

**Univerzita Palackého v Olomouci**

**Filozofická fakulta**

**Katedra bohemistiky**

**Česká filologie se zaměřením na editorskou práci ve sdělovacích  
prostředcích**

**Korpusy a jejich využití ve výuce českého jazyka  
na ZŠ A SŠ**

**Bakalářská práce**

**Olomouc 2015**

Vedoucí diplomové práce:

**PhDr. Petr Pořízka, Ph. D.**

Vypracovala:

**Kristina Záchová**

Prohlašuji, že jsem bakalářskou diplomovou práci vypracovala samostatně  
a v seznamu literatury jsem uvedla veškeré informační zdroje, které jsem použila.

V Olomouci dne 15. 4. 2015

.....

Na tomto místě bych ráda poděkovala vedoucímu své bakalářské práce PhDr. Petru Pořízkovi, Ph. D., za odborné konzultace, připomínky a čas, který věnoval mé bakalářské práci.

## Obsah

ÚVOD .....	6
I. TEORETICKÁ ČÁST .....	7
1. Český národní korpus (ČNK) .....	7
1.1 Typologie korpusů ČNK .....	7
1.2 Reprezentativnost korpusu ČNK .....	10
1.3 Možnosti práce s ČNK .....	10
2. Další korpusy mimo ČNK .....	11
2.1 Korpus DIALOG .....	11
2.2 Česká elektronická knihovna .....	11
2.3 Korpus českého verše .....	12
2.4 Vokabulář webový .....	13
2.5 Pražský závislostní korpus .....	14
3. Základní pojmy a postupy při práci s korpusy .....	15
3.1 Základní práce s korpusy .....	15
3.2 Pokročilejší práce s CQL dotazy .....	18
4. Výukové metody a stanoviska .....	19
II. PRAKTICKÁ ČÁST .....	20
5. Přehled manažerů ČNK a jejich využití .....	20
5.1 Kontext .....	20
5.2 Park .....	20
5.3 SyD .....	21
5.4 Morfio .....	22
5.5 KWords .....	24
5.6 Bonito .....	27
5.7 NoSketch Engine .....	27
6. Úkoly pro práci v kontextu .....	28

6.1 Úkoly určené pro základní školy .....	28
6.2 Úkoly určené pro střední školy .....	30
6.3 KonText jako pomůcka k jazykovým hrám pro učitele ZŠ .....	37
7. Cvičení se SyD (vyhledávání dubletních a variantních tvarů) .....	41
8. Využití ČEK a KČV .....	49
8.1 ČEK .....	49
8.1.1 Cvičení pro ČEK: .....	51
8.2 KČV = Nástroje pro práci a vytěžování jazykových dat z ČEK .....	55
8.2.1 Gunstick .....	55
8.2.2 Cvičebnice .....	56
8.2.3 Hex .....	56
8.2.4 Hry v KČV .....	58
9. Vokabulář webový .....	60
ZÁVĚR .....	64
ANOTACE .....	65
ZDROJE: .....	66
Bibliografie: .....	66
Softwarové nástroje: .....	68
PŘÍLOHY: .....	69
Příloha č. 1: Přehled regulárních výrazů pro CQL .....	69
Příloha č. 2 – Poziční tagset (České poziční morfologické značky) .....	71
Příloha č. 3: Vyhledávací vzory pro KonText a Bonito .....	73
Příloha č. 4 – Tahák pro studenty .....	75

## ÚVOD

Za téma své bakalářské práce jsem si vybrala Korpusy a jejich využití na základních a středních školách. Budu zde vycházet z jazykovědného oboru korpusová lingvistika, který úzce souvisí s velmi rychlým a intenzivním rozvojem informačních technologií.

Korpusová lingvistika v českém prostředí je datována od r. 1994, kdy na Filozofické fakultě Univerzity Karlovy v Praze vzniká multidisciplinární a s mnoha institucemi kooperující Ústav Českého národního korpusu pod vedením Františka Čermáka. Český národní korpus (ČNK) se poté stal velmi ceněným v evropském měřítku.<sup>1</sup> ČNK poskytuje mnohá využití a uplatnění nejen v jazykovědě a lingvistických výzkumech, ale i v dalších humanitních oborech (psychologie, sociologie). Je také databankou pro historiky, literární vědce, publicisty a žurnalisty. Korpus uplatňují překladatelé a v neposlední řadě pedagogové pro demonstraci lingvistických jevů.

S rozvojem počítačové gramotnosti již od útlého věku dětí se čím dál více nabízí alternativní metody výuky na školách s využitím moderních technologií. Budu se ve své bakalářské práci snažit naznačit možné příklady využití, jak s korpusovými programy a manažery pracovat, jak je zapojit k výkladu vyučujícího a jak skrze ně zkoušet znalosti žáků a studentů.

V teoretické části se zaměřím vysvětlení základních pojmů, historii a vývoj korpusové lingvistiky v ČR a také klasifikaci korpusů. Jednu kapitolu věnuji i výuce českého jazyka na školách a poukážu na nutnost rozvíjení jazykových schopností v mateřském jazyce. V praktické části budou k jednotlivým manažerům a programům (nejen pro vytěžování dat z ČNK) návrhy cvičení, úkolů a her pro základoškolskou a středoškolskou výuku češtiny.

---

<sup>1</sup> ČERMÁK, František. Korpusová lingvistika: stručný historický přehled. [online].

# I. TEORETICKÁ ČÁST

## 1. Český národní korpus (ČNK)

Jazykový korpus je dle internetové Příručky ČNK definován jako „*rozsáhlý soubor autentických textů (psaných nebo mluvených) převedený do elektronické podoby v jednotném formátu tak, aby v něm bylo možné jednoduše vyhledávat jazykové jevy, zejména slova a slovní spojení (kolokace). Od obyčejného textového archívu či databáze jej odlišuje především skutečnost, že bývá pečlivě sestaven za konkrétním, nejčastěji lingvistickým účelem (např. má reprezentovat současný mluvený či psaný jazyk nebo jeho část, např. publicistické texty). Korpus zobrazuje jazykové jevy v jejich přirozeném kontextu, a umožňuje tak vytvářet na reálných datech podložený jazykový výzkum v rozsahu, který byl dříve nemyslitelný.*“<sup>2</sup>

ČNK je blíže popsán tamtéž: „*Projekt Českého národního korpusu už od roku 1994 zajišťuje přístup k více než dvěma miliardám slov uspořádaných v korpusech synchronních i diachronních, mluvených i psaných, paralelních i jednojazyčných. Nejde ovšem jen o rozsah, ale především o kvalitu zpracování a uživatelské pohodlí při jejich vytěžování.*“ Cílem jazykových korpusů je mapování současné jazykové situace jejího komplexního synchronního i diachronního lingvistického zkoumání.

Korpusy ČNK mají různá zaměření či témata. Mohou být synchronní či diachronní, psané či mluvené, publicistické, korespondenční, tematicky zaměřené atd. (viz následující kapitolu).

### 1.1 Typologie korpusů ČNK

Jednotlivé jazykové korpusy lze definovat na základě jejich vlastností.

Primárně lze korpusy dělit na referenční (tedy korpusy neměnné, ke kterým lze odkazovat, a jsou replikovatelné. Příkladem zde může být SYN2000 aj.) a nereferenční (například SYN, který sdružuje všechny referenční korpusy, zaznamenává jejich změny a vývoj – např. v morfologických značkách a lemmatizaci).

---

<sup>2</sup> Cvrček – Richterová: Příručka ČNK, 2013, online.

Dále je možné rozlišovat korpusy synchronní a diachronní. Synchronní zaznamenávají úzus užívaného jazyka v jednom úzce specifikovaném období, a to bez vývojových změn. Jazyk je zde považován za živý, protože jej využívají žijící mluvčí. U psaných textů platí, že je synchronní, pokud je čtenářem stále čten a chápán. Naopak diachronní korpusy zachycují vývoj jazyka v určitém časovém období. Jediným reprezentantem tohoto typu je v ČNK DIAKORP, který zachycuje psané texty ve vývoji sedmi století. Jako alternativu k DIAKORPu je možné uvést webovou aplikaci Vokabulář webový (viz kapitolu č. 2.4).

Korpusy ČNK lze také dále dělit stejně jako podoby živého jazyka, a to na psaný a mluvený. Psané (řada SYN a specializované) jsou nejběžněji užívanými korpusy ČNK. Zprostředkovávají reprezentativní vzorek psaných a tištěných textů, a to z oblastí beletrie, publicistiky (řada SYN PUB) či odborných textů. Korpus mluvené češtiny (řada ORAL a specializované) zachycuje běžnou podobu lidské mluvy. Autentická promluva (dialog, monolog, připravený či nepřipravený) je zaznamenán a dle přísných transkripčních pravidel přepsán, mimo zvukového záznamu zde figuruje i fonetický a ortografický přepis. Tvorba tohoto typu je časově mnohem náročnější. Orální typy jsou zastoupeny řadou ORAL, Pražským a Brněnským mluveným korpusem. Speciálním typem mluveného korpusu je korpus DIALEKT, který zachycuje nářečí ve vývoji, jedná se o diachronní typ.

Jako poslední typ je nutné uvést specializované korpusy, které nereprezentují jazykový úzus jako celek, ale zaměřují se na jeho jednu konkrétní část či specifické využití. Do této skupiny patří KSK (=Korpus soukromé korespondence, který zachycuje pravděpodobně poslední generaci tradičně ručně psaných dopisů, a to ve dvou tisících exemplářích. Najdeme zde jazyk na přelomu mezi mluvenou a psanou podobou, specifické idiolekty i spontánnost vyjádření), žakovský korpus CzeSL-plain (Czech as a Second Language, plain = bez anotace; tedy korpus psaného jazyka žáků, kteří mají češtinu jako druhý jazyk, nejsou tedy rodilými mluvčími), LINK (původně LIngvistův Narozeninový Korpus, který byl vytvořen u příležitosti životního jubilea prof. Františka Čermáka; je sestaven pouze z odborných lingvistických textů. Slouží ke zkoumání jazykových specifik akademiků a lingvistů) či ORWELL (obsahuje text jediné knihy, a to českého



překladu románu 1984 od George Orwella. Tento korpus je morfologicky označován jako MTE i AJKA.).<sup>3</sup>

Posledním specifickým typem jsou paralelní korpusy, které srovnávají více jazyků v rámci jedné metodologie. Úžeji se dělí na srovnatelné (obsahují reprezentativní vzorek typově různých textů z více jazyků a kvantitativně jsou vyvážené) a překladové (obsahují originály a překlady stejných textů, po různých úsecích jsou zarovnány do alingmentů). V ČNK se nachází od roku 2005 pouze jediný paralelní korpus, a to INTERCORP, který obsahuje texty z 27 jazyků. Je typově vyvážený, synchronní, nereferenční (stále se rozrůstá) a průběžně morfologicky značkový. Obsahuje žurnalistické, beletristické i odborné texty, stejně jako legislativní zprávy z Evropské unie.

Pro úplnost uvádím klasifikaci z internetové Příručky ČNK<sup>4</sup>:

1. Synchronní psané korpusy
  - a. řada SYN (SYN2000, SYN2005, SYN2010, nereferenční SYN) a korpusy publicistických textů PUB (SYN2006PUB, SYN2009PUB, SYN2013PUB)
  - b. specializované (CZESL-PLAIN, CZESL-SGT, JEROME, LINK, ORWELL, KSK-DOPISEY, SCRIPT2012)
2. Synchronní mluvené korpusy
  - a. řada ORAL (ORAL2006, ORAL2008, ORAL2013)
  - b. specializované (BMK, PMK, SCHOLA2010)
3. Diachronní psaný korpus (DIAL)
4. Paralelní korpusy (Intercorp)

---

<sup>3</sup> Cvrček – Richterová: Příručka ČNK, 2013, online.

<sup>4</sup> Cvrček – Richterová: Příručka ČNK, 2013, online.

## **1.2 Reprezentativnost korpusu ČNK**

Žádný jazykový korpus nemůže být všestranně reprezentativní. Do korpusů je vždy zanesen pouze reprezentativní vzorek textů ze všech odvětví, aby co nejvěrněji a nej přesněji mohly reprezentovat daný jazyk v daném časovém období. Jeho reprezentativnost však přímo úměrně nemůže narůstat s velikostí a obsahem korpusu, což je nutná hypotéza pro další práci s ním a také pro zhodnocení získaných dat.

## **1.3 Možnosti práce s ČNK**

Každý uživatel internetu má volný přístup k Českému národnímu korpusu z webového rozhraní [www.korpus.cz](http://www.korpus.cz), a to i bez předchozí registrace. Uživatelé jsou však k dispozici pouze velmi omezené funkce vyhledávání v nereferenčním korpusu SYN, proto je lepší se pro plnou práci s touto databází registrovat.

Práci s výše uvedenými korpusy je možná vykonávat pomocí různých webových aplikací, tzv. korpusových manažerů, do kterých je nutné se přihlásit pomocí přístupových údajů. Online je možné s korpusy pracovat na výše zmíněném webu, a to pomocí základního manažeru Kontext (pro vyhledávání), Park, SyD, Morfio a KWords. Mimo ČNK budu pracovat i s dalšími databázemi a programy (např. Česká elektronická knihovna či Korpus českého verše). Blíže budou aplikace a manažery popsány v praktické části, včetně jednoduchých manuálů pro práci s nimi.

## **2. Další korpusy mimo ČNK**

Jazykové korpusy a databáze textů v českém jazyce vznikají a fungují i mimo Ústav Českého národního korpusu (ÚČNK). Jejich výčet a popis se nachází v následujících podkapitolách.

### **2.1 Korpus DIALOG**

Pražský Ústav pro jazyk český Akademie věd České republiky (ÚJČ AV ČR) spravuje DIALOG, což je multimediální korpus mluvené češtiny. Shromažďuje veřejné jazykové projevy dialogického typu – nahrávky a přepisy diskusních pořadů českých veřejnoprávních i komerčních televizí (např. Všechnopárty, Show Jana Krause, Sedmička, Krásný ztráty, Na plovárně atd.). Slouží výzkumu mediální komunikace a výzkumu mluvené češtiny v její současné veřejné podobě.<sup>5</sup> Má několik vývojových verzí, od nejstaršího: DIALOG 0.1 m (10 nahrávek z r. 1999 – 2005), DIALOG 0.2 (40 nahrávek), DIALOG 0.3 (70 nahrávek), DIALOG 1.0 (100 ks) a nejnovější a nejrozsáhlejší DIALOG 1.1 (150ks). Celý korpus DIALOG je plně lemmatizován a morfologicky anotován. I zde je pro uživatele vhodná registrace. DIALOG funguje ve spolupráci s Ústavem formální a aplikované lingvistiky Matematicko-fyzikální fakulty Univerzity Karlovy (MFF UK) v Praze. Je dostupný na: <http://ujc.dialogy.cz/>

### **2.2 Česká elektronická knihovna**

Česká elektronická knihovna (ČEK) je nejrozsáhlejší plnotextovou databází tohoto druhu na českém internetu. Vytvořil ji Ústav pro českou literaturu AV ČR. Obsahuje 1700 knih původní české poezie z 19. a 20. století (přináší kompletní veršované dílo nejvýznamnějších básníků české novodobé literatury (např. K. H. Mácha, K. J. Erben, K. Havlíček Borovský, J. Neruda, J. Vrchlický, O. Březina)). Knihovna je zdarma dostupná po přihlášení na <http://www.ceska-poezie.cz/cek/>.

Na webu ČEK je možné prohlížet a pro osobní potřeby stahovat a tisknout uvedené básnické texty. Pro složitější práci s texty ČEK (vyhledávání, statistické vyhodnocování, frekvenční slovníky atd.) se využívá nástrojů Korpusu českého verše (KČV), kterým se budu zabývat v následující kapitole.

---

<sup>5</sup> Korpus DIALOG: Úvod., online.

### 2.3 Korpus českého verše

Korpus českého verše (KČV) vznikl v Ústavu pro českou literaturu Akademie věd ČR (ÚČL AV ČR) v roce 2013. Je plně lemmatizovaný, foneticky, morfologicky, metricky a stroficky anotovaný korpus české poezie 19. a počátku 20. století. Vychází z textů plnotextové databáze Česká elektronická knihovna (<http://www.ceska-poezie.cz/cek/>) a obsahuje 1689 básnických sbírek.

Pro práci s KČV je možné využít několik různých online nástrojů. **Databáze českých meter** (DČM) u textů KČV zpřístupňuje popis meter a strof. Kromě bibliografických údajů lze zde vyhledávat a zpracovávat data podle veršového rozměru, metrického vzorce, rýmovanosti a strofičnosti textu, strofického a rýmového schématu, ustáleného názvu rozměru či pevných forem poezie.<sup>6</sup> DČM je volně přístupná na: <http://metro.ucl.cas.cz/kveta/>.

Webová aplikace **Gunstick** (=databáze českých rýmů; přístupná online na <http://www.versologie.cz/gunstick/>) generuje rýmové skupiny (dvojice a větší) pro zadané slovo (nikoli lemma). Rýmy jsou opět generovány z Korpusu českého verše (do r. 1920). Databáze obsahuje přes milion rýmových párů.<sup>7</sup> Jejich hravé využití předvedu v kapitole 7.2.1.

Pro registraci do KČV uživatel také získá přístup do interaktivní online **Cvičebnice** (dostupná na <http://www.versologie.cz/cvicebnice/>), která názorně doplňuje publikaci Úvod do teorie verše (Praha: Akropolis, 2013).<sup>8</sup> Cvičení zde jsou doplněna o výkladové pasáže, které středoškolské studenty informují o metrech, strofách a dalších versologických kategoriích. Praktické využití cvičebnice demonstruji v kapitole 7.2.2.

Další nástroje pro vytěžování versologických dat z KČV jsou **HEX – klíčová slova v české poezii**<sup>9</sup> (pro hledání klíčových slov u daného autora, básně či sbírky, nebo v daném období atd., <http://www.versologie.cz/hex/>, viz kapitolu 7.2.3),

---

<sup>6</sup> PLECHÁČ, P. – IBRAHIM, R.: **Databáze českých meter**, Praha 2013, online.

<sup>7</sup> PLECHÁČ, P. – IBRAHIM, R.: **Gunstick – databáze českých rýmů**, Praha 2013, online.

<sup>8</sup> PLECHÁČ, P. – IBRAHIM, R. – ŘÍHA, J.: **Úvod do teorie verše – cvičebnice**, Praha 2013, online.

<sup>9</sup> PLECHÁČ, P.: **Hex – klíčová slova v české poezii**, Praha 2013, online.

**Frekvenční slovníky české poezie**<sup>10</sup> (udávají frekvenci lemmat a tokenů (slovních tvarů) ve sbírkách, autorských subkorpusech i celém KČV, online na: <http://www.versologie.cz/slovníky/slovníky.html>) a **Eufonometr**<sup>11</sup> (kvalifikuje míru libozvučnosti libovolného zapsaného textu, tzv. eufonický koeficient, <http://www.versologie.cz/eufonometr/eufonometr.html>). Tyto aplikace jsou však příliš specifické a dále se jimi nebudu ve své práci zabývat.

## 2.4 Vokabulář webový

Diachronní korpus Vokabulář webový (online: [www.vokabular.ujc.cas.cz/](http://www.vokabular.ujc.cas.cz/)) široké veřejnosti zprostředkovává Ústav pro jazyk český (ÚJČ AV ČR) v Praze. Jedná se o bezplatnou internetovou databázi starých českých textů, nahrávek a obrazových materiálů k poznání historické češtiny. V současné době jej tvoří tyto lexikální zdroje: Malý staročeský slovník (MSS), pracovní heslář k lístkové kartotéce Staročeského slovníku (HesStčS), Slovník staročeský Jana Gebauera (GbSlov), Staročeský slovník, sešit 1–26 (StčS), Slovníček staré češtiny F. Šimka (ŠimekSlov), Index Slovníku staročeských osobních jmen Jana Svobody (IndexSvob), Elektronický slovník staré češtiny (ESSČ) a digitalizovaná lístková kartotéka Jana Gebauera.

Vokabulář umožňuje uživateli jednoduché vyhledávací rozhraní (záložka Vyhledávání) a plnou textovou databanku, čímž se blíží korpusovým metodám. Rozdílem může být metoda tzv. listování (záložka Listování), kdy uživatel prochází slovníkové zdroje Vokabuláře po listech. Výhodou jsou také audionahrávky. Do této databáze není nutná předchozí registrace.

Modul digitalizovaných mluvnic a obdobných příruček z 16. – 19. století (dostupné na: <http://vokabular.ujc.cas.cz/moduly/mluvnice/>) nabízí uživateli pohled na historické pojetí gramatiky. Slouží k nejen k studiu českého jazykového systému z diachronního (tj. vývojového) pohledu, ale taktéž k bádání o vývoji české gramatiky. V modulu odborné literatury je postupně zveřejňována Historická mluvnice jazyka českého Jana Gebauera.<sup>12</sup>

---

<sup>10</sup> PLECHÁČ, P.: *Frekvenční slovníky české poezie*, Praha 2013, online.

<sup>11</sup> PLECHÁČ, P.: *Eufonometr*, Praha 2013, online.

<sup>12</sup> ČERNÁ – LEHEČKA. *Vokabulář Webový*, online.

## 2.5 Pražský závislostní korpus

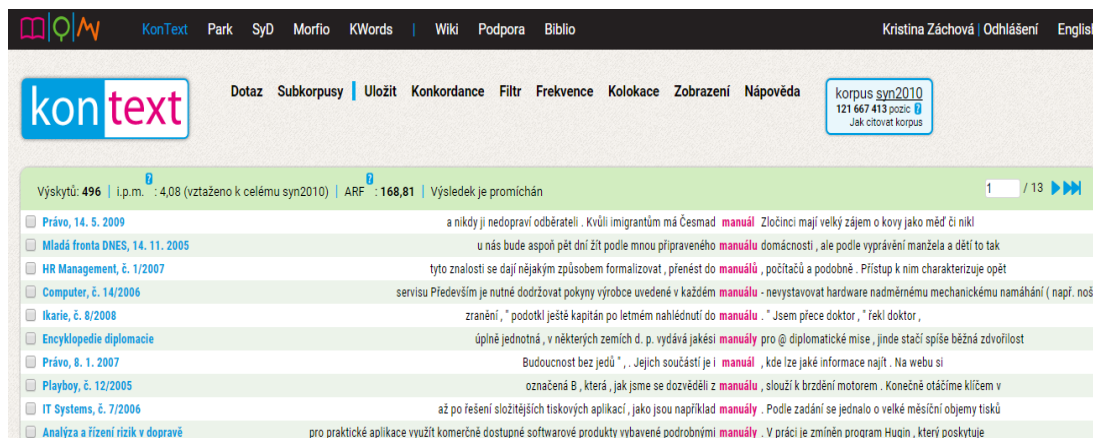
Pražský závislostní korpus (PDT – z angl. Prague Dependent Treebank) vzniká v Ústavu formální a aplikované lingvistiky (ÚFAL) v Praze ve spolupráci s Matematicko-fyzikální fakultou Univerzity Karlovy (MFF UK). Slouží k výzkumům především počítačové lingvistiky. Tento typ databáze vychází z myšlenek Pražské lingvistické školy a zaměřuje se na detailní anotace každého slova v textech. Provádí značkování morfologické, syntaktické (povrchové i hloubkové, vč. aktuálního členění) i sémantické/pragmatické, ale to velmi precizně – ručně. Zpracovává velké množství lingvisticky bohatých textů ze čtyř druhů tiskovin (Mladá fronta DNES, Lidové noviny, Česko-moravský profit a časopis Vesmír), nyní je v korpusu okolo dvou milionů slov.

Pražský závislostní korpus (verze PDT 3.0) je dostupný ke stažení na <https://lindat.mff.cuni.cz/repository/xmlui/handle/11858/00-097C-0000-0023-1AAF-3>, manuál a popis najdete online česky i anglicky zde: <http://ufal.mff.cuni.cz/pdt2.0/index-cz.html>.

### 3. Základní pojmy a postupy při práci s korpusem

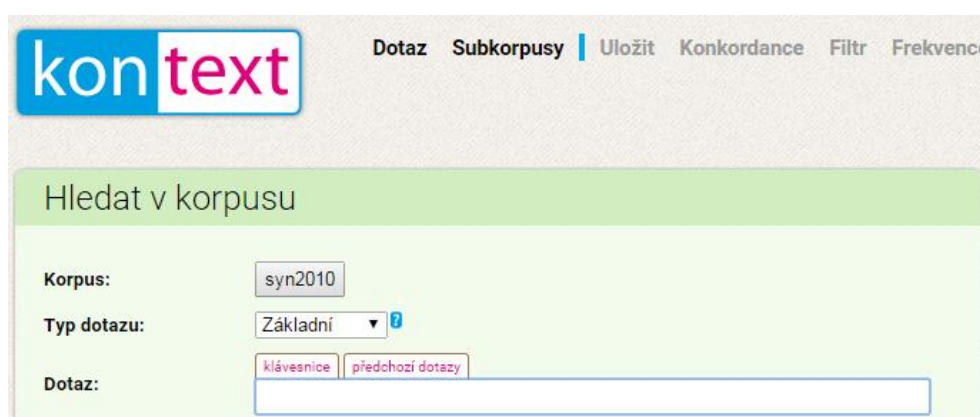
#### 3.1 Základní práce s korpusem

Pro práci s korpusem je třeba si osvojit základní pojmy z korpusevé lingvistiky. Pracuje-li se s psanými korpusem ČNK, je třeba uvážit, že jsou v nich obsažené psané texty z beletrie, odborných textů a žurnalistiky.



Obr. 1: Webové rozhraní KonText

Při využívání webového rozhraní KonText se vybere korpuse, se kterým chce uživatel pracovat. Následně zvolí, zda chce pracovat se základním vyhledáváním, pozičními atributy (lemma, slovní tvar či fráze) či s dotazy v jazyce CQL (Corpus Query Language). Poziční atribut lemma vyhledává slova podle slovníkových tvarů (infinitiv u sloves/nominativ maskulin u substantiv), atribut slovní tvar vyhledává pouze konkrétní formu (např. i frazeologismů). Potom se do příkazového řádku napíše požadovaný dotaz (slovo či CQL kód) a zadá vyhledávat.



Obr. 2: Vyhledávání v KonTextu

Výsledkem vyhledávání jsou konkordance = konkordanční seznamy, ty představují všechny výskyty vyhledaného slova. Vyhledávané slovo se označuje jako KWIC (=keyword in context), protože na konkordančních řádcích je obklopeno levým a pravým kontextem.

Např. Základním vyhledáváním je možné zobrazit substantivum *manuál*. Výsledek se zobrazí na konkordančním seznamu následujícím způsobem.

a nikdy ji nedopraví odběrateli . Kvůli imigrantům má Česmad **manuál** Zločinci mají velký zájem o kovy jako měď či nikl u nás bude aspoň pět dní žít podle mnou připraveného **manuálu** domácnosti , ale podle vyprávění manžela a dětí to tak tyto znalosti se dají nějakým způsobem formalizovat , přenést do **manuálů** , počítačů a podobně . Přístup k nim charakterizuje opět servisu Především je nutné dodržovat pokyny výrobce uvedené v každém **manuálu** - nevystavovat hardware nadměrnému mechanickému narušení , " podotkl ještě kapitán po letmém nahlédnutí do **manuálu** . " Jsem přece doktor , " řekl doktor , úplně jednotná , v některých zemích d. p. vydává jakési **manuály** pro @ diplomatické mise , jinde stačí spíše běžná zdvořilá Budoucnost bez jedů " , . Jejich součástí je i **manuál** , kde lze jaké informace najít . Na webu si označená B , která , jak jsme se dozvěděli z **manuálu** , slouží k brzdění motorem . Konečně otáčíme klíčem v až po řešení složitějších tiskových aplikací , jako jsou například **manuály** . Podle zadání se jednalo o velké měsíční objemy tisků se využít komerčně dostupné softwarové produkty vybavené podrobnými **manuály** . V práci je zmíněn program Hugin , který poskytuje po smyslu se netýká jen mě . Bylo nás dobrovolných **manuálů** pár tisíc a občas se mezi námi objevila i známá starší generaci i návod , jak ušetřit RECEPT NA REFORMU **Manuál** LN pro pacienty PRAHA Nejčastěji chodí k lékařům , nejčastěji , ale s možností voleb ) , prioritou clony/času a **manuál** , dále pět nejčastěji používaných scénických režimů - port nebo zdání střetu zájmů , " stojí například v Etickém **manuálu** amerického Kongresu . Paroubkova shovívavost k Rathovi knihy , zejména odborné a popularizační , instruktážní brožury , **manuály** , přehledy ) , periodické publikace - časopisy , noviny

**Obr. 3: Konkordanční řádky zobrazující lemma *manuál***

Úplně vlevo na každém řádku najdeme zdroj textu a rok původu. V horním řádku zleva se zobrazuje celkový počet výskytů slova, jeho relativní frekvence i. p. m. (výskytů na 1 milion slov ve vybraném korpusu) a ARF (průměrná redukováná frekvence), která reflektuje rozložení slov v celém korpusu.

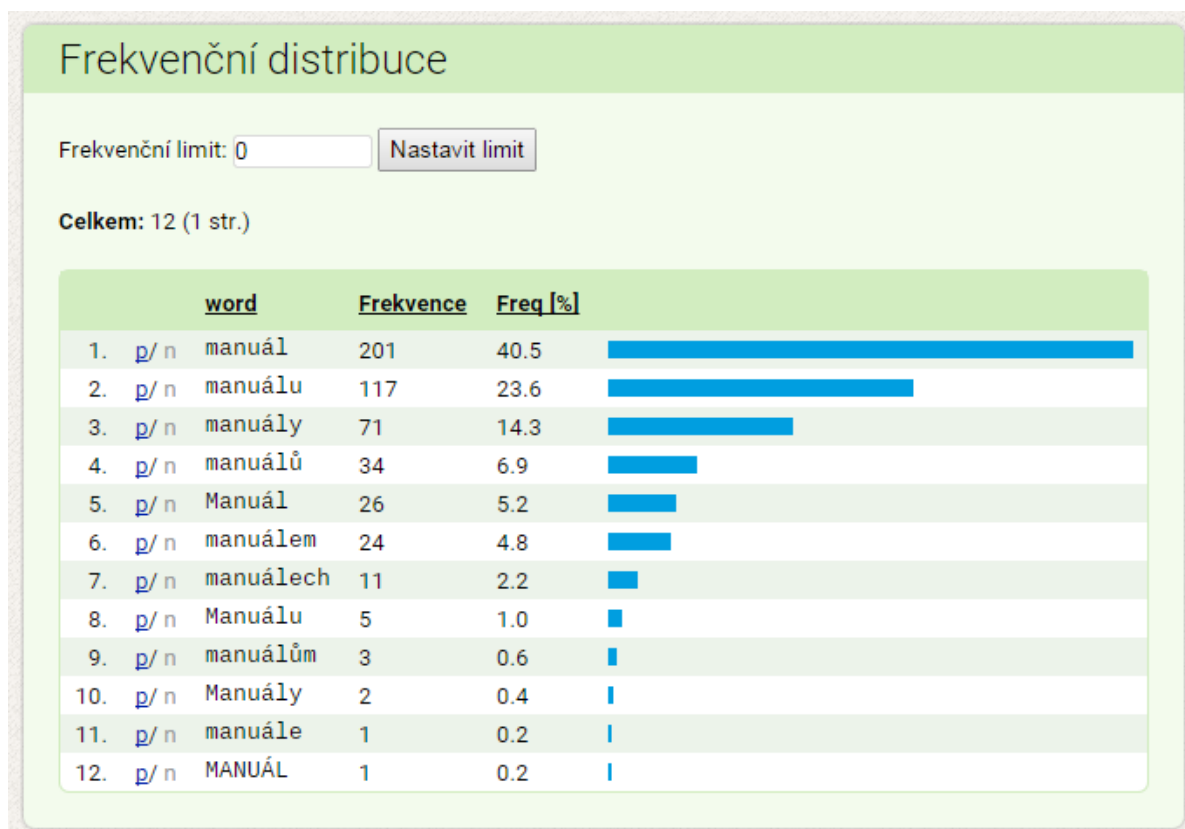
The screenshot shows the 'kon text' search engine interface. At the top, there is a search bar and navigation tabs: 'Dotaz', 'Subkorpusy', 'Uložit', 'Konkordance', 'Filtr', and 'Frekvence'. Below the tabs, there are options for file formats: 'CSV', 'XLSX', 'XML', 'TXT', and 'Vlastní...'. At the bottom, a green summary bar displays the following statistics: 'Výskytů: 186', 'i.p.m.: 1,53 (vztaženo k celému syn2010)', 'ARF: 63,24', and 'Výsledek je promíchán'.

**Obr. 4: Řádek zobrazující výskyty a frekvence**

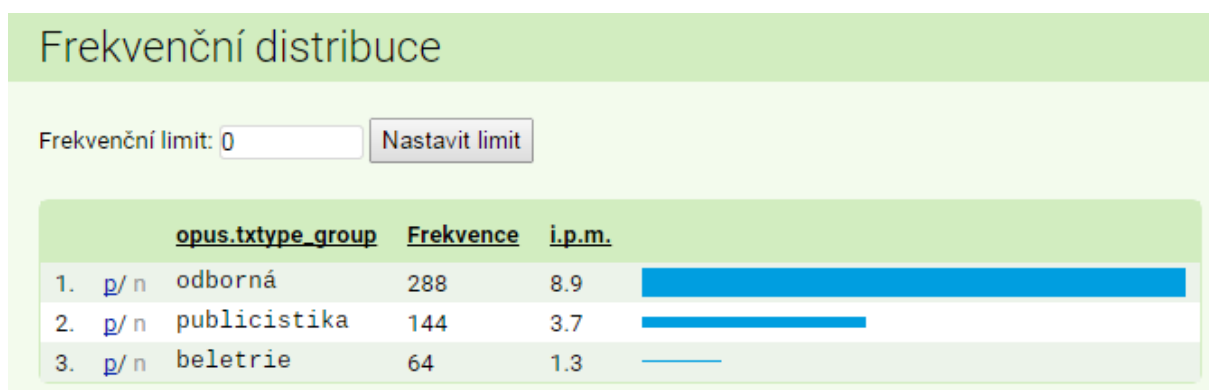
Pro detailnější zkoumání vyhledaných dat se využívá funkce Frekvence (horní záložka), která zobrazí frekvenci a počty výskytů KWIC v podobě jednotlivých lemmat, slovních druhů, typů textů a dokumentů. Frekvence lemmat zobrazí, jaká



lemmata se v konkordanci nacházejí a v jakém poměru, slovní tvary ukážou poměr a výskyt všech tvarů lemmat (všechny pády, rody, osoby, čísla apod.). Frekvence typů textů zobrazí procentuální zastoupení výskytů v beletrii, žurnalistice a odborných textů, dále v dílčích žánrech a dokumentech. Frekvenci dokumentů lze zobrazit i zvlášť.



Obr. 5: Frekvenční distribuce slovních tvarů lemmatu *manuál*



Obr. 6: Frekvenční distribuce lemmatu *manuál* v typech textů

### 3.2 Pokročilejší práce s CQL dotazy

Pro pokročilejší práci s korpusovými daty se využívá dotazů CQL. Uživatel si ve vyhledávacím formuláři vybere konkrétní korpus a typ dotazu – CQL. Dotazy poté zadává ve tvaru **[atribut="hodnota"]**. Mnohé dotazy je nutné doplnit o tzv. regulární výrazy, což jsou speciální zástupné znaky (např. tečka, hvězdička, různé typy závorek, lomítka atd.), jejichž seznam a klávesové zkratky uvádím v příloze č. 1.

*Např.:* Pro vyhledání lemmatu *brýle* lze využít dotaz [lemma="brýle"]. Pro hledání tvaru substantiva *brejle* se hodí dotaz [word="brejle"], jeho celé paradigma se zobrazí dotazem [word="brejl.\*"].

Na pozici atribut můžeme dosadit lemma, word (tj. slovní tvar), pos (tj. slovní tvar, z angl. *part of speech*) či tag (vyjádřený morfologickými značkami). Použije-li uživatel atributu tag, dosazuje na konkrétní pozice hledané morfologické značky (mluvnické kategorie – rod, číslo, pád, osoba, čas atd.), viz tabulku v příloze č. 2.

*Např.:* Pro nalezení všech substantiv, feminim v plurálu, v genitivu či akuzativu je možné použít následující dotaz: **[tag="N.FP(2|4).\*"]**. Pozice 5 (pád) má 2 možnosti, proto je oddělen závorkou a svíslítkem (viz přílohu č. 1 – Regulární výrazy). Jelikož jsou zbylé pozice blíže nespecifikované, dotaz je zakončen výrazem **.\***. Jeden nespecifikovaný údaj se nahrazuje tečkou.

Detailní postupy práce a příklady jsou rozepsány v Kurzu práce s ČNK na Wiki (=webové příručce): <http://wiki.korpus.cz/doku.php/kurz:uvod>. Mimo Kurz práce jsou zde také vypsány a názorně vysvětleny veškeré pojmy korpusové lingvistiky, historie a typologie korpusů ČNK atd. Já jsem zde uvedla pouze základní kroky práce s rozhraním KonText a ty typy dotazů, které budu využívat ve cvičeních v praktické části. Webová příručka pro práci s korpusy ČNK je online na adrese: <http://wiki.korpus.cz/doku.php/start>.

#### 4. Výukové metody a stanoviska

Ve výzkumu PISA, který u patnáctiletých žáků zjišťoval mimo jiné čtenářskou gramotnost, bylo zjištěno, že čeští žáci mají méně hodin českého jazyka než v jiných evropských zemích (Učitelství 02/2002) a jejich čtenářská gramotnost patří k nejslabším. Tento výzkum také konstatuje, že žáci v České republice nejsou vedeni ke kritickému hodnocení textů různých druhů, k zaujímání stanovisek a názorů, ani k samostatnému vyhledávání informací. Tyto klíčové dovednosti pro život by měly být u dětí cíleně rozvíjeny ve školách, především v hodinách mateřského jazyka.

V učitelství by to mělo znamenat, že bychom u dětí měli rozvíjet jazykové dovednosti, slovní zásobu i jejich využití v komunikaci a textech. Není nutné, aby se děti pouze učily definice a poučky, je vhodné je vést k poznání hrami, zajímavými cvičeními a využitím v praxi.<sup>13</sup>

Ve své bakalářské práci nabízím spoustu cvičení, her a možného využití českého jazyka v praxi. Dále poukazuji na to, jak jdou mnohé online nástroje a databáze využívat ve výuce češtiny, literatury i v praxi. Hravou formou jde rozvíjet slovní zásobu, opakovat jazykové jevy i zapojovat fantazii. Starší studenti na středních školách zas jistě ocení možnost alternativní výuky na názorných textech, práci se staršími texty a praktickou ukázkou toho, na kterou webovou stránku je možné se obrátit, pokud si nejsou jisti nějakým jazykovým jevem, pokud hledají nějaké konkrétní texty či využití některých slov a jejich zobrazení v kontextu.

U Korpusu českého verše se nabízí hravé formy s tvorbou vlastních básniček. Žáci i studenti také jistě využijí obsáhlou Českou elektronickou knihovnu, která nabízí k vlastní četbě sbírky "českých klasiků".

U většiny cvičení a her si žáci i studenti vyzkouší práci s texty, slovníky, internetovými databázemi, základy dotazovacího jazyka CQL, vyhledávání informací v online manuálech k programům apod. Některá cvičení jsou opřena i o dílčí práci s učebnicí, Pravidly českého pravopisu, mluvnicí či slovníkem, čímž se všestranně rozvíjí jejich schopnost vyhledávání informací s více zdroji.

---

<sup>13</sup> KOTEN, Hněvín, 2006, str. 36

## II. PRAKTICKÁ ČÁST

### 5. Přehled manažerů ČNK a jejich využití

Pro práci s korpusem ČNK je možné využívat několik různých softwarových nástrojů, takzvaných korpusových manažerů a vyhledávacích rozhraní. Na novém webu [www.korpus.cz](http://www.korpus.cz) (který funguje od r. 2012) najdeme několik rozhraní, která jsou uživatelsky přívětivější a přehlednější než předchozí verze manažerů.

#### 5.1 Kontext

Základní verzí je online vyhledávač dat z ČNK KonText, který je vyšší verzí manažeru Bonito a NoSketch Engine. Dá se v něm vyhledávat komplexně pomocí CQL (Corpus query language) ale i pomocí jednoduchých předdefinovaných dotazů a zadáním pozičních či strukturních atributů přímo z nabídky. Následně zobrazuje hledané výrazy v konkordančních řádcích (okolním kontextu). Následně lze výraz vyhodnotit, např. zjistit jeho frekvenční distribuci, asociační míry či další metajazyková data. Základní tabulka pozičních atributů pro práci s korpusem je v příloze mé práce. KonText je dostupný na [www.kontext.korpus.cz](http://www.kontext.korpus.cz).

KonText je základním online nástrojem pro vytěžování jazykových dat z Českého národního korpusu. Popis práce jsem uvedla v kapitole č. 3 – Základní pojmy a postup práce. Můžeme skrze něj vyhledávat jednotlivé výrazy i celá paradigmatata, syntaktické struktury, výrazy dle morfologických značek (tzv. tagů), porovnávat frekvence i rozebírat jazyk jednotlivých médií či beletristických děl.

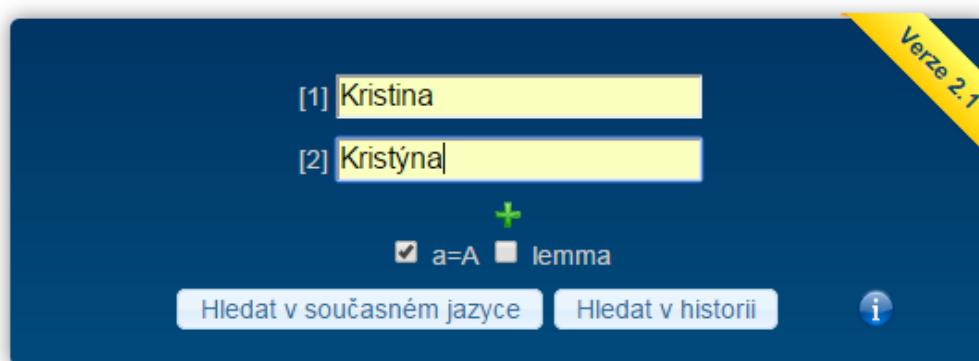
Tomuto nástroji se budu detailně věnovat také v praktické části, kde předvedu mnoho cvičení, úkolů a her na jeho využívání ve výuce.

#### 5.2 Park

Zkratka Park označuje starší rozhraní pro vstup do paralelního korpusu Intercorp. Ten obsahuje množství českých textů s překlady do jiných jazyků. Oba typy jsou vzájemně zarovnány do alingmentů pro snazší orientaci uživatele. Slouží k výuce cizích jazyků, k teoretickým studiím i pro překladatele. Jelikož jde o zastarávající program, v březnu 2015 bude jeho funkčnost ukončena. Park je online: <https://park.korpus.cz/>.

### 5.3 SyD

Nástroj SyD slouží ke srovnávání dvou a více slovních variant či spisovných dubletních tvarů. Tento všestranný průzkum lze provádět ze synchronního (v psaném i mluveném korpusu) i diachronního (v psaném a mluveném korpusu) pohledu. Dostupný je na: <https://syd.korpus.cz/>.



**Obr. 7: Internetové rozhraní SyD pro srovnání variant**

Práce s aplikací SyD je jednoduchá a není zde nutná registrace. Vyhledávací rozhraní má dvě textová pole, kam uživatel zadá varianty k porovnání. Zeleným plusem je možné přidat více polí pro varianty, červeným křížkem je lze zpětně odebrat. Vhodné je ponechat možnost  $a=A$ , která porovná varianty s různým zápisem velikostí písmen. Hledáme-li podle základního tvaru slova celé jeho paradigma, je nutné zatrhnout možnost *lemma*. Bez jejího označení hledáme pouze jeden konkrétní slovní tvar, tj. totéž co u KonTextu poziční atribut *word*.

*Např.:* V aplikaci srovnám dvě dubletní tvary svého křestního jména, a sice *Kristina* a *Kristýna*. Při hledání typu *word* (bez označení *lemmatu*) je poměr variant cca 38 % a 62 %. Zadám-li stejné srovnání s atributem *lemma*, procenta se relativně vyrovnají na 44 % a 56 %.

## 5.4 Morfio

Aplikace Morfio využívá pouze psané korpusy ČNK, a to k vyhledávání morfologických vztahů mezi jednotkami korpusu. Vybírá z textů dvojice slov, které vznikají stejným slovtvorným postupem, a vyhodnocuje jejich morfologickou produktivitu. Aplikace Morfio se nachází na <https://morfio.korpus.cz/>.

Morfio srovnává jeden a více slovtvorných postupů ve slovních druzích. Uvedu na příkladu.

# Morfio

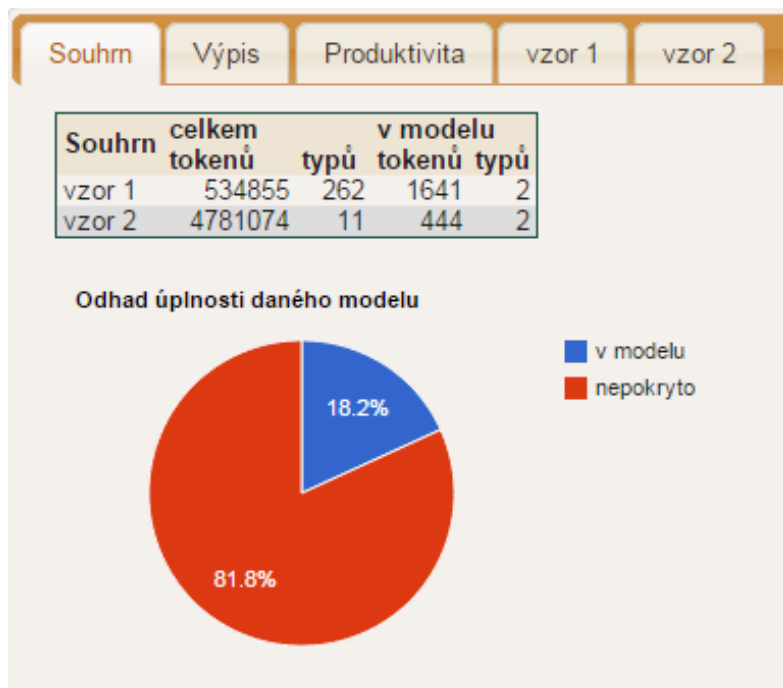
The screenshot shows the Morfio web interface with the following settings:

- Category: **společný**
- Morphological specifications: **podstatná jména** (N.\*), **slovesa** (V.\*)
- Patterns: **vzor 1:** [:Lab:][:Vok:][:Kons:], **vzor 2:** [:Lab:][:Vok:][:Kons:]
- Additional pattern: **Další vzor**
- Corpus: **SYN2010**, Frequency: **10**, Search for: **lemmata**, Evaluate: **lemmata**
- Letter size:  **ignorovat**
- Alternates: **Alternace**
- Buttons: **Hledat**, **Nové zadání**,  **zachovat záznam procesu**, **Nápověda**
- Link: **Odkaz na toto zadání: <http://morfio.korpus.cz/6TnyW0Jg>**

**Obr. 8: Internetové rozhraní Morfio**

*Např.:* Uživatel zadá kritéria minimálně pro dva vzory. Já jsem zvolila společný rys labiálu, libovolný vokál a libovolný konsonant, ze slovních druhů jsem vybrala substantiva a verba. Zbylá kritéria jsem ponechala v původním nastavení. V tomto vyhledávání je ke srovnání velký počet výskytů, i když substantiv je typově mnohonásobně více. Morfio našlo takto slovtvorně utvořených 262 lemmat substantiv o 534.855 výskytech (např. *pán, bál, pes, fáč* atd.), zatímco u verb je pouze 11 lemmat, ale 4.781.074 výskytů (např. *být, mít, pět, vít...*).

Výsledky je možné zobrazit v tabulce a grafu v záložce *Souhrn*, v záložkách *vzor 1* a *vzor 2* lze také zobrazit seznam všech vyhledaných lemmat včetně jejich výskytů.



**Obr. 9: Grafické zobrazení výsledků v Morfiu**

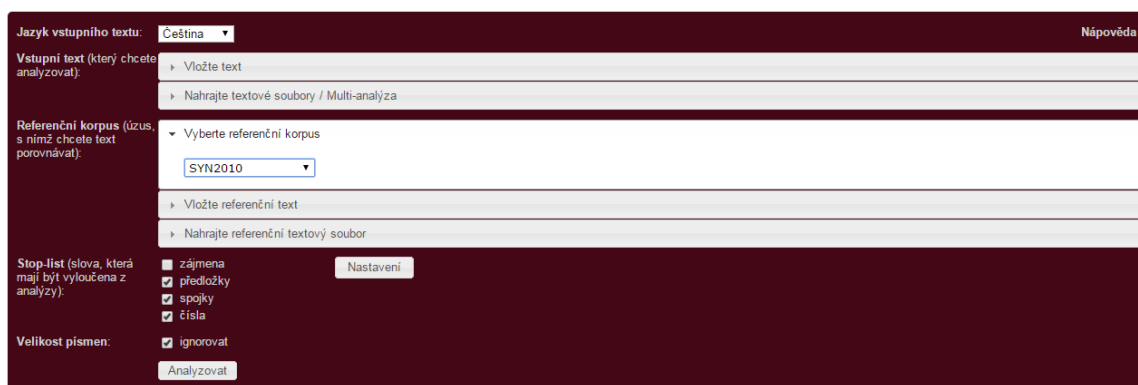
Z tohoto srovnání můžu udělat výsledek, že třípísmenná substantiva utvořená touto kombinací fonémů jsou typově četnější, takto slovtvorně vzniklá slovesa jsou typově chudší, zato jejich výskyt je velmi hojný. Důvodem osminásobně vyššího výskytu mohou být dva faktory – obecně vyšší četnost sloves v celkovém jazykovém úzu a také výskyt sloves z jazykového centra (mít, být).

Aplikace Morfio mi přijde příliš odborná na úroveň základních a středních škol, protože na tomto stupni školství jsou znalosti morfologie (a cizích termínů) velmi nízké. Pro tuto aplikaci tedy v praktické části nejsou žádná cvičení.

## 5.5 KWords

Program KWords vyhledává v námi zadaném textu klíčová slova (keywords), a to v komparaci s některým z referenčních korpusů ČNK (např. řady SYN, totalita aj.). Výsledkem procesu je tedy nejen soupis na klíčových slov v textu, ale také jejich atypičnost ve srovnání s dobovým územ z daného roku referenčního korpusu. Program KWords je volně dostupný na <https://kwords.korpus.cz/>.

**IKI** words



The screenshot shows the KWords web interface. At the top, there is a language dropdown menu set to 'Čeština' and a 'Nápověda' link. Below this, there are two main sections: 'Vstupní text (který chcete analyzovat):' and 'Referenční korpus (úzus, s nímž chcete text porovnávat):'. The first section has two input fields: 'Vložte text' and 'Nahrajte textové soubory / Multi-analýza'. The second section has a dropdown menu for 'Vyberte referenční korpus' with 'SYN2010' selected, and two more input fields: 'Vložte referenční text' and 'Nahrajte referenční textový soubor'. At the bottom, there is a 'Stop-list (slova, která mají být vyloučena z analýzy):' section with checkboxes for 'zájmena', 'předložky', 'spojky', and 'čísla'. There is also a 'Nastavení' button. Below that, there is a 'Velikost písmen:' section with a checked checkbox for 'ignorovat' and an 'Analyzovat' button.

**Obr. 10: Internetové rozhraní KWords pro vyhledávání klíčových slov v textech**

Do programu se zkopíruje text či nahraje textový dokument v češtině či angličtině. Dále si uživatel zvolí referenční korpus (případně nahraje vlastní referenční text), se kterým se prvotní text srovná a vyhodnotí klíčová slova. Dále je možné vyloučit z analýzy neflexivní slovní tvary (prepozita a konjunkce), čísla, případně pronomina.

KWords je vhodný program například na diplomové či seminární práce, které mnozí studenti odevzdávají již k maturitní zkoušce. Rozhodně by bylo vhodné tento program zmínit při tvorbě výstupních prací na středních školách.

Druhým možným využitím ve školství může být srovnání současného textu s historickým textem. Žáci mohou dostat úkol najít si nějaký současný text (z novin, časopisu, učebnice informatiky atd.) a srovnat jej např. s některým korpusem typu totalita. Po analýze se nabízí diskuze na téma, zda by před pár desítkami let lidé rozuměli našim současným textům zabývajícími se výzkumem, vědou, technologiemi, zdravým životním stylem atd.



Např.: Pro příklad jsem si vybrala jako vstupní text článek z online ženského magazínu – recenzi suchých šamponů<sup>14</sup>, který srovnám s referenčním korpusem SYN2010 a následně i s korpusem totalita 1952. V obou případech jsou klíčová slova stejná – *suchý*, *šampon* a *vlasy*. Dalšími četnými slovy jsou *drogerie*, *redaktorka* a značky jednotlivých produktů. Nejčastějším slovem v distribuci je verbum *jsem*. Výsledky se zobrazí v grafu se syntaktickými vazbami, v tabulce s počtem výskytů i distribuci se zobrazením v částech textu.

#### Tematická koncentrace

Pořadí	Tvar	FQ
4	vlasy	14
7	suchý	10
8	šampon	10

#### Klíčová slova

	Tvar	LL	DIN	Fq(text)	Fq(ref)	TC
1	syoss	71.069	100.0000	3	0	-
2	schauma	64.339	99.9990	3	2	-
3	klorane	62.751	99.9986	3	3	-
4	šampón	55.225	99.9933	3	14	-
5	vyčesat	55.225	99.9933	3	14	-
6	drogerií	68.342	99.9864	4	38	-
7	oriflame	64.334	99.9770	4	64	-
8	šampon	159.544	99.9755	10	171	+
9	šampony	63.750	99.9752	4	69	-
10	překvapena	42.435	99.9383	3	129	-
11	dove	40.041	99.9077	3	193	-
12	koupíte	78.632	99.8958	6	436	-
13	redaktorka	65.003	99.8901	5	383	-
14	mastné	51.246	99.8792	4	337	-
15	suchý	120.000	99.8183	10	1267	+
16	umýt	33.782	99.7368	3	551	-
17	vlasech	29.369	99.4500	3	1153	-
18	síti	47.150	99.3415	5	2302	-
19	vlasů	32.069	98.6647	4	3747	-
20	vlasy	111.618	98.6346	14	13412	+
21	kč	31.733	94.7713	6	22448	-
22	rozhodně	14.188	93.0812	3	14982	-
23	musím	13.296	91.9708	3	17487	-
24	byla	20.198	75.8364	9	172366	-
25	jsem	30.817	66.7207	19	528560	-

Obr. 11: Tabulkové zobrazení výsledků v aplikaci KWords

<sup>14</sup> ZEMEN, Brigita. Velký test: Jak používat suchý šampon a který je nejlepší?. *Blesk pro ženy* [online]. 2015 [cit. 2015-04-10]. Dostupné z: <http://prozeny.blesk.cz/clanek/pro-zeny-trendy-krasa/312700/velky-test-jak-pouzivat-suchy-sampon-a-ktery-je-nejlepsi.html>



Obr. 12: Grafické zobrazení výsledků v aplikaci KWords

Protože aplikace KWords nemá širší využití ve školní výuce, dále s ní ve své práci nebudu pracovat.

## 5.6 Bonito

Program Bonito byl velmi užitečným a výkonným vyhledávačem v korpusech, který vznikl v Centru pro zpracování přirozeného jazyka, pro nyní jej však nahradila uživatelsky přívětivější aplikace KonText. Nejen že KonText oproti Bonitu nabízí předdefinované dotazy a jednodušší ovládání. Navíc uživatelé odpadla nutnost instalace softwaru do vlastního počítače. Nyní již Bonito není podporováno a vyvíjeno, v březnu 2015 bude jeho funkčnost ukončena. Nyní je však ještě stále ke stažení zde: <https://ucnk.ff.cuni.cz/stahni.php>

## 5.7 NoSketch Engine

NoSketch Engine (NoSkE) je open-source rozhraní z webu <http://nlp.fi.muni.cz/trac/noske>. Vychází z komerčního placeného programu SkE (Sketch Engine, dostupný na adrese: <https://www.sketchengine.co.uk/>), je však doplněn o několik dalších funkcí. I tady byla nutná instalace do pc, viz stránku <http://nlp.fi.muni.cz/trac/noske>. Tento projekt kombinoval programy Manatee a Bonito. Poté na NoSkE navázal Kontext a teď jej plně nahrazuje. NoSkE taktéž zanikne v březnu 2015.

## 6. Úkoly pro práci v kontextu

### 6.1 Úkoly určené pro základní školy

#### Cvičení 1 (Kontext) – Druhy a typy slov s jejich příklady

Kantor si s žáky zopakuje 10 slovních druhů (podstatná jména, přídavná jména, zájmena, číslovky, slovesa, příslovce, předložky, spojky, částice, citoslovce) a typů slov (česká a cizí, pomnožná, hromadná, látková, citově zabarvená (lichotivá i hanlivá), zdobněliny, složeniny, spisovná a nespisovná, nářeční, slangová, frazeologismy, vulgarismy...).

**Úkol č. 1:** Vyhledejte pomocí rozhraní KonText příklady daných slov, která znáte. Soutěžte o to, kdo najde slovo s největším či nejmenším výskytem v daném korpusu. Na tabuli lze sepisovat body a nakonec vyhlásit vítěze.

**Řešení:** Žáci ZŠ zvolí atribut lemma v nereferenčním korpusu SYN a do příkazového řádku napíší hledané slovo (slovesa v infinitivu, zbylé ohebné druhy v nominativu (1. pádu) a mužském rodě).

#### Cvičení 2 (Kontext) – Frekvence nadřazených a podřazených slov

Vyučující si s žáky zopakuje nadřazená a podřazená slova a nechá je vyjmenovat nějaké příklady. Potom si každý žák zadá pod atributem lemma vyhledat jedno slovo nadřazené a druhým dotazem slovo k němu podřazené. Opíše si počty výskytů. Jsou nadřazená slova více či méně frekventovaná? Srovnejte to na více příkladech a zamyslete se, proč tak dané frekvence jsou.

**Řešení:** Já volím pro příklad tyto dvojice: *sport* – *foťbal*, *nádobí* – *vařečka* a *zaměstnání* – *učitel*. Výsledek z kontextu je následující: *sport* (395.320 výskytů) – *foťbal* (415.235 výskytů);

*nádobí* (23.278 výskytů) – *vařečka* (3.844 výskytů);

*zaměstnání* (154.187 výskytů) – *učitel* (261.599 výskytů).

U takovýchto proměnných výsledků se nabízí diskuze, zda jsou u nadřazených a podřazených slov nějaké frekvenční souvislosti. Např. *sport* je slovo obecné, konkrétně *foťbalem* se v tisku zabývá každý den alespoň jeden článek. *Zaměstnání* je

slovo obecnější, avšak figuruje v syntaktických vazbách (*hledat/žádat o zaměstnání...*), *učitel* může být osobou různých novinových kauz, svědkem události ve školství, proto může být v tisku frekventovanější.

### **Cvičení 3 (Kontext) – Synonyma**

Žákům je položena otázka, co jsou synonyma a proč jsou v jazyce potřebná. Pomocí atributu lemma následně vyhledají skupiny synonym a určí, proč jsou některá slova ze skupiny frekventovanější. Jaké by preferovali ve větách oni? Žáci mohou zpracovat uvedené příklady, doplnit synonymické řady nebo si vymyslet vlastní.

- a) Hledat a získat *práci – zaměstnání – povolání – profesi – job – fleka.*
- b) Koukat se na *televizi – telku – televizor.*
- c) Pracovat jako *učitel – kantor – vyučující – pedagog – didaktik.*
- d) *Dívka – holka – mladá dáma – mademoiselle – děvče – dívčina – děvčica – holčička – děvčátko – děvenka – dívka – kočka...*
- e) *Kluk – hoch – borec – chlapec – jinoch – klouček – mladý muž – mladík .*

**Řešení:** Daná synonyma mohou a nemusejí být v daných kontextech v synonymickém vztahu. Některá jsou obecná a mohou mít bohatší sémantiku (více významů), jiná jsou úžeji specifikovaná (např. užívaná jen v úzkém okruhu osob, v sociální skupině). U některých slovních tvarů jde o určitou jazykovou rovinu – slang, dialekt aj. Některá slova jsou více odborná a některá zastarávají (archaizují).

### **Cvičení 4 (Kontext) – Jsem také v korpusu?**

- a) Zajímavé může být hledat KonTextem svoje křestní jméno či příjmení. Zjistěte, či jméno nebo příjmení se v korpusu SYN vyskytuje nejčastěji a či nejméně často, případně vůbec?
- b) Mají-li žáci či studenti ve třídě nějaké přezdívky, zkuste je také vyhledat a porovnat.
- c) Přemýšlejte, zda některá jména mají alternativní tvary. Pokud nějaké máte ve třídě, vyhledejte je pomocí srovnávací aplikace SyD. (*např. Jan – Jeník – Honza, Kristina – Kristýna – Kristína, Josef – Pepa – Pepík, Benjamin – Benjamín – Ben, Virginie – Gina, Beáta - Bea...*)

- d) Je možné si také zadat do KonTextu název města nebo vesnice, kde bydlíte (případně nějaké z okolí, kde bydlí vaše babička, kamarádi atd.). Každý může hledat jinou obec a následně můžete srovnat, co je častější. Popřemýšlejte, proč tomu tak je. V jakých kontextech se názvy obcí vyskytují a v jakých typech textů?

## **6.2 Úkoly určené pro střední školy**

### **Cvičení 1 (Kontext) – Kde se v českém jazyce využívá ú/Ú a ů/Ů**

Učitel se studenty (či staršími žáky) rozvede diskuzi, kde všude je možné v českém jazyce používat grafémy „ú“ a „ů“. V učebnici (mluvnici či Pravidlech českého pravopisu) si studenti mohou následně ověřit svoje nápady. Na závěr si svoje myšlenky ověří pomocí aplikace KonText za využití nereferenčního korpusu SYN.

**Úkol č. 1:** Vycházím z hypotézy, že „ú“ se obecně používá na začátku slov. Je možné tento předpoklad ověřit pomocí CQL vyhledávání s atributem lemma. Studenti si vypíší pět nejčastějších slov (viz záložka Frekvence lemmat). Totéž si lze zjistit a vypsát u vlastních jmen (propríí. V jakých v typech textů se vlastní jména na "Ú" nacházejí?

**Úkol č. 2:** Může být „ú“ uprostřed slova? Za pomoci CQL dotazu s atributem word jde i tuto otázku vyřešit. U jakých typů slov se vyskytuje? Výsledek lze konfrontovat s učebnicí gramatiky.

**Úkol č. 3:** Zda může být „ú“ na konci slova, jde také vyhledat v KonTextu. O jaký typ slov se bude jednat?

**Úkol č. 4:** A naopak, může se „ů“ nacházet na začátku slova? Případně je možné jej někde nalézt s velkým písmenem? Využijte atributu word.

## Řešení:

**Úkol č. 1:** Zadám dotaz [lemma="ú.\*"], čímž se mi vyhledají všechny slovní tvary začínající na "ú". V konkordačních řádcích se zobrazí 14.644.726 výskytů. Základní tvary (lemmata) najdeme ve vrchní záložce Frekvence – Lemmata. Těch se v korpusu nachází pouze 8.548. Nejčastější jsou: *úřad, úspěch, úspěšný, útok a úplně*.

U proprií je postup stejný, jen dotaz se bude lišit ve velikosti písmene: [lemma="Ú.\*"]. Zobrazí je 677.987 výskytů, a to především v tiskovinách – denících (Hospodářské noviny, Mladá fronta Dnes, Deníky Bohemia, Právo...). Stejným způsobem jako u obecných jmen si student zobrazí frekvenční distribuci, která ukáže 3563 lemmat. Nejčastější výskyty jsou u slov *Ústí, Újezd a Ústecko*. Následuje velký výčet zkratk nejružnějších *úřadů*.

**Úkol č. 2:** Zvolím dotaz [word=".+ú.\*"]. V dotazu je tečka a plusu (jeden a více výskytů, viz přílohu Regulární výrazy), protože před „ú“ se musí nacházet alespoň jedno písmeno. KonText vyhledá 1.499.376 výskytů. V záložce Lemmata se dozvíme, že sem patří slova složená (složeniny, kompozita; např. *trojúhelník, Orlickoústecko, protiútok*), slova s předponami (deriváty; např. *zúčastnit se, zúžení, bezúplatný*), cizí propria (*Abú Dhabí, Súdán*) anebo slova cizího původu (*skútr, kúra, túra*).

**Úkol č. 3:** Zadám do KonTextu dotaz [word=".\*ú"]. Výskytů je pouze 36.866, slovních tvarů je 3758. I zde se nacházejí propria cizího původu (*Abú, Káthmándú, Honšú, Medvídek Pú*) a citoslovce (*tú, bú, hú*). Překvapivě se zde nacházejí slovesa (*budú, sú, dů*) a podstatná jména (*horú, mnú, pěknú...*) s dialektickou jihomoravskou koncovkou -ú. Nutné je také upozornit, že zda nejdeme i četné výskyty autentických chyb v textech (např. zkratka *KÚ (katastrální úřad)* psána jako *kú* je v korpusu 169x, *metrú* najdeme 93x v denících a tamtéž najdeme 60x slovo *bodú* a 55x *hráčú*).

**Úkol č. 4:** Pro tento úkol zvolím dorazy [word="ú.\*"] a [word=".\*Ú.\*"]. Ač si od základní školy pamatuji poučku, že „ú“ na začátek slova nepatří, KonText vyhledá 1.979 výskytů. Jedná se buď o chybný zápis citoslovce (*ú!*), překlapy (*úProti, pr úběžně*) anebo opět autentické chyby (*úvěrový, úmysl*). Zde je třeba přijmout fakt, že korpusy ČNK zaznamenávají autentické texty s reálně uskutečňovanými chybami. Nelze tvrdit, že by korpusy byly novou jazykovou příručkou, protože obsahují autentické pravopisné chyby.

## **Cvičení 2 (KonText): Politická slova a žurnalismy**

**Úkol č. 1:** Najděte v korpusu SYN pomocí aplikace KonText následující slova: *pravdoláskař*, *havlovat* a *havlovačka*. Znáte tato slova? Nacházejí se zde v korpusu všechna a v jakém typu textů? Z kontextu odhadněte jejich význam a využití. Zkuste je najít i pomocí webového vyhledávače Google.cz a ověřte si je.

**Úkol č. 2:** Znáte nějaké další žurnalismy (=neologismy<sup>15</sup> vzniklé v žurnalistice)? Vyhledejte je KonTextem.

### **Řešení:**

**Úkol č. 1:** Jediné nalezené slovo v korpusu SYN je *pravdoláskař*, a to se dvěma výskyty v jednom novinovém článku z Lidových novin. Je nutné přihlédnout k tomu, že všechna výše zmíněná slova mohou být považována za neologismy a žurnalismy. Všechna označení jsou spojovány s osobností Václava Havla a s jeho výroky, počiny a především dobou. Tyto pojmy se dříve objevily pouze v několika málo článcích a k jejich mediálnímu nadužívání došlo až po smrti Václava Havla (18. 12. 2011). Na tomto příkladě je potřeba si uvědomit, že nejnovější korpusy řady SYN (SYN2010 a SYN2013PUB) obsahují texty cca do konce roku 2009 a navíc, v publicistice obsahují pouze reprezentativní vzorek, nikoli všechny texty. V další vývojové řadě SYN snad bude možné dohledat i tyto obraty.

Slova jde vyhledat pomocí vyhledávače např. Google.cz, kde se zobrazí různé publicistické a zpravodajské články anebo internetové diskuze, kde lidé tyto pojmy rádi používají a zesměšňují. Zde dohledáme jejich význam a kontext.

**Úkol č. 2:** Pro inspiraci lze vyhledat třeba lemmata *tunelovat* (45 výskytů), *mítink* (1.006 výskytů), *šmejd* (186 výskytů), *eseróčko* (20 výskytů), *socdemáci* (3 výskyty), *filozofie firmy* (11 výskytů)... Všechny žurnalismy jsou zaznamenány především v publicistických zdrojích, lemma *šmejd* najdeme zastoupený i v beletrii.

---

<sup>15</sup> Pojem neologismus dle slovníku prof. Eduarda Lotka: „nevžitý, neustálený tvar, nové slovo, popř. jeho nové užití nebo jiná nová jednotka v jazyce vzniklá, resp. vznikající, popř. do něj převzatá odjinud z potřeby pojmenovat něco nového (např. grant, folkař), z důvodu strukturních (např. parník: parolod') apod., novotvar.“



### **Cvičení 3 (KonText): Korpus reflektuje autentické chyby a neznalost!**

Na tomto cvičení bych ráda poukázala na to, že nechybují pouze školáci a že Korpus není náhradou Pravidel českého pravopisu, mluvnice či jakékoli lingvistické učebnice. Korpus obsahuje autentické texty (odborné, publicistické, beletristické, korespondenci atd.), a proto zde najdeme i autentické chyby.

**Úkol č. 1:** Pomocí KonTextu (případně SyDu) vyhledejte a porovnejte následující dvojice slov. Víte sami, která ze dvou uvedených možností je správná? Srovnejte počty výskytů a následně si v nějaké jazykové příručce či učebnici ověřte, která možnost je správná. Použijte korpus SYN a základní vyhledávání.

- a) *bizardní* – *bizarní*;
- b) *tematický* – *tématický*;
- c) *pernamentka* – *permanentka* (nebo *permanentně* – *permanentně*);
- d) *protežovat* – *protěžovat*;
- e) *křestu* – *křtu* (2. pád slova *křest*),
- f) *svatba* – *svadba*;
- g) *tamější* – *tamnější*;
- h) *standart* – *standard*.

**Řešení:** Správné tvary jsou téměř vždy frekventovanější, ale vždy najdeme i výskyty chybných forem. Správné tvary jsou: *bizarní*, *tematický*, *permanentka*, *permanentně*, *protežovat*, *křtu*, *svatba*, *tamější*, *standard*.

**Úkol č. 2:** Podívejte se na následující dvojice slov. Pravopisnou neznalostí se zásadně změnil jejich význam. Vyhledejte jejich konkordance a podívejte se, jak se slova v daném tvaru chovají v kontextu. Pomocí kontextových příkladů vysvětlete významový rozdíl:

- a) *shlédnout* – *zhlédnout*;
- b) *spravovat* – *zpravovat* (případně *správa* – *zpráva*);
- c) *sbít* – *zbit*;
- d) *Matěj* – *matěj*;
- e) *holt* – *hold*.

**Řešení:** *Shlédnout (dolů) z rozhledny, do taliře – zhlédnout (co) film, divadelní představení; spravovat majetek, firmu, území – zpravovat (o čem, podávat zprávu); správa hradu, území – zpráva o hospodaření, o činnosti; sbít (dohromady) poličku, laťky, šibenici – zbit (koho) Tomáše, vojáka; matěj (chybné označení fotografa, mimo to využití ve frazeologismu *Ty jsi ale matěj!*, které se v korpusu nevyskytuje) – Matěj (označení konkrétní osoby); holt (příslovce) – vzdát, položit hold. Význam uživatel pozná z levých a pravých kontextů, v řešení uvádím náhodný výběr. Tentýž úkol lze tak řešit pomocí aplikace SyD, po zobrazení výsledků je možné si ke každé variantě zobrazit také její kolokace (slovní vazby). Zde není výsledek moc přesvědčivý, protože zde jsou nejčastější vazby s neflektivními slovními tvary, případně s lemmaty sloves, proto navrhuji řešení spíše s KonTextem.*

**Úkol č. 3:** Klasikou školních diktátů je psaní malých a velkých písmen v místních názvech. Učitel si se studenty zopakuje základní pravidla pro psaní velkých písmen v názvech. Zkuste si následující slova vyhledat KonTextem a následně se podívejte na Frekvenci – Slovní tvary. Kolik chybných zápisů najdete? Použijte nereferenční korpus SYN a základní vyhledávání.

- a) *Kostelec nad Černými lesy;*
- b) *Řád německých rytířů;*
- c) *Divadlo Bez zábradlí, Divadlo Na Fidlovačce, Divadlo komedie;*
- d) *stanice Náměstí Míru;*
- e) *Spojené státy americké.*

**Řešení:** V odrážkách uvádím správné zápisy názvů. Ve frekvenčních distribucích slovních tvarů uživatel najde všechny zaznamenané zápisy. Zde si jde povšimnout, že mnozí autoři článků ve velkých písmenech chybují, jiní tyto složité názvy vypisují celé velkými písmeny. Běžnému uživateli bych doporučila vyhledat si stránku dané obce či instituce, kde je jejich název napsán bezpečně správně.

#### **Cvičení 4 (KonText): Předpony a přípony**

V následujících úkolech si žáci a studenti vyzkoušejí morfologické spojování slov s typicky českými předponami a příponami (prefixy i sufixy), ale i s takzvanými prefixoidy a sufixoidy, které jsou v úzu českého jazyka čím dál častější. Nejprve se však pokuste sami ve třídě vymyslet příklady ke každému cvičení.

U CQL kódů užíjte atribut **lemma**. U každého úkolu si vypište příklady z co nejvíce slovných druhů (podstatná jména, přídavná jména, slovesa, zájmena, příslovce...). Popřemýšlejte, odkud nalezená cizí slova ze třetího a čtvrtého cvičení znáte (školní předměty, reklama atd.).

**Úkol č. 1:** Najděte pomocí CQL kódu slova s danými **předponami**. U řešení zvažte, zda se VŽDY jedná o předpony.

- a) *před-, za-, nad-, pod-;*
- b) *po-, za-, nade-, při-;*
- c) *záporné ne-, ni-;*
- d) *vymyslete sami nějaké další předpony.*

**Úkol č. 2:** Najděte pomocí CQL kódu slova s danými **příponami**. U řešení zvažte, zda se VŽDY jedná o přípony.

- a) *-ník, -ek;*
- b) *-tel, -ec, -ič,*
- c) *-eček, -íčko, -čka;*
- d) *-ami, -amy, -emi;*
- e) *vymyslete sami nějaké další přípony.*

**Úkol č. 3:** S pomocí učebnice či mluvnice zjistěte, co znamená pojem **PREFIXOID**. Najděte pomocí CQL kódu slova s danými **prefixoidy**. Užíjte atributu **lemma**. Vypište si příklady z co nejvíce slovných druhů. Popřemýšlejte, odkud nalezená cizí slova znáte (školní předměty, reklama atd.).

- a) *bio-, eko-;*
- b) *super-, ultra-, inter-, anti-;*
- c) *uni-, euro-;*
- d) *mega-, maxi-, mini-.*

**Úkol č. 4:** S pomocí učebnice či mluvnické zjistěte, co znamená pojem **SUFIXOID**. Najděte pomocí CQL kódu slova s danými **suffixoidy**.

- a) *-log, -logie;*
- b) *-man, -shop, -holik;*
- c) *-mánie, -fikace, -fobie, -filie.*

**Úkol č. 5:** Stejným způsobem vyhledejte česká kompozita (složeniny) podobného charakteru. Vnímáte je stejně jako cizí hybridní kompozita (viz cvičení výše)?

- a) *vele-, velko-, malo-;*
- b) *-muž, -žena.*

**Řešení:** V KonTextu vyhledáváme pomocí CQL kódu [lemma=".\*zadání .\*"]. U předpon a prefixoidů víme, že vždy musí být nějaká písmena za zadáním, ale může (ale nemusí) být něco před ním. Povinnou část tedy zapíšeme tečkou a plusem. Nepovinnou píšeme pouze s tečkou a hvězdičkou (protože zde je nula a více výskytů libovolných písmen). Vypadá to tedy takto: [lemma=".\*bio.+"]. U přípon jsou povinné a nepovinné části naopak. Proto vzor pro přípony a suffixoidy zadáme takto: [lemma=".\*log.\*"].

### 6.3 KonText jako pomůcka k jazykovým hrám pro učitele ZŠ

Aplikace KonText nemusí sloužit pouze jako nástroj pro vyhledávání a ověřování jazykových jevů samotnými studenty, ale může sloužit jako pomůcka učiteli. Nejen že zde najde autentické příklady užití slov a může si z uvedených vět a kontextů vytvořit nějaká cvičení pro výuku (případně diktát, desetiminutovku atd.), ale může zde hledat inspiraci pro jazykové hry, které se dají využít především na 1. stupni základní školy – na odlehčení učiva, na prohloubení znalostí, za odměnu či v rámci projektových dnů... Tyto hry děti nutí nad slovy více přemýšlet, rozvíjejí jejich slovní zásobu i kombinatorní schopnosti. Několik her uvádím k vyzkoušení, doporučuji pro 3. – 7. třídu ZŠ.

#### **Hra 1 (KonText): Vymysli nejdelší slovo!**

Žáci získají čas okolo 5 minut na to, aby se zamysleli, jaké znají nejdelší české slovo (podstatné jméno) v 1. pádě... Pokud se dětem hra zalíbí, můžete zkusit vymyslet nejdelší slovo v každém slovním druhu.

**Řešení:** Učitel by si měl pro žáky sám nějaká dlouhá substantiva nachystat. Pracuje s nereferenčním korpusem SYN, a to pomocí vyhledávání CQL s atributem lemma. Do závorky přidáme podmínku slovního druhu (atribut pos, viz přílohu č. 2). Hledá-li lemmata např. o nejméně 20 písmenech, bude dotaz vypadat takto: [lemma=".{20,}"]. Výsledkem jsou složená slova (=kompozita a hybridní kompozita) jako *konkurenceschopnost*, *beachvolejbalistka*, *nedisciplinovatelnost*, *multiinstrumentalista*, *Garmisch-Partenkirchen*... Starší žáci mohou také vyzkoušet najít co nejdelší slovo pomocí KonTextu.

### **Hra 2 (KonText): Z dlouhých slov utvoř další**

Tato hra rozvíjí dětskou slovní zásobu a kombinační schopnosti. Uživatel si vyhledá v korpusu (stejným způsobem jako ve Hře 1) dlouhá slova, stačí 10 – 15 písmen. Snaží se vybrat taková slova, která děti znají z běžného užívání (např.: *lokomotiva, materiádouška, sedmikráska, podprsenka, samoobsluha...*). Vybrané dlouhé slovo napíše dětem na tabuli a nechá je např. 10 minut vymýšlet nová slova (můžeme zadání omezit na 1. pády podstatných jmen, ale u mladších žáků to není nutné) z písmen dlouhého slova (př. *LOKOMOTIVA* – *motiv, vlk, lom, lok, kolo, kola, lak, oko...*). Úkol se pro děti ulehčí vizualizací – tj. děti mají slovo napsané na rozstříhaných kartičkách a mohou si z písmenek skládat slova na lavici. Neupravuje se ani délka vokálů, ani háčky. Písmena se mohou ze slova použít vždy jen jednou.

### **Hra 3 (KonText): Nahoru a dolů**

Pomocí CQL dotazu opět najdeme nějaké dlouhé slovo (*listonoška, telegram...*) a řekneme žákům, aby si dané slovo napsali na linkovaný papír shora dolů (svisle) na začátku řádku a zdola nahoru na konci řádku totéž slovo. Nyní nám vyjdou řádky s počátečními a koncovými písmeny do hry. Děti mají vymyslet taková slova, která by se do dané šablony hodila. Viz příklady.

L	LATA, LÁSKA, LEDVINA	A	L	LÍPA, LETUŠKA	A
O	OKOV, OSTROV	V	I	IBIŠEK	K
K	KOCOURI, KOSTI	I	S	SOUŠ, SKRÝŠ	Š
O	OCET	T	T	TEPLO, TĚLO	O
M	MOTO, MOLO, MÍSTO	O	O	OSTEN	N
O	OBJEM, OSM	M	N	NIKDO, NĚCO	O
T	TICHO, TOPIVO	O	O	OCELOT	T
I	IRÁK	K	Š	ŠTOS	S
V	VÍČKO	O	K	KAZI	I
A	ATOL, ARIEL	L	A	ALOBAL	L

**Obr. 13: Možné řešení jazykové hry Nahoru a dolů**

**Řešení:** Nejen, že pomocí CQL dotazu v KonTextu (viz Hru 1) můžeme hledat dlouhá slova, ale pokud dojdeme s žáky na nějakou kombinaci písmen, kde nejsme společnými silami vymyslet nějaké vhodné slovo, můžeme jej zkusit vyhledat KonTextem. Např. pokud nevíme slovo na I-K či O-M, zadáme dotaz: [word="i.\*k"] (*informatik, intrik, infostánek, infrazvuk, ibišek...*) či [word="o.\*m"] (*osm, objem, obchodním, ovšem, obrazem...*).

Obměnou může být zadání, kdy žáci tuto hru řeší pouze s využitím aplikace KonText a hraje se na nejrychlejšího řešitele.

#### **Hra 4 (KonText): Telegram**

Kantor opět využije dlouhá slova (např. z výše uvedených her, anebo vyhledá nová v KonTextu). Žákům jedno zadá a požádá je, aby si jej napsali na horní část papíru. Následně na jednotlivá písmenka vymýšlejí zprávu = telegram, aby dával smysl.

**Řešení:** *MASOPUST – Milá Anežko, Snad Odpustíš, Protože Umírám. Sbohem Tibor.*

### **Hra 5 (KonText): Co obsahuje OKO a DUB?**

Hra se hodí pro práci se slovy příbuznými a odvozenými. Žákům dáme nějaké krátké slovo (např. *oko, dub, les, lak...*), ke kterému mohou libovolně přidávat písmena (nejen předpony, přípony a koncovky), aby však sekvence písmen byla zachována (např. *lak – klakson, odlakovač, lakovky, šelak, vlak, povlak, plak; oko – Okoř, dokořán, dokonat, rokoko...*).

Po 10 minutách hry vyhodnotíme „slovního krále“. Může následovat diskuse, zda jsou slova opravdu příbuzná (a zda jsou tvořena stejným/podobným slovním základem) anebo jde o náhodnou podobnost slov.

**Řešení:** Chce-li se učitel dopředu připravit na danou hru a mít dostatek slov v záloze, vyhledá si slova v korpusu. Pracujeme s vyhledáváním CQL za využití atributu `word`, kde danou sekvenci písmen doplníme z obou stran regulárním výrazem `.*`. Dotazy tedy vypadají takto: `[word=".*lak.*"]` (*vlak, tlak, plakát, oblak, plaketa, galakonzert...*); `[word=".*oko.*"]` (*okovat, kokos, spokojen, eurokomisař, dokonce, okolí, Sokol ...*); `[word=".*dub.*"]` (*duben, Pardubice, dubleta, Dublin, dubstep, jedubaba, dubina...*); `[word=".*les.*"]` (*pokles, klest, blesk, alespoň, Boleslav, bolest, Palestina, těleso...*).



## 7. Cvičení se SyD (vyhledávání dubletních a variantních tvarů)

Tento nástroj se hodí pro výuku českého jazyka žáků ZŠ i studentů SŠ. Žáci/studenti pracují s webovou aplikací SyD (online na adrese [www.syd.korpus.cz](http://www.syd.korpus.cz)), která vyhledá a porovná variantní či dubletní tvary slov. Nutno zdůraznit, že dubleta se skládá ze dvou a více spisovných variant téhož slova (*polévka – polívka, mléko – mlíko...*), zatímco varianty jsou skupiny dvou a více výrazů, které nemusí být vždy spisovné. Může se jednat o spisovný tvar a k němu expresivum, dialektismus, slangismus atd.

Pro práci s aplikací SyD uživatel zadá dvě či více významově podobných slov. Může porovnávat z hlediska synchronního či diachronního jazyka anebo psané či mluvené podoby jazyka v korpusech. Výhodou oproti KonTextu je, že zde není nutná uživatelská registrace. Dle preferencí lze vybrat i možnost lemma (nebo je možné vyhledávat slova v základním nastavení, tj. pouze konkrétní slovní tvary – jako atribut word). Já většinou pracuji s oběma variantami (u jednotlivých cvičení jsou mé tipy napsány), ačkoliv výsledky s atributem lemma u mluvených korpusů (nejsou plně morfologicky otagované) nejsou stoprocentní (může dojít k homonymním tvarům slov).

**Poznámka:** U každého cvičení jsou uvedeny náhodné příklady. Je možné ve výuce aplikovat další varianty a demonstrovat si skrze tento nástroj další lingvistická pravidla.

### Cvičení 1 (SyD): Porovnávání dublet

**Úkol č. 1:** Napadají vás nějaké dvojice slov, které znamenají totéž, jsou synonymní, jsou si zvukově podobné? Zkuste je vyhledat pomocí aplikace SyD.

**Úkol č. 2:** V jazyce se mnohdy objevují dubletní tvary slov (liší jen v hláске či skupině hlásek, mají totožný význam a jsou spisovné). Napadají vás nějaká? Zkuste je vyhledat. Potom vyhledejte následující dvojice (použijte atribut word):

- a) *polévka – polívka*;
- b) *mléko – mlíko* (případně *mlíčko – mléčko*);
- c) *plenka – plínka*;
- d) *pointa – poenta* (*poenta* je nespisovná varianta, nikoli dubleta);
- e) *lisknout – lísknout* (*lísknout* je nespisovná varianta, nikoli dubleta).

**Úkol č. 3:** Dubletní tvary se také objevují u souhlásek, nejčastěji u s/z. Vyhledejte tyto dvojice a zjistěte, která slova jsou frekventovanější. Které jsou vám osobně bližší a milejší? (použijte atribut word)

- a) *univerzita* – *universita*;
- b) *gymnázium* – *gymnasium*;
- c) *diskuze* – *diskuse*;
- d) *kurz* – *kurs* (*diskurz* – *diskurs*, *konkurz* – *konkurs*).

**Úkol č. 4:** Srovnejte synonyma neohebných slovních druhů (použijte atribut word):

- a) *téměř* – *skoro*;
- b) *pořád* – *stále* – *neustále* – *furt* (*furt* je nespisovná varianta, nikoli dubleta);
- c) *možná* – *asi* – *nejspíš*.

**Řešení:**

**Úkol č. 1:** Individuální řešení.

**Úkol č. 2:** a) *Polévka* je častější v psaném jazyce, *polívka* převládá v mluveném; b) *mléko* je častější v psaném jazyce, *mlíko* v mluveném (*mlíčko* převládá v obou variantách, *mléčko* se v psaném vůbec nevyskytuje); c) *plenka* převažuje (pro mluvený korpus nejsou dostatečná data); d) *pointa* vždy převažuje, v mluveném se vyskytuje *poenta*. (pozn.: *Poenta* je správná a jediná spisovná výslovnost tohoto slova.); e) Pro srovnání obou variant nejsou v korpusu dostatečná data. Slovo *lisknout* se v korpusu SYN nachází pouze jednou – je možné to ověřit v aplikaci KonText.

**Úkol č. 3:** a) Počeštělý tvar *univerzita* převládá v psaném i mluveném úzu, ač se tvar *universita* (z angl. *university*) začíná opět dostávat do užívání; b) u tvaru slova *gymnázium* je to stejné jako u *univerzity*, převládá český tvar v obou případech (POZOR! Tvary *gymnazium* a *gymnasium* jsou nespisovné!); c) *diskuse* převládá v psaném úzu, *diskuze* v mluveném. Zastoupeny jsou však u obojího a obě varianty jsou korektní; d) Jazyková korektnost u obojího platí v tomto případě. *Kurz* (*konkurz* i *diskurz*) jsou uživatelsky oblíbenější v obou formách jazyka. Jen dublety *diskurz/diskurs* jsou frekvenčně téměř vyrovnané (*diskurz* 70 výskytů, tj. 56,91 %, *diskurs* 53 výskytů, 43,03 %).

**Úkol č. 4:** a) *Téměř* převažuje v psaných textech, *skoro* je hovorovější, převládá v mluveném jazyce. Zde stojí za zmínku pohled do diachronního srovnání, kde se ve starších textech tato dvě slova velmi proměňují v četnosti; b) v psaných textech převládá *stále*, *pořád* je za ním, *furt* je nejméně časté. V mluvené řeči je to naopak, *furt* má nejvyšší četnost, *stále* a *neustále* mají minimální výskyt.

### **Cvičení 2 (SyD): Srovnání pozdravů a souhlasů**

**Úkol č. 1:** Zamyslete se nad skupinou „souhlasů“: *ano – jo – ba – bať – ja*. Které je podle vás nejčastější v mluvené a které v psané češtině? Svůj názor si zkonfrontujte v programu SyD (použijte atribut word). Víte, o jaký slovní druh se jedná? Dohleďte si to v učebnici nebo v Pravidlech českého pravopisu. Gramatické kategorie také najdete na stránce programu SyD pod grafy.

**Úkol č. 2:** Podobně uvažujte nad pozdravy (použijte atribut word): *dobrý den – ahoj – zdravíčko – čau – ahojky – pá – papa...* Po srovnání si do skupiny přidejte 8. člen – neutrální pozdrav (běžný v internetové komunikaci), a sice *zdravím*. Jak Vám to pohne s výsledky? Zamyslete se nad důvodem vyššího výskytu. Vymyslete a přidejte do srovnání i další pozdravy.

### **Řešení:**

**Úkol č. 1:** Po zadání 5 variant zjistíme, že minimální frekvence dosahuje hanácký souhlas *bať* (v mluveném 2x, v psané 7x). V psaném korpusu převládá spisovné *ano* (téměř 40 tisíc výskytů), v mluveném vítězí hovorové *jo* (téměř 84 tis.). Za povšimnutí stojí vysoký výskyt slova *ja* v mluvených projevech (10x vyšší než v psaných textech), které v kontextech nevyjadřuje souhlas, ale hovorové zkrácení osobního zájmena *já* (*ja vim, ja doufám, ja říkám* atd.).

Podle aplikace SyD se jedná o částice. Dle tzv. akademické Mluvnice češtiny (svazek 2)<sup>16</sup> patří výrazy *ano/ne* a jejich ekvivalenty (*bohužel, nikoliv, zajisté* atd.) k částicím (hodnotícím – modalitním – jistotním).<sup>17</sup>

<sup>16</sup> PETR: Mluvnice češtiny (2) Praha, 1986.

<sup>17</sup> Pro srovnání více zdrojů k zařazení ANO/NE viz článek Miloslava Vondráčka v periodiku Naše řeč (z r. 1998): <http://nase-rec.ujc.cas.cz/archiv.php?art=7419> (online, cit. 1. 3. 2015).

**Úkol č. 2:** V psaném i mluveném jednoznačně vítězí pozdrav *ahoj* (cca 57 % a 44 % výskytů). V psaných projevech následuje formální *dobrý den* (26 %) a přátelské *čau* (pouze 7 %). Pouhých 48 výskytů (1 %) má *ahojky*.. V mluvených záznamech pozdravu *ahoj* velmi konkuruje *čau* (rozdíl o 69 výskytů, tj. 33 %), *dobrý den* má jen 18 %. V souhrnu mezi nejméně používané patří *ahojky*, *zdravičko* (ani zde každý kontext nebude ve smyslu pozdravu) a *papa*.

Slovo *zdravím* patří k četnějším slovům ve srovnání. V mnohých kontextech nebude bráno jako pozdrav, ale jako sloveso ve větě, což dokazuje podrobnější graf – nejvyšší procento výskytu je v odborné literatuře.

### **Cvičení 3 (SyD): Srovnání různých rovin jazyka (expresiva, deminutiva, slova cizího původu a dialektismy)**

**Úkol č. 1:** Vyzkoušejte srovnání skupiny slov obecných a expresivních. (použijte atribut lemma)

- a) *pít* – *chlastat* – *sosat*;
- b) *muž* – *chlap* – *chlapisko*;
- c) *spát* – *chrnět* – *ducat* – *chrápat*.

**Úkol č. 2:** Srovnávejte četnost obecných slov a deminutiv (=zdrobnělin) (použijte atribut lemma). Proč zdrobněliny nikdy nemohou být častější? U prvních tří bodů sledujte koncovky deminutiv a aplikujte je na další příklady. Jsou vždy zdrobněliny spisovné? A je příklad d) opravdu zdrobnělina?

- a) *chlap* – *chlapec* – *chlapeček*;
- b) *dům* – *domek* – *domeček*;
- c) *lidé* – *lidičkové* – *lidičci* – *človícci* – *človíčkové*;
- d) *spát* – *spinkat*.

**Úkol č. 3:** Hledejte dvojice slov, které se liší původem českým nebo cizím (použijte atribut word):

- a) *vyhledat* – *vygooglovat* – *vygůglovat* – *vygooglit* – *vygůglit*;
- b) *cvaknout* – *kliknout*;
- c) *vypnout* – *offnout*;
- d) *napsat* – *kontaktovat* – *mailovat*.

**Úkol č. 4:** Vyhledávejte variantní skupiny slov. Čím jsou daná slova atypická?

Přijdete

na další? (použijte atribut lemma)

- a) *krocan* – *morák*;
- b) *slepice* – *sléпка* – *slípka*;
- c) *přijdu* – *přídu* – *přindu*;
- d) *kalhoty* – *gatě* – *ga'ata* – *ka'ata*;
- e) *tramvaj* – *tram* – *šalina*.

**Řešení:**

**Úkol č. 1:** a) Spisovné *pít* převládá, expresivní a příznakové *chlstat* je v menšině. Expresivum *sosat* má pouze 5 výskytů v psaném jazyce. Eufemismus *bumbat* najdeme v psaném i mluveném kurpusu dohromady 15x. b) Tvar *muž* převládá v psaném, *chlap* v mluvené řeči. Expresivní *chlapisko* najdeme jen v minimu psaných textů (12 výskytů) a jednou v mluveném úzu. Pro zajímavost je možné dohledat toto lemma v KonTextu. c) Spisovné *spát* převládá v obou korpusech, *chrápat* následuje, ač může mít jiné významové konotace. Expresiva *chrnět* a *ducat* mají minimum výskytů.

**Úkol č. 2:** Deminutiva nemohou být frekventovanější, protože nesou funkční příznak. Neutrální a bezpříznaková slova jsou vždy více v jazykovém centru. Najdeme u nich tyto koncovky: *-ek*, *-eček*, *-ec*, *-ičci*, *-ičkové*, *-iček*... a) *Chlapec*, který dnes již není pocíťován jako zdrobnělina, má nejvíce výskytů v psané formě jazyka, *chlap* v mluvené. Vyskytují se všechny tři varianty poměrně hodně. b) V obou případech převládá neutrální *dům*, nejméně frekventovaný je dvakrát zdrobnělý *domeček*. c) V tomto případě je potřeba vyhledávat pomocí atributu lemma. Všechny tyto tvary jsem našla v uživatelském úzu internetových diskutujících. Každý tvar je v korpusech demonstrován alespoň jedním záznamem v některé vrstvě jazyka. Naprosto však převládá příznakově neutrální tvar *lidé*.

d) *Spát* vždy převládá téměř se 100 %. *Spinkat* má 84 výskytů v psaném a 38 výskytů v mluveném. Ano, *spinkat* je doopravdy zdobnělinou slovesa *spát*, viz článek V. Štangalové: Zdobněliny a opakování v dětské poezii v časopise *Naše řeč*<sup>18</sup>.

**Úkol č. 3:** Vybrala jsem pro srovnání tato slova, která jsou podle mě v běžném současném úzu většiny uživatelů jazyka. a) Všechny tvary jsem našla v internetových diskuzích a fórech, dle Slovníku současné češtiny jsou však přípustné pouze tvary *googlovat* (hovorový výraz), případně *googlit* (zde se jedná o slang).<sup>19</sup> Korpus zná pouze slovo *vygooglovat* (16 výskytů v psaném jazyce), převládá české *vyhledat*, které ale může mít více významů. b) V tomto případě převládá cizí slovo *kliknout*, v psaných projevech má 86 %. V mluvených korpusech mají obě slova shodně 7 výskytů. c) *Vypnout* má 100% převahu, *offnout* patří do slangu, ale nenajdeme jej v korpusu, ani ve Slovníku současné češtiny. d) U této trojice cítím sémantický rozdíl už ve spojitosti se syntaktickou vazbou (*napsat komu* (dativ), *kontaktovat koho* (akuzativ), *mailovat komu* (dativ)). Podobně se to liší i ve srovnání v SyDu, samozřejmě *napsat* je nejčastější, *kontaktovat* je užité zřídka, *mailovat* má pouze 8 výskytů (tedy nejméně) v psaných textech, v mluvených se nevyskytuje. V tomto případě má *napsat* nejvíce kolokací (*napsat komu, co, jak, kdy atd.*), naopak *mailovat* je nejvíce sémanticky specifické.

Možnou variantou pro důslednější srovnávání v SyDu může být změna hledaných výrazů, zkusíme tedy: *mailovat – napsat mail – kontaktovat mailem*. Touto změnou dojde k jednodušší sémantice, což ovlivní výsledek. *Mailovat* je tedy nejčastější (8 výskytů), *napsat mail* najdeme zřídka (2 výskyty). *Kontaktovat mailem* či *kontaktovat prostřednictvím mailu* se v psaných korpusech vůbec nevyskytuje.

**Úkol č. 4:** V synchronním i diachronním korpuse se obvykle vyskytují všechny varianty (vyjma slova *ga'ata*), obecná česká slova bez příznaku dialektičnosti jsou však mnohem častější a frekventovanější.

---

<sup>18</sup> ŠTANGALOVÁ, Věra. Zdobněliny a opakování v dětské poezii. *Naše řeč*. 1991, roč. 74, č. 2. Dostupné z: <http://nase-rec.ujc.cas.cz/archiv.php?art=6997>

<sup>19</sup> *Nechybujte.cz: Slovník současné češtiny* [online]. [cit. 2015-03-02]. Dostupné z: <http://www.nechybujte.cz/>

#### **Cvičení 4 (SyD): Jak známe frazeologismy?**

S učebnicí si žáci zopakují, co je to frazeologismus a jaké může mít podoby. Srovnají jejich různé varianty (použijte atribut word). Znáte další synonymické frazeologismy? Vyzkoušejte toto srovnání i na nich.

- a) *poradit se s Vaňkem – vzít do zaječích – vzít nohy na ramena;*
- b) *mít filipa – mít za ušima – mít pod čepicí;*
- c) *stokrát nic umořilo osla – stokrát nic umořilo vola;*
- d) *jít na kutě – jít do hajan.*

#### **Řešení:**

a) *Poradit se s Vaňkem* je již frazeologismem archaizujícím, v korpusu nemá jediný výskyt. Z dalších dvou variant je častější *vzít nohy na ramena*. Pro mluvený jazyk nejsou dostatečná data. b) Tyto tři frazeologismy v základním tvaru mají minimum výskytů, proto lze vyhledávat tvary *má filipa – má za ušima – má pod čepicí*, kde je výskytů více. Potom je nejfrekventovanější *mít za ušima* (6 výskytů). V KonTextu má nejvíce výskytů hledaný výraz *mít filipa*, a to především z častého výskytu kolokace *Jak je důležité míti Filipa*, název divadelní hry Oscara Wilda. c) V tomto případě je opět minimum výskytů, takže SyD není schopen vyhodnotit data a srovnat je. Můžeme však obě fráze vyhledat v korpusu SYN pomocí aplikace KonText. *Stokrát nic umořilo osla* má 74 výskytů, zatímco *stokrát nic umořilo vola* pouhých osm. d) Ať frázi využijeme se slovesem *jít* nebo pouze *do hajan* a *na kutě*, vždy má výskyty pouze spojení *na kutě/jít na kutě*. V KonTextu při základním hledání má *jít na kutě* 142 výskytů, zatímco *jít do hajan* má 55 konkordancí.

## **Cvičení 5 (SyD): SvD jako ortografická pomůcka**

**Úkol č. 1:** Vyberte správnou variantu slova. Srovnějte oba tvary pomocí SyD a dle četnosti se rozhodněte, který je správný. Platí to? Použijte atribut lemma.

- a) *síra* – *sýra*;
- b) *vinyl* – *vynil* – *vinil* – *vynyl*;
- c) *oběd* – *objed*;
- d) *vinný* – *viný* (případně *vina* – *vinna*);
- e) *sirový* – *syrový*.

**Úkol č. 2:** Jaké slovo vám dělá problémy? Čím si v pravopisu nejste jisti? Zkuste si to vyhledat.

### **Řešení:**

Ačkoli najdeme Pravidla českého pravopisu online na [www.pravidla.cz](http://www.pravidla.cz), i aplikaci SyD a korpusové texty můžeme takto využít. Zadáme do srovnání dva tvary, nad kterými se nemůžeme rozhodnout např. s problematikou *s/z*, *i/y*, *ě/je* atd. Při nerozhodnosti si můžete tvar ověřit právě v Pravidlech. Vždy ale používejte atribut lemma (!!!), jinak může dojít ke shodě homonymních tvarů různých slov (např. *síra* (chemický prvek) X *sýra* (genitiv od substantiva *sýr*)).

**Úkol č. 1:** Ve všech případech a – e je nejvíce zastoupený správný slovní tvar (a to téměř 100 %. tj. *síra*, *vinyl*, *oběd*, *vinný*, *syrový*). Jen u tvaru *sirový* jsou asi 3 chybné výskyty.

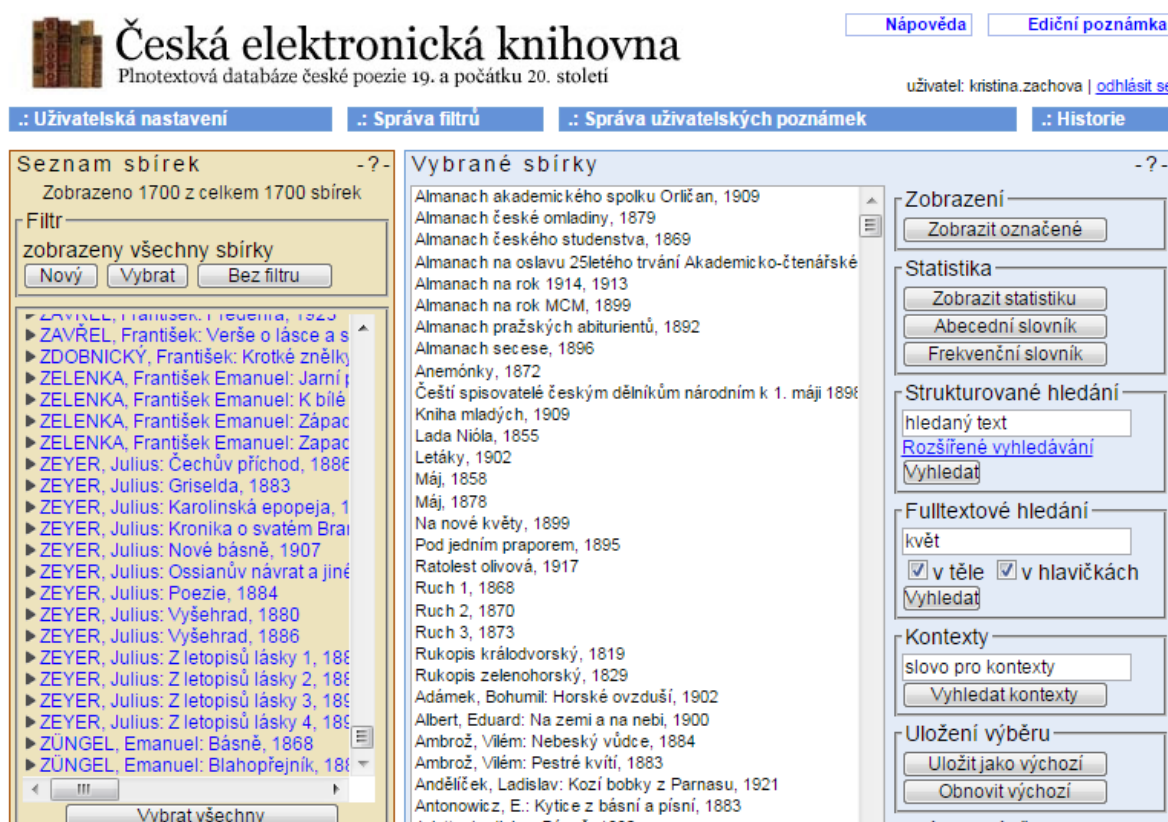
**Úkol č. 2:** Individuální řešení.



## 8. Využití ČEK a KČV

### 8.1 ČEK

Česká elektronická knihovna (ČEK) poslouží přihlášeným studentům především k samostatné četbě a vlastním rozborům básní. Texty sbírek je možné pro nekomerční účely (naoř. pro vlastní studium, četbu, práci se studenty) také tisknout. V tomto vidím velkou výhodu pro maturanty či studenty před recitační soutěží, protože ve školních knihovnách bývají sbírky jen v nízkém počtu výtisků. Sbírkou je zde možné vyhledávat podle názvu, jména autora či roku vydání.



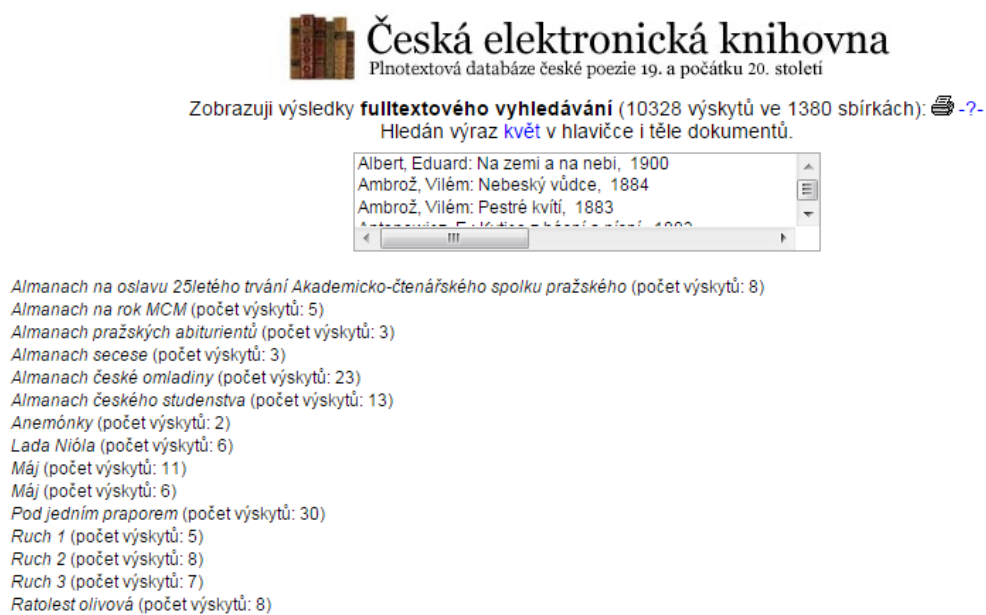
Obr. 14: Online rozhraní pro vytěžování dat z ČEK

Učitelé mohou využít tento zdroj k tomu, aby studentům zadali nějakou báseň či sbírku ke čtení a mohli je odkázat na tento web, ke kterému se všichni studenti bezpečně dostanou ze svých domovů.


Velkou výhodou České elektronické knihovny je spojení práce s literaturou (poezií) a jazykovědou. V následujících úkolech si studenti vyzkouší hledat slova v kontextu poezie, hledat jednotlivé básně a motivy, ale i hledat různé jazykové prvky a frekvenční distribuce typické pro tento typ textů.

Problém je, že se každý student musí nejprve sám přihlásit do databáze. Česká elektronická knihovna a poté ji lze využívat pouze v internetových vyhledávacích Internet Explorer 6.0 a Mozilla Firefox 1.0 (viz Technické požadavky: <http://www.ceska-poezie.cz/cek/scripts/technickepožadavky.php>). Ač to není v technických požadavcích uvedeno, ČEK bezpečně funguje i v současných verzích prohlížeče Opera, což může být optimální řešení pro studenty doma i ve škole.

Práce s ČEK je v Opeře uživatelsky přívětivá. Po přihlášení se zobrazí okno pro práci s daty. V levém sloupci je vhodné zvolit *Vybrat všechny sbírky*. Nyní uživatel pracuje s databází 1700 českých básnických sbírek, tj. s celou databází ČEK. V pravém sloupci pracuji s fulltextovým vyhledáváním (tedy vyhledáváním v nadpisech i jednotlivých básnických textech sbírek). Výsledky se zobrazí v dalším okně, na horních řádcích je počet výskytů v počtu sbírek. Následuje seznam sbírek s jednotlivými výskyty. Jednotlivé sbírky lze zvlášť zobrazovat, klíčová slova pak vidíme v kontextech a je možné je dále analyzovat.



**Česká elektronická knihovna**  
Plnotextová databáze české poezie 19. a počátku 20. století

Zobrazují výsledky **fulltextového vyhledávání** (10328 výskytů ve 1380 sbírkách):  -?-  
Hledán výraz **květ** v hlavičce i těle dokumentů.

Albert, Eduard: Na zemi a na nebi, 1900
Ambrož, Vilém: Nebeský vůdce, 1884
Ambrož, Vilém: Pestré kvítí, 1883
Antonie, Marie: Květy a květinářství, 1883

*Almanach na oslavu 25letého trvání Akademicko-čtenářského spolku pražského* (počet výskytů: 8)  
*Almanach na rok MCM* (počet výskytů: 5)  
*Almanach pražských abiturientů* (počet výskytů: 3)  
*Almanach secese* (počet výskytů: 3)  
*Almanach české omladiny* (počet výskytů: 23)  
*Almanach českého studenstva* (počet výskytů: 13)  
*Anemónky* (počet výskytů: 2)  
*Lada Nióla* (počet výskytů: 6)  
*Máj* (počet výskytů: 11)  
*Máj* (počet výskytů: 6)  
*Pod jedním praporem* (počet výskytů: 30)  
*Ruch 1* (počet výskytů: 5)  
*Ruch 2* (počet výskytů: 8)  
*Ruch 3* (počet výskytů: 7)  
*Ratořest olivová* (počet výskytů: 8)

**Obr. 15: Zobrazení výsledků pro výraz květ ve fulltextovém hledání ČEK**

### 8.1.1 Cvičení pro ČEK:

#### Cvičení 1 (ČEK): Hledání slov a motivů

**Úkol č. 1:** Studenti zadají do fulltextového vyhledávače postupně následující slova: *sebevražda, otcovražda, vražda otce, lámání kolem*.

**Úkol č. 2 (pro mladší žáci):** Do fulltextového vyhledávání zadáme slova jako *jaro, maminka/mamince/matička, květiny, štěstí* či *radost*. Tyto básně můžeme využít třeba na Den matek, MDŽ, jarní výzdobu třídy, školní vystoupení atd.

**Úkol č. 3:** Podle oblíbeného motivu si každý žák/student může vyhledat a vybrat báseň na recitační soutěž.

**Úkol č. 4:** Fulltextovým vyhledáváním najdete slovo *květ* a podívejte se na kontexty ve sbírce Lada Niola. Vysvětlete jednotlivé významy v kontextech.

#### **Řešení:**

**Úkol č. 1:** Všechny uvedené motivy najdeme v Máchově Máji.

*Sebevražda* se také vyskytuje 9x v sedmi sbírkách: Dyk, Viktor: Satiry a sarkasmy (počet výskytů: 2), Lešehrad, Emanuel: Meditace (počet výskytů: 2), Merhaut, Josef: Básně (počet výskytů: 2), Pachmayer, Josef: Písňe zločincovy (počet výskytů: 1), Ruda, Jiří: Báseň o Kladně (počet výskytů: 1), Sova, Antonín: Píseň o Rovnosti (počet výskytů: 1).

*Otcovražda* má 0 výskytů, zatímco spojení *vražda otce* je velmi častým motivem, v ČEK je 2837 výskytů v 786 básnických sbírkách. Tento motiv byl v tomto období častý, kompozitum *otcovražda* v té době asi nebylo v básnickém úzu.

*Lámání kolem* je z těchto motivů ve sbírkách nejčastější. Má okolo 12 tisíc výskytů ve 1425 sbírkách. Stojí za zvážení a diskuzi, zda by tyto motivy byly v této frekvenci také v současné české poezii? Je možné, že by některý motiv úplně vymizel?

**Úkol č. 2:** Individuální řešení.

**Úkol č. 3:** Individuální řešení.

**Úkol č. 4:** V řešení uvádím jednotlivé úryvky z Lady Nioly (8. sbírka shora, 6 výskytů substantiva *květ*) a přidávám vlastní interpretaci motivu *květ*.

*Předlouhým déštěm skleslý květ.* – motiv zhrzené naděje, smutek

*I chví se ještě ret,*

*jak v sněhu růže květ.* – metaforické přirovnání rtů ke květu

*Proto prchá květ z mých lící,*

*Proto pláču, drahá růže!*“ – ztráta naděje a radosti vyjádřen změnou barvy tváří.

*...mužiček v žlutých brslenkách a červené kamizolce, na hlavě měl místo kloboučku květ slunečnice, a byl navzdor tomu...* – popis lesního muže, reálný obraz květu

*... ale pleť bílá, jako květ vodní růže...* – básnické přirovnání

*Škoda, kdyby jiný byl utrhl květ její, neznaje ani krásu jeho ocenit.* – *Ale co tu slibů, přísah u těch děvčat, hned aby se jim na věky zadal. Blázínkové!* –“ – utrnutí květu dívky je zde coby obraz jejího mládí a krásy, dívka dospívá po boku muže.

## **Cvičení 2 (ČEK): Básnická slova a přechodníky**

**Úkol č. 1:** Víte, co jsou to tzv. poetismy (=básnická slova)? Pojem si vysvětlíte či využijte učebnici/výkladový slovník. Následně si některé příklady vyhledejte v básnických sbírkách (např. *duma, mha, luna, vesna, kol, lad...*)

**Úkol č. 2:** Nejprve si tipněte, která slova jsou v poezii (upozorňuji, že v 19. a 20. století) nejčastější? V pravé liště si potom zobrazte frekvenční slovník všech sbírek. Podívejte se, jaká slova jsou nejfrekventovanější.

**Úkol č. 3:** Studenti si s učebnicí gramatiky zopakují (zjistěte), co jsou to přechodníky a jak se tvoří. Každý si potom ze slovesa vytvoří libovolný přechodník a vyhledá jej fulltextovým vyhledáváním. Následně se zamyslete, proč jsou přechodníky tak často používané v poezii a proč se dnes v běžné komunikaci nepoužívají?

### **Řešení:**

**Úkol č. 1:** V internetové encyklopedii *Co je co?* je poetismus vysvětlen takto: „*básnická slova – stylisticky příznakové výrazy, náležející k lexikálním prostředkům poezie (vesna, perut; luna, дума).* V české literatuře jsou poetismy zastoupeny zejména v básnické tvorbě lumírovců. V moderní poezii přejímají úlohu poetismy neologismy, archaismy a cizí slova.“<sup>20</sup> Poetismy v zadání mají hojně výskyty ve sbírkách. Studenti mohou diskutovat o tom, proč takové básnické tvary vznikají a proč se neuvžívají v běžné komunikaci.

---

<sup>20</sup> Co je co?, online, 2002

**Úkol č. 2:** Nahoře v pravém sloupci vybereme *Frekvenční slovník*, v dalším okně zvolíme totéž. Frekvenční vyhodnocení všech sbírek probíhá delší dobu. Nejčastější jsou předložky a spojky (*a, v, se, na...*), *dále* tvary slovesa *být*, osobní zájmena (*já, ty*) a ze substantiv je nejčetnější *srdce*.

**Úkol č. 3:** Individuální řešení.

### **Cvičení 3 (ČEK): Rozbor básní Petra Bezruče (starší SŠ studenti)**

Ve výčtu básnických sbírek nalevo si najdete Bezručovu nejznámější sbírku **Slezské písně**. Zobraďte si první a poslední vydání knihy (tj. z let 1909 a 1967; nahoře zvolte Text sbírky s obsahem – pro lepší orientaci) a podívejte se, jak se sbírka změnila za 58 let (počet básní, řazení v obsahu atd.). Následně si v obou vydáních najdete snad nejznámější básně **Ostrava** a **Maryčka Magdónova** a porovnejte si obě verze – změny slov, řazení, úprava pravopisu atd. (Pozn.: Všechny změny jsou autorské.)

**Řešení:** Záměrně jsem zvolila tyto básně, protože z vlastní zkušenosti vím, že zde je spousta změn. Autor zde záměrně upravuje slova ve verších, aby lépe pasovaly do meter. Někde dochází k nezáměrným chybám v přepisu, dochází tedy k záměnám slov (*robili – pobili*). Oproti první verzi došlo také k ortografickým úpravám.

### **Cvičení 4 (ČEK): Doplně slova (starší žáci, studenti SŠ)**

Pokud mají žáci/studenti všichni čtenářskou zkušenost s nějakou básní (např. po přečtení v čítance či v ČEK), můžeme jim v hodině číst báseň znovu a v každém verši nějaké slovo vynechat. Žáci slova doplňují.

(*Ukázka cvičení:* K. J. Erben: Zlatý kolovrat (sbírka Kytice): *Okolo lesa pole \_\_\_\_\_, hoj, jede, \_\_\_\_\_ z lesa pán. Na \_\_\_\_\_ bujném jede koni, vesele \_\_\_\_\_ zvoní, jede \_\_\_\_\_ a \_\_\_\_\_.*)

### **Cvičení 5 (ČEK): Poznej báseň, urči sbírku a autora**

Pokud mají žáci/studenti všichni čtenářskou zkušenost s nějakou básní (např. po přečtení v čítance či v ČEK), přečteme jim v hodině báseň znovu, případně vybereme nějakou pasáž. Můžeme soutěžit o to, kdo z nejmenšího počtu veršů pozná danou báseň. Po straších žácích či studentech můžeme chtít určit i básnickou sbírku a autora. Toto cvičení jde jistě využít i při testech a opakování učiva.

### **Cvičení 6 (ČEK): Rozstříhaná básnička**

Vybereme mladším žákům z ČEK nějakou básničku (výběr můžeme provést na základě vhodného motivu či osobní znalosti přiměřeného textu). Nejprve ji rozstříháme po verších či více slovech a ve dvojicích/skupinkách se ji žáci snaží správně složit. Na závěr je jim ukázána původní verze básně. Starším studentům můžeme vybrat některý známý kanonický text (z Máje, Kytice, apod.) a vybranou pasáž/sloku rozstříhat na jednotlivá slova.

## **8.2 KČV = Nástroje pro práci a vytěžování jazykových dat z ČEK**

Jak jsem již výše zmínila, Českou elektronickou knihovnu lze nejen používat jako zdroj mnohých textů české poezie a pro vyhledávání motivů, ale lze z ní nově (od roku 2014) získávat i další data pomocí softwarových nástrojů Korpusu českého verše. I zde je nutné uživatelské přihlášení (zvláště než do ČEK), zobrazí se však v jakémkoli internetovém prohlížeči.

### **8.2.1 Gunstick**

Aplikace **Gunstick (=Databáze českých rýmů)** je zajímavou pomůckou pro mladší žáky, aby pochopili, jak a proč se slova rýmují. (K tomuto nástroji není potřeba přihlášení.)

U cvičení neuvádím řešení, protože jsou všechna individuální.

#### **Cvičení č. 1 (Gunstick): Proč se slova rýmují?**

Ukažte si, jak se slova rýmují. Najděte slova *milý, maminka, tráva, les, hrad, zpívá* atd. Každý žák si může najít jiné slovo. Následně z vyhledaných rýmových dvojic/skupiny napište společně básničku.

#### **Cvičení č. 2: Básnička mamince**

Ke Dni matek či MDŽ můžete maminkám vyrobit pěkné blahopřání a do něj vepsat vlastní básničku. Vhodné rýmy opět najdete v aplikaci Gunstick.

#### **Cvičení č. 3 (Gunstick): Básnička na jeden rým**

Pokud najdete slovo o více různých rýmech (hledejte slova obecnějšího charakteru – osoby, příroda, věci denní potřeby, pocity atd.), zkuste napsat básničku pouze s těmito rýmy (tedy podle veršového schématu AAAA...; např. rýmy: *láska – sázka – páska - vráska*).

#### **Cvičení č. 4 (Gunstick): Typy veršů**

Pomocí aplikace Gunstick vyhledejte rýmové dvojice ke čtyřem náhodným slovům dle vlastního výběru. Potom si v učebnici literární teorie najděte všechny typy veršů (střídavý, sdružený, obkročný atd.) a zkuste si napsat básničku podle veršových schémat (ABAB, AABB, ABBA...). Následně si spočítejte, zda vám sedí slabiky.

### **Cvičení č. 5 (Gunstick): Napiš sonet**

Šikovní studenti si v učebnici zjistí schéma sonetu, které je pevně dané. Pomocí Gunsticku si najdou vhodné rýmy a napíší sonet coby zvláštní literární úkol.

#### **8.2.2 Cvičebnice**

Cvičebnice KČV nabízí velké množství veršových příkladů ze sbírek ČEK, které testují znalosti žáků či studentů. Můžeme testovat znalosti meter, rozměrů či pevných forem. Časomíry jsou mimo učivo ZŠ i SŠ, proto je zde neuvádím. Při chybném řešení si můžeme zvolit Výklad (pravá lišta) k dané problematice, kde jsou vysvětleny a popsány všechny pevné formy, metra i rozměry. Výhodou je, že se příklady neopakují, řešení je ukázáno a vysvětleno ve výkladech. Nutné je předchozí přihlášení. Funguje ve všech internetových prohlížečích.

#### **8.2.3 Hex**

Software pojmenován po počítači z Úžasné Zeměplochy Terryho Pratcheta pracuje stejně jako aplikace KWords od ČNK. Hex slouží k vyhledávání klíčových slov v básních, a to o minimální frekvenci 3 četnosti (možno zvolit i více).

### **Cvičení č. 1 (Hex): Klíčová slova obecně**

Jaká klíčová slova se podle Vás mohou v básních objevit? Vyhledejte je a podívejte se na četnosti výskytů. (Tip: *láska, mládí, stáří, smrt, život, krása...*)

### **Cvičení č. 2 (Hex): Klíčové slovo válka**

Zvažte, kdy mohlo být v básních slovo *válka* nejčastější (uvažujeme nad časovým intervalem 1800 – 1938). Následně vyhledejte Hexem a potvrďte/vyvráťte své domněnky. Proč tomu tak je? Který autor na ni nejvíce poukazuje v letech 1920 – 1928? Dohledejte si jeho básně v ČEK a přečtěte si je.

Totéž vyzkoušejte se slovem *naděje*, přemýšlejte o ní v časovém intervalu 1800 – 1960 (RF slova).



**Řešení:**

Motiv vyhledávám Hexem o minimální četnosti 2 na sbírku. *Válka* je ve sbírkách ČEK poměrně častým klíčovým slovem, a to především v období 1920 – 1930. Ne vždy se jedná o reflexi první či druhé světové války, může jít také o vnitřní pocitové války, což může vysvětlit vysoký výskyt v roce 1905.

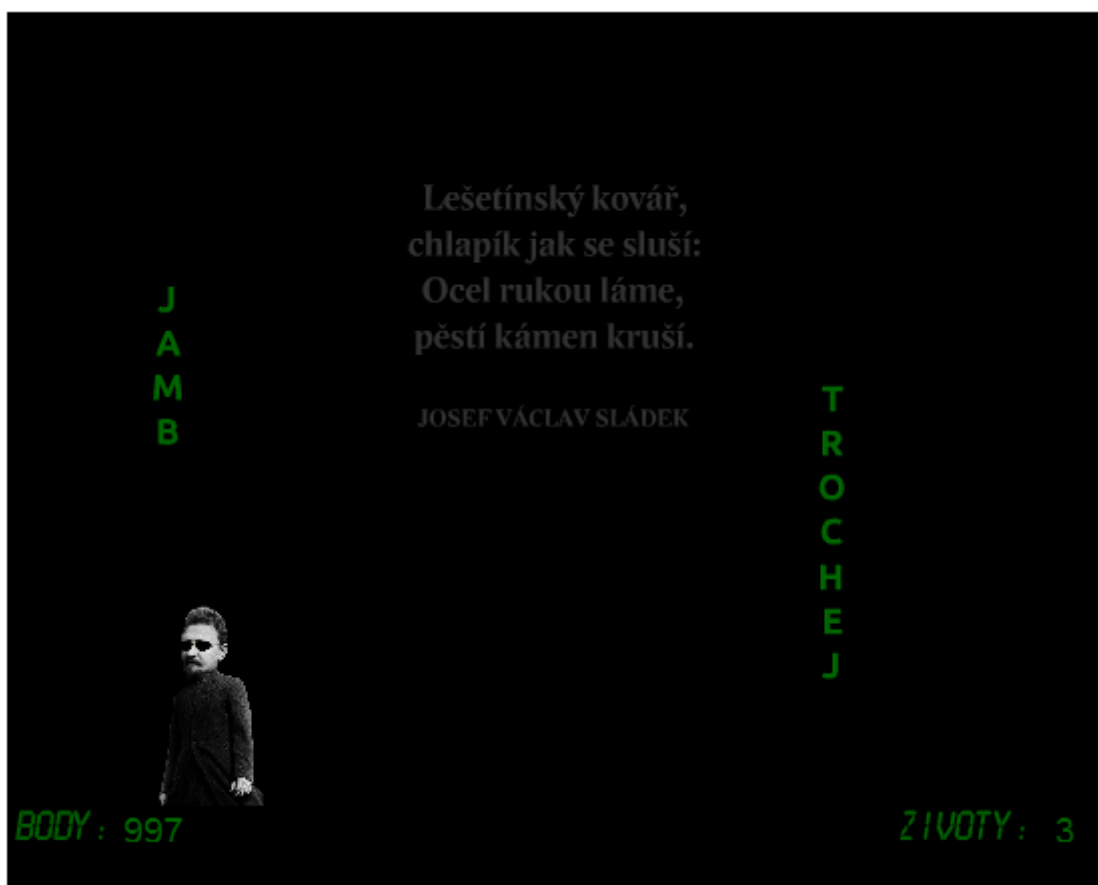
V online nástroji je pod grafem (s různými frekvencemi) možné dohledat konkrétní sbírky s jednotlivými sbírkami.

Klíčové slovo *naděje* je v české poezii poměrně rovnoměrně rozmístěno od roku 1835 do roku 1945. Nejpočetnější frekvence *naděje* jsou v roce 1872 a 1899, kde autoři vkládají *naděje* do nového století.

#### 8.2.4 Hry v KČV

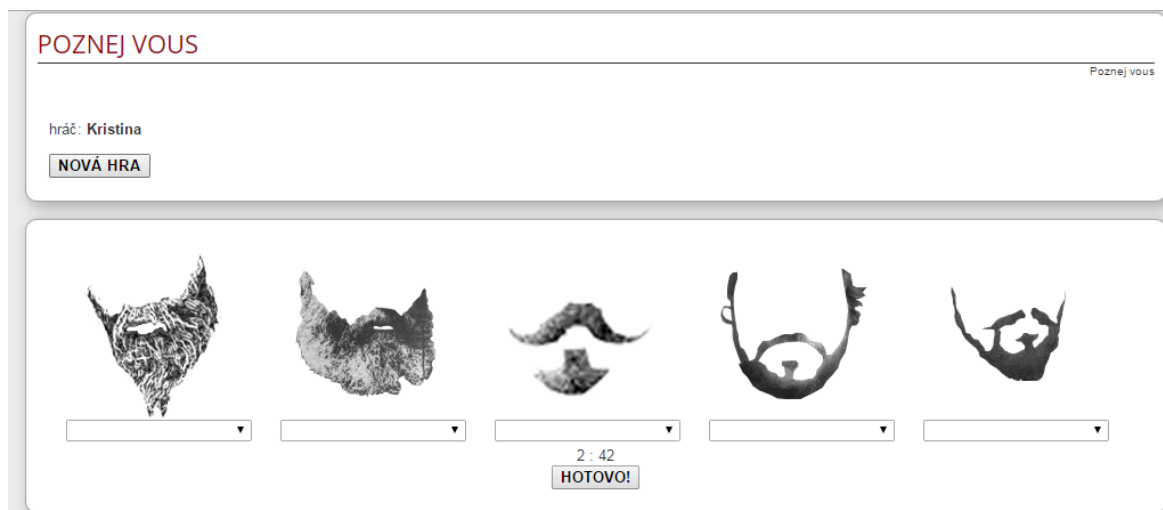
Versologický tým udělal v Korpusu českého verše také dvě drobné hry, kterými lze taktéž ozvláštnit výuku českého jazyka nebo literatury.

První hra se jmenuje **Metrics**. Spouští je online v internetovém prohlížeči, ovládá se šipkami vlevo a vpravo. Po zadání jména hráče se nám zobrazí herní pole, na němž je úryvek básnického textu k rozboru verše, padající typy veršů (*jamb* či *trochej*) a postava literárního vědce (dle úrovně – Josef Král, Jan Mukařovský, Miroslav Červenka). O každém z literárních vědců se může uživatel dočíst v odkaze na Slovník českých spisovatelů. Postava ve hře se pohybuje do stran pomocí šipek a chytá typy veršů dle předložené ukázky. Uživatel se v průběhu hry může dopustit maximálně tří omylů, každým se mu ubere jeden život, poté jeho hra skončí. Šikovný uživatel může projít třemi levely hry. Metrics je online dostupný bez přihlášení na odkazu: <http://www.versologie.cz/metrix/index.html#>.



Obr. 17: Online hra Metrics na procvičování meter

Druhou versologickou aplikací je hra **Poznej vous**. Je volně dostupná v internetovém prohlížeči. Po zadání uživatelského jména se nám zobrazí 5 obrázků různých vousů českých básníků. U každého obrázku je roleta s pěti jmény na výběr. Hra měří čas a vyhodnocuje znalosti. Při vyhodnocení nám ukáže správné řešení, ale i celé portréty známých českých literátů. Hru je možné několikrát zopakovat, protože obrázků je v databázi okolo dvaceti. Aplikace Poznej vous je přístupná zde: <http://versologie.cz/vous/>.



Obr. 17: Online hra Poznej vous

## 9. Vokabulář webový

Ve Vokabuláři webovém (VW) je možné zobrazovat staré české texty, slovníky, mluvnici nebo poslouchat audioknihy. Přidávám možná cvičení a využití staročeských textů.



Obr. 18: Vyhledávání na webu Vokabulář webový

### Cvičení č. 1 (VW): Co na to Jungmann?

Při probírání učiva Národního obrození můžeme otevřít tento digitalizovaný čtyřdílný slovník Josefa Jungmanna (VW – *Digitalizované slovníky – Seznam hesel – Zobrazit abecedu*). Každý žák si vybere počáteční písmeno svého jména/příjmení a projde si stránky hesel na toto písmeno. Kterému slovu byste dnes již nerozuměli? Které slovo vás nejvíce zaujalo?

## Cvičení č. 2 (VW): Pracujeme se slovníky

V záložce *Vyhledávání* najdeme hledání v sedmi staročeských slovnících (ne však v Jungmannově). Následně můžeme výsledek zobrazit v generálním hesláři (heslo s vysvětlivkami) či fulltextovém vyhledávání (tj. slovo vidíme v kontextu).

**Úkol č. 1:** Tímto cvičením demonstrujeme práci se slovníky. Zadejte libovolné slovo (např. *krása*, *být*, *hvězda*...), na kterém si v generátoru hesel ukážeme, jaké informace slovníky nabízí, což platí i pro papírové slovníky.

**Řešení:** U tohoto nástroje najdeme původní zdroj (ze kterého slovníku je heslo čerpáno), jeho flexe a rod, jazykový původ, významy a kontexty. K řešení přikládám obrázek.

**krása**, -y, fem., stsl. *krasa* venustas, pulchritudo. — 1. Světlost, lesk, záře, Licht, Glans. (Kristus) temnosti pek... scatione sui splendoris Otc. 323<sup>b</sup>. Měsíc  
všicku svú krafu od s... ásu = světlost, srov.: (měsíc) od slunce  
světél bývá ŠtítMus. 2... rafu, ovšem by měsíc krásen nebyl Štít.  
uč. 55<sup>a</sup>. Slunce (při sn... še ztratilo Krist. 100<sup>b</sup>, ztratilo svou krásu  
= svou světlost, zatmil... atí Kremism. 93<sup>a</sup>, slunce svú krafu potraty  
PassKlem. 107<sup>a</sup>. Slunečná krafa bieše toho času, že se marináři diviechu, jenž na tu  
světlost patriechu Baw. 193, *krása* a světlost (obé tu znamená věc touž). (Hospodin) jest  
přidal krafy splendorem Ol. Jud. 10, 4. (Nero) počě sě w krafi velikého plamene kochati  
Pass. 311, v krásě = v záři. Srov. krásný = světlý a krásnost = světlost. — 2. Barva jistá  
červená, světlečervená, Rōthe. Králová... zbleděvši a svú krafu proměnivši na své panně  
omdle in pallorem colore mutato Ol. Esth. 15, 10, *krása* oppos. ke zbleděti. Srov. krásný =  
červený a stpol. *krasa* = barva Kobyl.-Rost. 172, rus. *kraska* = červenost. — 3. Sličnost,  
ozdoba, Schönheit, Zierde. Daj nám za hříechy plakati... pro crafi tvé zbavenie Kunh.  
150<sup>a</sup>. Rušil s' tiem jeho (dietěte) crazzu, ž' si nedoždal svého času ApD. c. (Olympias)  
ovšem divné krafy bieše AlxV. 74. Vzal mu (Herkules Paridovi) tetku ovšem divné krafy  
t. 782. (Helena) jměla krafu přemnoženú t. 810. I krafi tvé pulchritudine tua ŽKlem. 44, 5,  
kraffu tvú ŽWtitb. tamt; kraffu oděn jest decore ŽKlem. 92, 1; krafi méj decori ŽWittb.  
29, 8, kraffie mé ŽKlem. tamt. krafa tě jest omýlila species Pror. Dan. 13, 56. krafy  
púščové speciosa ŽPod. 64, 13. krafa nebo twarnoft species Ol. Ex. 24, 17. Jenž t' (ženich)  
by byl v takěj věhlasi, tak učeny, v takěj krafy i múdry jako já sama Kat. v. 376. Divili jsú  
se jich kraafe KřižB. 95<sup>b</sup>. krafa pulchritudo Prešp. 1399, forma MamA. 23<sup>b</sup>, decor t. 20<sup>b</sup>,  
aspectus Veleš. — Jm. osobní: Johannes Krassa TomZ. 1385 n 16; »Krása« v Tomkově  
překladě Kosmy III, 23, začěž je v rkp. A Kirassa, 1 Kyrassa atd. (stran -i- v. Chrabří:  
Chyraber m. Chabr).

Poslední změna: 20. 01. 2015 Zdroj: J. Gebauer, Slovník staročeský

Obr. 19: Slovníkový záznam z vyhledávání slova *krása* na VW

### **Cvičení č. 3 (VW): Příroda – kde se vzala?**

**Úkol č. 1:** V Jungmannově slovníku se poprvé objevilo slovo *příroda*. Najděte jej pomocí cesty: *Digitalizované slovníky – 3. díl – Listování* – hledejte od strany 605... Když najdete slovo *příroda* a pokuste se přečíst si jeho význam. Pochopíte také, co znamenají odvozená slova: *přjroděna, přjrodina, přjrodniectwj, přjrodoskum, přjrodoslowj, přjrododěge* apod.? Najdete je na dalších dvou stranách. Jsou používány i dnes nebo jsou naprosto vypuštěny z jazyka či nahrazeny jiným slovem?

**Úkol č. 2:** Najděte ve slovnících (záložka *Vyhledávání*) pojem *příroda*. Pod jakým heslem je uvedeno v Malém staročeském slovníku? V jednom vyhledaném odkazu najdete staročeské slovo, které znamená *nevnímatelný, nepozorovatelný, nepostižitelný smysly*. Jaké to je?

#### **Řešení:**

**Úkol č. 1:** Heslo *příroda* se nachází na straně 609. Významy jsou: a) co se ku předešlému urodilo; b) přirozenost, vlastnost věci přirozená; c) moc původně ve všech bytnostech se nalézající, a rozdíl i proměny jich působící; d) soubor všech věcí na Zemi. Další výše uvedená slova jsou na stranách 609 – 610.

**Úkol č. 2:** V Malém staročeském slovníku se *příroda* sémanticky vykládá jako *přirozenost, pohlaví*. Ve fulltextovém vyhledávání k pojmu *příroda* najdeme heslo *nečutedlný, nečutelný* významu *nevnímatelný, nepozorovatelný, nepostižitelný smysly*.

#### **Cvičení č. 4 (VW): Co to znamená?**

Pro toto cvičení využijeme Staročeskou textovou banku. Vyhledávání v ní je jednoduché, zobrazení výsledků je nápadně podobné zobrazování konkordančních řádků v KonTextu.

Do vyhledávání zadejte následující slova. Z kontextu odhadněte jejich význam. Pro pomoc se můžete podívat do slovníků (*Vyhledávání*). V pravé liště se také podívejte na počty pozic (=tokenů). Ze všech příkladů najděte nejfrekventovanější slovo.

- a) *snide, vznide;*
- b) *luzi, dřevo;*
- c) *dievka, Písmo.*

#### **Cvičení č. 5 (VW): Přechodníky**

Při výuce okrajových jazykových jevů (archaismů) můžeme studentům ukázat přechodníkové tvary v kontextech staročeských textů. Ukažte si pomocí učebnice způsob tvoření přechodníků přítomných i minulých se všemi třemi koncovkami (namísto -ou- v koncovkách použijte -ú- vzhledem k povaze textů). Pak pomocí Staročeské textové banky vyhledejte všechny tvary přechodníků od sloves:

- a) *být*
- b) *číst*
- c) *říct*

#### **Cvičení č. 6 (VW): Poslechni si staročeské texty**

Dalším využitím Vokabuláře může být skupinový poslech starých českých textů (*Bible olomoucká, Tkadleček, Cestopis tzv. Mandevilla, Sbirka kázání* atd.), jejichž přednes najdeme v sekci Audioknihy. Je možné studentům ukázat daný text. Osobně to považuji za lepší způsob než je nutit číst transkribované texty.

## **ZÁVĚR**

Ve své bakalářské práci jsem se zabývala využitím korpusové lingvistiky na základních a středních školách. V teoretické části jsem popsala historii korpusové lingvistiky v České republice, její základní pojmy a postupy práce, Český národní korpus jako celek a včetně typologie. Tato část obsahuje i kapitolu o nutnosti zábavné formy výuky češtiny na základních školách pro další rozvoj žáků.

Hlavní část mé práce je prakticky zaměřená a slouží jako příručka pro práci s korpusy ve výuce. Navrhuji zde různá cvičení na osvojování si jazykových jevů a poznatků, některá z nich žákům nad školní osnovy rozšíří znalosti z lingvistiky jejich mateřského jazyka. Zadáni pracují s poznatky z morfologie, lexikologie, syntaxe i ortografie. Úkoly a jazykové hry směřuji především k vytěžování dat z Českého národního korpusu, jistou část cvičení také věnuji korpusům a online nástrojům mimo ČNK. Pracuji také s aplikacemi České národní knihovny, Korpusu českého verše či Vokabuláře webového, které lze využít spíš v hodinách literatury. Na těchto programech je možné názorně vidět, jak se lingvistika a literární teorie prolínají.

Výsledkem mojí práce je názorná příručka zaměřená na využití korpusové lingvistiky ve školství, která může posloužit jako učebnice pro aplikaci jazykovědných a literárněvědných poznatků.



## **ANOTACE**

**Autorka:** Kristina Záchová

**Katedra:** Katedra bohemistiky, Filozofická fakulta Univerzity Palackého Olomouc

**Název bakalářské práce:** Korpusy a jejich využití ve výuce na základních a středních školách

**Vedoucí práce:** PhDr. Petr Pořízka, Ph. D.

**Počet znaků:** 84.337 znaků

**Počet příloh:** 4

**Počet titulů použité literatury:** 16 (+14 softwarových nástrojů)

**Klíčová slova:** *korpus, ČNK, Český národní korpus, výuka, školství, KČV, Korpus českého verše, ČEK, Česká elektronická knihovna, Vokabulář webový, alternativní výuka, lingvistické softwary, příručka, čeština.*

### **Anotace bakalářské práce:**

Má bakalářská práce se zabývá popisem korpusů a lingvistických softwarů pro český jazyk. Vysvětluje základní práci s nimi a technické parametry pro užívání, může sloužit jako základní manuál korpusové lingvistiky pro učitele na základních a středních školách. Největší část práce je přímo zaměřená na výuku, obsahuje množství cvičení a praktických úkolů pro vytěžování jazykových dat z korpusů a práci s dalšími softwary (např. Vokabulář webový, Korpus českého verše, Česká elektronická knihovna atd.). Cvičení jsou zaměřena na ortografické znalosti, slovtvorné principy, lexikální znalosti i na znalosti literatury a literární teorie.

### **Annotation of a Bachelor Theses**

My Bachelor Theses deals with descriptions of corpuses and linguistic pieces of software for the Czech language. It explains basic work with them and technological parameters for using them, it can also serve as a basic manual for corpus linguistics for teachers of primary and secondary schools. The largest part of the paper is directly aimed at the process of teaching, it contains a lot of exercises and practical tasks for gaining the language data from the corpuses, and work with more pieces of software (e. g. Web Vocabulary, Corpus of the Czech Verse, Czech Electronic Library, etc.). The exercises are specialized in orthographic knowledge, morphological principles, lexical knowledge, and also in knowledge of literature and the theory of literature.

## ZDROJE:

### Bibliografie:

CVRČEK, Václav - RICHTEROVÁ, Olga (eds). [Internet]. *Příručka ČNK*; 2015 [cit 2015-02-19]. Dostupné z: <http://wiki.korpus.cz/> .

ČERMÁK, František – BLATNÁ, Renata (eds.): *Jak využívat Český národní korpus*. Praha: NLN, Nakladatelství Lidové noviny, 2005, 179 s. ISBN 80-710-6736-9.

ČERMÁK, František. *Korpusová lingvistika: stručný historický přehled*. [online]. [cit. 2015-02-19]. Dostupné z: [ucnk.ff.cuni.cz/doc/korp\\_lingv\\_prehled.rtf](http://ucnk.ff.cuni.cz/doc/korp_lingv_prehled.rtf)

ČERNÁ, Alena M. a Boris LEHEČKA. *O Vokabuláři*. [online]. [cit. 2015-02-20]. Dostupné: <http://vokabular.ujc.cas.cz/informace.aspx?t=ovokabulari&o=ovokabulari>.

IBRAHIM, Robert - PLECHÁČ, Petr. *Báseň a počítač*. Praha: Akademie věd České republiky, 2014. Věda kolem nás.

KOCEK, J; KOPŘIVOVÁ, M.; KUČERA, K. (eds.). *Český národní korpus. Úvod a příručka uživatele*. Praha: Filozofická fakulta UK, Ústav Českého národního korpusu, 2000.

KOTEN, Tomáš. *Škola? V pohodě! : metody, hry a formy práce pro realizaci učiva, pro dosažení očekávaných výstupů a rozvoj klíčových kompetencí*. Vyd. 1. Most: Hněvín, 2006. ISBN 80-86654-18-4.

LOTKO, Edvard. *Slovník lingvistických termínů pro filology*. 1. vyd. Olomouc: Vydavatelství Univerzity Palackého, 1999. 112 s. ISBN 8070679654.

MIKULÍKOVÁ, Vendula. *Výuka češtiny hravě* [elektronický zdroj]. 1. vyd. Olomouc: Univerzita Palackého v Olomouci, 2011. 1 CD-ROM. CD. ISBN 978-80-244-3011-9.

PETR, Jan. *Mluvnice češtiny. (2), Tvarosloví*. 1. vyd. Praha : Academia, 1986.

PLECHÁČ, P. Versologický tým: *Korpus českého verše*. [online]. [cit. 2015-02-24]. Dostupné z: <http://www.versologie.cz/kcv.html>

ŠTANGALOVÁ, Věra. *Zdrobněliny a opakování v dětské poezii*. *Naše řeč*. 1991, roč. 74, č. 2. Dostupné z: <http://nase-rec.ujc.cas.cz/archiv.php?art=6997>

VONDRÁČEK, Miloslav: *Citoslovce a částice – hranice slovního druhu*. *Naše řeč*. Praha: Ústav pro jazyk český AV ČR, v. v. i., 1998, roč. 81, č. 1. str. 29. Dostupné online z: <http://nase-rec.ujc.cas.cz/archiv.php?art=7419>

*Co je co?: Vaše encyklopedie* [online]. 2.9.2002 [cit. 2015-04-10]. Dostupné z: [http://www.cojeco.cz/index.php?id\\_desc=74162&s\\_lang=2&detail=1](http://www.cojeco.cz/index.php?id_desc=74162&s_lang=2&detail=1)

*Nechybujte.cz: Slovník současné češtiny* [online]. [cit. 2015-03-02]. Dostupné z: <http://www.nechybujte.cz/>

*Pravidla.cz: Pravidla českého pravopisu* [online]. [cit. 2015-04-11]. Dostupné z: <http://www.pravidla.cz/>

### **Softwarové nástroje:**

*Český národní korpus*. Ústav Českého národního korpusu FF UK, Praha. Dostupný z: <http://www.korpus.cz>. [cit. 2015-4-9].

*KonText*. FF UK. Praha 2011. Dostupný z <https://kontext.korpus.cz/> [cit. 2015-2-24].

*SyD*. FF UK. Praha 2011. Dostupný z: <http://syd.korpus.cz> [cit. 2015-2-24].

*Morfio*. FF UK. Praha 2011. Dostupný z <https://morfio.korpus.cz/> [cit. 2015-2-24].

*KWords*. FF UK. Praha 2011. Dostupný z <https://kwords.korpus.cz/> [cit. 2015-2-24].

*Korpus DIALOG 1.1*. Praha: Ústav pro jazyk český, AV ČR, 2012. Dostupný z: <http://ujc.dialogy.cz>. [cit. 2015-2-24].

ČERNÁ, Alena M. a Boris LEHEČKA. *Vokabulář Webový*. [cit. 2015-02-20]. Dostupné z: <http://vokabular.ujc.cas.cz/>.

PLECHÁČ, P. – IBRAHIM, R.: *Databáze českých meter* (Ústav pro českou literaturu AV ČR: Praha 2013), dostupný z <http://www.versologie.cz>. [cit. 2015-2-24]

PLECHÁČ, P. – IBRAHIM, R. – ŘÍHA, J.: *Úvod do teorie verše – cvičebnice* (Ústav pro českou literaturu AV ČR: Praha 2014), dostupný z <http://www.versologie.cz>. [cit. 2015-2-24].

PLECHÁČ, P. – IBRAHIM, R.: *Gunstick – databáze českých rýmů* (Ústav pro českou literaturu AV ČR: Praha 2013), dostupný z <http://www.versologie.cz>, [cit. 2015-2-24].

PLECHÁČ, P.: *Hex – klíčová slova v české poezii* (Ústav pro českou literaturu AV ČR: Praha 2014), dostupný z <http://www.versologie.cz>, [cit. 2015-2-24].

PLECHÁČ, P.: *Frekvenční slovníky české poezie* (Ústav pro českou literaturu AV ČR: Praha 2013), dostupný z <http://www.versologie.cz>, [cit. 2015-2-24].

PLECHÁČ, P.: *Eufonometr* (Ústav pro českou literaturu AV ČR: Praha 2013), dostupný z <http://www.versologie.cz>, [cit. 2015-2-24].

SVADBOVÁ, Blanka et al.: *Česká elektronická knihovna: Poezie 19. a počátku 20. století*. Ústav pro českou literaturu AV ČR: Praha 2005. [online]. Dostupný z: <http://www.ceska-poezie.cz/cek/>. [cit. 2015-04-11].

## PŘÍLOHY:

### Příloha č. 1: Přehled regulárních výrazů pro CQL

- **tečka** (.) – představuje jeden libovolný znak,
- **interval** ({n, k}) – představuje *n* až *k* opakování předchozího znaku nebo většího celku; je-li *k* vynecháno ({n,}), odpovídá intervalu nejméně *n* opakování, pokud má interval tvar {n}, odpovídá mu přesně *n* opakování;
- **hvězdička** (\*) – představuje libovolný počet (0 a více) opakování předchozího znaku nebo celku, je tedy ekvivalentní s {0,}
- **plus** (+) – představuje 1 nebo více opakování předchozího znaku nebo celku, totéž co {1,}
- **otazník** (?) – představuje žádný nebo jeden výskyt předchozího znaku nebo celku, identické s {0,1}
- **seznam** ([]) – představuje alternativu. Nabízí možnost vybrat jeden libovolný znak z těch, které jsou uvedeny v seznamu uvnitř hranatých závorek; pokud je prvním znakem seznamu stříška (^), jde o negovaný seznam a představuje tedy libovolný jeden znak kromě těch uvedených uvnitř hranatých závorek; v rámci seznamu je možné používat také pomlčku (-) jako operátor rozsahu (např. [a-z],[1-9]),
- **svislá čára** (|) – představuje také alternativu, ne ovšem mezi jednotlivými znaky, ale celými řetězci tvořícími celek,
- **kulaté závorky** – libovolnou část výrazu je možné seskupit do kulatých závorek, vytvořit tak celek a ovlivnit tím prioritu jeho vyhodnocování nebo na něj aplikovat výše zmíněné kvantifikátory, vztahující se jinak jen na jeden (předcházející) znak,
- **zpětné lomítko** (\) – pokud některému speciálnímu znaku předchází zpětné lomítko, ztrácí tento znak svůj zvláštní význam (což umožňuje např. vyhledávat konkrétní interpunkční znaménka).

Příklady využití regulárních výrazů najdete v následující tabulce:

Příklad	regulární výraz
všechny tvary slova <i>ptakopysk</i>	<code>ptakopys.*</code>
slovo <i>kdy</i> s malým nebo velkým počátečním písmenem	<code>[kk]dy</code>
tečka jako interpunkční znaménko	<code>\.</code>
infinitivy předponových sloves od <i>nést</i>	<code>.+nést</code>
různě dlouhé varianty citoslovce <i>ratata</i>	<code>ra(ta)+</code>
pravopisnou dubletu: <i>diskuze</i> psané i se <i>s</i>	<code>diskuse diskuze nebo disku[sz]e</code>
libovolná čísla skládající se ze tří nebo čtyř cifer	<code>[0-9]{3,4}</code>

### Přehled klávesových zkratk regulárních výrazů

	svislítko	AltGr + Shift + pod „Backspace“ nebo Alt + W
{}	složené závorky	AltGr + 9, AltGr + 0 nebo Alt + B, Alt + N
[]	hrnaté závorky	Alt + F, Alt + G
^	stříška	Alt + š (či 3)
\	zpětné lomítko	Ctrl + Alt + Q

## Příloha č. 2 – Poziční tagset (České pozíční morfologické značky)<sup>21</sup>

### ČESKÉ POZIČNÍ MORFOLOGICKÉ ZNAČKY

<p><b>1. SLOVNÍ DRUH</b></p> <p><b>A</b> adjektivum (přídavné jméno)</p> <p><b>C</b> číslovka nebo číselný výraz s číslicemi</p> <p><b>D</b> adverbium (příslovce)</p> <p><b>I</b> interjekce (citoslovce)</p> <p><b>J</b> konjunktive (spojka)</p> <p><b>N</b> substantivum (podstatné jméno)</p> <p><b>P</b> pronomen (zájmeno)</p> <p><b>V</b> verbum (sloveso)</p> <p><b>R</b> prepozice (předložka)</p> <p><b>T</b> partikule (částice)</p> <p><b>X</b> neznámý, neurčený, neurčitelný slovní druh</p> <p><b>Z</b> interpunkce, hranice věty</p>	<p><b>7</b> zájmeno reflexivní „se“, „si“, „ses“, „sis“</p> <p><b>8</b> zájmeno přivlastňovací „svůj“</p> <p><b>9</b> zájmeno vztažné „jenž“, „jíž“, „iž“, ... po předložce (n-: „něhož“, „níž“, ...)</p> <p><b>A</b> adjektivum</p> <p><b>B</b> sloveso, tvar přítomného nebo budoucího času</p> <p><b>C</b> adjektivum, jmenný tvar (např. „rád“, „schopen“)</p> <p><b>D</b> zájmeno ukazovací (např. „ten“, „onen“)</p> <p><b>E</b> zájmeno vztažné „což“</p> <p><b>F</b> součást předložky, která nikdy nestojí samostatně (např. „nehledě (na)“, „vzhledem (k)“)</p> <p><b>G</b> adjektivum odvozené od přítomného přechodníku</p> <p><b>H</b> krátké tvary osobních zájmen (např. „mě“, „mi“, „ti“, „mu“)</p> <p><b>I</b> citoslovce</p> <p><b>J</b> zájmeno vztažné „jenž“, „jíž“, ... bez předložky</p> <p><b>K</b> zájmeno tázací/vztažné „kdo“ vč. tvarů s „ž“ a „s“</p> <p><b>L</b> zájmeno neurčité (např. „všechny“, „sám“)</p> <p><b>M</b> adjektivum odvozené od minulého přechodníku</p> <p><b>N</b> substantivum</p> <p><b>O</b> samostatně stojící zájmena (např. „svůj“, „nesvůj“, „tentam“)</p> <p><b>P</b> zájmena osobní (např. „já“, „ty“, „on“) vč. tvarů s „-s“ (např. „tys“)</p> <p><b>Q</b> zájmeno tázací/vztažné „co“, „copak“, „cožpak“</p> <p><b>R</b> předložka (obecná, bez vokalizace)</p> <p><b>S</b> zájmeno přivlastňovací „můj“, „tvůj“, „jeho“, vč. plurálu</p> <p><b>T</b> částice</p> <p><b>U</b> adjektivum přivlastňovací (např. „-ův“, „-in“)</p> <p><b>V</b> předložka vokalizovaná (např. „ve“, „ku“)</p>	<p><b>W</b> zájmena záporná (např. „nic“, „nikdo“, „nikdy“, „žádný“)</p> <p><b>X</b> slovní tvar, který byl rozpoznán morfologickou analýzou, ale značka chybí</p> <p><b>Y</b> zájmeno tázací/vztažné „co“ spojené s předložkou („oč“, „nač“, „zač“)</p> <p><b>Z</b> zájmeno neurčité (např. „nějaký“, „některý“, „číkoll“, „cosi“)</p> <p><b>a</b> číslovka neurčitá (např. „mnoho“, „málo“, „tolik“, „několik“, „kdovíkolik“)</p> <p><b>b</b> příslovce, které se neskládá ani ho nelze negovat (např. „pozadu“, „naplocho“)</p> <p><b>c</b> kondicionál slovesa „být“ („by“, „bych“, „bys“, „bychom“, „byste“)</p> <p><b>d</b> číslovka druhová s adjektivním skloňováním (např. „dvoji“, „desaterý“)</p> <p><b>e</b> přechodník přítomný</p> <p><b>f</b> infinitiv</p> <p><b>g</b> příslovce, které se skládá a časuje</p> <p><b>h</b> číslovka druhová (např. „jedny“, „nejedny“)</p> <p><b>i</b> sloveso ve tvaru rozkazovacího způsobu</p> <p><b>j</b> číslovka druhová větší nebo rovna 4 v substantivním postavení (např. „čtyřero“, „desatero“)</p> <p><b>k</b> číslovka druhová větší nebo rovna 4 v adjektivním postavení, krátký tvar (např. „čtyřery“)</p> <p><b>l</b> číslovka základní s nesubstantivním skloňováním (např. „jeden“, „dva“, „tři“, „čtyři“, „pět“, „šest“, „tisíc“)</p> <p><b>m</b> přechodník minulý (např. „udělav“)</p> <p><b>n</b> číslovka základní větší nebo rovna 5</p> <p><b>o</b> číslovky násobné neurčitá (např. „mnohokrát“, „tolikrát“)</p>
<p><b>2. SLOVNÍ PODDRUH</b></p> <p># hranice věty</p> <p>% autorova signatura (např. „haš-99.-B.-S“)</p> <p>* slovo „krát“</p> <p>• spojka podřadící (včetně „aby“, „kdyby“ ve všech tvarech)</p> <p>} číslovka psaná římskými číslicemi (např. XIV)</p> <p>: interpunkce všeobecná</p> <p>= číslo zapsané číslicemi</p> <p>? číslovka „kolik“</p> <p>@ slovní tvar, který není morfologickou analýzou rozpoznán</p> <p>^ spojka souřadící</p> <p>4 zájmeno vztažné/tázací s adjektivním skloňováním (např. „jaký“, „který“, „o“)</p> <p>5 zájmeno „on“ ve tvarech po předložce (např. „něj“, „něho“)</p> <p>6 zájmeno reflexivní „se“ v dlouhých tvarech (např. „sebe“)</p>		

<sup>21</sup> Jak využívat Český národní korpus. Příloha knihy.

<p><b>p</b> přechází činné (vč. přidaného „-s“)  <b>q</b> archaické přechází činné (zakončení „-t“)  <b>r</b> číslovka řadová  <b>s</b> přechází trpné (vč. přidaného „-s“)  <b>t</b> archaický slovesný tvar přítomného a budoucího času (zakončení „-t“)  <b>u</b> číslovka tážací násobná „kolikrát“  <b>v</b> číslovka násobná (např. „pětkrát“)  <b>w</b> číslovka neurčitá s adjektivním skloňováním (např. „nejeden“, „tolikátý“)  <b>y</b> zlomek zakončený na „-ina“ (např. „pětina“)  <b>z</b> číslovka tážací řadová „kolikátý“</p>	<p><b>5. PÁD</b>  - neurčuje se  <b>1</b> nominativ (první pád)  <b>2</b> genitiv (druhý pád)  <b>3</b> dativ (třetí pád)  <b>4</b> akuzativ (čtvrtý pád)  <b>5</b> vokativ (pátý pád)  <b>6</b> lokál (šestý pád)  <b>7</b> instrumentál (sedmý pád)  <b>X</b> libovolný pád (1/2/3/4/5/6/7)</p>	<p><b>X</b> libovolný čas (F/R/P)  <b>10. STUPEŇ</b>  - neurčuje se  <b>1</b> 1. stupeň  <b>2</b> 2. stupeň  <b>3</b> 3. stupeň  <b>11. NEGACE</b>  - neurčuje se  <b>A</b> afirmativ (bez negativní předpony „ne-“)  <b>N</b> negace (tvar s negativní předponou „ne-“)</p>
<p><b>3. ROD</b>  - neurčuje se  <b>F</b> femininum (ženský rod)  <b>N</b> neutrum (střední rod)  <b>H</b> femininum nebo neutrum  <b>I</b> maskulinum inanimatum (rod mužský neživotný)  <b>M</b> maskulinum animatum (rod mužský životný)  <b>Q</b> femininum singuláru nebo neutrum plurálu (pouze u přechází a jmenných adjektiv)  <b>T</b> maskulinum inanimatum nebo femininum plurálu (pouze u přechází a jmenných adjektiv)  <b>X</b> libovolný rod (F/M/I/N)  <b>Y</b> maskulinum (animatum nebo inanimatum)  <b>Z</b> nikoli femininum, tj. M/I/N</p>	<p><b>6. PŘIVLASTŇOVACÍ ROD</b>  - neurčuje se  <b>F</b> femininum  <b>M</b> maskulinum animatum (pouze u adjektiv)  <b>X</b> libovolný rod (M/I/F/N)  <b>Z</b> nikoli femininum, tj. M/I/N</p>	<p><b>12. AKTIVUM/PASIVUM</b>  - neurčuje se  <b>A</b> aktivum  <b>P</b> pasivum  <b>13. NEPOUŽITO</b>  - neurčuje se  <b>14. NEPOUŽITO</b>  - neurčuje se  <b>15. VARIANTA</b>  - základní tvar  <b>1</b> varianta k základnímu tvaru  <b>2</b> řídka, archaická nebo knižní varianta k základnímu tvaru  <b>3</b> velmi archaický tvar, též hovorový  <b>4</b> velmi archaický nebo knižní tvar, pouze spisovný  <b>5</b> hovorový tvar (ve veřejných projevech)  <b>6</b> hovorový tvar (koncovka obecné češtiny)  <b>7</b> hovorový tvar, varianta k 6  <b>8</b> zkratky  <b>9</b> speciální použití (např. tvary zájmen po předložkách)</p>
<p><b>4. ČÍSLO</b>  - neurčuje se  <b>D</b> duál (pouze 7. pád feminin)  <b>P</b> plurál (množné číslo)  <b>S</b> singulár (jednotné číslo)  <b>W</b> pouze v kombinaci se jmenným rodem  <b>Q</b> (singulár pro feminina, plurál pro neutra)  <b>X</b> libovolné číslo (D/S/P)</p>	<p><b>7. PŘIVLASTŇOVACÍ ČÍSLO</b>  - neurčuje se  <b>P</b> plurál  <b>S</b> singulár  <b>X</b> libovolné číslo (S/P)  <b>8. OSOBA</b>  - neurčuje se  <b>1</b> 1. osoba  <b>2</b> 2. osoba  <b>3</b> 3. osoba  <b>X</b> libovolná osoba (1/2/3)  <b>9. ČAS</b>  - neurčuje se  <b>F</b> futurum (budoucí čas)  <b>P</b> přezens (přítomný čas)  <b>R</b> préteritum (minulý čas)  <b>H</b> přezens nebo préteritum (P/R)</p>	<p><b>13. NEPOUŽITO</b>  - neurčuje se  <b>14. NEPOUŽITO</b>  - neurčuje se  <b>15. VARIANTA</b>  - základní tvar  <b>1</b> varianta k základnímu tvaru  <b>2</b> řídka, archaická nebo knižní varianta k základnímu tvaru  <b>3</b> velmi archaický tvar, též hovorový  <b>4</b> velmi archaický nebo knižní tvar, pouze spisovný  <b>5</b> hovorový tvar (ve veřejných projevech)  <b>6</b> hovorový tvar (koncovka obecné češtiny)  <b>7</b> hovorový tvar, varianta k 6  <b>8</b> zkratky  <b>9</b> speciální použití (např. tvary zájmen po předložkách)</p>



## Studijní literatura

- Čermák, F.: Jazykový korpus: prostředek a zdroj poznání. *Slovo a slovník* 11, 1995, str. 119–140
- Čermák, F.: Konference. Staré známé, nebo neznámé slovo? *Česká literatura*, 1. sv. světě, 1 a 2, 2001, str. 27–30
- Čermák, F. – Kubiček, P.: Jazykový korpus a škola, *Český jazyk a literatura*, 41, 1997, str. 84–92
- Český národní korpus. *Úvod a příručka uživatele*. FF UK, Praha 2000
- Frekvenční slovník češtiny*. NLN, Praha 2004
- Havránek, B. – Jedlička, A.: *Česká mluvnice*. SPN, Praha 1960
- Pravidla českého pravopisu*. Academia, Praha 1993 – zkratka **PCP**
- Příruční slovník jazyka českého*. Státní pedagogické nakladatelství, Praha 1915, 17 – zkratka **PSJČ**
- Retrográdní slovník současné češtiny*. Academia, Praha 1986
- Slovník české frazeologie a idiomatické výrazy*. I. Přirovnání (1983), II. Výrazy nespojené (1988), III. Výrazy slovesné (1994). Academia, Praha – zkratka **SČFI**
- Slovník spisovné češtiny*. Academia, Praha 1994 – zkratka **SSČ**
- Slovník spisovného jazyka českého I-VIII*. Academia, Praha 1989 – zkratka **SSJČ**

## Vyhledávací vzory

- 1) slovní tvar  
např. *kočka*  
**kočka**  
1a slovní tvar s nespecifikovaným zakončením  
např. *kočka, kočky, kočkodan...*  
**kočk.\***  
1b slovní tvar s nespecifikovaným začátkem  
např. *odbočka, otočka, vánočka, modroočka...*  
**\*očka**
- 2) slovní tvar s variantou jednoho písmene (malé x velké, dlouhé x krátké, i x y)  
**[Kk]očka, probl[ee]m, b[yy]dlo**  
např. *kočka x kočka, problém x problēm, bidlo x bydlo ...*
- 3) slovní tvar s variantami více písmen  
např. slova složená ze tří písmen, a to souhlásky, samohlásky a souhlásky  
**[bcčdđfghjklmnpqrřsstřtvwxzž][aaeeěiioóuuúyy][bccčdđfghjklmnpqrřsstřtvwxzž]**
- 4) kombinace slovních tvarů  
např. *domáci/kočka*  
**domáci kočka**  
4a kombinace tvarů s volným zakončením  
např. *domáci kočkách...* (nikoli *domáci kočce, domácích koček*)  
**domác.\* kočk.\***
- 5) lemma (tag)  
A např. *kočka* (tj. *kočka, kočky, kočce...*)  
**[lemma="kočka"]**  
B např. neživotná substantiva mužského rodu  
**[tag="N.I.\*"]**  
5A lemma se stejným zakončením  
např. *obout, kout, plout, zaujmout* (nikoli slovesa zakončená na *-nout*)  
**[lemma="\*[-n]out\*"]**  
5Ab lemma se stejným začátkem  
např. *kolo, kolona, kolonie, kulbs, kolotoč, kolouch...*  
**[lemma="\*kolo.\*"]**

<sup>22</sup> Jak využívat Český národní korpus. str. 176 - 179

- 6) lemma v kombinaci s tagem – tvar téhož lemmatu  
např. lemma *kočka* v akuzativu singuláru  
[[lemma="kočka"] & (tag="...S4.\*")]
- 7) lemma v kombinaci s tagem – tvar jiného lemmatu  
např. lemma *kočka* s následujícím slovesem v prezentu – *kočka mňouká*  
[[lemma="kočka"] [tag="V.....P.\*"]  
7a vyhledání kombinace lemmatu a slovního tvaru  
např. *mít namířeno, měl uděáno...*  
[[lemma="mít"] [(tag="Vs.\*") & (word=".\*o\*")]]
- 8) kombinace lemmatu s tagem – přípony  
např. *divadlo, sedadlo, umyvadlo...*  
[[lemma=".\*dlo\*"] & (tag="N.\*")]
- 9) kombinace lemmatu s tagem – předpony  
např. *přidělat, přišít, přitáhnout...*  
[[lemma="př.\*"] & (tag="Vf.\*")]
- 10) lemma a jeho odvozeniny  
např. *přst, přstík, přsten, náprstek...*  
[[lemma=".\*prst.\*"]]
- 11) odvozeniny s různými slovními základy  
např. *přslušník, příbormík, příložník, příživník...*  
[[lemma="př.\*ník\*"]]
- 12) kombinace (posloupnost) lemmat  
např. *domácí kočka* (tj. *domácí kočky, domácí kočce...*)  
[[lemma="domácí"] [lemma="kočka"]]
- 12a vyhledání kombinace dvou lemmat a slovního tvaru  
např. *měla jsem namířeno, má být namířeno apod.*  
[[lemma="mít"] [lemma="být"] [(tag="Vs.\*") & (word=".\*o\*")]]
- 13) kombinace lemmat s jedním vloženým slovem  
např. *domácí léčení kočky*  
[[lemma="domácí"] [] [lemma="kočka"]]
- 14) kombinace (posloupnost) lemmat s více vloženými slovy  
např. *domácí zvířata a kočky, domácí/léčení kočky, domácím kočkám, domácí  
zvířata, zvlášť kočky*  
[[lemma="domácí"] [(0,3) [lemma="kočka"]]
- 14a přesný počet vložených slov  
např. *pes, když uvidí kočku; pes, event. kočka; psi, celá smečka koček*
- 14b opakující se lemmata nebo tagy několikrát za sebou (např. 3 substantiva v dativu)  
např. *hraběti Karlu Josefovi, panu Václavu Durychovi, doktoru Josefu Peškovi*  
[[tag="N...3.\*"]]{3}
- 15) kombinace (posloupnost) lemmat s přímým a inverzním slovoledem  
např. *domácí kočka, kočka domácí*  
[[lemma="domácí"] [lemma="kočka"] | ((lemma="kočka") [lemma="domácí"])]
- 16) kombinace (posloupnost) lemmat s přímým a inverzním slovoledem a s vloženými slovy  
např. *věnoval pozornost, věnovali jim velkou pozornost, pozornost byla věnována, pozornost jsem tomu nevěnoval*  
[[lemma="věnovat"] [(0,3) [lemma="pozornost"] | ((lemma="pozornost") [(0,3) [lemma="věnovat"]])]]
- 17) více lemmat (tagů) současně  
A např. *psy / kočkám / morčata* (l = různé konkordanční řádky)  
[[lemma="pes"] | [lemma="kočka"] | [lemma="morče\*"]]  
B např. infinitiv / přechodník /pasivní přičestí / vstáváje / vyznačena)  
[[tag="Vf.\*"] | [tag="Ve.\*"] | [tag="Vs.\*"]]
- 18) kombinace kolokací složených z lemmat a tagů  
např. *koupit auto, dům, chleba...*; *prodat auto, dům, objekt...*  
[[lemma="koupit"] [tag="N...4.\*"] | ((lemma="prodat") [tag="N...4.\*"])]
- 19) interpunkční znaménko  
např. *!*  
\!
- 19a kombinace interpunkčních znamének  
např. *...!!!*  
\.\.\.\!\!\!\!
- 19b kombinace interpunkčních znamének  
např. *...nebo*  
[word="."] [3] [word="\!"] [3]
- 20) kombinace tagů a lemmat s interpunkcí  
např. *byci, krávy a telata; lékaři, sestry a ošetřovatelé...*  
[[tag="N..P1.\*"] [word="."] [tag="N..P1.\*"] [word="a"] [tag="N..P1.\*"]]
- 20a lemma nebo tag s interpunkčními znaménky  
např. *...pozdřav?, ...umělec?*  
[[word="."] [tag="N..S1.\*"] [tag="N..S1.\*"] [word=".\!"] [word=".\?"]]

## Příloha č. 4 – Tahák pro studenty

### Tahák – Jak pracovat s korpusem

ČNK – [www.korpus.cz](http://www.korpus.cz), v horních záložkách programy **KonText**, **SyD**, **Morfio**, **KWords**...

V KonTextu pracujeme se základním vyhledáváním či atributem lemma, pokud není uvedeno jinak. Někde je nutné pracovat se vzorci CQL, které mají toto schéma:

**[atribut="hodnota"]**

☐...pravý Alt + F, pravý Alt + G

„**uvozovky**“... Shift + ů

**Užívané atributy:** **word** (konkrétní slovní tvar), **lemma** (slovníkový tvar = infinitiv, 1. pád u jmen), **pos** (slovní druh)

#### Regulární výrazy (zástupné znaky):

.	tečka	1 libovolný znak (písmeno, číslo)
.*	tečka + hvězdička	libovolný počet znaků (0 a více!)
()	klasické závorky	shlukují více částí, např. výčet variant
	střednice	NEBO, pro výčet variant (pravý Alt + W)