílohy slovníků mohou potom obsahovat přehled gramatiky, není-li již součástí úvodu, dále se v této části slovníku může objevit konverzační příloha a fráze, různé tabulky, obrázkové přílohy, popř. i reálie země, kde je cílový jazyk používán.

2.2.2. Slovníková stať

Další část slovníku nazýváme slovníková stať. Bezesporu se jedná o nejdůležitější část slovníku. Slovníková stať bývá v překladových slovnících rozdělena na dvě části, je-li slovník oboustranný. Někdy se uvádí, že slovníková stať se dělí na tolik částí, kolik je písmen v abecedě³⁷. V následujících kapitolách si představíme některé aspekty z problematiky zpracování slovníku, a sice výběr hesel a jejich řazení.

2.2.2.1. Výběr hesel do slovníku

Otázka výběru hesel (lemmat) je při sestavování slovníku jednou ze základních, ale zároveň i nejobtížnějších, neboť sestavovat slovníky přináší mnoho otázek. Co bude potenciální uživatel ve slovníku hledat? Která hesla lze naopak vynechat? Z dat nashromážděných během průzkumu je proto třeba vybrat jen ta, která nejlépe odpovídají stanoveným požadavkům. K tomuto úkolu by měl tvůrce slovníku přistupovat pečlivě a držet se alespoň čtyř kritérií, která nyní postupně zmíníme.

³⁷ Např. Martínez de Sousa in Brabcová : 2012 : 26

Tři z těchto kritérií, kterými jsou záměr slovníku, skupina uživatelů a rozsah slovníku, můžeme souhrnně označit jako kritéria vnější. Čtvrté pravidlo, které zahrnuje jazykové principy výběru, bychom klasifikovali jako vnitřní kritérium. Do tohoto pravidla zahrnujeme frekvenci použití, ale zároveň i objektivitu, aby byla vybraná slova pro daný jazyk reprezentativní.³⁸

Také J. Filipec se domnívá, že hlavním kritériem při výběru slov je typ slovníku a potřeba uživatele. Proto by hesla měla být volena vždy tak, aby odpovídala tematické náplni slovníku. „Problematika cílového uživatele má aspekt pragmatický i sociolingvistický. Sociolingvistické hledisko se týká makrostruktury i mikrostruktury slovníku. Jeho uživatelem může být celá národní společnost, vzdělané vrstvy, mládež školou povinná. V prvním případě jde o to, aby slovník obsahoval hesla společensky relevantní, klíčová, odpovídající společenské diferenciaci. Týká se to např. politických ekonomických, ekologických termínů, které se stávají běžnými, internacionalismů, professionalismů, perspektivních neologismů.“³⁹ Zahrnování odborných názvů, internacionalizmů a geografických názvů či vlastních jmen, čili proprií, se kterými se běžně v životě setkáváme, souvisí i s výběrem slov podle frekvence. Zde však J. Šabršula upozorňuje, že „v případě odborných názvů se může stát, že kvůli rozdílům v sociálně politickém vývoji různých jazyků, není nalezen příslušný ekvivalent.“⁴⁰

Filipec dodává, že „otázkou je i zřetel ke společenskému tabu (výběr slov jako např. *erekce, vložka, bordel*), vulgarismům (*prdel, řít, hovno*) typickým pro uvolněný jazyk.“⁴¹ Konkrétně v případě kapesních slovníků autoři často kladou důraz na užití slov běžně užívaných v každodenních situacích, hovorové či dokonce slangové výrazy, proto se podobné výrazy v těchto typech slovníků dají očekávat.

Dalším, neméně důležitým bodem při výběru hesel, je kritérium frekvenční, tj. výběr slov podle frekvence užití. „Frekvence je vědomě nadřazena stylu. Frekvence svým způsobem podmiňuje vše ostatní a nelze absolutizovat něco proto, že je to spisovné, a přitom třeba okrajové, před tím, co je mluvené a patří k nejfrekventovanějším slovům.“⁴²

³⁸ Haensch in Brabcová : 2012 : 27

³⁹ Čermák, Blatná, 1995 : 18

⁴⁰ Šabršula, 1989 : 241

⁴¹ Čermák, Blatná, 1995 : 22

⁴² Tamtéž : 99

K tématu frekvence Čermák dodává, že lexikograf by neměl vybírat slova, která jsou používána příležitostně, ani slova dobově omezená, ale měl by volit slova s vyšší frekvencí užití, objektivní a současná. Autor dále zdůrazňuje potřebu vyváženého čerpání z celého jazyka a s tím spojenou nutnost zrušení restrikce dané spisovností. Klade proto důraz na užití rovnoprávně psané i mluvené složky slovní zásoby jazyka odrážejícího skutečný úzus.

Z výše uvedeného vyplývá, že při sestavování slovníku je důležitý autor, který by měl z ovládat nejen slovní zásobu obou jazyků, ale také realie zemí, odkud daný jazyk pochází. Je třeba mít na paměti, že jazykové systémy se mohou značně lišit, což se promítá i do struktury slovní zásoby. Protože však slovník nemůže obsahovat naprosto všechny lexikální jednotky a výběr slov při jeho přípravě podléhá řadě kritérií, doporučuje F. Čermák dodržovat tyto zásady⁴³:

1) Lexémy je třeba vybírat podle **frekvence**. I tu je však třeba korigovat a složení hesel obsahově kontrolovat, aby byla slovníková hesla reprezentativní a vyvážená.

2) Zásadní je i **nominativní reprezentace**. Je žádoucí zachovat úroveň nominativní reprezentativnosti. To znamená, že ve slovníku nesmí být velké mezery v pojmech, i když jsou některé pojmy používány okrajově a s nižší frekvencí. Ani v malém slovníku nemají být pojmové mezery v tom smyslu, že tu budou prostě a nezodpovědně vynechány (např. při redukci ze slovníku většího) některé neokrajové pojmové okruhy. K tomu však může dojít i jen slepým spoléháním pouze na frekvenci, kdy některá základné pojmenování (a pojmy) mohou pro svou relativní řídkost vypadnout, jakkoli jsou třeba v dané kategorii pojmenování citu, rostliny či materiálu apod. základní (často jde o hyperonymum).

3) Zejména v případě malých slovníků, kde je třeba hesla vybírat úsporně, by lexikograf měl z třídy synonym volit lexémy **bezpříznakové** proti těm příznakovým. „Bepříznakový lexém je neutrální a nemá žádné citové zbarvení. Takto je bezpříznakové např. *malý, červený* proti *drobný, rudý*. Totéž hledisko, i když na jiné

⁴³ Čermák, Blatná, 1995 : 232-234

úrovni platí i v dalších případech variantnosti, ať už jde o regionalismy či tvary pro zastoupení morfologie.“

4) Slovník by měl svým obsahem **motivovat** uživatele k **růstu** jeho **slovní zásoby**. Měla by se tu objevit slova s kořenem, který má min. 3-5 derivátů. I když se uživatel s těmito deriváty setká zřejmě až později, takovýto kořen, který je k nim pak klíčem, by měl být ve své základní formě zachycen už v poměrně malém slovníku.

„Lexikografická dokumentace neboli excerpce je tudíž důležitým krokem při koncipování slovníku a sestává se ze sbírání reprezentativního materiálu jazyka. Zdroje pro sběr tohoto materiálu jsou klasická i odborná literatura, může jím být ale i tisk a mnoho dalších, důležitá je již zmiňovaná reprezentativnost vybraného materiálu. Jelikož je metoda excerpce velmi zdlouhavá, využívá se dnes spíše paralelních korpusů, tedy databází textů ve dvou či více různých jazycích, které jsou si navzájem obsahově i formálně podobné. Výhodou využívání těchto korpusů je nejen značná úspora času, ale také možnost porovnání velmi rozsáhlého souboru hesel.“⁴⁴

2.2.2.2. Řazení slov ve slovnících

Uspořádání hesel (heslových statí) překladového slovníku bývá nejčastěji alfabeticke. Jedná se zřejmě o nejběžnějším způsobem řazení, které má své výhody i nevýhody. Mezi výhody lze uvést, že dané heslo velmi snadno najdeme, jelikož má pevně dané pořadí. V oblastech, kde je rozvíjející člen až na druhém místě (např. chemie), se velmi dobře tvoří hnízdování. Nevýhodou je, že se za sebe dostanou termíny, které spolu vůbec nesouvisí, ale jen začínají na stejná písmena, což dokazuje následující příklad, v němž se míchají termíny týkající se psů, telat a dogmat⁴⁵:

<p><i>dog – pes, psovitá šelma</i> <i>dogie - tele bez matky, zatoulané tele, podřadné dobytče</i> <i>dogma – dogma</i> <i>dogman - milovník psů, znalec psů, kynolog</i> <i>dogmata - článek víry</i> <i>dogmatican - dogmatik</i> <i>dogmatician – dogmatik</i></p>
--

⁴⁴ Pfliegerová, 2013 : 29

⁴⁵ Řazení slov ve slovnících, JTPUnion.org [online]

Alfabetické řazení je založeno na pevném pořadí písmen abecedy a lze rozlišit dva typy, a sice řazení přímé a řazení retrográdní, kdy jsou hesla uspořádána podle abecedy, ale odzadu. „Podstata retrográdního slovníku záleží v tom, že se slova, obsažená v jistém slovníku nebo v jistém archívním materiálu, seřadí sice abecedně, ale postupuje se od konce slov. Tím vznikne index slov, který nazýváme retrográdním slovníkem. (Měli bychom tedy spíše mluvit o retrográdním indexu slov.)“⁴⁶

Abecední uspořádání hesel může být z praktických důvodů modifikováno. Vedle abecední organizace se tak někdy i přes své nedostatky užívá metoda tzv. hnízdování. Filipec jej definuje jako „sružování slovotvorně souvisejících lemmat a sublemmat, lexémů“⁴⁷. Hnízdování lze také charakterizovat jako seskupování odvozenin v heslové stati, kdy jsou příbuzná hesla uspořádána do skupin, což přináší úsporu místa ve slovníku. Hnízdování se používá, pokud máme více skupin slov se stejným základem. Slova rozřadíme do skupin a v rámci těchto skupin pak můžeme použít jiný typ řazení, např. alfabetické, které bylo vysvětleno výše. Následující příklad ukazuje, jak takové alfabetické hnízdování může vypadat. Heslo bylo vybráno z *Francouzsko-českého, česko-francouzského slovníku* vydavatelství FRAGMENT,⁴⁸ na který se blíže zaměříme ve druhé části této práce.

<p>mluv a langue ž, langage m; -čí porte-parole* m; -ení bavardage m, parler m; -it 1 (s kým o kom/ čem) parler (à avec q de q/ qc) (-it francouzsky parler français) 2 co říkat dire*⁴² qc -ka bavard, e m, ž, causeu r, -se, m, ž, -nice</p>
--

I hnízdování má své výhody i nevýhody. K výhodám patří informace o významových vztazích, nevýhoda může spočívat ve zvětšování hesláře odkazy a také v tom, že se do těchto hnízd mohou občas dostat i hesla nepříbuzná, čímž je narušena věcná či abecední souvislost. Takto hnízdovaná slova tak bývají pro nepoučené uživatele ztracena. Není tedy příliš vhodné užívat tuto metodu v malých kapesních slovnících, které bývají často

⁴⁶ Více k tématu retrográdní řazení viz Štindlová, J. : Retrográdní slovníky, SaS

⁴⁷ Čermák, Blatná, 1995 : 26

⁴⁸ Velký kapesní francouzsko český-česko-francouzský slovník. KPS-FRAGMENT : 2007

určeny začátečníkům, studentům základních či středních škol, či občasným uživatelům, neboť u se těchto nepředpokládá hlubší jazyková znalost a zkušenost.

Dalším případem jsou slovníčky, kde jsou slova seřazena podle pořadí, v jakém se vyskytují v textu. Tento způsob se vyskytuje zejména v glosářích. Princip funguje tak, že slovo, které se v textu vyskytuje nejdříve, je ve slovníku řazeno jako první, bez ohledu na abecední řazení nebo jeho význam. Tento druh slovníků najdeme například na konci časopisových článků, v učebnicích apod.

Dále můžeme jmenovat řazení etymologické, které spočívá v seskupování hesel majících společný kořen. Zmínit lze taktéž uspořádání, resp. subuspořádání statistické, kde jsou ekvivalenty v rámci hesla řazeny podle frekvence jejich výskytu. I tento druh řazení má své stinné a světlé stránky. Jako pozitivum lze vidět např. to, že uživatel sice na prvním místě najde ekvivalent, který je nejpoužívanější, to však nutně neznamená, že je to právě to slovo, které se hodí do daného kontextu. U řazení abecedního zase hrozí, že jako první uvidí uživatel slovo, jehož užití je vzácné a naopak nejběžnější slovo se ocitne až na posledním místě. Vhodným řešením by tedy mohlo být řazení, které vychází z frekvence, avšak autor na tuto skutečnost v odpovídající části slovníku upozorní a tento výběr vysvětlí, což by ostatně měl být běžný postup ve všech slovnících.

Ačkoli může být způsob řazení hesel v jednotlivých slovnících řešen různě, existují určitá hlediska, kterými se autoři slovníků obvykle řídí. Těmito hledisky jsou především⁴⁹:

- 1) celková přehlednost
- 2) uspořádání hesel podle významu slov
- 3) postupování od významu nejběžnějšího k významům souvisejícím
- 4) ohled na uživatele, pro něž je slovník určen

⁴⁹ Faltová, 2007 : 19

2.3. Mikrostruktura

V této kapitole si vysvětlíme pojem mikrostruktury slovníku a její jednotlivé složky. V dalších podkapitolách se potom budeme věnovat každé části, které mikrostrukturu tvoří.

Pod pojmem mikrostruktura slovníku chápeme formu a uspořádání informací uvnitř hesla.

Filipec mikrostrukturu popisuje jako „lingvisticky relevantní soubor informací o heslovém slově (vstupu). Je tedy jednotou langue a parole, jednotou paradigmaticko-syntagmatickou. V mikrostrukturu se sdružují a prostupují roviny fonologická, gramatická (morfologická a syntaktická), sémantická, pragmatická, stylistická a textová, popř. i kvantitativní (statistická, frekvenční), historická (etymologický údaj) a rovina mezilexémových vztahů.“⁵⁰ Filipec dále dodává, že „údaje o heslech jsou řazeny v ustáleném pořadí a ekonomicky s využitím zkratk a zvýrazňujících typografických znaků, jako jsou závorky, interpunkce, značky, odlišující číslice, různá velikost a typ a styl písma apod.“⁵¹ Jde tedy o „určitou skupinu rubrik, vztahujících se k danému pojmu, jež tento pojem dále doplňují, rozvíjejí, podle přesně stanovených strukturálních pravidel daného slovníku.“⁵²

F. Čermák popisuje schéma slovníkového hesla následovně⁵³:

Synchronní složky:

1. paradigmatické složky

- a) forma – grafická a akustická forma lexému
- b) funkce – označení třídivé příslušnosti, především slovních druhů
- c) frekvence a styl – frekvence by měla být uvedena bezprostředně za a) a i)
- d) sémantika – musí být komplementární ze sémantikou kontextu
- e) hyponymie – zachycuje přímo nadřazený pojem
- f) synonymie
- g) opozitnost
- h) úzus – doplněk hesla

⁵⁰ Čermák, Blatná, 1995 : 23

⁵¹ Čermák, Blatná, 1995 : 23

⁵² Kapusta, 2008 : 21

⁵³ Kapusta, 2008 : 112

2. syntagmatické složky

- i) forma (valence)
- j) funkce
- k) styl
- l) sémantika kontextu
- m) exemplifikace
- n) technický odkaz

Diachronní složka:

- o) etymologie – může a nemusí být obsažena

2.3.1. Typy hesel

Hesla lze dělit podle počtu lexikálních jednotek, a to na jednoslovná (obsahující pouze jeden lexém) a víceslovná hesla. Mezi víceslovná hesla bychom mohli zařadit i idiomy, frazémy a terminologická spojení. F. Čermák v souvislosti s typy hesel zmiňuje, že v hesláři by měly být minimálně tyto typy hesel⁵⁴:

- a) **Standardní hesla**, je však třeba respektovat zásadní rozdíl mezi synsémantiky a autosémantiky co do jejich kvantitativního zastoupení (synsémantika, tj. prepozice, partikule, konjunkce, a interjekce, pronomina a numerale budou zastoupena téměř vyčerpávajícím způsobem).
- b) **Vydělované části slov či morfémy** (např. slovesné prefixy)
- c) Odpovídající výběr **značek a zkratk současných a podle potřeby i minulých** (ČR, oÚ, VB, ČSFR)
- d) Odpovídající výběr **geografických popř. jiných propriálních** názvů a to i cizích – měl by se řídit frekvencí výskytu a potřebou identifikovat cizojazyčnou formou známého nebo informovat i průvodních konotacích.

⁵⁴ Čermák, Blatná, 1995 : 235-236

Čermák dále odlišuje i tzv. dublety, neboli dvojtvary, což jsou varianty na úrovni lexému, srov. čes. *mušet* : *musit*; *museum* : *muzeum*; *hřmít* : *hřmět* nebo angl. *speciality* : *specialty*. Dalším typem mohou být hesla, resp. formy čistě odkazové (technické). Takovým příkladem jsou různé nepravidelné tvary lexémů a supletivní formy některých slov. S tímto souvisí i supletivismus, jež lze charakterizovat jako tvary téhož slova, pro něž používáme tvary odvozené od jiného kořene, než je slovo výchozí. F. Čermák definuje supletivismus jako „jev alomorfie založené na etymologicky nepřibuzných základech.“⁵⁵ Ne vždy se však tyto nepravidelnosti ve slovnících objevují. Jedná se např. o substantiva *člověk – lidé*, *rok – let*, pronomina *já – mě*, verba *jde – šel*, či stupňovaná řada adjektiv *dobrý – lepší – nejlepší*. Tento typ hesel je užitečným klíčem k dešifraci textu zvl. pro cizince. Čermák rovněž poukazuje i na monokolokabilní, tj. nesamostatně existující komponenty frazémů (např. *holičky : nechat na holičkách*).⁵⁶

2.3.2. Výstavba hesla

Klíčovým bodem v otázce sestavování slovníků je výstavba jeho jednotlivých hesel. Pokud jde o výstavbu samotného hesla, i ta by měla být přehledná a snadná pro orientaci, tak jako výstavba celého slovníku. Jak ve své práci poznamenává Faltová „je však zřejmé, že požadavky na výstavbu hesel malého, středního a velkého slovníku se budou lišit.“⁵⁷ „Velký akademický slovník má strukturu mnohem složitější a faktor přehlednosti zvláště velkých hesel (i vícestránkových) se dostává do popředí. Vlastní náplň, a tedy i počet formálních součástí hesla, je tudíž dán pragmaticky, souvisí ovšem jak s typem jazyka, tak slovníku. Povaha či absence některých složek však jasně souvisí s nedostatkem hlubšího lingvistického vhledu, zvl. u části valenční.“⁵⁸

Nejdůležitější roli při výběru lexikálního materiálu hraje počáteční výběr forem. Heslář, který dá slovníku základní podobu, orientaci i rozsah, je závislý na explicitně vytyčených cílech, typu uživatele a jeho jazykových znalostech. Povaha hesláře musí být přitom zdůvodněna kvantitativně i kvalitativně, tj. rozsahem i složením. Z tohoto výběru vychází jak makrostruktura, tak mikrostruktura slovníku. Jde-li o slovník malého typu, je třeba výsledky analýzy podříditi zásadě úspornosti a musí být uváděny jen informace

⁵⁵ Čermák, 2011 : 214

⁵⁶ Čermák 2011 : 214

⁵⁷ Faltová, 2007 : 16

⁵⁸ Čermák, Blatná, 1995 : 240

nejnutnější. Heslo by mělo být vystavěno tak, aby zahrnovalo všechny důležité informace o příslušné lexikální jednotce na úrovni jednotlivých jazykových rovin. Slovník by tak měl spolehlivě poskytovat alespoň 5 základních informací, a to: fonologicko-fonetickou, morfologickou, syntaktickou, sémantickou a pragmaticko-encyklopedickou. Úhrn hesel pak vytváří heslový korpus, čili heslář.⁵⁹

2.3.3. Struktura a složky hesla

Z hlediska struktury má slovníkové heslo tyto dvě základní části:

- a) část první, které nazýváme záhlaví
- b) část druhá, představující hlavní část hesla, kde je popisován význam lexikální jednotky v různých aspektech

Stálým prvkem mikrostruktury je pravopisný údaj, který může zahrnovat i případné varianty grafické podoby daného hesla. Obecně lze říci, že mezi základní složky hesla patří: lemma, případně označení homonymie; transkripce, která bývá tradičně uváděna v hranatých závorkách. Klíčovou částí je překlad významu (resp. výklad v případě jednojazyčných výkladových slovníků). Dále se uvádí gramatická charakteristika ve formě slovnědruhového zařazení, popř. další morfologické údaje a valence. Mohou být zařazeny i údaje o původu hesla. Další složkou je stylisticko-pragmatická charakteristika, tj. obor užití lexikální jednotky. Uvádí se též příklady užití neboli exemplifikace, či citace tj. uvedení slova v typickém kontextu. Někdy bývá lexém začleněn do synonymních a opozitních vztahů. V závěru se může objevit odkaz na podobná hesla, popř. gramatickou, či konverzační část nebo obrázkovou přílohu. Součástí heslové stati mohou být i víceslovné lexikální jednotky, fráze, idiomy a ustálená spojení, popř. další jednotky, využívá-li slovník systém hnízdování. Tyto složky jsou nedílnou součástí slovníkového hesla. Jejich cílem je ilustrovat a konkretizovat význam slova a doplňovat a zpřesňovat jeho významové odstíny⁶⁰.

⁵⁹ Čermák, Blatná, 1995 : 24

⁶⁰ Faltová, 2007 : 16

2.3.3.1. Lemma neboli slovníkový vstup

Filipec⁶¹ uvádí, že ve většině případů jsou jednotlivá heslová slova zachycena pouze ve svém základním tvaru, tzv. lemmatu. Termín lemma, který se užívá zvláště v počítačové terminologii, je základním stavebním kamenem každého slovníku. Jedná se o reprezentativní tvar lexému, které uvádí heslovou stat'. Reprezentativní tvar heslových slov se uvádí v nominativu singuláru, u sloves v infinitivu ve specifických tvarech a u sublemmat může jít i o jiné tvary (neosobní slovesa, pomnožná substantiva aj.) Filipec dále dodává, že „vedle heslových slov se uvádějí v abecedním pořadí i slovtvorné morfémy a části slov (složenin) např. předpony *na-*, *ne-*. Uvádějí se také části složenin, konfixy, a to první (*jedno-*, *mikro-*, *tělo-*), druhé (*-fil*, *-log*) nebo dokonce i vnitřní (*-půl-*). Hesly mohou být i ustálená spojení slov, lexikalizovaná syntagmata a frazémy, uváděné pod klíčovým slovem spojení (*lelky : chytat lelky*). V lemmatu by dále měl být znázorněn i tzv. morfematický šev, a to v místě, kde se slovo ohýbá. Homonymní výrazy (hl. homografa) bývají obvykle odlišována číslicemi; u polyfunkčních slov je pak důležité dodávat i funkční zkratky (zvl. slovnědruhové) srov. *co*: 1 pron, 2 konj, 3 interj; *místo* 1 subst 2 adv.“⁶²

Vedle formy grafické by měla být součástí záhlaví i forma akustická, čili transkripce. Ta by měla respektovat prostředky Mezinárodní fonetické abecedy (I.P.A. resp. A.P.I.) včetně přízvuku. Jiné, často volnější, zjednodušené formy fonetické transkripce, jsou méně přesné, u menšího slovníku však mohou mít se zřetelem k laickému uživateli své oprávnění.

Další složkou záhlaví by mělo být stylové označení pomocí zkratk jednotných v celém slovníku (např. *anat.*, *bot.*, *jur.*, *med.*, aj.). Tato stylová diferenciacie, tj. odlišování jednotek příznakových a nepříznakových, je systémové zařazení lexikálních jednotek. Označením lexikální jednotky zkratkou tedy signalizujeme její příslušnost k odpovídajícímu dílčímu systému, jenž má vlastní specifickou soustavu pojmů.⁶³

⁶¹ Čermák, Blatná, 1995 : 21

⁶² Čermák, Blatná, 1995 : 241

⁶³ Čermák, Blatná, 1995 : 240

2.3.3.2. Gramatická charakteristika

Gramatická charakteristika nese informace o slovním druhu daného hesla. Obvykle je uváděna pomocí mezinárodní latinské zkratky. Tato informace „musí pokrývat všechny gramatické kategorie, musí být podána soustavně, přehledně a úsporně především v tom smyslu, že je jednoznačným klíčem k celé paradigmatické (nejen k morfologii). Označení paradigmatické informace je obecně nutné vždy, když nelze potřebnou informaci jednoznačně zjistit z formy lemmatu, což ani pro češtinu není většinou možné, srov. jen zdánlivou ‚stoprocentní‘ pravidelnost zakončení adjektiv na *-ný*, *-ský*, kontrastující se substantivy *vrátný*, *hostinský*. V některých případech je výhodnější i úspornější uvést subkategorizaci, např. u českých substantiv stačí zkratka rodu a (ne)životnosti *m/an*, *m/in*.“⁶⁴

Čermák se domnívá, že „dobrý slovník se mimo jiné pozná podle toho, že by měl fungovat jako základní příručka a autor by tedy rozhodně neměl spoléhat na to, že bude slovník užíván ve spojení s jinými příručkami a uživatel bude muset dohledávat upřesnění jinde. Aby se tomuto předešlo, je možnost vložit do překladového slovníku určitý gramatický dodatek a odkazovat na něj.“⁶⁵

2.3.3.3. Ekvivalenty

Dle F. Čermáka lze ekvivalenci definovat jako „vztah mezi určitým lexémem výchozího jazyka a odpovídající formou cílového jazyka, založený na jejich (téměř) identickém, resp. velmi podobném významu a funkci. Členy, tj. koncové termíny tohoto vztahu, se nazývají ekvivalenty. Jedná se centrální část slovníkového hesla (stati), která má poskytovat k lemmatu, resp. k určitému významu lemmatu, optimální ekvivalenty, a to jak významové, tak funkční.“⁶⁶ Autor nicméně dodává, že nelze vyžadovat, aby byly ekvivalenty v promluvě zcela zaměnitelné. „Míra explicitnosti, a tedy i exaktnosti, je závislá i na typu slovníku (ve slovnících malého typu je menší). Pro upřesnění kontextu bývají ekvivalenty často doprovázeny také glosou, která je od lemmatu odlišena jiným typem písma, popř. závorkou. Glosa může stát buď před, nebo za slovem. V případě více ekvivalentů by se při jejich řazení mělo vycházet od obecného ke specifickému a od více

⁶⁴ Čermák, Blatná, 1995 : 241

⁶⁵ Čermák, Blatná, 1995 : 241

⁶⁶ Čermák, Blatná, 1995 : 238

k méně frekventovanému, řídkému. Jejich případná hierarchizace může být značena číslováním (obvyklé u polysémie či polyfunkčnosti). Rozdíly u částečné ekvivalence jakéhokoli druhu, tj. kognitivní, denotativní, pragmatické, kulturní, stylové, je třeba co nejdůležitěji, ale zároveň co nejstručněji a jednotným způsobem glosovat, a to třeba i zápornou formou.⁶⁷ Kromě glosy, ekvivalentu a gramatické kategorie, lze význam slova upřesnit např. i parafrází (synonymum, antonymum, hyperonymum, hyponymum) nebo označením oboru či stylu pomocí zkratk, o kterých již v této práci byla řeč.

„Ekvivalent slovníkového hesla by však neměl být pouhou cizojazyčnou definicí, měl by také zohledňovat drobné významové nuance, které jsou způsobeny například použitím slova v ustálených spojeních.“⁶⁸ V jazycích se totiž kromě výrazů, jejichž překladový význam lze jednoznačně a správně určit, vyskytuje i celá řada výrazů, ustálených spojení a idiomů, ke kterým lze jen velmi těžko hledat přesný ekvivalent. Tyto je nutné překládat přenesením významu do systému, struktury a podstaty vnímání jazyka cílového.⁶⁹

F. Čermák se k tomuto problému vyjadřuje následovně: „Teoreticky lze lišit čtyři, resp. tři obecné případy ekvivalence lexémů; první sloupec představuje lexémy např. češtiny, druhý sloupec pak ekvivalenty v jiném jazyce (=situace česko-cizojazyčného slovníku)“⁷⁰:

A	1 : 1	(jednoznačná ekvivalence, izomorfismus)
B	1 : více	(částečná ekvivalence, anizomorfismus)
C	více : více	(částečná ekvivalence, anizomorfismus)
D	1 : 0	(absence lexémů a tedy i ekvivalence; lakuny)

K tomuto schématu pak Čermák doplňuje: „Ekvivalence by měla být především sémanticko-funkční povahy a měly by být založena na aproximativní identitě úzu. Prakticky je třeba počítat alespoň s pěti typy ekvivalence:

- 1) **jednoznačná ekvivalence**: lexém A odpovídá lexému B, případ poměrně ojedinelý

⁶⁷ Tamtéž : 242

⁶⁸ Pfliegerová, 2013 : 25

⁶⁹ Pfliegerová, 2013 : 25

⁷⁰ Čermák, Blatná, 1995 : 238

- 2) **opis**: ekvivalent chybí, lexém je nutno vysvětlit opisem
- 3) **částečná ekvivalence**: škála mezi lexémy A a B, lze z hlediska distribuce rozlišit ještě další podskupiny
 - a) vztah **širší a užší** a naopak (např. angl. *cut* a čes. *stříhat, krájet, sekat*.
Nebo naopak čes. *ržát* a angl. *neigh, whinny*)
 - b) vztah **obecné a specifické** a naopak (např. čes. *babička* a švéd. *farmor, momor* (babička z otcovy a matčiny strany) čes. *prst* a angl. *finger, toe* (prst u nohy a ruky)
 - c) u rovnocenných pojmů může jít také jen o **částečný průnik**, tj. situace, kdy se výchozí lexém a ekvivalent překrývají jen zčásti.
- 4) **ekvivalent vázaný**: vázaný na pevnou kombinaci tvarů, jde především o frazémy a idiomy
- 5) **ekvivalent funkčně neparalelní**; specifický případ částečné ekvivalence, např. české *líbit se* vůči angl. *like*, kde jsou odlišné subjekty: *I like it* a *Líbí se mi to*. Obdobně sloveso *manquer* ve francouzštině. Dalším případ: verbum aktivní-verbum pasivní, slovo-věta aj.

Ekvivalence je ovšem často komplikována aspekty kulturními, sociálními (včetně stylových) a pragmatickými. V zásadě by však ekvivalent měl optimálně podchytit maximum shodných potřebných informací sám o sobě, tj. bez glosy (tj. i syntakticko-funkční paralelismus, stylovou obdobu aj.)⁷¹

Z hlediska uživatele a jeho práce se slovníkem je možné jmenovat dva typy ekvivalentů – exemplifikační a dosazovací. V prvním případě jsou ekvivalenty založeny na uvedení v příkladech či spojení. V případě ekvivalentů dosazovacích je u slovníkového hesla uvedena řada synonym, ze kterých si uživatel na základě svých jazykových znalostí vybere ten nejadekvátnější.⁷² K tomu se však F. Havlová ve svém článku *Kapesní slovník cizích slov* vyjadřuje následovně: „Nelze uspokojivě vystihovati význam slov pomocí synonym; ani synonyma se skoro nikdy významově nekryjí. Tak např. význam slov *hojnost* (*hojný, hojně*), nebyl by úplně vystižen zdánlivými synonymy *mnohost* (*mnohý, mnoho*), *četnost* (*četný, četně*), neboť tato synonyma mají sice přibližně také význam ‚víc

⁷¹ Čermák, Blatná, 1995 : 238-240

⁷² Pfliegerová, 2013 : 28

než *dostačující množství*, ale schází jim další podstatný znak skupiny hojnost atd., totiž označení hojnosti nějakým způsobem žádoucí, takže hojnost znamená *‚více než postačující množství něčeho žádoucího‘*, tedy na př. *hojná úroda, hojné jídlo, hojný užitek* aj., nikoli však na př. *hojnost komárů, lupičů, hojné pohromy, hojná úmrtí* aj. Navíc zjišťujeme, že pod hlavičkou ‚synonyma‘ uvádějí lexikografické definice pouze ‚kvazisynonyma‘, u nichž se ve skutečnosti nepočítá s obráceným vztahem, tzn. že se synonymie ve slovnících neuplatňuje jako symetrický vztah, ale jako vztah s jedním preferenčním směrem.“⁷³

F. Čermák se k této problematice synonymie vyjadřuje takto: „Synonymie je jev v zásadě náhodné podobnosti významu dvou nebo více různých lexémů, tj. majících různou formu; je jedním z hlavních zdrojů jazykové stylistiky. Srov. např. angl. *buy: purchase* nebo čes. *těžký : obtížný*. Kritéria bližšího lišení typů synonym jsou velmi různá a málokdy zcela spolehlivá; hlavní problém je v určení stupně, resp. úplnosti podobnosti.“⁷⁴ Kromě absolutní (někdy též totální) synonymie zmiňuje autor také tautonyma, tj. synonyma z různých variet resp. jazyků, srov. čes. a moravské *vesnice : dědina*.

Oba výše zmiňované typy ekvivalentů, tj. exemplifikační a dosazovací, mají své výhody, jsou však v zásadě závislé na jazykových zkušenostech a citu uživatele. „Každému heslu výchozího jazyka zpravidla odpovídá několik ekvivalentů jazyka cílového, pro uživatele je tedy důležité, aby vždy pozorně sledoval kontext, do kterého je výraz zasazen, a podle něj správně zvolil překladový ekvivalent.“⁷⁵ Důležitou roli sehrávají v otázce ekvivalence obousměrné slovníky. „Pokud budeme studovat oboustranný překladový slovník, nelze se domnívat, že hesla v obou publikacích budou pouze mechanicky převrácena a abecedně seřazena. Překladové ekvivalenty se totiž mohou značně lišit v závislosti na tom, zda je konkrétní heslo v pozici jazyka cílového nebo výchozího. Pro zkušenějšího uživatele jazyka jsou pak tyto oboustranné slovníky velmi důležité, jelikož umožňují ověření adekvátnosti překládaného ekvivalentu.“⁷⁶

Pro shrnutí tedy lze říct, že lexikograf by rozhodně měl mít na paměti, že ekvivalenty ve slovníku mají co nejlépe korespondovat s výchozími hesly nejen po sémantické stránce,

⁷³ Havlová, 1995, Naše řeč

⁷⁴ Čermák, 2011: 237

⁷⁵ Pfliegerová, 2013 : 28

⁷⁶ Tamtéž

ale i vzhledem ke konotaci slov a pragmatické hodnotě.

2.3.3.4. Valence

Filipec popisuje valenci jako „schopnost dominujícího, klíčového lexému vyžadující jistý počet větných pozic (valenčních členů).“⁷⁷ Pod pojmem valence ve slovníku si tedy můžeme představit schopnost lexému na sebe vázat další větné členy. Čermák se domnívá, že „valenci je třeba uvádět u každého hesla, výjimkou jsou snad pouze citoslovce. Za valenci je třeba považovat všechny u daného lexému systémově predeterminované třídy a kategorie, signalizované obvykle přítomností některého formálního exponentu (pád, prepozice, konjunkce, adverbium, infinitiv, ale třeba i komparativ aj.), které spolu s ním musejí, popř. mohou vyskytnout. Obě valence (českou i cizojazyčnou) je velmi výhodné uvádět blízko nebo vedle sebe, srov. angl. *look, have a look (na něco at sth., do něčeho into sth.)*“⁷⁸ Filipec k tomu dodává ještě následující: „Je-li výklad významu, sémému klíčového slova v kontextu soběstačný, nevyžaduje valenční, a tedy externí doplnění. Např. ve spojeních *sláma chytla ‘začala hořet’, barva chytla ‘ulpěla na povrchu’*, jde tedy o významově soběstačné, smysluplné výpovědi, kdežto ve spojeních **kamarád chytl* nikoli. Tu je potřebné valenční doplnění, např. *Kamarád mě chytl rukou/za ruku*. Tyto případy dokládají souhru významu, valence a kontextu.“⁷⁹

2.3.3.5. Exemplifikace

Exemplifikace, tedy uvádění příkladů, citace, ale i případné začlenění synonym a antonym do této části hesla, vkládá lemma do jeho běžného kontextu, čímž ilustruje systémovou spojitelnost lexému s jinými lexémy. Exemplifikace stojí ve druhé části slovníkového hesla a tvoří lexikálně sémantický obsah slova tak, jak je užíván v různých situacích a kontextech. Cílem exemplifikace je popsat jednotlivé významy, příslušná jazyková spojení a terminologická užití.⁸⁰

„U synsémantických slov by měla být exemplifikace uvedena vždy, na rozdíl od slovnovýznamových, kde je ilustrace otázkou výběru, neboť jednoznačná ekvivalence zvl.

⁷⁷ Čermák, Blatná, 1995 : 37

⁷⁸ Tamtéž : 243

⁷⁹ Čermák, Blatná, 1995 : 37

⁸⁰ Faltová, 2011 : 18

termínů ji nepotřebuje. Uvedení příkladů je naopak nezbytné v případě mnohoznačných slov. Délka příkladu sahá od kontextu větného (popř. vícevětného) až ke kombinaci slov. Je pochopitelné, že exemplifikace neboli ilustrace v zásadě nepotřebuje překlad. Exemplifikace, neboli zasazení lemmatu do jeho přirozeného kontextu, jednak konkretizuje strukturní informaci z jiných složek hesla (valenci, funkci, ale i část sémantiky), a jednak ilustruje kolokabilitu, čili spojitelnost lexému s jinými lexémy. Je komplementární k ekvivalentu a pokrývá široce sémantiku, syntax, styl i kulturně-encyklopedický druh informace.⁸¹

Vzhledem k tomu, že exemplifikace je specifickou lexikografickou náhradou reálného kontextu užití, „musí být jak typická, tj. ilustrující typický (a nikoli často idiosynkratický autorský) úzus, tak konkretizační vůči veškeré informaci většiny hesla, zvláště však vůči informaci naznačené funkcí, valencí, kontextem a sémantikou.“⁸²

Michaela Lišková z Oddělení současné lexikologie a lexikografie v Ústavu pro jazyk český AV ČR však poukazuje na to, že při tvorbě slovníku často bývají vytýkány tyto nedostatky: „exemplifikace je chudá, chybí některé z typických spojení. Nebo naopak exemplifikace je příliš obsáhlá, a tudíž i nepřehledná, s opakujícími se doklady se stejnou spojitelností a sémantikou; jsou zařazovány okrajové případy užití lexikální jednotky; exemplifikace je nicneříkající (např. u frazému *Tichá voda břehy mele: No, pěkně se vybarvuje. Ale jak se říká, tichá voda břehy mele.*)“⁸³ O lexikální ustálenosti výrazu se lze přesvědčit jednak zjišťováním frekvence v jeho příslušném jazykovém prostředí, jednak kontrolou těchto dokladů frekvence vlastním jazykovým povědomím. Pokud jsou však výše zmiňované nedostatky eliminovány, nebo jsou minimální, může být slovník přínosný nejen těm, kdo se jazyk teprve učí. Slovník mohou ocenit i rodilí mluvčí, tj. ti, kteří jazyk již ovládají, a to např. učitelé a autoři učebnic. Není totiž vždy snadné si vzpomenout na dobrý a vystihující příklad užití slova, potřebujeme-li ho typicky ilustrovat, např. při psaní učebnice. I v těchto případech může slovník pomoci.

⁸¹ Čermák, Blatná, 1995 : 243

⁸² Tamtéž : 107

⁸³ Akademický slovník současné češtiny a softwarový nástroj pro jeho tvorbu, www.ujc.cz [online]

2.3.3.6. Víceslovné lexémy

Do této kategorie bychom zařadili různá slovní spojení, idiomy či frazémy, které lze najít v části hesla nazývané sublemma. Zde by měla být uváděna základní, pevná (vázaná) spojení, zvl. terminologické povahy, jako např. *obecní úřad* nebo *krajský soud*. Tato spojení slov mohou být umístěna rozptýleně u jednotlivých hesel či podhesel nebo na konci heslové stati. Jejich výběr je však sám o sobě značný problém, neboť tato spojení svým rozsahem představují množinu zřejmě větší než je sám lexikon jednoslovných lexémů.⁸⁴

Čermák k této problematice uvádí následující: „Jev frazeologie, resp. idiomatiky (oba pojmy, stejně jako frazém a idiom, se prakticky překrývají, první však je označením formálních aspektů jevu, druhý aspektů sémantických) je univerzální a vyskytuje se na všech rovinách jazyka spjatých s kombinatorikou významových jednotek. Počítá se do něj jak třeba přísloví na úrovni vět (*Pozdě bycha honit*), tak idiomatické jednoslovné kompozitum, srov. však např. i idiom vzniklý derivací, čes. *na-jít*, jehož význam ‚nalézt‘ z významů komponentů nevyplývá (*najít* ≠ *na+jít*), nebo přirovnání (obv. s pevným ‚jako‘: *mít hlad jako vlk*) Největší oblastí je však oblast kolokací, nevětných kombinací slov (*vzít do zaječích*).“⁸⁵ Výběr frazémů, resp. idiomů je jev do velké míry závislý na typu slovníku. „Některé frazémy, resp. idiomy jsou natolik frekventované, a při dešifraci textu krajně obtížné, že je třeba s nimi počítat i u malého slovníku, i když tady nejde v první řadě o jejich aktivní znalost.“⁸⁶

F. Čermák však upozorňuje na terminologickou problematiku v této oblasti. „Bez nároku na vyčerpání celé problematiky si je napřed třeba připomenout v této oblasti dosavadní neuspokojivou situaci terminologickou. Necháme-li stranou lingvisticky těžko obhajitelné, přesto však úporné a nepřijatelně zúžené lpění syntaktiků na jinak zcela obecném termínu *syntagma* pro větnou dvojici, pak na úrovni kombinací nevětných je třeba si připomenout, že se poměrně libovolně a bez valného terminologického zakotvení užívala hrst termínů, spíše však pseudotermínů historicky různého pozadí. Srov. *slovní spojení, sousloví, sdružené pojmenování, rčení, úsloví, víceslovné pojmenování, fráze a obrat*.“⁸⁷ Čermák navíc poukazuje na to, že Slovník spisovné češtiny většinu z těchto termínů definuje

⁸⁴ Čermák, 2011 : 211

⁸⁵ Čermák, 2011 : 211

⁸⁶ Čermák, Blatná, 1995 : 244

⁸⁷ Čermák, 2001 : 114-125

„bohužel téměř dokonalým kruhem“.⁸⁸ Autor ještě dodává, že „stranou při těchto úvahách zůstala oblast frazeologie s jejím dnes snad už obecněji přijímaným *frazémem* a *idiomem*; s neorganickým rusismem *frazeologismus* se setkáváme už jen zřídka.“⁸⁹

2.3.3.7. Technický odkaz ve slovníkovém hesle

Odkaz by měl sloužit k pomocnému provázání všeho, co abecední princip v hesle rozděluje či zatemňuje, a to např. gramatickým odkazům vidů, komparativů apod. Při odkazování je zásadně důležitá forma, explicitnost, přesnost a hlavně důslednost odkazu; nelze odkazovat vágně, ani kruhem, což je jeden z nejčastějších nedostatků ve slovníkovém výkladu.⁹⁰ Jinými slovy je třeba se vyhnout případům, kdy „k definici slova A je užito slova B (C,D...) a k definici slova B (C,D...) je naopak užito slova A. např. substantivum *gram* je definováno jako *tisícina kilogramu* a s ním související substantivum *kilogram* je naopak definováno jako *tisíc gramů*. Výsledným zjištěním uživatele slovníku je pouze to, že gram je tisícina ‚čehosi‘. Často tímto způsobem bývá propojeno více jednotek, jak můžeme vidět v následujícím příkladu“⁹¹:

- **oděv** - šaty, oblek
- **oblek** - věc ušitá z látky apod. k oblékání, oděv, šaty
- **šaty** - svrchní oděv, oblek

2.4. Homonymie a polysémie

S problematikou slovníkových hesel úzce souvisí také otázka homonymie a polysémie. Homonymie je „jev náhodné podobnosti formy dvou nebo více lexémů (jejich význam je ovšem různý) a při analýze textu může někdy působit potíže, srov. fr. *fin* (*jemný*) a *fin* (*konec*) nebo čes. *travička1* a *travička2*. Homonyma se dále člení podle formy grafické (homografa – angl. *read*, *read*) a fonické formy (homofona – čes. *mýt*, *mít*).“⁹² Některá slova jsou však zároveň homofona i homografa, např. angl. nepravidelná anglická slovesa *put*, *cut*. Dodejme, že homonymum může nicméně vzniknout i pouhým rozdílem

⁸⁸ Tamtéž + zde i bližší popis jednotlivých definic

⁸⁹ Tamtéž

⁹⁰ Čermák, 2001 : 66

⁹¹ Čermák, 2001 : 67

⁹² Čermák, 2011 : 212

v rodě, např. fr. *le mode* a *la mode*, popř. odlišností v přízvuku, srov. angl. *rebel* (subst.) a *re'bel* (verb.) Filipec rozlišuje čtyři typy homonymie, a to⁹³:

- 1) vlastní lexikální homonymie, která se označuje číselným indexem, obvykle arabskou číslicí, u heslového slova: *kolej*¹⁻²; *klika*¹⁻³, *zapírat*¹⁻³, *masově*¹⁻²
- 2) slovnědruhová homonyma, která se uvádějí buď s číselným indexem (k odlišení od lexikálních římskou číslicí) nebo v souborné heslové stati dělené speciální značkou: *hanba* I ž, II přísl; III citosl.
- 3) morfologická homonyma, u nichž jde veskrze o významovou souvislost, se uvádějí buď jako heslo a podheslo v jedné heslové stati: *beran*, nebo při nadřazení významového hlediska v mezích polysémie jako samostatné sémémy: *bačkora* 1. ž. ‘měkká domácí obuv’; 2. *m* i ž ob. hanl. ‘člověk měkký, slabý’
- 4) slovotvorná homonyma (s jedním morfémem významově disjunktivním, s druhým stejného významu) se chápou jako polysémní: *čistič* m 1. ‘kdo čistí’; 2. ‘čistící zařízení, prostředek’. Slovtvorná homonyma bez významové souvislosti se označují jako vlastní lexikální homonyma: *travička1*, zdrob. k tráva; *travička2* ž k travič.

Vztah homonymie a polysémie však nebývá vždy jednoznačný, neboť je někdy obtížné rozhodnout, zda jde o dvě slova (homonyma), anebo jediné, ale s dvěma významy.

Polysémii (též mnohoznačnost, popř. mnohofunkčnost) charakterizuje F. Čermák jako „jev běžný, ba zákonitý v každém jazyce u jeho centrálních, nejčastěji užívaných lexémů a je to zároveň i jeden z nejčastějších způsobů rozšiřování slovníku, lexikonu (‘ideální’ slovník, v němž by jedné formě vždy odpovídal jeden význam, nemá žádný z přirozených jazyků). Takový jazyk by pro přílišné množství lexémů těžko fungoval.“⁹⁴ Autor dále dodává, že polysémie je „označení spektra dvou či více významů téhož lexému, u kterých je společný průnik a prokazatelná synchronní souvislost jednoho s druhým; kde taková polysémní souvislost není, jde obvykle o dvě slova homonymní.

⁹³ Čermák, Blatná, 1995 : 24

⁹⁴ Čermák 2011 : 214

Lexém mající pouze význam jediný pak nazýváme monosémní.⁹⁵ Hranice mezi polysémií a homonymií však nejsou jasně vymezené. To dokládá Čermák na následujícím příkladu čes. slova *koruna*:⁹⁶

1 – horní část stromu

2 – horní ozdoba krále (tj. jeho hlavy)

3 – mince (pův.) s obrazem krále/ jeho hlavy

Významová souvislost u jednotlivých významů polysémie se obvykle nazývá motivací. Ta se zakládá na odvození jednoho významu z jiného, a to zpravidla metaforickým či metonymickým procesem. V případě vztahu 1-2 se jedná o podobnost a tedy metaforu, v případě vztahu 2-3 zase o sousedství, tedy metonymii.

Čermák dále dodává, že „bude-li lexém polysémní, je třeba uvážit přesun, častěji však rozdělení různých oddílů hesla k jednotlivým významům a jejich zmnožení pro každý význam. Řazení významů by mělo být v synchronním slovníku preferenčně od nejfrekventovanějšího k řídkým.“⁹⁷

2.5. Metajazyk ve slovníku

Z etymologického pohledu lze říci, že slovo *metajazyk* obsahuje řeckou předponu *meta-* která má význam ‚přes, po, za‘. Je to tedy, jazyk za jazykem, jazyk popisující jazyk jakožto předmět popisu. Protikladem k metajazyku je vlastní jazyk, který se nazývá *objektový*. Metajazyk je mimo jiné obsažen i v některých hrách, např. v křížovkách, do nichž se na základě metajazykové legendy vpisuje slovo objektového jazyka.

„V lingvistice jsou součástí metajazyka všechny popisné, výkladové a teorií užívané prostředky, které slouží jednak k popisu jazykových prvků, tj. nominací, jednak k popisu vztahů mezi nimi, tj. syntaxe. V lexikografii se metajazyk používá v příkladových slovních spojeních a větách. Do metajazyka tak patří všechny složky od gramatických

⁹⁵ Tamtéž

⁹⁶ Čermák 2011 : 214

⁹⁷ Čermák, Blatná, 1995 : 108

údajů, jako jsou zkratky označující koncovku genitivu, gramatický rod, u sloves dokonavost apod., až po slovníkovou definici.“⁹⁸

V lexikografické praxi se pro metajazyk již tradičně používají některé symboly, znaky, jakož i styly písma. Tak například heslové slovo neboli lemma bývá zpravidla uvedeno tučným písmem, slovníková definice zase kurzívou. Dále jsou uváděny různé gramatické, oborové, či stylistické zkratky, jako např. *zast.*, *bás.*, *lék.*, Co se týče znaků a symbolů, můžeme si všimnout, že výslovnost se nejčastěji uvádí v hranatých závorkách []. Ve špičatých závorkách < > bývá zase uveden původ slova, tildou ~ je zastoupeno heslové slovo u sublemmat atd.

Čermák zdůrazňuje, že je třeba metajazyk do slovníku zakomponovat velmi pečlivě, neboť „i když nejde primárně o jazyk definic a terminologii výkladu, nejde na druhé straně ani jen o jazyk zkratk a značek. Týká se kromě ekvivalentu veškerého podání, výkladu informace pro uživatele v podobě glos, poznámek, opisů, aj.“⁹⁹

Situaci vystihuje tvrzení české lingvistky Renaty Blatné, která se domnívá, že „situaci lexikografa by bylo možno přirovnat např. k malíři, který by měl vysvětlit svůj tvůrčí záměr obrazu opět pomocí kresby.“¹⁰⁰

2.6. Desatero slovníkářovo

Na závěr si uveďme slovníkářovo desatero¹⁰¹ sestavené významným českým lingvistou Františkem Čermákem, který je specialistou právě na lexikologii a lexikografii, jakož i na korpusovou lingvistiku. Tato pravidla prakticky shrnují výše popsané body a témata.

- 1) Slovník do cílového jazyka má dělat kvalifikovaný rodilý mluvčí.
- 2) Žádný slovník není úplný, jde vždy o výběr. Ten však nesmí být ledajaký
- 3) Slovník má být v konečné podobě na jiných zdrojích informace nezávislý, i když se implicity a jako předchozí znalost obvykle předpokládají (ideální je vycházet z existujících popisů obou jazyků).

⁹⁸ Čermák, Blatná, 1995 : 73-74

⁹⁹ Čermák, Blatná, 1995 : 235

¹⁰⁰ Tamtéž : 74

¹⁰¹ Čermák, Blatná, 1995 : 246

- 4) Nepřeložitelnost neexistuje, existuje jen omezená znalost slovníkáře: problémy se neházejí pod stůl, ale řeší dobrým příkladem a kontextem, když je nelze zobecnit jako význam.
- 5) Žádný slovník není bez chyb. Slovník je umění možného.
- 6) Slovu neodpovídá slovo, ale heslu heslo s celým aparátem a vztahy.
- 7) Ekvivalent je především sémanticko-funkční, založený na aproximativní identitě úzu i systémové příslušnosti toho, co se srovnává v obou jazycích. Skutečný úzus lze zjistit jen studiem typické kolokability.
- 8) Heslo a jeho informace má vyhovět potřebě uživatele a povaze sledovaného jazyka, a proto může být svou povahou jak nadslovní, tak podslovní. Je v průsečíku relevantních vztahů paradigmatických a syntagmatických.
- 9) Slovník je přítel, a ne učeně odpudivý povýšenec či podvodník: přes velkou sumu vloženého umu a znalostí musí vycházet vstřícně, a ne mást či jen rozmarně napovídat, nakusovat.
- 10) Lepší jeden slovník než deset gramatik, je-li dobře udělán.

3. Korpus

V tomto oddílu se budeme věnovat korpusu a otázkám s ním souvisejícím. Stručně srovnáme situaci korpusů dříve a dnes a uvedeme si, jaká jsou jeho možná využití. Následně bude pozornost zaměřena na situaci korpusu v českém kontextu, zmíníme proto projekt Český národní korpus. Na závěr se pokusíme shrnout možné výhody a nevýhody korpusu a z nich vyplývající možnosti a meze korpusu.

František Čermák, přední český lingvista a jeden z hlavních představitelů korpusové lingvistiky u nás, charakterizuje korpus jako „rozsáhlý, elektronicky uložený, zpracováváný a přístupný soubor jazykových dat cíleně nashromážděný jako referenční zdroj pro vědecké studium jazyka a zpracování užitých jazykových nástrojů, příruček.“¹⁰² Jinými slovy Čermák korpusem míní „velmi rozsáhlý, strukturovaný soubor textů (resp. přepisů záznamu mluvy v případě mluveného jazyka) různého druhu (obvykle i s vnesenou anotací), obhospodařovaný počítačem.“¹⁰³ Francis¹⁰⁴ zase definuje korpus jako „soubor textů, který se považuje za reprezentativní pro daný jazyk, nářečí nebo jinou podmnožinu jazyka“¹⁰⁵ Obecně se korpusem může rozumět i jakýkoli soubor textů (např. korpus textů určitého autora, žánru apod.) Korpus, tedy soubor dokladů autentických reálných jazykových dat, je pro lingvisty všeho druhu obecným a základním zdrojem, popř. testovacím prostředím, který, jak dodávají Cvrček i Kovářiková, „výrazně převyšuje jakýkoli jiný dostupný zdroj dat, a to co do kvantity, kvality i vytežitelnosti.“¹⁰⁶ „S jeho pomocí je možné vyhledávat slova a slovní spojení v kontextu a zjistit jejich frekvenci v korpusu i původní textový zdroj. Umožňuje i další zpracování nalezeného (např. abecední třídění apod.). U některých korpusů lze vyhledávat i podle slovních druhů.“¹⁰⁷ „Hlavní předností korpusu je i schopnost vypovídat o frekvenci (četnosti) jevů a jejich typickém úzu, což je informace jen pomocí badatelovy intuice nezjistitelná. Jelikož do korpusu vstupují texty jako celek, poskytuje na rozdíl třeba od lístkového katalogu nevyběrové informace o všech typech jazykových jevů (navíc v rozsahu, který byl dříve nemyslitelný).“¹⁰⁸

¹⁰² Čermák, Klímová, Petkevič, 2000 : 15

¹⁰³ Čermák, 2005 : Korpus, informace a lingvistika [online]

¹⁰⁴ W. Nelson Francis : Language corpora B.C. In Čermák, Klímová, Petkevič : *Studie z korpusové lingvistiky*, Praha, 2000.

¹⁰⁵ Čermák, Klímová, Petkevič, 2000 : 167

¹⁰⁶ Cvrček, Kovářiková, 2011 : 115

¹⁰⁷ Co je korpus? www.ucnk.ff.cuni.cz [online]

¹⁰⁸ www.wiki.korpus.cz [online]

„Co se týče historie, lingvisté v minulosti sbírali své jazykové doklady tak, že si je vypisovali na jednotlivé lístečky - excerpta, zapisovali zdroj, odkud daný výskyt získali a celý kontext, ve kterém se slovo nebo gramatický jev vyskytl. Potom setřídili excerpta abecedně. Tím vznikl archiv.“¹⁰⁹ Vznik prvních korpusů a korpusové lingvistiky se datuje do počátku 60. let 20. stol. Toto období znamenalo důležitý obrat, neboť bylo zahájeno několik velkých projektů korpusů. Od té doby rozsah i počet korpusů roste a korpusová lingvistika stále vzkvétá, především díky rozvoji informatiky. I mnoho lingvistických škol založilo své popisy jazyků a své jazykové teorie právě na pozorování a sbírání přirozených dat. Podobného názoru jsou i autoři Cvrček a Kovaříková: „Od doby vzniku prvního korpusu (Brown Corpus) v 60. letech bylo vytvořeno mnoho nových korpusů, obecných i specializovaných, roste i jejich uplatnění v jazykových příručkách různého druhu (sestavení moderního slovníku se bez korpusové základny dnes už neobejde. Využití korpusů v jazykovědném výzkumu si ve světovém měřítku už vydobylo poměrně pevnou pozici, stejně tak jako korpusová lingvistika, která se postupně etablovala jako samostatná jazykovědná disciplína (na úrovni např. kognitivní lingvistiky, generativní lingvistiky nebo psycholingvistiky), která se zabývá všemi běžnými rovinami jazykového popisu.“¹¹⁰

Co se týče rozsahu dat obsažených v korpusech, M. Šulc uvádí, že v začátcích „velikost korpusů vycházela spíše z lidských, technických a finančních možností prvních korpusových lingvistů a pohybovala se od několika set tisíc k přibližně dvěma milionům slovních výskytů. V dnešní době je za (pravděpodobně dočasný) standard považována velikost 100 milionů slovních výskytů, kterou má snad nejcitovanější korpus 90. let, British National Corpus (BNC). I u předních zahraničních korpusů se pohybují v řádu desetimiónů a stamiónů slov. Tuto velikost nabízí i ČNK a jeho první varianta reprezentativního korpusu současné psané češtiny SYN2000, zpřístupněná veřejnosti v roce 2000.“¹¹¹ „Celkové množství korpusů ve světě nebo i jen v Evropě lze dnes už těžko odhadovat. Pro Evropu v každém případě platí, že je obtížné vůbec najít jazyk, který by korpus neměl, nebo pro který by nebyl korpus připravován.“¹¹²

¹⁰⁹ Český národní korpus. Schmiedtová, 1999. www.ikaros.cz [online]

¹¹⁰ Cvrček, Kovaříková, 2011 : 113-134

¹¹¹ M. Šulc, 2011 : 53-61

¹¹² ČNK základní charakteristika a širší souvislosti. Knihovnická revue, 2004 : 152-168.

Čermák¹¹³ je toho názoru, že korpusy jsou dnes ve většině vyspělých zemí již samozřejmostí a jejich význam stále roste. Jsou jednak aktivně využívány, jednak jsou podporovány výzkumy na nich založené. Jak jsme zmínili výše, dnešní korpusy jsou úzce spjaty s počítačovou technikou, což vede nejen k obrovskému zrychlení a usnadnění lingvistovy práce, ale také k jejímu zkvalitnění. O tom je přesvědčen i americký jazykovědec Wallace Chafe: „Data takto získaná, která lze průběžně snadno modernizovat a doplňovat, představují dnes absolutně nejbohatší a nejrealističtější zdroj poznání jazyka vůbec. Možnostmi kvalitativními a kvantitativními tento zdroj předčí všechny lingvistovy pracně vybudované kartotéky a archivy minulosti. Proto je také korpus předpokladem ke skutečné revoluci v práci s jazykem, a proto se zcela právem mluví o posledním desetiletí tohoto století také jako o dekadě korpusové lingvistiky.“¹¹⁴ Nicméně J. Králík upozorňuje na fakt že „univerzální, všeobecně uznávaná představa o využití korpusů dosud neexistuje, neboť neexistuje ani představa o všech možných budoucích požadavcích na korpusy jako takové.“¹¹⁵

Na tomto místě lze uvést též tvrzení Cvrčka a Kovářkové, autorů článku *Možnosti a meze korpusové lingvistiky*, kteří se domnívají, že „korpus svými metodologickými východisky navazuje na strukturalismus, přináší ovšem do jazykovědného výzkumu specifické nástroje a metody.“¹¹⁶ Situaci přirovnávají k jedné z oblíbených metafor korpusových lingvistů, která hovoří o tom, že „korpus by se otevřením nových perspektiv pro badatele dal přirovnat k mikroskopu. Ačkoli můžeme chápat mikroskop jako pouhý nástroj, s jeho vynálezem byly vytvořeny podmínky pro vznik zcela nových disciplín (např. mikrobiologie), které zkoumají dotud nepozorovanou a nepozorovatelnou realitu, a tím nalézají zcela svébytný předmět popisu, a tedy i ‚raison d’être‘. Stejně tak korpus svým neobyčejným rozsahem předkládá badateli vhléd do oblastí jazykových vztahů, které na izolovaném materiálu jedné věty nebo jednoho textu nebyly viditelné.“¹¹⁷

¹¹³ Čermák, 2005 : 2 [online]

¹¹⁴ Čermák, Klímová, Petkevič, 2000 : 57

¹¹⁵ Králík: 2001, SaS

¹¹⁶ Cvrček, Kovářková, 2011 : 113-134

¹¹⁷ Cvrček, Kovářková, 2011 : 113-134

3.1. Typologie korpusů

Francis rozlišuje v základním dělení korpusů „obecný (ideálně nespécializovaný, psaný, mluvený a synchronní) od specializovaného (např. terminologického, dialektového), jejichž obojí existence je nutná, oba tvoří základní subkorpusy korpusu národního.“¹¹⁸

Jiné dělení představuje typologie podle italské lingvistky Isabelly Chiari, která dělí korpusy z hlediska kvantitativního a kvalitativního¹¹⁹

A) Kvantitativní - z hlediska velikosti korpusu

velikost korpusu	počet slov	příklad (sub)korpusu ¹²⁰
nedostatečný (nereprezentativní)	do 15 tis.	---
malý	15 tis. - 100 tis.	ORWELL
střední	100 tis. - 1 mil.	Pražský mluvený korpus, Brněnský mluvený korpus
středně velký	1 mil. - 50 mil.	Brown Corpus, DIAKORP
standardní	50 mil - 100 mil.	JEROME
velký	přes 100 mil.	SYN2000, British national corpus, FRANTEXT

B) Kvalitativní - z hlediska obsahu

- a. synchronní vs. diachronní
- b. psaný vs. mluvený
- c. smíšený (multižánrový) vs. žánrově vymezený (specifický, terminologický)
- d. jednojazyčný vs. vícejazyčný (paralelní)
- e. referenční (uzavřený, neměnný) vs. nereferenční (neuzavřený, upravuje se)

Kromě výše zmíněných skupin vyčleňují autoři Čermák a Schmiedtová rovněž korpusy studijní (*learners' corpora*) složené z textů psaných studenty cizích jazyků. Dle autorů však existuje také nepřehledná řada čistě tematických korpusů vytvářených pro potřeby jediného oboru či odvětví.

¹¹⁸ Čermák, Klímová, Petkevič, 2000 : 169

¹¹⁹ Chiari, 2007 : 45

¹²⁰ Více k přehledu korpusů na: <https://ucnk.ff.cuni.cz/struktura.php>, přehled korpusů jazyků slovanských a neslovanských také na: https://ucnk.ff.cuni.cz/jine_korpusy.php

3.2. Český národní korpus

Své místo má v otázce korpusu i česká lingvistika. V českém kontextu je v tomto ohledu důležitý projekt *Českého národního korpusu* (ČNK), na němž spolupracuje deset univerzitních pracovišť a akademie (Ústav pro jazyk český a Ústav pro českou literaturu). Tento akademický projekt funguje již dvacet let, neboť Ústav Českého národního korpusu byl založen roku 1994 a jeho pravidelná práce na půdě Filozofické fakulty Karlovy univerzity začala 1. října 1996, kdy „byl vytvořen tým stálých zaměstnanců a ústav mohl zahájit svou činnost. Cílem práce ÚČNK je budování, rozvoj a údržba Českého národního korpusu. V tomto smyslu je korpus prostředkem nejen ke kultivování a poznávání českého jazyka, ale umožňuje i zkoumání zákonitostí lidského myšlení, chování a dalších jevů a šíře i studium naší současnosti a kultury, jak je odráží jen jazyk.“¹²¹ ČNK je přístupný nejen lingvistům, ale i ostatním zájemcům z nejrůznějších oborů, v neposlední řadě pak i studentům. Jedná se o rozsáhlý a velmi dobře propracovaný jazykový korpus, který zajišťuje přístup k více než dvěma miliardám slov uspořádaných v korpusech synchronních i diachronních, mluvených i psaných, paralelních i jednojazyčných, jak dokládají i slova F. Čermáka, zakladatele Ústavu ČNK: „ČNK vznikl jako korpusový projekt kladoucí si za cíl komplexní mapování češtiny synchronní i diachronní a psané i mluvené a jeho jednotlivé korpusy odrážejí také svými názvy to které zaměření, tedy SYN2000 (reprezentativní psaný, synchronní, 100 miliónů), DIA (postupně rostoucí psaný diachronní za všechna období, srov. K. Kučera, který vytvořil koncepci diachronního korpusu, 1998, 1999), ORAL (kde jde zvláště o PMK a BMK, tj. synchronní Pražský mluvený korpus a Brněnský mluvený korpus. Korpusy nářeční (DIAL) jsou v počátcích.“¹²² ČNK kromě různých druhů subkorpusů obsahuje také nástroje pro práci s korpusovými daty, ale i manuál a poradnu pro uživatele.

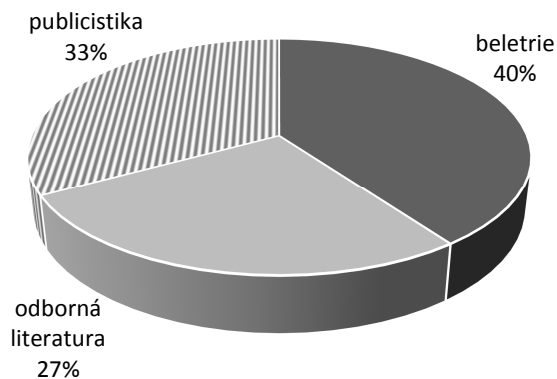
¹²¹ Schmiedtová: 1999 : Český národní korpus. [online]

¹²² Čermák, 2004 : Korpusová lingvistika : stručný historický přehled. In *Český národní korpus* [online].

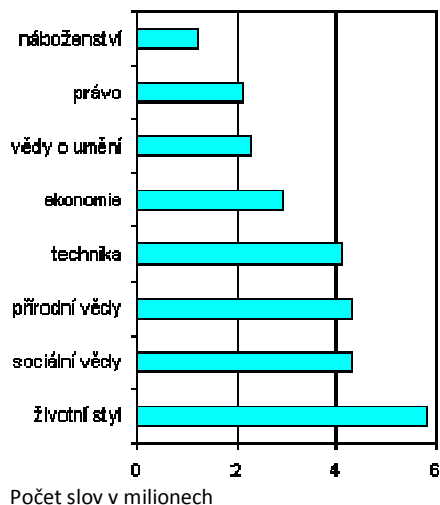
3.2.1. Korpus SYN2005

Představme si nyní konkrétní korpus ČNK, který nám ve druhé části práce bude sloužit jakožto výchozí materiál.

Nejprve se podívejme na stručnou charakteristiku korpusu SYN2005. Jedná se o synchronní reprezentativní korpus současné psané češtiny, který čítá 100 milionů slov. Co se týče jeho složení, tento korpus obsahuje texty z beletrie (40%), publicistiky (33%) a odborné literatury (27%). Shrnutí zobrazuje následující graf:



Ústav Českého národního korpusu uvádí, že veškeré publicistické texty jsou z let 2000 - 2004, odborná literatura pochází z let 1990 - 2004, beletrie pak může být dokonce i starší, nicméně v obou případech byl však kladen důraz na to, aby starších textů bylo co nejméně. Složení odborné a další specializované literatury podle tematického zaměření zobrazuje následující graf¹²³:



¹²³ Korpus SYN2005. www.ucnk.ff.cuni.cz [online].

3.3. Výhody a nevýhody korpusu

Co se týče výhod spojených s korpusem, Cvrček a Kovářiková uvádí následující důvody, proč je korpus, resp. korpusová data, pro lingvistická zkoumání cenná¹²⁴:

1. korpus obsahuje reálná autentická jazyková data, která jsou zde uložena ve svém přirozeném kontextu, jinými slovy korpus obsahuje takové údaje, které jsou vzorkem skutečného úzu, a to jak psaného, tak mluveného
2. korpus principiálně není budován selektivně, rozhodnutí o zařazení textů není subjektivní (možná subjektivnost je potlačena množstvím textů)
3. data v korpusu odráží, jak mluvčí jazyk skutečně používají, jsou skutečné texty, resp. promluvy, to znamená, že se neopírají o deklarativní postoje mluvčích typu „já si myslím“
4. korpus je navíc sestavován jako nespecifický zdroj informací o jazyce, tzn. není vytvářen s ohledem na jeden konkrétní výzkumný úkol (např. zkoumání morfologie nebo tvorbu slovníku), a může tedy sloužit k výzkumu nejrůznějších jevů v různých oblastech, kromě lingvistiky např. i v didaktice, literární vědě, psychologii či sociologii, ale i oborům zdánlivě s jazykem nesouvisejících, jako je matematika či informatika.
5. pomocí korpusu lze realizovat i zkoumání, která s jiným jazykovým materiálem nejsou možná
6. data jsou pohotově k dispozici každému uživateli
7. korpusový prohlížeč je schopný vyhledávat i složité dotazy během několika vteřin
8. velké a nespecializované korpusy v ČNK jsou statické, texty v nich neubývají ani nepřibývají; proto je možné každé vyhledávání zopakovat a ověřit, a to i po několika měsících či letech, jinými slovy korpus umožňuje opakování experimentu za identických podmínek

¹²⁴ Cvrček, Kovářiková, 2011 : 113-134

9. data lze pomocí subkorpusů jednoduše rozčlenit podle zadaných údajů o textu. V ČNK jsou to mj. tyto: registr, datum vzniku a pohlaví autora/autorky, u překladů zdrojový jazyk, médium textu (kniha, noviny) a bibliografické údaje

10. korpus umožňuje získat údaje o frekvenci neboli typičnosti jevů

11. kontinuální budování korpusů plní i velmi žádoucí vedlejší funkci – tvoří bázi pro mapování jazykového vývoje a slouží jako archiv dobového jazykového úzu

Čermák k těmto poznatkům doplňuje ještě následující body¹²⁵:

12. cennou je i možnost propojit korpusy dvou nebo více jazyků do paralelního korpusu, který umožňuje stejnými korpusovými metodami získat hlubší vhled do srovnání jazyků

13. Jako zásadní se dnes jeví především vztah korpusu a moderní lexikografie, a to jak co do nesrovnatelné nabídky datové tak nové metodologie založené na syntagmaticnosti

Pokud jde o nevýhody, je třeba podotknout, že ani korpusová lingvistika není bez limitů. Cvrček a Kovářiková uvádí tyto¹²⁶:

1. Základním problémem je dán rozsahem dat (není jich nikdy dost) a jejich povahou
2. Korpusové lingvistice je často vytýkáno to, že pracuje pouze se vzorkem promluv, a není proto schopna postihnout jevy potenciální. Korpus nikdy nemůže obsáhnout všechny možnosti, kontexty, určité konstrukce tak mohou chybět. Nicméně absence konstrukce ještě neznamená, že tato není možná, nicméně je tento jev vždy signifikativní a měli bychom jej proto brát v potaz. S tímto souvisí i další nevýhoda, a sice že korpus nemůže poskytnout záporný důkaz.

¹²⁵ Čermák, 2004 : Korpusová lingvistika : stručný historický přehled. In *Český národní korpus* [online].

¹²⁶ Cvrček, Kovářiková, 2011 : 113-134

3. V korpusu mohou být chyby, ty by však v kvalitním korpusu měly být odlišeny od korektních konstrukcí.
4. Závažná je jistě i námitka, že na rozdíl od mluveného jazyka v psaném jazyce zdaleka všechno není a některé typy informace v něm nejsou vůbec. O poměru mezi jazykem mluveným a slyšeným versus psaným a čteným se vyjadřuje i M. Šulc, podle jehož názoru budou korpusy vždy nevyvážené „vzhledem k poměru, ve kterém se vyskytují v běžném životě (odhady hovoří o 90-95 % mluveného jazyka a pouhých 10, ale možná spíš jen 5 % jazyka psaného). Jistě budou pomocníkem pro rozlišení frekvence užití v oblasti mluveného a psaného jazyka, jistě budou velmi cenným zdrojem pro specializovaná studia mluveného jazyka, např. morfologická či syntaktická, která jsou velmi potřebná, avšak pro lexikografická studia podobné korpusy nebudou ještě velmi dlouhou dobu dostatečně veliké.“¹²⁷
5. Další limit korpusové lingvistiky je dle Cvrčka a Kovářkové spojen s procesem získávání dat. Je zjevné, že promluvy z některých vypjatých situací (partnerská hádka, poslední slova umírajícího, intimní rozhovor milenců apod.) nebudou v korpusu nikdy dostatečně zastoupeny. Můžeme se proto asi rozloučit s myšlenkou, že bychom měli někdy relativně úplnou korpusem podloženou představu např. o inventáři českých nadávek, o jazyce rozčilených lidí nebo o způsobech komunikace v emocionálně excitovaných situacích apod.
6. S výše zmíněným souvisí i otázka reprezentativnosti korpusových dat, která je rovněž zásadní. Pojmem reprezentativností chápeme takový vzorek, který zachycuje stejnoměrně jak kvalitativní, tak kvantitativní rysy jazyka.

Jak se domnívají Čermák, Králík, Kučera, autoři článku *Recepte současné češtiny a reprezentativnost korpusu*, „reprezentativnost bude do značné míry určována strukturou korpusu, tj. poměrným zastoupením textů z různých komunikačních oblastí reprezentativnost korpusu je problémem, který dosud nebyl uspokojivě (tj. dostatečně zdůvodněně a nesporně) vyřešen ani v žádném z četných korpusových projektů

¹²⁷ M. Šulc, 2011 : 53-61

existujících v různých zemích světa; v řadě vlivných koncepcí se dokonce rezignovaně přiznává nemožnost reprezentativnosti v pravém slova smyslu vůbec dosáhnout.“¹²⁸

Na druhou stranu, autoři Cvrček a Kovářková se domnívají, že „jsou výzkumné úkoly, u nichž je s ohledem na jejich cíl možné dospět k všeobecně přijatelnému řešení (např. pro studium jazyka K. Čapka je vhodným nástrojem korpus všech jeho publikovaných textů). Ať už je tedy koncept reprezentativnosti jakýkoli, vždy existuje možnost vytvořit si vlastní výběrový korpus (virtuální subkorpus) podle požadavků daného výzkumu. Zdaleka ne každý výzkum si totiž klade za cíl popisovat celý jazyk.

Z toho plyne, že:

- a) všechny výtky vůči reprezentativnosti jsou výtkami proti konkrétnímu sestavení daného korpusu, nikoli proti principu korpusové lingvistiky
- b) je plně v moci badatele se s tímto nedostatkem vyrovnat sestavením vlastního materiálu na základě nabídky poskytovatele korpusu.“¹²⁹

V těchto bodech souhlasí i výše zmiňovaní autoři Čermák, Králík, Kučera, kteří se domnívají, že „lze předpokládat, že zájem řady uživatelů se bude soustřeďovat jen k dílčím oblastem jazykového úzu (mnohdy jistě půjde právě o zvláštnosti, o periferii), a že tedy spíš než strukturovanost ČNK jako celku tito zájemci ocení to, že pro své speciální účely a potřeby budou moci z textů obsažených v ČNK vytvářet své vlastní a sledovanému cíli podřízené virtuální korpusy.“¹³⁰

3.4. Další druhy materiálu pro lingvistické účely

Další materiály, které máme jako lingvisté v současnosti k dispozici, lze získat i jinými postupy, pomocí nichž následně lze realizovat výzkum. Cvrček a Kovářková jmenují tyto:

¹²⁸ F. Čermák, J. Králík, K. Kučera, 1997 : 117-124

¹²⁹ Cvrček, Kovářková, 2011 : 113-134

¹³⁰ Čermák, Králík, Kučera, 1997 : 117-124

1) Lingvistická introspekce

„Ačkoli byla donedávna hlavním zdrojem informací o jazyce data intuitivní, získaná většinou

introspekci samotného badatele, nelze tento přístup považovat za spolehlivý, jelikož spoléhání se na vlastní jazykové povědomí vede k subjektivním výsledkům (které budou navíc ovlivněné znalostí předchozích popisů jazyka). Oproti intuici má korpusový materiál výhodu, že je objektivní a dle autorů přináší mimo jiné i poznatky, na které by lingvista „z hlavy“ nepřišel.“¹³¹

K této problematice uvádí Cvrček a Kovářiková příklad různých variant frazému *Vlk se nažral a koza zůstala celá*. Ačkoli jsou frazémy často vnímány jako neměnný celek, díky korpusu je možné zjistit, že jsou užívány daleko volněji, než by se dalo čekat. Korpusu odhalil užití frazému v následujících obměnách¹³²:

Jenže vlk se nikdy nemůže nažrat tak, aby koza zůstala celá.

... byrokratický vlk nažral a podnikatelská koza zůstala celá.

Aby se prostě vlastenecký vlk nažral a národní koza zůstala celá.

Aby se vlk federální kasy nažral a domácí špeditérská koza zůstala celá.

Autoři uvádí i další příklad, který přináší zajímavý postřeh týkající se synonym šedivý a šedý. Přestože jsou tato dvě slova vnímána jako velmi blízká, díky korpusu lze odhalit, „že se liší víc, než bychom u takto blízkých synonym čekali; velkou roli tu totiž hrají terminologie a kolokace.“¹³³

lemma šedivý: vlasy, oblek, mraky, obloha, nebe, vousy, kalhoty, šaty, knír, bradka, myš, mlha, popel

lemma šedý: vlasy, oblek, plášť, kalhoty, obloha, mraky, ale také eminence, ekonomika, zákal, hmota (mozku), zóna, kůra mozková, litina

Autoři doplňují, že „kombinace lexému *šedivý* a slov *eminence, ekonomika, zákal* apod. se nevyskytují. Takovéto příklady by se však pomocí metody introspekce nacházely jen velmi obtížně.

¹³¹ Cvrček, Kovářiková, 2011 : 113-134

¹³² Tamtéž

¹³³ Cvrček, Kovářiková, 2011 : 113-134

Nicméně skutečnost, že intuitivní data jsou nedostatečně spolehlivá, samozřejmě neznamená, že by tento přístup měl být zcela zavržen. Naopak, jsou velmi důležitou součástí lingvistického výzkumu v několika jeho fázích. V první řadě lingvistickou introspekci využíváme pro zvolení vhodné a smysluplné výzkumné otázky. Zároveň je introspekce cenným vodítkem při hodnocení výsledků nalezených pomocí korpusu.“¹³⁴

2) Dotazníky

Některé nevýhody plynoucí z introspekce, by dle autorů mohla vyřešit metoda dotazníků. I ta má však v některých případech stinné stránky. „Problém mísení rolí lingvisty a mluvčího, s kterým se setkáme u introspekce, by bylo možné vyřešit obrácením se na informanty, ostatní mluvčí jazyka, nejlépe nefilology. V této otázce je výhodné odlišit postoje deklarativní, to jsou ty, které mluvčí projevují při dotazníkové evaluaci jevů, a postoje reálné, tj. ty, které se projevují právě v úzu, např. výběrem jazykových prostředků pro stylizaci, jinými slovy jazykové chování. Deklarativní postoje jsou dnes nejčastěji sledovány pomocí dotazníků. Nicméně i tato metoda (stejně jako introspekce) je náchylná k ovlivňování zkoumaného předmětu skrze pozorování, ke kterému může dojít při aranžování výzkumu (formulace dotazů atp.).“¹³⁵

3) Příležitostný sběr materiálu

Autoři uvádí, že k příležitostnému sbírání materiálu může dojít např. ve chvíli, kdy „badatele něco zaujme v náhodně čteném textu, či zaslechnutém hovoru. Nevýhoda tohoto přístupu však spočívá v tom, že se pozornost přirozeně soustředí především na zvláštnosti nebo neobvyklé jevy. Takový přístup je pak subjektivně deformující, neboť to, co je obvyklé, frekventované, pravidelné nebo centrální, většinou pozornost nepřitahuje. Výsledky, třebaže mnohdy zajímavé, jsou touto metodou tedy nutně poznamenány.“¹³⁶

3.5 Lexikografické využití korpusu

S korpusem souvisí odvětví lingvistiky, které nazýváme korpusová lingvistika, tedy ta část lingvistiky, která „systematicky pracuje s korpusem a jeho nástroji, resp. studuje

¹³⁴ Tamtéž

¹³⁵ Cvrček, Kovářková, 2011 : 113-134

¹³⁶ Tamtéž

zásady a praxi práce s ním s cílem lepšího poznání funkce a struktury jazyka.¹³⁷ Vzhledem k tomu, že se v naší práci zabýváme slovníky a jejich výstavbou, uvedme si nyní několik poznatků, týkajících se vztahu korpusové lingvistiky a lexikografie.

Významný pokrok v lexikografii znamenal vydání slovníku COBUILD Dictionary of English Language, který sestavil John Sinclair a jeho kolegové. „Návrh podoby jednotlivých hesel se opíral primárně o korpusové doklady, což kontrastuje s klasickou lexikografickou praxí, která jako hlavního zdroje dat často užívá jiné slovníky a jen občas je nahrazuje méně systematickou exerceí z korpusu.“¹³⁸ Dle F. Karlssona¹³⁹ může dále korpusová lingvistika (mimo to, co ukázal už projekt COBUILD), lexikografii přispět i různými statistickými metodami jako je vzájemná informace, kterou lze využít pro vyhledávání signifikantních kolokací. „Vzájemná informace srovnává pravděpodobnost společného výskytu např. dvou slov (říkáme jí společná pravděpodobnost) s náhodou, tj. s pravděpodobností, že daná slova nalezneme samostatně. Na základě toho lze dospět k několika lexikograficky užitečným postřehům. Za prvé jsou korelace základem pro výběr příkladových vět ukazujících běžný úzus.“¹⁴⁰ Dobrý slovník by měl samozřejmě reflektovat spíše frekventované a typické příkladové věty než věty řídké, zastaralé, nebo autorem vymyšlené, které bychom označili jako autorské inovace.

Karlsson dále uvádí, že „tyto korelace ukazují gramaticky a morfosyntakticky signifikantní kolokace a pomáhají i při rozhodování o tom, které významy jsou centrální a měly by být blíže k počátku slovníkového hesla.“¹⁴¹

F. Karlsson přináší shrnutí toho, jakým způsobem může být korpus užitečný na poli lexikografie¹⁴²:

- 1) Umožňuje vytvářet frekvenční seznamy lemmat. To je dobrá pomůcka, když uvažujeme o tom, které slovo zahrnout do slovníku.

¹³⁷ Čermák, Klímová, Petkevič, 2000 : 15

¹³⁸ Čermák, Klímová, Petkevič, 2000 : 427

¹³⁹ Tamtéž : 428.

¹⁴⁰ Čermák, Klímová, Petkevič, 2000 : 428

¹⁴¹ Tamtéž.

¹⁴² Čermák, Klímová, Petkevič, 2000 : 434-435

- 2) Umožňuje srovnání toho, jak se slovník účastní složení textových typů na úrovni lemmat, což lépe osvětluje takové pojmy jako jádro slovní zásoby a specializovaná slovní zásoba. Srovnání též pomáhá při výběru slov pro slovník.
- 3) Lemmatizace ve spojení se slovními druhy a dalšími morfologickými a syntaktickými rysy poněkud zvyšuje úroveň abstrakce a pomáhá lexikografovi při strukturaci korpusových dat.

Podobný názor mají i autoři Čermák a Schmiedtová, kteří navíc přínos korpusu oproti tradiční excerpci spatřují již díky konkordanci (tj. „výpis všech řádek s výskytem hledaného jevu v kontextu), která je základním výsledkem hledání v korpusu, jim především předkládá data podstatně lépe podložená a zasazená do dostatečného, libovolně rozšiřitelného kontextu; vedle toho korpusový program poskytuje i nejdůležitější frekvenční a statistické charakteristiky příslušných dat a další techniky průzkumu souvislostí mezi slovy. Na základě analýzy všech těchto údajů mohou lingvisté velmi snadno získat představu o tom, co a jak se používá typicky a co okrajově (zaměření na typičnost užití je samozřejmě mimořádně důležité pro lexikografii, ale zdaleka nejen pro ni).“¹⁴³

F. Čermák dělí účely čerpání z korpusů do následujících šesti okruhů¹⁴⁴:

- 1) Ověřování existence prvku, lexému či jeho aspektu
- 2) rozhodování o odstupňovanosti variant a úzu forem
- 3) Časová datace dokladu, zvl. při posuzování synchronie
- 4) Zjišťování
 - a. formální, sémantické informace a prvku
 - b. jeho paradigmatické a syntagmatické povahy
 - c. povahy a jeho okolí (struktury jeho kontextu)
- 5) Posouzení optimální nebo autentické citace/ exemplifikace
- 6) Kontrola, popř. analýza metajazyka

¹⁴³ ČNK základní charakteristika a širší souvislosti. Knihovnická revue, 2004 : 152-168.

¹⁴⁴ Tamtéž : 27-29

Autor dále uvádí, že na využití korpusů a jeho produktů se vedle lexikologů a lexikografů mohou podílet i:

Lingvisté

- A) teoretici: jazyková teorie obecně, počítačová lingvistika, analýza a syntéza mluvy
- B) aplikovaní: terminologie, překlad tradiční i strojový, jazyková propedeutika a tvorba učebnic

Nelingvisté

- A) Obsahoví specialisté: historici, literární kritici, terminologové
- B) Komunikační specialisté: armáda aj.
- C) Informační a kognitivní specialisté a specialisté na umělou inteligenci

4. Frekvenční slovník

Následující kapitolu věnovanou frekvenčním slovníkům jsme zařadili z toho důvodu, že druhá část práce úzce souvisí s frekvenčními slovníky, frekvenčními seznamy a korpusovou frekvencí obecně. V empirické části se totiž budeme zabývat analýzou hesel vybraného souboru slovníků, a to z hlediska nejen jejich vzájemného srovnání, ale především z hlediska frekvence. Budeme se snažit zjistit, do jaké míry se autoři slovníků řídili frekvenčními seznamy, popř. zda při výběru hesel volili jiná kritéria. Cílem této kapitoly je tedy frekvenční slovníky stručně představit, vysvětlit k čemu slouží a jaké je jejich možné užití.

Frekvenční slovník bychom zjednodušeně mohli nazvat také slovník nejužívanějších slov. Frekvence slov (slovních tvarů, slovních spojení) je jedním z kritérií pro popis a uspořádávání slovní zásoby. Přiřazením frekvenčních údajů, neboli kvantitativních údajů o frekvenci, lze vytvořit frekvenční slovník. Takový slovník vychází z korpusu, a to tak, že jednotlivá slova jsou řazena na základě jejich četnosti (frekvence) v korpusu, přičemž na prvním místě je uvedeno slovo s nejvyšší četností. Frekvenční slovníky bývají obvykle jednojazyčné. Mohou být uplatňovány jak pro účely teoretické (např. vyvozování různých lingvistických teorií, popis jazyka, apod.), tak praktické (např. užití pedagogické - výuka cizích jazyků; technické - izolování klíčových slov, atd.)¹⁴⁵

První Český frekvenční slovník vytvořil kolektiv autorů Jelínek, Bečka, Těšitelová v roce 1961. Ačkoli tento slovník obsahoval 1,6 milionů slov, byl vytvořen manuálně. Nicméně, jak se domnívá Hladká, konec 20. stol. zaznamenal významný pokrok v korpusové lingvistice, a tím i pokrok ve frekvenčním lexikografickém popisu češtiny.¹⁴⁶ Frekvenční slovníky, stejně jako slovníky překladové či výkladové, jsou tedy v dnešní době založeny na korpusu, nebo by tomu tak alespoň mělo být. Další vydání *Českého frekvenčního slovníku* následovalo v roce 2004. Lze se domnívat, že se jednalo o jedno z průkopnických, avšak význačných děl na poli korpusové lingvistiky v českém prostředí, neboť shromážděné frekvenční údaje se opírají o materiál ČNK. Tento slovník je tedy, na rozdíl od svého předchůdce, založen na nové a spolehlivější bázi, a tou je Český národní korpus – projekt neustále se rozvíjející, který dnes čítá více než 2,2 miliardy slov. Takto velký reprezentativní korpus umožňuje zajistit vysokou spolehlivost

¹⁴⁵ Těšitelová, 1997, č. 4

¹⁴⁶ Hladká, Martincová: *Tradice a současnost české lexikografie*. www.phil.muni.cz [online]

předkládaných dat. Dle autorů je Frekvenční slovník češtiny „užitečným zdrojem informací nejenom pro širokou veřejnost, ale také pro lexikografy jako vhodný základ pro tvorbu heslářů. Je založen na textovém korpusu, jehož složení bylo proporčně vyváženo tak, aby věrně zachycovalo současný psaný jazyk. V hlavní části slovníku najde čtenář 50 000 nejběžnějších apelatív spolu s údaji o jejich frekvenci, 2 000 nejčastějších vlastních jmen a 1 000 nejčastějších zkratek. Z dodatků se dále dozvíte, jaká interpunkční znaménka jsou nejpoužívanější, která písmena jsou v českém textu nejfrekventovanější“¹⁴⁷

Mimo to však existují také frekvenční slovníky mluveného jazyka, jakým je i *Frekvenční slovník mluvené češtiny*, který vyšel naposledy v roce 2007 a jedná se o vůbec první slovník svého druhu. I tento slovník vychází z korpusu, v tomto případě z Pražského mluveného korpusu. Představuje „autentickou mluvenou češtinu, která tu stojí v protikladu, často příkrém, k češtině spisovné a psané. Ukazuje, jak lidé skutečně mluví, tj. bez kodifikačních zásahů a deformací.“¹⁴⁸

„Frekvenční slovníky představují významný krok na cestě k hlubšímu poznání slovní zásoby jazyka, její struktury a fungování. Vývojové změny v jazyce jsou totiž nejzřetelnější právě u frekvence slov, jejíž obraz se mění módností publicistického vyjadřování, vývojem vědy a technického myšlení, vlivem oblíbených písní atd.“¹⁴⁹ Frekvence je tedy pro zkoumání jazyka podstatná zejména tím, že podává informaci o centru a periférii jazykových jevů.

Vzhledem k tomu, že v naší práci se zabýváme analýzou vybraných kapesních slovníků, bude pro nás frekvence důležitým východiskem. Jedním z našich úkolů bude zjistit, do jaké míry se korpusová frekvence (ze které vycházejí i frekvenční slovníky češtiny) odráží na výběru slovníkových hesel, jinými slovy do jaké míry autoři slovníků frekvenci zohledňují, resp. Toto kritérium přehlídí a proč tomu tak je. Těmto otázkám se budeme věnovat v empirické části práce.

¹⁴⁷ Publikace ÚČNK, www.ucnk.ff.cuni.cz [online]

¹⁴⁸ Publikace ÚČNK, www.ucnk.ff.cuni.cz [online]

¹⁴⁹ Kraus, Frekvenční slovník češtiny, Naše řeč

Část II.

Empirická část

1. Úvod

Tato část práce si klade za cíl provést rešerši na vybraném souboru pěti česko-francouzských kapesních slovníků. Jak jsme již naznačili v první části, při výzkumu budeme analyzovat hesla vybraných slovníků, a to jednak z hlediska jejich vzájemného srovnání, jednak z hlediska frekvence. Seznam hesel získaných ze souboru pěti slovníků budeme konfrontovat s frekvenčním seznamem vzniklým na základě korpusu SYN2005. Tyto výsledky nám umožní analyzovat, do jaké míry se autoři vybraných slovníků (ne)řídili korpusovou frekvencí, popř. jaká další kritéria při výběru hesel zohlednili a co je k tomuto výběru mohlo vést. Dále se zaměříme na slova nejfrekventovanější, nebo naopak málo frekventovaná, a jejich zastoupení ve slovnících. Zde by se dalo předpokládat, že kapesní slovníky by měly zahrnovat především slova s vyšší frekvencí, a slova s nízkou frekvencí by naopak obsažena být neměla. Rovněž by se dalo očekávat, že slovníky, které jsme pro účely naší rešerše vybrali a které mají podobný rozsah a jsou zaměřeny na podobné uživatele, si budou podobné jednak po stránce obsažených hesel, jednak z hlediska celkové výstavby slovníku. Zda jsou tyto hypotézy pravdivé, jsme se snažili vyzkoumat v naší studii.

K provedení průzkumu je tedy zapotřebí soubor pěti slovníků a korpus SYN2005. Tento materiál si představíme v následujících dvou kapitolách. Dále bude popsána použitá metoda a následně již uvedeme samotnou studii. Na závěr představíme výsledky vyplývající z našeho výzkumu.

2. Materiál

2.1. Slovníky

Slovníky, kterými se budeme zabývat ve druhé části této práce, můžeme zařadit mezi slovníky malé překladové, konkrétně dvojjazyčné česko-francouzské, francouzsko-české. Rešerše byla provedena na pěti současných slovnících nabízených na trhu jako slovníky kapesní. Protože jejich názvy se z pochopitelných důvodů téměř neliší, budeme na ně odkazovat prostřednictvím názvů nakladatelství (FRAGMENT, FIN, OLOMOUC, LINGEA, LEDA). Cílem každé z následujících kapitol je představit vybrané slovníky, zhodnotit jejich kladné a záporné aspekty a poukázat na prvky, kterými se daný slovník odlišuje nebo vyniká ve srovnání s ostatními tituly. Tyto aspekty budeme analyzovat na základě teoretických informací, kterými jsme se zabývali v první části této práce. Pozornost bude věnována rozsahu, tj. zda daný slovník odpovídá vymezeným kritériím počtem hesel. Zvolili jsme právě počet hesel, nikoli počet stránek, který sice bývá na obalech slovníků uváděn, avšak vzhledem k odlišným formátům slovníků není tento údaj zcela relevantní. Dále se zaměříme na to, které složky makrostruktury jsou ve slovnících obsaženy, tzn. zda slovník kromě hlavního textu obsahuje i údaje jako jsou obecné informace o publikaci, předmluva, rejstřík. Dále nás bude zajímat, jakým způsobem je vysvětlen metajazyk, či případné přílohy, které tvoří součást slovníku. Pozornost bude věnována i řazení hesel, jakož i mikrostruktuře jednotlivých slovníků. Tyto poznatky nám pomohou určit, do jaké míry byly splněny cíle, které si autoři slovníků vytyčili, tedy zda dílo odpovídá zamýšlenému účelu, a nakolik jsou slovníky přehledné, což má zásadní vliv pro dobrou orientaci ve slovníku.

Výše zmíněné rysy jsme analyzovali, zjištění shrnuje následující tabulka:

Tabulka č. 1: Celkové shrnutí pěti analyzovaných slovníků

	FRAGMENT	FIN	OLOMOUC	LINGEA	LEDA
Celý název slovníku	Velký kapesní slovník	Kapesní slovník	Kapesní slovník	Kapesní slovník (nejen na cesty)	Praktický slovník kapesního formátu
Počet hesel celkem* ¹	37 500	34 000	30 000	34 000	45 800
Jméno autora	Mgr.V.Borovanová Mgr.P.Kameníková Mgr. D.Schránilová	FIN Publishing – autor neuveđen	I. Gailly a kol.	Kolektiv pracovníků Lingea	V. Rejtharová
Rejstřík	✓	✗	✗	✓	✗
Předmluva	✓	✗	✓	✓	✓
Vysvětlení symbolů a znaků	✓	✓	✓	✓	✓
Vysvětlení zkratk	✓	✓	✓	✓	✓
Další dodatky	-velké srovnání kapesních slovníků	- fr.abeceda - poznámky k výslovnosti	- poznámky k výslovnosti -černobílé obrázky - barevné ilustrace	- adresy velvyslanectví a konzulátů -mapy - geografické info	- fr.abeceda - poznámky k výslovnosti
Uveden zdroj excerpce hesel	✓	✗	✗	✓	✓
Obousměrný	✓	✓	✓	✓	✓
Výslovnost IPA	✗	✗	✗	✓	✗
Výslovnost počestěná verze	ve fr. části* ² + některá slova i v české části	v obou částech	ve fr. části	ve fr. části	ve fr. části
Gramatika	✓	✓	✓	✓	✓
Konverzační část	✓	✗	✓	✓	✗
Rok vydání	2007	2005	2004	2008	2006
Formát	90 x 120 mm	85 x 120 mm	110 x 150x mm	85 x 118 mm	105 x 148mm
Doporučená cena	169,-	180,-	289,-	179,-	250,-

*¹ Celkový počet hesel dohromady v obou částech, tj. součet hesel ve fr. i české části

*² fr. část = část francouzsko-česká, tzn. hesla ve francouzštině, ekvivalenty v češtině

2.1.1. FRAGMENT

Jak vidíme ve výše uvedené tabulce, slovník nakladatelství FRAGMENT má rozměry odpovídající zhruba formátu A7 a obsahuje 37 500 hesel. Ačkoli svým vzhledem připomíná slovník kapesní, na základě našich předchozích vymezení bychom ho zařadili spíše mezi slovníky střední velikosti. To je možná i jeden z důvodů, proč jej autoři pojmenovali jako „Velký kapesní slovník“. Slovník je nicméně velmi lehký a skladný, což je jistě výhodou, jelikož předpokládaný uživatel, tj. turista či student, jej tak může pohodlně nosit při sobě.

Slovník byl publikován roku 2007 a na jeho sestavování se podílelo několik autorů, jmenovitě Mgr. V. Borovanová, Mgr. P. Kameníková, Mgr. D. Schránílová, jejichž jména jsou uvedena v úvodu publikace. Dále v této části slovníku nechybí ani informace o zdroji pro excerpci slovníkových hesel, jimiž byly nejen přední francouzské slovníky, jako Le Petit Robert, Larousse či gramatiky českých autorů J. Taišlové či tria Hendrich-Tláskal-Radina. Uvedení zdrojů je jistě pozitivem, autoři se však netají ani tím, že čerpali inspiraci i v česko-francouzských kapesních slovníku jiných autorů. Kromě těchto materiálů autoři při sestavování slovníku vycházeli i z vlastních statistických údajů o frekvenci, které však nejsou blíže specifikovány.

Kromě předmluvy v úvodu se slovník dále skládá z návodu k použití, který najdeme hned na přední straně desek, a který vysvětluje různé typy písma, speciální symboly, značení a strukturu hesel. Dále na vnitřních deskách nalezneme velké srovnání slovníků. Slovník obsahuje rejstřík, což je ve slovníku obsahujícím několik částí podstatné. Kromě hlavního slovníku je součástí slovníku též gramatická část zahrnující výslovnost, slovní druhy a časování sloves. Ve slovníku je zahrnuta i frazeologie se zaměřením na konverzační obraty z prostředí cestování, či obraty společenského styku. Mimo to se slovník věnuje i obchodní korespondenci a poskytuje rady, jak napsat dopis či email ve francouzštině. V doplňkové části jsou pak navíc uvedeny geografické názvy, míry a číslovky.

Tento slovník se od ostatních čtyř odlišuje zejména jedním rysem, a tím je metoda hnízdování. V praxi to znamená, že jednotlivá lemmata jsou sice řazena abecedně, ale na základě příbuznosti obsahují navíc i další podhesla, ať už se jedná o frazémy nebo jednoslovné lexémy. Heslové odstavce jsou tedy obsáhlejší. Tato podhesla nejsou

započítávána do celkového počtu hesel. Podhesla přidružená touto metodou jsou sice opět řazena alfabetycky, dochází však k tomu, že např. pod heslem *monogamie* se skrývá i heslo *monokl*, *monopol*, heslo *morfium* je zase třeba hledat pod lemmatem *morfologie* a heslo *mrakodrap* je dokonce zařazeno ve stejném vstupu jako *mračit se*. Ačkoli se jedná o řazení na základě morfologické příbuznosti, pro nezkušeného uživatele může tato metoda představovat potíže a hrozí, že se v takto uzpůsobeném slovníku nebude dobře orientovat. Můžeme se tedy domnívat, že používání tohoto typu slovníku vyžaduje větší pozornost a jisté jazykové kompetence, kterými uživatel tohoto typu slovníku často nedisponuje. Co se dále týče makrostruktury, můžeme zmínit, že na stránkách ve slovníkové stati je vždy uvedeno orientační písmeno, jakož i první a poslední heslo na stránce, což je jistě kladným bodem pro rychlou a přehlednou orientaci.

Pokud jde o mikrostrukturu, můžeme si všimnout, že výslovnost jednotlivých hesel je systematicky uváděna u hesel ve francouzské části, v české části ji najdeme jen u slov s obtížnou výslovností. Transkripce je však zjednodušená - počeštěná, nejsou tedy užívány znaky mezinárodní abecedy IPA. Je-li heslo terminologicky či stylově specifické, je tato skutečnost označena příslušnou zkratkou. Z morfologie je uváděn pouze rod u substantiv, jiné gramatické označení uváděno není, nedochází-li ke změně slovního druhu. Čísly jsou v heslech značeny polysémnní výrazy, číselný horní index označuje homonyma a hvězdička s číslem potom odkazuje na příslušné nepravidelné sloveso časované v přehledu. Černý puntík upozorňuje uživatele na změnu slovního druhu.

Závěrem lze říci, že daný slovník je skladného formátu, a nezabere tak mnoho místa v batohu zamýšleného uživatele, kterým je primárně cestovatel či student. Slovník je lehký, a to navzdory tomu, že obsahuje poměrně obsáhlou přílohovou část a značný počet hesel, který je vyšší než je průměr u kapesních slovníků. Slovník si klade za cíl maximálně pomoci při užití jak psané a mluvené, tak hovorové a spisovné francouzštiny. Dále je jeho cílem obsáhnout co nejvyšší slovní zásobu, což je pravděpodobně jeden z důvodů, proč autoři zvolili metodu hnízdování. Tento systém však nemusí být vždy nejvhodnějším řešením, uvážíme-li, že slovník je primárně určen pro osoby, které jsou v daném jazyce začátečníky a často i laiky v oboru lexikologie a lingvistiky vůbec. Jako negativum bychom také mohli hodnotit fakt, že informace o obsaženém počtu hesel je uvedena ne příliš přehledně (je uváděn počet hesel zvlášť pro českou a francouzskou část), což může uživatele mást zvláště při porovnávání slovníku s ostatními slovníky.

2.1.2. FIN

Slovník nakladatelství FIN Publishing je svým formátem rovněž malý, lehký, a tudíž skladný a vhodný na cesty, do batohu či kapsy. Svým vzhledem tedy odpovídá označení „kapesní“. Slovník obsahuje přibližně 17 000 hesel v každé části, čímž rovněž převyšuje obvyklý počet hesel u kapesních slovníků. Publikace byla vydána roku 2005 a jedná se o první vydání tohoto titulu. To je ale zhruba vše, co se k obecným údajům o slovníku dočteme. Není zde žádná zmínka o tom, kým byl slovník sestaven, lektorován či redigován, ani z jakých materiálu autoři při vytváření slovníku vycházeli. Nenalezneme zde předmluvu ani rejstřík. V úvodní části slovníku tak najdeme pouze vysvětlení zkratk a značek použitých ve slovníku, znaky pro přepis francouzské výslovnosti a francouzskou abecedu, ač bez doprovodných slovních příkladů. Chybí zde i orientační písmeno na okrajích stránek, což by jistě usnadnilo orientaci. V horní části stránek je však uváděno alespoň první a poslední heslo na stránce. Pod pojmem „gramatická příloha“ bychom možná čekali podrobnější informace týkající se různých gramatických kategorií, jako je tomu u ostatních čtyřech slovníků, v tomto případě však z gramatiky nalezneme pouze časování nepravidelných sloves. Tento oddíl se nachází uprostřed slovníku jako předěl mezi francouzskou a českou částí. Co se dále týče makrostruktury, můžeme si všimnout, že hesla jsou řazena tradičně podle abecedy, na rozdíl od předchozího slovníku však nedochází k hnízdování. Pojí-li se k hlavnímu heslu i podhesla, tato již alfabetycky řazena nejsou, jak dokládá následující příklad¹⁵⁰.

místo ① část prostoru *endroit* [ãdr^ua] *m*, *place* [plas] *f*, *lieu* [ljø] *m*; ~ **dopadu** *point m d'incidente*; ~ **k sezení** *place asise*; ~ **nehody** *lieu d'accident*; **stavební** ~ *emplacement m à bâtir* ② *obec* *localité* [lokalite] *f*; **výletní** ~ *station f touristique*; **v -ě bydlíště** *dans le lieu d'habitation* ③ část *endroit m*; **prázdné/ volné** ~ **v knize** *espace m blanc* ④ pořadí *place f*, *rang* [rã] *m*; **zaujímat první** ~ *occuper/ tenir la première place/ le premier rang* ⑤ *zaměstnání* *place f*, *poste* [post] *m*, *emploi* [ãpl^ua] *m*; **pracovní** ~ *poste stable*; **volné** ~ *place vacante*; **bez -a** *sans emploi*

Princip, dle kterého jsou tato podhesla řazena, ve slovníku není specifikován. Nabízí se řazení dle korpusové frekvence. Po zadání těchto spojení do korpusu SYN2005 jsme

¹⁵⁰ FIN Publishing, 2005 : 497

však zjistili, že ani tato metoda neodpovídá, neboť se ukázalo, že spojení s nejvyšším počtem výskytů (255) je *místo nehody*, uváděné na třetí pozici, zatímco spojení *místo dopadu*, které je na první pozici, mělo počet výskytů pouze 53.

Jak si v uvedeném příkladu můžeme rovněž povšimnout, výslovnost je uváděna v tradičních hranatých závorkách. Ani zde se však nejedná o přepis výslovnosti pomocí znaků mezinárodní abecedy, ale je opět užita počeštěná verze. Jako pozitivum lze však hodnotit fakt, že výslovnost je uváděna v obou částech, tj. ve francouzské i české sekci, což je praktické a začátečník tuto skutečnost jistě ocení. Navíc je výslovnost stručně vysvětlena i v úvodní části slovníku. Pokud jde o mikrostrukturu, z morfologických kategorií je uváděn pouze rod substantiv, jak můžeme ostatně také zaznamenat ve výše uvedeném příkladu. Ostatní slovní druhy jsou vyznačeny pouze v případě, kdy dané heslo může být více slovními druhy. Dále je uživatel pomocí zkratk informován o terminologické či stylové příslušnosti hesla. I v tomto slovníku je pomocí hvězdičky odkazováno na přehled nepravidelných sloves. Pomocí exponentu jsou potom značeny homonymní výrazy.

Na závěr si uvedme několik stinných stránek tohoto slovníku. Za nedostatek oproti ostatním slovníkům, které jsou součástí naší rešerše, je možné považovat absenci konverzační části. Na deskách se sice píše, že slovník „je každodenním ideálním pomocníkem na cestách, doma, ve škole i na úřadě.“¹⁵¹ O tomto tvrzení by se však dalo polemizovat, neboť pro školní potřeby ve slovníku není zastoupeno mnoho gramatických údajů, pro účely cestování a vyjednávání na úřadě zase chybí užitečné základní konverzační obraty. Slovník se tak jeví poněkud stroze, sloužící pouze jako základní pomůcka pro překlad a dorozumění se, a může tedy sloužit spíše příležitostným uživatelům francouzštiny. Na obálce je dále uvedeno, že slovník je prvním vydáním tohoto titulu, doufejme tedy, že následující vydání budou doplněna o chybějící informace.

¹⁵¹ FIN Publishing, 2005

2.1.3. OLOMOUC

Tento obousměrný slovník obsahuje téměř 30 000 hesel, a ačkoli nese název „kapesní“, svým rozsahem, formátem, ale především váhou odpovídá spíše slovníku střední velikosti. Titul byl publikován roku 2004 a jako autorka je uváděna Iva Gailly. Ve slovníku však nalezneme i další jména osob, které se na tvorbě slovníku podílely nebo jej lektorovaly. Na vnitřních stranách desek nalezneme vysvětlení struktury hesel, seznam použitých zkratk a značek a poznámky k výslovnosti. V úvodní části slovníku se pak kromě předmluvy objevují také informace o použité literatuře. Dozvíme se tak, že autoři čerpali z francouzských a českých gramatik i slovníků. Co zde však nenalezneme je rejstřík, což lze považovat za závažný nedostatek, vzhledem k tomu, z kolika sekcí a různých příloh se slovník skládá. Na okrajích stránek rovněž chybí orientační písmeno, které by usnadnilo vyhledávání, opět je uvedeno pouze první a poslední heslo na stránce. Jak bylo naznačeno, tento slovník se od ostatních odlišuje zejména svým formátem a váhou, a to proto, že jako jediný obsahuje ilustrace, a to nejen ilustrace černobílé demonstrující lexémy v hlavní části slovníku, ale také barevné ilustrace v přílohové části ve formě celostránkových obrázků věnujících se různým odvětvím jako je gastronomie, anatomie, botanika, zoologie, bydlení a mnoho dalších. Tato šedesátistránková příloha má za cíl ještě více rozšířit použitou slovní zásobu o další odborné výrazivo. Kromě toho se v přílohové části nachází i rozsáhlý gramatický přehled, který se věnuje většině slovnědruhových kategorií, nepravidelným slovesům, zařazena je i konverzační část zaměřená na fráze každodenního užití, která uživatelům poslouží jako základní pomůcka při orientaci v jazyce a dorozumění se v běžných situacích. V příloze dále nalezneme i zrcadlové informace o frankofonních zemích. Uživatel se tak po prostudování tohoto slovníku může směle pustit do konverzace ve francouzském jazyce.

Dále můžeme zmínit, že ani v tomto slovníku není uplatněna metoda hnízdování, hesla jsou řazena alfabetycky a podhesla se vyskytují jen zřídka. Mikrostruktura je tedy pojata velice jednoduše, neboť celé heslo se nejčastěji skládá pouze z lemmatu, jeho ekvivalentu a u substantiv také označení rodu. Výslovnost nalezneme pouze ve francouzské části, a to opět zjednodušeně pomocí počestěných znaků, které jsou však vysvětleny v poznámkách k přepisu francouzské výslovnosti nacházejících se v úvodu publikace. Případná specifika či stylová příslušnost jsou v hesle označena zkratkami. Hvězdičkou je opět odkazováno na seznam nepravidelných sloves.

Pro shrnutí lze říci, že výrazným znakem tohoto slovníku je jeho obsáhlá příloha v podobě doprovodných černobílých i celostránkových barevných obrázků usnadňujících zapamatování. Tuto část lze hodnotit jako velmi zajímavou a užitečnou, nicméně toto doplnění výrazně zvyšuje rozsah a především váhu slovníku, což je druhým výrazným znakem, ovšem vzhledem k tomu, že se jedná o slovník kapesního typu, je tento rys spíše negativní. Adresáty slovníku jsou studenti středních a jazykových škol, vzhledem k formátu slovníku se však zdá praktičtější slovník studovat spíše doma.

2.1.4. LINGEA

Tento obousměrný kapesní slovník s podtitulem „nejen na cesty“ obsahuje celkově 34 000 hesel a jedná se již o druhé vydání, které vyšlo roku 2008. Tato publikace byla zpracována autorským kolektivem pracovníků Lingea s.r.o., avšak bližší informace o autorech, kteří se na zpracování slovníku podíleli, slovník neudává. Chybí zde i struktura hesla znázorňující, jak se ve slovníkových heslech orientovat. Slovník podává alespoň vysvětlení metajazyka, tj. seznam použitých zkratk a poznámky k výslovnosti. V úvodu slovníku nalezneme také rejstřík a předmluvu. Dále zde najdeme orientační mapy a adresy na velvyslanectví a konzuláty ve Francii, Švýcarsku, Lucemburku a Kanadě. Slovník obsahuje jak konverzační, tak poměrně rozsáhlou gramatickou část, která se zabývá vybranými slovními druhy, ale i větnou stavbou. Součástí gramatické sekce je rovněž přehled nepravidelných sloves. V konverzační části potom najdeme všeobecné výrazy, ale i obraty z každodenního života či fráze zaměřené na cestování. Naproti tomu zde chybí seznam literatury, ze které autoři při vytváření slovníku vycházeli. Tak jako u dvou předchozích titulů, tj. slovníků FIN a OLOMOUC, i zde chybí orientační písmeno na okraji stránek a jsou vždy uváděna pouze první a poslední hesla na stránce. Hesla jsou řazena tradičně dle abecedy bez použití metody hnízdování. Dále si můžeme všimnout, že výskyt podhesel není příliš častý. Pokud jde o výslovnost, vyskytuje se sice pouze ve francouzské části, nicméně tento slovník jako jediný z našeho souboru užívá pro přepis znaky mezinárodní abecedy, a ne transkripci zjednodušenou. Tento výběr lze jistě hodnotit kladně, na druhou stranu je otázkou, zda uživatel, kterým, podle slov autorů, má být především cestovatel či středoškolský student, se dokáže rychle zorientovat v tom, jaký je rozdíl mezi [ɛ] a [e], popř. co znamená znak [ʒ] atd. Rovněž zde nalezneme

informaci o stylové příslušnosti daných hesel. Horní indexem jsou značeny homonymní výrazy a hvězdička u slovesa nás opět upozorňuje na jeho nepravidelnost.

Nakladatelství Lingea uvádí, že se jedná o „slovníky malé velikosti, ale velké rozsahem, patřícím k nejrozsáhlejším kapesním slovníkům na českém trhu.“¹⁵² S tímto tvrzením lze rozhodně souhlasit. Dodejme ještě, že z našeho souboru pěti slovníků se jedná o nejmenší, což je pro kapesní slovník zcela určitě výhodou. Slovník tak může být bez problému využit na cestách, v práci či při studiu, tak jak bylo autory zamýšleno.

2.1.5. LEDA

Praktický slovník kapesního formátu nakladatelství LEDA vyšel roku 2006 a jedná se o první vydání této publikace. Z obecných informací o slovníku se dozvíme jméno odpovědné redaktorky, kterou je PhDr. Jarmila Dušková. Slovník lektorovala Dr. Martha Trégouet. V úvodní části slovníku nalezneme předmluvu, v níž autoři mimo jiné uvádí, že při své práci vycházeli z francouzských výkladových slovníků *Le Petit Robert* a *Le Petit Larousse*. Dále slovník uvádí informaci o počtu hesel. Ve francouzské části je obsaženo 21 200 hesel, v české potom 24 600. Ačkoli je slovník obousměrný, autoři uvádí, že primárně je určen českému uživateli. Na vnitřních deskách slovníku je vysvětleno, jak rozumět jednotlivým znakům, popisům a typům písma. Najdeme zde i instrukce ke struktuře hesla, jakož i použité symboly a zkratky. Do slovníku však nebyl zařazen rejstřík. Slovník nabízí informace o francouzské výslovnosti a z gramatiky potom přehled nepravidelných sloves. Konverzační část však slovník neobsahuje. Kladně lze hodnotit značení orientačního písmene na okraji stránek, jakož i uvádění prvního a posledního hesla na stránce. K makrostruktuře dále dodejme, že je použito tradiční alfabetské řazení bez metody hnízdování a vyskytuje se zde minimum podhesel. Pokud jde o mikrostrukturu slovníku, můžeme zmínit, že transkripce je uváděna pouze ve francouzské části, a to pomocí počeštěných znaků pro přepis francouzské výslovnosti. Homonyma jsou značena číselným indexem a hvězdičkou u sloves jsou opět značena nepravidelná slovesa. Co se týče gramatické charakteristiky, je vyznačen pouze rod u substantiv. Slovník podává informace o stylistickém či oborovém zařazení hesel, a to pomocí zkratk.

¹⁵² LINGEA, s.r.o., 2008

Jedním ze specifických rysů tohoto slovníku je, že kromě současné slovní zásoby klade důraz také na hovorové a slangové výrazy. Zároveň však autoři uvádí, že jsou zde zastoupeny i základní odborné termíny. Slovník je určen primárně českému uživateli, což je důvod, proč je počet hesel v české části vyšší.

2.2 Korpus SYN2005

Jak již bylo zmíněno v empirické části, v naší studii bude pro srovnání použit korpus SYN2005. Tento korpus byl vybrán proto, že, jak již jeho název napovídá, byl zveřejněn v roce 2005. Většina zkoumaných slovníků, které tvoří součást našeho souboru, vyšla rovněž v tomto období, proto předpokládáme, že v slovní zásoba je v tomto ohledu pro srovnání odpovídající.

Jak jsme rovněž naznačili v úvodu empirické části, naše studie se opírá o analogickou rešerši M. Šulce¹⁵³, který však při svém výzkumu konfrontoval výsledky s frekvenčním seznamem získaným na základě korpusu staršího, a sice FSC2000. Složení korpusu FSC2000 je 60% publicistických textů, 25% odborné literatury a 15% beletrie. Tento poměr se od námi vybraného korpusu liší především v poměru publicistických textů, na druhou stranu korpus SYN2005 byl vybrán především vzhledem k jeho roku zveřejnění, a tedy s předpokladem, že by frekvenční seznam měl více odpovídat slovní zásobě obsažené v našem souboru slovníků.

¹⁵³ Šulc : Kapesní slovníky a frekvence lemmat v korpusu. 2008.

3. Metoda

Nyní, když jsme si představili materiál, se kterým budeme pracovat, uveďme si, jak jsme při studii postupovali.

Nejprve jsme provedli excerpci všech hesel, která začínají písmenem „m“ ze všech pěti vybraných slovníků. V případě slovníku FRAGMENT, kde je použita metoda hnízdování, byl však počet hesel podstatně nižší v porovnání s ostatními slovníky, proto jsme se posléze rozhodli v našem průzkumu zohledňovat hesla a podhesla dohromady, nikoliv analyzovat pouze hlavní hesla. Pokud bychom totiž pracovali pouze s hlavními hesly a při výpočtech nebrali v úvahu také podhesla, byly by výsledky naprosto odlišné, a to do té míry, že slovník s nejnižším počtem (pouze) hesel by byl slovník FRAGMENT. Při současné analýze se pořadí změnilo a výsledky jsou zároveň vyrovnanější.

Údaje získané pomocí excerpcce jsme poté zaznamenali do databáze a seznamy hesel a podhesel mezi sebou následně porovnali tak, abychom získali seznam hesel, která se vyskytují ve všech pěti slovnících. Tato hesla budeme dále nazývat hesla *nespecifická*. Druhou skupinu tvoří hesla *specifická*, tedy ta, která nejsou obsažena ve všech pěti slovnících.

Na základě údajů z korpusu SYN2005 byl vytvořen frekvenční seznam slov začínající na písmeno „m“. Tento frekvenční seznam jsme získali následujícím způsobem:

Nejprve jsme na webových stránkách Českého národního korpusu v nabízeném seznamu vybrali korpus SYN2005. Vyhledávání jsme vymezili pouze na lemmata, tj. základní tvar slova, který pod sebou seskupuje všechny tvary daného slova. Poté jsme do pole zadali m.*, což je tvar, kterým se dotazujeme na všechna lemmata o libovolné délce začínající písmenem m. Slovní druh jsme ponechali nespecifikovaný. Na základě tohoto dotazu korpus zobrazil 4 357 095 tokenů. Tato hesla jsme si zobrazili v seznamu dle nejvyšší frekvence a tento seznam, čítající 21 009 hesel, jsme následně uložili do .xlsx souboru. Na základě tohoto seznamu se následně odvíjel náš další výzkum.

Frekvenční údaje z korpusu byly zaznamenány do tabulky, která obsahuje jednak jednotlivá hesla, jednak jejich korpusovou frekvenci. Následně jsme porovnávali frekvenční seznam s celkovým seznamem hesel všech pěti slovníků, dále se seznamem hesel nespecifických i specifických. Při tom byly vyděleny skupiny slov, které se

vyskytují ve slovnících, ale v korpusu nejsou obsažena. Tato slova jsou tedy bez frekvence. Tak jako v rešerši Šulce, i my jsme sestavili skupinu slov, jejichž frekvence je vyšší než 1 000, přesto bychom je v žádném z pěti slovníků nenalezli. Seznamy těchto slov budou uvedeny v závěrečném shrnutí výsledků.

Nutno dodat, že jsme se při výzkumu potýkali s problémem jednoslovných a víceslovných hesel. Některá hesla, jak základní, tak podhesla, ve slovnících tvoří víceslovné lexémy, naproti tomu frekvenční seznam na základě korpusu SYN2005 zobrazuje pouze lexémy jednoslovné. Do naší rešerše jsme tedy zahrnuli pouze jednoslovné lexémy, které je tedy možno porovnávat s výsledky z korpusu.

3.1. Srovnání jednotlivých slovníků

Tak jako v analogické studii M.Šulce, i v našem případě byla prvním krokem výzkumu excerpcí heslových slov s prvním grafémem „m“ z pěti vybraných česko-francouzských kapesních slovníků (FRAGMENT, FIN, OLOMOUC, LINGEA, LEDA).

Druhým krokem bylo získání seznamu hesel, která jsou obsažena ve všech pěti slovnících, tj. hesel nespecifických, která o rozdílech mezi slovníky nic nevyovídají. Seznam těchto hesel čítá 250 slov.

Tabulka 2: Hesla specifická vs. hesla nespecifická a jejich podíly v jednotlivých slovnících (zaokrouhleno na celá čísla)

Slovník	Počet hesel od „m“	Nespecif. hesla	Podíl nespecif. hesel	Specif. hesla	Podíl specif. hesel
FRAGMENT	738	250	34%	488	66%
FIN	450	250	56%	200	44%
OLOMOUC	562	250	44%	312	56%
LINGEA	685	250	36%	435	64%
LEDA	523	250	48%	273	52%

Tabulka č. 2 ukazuje, jak velká část heslových slov u jednotlivých slovníků zahrnuje nespecifická hesla (34% – 56%). Slovník s nejvyšším podílem nespecifických hesel je FIN, neboť v porovnání s ostatními slovníky je ve slovníku FIN počet hesel začínajících písmenem „m“ nejnižší. Obdobně to platí i v případě specifických hesel, jejichž podíl je naopak nejvyšší u slovníku s nejvyšším počtem hesel. Zde jsou však rozdíly méně

markantní, neboť na druhém místě následuje slovník LINGEA, tedy slovník který má druhý nejvyšší počet hesel a rovněž i podíl specifických hesel je jen o dvě procenta nižší oproti slovníku FRAGMENT. To nás přivádí k druhému údaji, který lze z tabulky dále vyčíst, a sice počty hesel od písmene „m“. Zde je třeba připomenout, že jsme při rešerši pracovali pouze s jednoslovnými výrazy, a to z toho důvodu, abychom mohli data porovnávat s frekvenčním seznamem získaným na základě korpusu SYN2005, který uvádí pouze jednoslovné výrazy. Celkový počet hesel s počátečním písmenem „m“ by tedy ve skutečnosti byl ještě vyšší (FRAGMENT 877, FIN 729, OLOMOUC 675, LINGEA 797, LEDA 708). Nicméně pokud bereme v potaz pouze jednoslovné výrazy, slovník obsahující nejvíce hesel a podhesel s počátečním písmenem „m“ je FRAGMENT, následovaný slovníkem LINGEA. Naopak nejméně hesel začínajících písmenem „m“ obsahuje slovník FIN.

Tabulka č. 3 ukazuje, jak velký podíl v jednotlivých slovnících zabírají hesla začínající grafémem „m“. Opět byly zohledněny pouze jednoslovné lexémy. Podíly se pohybují v rozmezí 1,14 % – 2,01 %, v průměru je tedy heslům začínajících písmenem „m“ věnováno 1,66 % slovníku. Jak můžeme vidět v tabulce, tento průměr převyšují tři slovníky (FRAGMENT, OLOMOUC, LINGEA), naopak na průměr nedosahují dva slovníky, a sice FIN a LEDA, a to navzdory tomu, že slovník LEDA obsahuje ze zkoumaných slovníků nejvyšší počet stránek.

Tabulka 3: Počet hesel celkem, počet hesel začínajících „m“, podíl hesel od „m“ (zaokrouhleno na dvě desetinná místa)

Slovník	FRAGMENT	FIN	OLOMOUC	LINGEA	LEDA
Hesla celkem	37 500	34 000	30 000	34 000	45 800
Hesla „m“	738	450	562	685	524
Podíl	1,97 %	1,32 %	1,87 %	2,01 %	1,14 %

3.2. Frekvence vs. slovníková hesla

Třetím krokem naší rešerše bylo porovnávání slovníku z hlediska frekvence hesel, resp. na základě frekvenčního seznamu, který jsme získali v korpusu SYN2005. Díky těmto údajům byla vytvořena tabulka č. 4, která zobrazuje jednotlivá frekvenční pásma a heslová zastoupení jednotlivých slovníků v těchto pásmech. Můžeme si povšimnout, že

sloupec s celkově nejnižšími čísly, je sloupec s nespécifickými hesly, tj. těmi výrazy, které jsou zastoupeny ve všech pěti zkoumaných slovnících. V každém sloupci tabulky jsme označili tři pole s nejvyššími hodnotami pro daný sloupec. Na základě tohoto označení pak můžeme dojít ke zjištění, že pásma s nejvyšším zastoupením hesel jsou shodně pro všechny slovníky pásma s frekvencí 1000 – 2 000, vysoké hodnoty jsou potom i ve frekvenčním pásmu 100-199. Četné hodnoty se vyskytují taktéž ve frekvenčním pásmu 10 – 99. Výsledky jsou ovšem poněkud překvapivé, neboť v pásmech s frekvencí nad 1 000, tzn. u slov nejfrekventovanějších, jsou hodnoty nízké, v některých případech dokonce nižší, než u frekvenčního pásma 1-9. Takové výsledky nám podávají poměrně jasný důkaz o tom, do jaké míry se tvůrci zkoumaných slovníků řídili frekvenčními seznamy.

Tabulka 4: Počet hesel v jednotlivých frekvenčních pásmech

Frekvence	Nspecif. hesla	FRAGMENT	FIN	OLOMOUC	LINGEA	LEDA
nad 50 tis.	9	13	10	12	11	13
40-50 tis.	4	4	4	4	4	4
30-40 tis.	4	4	4	4	4	4
20-30 tis.	6	6	8	8	8	8
10-20 tis.	24	28	26	28	31	29
9-10 tis.	1	1	1	1	1	1
8-9 tis.	3	3	3	3	3	3
7-8 tis.	4	6	5	5	5	5
6-7 tis.	5	5	5	5	5	5
5-6 tis.	10	17	14	16	19	14
4-5 tis.	4	10	7	9	10	8
3-4 tis.	26	37	34	33	39	29
2-3 tis.	27	53	37	45	48	34
1-2 tis.	33	79	49	63	72	55
900-999	7	22	10	15	20	17
800-899	7	18	9	13	13	10
700-799	10	28	15	21	27	22
600-699	10	27	15	21	28	17
500-599	11	35	19	20	32	26
400-499	8	48	27	37	40	24
300-399	11	40	21	24	45	30
200-299	10	65	33	37	57	42
100-199	14	89	48	62	81	53
10-99	1	74	27	61	45	47
1-9	0	5	6	8	8	4
0 (bez frekvence)	1	22	12	8	30	22

Tabulka č. 5 pak podává zjednodušený přehled o frekvenčních pásmech a zároveň znázorňuje i procentuální podíl těchto pásem, aby bylo možné získat rychlý přehled o tom, do jaké míry byly při výstavbě slovníku brány v potaz frekvenční údaje. Poslední řádek tabulky potom zobrazuje počet a podíl lemmat, která se v korpusu nevyskytují vůbec, ale do daného slovníku byla přesto zařazena.

Tabulka 5: Shrnutí - lemmata v základních frekvenčních pásmech a jejich podíl pro jednotlivé slovníky (zaokrouhloeno na jedno des. místo)

Frekvence	Nespecif.hesla	FRAGMENT	FIN	OLOMOUC	LINGEA	LEDA
nad 10 000	47	55	52	56	58	58
	18,8 %	7,5 %	11,6%	9,9 %	8,5 %	11,1%
1 000 – 10 000	113	211	169	180	202	154
	45,2 %	28,6 %	30,1 %	32 %	29,5 %	29,4%
100 - 999	88	372	197	250	343	241
	35,2 %	50,4 %	43,8%	44,5%	50,1 %	45,9 %
1-99	1	79	33	69	53	51
	0,4 %	10,7 %	7,3 %	12,3 %	7,7 %	9,7%
0	1	22	12	8	30	22
	0,4 %	2,9 %	2,7%	1,4 %	4,4%	4,2 %

Na základě předpokladu, že kapesní slovníky by měly zahrnovat především slova s vysokou frekvencí, a naopak obsahovat jen minimum slov s nízkou frekvencí, nyní můžeme zhodnotit, do jaké míry je toto tvrzení pravdivé. V tabulce vidíme, že pásmo s nejvyššími hodnotami je to s frekvencí v rozmezí 100 – 999, následované pásmem s frekvencí 1 000 – 10 000. To je poněkud v rozporu s počátečním očekáváním, neboť nejvyšší hodnoty by se daly předpokládat pro pásma s frekvencí od 1000 výše. Ještě více překvapivý se jeví fakt, že pásmo s frekvencí 1 – 99 obsahuje podobné, někdy dokonce i vyšší hodnoty, než pásmo s frekvencí slov nad 10 000. Takový výsledek je zarážející a v rozporu s naší počáteční hypotézou. Nejlépe z tohoto hodnocení vychází slovník FIN, který nejlépe odráží nejfrekventovanější slova a zároveň má i nejnižší zastoupení poměru slov nejméně frekventovaných.

Celkově lze říci, že zkoumané výsledky vyvrací předpokládaný stav, čili námi zkoumané kapesní slovníky neodráží především nejfrekventovanější slova, ale naopak velmi často obsahují hesla s nízkou frekvencí, kterých by dle předpokladů mělo být obsaženo jen minimum. Následující rámeček přináší bližší údaje o posledním řádku předchozí tabulky, tj. o heslech bez frekvence. Jsou zde totiž zaznamenána hesla a podhesla (pouze

jednoslovné lexémy), která se nevyskytují v korpusu, přesto je lze však nalézt v některém z pěti slovníků. Tato hesla v korpusu nenajdeme nejčastěji z toho důvodu, že do slovníku byl zařazena v jiném gramatickém tvaru, např. plurál, femininum, popř. se jedná o variantu téhož lexému.

magnezismus	marketink	memoáry	mínus	mozolový
makaróny	masmédia	menší	miš-maš	mravy
maloobchodnice	masochizmus	mil.	mizantrop	muslimka
malovýrobci	max.	milenci	mladší	mydliny
manikér	medailon	min.	Moravan	mžikem

Pokud bychom se výsledky z výše uvedeného rámečku pokusili analyzovat, lze říci následující:

- a. Autoři zařadili do slovníků i několik zkratk (*max.*, *mil.*, *min.*), což pro uživatele může být přínosné, přestože se tyto výrazy v korpusu nevyskytují. Na druhou stranu v korpusu bychom nenašli lemma monolyže, do slovníků však toto heslo vybráno bylo, ačkoliv je jeho zařazení do takto malého slovníku poněkud překvapivé. Obdobně lze hodnotit i heslo *Moravan* či *mozolový*, jejichž užitečnost je v tak malém slovníku spíše pochybná.
- b. Zařazení hesel *menší* a *mladší*, tedy druhého stupně adjektiv je poněkud sporné, a to jednak vzhledem k velikosti slovníku, jednak z toho důvodu, že většina zkoumaných slovníků zahrnuje i gramatickou část, kde lze informaci o stupňování adjektiv dohledat. Podobně hodnotíme i uvádění lemmat ženského rodu, a sice *muslimka*, *maloobchodnice* uváděných jako hlavní heslo. Zde opět platí, že informaci o přechylování substantiv do ženského rodu lze dohledat v příslušné části slovníku, popř. by tato slova měla být spíše zařazena jako podheslo.
- c. Neobvyklé je i uvádění hesel v plurálu (*malovýrobci*, *mydliny*, *masmédia*, *milenci*, *makaróny*, *mravy*, *memoáry*), vezmeme-li v úvahu, že dané lexémy existují i v singuláru, avšak slovníky je v této formě často neuvádí. Na druhou stranu tvůrce slovníku k tomuto rozhodnutí vedl pravděpodobně fakt, že některá z těchto slov se v plurálu vyskytují častěji než v singuláru.

- d. Pokud jde o hesla *mínus*, *mizantrop*, *magnetizmus*, *marketink*, *masif*, *masochizmus*, jedná se o varianty lexému, v korpusu tato hesla najdeme, zadáme-li *minus*, *misanthrop*, *magnetizmus*, *marketing*, *masiv*, *masochismus*.

Na základě zjištěných poznatků lze konstatovat následující:

Můžeme zaznamenat, že počty lemmat s frekvencí nižší než 100 se pohybují v rozmezí od 33 do 79. Při prozkoumání těchto hesel si lze povšimnout, že některá náleží do kategorie okrajových hesel, nebo jejich přítomnost v malém slovníku, určeném primárně turistům a studentům středních škol či jazykových kurzů, není zcela adekvátní. Jedná se o následující případy:

- 1) Feminina přechýlená z maskulin a jejichž užitečnosti bychom vzhledem k rozsahu slovníku mohli pochybovat (*mulatka* FQ=79, *masérka* FQ=39, *maturantka* FQ=26, *muzikantka* FQ=18, *matematička* FQ=17, *maniačka* FQ=3).
- 2) Zařazení některých derivátů, konkrétně substantiv odvozených od sloves (*matení* FQ=42, *mstivost* FQ=13, *metení* FQ=2), resp. sloves odvozených od adjektiv (*měkčit* FQ=5) hodnotíme jako pochybné, a to zprvč vzhledem k rozsahu slovníku, zadruhé s ohledem na cílovou skupinu uživatelů, s čímž souvisí i klesající pravděpodobnost potřeby vyjádřit se právě tímto derivátem.
- 3) Kompozita, jejichž užitečnost je v takto malém slovníku pochybná (*místopřísežný* FQ=18, *maloburžoazie* FQ=11). Podobně i další lexémy, u nichž je otázkou, zda mohou být pro školáka nebo turistu potřebná (*mrva* FQ=87, *mirabelka* FQ=28)
- 4) Lexémy, které jsou pro zamýšleného adresáta (školáka, turistu) příliš odborné (*misál* FQ=98, *meliorace* FQ=88, *metalurgický* FQ=67, *modistka* FQ=63, *měchýřek* FQ=24, *mník* FQ=18, *monokiny* FQ=3).
- 5) Lexémy stylisticky vymezené či výrazně expresivní, které by měli nerodilí mluvčí používat opatrně, neboť vyžadují širší vysvětlení týkající se situačního užití (*megera* FQ=44, *matura* FQ=26, *mačo* FQ=3).

Dále lze ještě dodat následující zjištění:

- I. Pokud jde o zájmena, ve všech pěti slovnících, které byly předmětem našeho zkoumání, jsou zastoupena minimálně zájmena *my* a *můj*, popř. i jejich jednotlivá paradigmata (*moje, má, mé*), či paradigmata lexému *já* (*mi, mě*). Další zájmena uvádí pouze slovník LINGEA, a sice zájmena neurčitá *málokdo, málokde, máloco, málokterý*.

- II. Adverbia odvozená od adjektiv (*mírumilovně, materiálně, masově*). Jednak je otázkou, zda jsou v takto malém slovníku potřebná či užitečná, jednak by se překlad těchto adverbí dal odvodit od příslušných ekvivalentů adjektiv (vzhledem k tomu, že se nejedná o adverbia nepravidelná), a to na základě gramatické přílohy, kterou většina analyzovaných slovníků obsahuje.

- III. Co se týká slovních druhů obecně, podle očekávání jsou nejvíce zastoupené otevřené slovní druhy (podstatná jména, přídavná jména, slovesa, příslovce). Uzavřené slovní druhy (zájmena, číslovky, předložky, spojky, částice) jsou v menšině. Následující tabulky přináší přehled, která hesla ze skupiny uzavřených slovních druhů byla do slovníků zahrnuta (výsledky korpusu SYN2005):

Tabulka 6a: Zastoupení uzavřených slovních druhů ve slovnících – číslovky

číslovka	frekvence	FRAGMENT	FIN	OLOMOUC	LINGEA	LEDA
<i>mnoho</i>	41953	✓	✓	✓	✓	✓
<i>málo</i>	8046	✓	✓	✓	✓	✓
<i>moc</i>	5257	✓	✓	✓	✓	✓
<i>mnohokrát</i>	2192	✓	✗	✓	✓	✗
<i>mockrát</i>	536	✗	✗	✗	✓	✗
<i>mnohonásobný</i>	522	✗	✗	✗	✓	✗
<i>méně</i>	124	✓	✓	✓	✓	✓
<i>míň</i>	60	✓	✗	✗	✓	✗
<i>milionkrát</i>	53	✗	✗	✗	✗	✗
<i>milióntina</i>	47	✗	✗	✗	✗	✗
<i>miliónkrát</i>	45	✗	✗	✗	✗	✗
<i>miliardtina</i>	24	✗	✗	✗	✗	✗
<i>milióntý</i>	14	✗	✗	✗	✗	✗
<i>miliardkrát</i>	11	✗	✗	✗	✗	✗
<i>málokrát</i>	6	✗	✗	✗	✗	✗
<i>mnohotisíckrát</i>	2	✗	✗	✗	✗	✗
<i>méněkrát</i>	2	✗	✗	✗	✗	✗

Tabulka 6b: Zastoupení uzavřených slovních druhů ve slovnících - zájmena

zájmeno	frekvence	FRAGMENT	FIN	OLOMOUC	LINGEA	LEDA
<i>můj</i>	21 6336	✓	✓	✓	✓	✓
<i>málokdo</i>	2106	✓	✗	✗	✓	✗
<i>málokterý</i>	691	✓	✗	✗	✓	✗
<i>my</i>	311	✓	✓	✓	✓	✓
<i>máloco</i>	136	✗	✗	✗	✓	✗

Tabulka 6c: Zastoupení uzavřených slovních druhů ve slovnících – předložky

předložka	frekvence	FRAGMENT	FIN	OLOMOUC	LINGEA	LEDA
<i>mezi</i>	117800	✓	✓	✓	✓	✓
<i>mimo</i>	19332	✓	✓	✓	✓	✓
<i>místo</i>	15187	✓	✓	✓	✓	✓
<i>mimoto</i>	722	✗	✓	✓	✗	✓

Poslední poznámka se týká kategorie *proprií*. Na základě rešerše bylo zjištěno, že ve zkoumaných slovnících jsou uváděna vlastní jména jako např. *Marie, Markéta*. Pokud bychom předpokládali, že autora k tomuto výběru vedly údaje o nejčastěji užívaných jménech, zůstává však otázkou, proč do slovníku nezařadil rovněž i dvě nejčastější jména pro muže. Pokud jde totiž o vlastní jména mužská, jediné jméno v maskulinu, které v celkovém seznamu hesel všech pěti slovníků figuruje, je Mikuláš. To však hodnotíme jako specifické, vzhledem k jeho souvislosti se svátkem, jež se s tímto jménem pojí. Dále jsou z kategorie vlastních jmen uváděna i některá vlastní jména geografická (*Macao, Makedonie, Malajsko/Malajsie, Macocha, Maroko*). Nejhojněji jsou tato hesla zastoupena ve slovníku LEDA a LINGEA. Nicméně oba tyto slovníky obsahují přílohou část, v níž by uživatel měl naléznout geografické informace, proto se lze domnívat, že v hlavní části slovníku by již tato hesla nemusela být zastoupena a získalo by se tak místo pro jiná hesla, např. ta, která mají v korpusu frekvenci vyšší než 1 000, ale žádný z analyzovaných slovníků je nezařadil. Jejich seznam si nyní představíme.

Tabulka 7: Seznam slov s frekvencí vyšší než 1 000, která se nevyskytují v žádném z pěti slovníků

Nezařazené lemma	Frekvence	Nezařazené lemma	Frekvence
myslit	52142	music	1473
mm	7831	místostarosta	1404
městečko	4990	m^2	1403
mívat	3690	miss	1342
mj	3191	mariánský	1311
<i>modlit*</i>	3013	maličko	1210
metrum	2743	materiálový	1164
matčín	2021	monsieur	1153
metropol*	1970	madame	1145
ml	1969	mg	1132
myšlenkový	1641	mostecký	1073
mávnout	1572	meziročně	1033

*hesla jsou ve slovnících uváděna jako *modlit se*, resp. *metropole*

Na základě tabulky č. 7 si můžeme povšimnout, že co se týče slov s frekvencí vyšší než 1000 neobsažených v žádném z pěti zkoumaných slovníků, často se jedná o (délkové) míry: *mm*; *ml*; m^2 ; *mg* či zkratky (*mj*). Tato slova mohou figurovat v doplňkových či konverzačních částech slovníků, tudíž tato slova nebyla zařazena mezi hlavní hesla či podhesla.

Zdrobněliny (*městečko*; *maličko*) a cizí slova (*music*; *miss*; *monsieur*; *madame*, popř. i *metrum*) nebyly zařazeny, můžeme se domnívat, že autoři slovníků se tak rozhodli vzhledem k rozsahu slovníku a užitečnosti těchto slov z hlediska cílových mluvčích. Podobně je tomu v případě adjektiv jmen vlastních (*mariánský*; *mostecký*). Naproti tomu adjektiva *myšlenkový*; *materiálový* a zejména *matčín* by zařazena být mohla i do kapesních slovníků.

Pokud bychom výše uvedené výsledky chtěli vyjádřit v procentech, lze výsledky shrnout takto:

Z množiny 312 slov, která lze na základě frekvence hodnotit jako velmi frekventovaná (tj. slova s frekvencí nad 1 000), mají jednotlivé slovníky následující zastoupení:

Tabulka č. 8: Zhodnocení hesel s frekvencí vyšší než 1 000

Frekvence	FRAGMENT	FIN	OLOMOU C	LINGEA	LEDA
nad 1 000	266	221	236	260	212
	85 %	71 %	76 %	83 %	68 %

V tabulce vidíme, že nejlepší zastoupení má slovník FRAGMENT, který odráží 85 % nejfrekventovanějších slov. Naopak nejhůře je na tom dle výsledků slovník LEDA, který odráží pouze 68 %, jinými slovy tento slovník opomíjí 32 % nejfrekventovanějších slov, což je zvláště u malého slovníku, který by měl primárně obsahovat právě slova nejfrekventovanější, závažný nedostatek.

Závěr

Na základě údajů ve výše uvedených tabulkách lze konstatovat, že autoři žádného z analyzovaných slovníků se při jejich výstavbě neřídili pouze korpusovou frekvencí. Pokud by tomu tak bylo, nemohlo by se stát, že by se množství lemmat s korpusovou frekvencí nižší než 100 pohybovalo v rozmezí 7% – 12%, což jsou čísla markantní. Na druhou stranu je pravda, že toto rozmezí není mezi jednotlivými slovníky - poměrně k jejich velikosti - výrazně rozdílné (FRAGMENT 10,7 %, FIN 7,3 %, OLOMOUC 12,3 %, LINGEA 7,7 %, LEDA 9,7%). To platí i pro hodnoty u ostatních frekvenčních pásem. Můžeme tedy konstatovat, a to i s ohledem i na celkovou výstavbu a částí obsažených v jednotlivých slovnících, že hypotéza o podobnosti slovníků, kterou jsme zmínili v úvodu tohoto oddílu, byla v podstatě potvrzena.

Šulc se ve své studii domnívá, že takové množství lemmat s frekvencí nižší než 100 rozhodně není zanedbatelné a na vině je dle něj nedostatečná profesionalita autorů daných slovníků. S tímto tvrzením se v naší studii můžeme ztotožnit a dodat, že tvůrci slovníků korpusová data nezohledňují ani navzdory dnešní dostupnosti a stále širším možnostem využití korpusů. Šulc ve své studii rovněž poukazuje na to, že na vině jsou však i profesionální lexikografové, a to z důvodu jejich nezájmu o kapesní slovníky, přestože se jedná o nejrozšířenější typ překladových slovníků. Nabízí se však také úvaha, zda se nakladatelé v dnešní době nezačínají orientovat novým směrem a neinvestují spíše do slovníků elektronických či online slovníků, popř. různých mobilních aplikací, než aby doplňovali či aktualizovali tištěné kapesní slovníky.

Dále lze říci, že zohlednění korpusové frekvence a frekvenčního seznamu by sice mělo být prvním krokem při tvorbě slovníku, na druhou stranu je třeba brát v potaz, že frekvence není jediným východiskem při tvorbě slovníku, a proto by tento rys neměl být rozhodující. Jak jsme zmínili již dříve, při výstavbě slovníku je třeba zohlednit i další kritéria, jako je rozsah slovníku a cílový uživatel. Tato kritéria by pak měla být do slovníku zapracována jednak úpravou hesláře odvozeného z frekvenčního seznamu, jednak zařazením přílohových částí, kde budou uvedena potřebná doplnění (gramatická doplnění) a ucelený přehled odpovídající potřebám adresátů (konverzační sekce zaměřené na každodenní situace, turistiku, číslovky, barvy, dny v týdnu, měsíce aj.) V tomto ohledu většina zkoumaných slovníků účel uspokojivě splňuje.

Na základě zjištěných poznatků lze rovněž zmínit, že ne vždy platí poměr čím je slovník objemnější, tím i obsáhlejší. V našem souboru byl totiž s převahou nejobjemnější i vzhledově největší slovník nakladatelství OLOMOUC, avšak tento slovník nebyl nejobsáhlejší, alespoň co se hlavních hesel začínajících písmenem „m“ týče. Nejvíce hlavních lemmat bylo obsaženo naopak v nejmenším slovníku našeho souboru, a sice ve slovníku LINGEA. V tomto případě je přísloví „zdání klame“ velmi výstižné. Nicméně je třeba připomenout, že slovník FRAGMENT využívá při organizaci metody hnízdování, proto jsme v naší studii pracovali nejen s hlavními hesly, ale i podhesly, aby bylo výsledky možno srovnávat objektivněji.

Na závěr ještě dodejme, že označení „*kapesní*“ je poněkud zavádějící, jelikož počet hesel všech pěti vybraných slovníků se pohybuje kolem 35 000 (s výjimkou slovníku LEDA, kde je počet hesel 45 800), což odpovídá slovníkům středního rozsahu. Dále může pro uživatele představovat problém také fakt, že celkový počet hesel jednotlivých slovníků bývá uváděn nejednotně, tj. některé slovníky udávají informaci o celkovém počtu hesel, jinde se dozvíme počet hesel odděleně pro českou a francouzskou část, což může vést k nepřehlednosti. Kromě toho může být matoucí i fakt, že jednotlivá nakladatelství označují své slovníky různě (kromě označení *kapesní* slovník se můžeme setkat i s označením *praktický*, *šikovný*, *mini*, *velký kapesní*, popř. se se označení „*kapesní*“ v názvu neobjevuje vůbec), byť se stále jedná o slovník téhož typu.

Résumé

mots-clés

dictionnaire, corpus, fréquence, lemme, lexicographie

Les dictionnaires bilingues de format poche ont fait l'objet du travail présenté, qui porte plus particulièrement sur la question du choix des entrées. Dans cette perspective, le travail est constitué de deux parties principales, à savoir la partie théorique et la partie empirique. La première partie présente des remarques théoriques et réflexions sur le sujet, telles que la structure des dictionnaires. Ensuite la seconde partie intègre les observations auxquelles nous sommes arrivées dans la première partie. La seconde partie présente également une étude de dictionnaires choisis et les résultats qui s'ensuivent. Dans cette étude, nous avons cherché à déterminer dans quelle mesure la fréquence du corpus est reflétée dans ces dictionnaires en ce qui concerne le choix de lemmes et pour quelles raisons les auteurs de dictionnaires respectent ou non cette fréquence.

A cette fin, nous avons choisi cinq dictionnaires tchèque-français contemporains de format poche (concrètement de maisons d'édition FRAGMENT, FIN, OLOMOUC, LINGEA, LEDA). Quant aux données, nous avons choisi d'effectuer la recherche sur les entrées commençant par la lettre „m“. Pour cet objet, les listes d'entrées respectives ont été extraites des dictionnaires et comparées avec la liste tirée du corpus SYN2005.

Les résultats de notre recherche établissent que la fréquence n'a manifestement pas été le facteur décisif en ce qui concerne le choix de lemmes. D'un part, cela peut être dû au défaut de professionnalisme des auteurs des dictionnaires étudiés, c'est-à-dire à une connaissance insuffisante ou un manque d'inexpérience, en dépit de la disponibilité des corpus et les larges possibilités qu'ils offrent. D'autre part c'est aussi par manque de lexicographes professionnels et leur du fait de leur faible intérêt pour cette catégorie de dictionnaires, bien qu'il s'agisse des dictionnaires les plus répandus. Enfin, il faut également prendre en considération qu'à présent les lexicographes professionnels s'orientent plutôt vers les dictionnaires électroniques ou en ligne voire les applications pour les téléphones intelligents et que ces dernières sont devenues le nouveau centre d'intérêt prioritaire du développement et de l'investissement. Néanmoins nous avons conclu qu'en plus de la fréquence, d'autres traits jouent un rôle important lors du procédé de composition des dictionnaires. Il s'agit notamment de nombre de caractère de chaque

dictionnaire ainsi que de ses utilisateurs potentiels, leur connaissance et leur expérience linguistique supposées.

En plus du choix des lemmes, nous avons également évalué la structure globale de ces dictionnaires et nous avons pu constater qu'à cet égard les dictionnaires analysés semblaient satisfaisants.

Néanmoins nous voudrions formuler une dernière remarque concernant la dénomination des dictionnaires examinés. Il se trouve que même s'il s'agit de publications du même genre, les dénominations de ces produits linguistiques sont très diverses, selon les choix de leurs maisons d'édition respectives. Cela peut mener à la confusion et il serait donc souhaitable d'unifier ces dénominations.

Seznam použité literatury

Slovníky

ČESKO-FRANCOUZSKÝ, FRANCOUZSKO-ČESKÝ KAPESNÍ SLOVNÍK S VÝSLOVNOSTÍ. Nakladatelství OLOMOUC, Olomouc, 2004.

FRANCOUZŠTINA-ČEŠTINA, PRAKTICKÝ SLOVNÍK, LEDA, Praha, 2006.

FRANCOUZSKO ČESKÝ-ČESKO-FRANCOUZSKÝ KAPESNÍ SLOVNÍK. FIN Publishing, Praha, 2005.

FRANCOUZSKO ČESKÝ-ČESKO-FRANCOUZSKÝ KAPESNÍ SLOVNÍK, LINGEA, s.r.o., Brno, 2008.

VELKÝ KAPESNÍ FRANCOUZSKO ČESKÝ-ČESKO-FRANCOUZSKÝ SLOVNÍK. KPS-FRAGMENT, Praha, 2007.

Ostatní literatura

BRABCOVÁ, E.: *Česko-španělské a španělsko-české slovníky* tištěné. Bakalářská práce, Jihočeská univerzita, České Budějovice, 2010.

CVRČEK, V., KOVÁŘÍKOVÁ, D.: *Možnosti a meze korpusové lingvistiky*. Naše řeč, 94, 2011, č. 3, s. 113–133.

ČEJKA, M.: *Česká lexikologie a lexikografie*. Masarykova univerzita, Brno 1992.

ČERMÁK, F.: *JAZYK A JAZYKOVĚDA. Přehled a slovníky*. Karolinum, Praha, 2011.

ČERMÁK F., BLATNÁ R. (eds.): *Manuál lexikografie*. Nakladatelství H&H, Jinočany 1995.

ČERMÁK, F.: *Zpracování česko-cizojazyčného slovníku*. Metodické listy, ÚJOP 4, 1-16, 1989 (rozšířená verze Překladové lexikografie, in Manuál lexikografie 1995, 230-248)

ČERMÁK, F. – KLÍMOVÁ, J. – PETKEVIČ, V. (eds.): *Studie z korpusové lingvistiky*. Karolinum, Praha 2000.

ČERMÁK, F., KRÁLÍK, J., KUČERA, K.: *Recepce současné češtiny a reprezentativnost korpusu (výsledky a některé souvislosti jedné orientační sondy na pozadí budování Českého národního korpusu)*. Slovo a slovesnost, 58, 1997, č. 2, s. 117-124.

ČERMÁK, F- SCHMIEDTOVÁ, V.: *Český národní korpus – základní charakteristika a širší souvislosti*. Knochovnická revue, 15, č. 3. 2004.

FALTOVÁ, M.: *Francouzsko-české a česko-francouzské překladové terminologické slovníky z oblasti práva, administrativy a ekonomie: srovnávací studie*. Bakalářská práce, Jihočeská univerzita, České Budějovice, 2007.

FILIPEC, J.: *K úkolům české lexikologie*. Slovo a slovesnost, r. 29, 1968.

FILIPEC, J.: *Naše současná společnost, slovní zásoba a slovníky*. Naše řeč, č. 75, 1992.

FOŘTOVÁ, E.: *Slovensko-česká mezijazyková homonymie – Lexikální analýza na bázi Slovensko českého slovníku (U-Ž)*. Bakalářská práce, Univerzita Karlova, Praha, 2014.

HAVLOVÁ, F.: *Kapesní slovník cizích slov*. Naše řeč, ročník 38, č. 3-4, 1995.

HLADKÁ, Z.: *Lexikografie dnes a zítra*. Naše řeč, ročník 79, č. 3, 1996.

KAPUSTA, O.: *Česko-francouzské slovníkářství ve 20. století na pozadí vývoje lexikografie jako vědní disciplíny*. Diplomová práce, Univerzita Palackého, Olomouc, 2008.

KRÁLÍK, J.: *Vyvážení zdrojů Synchronního korpusu češtiny SYN2000*. Slovo a slovesnost, 62, 2001, č. 1, s. 38-5.

KRAUS, J.: *Frekvenční slovník češtiny*. Naše řeč, 53, 1970, s. 4-5.

NĚMEC, I.: *Vědecké a technické možnosti rozvoje české lexikografie*, Slovo a slovesnost, ročník 53 (1992), číslo 1, 1992, s. 48-55.

PFLEGEROVÁ, L.: *Moderní česko-italské a italsko-české slovníky*. Bakalářská práce, Jihočeská univerzita, České Budějovice, 2013.

SAKSOVÁ, V.: *Lexikografická a frekvenční analýza slovní zásoby pro česko-čínský technický slovník*. Olomouc, 2014.

ŠABRŠULA, J.: *Základy francouzské lexikologie*. SPN, Praha, 1983.

ŠEVČÍKOVÁ, P.: *Laický nářeční slovník z lexikografického hlediska*. Bakalářská práce, Masarykova univerzita, Brno, 2010.

ŠTNDLOVÁ, J.: *Retrogradní slovníky*, Slovo a slovesnost, ročník 21 (1960), číslo 4, str. 270-272

ŠULC, M.: *Kapesní slovníky a frekvence lemmat v korpusu*. Časopis pro moderní filologii, Roč. 90, 2008, s. 19-27.

ŠULC, M.: *Tematická reprezentativnost korpusů*. Slovo a slovesnost, 62, 2001, č. 1, str. 53-6.

TĚŠITELOVÁ, M.: *Sto let od vydání prvního frekvenčního slovníku*. Naše řeč, 80, č. 4.

ZÍSKAL, A.: *Příspěvky k lexikografické teorii a praxi*, Slovo a slovesnost, ročník 4, číslo 1, 1938 s. 19-27.

Internetové zdroje

AKADEMICKÝ SLOVNÍK SOUČASNÉ ČEŠTINY A SOFTWARE NÁSTROJ PRO JEHO TVORBU. Proměna jazyka a jeho výzkumu v době médií a technologií. Sborník z Konference studentů českého jazyka, ÚJČ 2013. [online] [cit. 2016-01-29]. Dostupné z: <<http://www.ujc.cas.cz/miranda2/export/sitesavcr/ujc/sys/galerie-download/sbornik-brling.pdf>>

CO JE SLOVNÍK? www.ajslovicka.cz [online]. [cit. 2015-10-10]. Dostupné z: <<http://www.ajslovicka.cz/clanky/anglicko-slovniky.html>>.

CO JE KORPUS? [www.ucnk.ff.cuni.cz](https://ucnk.ff.cuni.cz/co_je_korpus.php) [online] [cit. 2015-10-29]. Dostupné z: <https://ucnk.ff.cuni.cz/co_je_korpus.php>.

ČERMÁK, F.: Korpusová lingvistika : stručný historický přehled. In Český národní korpus [online]. Praha : Ústav Českého národního korpusu, 2004 [cit. 2015-10-29]. Dostupné z: <http://korpus.cz/doc/korp_lingv_prehled.rtf>.

ČERMÁK, F: Korpus, informace a lingvistika. [online]. In HASIL, J.: Přednášky z 48. běhu Letní školy slovanských studií. Praha, Filozofická fakulta Univerzity Karlovy, 2005 s. 15-24. [cit. 2015-10-29]. Dostupné z: <<http://korpus.cz/doc/korp-info-lingv.rtf>>.

ČERMÁK, F.: SÉMIOTIKA SLOVNÍKU. 2001, 114-125 [online]. [cit. 2015-10-13]. Dostupné z: <https://ucnk.ff.cuni.cz/doc/sem_slov.rtf>.

ČERMÁK, F., DOLEŽALOVÁ-SPOUSTOVÁ, D., HLAVÁČOVÁ, J., HNÁTKOVÁ, M. JELÍNEK, T., KOCEK, J., KOPŘIVOVÁ, M., KŘEN, M., NOVOTNÁ, R., PETKEVIČ, V., SCHMIEDTOVÁ, V. SKOUMALOVÁ, H., ŠULC, M., VELÍŠEK, Z.: SYN2005: žánrově vyvážený korpus psané češtiny. Ústav Českého národního korpusu FF UK, Praha 2005. [cit. 2015-10-29]. Dostupné z: <<https://www.korpus.cz>>.

HLADKÁ, Z., MARTINCOVÁ, O.: *Tradice a současnost české lexikografie*. [online]. [cit. 2015-10-13]. Dostupné z: <https://www.phil.muni.cz/cest/lide/hladka/CJA014_Tradice.rtf>.

JUNGMANN, J.: Slovník česko-německý. Praha, 1835. [online] [cit. 2016-02-07] Dostupné z: <http://kramerus.nkp.cz/kramerus/MShowPageDoc.do?id=4552304&mcp=&idpi=23998924&a_uthor=>>.

KRAUS, J: Frekvenční slovník češtiny

ŘAZENÍ SLOV VE SLOVNÍCÍCH. www.JTPunion.org. Cit. [cit. 2015-10-10].. Dostupné z: <www.jtpunion.org/spip/article.php3?id_article=2164>.

SCHMIEDTOVÁ, V.: Český národní korpus. Praha, 1999. [online] [cit. 2015-10-10].. Dostupné z: <<http://ikaros.cz/cesky-narodni-korpus>>.