

JIHOČESKÁ UNIVERZITA V ČESKÝCH BUDĚJOVICÍCH  
FILOZOFICKÁ FAKULTA  
ÚSTAV BOHEMISTIKY

DIPLOMOVÁ PRÁCE

**VYBRANÉ NÁŘEČNÍ JEVY V KORPUSU DIALEKT A JEJICH  
PREZENTACE VE VÝUCE ČEŠTINY**

Vedoucí práce: Mgr. Markéta Maturová, Ph.D.

Autorka práce: Bc. Kristýna Nachlingerová

Studijní obor: Bohemistika

Ročník: 2.

2022

Prohlašuji, že jsem autorkou této kvalifikační práce a že jsem ji vypracovala pouze s použitím pramenů a literatury uvedených v seznamu použitých zdrojů.

České Budějovice, 9. 5. 2022

---

Kristýna Nachlingerová

Na tomto místě bych ráda co nejsrdečněji poděkovala své školitelce, Mgr. Markétě Maturové, Ph.D., za čas, který mi věnovala nejen během psaní této diplomové práce, ale zejména za její vstřícný a ochotný přístup, jenž mě provázel celým studiem.

Dále děkuji Mgr. Haně Goláňové, Ph.D., z Ústavu Českého národního korpusu Filozofické fakulty Univerzity Karlovy za cenné rady a čas, který strávila nad touto prací, jež by bez ní nikdy nevznikla.

Nakonec děkuji své rodině za nekonečnou podporu během celého studia i mimo něj.

## **ANOTACE**

Tématem této diplomové práce je analýza vybraných nářečních jevů v korpusu DIALEKT a možnosti jejich prezentace ve výuce českého jazyka na středních školách. Teoretická část práce se zaměřuje na představení mluvených korpusů Českého národního korpusu, zejména pak na korpus DIALEKT a jeho charakteristiku, pozornost je věnována ale i zahraničním nářečním korpusům a mapovým aplikacím. Následně jsou popsány charakteristické jevy jihočeské a západočeské nářeční oblasti, právě na tyto oblasti se analýza dat v korpusu zaměřuje. Praktická část práce je pojata didakticky a skládá se z návrhů cvičení zaměřených na vybrané jevy, které poslouží k prezentaci oblastních nářečních rysů. Cílem práce je korpus DIALEKT nejen představit, ale zejména ukázat, jak jej lze pomocí vytvořených cvičení využít při výuce českého jazyka.

## **ANNOTATION**

The topic of this thesis is an analysis of chosen dialectical phenomena in the DIALEKT corpus and the possibilities of its presentation in Czech language lessons at high schools. The theoretical part focuses on the introduction of Czech National Corpus' spoken corpora, mainly on the DIALEKT corpus and its characteristics, however, it pays attention to foreign dialectical corpora and map applications as well. Afterwards, the distinctive phenomena of South Bohemian and West Bohemian dialectical region are described, the analysis of the data in the corpus is focused just on these two regions. The practical part is approached didactically and it consists of exercise suggestions focused on the chosen dialectical phenomena which should serve as a presentation of regional dialectical features. The aim of the thesis is to introduce the DIALEKT corpus and mainly to show, with the help of the exercises, how it could be used in Czech lessons.

# **OBSAH**

ÚVOD.....	9
METODOLOGIE .....	11
SEZNAM ZKRATEK .....	12
<b>1 MLUVENÉ KORPUSY .....</b>	<b>14</b>
1.1 Druhy mluvených korpusů.....	14
1.1.1 Obecné mluvené korpusy a jejich charakteristiky .....	14
1.1.2 Specializované mluvené korpusy a jejich charakteristiky .....	15
1.2 Výstavba mluvených korpusů.....	17
1.3 Transkripce v jednotlivých mluvených korpusech .....	19
1.4 Využití mluvených korpusů.....	21
1.4.1 Využití korpusu DIALEKT .....	22
1.4.2 Využití korpusu ORAL.....	23
1.4.3 Korpus SCHOLA2010.....	24
<b>2 KORPUS DIALEKT .....</b>	<b>25</b>
2.1 Charakteristika .....	25
2.1.1 Sběr dat .....	25
2.1.2 Složení korpusu.....	26
2.1.3 Dvouúrovňová transkripce.....	27
2.2 Nadstavbové aplikace založené na korpusu DIALEKT .....	30
2.2.1 Archiv diferenčních hlásek nářečí českého jazyka .....	30
2.2.2 Mapová aplikace pro korpusy mluvené češtiny .....	31
<b>3 ČESKÝ JAZYKOVÝ ATLAS .....</b>	<b>34</b>
3.1 Terénní výzkum .....	35
3.2 Zpracování materiálu .....	36
3.3 Elektronická verze atlasu .....	37

4	ZAHraniČNÍ NÁŘEČNÍ KORPUSY A APLIKACE DOSTUPNÉ PRO VEŘEJNOST.....	38
4.1	Databáze a archivy .....	38
4.2	Korpusy.....	40
4.2.1	Slovenský nářeční korpus .....	40
4.2.2	Německé nářeční korpusy.....	41
4.2.3	Norský nářeční korpus .....	42
4.3	Mapové aplikace .....	43
4.3.1	Mapová aplikace německých nářečí .....	43
4.3.2	Mapové aplikace polských nářečí.....	44
4.3.3	Mapová aplikace bulharských nářečí.....	46
4.3.4	Britská zvuková mapa.....	47
4.4	Jazykové atlasy .....	48
4.4.1	Německý jazykový atlas .....	48
4.4.2	Bulharský jazykový atlas .....	48
5	CHARAKTERISTIKA ZÁPADOČESKÉ A JIHOČESKÉ NÁŘEČNÍ OBLASTI	
	50	
5.1	Nářeční rozrůznění českého jazyka .....	50
5.2	Společné znaky západočeské a jihočeské nářeční oblasti.....	51
5.2.1	Charakteristické znaky západočeské nářeční oblasti .....	53
5.2.2	Charakteristické znaky jihočeské nářeční oblasti .....	55
5.3	Nářeční jevy západočeské a jihočeské v korpusu DIALEKT.....	56
6	CVIČENÍ ZAMĚŘENÁ NA NÁŘEČNÍ JEvy V KORPUSU DIALEKT.....	58
6.1	HLÁSKOSLOVÍ .....	65
6.2	TVAROSLOVÍ.....	94
6.3	SLOVNÍ ZÁSOBA.....	120
6.3.1	JÍDLO .....	120

6.3.2	SVÁTKY, ZVYKY, TRADICE.....	127
6.3.3	HOSPODÁŘSTVÍ.....	134
	ZÁVĚR.....	139
	SEZNAM POUŽITÝCH ZDROJŮ.....	143
	SEZNAM PŘÍLOH.....	151
	PŘÍLOHA – SEZNAM CVIČENÍ.....	152



## ÚVOD

Jazykové korpusy představují rozsáhlý zdroj materiálu z dob současných i minulých. Je to elektronický, moderní a jednoduchý způsob k zachování a zkoumání jak mluveného slova, tak i psaných textů. Materiál je neustále aktualizován, doplňován a ukládán na jedno místo, jež je dostupné všem, kteří mají zájem s ním pracovat. V České republice zpřístupňuje tyto texty od roku 1994 Český národní korpus, jehož provozovatelem je Ústav Českého národního korpusu a Ústav teoretické a počítačové lingvistiky Filozofické fakulty Univerzity Karlovy. Nepodílí se na něm však pouze odborníci z těchto ústavů, ale například i studenti z jiných vysokých škol, na nichž se práce s korpusy vyučuje, je hojně využívána právě k prezentování a zkoumání jazykového materiálu z různých perspektiv. Jazykové korpusy prezentují jazyk tak, jak se opravdu užívá nebo užíval, bez jakýchkoliv příkras. Předkládaná diplomová práce *Vybrané nářeční jevy v korpusu DIALEKT a jejich prezentace ve výuce češtiny* se proto opírá o korpus právě z důvodu rozsáhlé a autentické materiálové základny, která nabízí široké možnosti zkoumání.

Diplomová práce vychází z článku *Nářeční korpus DIALEKT a jeho využití ve výuce češtiny* Hany Goláňové a Martiny Waclawičové, které se ve velké míře na tvorbě tohoto korpusu podílejí, ale také z vlastního zájmu o dané téma a ze zkušenosti autorky této práce, která má možnost se na korpusu DIALEKT jakožto editorka podílet. Korpus DIALEKT obsahuje jazykový materiál zaměřený na nářeční promluvy, jedná se tedy o autentický mluvený jazyk. Má jak synchronní, tak diachronní charakter – umožňuje zkoumat současný stav nářečí, ale i jeho vývoj. Tato práce se zaměřuje na jihočeskou a západočeskou nářeční skupinu.

Práce je rozdělena do dvou částí a celkem 6 kapitol. První část, teoretická, obsahuje 5 kapitol popisujících obecně mluvené korpusy, jejich stavbu, zpracování dat a využití. Další kapitola popisuje detailně korpus DIALEKT, na jehož materiálu byl založen celý výzkum. Pozornost je věnována zejména způsobu sběru dat, složení korpusu a transkripci, tedy způsobu přepisu mluveného slova do psané podoby. Dále jsou zmíněny nadstavbové aplikace, které jsou na tomto korpusu založeny a s nimiž se dá různými způsoby při výzkumu nářečí pracovat. Třetí kapitola praktické části je zaměřena na Český

jazykový atlas, jenž zaujímá v dialektologické tradici významné místo, snahou je ukázat, jak probíhal výzkum pro získání materiálu, jak byl tento materiál zpracován a jak lze s atlasem pracovat. Další kapitola mapuje zahraniční jazykové atlasy, korpusy a další aplikace, stručně je popisuje a ukazuje možnosti jejich využití. Poslední kapitola teoretické části je spolu s kapitolou zaměřující se na vybraný korpus stěžejní pro celou diplomovou práci, zejména ale její praktickou část. Soustředí se na popis charakteristických nářečních rysů jihozápadočeské nářeční skupiny, jejích oblastí, typů a okrajových úseků. Tyto rysy jsou dále zpracovány právě v praktické části práce, která na tuto kapitolu navazuje a snaží se ověřit, zda jsou rysy opravdu charakteristické, v korpusu dochované a stále užívané u nejstarší generace mluvčích.

Praktická část práce se skládá z jedné rozsáhlé kapitoly a je pojata didakticky. Zaměřuje se na možnosti prezentace vybraných jihozápadočeských nářečních rysů ve výuce českého jazyka v korpusu DIALEKT pomocí cvičení. Úvodem kapitola ukazuje základy práce s korpusem ve vyhledávači KonText, následují jednotlivá cvičení zaměřená na vybrané nářeční jevy. Cvičení prezentují jejich užívání a zároveň různé způsoby práce s korpusem, ale i její limity.

Cílem předkládané diplomové práce je tedy popis mluvených korpusů poskytovaných Českým národním korpusem se zaměřením na nářeční korpus DIALEKT. Práce dále mapuje zahraniční nářeční korpusy, atlasy a aplikace, popisuje jihozápadočeskou nářeční skupinu, jejíž vybrané znaky jsou v praktické části práce vyhledávány a prezentovány jako cvičení využitelná ve výuce českého jazyka. Výsledky výzkumu jsou shrnuty v závěru práce a ukazují, nakolik je práce s korpusy vhodná a užitečná pro prezentaci jazykového materiálu z různých pohledů.

## METODOLOGIE

Předkládaná diplomová práce *Vybrané nářeční jevy v korpusu DIALEKT a jejich prezentace ve výuce češtiny* se dělí na 2 části, teoretickou a praktickou. První část obsahuje 5 kapitol, které teoreticky popisují mluvené korpusy, korpus DIALEKT, Český jazykový atlas, zahraniční nářeční korpusy a aplikace dostupné pro veřejnost, nakonec charakteristické rysy západočeské a jihočeské nářeční skupiny. Ke zpracování těchto kapitol byly využity jak knižní publikace, tak i časopisecké články, online slovníky a další dostupné zdroje. Informace získané ze studia těchto pramenů, které jsou uvedeny na konci práce v bibliografii, byly využity k praktické části, tedy analýze jazykového materiálu excerpovaného z mluveného nářečního korpusu.

Praktická část je založena na datech z korpusu DIALEKT, skládá se ze 3 podkapitol a zaměřuje se na tvorbu cvičení, pomocí nichž lze vysledované nářeční jevy z nahrávek prezentovat ve výuce. Konkrétně se cvičení zaměřují na západočeskou a jihočeskou nářeční skupinu. V první podkapitole je popsána metodologie tvorby cvičení, dále jsou uvedena cvičení samotná s podrobným obrázkovým postupem, v příloze je uveden seznam všech cvičení. Kromě korpusových dat jsou začleňována data z ČJA a SSJČ.

V závěru práce jsou teoretické poznatky syntetizovány s praktickými cvičeními a je zhodnoceno, nakolik jsou výsledky reprezentativní a směrodatné pro charakteristiku a další výzkum vybraných nářečních skupin. K citování zdrojů byla použita citační norma ČSN ISO 690 způsobem poznámek pod čarou, jejíž forma byla převzata z citačního portálu Citace PRO<sup>1</sup>. Práce využívá data, která poskytuje výzkumná infrastruktura LINDAT/CLARIAH-CZ (<https://lindat.cz>) podporovaná Ministerstvem školství, mládeže a tělovýchovy České republiky (projekt č. LM2018101).

---

<sup>1</sup> Citační portál Citace PRO je dostupný na adrese <https://www.citacepro.com/>.

## SEZNAM ZKRATEK

AGD	Archiv mluvené němčiny (Archiv für Gesprochenes Deutsch)
ASJ	Atlas slovenského jazyka
AV ČR	Akademie věd České republiky
aj.	a jiné
apod.	a podobně
ČJA	Český jazykový atlas
ČNK	Český národní korpus
ČR	Česká republika
DGD	Databáze mluvené němčiny (Datenbank für Gesprochenes Deutsch)
FF JU	Filozofická fakulta Jihočeské univerzity
FF MU	Filozofická fakulta Masarykovy univerzity
FF UK	Filozofická fakulta Univerzity Karlovy
KN-SNK	Korpus nářečí Slovenského národného korpusu
kol.	kolektiv autorů
mil.	milion
ms	milisekunda
např.	například
NESČ	Nový encyklopedický slovník češtiny
PF JU	Pedagogická fakulta Jihočeské univerzity
PSJČ	Příruční slovník jazyka českého
s.	strana
SAV	Slovenská akadémia vied
SNČJ	Slovník nářečí českého jazyka
SPJČ	Slovník pomístních jmen v Čechách
SPJMS	Slovník pomístních jmen na Moravě a ve Slezsku
SR	Slovenská republika
srov.	srovnej
SSJČ	Slovník spisovného jazyka českého
tis.	tisíc
tj.	to jest

tzv.  
ÚJČ

takzvaný  
Ústav pro jazyk český Akademie věd ČR, v. v. i.

# 1 MLUVENÉ KORPUSY

„Korpus mluveného jazyka (zkráceně též mluvený korpus) je jeden z typů korpusů, který se soustřeďuje na zachycení mluvy, mluveného jazyka. Zvukový záznam mluvy je v nich přepsán podle určitých transkripčních pravidel. Tento přepis (transkript) je zpravidla spojen se zvukovým záznamem.“<sup>2</sup> Korpusy mluvené i psané zpřístupňuje v České republice *Český národní korpus* (dále jen ČNK).

## 1.1 Druhy mluvených korpusů

ČNK umožňuje přístup k několika druhům mluvených korpusů, jedná se o korpusy synchronní, ukazující stav jazyka v určité době, tedy bez ohledu na historický vývoj.<sup>3</sup> Mluvené korpusy reprezentují zejména současnou mluvenou češtinu. ČNK disponuje ale i řadou korpusů specializovaných, které obsahují materiál starší. Všechny níže zmiňované korpusy lze vyhledávat pomocí nástroje KonText.

### 1.1.1 Obecné mluvené korpusy a jejich charakteristiky

Největšími obecnými mluvenými korpusy jsou ty z řady ORAL – ORAL2013, ORAL 2008 a ORAL2006. Sjednocený korpus ORAL ve verzi z roku 2017 obsahuje zhruba 5,4 mil. slov (tokenů<sup>4</sup>). Jedná se o korpus neformální mluvené češtiny, obsahuje nahrávky z celého území České republiky, nejvíce dat však pochází z její české části. Nahrávky pro tento korpus byly pořizovány v letech 2002–2011. Korpus ORAL vznikl opravením a spojením výše vypsáných korpusů, k nim byla navíc připojena data z korpusu ORAL-Z, jenž zachycuje na rozdíl od ostatních situace formální. Korpus navíc obsahuje lemmatizaci<sup>5</sup> a morfologické značkování, tzv. tagování, přičemž „pojetí lemmatu je širší, než je tomu u jazyka psaného. Prioritou je možnost vyhledat všechny tvary určitého slova, které mohou být zaznamenány s redukovanou výslovností, ale i nářeční tvary, jež by

---

<sup>2</sup> CVRČEK, Václav a Olga RICHTEROVÁ. *Pojmy:mluveny* [online].

<sup>3</sup> Korpus DIALEKT stojí na pomezí korpusu synchronního a diachronního, neboť částečně sleduje i historický vývoj mluvy. Je totiž koncipován ve dvou časových vrstvách a nejstarší nahrávky mají v podstatě diachronní charakter.

<sup>4</sup> Za token je v korpusové lingvistice většinou považováno jedno grafické slovo, respektive jeho konkrétní realizace, jedná se o nejmenší jednotku textu.

<sup>5</sup> Lemma představuje zpravidla slovníkovou podobu hesla, v lemmatizovaných korpusech se jednomu slovnímu tvaru či skupině tvarů přiřazuje lemma.

mohly být samotné lemma.“<sup>6</sup> Systém tagování je stejný jako u korpusů psaných. Sociolingvistické charakteristiky jsou u korpusů řady ORAL velmi podrobné, mezi strukturální atributy patří pohlaví, nejvyšší dosažené vzdělání, věk, nářeční oblast současného pobytu a další.

ORTOFON je druhým největším obecným korpusem mluveného jazyka, navazuje na korpusy řady ORAL, tedy na korpusy neformální mluvené češtiny. Mluvčí v nahrávkách dialogického charakteru pocházejí opět z celé České republiky a jejich projevy jsou spontánní, nepřipravené. ORTOFON se však od korpusů skupiny ORAL odlišuje víceúrovňovou transkripcí – přepis probíhá na rovině ortografické, fonetické a navíc ještě metajazykové<sup>7</sup>. V roce 2020 byla zveřejněna druhá verze korpusu, která nyní obsahuje zhruba 2,1 mil. slov (tokenů) dohromady v 615 nahrávkách. Druhá verze korpusu, na rozdíl od té první, není vyvážená, neboť se snaží zpřístupnit co nejvíce materiálu, nehledí tedy na vyvažování v rámci sociolingvistických kategorií (pohlaví, věk, dosažené vzdělání, nářeční oblast pobytu do 15 let) jako verze první.<sup>8</sup>

Posledním z obecných mluvených korpusů češtiny je korpus ORATOR o velikosti asi 580 tis. slov (tokenů). Nahrávky z let 2005–2019 obsahují monologické projevy, na které byli rodilí mluvčí dopředu připraveni (přednášky, přípitky, uvítání hostů). Stejně jako ORAL je i tento korpus lematizován a morfologicky značkován, má taktéž stejná pravidla pro přepis, který je jednoúrovňový, zachycený ortografickým způsobem.<sup>9</sup> Vzhledem k zaměření se zde vyskytují strukturální atributy, které u jiných korpusů nenajdeme – typ situace, zaměření projevu, přítomnost publika, speciální příznaky projevu a další.<sup>10</sup>

### **1.1.2 Specializované mluvené korpusy a jejich charakteristiky**

ČNK zpřístupňuje řadu korpusů specializovaných, z nichž nejobsáhlejší mluvený korpus je SCHOLA2010 o velikosti přibližně 790 tis. slov (tokenů). Jedná se o korpus

---

<sup>6</sup> CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:lemtag\_mluv* [online].

<sup>7</sup> Metajazyková rovina zaznamenává doplňující neverbální zvuky mluvčího či okolní zvuky v nahrávce slyšitelné.

<sup>8</sup> CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:ortofon* [online].

<sup>9</sup> CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:orator* [online].

<sup>10</sup> CVRČEK, Václav a Olga RICHTEROVÁ. *Pojmy:atributy\_strukturni* [online].

zaznamenávající mluvený jazyk vyučovacích hodin, obsahuje mluvu žáků a učitelů během vyučování, jde tedy o materiál z prostředí formálního. Nahrávky byly pořizovány v letech 2005–2008 na různých školách v České republice, jsou zde patrné prvky obecné češtiny a v některých oblastech značný vliv dialektů. Do nahrávání byly zahrnuty třídy základních a středních odborných škol, dále také víceletých i čtyřletých gymnázií. Pořizování nahrávek probíhalo v následujících předmětech – český jazyk a literatura, matematické a přírodovědné předměty, společenskovědní a výchovné předměty, nakonec informatika, technické a profesně-pracovní vyučování. Korpus je vzhledem k počtu nahrávek jednotlivých předmětů vyvážený.<sup>11</sup>

PMK neboli Pražský mluvený korpus je dalším ze specializovaných mluvených korpusů a zároveň prvním korpusem mluvené češtiny, obsahuje zhruba 675 tis. slov (tokenů). Zachycuje mluvenou češtinu z Prahy a jejího okolí. Korpus obsahuje promluvy formální i neformální, v nahrávkách vystupují jako mluvčí (tzv. informátoři) muži i ženy ve věku od 20 let se vzděláním základním, středoškolským i vysokoškolským. Formální promluvy mají charakter monologu odpovědi na obecné otázky kladené exploraátorem (nahrávajícím mluvčím) – otázky mají nespisovně-spisovnou povahu, aby nebyly odpovědi nijak ovlivněny. Neformální promluvy mají pak formu dialogu mezi dvěma mluvčími na vlastní a předem blíže neurčené téma. PMK je dosud bez lemmatizace a morfologického značkování, proto je potřeba při práci s tímto korpusem na to brát zřetel.<sup>12</sup>

BMK neboli Brněnský mluvený korpus o velikosti asi 490 tis. slov (tokenů) je prvním mluveným korpusem z moravské části České republiky, obsahuje 250 nahrávek z let 1994–1999 a prezentuje specifickou mluvu města Brna. „V běžné mluvě Brňanů se prolíná zejména středomoravský interdialekt s pronikající obecnou češtinou [...], v oblasti slovní zásoby jsou patrné relikty někdejšího soužití brněnské češtiny s německým jazykem a vliv brněnského slangu (hantecu).“<sup>13</sup> Korpus vznikl podobně jako PMK, aby byl sociolingvisticky vyvážený, tj. obsahoval nahrávky formální i neformální, v nichž mluví muži i ženy různých věkových kategorií a různého vzdělání. Promluvy zde mají podobnou charakteristiku jako v PMK, avšak formální otázky byly kladeny spisovně. Neformální promluva měla opět formu dialogu dvou nebo více lidí,

---

<sup>11</sup> CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:schola2010* [online].

<sup>12</sup> CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:pmk* [online].

<sup>13</sup> CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:bm* [online].



„jeden z účastníků dialogu byl zpravidla i respondentem ve formální promluvě, což umožňuje sledovat rozdíly češtiny užívané v neoficiálních a polooficiálních situacích.“<sup>14</sup> I přes pokusy BMK morfologicky označkovat je tento korpus dosud přístupný bez tagování.

Korpus prezidentských projevů SPEECHES je dalším z mluvených korpusů češtiny. Stojí na pomezí psaného korpusu SYN (zachycuje veřejný psaný jazyk) a mluveného korpusu ORAL (neformální a nepřipravený jazyk), neboť se zaměřuje na připravené texty určené k přednesu. Korpus o velikosti něco přes 215 tis. slov (tokenů) zachycuje 151 prezidentských projevů z let 1918–2015, nahráváno bylo celkem 14 prezidentů. Obsahuje několik strukturních atributů, kterými projevy podrobně popisuje (mluvčí, datum pronesení, titulky, příležitost pronesení projevu apod.), navíc je lemmatizován a morfologicky označován podobně jako korpus SYN2010.<sup>15</sup>

LINDSEI\_CZ je předposledním z řady specializovaných mluvených korpusů češtiny a obsahuje nahrávky spontánní mluvené angličtiny pokročilých žáků s češtinou jako prvním jazykem. Korpus vznikl jako část mezinárodního projektu LINDSEI, který „doplňuje psaný žakovský korpus International Corpus of Learner English (ICLE) o korpus mluvené pokročilé žakovské angličtiny.“<sup>16</sup> Celý korpus se skládá z 11 subkorpusů, každý z nich je zaměřený na jeden jazyk a obsahuje 50 rozhovorů. Český subkorpus o velikosti přibližně 120 tis. slov (tokenů) byl vytvářen v letech 2012–2015 a obsahuje nahrávky studentů 3. a vyšších ročníků anglistiky z FF UK.

DIALEKT je posledním a zároveň nejmenším ze specializovaných mluvených korpusů češtiny, avšak pro tuto práci je stěžejní, neboť právě z jeho dat diplomová práce vychází. Podrobně je korpus popsán v kapitole č. 2 od s. 25.

## 1.2 Výstavba mluvených korpusů

Mluvené korpusy nejsou vystavěny dle jednotné struktury a jejich uspořádání se liší. „Každý korpus se tak člení na různé strukturní jednotky a ke každé jednotce (v závislosti

---

<sup>14</sup> Tamtéž.

<sup>15</sup> CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:speeches* [online].

<sup>16</sup> CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:lindsei\_cz* [online].

na anotaci) náleží jeden nebo více strukturních atributů.“<sup>17</sup> Strukturní jednotky mluvených korpusů jsou taktéž odlišné od jednotek korpusů psaných, a to kvůli rozdílné povaze dat.

V mluvených korpusech ČNK jsou užívány tři strukturní jednotky, a sice *doc* – dokument, ucelená sonda obsahující dialog; *sp* – jednotka obsahující informace o mluvčím<sup>18</sup>; *seg* – jednotka členění mluveného textu v rámci *sp*. Jako strukturní jednotku lze obecně pojmut i *pozici* – jednotka korpusu v rámci tokenizace<sup>19</sup> – běžně se však o ní jako o strukturujícím prvku neuvažuje. Strukturní jednotky obsahují různé strukturní atributy. Všechny jednotky obsahují atribut *id*, ten slouží jako identifikátor textu. Mnohé atributy se u mluvených korpusů liší.

Korpusy řady ORAL mají společné následující základní atributy – v jednotce *doc* rok nahrání sondy, počet mluvčích v sondě, typ promluvy; v jednotce *sp* číselné označení mluvčího v rámci sondy, pohlaví, věková kategorie, věk, vzdělání, nářeční oblast místa pobytu mluvčího v dětství. Spojený korpus ORAL obsahuje navíc další řadu atributů, které se vztahují k nahrávce (*doc*), například označení formálnosti situace, zda byl rozhovor dopředu připraven, zda je k transkriptu nahrávky dostupný zvuk, vztah mluvčích v nahrávce či za jaké situace došlo k hovoru. Dále se zde objevují atributy vztahující se k mluvčímu (*sp*) – automaticky přiřazená přezdívka, pohlaví, nejvyšší dosažené vzdělání, věk, nářeční oblast současného pobytu a pobytu v dětství, zaměstnání aj.

Korpusy s víceúrovňovou transkripcí opět obsahují jiné strukturní atributy než korpusy výše zmíněné. Korpus ORTOFON v jednotce *doc* obsahuje jako atributy počet generací mluvčích v nahrávce a pohlaví. Podstatně více atributů se ale objevuje v jednotce *sp* – velikost sídla, kde mluvčí žije v současné době, kde žil v dětství a kde žil nejdelší dobu, obor vzdělání mluvčího aj. Tento korpus navíc obsahuje strukturu – *para*, do které jsou zahrnuta slova či skupiny slov vyslovených s určitým doprovodným rysem (nahlas, šeptem apod.), dále strukturu *pw*, která obsahuje pozice tvořící jeden přízvukový takt,

---

<sup>17</sup> CVRČEK, Václav a Olga RICHTEROVÁ. *Pojmy:atributy\_strukturni* [online].

<sup>18</sup> Obvykle se strukturní jednotka *sp* nazývá jako promluva, avšak nelze užití tohoto pojmu ztotožňovat s jeho tradičním lingvistickým vymezením. Hranice jednotky *sp* odpovídá hranicím replik – souvislých promluv jednoho mluvčího.

<sup>19</sup> Tokenizací se rozumí proces rozdělení textu na tzv. tokeny – nejmenší jednotky textu, grafická slova.

strukturu *overlap* označující pozice řečené v překryvu s mluvou jiného mluvčího, nakonec strukturu *unclear*, jež označuje pozice, u kterých si editor při transkripci nebyl jist jejich obsahem. Atributy korpusu DIALEKT jsou podrobně rozebrány ve druhé kapitole.

Atributy korpusu ORATOR, který je zaměřen na dopředu připravené monologické projevy, odpovídají zaměření korpusu a jsou následující – rok pořízení nahrávky, místo nahrávky, typ situace (ceremoniál, jednání, přednáška aj.), zaměření projevu (oficiální, politické apod.), žánr projevu, přítomnost publika, komu je projev určen, speciální příznaky projevu, zdroj nahrávky a další (tyto atributy se vztahují k jednotce *doc*). O mluvčím (jednotka *sp*) zde pak najdeme atributy označující přezdívku, pohlaví či procentuální zastoupení řeči.<sup>20</sup>

### 1.3 Transkripce v jednotlivých mluvených korpusech

Transkripce neboli přepis mluvené řeči písmem probíhá u každého z mluvených korpusů trochu jiným způsobem. Různě se přepisují některé jazykové jevy i interpunkce, zavádějí se navíc další znaky a symboly.

Transkripce nahrávek v korpusech řady ORAL byla upravena a sjednocena, někde se ale stále objevují odchylky. Z jazykových jevů je transkripce sjednocená u všech korpusů této řady v přepisu ustálené zjednodušené výslovnosti (*vemu*, *neska*, *myslim*). Dle výslovnosti se přepisují slova začínající na *j-* (*méno/jméno*), slova s protetickým *v-* (*vokno*), souhlásková skupina *šť'* (*ešče*) a nářeční varianty, které lze v korpusu taktéž nalézt (*zme*, *vo našom*). Dle pravopisu se ve všech korpusech řady ORAL přepisuje asimilace znělosti, pokud se jedná o pravidelné jevy bez variant (*dub*, *sbírat*). Co se kvantity týče, zachycuje se krácení i dloužení v základech slov i v koncovkách (*rohlik*, *klucí*) ve všech korpusech.<sup>21</sup>

Ze zásad pro přepis v korpusech řady ORAL vycházejí pravidla pro přepis v korpusu SCHOLA2010, která však bylo nutné upravit, aby mohly být zaznamenány například

---

<sup>20</sup> CVRČEK, Václav a Olga RICHTEROVÁ. *Pojmy:atributy\_strukturni* [online].

<sup>21</sup> CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:oral:pravidla* [online].

i nářeční jevy. V souladu s pravidly pravopisu se zaznamenávají hlásky *y, i* (přestože by bylo možné zaznamenat i tvrdou výslovnost *y/ý* v nahrávkách východomoravských) a asimilace znělosti, zdvojené hlásky, cizí vlastní a obecná jména. Důsledně dle výslovnosti jsou však zaznamenávány regionálně příznakové případy asimilace, které lze i v tomto korpusu najít (*abyzme, zme, zhodné* atd.). Dále se dle reálné výslovnosti zaznamenává artikulační asimilace a zjednodušování souhláskových skupin, výslovnost protetického *v-* či kvantita samohlásek (zachycováno je typické východomoravské a slezské krácení, taktéž je podle dialektologických pravidel zapisováno původní dlouhé *ú* typické pro východomoravskou oblast).<sup>22</sup>

Transkripce v korpusu ORTOFON probíhá na několika rovinách, stejně tak v korpusu DIALEKT. K přepisu nahrávek se používá transkripční program ELAN, který umožňuje propojit několik rovin přepisu přímo se zvukem. Přepis probíhá na rovině ortografické, fonetické a metajazykové. Ortografická rovina se snaží zápis co nejvíce přiblížit pravidlům českého pravopisu, avšak jsou zde od pravidel jisté odchylky, dle skutečné realizace se zachycuje např. protetické *v* nebo koncovky (*s malejma dětma*). Na rovině fonetické probíhá přepis přesně dle toho, co v nahrávce zazní. Neužívá se zde ale běžných znaků mezinárodní fonetické transkripce, nýbrž objevují se symboly běžné v dialektologických prepisech, pomocí nichž jsou zaznamenávány hláskoslovné nářeční jevy. Mezi takové symboly patří například tvrdé, měkké a dlouhé slabikotvorné *l* (*l, l', l̃*), dlouhé slabikotvorné *r* (*r'*), výjimečně se zaznamenává i výslovnost tvrdého *y/ý*, obouretného *v* (*w*) a palatalizovaného *s, z, dz, c* (*ś, ź, dź, ć*) (typické pro slezskopolskou nářeční podskupinu). Poslední rovinou přepisu je rovina metajazyková, ta zachycuje neverbální projevy mluvčích a ostatní okolní zvuky.<sup>23</sup> Víceúrovňové transkripci v korpusu DIALEKT se věnuje podkapitola č. 2.1.3 na straně 27.

Přepis v korpusu LINDSEI\_CZ vychází z transkripčního manuálu vydaného Centrem pro anglickou korpusovou lingvistiku, jedná se o přepis ortografický, přizpůsobený pravidlům pravopisu.<sup>24</sup>

---

<sup>22</sup> CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:zasady\_pro\_prepis\_vyucovacich\_hodin* [online].

<sup>23</sup> CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:ortofon:pravidla* [online].

<sup>24</sup> CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:lindsei\_cz* [online].

Transkripce v korpusech PMK a BMK je prováděna stejným způsobem. Přepis usiluje o zachycení co nejvěrnější podoby mluveného jazyka, není zde však užíván přepis fonetický, dochází ke kombinaci pravopisných pravidel a fonetického zápisu.<sup>25</sup>

#### 1.4 Využití mluvených korpusů

Mluvené korpusy poskytují velmi široké využití zejména díky autentickému materiálu, který poukazuje na opravdové užívání mluvy v České republice. Slouží však nejen k záznamu mluveného jazyka tak, jak se v dané době používá, ale hlavně ke zkoumání řeči. Mluvené korpusy jsou předmětem analýz jazyka jako takového na rovině foneticko-fonologické, morfologicko-syntaktické, lexikálně-sémantické i rovině pragmatické<sup>26</sup>, nicméně lingvistické analýzy nejsou jediným způsobem, jak se dají tyto korpusy využít. Data z mluvených korpusů mohou být podkladem pro tvorbu slovníků, např. SNČJ (*Slovník nářečí českého jazyka*), a mohou být srovnávána s jazykovými atlasy. Korpusy zachycující čtený připravený text (např. korpus ORATOR a SPEECHES) napomáhají ke zdokonalování programů k rozpoznávání řeči.<sup>27</sup>

Korpusy psané i mluvené jsou v posledních letech stále více zařazovány do výuky češtiny i cizích jazyků. On-line prostředí a přívětivé webové rozhraní je pro studenty velmi atraktivní a výuku zpěstří, proto je práce s nimi tak přitažlivá. Zapojení techniky do výuky však není zdaleka jediným důvodem, proč je užívání korpusů ve výuce užitečné.

Hlavní výhoda zařazení korpusů do výuky spočívá zejména v tom, že poskytují naprosto autentický a reprezentativní materiál. Lze z nich tedy poznat, jak se reálně jazyk užívá, ne jak by se užívat měl. Dalším plusem je množství materiálu, ke kterému máme na jednom místě přístup. Práce s různými korpusovými nástroji umožňuje analyzovat jazyk do hloubky a po správném zadání dotazu během několika vteřin zobrazit výsledky vyhledávání, odpadají zde tedy zdoluhavé analýzy a práce s odbornou literaturou. Díky tomu, že korpus zobrazuje slova v kontextu, můžeme také význam těch, které neznáme, odhadnout právě ze slov, jež jej obklopují. Korpus však může být pouze zajímavým

---

<sup>25</sup> CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:bmK* [online].

<sup>26</sup> Z nahrávek lze vypořadovat, jak účastníci nahrávání užívají řeč v sociálním kontextu.

<sup>27</sup> CVRČEK, Václav a Olga RICHTEROVÁ. *Pojmy:mluveny* [online].

zpestřením a rozhodně není nutné, aby nahradil všechny stávající metody výuky, učitel může například využít korpusová data k tvorbě vlastních výukových materiálů.

Mluvené korpusy poskytují uživatelům nahrávky a přepisy autentického běžně užívaného jazyka, ve kterých je možné vyhledávat spoustu jazykových jevů. Zařazení práce s mluvenými korpusy do výuky přinese studentům přístup k velkému množství materiálu, který by byl ve své povaze jinde jen těžko sehnatelný. V podkapitolách níže je představeno, jaké korpusy mluveného jazyka lze během výuky využít a co na nich pomocí korpusových nástrojů ukázat.

#### 1.4.1 Využití korpusu DIALEKT

Specializovaný korpus DIALEKT obsahující nahrávky ze všech nářečních oblastí České republiky sice není zatím tak velký jako obecné mluvené korpusy, nicméně jeho specializace umožňuje zařazení do výuky nářeční problematiky. Díky tomu „se nářeční korpus může stát užitečnou a názornou pedagogickou pomůckou, kterou lze využít při přípravě výuky nebo pro zadání domácího cvičení, které studenti mohou řešit ať už tradičně na papíře nebo s využitím počítače.“<sup>28</sup> Vyhledávat lze jak nářeční slovní zásobu a tvořit například i frekvenční seznamy, ale také jevy hláskoslovné, tvaroslovné a v některých případech i syntaktické.

Nejen korpus DIALEKT lze využít k prezentaci nářečí ve výuce, stejně dobře mohou posloužit i nadstavbové aplikace na něm založené. Pomocí barevně zpracované *Mapy nářečí České republiky* se studenti snadno naučí nářeční oblasti spolu s jejich podskupinami, úseky a typy. K tomu může pomoci i *Mapka – Mapová aplikace pro korpusy mluvené češtiny*, která je na *Mapě nářečí České republiky* založená. Z té se studenti navíc mohou dozvědět i charakteristiky všech nářečních oblastí, mohou si pustit nahrávky či vyhledat své město a tím zjistit, do které nářeční oblasti a podskupiny spadá. Pokud si student pustí nahrávku a zobrazí její přepis, setká se s nářečními hláskami, které pak může vyhledat v *Archivu diferenčních hlásek nářečí českého jazyka*, aby zjistil, o jakou hlásku se přesně jedná a kde ještě se s ní může setkat.

---

<sup>28</sup> GOLÁŇOVÁ, Hana a Martina WACLAWIČOVÁ. Nářeční korpus DIALEKT a jeho použití ve výuce češtiny, s. 132.

Návrh cvičení, pomocí nichž lze korpus DIALEKT do výuky začlenit a vyhledávat tak různé nářeční jevy, je obsažen v praktické části této práce, viz kapitola 6 od strany 58.

#### 1.4.2 Využití korpusu ORAL

Sjednocený korpus ORAL (obsahující korpusy ORAL2006, ORAL2008 a ORAL2013) je největším obecným mluveným korpusem. Obsahuje nahrávky neformálních rozhovorů rodilých mluvčích českého jazyka z celého území České republiky. Představuje tedy češtinu tak, jak se opravdu užívá, autentickou a užívanou v přirozeném prostředí mluvčích. Díky povaze dat lze zkoumat jak obecnou češtinu (nahrávky z oblasti českých nářečí v užším smyslu poskytuje korpus ORAL2006 a ORAL2008), tak i některé morfologické nebo lexikální jevy teritoriálních dialektů (korpus ORAL2013 již obsahuje nahrávky z území celé České republiky).<sup>29</sup>

Vyhledávat v korpusech řady ORAL lze nejen pomocí nástroje KonText, data jsou navíc zahrnuta i do průzkumníku variant SyD, ve kterém reprezentují právě mluvenou češtinu. Jeho zařazení je vhodné zejména při výuce cizinců, pro jejichž porozumění je nutné znát varianty užívané v přirozeném mluveném jazyce. Nabízí se zde otázka, kdy obecnou češtinu prezentovat – vhodnější je dle některých lingvistů (např. Milana Hrdličky) zařazovat tuto problematiku až ve výuce pokročilejších studentů, navíc odděleně od spisovné češtiny.<sup>30</sup> Pomocí nástroje SyD pak student snadno rozezná, která varianta je spisovná a která obecně česká, SyD totiž zobrazuje procentuální zastoupení variant v psaném i mluveném jazyce, což není samozřejmě přínosné pouze pro studenty češtiny jako cizího jazyka, avšak pro nerodilé mluvčí představuje korpus ucelenou a rozsáhlou databázi informací.

V nástroji SyD můžeme porovnávat synonyma určitého slova, hláskoslovné i morfologické varianty. Při tvorbě cvičení či zapojení nástroje do výuky je ale nezbytné, aby vyučující předem ověřil výsledky vyhledávání. Mluvené korpusy neobsahují tolik dat jako ty psané, a tak se může stát, že pro daný jev nebudou nalezena dostatečná data,

---

<sup>29</sup> CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:oral* [online].

<sup>30</sup> VALIŠOVÁ, Pavlína. Data-driven learning a výuka češtiny jako cizího jazyka, s. 31.

vyhledaný výsledek proto může přinášet zkreslené informace o hledaném jevu. S menším objemem dat je však nutné počítat při práci se všemi mluvenými korpusy.

### 1.4.3 Korpus SCHOLA2010

Korpus SCHOLA2010 ze školního prostředí je jedinečný svým didaktickým zaměřením. Obsahuje nahrávky z vyučovacích hodin, výjimečný je také proto, že na rozdíl od ostatních korpusů obsahuje mluvu nejen dospělých, ale i dětí a mládeže.<sup>31</sup> Patří mezi korpusy akviziční, „které slouží primárně pro studium procesů spojených s osvojováním řeči a s jejím vyučováním jako prvního i druhého jazyka.“<sup>32</sup>

Využití korpusu SCHOLA2010 není však jen pro žáky, ale zejména pro pedagogy, kteří výsledky výzkumu v tomto korpusu mohou pak aplikovat ve vyučování. V nahraném materiálu se střetává formální a neformální jazyk žáků i učitelů, vyučující může identifikovat hlavní chyby, které žáci v mluvě dělají (např. hyperkorektní či nedbalá výslovnost), analyzovat, nakolik žáci používají obecnou češtinu a nářeční prvky v mluvě oproti spisovné češtině. Všechny tyto informace se dají ve výuce zužitkovat tak, aby studenti lépe vyslovovali<sup>33</sup>, nechybovali ve slovesných tvarech apod.

---

<sup>31</sup> CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:schola2010* [online].

<sup>32</sup> Tamtéž.

<sup>33</sup> Výslovnost cizích vlastních a obecných jmen se uvádí do hranatých závorek, díky čemuž je možné vyhledat výslovnost správnou, nebo naopak chybnou. To může být přínosné i při výuce literatury, lze ověřit výslovnost jmen některých cizích spisovatelů, kteří jsou v korpusu zachyceni.



## 2 KORPUS DIALEKT

Korpus DIALEKT je jeden ze specializovaných korpusů mluveného jazyka vzniklých v ČNK. Tato kapitola se věnuje jeho charakteristice, způsobu sběru dat, složení a dvouúrovňové transkripci, nakonec také aplikacím na tomto korpusu založených.

### 2.1 Charakteristika

„Korpus DIALEKT prezentuje tradiční teritoriální dialekty zachycené na území celé České republiky.“<sup>34</sup> Některé nahrávky pocházejí i z území Polska. Korpus byl ve své první verzi zpřístupněn veřejnosti roku 2017 a obsahoval zhruba necelých 100 tis. slov (tokenů) bez interpunkce ve 324 nahrávkách, ve druhé rozšířené verzi z prosince 2021 obsahuje kolem 223 tis. slov (tokenů) v 972 nahrávkách. Hlavním cílem tohoto projektu je zpřístupnit nářeční jazyková data pro odbornou i neobornou veřejnost, rozšiřovat povědomí o dialektech v Čechách, na Moravě i ve Slezsku a pomoci ve školách při výuce nářeční problematiky.

#### 2.1.1 Sběr dat

V korpusu DIALEKT se objevují nahrávky, jejichž sběr započal již koncem 50. let minulého století, a to pro potřeby ČJA a nářečního slovníku, na kterém pracovalo dialektologické oddělení Ústavu pro jazyk český (dále jen ÚJČ) Akademie věd České republiky, v. v. i (dále jen AV ČR). Tento materiál je taktéž publikován v *Dodatecích Českého jazykového atlasu* z roku 2011 a do korpusu DIALEKT se dostal až druhotně. Nahrávky byly pořizovány i jednotlivci, jak je tomu i u těch novějších, kromě soukromého sběru pocházejí nahrávky ze sběru sond Ústavu Českého národního korpusu FF UK či, ve většině případů, z univerzitních institucí (FF UK, FF JU, PF JU, FF MU).

Korpus DIALEKT se snaží zachytit nejstarší stav dialektů, nahrávány jsou proto osoby ve věkové kategorii nad 60 let<sup>35</sup>, kteří byli narozeni od konce 19. století do první poloviny 20. století. Mluvčími neboli informátory jsou lidé zpravidla žijící na venkově, kteří znají

---

<sup>34</sup> CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:dialekt* [online].

<sup>35</sup> Oproti korpusu ORAL, který vymezuje dvě věkové kategorie – do 35 let a nad 35 let.

starousedlý způsob života a mají vztah k řemeslům a zemědělství. Od toho se odvíjejí i témata nahrávek – život na venkově, zvyky a tradice, folklor, vyprávění o dobových událostech. Materiál v nahrávkách má formu řízeného rozhovoru, ten je ale spíše monologického rázu, explorátor se krátkými otázkami táže informátora, který vypráví. Nahráváno je v domácím prostředí, promluvy jsou nepřipravené a neformální, aby byla zachována autenticita jazykového projevu.<sup>36</sup>

## 2.1.2 Složení korpusu

Nahrávky v korpusu DIALEKT se dělí do dvou vrstev – starší vrstva obsahuje nahrávky pořizené od 50. do 80. let 20. století, nahrávky v novější vrstvě vznikaly od počátku 90. let 20. století až do současnosti. Sondy jsou pořizovány po celé České republice a jsou tak zastoupeny všechny nářeční oblasti, což lze vidět na obrázku č. 1 níže. Navíc jsou v korpusu obsaženy nahrávky z Kladska a okolí polského Husince.



Obrázek 1 – Mapa nářečí českého jazyka, zdroj: <https://wiki.korpus.cz/doku.php/cnk:dialekt:mapa>

Korpus DIALEKT je lemmatizován a morfologicky značkován stejně jako korpusy ORAL a ORTOFON. Tagování slovních druhů není ale vždy spolehlivé kvůli variabilitě nářečního materiálu a nářečním koncovkám, bylo proto nutné lemmatizaci a značkování upravit.

<sup>36</sup> CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:dialekt* [online].

„Korpus DIALEKT obsahuje velmi pestrý jazykový materiál co do typů mluvčích a míst nahrávání a tuto různorodost zachycuje sociolingvistické značkování.“<sup>37</sup> Korpus obsahuje mnoho strukturních atributů týkajících se jak mluvčích, tak i nahrávky samotné. U nahrávky se uvádí její identifikační označení, zdroj, rok a místo pořízení, nářeční vrstva (stará, nová), typ sídla (městská, venkovská lokalita); místo pořízení – velikost sídla, stát, region, nářeční oblast, nářeční podskupina, nářeční úsek, nářeční typ, okres; dále typ promluvy podle dialogičnosti (dialog, monolog, převážně dialog, převážně monolog), téma promluvy, celkový počet mluvčích a přítomnost explorátora v nahrávce. Strukturní atributy uváděné k mluvčímu jsou následující – identifikátor segmentu, automaticky vygenerovaná přezdívka, zda je či není mluvčí explorátor, pohlaví mluvčího, rok narození, věková kategorie (u nářečních mluvčích se uvádí pouze jedna – nad 35 let, vybírání byli informátoři především ve věkové kategorii nad 60 let<sup>38</sup>), vzdělání, nejdelší zaměstnání; místo nejdelšího pobytu mluvčího – obec, stát, region, nářeční oblast, nářeční podskupina, nářeční úsek, nářeční typ; nakonec je uvedeno místo pobytu mluvčího do 15 let i s nářeční oblastí.<sup>39</sup>

### 2.1.3 Dvouúrovňová transkripce

Transkripce v korpusu DIALEKT probíhá v programu ELAN, který umožňuje zpracování přepisu ve dvou rovinách. Program ELAN byl vytvořen holandským institutem Max Planck Institute for Psycholinguistics.<sup>40</sup> V programu jsou paralelně užity dvě transkripční roviny, dialektologická a ortografická, které jsou časově spojeny se zvukovým záznamem, mimo ně je možné zachytit další metalingvistické informace o mluvčích a nahrávce samotné. Za použití korpusového manažeru KonText lze v korpusu DIALEKT vyhledávat jak na rovině dialektologické, tak i ortografické.<sup>41</sup>

Přepis na rovině dialektologické je základní, vychází z pravidel pro transkripci dialektologických textů, která byla stanovena v následujících publikacích – *Pravidla pro vědecký přepis dialektických zápisů českých a slovenských* (1951), *České nářeční texty*

---

<sup>37</sup> GOLÁŇOVÁ, Hana a Martina WACLAWIČOVÁ. Nářeční korpus DIALEKT a jeho použití ve výuce češtiny. *Český jazyk a literatura*, s. 129.

<sup>38</sup> Je zde absence generačního rozpětí, v obou vrstvách nahrávek je zastoupena jen nejstarší generace mluvčích.

<sup>39</sup> CVRČEK, Václav a Olga RICHTEROVÁ. *Pojmy: atributy\_strukturni* [online].

<sup>40</sup> Elan. Dostupné z: <https://archive.mpi.nl/tla/elan>.

<sup>41</sup> GOLÁŇOVÁ, Hana a Martina WACLAWIČOVÁ. The DIALEKT Corpus and Its Possibilities. *Jazykovedný časopis*, s. 338–339.

(1976) a *Dodatky Českého jazykového atlasu* (2011). Dialektologická transkripce představuje věrné zachycení toho, co mluvčí pronesl, nacházejí se v ní nářeční hlásky, pro které se užívají speciální znaky.<sup>42</sup> Jejich seznam poskytuje ČNK v *Archivu diferenčních hlásek českého jazyka*, nadstavbové aplikaci založené na korpusu DIALEKT. Podrobněji se této aplikaci věnuje podkapitola č. 2.2.1 na s. 30.

Interpunkce se na dialektologické rovině zapisuje ve shodě s pravidly pravopisu, stejně tak je dodržována i hranice jednotlivých slov. Na této rovině se užívá standardních interpunkčních znamének (i k zaznamenání zvláštních syntaktických rysů projevu), na začátku věty se však píše malé písmeno. Pokud je výpověď neukončená, označuje se třemi tečkami, ty se připojují za poslední slovo výpovědi bez mezery. Pokud se jedná o nedorečená slova, označují se hvězdičkou připojenou opět bez mezery. Co se samotných hlásek týče, jejich kvantita i kvalita je zaznamenávána dle skutečné výslovnosti. Rozlišuje se např. *y-i* (výslovnost *y/ý* je typická pro nahrávky z východomoravské a slezské nářeční oblasti), dále měkké, tvrdé a dlouhé slabikotvorné *l* (*l', l, l̂*), dlouhé slabikotvorné *r* (*r̂*), zaznamenává se protetické *v*, dle běžných dialektologických pravidel se zapisuje dlouhé *ú*. Z dalších nářečních hlásek se zaznamenávají široké samohlásky (*ä, e, o*), zaokrouhlené samohlásky (*ǝ*) a úzká výslovnost dvojhlásek *ou* a *ej*. Zaznamenává se i výslovnost redukováná, jde o případy, ve kterých je výslovnost samohlásky jen naznačena. Pro její označení se užívá znak *a*, jímž se dále označuje také znělá výslovnost souhlásky na konci slova. Dle skutečné realizace se na dialektologické rovině zachycuje asimilace znělosti mezislovní i uvnitř slova, dále asimilace artikulační a zjednodušené souhláskové skupiny.<sup>43</sup>

Ortografická rovina představuje druhou úroveň přepisu, na které se zápis blíží standardnímu pravopisu s výjimkou některých rysů. Zachovává se nářeční či regionální lexikum, jevy morfologické (koncovky ohebných slovních druhů), zapisuje se i protetické *v-*. Kvantita se na ortografické rovině přepisuje podle ortografických pravidel, a to i v případě slezského krácení. Interpunkce se na ortografické rovině neznačí ve shodě s pravidly pravopisu jako na rovině dialektologické, je zde užíváno interpunkce pauzové, označují se pauzy a předěly.<sup>44</sup> Pauzy se na ortografické rovině značí dvěma tečkami,

---

<sup>42</sup> CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:dialekt:pravidla* [online].

<sup>43</sup> Tamtéž.

<sup>44</sup> Podrobněji viz <https://wiki.korpus.cz/doku.php/cnk:dialekt:pravidla>.

předěl pak tečkou jednou, vykřičník a otazník jsou používány ve shodě s dialektologickou rovinou. Tři tečky se na rovině ortografické pro označení přerušení či nedokončení repliky neuvádějí. Další speciální znaky jsou zavedeny pro označení přerušení a opětového navázání promluvy v případě, že nejedná o překryv (současný projev více mluvčích najednou).<sup>45</sup>

V zásadách pro přepis existují i pravidla společná pro obě roviny. Jedná se o zápis neverbálních a responzních zvuků – *hmm* pro souhlasné zvuky, *emm* pro odmítavé. Hezitační zvuky se označují pomocí @. Jak již bylo výše zmíněno, velká písmena se nepoužívají na začátku vět, ale pouze u proprií. Příjmení jsou však z důvodu anonymizace kódována a nahrazována jinými, vybíráno je z následujících příjmení dle zařazení k deklinačnímu typu – *Novák, Hruška, Král, Tichý, Ševců, Nevole, Poupě*. Na obou rovinách je ale i u příjmení zachována v přepisu vyslovená koncovka. Podobně je tomu u přezdivek, které se nahrazují jako *Zeleňák, Mazal* či *Červený*. Na obou rovinách se také označuje, zda jsou slova vyslovena s určitou charakteristikou. Užívá se zkratky, která se píše společně s projevem do špičatých závorek. Mezi nejčastěji užívané patří např. *CP* pro cizojazyčný projev, *KR* pro křik, *NH* pro nahlas pronesený projev, *PR* pro přerěknutí, *PT* pro projev pronesený potichu nebo *SM* pro pronesení se smíchem.<sup>46</sup>

Při práci s transkripčním programem ELAN se do přepisu zahrnují ještě další dvě roviny – meta a META. Rovina meta se váže vždy k určitému mluvčímu a je označena stejným číslem jako rovina dialektologická a ortografická související s projevem daného mluvčího. Mezi údaje uváděné na rovině meta se řadí např. dlouhá pauza, nadechnutí, odkašlání, pláč nebo smích. Na rovině META se pak zaznamenávají zvuky týkající se nahrávky jako celku – hlasitý hovor v pozadí, klepání, nesrozumitelný hovor více mluvčích, ruch z ulice, zvonění telefonu, zvuk z rádia či televize a další.

---

<sup>45</sup> Podrobněji viz <https://wiki.korpus.cz/doku.php/cnk:dialekt:pravidla>.


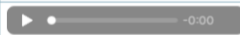





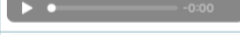


<sup>46</sup> Podrobněji viz <https://wiki.korpus.cz/doku.php/cnk:dialekt:pravidla>.

## 2.2 Nadstavbové aplikace založené na korpusu DIALEKT

ČNK kromě samotného korpusu DIALEKT zpřístupňuje další aplikace, které jsou na něm založené a obsahují data z něj. Jedná se o *Archiv diferenčních hlásek nářečí jazyka českého* a o aplikaci *Mapka*.

### 2.2.1 Archiv diferenčních hlásek nářečí českého jazyka

*Archiv diferenčních hlásek nářečí českého jazyka*<sup>47</sup> obsahuje seznam většiny nářečních hlásek typických pro české dialekty. Archiv má formu tabulky, ve které jsou hlásky rozděleny na samohlásky a souhlásky. U každé hlásky je vždy uvedeno, jak se může realizovat, nářeční oblast a případně bližší geografické určení, pro které je daná realizace typická. Dále jsou v tabulce obsaženy ke každé hlásce zvukové ukázky, které byly převzaty přímo z korpusu DIALEKT společně i s přepisem, nakonec je uvedeno místo a rok pořízení nahrávky, ze které pochází ukázka. Archiv slouží nejen jako ucelený seznam nářečních hlásek vhodný pro práci odborníků. Lze jej velmi dobře využít při výuce češtiny k prezentaci typických regionálních hlásek spolu s jejich autentickou ukázkou.

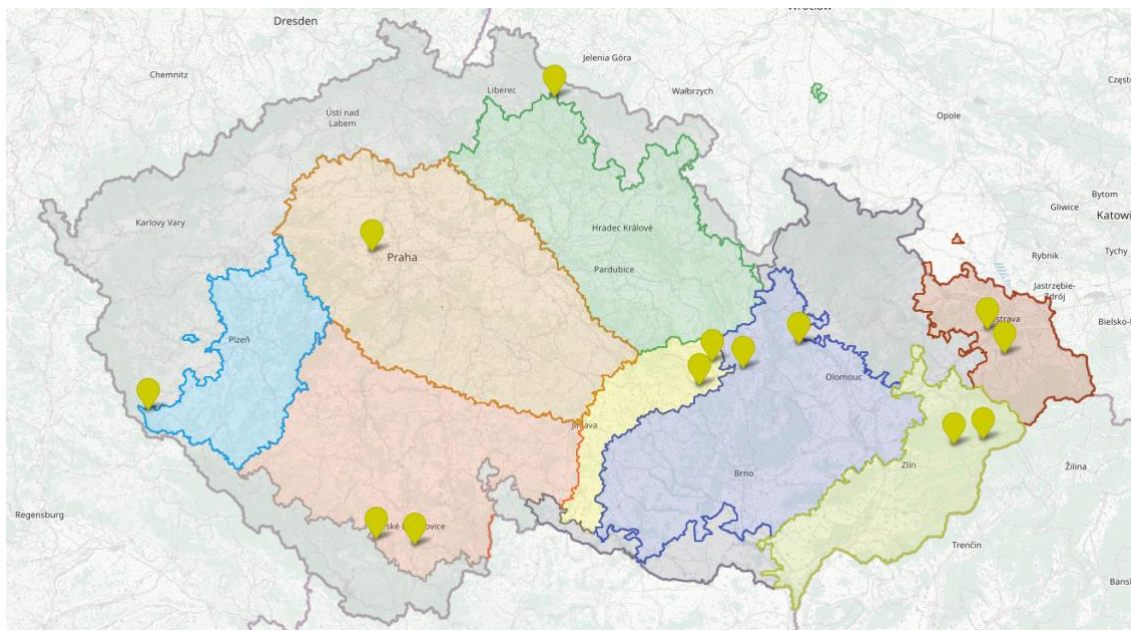
ou	zavřené °u	Chodsko (západočeská nářeční oblast), Doudlebsko (jihočeská nářeční oblast)	 -0:01	takov°u plotn°u	Postřekov; 1963 (okr. Domažlice)
			 -0:00	s°usedí	Postřekov; 1963 (okr. Domažlice)
	zavřené °u	Chodsko (západočeská nářeční oblast), Doudlebsko (jihočeská nářeční oblast)	 -0:01	mil to seb°u	Postřekov; 1963 (okr. Domažlice)
			 -0:01	já sem jedn°u víděla	Postřekov; 1963 (okr. Domažlice)
			 -0:02	se vo°upaji, nast°uhaji	Postřekov; 2007 (okr. Domažlice)
			 -0:01	ššelijak°j učitel	Postřekov; 1963 (okr. Domažlice)
ej	zavřené °j	Chodsko (západočeská nářeční oblast), Doudlebsko (jihočeská nářeční oblast)	 -0:01	házeli ti v°jce	Postřekov; 2007 (okr. Domažlice)
	zavřené j	Chodsko (západočeská nářeční oblast), Doudlebsko (jihočeská nářeční oblast)	 -0:00	každ°j misilil	Postřekov; 1963 (okr. Domažlice)
			 -0:00	ma°j kluk	Postřekov; 2007 (okr. Domažlice)
			 -0:02	bil viváler°j na slamjenkáj, nechal se viknout	Soběnov; 1964 (okr. Český Krumlov)

Obrázek 2 – Archiv diferenčních hlásek nářečí českého jazyka, zdroj:  
<https://wiki.korpus.cz/doku.php/cnk:dialekt:archivhlasek>

<sup>47</sup> Dostupný na adrese <https://wiki.korpus.cz/doku.php/cnk:dialekt:archivhlasek>.

## 2.2.2 Mapová aplikace pro korpusy mluvené češtiny

*Mapka*<sup>48</sup> neboli *Mapová aplikace pro korpusy mluvené češtiny* je nová aplikace zpřístupněná pro veřejnost v roce 2020, její poslední aktualizace proběhla koncem roku 2021. Je založená na korpusu DIALEKT. Budou do ní časem připojena data z korpusu ORTOFON a snad i ze spojených korpusů řady ORAL.



Obrázek 3 – Mapová aplikace pro korpusy mluvené češtiny, zdroj: <https://www.korpus.cz/mapka/>

Hlavní funkcí *Mapky* je zobrazení míst nahrávek v rámci nářečních oblastí, jež jsou začleněny do korpusu DIALEKT, jako jehož vizuální pomůcka také *Mapka* slouží (aplikace navíc nabízí široké možnosti územního členění, dokonce poskytuje zobrazení administrativního členění ČR na kraje a okresy, v nejnovější verzi z prosince 2021 bylo doplněno i historické členění, díky němuž je možné na mapě ČR zobrazit německé jazykové ostrovy a slezskou zemskou hranici). Ke každé z osmi základních nářečních oblastí jsou uvedeny na mapě obecné rysy regionu, specifické rysy nářeční oblasti a seznam obcí s nahrávkami v korpusu DIALEKT spolu s autentickými ukázkami. Každá nářeční oblast obsahuje jednu nahrávku z nové vrstvy a jednu ze staré, k oběma je k dispozici transkripce (dialektologická i ortografická). Ke každé z uvedených ukázek je připojen rozbor většiny nářečních jevů, které se v ukázce nacházejí, jde o komplexní

<sup>48</sup> Dostupná na adrese <https://www.korpus.cz/mapka/>.

rozbor vztahující se ke všem rovinám (hláskosloví, tvarosloví, skladba věty a souvětí, slovní zásoba). K ukázkám, nikoliv však ke všem lokalitám, které je možné na mapě zobrazit, je navíc připojen i zvukový záznam, jenž lze přehrát (nahrávky se týkají témat, díky kterým se jak žáci a studenti, tak i neodborná veřejnost pracující s *Mapkou* může dozvědět mnoho zajímavých informací o dřívějším životě či různých událostech v naší republice).<sup>49</sup>

Práce s aplikací *Mapka* je velmi jednoduchá a intuitivní. Vyhledávání lze omezit jen na určitou oblast nebo například dle administrativního členění. Omezit lze i výběr lokalit, který je automaticky přednastaven a zobrazuje lokality s ukázkami nářečních promluv (ukázky budou s postupným rozšiřováním korpusu DIALEKT v aplikaci přibývat). Vyhledávat lze obce nebo i jejich části, ke každé obci se pak zobrazí základní informace včetně nářeční oblasti. Vybrané obce lze na mapu připnout, což vyhledávání usnadňuje a umožňuje tak vytvořit mapu vlastní, kterou je možné v aplikaci pro pozdější práci také stáhnout. Dále se nabízí možnost tyto obce uložit do počítače jako vlastní vrstvu a dle potřeby ji znovu nahrát a pracovat s ní. Na obrázku 4 níže je vidět vlastní mapa doudlebského úseku, na níž jsou modře zobrazena místa zahrnutá do celkového sběru nářečního materiálu, žlutě místa, ze kterých pochází nahrávky dostupné v aplikaci.<sup>50</sup>

Aplikace obsahuje v menu odkazy na wiki ČNK, konkrétně na stránky týkající se popisu nářečního členění České republiky, korpusu DIALEKT a transkripce v něm, výše zmíněného archivu, korpusu ORTOFON a také odkaz na aplikaci KonText. Dále se lze přes mapovou aplikaci dostat k vyhledávání v *Českém jazykovém atlase* a jeho novější elektronické verzi i starší verzi k tisku, dále i k elektronické verzi *Slovníku nářečí českého jazyka* (dále jen SNČJ), *Slovníku pomístních jmen v Čechách* (dále jen SPJČ) a *Slovníku pomístních jmen na Moravě a ve Slezsku* (dále jen SPJMS).<sup>51</sup>

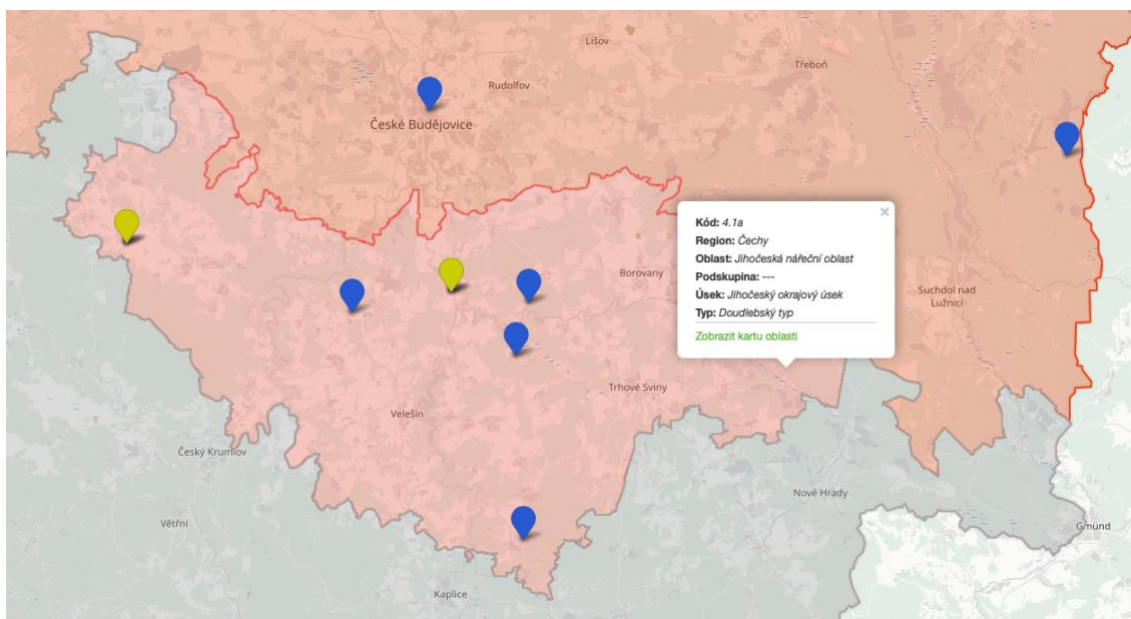
---

<sup>49</sup> GOLÁŇOVÁ, Hana, WACLAWIČOVÁ, Martina a Jakub PEJCHA. *Mapka: Mapová aplikace pro korpusy mluvené češtiny* [online].

<sup>50</sup> Tamtéž.

<sup>51</sup> Tamtéž.





Obrázek 4 – Vlastní mapa doudlebského úseku s místy nahrávání, zdroj: <https://www.korpus.cz/mapka/#>

### 3 ČESKÝ JAZYKOVÝ ATLAS

Jazykový atlas je „soubor map zobrazující jazykové jevy v jejich zeměpisném rozšíření v předem vymezené oblasti. Každému jevu je věnována zpravidla jedna mapa.“<sup>52</sup> Jazykové atlasy se tím odlišují od mapových aplikací, které prezentují nářeční jevy na mapě určitého území jako celek. Naopak jazykové atlasy jsou vytvářeny metodou jazykového zeměpisu. Je to „dílčí jazykovědná disciplína uplatňovaná zejména v dialektologii a zabývající se výzkumem rozdílů sledovaných nářečí nebo jiných jazykových jevů a jejich územního rozšíření.“<sup>53</sup> Pomocí této metody se zkoumané jevy zakreslují do mapy, jejichž soubor pak vytváří právě jazykový atlas, který nejen zobrazuje hranice zakreslovaných jevů, ale snaží se daný jev zasadit do širších souvislostí.<sup>54</sup>

*Český jazykový atlas* (dále jen ČJA) je jedno ze stěžejních a nejpodrobnějších děl české dialektologie. ČJA má 5 svazků plus *Dodatky* (sumarizující svazek) jak v podobě knižní, tak i elektronické, která poskytuje on-line verzi i verzi atlasu k tisku.<sup>55</sup> ČJA je výsledkem velmi rozsáhlého výzkumu, který provádělo dialektologické oddělení ÚJČ AV ČR. První svazek atlasu byl vydán roku 1992, zaměřuje se na lexikum z okruhu místního, domácího prostředí a člověka samotného. Druhý svazek vyšel roku 1997 a týká se zahrady a sadu, živočichů, lesa a rostlin, krajiny, času, počasí, podoby vesnice dříve a nyní, nakonec zábavy a zvyků. Třetí svazek vyšel roku 1999, je zaměřen na polní zemědělské práce, hospodářskou usedlost, zemědělské nářadí a nástroje, dobytek a drůbež. Čtvrtý svazek se již nezaměřuje na lexikum, jedná se o svazek morfologický, byl vydán roku 2002, stejně tak pátý svazek, který se týká hláskosloví, syntaxe, adverbíí, pojednává také o doplňkovém výzkumu městské sítě a popisuje izoglosy v českých nářečích, vyšel roku 2005. Poslední částí ČJA jsou *Dodatky*, které byly vydány roku 2011, jedná se o svazek závěrečný a sumarizující. *Dodatky* obsahují úvodní nářeční mapu, seznam pravidelných regionálních obměn a mapy znázorňující jejich geografické rozšíření, seznam obcí a měst a jejich charakteristiky, dotazníky, které byly podkladem pro nářeční výzkum, ukázky přepisů a audiozáznamů nářečních promluv, rejstřík položek, nakonec bibliografii české dialektologie od roku 1968.

---

<sup>52</sup> KLOFEROVÁ, Stanislava. Jazykový atlas.

<sup>53</sup> KLOFEROVÁ, Stanislava. Jazykový zeměpis.

<sup>54</sup> Tamtéž.

<sup>55</sup> Elektronická verze ČJA je dostupná na adrese <https://cja.ujc.cas.cz/e-cja/>, verze k tisku na adrese <https://cja.ujc.cas.cz/>.

### 3.1 Terénní výzkum

Terénní výzkum pro potřeby ČJA probíhal v 60. a 70. letech minulého století tzv. korespondenční metodou – pomocí dotazníku, který se snažil postihnout všechny roviny jazykového systému, tedy hláskosloví, tvarosloví, skladbu i slovní zásobu. Základem pro dotazník, který tvořilo 2649 položek, byly tzv. inventáře diferenčních nářečních jevů, jejichž materiál byl upraven, čímž vznikl *Dotazník pro výzkum českých nářečí*. Výzkum v terénu probíhal pomocí tohoto dotazníku v síti venkovských lokalit (420 obcí, z toho 223 v Čechách, 197 na Moravě a ve Slezsku, zahrnuto je i 5 polských obcí)<sup>56</sup> i v řidší síti měst (57 měst, z toho 30 v Čechách a 27 na Moravě a ve Slezsku).<sup>57</sup>

Samotný výzkum bylo nutné přizpůsobit lokalitě, ve které probíhal. Byl realizován dialektologickým oddělením ÚJČ AV ČR, a to dialektology v Praze i v Brně. Výzkum prováděly skupiny explorátorů, nejčastěji odborníků, kteří dopodrobna znali nářečí. Bádání ve venkovských lokalitách probíhalo ve dvou fázích (1964–1966 a 1967–1972). Výzkum byl zaměřen na tradiční venkovská nářečí užívaná starší generací, tam je v mluvě nářečí stále zachováno. S dotazníkem se pracovalo zpravidla v domácím prostředí metodou tzv. řízené konverzace, kdy informátor odpovídal na různé typy dotazů tak, aby mohl explorátor zaznamenat požadované odpovědi, rozhovor byl však volnější a informátor mohl přidat i další informace, které byly taktéž zaznamenány. Záznamy se prováděly přímo do tištěných formulářů, všechny materiál byl uložen a foneticky transkribován dle *Pravidel pro vědecký přepis dialektických zápisů českých a slovenských* (1951). Po výzkumu na venkově se přistoupilo ke zkoumání v městských lokalitách, jež probíhalo v letech 1973–1976. Byla zde snaha zachovat stejný metodický postup jako při výzkumu venkovských lokalit, zároveň vybírat informátory, kteří nebudou ovlivněni spisovným jazykem. Ve městech byl výzkum prováděn menšími skupinami explorátorů. Jak při bádání na venkově, tak i ve městech byly v některých případech pořizovány magnetofonové nahrávky, jejichž přepisy se staly základem publikace *České nářeční texty* (1976).<sup>58</sup>

---

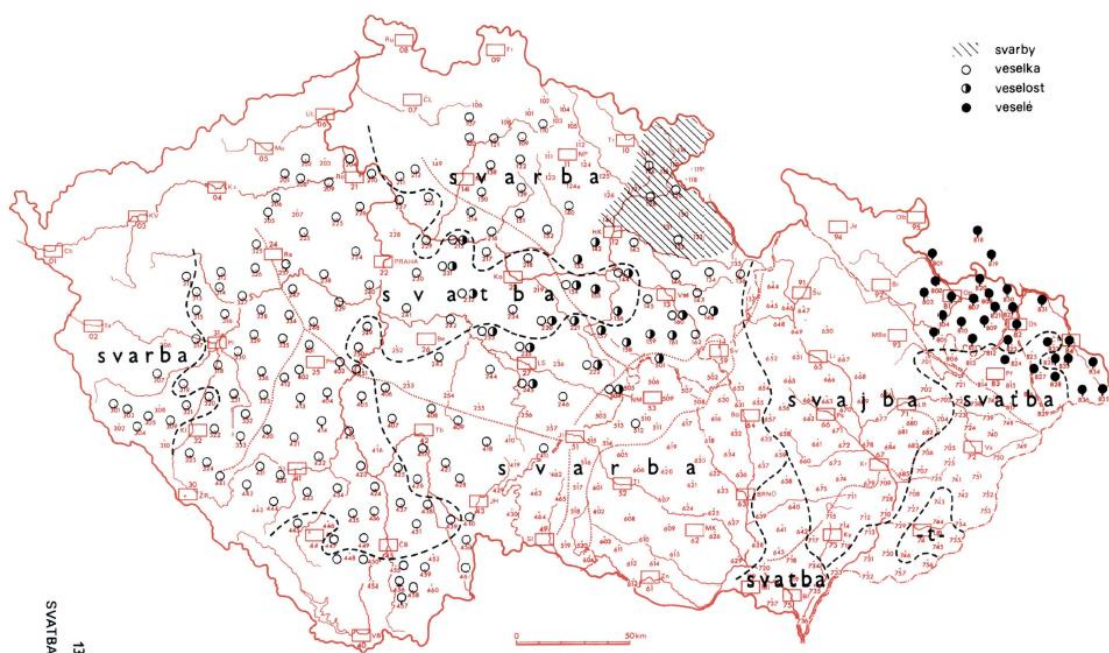
<sup>56</sup> Z celkového počtu zahrnutých lokalit tvoří 107 obcí tzv. opěrnou síť, to znamená, že byl dotazník použit v úplné formě. Některé položky byly v ostatních lokalitách, které tvoří tzv. základní síť, vypuštěny.

<sup>57</sup> BALHAR, Jan a kol. *Český jazykový atlas 1* [online], s. 11–18.

<sup>58</sup> Tamtéž, s. 18–21.

### 3.2 Zpracování materiálu

Materiál v ČJA je zpracován kartograficky, mapy v atlasu však nejsou pouze nápisové, jaké se užívaly dříve a v nichž byly jevy ve fonetickém zápisu vpisovány přímo do bodů na mapě. Místo nich jsou užívány mapy se symboly, lingvistické jevy se tedy zobrazují na mapě pomocí grafických značek, avšak v kombinaci právě s nápisovými, došlo tak k syntetizujícímu zobrazení jazykového materiálu.<sup>59</sup>



Obrázek 5 – Ukázka mapy v ČJA, zdroj: <https://cja.ujc.cas.cz/CJA1/>

Obrázek 5 výše představuje jednu z map v ČJA, konkrétně mapu č. 87 lexikální položky *svatba* v prvním svazku atlasu.<sup>60</sup> Hlavním grafickým prostředkem jsou izoglosy, které vymezují areál užívaného ekvivalentu lexikální položky. Do daného areálu se umísťuje nápis v podobě ekvivalentu, který se právě v té oblasti užívá. V mapách se užívá i šrafování, zejména u složitějších map, kdy se areály výskytů určitých variant prolínají. Dále mapa obsahuje grafické značky, které zachycují užití ekvivalentů v určitém bodu na mapě, značka je vždy uvedena nad číselným označením bodu sítě. Mimo to se tyto značky užívají k zachycení menších areálů. Všechny užití symboly jsou vysvětleny

<sup>59</sup> Tamtéž, s. 23–24.

<sup>60</sup> Na mapě je možné vidět typický západočeský hláskoslovný jev, a to změnu *d* na *r*, více k charakteristickým znakům této oblasti viz podkapitola 5.2.1 na s. 53.

v legendě, již každá mapa obsahuje. Ta však vysvětluje jen význam doplňujících symbolů, neboť hlavní areály jsou vyznačeny přímo v mapě.<sup>61</sup>

### 3.3 Elektronická verze atlasu

ČJA poskytuje 2 elektronické verze. První verze vznikla v letech 2012–2014 pod vedením dialektologického oddělení ÚJČ AV ČR a obsahuje zdigitalizované jednotlivé díly atlasu, ve kterých lze ale pouze listovat.<sup>62</sup> Druhá elektronická verze atlasu vznikla právě pro možnost vyhledávání jednotlivých položek či přímo jazykových dokladů. Z úvodní stránky této verze se lze dostat na heslář – seznam všech položek seřazených tak, jak jdou za sebou v tištěné a první elektronické verzi atlasu. K dispozici je i rejstřík, obsahující všechny položky seřazené abecedně. K dispozici jsou také odkazy na slovníky – PSJČ, SSJČ, SNČJ, SPJČ a SPJMS. Při samotném vyhledávání v elektronické verzi se dostaneme na danou položku a objeví se všechny informace tak, jak jsou i v knižní verzi či verzi k tisku spolu s mapou, na které je položka se svými ekvivalenty zakreslena.<sup>63</sup>

---

<sup>61</sup> BALHAR, Jan a kol. *Český jazykový atlas 1* [online], s. 36–38.

<sup>62</sup> Tato elektronická verze ČJA je dostupná na adrese <https://cja.ujc.cas.cz/>.

<sup>63</sup> Tato verze ČJA je dostupná na adrese <https://cja.ujc.cas.cz/e-cja/>.

## 4 ZAHRANIČNÍ NÁŘEČNÍ KORPUSY A APLIKACE DOSTUPNÉ PRO VEŘEJNOST

Zahraniční instituce zpřístupňují veřejnosti jak korpusy, tak i databáze, mapové aplikace a jazykové atlasy. Většina z těchto projektů je vytvářena odbornou veřejností, tedy lingvisty, u projektů zaměřených na nářečí zejména dialektology, ale některé z nich vznikají i díky zájmu z řad veřejnosti neodborné, která se také podílí na jejich zpracování. Níže uvedené zahraniční projekty jsou pouze výběrovou ukázkou a mohou sloužit jako inspirace.

### 4.1 Databáze a archivy

CLARIN (*Common Language Resources and Technology Infrastructure*) je databáze zpřístupňující široké veřejnosti i odborníkům jazykové zdroje, provozuje ji Evropské konsorcium pro výzkumnou infrastrukturu neboli ERIN (*European Research Infrastructure Consortium*). Snaží se je shromáždit na jedné platformě a on-line, aby byly všem, kteří je využívají, neustále k dispozici. Poskytuje seznam<sup>64</sup> 90 mluvených korpusů všech druhů v 15 jazycích včetně češtiny. Dělí je na korpusy s transkripcí a audiozáznamy, korpusy pouze s transkripcí a ostatní korpusy. Z této databáze, do které byl zahrnut i ČNK, byly vybrány některé z níže popsaných korpusů.<sup>65</sup>

Dalším projektem, který veřejnosti zpřístupňuje nářeční materiál, je archiv IDEA (*The International Dialects of English Archive*), který vznikl roku 1998 jako první internetový archiv poskytující nahrávky dialektů angličtiny z celého světa, se sběrem materiálu se pokračuje dodnes. *Paul Meier Dialect Services* je provozovatelem archivu v čele s Paulem Meierem, zakladatelem a hlavním přispěvatelem. Mimo něj se sběrem materiálu pro archiv zabývají profesori z různých univerzit či členové VASTA (*Voice and Speech Trainers Association*). S více než 170 hodinami nahrávek ze 120 zemí světa je IDEA největším archivem takového typu. Vyhledávání v archivu probíhá buď pomocí mapy (ta obsahuje vyznačená místa nahrávání, přes která lze nahrávku přehrát, viz obrázek 6 níže), nebo přes seznam kontinentů a států, ve kterých nahrávání probíhalo. U každého

---

<sup>64</sup> Seznam je dostupný na webové stránce <https://www.clarin.eu/resource-families/spoken-corpora>.

<sup>65</sup> Spoken Corpora. Dostupné z: <https://www.clarin.eu/resource-families/spoken-corpora>.

mluvčího je uvedena národnost, věk a povolání, skrze tyto charakteristiky lze v archivu taktéž hledat, navíc lze vyhledávat i jednotlivá slova z transkribovaných nahrávek.<sup>66</sup>



Obrázek 6 – Mapové vyhledávání v archivu IDEA, zdroj: <https://www.dialectsarchive.com/globalmap>

Další databázi dialektů zpřístupňuje projekt ScanDiaSyn (*Scandinavian Dialect Syntax*) ve spolupráci s *Nordic Centre of Excellence*. Na projektu pracuje současně 10 výzkumných skupin napříč všemi skandinávskými státy, provozuje ho Filozofická fakulta univerzity v Tromsø. V rámci projektu byla vytvořena syntaktická databáze norských dialektů. V databázi jsou obsaženy promluvy 924 mluvčích norštiny z 207 míst, které obsahují věty s různými syntaktickými jevy. Všechny věty jsou označeny a na základě tohoto značkování lze generovat nářeční mapy s izoglosami daných syntaktických jevů. Vyhledávat lze také dle místa, věku či pohlaví informátorů, a nebo podle samotných syntaktických fenoménů. Ve stejném projektu byl mimo databázi vytvořen také nářeční korpus norštiny o velikosti zhruba 2,75 mil. slov (tokenů), který

<sup>66</sup> IDEA – *International Dialects of English Archive* [online]. Dostupné z: <https://www.dialectsarchive.com/>.

obsahuje nahrávky spontánních projevů v dialektech severních germánských jazyků napříč skandinávskými zeměmi.<sup>67</sup>

## 4.2 Korpusy

### 4.2.1 Slovenský nářeční korpus

*Korpus nářečí Slovenského národného korpusu* (KN-SNK) je realizován oddělením Slovenského národného korpusu Jazykovedného ústavu Ľ. Štúra Slovenskej akadémie vied (dále jen SAV). KN-SNK byl vytvářen od roku 2013, první verze byla spuštěna roku 2014 a poslední přístupná verze (dialekt-5.0) je z března roku 2022, čítá zhruba 980 tis. slov (tokenů). Korpus se snaží o shromáždění existujících, zejména již publikovaných prepisů nářečních zvukových záznamů, které pocházejí ze starších publikací – monografií týkajících se jednotlivých slovenských nářečí, vysokoškolských příruček, disertačních prací, periodických článků<sup>68</sup>, část obsažených dat poskytlo Dialektologické oddelenie Jazykovedného ústavu Ľ. Štúra SAV. Korpus tedy nevzniká prepisem aktuálně pořizovaných nahrávek, na rozdíl od korpusu DIALEKT.<sup>69</sup>

Vyhledávání v KN-SNK probíhá na základě jednotlivých slov, korpus není lemmatizovaný ani morfologicky označovaný, vyhledávání je pomocí nástroje NoSketch Engine<sup>70</sup> zpřístupněno registrovaným uživatelům, možné je vyhledávat i bez registrace, avšak s velmi omezenými možnostmi. Korpus obsahuje speciální strukturní atributy, metainformace o textech i informátorech a explorátorech, které jsou velmi podobné datům v českém korpusu DIALEKT.<sup>71</sup>

---

<sup>67</sup> Nordic Dialect Corpus and Syntax Database. Dostupné z: <http://www.tekstlab.uio.no/scandiasyn/index.html>.

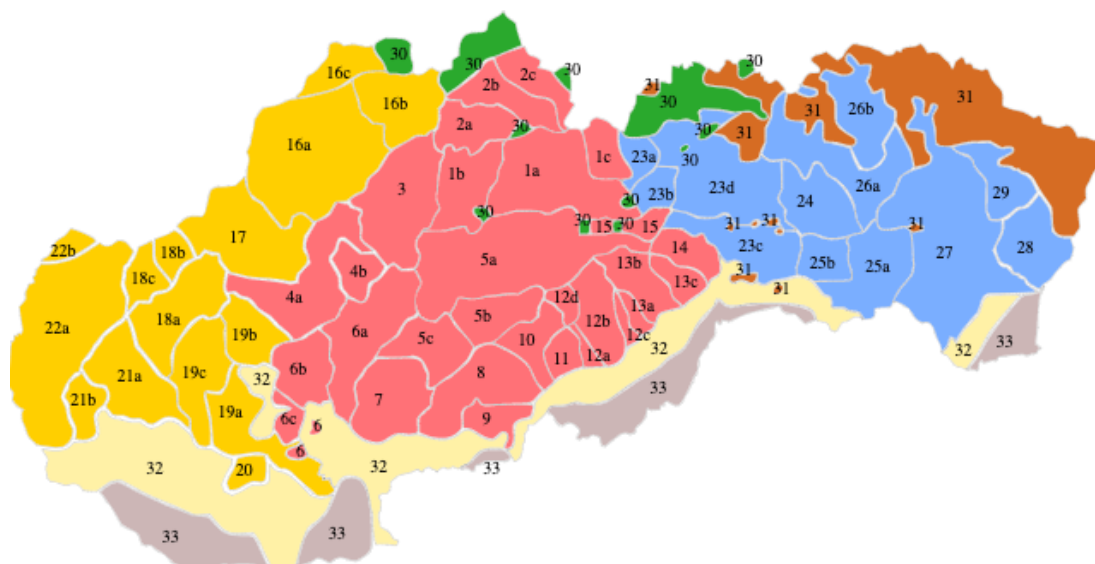
<sup>68</sup> Podrobný seznam textových zdrojů viz [https://korpus.sk/dialect\(2d\)zdroje.html](https://korpus.sk/dialect(2d)zdroje.html).

<sup>69</sup> Korpus nářečí Slovenského národného korpusu. Dostupné z: <https://korpus.sk/dialect.html>.

<sup>70</sup> Vyhledávání je dostupné po registraci na adrese <https://bonito.korpus.sk/>.

<sup>71</sup> Korpus nářečí Slovenského národného korpusu. Dostupné z: <https://korpus.sk/dialect.html>.





Obrázek 7 – Mapa nářečí na území SR, zdroj: <https://korpus.sk/dialect.html>

#### 4.2.2 Německé nářeční korpusy

Mluvené korpusy německého jazyka<sup>72</sup> zpřístupňují mnoho nářečních korpusů, všechny z nich jsou označeny názvem *Deutsche Mundarten*. Prvním je nářeční korpus *DDR*, který byl vytvořen zaměstnanci Ústavu pro německý jazyk a literaturu německé Akademie věd. Tento korpus obsahuje 385 hodin<sup>73</sup> nahrávek z let 1960–1968, které zachycují vzorky východoněmeckých dialektů, jde především o rozhovory a vyprávění příběhů. Ve spolupráci s Archivem mluvené němčiny (AGD – *Archiv für Gesprochenes Deutsch*) bylo 117 nahrávek přepsáno, jejich transkript je na stránkách Archivu k dispozici, dále je uložen v Databázi mluvené němčiny (DGD – *Datenbank für Gesprochenes Deutsch*), kde s ním lze po registraci pracovat.<sup>74</sup>

Dalším nářečním korpusem německého jazyka je korpus *Kreis Böblingen*, byl vytvořen Ulrichem Engelem, německým filologem. Korpus obsahuje 73 nahrávek z let 1965–1967, které zachycují dialekt v okrese Böblingen. Jedná se opět o příběhy, rozhovory a čtený jazyk mluvčích daného nářečí. Všechny záznamy jsou k dispozici v AGD a DGD.<sup>75</sup>

<sup>72</sup> Seznam všech mluvených korpusů německého jazyka viz [http://agd.ids-mannheim.de/korpus\\_index.shtml#2](http://agd.ids-mannheim.de/korpus_index.shtml#2).

<sup>73</sup> Data o počtu slov určujících velikost korpusu nejsou, tak jako u ostatních zmíněných, k dispozici.

<sup>74</sup> Deutsche Mundarten: DDR. Dostupné z: [http://agd.ids-mannheim.de/DR--\\_extern.shtml](http://agd.ids-mannheim.de/DR--_extern.shtml).

<sup>75</sup> Deutsche Mundarten: Kreis Böblingen. Dostupné z: [http://agd.ids-mannheim.de/BB--\\_extern.shtml](http://agd.ids-mannheim.de/BB--_extern.shtml).

Dalším z korpusů je *Zwirner-Korpus*, jenž je jedním z nejobsáhlejších nářečních korpusů a byl vytvořen v rámci projektu Německého archivu jazyků. Korpus obsahuje 5796 zvukových záznamů o délce 1077 hodin z let 1955–1972. Byla zde snaha zaznamenat německé dialekty co nejdůkladněji, sběr dat probíhal v tehdejších spolkových zemích, ve Francii či v Rakousku. Mluvčí v nahrávkách byli z mladších i starších generací. Zaznamenány jsou různé druhy mluvených textů, vyprávění, ukázky starších děl, ale i tzv. Wenkerovy věty<sup>76</sup>. Nahrávky a jejich přepisy jsou opět digitalizovány v AGD tentokrát i s lemmatizací a dalšími vysvětlivkami. Korpus je taktéž dostupný v DGD.<sup>77</sup>

*Südwestdeutschland und Vurarlberg* je dalším z nářečních korpusů, jenž byl vytvořen v rámci projektu Tübingen Language Center v jihozápadním Německu, jeho záznamy tentokrát sloužily zejména k doplnění výše zmíněného *Zwirner-Korpus*. Obsaženo je 242 nahrávek pořízených v jihozápadním Německu v letech 1966–1970. Do nahrávání byli zařazeni i lidé pocházející z Ukrajiny, kteří se usadili v německém Württembergu. Záznamy jsou taktéž digitalizovány v AGD a uloženy v DGD.<sup>78</sup> Dalším z korpusů, který měl opět doplnit nářeční *Zwirner-Korpus*, je korpus *Schwarzwald* (vytvořený v rámci stejného projektu) čítající 126 zvukových záznamů o délce zhruba 37 hodin. Nahrávání probíhalo v letech 1964–1974 v okresech Freudenstadt a Wolfach v jihozápadním Německu, nahrávky jsou opět dostupné v AGD i DGD.<sup>79</sup>

#### 4.2.3 Norský nářeční korpus

Ve spolupráci čtyř norských univerzit (Univerzita v Oslu, Univerzita v Bergenu, Norská arktická univerzita v Tromsø a Norská univerzita vědy a technologie) a dalších organizací byl vytvořen projekt zvaný LIA (*Language Infrastructure made Accessible*). Projekt vznikl v letech 2014–2019 se snahou sebrat co nejvíce nahrávek, které byly vytvořeny v posledních 60 letech za účelem nářečního a onomastického výzkumu. Nahrávky byly digitalizovány Norskou národní knihovnou a ty zajímavé s kvalitním zvukovým

---

<sup>76</sup> Wenkerovy věty čítají čtyřicet vět s běžně užívanými slovy, pomocí jejichž odpovědí se dialekty analyzovaly. Vytvořil je německý dialektolog Georg Wenker, byly rozesílány do jednotlivých jazykových oblastí v Německu, odpovědi na ně pak přicházely přímo v užívaných dialektech.

<sup>77</sup> Deutsche Mundarten: Zwirner-Korpus. Dostupné z: [http://agd.ids-mannheim.de/ZW--\\_extern.shtml](http://agd.ids-mannheim.de/ZW--_extern.shtml).

<sup>78</sup> Deutsche Mundarten: Südwestdeutschland und Vorarlberg. Dostupné z: [http://agd.ids-mannheim.de/SV--\\_extern.shtml](http://agd.ids-mannheim.de/SV--_extern.shtml).

<sup>79</sup> Deutsche Mundarten: Schwarzwald. Dostupné z: [http://agd.ids-mannheim.de/SW--\\_extern.shtml](http://agd.ids-mannheim.de/SW--_extern.shtml).

záznamem dále zpracovány v rámci LIA. Projekt zpřístupňuje dva typy nahrávek – s nářečním materiálem pocházející přímo z Norska a nahrávky pocházející z Ameriky.<sup>80</sup>

Korpus byl ve své první verzi zpřístupněn v září roku 2019, obsahuje zhruba 3,5 mil. slov (tokenů) a v nahrávkách vystupuje 1374 mluvčích z více než 200 míst v Norsku. Nahrávky mají dialogickou povahu a týkají se zejména starých zvyků, způsobů obživy, života na venkově apod. Jazyk v nich je neformální a nahrávání většinou probíhalo u informátorů doma. Korpus je morfologicky označován.<sup>81</sup>

### 4.3 Mapové aplikace

#### 4.3.1 Mapová aplikace německých nářečí

Kromě německých nářečních korpusů je zpřístupněna veřejnosti také německá mapová aplikace pod názvem *Dialektkarte*<sup>82</sup>. Mapa obsahuje krátké nahrávky 298 mluvčích z celého Německa, ale také z území Rakouska, Švýcarska a severní Itálie. Jako mapové rozhraní je využita Google mapa, která zobrazuje všechna místa nahrávání pomocí červených bodů, dostupná je i funkce Street View. Ke každé nahrávce jsou uvedeny informace o daném dialektu, místě nahrávání a dále věku, pohlaví a profesi informátora. Krátkou nahrávku lze přehrát a k dispozici je její přepis v nářeční mluvě, ale i verze ve spisovné němčině, viz obrázek č. 8.<sup>83</sup>

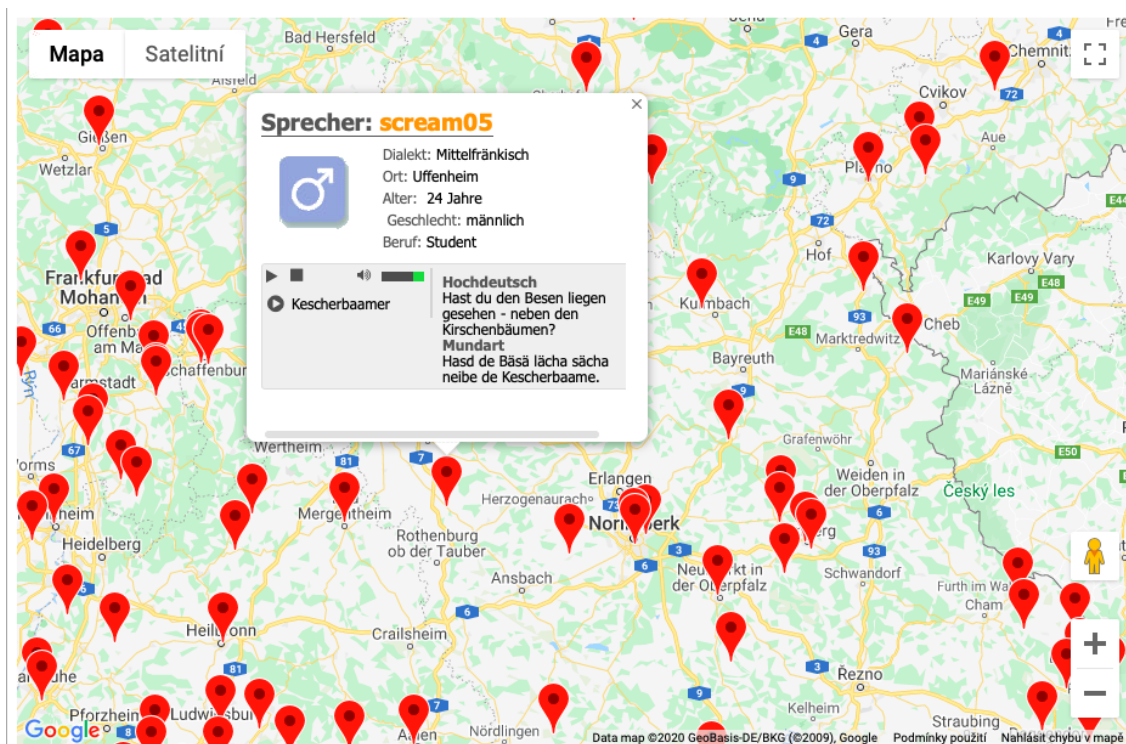
---

<sup>80</sup> LIA. Dostupné z: <https://www.hf.uio.no/iln/english/research/projects/language-infrastructure-made-accessible/index.html>.

<sup>81</sup> LIA Norwegian – Corpus of historical dialect recordings. Dostupné z: <https://www.citacepro.com/dokument/N85Ps0xlyOgMbNFo?kontrola=1>.

<sup>82</sup> Z dostupných zdrojů se nepodařilo zjistit, kdo mapovou aplikaci *Dialektkarte* provozuje.

<sup>83</sup> SPERBER, Andreas. *Dialektkarte* [online]. Dostupné z: <https://www.dialektkarte.de/>.



Obrázek 8 – Německá nářeční mapa, zdroj: <https://www.dialektkarte.de/>

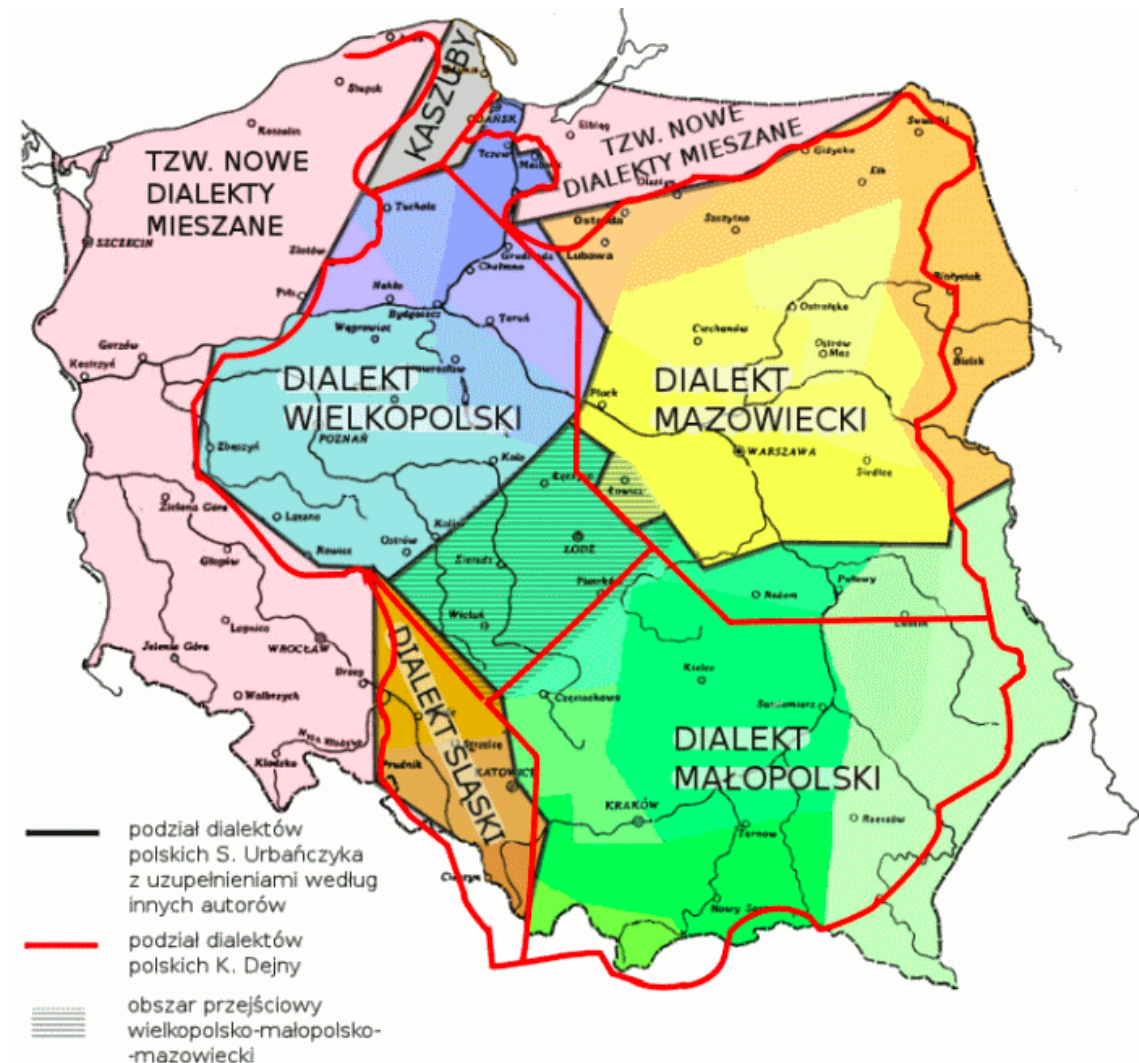
### 4.3.2 Mapové aplikace polských nářečí

Mapová aplikace polských nářečí je volně dostupná pro veřejnost a byla zrealizována Ministerstvem kultury a národního dědictví, podílel se na ní kolektiv autorů pod vedením lingvistky Haliny Karaś. Webová stránka, na níž se mapová aplikace nachází, je pojata jako všestranné kompendium popisující polské dialekty.<sup>84</sup>

V mapové aplikaci (viz obr. 9 níže) jsou pomocí barev a hranic rozděleny základní nářeční oblasti, z nichž každou lze rozkliknout pro její podrobnější rozdělení. Oblast je vždy doplněna komentářem týkajícím se např. geografie daného regionu a ke všem dialektům je připojena minimálně jedna nahrávka. Nahrávky obsahují transkripci a základní informace týkající se informátora a roku pořízení. Dostupné jsou pro většinu oblastí i slovníky shrnující několik slov běžně užívaných právě v tom daném regionu.<sup>85</sup>

<sup>84</sup> Mapa dialektów. Dostupné z: <http://www.dialektologia.uw.edu.pl/index.php?l1=opis-dialektow>.

<sup>85</sup> Tamtéž.



Obrázek 9 – Mapová aplikace polských nářečí, zdroj: <http://www.dialektologia.uw.edu.pl/index.php?l1=opis-dialektow>

Polským dialektům se dále věnuje i interaktivní mapa, jež zachycuje tato nářečí na území Litvy. Jejím provozovatelem je Filozofická fakulta univerzity ve Vilniusu. Celá platforma se snaží komplexně poukázat na souvislosti mezi polštinou, jejími dialekty a dalšími jazyky v Litvě. Na mapě je vyznačeno několik bodů (viz obr. 10) z oblastí, odkud pocházejí nahrávky. Po rozkliknutí vybraného bodu se objeví základní informace o daném místě, o informátorovi, odkazy na odbornou literaturu, nahrávka i s jejím přepisem a případně i fotografie z místa.<sup>86</sup>

<sup>86</sup> Gwary polskie na Litwie. Dostupné z: <http://www.lenkutarmes.flf.vu.lt/>.





Obrázek 10 – Polská nářečí v Litvě, zdroj: <http://www.lenkutarmes.flf.vu.lt/>

### 4.3.3 Mapová aplikace bulharských nářečí

Mapovou nářeční aplikaci bulharského jazyka poskytuje veřejnosti Institut pro bulharský jazyk, který se zabývá výzkumem a popisem jazyka jak z pohledu synchronního, tak i diachronního, spolu s Ústavem bulharské dialektologie a jazykového zeměpisu.<sup>87</sup>

Interaktivní mapa bulharštiny vyobrazuje nářeční oblasti spolu s jejich hranicemi a představuje jednotlivé dialekty. K nim jsou k dispozici popisy a v případě některých také nahrávky, všechny detaily lze na mapě zobrazit po stisknutí jednotlivých ikon. Kromě hranic nářečních oblastí jsou v mapě patrné i izoglosy, jež reprezentují dodatečné typologické charakteristiky dialektů. Příklady užití v popisech dialektů byly převzaty z jazykového atlasu, viz níže.<sup>88</sup>

<sup>87</sup> Map of the Dialectal Division of the Bulgarian Language. Dostupné z: <https://ibl.bas.bg/en/struktura/balgarska-dialektologiya-i-lingvistichna-geografiya/karta-na-dialektnata-delitba-na-balgarskiya-ezik/>.

<sup>88</sup> Tamtéž.



Obrázek 11 – Bulharská nářeční mapa, zdroj: [https://ibl.bas.bg/bulgarian\\_dialects/](https://ibl.bas.bg/bulgarian_dialects/)

#### 4.3.4 Britská zvuková mapa

*The British Library* zpřístupňuje pro veřejnost projekt *Sounds*, který zachycuje kolem 90 000 tis. zvukových nahrávek hudby, mluveného slova, přírodních zvuků apod. K daným okruhům nahrávek jsou zpřístupněny také zvukové mapy. Jedna z nich se týká právě akcentu a dialektů. Dostupné nahrávky pocházejí z *BBC Voices* (nahrávky od počátku 21. století), *Millennium Memory Bank* (nahrávky obsahující životopisná vyprávění z let 1988 a 1989), *Survey of English Dialects* (nahrávky z let 1950–1961 zachycující lidový jazyk, sběr probíhal ve 313 vesnických lokalitách napříč Anglií), nakonec z *Berliner Lautarchiv British & Commonwealth recordings* (821 digitálních záznamů z let 1915–1938 zaznamenávajících písně, lidové pohádky a vyprávění v různých jazycích a dialektech).<sup>89</sup>

<sup>89</sup> Sound Map – Accents & dialects. Dostupné z: <https://sounds.bl.uk/Sound-Maps/Accents-and-Dialects>.

## 4.4 Jazykové atlasy

Jazykové atlasy, které byly již výše obecně charakterizovány na počátku třetí kapitoly na s. 34, mohou být buď národní, nebo součástí nadnárodních projektů. Mezi národní jazykové atlasy řadíme například zmíněný ČJA. Nadnárodní jazykové atlasy představují společná díla evropských dialektologů, patří mezi ně například *Evropský jazykový atlas* či *Slovanský jazykový atlas*. Ten je „jedním z nejvýznamnějších projektů slovanské jazykovědy“.<sup>90</sup> Některé z národních jazykových atlasů jsou stejně jako ten český dostupné v elektronické podobě. Jako příklad byl vybrán *Německý jazykový atlas* a *Bulharský jazykový atlas*, které jsou popsány níže.

### 4.4.1 Německý jazykový atlas

*Německý jazykový atlas* (*Atlas zur deutschen Alltagssprache*) je pod vedením profesorů z univerzity v Salzburku a univerzity v Liège vytvářen od roku 2003 a navazuje na již dříve vytvořený projekt *Wortatlas der deutschen Kolssprachen*, jenž vycházel od roku 1977 až do konce 2. tisíciletí. Jazykový atlas si klade za cíl zachytit rozmanitost německého jazyka, pomocí on-line dotazníků (tzv. korespondenční metodou)<sup>91</sup> shromažďuje současné informace o regionálních variantách jazyka a rovněž je porovnává se starými kartami z výše zmíněného projektu. Díky tomu lze v atlasu vysledovat tendence vývoje němčiny. Atlas má doposud 11 dílů dostupných on-line na webových stránkách a nyní probíhá sběr a zpracování materiálu z dotazníků ke zpřístupnění dílu dvanáctého (dotazník je možné vyplnit přímo na stránkách atlasu). K atlasu byl vytvořen tematický rejstřík usnadňující vyhledávání.<sup>92</sup>

### 4.4.2 Bulharský jazykový atlas

*Bulharský jazykový atlas* je poskytován v knižní podobě a internetové verzi k tisku a byl, stejně jako mapová nářeční aplikace, vytvořen pod vedením Ústavu bulharské

---

<sup>90</sup> PŘADKOVÁ, Petra. *Slovanský jazykový atlas* [online], s. 3.

<sup>91</sup> Dotazníky jsou dostupné on-line pro širokou veřejnost, čímž lze shromáždit materiál ze všech potřebných míst. Otázky se tážou na užití výrazu tak, jak je to pro danou oblast typické. Navíc je mimo odpovědi na otázky možné poskytnout doplňující informace.

<sup>92</sup> ELSPAB, Stephan a Robert MÖLLER. *Atlas zur deutschen Alltagssprache* [online]. Dostupné z: <https://www.atlas-alltagssprache.de/>.



dialektologie a jazykového zeměpisu. Atlas má čtyři vydání, z nichž každé se zaměřuje na jednu oblast dialektů, vycházet začal v roce 1964. V roce 1972 byl vydán atlas mapující bulharské dialekty v Egejské Makedonii (dnešní oblast severního Řecka), o 14 let později vyšel další ve stejném formátu, avšak zaměřený na regiony Tzaribrod a Bosilegrad. V roce 2001 vyšlo zobecněné vydání jazykového atlasu zaměřující se na fonetiku, akcentaci a lexikologii, v roce 2016 pak vydání týkající se morfologie.<sup>93</sup>

---

<sup>93</sup> Publications. Dostupné z: <https://ibl.bas.bg/en/struktura/balgarska-dialektologiya-i-lingvistichna-geografiya/publikatsii/>.

## 5 CHARAKTERISTIKA ZÁPADOČESKÉ A JIHOČESKÉ NÁŘEČNÍ OBLASTI

Západočeská a jihočeská nářeční oblast jsou předmětem praktické části této diplomové práce, jež obsahuje návrh cvičení, pomocí kterých lze v korpusu DIALEKT prezentovat různé nářeční jevy specifické právě pro tyto dvě oblasti. V kapitole je teoreticky vymezena charakteristika nářečního rozrůznění českého jazyka, následně jsou popsány hlavní znaky obou oblastí spolu s jevy, na které je praktická část zaměřena.

### 5.1 Nářeční rozrůznění českého jazyka

Česká nářečí se dle *Nového encyklopedického slovníku češtiny* (dále jen NESČ) dělí do čtyř skupin.<sup>94</sup> První, největší skupinu, tvoří nářečí česká v užším smyslu. Tato oblast zaujímá západní území České republiky spolu s úseky hraničícími s jihozápadní Moravou. Na území samotné Moravy se rozkládají dvě oblasti. Ve střední části je to nářeční oblast středomoravská, také označovaná jako hanácká. Na východě při hranicích se Slovenskem je pak skupina východomoravských nářečí.<sup>95</sup> Severovýchodní okraj Moravy, území českého Slezska a přilehlé části Polska zaujímá čtvrtá nářeční skupina – slezská, nářečí jsou označována také jako lašská.<sup>96</sup> Dříve byla vyčleňována skupina slezsko-polských nářečí, současní dialektologové ji již samostatně nevydělují.<sup>97</sup>

Každá z výše zmíněných nářečních oblastí se dále dělí na menší nářeční jednotky, jako jsou podskupiny, úseky a typy, podrobný přehled dělení poskytuje *Mapa nářečí českého jazyka* vytvořená ČNK (viz obrázek 1 na s. 26), jež dodržuje stejný princip dělení a číslování oblastí jako ČJA či SNČJ a o něž se opírá i korpus DIALEKT a aplikace *Mapka*.

---

<sup>94</sup> Stejně stanovisko k rozdělení nářečních skupin do čtyř základních oblastí zaujímá i ČJA.

<sup>95</sup> Východomoravská nářečí byla dříve nazývána také jako moravskoslovenská, od této terminologie se ale upustilo a přijato bylo historiografické hledisko, které je přesnější a podle kterého jsou pojmenovány všechny české nářeční skupiny (viz Jaromír Bělič, *Nástin české dialektologie*, 1972, s. 12).

<sup>96</sup> KLOFEROVÁ, Stanislava. Nářeční skupina.

<sup>97</sup> Samostatně vyčleňuje slezsko-polská nářečí například Jaromír Bělič (*Nástin české dialektologie*, 1972).

## 5.2 Společné znaky západočeské a jihočeské nářeční oblasti

Západočeská a jihočeská nářeční oblast jsou součástí nářečí českých v užším smyslu. Obě oblasti mají znaky charakteristické pro celou skupinu českých nářečí a přilehlé jihozápadní části Moravy, tedy znaky obecněčeské. ČJA řadí mezi typické hláskoslovné jevy diftongizaci  $y > ej$  a  $ú > ou$  (*bejt, nabouchal*), úžení  $é > í$  (*fšelijakí*), přítomnost protetického *v-* před *o-* (*zavorali*), změna  $šč > št'$  (*eš'te*), změna skupiny *-nsk-*  $>$  *-nck-* (*vojenckej*).<sup>98</sup> NESČ navíc uvádí splynutí *y* s *i* (*bil = byl i bil*) a progresivní asimilaci skupiny *sh-* (*schánět, schořet*).<sup>99</sup> *Mapka* dále nezapomíná zmínit provedení přehlásky 'a > e (*práce*) a 'u > i, 'ú > í (*práci, klíč*).<sup>100</sup>

Mezi tvaroslovné znaky obecněčeské patří koncovka *-ma* v 7. pádu plurálu ve všech rodech (*barákama, holkama, japkama*), užívání neživotných tvarů adjektiv a zájmen v 1. pádu plurálu ve všech rodech (*starí chlapi, mladí holki*), životné tvary rodových zájmen v 1. pádu plurálu u feminin (*takoví mladí holki*), tvar *sem* slovesa *být* v 1. osobě singuláru (*já sem*), elipsa pomocného slovesa v 1. osobě v préteritu (*já mu dávala*) a absence koncového *-l* v příčestích minulých u maskulin (*řek, narost*).<sup>101</sup>

Výše zmíněnými znaky se vyznačuje zejména středočeská podskupina, právě z jejíž mluvy vznikl interdialekt obecná čeština, „oblast nemá žádné výrazné znaky, které by ji proti sousedním nářečním oblastem jako celek nějak vymezovaly,“<sup>102</sup> uvádí Arnošt Lamprecht. Je sice pravda, že středočeské nářeční znaky nejsou tak výrazné, avšak *Mapka* popisuje i specifické rysy nářeční oblasti středočeské, z nichž některé jsou shodné se severočeskými, zejména ty hláskoslovné, tvaroslovné se shodují jak se severočeskými, tak i jihozápadočeskými, ale uvádí se i jevy užívané pouze na části středočeského území. Specifické znaky středočeských nářečí potvrzuje i výzkum Pavla Jančáka.<sup>103</sup>

<sup>98</sup> BALHAR, Jan a kol. *Český jazykový atlas. Dodatky* [online], s. 345.

<sup>99</sup> BACHMANNOVÁ, Jarmila. Česká nářeční skupina.

<sup>100</sup> GOLÁŇOVÁ, Hana, WACLAWIČOVÁ, Martina a Jakub PEJCHA. *Mapka: Mapová aplikace pro korpusy mluvené češtiny* [online].

<sup>101</sup> Tamtéž.

<sup>102</sup> LAMPRECHT, Arnošt. *České nářeční texty*, s. 17.

<sup>103</sup> GOLÁŇOVÁ, Hana, WACLAWIČOVÁ, Martina a Jakub PEJCHA. *Mapka: Mapová aplikace pro korpusy mluvené češtiny* [online].

Západočeská a jihočeská nářeční oblast má však specifické znaky, které se odlišují od znaků obecněčeských. Dialektologové berou většinou tyto dvě oblasti jako jednu společnou oblast jihozápadočeskou, a to právě z důvodu mnoha společných nářečních rysů.<sup>104</sup> Nejpodrobněji popsal tyto znaky, které jsou uvedeny níže, Jaroslav Voráč v dílech *Česká nářečí jihozápadní I* (1955) a *Česká nářečí jihozápadní II* (1976), dále jsou doplněny informace z děl dalších dialektologů, kteří z jeho díla vycházejí. K rozdělení jihozápadočeské oblasti poprvé přistoupil ČJA z praktických důvodů, neboť se jedná o oblast velmi rozsáhlou. Dle ČJA přistoupil k tomuto rozdělení i ČNK v korpusech mluveného jazyka ORAL, ORTOFON a DIALEKT.

Z hláskoslovných rysů společných pro obě oblasti uvádí J. Voráč následující – depalatalizace  $e > a$  (*jahla*), vyrovnání ve prospěch přehlásky  $a > ě$  (*e*) (*začel*), nadměrná dvojhhláska *ou* za *ú* (které vzniklo z *ó*) (*roužnout*), zachování *í* po sykavkách (*c, z, s*) (*cíťit*), výslovnost slabikotvorného *r, l* s průvodním vokálem (*sirp*), redukce samohlásky *i, y* ve skupinách *il, yl, ly, ry* (*mlnář*), absence hlásky *g* (*ňihdo*), výslovnost skupiny *mě* jako *mje* (*mjesto*), změnu  $d > r$  mezi dvěma samohláskami (*storola*), disimilace skupin *sč* a *stř* na *chč* a *chtř* (*chtřilet*), změna *vi* na *j(i)* po samohláskách (*jalojce*), protetické *h-* a *i-, í-* na počátku slov (*itrocel, hud'it*).<sup>105</sup> *Mapka* dále uvádí užívání krátké samohlásky v infinitivech sloves (*pect*), tvary slovesa *muset se š* (*to muší d'elat*) a zpětnou (regresivní) asimilaci znělosti před souhláskami *r, l, j, m, n, ň, v* (*padesát let, třebaz v noci*),<sup>106</sup> o té se zmiňuje i ČJA (J. Voráč ji ale mezi společnými znaky neuvádí), jenž přidává jako další společný jev změnu skupiny *kd-* v *hd-* u zájmen a příslovcí (*tam nehde, je tu ňehdo*).<sup>107</sup>

Mezi tvaroslovné rysy společné pro obě oblasti řadí J. Voráč ustrnulý tvar přivlastňovacích přídavných jmen na *-ovo* (*taťínkovo bratr*) a rodinných jmen na *-ojc, -ovic* (*Novákojc, Novákovíc*), vyrovnání dlouhých a krátkých tvarů sloves vzoru *uměti* a *sázeti* ve prospěch krátkých ve 3. osobě plurálu (*umí, sází* místo *umějí, sázejí*), nakonec uvádí některé společné tvary imperativů (*noš, řekňte*).<sup>108</sup> *Mapka* přidává další znaky

<sup>104</sup> Společným jihozápadočeským nářečním jevům se věnuje několik cvičení v praktické části, viz oddíl *Hláskosloví*, cv. 1 (s. 65) a cv. 2 (s. 71); oddíl *Tvarosloví*, cv. 9 (s. 94), cv. 10 (s. 97), cv. 11 (s. 100), cv. 12 (s. 103) a cv. 13 (s. 106).

<sup>105</sup> VORÁČ, Jaroslav. *Česká nářečí jihozápadní. Studie jazykově zeměpisná*, s. 9–10.

<sup>106</sup> GOLÁŇOVÁ, Hana, WACLAWIČOVÁ, Martina a Jakub PEJCHA. *Mapka: Mapová aplikace pro korpusy mluvené češtiny* [online].

<sup>107</sup> BALHAR, Jan a kol. *Český jazykový atlas. Dodatky* [online], s. 345.

<sup>108</sup> VORÁČ, Jaroslav. *Česká nářečí jihozápadní. Studie jazykově zeměpisná*, s. 10.

charakteristické pro obě skupiny – kromě ustrnulého tvaru přivlastňovacích přídavných jmen na *-ovo* uvádí i tvary na *-ino/-íno* (*sestřino/sestřino d'eti*). Dále uvádí koncovku *-í* v 1. pádu plurálu podstatných jmen, zájmen a číslovek (*montéři*), jež je typická spíše pro západočeský úsek, rozlišování tvarů 2. pádu singuláru oproti 3. a 6. pádu singuláru feminin u skloňování adjektivního tvrdého a skloňování rodových zájmen (*do tí Prahi, ftej druhej tříde*), tvar *mi* zájmena *já* ve 3. pádu (*husi mi utekli*), vyrovnání koncovky měkkého skloňování (typ *o něm*) maskulin a neuter rodových zájmen v 7. pádu singuláru se skloňováním tvrdým (typ *o tom*) (*cht'el eš'ťe vo ňom ňeco ved'et*), tvar *si* slovesa *být* ve 2. osobě singuláru (*ti si tadi vocut*), nakonec tvar *bi si* v kondicionálu 2. osoby singuláru (*abi si šla*).<sup>109</sup>

### 5.2.1 Charakteristické znaky západočeské nářeční oblasti

Západočeská nářeční oblast se rozkládá v západním cípu nářečí českých v užším smyslu, oblast sousedí na jihovýchodě s nářečími jihočeskými, na severovýchodě pak s oblastí středočeskou.

Mezi charakteristické znaky hláskoslovné typické pro tuto oblast patří vkladné (epentetické) *j* před měkkými zubnicemi (*d', t'*) (*nechojt'*), dochází zde ke krácení samohlásky *-á* v 1. osobě singuláru slovesných tvarů a v tvarech zájmen (*dam, vam*), změna přípony *-oval* > *-uval* (*kupuval*) a redukovaná výslovnost samohlásky ve slabikách *ní* a *ní* (*pjekna počasí, prvňá d'eti*).<sup>110</sup> Některé hláskoslovné znaky jsou typické ovšem jen pro část západočeského území. Pro všechny okrajové úseky je příznačná zavřená výslovnost dvojhlásek *ou* a *ej* (*sejd<sup>o</sup>u se, zlat<sup>u</sup> svarbu*). Pro okrajový úsek domažlický (chodský) je charakteristická široká výslovnost samohlásky *á* jako *ä* (*po svätčích*), délka v dvojslabičných substantivech, zájmených tvarech, přičestích minulých a infinitivech (*d'ífče, naše, vorál, vorát*), typicky chodské je protetické *h-* na počátku slov (*hucháč, hřešato*), změna *d* > *r* (*storola, bure*), zjednodušení skupiny *vj* > *j* (*sjet*) a výslovnost slabikotvorných souhlásek *r, l* s průvodním vokálem (*sarce, valk*). Pro okrajový úsek manětínský je typická výslovnost skupiny *mě* jako *mje* (*mjesto*), tento znak se objevuje i na Stříbrsku, kde jsou uchovávány další archaické znaky, například pozůstatky tvrdého

<sup>109</sup> GOLÁŇOVÁ, Hana, WACLAWIČOVÁ, Martina a Jakub PEJCHA. *Mapka: Mapová aplikace pro korpusy mluvené češtiny* [online].

<sup>110</sup> Tamtéž.

*l* (*do stodol*), změna skupiny *-ovi* > *-oj* (*bratroj*) či zjednodušení skupiny *dn* > *nn* > *n* (*jennou, ve ne*).<sup>111</sup>

Západočeská nářeční oblast jako celek se však vyznačuje zejména výraznými rysy tvaroslovnými. Mezi ně se řadí koncovka *-om*<sup>112</sup> ve 3. pádu plurálu substantiv ve všech rodech (*t'ém muzikantom, tu'ím d'etom*), analogické vyrovnání tvarů u podstatných jmen vzorů *růže* a *píseň* podle vzoru *kost* (*jít eš'te do nemoc'nici*), tvar zájmena *mi* namísto *m'ne* ve 4. pádu (*vo'ni mi požádali*), ve 3. pádu pak tvar *t'e* namísto *t'i* (*já t'e pomůžu*) a tvar ukazovacího zájmena typu *tuten* (*tuten kluk, tuti krámi*). Tvary ukazovacích zájmen zde mají analogické koncovky podle měkkého skloňování (vzor *náš*) (*vot t'ich šest'i let*), tvar přičestí trpného bývá zakončen na *-eno* (*navareno, je zavřeno*). Charakteristické je dále odsouvání koncového *-k* u slovtvorného formantu *-pak* (*copa bich se bál, jakipa to bulo*).<sup>113</sup> Pro okrajové úseky je příznačným tvaroslovným znakem tvar *je* zájmena *ono* ve 4. pádu (*přivedla d'ít'e, a nehlídala si je*). Dříve se na Chodsku vyskytovala podoba přičestí *bul* namísto *byl*, podle které byli Chodové nazýváni jako Buláci, dnes ale již není dochována.<sup>114</sup>

Západočeská nářeční oblast má i svá lexikální specifika, k nimž přispívá zejména Chodsko. Je to totiž region specifický nejen pro své nářečí, ale i etnografii, stále se zde uchovávají lidové tradice. V místním lexiku se tedy vyskytují tzv. etnografismy (dialektology bývají označovány také jako specifické dialektismy), které označují „předměty a jevy specifické pouze pro život v určitém regionu (např. názvy částí kroje).“<sup>115</sup> Slovní zásoba z této oblasti je zpracována v *Dialektickém slovníku chodském* (1907), jehož autorem je národopisec a dialektolog Jan František Hruška, dále v *Chodském slovníku*, jenž byl sestaven Josefem Kotalem ku příležitosti 40. výročí úmrtí etnografa a sběratele lidových písní Jindřicha Jindřicha. Tento slovník byl původně součástí velkého národopisného díla *Chodsko* (1956). Z novějších publikací je chodské

---

<sup>111</sup> BACHMANNOVÁ, Jarmila. Česká nářeční skupina.

<sup>112</sup> Samohlásky v koncovkách ve 3. a 6. pádech jsou kráceny, pravidelně u jmen rodu ženského v případě, že nepřijímají koncovku *-om* (*ženam, ženach*) (JANČÁK Pavel. K jazykovému zeměpisu Čech. Některé případy západočeského krácení v koncovkách).

<sup>113</sup> GOLÁŇOVÁ, Hana, WACLAWIČOVÁ, Martina a Jakub PEJCHA. *Mapka: Mapová aplikace pro korpusy mluvené češtiny* [online].

<sup>114</sup> BACHMANNOVÁ, Jarmila. Česká nářeční skupina.

<sup>115</sup> HLADKÁ, Zdeňka. Dialektismus.

lexikum obsaženo v díle *Plzeňsko-český slovník: pročpa tudlecto říkáme?* (2017), který vznikl z iniciativy neodborné veřejnosti a jehož autorem je Jan Hajšman.

### 5.2.2 Charakteristické znaky jihočeské nářeční oblasti

Jihočeská nářeční oblast hraničí na západě se západočeskou oblastí, na severu s nářečím středočeským, na východě pak s oblastí českomoravskou, zejména se žďársko-bystřickým úsekem.

Charakteristickým hláskoslovným jevem této oblasti je tzv. jihočeská disimilace<sup>116</sup> – rozrůznění zdvojených sykavek (*aj sme se na tom dohodli, u naj se to d'elalo*).<sup>117</sup> Typická je i výslovnost skupiny *mě* jako *mje*<sup>118</sup> (i na Manětínsku) (*mjesto*). Další hláskoslovné jevy jsou příznačné pro Doudlebsko, okrajový úsek této oblasti. Charakteristická je výslovnost dvojhlásek *ej* a *ou* s různým stupněm zavřenosti (*jedn<sup>o</sup>u, takov<sup>u</sup>, vikinut<sup>e</sup>j*), v koncovkách se zachovává délka samohlásek (*nevím*), absence protetického *v-* (*otpoledne*), rozložená výslovnost měkkých retnic (*bj, pj, mj, vj*) (*babjička, krmjili*), předpona *ne-* namísto *ňe-* u zájmen a zájmenných příslovčí (*negde*) a progresivní asimilace znělosti ve skupinách *kv-* a *tv-* (*kfočna, potfora*).<sup>119</sup> NESČ navíc uvádí mezi hláskoslovnými rysy Doudlebska stopy po průvodních vokálech u slabikotvorných likvid a stopy po archaickém tvrdém *l*.<sup>120</sup>

Z tvaroslovného hlediska nemá jihočeská nářeční oblast tolik specifických rysů jako ta západočeská. Typicky jihočeský je 7. pád plurálu podstatných jmen typu *s pekaři, s poli*. Imperativ sloves se objevuje s měkkou souhláskovou skupinou (*noš, odpušť*).<sup>121</sup> NESČ uvádí ještě změnu skupiny *-ovi* na *-ou* (*Horažďouce, Novákouc*)<sup>122</sup>, tento jev nicméně J. Voráč nezmiňuje, uvádí pouze změnu slabiky *-vi-* na *-ji-* nebo *-j-* po

<sup>116</sup> Disimilace je změna hlásky v proudu řeči, která vede k rozlišení hlásky, naopak asimilace spočívá ve vzájemném přizpůsobování hlásek.

<sup>117</sup> GOLÁŇOVÁ, Hana, WACLAWIČOVÁ, Martina a Jakub PEJCHA. *Mapka: Mapová aplikace pro korpusy mluvené češtiny* [online].

<sup>118</sup> Výslovnost skupiny *mě* jako *mje* je pozůstatkem měkkých retnic, což je jeden z mnoha znaků, které mají tyto dvě oblasti společné s okrajovými nářečemi v rámci celého jazykového systému České republiky.

<sup>119</sup> Tamtéž.

<sup>120</sup> BACHMANNOVÁ, Jarmila. Česká nářeční skupina.

<sup>121</sup> GOLÁŇOVÁ, Hana, WACLAWIČOVÁ, Martina a Jakub PEJCHA. *Mapka: Mapová aplikace pro korpusy mluvené češtiny* [online].

<sup>122</sup> BACHMANNOVÁ, Jarmila. Česká nářeční skupina.

samohláskách (*jalojce, jalojice*).<sup>123</sup> I Doudlebsko má opět svá tvaroslovná specifika – tvar *mje* nebo *me* zájmena *já* ve 4. pádu singuláru (*přivítal me*) a tvar *som* slovesa *být* v 1. osobě singuláru v přítencu (*šel som pomoct*).<sup>124</sup> NESČ opět *Mapku* doplňuje a dodává jako tvaroslovný rys koncovku *-om* ve 3. pádu plurálu maskulin (*sousedom*), což je typický rys západočeských nářečí, dále výskyt životných tvarů adjektiv v přívlastkovém postavení (*staří přátelé*), tvar *je* zájmena *ono* ve 4. pádu singuláru (příznačné i pro okrajové úseky západočeských nářečí) (*prase povalili a hodili je do trokú*) či existenci zdvojených dativních tvarů (*to ti tebe povídám*). Na Prachaticku se dříve užíval infinitiv zakončený na *-ti*, dnes jen přežívá ve formě *-t'* (*chtít', rosčilit' se*), z toho důvodu se nářečí nazývalo jako tiťácké.<sup>125</sup>

I jihočeská nářeční oblast má opět svá lexikální specifika. Stejně jako v západočeské oblasti přispívá etnografismy Chodsko, zde je to etnografický region Doudlebska. Slovní zásobu Doudlebska zachycují díla dialektologa Zbyňka Holuba *Lexikon nejjižnějšího úseku českých nářečí* (2003), v dalším vydání z roku 2004 pod názvem *Doudlebské nářečí a slovník*, dále dílo jazykovědce Františka Jílka *Jihočeský člověk a jeho řeč* (1961).

### 5.3 Nářeční jevy západočeské a jihočeské v korpusu DIALEKT

Nářeční jevy typické pro západočeskou a jihočeskou oblast uvedené v podkapitole 5.2 výše lze vyhledat v korpusu DIALEKT a pracovat s nimi. Vyhledat se dají jak jednotlivé hlásky a lemmata, v nichž se nejčastěji vyskytují, tak i tvary slov typické pro dané oblasti. Hledání v korpusu je možné omezit pouze na určité nářeční oblasti, díky čemuž se objeví jen výsledky z nahrávek pořízených právě ve vybraných lokalitách, lze si tak vytvořit tzv. subkorpus zahrnující nahrávky pouze z dané oblasti. Práce v korpusu DIALEKT probíhá přes rozhraní KonText<sup>126</sup>, v němž lze vyhledávat buď na rovině dialektologické, nebo ortografické, přičemž klíčová je právě rovina dialektologická, v níž jsou zaznamenány jak všechny nářeční hlásky (viz *Archiv diferenčních hlásek nářečí českého jazyka*), tak i tvaroslovné jevy.

<sup>123</sup> VORÁČ, Jaroslav. *Česká nářeční jiozápadní. Studie jazykově zeměpisná*, s. 9.

<sup>124</sup> GOLÁŇOVÁ, Hana, WACLAWIČOVÁ, Martina a Jakub PEJCHA. *Mapka: Mapová aplikace pro korpusy mluvené češtiny* [online].

<sup>125</sup> BACHMANNOVÁ, Jarmila. *Česká nářeční skupina*.

<sup>126</sup> Dostupné na adrese [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v1\\_dial](https://www.korpus.cz/kontext/query?corpname=dialekt_v1_dial).



Podrobně se popisu vybraných jevů v korpusu věnuje článek H. Goláňové a M. Waclawičové *Vybrané hláskové jevy severovýchodočeské nářeční oblasti v korpusu DIALEKT*, jenž byl východiskem pro následující kapitolu i celou diplomovou práci. V 6. kapitole je na vybraných jevech z oblasti jihozápadočeské, západočeské, jihočeské, dále z Chodska a Doudlebska ukázáno, jak je lze v korpusu vyhledat a co nám dané informace o nářečí říkají.

## 6 CVIČENÍ ZAMĚŘENÁ NA NÁŘEČNÍ JEVY V KORPUSU DIALEKT

Praktická část diplomové práce se zaměřuje na tvorbu cvičení, pomocí kterých lze vyhledávat vybrané nářeční jevy jihočeské a západočeské oblasti v korpusu DIALEKT.<sup>127</sup> Cílem cvičení je demonstrovat na živém jazykovém materiálu typické nářeční znaky a ukázat, jak je lze prezentovat v hodinách českého jazyka při výuce nářeční problematiky na středních školách. Níže je popsáno, jak probíhala tvorba cvičení, jak se s nimi má pracovat, dále cvičení samotná s postupem a obrázky, nakonec komentář výsledků. Seznam všech cvičení je uveden v příloze této práce.

Práce obsahuje celkem 24 cvičení, která jsou rozdělena do 3 kategorií – první kategorie se zaměřuje na jevy hláskoslovné, druhá na tvaroslovné a poslední na nářeční slovní zásobu, každá kategorie obsahuje 8 cvičení, ta jsou sestavena k vyhledávání v rozhraní KonText. Cvičení jsou v každém oddílu seřazena dle oblastí – nejprve jevy společné pro celou jihozápadočeskou oblast, následně se zaměřením na oblast západočeskou, pak jihočeskou, dále jevy společné pro Chodsko a Doudlebsko, poté samostatně jevy typické pro chodský region, nakonec pro doudlebský typ. Tímto způsobem jsou řazena i data v teoretické části práce, stejně postupuje také aplikace *Mapka*. V lexikálním oddílu jsou cvičení rozdělena do třech tematických celků. Všechna data pocházejí z druhé verze korpusu DIALEKT a jsou aktuální k 1. 5. 2022.

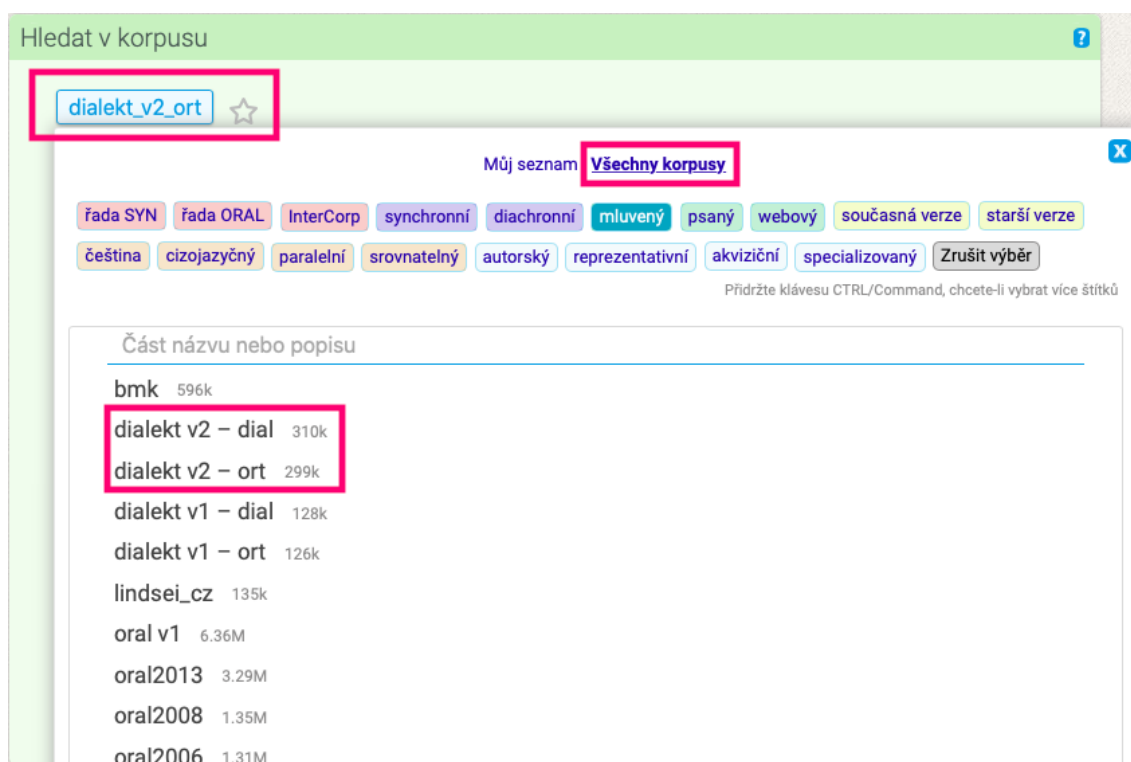
Pro tvorbu cvičení byly využity přepisy nahrávek ze západočeské a jihočeské oblasti obsažených v korpusu DIALEKT, z nichž byly vybrány typické nářeční jevy, které jsou teoreticky popsány v kapitole č. 5. Tyto znaky byly promítnuty do korpusu pomocí výše zmíněného vyhledávače a na základě výsledků byla sestavena cvičení tak, aby měla dostatečnou výpovědní hodnotu a skutečně ukázala, které nářeční jevy jsou stále typické a dodnes živé a které lze naopak považovat za již neužívané. Pro cvičení zaměřená na lexikum byla využita jak výše zmíněná korpusová data, tak data z ČJA. Cvičení jsou sestavena tak, aby provedla uživatele základní prací s korpusem, aby seznámila s hlavními regulárními výrazy a základními funkcemi vyhledávače KonText, taktéž aby

---

<sup>127</sup> Typické nářeční jevy západočeské a jihočeské jsou popsány i v aplikaci *Mapka*, některé z nich jsou zachyceny přímo v ukázkových kartách. Doporučujeme vybrané nářeční jevy v aplikaci vyhledat, poté pomocí cvičení v korpusu jejich užívání ověřovat.

seznámila s prací s ČJA. Postupně jsou ve cvičeních ukázány postupy od nejjednodušších po složitější, aby byla práce co nejvíce pestrá a ukázala možnosti, které práce s korpusovým manažerem nabízí.

Každé cvičení obsahuje zadání, přesný postup, jak daný jev v korpusu vyhledat, společně s obrázky pro jednodušší orientaci, nakonec řešení spolu s komentářem, co výsledek o zadaném jevu vypovídá. Pro zpracování cvičení je nutné provést bezplatnou registraci na webových stránkách ČNK<sup>128</sup>, aby bylo možné vyhledávat v korpusovém manažeru KonText<sup>129</sup>. Při vyhledávání je nutné specifikovat typ korpusu na *dialekt – v2 dial* nebo *dialekt – v2 ort*<sup>130</sup> dle povahy dat, některé jevy se objevují i na rovině ortografické, některé ale pouze na rovině dialektologické, viz obrázek č. 12 níže. V postupu u každého cvičení je uvedeno, na které rovině má vyhledávání probíhat.



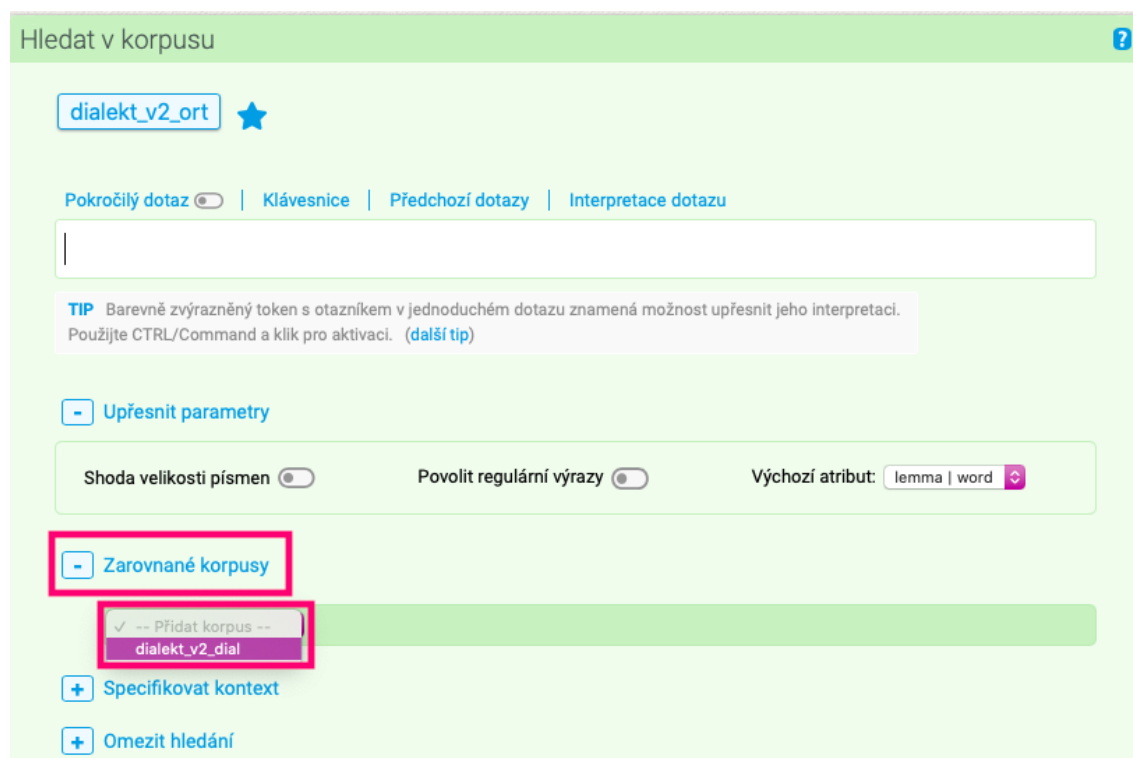
Obrázek 12 – Omezení vyhledávání na korpus DIALEKT, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_ort](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_ort)

<sup>128</sup> Dostupné z: <https://www.korpus.cz/>.

<sup>129</sup> Dostupné z: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial).

<sup>130</sup> Označení v2 znamená, že vyhledávání bude probíhat ve druhé verzi korpusu, *dial* značí rovinu dialektologickou, *ort* rovinu ortografickou.

Korpusový manažer KonText navíc nabízí možnost zobrazit si při vyhledávání obě roviny vedle sebe jako tzv. paralelní korpusy. Pokud budeme vycházet při vyhledávání z roviny dialektologické (*dialekt – v2 dial*), vybereme jako paralelní korpus *dialekt – v2 ort* (a naopak). Paralelní korpus vybereme pomocí tlačítka *Zarovnané korpusy*, následně *Přidat korpus* (viz obrázek 13 níže).



Obrázek 13 – Přidání zarovnaného korpusu při vyhledávání, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_ort](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_ort)

Další možností je použít korpusové nastavení tak, aby bylo slovo na obou rovinách zobrazeno jako KWIC. V horním řádku zvolíme záložku *Zobrazení* a možnost *Korpusová nastavení...*, viz obrázek 14. Následně se objeví tabulka, ve které vybereme jako druhý poziční atribut, který se bude zobrazovat, rovinu *ort/dial*, viz obrázek 15. Tím zajistíme, že se nám vedle KWIC bude objevovat i jeho ekvivalent na druhé rovině.

KonText interface showing search results for a dialect corpus. The interface includes a search bar, navigation tabs, and a list of search results. The 'Zobrazení' (Display) and 'Korporusová nastavení' (Corpus settings) options are highlighted in red boxes.

Obrázek 14 – Volby zobrazení a další korpusová nastavení v manažeru KonText, zdroj: [https://www.korpus.cz/kontext/view?maincorp=dialekt\\_v2\\_dial&viewmode=kwic&pagesize=40&attrs=word%2Cort&attr\\_vmode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~1AGq0mayoow0](https://www.korpus.cz/kontext/view?maincorp=dialekt_v2_dial&viewmode=kwic&pagesize=40&attrs=word%2Cort&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~1AGq0mayoow0)

Korporusová nastavení pro dialekt\_v2\_dial dialog box. The 'Poziční atributy' (Positional attributes) tab is selected. The 'ort' attribute is checked under the 'Zobrazovat' (Display) column.

	Zobrazovat	Hlavní
word [dial]	<input checked="" type="checkbox"/>	<input checked="" type="radio"/>
lc [lowercase dial]	<input type="checkbox"/>	<input type="radio"/>
lemma	<input type="checkbox"/>	<input type="radio"/>
lemma_lc [lowercase lemma]	<input type="checkbox"/>	<input type="radio"/>
tag	<input type="checkbox"/>	<input type="radio"/>
pos [part of speech]	<input type="checkbox"/>	<input type="radio"/>
case [grammatical case]	<input type="checkbox"/>	<input type="radio"/>
ort	<input checked="" type="checkbox"/>	<input type="radio"/>
lc [lowercase ort]	<input type="checkbox"/>	<input type="radio"/>
uid	<input type="checkbox"/>	<input type="radio"/>

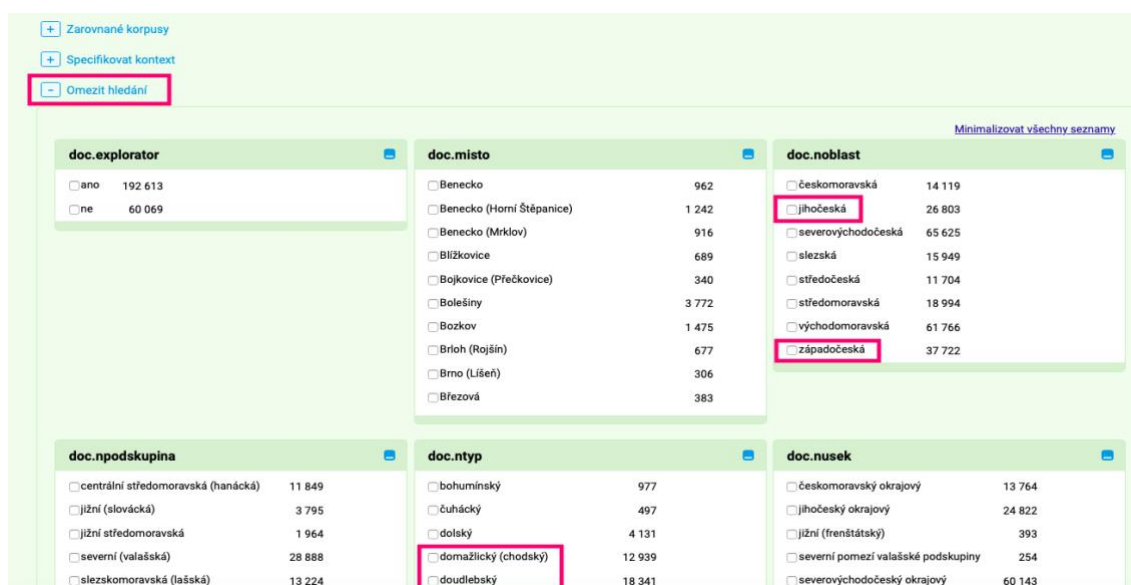
Obrázek 15 – Výběr roviny ort jako dalšího pozičního atributu ke zobrazení, zdroj: [https://www.korpus.cz/kontext/view?maincorp=dialekt\\_v2\\_dial&viewmode=kwic&pagesize=40&attrs=word%2Cort&attr\\_vmode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~1AGq0mayoow0](https://www.korpus.cz/kontext/view?maincorp=dialekt_v2_dial&viewmode=kwic&pagesize=40&attrs=word%2Cort&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~1AGq0mayoow0)

V některých případech bude třeba povolit pod vyhledávacím oknem tzv. regulární výrazy, jedná se o speciální znaky se specifickým významem, jejichž pomocí lze vyhledávání omezit a hledat například jen části určitých slov. V jiných případech bude zase třeba vyhledávat pomocí pokročilého dotazu za pomoci tzv. CQL dotazovacího jazyka, zvláště pokud se bude jednat o složitější zadání. To nám umožní dotaz za použití dalších konkrétních znaků specifikovat, abychom nemuseli výsledky procházet a selektovat ručně. Důležitý bude pro některý typ cvičení taktéž výchozí atribut, nejčastěji budeme vyhledávat *word* (realizace určitého *lemmatu*) (viz obrázek 16).



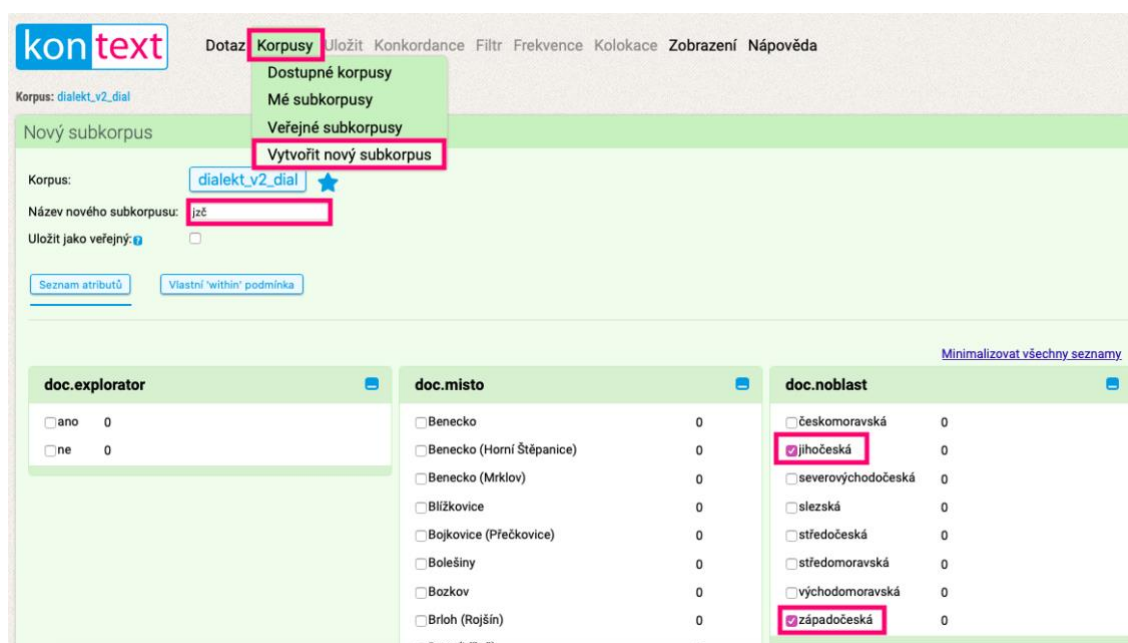
Obrázek 16 – Povolení regulárních výrazů, pokročilého dotazu, volba výchozího atributu při vyhledávání v korpusu, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_ort](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_ort)

Abychom zajistili, že korpus bude vyhledávat pouze výsledky z jihočeské nebo západočeské nářeční oblasti, musíme omezit vyhledávání a specifikovat, z jaké oblasti mají výsledky pocházet. Po stisknutí tlačítka *omezit vyhledávání* se otevře několik menších tabulek s možnostmi, nejprve je nutné v tabulce *doc.noblust* vybrat jihočeskou, západočeskou nebo obě oblasti. Dále lze specifikovat i typ nebo úsek (*doc.ntyp*, *doc.nusek*) a vyhledávat tak pouze v nahrávkách z Chodska nebo Doudlebska, jak je vidět na obrázku 17 níže.



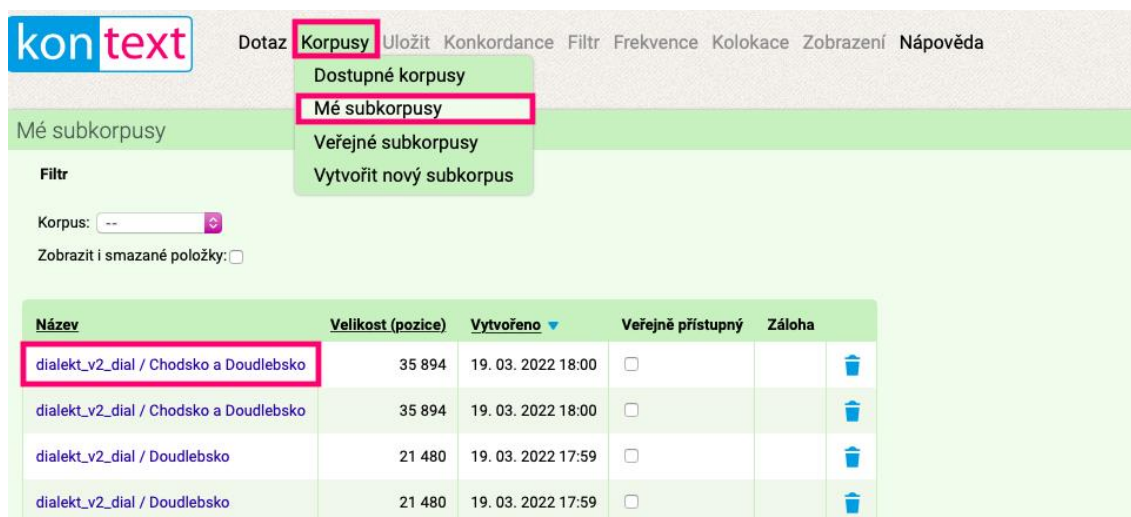
Obrázek 17 – Omezení vyhledávání na oblast a typ, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_ort](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_ort)

Dalším způsobem, jak omezit hledání na oblast či typ, je vytvoření tzv. subkorpusu. V horní liště nabídek vybereme záložku *Korpusy* a dále možnost *Vytvořit nový subkorpus*. Následně zvolíme korpus, tedy *dialekt – v2 dial* a subkorpus pojmenujeme tou oblastí, ve které budeme vyhledávat, tu následně zaškrtneme stejně jako při omezení hledání (viz obrázek 17). Pro pojmenování subkorpusů můžeme použít zkratky, například název *jzč* pro jihozápadočeskou oblast, *zč* pro západočeskou, *jč* pro jihočeskou, následně subkorpus *Chodsko* pro chodský typ jihozápadočeské oblasti, *Doudlebsko* pro doudlebský typ jihočeské oblasti, nakonec subkorpus *Chodsko a Doudlebsko* pro oba typy najednou. Vytvořením subkorpusu odpadá nutnost omezovat hledání na danou oblast. Postup je znázorněn níže na obrázku 18.



Obrázek 18 – Vytvoření subkorpusu, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial&usesubcorp=Chodsko%20a%20Doudlebsko](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial&usesubcorp=Chodsko%20a%20Doudlebsko)

Pokud budeme při vyhledávání využívat možnost subkorpusu namísto omezení hledání, na horní liště opět zvolíme záložku *Korpusy*, dále *Mé subkorpusy*, vybereme ten, který potřebujeme (viz obrázek 19 níže). Následně budeme přesměrováni na základní stránku manažeru KonText, ve kterém bude daný subkorpus už automaticky nastaven.



Obrázek 19 – Postup při výběru subkorpusu pro vyhledávání, zdroj: <https://www.korpus.cz/kontext/subcorpus/list>

Při práci s ČJA, která je do cvičení zařazena ve všech oddílech, budeme vyhledávat v elektronické verzi atlasu i v naskenované verzi k tisku.<sup>131</sup> Někdy nastane situace, kdy elektronická verze atlasu nevyhledá žádný doklad přesto, že je forma v atlasu zachycena. Na tomto místě je třeba ověřovat položky v abecedně řazeném Rejstříku jazykových dokladů, který je dostupný jak v elektronické verzi, tak i ve verzi k tisku, a to v *Dodatcích*.<sup>132</sup> V odkazech na ČJA je uvedena zkratka dílu (např. ČJA 1 – díl první), strana, na níž se ve verzi k tisku objevuje daná položka či jazykový doklad, a strana, na které jsou výrazy mapovány.

<sup>131</sup> Elektronická verze dostupná z: <https://cja.ujc.cas.cz/e-cja/>, naskenovaná verze k tisku dostupná z: <https://cja.ujc.cas.cz/>.

<sup>132</sup> Dostupné z: <https://cja.ujc.cas.cz/CJA-dodatky/>.

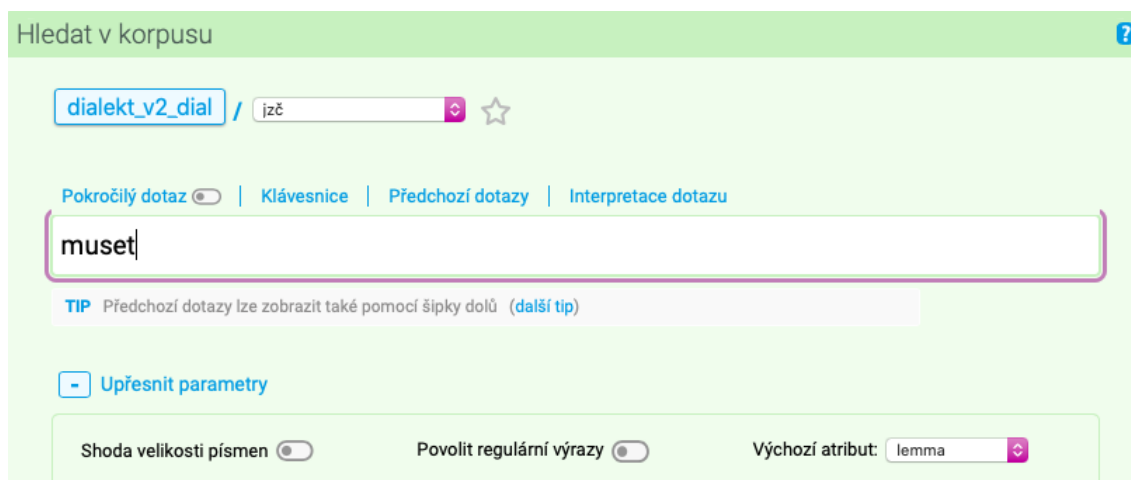


## 6.1 HLÁSKOSLOVÍ

### Cvičení č. 1 – Výslovnost tvarů slovesa *muset* se *š* v západočeské a jihočeské nářeční oblasti

Zadání: Sloveso *muset* je v západočeské a jihočeské nářeční oblasti tradičně v různých tvarech vyslovováno se *š* namísto *s*. Vyhledejte všechny slovní tvary slovesa *muset* v korpusu *DIALEKT*. Srovnejte výskyty spisovné a nářeční výslovnosti v těchto tvarech – převažuje nářeční výslovnost nad tou spisovnou? Ve které oblasti se objevuje častěji, jihočeské nebo západočeské?

Postup: Výslovnost slovních tvarů slovesa *muset* budeme vyhledávat na rovině dialektologické, zvolíme tedy korpus *dialekt* – v2 *dial* (výslovnost se na rovinu ortografickou převádí do spisovné podoby). Omezíme hledání na jihočeskou a západočeskou nářeční oblast (viz obrázek 17), nebo pro vyhledávání zvolíme subkorpus *jzč* (viz obrázek 19). Regulární výrazy nebudeme při vyhledávání potřebovat. Jako výchozí atribut pro vyhledávání zvolíme *lemma*. Podoba dotazu viz obrázek 20 níže.



Hledat v korpusu

dialekt\_v2\_dial / jzč

Pokročilý dotaz | Klávesnice | Předchozí dotazy | Interpretace dotazu

muset

TIP Předchozí dotazy lze zobrazit také pomocí šipky dolů (další tip)

Upřesnit parametry

Shoda velikosti písmen Povolit regulární výrazy Výchozí atribut: lemma

Obrázek 20 – Podoba dotazu k vyhledání slovních tvarů lemmatu *muset*, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial&usesubcorp=jzč](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial&usesubcorp=jzč)

Po vyhledání takto zadaného dotazu se objeví tzv. konkordance (všechny doklady hledaného jevu), uprostřed konkordančního řádku se nachází růžově označené KWIC (key word in context), tzv. klíčové slovo, tedy hledaný jev. Na levé straně nahoře je počet výskytů, pod ním se nachází kód nahrávky, ze které výskyt pochází, viz obrázek 21 níže).

Výskytů: 209 | i.p.m.: 2 775,49 (vztaheno k subkorpusu "jzč") | ARF: - | Výsledek je promíchán

Vyber řádků: základní

<input type="checkbox"/>	0865-C-JIC-2009-N	mám zážitek škaredej.   +   no a ráno sem to zase   <b>musela/musela</b>   fšecko viklidit, podlahu umejit a mohla sem jit potom
<input type="checkbox"/>	0863-C-JIC-2011-N	se to delalo .   +   (rušivý zvuk)   +   (nadechnuti)   +   (bouchnuti)   +   tak to   <b>musel/musel</b>   vzid do klesti, vofej, no a teřkon tu
<input type="checkbox"/>	0885-C-JIC-2011-N	zatach a bilo to hotovi .   +   (nadechnuti)   +   ale muselo .   <b>musel/musel</b>   to mid na nakej struie nebo neco, že to
<input type="checkbox"/>	0891-C-JIC-2011-N	tam bil sám, tak to ...   +   (zvuk ptáka)   +   (nadechnuti)   +   (zvuk ptáka)   +     <b>musel/musel</b>   d'elat fšecno .   +   (pousmán)   +   (zvuk ptáka zvuk ptáka)   +
<input type="checkbox"/>	0289-C-ZAC-2011-N	aři korunu .   +   (nadechnuti)   +   (rušivý zvuk)   +   (rušivý zvuk)   +   (nadechnuti)   +   a Karel potom   <b>musel/musel</b>   jit topid do kasáren, celou zimu topil f kasárnech
<input type="checkbox"/>	0025-C-JIC-1964-SA	zimu se musela pohnojit, no, po jaru se   <b>musel/musel</b>   ten hnuj schrabat, anebo diš to bil kompost,
<input type="checkbox"/>	0877-C-JIC-2011-N	šesnáctim roce, hmm .   +   (zvuk ptáka)   +   (nadechnuti)   +   no a to   <b>musel/musel</b>   bejd hlug hroznej, ne, s toho, s
<input type="checkbox"/>	0025-C-JIC-1964-SA	ten hnuj schrabat, anebo diš to bil kompost,   <b>musel/musel</b>   se, feřkeme, na zimu rozházet, no,
<input type="checkbox"/>	0025-C-JIC-1964-SA	to se musela ta louka, tedi na zimu se   <b>musela/musela</b>   pohnojit, no, po jaru se musel ten hnuj
<input type="checkbox"/>	0025-C-JIC-1964-SA	míel pachtovaneřich vot opce .   +   no, a to se   <b>musela/musela</b>   ta louka, tedi na zimu se musela pohnojit,
<input type="checkbox"/>	0589-C-ZAC-2005-N	vite, abí to ti opčáni podepsali, tak sem   <b>musela/musela</b>   já .   +   (mlasknuti)   +   jo, jo, jo .   +   vobjehnout
<input type="checkbox"/>	0903-C-JIC-2009-N	bála .   +   to bila taková železná tič a to sem   <b>musela/musela</b>   to stavidlo pusřid dólů, abí ta voda tam neřla
<input type="checkbox"/>	0376-C-JIC-2009-N	bilo třináč a půl .   +   a půl roku sem eřfe   <b>musela/musela</b>   chodíř do Todře do řkoli, do feč řřmáři .

Obrázek 21 – Popis konkordance po vyhledání dotazu, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr\\_ymode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~xOe8Kqc6Mycq](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr_ymode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~xOe8Kqc6Mycq)

Korpus vyhledal pro tento dotaz celkem 209 slovních tvarů lemmatu *muset*. V nich jsou ale zahrnuty varianty se *s* i *š*. Abychom nemuseli manuálně procházet výsledky a hledat varianty, v nichž se vyskytuje *š*, použijeme frekvenční distribuci slovních tvarů pomocí tabulkového zobrazení. V horní liště zvolíme záložku *Frekvence*, následně *Slovní tvary [A = a]*, zobrazí se graf, nad nímž vybereme možnost *Tabulkové zobrazení*. V tabulce se zobrazí dohromady 37 slovních tvarů, prvních 10 nejčastějších viz obrázek 22 níže.

Frekvenční seznam

Zobrazení v grafech | **Tabulkové zobrazení**

Zobrazit konfidenční intervaly  ⓘ

1 / 1 (celkem: 37 položek)

	Filtr	word [dial]	Freq	i.p.m.
1	p / n	mušeli	39	517,91
2	p / n	museli	29	385,12
3	p / n	mušel	24	318,72
4	p / n	musela	15	199,02
5	p / n	muselo	13	172,64
6	p / n	mušela	13	172,64
7	p / n	mušelo	13	172,64
8	p / n	muší	11	146,08
9	p / n	musel	7	92,96
10	p / n	muší	7	92,96

Obrázek 22 – Tabulkové zobrazení frekvenční distribuce slovních tvarů lemmatu *muset*, zdroj: [https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr\\_ymode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~xOe8Kqc6Mycq&fcrit=word%2Fie%20~0%3E0&flimit=1&freq\\_sort=&page=1&fit\\_include\\_empty=0&freqlevel=1&freq\\_type=tokens](https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr_ymode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~xOe8Kqc6Mycq&fcrit=word%2Fie%20~0%3E0&flimit=1&freq_sort=&page=1&fit_include_empty=0&freqlevel=1&freq_type=tokens)

Z tabulkového zobrazení frekvenční distribuce slovních tvarů se ukazuje, že 18 z 37 tvarů lemmatu *muset* je vyslovováno nářečně, tedy se *š*. Z celkového počtu 209 výskytů je to

dohromady 112, což je 53,6 %. Z těchto výsledků vidíme, že nářeční varianta lehce převažuje nad spisovnou. Mezi nejčastěji užívané slovní tvary patří 39krát tvar *mušeli*, 24krát *mušel*, 13krát tvary *mušela* a *mušelo*, 7krát tvar *muší*, po 2 výskytech tvary *mušíte*, *mušíme*, *muším*, nakonec po 1 tvary *nemušelo*, *mušíš*, *nemuším*, *nemušíž*, *mušíte*, *muším*, *nemuší*, *mušil*, *mušet* a *nemušeli*.

Pokud chceme zjistit, v jaké oblasti se nářeční výslovnost objevuje nejčastěji, nemůžeme použít frekvenční distribuci v tomto stejném dotazu, neboť by se ukázaly výsledky nářeční i spisovné výslovnosti dohromady. Nemusíme ale ani manuálně procházet výsledky, zadáme namísto toho nový dotaz. V něm budeme vyhledávat všechny slovní tvary o libovolné délce, v nichž se bez ohledu na pozici ve slově vyskytuje *-muš-*. Nesmíme zapomenout pro tento nový dotaz povolit regulární výrazy a jako výchozí atribut zvolit *word*. Podoba dotazu viz níže na obrázku 23.

Hledat v korpusu

dialekt\_v2\_dial / jzč

Pokročilý dotaz | Klávesnice | Předchozí dotazy | Interpretace dotazu

**\*muš.\***

**TIP** Barevně zvýrazněný token s otázkem v jednoduchém dotazu znamená možnost upřesnit jeho interpretaci. Použijte CTRL/Command a klik pro aktivaci. (další tip)

**-** Upřesnit parametry

Shoda velikosti písmen Povolit regulární výrazy Výchozí atribut: word

Obrázek 23 – Podoba dotazu k vyhledání slovních tvarů s *-muš-*, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial)

Pro takto položený dotaz vyhledá korpus 118 slovních tvarů, což ale neodpovídá prvnímu výsledku vyhledávání, kdy frekvenční distribuce ukázala, že tvarů s nářeční výslovností je v korpusu 112. Je tedy nutné výsledky zkontrolovat – opět použijeme frekvenční distribuci lemmat pomocí tabulkového zobrazení (tabulka viz obrázek 24 níže).

Frekvenční seznam

Zobrazení v grafech    Tabulkové zobrazení

Zobrazit konfidenční intervaly  ⓘ

1 / 1 (celkem: 3 položky)

Filtr	lemma	Freq	i.p.m.
1	p / n muset musit	112	1 487,34
2	p / n muž	5	66,4
3	p / n mužský	1	13,28

Obrázek 24 – Frekvenční distribuce lemmat, zdroj: [https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr\\_ymode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~RoU2iYkuC6Io&fcrit=word%2Fie%200~0%3E0&flimit=1&freq\\_s](https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr_ymode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~RoU2iYkuC6Io&fcrit=word%2Fie%200~0%3E0&flimit=1&freq_s)

Z tabulky vidíme, že jsou do vyhledávání zahrnuty slovní tvary lemmat *muž* a *mužský*. Použijeme k tomu negativní filtr – v tabulce frekvenční distribuce vyhledáme slovní tvar, zmáčkneme u něj písmeno *n*, tím se zobrazí pouze výskyty bez tohoto lemmatu. Tato akce nás vrátí zpět na konkordanční řádky výskytů, musíme tedy znovu zobrazit frekvenční distribuci a stejný postup opakovat u druhého lemmatu. Tímto způsobem se dostaneme ke konkordančním řádkům pouze se slovními tvary slovesa *muset* s nářeční výslovností.

Nyní chceme zjistit, zda je nářeční výslovnost tvarů slovesa *muset* užívána častěji v jihočeské, nebo západočeské nářeční oblasti. Dále nás zajímá rozložení výskytů ve vrstvách nahrávek. Obě tyto informace můžeme jednoduše zjistit v jednom kroku, a to opět pomocí frekvenční distribuce. V horní liště zvolíme *Frekvence*, dále *Vlastní...*, otevře se tabulka, v níž vybereme možnost *Podle typů textů*. Zvolíme strukturní atributy *doc.noblást* a *doc.rok* a vytvoříme frekvenční seznam. Pro přehlednost opět použijeme tabulkové zobrazení. Otevřou se nám dvě tabulky, v nichž se objevuje distribuce slovních tvarů v nářečních oblastech a v letech pořízení nahrávek, viz obrázek 25 níže.

1 / 1 (celkem: 2 položky)			
Filtr	doc.noblast	Freq	i.p.m.
1	p / n jihočeská	72	2 334,18
2	p / n západočeská	40	899,77

1 / 1 (celkem: 11 položek)			
Filtr	doc.rok	Freq	i.p.m.
1	p / n 2009	25	1 297,49
2	p / n 2017	23	3 130,01
3	p / n 2011	21	2 216,12
4	p / n 2007	18	1 359,93
5	p / n 2008	9	1 012,06
6	p / n 1964	4	2 846,98
7	p / n 1963	4	2 768,17
8	p / n 195X	3	6 302,52
9	p / n 1966	3	6 134,97
10	p / n 1967	1	1 114,83
11	p / n 2005	1	192,06

Obrázek 25 – Frekvenční distribuce slovních tvarů podle nářečních oblastí a roků pořízení nahrávek, zdroj: [https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr\\_vmode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~x8kaAg6K242I&](https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~x8kaAg6K242I&)

Z první frekvenční tabulky zjistíme, že z celkového počtu 112 slovních tvarů jich je 72 zaznamenáno v jihočeské nářeční oblasti, to je 64,3 %. Zbytek výskytů (40) je z oblasti západočeské. To ukazuje na fakt, že nářeční výslovnost tvarů slovesa *muset* by mohla být typičtější pro jihočeskou oblast. Ze druhé tabulky můžeme vidět, že nejvíce slovních tvarů pochází z nové vrstvy nahrávek – 25 výskytů z roku 2009, 23 z roku 2017, 21 z roku 2011, 18 z roku 2007, 9 z roku 2008, 1 z roku 2005. Dostatek výskytů v nové časové vrstvě nahrávek tedy ukazuje na to, že se jedná o stále živý a dosud užívaný nářeční rys.

Po shrnutí všech výše zmíněných výsledků můžeme nářeční výslovnost tvarů slovesa *muset* se *š* charakterizovat jako typický nářeční jev zejména pro jihočeskou nářeční oblast, přesto zachovaný i v západočeské nářeční oblasti. Je to jev typický pro nejstarší generaci mluvčích, avšak ze staré vrstvy nahrávek není v korpusu pro tento rys dostatek materiálu, nicméně důležitým výsledkem je to, že v současnosti je to rys stále živý.

Pro srovnání lze sloveso *muset* vyhledat v ČJA. Vyhledat ho můžeme snadno v elektronické verzi atlasu<sup>133</sup>. Otevře se karta s danou položkou, z níž se můžeme dočíst,

<sup>133</sup> Dostupné z: <https://cja.ujc.cas.cz/e-cja/>.

že položka je pod číslem 358 zařazena do ČJA 4 (4. díl atlasu – morfologická část). Z připojené mapy je patrné, že se sloveso *mušet* vyskytuje dubletně s *mušet* v jihočeské nářeční oblasti, v západočeské je navíc doložen tvar *mušit*. Ve verzi atlasu k tisku<sup>134</sup> nalezneme položku na s. 506, mapa viz s. 505.

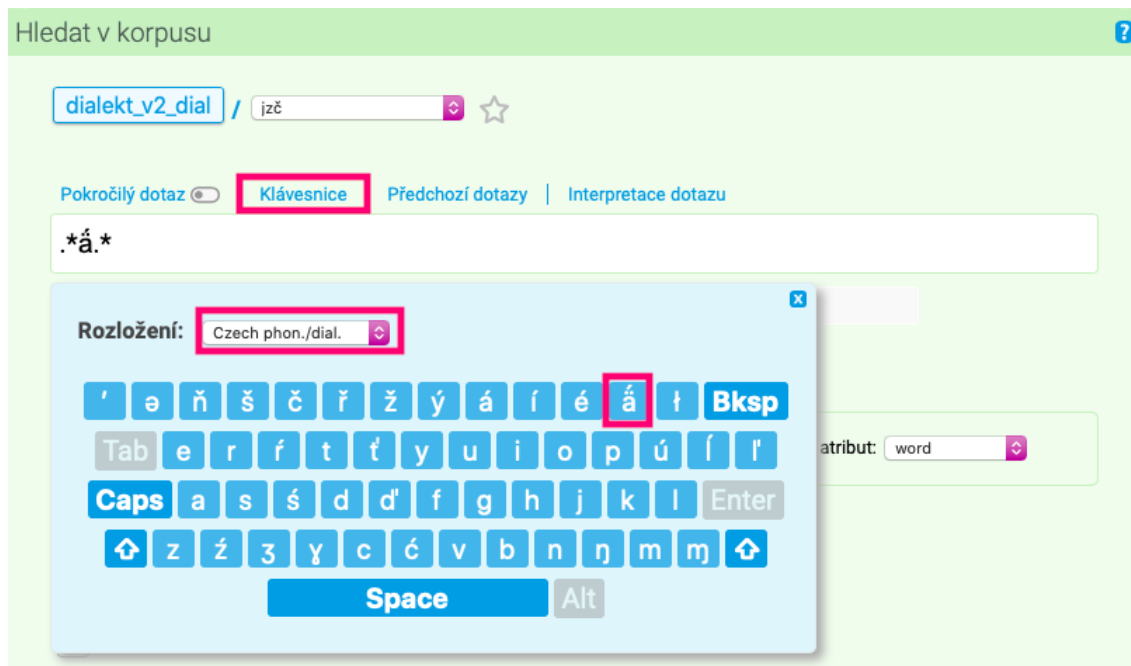
---

<sup>134</sup> Dostupné z: <https://cja.ujc.cas.cz/>.

## Cvičení č. 2 – Široká výslovnost *á* v západočeské a jihočeské nářeční oblasti

Zadání: Pro jihozápadočeskou nářeční skupinu je z hlediska hláskosloví typická široká výslovnost dlouhého *á*. Vyhledejte tento jev v korpusu DIALEKT. Jaké je zastoupení jevu z hlediska nářečních skupin?

Postup: Širokou výslovnost dlouhého *á*<sup>135</sup> budeme v korpusu vyhledávat na rovině dialektologické, jako korpus zvolíme tedy *dialekt – v2 dial* (na rovině ortografické je výslovnost zaznamenána ve spisovném jazyce). Omezíme hledání na západočeskou a jihočeskou nářeční oblast (viz obrázek 17), nebo pro vyhledávání zvolíme subkorpus *jzč* (viz obrázek 19). Při vyhledávání budeme používat regulární výrazy. Výchozím atributem pro vyhledávání bude *word*. Široká výslovnost dlouhého *á* má vlastní speciální znak, značí se pomocí písmene *a* s přehláskou a čárkou zároveň. V korpusu vyhledáme daný znak pomocí české klávesnice a vložíme. Nakonec přidáme regulární výrazy (tečku a hvězdičku), aby korpus vyhledal všechny slovní tvary o libovolné délce, v nichž se vyskytuje široká výslovnost *á* neohledně na pozici ve slově. Podoba dotazu je vidět níže na obrázku 26, stejně jako způsob, jak vyhledat speciální znak na české klávesnici.



Obrázek 26 – Podoba dotazu a zadání speciálního znaku pro širokou výslovnost *á*, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial&usesubcorp=jzč](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial&usesubcorp=jzč)

<sup>135</sup> Širokou výslovnost dlouhého *á* si lze poslechnout v *Archivu diferenčních hlásek nářečí českého jazyka*, viz <https://wiki.korpus.cz/doku.php:cnk:dialekt:archivhlasek>.

Pro takto položený dotaz vyhledá korpus celkem 404 výskytů slovních tvarů, což je vysoké číslo. Pokud se podíváme na frekvenční distribuci slovních tvarů pomocí tabulkového zobrazení, zjistíme, že nejčastějším je slovní tvar *tám* (83 výskytů), následuje tvar *dá* (9 výskytů), *pekáč* (7 výskytů), po 6 výskytech tvar *válki* a *nás*, po 4 výskytech tvary *nám*, *řiká*, *máte* či *sedlák*.

Protože je v tomto případě počet výskytů vysoký, použijeme frekvenční distribuci i pro zjištění zastoupení nářečního jevu v jednotlivých oblastech a letech. V horní liště opět vybereme záložku *Frekvence*, následně zvolíme *Vlastní...*, poté vybereme frekvenční distribuci podle typů textů. Následně musíme vybrat ty strukturní atributy, které chceme do frekvenční distribuce zahrnout, tedy *doc.noblast* (nářeční oblast) a *doc.rok* (rok pořízení nahrávky) (viz obrázek 27 níže), a vytvoříme frekvenční seznam.

Frekvenční distribuce

Běžná **Podle typů textů** Závislost dvou atributů

Frekvenční limit: 1

Zahrnout i kategorie bez výskytů:

Strukturní atributy:

doc.explorator	<input type="checkbox"/>	sp.explorator	<input type="checkbox"/>	para.type	<input type="checkbox"/>
doc.id	<input type="checkbox"/>	sp.id	<input type="checkbox"/>		
doc.misto	<input type="checkbox"/>	sp.mistodet	<input type="checkbox"/>		
<b>doc.noblast</b>	<input checked="" type="checkbox"/>	sp.mistonej	<input type="checkbox"/>		
doc.npodskupina	<input type="checkbox"/>	sp.noblastdet	<input type="checkbox"/>		
doc.ntyp	<input type="checkbox"/>	sp.noblastnej	<input type="checkbox"/>		
doc.nusek	<input type="checkbox"/>	sp.npodskupinanej	<input type="checkbox"/>		
doc.okres	<input type="checkbox"/>	sp.ntypnej	<input type="checkbox"/>		
doc.pocetml	<input type="checkbox"/>	sp.nuseknej	<input type="checkbox"/>		
doc.promluvatyp	<input type="checkbox"/>	sp.okresnej	<input type="checkbox"/>		
doc.region	<input type="checkbox"/>	sp.pohlavi	<input type="checkbox"/>		
<b>doc.rok</b>	<input checked="" type="checkbox"/>	sp.prezdivka	<input type="checkbox"/>		
doc.sidlotyp	<input type="checkbox"/>	sp.profese	<input type="checkbox"/>		
doc.sidlovelikost	<input type="checkbox"/>	sp.regionnej	<input type="checkbox"/>		
doc.stat	<input type="checkbox"/>	sp.rok	<input type="checkbox"/>		
doc.tema	<input type="checkbox"/>	sp.soundfile	<input type="checkbox"/>		
doc.vrstva	<input type="checkbox"/>	sp.statnej	<input type="checkbox"/>		
doc.wordcount	<input type="checkbox"/>	sp.vek	<input type="checkbox"/>		
doc.zdroj	<input type="checkbox"/>	sp.veknum	<input type="checkbox"/>		
		sp.vzdelanitypl	<input type="checkbox"/>		
		sp.vzdelanitypll	<input type="checkbox"/>		

Vytvořit frekvenční seznam

Obrázek 27 – Frekvenční distribuce podle typů textů, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr\\_ymode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~hS06iKQgEua4](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr_ymode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~hS06iKQgEua4)

Po vytvoření frekvenčního seznamu zahrnujícího nářeční oblast a rok se zobrazí dva grafy, pro lepší přehlednost a konkrétní číselné údaje zvolíme tabulkové zobrazení (viz obrázek 28 níže). Z první tabulky se dočteme, že nejvíce výskytů široké výslovnosti *á*,



konkrétně 389 ze 404, pochází ze západočeské oblasti. To je 96,3 %, zbytek výskytů zastupuje oblast jihočeská. Tento nářeční jev je tedy charakteristický spíše pro západočeskou oblast jihozápadočeské nářeční skupiny, nicméně i doklady o jevu v jihočeské oblasti jsou důležité. Co se roku pořízení nahrávky týče, nejvíce dokladů pochází z nové vrstvy, celkem v letech 2005, 2007, 2008, 2009 a 2011 vyhledal korpus 344 výskytů, což je 85,1 %. Je to opět proto, že ze staré vrstvy je pro tento rys málo materiálu.

Frekvenční seznam

Zobrazení v grafech    Tabulkové zobrazení

Zobrazit konfidenční intervaly  ⓘ

1 / 1 (celkem: 2 položky)

Filtr	doc.noblast	Freq	i.p.m.
1	p / n    západočeská	389	8 750,22
2	p / n    jihočeská	15	486,29

1 / 1 (celkem: 8 položek)

Filtr	doc.rok	Freq	i.p.m.
1	p / n    2009	120	6 227,94
2	p / n    2007	117	8 839,53
3	p / n    2011	59	6 226,26
4	p / n    2008	34	3 825,38
5	p / n    1972	30	59 171,06
6	p / n    1964	15	10 676,16
7	p / n    1963	15	10 380,62
8	p / n    2005	14	2 696,46

Obrázek 28 – Tabulkové zobrazení frekvenční distribuce podle typů textů, zdroj: [https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr\\_ymode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~hS06iKQgEua4&fcrit=doc.noblast%200&flimi](https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr_ymode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~hS06iKQgEua4&fcrit=doc.noblast%200&flimi)

Širokou výslovnost *á* lze tedy charakterizovat jako nářeční jev v dnešní době stále užívaný a živý u nejstarší generace mluvčích, a to hlavně v západočeské nářeční oblasti.

### Cvičení č. 3 – Vkladné -j- před zubnicemi *t'*, *d'* v západočeské nářeční oblasti

Zadání: *Specifickým znakem západočeské oblasti je vkladné neboli epentetické -j- před měkkými zubnicemi, kterými jsou t' a d', vznikají tedy spojení jt' a jd'. Vyhledejte v korpusu DIALEKT tuto vkladnou hlásku a uveďte, zda se s ohledem na počet výskytů jedná o stále živý nářeční rys. Zjistěte, v jakých tvarech se nejčastěji vyskytuje, zaměřte se přitom na slovní druhy.*

Postup: Vyhledávání vkladného -j- před měkkými zubnicemi bude probíhat na rovině dialektologické, korpus tedy zvolíme *dialekt – v2 dial*, neboť na ortografické rovině se tento rys nepromítá. Omezíme vyhledávání na západočeskou nářeční oblast bez bližšího určení úseku (viz obrázek 17), nebo vybereme subkorpus *zč* (viz obrázek 19). Pro vyhledání vkladného -j- pouze před zubnicemi *t'* a *d'* se nabízí dva velmi jednoduché postupy obsahující regulární výrazy, které je proto nutné při zadávání dotazu povolit.

(1) V prvním postupu použijeme dva různé dotazy. Budeme vyhledávat všechna slova o libovolné délce, v nichž se vyskytuje *jt'* nebo *jd'*. Regulární výrazy, které použijeme, budou tečka (zastupující jeden libovolný znak) a hvězdička (libovolné opakování předchozího znaku), vyhledávat budeme atribut *word* (slovní tvar). Podoba obou dotazů je vidět na obrázku 29 níže. Takto pomocí regulárních výrazů zařídíme, že hledaná slova budou mít libovolnou délku a skupina daných hlásek se může objevit v jakékoliv části slova.

Hledat v korpusu

dialekt\_v2\_dial ★

Pokročilý dotaz | Klávesnice | Předchozí dotazy | Interpretace dotazu

.\*jt'.\*

Hledat v korpusu

dialekt\_v2\_dial ★

Pokročilý dotaz | Klávesnice | Předchozí dotazy | Interpretace dotazu

.\*jd'.\*

Obrázek 29 – Podoba dotazů k vyhledání vkladného -j- před měkkými zubnicemi, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial)

Po zadání dotazů zjistíme, že pro vkladné *-j-* před *t'* vyhledal korpus v západočeské skupině 16 výskytů (viz obrázek 30 níže), ale pouze v 10 z nich se jedná o epentetické *-j-*, v ostatních případech jde o slova, v nichž se *-j-* ve spisovné podobě nachází (např. *přijít, pojt'*)<sup>136</sup>. Z 10 relevantních výskytů se 4 objevily ve slovesném tvaru, konkrétně v rozkazovacím způsobu *nechojt'*, dále po 2 výskytech v příslovci *tejt'* a práci částici *ajt'*, po 1 výskytu se objevilo substantivum *chujt'* a spojka *bujt'*. Všechny z 10 výskytů vkladného *-j-* pocházejí z nové vrstvy nahrávek, což je důležitý výsledek ukazující na fakt, že se tento nářeční jev vyskytuje stále i v současné době.

The screenshot shows a search interface with the following details:

- Search results: 16 (Výskytů: 16)
- Language: i.p.m. (Spočítat)
- ARF: 4,28
- Result type: Výsledek je setříděn
- Page: 1 / 1
- Filter: Výběr řádků: základní

The table below summarizes the visible search results:

ID	Text	Word	Meaning
0021-C-ZAC-1963-SA	jeden povídá :   +   "copa ? ta má řáskou dřívou	chujt'/chuf	! "   +   no, ale koštovali a řáák jim nelezla
0023-C-ZAC-1970-SA	.   +   ti miadí tu nejvedej řic, co je .   +	ajťaf	si toho každěj váží !   +   no ,tak diš sem
0258-C-ZAC-2007-N	mámo .   +   jedna žencká řekla , jedna bába , a	tejt'/tef	ji chlápec teki štúdoval a řeco tenkrát neveďel řeco ,
0300-C-ZAC-2014-N	(nadechnutí)   +   (klepání)   +   hele ho , hele ho , pocem ,	pojť/pojď	se podívat na toho krášíl . "   +   (rušivý zvuk)   +   to sou
0301-C-ZAC-2014-N	sem a kamaráďi mi řikaj :   +   (nadechnutí)   +   (nadechnutí)   +	pojť/pojď	si sednout , je svačina . "   +   (nadechnutí)   +   já pošm
0302-C-ZAC-2014-N	, von bi mistr nadával , že se zdržuju ,	přijť/přijď	, aš tadi nebudé . "   +   (nadechnutí)   +   (nadechnutí)   +   já řel
0302-C-ZAC-2014-N	(nadechnutí)   +   tag von řel prič a já :   +   (nadechnutí)   +	přijť/přijďte	dvá , chlápci , a zase dvá . "   +   (nadechnutí)   +
0302-C-ZAC-2014-N	dobrá , že jo , no . "   +   (nadechnutí)   +	nechojť/mechoď	tam , dostaneš přes prdel . "   +   (hlasitý hovor v pozadí)   +   (nadechnutí)   +
0302-C-ZAC-2014-N	. "   +   (hlasitý hovor v pozadí)   +   (nadechnutí)   +   a neřek řigdo :   +	nechojť/mechoď	tam , dostaneš přes prdel . "   +   že ne ?   +
0302-C-ZAC-2014-N	*   +   že ne ?   +   neřek , neřikal řigdo :   +	nechojť/mechoď	tam , dostaneš přes prdel , ne , jo ,
0302-C-ZAC-2014-N	bjěš ! "   +   neřek jeden druhému , neřek :   +	nechojť/mechoď	tam , tám , tu řelvu tam má , ale
0615-C-ZAC-2009-N	ďeal u správě silňidz hildáče .   +   *k přišel domú a	bujť/buď	si dřim chvíli ...   +   hmm .   +   hmm .   +   a pag
0627-C-ZAC-2009-N	řech out tám stálo seřazeních , né .   +   nó a	tejt'/tef	voři ...   +   tef přijel ten dřip a voři vobá ti
0628-C-ZAC-2009-N	:   +   "Jozěf , pocem . "   +   ne .   +	pojť/pojď	se podívat , má @ , má @ se nam
0628-C-ZAC-2009-N	hmm , hmm .   +   má se nam teit kráva ,	pojť/pojď	se podívat , co tomu řekneš .   +   hmm .   +   no
0628-C-ZAC-2009-N	sudi s přeřicí . "   +   hmm .   +   hmm .   +	ajťaf	na to nepřidou ... "   +   hmm .   +   né .   +   no

Obrázek 30 – Výsledky vyhledávání vkladného *-j-* před *t'*, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr\\_ymode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~WY4kuK8e22Uy](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr_ymode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~WY4kuK8e22Uy)

Výsledek lze názorně zobrazit přes frekvenční distribuci – v horní liště zvolíme záložku *Frekvence*, pak *Slovní tvary* a následně *Tabulkové zobrazení* (viz obrázek 31 níže).

<sup>136</sup> Doporučujeme pro přehlednost nevyhovující slova z výskytů odstranit pomocí negativního filtru.

Frekvenční seznam

Zobrazení v grafech    Tabulkové zobrazení

Zobrazit konfidenční intervaly

1 / 1 (celkem: 8 položek)

Filtr	word [dial]	Freq	i.p.m.
1 p / n	nechojť	4	12,89
2 p / n	pojť	4	12,89
3 p / n	tejť	2	6,45
4 p / n	ajť	2	6,45
5 p / n	bujť	1	3,22
6 p / n	přijťte	1	3,22
7 p / n	přijť	1	3,22
8 p / n	chujť	1	3,22

Obrázek 31 – Frekvenční distribuce výsledků, zdroj: [https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attr\\_ort=ort&attr\\_vmode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~kMqsySKKq6Ms&fcrit=word%2Fie%200~0%3E0&flimit=1&fq\\_sort=&fpage=1&f](https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attr_ort=ort&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~kMqsySKKq6Ms&fcrit=word%2Fie%200~0%3E0&flimit=1&fq_sort=&fpage=1&f)

Pokud zadáme do vyhledávání druhý dotaz, zjistíme, že vkladné *-j-* před *d'* není v západočeské skupině vůbec zaznamenáno, což nicméně nemusí ukazovat na absenci tohoto jevu. Už v prvním případě byl počet výskytů nízký a v tomto případě může být nulový i z důvodu nedostatku materiálu.

(2) Druhým způsobem, jak vkladné *-j-* před měkkými zubnicemi v korpusu vyhledat, je opět použití regulárních výrazů, ale tentokrát jednoho nového, díky němuž nám bude stačit jen jeden dotaz. Tím výrazem jsou hranaté závorky (*/*). Pomocí nich lze uvést alternativy, které má korpus prozkoumat, takže pokud do závorek napíšeme *t'* i *d'*, do vyhledávání budou zahrnuty obě možnosti. Podoba dotazu je vidět na obrázku 32 níže.

Hledat v korpusu ?

**dialekt\_v2\_dial** ★

Pokročilý dotaz | [Klávesnice](#) | [Předchozí dotazy](#) | [Interpretace dotazu](#)

Obrázek 32 – Podoba dotazu k vyhledání vkladného *-j-* před zubnicemi v rámci jednoho dotazu, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial)

Takto zadaný dotaz vyhledá v západočeské skupině opět 16 výskytů, čímž lze i snadno ověřit správnost prvního postupu, a sice absenci vkladného *-j-* před *d'*, neboť výsledky

jsou stejné jako při prvním hledání pomocí dvou dotazů. Jak již bylo zmíněno výše, oproti jiným hledaným jevům je počet výskytů nízký nebo žádný, naopak výskytů bez vkladného *-j-* bylo nalezeno poměrně dost. Důležitým výsledkem pro naše hledání je, že byl jev v nové vrstvě vůbec doložen. Doklady z korpusu ukazují, že jev je v menší míře doložen i v jihočeské nářeční oblasti, ale typický je spíše pro oblast západočeskou<sup>137</sup>. Na tomto místě se nabízí opět tento jev zkoumat při uveřejnění další verze korpusu, ve které bude dostupné větší množství materiálu.

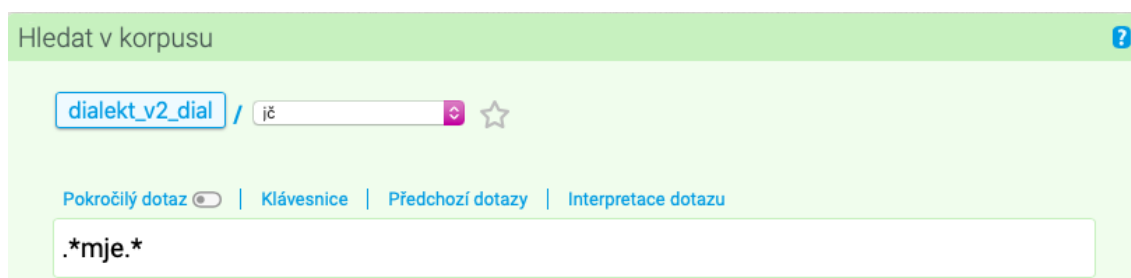
---

<sup>137</sup> Viz ČJA 4, položka 303, s. 424, mapa na s. 425. Disimilovaná výslovnost retnic je podle ČJA typická pro celou jihozápadočeskou polovinu Čech s přesahem do jihozápadní části středočeských nářečí.

#### Cvičení č. 4 – Výslovnost skupiny *mňe* jako *mje* v jihočeské nářeční oblasti

Zadání: *Typickým hláskoslovným rysem pro jihočeskou nářeční oblast je výslovnost skupiny mňe jako mje. Vyhledejte tento jev v korpusu DIALEKT a zjistěte, ve kterých slovech se vyskytuje nejčastěji. Zhodnoťte, zda se tento jev používá v současnosti a jaká je četnost jeho výskytu jevu oproti spisovné výslovnosti.*

Postup: Výslovnost skupiny *mňe* jako *mje* budeme vyhledávat na rovině dialektologické, zvolíme tedy korpus *dialekt* – v2 *dial* (na rovině ortografické je výslovnost převáděna do spisovné podoby). Hledání omezíme na jihočeskou nářeční oblast (viz obrázek 17), nebo pro vyhledávání zvolíme subkorpus *jč* (viz obrázek 19). Povolíme regulární výrazy, jako výchozí atribut pro vyhledávání zvolíme *word*. K vyhledávání budeme potřebovat z regulárních výrazů tečku a hvězdičku, pomocí kterých zajistíme, že korpus vyhledá všechny slovní tvary o libovolné délce, v nichž se nehledě na pozici ve slově bude vyskytovat skupina *mje*. Podoba dotazu viz níže na obrázku 33.



Hledat v korpusu

dialekt\_v2\_dial / jč

Pokročilý dotaz | Klávesnice | Předchozí dotazy | Interpretace dotazu

.\*mje.\*

Obrázek 33 – Podoba dotaz k vyhledání skupiny -mje-, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial&usesubcorp=jč](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial&usesubcorp=jč)

Pro takto položený dotaz vyhledá korpus v jihočeské nářeční oblasti 132 výskytů slovních tvarů, v nichž se objevuje skupina *-mje-* (doklady v korpusu nicméně ukazují i na ojedinělý výskyt jevu v západočeské oblasti<sup>138</sup>). Abychom zjistili, ve kterých slovních tvarech se tato skupina vyslovuje nejčastěji, zvolíme frekvenční distribuci slovních tvarů pomocí tabulkového zobrazení. Z frekvenční tabulky se dozvíme, že nejčastěji je skupina vyslovována ve slovním tvaru *mjeli* (41 výskytů) a *mjel* (21 výskytů). Následuje tvar *mje* (18 výskytů), *mjela* (12 výskytů), *nemjeli* (6 výskytů), *nemjel* (4 výskytů), *zimje* (3 výskytů), po 2 výskytů ve tvarech *lomje*, *nemjela*, *mjesta* a *mjelo*. Další slovní tvary

<sup>138</sup> Viz ČJA 1, PRO mapa B, s. 54–55, B3. Zde se ukazuje možnost výskytu jevu právě i v západočeské nářeční oblasti. Část PRO (Pravidelné regionální obměny) je zařazena do každého dílu ČJA.

už byly zastoupeny pouze jednou, například *mámje*, *umjela*, *pamjeťnice*, *zemjed'elství* či *slamjenki* (viz obrázek 34).

	Filtr	word [dial]	Freq	i.p.m.
1	p / n	mjeli	46	1 491,28
2	p / n	mjel	21	680,08
3	p / n	mje	18	583,54
4	p / n	mjela	12	389,03
5	p / n	nemjeli	6	194,51
6	p / n	nemjel	4	129,68
7	p / n	zimje	3	97,26
8	p / n	lomje	2	64,84
9	p / n	nemjela	2	64,84
10	p / n	mjesta	2	64,84
11	p / n	mjelo	2	64,84
12	p / n	mámje	1	32,42
13	p / n	krámje	1	32,42
14	p / n	umjela	1	32,42
15	p / n	pamjeťnice	1	32,42
16	p / n	mjestu	1	32,42
17	p / n	umjet	1	32,42
18	p / n	zemjed'elství	1	32,42
19	p / n	soukromje	1	32,42
20	p / n	slamjenkáy	1	32,42
21	p / n	zímje	1	32,42
22	p / n	slamjenki	1	32,42
23	p / n	slamjence	1	32,42
24	p / n	umjel	1	32,42
25	p / n	neumjel	1	32,42

Obrázek 34 – Frekvenční distribuce slovních tvarů se skupinou -mje- pomocí tabulkového zobrazení, zdroj: [https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr\\_vmode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~tmQY8sImYkAA&fcr](https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~tmQY8sImYkAA&fcr)

Pro zjištění rozložení výskytů hledaného jevu ve vrstvách nahrávek můžeme opět zvolit frekvenční distribuci, a to vlastní podle typů textů. Jako strukturní atribut zvolíme *doc.rok*, následně se zobrazí tabulky výskytů podle roků pořizení nahrávek, v nichž se nářeční jev vyskytl (viz obrázek 35).

Frekvenční seznam

Zobrazení v grafech    Tabulkové zobrazení

Zobrazit konfidenční intervaly  ⓘ

1 / 1 (celkem: 9 položek)

	Filtr	doc.rok	Freq	i.p.m.
1	p / n	2009	55	5 291,51
2	p / n	2011	21	3 943,66
3	p / n	2008	20	8 944,54
4	p / n	2014	17	9 294,07
5	p / n	1964	9	8 272,06
6	p / n	2017	4	544,37
7	p / n	1967	3	3 344,48
8	p / n	1963	2	3 016,59
9	p / n	1966	1	3 875,97

Obrázek 35 – Frekvenční distribuce výskytů podle roku pořízení nahrávky, zdroj: [https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr\\_ymode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~tmQY8sImYkAA&fcrit=doc.rok%200&flimit=1&](https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr_ymode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~tmQY8sImYkAA&fcrit=doc.rok%200&flimit=1&)

Z tabulky výše je vidět, že nejvíce nahrávek je zastoupeno v nové vrstvě (117 ze 132), to z toho důvodu, že v je v ní obsaženo více materiálu. Vzhledem k počtu výskytů v této vrstvě nářečního materiálu se jev ukazuje jako stále živý.

Abychom zjistili, jaký je podíl nářeční výslovnosti oproti spisovné, musíme změnit dotaz a vyhledat *\*mňe.\** namísto *\*mje.\**, přitom zachováme výchozí atribut i subkorpus nebo omezení na jihočeskou oblast. Pro takto položený dotaz vyhledá 173 výskytů slovních tvarů se skupinou *-mňe-*, viz obrázek 36.

Výskytů: 173 | i.p.m.: 5 608,51 (vztaheno k subkorpusu "jc") | ARF: - | Výsledek je setříděn

Výběr řádků: základní

<input type="checkbox"/>	0024-C-JIC-1967-SA	sem se vdála , no , tak maminka počít na	mňe/mňe	chvátala , a sice tafínek , abí mňe maminka naučila
<input type="checkbox"/>	0024-C-JIC-1967-SA	počít na mňe chvátala , a sice tafínek , abí	mňe/mňe	maminka naučila péct chleba .   +   a já sem se do
<input type="checkbox"/>	0024-C-JIC-1967-SA	se do toho moc nehmla .   +   potom ale sem ho	mňela/mňela	pect , fšecho sem si uřelala , a diš sem
<input type="checkbox"/>	0024-C-JIC-1967-SA	, fšecho sem si uřelala , a diš sem ho	mňela/mňela	dávad do pece , tag vona mi dičí přřřla ho
<input type="checkbox"/>	0024-C-JIC-1967-SA	sem neřla .   +   tafínek jí sem nedovolil .   +   a mi	mňeli/mňeli	pec na šes chlebůch .   +   no a šest sem jich
<input type="checkbox"/>	0024-C-JIC-1967-SA	ho tam nanes , tak sem ho tam řsadila ,	mňela/mňela	sem strach , abí mi nezřil řákej řen řirovej .
<input type="checkbox"/>	0024-C-JIC-1967-SA	že sem celá uřtrařená s chlebem , no , tak	mňe/mňe	jako pořeřil , no , tak sem bila řáda .
<input type="checkbox"/>	0025-C-JIC-1964-SA	, tak sem bila řáda .   +   no , mi sme	mňeli/mňeli	, řekřeme , takoví malí hospodářřtví , řegdo mňel louku
<input type="checkbox"/>	0025-C-JIC-1964-SA	sme mňeli , řekřeme , takoví malí hospodářřtví , řegdo	mňel/mňel	louku svojí , a já sem louku svojí nemřel .
<input type="checkbox"/>	0025-C-JIC-1964-SA	řegdo mňel louku svojí , a já sem louku svojí	nemňel/nemňel	.   +   tak sem mňel pachtovanou louku vot opce .   +   urřřteřch

Obrázek 36 – Vyhledávání skupiny *-mňe-* v jihočeské nářeční oblasti, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr\\_ymode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~Oce8KKW4k48w](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr_ymode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~Oce8KKW4k48w)

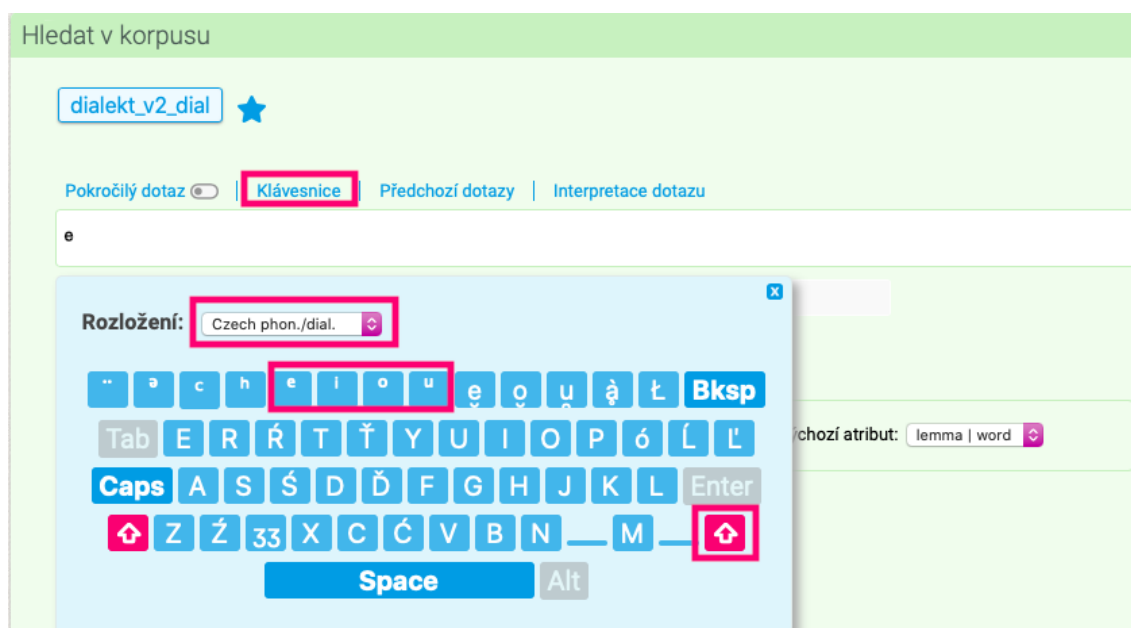
Celkem je tedy skupina *-mňe-/mňe-* zastoupena v korpusu 305krát, z toho 132 výskytů, tedy 43,3 %, zastupuje nářeční výslovnost *-mje-*. To je téměř polovina, což potvrzuje, že se jedná o jev častý a užívaný vedle spisovné výslovnosti velmi frekventovaně.



## Cvičení č. 5 – Zavřená výslovnost dvojhlásek *ej* a *ou* na Doudlebsku a Domažlicku

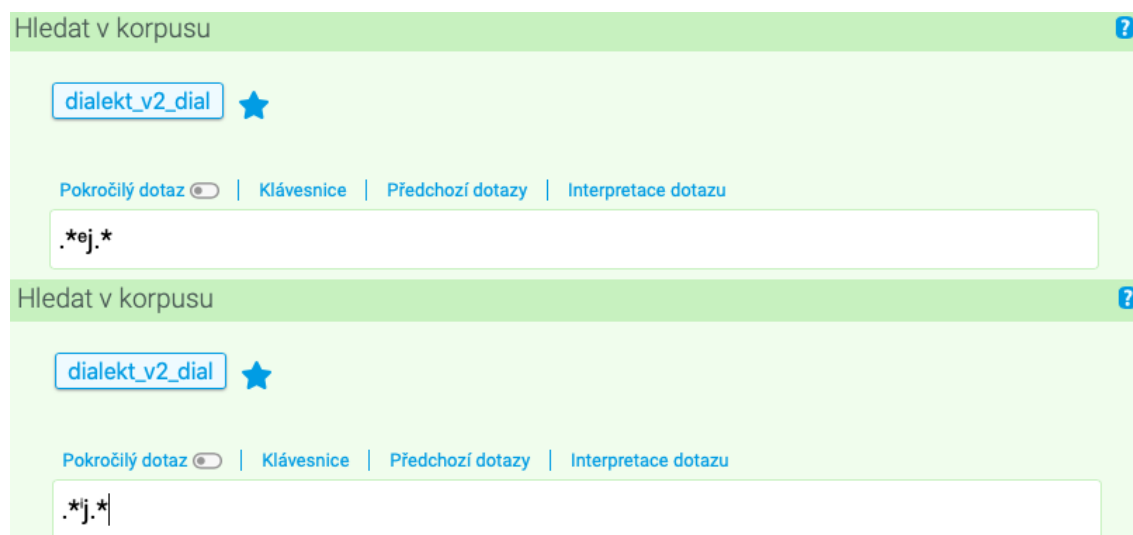
Zadání: Zavřená výslovnost dvojhlásek *ej* a *ou* je nářečním rysem typickým pro doudlebský okrajový úsek jihočeské nářeční oblasti, ale i domažlický úsek oblasti západočeské. Vyhledejte tyto dvojhlásky v korpusu *DIALEKT se zavřenou výslovností a srovnejte počet výskytů s výslovností spisovnou*. Tip pro vyhledávání zavřené výslovnosti – první hláska z dvojhlásky se zaznamenává pomocí horního indexu (tedy např. jako *ej* nebo *ou*).

Postup: Zavřenou výslovnost dvojhlásek *ej* a *ou* je nutné vyhledávat na rovině dialektologické, jako korpus tedy zvolíme *dialekt – v2 dial* (zavřená výslovnost hlásek se na rovině ortografické nepromítá, výslovnost se zde převádí do spisovné podoby). Protože je tento rys typický pro Doudlebsko a Chodsko, opět musíme buď omezit vyhledávání (viz obrázek 17), nebo vybrat subkorpus *Chodsko a Doudlebsko* (viz obrázek 17) a povolit regulární výrazy. Vyhledávat budeme jednotlivé dvojhlásky s jejich zavřenou výslovností, která může mít různou podobu. Z *Archivu diferenčních hlásek nářečí českého jazyka* (viz podkapitola 2.2.1) se dozvíme, že zavřená dvojhláska *ej* může přecházet až do *ij*, naopak *ou* do *uu* (zde je možné si výslovnost i poslechnout). V případě zavřené výslovnosti je první z hlásek zaznamenána pomocí horního indexu, aby byla odlišena od výslovnosti spisovné. Pro zadání horního indexu do vyhledávacího okna je třeba vybrat českou klávesnici a pomocí tlačítka shift zadat horní index, viz obrázek 37.



Obrázek 37 – Zadání horního indexu pomocí české klávesnice v korpusovém manažeru KonText, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial)

Nejprve začneme s vyhledáváním dvojhlásky *ej*, jejíž zavřenou podobu můžeme vyhledat pomocí dvou zápisů – *ej* a *ʲj*. Již před zadáním dotazu jsme povolili regulární výrazy, neboť je třeba sestavit dotaz tak, aby vyhledal všechna slova, ve kterých se tato dvojhláska objevuje, nehledě na jejich délku a umístění dvojhlásky. K tomu použijeme opět tečku zastupující jeden libovolný znak a hvězdičku, jež znamená jeho libovolné opakování. Výchozím atributem bude *word*, podoba dotazu viz obrázek 38 níže.



Obrázek 38 – Vyhledávání dvojhlásky *ej* se zavřenou výslovností, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial)

V případě vyhledávání dvojhlásky *ej* nalezne korpus v domažlickém a doudlebském úseku 18 výskytů, a to hlavně v koncovkách přídavných jmen, zájmen a číslovek (např. *takov<sup>e</sup>j*, *ženat<sup>e</sup>j*, *vožral<sup>e</sup>j*, *druh<sup>e</sup>j*), dále i v podstatných jménech (*ʲjden*, *v<sup>e</sup>jce*), viz obrázek 39. Z těchto výsledků jich 12 pochází ze staré vrstvy nahrávek, 6 z vrstvy nové. Co se jednotlivých úseků týče, 11 výskytů bylo nalezeno na Domažlicku (5 ve nové a 6 v nové vrstvě nahrávek), zbylých 7 v doudlebském okrajovém úseku (všechny ve staré vrstvě nahrávek).

Pokud vyhledáme dvojhlásku *ʲj*, tedy výslovnost ještě zavřenější, nalezne korpus 38 výskytů, opět nejvíce v koncovkách přídavných jmen a zájmen (*pln<sup>ʲ</sup>j*, *star<sup>ʲ</sup>j*, *mal<sup>ʲ</sup>j*, *takov<sup>ʲ</sup>j*, *každ<sup>ʲ</sup>j*), 1 výskyt v předponě přídavného jména třetího stupně (*n<sup>ʲ</sup>jveč<sup>ʲ</sup>i*), 1 výskyt v částici (*st<sup>ʲ</sup>j<sup>n</sup>e*) a 1 ve slovesném tvaru (*b<sup>ʲ</sup>jvávali*), viz obrázek 40. Z celkového počtu výskytů jich pouze 6 pochází ze staré vrstvy nahrávek (z toho 4 výskyty z Domažlicka, 2 z Doudlebska), 32 výskytů je tedy z vrstvy nové, z nichž všechny kromě 2 pochází

z Domažlicka. To je opět dáno větším množstvím nářečního materiálu, který pochází z nové vrstvy.

Pokud změním dotaz a vyhledáme dvojhlásku *ej* bez zavřené výslovnosti, nalezneme korpus 603 výskytů, v nichž je ale zahrnuta i podoba *ej* a *ij*. Z celkového počtu je dvojhláska zaznamenána 56krát se zavřenou výslovností, což je 9,3 %. Zavřená výslovnost se objevuje v západočeské oblasti ve 45 výskytech, tedy zhruba v 80 %, což nasvědčuje, že se jedná nářeční rys dosud zastoupený především na Chodsku, na Doudlebsku už tolik ne. Jedná se tedy o nářeční jev v korpusu zaznamenaný, ale dnes už spíše na ústupu.

Výskyt	Text	Výslovnost
<input type="checkbox"/> 0017-C-ZAC-1963-SA	sebu a vodváli, a b'li tám zase, zase	<b>takovj/takovej</b> žbán nebo hrnec nebo ten, jak se tomu říká
<input type="checkbox"/> 0017-C-ZAC-1963-SA	co s toho dvora bal... vite, von bal	<b>ženatj/ženatej</b> jíde. ten tám teki vorál, a tak ten
<input type="checkbox"/> 0017-C-ZAC-1963-SA	. jenomže potom, di chodili do školi,	<b>fieljakj/všeljakaj</b> učitel je na ře chfil a volti mu je dali
<input type="checkbox"/> 0019-C-ZAC-1964-SA	každej mu řeco dāl, tak mřel žvobřif potom na	<b>ceřj/celej</b> t'jden. tūto tag balo, to je fakt pravda
<input type="checkbox"/> 0019-C-ZAC-1964-SA	mu řeco dāl, tak mřel žvobřif potom na ceřj	<b>t'jden/tejden</b> . tūto tag balo, to je fakt pravda,
<input type="checkbox"/> 0030-C-JIC-1967-SA	to řelalo tak, že toho umalce vřicki museli celho	<b>umřj/umeřj</b> . a na to právjje ta Hóřkřvā bīla. ona
<input type="checkbox"/> 0030-C-JIC-1967-SA	pořihali na toho muže ři Mejšički, a řiř přišel	<b>vořalj/vořalej</b> domū, tak mu to pjeķie... (nadechnuti) to
<input type="checkbox"/> 0031-C-JIC-1964-SA	pec už jako k tomu schopnā, tak se to	<b>vismj/řikj/vysmejřilo</b> , popel se vřhroul. (nadechnuti) no to bila
<input type="checkbox"/> 0031-C-JIC-1964-SA	pak se vzalo pometlo, pometlo ze slāmi, na	<b>t'jřku/tejřku</b> se dalo @ slāma, vře, a ta slāma
<input type="checkbox"/> 0031-C-JIC-1964-SA	a ta slāma se namořila jako drobed do vodi a	<b>vismj/řikj/vysmejřilo</b> se to řhuře ven. vipucovala se hullt ta pekāma
<input type="checkbox"/> 0031-C-JIC-1964-SA	abi se vřhnoul, no. pak se tam ten	<b>vikinutj/vřkynutej</b> chleba dāl. sālzel, no, řikalo se sālzet
<input type="checkbox"/> 0031-C-JIC-1964-SA	lopatu a řup s řim do peci. (ruřivj zvuk) zase	<b>druhj/druhej</b> řhuře a zase do peci. takovj bochřiki bilo,

Obrázek 39 – Vyhledávání zavřené výslovnosti dvojhlásky *ej* jako *ej*, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr\\_ymode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~S4E4SSWias0](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr_ymode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~S4E4SSWias0)

Výskyt	Text	Výslovnost
<input type="checkbox"/> 0017-C-ZAC-1963-SA	ten tām teki vorāl, a tak ten jich našel	<b>takovj/takovij</b> kořik, na ruku jak se nosi, bramburovij,
<input type="checkbox"/> 0017-C-ZAC-1963-SA	na ruku jak se nosi, bramburovij, rukovāk,	<b>plnj/plnij</b> . no a to kařđj misřil: "jej,
<input type="checkbox"/> 0017-C-ZAC-1963-SA	, bramburovij, rukovāk, plnj. no a to	<b>kařđj/kařđij</b> misřil: "jej, ten bude bořāř!"
<input type="checkbox"/> 0018-C-ZAC-1972-SA	potom je mřāřinka, potom je pacholek, to je	<b>nřvjřej/nřvjřej</b> řeledin, tedi vorāl, a potom je pohūnek.
<input type="checkbox"/> 0031-C-JIC-1964-SA	a dalo se na slāmjenki. diř bil	<b>vřvālrij/vřvālrij</b> na slāmjenkřj, nechal se vikinout a pak sme ho
<input type="checkbox"/> 0031-C-JIC-1964-SA	teda, no a ten bochřik, jag bil na	<b>t'rij</b> slāmjenće, tak obrāřil se takle na tu lopatu a
<input type="checkbox"/> 0235-C-ZAC-2007-N	bilo takovi pořasi a vona řizla takle. mi mřivali	<b>takovj/takovij</b> chřif nahore, vespod bili prasřka a nař řim bil
<input type="checkbox"/> 0235-C-ZAC-2007-N	nahore, vespod bili prasřka a nař řim bil eře	<b>takovj/takovij</b> chřif, tam se řizlo po takovim malim řebřičku,
<input type="checkbox"/> 0235-C-ZAC-2007-N	tak vona se tām řla pořivāt, a řiķo přišel	<b>takovj/takovij</b> vřchar velki, a jak vona tām bila nahore,
<input type="checkbox"/> 0236-C-ZAC-2007-N	tak je to řida. tak? to řechno ři	<b>nepovřdj/nepovřdij</b> . to je ūřasni slovo. (smřich) jā si to
<input type="checkbox"/> 0236-C-ZAC-2007-N	. jā si to napřisu. (smřich) to řechno	<b>nepovřdj/nepovřdij</b> . to jā se s řim vadřim a řechno mu
<input type="checkbox"/> 0239-C-ZAC-2007-N	f Postřekovje bila mřlogři. (nadechnuti) von tu bal teki	<b>řākij/nřjākij</b> řalāl, ale ten ūž bal starj, tak ten
<input type="checkbox"/> 0239-C-ZAC-2007-N	tu bal teki řākij řalāl, ale ten ūž bal	<b>starj/starj</b> , tak ten řegři, teki si tu ūřal mři
<input type="checkbox"/> 0239-C-ZAC-2007-N	nemoch. (řuch z ulice) hmm, hmm. to von bil	<b>starj/starj</b> , to bil rodāk postřekovřkj a von řegře břidle*.
<input type="checkbox"/> 0239-C-ZAC-2007-N	hmm. to von bil starj, to bil rodāk	<b>postřekovřkj/postřekovřkj</b> a von řegře břidle*, řegře řelal, ale to

Obrázek 40 – Vyhledávání zavřené výslovnosti dvojhlásky *ej* jako *ij*, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr\\_ymode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~iQg8as8eqkow](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr_ymode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~iQg8as8eqkow)

Stejný postup zopakujeme při vyhledávání zavřené výslovnosti dvojhlásky *ou*, kterou opět můžeme vyhledat dvěma způsoby – jako *ou* a jako *u*. V prvním případě vyhledá korpus 49 výskytů, 22 ve staré vrstvě nahrávek, 27 v nové, což je důležitý výsledek. Zavřená výslovnost se objevila například ve slovech *sousedři*, *sou*, *dou*, *sebou*, *louce*,

*pjekn<sup>o</sup>u*. Ve druhém případě korpus vyhledal 50 výskytů, tedy o jeden více než v prvním případě. Z nich pouze 9 pochází ze staré vrstvy nahrávek, navíc všechny z nich z domažlického úseku. V nové vrstvě se tedy zavřenější výslovnost objevuje 41krát, z toho pouze jeden výskyt pochází z Doudlebska. Zavřenější výslovnost se objevila například ve slovech *s<sup>u</sup>*, *seb<sup>u</sup>*, *jedn<sup>u</sup>*, *cel<sup>u</sup>*, *star<sup>u</sup>*, *vik<sup>u</sup>upili*, *m<sup>u</sup>uku*.

Pokud opět změním dotaz na *.<sup>o</sup>u.\**, abychom vyhledali spisovnou výslovnost, korpus nalezne 673 výskytů, v nich je ale opět zahrnuta varianta *<sup>o</sup>u* i *<sup>u</sup>*. Z celkového počtu výskytů je tedy zavřená výslovnost zastoupena v 99 případech, což je 14,7 %. Pouze 8 výskytů zavřené výslovnosti se objevilo v doudlebském okrajovém úseku, tudíž je tento nářeční jev opět zastoupen převážně na Domažlicku, přesněji v 91 případech, což je 91,9 % z celkového počtu 99 výskytů zavřené výslovnosti. Na Doudlebsku lze nářeční jev charakterizovat jako mizející proto, že je málo doložený.

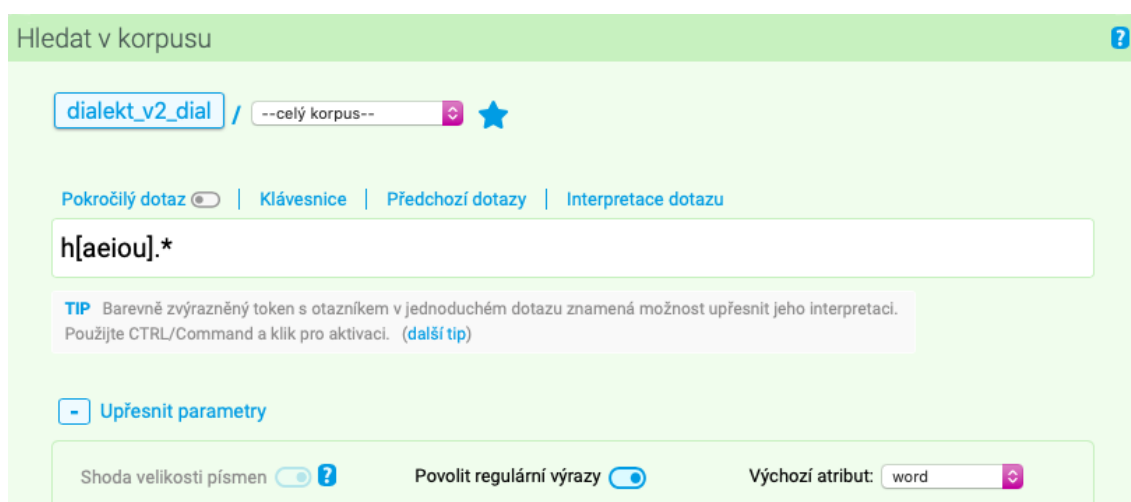
Pokud srovnáme všechny výsledky vyhledávání, vidíme, že i přes nižší podíl s ohledem na výslovnost spisovnou lze zavřenou výslovnost daných dvojhlásek označit jako dosud dochovaný rys u nejstarší generace mluvčích, a to především na Chodsku. Tomu odpovídá i celkový počet výskytů zavřené výslovnosti dvojhlásek v chodském úseku, korpus jich vyhledal 136, oproti tomu na Doudlebsku jich bylo nalezeno pouze 19, což ukazuje, že zde je tento nářeční rys již na ústupu.

## Cvičení č. 6 – Náslovné *h-* na Domažlicku

Zadání: Náslovné *h-* ve slovech začínajících na samohlásku je typické pro chodský úsek západočeské nářeční skupiny. Vyhledejte v korpusu *DIALEKT*, zda se toto *h-* na Chodsku stále objevuje a jestli je doloženo i v nové vrstvě nářečního materiálu.

Postup: Náslovné *h-* budeme vyhledávat v korpusu *dialekt – v2 dial* (na rovině ortografické se slova s náslovným *h-* převádí do spisovné podoby) a je nutné omezit hledání na oblast západočeskou a úsek domažlický (chodský) (viz obrázek 17), nebo vybrat subkorpus *Chodsko* (viz obrázek 19). Při vyhledávání budou užívány regulární výrazy, je tedy třeba je povolit, aby pomocí nich mohl korpus vyhledávat.

Dotaz by na první pohled mohl vypadat stejně jako v předchozím cvičení, tedy *h.\** – vyhledali bychom všechna slova, která začínají na *h-*. Nicméně potřebujeme, aby za ním následovala samohláska. Abychom nemuseli zvlášť vyhledávat dotazy jako *ho.\**, *hu.\**, *ha.\**, přidáme do dotazu další regulární výraz, pomocí něhož zajistíme, že dotaz vyhledá všechna slova na *h-*, za nímž následuje samohláska. Tímto zamezíme i tomu, že budou do vyhledávání zahrnuty výskyty hezitačního zvuku *hmm*, který se v nahrávkách velmi často objevuje. Rovněž jako výchozí atribut zvolíme *word*, aby korpus nevyhledával například slovní tvary *víc* lemmatu *hodně*. Podoba dotazu je vidět níže na obrázku 41.



Obrázek 41 – Vyhledávání náslovného *h-* před samohláskami, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial)

Písmeno *h* na počátku dotazu značí, že hledáme slova na něj začínající. Potřebujeme však vyhledat slova, v nichž bude za tímto *h* následovat samohláska. Hranatá závorka je dalším

regulárním výrazem, který nabízí výběr jednoho libovolného znaku uvnitř těchto závorek, proto jsou do nich napsány všechny samohlásky kromě y (to je na rovině dialektologické zaznamenáno pouze v případě tvrdé výslovnosti). Následuje opět tečka pro jeden libovolný znak a hvězdička pro jeho libovolné opakování. Pro takto položený dotaz vyhledal korpus 104 výskytů slov, z nichž ale nejsou všechna s náslovným *h-*, nýbrž jsou to slova, která na *h-* ve spisovném jazyce začínají.

Výběr roviny *ort* jako dalšího pozičního atributu, který je vedle KWIC zobrazen, nám umožňuje lepší orientaci ve výsledcích, nicméně stále není nutné je procházet ručně, ale lze použít další nástroj, tentokrát filtr. Ten umožňuje výsledky dotazů selektovat na základě určité dodatečné podmínky. Pomocí pozitivního filtru můžeme ze všech výskytů zobrazit jen určitou skupinu, naopak pomocí negativního můžeme jinou skupinu z celkových výskytů vyjmout. V tomto případě použijeme negativní filtr, jímž vyjmemme na rovině ortografické všechna slova, která začínají na *h-*. V horním řádku vybereme záložku *Filtr* a následně *Negativní...*, objeví se tabulka, do hledacího řádku zapíšeme *h.\** a upřesníme parametry – povolíme regulární výrazy a jako poziční atribut vybereme *ort*. Rozsah hledání navíc musíme změnit na *0* včetně KWIC, aby byl negativní filtr aplikován pouze na klíčové slovo (viz obrázek 42).

Aplikovat operaci "negativní filtr" ✕

Pokročilý dotaz  | Klávesnice | Předchozí dotazy | Interpretace dotazu

**TIP** U zarovnaných korpusů je rychlejší hledat jen v textech společných pro všechny prohledávané jazyky. Najít tyto texty lze pomocí funkce 'omezit hledání - zúžit výběr' [\(další tip\)](#)

Rozsah hledání od:  do:  včetně KWIC

Shoda velikosti písmen  ?

Povolit regulární výrazy

Výchozí atribut:

Obrázek 42 – Aplikace negativního filtru na poziční atribut *ort*, zdroj: [https://www.korpus.cz/kontext/view?maincorp=dialekt\\_v2\\_dial&viewmode=kwic&pagesize=40&attrs=word%2Cort&attr\\_vmode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~1AGq0mayoow0](https://www.korpus.cz/kontext/view?maincorp=dialekt_v2_dial&viewmode=kwic&pagesize=40&attrs=word%2Cort&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~1AGq0mayoow0)

Po aplikování negativního filtru se objeví pouze ty výskyty slov, která začínají na ortografické rovině na samohlásku, na dialektologické jsou ale vyslovována

s náslovným *h-*, dohromady jich je 8 (viz obrázek 43). Nejvíce výskytů z nich má tvar *horingle* – 3krát, následuje 2krát tvar *horíngle*, po 1 výskytu tvar *horinglički*, *huš* a *hucháč*. Výskyt byl tedy zaznamenán celkem pouze u 3 slov, neboť právě u slova *horingle* se jedná jen o jeho další tvary.<sup>139</sup>

Výskytů: 8   <a href="#">lp.m.: Spočítat</a> (vztaheno k podmnožině určené vybranými typy textů)   ARF: 2,27   Výsledek je seřazen			
Výběr řádků: základní			
<input type="checkbox"/>	0017-C-ZAC-1963-SA	nebo hrnec nebo ten , jak se tomu říká ,	<b>hucháč/hucháč</b> , s pežežma .   +   řčko uš pošestí , pošestí uš
<input type="checkbox"/>	0240-C-ZAC-2007-N	, tak se to vom* , vocukruje a vomastí a	<b>huš/uzš</b> to můžete jist .   +   hmm .   +   to je dobrota .
<input type="checkbox"/>	0256-C-ZAC-2007-N	vona ji říká :   +   (rušivý zvuk)   +   " ti máš rákí novi	<b>horingle/oringle</b> ."   +   (smích)   +   a vona nevedela Kateřina , co to
<input type="checkbox"/>	0256-C-ZAC-2007-N	jo ?   +   tak to teda ...   +   mi sme jagživi říkali	<b>horingle/oringle</b> .   +   Kateřina koukala .   +   jó ? no , to je
<input type="checkbox"/>	0256-C-ZAC-2007-N	.   +   hoďie je to tu skomolení .   +   mi sme říkali	<b>horingle/oringle</b> , anebo dij bili děti malí , tag bili horinglički
<input type="checkbox"/>	0256-C-ZAC-2007-N	horingle , anebo dij bili děti malí , tag bili	<b>horinglički/oringličky</b> a nič jinoho sme ne* ...   +   ne , ičko mají
<input type="checkbox"/>	0262-C-ZAC-2007-N	já řekla :   +   " řetku , já bi sém chfela	<b>horingle/oringle</b> ."   +   dál mámě sto korun , mámě naši sto
<input type="checkbox"/>	0262-C-ZAC-2007-N	do mňesta sme šli pjěški , z mňesta ...   +   ale	<b>horingle/oringle</b> stáli třiced dva korun , to teki fur vim .

Obrázek 43 – Výskyty náslovného *h-* po aplikaci negativního filtru na ortografickou rovinu, zdroj: [https://www.korpus.cz/kontext/view?maincorp=dialekt\\_v2\\_dial&viewmode=kwic&pagesize=40&attrs=word%2Cort&attr\\_mmode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~9u](https://www.korpus.cz/kontext/view?maincorp=dialekt_v2_dial&viewmode=kwic&pagesize=40&attrs=word%2Cort&attr_mmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~9u)

Z výsledků vyhledávání je patrné, že náslovné *h-* se v domažlickém úseku západočeské skupiny objevuje, a to hlavně v nové vrstvě nahrávek, ze staré vrstvy pochází pouze 1 výskyt (z důvodu absence materiálu, jinak by byl jev u nejstarší generace mluvčích více rozšířený). Tento rys lze charakterizovat na Domažlicku<sup>140</sup> jako méně rozšířený, v současné době již na ústupu.

<sup>139</sup> Viz ČJA5, heslo 243 – *oringle*, s. 352, mapa s. 351, kde jsou zaznamenány i případy na *hu-*.

<sup>140</sup> Kromě domažlického okrajového úseku se v korpusu objevují 2 výskyty *h-* z jihočeské nářeční oblasti, více o tomto jevu viz ČJA 5, položky 238–247. Možnost výskytu v jihočeské oblasti ukazuje, že rozšíření nářečního jevu bylo v minulosti mnohem širší.

## Cvičení č. 7 – Protetické v- před o- na Doudlebsku

Zadání: Absence protetického v- před o- je nářeční znak typický pro doudlebský okrajový úsek jihočeské nářeční podskupiny. Vyhledejte tento jev v korpusu DIALEKT a ověřte, zda se na Doudlebsku protetické v- opravdu nevyskytuje.

Postup: Vyhledávat budeme v korpusu *dialekt – v2 ort*, protože protetické v- se zaznamenává i na ortografické rovině, ta je navíc přehlednější, neboť se na ní nepromítají nářeční hlásky. Hledání omezíme na jihočeskou oblast a doudlebský typ (viz obrázek 17), nebo vybereme subkorpus *Doudlebsko* (viz obrázek 19), povolíme regulární výrazy. Hledáme všechna slova, ve kterých se před o- vyskytuje v-, do vyhledávání zadáme dotaz *vo.\**. Tečka a hvězdička jsou jedny z výše zmíněných regulárních výrazů, tečka značí jeden libovolný znak, hvězdička znamená opakování předcházejícího znaku. Tímto způsobem zajistíme, že korpus vyhledá v doudlebském okrajovém úseku všechna slova, která začínají na *vo-*, nehledě na jejich délku (viz obrázek 44).

Obrázek 44 – Vyhledávání protetického v- před o-, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_ort](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_ort)

Vyhledávání ukázalo 114 slovních tvarů v- před o- na počátku slova. Dalším krokem je projít výsledky nejlépe pomocí frekvenční distribuce slovních tvarů (*Frekvence – Slovní tvary [A=a] – Tabulkové zobrazení*) a zjistit, ve kterých výskytech je v- opravdu protetické, neboť se zobrazují i slova, která začínají na *vo-*. Ze 114 výskytů se protetické v- před o- vyskytlo v 52 slovních tvarech, což tvoří téměř polovinu výskytů, přesněji 45,6 %. Bylo to například ve slovech *vožralej* (viz obrázek 45 níže), *vodpoledne*, *vodtud*,



*vatáče*, *vobilí*. Nejvíce výskytů bylo však ve tvarech zájmen – 10krát zájmeno *voni*, 6krát zájmeno *von*; dále pak 5krát předložka *vod* a 3krát předložka *vo*.

Výskytů: 114	l.p.m.: Spočítat	ARF: 11,27	Výsledek je setříděn
Výběr řádků: základní			
0028-C-JIC-1963-SA	tak povídá . to víte ..	+	taky se přišlo do
0028-C-JIC-1963-SA	to bylo vždycky při každé příležitosti . se povídalo o		<b>volí/volí</b> .. to bylo vždycky při každé příležitosti . se povídalo
0030-C-JIC-1967-SA	ona na to byla . protože .. ona potřebovala tu		<b>volech/volech</b> .. o dobytku . o hospodářství . jak rostlo obilí
0030-C-JIC-1967-SA	.. protože .. ona potřebovala tu vodu .. tu umrlčí		<b>vodu/vodu</b> .. tu umrlčí vodu . jak ona říkala ..   +   nu
0030-C-JIC-1967-SA	. jak ona říkala ..	+	<b>vodu/vodu</b> . jak ona říkala ..   +   nu vzala si tu umrlčí
0030-C-JIC-1967-SA	.. ale po tej vodě jak jsme mu dali tu		<b>vodu/vodu</b> do džbánku .. když byla . tam . už hotová
0030-C-JIC-1967-SA	počítaly na toho muže tý Mejšický .. a když přišel		<b>vožraje/vožral*</b> .. domů .. tak mu to pěkně ..   +   (nadechnutí)   +   (nadechn
0030-C-JIC-1967-SA	mu to namíchaly do něčeho .. nebo jestli mu tu		<b>vodu/vodu</b> daly vypít .. to víte že to byla ošklivota ..
0030-C-JIC-1967-SA	měl .. ale .. já nevím .. ale po tej		<b>vodě/vodě</b> jak jsme mu dali tu vodu vypít .. tak přestal
0030-C-JIC-1967-SA	.. ale po tej vodě jak jsme mu dali tu		<b>vodu/vodu</b> vypít .. tak přestal pít ..   +   a já jsem ji
0031-C-JIC-1964-SA	.. no a pak se navečer zadělalo ..	+	<b>vođa/vođa</b> .. poslilo se to .. teplá voda .. a kmin

Obrázek 45 – Výsledky dotazu *vo.\**, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=dial&attr\\_vmode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~Zq8SsklymOWg](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=dial&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~Zq8SsklymOWg)

Četnost výskytů nelze ale v tomto případě považovat za finální výsledek, neboť nevíme, jaký je poměr oproti výskytům slov začínajících na *o-*, v nichž protetické *v-* chybí, tyto hodnoty je nutné následně porovnat. Pokud změním dotaz na *o.\**, korpus vyhledá na Doudlebsku 284 slovních tvarů, které začínají na *o-*, za nímž následuje libovolný počet hlásek. Jedná se například o slovní tvary zájmen (*ona*, *on*, *oni*), dále slovesa (*odešel*, *oženit*, *odrhali*) či substantiva (*obilí*, *oškvarky*, *održky*) a další (viz obrázek 46 níže).

Výskytů: 284	l.p.m.: Spočítat	ARF: 24,89	Výsledek je setříděn		
Výběr řádků: základní					
<input type="checkbox"/>	0028-C-JIC-1963-SA	šel humny ..	+	no a .. tak humny jsem to	<b>obešel/obešel</b> .. spotichma .. aby to žádný tak jako nepozoroval ..
<input type="checkbox"/>	0028-C-JIC-1963-SA	spotichma .. aby to žádný tak jako nepozoroval .. a		<b>oni/oňi</b> už tam krmili .. u nich . doma ..   +   oni	
<input type="checkbox"/>	0028-C-JIC-1963-SA	oni už tam krmili .. u nich . doma ..	+	<b>oni/oňi</b> mě žádný neviděli než @ ta moje žena .. ta	
<input type="checkbox"/>	0028-C-JIC-1963-SA	co chci .. jsem seděl dál ..	+	<b>ona/ona</b> potom tam někde . splašila . děti .. ty byly	
<input type="checkbox"/>	0028-C-JIC-1963-SA	pole .. šli pro něj . pro tátu a ..		<b>on/on</b> přišel domů . a přivítal mě a .. povídalo se	
<input type="checkbox"/>	0028-C-JIC-1963-SA	mě a .. povídalo se zas dál .	+	<b>o/o</b> pouti . kterak to bylo .. a jak jsme měli	
<input type="checkbox"/>	0028-C-JIC-1963-SA	. to bylo vždycky při každé příležitosti . se povídalo		<b>o/o</b> volech .. o dobytku . o hospodářství . jak rostlo	
<input type="checkbox"/>	0028-C-JIC-1963-SA	vždycky při každé příležitosti . se povídalo o volech ..		<b>o/o</b> dobytku . o hospodářství . jak rostlo obilí .. jak	
<input type="checkbox"/>	0028-C-JIC-1963-SA	příležitosti . se povídalo o volech .. o dobytku .		<b>o/o</b> hospodářství . jak rostlo obilí .. jak je tam a	
<input type="checkbox"/>	0028-C-JIC-1963-SA	volech .. o dobytku . o hospodářství . jak rostlo		<b>obilí/obilí</b> .. jak je tam a tam na tom gruntě ..	
<input type="checkbox"/>	0028-C-JIC-1963-SA	jak je tam a tam na tom gruntě .. pěkný		<b>obilí/obilí</b> .. jak zase je tam horší . anebo co ..	
<input type="checkbox"/>	0028-C-JIC-1963-SA	tam dál moučky .	+	<b>obyčejně/občejně</b> .. o ženění se ani nepovídalo ..   +   no až potom	

Obrázek 46 – Výsledky vyhledávání dotazu *o.\**, zdroj: [https://www.korpus.cz/kontext/view?maincorp=dialekt\\_v2\\_ort&viewmode=kwic&pagesize=40&attrs=word&attr\\_vmode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~5YgSMGywaew](https://www.korpus.cz/kontext/view?maincorp=dialekt_v2_ort&viewmode=kwic&pagesize=40&attrs=word&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~5YgSMGywaew)

Celkový počet slovních tvarů, které začínají na *o-* nebo *vo-*, je tedy 336, 52 z nich je s protetickým *v-*, což je 15,5 %. Z této informace se dá již soudit, že protetické *v-* je sice na Doudlebsku zastoupeno, avšak ne ve velké míře a jeho absenci lze proto stále považovat za jev charakteristický pro chodský nářeční typ. Navíc pouze jeden z výskytů pochází ze staré vrstvy nahrávek, tudíž lze soudit, že protetické *v-* se dostává

na Doudlebsko vlivem blízkosti krajského města. Z něj se tam šíří městská mluva, v níž se protetické *v*- běžně objevuje.<sup>141</sup>

---

<sup>141</sup> Položka v ČJA viz ČJA 5, položka 258, s. 370, mapa s. 371.

## Cvičení č. 8 – Rozložená výslovnost měkkých retnic *p, b, m, v* na Doudlebsku

Zadání: Rozložená výslovnost měkkých retnic je rys typický pro doudlebský okrajový úsek jihočeské nářeční oblasti. Jedná se o retnice *p, b, m* a *v*, které se před *i* vyslovují jako *pji, bji, mji* a *vji*. Vyhledejte tyto retnice v korpusu *DIALEKT* se zavřenou i spisovnou výslovností a srovnajte výsledky.

Postup: Vyhledávání rozložené výslovnosti měkkých retnic bude probíhat na rovině dialektologické, zvolíme tedy korpus *dialekt* – v2 *dial*. Na rovině ortografické nebude tento jev zaznamenán, neboť se zde výslovnost převádí na spisovnou. Vyhledávání je nutné omezit na jihočeskou nářeční oblast a doudlebský typ (viz obrázek 17), nebo vybrat subkorpus *Doudlebsko* (viz obrázek 19). Dále povolíme regulární výrazy, které budeme při vyhledávání používat. Jako výchozí atribut zvolíme *word* (slovní tvar). K vyhledání tohoto nářečního jevu se nabízí opět dva jednoduché postupy.

(1) První možností, jak vyhledat jednotlivé retnice s rozloženou výslovností, je zadat každou z nich ve spojení s *j* a *i* spolu s regulárními výrazy (tečkou a hvězdičkou). Takto zajistíme, že budou vyhledána všechna slova s daným spojením hlásek bez ohledu na to, kde ve slově se hlásky objevují a jak jsou slova dlouhá. Podoba dotazu viz obrázek 47 níže, v ostatních dotazech jen změníme retnici a postupně vyhledáme všechna spojení. Následně vyhledáme ty samé dotazy s malou obměnou – vynecháme z dotazu *-j-*, díky čemuž vyhledá korpus jednotlivé retnice se spisovnou výslovností (viz obrázek 48).



Hledat v korpusu

dialekt\_v2\_dial ★

Pokročilý dotaz | Klávesnice | Předchozí dotazy | Interpretace dotazu

\*pji.\*

Obrázek 47 – Podoba dotazu k vyhledání rozložené výslovnosti retnice *p* před *i*, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial)



Obrázek 48 – Podoba dotazu k vyhledání spisovné výslovnosti retnice *p* před *i*, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial)

Po zadání obou dotazů s *p* zjistíme, že v korpusu není vůbec zachycena rozložená výslovnost této retnice, naopak pro spisovnou výslovnost vyhledal korpus 40 výskytů. Ty samé dotazy vyhledáme s retnicí *b* – rozloženou výslovnost našel korpus 6krát, a to ve slovech *babjička*, *babjičinej*, *probjil*, *bjila* a 2krát ve slově *bjičišťata* (toto slovo bylo v obou případech zachyceno v nové vrstvě nahrávek, ostatní výskyty pocházejí z vrstvy staré). Spisovná výslovnost této retnice před *i* byla zaznamenána celkem 699 krát, tudíž podíl rozložené výslovnosti oproti spisovné je zanedbatelný (0,9 %). Ten samý postup opakujeme s retnicí *m* – rozložená výslovnost se objevila v korpusu jen jednou, a to ve slově *krmjili* v nahrávce ze staré vrstvy, naopak spisovná výslovnost 155krát. V tomto případě není rozložená výslovnost oproti spisovné opět zastoupena ani jedním procentem (podíl je 0,6 %). Poslední retnice, jejíž výslovnost byla zkoumána, je *v*, která nebyla (stejně jako *p*) s rozloženou výslovností vůbec zaznamenána, se spisovnou 250krát.

(2) Další možný postup, pomocí něhož lze ověřit výskyt rozložené výslovnosti retnic před *i* na Doudlebsku, nabízí vyhledávání pomocí jednoho dotazu. K tomu použijeme kromě již výše zmíněných regulárních výrazů ještě hranaté závorky (*[ ]*). Do závorek vepíšeme všechny retnice, které zkoumáme, budou fungovat jako alternativy, které bude korpus vyhledávat. Podoba dotazu je vidět níže na obrázku 49.



Obrázek 49 – Podoba dotazu k vyhledání rozložené výslovnosti retnic v rámci jednoho dotazu, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial)

Pomocí takto sestaveného dotazu zobrazí korpus všechny výskyty rozložené výslovnosti retnic na Doudlebsku, a to bez nutnosti hledat každou retnici zvlášť. Samotné vyhledávání je tedy úspornější, naopak třídění výsledků časově náročnější, ale poslouží i pro kontrolu a ověření správnosti postupu předchozího. Korpus vyhledal 7 výskytů, což odpovídá i výsledkům – 6krát rozložená výslovnost retnice *b*, 1krát retnice *m*, ostatní 2 retnice nebyly zaznamenány. Výsledky tohoto dotazu viz obrázek 50 níže.

Obrázek 50 – Výsledky vyhledávání rozložené výslovnosti retnic před *i*, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr\\_ymode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~mUI4S8c0AyCw](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr_ymode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~mUI4S8c0AyCw)

Z výsledků obou výše užitých postupů je patrné, že rozložená výslovnost retnic před *i* už nepatří na Doudlebsku mezi živé nářeční rysy. Nasvědčuje tomu jednak fakt, že pouze 2 výskyty retnice *b* byly zaznamenány s rozloženou výslovností v nové vrstvě nahrávek, dále to, že retnice *m* se objevila jen jednou, a to ve staré vrstvě, nakonec úplná absence rozložené výslovnosti retnice *p* a *v*. V tomto případě se navíc nelze 100% spolehnout na nedostatek materiálu jakožto důvod pro absenci hledaného jevu, neboť výskyty spisovné výslovnosti těchto retnic jsou v některých případech opravdu vysoké. Obecně lze tedy zhodnotit tento nářeční rys jako archaický, dnes již na ústupu vzhledem tomu, že jsou v nové vrstvě nahrávek doloženy jen dva výskyty rozložených retnic s *-ji-*.<sup>142</sup> Jinak je tomu v případě rozložené retnice *mje-*, jejíž problematice se věnuje cvičení 4 na s. 78.

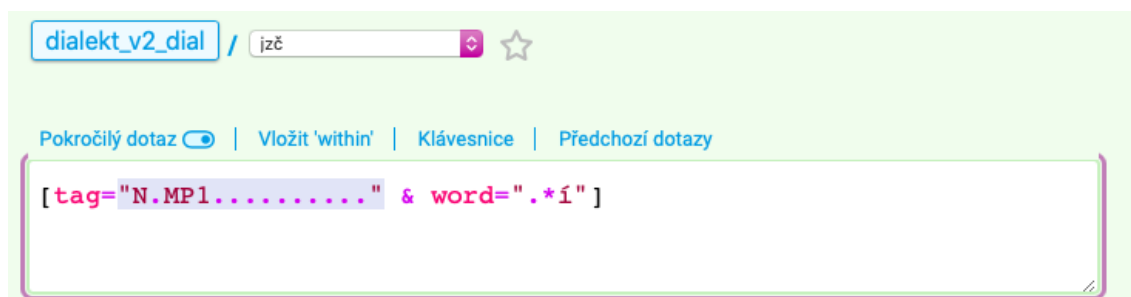
<sup>142</sup> Více k nářečnímu jevu viz ČJA 5, položka 227a, s. 332, mapa s. 333.

## 6.2 TVAROSLOVÍ

### Cvičení č. 9 – Koncovka *-í* v 1. pádu plurálu životných maskulin v západočeské a jihočeské nářeční oblasti

Zadání: Koncovka *-í*<sup>143</sup> v 1. pádu plurálu maskulin životných<sup>144</sup> namísto krátkého *-i* je typickým jevem západočeské a jihočeské nářeční oblasti. Vyhledejte tuto koncovku v korpusu DIALEKT a uveďte několik příkladů slov, ve kterých se vyskytuje. Zhodnoťte, zda je tento jev používanější v jihočeské, nebo západočeské oblasti. Ověřte, zda je častější ve srovnání se spisovnou koncovkou *-i* a jestli je stále užívaný u nejstarší generace mluvčích.

Postup: Vyhledávání bude probíhat v korpusu *dialekt* – v2 *dial*, neboť na ortografické rovině se nepromítá dloužení či krácení a samohlásky jsou zaznamenávány ve spisovné podobě. Dále je nutné omezit hledání na západočeskou a jihočeskou nářeční oblast (viz obrázek 17), nebo vybrat subkorpus *jzč* (viz obrázek 19). Z důvodu složitosti zadání povolíme pokročilý dotaz – jeho podoba viz níže na obrázku 51.



Obrázek 51 – Podoba pokročilého dotazu k vyhledání koncovky *-í* v 1. p. pl. maskulin životných, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial)

Při vyhledávání pomocí pokročilého dotazu se pracuje s tzv. CQL jazykem, jež je kombinací regulárních výrazů, pozičních atributů a tagů neboli morfologických značek. Každý takový dotaz má pevný formát – zapisuje se do hranatých závorek, dále je nutné uvést poziční atribut (lemma, word, tag, slovní druh apod.). Následuje znak = a po něm dvoje horní uvozovky, do nichž se zapisuje, co konkrétně bude předmětem vyhledávání. Je tedy nutné specifikovat dotaz tak, aby vyhledal podstatná jména rodu mužského životného v 1. pádu plurálu, která končí na *-í*. Jako první poziční atribut byl tedy zvolen

<sup>143</sup> V tomto případě dochází k prolínání roviny hláskoslovné a tvaroslovné, dále také cv. 10, s. 97.

<sup>144</sup> *Mapka* uvádí kromě životných maskulin koncovku *-í* i u zájmen a číslovek, zde se zaměřujeme pouze na podstatná jména, nicméně i v korpusu DIALEKT jsou doklady tohoto tvaru právě u zájmen a číslovek.

tag, který umožňuje vybrat morfologické charakteristiky daných slov. Tag obsahuje celkem 15 pozic, jež lze specifikovat. První pozice určuje slovní druh, tedy *N* jako podstatná jména. Na druhém místě stojí tečka, druhá pozice je totiž bližší určení slovního druhu, které není v tomto případě nutné. Třetí pozici zastupuje jmenný rod, tedy *M* pro životná maskulina. Čtvrtá pozice je označena písmenem *P* pro plurál, pátá pozice číslem *1* pro označení pádu. Dalších 10 pozic je označeno tečkou, neboť opět nejsou pro vyhledávání potřebné (přivlastňovací rod, čas, stupeň, negace apod.). Pomocí znaku & je k tagu připojen další poziční atribut – word, jehož pomocí vyhledáme určitý tvar slova. Tvar zapíšeme pomocí regulárního výrazu, který již známe z předchozích cvičení, tedy *. \*í* – tečka označující libovolný znak, hvězdička jeho opakování, *-í* jako koncovka. Zadání tohoto pokročilého dotazu je sice složitější než vyhledávání pouze pomocí regulárních výrazů, nicméně lze takto dotaz blíže specifikovat a předejít tak tomu, že se vyhledají všechna slova končící na *-í*. Pak by musela proběhnout selekce výsledků, která by byla časově náročnější.<sup>145</sup>

Celkem korpus vyhledal 116 výskytů substantiv, která mají tvar 1. pádu plurálu maskulin životných a končí na *-í*. Mezi nimi se objevila např. slova jako *chlapci*, *sousedí*, *pohuncí*, *kluci*, *zedníci* (viz obrázek 52). Z celkového počtu výskytů jich 15 pochází z jihočeské oblasti (všechny z nové vrstvy nahrávek), 101 z oblasti západočeské (z nich 6 ze staré, 95 z nové vrstvy).

Výskyt	Výskyt	Výskyt	Výskyt
0017-C-ZAC-1963-SA	to bálo šestkrát . . . a tenkrát mí děti , dva	chlapci/chlapci	takoví , menčí bali , nákejty deset let jeden a
0017-C-ZAC-1963-SA	feh Brichtů s toho dvora u Krupů , to s'ú	s'usedí/sousedí	. . . tag zase vipravovala , di bala taková dříče ,
0017-C-ZAC-1963-SA	di bala taková dříče , tak tám tera slúžili	pohunci/pohunci	, vite , pás se dobitek . . . di bala taková
0019-C-ZAC-1964-SA	. . . no , diš sme se vitancovali , tak nás	chlapci/chlapci	doprovodili a šli sme domú . . . nebo chodil na Štedrej
0021-C-ZAC-1963-SA	za prvíi svjetoví válki tadi na Kralovicku tak přišli ti	komisaři/komisaři	, ti , jak rekvirovali vobili , né , do
0021-C-ZAC-1963-SA	takle do kapsi do zimníku a jeli . . . a ti	komisaři/komisaři	, diš jeli po silnici , tak si řekli :
0245-C-ZAC-2007-N	potom us se chodilo , jó , a potom nás	chlapci/chlapci	šupali . . . (nadechnuti) . . . (nadechnuti) . . . teda diš sme dicki f pořídli
0245-C-ZAC-2007-N	uš sou jenom ti klapački , a tak se chodí	chlapci/chlapci	a sou tu @ štiri parti . . . (ruch z ulice) . . . (nadechnuti) . . . hmm
0245-C-ZAC-2007-N	dicki důchotkře a nebo holt ... (nadechnuti) . . . (smích) . . . (smích) . . . tudle	souseďi/souseďi	se k nám nasřehovali , tak ti maj tři děti
0246-C-ZAC-2007-N	(ruch z ulice) . . . 1 . . . (ruch z ulice) . . . (nadechnuti) . . . jo , to plafili jenom	chlapci/chlapci	, fém muzikantom . . . hmm . . . (nadechnuti) . . . a potom ,
0246-C-ZAC-2007-N	nevím uš , ale přěšne , tak potom ti @	voráči/voráči	, teda to bili ti chlapci , chodili vot statku
0246-C-ZAC-2007-N	tak potom ti @ voráči , teda to bili ti	chlapci/chlapci	, chodili vot statku , vot , gde mřeli dřěfata
0246-C-ZAC-2007-N	. . . (nadechnuti) . . . no a tak @ chodili práve voráči ,	voráči/voráči	. . . (ruch z ulice) . . . no , za nás uš to , uš

Obrázek 52 – Výsledky vyhledávání koncovky *-í* v 1. p. pl. maskulin životných pomocí pokročilého dotazu, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr\\_ymode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~soMcwMoOiIi](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr_ymode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~soMcwMoOiIi)

<sup>145</sup> Zadávání pokročilých dotazů pomocí tagu není v korpusu DIALEKT spolehlivé, neboť tagování (určování slovních druhů u nářečí, kde je velká variabilita) vždy plně neodpovídá. Doporučujeme tedy výsledky pomocí běžného dotazu s regulárními výrazy přesto ověřit, byť je to časově náročnější.

Pokud pokročilý dotaz zadáme znovu, ale namísto *-í* vyhledáme poziční atribut word s koncovkou *-i*, získáme 101 výskytů maskulin životných v 1. pádu plurálu končících na *-i*. Z nich je 43 výskytů doloženo v západočeské oblasti (2 ve staré vrstvě nahrávek, 41 v nové), 58 v oblasti jihočeské (7 výskytů ve staré vrstvě, 51 v nové).

Celkový počet výskytů je 217, z nich tvoří koncovka *-í* 53,5 %, tedy zhruba polovinu. Pokud se podíváme na podíl koncovky v jednotlivých oblastech, v západočeské je to 101 ku 43, tedy 70,1 %. Pokud srovnáme výsledky z jihočeské oblasti, zde je to 15 ku 58 ve prospěch spisovné koncovky, koncovka *-í* činí tedy pouze 20,5 % z celkového počtu 73 výskytů.

Výsledky vyhledávání ukazují, že v západočeské oblasti se jedná o nářeční rys živý u nejstarší generace mluvčích, ale zároveň v současnosti užívaný. Rozhodně se zde jedná o jev frekventovanější, nicméně výsledky ukazují, že i v jihočeské oblasti se v menší míře vyskytuje.<sup>146</sup>

---

<sup>146</sup> Viz ČJA 5, položka 191, s. 288, mapa s. 289. Dále viz ČJA 4, položka 90 a 91, s. 154, mapa s. 155. Položka 91 potvrzuje větší výskyt nářečního jevu v západočeské nářeční oblasti, čemuž odpovídají i výsledky hledání.

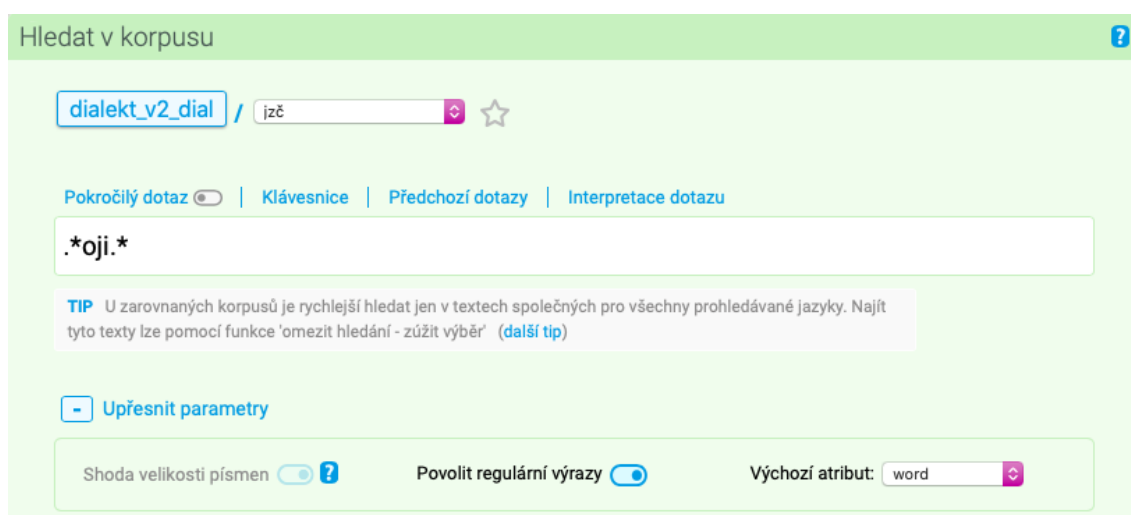


## Cvičení č. 10 – Změna skupiny *-ovi* na *-oji* v západočeské a jihočeské nářeční oblasti

Zadání: Západočeská a jihočeská nářeční oblast se vyznačuje mnoha tvaroslovnými rysy. Jedním z nich je změna skupiny *-ovi-* na *-oji-* zejména v místních názvech. Vyhledejte tento jev v korpusu *DIALEKT* a okomentujte četnost jeho výskytů. Je to jev typický pro západočeskou, nebo spíše jihočeskou oblast? Nachází se i v nové časové vrstvě nahrávek?

Postup: Změnu skupiny *-ovi-* na *-oji-* v západočeské a jihočeské nářeční oblasti budeme vyhledávat na rovině dialektologické, neboť pokud se změna nevyskytuje v koncovce slova, nebude na ortografické rovině zachycena. Jako korpus zvolíme tedy *dialekt* – v2 *dial*. Omezíme hledání na západočeskou a jihočeskou nářeční oblast (viz obrázek 17), nebo vybereme k vyhledávání subkorpus *jzč* (viz obrázek 19). Pro sestavení dotazu budeme potřebovat regulární výrazy, tudíž je povolíme, jako výchozí atribut zvolíme *word*.

Naším cílem je v korpusu vyhledat v západočeské a jihočeské nářeční oblasti všechna slova o libovolné délce, která obsahují část *-oji-* nehledě na její pozici ve slově. Z regulárních výrazů budeme potřebovat tečku (jakýkoliv jeden znak) a hvězdičku (libovolné opakování předchozího znaku). Podoba dotazu viz obrázek 53 níže.



Hledat v korpusu

dialekt\_v2\_dial / jzč

Pokročilý dotaz | Klávesnice | Předchozí dotazy | Interpretace dotazu

**\*oji.\***

**TIP** U zarovnaných korpusů je rychlejší hledat jen v textech společných pro všechny prohledávané jazyky. Najít tyto texty lze pomocí funkce 'omezit hledání - zúžit výběr' ([další tip](#))

**-** Upřesnit parametry

Shoda velikosti písmen | Povolit regulární výrazy | Výchozí atribut: word

Obrázek 53 – Podoba dotazu k vyhledání slovních tvarů s *-oji-*, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial&usesubcorp=jzč](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial&usesubcorp=jzč)

Pro takto položený dotaz vyhledal korpus na dialektologické rovině 44 výskytů slovních tvarů s *-oji-* (viz obrázek 54 níže). Hned na první pohled z konkordančních řádků

ale můžeme vidět, že zdaleka ne všechny tvary odpovídají změně *-ovi-* na *-oji-*, proto je třeba výsledky manuálně projít. Další možností je sestavit frekvenční distribuci slovních tvarů (*Frekvence – Slovní tvary [A=] – Tabulkové zobrazení*), ale i v té je nutné slovní tvary vytrítit.

Výskyt: 44	lp.m.: 584,31	vztaheno k subkorpusu "jzč"	ARF:-	Výsledek je setříděn
<input type="checkbox"/>	0022-C-ZAC-1965-SA	, teda přiměňi Noha, bil tesařem a dělali ve	Stražejich/Stražovicích	.   +   a díž bili u vobjeda, tak viděli,
<input type="checkbox"/>	0025-C-JIC-1964-SA	se musela ta louka, tedí na zimu se musela	pohnojil/pohnojit	, no, po jaru se musel ten hnúj schrabat
<input type="checkbox"/>	0028-C-JIC-1963-SA	hastman nekam ftaý do toho jezera.	vystrojil/vystrojil	a šel sem tam.   +   abí me žáneý nevidel,
<input type="checkbox"/>	0028-C-JIC-1963-SA	" no, jak chcete ví."	tátovi/tátovi	to asi bilo tehdi takí uš jasní, kerak to
<input type="checkbox"/>	0029-C-JIC-1964-SA	zamet som dvořeček, šel som pomoct g dobitku,	napojil/napojit	, nakrmit pomost.   +   díž bilo šez hođin, selka
<input type="checkbox"/>	0030-C-JIC-1967-SA	do truhlí .   +   (nadechnutí)   +   nu, tak pjeķie šla takí	strojil/strojil	toho Mařeje Kožíšníkovo do truhlí, no a to u
<input type="checkbox"/>	0039-M-JIC-1964-SA	kamarádka pá:   +   " já domu nepudu ! já se	bojím/bojím	, poř spátí ! "   +   chfela uřítak.   +   a huš
<input type="checkbox"/>	0239-C-ZAC-2007-N	.   +   (ruch z ulice)   +   hmm .   +   a to ste se vdávala f	krojil/krojil	, že jo, ešfe ?   +   (ruch z ulice)   +   no, to
<input type="checkbox"/>	0244-C-ZAC-2007-N	tea tři kilometri ot sebe, a každá veselice má	svoji/svoji	tipickou krajku.   +   hmm.   +   jo.   +   @ vite
<input type="checkbox"/>	0247-C-ZAC-2007-N	už bili teda jenom v* @ ženi bili voblečeni f	krojil/krojil	a mušski bili, muski bili normálne.   +   (nadechnutí)   +
<input type="checkbox"/>	0249-C-ZAC-2007-N	tak to, a díž bil tvrď", tak se	ukrojil/ukrojil	takověj krajic.   +   a strčilo se to do toho,
<input type="checkbox"/>	0249-C-ZAC-2007-N	kinutí.   +   a potom se vzali tí bramburoví knedlíki,	rozkrojilo/rozkrojilo	se to a napíchlí na vidličku a do toho máčedla
<input type="checkbox"/>	0250-C-ZAC-2007-N	takí maji svoji mlvu.   +   no.   +   tam teki maji	svoji/svoji	, jo.   +   nó, tam určite sou op* @
<input type="checkbox"/>	0260-C-ZAC-2007-N	tu !řuku propil.   +   (ruch z ulice)   +   (srkání)   +   2.   +   (pousmání)   +   propil	jalojici/jalojici	, tu vidím dođnes, jak jí vedé @.

Obrázek 54 – Výsledky vyhledávání slovních tvarů s *-oji-* v jihozápadočeské nářeční oblasti, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr\\_ymode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&fromp=1&q=~cWuQCI4uE42i](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr_ymode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&fromp=1&q=~cWuQCI4uE42i)

Ať už zvolíme jakýkoliv postup selekce materiálu, nakonec dojdeme k tomu, že změna *-ovi-* na *-oji-* je v jihozápadočeské skupině zachycena 8krát, a to ve tvarech *Stražojicích* (na ortografické rovině *Strážovicích*), *tátovi* (na ort. *tátovi*), *jalojici* (na ort. *jalovici*), *Novákojdz* (na ort. *Novákojic*), *Králojice* (na ort. *Královice*), *Straškojice* (na ort. *Strážkovicice*), *Řevnojice* (na ort. *Řevnovice*) a *Straškojicích* (na ort. *Strážkovicích*). Z celkového počtu 44 výskytů slovních tvarů s *-oji-* se jedná o změnu z *-ovi-* pouze v 18,2 % případů. To není sice vysoký podíl, ale ani to není zanedbatelný výsledek, naopak ukazuje, že se jedná stále o jev charakteristický.

Z výše zmíněných 8 slovních tvarů se změnou *-ovi-* na *-oji-* jich 5 pochází z jihočeské nářeční oblasti (62,5 %) (*tátovi*, *Králojice*, *Straškojice*, *Řevnojice*, *Straškojicích*), zbylé 3 z oblasti západočeské. Pouze 2 výskyty pocházejí ze staré vrstvy nahrávek (*Strážojicích*, *tátovi*), doklady z nové časové vrstvy ukazují na to, že se nářeční jev užívá i v současnosti.

Závěrem lze změnu *-ovi-* na *-oji-*<sup>147</sup> v západočeské a jihočeské nářeční oblasti zhodnotit jako jev ne častý, nicméně důležitou roli hraje, že byl zachycen, a to ve staré i v nové vrstvě nahrávek, což ukazuje na jeho stále užívání i v dnešní době. Dále lze dle výsledků jev považovat za charakteristický spíše pro jihočeskou oblast s ohledem na vyšší počet

<sup>147</sup> Viz ČJA 5, položka 314, s. 436, mapa s. 437; položka 315, s. 438, mapa s. 439.

výskytů (v tomto případě jsou ale čísla těsná a s dalším jazykovým materiálem se mohou změnit).

## Cvičení č. 11 – Nesklonný tvar rodinných jmen na *-ojc* v západočeské a jihočeské nářeční oblasti

Zadání: *Typickým tvaroslovným jevem pro západočeskou a jihočeskou nářeční oblast je nesklonný tvar rodinných jmen zakončený na -ojc namísto spisovného -ovic. Vyhledejte tento jev v korpusu DIALEKT a zhodnoťte míru užívání vzhledem k počtu výskytů. Objevuje se tento jev více v západočeské, nebo jihočeské oblasti? Je zachycen i v nové časové vrstvě nahrávek?*

Postup: Vyhledávání nesklonných rodinných jmen na *-ojc* můžeme v korpusu DIALEKT vyhledávat jak na rovině dialektologické, tak na rovině ortografické, na níž je tato koncovka zachována. Vybereme tedy buď korpus *dialekt – v2 dial*, nebo *dialekt – v2 ort*. Hledání omezíme na západočeskou a jihočeskou nářeční oblast bez bližšího určení (viz obrázek 17), nebo rovnou vybereme subkorpus *jzč* (viz obrázek 19). Povolíme regulární výrazy, jako výchozí atribut zvolíme *word*.

Vyhledávat budeme všechny slovní tvary ze západočeské a jihočeské nářeční oblasti, které končí na *-ojc*. Z regulárních výrazů budeme potřebovat tečku (jeden libovolný znak) a hvězdičku (libovolné opakování předchozího znaku). Podoba dotazu viz obrázek 55 níže.

Hledat v korpusu

dialekt\_v2\_dial / jzč

Pokročilý dotaz | Klávesnice | Předchozí dotazy | Interpretace dotazu

\*ojc

TIP Barevně zvýrazněný token s otazníkem v jednoduchém dotazu znamená možnost upřesnit jeho interpretaci. Použijte CTRL/Command a klik pro aktivaci. (další tip)

Upřesnit parametry

Shoda velikosti písmen Povolit regulární výrazy Výchozí atribut: word

Obrázek 55 – Podoba dotazu k vyhledání nesklonných rodinných jmen na *-ojc*, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial&usesubcorp=jzč](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial&usesubcorp=jzč)

Pro takto položený dotaz vyhledal korpus 41 slovních tvarů zakončených na *-ojc*. Pro lepší přehled slovních tvarů a zároveň kontrolu použijeme frekvenční distribuci. Na

horní liště zvolíme možnost *Frekvence*, následně *Slovní tvary [A=a]*, zvolíme tabulkové zobrazení. Slovní tvary a počet jejich výskytů viz obrázek 56 níže.

Frekvenční seznam

Zobrazení v grafech    Tabulkové zobrazení

Zobrazit konfidenční intervaly  ⓘ

1 / 1 (celkem: 6 položek)

Filtr	word [dial]	Freq	i.p.m.
1	p / n novákojc	23	305,44
2	p / n hruškojc	11	146,08
3	p / n králojc	3	39,84
4	p / n ernestojc	2	26,56
5	p / n strajškojc	1	13,28
6	p / n vranojc	1	13,28

Obrázek 56 – Frekvenční distribuce slovních tvarů na *-ojc*, zdroj: [https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attr=ort&attr\\_ymode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~f6aSeiUm4UOe&fcrit=word%2Fie%200~0%3E0&flimit=1&fq\\_s](https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attr=ort&attr_ymode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~f6aSeiUm4UOe&fcrit=word%2Fie%200~0%3E0&flimit=1&fq_s)

Z tabulky frekvenční distribuce slovních tvarů můžeme vidět, že tvarů na *-ojc* je v korpusu v jihozápadočeské oblasti doloženo jen 6. Z toho navíc 2 neodpovídají nesklonnému rodinnému jménu – *Straškojc*, *Vranojc* – jedná se zde o jména místní. Pokud tyto tvary vyjmeme pomocí negativního filtru (zvolíme písmeno *n* ve sloupci *Filtr*), zůstanou 4 slovní tvary a celkem 39 výskytů, které všechny odpovídají hledanému jevu. Z nich je nejčastěji zastoupeno jméno *Novákojc* (23 výskytů), dále *Hruškojc* (11 výskytů), *Králojc* (3 výskyty) a *Ernestojc* (2 výskyty). Je důležité brát v potaz, že příjmení Novák, Hruška a Král jsou zástupná (viz podkapitola 2.1.3, s. 29), jinak by pestrost příjmení byla mnohem vyšší.

Nyní vytvoříme nové tabulky frekvenční distribuce, a to podle nářeční oblasti a podle roku pořízení nahrávky, abychom zjistili zastoupení nářečního jevu v rámci oblastí a časových vrstev nahrávek. Na horní liště opět zvolíme záložku *Frekvence*, dále *Vlastní...*, pak *Podle typů textů* a jako strukturní atributy vybereme *doc.noblasi* a *doc.rok*. Zvolíme tabulkové zobrazení pro větší přehlednost, viz obrázek 57 níže.

Frekvenční seznam

Zobrazení v grafech    Tabulkové zobrazení

Zobrazit konfidenční intervaly  ⓘ

1 / 1    (celkem: 1 položka)

Filtr	doc.noblast	Freq	i.p.m.	
1	p / n	západočeská	40	899,77

1 / 1    (celkem: 6 položek)

Filtr	doc.rok	Freq	i.p.m.	
1	p / n	2005	24	4 622,05
2	p / n	2008	4	450,05
3	p / n	2009	4	207,06
4	p / n	2014	4	713,39
5	p / n	2011	2	211,06
6	p / n	195X	2	4 201,68

Obrázek 57 – Frekvenční distribuce slovních tvarů na *-ojc* podle nářeční oblasti a roku pořízení nahrávky, zdroj: [https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr\\_vmode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~3kWi22y](https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~3kWi22y)

Z tabulky frekvenční distribuce výše můžeme vidět, že nesklonný tvar rodinných jmen na *-ojc* není vůbec doložen v jihočeské nářeční oblasti, všech 40 výskytů pochází z oblasti západočeské. Co se týče zastoupení ve vrstvách nahrávek, pouze 2 výskyty pocházejí ze staré vrstvy nahrávek (195X), zbylých 38 výskytů z vrstvy nové, což ukazuje na časté užívání jevu v současnosti.

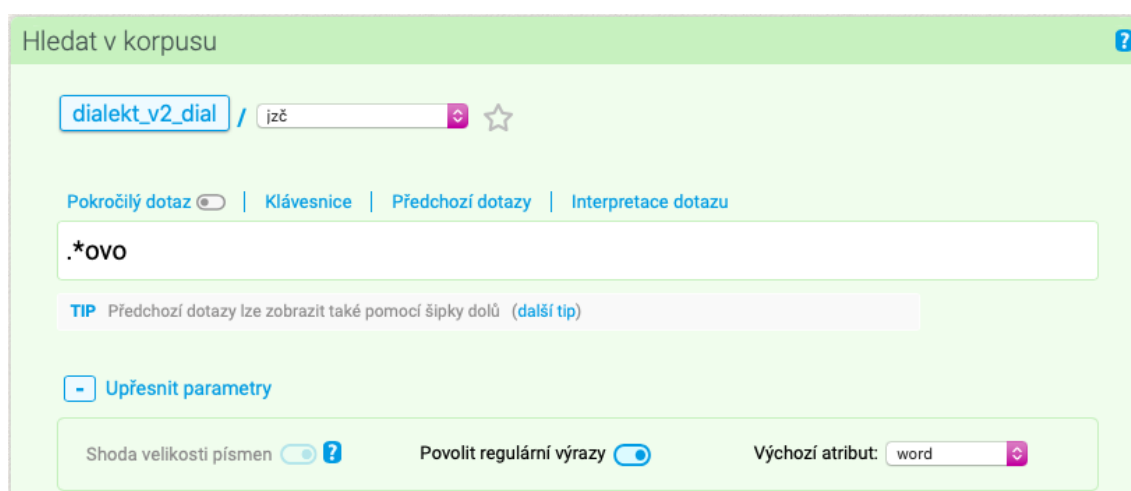
Závěrem lze tedy nesklonný tvar rodinných jmen na *-ojc* shrnout jako v korpusu častěji doložený v západočeské nářeční oblasti, neboť nebyl v korpusu DIALEKT ani jeden doklad pro tento jev v oblasti jihočeské, nicméně to může být z důvodu nedostatku materiálu. Navíc lze tvrdit, že se jedná o jev užívaný i v současnosti, a to díky vysokému výskytu v nové časové vrstvě nahrávek.

## Cvičení č. 12 – Tvar přivlastňovacího zájmena na *-ovo* v západočeské a jihočeské nářeční oblasti

Zadání: V západočeské a jihočeské nářeční oblasti je charakteristickým rysem ustrnulý tvar přídavných jmen se zakončením na *-ovo* namísto spisovného *-ův*. Vyhledejte tento tvaroslovný jev v korpusu *DIALEKT* a zjistěte, zda pro něj existují doklady. Pokud ano, lze tento jev označit s ohledem na počet výskytů jako často užívaný? Pro kterou nářeční oblast je typičtější?

Postup: Ustrnulý tvar přivlastňovacích zájmen na *-ovo* můžeme v korpusu vyhledávat jak na rovině dialektologické, tak na rovině ortografické, neboť na obou těchto rovinách je nářeční koncovka zachována, jako korpus tedy můžeme zvolit jak *dialekt – v2 dial*, tak *dialekt – v2 ort*. Hledání omezíme na západočeskou a jihočeskou nářeční oblast (viz obrázek 17), nebo pro vyhledávání zvolíme subkorpus *jzč* (viz obrázek 19). Při vyhledávání budeme potřebovat regulární výrazy, musíme je tedy povolit a jako výchozí atribut zvolíme *word*.

Cílem je v korpusu vyhledat všechny slovní tvary, které končí na *-ovo*. Jako regulární výrazy použijeme tečku (zastupující jeden libovolný znak) a hvězdičku (libovolné opakování předchozího znaku). Podoba dotazu viz níže na obrázku 58.



Hledat v korpusu

dialekt\_v2\_dial / jzč

Pokročilý dotaz | Klávesnice | Předchozí dotazy | Interpretace dotazu

\*ovo

TIP Předchozí dotazy lze zobrazit také pomocí šipky dolů (další tip)

Upřesnit parametry

Shoda velikosti písmen Povolit regulární výrazy Výchozí atribut: word

Obrázek 58 – Podoba dotazu k vyhledání ustrnulých tvarů přivlastňovacích zájmen na *-ovo*, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial&usesubcorp=jzč](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial&usesubcorp=jzč)

Pro takto zadaný dotaz korpus vyhledal 18 slovních tvarů na *-ovo*, viz obrázek 59 níže. Pokud se podíváme na konkordance, tak ale hned na první pohled můžeme vidět, že ne

ve všech případech se jedná o ustrnulý přivlastňovací tvar, a to v případě slovních tvarů *slovo*, *kremovo* a *hotovo*. Proto je nutné výsledky dále třídit.

Výskyt	IP.m.	Forma	Frekvence	IP.m.
0027-C-JIC-1965-SA		přivedli a, a zabili a prodali maso a bilo	hotovo/hotovo	1
0236-C-ZAC-2007-N		tak ?   +   to čecho jí nepovíjí   +   to je úžasní	slovo/slovo	5
0230-C-ZAC-2007-N		na jižní Moravě fíčíhoj to tak říkají, tam neznají	slovo/slovo	2
0236-C-ZAC-2007-N		čera si spomněla máma, jak mluvila, říká to	slovo/slovo	2
0236-C-ZAC-2007-N		smíjejou se, protože třeba na to	slovo/slovo	1
0263-C-ZAC-2007-N		napřel tám chodilo hromadu lidí   +   (nesrozumitelný hovor více mluvčích najednou)   +	Jirkovo/Jirkovo	1
0589-C-ZAC-2005-N		to je možná   +   nákej, vim, že nákej	ďedovo/ďedovo	1
0592-C-ZAC-2005-N		sme dobrí sousedi a i příbuzní po	taťinkovo/taťinkovo	1
0602-C-ZAC-2005-N		tohleto chci nafjalovo, @ obivák chci na	kremovo/kremovo	1
0604-C-ZAC-2005-N		to je možná   +   nákej, vim, že nákej	taťinkovo/taťinkovo	1
0607-C-ZAC-2005-N		neří viňenek, je to fech Novákojc   +   hmm   +   Jarovo/Jarovo	Jarovo/Jarovo	1
0610-C-ZAC-2009-N		hmm   +   a terko s*, strejda, jako	taťinkovo/taťinkovo	1
0614-C-ZAC-2005-N		Novákojzd bidleli a ta paň řidíci, jako žena toho	Novákoj/Novákoj	1
0614-C-ZAC-2005-N		hmm   +   hmm   +   a za prostředním oknem bila háká	Nezvalovo/Nezvalovo	1
0615-C-ZAC-2009-N		pak přišel nákej Novák, to bil kam*	tátovo/tátovo	1
0628-C-ZAC-2009-N		tudle bil, to bil nejvěčí dvír, tudle	Novákoj/Novákoj	1
0684-C-JIC-2017-N		tám bili   +   ten t*, otec jeho	mužovo/mužovo	1
0847-C-JIC-2008-N		říkal jejejejejeje, druhéj ruruururuuru, zezezezezezezez a vono to	slovo/slovo	1

Obrázek 59 – Výsledky vyhledávání ustrnulých přivlastňovacích tvarů na -ovo, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr\\_ymode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~Xuy2i8eSY4A2](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr_ymode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~Xuy2i8eSY4A2)

Abychom mohli snadno selektovat výsledky, vytvoříme tabulku frekvenční distribuce slovních tvarů (viz obrázek 60 níže). Z té můžeme vidět, že slovní tvar *slovo* se objevuje hned na prvním místě s 5 výskyty, tvar *kremovo* a tvar *hotovo* oba s 1 výskytem. Ostatní tvary v tabulce odpovídají ustrnulému přivlastňovacího tvaru. Pro zobrazení pouze těchto výsledků použijeme u nežádoucích tvarů negativní filtr.

Frekvenční seznam

Zobrazení v grafech    Tabulkové zobrazení

Zobrazit konfidenční intervaly  ⓘ

1 / 1 (celkem: 12 položek)

Filtr	word [dia]	Freq	i.p.m.
1	p / n slovo	5	66,04
2	p / n novákovovo	2	26,56
3	p / n taťinkovo	2	26,56
4	p / n taťinkovo	1	13,28
5	p / n kremovo	1	13,28
6	p / n mužovo	1	13,28
7	p / n jarovo	1	13,28
8	p / n nezvalovo	1	13,28
9	p / n ďedovo	1	13,28
10	p / n jirkovo	1	13,28
11	p / n tátovo	1	13,28
12	p / n hotovo	1	13,28

Obrázek 60 – Frekvenční distribuce slovních tvarů na -ovo, zdroj: [https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr\\_ymode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~Xuy2i8eSY4A2&fcrit=word%2Fie%200~0%3E0&flimit=1&freq\\_](https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr_ymode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~Xuy2i8eSY4A2&fcrit=word%2Fie%200~0%3E0&flimit=1&freq_)



Pokud z výskytů vyjmeme tvar *slovo*, *kremovo* a *hotovo*, dostaneme 11 slovních tvarů na *-ovo*, které odpovídají ustrnulému přivlastňovacímu tvaru. Z nich je nejvíce, tedy 2krát, zastoupen slovní tvar *Novákovo* a *taťinkovo*, pak po jednom výskytu tvary *tátovo*, *taťinkovo*, *mužovo*, *Jarovo*, *Nezvalovo*, *dědovo* a *Jirkovo*.

Co se počtu výskytů týče, tak jich sice není tolik, nicméně důležité je, že nářeční jev je v korpusu vůbec doložen. Pokud se podíváme na konkordanční řádky po vyjmutí tvarů, které nesouhlasí s ustrnulým tvarem na *-ovo*, v levém sloupci můžeme vidět, že všechny výskyty pocházejí z nové vrstvy nahrávek, což ale neznamená, že by se v minulosti tento nářeční jev nevyskytoval<sup>148</sup>, spíše se jedná o nedostatek jazykového materiálu. Výskyt jevu v nové vrstvě nahrávek však ukazuje, že se jedná o stále živý nářeční rys.

Závěrem můžeme ustrnulý tvar přivlastňovacích tvarů na *-ovo* zhodnotit jako charakteristický pro západočeskou oblast a zejména, což je důležité, jev stále užívaný.

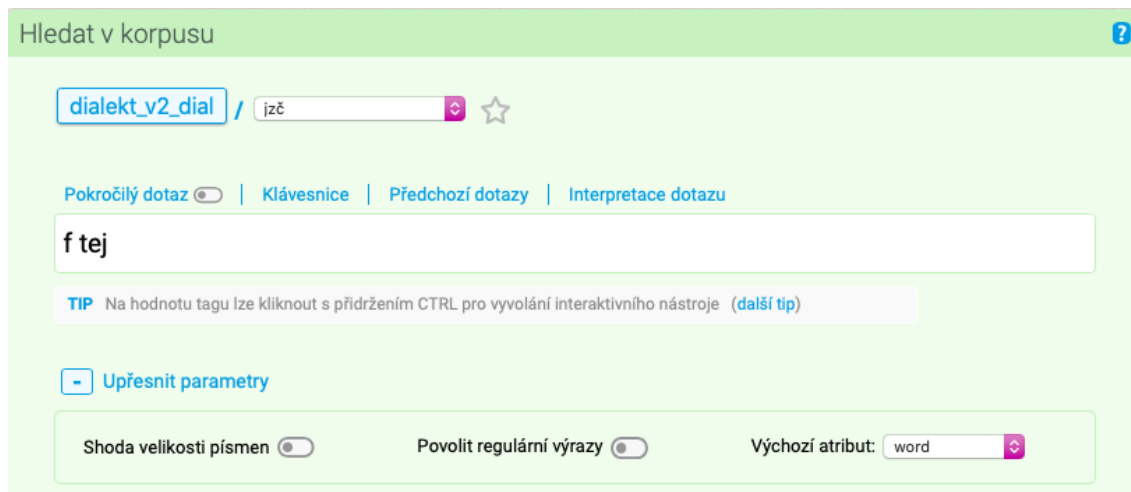
---

<sup>148</sup> Viz ČJA 4, položka 209, s. 312, mapa s. 315.

### Cvičení č. 13 – Slovní spojení *f tej* a jeho ekvivalenty v západočeské a jihočeské nářeční oblasti

Zadání: Pro jihozápadočeskou nářeční skupinu je z hlediska tvarosloví typické užívání ve 3. a 6. pádu podoby *tej* s příslušnou předložkou. Pro naše cvičení si zvolíme jedno předložkové spojení, a to *f tej*. Toto spojení je ve vybraných nářečních oblastech důsledně odlišováno od pádu druhého (*do tí*). Vyhledejte toto slovní spojení v korpusu DIALEKT a zhodnoťte počet výskytů. Najděte i spisovný a další hovorový ekvivalent tohoto spojení a zjistěte, která varianta převažuje. Je to jev užívanější v západočeské, nebo jihočeské skupině?

Postup: Slovní spojení *f tej* budeme vyhledávat na rovině dialektologické, na rovině ortografické bude spojení přepsáno jako *v tej*. Jako korpus tedy zvolíme *dialekt – v2 dial*. Vyhledávání omezíme na západočeskou a jihočeskou nářeční oblast (viz obrázek 17), nebo použijeme subkorpus *jzč* (viz obrázek 19). Regulární výrazy nebudeme při vyhledávání potřebovat, jako výchozí atribut zvolíme *word*. Podoba dotazu viz níže na obrázku 61.



Hledat v korpusu

dialekt\_v2\_dial / jzč

Pokročilý dotaz | Klávesnice | Předchozí dotazy | Interpretace dotazu

f tej

TIP Na hodnotu tagu lze kliknout s přidržením CTRL pro vyvolání interaktivního nástroje (další tip)

[-] Upřesnit parametry

Shoda velikosti písmen Povolit regulární výrazy Výchozí atribut: word

Obrázek 61 – Podoba dotazu k vyhledání slovního spojení *f tej*, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial&usesubcorp=jzč](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial&usesubcorp=jzč)

Pro takto položený dotaz vyhledal korpus 29 výskytů *f tej*, a to například ve spojení *f tej d'íži*, *f tej d'ílňe*, *f tej ulici*, *f tej dobje*, *f tej chalupje*, *f tej vesňici* nebo *f tej Lahuti*, viz obrázek 62 níže.

Výskytů: 29 | l.p.m.: 385,12 (vztaženo k subkorpusu "JZ") | ARF: - | Výsledek je seříděn 1 / 1

Vyber řádků: základní

<input type="checkbox"/>	0031-C-JIC-1964-SA	míčačka, vite, velká, a s tou míčačkou	/v tej/tej	džji se to míšelo, aš to huld bilo pjejni
<input type="checkbox"/>	0279-C-ZAC-2011-N	(nadechnuti)   +   a každá, fšechni ti dfeť, co bili	/v tej/tej	první tříde a druhé tříde, tak na ři fšpominaj
<input type="checkbox"/>	0279-C-ZAC-2011-N	ročníkí, první, druhé a třetí třídí, a	/v tej/tej	druhé tříde bila štvrtá, plátá.   +   (rušivý zvuk)   +   (nadechnuti)   +   (nadechnuti)
<input type="checkbox"/>	0279-C-ZAC-2011-N	, pomáhát, j, tag díž uš sme bili	/v tej/tej	štvrté, páté tříde, tak sme museli taki řeco
<input type="checkbox"/>	0300-C-ZAC-2014-N	vzal a ešfe tam mám spáteční lístek f tom,	/v tej/tej	.*   +   (nadechnuti)   +   (nadechnuti)   +   hmm.   +   taksikař to bal?
<input type="checkbox"/>	0301-C-ZAC-2014-N	dóle, jo.   +   (nadechnuti)   +   já řval jako bejk,	/v tej/tej	dřilše a přišla, přišla s ošefovni sestra a zustala
<input type="checkbox"/>	0599-C-ZAC-2005-N	u* mam pitomou hlavu.   +   no a tám sme bili	/v tej/tej	Starej Rolli rok.   +   a pak sme se sřehovali do
<input type="checkbox"/>	0591-C-ZAC-2005-N	tak posledňi staveňi Tímákowa.   +   hmm.   +   jo.   +   (mlasknuti)   +	/v tej/tej	ulici.   +   (nadechnuti)   +   no.   +   no a...   +   (mlasknuti)   +   potom
<input type="checkbox"/>	0614-C-ZAC-2005-N	(bouchnuti)   +   hmm.   +   no a šli sme fšechni ři.   +	/v tej/tej	, to bilo víření bjehem...   +   (nadechnuti)   +   (bouchnuti)   +   (mlasknuti)   +   sme
<input type="checkbox"/>	0614-C-ZAC-2005-N	@ to bilo v dobře jednadvacátiho, gdi sme aři	/v tej/tej	škole nebili.   +   hmm.   +   hmm.   +   (rušivý zvuk)   +   hmm.
<input type="checkbox"/>	0618-C-ZAC-2009-N	hmm.   +   tak tám bilo, spoustu out.   +   a	/v tej/tej	velkej seřfci biđili, biđili řáki ti veřtele.   +   a
<input type="checkbox"/>	0628-C-ZAC-2009-N	nemížu řic.   +   hmm.   +   hmm.   +   ale přehazovat síámu	/v tej/tej	perle.   +   nu, možná že vo tom ste síšřela
<input type="checkbox"/>	0666-C-ZAC-2009-N	dobje?   +   prosim?   +   bi*, bil řelocvig ve školé	/v tej/tej	dobje, nebo...   +   j, bil, j,
<input type="checkbox"/>	0741-C-ZAC-2008-N	, víš, no.   +   to jo.   +   hmm.   +	/v tej/tej	dobje prošfe k tomu... a ti kluci, to

Obrázek 62 – Výsledky vyhledávání slovního spojení *f tej*, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr\\_vmode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~kKImYyOg8UgG](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~kKImYyOg8UgG)

Pokud se podíváme na výsledky vyhledávání z hlediska nářečních skupin a vrstev nahrávek, nejjednodušším způsobem je vytvořit tabulku frekvenční distribuce. Na horní liště vybereme záložku *Frekvence*, následně *Vlastní...*, poté *Podle typů textů*. Jako atributy zvolíme *doc.noblast* a *doc.rok*, pro přehlednost vybereme tabulkové zobrazení. Frekvenční tabulky viz obrázek 63 níže.

Frekvenční seznam

Zobrazení v grafech Tabulkové zobrazení

Zobrazit konfidenční intervaly

1 / 1 (celkem: 2 položky)

Filter	doc.noblast	Freq	l.p.m.
1	p / n jihočeská	15	486,29
2	p / n západočeská	14	314,92

1 / 1 (celkem: 7 položek)

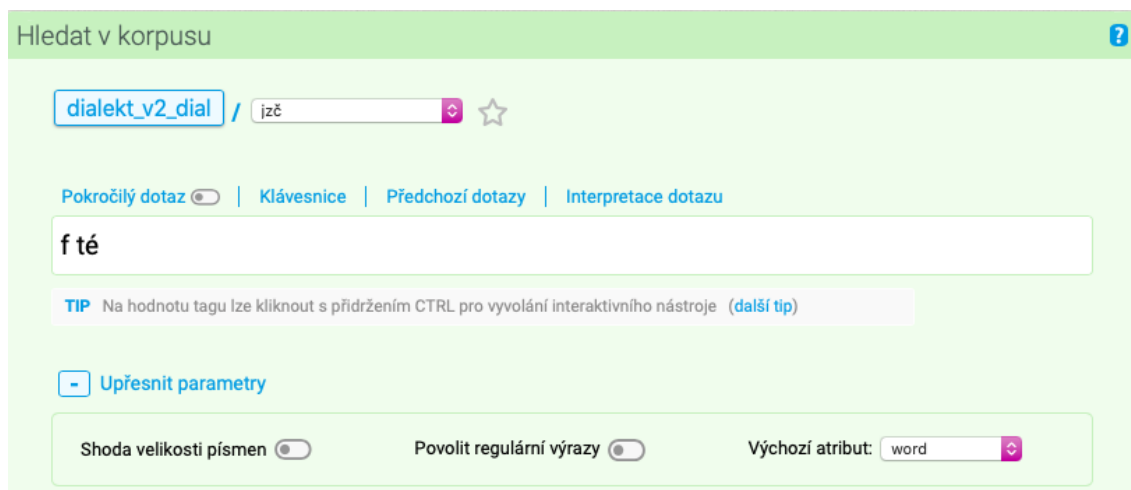
Filter	doc.rok	Freq	l.p.m.
1	p / n 2009	11	570,89
2	p / n 2008	4	450,05
3	p / n 2011	4	422,12
4	p / n 2005	4	770,42
5	p / n 2017	3	408,27
6	p / n 2014	2	356,07
7	p / n 1964	1	711,74

Obrázek 63 – Frekvenční distribuce slovního spojení *f tej* podle nářeční oblasti a roku pořízení nahrávky, zdroj: [https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr\\_vmode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~kKImYyO](https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~kKImYyO)

Z první tabulky frekvenční distribuce podle nářeční oblasti můžeme vidět, že slovní spojení *f tej* má v rámci jihozápadočeské oblasti rovnoměrné zastoupení, a to 15 výskytů v jihočeské oblasti, 14 v západočeské. Druhá frekvenční tabulka podle roku pořízení nahrávky už ale rovnoměrné výsledky nenabízí – ze staré vrstvy nahrávek pochází pouze

1 výskyt, a to z důvodu nedostatku materiálu. Důležitým výsledkem je však zastoupení v nové vrstvě, což ukazuje, že se jedná o nářeční jev v současnosti živý.

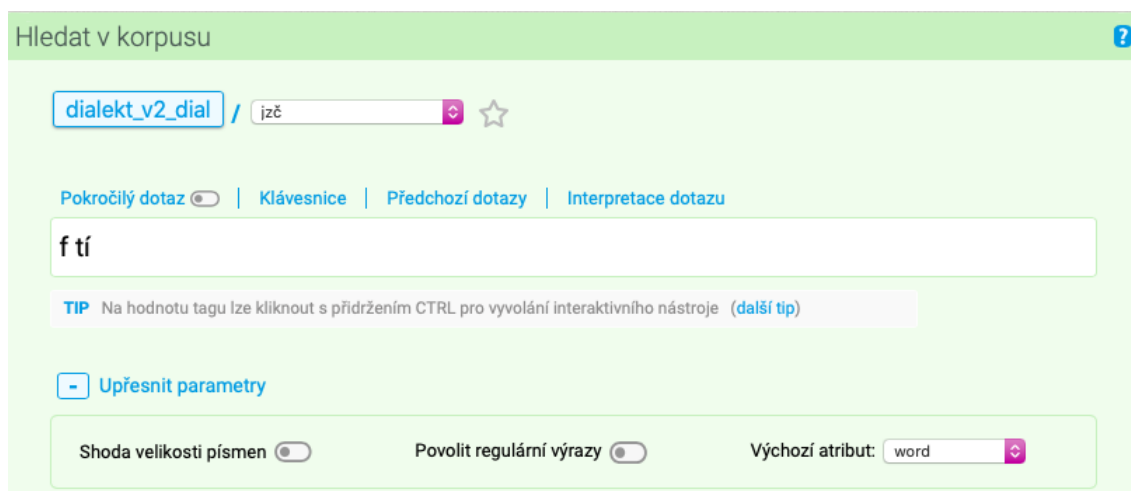
Pokud chceme vyhledat spisovný ekvivalent k tomuto slovnímu spojení, budeme vyhledávat *f té* namísto *f tej*, položíme tedy nový dotaz. Ponecháme omezení hledání nebo subkorpus a taktéž výchozí atribut. Podoba nového dotazu viz níže na obrázku 64.



The screenshot shows a search interface titled "Hledat v korpusu". At the top, there is a search bar with "dialekt\_v2\_dial" and "jzč" selected. Below the search bar, there are links for "Pokročilý dotaz", "Klávesnice", "Předchozí dotazy", and "Interpretace dotazu". The main search input field contains the text "f té". Below the input field, there is a tip: "TIP Na hodnotu tagu lze kliknout s přidržením CTRL pro vyvolání interaktivního nástroje (další tip)". There is also a section for refining parameters, "Upřesnit parametry", which includes three toggle switches: "Shoda velikosti písmen", "Povolit regulární výrazy", and "Výchozí atribut" (set to "word").

Obrázek 64 – Podoba dotazu k vyhledání slovního spojení *f té*, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial&usesubcorp=jzč](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial&usesubcorp=jzč)

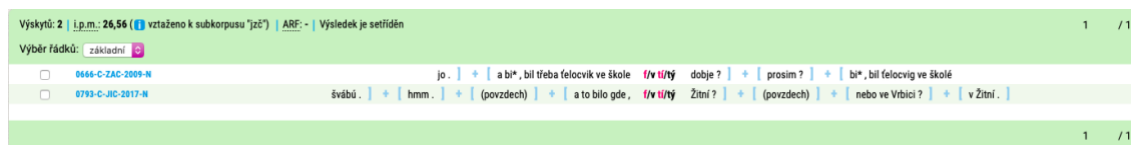
Po vyhledání dotazu zjistíme, že v západočeské a jihočeské oblasti není v korpusu slovní spojení *f té*, tedy spisovná varianta, doloženo ani jednou. Dále vyhledáme další nespisovnou variantu slovního spojení, a sice *f tí*, podoba dotazu viz obrázek 65 níže.



The screenshot shows the same search interface as in image 64, but with the search input field containing the text "f tí". All other elements, including the search bar, navigation links, tip, and parameter settings, remain the same.

Obrázek 65 – Podoba dotazu k vyhledání slovního spojení *f tí*, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial&usesubcorp=jzč](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial&usesubcorp=jzč)

Pro takto položený dotaz vyhledal korpus 2 výskyty *f tí*, konkrétně spojení *f tí dobře* a *f tí Žitní* (viz obrázek 66 níže). První výskyt pochází ze staré vrstvy nahrávek a ze západočeské oblasti, druhý z nové vrstvy a oblasti jihočeské.



Obrázek 66 – Výsledky vyhledávání slovního spojení *f tí*, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr\\_ymode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~0cwMCEygUesq](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr_ymode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~0cwMCEygUesq)

Z výsledků výše položených dotazů vidíme, že nářeční varianta slovního spojení *f tej* nemá v korpusu doloženou spisovnou variantu, pouze 2krát nespisovnou variantu *f tí*. To ukazuje, že je to rys výrazně převažující nad spisovnou podobou, jež není v korpusu zachycena. Rozložení je v rámci oblastí rovnoměrné<sup>149</sup>, výsledky ukazují, že se jedná ojev dodnes velmi živý a frekventovaný.

<sup>149</sup> V korpusu DIALEKT jsou zachyceny i případy dalších předložkových spojení tohoto typu, jejichž rozložení v rámci západočeské a jihočeské nářeční oblasti je taktéž poměrně rovnoměrné, stejně jako u předložkového spojení *f tej*.

## Cvičení č. 14 – Koncovka *-om* ve 3. pádu plurálu substantiv všech rodů v západočeské nářeční oblasti

Zadání: *Typickým tvaroslovným rysem pro západočeskou oblast je koncovka -om ve 3. pádu plurálu substantiv všech rodů. Vyhledejte v korpusu DIALEKT počet výskytů koncovky v západočeské oblasti a zjistěte, jaký je její poměr oproti spisovným koncovkám ve všech rodech. Dá se s ohledem na výsledky považovat tato koncovka za živý rys u nejstarší generace mluvčích? Je koncovka užívána více než -í v 1. pádu plurálu maskulin?*

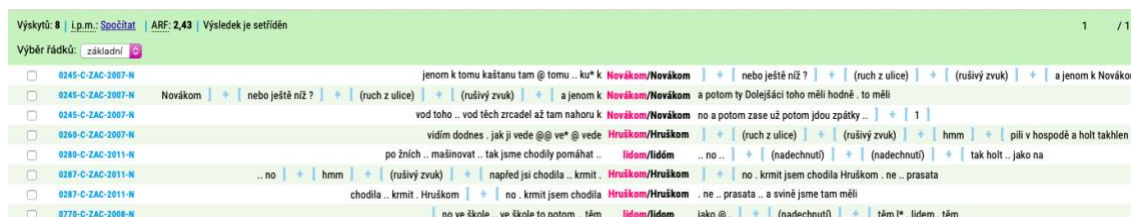
Postup: Vyhledávání bude probíhat na rovině ortografické, neboť koncovky ohebných slovních druhů se na této rovině zaznamenávají ve shodě s rovinou dialektologickou, korpus tedy vybereme *dialekt – v2 ort*. Hledání je opět nutné buď omezit na západočeskou nářeční oblast (viz obrázek 17), nebo zvolit subkorpus *zč* (viz obrázek 19), úsek dále nspecifikujeme, aby byly obsaženy výsledky z celé oblasti. V tomto cvičení budou ukázány dva postupy vyhledávání, z toho jeden způsob bohužel nepřináší přesné výsledky, a to ten pomocí pokročilého vyhledávání a jazyka CQL. Pro správný výsledek budeme muset použít vyhledávání pomocí regulárních výrazů a ke konečnému výsledku dospět využitím negativních filtrů.

Pokročilý dotaz sestavíme pomocí jazyka CQL tak, aby vyhledal substantiva s koncovkou *-om* ve 3. pádu plurálu bez ohledu na rod. Zaprvé bude nutné použít tzv. *tag* a specifikovat morfológické charakteristiky hledaných slov. Dále do stejného dotazu zařadíme i *word*, pomocí něhož zajistíme, že korpus vyhledá tvary pouze s danou koncovkou. Značkám v dotazu zadaném pomocí pokročilého jazyka se věnuje v tomto oddílu cvičení 9, viz s. 94. Podoba dotazu viz níže na obrázku 67.



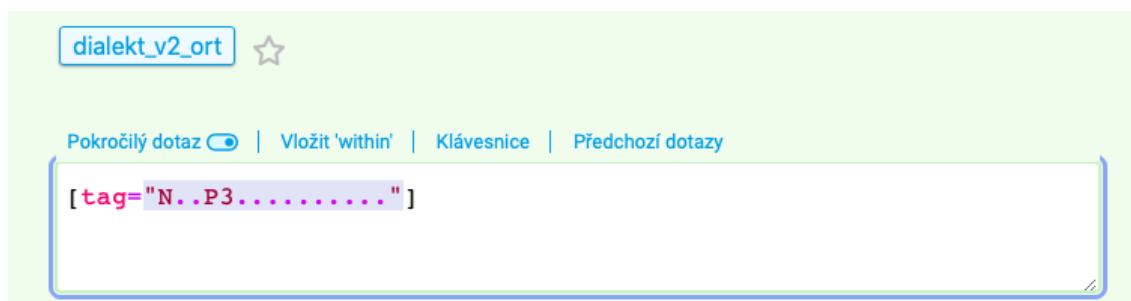
Obrázek 67 – Podoba pokročilého dotazu k vyhledání koncovky *-om* ve 3. p. pl. všech rodů, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_ort](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_ort)

Po zadání dotazu korpus vyhledal pouze 8 výskytů koncovky *-om* ve 3. pádu plurálu bez ohledu na rod substantiva. Z toho 2krát tvar *lidom*, 3krát tvar *Hruškom* a 3krát tvar *Novákom*<sup>150</sup>, viz obrázek 68.



Obrázek 68 – Výsledky vyhledávání koncovky *-om* ve 3. p. pl. všech rodů pomocí pokročilého dotazu, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attr\\_vmode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~bU2mG24e4YEw](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~bU2mG24e4YEw)

Zobrazené výsledky lze dále srovnat s ostatními koncovkami ve 3. pádu plurálu, navíc není nutné k tomuto kroku zadávat nový pokročilý dotaz, ale jednoduše smazat z původního dotazu tu část, ve které omezujeme hledání na slovní tvary končící na *-om* (stačí tedy smazat část `& word=".*om"` a ponechat jen tag, za kterým je potřeba uzavřít hranatou závorkou, viz obrázek 69).



Obrázek 69 – Podoba pokročilého dotazu po úpravě, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_ort](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_ort)

Pomocí takto zadaného dotazu vyhledá korpus 24 výskytů substantiv ve 3. pádu množného čísla bez dalšího bližšího určení, v nichž jsou zahrnuty všechny koncovky. V tomto případě ale nelze z důvodu chybného tagování výsledky srovnávat a zacházet s nimi jako s finálními, neboť jsou nepřesné.

<sup>150</sup> Tvary příjmení *Hruška* a *Novák* jsou zástupná.

Výskytů: 24	lp.m.: Spočítat	ARF: 6,37	Výsledek je setříděn	1 / 1
0018-C-ZAC-1972-SA	tij pískač... a houdek ...	no tak přijdú k	vrátím/vrátím	.. tak je zavěno ...
0238-C-ZAC-2007-N	no to jsme dělaly pořád ... a to se tém	Němům/Němům	tuze líbilo ...	(nadechnuti)   (nadechnuti)   a ty přijeli ... každé
0245-C-ZAC-2007-N	jenom k tomu kaštanu tam @ tomu ... ku* k	Novákem/Novákem	nebo ještě níž ?	(ruch z ulice)   (rušivý zvuk)   a jenom k Nc
0245-C-ZAC-2007-N	Novákem   nebo ještě níž ?	(rušivý zvuk)   a jenom k	Novákem/Novákem	a potom ty Dolejšíci toho měli hodné ... to měli
0245-C-ZAC-2007-N	vod toho ... vod těch zrcadel až tam nahoru k	Novákem/Novákem	no a potom zase už potom jdou zpátky ...	1
0249-C-ZAC-2007-N	je taky tuze dobrý	aha!   no   a jak říkáte	kramberákům/kramberákům	těba ?   hmm ... tady se říká vošouchy ...
0249-C-ZAC-2007-N	..   to se říkalo buchty ... no ...	no ...   a	buchtám/buchtám	se říkalo dílečky ...
0249-C-ZAC-2007-N	vidim dodnes ... jak jí vede @ ve* @ vede	Hruškům/Hruškům	(ruch z ulice)   (rušivý zvuk)   hmm   pili v hospodě a holt ta	
0249-C-ZAC-2007-N	(hluk v pozadí)   no   (štěkání psa)   to mě když tam poslali ... U	Hruškům/Hruškům	se u nich ř* -   (hluk v pozadí)   hmm ... U Hruškům	
0249-C-ZAC-2007-N	Hruškům se u nich ř* -   (hluk v pozadí)   hmm ... U	Hruškům/Hruškům	jo   (štěkání psa)   já říkám   v jaký boudě bydlí ! ...	
0280-C-ZAC-2011-N	po žních ... mašínovat ... tak jsme chodily pomáhat ...	lidom/lidóm	.. no ...   (nadechnuti)   (nadechnuti)   tak holt ... jako na	
0280-C-ZAC-2011-N	.. no   hmm   (rušivý zvuk)   napřed jsi chodila ... krmít ...	Hruškům/Hruškům	no ... krmít jsem chodila Hruškům ... ne ... prasata	
0287-C-ZAC-2011-N	chodila ... krmít ... Hruškům   no ... krmít jsem chodila	Hruškům/Hruškům	ne ... prasata ... a svině jsme tam měli	
0300-C-ZAC-2014-N	hmm   tak jste to potom nahlásil na policajtech ? ...	polcájkům/polcájkům	že jste 2 ukradli 2   jo ... člověče ...	
0424-C-ZAC-2005-N	posát nějaký ... tady zvláštní slova že třeba @	hřibkům/hřibkům	se říkalo hřibky ...   a ještě ještě něco ... že	
0426-C-ZAC-2009-N	hmm   a povídá ... hele ... jezdi prej lidi ke	kasárnám/kasárnám	hmm   jo ...   copak kdybysme se tam taky jeli podívat	

Obrázek 70 – Výsledky vyhledávání 3. p. pl. všech substantiv, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attr\\_vm=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~fnyuuMKUAW4Q](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attr_vm=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~fnyuuMKUAW4Q)

Díky výše uvedenému způsobu vyhledávání jsme narazili na slabou stránku nářečního korpusu, kterou je tagování. Přestože vyhledávání pomocí pokročilého dotazu se zdá jako důkladnější a ukazující přesné výsledky bez nutnosti další filtrace, není tomu vždy tak. Nářeční korpus bohužel není ve všech případech správně otagovaný, jak již bylo zmíněno v podkapitole 2.1.2 na straně 26, materiál je variabilní a samotná lemmatizace a morfologické značkování bylo komplikované, s čímž je nutné při vyhledávání a interpretaci dotazu počítat. Proto je nutné výsledky dosažené pomocí CQL pokaždé ověřovat, např. dotazem s využitím regulárních výrazů.

Už v případě vyhledávání koncovky *-om* ve 3. pádu plurálu všech rodů tak, jak bylo uvedeno na začátku tohoto cvičení, se objevuje první chyba – pokud pomocí stejného dotazu vyhledáme substantiva s koncovkou bez ohledu na pád, korpus ukáže o jeden výskyt víc. Konkrétně je to slovní tvar *lidóm* na rovině dialektologické, jehož protějšek na ortografické rovině je tvar *lidom*, ten je také ve 3. pádu plurálu, ale v původním vyhledávání nebyl nalezen.

Pokud vyhledáme v západočeské oblasti na ortografické rovině všechna slova končící na *-om* bez pokročilého dotazu, ale pouze pomocí regulárních výrazů (dotaz bude mít podobu *.\*om*), vyhledá korpus 370 výskytů. Hned na první pohled je jasné, že nejsou do vyhledávání zahrnuta pouze substantiva s touto koncovkou, neboť mezi prvními výskyty se objevuje několikrát slovní tvar *potom*. Tady přichází na řadu filtrace výskytů, nejjednodušším způsobem je zobrazení frekvence slovních tvarů pomocí tabulkového zobrazení. Zde kromě tvarů *Hruškům*, *Novákem* a *lidom*, které vyšly v původním



vyhledávání, vidíme 3x tvar *dětom*, po 1 tvar *papírom*, *muzikantom*, *pánom*, *Američanom*, *Němcom*, *Amerikánom*, *zrcadlom*, viz obrázek 71 níže.

Zobrazení v grafech    Tabulkové zobrazení

Zobrazit konfidenční intervaly  ⓘ

1 / 1 (celkem: 22 položek)

Filtr	word [ort]	Freq	i.p.m.
1	p / n potom	218	730,22
2	p / n tom	65	217,73
3	p / n jenom	60	200,98
4	p / n hruškom	3	10,05
5	p / n novákom	3	10,05
6	p / n dětom	3	10,05
7	p / n lidom	2	6,07
8	p / n ňom	2	6,07
9	p / n přitom	1	3,35
10	p / n jednom	1	3,35
11	p / n papírom	1	3,35
12	p / n teďkom	1	3,35
13	p / n muzikantom	1	3,35
14	p / n ičom	1	3,35
15	p / n strom	1	3,35
16	p / n pánom	1	3,35
17	p / n američanom	1	3,35
18	p / n němcom	1	3,35
19	p / n někom	1	3,35
20	p / n amerikánom	1	3,35
21	p / n vtom	1	3,35
22	p / n zrcadlom	1	3,35

Obrázek 71 – Tabulkové zobrazení frekvence slovních tvarů s koncovkou *-om* v západočeské oblasti, zdroj: [https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attrs=dial&attr\\_vmode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~JSsguYs4IiKi&fc](https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attrs=dial&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~JSsguYs4IiKi&fc)

Teprve na tomto místě lze o datech s jistotou dělat závěry, neboť je nutné pracovat se správnými výsledky. Počet výskytů pro koncovku *-om* ve 3. pádu plurálu u substantiv všech rodů je tedy vyšší. Konkrétně 18 substantiv bylo vyhledáno s touto koncovkou, to je o 10 více, než ukázalo původní hledání pomocí pokročilého dotazu, tento nářeční jev je tedy ještě častější, než se na první pohled zdálo. Zde je jasně vidět, že na morfologické značkování nelze 100% spoléhat a je nutné výsledky ověřovat jiným způsobem.

Nyní je možné vyhledávat výskyt se spisovnou koncovkou, tedy *-ům*. V původním dotazu změním pouze koncovku z *-om* na *-ům*, dotaz tedy bude mít podobu *.\*ům*, omezení na západočeskou oblast zůstane, výchozím atributem zůstane *word*. Pro takto

položený dotaz vyhledá korpus 10 výskytů (*Hruškům, Němcům* – opět zástupná příjmení; *bramborákům, vratům, policajtům, hříbkům, Američanům, sedlákům*), jeden výskyt však neodpovídá vyhledávanému nářečnímu jevu (*dům*). Z těchto výsledků je vidět, že nářeční koncovka převyšuje nad koncovkou spisovnou (18 ku 9). Počet výskytů hledané koncovky nedosahuje takového počtu jako koncovka *-í* v 1. pádu plurálu životných maskulin (viz cvičení 9, s. 94), kterých bylo 101.

Závěrem lze z výsledků soudit, že i přes nízký počet výskytů oproti koncovce *-í* lze koncovku považovat za jev živý. Koncovka je typická pro západočeskou oblast, nicméně v malé míře se vyskytuje i v oblasti jihočeské, existují pro to doklady v korpusu (na Doudlebsku), *Mapce* (na ukázkové kartě je příklad koncovky v nahrávce z Doudleb) i v ČJA<sup>151</sup>.

---

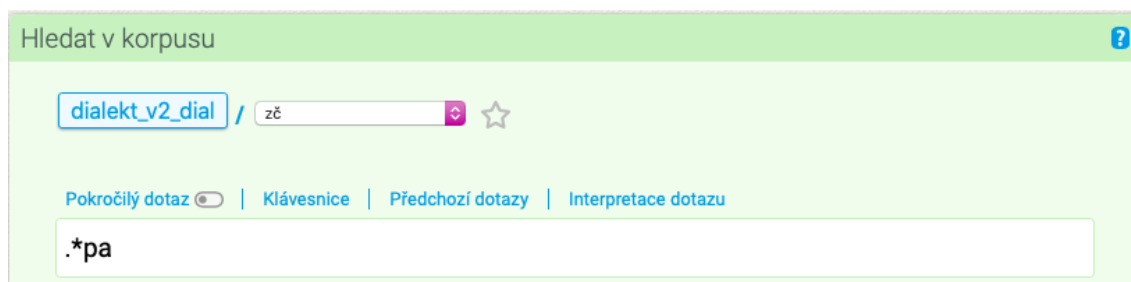
<sup>151</sup> Viz ČJA 4, kapitola Dativ plurálu, položky 150–164, s. 226–246.

## Cvičení č. 15 – Odsouvání koncového *-k* u slootovorného formantu *-pak* v západočeské nářeční oblasti

Zadání: Odsouvání koncového *-k* u slootovorného formantu *-pak* je tvaroslovný rys typický pro západočeskou nářeční oblast<sup>152</sup>. Vyhledejte tento nářeční jev v korpusu *DIALEKT* v dané oblasti a zjistěte, zda se jedná o jev užívaný i v současnosti. Srovnejte počet výskytů se spisovnou podobou.

Postup: Vyhledávání slootovorného formantu *-pak* bez koncového *-k* bude probíhat na rovině dialektologické, neboť na rovinu ortografickou je tento jev přepisován ve spisovné podobě, korpus vybereme tedy *dialekt – v2 dial*. Dále je nutné omezit vyhledávání na západočeskou nářeční oblast (viz obrázek 17), nebo zvolit subkorpus *zč* (viz obrázek 19). Nakonec je třeba povolit regulární výrazy, které budeme při vyhledávání potřebovat. Výchozím atributem pro vyhledávání bude *word*.

Podoba dotazu je vidět na obrázku 72 níže. Mezi použitými regulárními výrazy je tečka zastupující jeden libovolný znak, pak hvězdička, která zajistí libovolný počet opakování předchozího znaku. Takto položený dotaz zajistí, aby byla vyhledána v západočeské oblasti všechna slova o libovolné délce, která končí na *-pa*.



Obrázek 72 – Podoba dotazu k vyhledání formantu *-pak* bez koncového *-k*, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial&usesubcorp=zč](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial&usesubcorp=zč)

Po zadání dotazu korpus vyhledá 44 slovních tvarů, které končí na *-pa*. Hned prvním výskytem je ale slovní tvar *chalupa*, který neodpovídá hledanému jevu. Na tomto místě je potřeba filtrovat výsledky, a to buď manuálně, nebo pomocí frekvenční distribuce – na horní liště zvolíme záložku *Frekvence*, následně *Slovní tvary [A=a]* a *Tabulkové zobrazení*. Frekvenční seznam slovních tvarů je vidět níže na obrázku 73.

<sup>152</sup> V korpusu se nevyskytují doklady pro tento nářeční rys v jihočeské nářeční oblasti. Dále viz ČJA 5, položka 352, s. 482, mapa s. 483 (dokládá západočeskou nemapovanou obměnu *copa*).

Frekvenční seznam

Zobrazení v grafech      Tabulkové zobrazení

Zobrazit konfidenční intervaly  !

1 / 1 (celkem: 10 položek)

	Filter	word [dial]	Freq	i.p.m.
1	p / n	copa	19	427,39
2	p / n	jakpa	7	157,46
3	p / n	gdepa	4	89,98
4	p / n	depa	4	89,98
5	p / n	chalupa	4	89,98
6	p / n	kampa	2	44,99
7	p / n	čipa	1	22,49
8	p / n	jakípa	1	22,49
9	p / n	votkatpa	1	22,49
10	p / n	čemupa	1	22,49

Obrázek 73 – Tabulkové zobrazení frekvenční distribuce slovních tvarů končících na *-pa*, zdroj: [https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr\\_vmode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~TUcyGoe4gw08&fcrit=word%2F](https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~TUcyGoe4gw08&fcrit=word%2F)

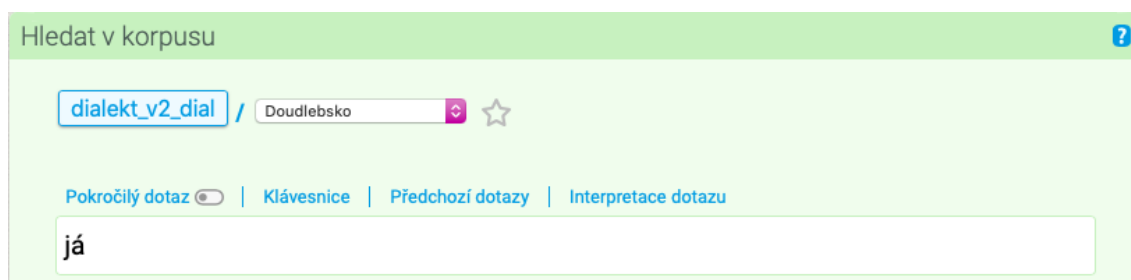
Z tabulkového zobrazení frekvenční distribuce je jasně vidět, že kromě 4 výskytů slovního tvaru *chalupa* odpovídají ostatní výskyty hledanému jevu, celkem tedy našel korpus slovtvorný formant *-pak* bez koncového *-k* 40krát, nejvíce ve tvaru *copa*, jenž je pro západočeskou oblast opravdu typický. Ze 40 výskytů jich je 37 doloženo v nové časové vrstvě nahrávek, to ukazuje, že je to jev stále živý, díky vysokému výskytu slova *copa* zřejmě dochází k jeho lexikalizaci.

Pokud vyhledáme formant *-pak* ve spisovné podobě, nalezneme korpus 41 výskytů, které všechny odpovídají hledanému jevu. Z nich je 35krát zastoupen slovní tvar *pak*, z dalších 2krát tvar *copak*, po 1 tvary *depak*, *jakpak*, *gdepak*, *dépak*. Z těchto výsledků je vidět, že tvary *copak*, *kdepak*, *jakpak* jsou užívány ve větší míře bez koncového *-k*. Jedná se tedy o jev dochovaný, zároveň v hojně míře užívaný i v současnosti.

## Cvičení č. 16 – Tvary zájmena *já* na Doudlebsku

Zadání: Pro doudlebský typ jihočeské nářeční oblasti jsou specifické některé tvary zájmena *já*, a to zejména tvary v jednotném čísle. Pokuste se vyhledat všechny tvary zájmena *já* v této oblasti pomocí jednoho dotazu a okomentujte výsledky s ohledem na frekvenci slovních tvarů. Které z nich se dají považovat za nářeční tvary?

Postup: Vyhledávání bude probíhat na dialektologické rovině, korpus zvolíme tedy *dialekt – v2 dial*, na ortografickou rovinu se totiž tvary zájmena *já* přepisují ve spisovné podobě. Dále je nutné buď omezit vyhledávání pouze na doudlebský typ jihočeské nářeční oblasti (viz obrázek 17), nebo pro vyhledávání zvolit subkorpus *Doudlebsko* (viz obrázek 19). Regulární výrazy není nutné povolovat, při vyhledávání je nebudeme potřebovat. Výchozím atributem bude *lemma* a do vyhledávacího okna zapíšeme zájmeno *já* (viz obrázek 74 níže). Tímto zajistíme, že korpus vyhledá všechny slovní tvary lemmatu *já*, které jsou v nahrávkách na Doudlebsku zachyceny.



Obrázek 74 – Podoba dotazu pro vyhledání slovních tvarů lemmatu *já*, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial&usesubcorp=Doudlebsko](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial&usesubcorp=Doudlebsko)

Korpus pomocí takto položeného dotazu vyhledá na Doudlebsku 226 slovních tvarů lemmatu *já*. Hned na první pohled je vidět, že nejčastěji bude zastoupen samotný tvar *já*. Abychom nemuseli výsledky ručně procházet a filtrovat, použijeme již výše zmíněný postup, a sice tabulkové zobrazení frekvence slovních tvarů (viz obrázek 75).

Frekvenční seznam

Zobrazení v grafech    **Tabulkové zobrazení**

Zobrazit konfidenční intervaly  ⓘ

1 / 1 (celkem: 8 položek)

Filtr	word [dial]	Freq	i.p.m.
1	p / n já	134	6 238,36
2	p / n mi	48	2 234,64
3	p / n mje	16	744,88
4	p / n mñe	9	418,99
5	p / n me	7	325,88
6	p / n ja	6	279,33
7	p / n mnou	5	232,77
8	p / n á	1	46,55

Obrázek 75 – Tabulkové zobrazení slovních tvarů lemmatu *já*, zdroj: [https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr\\_ymode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~caQAsyyUCwwS&fcrit=word%2Fie%200~0%3E0&flimit=1&freq\\_](https://www.korpus.cz/kontext/freqs?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr_ymode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~caQAsyyUCwwS&fcrit=word%2Fie%200~0%3E0&flimit=1&freq_)

Z frekvenční distribuce zjistíme, že z celkového počtu výskytů tvoří více než polovinu slovní tvar *já* (134 výskytů), 6krát krátký tvar *ja* a 1 tvar *á* bez počátečního *j*-. Se spisovnou výslovností na rovině dialektologické jsou dále zaznamenány tvary *mi*, *mñe* a *mnou*. Z nářečního hlediska je zajímavý tvar *mje* pro 3. a 6. pád, tvar *me* pro 4. pád a tvar *mi* pro 3. pád, který je obecně častější v mluveném jazyce, je to ale rys typický pro západočeskou i jihočeskou nářeční oblast (pro srovnání viz korpusový nástroj SyD<sup>153</sup>).

Ve slovním tvaru *mje* můžeme vidět rozloženou výslovnost retnice *m*, která je vyslovována s vkladným *-j-*. Tento případ byl z hláskoslovného hlediska vyhledáván a komentován v předchozím oddílu, konkrétně ve cvičení 8 na straně 91. Pokud si zobrazíme pozitivní filtr daného tvaru (pomocí písmene *p*), zobrazí se nám výskyty tohoto tvaru jako KWIC v konkordančních řádcích. Pokud projdeme výsledky, zjistíme, že 4 ze 16 výskytů odpovídají spisovnému tvaru *mně*. Zbylých 12 tvarů odpovídá spisovnému tvaru *mě* (z toho 1 ve staré vrstvě nahrávek, 11 v nové), jemuž dále odpovídá na dialektologické rovině 7 výskytů slovního tvaru *me* (z toho 5krát ve staré vrstvě nahrávek, 2krát v nové).

<sup>153</sup> Dostupné z: <https://syd.korpus.cz/>.

Z výsledků můžeme vidět, že spisovný tvar *mě* je v korpusu v 19 případech z 24 (79,2 %) zachycen s nářeční výslovností, a to jako *mje* nebo *me*. Spisovný tvar *mně* je ve 4 případech z 8 (50 %) zachycen s nářeční výslovností *mje*. Závěrem výsledky ukazují zejména na zachycení nářečního tvaru *me* a nářeční výslovnosti *mje* na Doudlebsku, jsou to tvary z jazykového hlediska zajímavé a důležité.<sup>154</sup>

---

<sup>154</sup> Více k těmto nářečním tvarům viz ČJA 4, položka 228, s. 340, mapa s. 343; položka 229, s. 342, mapa s. 343; položka 230, s. 342, mapa s. 345.

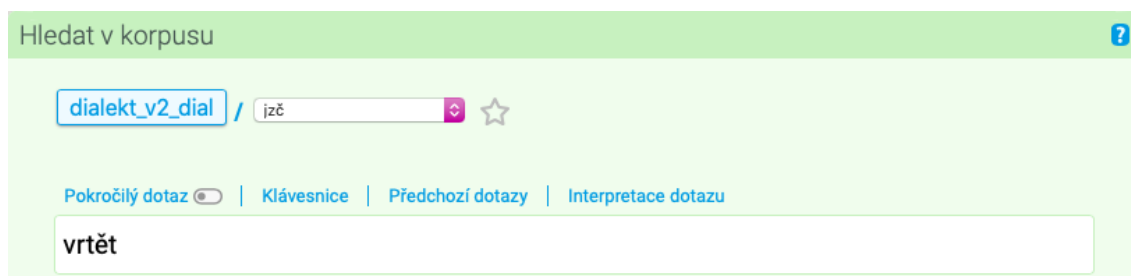
## 6.3 SLOVNÍ ZÁSoba

### 6.3.1 JÍDLO

#### Cvičení č. 17 – Sloveso *vrťet* v západočeské a jihočeské nářeční oblasti

Zadání: Sloveso *vrťet* má několik významů, tím základním je podle SSJČ, *krouživě n. sem tam pohybovat, kroutit*<sup>155</sup>. V západočeské a jihočeské nářeční oblasti je však toto sloveso užíváno v jiném významu. Vyhledejte slovní tvary lemmatu *vrťet* a z výsledků zjistěte další význam slovesa. Je slovo typičtější pro jihočeskou, nebo západočeskou nářeční oblast? Výsledky ověřte zpětně v SSJČ, v němž je zachycen i oblastní význam slovesa.

Postup: Chceme vyhledat všechny slovní tvary lemmatu *vrťet*, vyhledávat můžeme na rovině ortografické, v tom případě omezíme vyhledávání na západočeskou a jihočeskou nářeční oblast (viz obrázek 17). Pokud budeme vyhledávat na rovině dialektologické, můžeme zvolit vytvořený subkorpus *jzč* (viz obrázek 19). Ať už zvolíme jakýkoliv ze způsobů, vyhledávat budeme tvar *vrťet* a jako výchozí atribut zvolíme *lemma*. Podoba dotazu viz obrázek 76 níže.



Hledat v korpusu

dialekt\_v2\_dial / jzč

Pokročilý dotaz | Klávesnice | Předchozí dotazy | Interpretace dotazu

vrťet

Obrázek 76 – Podoba dotazu k vyhledání slovních tvarů lemmatu *vrťet*, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial&usesubcorp=jzč](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial&usesubcorp=jzč)

Korpus vyhledá v západočeské a jihočeské oblasti 6 slovních tvarů lemmatu *vrťet*, viz obrázek 77 níže. Nyní je nutné tyto výskyty projít a zjistit z nich další významy slova. Pro širší kontext je možné KWIC rozkliknout, pak se objeví více částí nahrávky, z nichž je možné vyčíst další informace. Nejlépe je význam slova patrný z druhého a pátého výskytu, z nichž se dozvídáme, že se vrťelo máslo, *vrťet* znamená tedy v tomto významu tlouct nebo stloukat máslo.

<sup>155</sup> Dostupné z: <https://ssjc.ujc.cas.cz/search.php?db=ssjc>.



Výskytů: 6   <a href="#">lp.m.: 79,68</a> (vztaženo k subkorpusu "jzč")   ARF: -   Výsledek je seříděn				1	/ 1
Vyběr řádků: základní					
<input type="checkbox"/>	0022-C-ZAC-1965-SA	u vobjeda, tak vířeli, že ta selka bude	vrřel/vrřít	.	tak si nalala do belku mlíko nebo smetanu a
<input type="checkbox"/>	0022-C-ZAC-1965-SA	drobet a nasipala to do belku	(nadechnutí)	nu a	vrřela/vrřela a za chvíli mňela svrceno a vřáhla přj mšla hrozou
<input type="checkbox"/>	0022-C-ZAC-1965-SA	přišel domů přj a říkal:	" mámo, budeme	vrřel/vrřít	." a diď já prej toho eše mam
<input type="checkbox"/>	0022-C-ZAC-1965-SA	nestojí."	"jen sem vem belk a pudem	vrřel/vrřít	." tak vona ho poslechla a vzala si tu
<input type="checkbox"/>	0022-C-ZAC-1965-SA	von vzal ten prášek a nasipal to tám a vona	vrřela/vrřela	(nadechnutí)	navřela takovou spoustu mšla, že to jagživi
<input type="checkbox"/>	0796-C-JIC-2017-N	mňeli krávu a ta nám dala mlíčko a tak sme	vrřeli/vrřeli	voď ří.	hmm a ták sme si pomastili

Obrázek 77 – Výsledky vyhledávání slovních tvarů lemmatu *vrřet* v jihozápadočeské nářeční oblasti, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr\\_ymode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~Hk8US8WsuKYs](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=ort&attr_ymode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~Hk8US8WsuKYs)

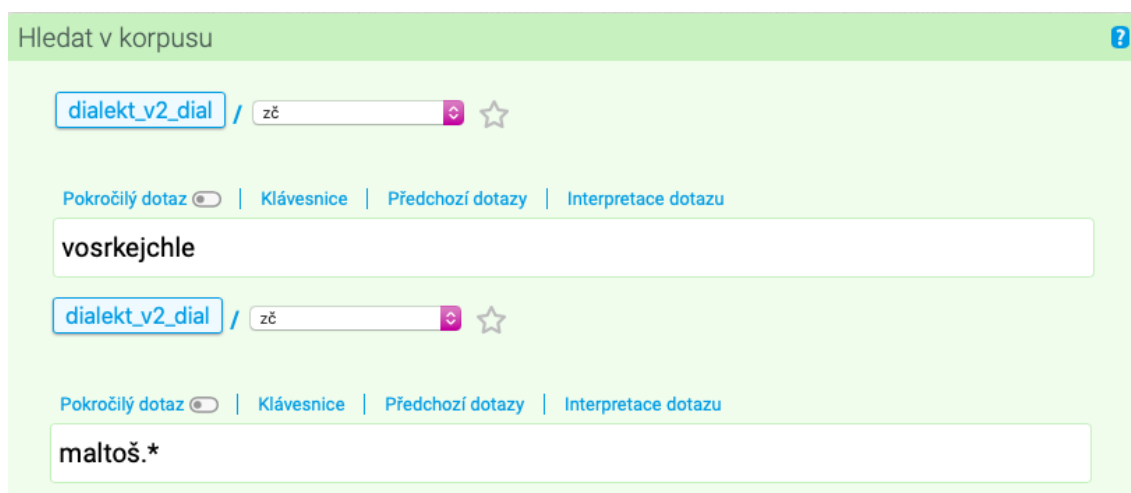
Z konkordančních řádků je vidět, že všechny slovní tvary slovesa *vrřet* byly zaznamenány s výše zmíněným nářečním významem. V ČJA 1 můžeme daný výraz nalézt pod položkou 113 *stloukat máslo* na s. 232. Z mapy k této položce na s. 233 je jasně vidět areál, v němž je sloveso *vrřet* v tomto významu užíváno – jihozápadočeský areál s přesahem do areálu západomoravského.

Pokud se ale podíváme na kódy nahrávek, zjistíme, že 5 výskytů ze 6 pochází z jedné nahrávky (0022-C-ZAC-1965-SA), a to ze západočeské nářeční oblasti a ze staré vrstvy nahrávek. Důležité je, že 1 výskyt slova v tomto významu také pochází z jihočeské nářeční oblasti a z nové vrstvy nářečního materiálu. Na tomto příkladu vidíme, že obecně známé výrazy mohou mít také oblastní, nebo úzce nářeční významy (někdy mohou mít etnografický charakter). Často jde o přenášení významu na základě podobnosti (zde na základě pohybu, neboť *vrřet* ve významu *kroutit a otáčet* napodobuje pohyb při *stloukání másla*). To potvrzuje i SSJČ, který vedle základního významu zmíněného v zadání uvádí i oblastní, v tomto případě zároveň nářeční význam *spojení vrřeti máslo – stloukat*.

## Cvičení č. 18 – Vosrkejchle a maltošna v západočeské nářeční oblasti

Zadání: V západočeské nářeční oblasti se užívá označení *vosrkejchle* a *maltošna* pro určitá jídla. Vyhledejte tyto jevy v korpusu *DIALEKT* a zjistěte jejich význam. Z jakých vrstev nahrávek tyto jevy pocházejí? Zjistěte, zda jsou zachyceny v SSJČ a v ČJA.

Postup: Nejprve budeme vyhledávat zadané slovní tvary v korpusu *DIALEKT*. Jednotlivé jevy vyhledáme na rovině dialektologické, jako korpus tedy zvolíme *dialekt – v2 dial* s omezením na západočeskou oblast (viz obrázek 17), nebo zvolíme subkorpus *zč* (viz obrázek 19). Výchozím atributem bude *word*, v případě druhého dotazu budeme potřebovat regulární výrazy. Podoba obou dotazů viz obrázek 78 níže.



Hledat v korpusu

dialekt\_v2\_dial / zč

Pokročilý dotaz | Klávesnice | Předchozí dotazy | Interpretace dotazu

vosrkejchle

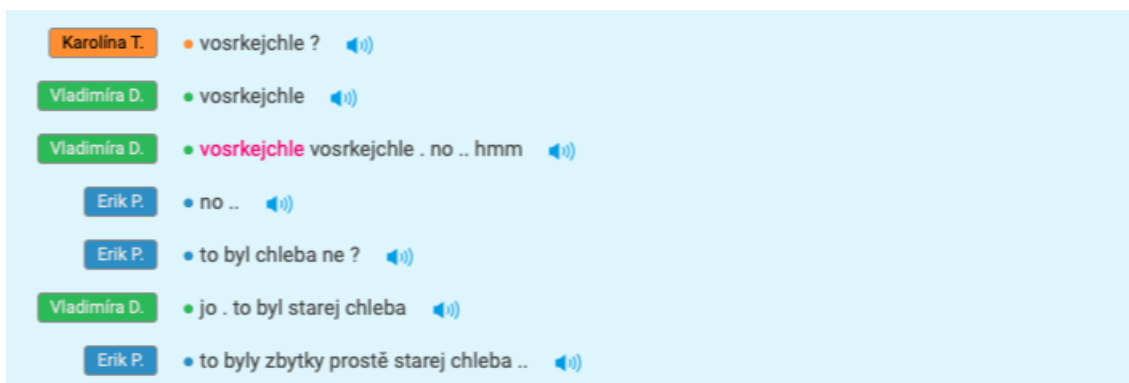
dialekt\_v2\_dial / zč

Pokročilý dotaz | Klávesnice | Předchozí dotazy | Interpretace dotazu

maltoš.\*

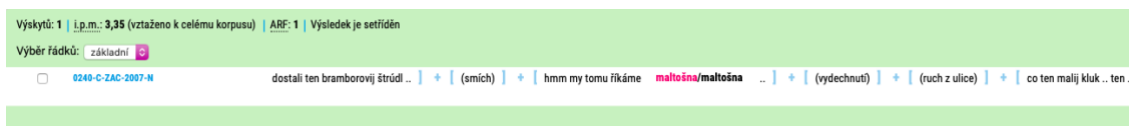
Obrázek 78 – Podoba dotazů k vyhledání slovních tvarů *vosrkejchle* a *maltošna*, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_ort](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_ort)

Pro slovní tvar *vosrkejchle* vyhledá korpus 4 výskyty, z nichž všechny pocházejí z nové vrstvy nahrávek, navíc ze stejné nahrávky (kód 0249-C-ZAC-2007-N). Z konkordančních řádků zjistíme, že tímto názvem se označoval chleba. Pokud si rozklikneme KWIC ve třetím výskytu, z promluv zjistíme, že to byl konkrétně chleba starý, viz obrázek 79 níže.



Obrázek 79 – Ukázka promluv s výskytem slovního tvaru *vosrkejchle*, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=dial&attr\\_ymode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~c4CmeOCGeycm](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=dial&attr_ymode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~c4CmeOCGeycm)

Po vyhledání druhého slovního tvaru, tedy *maltoš.\**, vyhledá korpus v západočeské oblasti pouze 2 výskyty, které taktéž pochází z 1 nahrávky z nové časové vrstvy (kód 0240-C-ZAC-2007-N, Domažlice), ve staré vrstvě výskyty chybí z důvodu absence materiálu. Přímou z konkordančního řádku lze vyčíst význam slova, a sice bramborový štrúdl (viz obrázek 80 níže). V nahrávce se přímo píše *bramborovij štrúdl* – to ukazuje na zavřenou výslovnost dvojhlásky *ej* (tomuto jevu se věnuje v oddílu hláskosloví cvičení č. 5 na s. 81).



Obrázek 80 – Výsledek vyhledávání slovního tvaru *maltošna*, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=dial&attr\\_ymode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~KsQuYgiaggqO](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=dial&attr_ymode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~KsQuYgiaggqO)

V dalším postupu budeme vyhledávat slova v ČJA. Pokud použijeme elektronickou verzi, už při zadávání dotazu nám atlas žádné jazykové doklady nenabízí, po zobrazení hesla taktéž žádné nevyhledá. Absenci výrazů ověříme v Rejstříku jazykových dokladů – zde zjistíme, že slova opravdu nejsou v atlasu zaznamenána. Při terénním výzkumu v 60. a 70. letech pro potřeby ČJA mohla být sice slova zachycena, nicméně nebylo možné zařadit všechny typy reálií.<sup>156</sup> Vybrané výrazy jsou chodské, tedy úzce lokální, to může být důvod, proč nejsou zaznamenány v SSJČ.

<sup>156</sup> Reálie vztahující se k chlebu viz ČJA 1, položky 132–136, s. 260–268.

## Cvičení č. 19 – Podstatné jméno *buchty* v jihočeské nářeční oblasti

Zadání: Slovo *buchty* se v jihočeské nářeční oblasti užívá i v jiném smyslu než sladké pečivo. Vyhledejte tento tvar v korpusu DIALEKT a zjistěte další význam tohoto slova. Uveďte místo a kód nahrávky, v níž byl tvar zaznamenán. Slovo je také zachyceno v ČJA. Uveďte číslo mapy a název položky v ČJA, pod kterou lze toto slovo v atlasu vyhledat.

Postup: Vyhledávat budeme v korpusu tvary slova *buchta* na rovině dialektologické, jako výchozí korpus zvolíme tedy *dialekt – v2 dial*. Je nutné buď omezit hledání na jihočeskou oblast (viz obrázek 17), nebo vyhledávat v subkorpusu *jč* (viz obrázek 19). Do vyhledávače zadáme tvar *buch.\**, atributem pro vyhledávání bude *word*. Výsledek vyhledávání viz obrázek 81 níže.

Výskytů: 4 | i.p.m.: 129,68 | vztaženo k subkorpusu "jč" | ARF: - | Výsledek je setříděn

Výběr řádků: základní

<input type="checkbox"/>	0791-C-JIC-2017-N	nó dobře, sebrali ste mñe cukr, tak nebudou buchti/buchty, že a nebude v ñem kv*, kinout kvasñice
<input type="checkbox"/>	0840-C-JIC-2008-N	(nadechnutí) + (nadechnutí) + no gde bili bohačì, tag nám dali buchtu/buchtu, vječinou mjeli pečenou. (hluk v pozadí)
<input type="checkbox"/>	0844-C-JIC-2008-N	maso, knedlíki, se nefíkali knedlíki, to bili buchti/buchty, a zelì. (nadechnutí) (mlasknutí)
<input type="checkbox"/>	0886-C-JIC-2011-N	starostì prose, abi pañimáma furd mñela s čeho d'ela buchti/buchty, no. (zvuk ptáka) (nadechnutí)

Obrázek 81 – Výsledky vyhledávání slovního tvaru *buchty* v jihočeské nářeční oblasti, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=word&attr\\_vmode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~VwKlEkuGMm80](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=word&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~VwKlEkuGMm80)

Výsledky vyhledávání je nutné projít a zjistit, zda se v konkordančním řádku dozvíme další významy slova. Ve 3. výskytu můžeme vidět, že knedlíky se nepojmenovávaly jako knedlíky, ale buchty. Slovo *buchty* se tedy na Doudlebsku jihočeské nářeční oblasti užívá i ve významu knedlíky. Na levé straně vidíme zároveň kód nahrávky, v níž byl slovní tvar zachycen – 0844-C-JIC-2008-N. Pokud ho rozklikneme, otevře se tabulka s několika sociologickými údaji o nahrávce samotné a o mluvčích (viz obrázek 82 níže). Z tabulky lze vyčíst, že nahrávka byla pořizena v doudlebském okrajovém typu, konkrétně přímo v Doudlebech.

doc.id	0844-C-JIC-2008-N	doc.wordcount	175
doc.zdroj	Soukromý sběr nahrávek	doc.rok	2008
doc.vrstva	nová	doc.misto	Doudleby
doc.sidlotyp	venkovská lokalita	doc.sidlovelikost	méně než 500
doc.stat	Česká republika	doc.region	Čechy
doc.noblast	jihočeská	doc.npodskupina	žádná
doc.nusek	jihočeský okrajový	doc.ntyp	doudlebský
doc.okres	České Budějovice	doc.promluvatyp	převážně monolog
doc.tema	Čím se hostili koledníci	doc.pocetml	2
doc.explorator	ano	sp.id	0844-C-JIC-2008:7
sp.prezdivka	Denisa Š.	sp.explorator	ne
sp.pohlavi	Z: žena	sp.rok	1911
sp.veknum	97	sp.vek	V: nad 35 let
sp.vzdelanotyp	B: základní či střední	sp.vzdelanotypII	ZŠ
sp.profese	prodavačka	sp.mistonej	Doudleby
sp.okresnej	České Budějovice	sp.statnej	Česká republika
sp.regionnej	Čechy	sp.noblastnej	jihočeská
sp.npodskupinanej	žádná	sp.nuseknej	jihočeský okrajový
sp.ntypnej	doudlebský	sp.mistodet	Doudleby
sp.noblastdet	jihočeská	sp.soundfile	6/d/b76661de.mp3
para.type			

Obrázek 82 – Tabulka se sociologickými údaji k nahrávce, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=dial&attr\\_vmode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=-VwKlekuGMm80](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=dial&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=-VwKlekuGMm80)

Nyní je nutné vyhledat daný tvar v ČJA – do vyhledávacího okna napíšeme slovo *buchta*, viz obrázek 83 níže. Vyhledávač rovnou nabízí zachycené tvary, tudíž rovnou můžeme zjistit, zda je do ČJA slovo zařazeno.

# ČESKÝ JAZYKOVÝ ATLAS

Obrázek 83 – Vyhledávání jazykového dokladu *buchta* v ČJA, zdroj: <https://cja.ujc.cas.cz/e-cja/>

Po vyhledání zadaného jazykového dokladu se otevře karta s položkou *knedlík*, viz obrázek 84 níže. Nad názvem vidíme nápis *ČJA 1*, to znamená, že slovo je zařazeno do prvního dílu atlasu, kterému odpovídají tematické okruhy místní prostředí, domácí prostředí, člověk. Před názvem je číselné označení, v tomto případě 100, pod názvem už můžeme vidět doložené tvary. Tímto způsobem lze ověřit, že tvar *buchta* ve významu *knedlík* je v ČJA opravdu dochován.

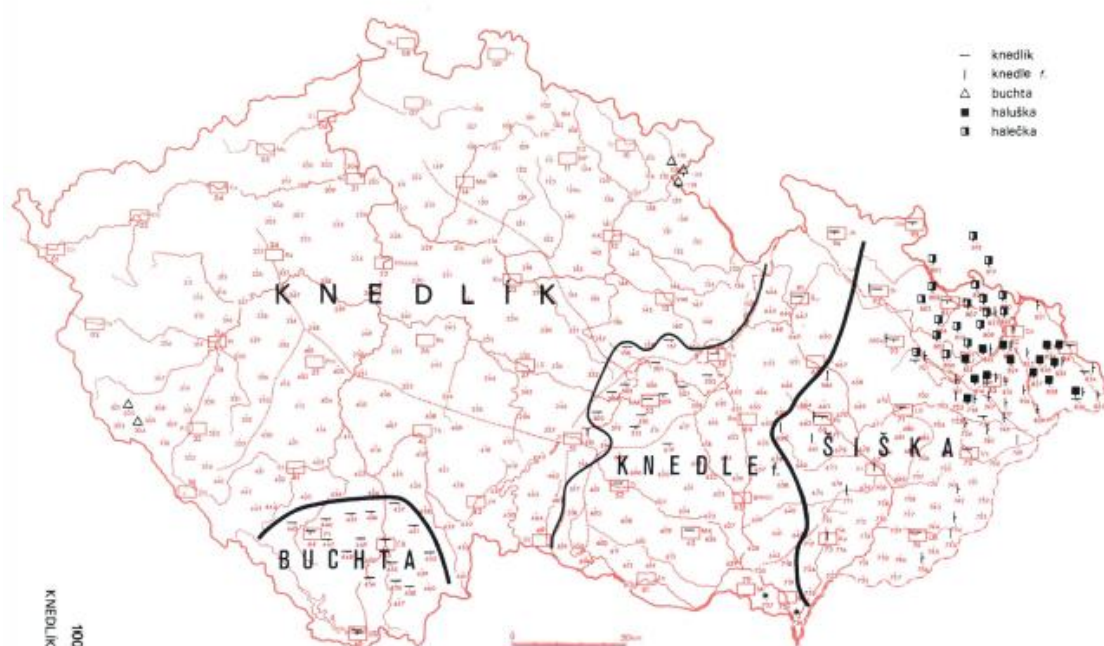
ČJA 1

100 knedlík (406a)

- 1 M knedlík  
knedle *f.* (knědla 755, 756)  
buchta (vařená buchta 303)  
šiška (též 79)  
halaška  
halečka
- S knofel 737, -ěl 736 (*obě nedubl.*)
- N šliška 831, kluska 817, knedlík kinutej, kisaněj, houskovej, rohlíkovej, tvrdej; knedle housková, zemlová, žemličková, rohlíková; šiška kvasničová, kvasená, kvašená, kisaná, kislá, rohlíková, z rohlíkama; halečka bječkova; buchta moučná, velká, vařená

Obrázek 84 – Lexikální jednotka knedlík v elektronické verzi ČJA, zdroj: <https://cja.ujc.cas.cz/e-cja/heslo/1/100>

Pokud chceme vyhledávat ve verzi ČJA k tisku, dostaneme se k ní přes webovou stránku <https://cja.ujc.cas.cz/>. Titulní strana nabízí možnost výběru z několika dílů, pro toto cvičení je důležitý díl první, který se, jak již bylo zmíněno, týká témat místní prostředí, domácí prostředí a člověk. V levé části stránky se nachází obsah, v němž najdeme i soupis zařazených položek. *Knedlík* najdeme pod číslem 100 na straně 210, mapu na straně 213 (viz obrázek 85 níže). Právě pomocí mapy lze snadno ukázat, že ekvivalent buchta je typický pro Doudlebsko, což potvrzují i doklady v korpusu. Pokud vyhledáme slovníkové heslo buchta v SSJČ, zjistíme, že hned u prvního významu je zaznamenáno nářeční slovní spojení ‚vařené buchty‘ ve významu knedlíky (označeno zkratkou nár.)



Obrázek 85 – Mapa lexikální položky knedlík v ČJA, zdroj: <https://cja.ujc.cas.cz/CJA1/>

### 6.3.2 SVÁTKY, ZVYKY, TRADICE

#### Cvičení č. 20 – Svátek sv. Štěpána v západočeské a jihočeské nářeční oblasti

Zadání: Svátek svatého Štěpána, tedy 26. prosince, má v západočeské a jihočeské nářeční oblasti své specifické pojmenování. S pomocí ČJA a korpusu vyhledejte, jak se tento svátek v dané oblasti nazývá a ověřte, zda je doložen v korpusu DIALEKT v tom samém významu.

Postup: Nejrychlejším a nejjednodušším způsobem, jak vyhledat nářeční ekvivalent pro svátek svatého Štěpána, je použít elektronickou verzi ČJA. V té se pokusíme vyhledat jazykové doklady pro tuto položku. Pokud začneme do vyhledávacího pole psát začátek spojení, tedy svátek, už nám ČJA sám nabízí odpovídající položku *svátek sv. Štěpána*, viz obrázek 86 níže.

## ČESKÝ JAZYKOVÝ ATLAS



Obrázek 86 – Vyhledávání položky svátek sv. Štěpána v ČJA, zdroj: <https://cja.ujc.cas.cz/e-cja/>

Po zobrazení hesla se otevře karta se všemi informacemi k dané položce – ČJA 2, položka 225, s. 468, mapa s. 471. Dozvíme se, že pro *svátek sv. Štěpána* jsou doloženy 3 základní nářeční ekvivalenty, z nichž západočeské oblasti odpovídá ekvivalent *prostřední svátek*. Toto slovní spojení se užívá jako sousloví pouze nářečně, SSJČ ho neuvádí.

Nyní přichází na řadu vyhledávání ekvivalentu v korpusu DIALEKT. Vyhledávat můžeme na ortografické rovině s omezením na západočeskou nářeční oblast (viz obrázek 17), podoba dotazu viz obrázek 87 níže.

Hledat v korpusu

dialekt\_v2\_ort ★

Pokročilý dotaz  | Klávesnice | Předchozí dotazy | Interpretace dotazu

prostřední svátek

Obrázek 87 – Podoba dotazu k vyhledání sousloví prostřední svátek, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_dial](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_dial)

Pro takto zadaný dotaz vyhledal korpus v západočeské nářeční oblasti 1 výskyt, který svým významem odpovídá svátku svatého Štěpána, což je jasně vidět přímo z konkordančního řádku bez nutnosti zkoumat další kontext, viz obrázek 88 níže. Výskyt pochází ze staré vrstvy nahrávek, konkrétně z roku 1964 (kód nahrávky 0019-C-ZAC-1964-SA) z Chodské Lhoty, jedná se tedy o okrajový domažlický úsek. Přestože korpus vyhledal pro ekvivalent pouze jeden výskyt, důležité je, že je sousloví v korpusu zaznamenáno.

Výskytů: 1 | Lp.m.: Spočítat | ARF: 1 | Výsledek je seříděn

Vyběr řádků: základní

0019-C-ZAC-1964-SA do kostela .. to byl velkej svátek .. + a na prostřední/prostřední svátek/svátek .. to bylo Štěpána .. tak to byla muzika ..

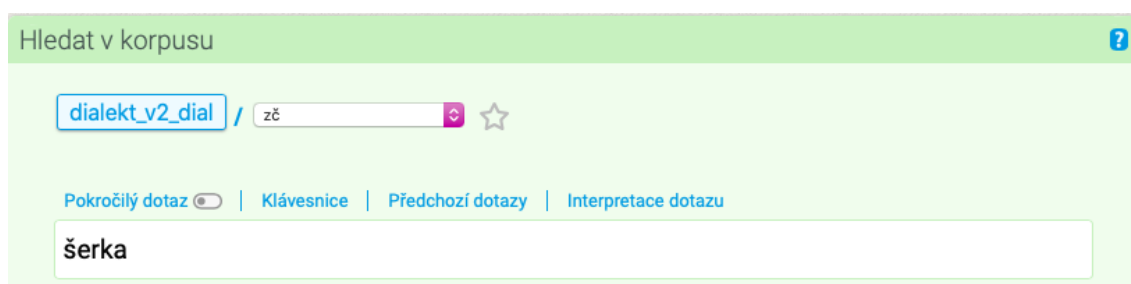
Obrázek 88 – Výsledky vyhledávání sousloví prostřední svátek, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=dial&attr\\_vmode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~5SCcecYeYOqa](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=dial&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~5SCcecYeYOqa)



## Cvičení č. 21 – Šerka v západočeské nářeční oblasti

Zadání: *Západočeská nářeční skupina a zejména její okrajový domažlický úsek je významným etnografickým regionem, v němž se zachovávají různé zvyky a tradice. K nim patří i tradiční a velmi rozmanité oblečení, které se nosilo k jistým příležitostem. Zjistěte, které oblečení se nazývalo **šerka**. Je tento nářeční lexikální jev zachycen v ČJA a SSJČ?*

Postup: Nejprve vyhledáme slovní tvar *šerka* v korpusu DIALEKT. Vyhledávat budeme na rovině dialektologické, jako korpus pro vyhledávání zvolíme tedy *dialekt – v2 dial*. Hledání buď omezíme na západočeskou nářeční oblast (viz obrázek 17), nebo zvolíme pro vyhledávání subkorpus *zč* (viz obrázek 19). Regulární výrazy nebudeme při vyhledávání potřebovat, jako výchozí atribut zvolíme *lemma*. Podoba dotazu viz obrázek 89 níže.



Hledat v korpusu

dialekt\_v2\_dial / zč

Pokročilý dotaz | Klávesnice | Předchozí dotazy | Interpretace dotazu

šerka

Obrázek 89 – Podoba dotazu k vyhledání slovních tvarů lemmatu *šerka*, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_ort](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_ort)

Pro takto zadaný dotaz vyhledal korpus 4 slovní tvary lemmatu *šerka* (viz obrázek 90 níže). Všechny výskyty pochází z nové časové vrstvy a ze 3 různých nahrávek, jejichž mluvčí byli ve věku 71 a 86 let v roce nahrávání. To ukazuje, že jde o nářeční lexikum užívané, v korpusu zaznamenané u starší generace. Zároveň byly všechny nahrávky pořízeny v Postřekově na Domažlicku.

V nahrávce 0239-C-ZAC-2007-N se pod názvem *šerka* označuje barevná sukňe – *a tuti mňeli potom barevní sukňe , to bilo jako šerki , sme říkali , a jupku přes to a přes*. Z konkordančních řádků nahrávky 0251-C-ZAC-2007-N a 0252-C-ZAC-2007-N není význam slova zcela patrný. V ČJA není zachycena položka *sukňe* nebo *kroj*, není tedy možné výsledky vyhledávání ověřit, nicméně SSJČ charakterizuje *šerku* následujícím způsobem – ‚pevná hrubá látka po domácku tkaná ze lnu a z vlny‘. Ve druhém, nářečním významu ‚kus oděvu z takové látky, zprav. sukňe‘.

Výskytů: 4   i.p.m.: 89,98 (vztaženo k subkorpusu "zč")   ARF: -   Výsledek je setříděn	
Výběr řádků: základní	
<input type="checkbox"/>	0239-C-ZAC-2007-N a tuti mñeli potom barevní sukñe , to bilo jako šerki/šerky , sme říkali , a jupku přes to a přes
<input type="checkbox"/>	0251-C-ZAC-2007-N , že jo . to bilo , diš se nosili šerki/šerky . a šerki bílí @ ... (nadechnuti) nó to je
<input type="checkbox"/>	0251-C-ZAC-2007-N . to bilo , diš se nosili šerki . a šerki bílí @ ... (nadechnuti) nó to je sváteční , ale
<input type="checkbox"/>	0252-C-ZAC-2007-N to je zase ešfe jinj kroj , neš je ta šerka/šerka , to je zase různí . (nadechnuti) (nadechnuti)

Obrázek 90 – Výsledky vyhledávání slovních tvarů lemmatu šerka, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=dial&attr\\_vmode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~xgsMswee0g4a](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=dial&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~xgsMswee0g4a)

Šerka je tedy označení pro barevnou sukni jakožto součást kroje. Z nahrávky 0239-C-ZAC-2007-N můžeme vyčíst, že součástí byla dále i jupka (SSJČ – ‚lehký ženský kabátek‘) – *a jupku přes to a přes*. Pokud rozklikneme z této nahrávky KWIC, můžeme z přepisu zjistit, že součástí byl i barevný šátek – *a přes to zase šáteg barevní*. Nahrávka 0251-C-ZAC-2007-N zachycuje jako součást kroje např. vyplétaný límec – *a ten sváteční kroj potom mñel třeba ten límec viplejtanej , že jo*. Další části kroje zmiňuje nahrávka 0251-C-ZAC-2007-N – *vipíglovaní rukáfce a potom mají takovi barevní sukñe fšelijákí zást'erki vornátoví a takoví černí šñerovački a do toho se nosí takovj šátek*.<sup>157</sup>

V dalším postupu se pokusíme nářeční lexikum *šerka* vyhledat v ČJA. Použijeme k tomu elektronickou verzi, ve které budeme slovo vyhledávat jako jazykový doklad. Už při zadávání slova do vyhledávacího okna vidíme, že je v atlasu dochováno, neboť je nám samo nabízeno. Když ale zobrazíme heslo, dozvíme se, že se slovo objevuje jako ekvivalent lexikální položky *popruh u trakaře* (viz ČJA 3, položka 165, s. 384, mapa s. 383). Z mapy zjistíme, že je to ekvivalent typický pro východní podskupinu slezských nářeční. Slovo je tedy v ČJA zachyceno, avšak s jiným významem, jedná se zde o zcela jinou realii.

<sup>157</sup> Na tomto příkladu můžeme vidět typické hláskoslovné jevy, které se v západočeské nářeční oblasti objevují – široká výslovnost dlouhého *á* (viz cvičení 2, s. 71) a zavřená výslovnost dvojhlásky *ej* (viz cvičení 5, s. 81).

## Cvičení č. 22 – Draní peří v jihočeské nářeční oblasti

Zadání: **Draní peří** zejména dříve představovalo na vesnici typickou činnost. **Drát peří** je i položka zařazená v ČJA. Ve kterém z jeho dílů toto spojení najdeme? Vyhledejte příslušný ekvivalent pro jihočeskou nářeční oblast a ověřte, zda je v korpusu DIALEKT zachycen. S touto položkou souvisí dále substantivní pojmenování činnosti ve významu ‚společné draní peří‘. Vyhledejte i tuto položku v ČJA, uveďte číslo, pod kterým je zařazena, a stranu ve verzi ČJA k tisku, na niž se nachází mapa s příslušnými ekvivalenty. Je tvar typický pro jihočeskou oblast doložen v korpusu DIALEKT?

Postup: Nejprve je nutné vyhledat, ve kterém dílu ČJA je zařazena lexikální položka *drát peří*. To můžeme zjistit snadno z elektronické verze atlasu tím, že vyhledáme dané výrazy jako položku atlasu, viz obrázek 91 níže.

# ČESKÝ JAZYKOVÝ ATLAS



Obrázek 91 – Vyhledávání položky *drát peří* v ČJA, zdroj: <https://cja.ujc.cas.cz/e-cja/>

Pokud zobrazíme heslo, dozvíme se z karty, že spadá do druhého dílu ČJA (ČJA 2), kterému odpovídají okruhy zahrada a sad, živočišstvo, les a rostlinstvo, krajina, čas a počasí, vesnice dříve a nyní, zábavy a zvyky. Položka je do atlasu zařazena pod číslem 200 na s. 425. Z informací se v oddílu č. 3 dozvíme, že pro jihočeský areál je typický ekvivalent *odrhat*, a to zejména pro Doudlebsko, viz mapa v ČJA na s. 427.

Dále zadáme tento nářeční výraz do korpusu DIALEKT. Vyhledávat budeme na rovině dialektologické s omezením na jihočeskou oblast (viz obrázek 17), nebo zvolíme pro vyhledávání subkorpus *jč* (viz obrázek 19). Povolíme regulární výrazy, výchozím atributem bude *word*. Do vyhledávacího pole zadáme dotaz *odrh.\**. Pro takový dotaz vyhledá korpus 5 výskytů, viz obrázek 92 níže.

Výskytů: 5 | i.p.m.: 162,1 (vztaženo k subkorpusu "jč") | ARF: - | Výsledek je setříděn

Výběr řádků: základní

<input type="checkbox"/>	0310-C-JIC-2014-N	drhad žencki peří ?	+	no jo , každou zimu sme	odrhali/odrhaly	, diš sme blii dvje holki , tag babička musela		
<input type="checkbox"/>	0840-C-JIC-2008-N	biló @ konedz , vječinou tag do jedenáci hodín se	odrhalo/odrhalo	, potom nám udelali kafe .	+	(nadechnuti)   +   (hluk v pozadí)		
<input type="checkbox"/>	0908-C-JIC-2009-N	uš se nemohli dočkat :	+	(nadechnuti)   +	" gda budete	odrhát/odrhát	uš ? "   +   (smích)   +   no .   +   no jo .   +   a	
<input type="checkbox"/>	0928-C-JIC-2009-N	škrabávali jako děti řipu .	+	krmnou řipu .	+	sme jí	odrhli/odrhli	rejžákem , potom sme jí překrojili na půl a na
<input type="checkbox"/>	0977-C-JIC-2009-N	a to dicky žencki říkali :	+	" gda už budete	odrhát/odrhát	? "   +	máma nas vařila kobihu .   +	ti misi blii

Obrázek 92 – Výsledky vyhledávání lemmatu *odrhát*, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attr\\_vmode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~zkYCOGWA4CSS](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~zkYCOGWA4CSS)

Z 5 výskytů, které korpus pro dotaz *odrh.\** v jihočeské nářeční oblasti vyhledal, 1 neodpovídá svým významem slovu, které vyhledáváme – *sme jí odrhli rejžákem* (0928-C-JIC-2009-N). Celkem tedy korpus vyhledal 4 tvary slovesa *odrhát*. Všechny pochází z nové časové vrstvy nahrávek, 3 z nich přímo z Doudleb, 1 z Kamenného Újezdu (Kosov), všechny tedy spadají do doudlebského okrajového typu. Důležité je, že byly tvary slovesa v korpusu zachyceny, doklady z nové vrstvy ukazují, že se jedná o slovo stále užívané. V SSJČ sloveso doloženo není.

V dalším postupu vyhledáme v ČJA lexikální položku, která s položkou *drát peří* souvisí a znamená ‚společné draní peří‘. Z předchozího vyhledávání víme, že původní položka je zařazena do ČJA 2, proto v něm budeme vyhledávat i položku novou, a to ve verzi atlasu k tisku. V obsahu najdeme lexikální položku s číslem 200 – *drát peří* (s. 425). Pod číslem 199 je zařazena položka *dračky* (s. 422)<sup>158</sup> – pokud si ji rozklikneme, z charakteristiky se dozvíme, že opravdu tato položka znamená ‚společné draní peří‘. Mapu najdeme na s. 423, z ní lze vyčíst, že pro jihočeský areál je typický nářeční ekvivalent *održky*. V SSJČ není výraz tento výraz uveden.

Nyní přichází na řadu opět vyhledávání v korpusu DIALEKT, v němž vyhledáme *održky* na rovině dialektologické pomocí regulárních výrazů – jako korpus tedy zůstane *dialekt* – v2 *dial* s omezením na jihočeskou oblast (viz obrázek 17), nebo zvolíme subkorpus *jč* (viz obrázek 19). Povolíme regulární výrazy, výchozím atributem bude *word*. Dotaz bude mít podobu *odrš.\** a korpus pro něj vyhledá 6 výskytů, viz obrázek 93 níže.

<sup>158</sup> Kromě výše zmíněných položek má s dráním peří souvislost položka *doderky*, viz ČJA 2, položka 201, s. 426, mapa s. 429. Ta ale nemá pro jihočeský areál specifický nářeční výraz.

Výskytů: 6 | Lp.m.: 194,51 (vztaženo k subkorpusu "JE") | ARF: - | Výsledek je seříděn

Vyběr řádků: základní

<input type="checkbox"/>	0840-C-JIC-2008-N	nó .   +   celj do pjejni peřini , tak to bjvávali <b>održki/održky</b> .   +   (hluk v pozadí)   +   (nadechnuti)   +   (nadechnuti)   +   (nadechnuti)
<input type="checkbox"/>	0900-C-JIC-2009-N	chci eše zeptat , jak to bilo , gdiž bili <b>održki/održky</b> , u feč Králú u vás , takovi velki .
<input type="checkbox"/>	0900-C-JIC-2009-N	a kolik tam chođilo třeba feč sousedek , na ti <b>održki/održky</b> ?   +   kolikrát náz bilo aš pjetadvacet .   +   no , to
<input type="checkbox"/>	0900-C-JIC-2009-N	no .   +   (bouchnuti)   +   (nadechnuti)   +   přectaf si , tam bejvali veseli <b>održki/održky</b> a to žencki uš se nemohli dočkat :   +   (nadechnuti)   +   (nadechnuti)
<input type="checkbox"/>	0901-C-JIC-2009-N	vám chođilo tolig lidí na ti hrátki , na ti <b>održki/održky</b> ?   +   k fem Králúm .   +   no .   +   jó , chođili
<input type="checkbox"/>	0977-C-JIC-2009-N	komu se eše v Doudlebech taki hođie chođilo na ti <b>održki/održky</b> nebo na hrátki ?   +   (hlasitý hovor v pozadí)   +   nó tak ...   +   já aři

Obrázek 93 – Výsledky vyhledávání slovních tvarů lemmatu *održky*, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=dial&attr\\_vmode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~rMkMIw6WaAke](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=dial&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~rMkMIw6WaAke)

Slovo *održky* bylo tedy ve významu ‚společné draní peří‘ nalezeno v korpusu v jihočeské nářeční oblasti 6krát, všechny výskyty pocházejí z nové časové vrstvy nahrávek a všechny tyto nahrávky byly pořízeny přímo v Doudlebech. Důležité je, že bylo slovo zachyceno v korpusu v takovém počtu a v nové vrstvě, to ukazuje na jeho frekventované užívání i v současnosti.

### 6.3.3 HOSPODÁŘSTVÍ

#### Cvičení č. 23 – Nářeční ekvivalenty slova *kozel* v západočeské a jihočeské nářeční oblasti

Zadání: *Názvy dobytka mají nářečně specifická pojmenování, jedním z nich je i kozel. Vyhledejte toto slovo v ČJA a zjistěte jeho ekvivalenty pro západočeskou a jihočeskou nářeční skupinu. Ve kterém dílu atlasu tuto položku najdeme? Uveďte číslo položky a stranu, na níž je ve verzi atlasu k tisku mapa. Ověřte přítomnost těchto nářečních podob v korpusu DIALEKT.*

Postup: Lexikální položku *kozel* můžeme vyhledat jak v elektronické verzi atlasu, tak i v té k tisku. Při použití elektronické verze zapíšeme do vyhledávacího okna dané slovo a vyhledáme ho jako položku. Po zobrazení hesla se otevře karta s danou lexikální položkou (viz obrázek 94 níže), která je zařazena do ČJA 3, jemuž odpovídají okruhy polní zemědělské práce, hospodářská usedlost, zemědělské náradí a nástroje, dobytek, drůbež. Položka je označena číslem 208.

#### ČJA 3

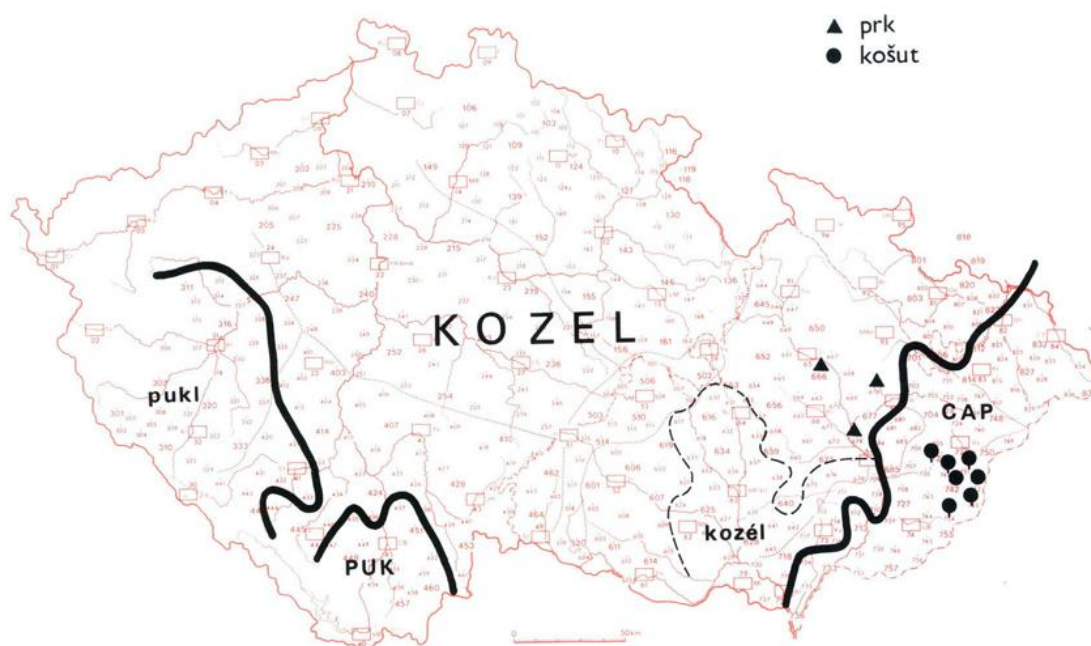
##### 208 kozel (869)

- 1 M    kozel — kozél  
      pukl (pokl 306, 421, též pugl 435, též pugel 447) — puk (púk 457)  
      prk 666, 676, 678 (prkál 676)  
      cap  
      košut
- 2        Zkoumalo se nář. pojmenování pro kozla, tj. pro samce kozy domácí. Pojmenování se liší zejména lexikálně a dále hláskoslovně (*kozel* × *kozél*). Z mapování jsme vyloučili prokazatelně deminutivní podobu *kozlík*. Ekvivalent *prk*, označující původně ‚kozí pach‘, mívá expresivní zabarvení; potvrdila jej i KLA.  
      Výrazy *puk* a *pukl* jsou odvozeny z něm. nář. slov *Bock* a *Bockel*, která byla na našem území hláskově adaptována; rozdíl *pukl* × *puk* vedeme na naší mapě jako hláskoslovný.
- 3        Pojmenování *kozel*, shodné se spisovným jazykem, bylo zaznamenáno na největší části zkoumaného území: je rozšířeno po celých Čechách (kromě záp. a již. okrajů) a na západní a střední Moravě a na záp. Slezsku; na širším Brněnsku s přesahy ke Kyjovu a k Mikulovu je doloženo v podobě *kozél*.  
      V zč. nář. a na Strakonicku vytváří kompaktní areál výraz *pukl*, podoba *puk* je charakteristická pro širší Českobudějovicko.  
      Oblast slova *cap* se rozkládá hlavně ve vm. nář. (s výjimkou dolských nář.), dále na Holešovsku a ve slez. nář. od řeky Odry na východ. Mezi Vsetínem a Uherským Brodem se většinou nedubletně vyskytuje lexém *košut*.

Obrázek 94 – Karta lexikální položky *kozel* v ČJA, zdroj: <https://cja.ujc.cas.cz/e-cja/heslo/3/208>

Pokud vyhledáváme ve verzi ČJA k tisku, je nutné si nejprve zařadit slovo do správného tematického okruhu, kterému odpovídá právě ČJA 3. V něm v obsahu pod číslem 208 nalezneme na s. 462 položku *kozel*, jejíž mapa se nachází na s. 461 (viz obrázek 95 níže) spolu s mapou pro lexikální položku (*koza je*) *březí*. Z elektronické verze atlasu i z verze

k tisku lze jasně vyčíst, že pro západočeskou nářeční oblast je doložený nářeční výraz *pukl*, pro jihočeskou (Českobudějovicko a Doudlebsko) ekvivalent *puk*.<sup>159</sup>



208  
KOZEL

Obrázek 95 – Mapa lexikální položky *kozel* v ČJA, zdroj: <https://cja.ujc.cas.cz/CJA3/>

Nyní přichází na řadu vyhledávání ekvivalentů v korpusu DIALEKT. Vyhledávat budeme na rovině dialektologické, vybereme subkorpus *jzč* (viz obrázek 19). Povolíme regulární výrazy, atributem pro vyhledávání bude *word*. Zadáme dotaz *puk.\** a vyhledáme. Korpus ukáže 2 výsledky, z toho 1 pro náš hledaný výraz – *puka* – v nahrávce 0824-C-JIC-2008-N z Doudleb – *ale to ale to bəlo sekanou , mi sme mívali puka malího t\* ot , nechala nam máma dicky kozle*. Nářeční ekvivalent pro jihočeskou skupinu je tedy v korpusu zachycen, ekvivalent *pukl* pro oblast západočeskou doložen ale není. V SSJČ není naopak doloženo slovo *puk*, ale *pukl* ano – ‚nář. kozel‘.

Na tomto cvičení, stejně jako na některých výše, se ukazují limity korpusu. Vyhledávány byly jevy, které jsou v ČJA doloženy z terénního výzkumu, avšak v korpusu nejsou

<sup>159</sup> Viz SSJČ, *pukl* – nář. kozel, ekvivalent *puk* není zaznamenán.

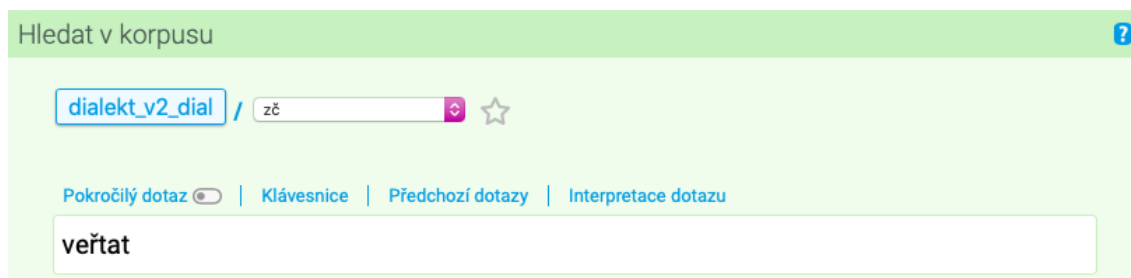
doloženy vůbec. To se však může změnit, až bude zveřejněn další verze korpusu rozšířena nářeční materiálová základna.



## Cvičení č. 24 – Výraz *veřtat* v západočeské nářeční oblasti

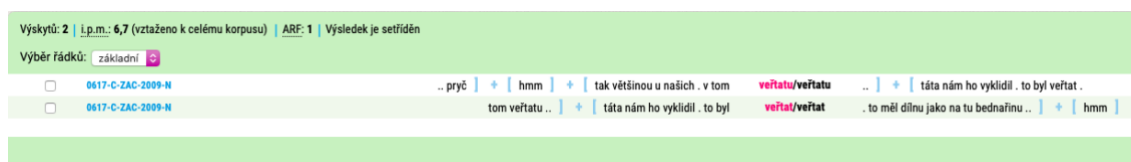
Zadání: *Součástí hospodářské usedlosti býval i tzv. veřtat, což je nářeční výraz. Vyhledejte ho v korpusu DIALEKT, zjistěte oblast výskytu a uveďte, jaká část domu to byla. Je tento ekvivalent doložen v ČJA? Pokud ano, pod jakou lexikální položkou je zařazen?*

Postup: *Veřtat* budeme vyhledávat v korpusu dialekt na rovině dialektologické. Jako korpus zvolíme *dialekt – v2 dial* a vyhledávání omezíme na západočeskou nářeční oblast (viz obrázek 17), nebo zvolíme subkorpus *zč* (viz obrázek 19). Výchozím atributem pro vyhledávání bude *lemma*. Podoba dotazu viz obrázek 96 níže.



Obrázek 96 – Podoba dotazu k vyhledání slovních tvarů lemmatu *veřtat*, zdroj: [https://www.korpus.cz/kontext/query?corpname=dialekt\\_v2\\_ort](https://www.korpus.cz/kontext/query?corpname=dialekt_v2_ort)

Pro takto zadaný dotaz vyhledá korpus v západočeské oblasti dva výskyty, oba v nové vrstvě nahrávek a v jedné nahrávce s kódem 0617-C-ZAC-2009-N z Losiné, viz obrázek 97 níže. Z druhého výskytu je patrný význam slova *veřtat*, tedy dílna, v tomto případě konkrétně na bednařinu – *to bil veřtat, to mňel d'ílnu jako na tu bednařinu*.



Obrázek 97 – Výsledky vyhledávání slovních tvarů lemmatu *veřtat*, zdroj: [https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=dial&attr\\_vmode=visible-kwic&base\\_viewattr=word&refs=%3Ddoc.id&q=~V4G4Co6g6U6e](https://www.korpus.cz/kontext/view?viewmode=kwic&pagesize=40&attrs=word&attrs=dial&attr_vmode=visible-kwic&base_viewattr=word&refs=%3Ddoc.id&q=~V4G4Co6g6U6e)

V dalším postupu budeme vyhledávat nářeční výraz v ČJA. Pokud zadáme slovo do vyhledávače v elektronické verzi atlasu, není nám nabídnuta žádná možnost, to znamená, že slovo není doloženo. To ale není pravda, zde se ukazuje jeden z limitů

této verze atlasu. Ekvivalent *veřtat* je totiž zařazen pod položku *kovárna* (ČJA 2, položka 172, s. 386, mapa s. 387), viz obrázek 98 níže.

#### ČJA 2

##### 172 *kovárna* (1984)

- 1 M *kovárna* — *kovarna*  
    *kovárňa* — *kovárňe*  
    *kováreň* (těž 506)  
    *kuzňa* (kužňa 818, 819, 830–836)
- N *veřtat*, *vařtat*, *d'ílna*
- 2 Náf. pojmenování pro kovářskou dílnu vykazují především rozdíly morfologické povahy (*kovárna* × *kováreň* × *kovárňa*) a rozdíly hláskoslovné (např. *kovárňa* × *kovárně*, *kovárna* × *kovarna*, *kuzňa* × *kužňa*). Jediný rozdíl lexikální (*kovárna* × *kuzňa*) je okrajový.  
    V běžné komunikaci, vázané více na situační kontext, lze místy užít v sledovaném významu těž výrazů *veřtat* a *dílna*, jež mají význam obecnější. Jde o užití zástupné, proto jsme je na mapě nezachytili.
- 3 Kromě slezských nářečí, z nichž se dokládá pojmenování *kuzňa*, se na celém ostatním teritoriu vyskytují varianty výrazu *kovárna*. Starou zemskou hranicí česko-moravskou sleduje rozdíl *kovárna* × *kovárňa/kovárně* (na Moravě). Varianta *kováreň* je v pruhu obcí na nejzápad. Moravě. Hláskoslovná obměna *kovarna* je omezena na zč. náf. úsek.

Obrázek 98 – Karta lexikální položky *kovárna* v ČJA, zdroj: <https://cja.ujc.cas.cz/e-cja/heslo/2/172>

V první části pod názvem lexikální položky můžeme nalézt ekvivalenty, které jsou označeny dvěma písmeny. K písmenu *M* náleží mapované varianty, k písmenu *N* nemapované, které nabývají považovány za plná nářeční synonyma, a protože nejsou mapované, tak nejsou ani jako jazykové doklady do vyhledávání zařazeny. Mezi takové patří právě i slovo *veřtat*, které je nicméně doloženo v korpusu, i když pouze 2krát a jen v 1 nahrávce. Bylo by vhodné zařadit i nemapované varianty do jazykových dokladů elektronické verze ČJA, pomocí korpusu pak lze ověřit, zda byla slova opravdu zachycena. V SSSJČ je slovo doloženo společně s výrazem *veřtat* pod jedno heslo – ‚zast. ob. dílna, řemeslnický stůl‘ či ‚náf. tkalcovský stav‘.

## ZÁVĚR

Korpusová lingvistika se v posledních letech velmi rozmáhá zejména pro její možnost využívat a zkoumat aktuální jazykovou situaci, v případě nářečí i situaci minulou. To byl hlavní důvod, proč je předkládaná diplomová práce zaměřena právě tímto směrem. Možnost zkoumat stav českých nářeční v minulosti nabízí mnohé odborné publikace či *Český jazykový atlas*. Korpus DIALEKT však přináší nářeční jazykový materiál i ze současnosti, přestože není zatím příliš obsáhlý.

Kapitoly 1–5 jsou zaměřeny teoreticky. V první kapitole byly obecně představeny mluvené korpusy jakožto materiálové úložiště dřívějšího i aktuálního a zejména autentického mluveného jazyka. Mluvených korpusů poskytuje ČNK několik, jejich stavba a struktura je v téže kapitole stručně rozebrána. Korpusy mluveného jazyka se ukazují, stejně jako korpusy psané, jako velmi složité komplexy náročné na sestavení i z hlediska časového, neboť vyžadují transkript mluveného slova do psané podoby. Na tomto procesu se autorka práce pro korpus DIALEKT podílí, což byl další důvod ke zvolení tématu souvisejícího s korpusovou lingvistikou. První kapitola se taktéž věnuje možnosti využití mluvených korpusů, ale pouze ve stručnosti, neboť na toto téma vzniklo již několik kvalifikačních prací a odborných článků.

Důležitá pro teoretickou část této práce je druhá kapitola, která se zabývá nářečním korpusem DIALEKT. Ten obsahuje nahrávky od 50. let minulého století po současnost, má tedy i diachronní charakter a lze ho využít právě pro zkoumání vývoje nářečí. Velkou výhodou je pestrost sociolingvistického značkování, korpus poskytuje mnoho informací o mluvčích a nahrávkách samotných, čímž lze zkoumat i mimojazykové aspekty. Složitý je přepis na dvou rovinách, na nějž je nutné dát pozor při vyhledávání – rovina ortografická je sice přehlednější, neboť nevyužívá fonetický přepis, ale spoustu znaků na ní není doloženo, s touto skutečností je třeba počítat. Na korpusu DIALEKT jsou založeny další nadstavbové aplikace poskytované ČNK, z nichž významné místo zaujímá zejména *Mapka*. Po výzkumu zahraničních nářečních korpusů, atlasů a dalších aplikací (viz kapitola 4) bylo zjištěno, že *Mapka* mezi nimi vyniká jak grafikou, tak jednoduchým interaktivním použitím i rozsahem materiálu. Co se týká zdrojů obsahujících nářeční materiál, ve třetí kapitole práce byl popsán ČJA, s nímž je v praktické části také pracováno, a jsou ukázány výhody i nevýhody především jeho elektronické verze.

Jak již bylo zmíněno, práce se mimo jiné zaměřila na zmapování databází a archivů zahraničních nářečních korpusů, mapových aplikací i jazykových atlasů. Významnou roli hraje databáze CLARIN s mnoha jazykovými zdroji. S ohledem na zahraniční nářeční korpusy lze DIALEKT charakterizovat jako důkladněji propracovaný, pokud ho srovnáme se slovenským nářečním korpusem, který není lemmatizovaný ani morfologicky značkový, tudíž nabízí užší možnosti využití. Na podobné úrovni jsou naopak mapové aplikace. Německá *Dialektkarte* poskytuje stejně jako *Mapka* mnoho materiálu (nahrávky s přepisy), bohužel nebylo možné z dostupných zdrojů dohledat jejího provozovatele. Nejvíce podobná *Mapce* je mapová aplikace polských nářečí, v níž jsou vyděleny a důkladně popsány nářeční oblasti, ke každému dialektu je připojena minimálně jedna nahrávka, tudíž co do materiálu se jedná taktéž o aplikaci obsáhlou. Z jazykových atlasů je třeba zmínit německý *Atlas zur deutschen Alltagsprache*, který má 11 dílů a další právě vzniká, co do obsahu je tedy mnohem rozsáhlejší než ČJA.

Vedle druhé kapitoly zaměřující se na korpus DIALEKT je neméně důležitá pátá kapitola, jejímž cílem bylo popsat charakteristické znaky jihozápadočeských nářečí, neboť je s nimi dále pracováno v další kapitole. Pro potřeby této kapitoly byly využity nejrůznější zdroje od odborných publikací přes časopisecké články po encyklopedické slovníky. Hlavním východiskem byla práce Jaroslava Voráče *Česká nářečí jihozápadní*. Poznatky z jeho díla byly dále doplněny publikacemi dalších odborníků, *Mapkou*, ČJA či NESČ.

Poznatky z druhé a páté kapitoly jsou syntetizovány a využity v šesté kapitole, která tvoří praktickou část této diplomové práce. Ta byla zaměřena didakticky, a sice na tvorbu cvičení, pomocí nichž lze vybrané charakteristické nářeční jevy z jihozápadočeské nářeční podskupiny prezentovat právě na korpusových datech. Cílem zároveň bylo na cvičeních ukázat základy práce s korpusovým vyhledávačem KonText. Do práce bylo zařazeno celkem 24 cvičení rozdělených do 3 oddílů – hláskosloví, tvarosloví, lexikum – s rovnoměrným rozložením nářečních oblastí, přičemž důraz je kladen i na významné okrajové nářeční typy – Chodsko a Doudlebsko. Pro cvičení byly vybrány znaky, které jsou v korpusu doloženy a je možné na nich vybrané jevy názorně prezentovat.

Diplomová práce se v praktické části snaží poukázat zejména na výhody využívání korpusových dat při výuce českého jazyka, a to na středních školách. Nabízí se zde i využití cvičení na školách vysokých při studiu dialektologie a korpusové lingvistiky.

Zde však vyvstávají určité limity, které je nutné zmínit. Výuka českého jazyka na středních školách se v současnosti zaměřuje zejména na literární vývoj a v hodinách mluvnice není tolik prostoru pro výklad nářečního rozrůznění. Nabízí se také zařazení práce s korpusem do hodin informatiky, při níž by prezentace nářečních znaků tímto způsobem splňovala mezipředmětovou spolupráci, avšak vyžadovala by kooperaci vyučujících napříč předměty.

V průběhu tvorby cvičení se ukázaly další limity, a to samotného korpusu či ČJA. Jedním z nich je nedostatek korpusového materiálu, neboť DIALEKT zatím není tak rozsáhlý. To můžeme vidět například v oddílu hláskosloví, konkrétně ve cv. 3 – vkladné *-j-* před *-d'* v západočeské oblasti není doloženo, stejně tak rozložená výslovnost retnice *p* a *v* na Doudlebsku, viz oddíl hláskosloví, cv. 8. V tvaroslovném oddílu se na cv. 14 ukazuje nedostatek dotazů zadávaných pomocí CQL dotazovacího jazyka. Korpus DIALEKT není ve všech případech správně morfologicky označovaný, proto je nutné pokročilé dotazy kontrolovat jiným způsobem. I když se dotazovací jazyk zdá jako přesný a specifitější, nepřesné značkování může způsobit chybné výsledky. Dalším nedostatkem CQL jazyka je jeho složitost, při zadávání dotazů je třeba užívat různé značky a symboly se specifickým významem, bez jejich znalosti nelze dotazovací jazyk správně používat.

Nakonec se limity korpusu a zároveň ČJA ukázaly při tvorbě cvičení zařazených do lexikálního oddílu. Při zkoumání nářeční slovní zásoby nevyhledával korpus tolik výskytů oproti hláskosloví a tvarosloví, která jsou v korpusu zaznamenána s mnohem vyšší frekvencí, a to z důvodu, že staré reálie mizí, nepoužívají se. Bylo zjištěno, že některá slova nejsou zaznamenána v ČJA přesto, že v korpusu je pro ně doklad, například *vosrkejchle* ve cv. 18, které bylo zaznamenáno v nové vrstvě nahrávek, ale terénní výzkum pro potřeby atlasu probíhal v 60. a 70. letech. ČJA nemohl zachytit všechna slova a slovní typy, a proto vzniká SNČJ<sup>160</sup>. Situace může být i opačná, jak se ukázalo ve cv. 23, v němž byly vyhledávány ekvivalenty slova *kozel*. Pro západočeskou oblast ČJA uvádí ekvivalent *pukl*, který ale není v korpusu doložen. To může být způsobeno nedostatečnou materiálovou základnou, tato situace se ale může s další verzí korpusu změnit. V případě cv. 24 při vyhledávání slova *veřtat* v ČJA došlo k tomu,

---

<sup>160</sup> *Slovník nářečí českého jazyka*, dostupné z: <https://sncj.ujc.cas.cz/>.

že elektronická verze atlasu slovo nevyhledala, ale slovo je zaznamenáno jako nemapovaný výraz. Na všech těchto případech se ukazují jak nedostatky korpusu i atlasu, tak i výhody toho, že pro zkoumání nářečních jevů je k dispozici více zdrojů, pomocí nichž je možné informace ověřovat a konfrontovat.

Korpusová data nabízí široké možnosti zkoumání díky různě zaměřeným korpusům a jsou velmi vhodná pro výzkum nářeční situace v minulosti i přítomnosti. Žáci v dnešní době žijí sociálními sítěmi, aplikacemi a internetem, proto by pro ně mohlo být on-line prostředí práce s korpusem atraktivní. Je však nutné mít na paměti, že žáci či studenti musí mít o této práci základní znalosti, které byly do cvičení postupně začleněny. Cílem práce bylo nejenom představit mluvené korpusy se zaměřením na DIALEKT a jak v něm lze vyhledávat jihozápadočeské nářeční znaky, ale zejména ukázat možné využití korpusu ve výuce českého jazyka, samozřejmě dle možností učitele i žáků či studentů. Doufáme, že práce bude užitečná at' už jako zdroj cvičení, tak i jako inspirace pro tvorbu vlastních cvičení či pouze pro prezentaci nářečního materiálu.

## SEZNAM POUŽITÝCH ZDROJŮ

### Knižní publikace

BĚLIČ, Jaromír. *Nástin české dialektologie*. Praha: Státní pedagogické nakladatelství, 1972.

LAMPRECHT, Arnošt. *České nářeční texty*. Praha: Státní pedagogické nakladatelství, 1976.

VORÁČ, Jaroslav. *Česká nářečí jihozápadní. Studie jazykově zeměpisná. Část první*. Praha: Nakladatelství Československé akademie věd, 1955.

VORÁČ, Jaroslav. *Česká nářečí jihozápadní. Studie jazykově zeměpisná. Část druhá*. Praha: Akademie, 1976.

### Články v periodikách a sbornících

GOLÁŇOVÁ, Hana a Martina WACLAWIČOVÁ. Nářeční korpus DIALEKT a jeho použití ve výuce češtiny. *Český jazyk a literatura*. 2018, **69**(3), 127–133. ISSN 0009-0786.

GOLÁŇOVÁ, Hana a Martina WACLAWIČOVÁ. The DIALEKT Corpus and Its Possibilities. *Jazykovedný časopis*. 2019, **70**(2), 336–344. ISSN 1338-4287.

JANČÁK, Pavel. K jazykovému zeměpisu Čech. Některé případy západočeského krácení v koncovkách. *Naše řeč* [online]. 1970, **53**(3), 129–137. ISSN 2571-0893. Dostupné z: <http://nase-rec.ujc.cas.cz/archiv.php?lang=en&art=5495>.

VALIŠOVÁ, Pavlína. Data-driven learning a výuka češtiny jako cizího jazyka. *CASALC Review*. 2012, **2**(2), 22–39. ISSN 1804-9435.

## Slovníky

*Slovník nářečí českého jazyka* [online]. Brno: Dialektologické oddělení Ústavu pro jazyk český AV ČR, v. v. i., 2016. Dostupné z: <https://sncj.ujc.cas.cz/>.

*Slovník spisovného jazyka českého* [online]. Praha: Ústav pro jazyk český AV ČR, v. v. i, 2011. Dostupné z: <https://ssjc.ujc.cas.cz/search.php?db=ssjc>.

## Hesla v Novém encyklopedickém slovníku češtiny

BACHMANNOVÁ, Jarmila. Česká nářeční skupina. In: *CzechEncy – Nový encyklopedický slovník češtiny* [online]. Brno: Masarykova univerzita, 2017. Dostupné z: <https://www.czechency.org/slovník/ČESKÁ%20NÁŘEČNÍ%20SKUPINA>.

HLADKÁ, Zdeňka. Dialektismus. In: *CzechEncy – Nový encyklopedický slovník češtiny* [online]. Brno: Masarykova univerzita, 2017. Dostupné z: <https://www.czechency.org/slovník/ČESKÁ%20NÁŘEČNÍ%20SKUPINA>.

KLOFEROVÁ, Stanislava. Jazykový atlas. In: *CzechEncy – Nový encyklopedický slovník češtiny* [online]. Brno: Masarykova univerzita, 2017. Dostupné z: <https://www.czechency.org/slovník/JAZYKOVÝ%20ATLAS>.

KLOFEROVÁ, Stanislava. Jazykový zeměpis. In: *CzechEncy – Nový encyklopedický slovník češtiny* [online]. Brno: Masarykova univerzita, 2017. Dostupné z: <https://www.czechency.org/slovník/JAZYKOVÝ%20ZEMĚPIS#bibitem8>.

KLOFEROVÁ, Stanislava. Nářeční skupina. In: *CzechEncy – Nový encyklopedický slovník češtiny* [online]. Brno: Masarykova univerzita, 2017. Dostupné z: <https://www.czechency.org/slovník/NÁŘEČNÍ%20SKUPINA>.



## Archivy, databáze, jazykové atlasy, korpusy, mapové aplikace

BALHAR, Jan a kol. *Český jazykový atlas* [online]. Elektronické vydání, HTML verze. Brno: Dialektologické oddělení Ústavu pro jazyk český AV ČR, v. v. i., 2018. ISBN 978-80-88211-06-8. Dostupné z: <https://cja.ujc.cas.cz/e-cja/>.

BALHAR, Jan a kol. *Český jazykový atlas 1* [online]. 2., elektronické, opravené a doplněné vydání. Brno: Dialektologické oddělení Ústavu pro jazyk český AV ČR, v. v. i., 2012. ISBN 978-80-86496-67-2. Dostupné z: <https://cja.ujc.cas.cz/CJA1/>.

BALHAR, Jan a kol. *Český jazyková atlas 2* [online]. 2., elektronické, opravené a doplněné vydání. Brno: Dialektologické oddělení Ústavu pro jazyk český AV ČR, v. v. i., 2012. ISBN 978-80-86496-68-9. Dostupné z: <https://cja.ujc.cas.cz/CJA2/>.

BALHAR, Jan a kol. *Český jazyková atlas 3* [online]. 2., elektronické, opravené a doplněné vydání. Brno: Dialektologické oddělení Ústavu pro jazyk český AV ČR, v. v. i., 2014. ISBN 978-80-86496-80-1. Dostupné z: <https://cja.ujc.cas.cz/CJA3/>.

BALHAR, Jan a kol. *Český jazyková atlas 4* [online]. 2., elektronické, opravené a doplněné vydání. Brno: Dialektologické oddělení Ústavu pro jazyk český AV ČR, v. v. i., 2014. ISBN 978-80-86496-81-8. Dostupné z: <https://cja.ujc.cas.cz/CJA4/>.

BALHAR, Jan a kol. *Český jazyková atlas 5* [online]. 2., elektronické, opravené a doplněné vydání. Brno: Dialektologické oddělení Ústavu pro jazyk český AV ČR, v. v. i., 2016. ISBN 978-80-86496-95-5. Dostupné z: <https://cja.ujc.cas.cz/CJA5/>.

BALHAR, Jan a kol. *Český jazykový atlas. Dodatky* [online]. 2., elektronické, opravené a doplnění vydání. Brno: Dialektologické oddělení Ústavu pro jazyk český AV ČR, v. v. i., 2016. ISBN 978-80-88211-01-3. Dostupné z: <https://cja.ujc.cas.cz/CJA-dodatky/>.

Deutsche Mundarten: DDR. *Archiv für Gesprochenes Deutsch* [online]. Mannheim: Leibniz-Institut für Deutsche Sprache, 2020. Dostupné z: [http://agd.ids-mannheim.de/DR--\\_extern.shtml](http://agd.ids-mannheim.de/DR--_extern.shtml).

Deutsche Mundarten: Kreis Böblingen. *Archiv für Gesprochenes Deutsch* [online]. Mannheim: Leibnitz-Institut für Deutsche Sprache, 2020. Dostupné z: [http://agd.ids-mannheim.de/BB--\\_extern.shtml](http://agd.ids-mannheim.de/BB--_extern.shtml).

Deutsche Mundarten: Schwarzwald. *Archiv für Gesprochenes Deutsch* [online]. Mannheim: Leibnitz-Institut für Deutsche Sprache, 2020. Dostupné z: [http://agd.ids-mannheim.de/SW--\\_extern.shtml](http://agd.ids-mannheim.de/SW--_extern.shtml).

Deutsche Mundarten: Südwestdeutschland und Vorarlberg. *Archiv für Gesprochenes Deutsch* [online]. Mannheim: Leibnitz-Institut für Deutsche Sprache, 2020. Dostupné z: [http://agd.ids-mannheim.de/SV--\\_extern.shtml](http://agd.ids-mannheim.de/SV--_extern.shtml).

Deutsche Mundarten: Zwirner-Korpus. *Archiv für Gesprochenes Deutsch* [online]. Mannheim: Leibnitz-Institut für Deutsche Sprache, 2020. Dostupné z: [http://agd.ids-mannheim.de/ZW--\\_extern.shtml](http://agd.ids-mannheim.de/ZW--_extern.shtml).

Elan. *The Language Archive* [online]. Nijmegen: Max Planck Institute for Psycholinguistics. Dostupné z: <https://archive.mpi.nl/tla/elan>.

ELSPAß, Stephan a Robert MÖLLER. *Atlas zur deutschen Alltagssprache* [online]. 2003. Dostupné z: <https://www.atlas-alltagssprache.de/>.

GOLÁŇOVÁ, Hana, WACLAWIČOVÁ, Martina a Jakub PEJCHA. *Mapka: Mapová aplikace pro korpusy mluvené češtiny* [online]. Verze 1.2. Praha: ÚČNK FF UK, 2021. Dostupné z: <https://www.korpus.cz/mapka/>.

Gwary polskie na Litwie. *Gwary polskie na Litwie* [online]. Vilnius: Uniwersytet Wileński, Wydział Filologiczny, 2016. Dostupné z: <http://www.lenkutarmes.flf.vu.lt/>.

*IDEA – International Dialects of English Archive* [online]. 2021. Dostupné z: <https://www.dialectsarchive.com/>.

Korpus nářečí Slovenského národného korpusu. *Slovenský národný korpus* [online]. Bratislava: Jazykovedný ústav Ľ. Štúra Slovenskej akadémie vied, 2018. Dostupné z: <https://korpus.sk/dialect.html>.

LIA. *University of Oslo* [online]. Oslo, 2020. Dostupné z: <https://www.hf.uio.no/iln/english/research/projects/language-infrastructure-made-accessible/index.html>.

LIA Norwegian – Corpus of historical dialect recordings. *National Library of Norway* [online]. Oslo: University of Oslo, 2021. Dostupné z: <https://www.nb.no/sprakbanken/en/resource-catalogue/oai-tekstlab-uio-no-lia-norsk/>.

Map of the Dialectal Division of the Bulgarian Language. *Institute for Bulgarian Language* [online]. 2014. Dostupné z: <https://ibl.bas.bg/en/struktura/balgarska-dialektologiya-i-lingvistichna-geografiya/karta-na-dialektnata-delitba-na-balgarskiya-ezik/>.

Mapa dialektów. *Dialekty i gwary polskie. Kompendium internetowe*. [online]. Ministerstwo Kultury i Dziedzictwa Narodowego, 2010. Dostupné z: <http://www.dialektologia.uw.edu.pl/index.php?11=opis-dialektow>.

Nordic Dialect Corpus and Syntax Database. *ScanDiaSyn* [online]. 2009. Dostupné z: <http://www.tekstlab.uio.no/scandiasyn/index.html>.

Sound Map – Accents & dialects. *British Library Sounds* [online]. The British Library, 2009. Dostupné z: <https://sounds.bl.uk/Sound-Maps/Accents-and-Dialects>.

SPERBER, Andreas. *Dialektkarte* [online]. Dostupné z: <https://www.dialektkarte.de/>.

Spoken Corpora. *Clarín* [online]. Utrecht: CLARIN ERIC, 2021. Dostupné z: <https://www.clarin.eu/resource-families/spoken-corpora>.

## Korpusové manažery

CVRČEK, Václav a Pavel VONDŘIČKA. *SyD – Korpusový průzkum variant*. Praha: Filozofická fakulta Univerzity Karlovy, 2011. Dostupné z: <https://syd.korpus.cz/>.

MACHÁLEK, Tomáš. *KonText – rozhraní pro vyhledávání v korpusech*. Praha: Filozofická fakulta Univerzity Karlovy, 2014. Dostupný z: <https://www.korpus.cz/kontext/>.

## Webové stránky

CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:bmK* [online]. Příručka ČNK, 2015. Dostupné z: <https://wiki.korpus.cz/doku.php/cnk:bmK>.

CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:dialekt* [online]. Příručka ČNK, 2020. Dostupné z: <https://wiki.korpus.cz/doku.php/cnk:dialekt>.

CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:dialekt:pravidla* [online]. Příručka ČNK, 2018. Dostupné z: <https://wiki.korpus.cz/doku.php/cnk:dialekt:pravidla>.

CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:lemtag\_mluv* [online]. Příručka ČNK, 2017. Dostupné z: [https://wiki.korpus.cz/doku.php/cnk:lemtag\\_mluv](https://wiki.korpus.cz/doku.php/cnk:lemtag_mluv).

CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:lindsei\_cz* [online]. Příručka ČNK, 2017. Dostupné z: [https://wiki.korpus.cz/doku.php/cnk:lindsei\\_cz](https://wiki.korpus.cz/doku.php/cnk:lindsei_cz).

CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:oral* [online]. Příručka ČNK, 2019. Dostupné z: <https://wiki.korpus.cz/doku.php/cnk:oral>.

CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:oral:pravidla* [online]. Příručka ČNK, 2017. Dostupné z: <https://wiki.korpus.cz/doku.php/cnk:oral:pravidla>.

CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:orator* [online]. Příručka ČNK, 2020. Dostupné z: <https://wiki.korpus.cz/doku.php/cnk:orator>.

CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:ortofon* [online]. Příručka ČNK, 2020. Dostupné z: <https://wiki.korpus.cz/doku.php/cnk:ortofon>.

CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:ortofon:pravidla* [online]. Příručka ČNK, 2017. Dostupné z: <https://wiki.korpus.cz/doku.php/cnk:ortofon:pravidla>.

CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:pmk* [online]. Příručka ČNK, 2018. Dostupné z: <https://wiki.korpus.cz/doku.php/cnk:pmk>.

CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:schola2010* [online]. Příručka ČNK, 2017. Dostupné z: <https://wiki.korpus.cz/doku.php/cnk:schola2010>.

CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:speeches* [online]. Příručka ČNK, 2015. Dostupné z: <https://wiki.korpus.cz/doku.php/cnk:speeches>.

CVRČEK, Václav a Olga RICHTEROVÁ. *Cnk:zasady\_pro\_prepis\_vyuovacich\_hodin* [online]. Příručka ČNK, 2015. Dostupné z: [https://wiki.korpus.cz/doku.php/cnk:zasady\\_pro\\_prepis\\_vyuovacich\\_hodin](https://wiki.korpus.cz/doku.php/cnk:zasady_pro_prepis_vyuovacich_hodin).

CVRČEK, Václav a Olga RICHTEROVÁ. *Pojmy:atributy\_strukturni* [online]. Příručka ČNK, 2020. Dostupné z: [https://wiki.korpus.cz/doku.php/pojmy:atributy\\_strukturni](https://wiki.korpus.cz/doku.php/pojmy:atributy_strukturni).

CVRČEK, Václav a Olga RICHTEROVÁ. *Pojmy:mluveny* [online]. Příručka ČNK, 2020. Dostupné z: <https://wiki.korpus.cz/doku.php/pojmy:mluveny?rev=1582542589&vecdo=cite>.

*Internetová jazyková příručka*. Praha: Ústav pro jazyk český AV ČR, v. v. i., 2008–2021. Dostupné z: <https://prirucka.ujc.cas.cz/>.

Publications. *Institute for Bulgarian Language* [online]. 2014. Dostupné z: <https://ibl.bas.bg/en/struktura/balgarska-dialektologiya-i-lingvistichna-geografiya/publikatsii/>.

PŘADKOVÁ, Petra. *Slovanský jazykový atlas* [online]. Praha: Středisko společných činností AV ČR, v. v. i., 2020. ISSN 2464-6245. Dostupné z: <https://www.academia.cz/uploads/media/preview/0001/07/4c0db9646ee930e4a147114ff83bae5c5d46d766.pdf>.

Všechny online zdroje byly zkontrolovány k datu 6. 5. 2022.

## **SEZNAM PŘÍLOH**

Příloha – Seznam cvičení

## PŘÍLOHA – SEZNAM CVIČENÍ

Cvičení č. 1 – Výslovnost tvarů slovesa <i>mušet</i> se <i>š</i> v západočeské a jihočeské nářeční oblasti.....	65
Cvičení č. 2 – Široká výslovnost <i>á</i> v západočeské a jihočeské nářeční oblasti .....	71
Cvičení č. 3 – Vkladné <i>-j-</i> před zubnicemi <i>t', d'</i> v západočeské nářeční oblasti.....	74
Cvičení č. 4 – Výslovnost skupiny <i>mňe</i> jako <i>mje</i> v jihočeské nářeční oblasti .....	78
Cvičení č. 5 – Zavřená výslovnost dvojhlásek <i>ej</i> a <i>ou</i> na Doudlebsku a Domažlicku ...	81
Cvičení č. 6 – Náslovné <i>h-</i> na Domažlicku .....	85
Cvičení č. 7 – Protetické <i>v-</i> před <i>o-</i> na Doudlebsku .....	88
Cvičení č. 8 – Rozložená výslovnost měkkých retnic <i>p, b, m, v</i> na Doudlebsku.....	91
Cvičení č. 9 – Koncovka <i>-í</i> v 1. pádu plurálu životných maskulin v západočeské a jihočeské nářeční oblasti .....	94
Cvičení č. 10 – Změna skupiny <i>-ovi</i> na <i>-oji</i> v západočeské a jihočeské nářeční oblasti	97
Cvičení č. 11 – Nesklonný tvar rodinných jmen na <i>-oje</i> v západočeské a jihočeské nářeční oblasti.....	100
Cvičení č. 12 – Tvar přivlastňovacího zájmena na <i>-ovo</i> v západočeské a jihočeské nářeční oblasti.....	103
Cvičení č. 13 – Slovní spojení <i>f tej</i> a jeho ekvivalenty v západočeské a jihočeské nářeční oblasti.....	106
Cvičení č. 14 – Koncovka <i>-om</i> ve 3. pádu plurálu substantiv všech rodů v západočeské nářeční oblasti .....	110
Cvičení č. 15 – Odsouvání koncového <i>-k</i> u slovotvorného formantu <i>-pak</i> v západočeské nářeční oblasti .....	115
Cvičení č. 16 – Tvary zájmena <i>já</i> na Doudlebsku .....	117
Cvičení č. 17 – Sloveso <i>vrtět</i> v západočeské a jihočeské nářeční oblasti.....	120
Cvičení č. 18 – <i>Vosrkejchle</i> a <i>maltošna</i> v západočeské nářeční oblasti .....	122
Cvičení č. 19 – Podstatné jméno <i>buchty</i> v jihočeské nářeční oblasti.....	124
Cvičení č. 20 – Svátek sv. Štěpána v západočeské a jihočeské nářeční oblasti.....	127
Cvičení č. 21 – <i>Šerka</i> v západočeské nářeční oblasti .....	129



Cvičení č. 22 – Draní peří v jihočeské nářeční oblasti .....	131
Cvičení č. 23 – Nářeční ekvivalenty slova <i>kozel</i> v západočeské a jihočeské nářeční oblasti .....	134
Cvičení č. 24 – Výraz <i>veřtat</i> v západočeské nářeční oblasti .....	137