

KORPUSY V JAZYKOVÉM VYUČOVÁNÍ II.

Kateřina Šormová, Karel Šebesta a kol.

Publikace *Korpusy v jazykovém vyučování* je určena učitelům, studentům učitelství i dalším zájemcům o vyučování jazyka. V první části se čtenář seznámí s teoriemi osvojování jazyka a jazykovým výzkumem, jsou mu představeny nejznámější a největší korpusy angličtiny, češtiny a výběrově i dalších jazyků s důrazem na korpusy akviziční, zejména na projekt AKCES. Pozornost je zaměřena také na typy korpusů a jejich parametry, tedy velikost, vyváženost, autentičnost, strukturní a poziční atributy. V druhé části publikace je představena práce s korpusy Českého národního korpusu, vyhledávání pomocí rozhraní KonText a pomocí aplikací SyD, Morfio, KWords, Treq a ukázka zapojení frekvenční analýzy do jazykového vyučování. Poslední část publikace tvoří pracovní listy věnované využití korpusových dat v jazykovém vyučování. Součástí pracovních listů je i podrobný popis metodického postupu a řešení zadaných cvičení.

Ediční řada Varia



FILOZOFICKÁ FAKULTA
Univerzita Karlova

KORPUSY V JAZYKOVÉM VYUČOVÁNÍ II.

Kateřina Šormová, Karel Šebesta a kol.

KATALOGIZACE V KNIZE - NÁRODNÍ KNIHOVNA ČR

Šormová, Kateřina, 1980-

Korpusy v jazykovém vyučování II. / Kateřina Šormová, Karel Šebesta a kol.. - Vydání první.

- Praha : Filozofická fakulta Univerzity Karlovy, 2019. - 1 online zdroj. - (Varia ; 78. svazek)

Anglické resumé

Obsahuje bibliografii a bibliografické odkazy

ISBN 978-80-7308-932-0 (online ; pdf)

* 81'322.2 * 81'232 * 81'24 * 811.162.3 * 811.162.3'243 * (048.8:082)

- jazykové korpusy

- jazyková akvizice

- jazyková výuka

- čeština

- čeština jako cizí jazyk

- kolektivní monografie

81 - Lingvistika. Jazyky [11]

Kniha vznikla díky podpoře projektu Zvýšení kvality vzdělávání a začleňování žáků s OMJ spojené s jeho radikální inovací (CZ.07.4.68/0.0/0.0/16_037/0000299)

a Kreativita a adaptabilita jako předpoklad úspěchu Evropy v propojeném světě (CZ.02.1.01/0.0/0.0/16_019/0000734).



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
OP Praha – půl růstu ČR



Tato publikace vznikla za podpory projektu Univerzity Karlovy Progres Q10, Jazyk v proměnách času, místa, kultury.

Recenzovali

Mgr. Svatava Škodová, Ph.D.

doc. PhDr. Milan Hrdlička, CSc.

Mgr. Jarmila Valková, Ph.D.

© Tomáš Gráf, Věra Hejhalová, Barbora Baštová, Karel Šebesta, Kateřina Šormová, 2019

© Univerzita Karlova, Filozofická fakulta, 2019

Za obsah a jazykovou správnost odpovídají autoři

Všechna práva vyhrazena

ISBN 978-80-7308-931-3 (print)

ISBN 978-80-7308-932-0 (online : pdf)

Obsah

Úvod

7

Seznámení s korpusem a korpusem a korpusem lingvistikou

| | | |
|--------|--|----|
| 1. | Korpusy, jejich typy; korpusy češtiny (Karel Šebesta) | 11 |
| 2. | Teorie osvojování jazyka a korpusový výzkum (Tomáš Gráf) | 16 |
| 2.1. | Korpus a jazykové testování | 17 |
| 2.2. | Žakovský korpus a jazyk pro specifické účely | 19 |
| 2.3. | Analýza žakovského jazyka | 21 |
| 2.3.1. | Gramatika | 22 |
| 2.3.2. | Slovní zásoba | 24 |
| 2.3.3. | Frazeologie | 27 |
| 2.3.4. | Diskurz | 29 |
| 2.3.5. | Pragmatika | 31 |
| 2.3.6. | Plynulost | 32 |
| 3. | Významné parametry korpusů (Karel Šebesta) | 34 |
| 3.1. | Velikost | 34 |
| 3.2. | Vyváženost | 37 |
| 3.3. | Autentičnost jazyka | 40 |
| 3.4. | Strukturní atributy | 43 |
| 3.5. | Poziční atributy – lingvistická (a chybová) anotace | 44 |
| 4. | Práce s korpusem (Věra Hejhalová) | 47 |
| 4.1. | Český národní korpus (ČNK) | 47 |
| 4.1.1. | Materiál/Obsah | 47 |
| 4.1.2. | Vyhledávač neboli korpusový manažer | 47 |
| 4.1.3. | Vyhledávání | 48 |
| 4.1.4. | Volba korpusu | 49 |
| 4.1.5. | Typ dotazu | 49 |
| 4.1.6. | Zadání dotazu | 51 |
| 4.1.7. | Specifikace kontextu | 52 |
| 4.1.8. | Omezení hledání | 53 |
| 4.1.9. | Výsledky hledání | 54 |

| | | |
|--------|--------------------|----|
| 4.2. | Korpusové aplikace | 60 |
| 4.2.1. | Aplikace SyD | 61 |
| 4.2.2. | Aplikace Morfio | 65 |
| 4.2.3. | Aplikace KWords | 67 |
| 4.2.4. | Aplikace Treq | 70 |

Využití korpusů ve výuce

| | | |
|--------|--|-----|
| 5. | Náměty pro strategie vyučování a pro tvorbu pracovních listů | 75 |
| 5.1. | Frekvenční analýza (<i>Kateřina Šormová</i>) | 75 |
| 5.2. | Pracovní listy (<i>Barbora Bařtová</i>) | 78 |
| 5.2.1. | Prostředí KonText | 79 |
| 5.2.2. | SyD | 120 |
| 5.2.3. | Morfio | 137 |
| 5.2.4. | KWords | 158 |

| | |
|---------------------------|-----|
| Použitá literatura | 171 |
| Seznam citovaných korpusů | 173 |
| Profily autorů | 183 |
| Rejstřík | 185 |
| Resumé/Summary | 187 |

Úvod

Milí čtenáři,

právě jste otevřeli rozšířené a doplněné druhé vydání publikace Korpusy v jazykovém vyučování. Je určena učitelům, studentům filologických oborů, jazykovým lektorům i všem dalším zájemcům o jazyk a jeho vyučování. Při přípravě této příručky jsme měli na mysli dva cíle. Prvním bylo představení jazykového korpusu jako nástroje, jehož prostřednictvím můžeme zajímavým způsobem zkoumat jazyk, pohlížet na něj z různých úhlů pohledu a v určitém smyslu si s ním i hrát. Zejména se zaměřujeme na korpusy akviziční, obsahující jazyková data z projevů mluvčích, kteří se jazyk teprve učí. Druhým cílem bylo poskytnout Vám inspiraci v tom, jak můžeme jazykový korpus využít při výuce češtiny jako cizího/druhého jazyka. Z toho důvodu tvoří poslední kapitola příručky sada dvaceti dvou pracovních listů, které by Vám měly posloužit jako návodná ukázka aktivit, které lze na základě práce s korpusem do vyučování přinést. Budeme rádi, pokud Vás přivedou k tvorbě Vašich vlastních aktivit a pracovních listů založených na korpusových datech. Přejeme Vám příjemné čtení a radost z práce s jazykem.

Autoři

Seznámení s korpusem a korpusem lingvistikou

1. Korpusy, jejich typy; korpusy češtiny

Corpus znamená latinsky tělo. Ve společenskovedním výzkumu se tohoto výrazu (resp. českého korpus) užívá přeneseně pro označení základního souboru dat, s nímž se dále pracuje. Pokud např. zkoumáme komunikaci při vyučovacích hodinách, můžeme pro tuto potřebu shromáždit určitý počet nahrávek vyučovacích hodin a tento materiál používat při výzkumu jako korpus.

Jazykový korpus¹ je rovněž takovým souborem (jazykových) dat, s nimiž je možné dále pracovat. Nejde ovšem o soubor jakýchkoli dat, ale o soubor vybudovaný (jazyková data do něj jsou vybrána, zpracována a uspořádána) podle přesných a explicitních lingvistických kritérií tak, aby reprezentoval určitou oblast jazykové praxe, jazykového úzu.

Elektronický či počítačový jazykový korpus (dále jen korpus) bývá vymezován několika různými znaky. Klasickou definici jazykového korpusu uvádějí F. Čermák a V. Schmiedtová: *Korpus se obvykle vymezuje jako strukturovaný, unifikovaný (a často též označovaný) rozsáhlý soubor jazykových dat, který je elektronicky uložený i zpracováváný; skládá se zpravidla z jednotlivých textů a jako celek si činí nárok na reprezentativnost vzhledem k vytčenému cíli* (Čermák – Schmiedtová 2004: 154). Obecně dnes platí, že jde o soubor strukturovaný, unifikovaný (aby bylo možné použít pro jeho zpracovávání a vyhledávání v něm automatických nástrojů) a elektronicky uložený a zpracováváný.

Vedle široce přístupných korpusů nekomerčních známe i korpusy komerční. Bývají to často korpusy speciální, které budují např. velká nakladatelství specializující se na vydávání slovníků, gramatik a učebnic pro jazykovou výuku, zvl. angličtiny (*Longman, Cambridge University Press* apod.).

Podle funkce a obsahu (podle zahrnutého jazykového materiálu a jeho vztahu k obvyklému, normálnímu užívání jazyka) lze rozlišit dva hlavní typy korpusů:

- a. korpusy obecné, tedy takové, které byly vytvořeny s cílem nabídnout pro výzkumné a jiné účely dostatečně velký vzorek normálního, obvyklého užívání příslušného jazyka jako celku a o nichž lze předpokládat, že normální, obvyklý jazykový úzus reprezentují;²

1 Seznam všech korpusů zmíněných v této publikaci naleznete na straně X.

2 Reprezentativnost ve vztahu k jazyku jako celku je pouze relativní, můžeme ale říci, že tvůrci korpusů k reprezentativnosti jako cíli směřují. Srov. vyjádření F. Čermáka (in Čer-

- b. korpusy speciální, které takovou ambici, reprezentovat normální, obvyklý jazykový úzus, nemají, jsou zaměřeny často na užívání jazyka nějakou zvláštní skupinou či jednotlivcem, příp. zachycují jazyková data, o nichž z nějakého jiného důvodu nelze předpokládat, že normální užívání jazyka v jeho celku reprezentují.

Jedním z rozšířených typů speciálních korpusů jsou korpusy akviziční, tedy takové, které *slouží primárně studiu procesů osvojování jazyka, včetně tzv. pozdějšího jazykového vývoje a užívání jazyka mluvčími, kteří (daný) jazyk neovládají na úrovni odpovídající úrovni dospělého rodilého mluvčího, sekundárně pak mohou plnit a plní řadu důležitých funkcí v didaktice jazyka: jsou významným zdrojem dat při tvorbě učebnic a učebních pomůcek, jako jsou slovníky nebo gramatiky, při přípravě testů a jazykových cvičení různého typu a uplatňují se i přímo jako didaktický nástroj v jazykové výuce* (Šebesta – Škodová 2012: 6).

Pokud jde o povahu jazykových dat akvizičních korpusů, jde v naprosté většině případů o doklady užívání jazyka např. dětmi předškolního věku, mládeží ve školním věku, nerodilými mluvčími apod. K tomu se přiřazují (především z praktických důvodů) data spojená s péčí o osvojování jazyka, tedy nahrávky vyučovacích hodin, jazyková data z učebnic a dalších učebních materiálů apod.

Akviziční korpusy takto vymezené mohou sloužit i výzkumům jinak zaměřeným: např. studiu jazyka mládeže jako sociální skupiny, popř. studiu jazyka některých subkultur mládeže či jejích sociálně vymezených podskupin, studiu vývojových trendů jazyka obecně (např. při srovnání jazyka mládeže s jazykem starších generací) apod.

Podtypem akvizičních korpusů jsou korpusy žákovské, sloužící studiu osvojování jazyka nerodilými mluvčími. Obsahují data z užívání jazyka nerodilými mluvčími a také jazyková data z výuky cizího jazyka.

Trochu volněji se k akvizičním korpusům řadí korpusy zaměřené na studium tzv. zděděného jazyka (*heritage language*) a jeho osvojování, tedy osvojování a užívání jazyka příslušníky krajanových komunit v zahraničí. S akvizičními korpusy je spojuje skutečnost, že i ony slouží studiu osvojování jazyka a sekundárně didaktickým účelům (např. praxi různých krajanových škol a kurzů pro krajany, Českých škol bez hranic, učebních pomůcek pro ně apod.).

Od akvizičních korpusů se odlišují tím, že korpusy zaměřené na jazyk krajanů mohou být budovány a využívány také pro potřeby studia jazykové atrice, rozpadu zděděného jazyka, tedy jako korpusy někdy označované

mák a kol., eds. 2011: 16) o tom, že lingvistický korpus je takový korpus, „který umožňuje vyvážené a reprezentativní zkoumání relativně celého jazyka, nikoliv jeho části, často v nezdůvodněných proporcích“.

termínem korpusy atriční. Krajanské korpusy jsou také důležitým zdrojem dat pro výzkumy jinak zaměřené, poskytují např. unikátní možnosti studia nářečí, jejichž prvky bývají někdy v krajanském jazyce zachovány, pro studium jazykového kontaktu obecně apod.

Z praktických důvodů je ovšem účelné korpusy akviziční a atriční neoddělovat příliš striktně: při budování i využívání obou těchto typů korpusů je nutné se vyrovnávat s podobnými problémy, jako jsou např. nestandardní jazyková data, obtížnost jejich získávání, vysoký podíl manuální práce při jejich budování, a užívá se k tomu podobných specificky pro tyto účely vyvinutých korpusových nástrojů (viz dále).

Podstatně volnější vztah k akvizičním korpusům mají korpusy sloužící studiu jazykových poruch a jejich terapie. Je možné označit je jako korpusy terapeutické (nejde o termín zavedený). S akvizičními korpusy je sblížuje skutečnost, že zachycují data z užívání jazyka mluvčími, kteří daný jazyk neovládají na úrovni dospělého rodilého mluvčího; při jejich budování se tedy mohou využít podobné nástroje a pravidla (např. přepisů) jako u ostatních akvizičních korpusů. Je proto z praktických důvodů účelné budovat je ve spolupráci s korpusy akvizičními.

Pro výuku češtiny jako cizího jazyka můžeme využívat především obecných korpusů češtiny a také specificky pro tento účel budovaných žákovských korpusů češtiny jako cizího/druhého jazyka. V jisté míře se mohou uplatnit i ostatní typy korpusů akvizičních, především akviziční korpusy češtiny jako prvního jazyka (L1). V další části kapitoly se tedy budeme věnovat prioritně oběma prvně uvedeným typům korpusů, korpusům obecným a žákovským; v některých případech, kde to pro nás může být zajímavé, se dotkneme i akvizičních korpusů prvního jazyka.

Počátky počítačových korpusů se zpravidla kladou do 60. a 70. let, a to jak korpusů obecných (jako první se většinou uvádí tzv. *Brown Corpus*, celým názvem *Brown University Standard Corpus of Present-Day Edited American English*, budovaný Henrym Kučerou a Nelsonem Francisem od r. 1961 a zveřejněný 1964, a *Lancaster-Oslo/Bergen Corpus (LOB Corpus)*, vytvořený v letech 1970–1978 a zpracovávající jako britský protějšek *Brown Corpusu* angličtinu britskou), tak i speciálních – jako jeden z prvních korpusů speciálních se uvádí např. *American Heritage Intermediate Corpus*, zahrnující vzorky z nejčtenějších knih americké mládeže ve školním věku v roce 1969.

První obecné počítačové korpusy ovšem navazují na podstatně delší tradici systematických sběrů jazykových dat a tvorby jazykových korpusů neelektronických. Odkazuje se v této souvislosti většinou na sběry citací, zpravidla pro lexikografické účely.³

3 Přehled korpusových počátků pro oblast angličtiny i češtiny podává např. Michal Šulc (Šulc, 1999).

To platí i pro korpusy češtiny, také česká korpusová lingvistika navazuje na dlouhou a bohatou tradici jinak než elektronicky zpracovávaných sběrů. První snahy o vytvoření centrální národní instituce, která by koncentrovaně rozvíjela korpusové aktivity pro češtinu, se objevují přibližně od druhé poloviny 80. let. V r. 1994 vznikl na FF UK pod vedením F. Čermáka Ústav Českého národního korpusu, který budování korpusů českého jazyka, obecných i speciálních, zajišťuje.

V současné době nabízí:

- synchronní korpusy psaného jazyka řady SYN,
- specializované korpusy psaného jazyka různého zaměření, včetně akvizičních (viz dále),
- synchronní korpusy mluveného jazyka řady ORAL a nový korpus ORTOFON,
- specializované korpusy mluveného jazyka různého zaměření, včetně akvizičních (viz dále),
- diachronní korpus *DIAKORP*,
- paralelní korpus *InterCorp*,
- srovnatelné korpusy *Aranea*, psané jednojazyčné webové a speciální korpusy jiných jazyků, než je čeština.

Pro učitele češtiny jako cizího jazyka mají význam vedle korpusů akvizičních primárně obecné korpusy češtiny, tedy korpusy řady SYN a ORAL, resp. ORTOFON, pro vyšší a nejvyšší úroveň a pro některé speciální účely je využitelný korpus paralelní – *InterCorp*.

Vedle korpusů Českého národního korpusu je pro učitele významný rovněž *Pražský závislostní korpus*, budovaný Ústavem formální a aplikované lingvistiky MFF UK; z něj vycházejí i některé didaktické pomůcky zaměřené na výuku syntaxe.

Korpusy akviziční navazují na neméně dlouhou tradici systematického sběru jazykových dat pro účely studia osvojování a pozdějšího vývoje jazyka dětí předškolního i školního věku, studia jazyka mládeže jako specifické jazykové variety, jazyka nerodilých mluvčích, jazyka mluvčích postižených jazykovou poruchou apod.⁴

V roce 2005 byl pod vedením K. Šebesty zahájen projekt *AKCES* (Akviziční korpusy češtiny) při Ústavu českého jazyka a teorie komunikace FF UK, na jeho vytváření se však významně podílejí i další pracoviště.⁵

4 K této akviziční tradici sběrů v mezinárodním i domácím výzkumu a ke vzniku prvních akvizičních korpusů počítačových podrobněji viz Šebesta, 2010.

5 Ústav teoretické a počítačové lingvistiky, Ústav formální a aplikované lingvistiky MFF UK, Ústav Českého národního korpusu, Ústav jazykové a odborné přípravy UK a další.

Projekt AKCES je pojat poměrně široce:⁶ zahrnuje postupné vytváření databází a návazně budování korpusů psaných i mluvených zaměřených na:

- a. jazyk dětí předškolního věku,
- b. jazyk dětí a mládeže od 5 do 24 let,
- c. jazyk nerodilých mluvčích češtiny,
- d. jazyk sociokulturně znevýhodněných skupin,
- e. jazyk krajanských komunit,
- f. jazyk ve vzdělávacím kontextu.
- g. Poskytuje podporu rovněž korpusům terapeutickým.

Jednotlivé korpusy jsou zveřejňovány zčásti v rámci Českého národního korpusu, zčásti v rámci Centra jazykové výzkumné infrastruktury LINDAT/CLARIN⁷ Ústavu formální a aplikované lingvistiky MFF UK a nově též na stránkách Ústavu formální a aplikované lingvistiky FF UK.

Pro výuku nerodilých mluvčích jsou relevantní zejména:

- korpusy žákovské (jazyka nerodilých mluvčích); z nich byly dosud v rámci Českého národního korpusu⁸ zveřejněny korpusy řady CZESL (polovina korpusu CZESL-PLAIN, korpusy CZESL-SGT a CZESL-MAN – druhý z nich není dosud veřejně přístupný), v rámci Centra LINDAT/CLARIN pak korpus AKCES 5 a další korpusy z řady CZESL derivované, např. CZESL-UD; volně se k nim řadí i česká část korpusu MERLIN, zpracovaná v rámci samostatně financovaného německo-italsko-českého projektu;⁹
- korpusy jazyka žáků ze sociokulturně znevýhodňujícího prostředí – z korpusů zveřejněných v rámci Českého národního korpusu sem patří cca čtvrtina korpusu CZESL-PLAIN, připravuje se k zveřejnění srovnávací korpus SKRIPT2015, jehož polovinu tvoří data mluvčích ze sociokulturně znevýhodněných komunit; z korpusů zveřejněných v rámci LINDATu jsou to korpusy AKCES 4 a mluvený korpus ROMi_1.0;
- korpusy písemných i mluvených projevů českých žáků – v rámci Českého národního korpusu byl zveřejněn korpus písemných prací žáků SKRIPT2012; jejich mluvené projevy jsou zachyceny zčásti ve školních dialogích přepsaných a zveřejněných v korpusech SCHOLA 2010 (na stránkách ČNK) a AKCES 2 ver 2.

6 Podrobnější informace o celém projektu jsou na stránkách <http://akces.ff.cuni.cz/>.

7 <https://lindat.mff.cuni.cz/cs/>

8 https://kontext.korpus.cz/first_form

9 <https://merlin-platform.eu/>

2. Teorie osvojování jazyka a korpusový výzkum

Nauka o jazykové akvizici se zabývá zkoumáním procesů osvojování mateřského či cizího jazyka. Jakkoliv doklady o spekulacích o podstatě tohoto procesu nacházíme již ve starověku, jako vědní obor a součást aplikované lingvistiky se ustavuje až v polovině 20. století, kdy nachází své místo v univerzitních studijních programech.

Jako každý vědecký obor pracuje i nauka o jazykové akvizici s konkrétními daty a od 80. let 20. století tato data stále častěji pocházejí z tzv. akvizičních korpusů. Jde o speciální typ jazykových korpusů, které si nejlépe představíme jako počítačově zpracované a prohledávatelné sbírky textů, jejichž původci jsou ti, kdo si jazyk osvojují: děti osvojující si mateřštinu nebo studenti (ať už v dětském, či dospělém věku) osvojující si další, zpravidla cizí, jazyk. Existují přitom různé typy těchto korpusů v závislosti na médiu (psaný či mluvený korpus), typu jazyka (mateřský či cizí) či charakteristice mluvčího (dítě, dospělý, mluvčí s řečovou poruchou, rodilý či nerodilý mluvčí, mluvčí ze sociokulturně či jinak znevýhodněné komunity aj.). Společně pak mají tyto korpusy to, že zachycují tzv. mezijazyk, tedy jazyk ve vývojové fázi, která ještě nedosahuje úrovně obvyklé pro dospělého rodilého mluvčího. Akviziční korpusy jsou tak sbírkou dokladů o tom, jak osvojování probíhá a jaké jsou typické rysy osvojovaného jazyka. Umožňují tak hledat odpověď na řadu klíčových otázek, kterými se nauka o akvizici zabývá, jako jsou například role jazykového transferu (tj. vlivu mateřštiny na osvojování cizího jazyka) a odlišnosti mezi mezijazykem a jazykem rodilého mluvčího, jakož i odlišnosti ve vyjadřování různých mluvčích. Je tak možné srovnávat jazykový projev studentů s různou mateřštinou, kteří si ale osvojují stejný cizí jazyk (např. angličtinu).

Tradiční výzkum v oblasti jazykové akvizice často pracoval s omezenými vzorky dat a nezřídka i s intuicí badatelů a učitelů cizích jazyků. Často tak byly výsledky výzkumu přijímány s výhradou, že se zjištěné skutečnosti mohou týkat právě jen onoho omezeného vzorku mluvčích a nedají se zobecnit. Akviziční korpusy naproti tomu poskytují velké množství dat sesbírané od desítek, stovek či tisíců mluvčích a poskytuje tak rozsáhlou bázi pro empirické ověřování hypotéz s velkou možností zobecnění závěrů.

Specifickým a zřejmě i nejběžnějším typem akvizičních korpusů jsou korpusy žákovské (angl. learner corpora). Jsou vystavěny z textů nerodi-

lých mluvčích, kteří se učí cizí jazyk. Jazyk, který vyprodukuje, pak označujeme jako jazyk žákovský (jakkoliv nezáleží na věku studenta; žákem je označován každý, kdo se učí). Zdrojem textů mohou být psané školní práce, materiály z jazykových zkoušek, přepsané nahrávky mluveného projevu či záznamy řečových úloh sestavených za účelem vytvoření samotného korpusu. V květnu 2018 čítal seznam *Learner corpora around the world*¹⁰ 167 žákovských korpusů. Další korpusy budou jistě i nadále vznikat.

Dalším typem akvizičních korpusů jsou korpusy osvojování mateřštiny. Nejznámější z nich, korpus CHILDES¹¹ (*Child Language Data Exchange System*), začal vznikat již v r. 1984 a shromažďuje data ze 130 specializovaných korpusů dětské řeči v 26 jazycích. Korpus se posléze připojil k projektu *Talk-Bank*, který obsahuje korpusy akvizice mluvčích s poruchami řeči (afázie), korpusy osvojování cizích jazyků, korpusy jazyka ve školních třídách a korpusy pro konverzační analýzu.

Akviziční korpusy se vybavují různými systémy anotace. Ty přejímají buď z obecné korpusové lingvistiky (např. anotace slovněduhová), nebo vytvářejí vlastní anotační systémy zachycující specifické rysy akvizice (např. jazykové odchylky, chyby či jiná specifika). Takto vybavené korpusy pak umožňují sofistikované metody vyhledávání a na nich založené analytické postupy, které dokážou odkrýt charakteristické rysy vyvíjejícího se jazyka, na jejichž základě je nežádka možné formulovat i pedagogické závěry.

2.1. Korpus a jazykové testování

Využití korpusů pro účely jazykového testování je jedním z nejnovějších příkladů aplikace jazykových korpusů a je to bezesporu oblast, kde se do budoucna očekává velká aktivita, a to především v oblasti definování a popisování úrovní pokročilosti. Současný stav výzkumu v této oblasti vychází především z možností sběru dat při jazykových zkouškách. Takto získané vzorky žákovského jazyka, u nichž je školenými examinátoři stanovena jazyková úroveň např. dle Společného evropského referenčního rámce (SERR), se analyzují například z hlediska četnosti určitých jevů a na základě těchto analýz je pak možné popsat, které jazykové prostředky jsou charakteristické pro danou úroveň pokročilosti. Lingvistům se tak daří reagovat na časté kritické hlasy, které si všímají, že Společný evropský referenční rámec není dostatečně specifický a v rámci popisu dovedností na jednotlivých

10 Centre for English Corpus Linguistics (date of access): Learner Corpora around the World. Louvain-la-Neuve: Université catholique de Louvain, <https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html>.

11 <https://childes.talkbank.org>

úrovních užívá příliš obecný jazyk. Výsledky tohoto výzkumu mají potenciál přispět k lepší srovnatelnosti.

Zatím nejvýznamnějším a nejrozsáhlejším počinem v této oblasti je projekt *English Profile*¹² (organizovaný University of Cambridge, Cambridge University Press a Cambridge English Language Assessment a podpořený Evropskou komisí). Z dat z jazykových zkoušek na všech úrovních angličtiny dle SERR byl sestaven korpus a na jeho základě pak seznamy slovní zásoby (*English Vocabulary Profile*) a gramatiky (*English Grammar Profile*) typické pro jednotlivé úrovně SERR. Tuto bezplatnou databázi mohou využívat učitelé, studenti či autoři jazykových materiálů a ověřit si, jaký konkrétní jazyk mluvčí různých úrovní angličtiny dokážou používat. Očekává se, že tento výzkum se zpětně promítne i do vývoje dokonalejších, empiricky založených technik zjišťování jazykové pokročilosti.

Největším žákovským korpusem, který vznikl na základě sběru materiálů z testů, je *Cambridge Learner Corpus*, který obsahuje v současnosti přes 57 milionů slov psaného žákovského jazyka od více než čtvrt milionu mluvčích. Korpus je přitom i nadále rozšiřován. Dalším významným projektem je *The Longman Learners' Corpus*, který poskytl podklady pro sestavení jazykových doporučení v žákovském slovníku *Longman Active Study Dictionary*. Mnoho nových výzkumných projektů nyní vzniká na základě mluveného korpusu žákovské angličtiny *Trinity-Lancaster Corpus*, který je sestaven z nahrávek a přepisů ústních zkoušek z angličtiny při zkouškách *Graded Examinations in Spoken English* organizovaných Trinity College v Londýně. Zatím zde však jde pouze o úrovně B1–C1.

Pro účely vývoje jazykových testů se však využívají i korpusy jazyka rodilých mluvčích. Například *T2K-SWAL Corpus* je souborem akademických textů, které se analyzují z pohledu analýzy diskurzu a žánrů. Výsledek tohoto výzkumu vedl k sestavení *Test of English as a Foreign Language, TOEFL iBT*.

Největší výhodou využití korpusové lingvistiky v oblasti jazykového testování je bezesporu to, že jazykové korpusy (ať už žákovského jazyka, nebo jazyka rodilých mluvčích) umožňují empiricky ověřit intuici tvůrců testových materiálů a připravit tak testy, které ověřují odpovídající jazykovou úroveň. To se může dít nejen na základě výzkumu toho, co mluvčí skutečně v jazyce dokážou, ale i toho, kde se odchylují od běžných norem (např. je možné zkoumat, jaké chyby jsou typické pro různé úrovně pokročilosti). Tvůrci jazykových testů se přitom mohou opřít o seznamy slov a jazykových struktur vygenerovaných z žákovských korpusů pro jednotlivé úrovně pokročilosti a mohou na datech z korpusu rovněž zkoumat, jak různé typy úloh ovlivňují výkon v testu a jak se určité gramatické či lexikální jevy

12 <http://www.englishprofile.org>

v souvislosti s rostoucí pokročilostí proměňují. Cílem tohoto výzkumu je přispět ke zlepšování podoby a vlastností jazykových testů.

Pro účely jazykového testování se však dají využít i korpusy běžné (pro angličtinu např. korpusy *British National Corpus* či *Corpus of Contemporary American English*), a to jako zdroj textů, které si mohou učitelé dle zadaných kritérií vyhledat a vytvořit si tak testy na slovní zásobu či gramatiku. Takto získaná korpusová data jsou autentická a jazykově obvykle i správná (co je v korpusu v daném registru či žánru frekventované, to zpravidla odpovídá zavedenému úzu/normě, a tudíž odráží kodifikaci).

2.2. Žákovský korpus a jazyk pro specifické účely

Jazykem pro specifické účely označujeme jazyk, který je užíván například v pracovním či akademickém prostředí a který se řídí určitými pravidly příslušného funkčního stylu (např. styl odborný, publicistický a administrativní), nezřídka je charakterizován i specifickou slovní zásobou. Jeho znalost umožňuje uživatelům začlenit se do skupiny lidí, odborníků či profesionálů, kteří takový jazyk typicky používají a kteří mají specifické komunikační potřeby a zvyklosti.

Jazyk pro specifické účely je od 60. let 20. století intenzivně zkoumán, což má mj. dopad i na jazykové vyučování, a to obzvláště pro angličtinu, která plní roli jazyka mezinárodní komunikace právě v profesní oblasti. Je také v tomto ohledu nejlépe popsána. Existují tak četné učebnice odborného jazyka, slovníky, aplikace a specializované jazykové kurzy. Rozlišuje se přitom mezi angličtinou pro pracovní účely (*English for Occupational Purposes*) a angličtinou pro účely akademické (*English for Academic Purposes*). Angličtina pro pracovní účely se nezaměřuje pouze na specifickou slovní zásobu (např. terminologii), ale především na to, jakým způsobem se v těchto oblastech lidské činnosti komunikuje, aby komunikace byla efektivní a pro dané účely vyhovující. To se týká jak jazyka psaného (např. obchodní korespondence, instruktážní texty), tak mluveného (např. komunikace lékařů s pacienty, komunikace v leteckém provozu).

Výzkum angličtiny pro akademické účely se zabývá celou škálou užití angličtiny v akademickém prostředí od interakcí mluvených (např. prezentace na konferencích) až po psaní akademických textů. Tato oblast je natolik široká, že se dále rozlišuje mezi angličtinou pro obecné a pro specificky oborové akademické účely. Do této oblasti se nezřídka zahrnují i specifické jazykové a studijní dovednosti a návyky (např. pořizování poznámek, práce s literaturou, organizace textu), ale například i prezentační, rétorické techniky. Pedagogické úsilí v této nesmírně stratifikované oblasti pak směřuje především k tomu, aby byly naplněny potřeby studentů a uživatelů jazyka v té které specializaci.

Korpusová lingvistika hraje v popisu jazyka pro specifické účely nezastupitelnou roli. Disponuje totiž technikami, jak shromáždit dostatečně velké vzorky reprezentativních ukázek odborného jazyka, na jejichž základě je možné popsat jeho typické rysy. To je důležité jak pro tvorbu pedagogických materiálů, tak pro zkoumání projevu těch, kdo si tento jazyk osvojují v rámci jiného jazyka, než je jejich mateřština. Zkoumají se např. frekvenční vzorce v jednotlivých stylech a je tak možné velmi přesně určit, která specifická slovní zásoba je pro daný styl charakteristická. Na základě korpusu *The Academic Corpus*¹³ tak například vznikl seznam akademické slovní zásoby pro angličtinu (*Academic Word List*¹⁴), který je pro mezinárodní studenty neocenitelnou pomůckou při studiu akademické angličtiny a při psaní akademických textů. Využití pak samozřejmě nachází i při výuce a sestavování jazykových testů.

Vedle textů napsaných rodilými mluvčími se však analyzují i akademické texty nerodilých mluvčích. Vznikají tak speciální žákovské korpusy akademického žákovského jazyka. I zde zcela převládají korpusy angličtiny. Jde obvykle o databáze psaných úloh (esejů, písemných zkoušek atp.) vzniklých na univerzitách. Jazyk, který v daném kontextu studenti používají, je charakteristický jak užitím terminologie, tak užitím specifických typů argumentace a postupů při výkladu. Jedním z prvních takových korpusů je *International Corpus of Learner English (ICLE)*, který začal vznikat v Belgii, na *Université catholique de Louvain*, v 80. letech 20. st. Práce na něm stále pokračuje, ač korpus v současné době obsahuje eseje v rozsahu 3,7 milionu slov od pokročilých studentů angličtiny z 16 různých zemí, respektive s 16 rozdílnými mateřskými jazyky.

Na stejném pracovišti se nyní rodí nový korpus akademické angličtiny s názvem *Varieties of English for Academic Purposes (VESPA)*. Korpus je tvořen národními subkorpusy mluvčích s rozdílnou mateřštinou. Je tak možné nejen zkoumat specifika vyvíjejícího se akademického jazyka u studentů, ale zároveň srovnávat, do jaké míry může být jeho podoba ovlivněna mateřštinou. Dalšími významnými korpusy v této oblasti jsou *British Academic Written English (BAWE)* a americký *Michigan Corpus of Upper-level Student Papers (MICUSP)* a jejich mluvené protějšky *British Academic Spoken English (BASE)* a *Michigan Corpus of Academic Spoken English (MICASE)*. Tyto čtyři korpusy jsou přitom ze zhruba 80 % tvořeny psaným projevem rodilých mluvčích, což umožňuje srovnávat jazyk cizojazyčných studentů s jazykem rodilých mluvčích. To je v žákovské korpusové lingvistice běžný přístup a nezřídka jsou žákovské korpusy doplňovány o kontrolní korpusy textů rodilých mluvčích, které jsou koncipovány tak, aby měly stejnou strukturu

13 <https://www.victoria.ac.nz/lals/resources/academicwordlist/information/corpus>

14 Pro češtinu je dostupný nástroj *Akalex*, viz www.korpus.cz/akalex.

a podobaly se kvůli srovnatelnosti v co největší možné míře. Korpus ICLE je takto doplněn o *Louvain Corpus of Native English Essays (LOCNESS)*, tedy korpus stejně zadaných esejů psaných rodilými mluvčími. Takový přístup však skrývá i jedno zásadní úskalí, a tím je implicitní předpoklad, že text rodilého mluvčího je z podstaty věci dokonalejší. Stranou přitom zůstává, zda jsou vybraní autoři z řad rodilých mluvčích kvalitními autory.

2.3. Analýza žákovského jazyka

V 50. letech 20. století se někteří lingvisté¹⁵ zabývali výzkumem jazyka za účelem zdokonalení metod jeho výuky. Vycházeli přitom z předpokladu, že je snazší naučit se jazyk, který je podobnější žákově mateřštině, a že podrobný popis podobností a rozdílů mezi mateřským jazykem a jazykem cizím je užitečnou pomůckou pro výuku a pro tvorbu jazykových učebnic. Předpokládali, že žáci budou chybovat v těch aspektech, které jsou rozdílné, a naopak snazší pro ně budou ty rysy, které jsou si mezi jazyky podobné. Domnívali se, že když se výuka bude zaměřovat především na tyto odlišnosti, žáci v nich nebudou chybovat. Skutečnost se ukázala být složitější a postupným výzkumem se ukázalo, že výuka vycházející ze srovnávání jazyků není zárukou bezchybného projevu žáků.

Toto období bylo rovněž počátkem zájmu o žákovský jazyk, a to především o chyby, kterých se žáci dopouštějí při psaní či mluvení. Chyby byly analyzovány ze všech stran a lingvisté se je pokoušeli pochopit v naději, že objeví důvod, proč žáci chybojí, a budou tak moci vyvinout dokonalejší výukové metody, které žákům umožní nechybovat. Stále více lingvistů však upozorňovalo, že chyby jsou nedílnou součástí učení a že představují doklad o vývoji pokročilosti a o tom, že učení se cizímu jazyku je systematický proces. Tuto teorii završil v roce 1972 americký lingvista Larry Selinker formulací teorie tzv. mezijazyka.

Selinkerův mezijazyk je dynamický, proměnlivý systém, který se v žákově myslí formuje na základě jeho zkušeností s jazykem a který mu umožňuje cizí jazyk používat. To je možné díky pravidlům, která si vytváří sám žák, a to na základě výuky, učení se, kontaktu s osvojovaným jazykem a dalších faktorů. Existence mezijazyka je základním principem, z něhož vychází výzkum jazykové akvizice. Data pro tento výzkum pak pocházejí stále častěji z žákovských korpusů. Ty umožňují analyzovat osvojování gramatiky, slovní zásoby, frazeologie, diskurzu, stylistiky, pragmatiky a s nástupem videokorpusů v budoucnosti i výzkum nonverbální komunikace.

15 Např. Robert Lado, Charles Fries, Pitt Corder.

2.3.1. Gramatika

V centru zájmu při zkoumání žákovského jazyka stojí gramatika a slovní zásoba, které jsou vnímány jako dva základní komponenty znalosti cizího jazyka. Cílem žákovské korpusové lingvistiky v této oblasti je popsat, jakým způsobem studenti používají gramatiku cizího jazyka, který se učí, a stanovit příslušné pedagogické postupy, které z těchto informací vyplývají. Žákovská korpusová lingvistika má tedy velmi blízko k pedagogice a didaktice a přispívá k vytváření tzv. pedagogických gramatik, tzn. popisů cizojazyčných gramatických systémů při současném zohlednění možných postupů při jejich výuce.

K tomu, aby mohla být zkoumána gramatika v žákovském jazyce za použití žákovských korpusů, musí být tyto korpusy doplněny o anotace. Anotace jsou značky, které lingvisté doplňují do korpusu proto, aby se v něm dal vyhledávat nejen konkrétní řetězec znaků (např. slovo), ale i specifická jazyková informace např. o slovním druhu či přítomnosti chyby. Rozlišujeme tak především značkování morfologické a chybové. Zatímco morfologické značkování je možné provést víceméně automaticky s využitím počítačových programů, značkování chybové se provádí ručně a je značně náročné na čas. Při morfologickém značkování je každému slovu přiřazena informace o slovním druhu a příslušném jazykovém tvaru. Díky tomu můžeme při vyhledávání v korpusu zadat např. české slovo *bez* a stanovit přitom, chceme-li, aby výsledky vyhledávání zahrnuly slovo *bez* jako podstatné jméno, či jako předložku. Pro žákovské korpusy jsou pak typické značky chybové, které se přiřazují všem výskytům chyb v jazykovém projevu, ať už jsou na úrovni gramatické, pravopisné, lexikální, či jiné.

Chybová anotace se často vkládá přímo do textu. Označuje se přitom chybný tvar a často se doplňuje i správná varianta (tzv. emendace). Jako příklad uveďme systém užívaný pro chybové značkování korpusu *ICLE* (viz výše). Tento systém využívá značky složené z písmen. Ty se vkládají před chybně užitý tvar. Na prvním místě bývá označení typu chyby (tj. jde-li o pravopis, gramatiku, slovní zásobu, lexikálně-gramatický jev či slovosled). Ve větě označené Příklad 1 je ukázáno, jakým způsobem je možné označkovat chybné užití gramatického času v anglické větě. Student ve větě chybně užil přítomný čas, proto je před chybným tvarem slovesa *live* vložena značka (GVT). Ta označuje, že jde o gramatickou chybu (G) týkající se třídy sloves (V, anglicky *verb*) a kategorie času (T, anglicky *tense*). Po slovese je vložena emendace, tedy návrh anotátora na to, jak by mohla znít správná věta. V tomto případě anotátor navrhuje, aby místo tvaru přítomného času slovesa *live* bylo užito předpřítomného průběhového času.

původní věta: *I live in Prague for ten years*

anotovaná věta: *I (GVT) live shave been livings in Prague for ten years*

Příklad 1: Ukázka chybové anotace anglické věty – gramatika

V takto označovaném korpusu je pak možné jednotlivé chybové značky vyhledávat a nechat si zobrazit třeba právě všechny věty s chybou v použití časů. Ty se pak dají dále analyzovat, případně je možné jich využít pro tvorbu cvičení pro výuku. Specifické uplatnění tyto postupy nacházejí při tzv. kontrastivní analýze mezijazyka, při níž se porovnává psaný nebo mluvený projev žáků s odlišnými mateřskými jazyky (např. angličtina Čechů a Španělů). Dají se tak odhalit podobnosti či rozdíly a zkoumat, z čeho tyto rozdíly pramení. Častým zdrojem chyb přitom bývá tzv. jazykový transfer, kdy si žák přenáší do svého cizojazyčného projevu prvky ze své mateřštiny. Jako ukázka nám znovu může posloužit Příklad 1, u něhož se můžeme domnívat, že chybné užití přítomného času v angličtině pramení z toho, že je v odpovídající větě v češtině používán právě přítomný čas (*Bydlím v Praze už deset let.*).

Je-li označován celý korpus, je možno přistoupit k analýze chyb. Chyby se vyhledají a roztrídí do kategorií podle různých kritérií. S použitím statistických nástrojů je pak možné ověřovat hypotézy o souvislosti mezi chybovostí a nejrůznějšími žákovskými proměnnými, jako jsou např. jazyková úroveň, věk, mateřský jazyk, délka studia nebo pobyt v zahraničí.

Na základě chybové analýzy je pak možné sestavit katalogy chyb, které mohou být návodné pro výuku nebo se dají využít při psaní výukových materiálů. Jazykové učebnice a slovníky určené pro žáky angličtiny tak dnes nezřídka obsahují informace o častých chybách. Jsou návodné nejen pro žáky, ale i pro učitele. Existuje dokonce i slovník chyb v angličtině,¹⁶ který obsahuje více než 2500 příkladů běžných chyb, vysvětlení daných jevů a ukázky správného užití z *British National Corpus*.

Zkoumat se ale nemusí pouze chyby. Srovnáváním jazyka rodilých mluvčích s jazykem žákovským je možné zjistit, co je pro žákovský jazyk charakteristické například frekvenčně, tj. které jevy se objevují častěji a které méně často. Mluví se pak o nadužívání či naopak menší míře používání určitých gramatických jevů. Takové analýzy jsou pak užitečné například pro výuku jazyka pro specifické účely (např. akademická angličtina, viz výše).

Jedním z nejzajímavějších a pro učitele angličtiny zřejmě i nejužitečnějších využití žákovských korpusů v oblasti gramatiky je projekt English

16 Turton, N.D. and J.B. Heaton (1996). *Longman Dictionary of Common Errors*. Harwich: Longman.

Grammar Profile.¹⁷ Jde o volně přístupný internetový zdroj, který je nadstavbou Společného evropského referenčního rámce (Council of Europe, 2001). Umožňuje vyhledávat gramatické jevy podle jazykové úrovně žáků. Projekt vznikl sestavením žákovského korpusu textů z jazykových zkoušek pro jednotlivé úrovně angličtiny dle Společného evropského referenčního rámce (*Cambridge Learner Corpus*) a anotací gramatických jevů, které se v něm vyskytují. Ty pak byly rozřazeny podle pokročilosti mluvčích a bylo stanoveno, které gramatické jevy jsou charakteristické pro danou úroveň pokročilosti. Prostřednictvím projektu tak lze zkoumat nejen vývoj osvojování anglické gramatiky od začátečnicků až po velmi pokročilé, ale projekt lze využít i prakticky, například pro sestavování osnov pro výuku či pro tvorbu testů, učebnic a jiných jazykových materiálů. Pomocí uživatelsky přívětivého a přehledného internetového rozhraní je možné si vybrat buď konkrétní jazykovou úroveň a prozkoumat, které gramatické jevy jsou pro ni charakteristické, nebo je možné naopak vyhledat konkrétní gramatický jev a zjistit, od kterých úrovní se v žákovské angličtině začíná typicky objevovat. Vše je doprovázeno množstvím autentických ukázek, a to v neopravené i opravené verzi, takže je možné sledovat, v čem se v daném jevu typicky chybuje.

2.3.2. Slovní zásoba

Korpusové studie slovní zásoby patří v korpusové lingvistice mezi nejčastější. Často přitom pracují s frekvencí výskytu jednotlivých slov, což umožňuje sestavit seznamy slov, která jsou více či méně běžná v určitých typech textů či stylech. Toho lze v kontextu vyučování využít například pro sestavování seznamů klíčové slovní zásoby,¹⁸ která – jak označení napovídá – obsahuje běžná slova, jež jsou nepostradatelná pro komunikaci a jsou známa většině roditelých mluvčích. Přitažlivost tohoto konceptu spočívá v myšlence, že žák, který ovládá toto lexikální minimum, by měl znát podstatnou slovní zásobu potřebnou pro většinu běžných situací. Příkladem takového lexikálního minima je třeba definiční slovník využívaný v *Oxford Advanced Learner's Dictionary*. V tomto slovníku pro pokročilé studenty angličtiny jsou všechna slova definována na základě klíčové slovní zásoby čítající 3000 slov. Pokud student tato slova ovládá, měl by být schopen – podobně jako tento slovník – vyjádřit vlastně cokoliv. Skutečnost je však naneštěstí o něco složitější, což je dáno mimo jiné tím, že řada běžných, každodenních slov, která jsou pro uživatele jazyka důležitá (např. židle, stůl, kalhoty), ve skutečnosti v textech nepatří mezi ta nejčastější, a neobjeví se tak v seznamech vytvářených z korpusu právě na základě frekvence.

17 <http://www.englishprofile.org/english-grammar-profile/egp-online>

18 Angl. *core vocabulary*.

Na tento nedostatek výborně reaguje jeden z nejužitečnějších korpusových projektů pro učitele i studenty angličtiny, *English Vocabulary Profile*.¹⁹ Pochází od stejných tvůrců jako *English Grammar Profile* a také využívá jako zdroj dat *Cambridge Learner Corpus*. Jde o unikátní projekt, který doplňuje Společný evropský referenční rámec o slovní zásobu, která je typicky využívána na jednotlivých úrovních pokročilosti. Jde přitom nejen o slova, ale i o slovní obraty, fráze a idiomy. Jak však sami tvůrci projektu tvrdí, nejde jen o seznam slov a slovních spojení. Uživatelsky přívětivé prostředí disponuje řadou funkcí. Pro každou úroveň pokročilosti je možné si vygenerovat seznam slovní zásoby, kterou tito mluvčí typicky ovládají. Je však možné nechat si vygenerovat i velice specifické seznamy slovní zásoby podle slovního druhu, typu jednotky (slovo, idiom, fráze, frázové sloveso), některých gramatických kategorií (např. počitatelnost), rejstříku (např. hovorová slova), témat (např. cestování), ale i použitých předpon a přípon. Můžeme tak snadno zjistit, jakou slovní zásobu k tématu cestování by měl znát student na úrovni např. A2. Při rozkliknutí vygenerovaných slov rovněž vidíme základní definice a příklady užití jednak ze slovníku, jednak ze samotného žakovského korpusu. Můžeme si rovněž slovo v databázi vyhledat, chceme-li zjistit, k jaké úrovni přináleží. Vzhledem k tomu, že slova mají často řadu více či méně příbuzných významů, poskytuje *English Vocabulary Profile* i údaje o tom, které významy jsou typicky součástí slovní zásoby studentů na dané úrovni. Výhodou je i to, že lze přepínat mezi britskou a americkou angličtinou, a v neposlední řadě i to, že ke slovům jsou přidány i nahrávky jejich výslovnosti. Projekt *English Profile* se svou gramatickou a lexikální částí je v současnosti jednou z nejpřesvědčivějších ukázek, jak je možné bezprostředně propojit žakovskou korpusovou lingvistiku s vyučováním a učením se cizího jazyka.

Seznamy slov podle úrovně pokročilosti nacházejí široké uplatnění při psaní zjednodušených čítanek a slovníků pro různé pokročilé žáky. Zjednodušené čítaneky využívají omezeného množství slov pro převyprávění celých knih a umožňují tak, aby se při četbě žák opakovaně setkával se stejnou slovní zásobou, což – jak dokládá výzkum osvojování slovní zásoby – výrazně napomáhá učení. Zjednodušené čítaneky pak mají i tu výhodu, že neodrazují žáka tím, že jsou po stránce slovní zásoby příliš složité. Uvádí se, že by na stránce zjednodušené knihy měla být maximálně 3 až 4 neznámá slova.

Seznamy slov se vytvářejí i pro potřeby výuky jazyka pro specifické účely (viz výše např. *Academic Word List*). Bývají také založeny na frekvenci výskytu, ale odstraňují se obecná frekventovaná slova, tak aby zůstala jen slova odborná. Obecně lze tedy říci, že seznamy slov vytvořené na základě různých korpusů významně přispívají k doplňování učebních sylabů

19 <http://www.englishprofile.org/wordlists>

o složku slovní zásoby na základě pokročilosti či specifických lexikálních potřeb studentů. V kontextu vyučování žáků s odlišným mateřským jazykem (dále jen OMJ) je toto jednou z cest, jak těmto žákům zprostředkovat slovní zásobu, kterou potřebují znát v jednotlivých školních předmětech.

Prostřednictvím žakovských korpusů je možné studovat užití slovní zásoby v žakovském jazyce. Podobně jako u gramatiky i zde hrají roli jazykové chyby. Ty jsou označeny speciální značkou a z takto vybaveného korpusu je následně možné vygenerovat seznam všech chyb, které se v korpusu v užití slovní zásoby objevují. Stejně jako u značkování gramatických chyb se u chyb lexikálních mohou značky vkládat přímo do textu, a to včetně emendace (tj. uvedení správné varianty). Ve větě uvedené jako Příklad 2 je ukázka značkování chyby v užití anglických sloves *borrow* a *lend*, která studenti často zaměňují zřejmě kvůli tomu, že čeština v obou případech využívá stejné sloveso a význam rozlišuje vyjádření zvratnosti (tj. půjčit komu a půjčit si). Před chybně užitým tvarem slovesa *borrowed* je značka (LS), která znamená, že jde o chybu lexikální (L, z angl. *lexical*) týkající se jednoho slova (S, z angl. *single*). Po slovesu je vložena emendace vyjadřující domněnku anotátora, že student chtěl ve skutečnosti použít sloveso *lend*.

původní věta: *she borrowed me some money*

anotovaná věta: *she (LS) borrowed \$lent\$ me some money*

Příklad 2: Ukázka chybové anotace anglické věty – slovní zásoba

Na základě chybové anotace slovní zásoby je tak možné zjistit, kde v užití slov žáci nejčastěji chybují. Na tyto chyby se lze posléze soustředit při výuce nebo tvorbě cvičení či jazykových materiálů. I zde se nezřídka ukazuje, že chyby jsou často důsledkem jazykového transferu, při němž student použije v cizím jazyce slovo, o němž se domnívá, že je užíváno stejně jako v jeho mateřštině. Tato slova označujeme jako zrádná, často též termínem *faux amis* (tj. falešní přátelé). Příkladem zrádného slova pro Čechy studující angličtinu může být slovo *klimatizace*, kterému v angličtině neodpovídá výraz *climatization*, jak se studenti někdy domnívají, ale *air-conditioning*.

Slovní zásoba se však v žakovských korpusech nezkoumá jen s ohledem na chybovost. Mezi často sledované parametry patří například frekvence, lexikální hustota, lexikální šíře či lexikální složitost. Frekvenční studie popisují, jak často studenti používají jednotlivá slova a jak se liší frekvence slov v žakovském jazyce a v jazyce rodilých mluvčích. Je tak možné například odhalit, která běžně užívaná slova se v žakovském jazyce objevují méně a naopak, která slova žáci užívají příliš. I to lze následně využít při výuce a tvorbě materiálů. Studie lexikální hustoty a lexikální šíře se pak zaměřují například na to, jaký je v textu poměr lexikálních (obsahových) slov

a slov funkčních a nakolik se slova v textu opakují. K tomu je možné využít nejrůznější nástroje volně dostupné na internetu. Lexikální složitost pak popisuje, do jaké míry žáci využívají méně frekventovaných slov. Prostřednictvím těchto testů je pak možné posuzovat pokročilost žáka s ohledem na jeho znalost slovní zásoby.

2.3.3. Frazeologie

Frazeologie zkoumá ustálená slovní spojení. Ta mohou být klasifikována na základě různých kritérií. Ve výzkumu slovní zásoby za použití korpusů se zřejmě nejčastěji zkoumají tzv. kolokace. Jde o spojení zpravidla dvou, ale i více slov, která se v korpusu často vyskytují společně a která spolu významově souvisejí a vytvářejí dohromady lexikální jednotku (např. *domácí zvíře, šedé vlasy, hustá mlha, temná noc* atp.). Některá slova mají přitom tendenci spojovat se jen s určitými jinými slovy, např. je běžné označit něco jako *krajně nevhodné*, ale spojení *krajně vhodné* rodilí mluvčí hodnotí jako neobvyklé. Rodilí mluvčí si tak zjevně na základě své jazykové zkušenosti vytvářejí mezi určitými slovy asociace, které jsou pak v jazyce opakovaným používáním upevňovány. To do značné míry charakterizuje, nakolik je text z pohledu rodilého mluvčího vnímán jako přirozený, jakkoliv ani intuice rodilého mluvčího zde nemusí být přesná a pro spolehlivé určení kolokací a jejich frekvence je nutný výzkum na základě rozsáhlých korpusů.

Ve výuce cizích jazyků hrají kolokace značnou roli. Výzkum ukázal, že žákům pomáhají při učení slovní zásoby a že díky užívání víceslovných spojení, která však produkují jako jednu naučenou lexikální jednotku, získávají čas potřebný pro plánování své promluvy. To má pozitivní dopad na plynulost jejich projevu. V neposlední řadě si při učení kolokací osvojují spojení, která jsou v jazyce rodilých mluvčích přirozená, což přispívá ke zvyšování jejich jazykové úrovně. Z těchto důvodů se informace o kolokacích stále častěji zařazují do jazykových učebnic a žákovských slovníků.

Užitečným nástrojem pro studenty a učitele, kteří chtějí pracovat s kolokacemi v anglickém jazyce, je nástroj *Just the Word*,²⁰ který umožňuje zadat jednotlivé slovo a nechat vypsát všechny jeho kolokace. Ty se generují z Britského národního korpusu (*British National Corpus*) a jsou rozděleny do kategorií dle typu spojení (např. adjektivum + substantivum, adverbium + adjektivum, sloveso + substantivum atp.). Nástroj je možné použít i jako pomůcku při psaní, jejímž prostřednictvím si v případech nejistoty můžeme ověřit, jaké spojení je časté, a tudíž přirozené, a můžeme si i vybrat z alternativ. Dozvídáme se přitom, kolikrát se zjištěná spojení v korpusu objevují, a pomocí odkazů si můžeme zobrazit i příslušné konkordanční (výsledkové)

20 www.just-the-word.com

stránky a prohlédnout si konkrétní příklady užití v korpusu. Nástroj tak lze velmi dobře využít i pro sestavování kolokačních cvičení a testů.

Mezi víceslovná spojení patří i frazémy (často též v kontextu vyučování angličtiny označované jako idiomy), jimiž rozumíme ustálená víceslovná spojení, jejichž význam není možné si odvodit z významů slov, ze kterých se frazém skládá (např. *kápnout božskou, spadnout z višně, natáhnout bačkory* atp.). Pro studenty cizích jazyků jde o náročnou součást slovní zásoby, kterou si mohou osvojit pouze přesným zapamatováním a musí navíc i vědět, kdy a v jakém kontextu lze tato spojení použít. Podobně neodvoditelná jsou v angličtině tzv. frázová slovesa, u nichž je význam slovesa v základu změněn či upraven částicí, která se s ním pojí (např. *take = vzít, take off = vzlétnout*).

Důležitost frazeologie ve výuce angličtiny zohledňuje i výše zmíněný internetový zdroj *English Vocabulary Profile*, který ve svém pokročilém vyhledávání umožňuje pracovat zvláště s frázemi, frazémy a frázovými slovesy. Je možné si vygenerovat seznam všech slovních spojení z těchto skupin pro jednotlivé úrovně pokročilosti nebo je možné zadat libovolné slovo a nechat si vypsát, v jakých víceslovných spojeních se na příslušné úrovni vyskytuje. Této funkce mohou využít nejen žáci, ale i učitelé při plánování výuky či sestavování jazykových cvičení a testů.

I u kolokací a frazémů se setkáváme s tzv. faux amis, tedy zrádnými kolokacemi a zrádnými frazémy. Žáci je nejčastěji používají proto, že se domnívají, že stejné spojení jako v jejich mateřštině funguje i v jazyce cizím. Mohou tak doslova přeložit frazém, který v cílovém jazyce neexistuje (např. nelze doslovně přeložit do angličtiny český frazém *šplouchá ti na máják* jako *it splashes on your lighthouse*). Příkladem zrádné kolokace může být třeba pokus o překlad českého *silný kuřák* do angličtiny jako *strong smoker*; správné je totiž *heavy smoker*. Je pravděpodobné, že zrádnost by fungovala i v opačném směru, tedy pro Angličana učícího se češtinu, který by spojení *těžký kuřák* v češtině při neznalosti správné kolokace zřejmě použil.

Kromě kolokací a frazémů lze text rozparcelovat i na tzv. n-gramy, tedy sekvence zpravidla dvou a více slov, mezi nimiž nemusí být nutně nějaký lingvistický vztah. Takové n-gramy, které se v textech nejčastěji opakují, obvykle zahrnují kromě gramatických slov i slova plnovýznamová a vypovídají něco o daném jazyce, např. o jeho typických ustálených jednotkách nebo tendenci k formulaickému vyjadřování.

V žakovských korpusech se stejně jako u gramatiky a slovní zásoby chyby ve frazeologii značkují. V již zmíněném systému by pak, jak vidíme z věty označené Příkladem 3, byla kolokační chyba označena značkou (LP), kde L označuje *lexical* a P pak *phrase* (fráze).

původní věta: *he is a strong smoker*
anotovaná věta: *he is a (LP) strong sheavys smoker*

Příklad 3: Ukázka chybové anotace anglické věty – frazeologie

Jeden typ výzkumu frazeologie v žákovském jazyce vychází z anotací chybovosti, při níž podobně jako u gramatiky a slovní zásoby lze s použitím anotovaných korpusů vyhledat všechny chybné výskyty. Řada anotátorů však chybné kolokace nechápe vždy jako chyby, ale spíše jako odchylky od běžné normy. Vycházejí z toho, že např. spojení *těžký kuřák* je spíše nepřirozené nežli úplně chybné. Výzkum frazeologie tak přispívá k pochopení vývoje přirozenosti žákovského jazyka a jeho přiblížení se k jazyku rodilých mluvčích. Taková srovnávání se uplatňují i v jiném typu výzkumu, na jehož základě je možné sledovat, jak časté jsou v žákovském textu a v textu rodilého mluvčího kolokace, případně n-gramy, a v čem jsou si tyto dvě jazykové variety blízké či naopak vzdálené. Je též nasnadě, že pokročilejší studenti si osvojili frazeologii cizího jazyka ve větší míře a s menší chybovostí, a výzkum tak umožňuje sledovat vývoj frazeologie u jednotlivých úrovní pokročilosti. Dalším specifickým typem výzkumu s pedagogickým dopadem je pak zkoumání frazeologie v různých žánrech. Toho se často využívá při výuce akademického jazyka a sestavování příslušných materiálů pro výuku.

2.3.4. Diskurz

Diskurzem rozumíme souvislý textový útvar, který přesahuje rámec jedné věty. U mluveného jazyka hovoříme obvykle o promluvách, u jazyka psaného pak o textech. Výzkum diskurzu je motivován přesvědčením lingvistů především v druhé polovině 20. století, že jazyk není dostatečně zkoumat jen s ohledem na nejnižší jazykové jednotky (např. hlásky, slabiky, morfémy, slova, fráze, věty), ale že je nutné vzít v potaz i takové aspekty, jakými jsou například struktura textu, význam v kontextu, role sociálního a kulturního kontextu na podobu textu, prostředky textové soudržnosti, informační struktura a obecně tedy veškeré jazykové i mimojazykové (nonverbální) prostředky, které s těmito aspekty souvisejí.

Zkoumání diskurzu v rámci korpusové lingvistiky je metodologicky zajímavé. Na jedné straně je komplikované tím, že výsledky vyhledávání v korpusu jsou prezentovány v úsečné podobě prostřednictvím konkordančních řádků, na druhé straně však korpusy díky své velikosti umožňují pracovat s obrovskými počty textů najednou a vyhledávat v nich jednotlivé diskurzní prvky v takovém rozsahu, že je možné výsledky zobecňovat a hledat tendence platné napříč texty.

Ukázkovým příkladem využití korpusu v tomto směru je významná gramatika anglického jazyka *Longman Grammar of Spoken and Written English*.²¹ Vychází z korpusu *Longman Spoken and Written Corpus*, který obsahuje 40 milionů slov z 37244 textů čtyř základních typů: konverzace, beletrie, žurnalistika a akademický text. Na jeho základě byl velmi přesně popsán například jazyk konverzace: bylo shledáno, že je bohatý na nevětne struktury, neúplné věty (vynechávají se například slova s nízkou informační hodnotou) a ustálená víceslovná spojení, která plní různé funkce (např. mluvčí indikuje, že bude pokračovat v promluvě, že se vrací k původnímu tématu atp.). Obsahuje zároveň i tzv. prostředky řečového managementu, které mluvčímu umožňují hovořit plynule. Sem patří např. výplňková slova a výplňkové pauzy, jejichž funkcí může být získání času pro plánování další promluvy nebo indikace, že mluvčí ještě současnou promluvu nedokončil. Tento výzkum inspiroval didaktiku cizích jazyků k tomu, aby se zaměřila nejen na tradiční výuku gramatiky založenou na psaném jazyce, ale i na jazykové prostředky charakteristické právě pro jazyk mluvený. Žákovská korpusová lingvistika pak v této oblasti zkoumá, do jaké míry tyto prostředky žáci využívají a jaké prostředky se naopak v jejich projevech vyskytují omezeně.

Dalším z ústředních témat je výzkum textové koherence. Jako koherenci označujeme sémantické, obsahové vztahy uvnitř diskurzu, které zajišťují jeho soudržnost, smysluplnost a plynulost. Mezi jednotlivými částmi koherentního textu totiž musí být patrná souvislost. Toho je možné dosáhnout používáním kohezních prvků, které explicitně naznačují, jak jsou jednotlivé úseky textu propojeny. Mezi tyto prvky patří například reference (odkazování na to, co již bylo řečeno, nebo na to, co bude v textu následovat), konektory (spojovací výrazy vyjadřující logický vztah mezi jednotlivými částmi), substitute (použití jiného slova, které jasně odkazuje na to, co již bylo vyřčeno) a tematicko-rematická struktura textu (tzv. aktuální členění větné). Výzkum koherence za použití korpusů je komplikovaný, neboť korpus neumožňuje zobrazit dlouhé úseky textu. Nejčastěji se proto pracuje právě s explicitními kohezními prvky (obzvláště konektory), které je možno v korpusu vyhledávat. V žákovském jazyce se pak často zjišťuje, jak žáci tyto prostředky dokážou využít a v jakých funkcích a jak jejich prostřednictvím vytvářejí strukturu textu. To má význam především při studiu a výuce akademického psaní. I v této oblasti se využívají kontrastivní studie, které srovnávají texty žákovské s texty rodilých mluvčích.

Pro vědecký text je též typické využívání rétorických prostředků pro zeslabování jistoty tvrzení (tzv. hedge) nebo jeho zesílování (tzv. booster). Ty se realizují použitím různých prostředků, např. modálních sloves (např. vy-

21 Biber, D., Stigg, J., Leech, G., Conrad, C. a E. Finegan. (1999) *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education Limited.

sledky mohou naznačovat, že...), adverbii (např. *tento výsledek nejspíš vypovídá o...*) aj. I toto je oblast, kterou si nerodilý mluvčí píšící vědecký text v cizím jazyce musí osvojit. V žákovských korpusech akademického jazyka pak můžeme zkoumat, do jaké míry se to daří a v čem spočívají odlišnosti od textů rodilých mluvčích či kompetentnějších autorů. Častým zjištěním je, že méně zkušení autoři nadužívají určitých prostředků modalizace vědeckého textu a nevyužívají dostatečně jiných možností.

V neposlední řadě lze žákovský text korpusově analyzovat i z pohledu zastoupení osobních zájmen a činného či trpného rodu a zjistit tak, do jaké míry dokáže žák prezentovat objektivní, neosobní vědecký text a jaké prostředky dokáže využít, když jako autor do textu potřebuje vstoupit například prezentováním svého postoje.

2.3.5. Pragmatika

Pragmatika studuje jazyk z pohledu mluvčího a příjemce a také to, jakým způsobem tyto uživatelé jazyka dosahují svých komunikačních záměrů. Zkoumá tedy, jaký záměr má mluvčí, když něco sděluje, a jak je jeho sdělení interpretováno příjemcem. To závisí na takových faktorech, jako je například kontext promluvy či vztah mezi mluvčím a příjemcem, roli hraje i příslušná jazyková kultura. Tyto faktory pak spolu s významem slov ovlivňují význam a dopad celého sdělení. Podle teorie řečových aktů, která stojí v centru pozornosti pragmatiky, je možné promluvou nejen něco prostě sdělit, ale i něco chtít, k něčemu se zavázat, k něčemu vyjádřit svůj postoj, případně vyřčením způsobit nějakou změnu. V žákovském jazyce je pak možné tyto řečové akty identifikovat a zjistit, jaké prostředky pro jejich vyjádření mluvčí používá. I zde je výhoda korpusů především ve zpřístupnění velkého objemu dat k výzkumu, který může být navíc prohledáván automatickými algoritmy za účelem identifikovat vzorce, které se v textu objevují. K diskurzivní a pragmatické analýze korpusů musejí být korpusy anotovány. To je však značně časově náročný a ne vždy příliš snadný úkol, především vzhledem k nutnosti správně vystihnout a interpretovat původní záměr mluvčího.²² Zkoumají se tak například funkce tzv. diskurzivních částic, což jsou prostředky, které umožňují v rozhovoru udržovat kontakt mezi mluvčími (např. angl. *you know* nebo české *víte*) a jsou v korpusech snadno vyhledatelné. Výzkum též dokládá, že užívání diskurzivních částic přispívá ke zvýšení plynulosti mluveného projevu a zvyšuje jeho přirozenost.

Z hlediska funkčního se značkují například korpusy zabývající se výzkumem specifických textových žánrů a útvarů. Existují např. korpusy

22 Jeden z příkladů pragmatické anotace korpusu je nástroj DART, více viz např. Weisser, M. (2018) *How to Do Corpus Pragmatics on Pragmatically Annotated Data Speech acts and beyond*. Studies in Corpus Linguistics: John Benjamins.

abstraktů (tj. krátkých shrnutí akademického textu). V nich jsou identifikovány a označovány příslušné rétorické kroky (např. zasazení tématu do kontextu, stanovení cíle studie atp.), následně pak mohou být tyto rétorické postupy z korpusu extrahovány a je možné zkoumat, jakými jazykovými prostředky jsou realizovány. Žákovský jazyk se zde znovu obvykle srovnává s jazykem rodilých mluvčích, ale provádějí se i srovnání mezi různými vědeckými obory ve snaze zefektivnit tento typ psané komunikace.

Studium pragmatiky ukazuje, že žák cizího jazyka musí kromě gramatiky a slovní zásoby ovládnout i pravidla komunikace, která jsou běžná pro sociokulturní prostředí, v němž se tento jazyk používá. Nejde jen o to, aby žák dosáhl svého komunikačního cíle, ale také aby svým jazykovým projevem neporušoval běžné (např. zdvořilostní) normy a dokázal pro svá sdělení volit prostředky, které odpovídají konkrétní situaci, v níž se komunikace odehrává. Jakkoliv je tato oblast důležitá pro výzkum i výuku, korpusové studie v této oblasti jsou zatím omezené.

2.3.6. Plynulost

S nástupem korpusů zachycujících mluvený jazyk, jež jsou s ohledem na nutnost nahrát a přepsat texty náročnější na sestavení než korpusy psané, se korpusově začínají zkoumat i charakteristiky řeči. Jednou z takových oblastí je výzkum plynulosti. Plynulost je obtížně definovatelný pojem, pod kterým si však nejsnáze představíme produkci řeči v dostatečném tempu a bez velkého množství prvků, které bezprostředně nesouvisejí s významem sdělení, jako jsou např. různé projevy váhání.

Tempo mluvy lze v korpusech měřit pouze tehdy, pokud jsou dostupné i původní nahrávky. Uvádí se obvykle ve slovech za minutu či slabikách za vteřinu. U pokročilých studentů angličtiny zjišťujeme průměrné tempo mluvy asi 150 slov za minutu, u rodilých mluvčích pak více než 180 slov za minutu.

Mezi jevy, které tzv. narušují plynulost projevu, se řadí kupříkladu vyplněné a nevyplněné (tiché) pauzy (např. *byl jsem eh unavený*), opakované segmenty (např. *nevím co se co se přesně stalo*), nedokončené segmenty (např. *byl jsem šel jsem tam pěšky*), opravy (např. *vzala mě nás na výlet*), fragmentovaná slova (např. *to nemu nemuselo být*) a prodlužování slabik (např. *řekl mi zeee*). Také tyto jevy se v korpusech značkují. Pak je možné zjistit nejen jejich frekvenci, ale i to, na jakém místě se v promluvě vyskytují. Jedním z rozdílů mezi žákovským jazykem a jazykem rodilého mluvčího je to, že žáci častěji umísťují tyto prvky mezi jednotlivé části frází, které rodilí mluvčí zpravidla vyřknou najednou (např. *velký eh význam má*). Vypovídá to zřejmě o tom, že rodilí mluvčí jsou schopni rychleji plánovat promluvu a produkují tak delší celky a zároveň váhají spíše na jejich začátcích než v jejich průběhu.

Všechny tyto prvky lze srovnávat na různých úrovních pokročilosti a také mezi žáky a rodilými mluvčími. Zkoumat lze i to, jaký efekt má na plynulost a její součásti typ úlohy, který žáci řešili (např. šlo-li o monolog, dialog, projev s předchozí přípravou, popis obrázku atp.). Tento výzkum pak krom studia pokročilosti a rozdílů mezi žakovským a rodilým jazykem přispívá i k poznání, jakým způsobem je produkována řeč. V didaktice má tento výzkum význam především pro oblast testování, neboť poskytuje popis ukazatelů charakteristických pro plynulost projevu, což se u ústních jazykových zkoušek nezřídka hodnotí a zároveň je i jedním z parametrů popisovaných Společným evropským referenčním rámcem.

V této kapitole jsme se pokusili nastínit, co jsou akviziční korpusy a jak se využívají při výzkumu žakovského jazyka za účelem jeho poznání a navržení možných didaktických postupů. Akviziční korpusy umožňují zkoumat žakovský jazyk z různých pohledů, srovnávat, jak se mění na jednotlivých úrovních pokročilosti a jaké jazykové prostředky a projevy tyto úrovně charakterizují. Dále umožňují srovnávání s jazykem rodilých mluvčích jako přirozeným cílem velké části žáků cizích jazyků a v neposlední řadě poskytují i empiricky ověřená zjištění využitelná při tvorbě jazykových materiálů (učebnic, slovníků, aplikací), jazykových testů a didaktických postupů. Jakkoliv je žakovská korpusová lingvistika oborem mladým, přinesl již výzkum ve všech těchto oblastech cenné výsledky a především prokazuje, že akviziční korpusy mají velký výzkumný i pedagogický potenciál, z něhož mohou čerpat nejen badatelé, ale i učitelé, studenti a tvůrci jazykových testů a sylabů.

3. Významné parametry korpusů

U korpusů obecných se předpokládá, že směřují k naplnění určitých závazných parametrů. Jde zejména o velikost a vyváženost korpusu a o autentičnost jazykových dat, která korpus zahrnuje. Uvedené tři parametry zajišťují jeho reprezentativnost ve vztahu k normálnímu, obvyklému jazykovému úzu.

Dalším důležitým parametrem je soustavné zaznamenávání informací o původu jednotlivých složek (textů), které jsou v něm zařazeny, o jejich autorech, žánru a dalších údajů relevantních pro práci s korpusem. Tyto metatextové informace jsou přiřazeny všem strukturním jednotkám korpusu formou strukturních atributů.

K uvedeným parametrům přistupuje lingvistické značkování (lingvistická anotace), tj. připojování lingvistické interpretace k jednotlivým tokenům (výskytům, textovým pozicím) formou pozičních značek. Každé slovo tedy má přidělený základní slovníkový tvar a dále značku, obvykle s morfologickými či méně často se syntaktickými informacemi.

Korpusy speciální se vyznačují tím, že některý či některé z těchto parametrů porušují, nenaplnují, případně ve snaze o jejich naplnění postupují specifickým způsobem, některé parametry rozšiřují (strukturní atributy, lingvistické značkování) apod.²³

3.1. Velikost

Obecné korpusy usilují o maximální možnou velikost. Čím větší objem jazykového materiálu zahrnují, tím větší je pravděpodobnost, že odrážejí reálný jazykový úzus a že budou využitelné i u jevů méně frekventovaných. Větší velikost rovněž umožňuje, aby byly v korpusu zahrnuty všechny relevantní variety daného jazyka v odpovídajících proporcích, tedy aby byl vyvážený a aby bylo možné považovat ho za reprezentativní vůči normálnímu, ob-

23 K tematice obecných korpusů a korpusové lingvistiky podrobněji např. Mc Enery - Hardie (2012), k akvizičním korpusům, zvl. pro češtinu, Šebesta (2010), Šebesta - Škodová (2012), Štindlová (2011), k aktuálním datům o nabídce korpusů pro češtinu a korpusů žákovských <https://wiki.korpus.cz/doku.php/cnk:uvod>, resp. Learner corpora around the world; z těchto zdrojů tato i první kapitola převážně čerpají.

vyklému užívání jazyka. Díky tomu mohou tyto korpusy dobře sloužit pro tvorbu slovníků, gramatik, thesaurů a dalších referenčních knih.

Velikost synchronních korpusů v posledních desetiletích rychle roste. Zatímco v počátcích korpusové lingvistiky, v 60. a 70. letech minulého století, se velikost korpusů uváděla v milionech tokenů (konkrétní výskyt každého slovního tvaru v textu),²⁴ u dnešních korpusů se uvádí v miliardách. Např. Český národní korpus jako zastřešující projekt zajišťující budování korpusů českého jazyka na Univerzitě Karlově zahrnuje k datu vzniku tohoto textu relativně nově český synchronní korpus SYN verze 6 o velikosti cca 4 miliardy slov, patří tedy mezi tzv. korpusy velké (pohybující se v řádu stovek milionů až v miliardách textových slov). Vedle toho rozlišujeme dnes ještě korpusy střední (obsahující řádově desítky milionů textových slov) a korpusy malé (pohybující se v řádu stovek tisíc textových slov).

I sebevětší korpus je ovšem menší než web – největší existující a volně dostupný soubor textů. I web lze využívat jako korpus; byla vyvinuta rozhraní, která k využití webu jako korpusu slouží – např. *WebCorp*.²⁵ Web jako korpus má některá specifika, která se mohou jevit jako výhody, nebo nevýhody, v závislosti na cíli, který při vyhledávání sledujeme. Většinou je nepochybnou výhodou webu jako korpusu jeho velikost. Některé výrazy či kolokace, které jsou málo frekventované, mohou v obyčejném korpusu úplně chybět nebo mít jen mizivé zastoupení, totéž platí i o některých druzích málo frekventovaných pravopisných či jiných formálních odchylek či variant, málo frekventovaných typů užití apod., které na webu – vzhledem k jeho velikosti – můžeme najít s vyšší úspěšností než v korpusu.

Hlavním problémem webu jako korpusu je skutečnost, že nám podává jazykový materiál v nerozlišené směsi textů různých typů, není tedy možné posoudit ani jeho skladbu, ani kontext užití hledaných jevů. To se může jevit jako značná potíž např. v případech, kdy hledáme velmi frekventované jazykové jevy a potřebujeme počet nalezených řádků redukovat podle nějakých lingvisticky relevantních kritérií. Dalším problémem je nestálost webu – web se neustále a průběžně mění, výzkum, který na něm byl založen, tedy není možné replikovat nebo jen velmi obtížně. Výhodou standardního obecného korpusu ve srovnání s webem je tedy jeho promyšlené složení z různých typů textů v náležitých proporcích, větší nebo menší míra jeho vyváženosti, jeho relativní ustálenost a zaznamenávání metatextových informací.

Žákovské korpusy se co do velikosti od korpusů obecných výrazně liší, přesněji řečeno, jsou v tomto ohledu podstatně pestřejší. Žákovské korpusy

24 Už zmiňovaný *Brown Corpus* obsahoval jeden milion slov, stejnou velikost měl i rovněž zmiňovaný korpus britské angličtiny *LOB corpus*.

25 <http://www.webcorp.org.uk/live/>

s objemem několika (desítek) milionů textových slov najdeme v současnosti pouze u angličtiny jako cizího jazyka. Dosud největší známý korpus je *Chungdahm Corpus*, tvořený eseji, které napsali korejšti studenti angličtiny na Chungdahm institutu, národním řetězci škol anglického jazyka, a chybově anotovaný tutor. Eseje jsou uloženy jako stále doplňovaná databáze, nikoli korpus; z této databáze byl vyčleněn pro výzkumné účely velký objem dat jako *Chungdahm Corpus* o velikosti přibližně 131 milionů slov (přesněji 130 754 tisíc slov) ve více než 860 esejích o průměrné délce 152 slov (Han et al. 2010).

Korpus této velikosti je však i u angličtiny spíše výjimkou. Co do velikosti následují *The Cambridge Learner Corpus* obsahující cca 50 milionů textových slov (také ten je založen na esejích napsaných v rámci zkoušek angličtiny) a *HKUST (The Hong Kong University of Science & Technology Learner Corpus)* o velikosti cca 25 milionů slov. Ostatní korpusy angličtiny i korpusy dalších jazyků se pohybují v podstatně nižších hodnotách – jen výjimečně dosahují hodnoty 2 milionů slov nebo ji překračují.

Mezi velikostně nejmenšími uváděnými korpusy jsou např. *The Pilot Arabic Learner Corpus*, korpus arabštiny užívané mluvčími angličtiny jako prvního jazyka, obsahující psané vyprávěcí texty mírně pokročilých a pokročilých studentů arabštiny o celkovém rozsahu 9 000 slov, nebo *The PIKUST*, korpus slovinštiny jako cizího jazyka, který obsahuje texty psané mluvčími 18 různých prvních jazyků, většinou chorvatštiny, srbštiny a ruštiny, o celkovém rozsahu 35 tisíc textových slov.

Velikost žákovských korpusů sice postupně roste, podobně jako velikost korpusů obecných, ale podstatně nižším tempem. Menší objem žákovských korpusů je dán obtížností sběru dat a finanční i časovou náročností jejich přepisu a dalšího zpracování.

První problém je spojen se skutečností, že je objem řečové produkce nerodilých mluvčích neporovnatelně menší než objem dat produkovaných mluvčími rodilými. Navíc se jedná o produkci obtížně dostupnou – pokud jde např. o písemné práce psané pro školu, disponuje jimi škola a nemusí je pro potřeby korpusu zpřístupnit. U mluvených projevů se sběry setkávají s podobnými problémy jako sběry pro mluvené korpusy jazyka rodilých mluvčích, jen s tím rozdílem, že projev nerodilého mluvčího bývá mnohonásobně kratší než projev rodilého mluvčího a celkový objem mluvených projevů nerodilých mluvčích je rovněž neporovnatelně menší.

K obtížnostem sběru přistupuje fakt, že projevy nerodilých mluvčích se sbírají v podobě nahrávek mluvených projevů nebo v podobě rukopisů, v obou případech se tedy musí texty přepisovat. Zatímco u mluvených projevů lze dnes, jde-li o kvalitní nahrávky standardní mluvy, využít automatického přepisu, u nerodilých mluvčích jde většinou o nestandardní jazyková data, která automatické zpracování v dobré kvalitě neumožňují, přepisuje se tedy manuálně, příp. se automatický přepis následně upravuje.

Je příznačné, že větších velikostí dosahují ty žákovské korpusy, které se mohou opřít o spolupráci se školami nebo které jsou přímo s jejich činností spojeny – srov. výše uvedený *Chungdahm Corpus*, zpracovávající produkci žáků řetězce škol *Chungdahm, Cambridge Learner Corpus*, rovněž obsahující produkci žáků při zkouškách, nebo *HKUST*, obsahující písemné práce vzniklé v rámci maturitních zkoušek z angličtiny. Totéž platí i o větších žákovských korpusech mluvených, jako je např. mluvená složka korpusu *BICCEL (Bilingual Corpus of Chinese English Learners)*, která vychází z nahrávek při národním testu mluvené angličtiny, korpus *NICT JLE (Japanese Learner English)*, také čerpající z testů mluvené angličtiny, nebo korpus *SWECCL (Spoken and Written English Corpus of Chinese Learners)*.²⁶

České žákovské korpusy řady *CZESL* patří svou velikostí v porovnání s jinými jazyky kromě angličtiny, která má v tomto ohledu mimořádné postavení, ke korpusům standardním, či spíše větším.

Pro badatelské i pedagogické účely se z korpusů této řady nejspíše nabízí automaticky emendovaný, anotovaný (také chybově) a lemmatizovaný korpus *CZESL-SGT*. Zahrnuje 8617 textů od 1965 různých autorů (to znamená, že od jednoho autora jsou v korpusu zařazeny v průměru zhruba čtyři texty). Objem jazykových dat činí 1148 tisíc tokenů (pozic), z toho 958 tisíc slov a 111 tisíc vět. Druhým korpusem téže řady s provedenou emendací, anotací a lemmatizací je *CZESL-MAN*; od *CZESL-SGT* se liší tím, že byl emendován manuálně a chybová anotace byla provedena poloautomaticky s vysokým podílem manuálního zpracování. Manuální emendace a anotace je spolehlivější než emendace a anotace automatická, ale také je časově výrazně náročnější. *CZESL-MAN* je proto zatím podstatně menší než korpus *CZSL-SGT* (obsahuje cca 124 tisíc slov) a není volně přístupný.

3.2. Vyváženost

Vyváženost obecných korpusů je vlastnost, o níž se tvůrci korpusů v nějaké míře snaží, míra jejího dosažení záleží ale na tom, jaká kritéria zvolí a jak se vyrovnají s obtížemi, které jsou s tím spojeny. Vzhledem k velmi variabilnímu možnostem využití korpusů není možné vyváženost korpusu opírat o vnitřní, lingvistická hlediska, ale o kritéria vnější.

Obecné korpusy v Českém národním korpusu se opíraly zpočátku o zřetel k recepci – ty typy textů, které byly častěji čteny větším počtem mluvčích daného jazyka, byly v korpusu zastoupeny ve větší proporcii než typy textů, které se četly méně. Struktura typů textů a proporce jejich zastou-

26 K mluveným korpusům angličtiny jako L2 viz např. Kolesnikova, O. – González-González, O.-A., 2016.

pení v synchronních korpusech současné psané češtiny byly tedy zpočátku stanoveny na základě výzkumů čtenářské recepcce.

V prvním korpusu řady SYN, SYN2000, tvořila většinu (dvě třetiny) textů publicistika, přibližně čtvrtinu texty odborné, zbytek texty imaginativní (beletrie, především próza). V korpusu SYN2005 se tyto proporce na základě nových výzkumů recepcce změnilly – publicistika představovala pouze 1/3 velikosti korpusu, odborná literatura 27 %, podíl beletrie vzrostl na 40 %; podobné proporce byly zvoleny pro korpus SYN2010. Korpus SYN2015 toto rozdělení opustil, tři textové makrotypy (publicistika, odborná literatura a beletrie) jsou v něm zastoupeny rovným dílem. Všechny uvedené korpusy (včetně korpusu SYN2015) se pokládají za reprezentativní v tom smyslu, že zahrnují co nejširší spektrum různých typů veřejných psaných (tištěných) komunikátů, které jako celek reprezentují současnou psanou češtinu, nezačycují ji však v proporcích odpovídajících přesně jazykové populaci.²⁷

V řadě SYN byly kromě toho zveřejněny rovněž korpusy české publicistiky (SYN2006PUB, SYN2009PUB, SYN2013PUB), které reprezentativní ve vztahu k obecnému jazykovému úzu nejsou (zahrnují pouze texty z oblasti publicistiky). Zatím poslední korpus SYN verze 6 zahrnuje všechny starší synchronní psané korpusy řady SYN včetně korpusů české publicistiky a také nový publicistický materiál, vzhledem k velké převaze publicistické složky ho tedy rovněž nelze označit za vyvážený a reprezentativní.

Vyváženost (a také velikost) a s nimi spojená reprezentativnost se řeší samostatně a dosahuje jiných hodnot u korpusů psaného a korpusů mluveného jazyka. Materiál pro psané korpusy se snáze získává (pro korpusy řady SYN získává tvůrce naprostou většinu materiálu, přibližně 90 %, přímo v elektronické podobě od nakladatelů a vydavatelů na základě smluv), zatímco získat materiál pro mluvené korpusy je obtížnější (data se sbírají formou audio či videonahrávek; větší podíl má běžná mluva) a náročné je i jeho další zpracování, především přepis. Proto se u mluvených korpusů pohybujeme v objemech podstatně nižších.

Pro češtinu je zatím největším mluveným korpusem ORAL, zahrnující dřívější korpusy řady ORAL (ORAL2006, ORAL2008, ORAL2013) a přepisy nových nahrávek (ORAL-Z). Přepisy nahrávek převážně neformálních rozhovorů, které ho tvoří, mají celkový objem 5,4 milionu textových slov, v porovnání se stamilionovými objemy psaných korpusů tedy podstatně menší. Korpus také není vyvážený z teritoriálního hlediska (podíl rozhovorů z Čech je vyšší).

Vyvážený co do relevantních sociologických kritérií (pohlaví, věk, vzdělání, oblast pobytu v dětství) a z tohoto hlediska reprezentativní je nejnovější mluvený korpus češtiny ORTOFON o objemu 1 milion textových slov.

27 <http://wiki.korpus.cz/doku.php/cnk:syn2015>

S vyvážeností volně souvisí i volba komponent, z nichž se korpus skládá; někdejší *Brown Corpus* a další korpusy jím v tomto ohledu inspirované (*Lancaster-Oslo-Bergen Corpus*, *British National Corpus*) se skládaly ze vzorků textů o standardním rozsahu (*Brown Corpus* např. obsahoval 500 vzorků textů o 2000 slovech). Soudobé korpusy, včetně obecných korpusů ČNK, se skládají z celých textů.

Lze říci, že reprezentativnost korpusu ve vztahu k běžnému, normálnímu jazykovému úzu je hodnota, k níž obecné korpusy snahou o velikost a vyváženost směřují, i když jí nikdy plně nedosahují. Při práci s korpusem ovšem sledujeme často jen určitou specifickou oblast jazyka a jeho užívání, posuzujeme tedy přirozeně i reprezentativnost příslušného korpusu z hlediska jeho výzkumného cíle a podle populace textů, na niž je náš výzkum zaměřen.

U žákovských korpusů o vyváženost ve smyslu úplného a proporcionálního zastoupení jednotlivých variet jazyka nerodilých mluvčích neusilujeme. Je to dáno tím, že u cizího jazyka, jeho osvojování a užívání se setkáváme s neporovnatelně vyšší variabilitou než u jazyka prvního a tato variabilita je vázána na mimořádně vysoký počet významných proměnných, které při výzkumu i při výuce potřebujeme brát v úvahu. Přinejmenším za současných technických podmínek a při současné velikosti žákovských korpusů není možné uvažovat o vyváženosti jakéhokoli žákovského korpusu ve vztahu k celé populaci.

Není to ani cílem jejich tvůrců. Hlavním cílem žákovských korpusů v tomto směru je umožnit studium variability žákovského jazyka (mezi-jazyka) a jeho užívání a studium vztahu této variability k co největšímu počtu sledovaných a pečlivě zaznamenávaných proměnných. Při práci s žákovským korpusem tedy neočekáváme, že reprezentuje mezijazyk nerodilých mluvčích jako celek, to není dost dobře možné, ale že spolehlivě zaznamenává a umožňuje studovat užívání některé jeho variety ve vztahu k co největšímu počtu faktorů. Při posuzování kvality žákovského korpusu tak přihlížíme nejen k jeho velikosti, ale k tomu, zda obsahuje tu varietu mezijazyka žáků, která nás zajímá, a také k tomu, kolik faktorů (a které) ovlivňujících její užívání zaznamenává a jak podrobně. Nejčastěji to bývá první jazyk žáků (např. čeština žáků s prvním jazykem čínským) a úroveň ovládnutí cílového jazyka (zřídka jde o začátečníky, častěji o mírně pokročilé nebo pokročilé), téma (blízké, známé, citově zabarvené vs. vzdálené, neznámé, citově neutrální), žánr, způsob sběru apod.

Korpusy řady CZESL²⁸ byly budovány se snahou získat jazyková data v takové skladbě, aby v nich byly zastoupeny všechny úrovně ovládnutí jazyka, od začátečníků po pokročilé, a aby v repertoáru prvních jazyků studentů byly dostatečně zastoupeny tři jejich skupiny: jazyky slovan-

28 CZESL-PLAIN, CZESL-SGT, CZESL-MAN (sekundárně i jejich deriváty).

ské, indoevropské neslovanské a jazyky neindoevropské, češtině vzdálené. Jednotlivé úrovně ovládnání jazyka (podle SERR) a skupiny prvních jazyků nejsou v těchto korpusech zastoupeny rovnoměrně, korpusy tedy nejsou z těchto hledisek vyvážené; pro studium variability užívání jazyka rodilými mluvčími ve vazbě na tyto i další parametry je však lze velmi dobře využívat.

3.3. Autentičnost jazyka

Třetím významným parametrem obecných korpusů je autentičnost jazykových dat, která jsou v nich uložena. Jde – vedle vyváženého zastoupení všech relevantních variet daného jazyka a vedle dostatečné velikosti korpusu – o třetí důležitou podmínku pro to, abychom mohli s korpusem pracovat jako s reprezentativním vzorkem normálního, obecného jazykového úzu.

Za autentické se pokládají takové projevy, které vznikly z autentické komunikační potřeby rodilého mluvčího v obvyklé, standardní situaci, nikoli projevy, které byly uměle elicitovány např. pro potřeby lingvistického výzkumu. U obecných korpusů je autentičnost dat zajištěna tím, že jsou do nich zařazovány projevy vzniklé v reálných komunikačních situacích.

Trochu složitější je situace u korpusů mluveného jazyka. Diskutovaná je především hranice toho, co máme považovat za jazyk mluvený. Prototypicky mluvené jsou nepochybně neformální nepřipravené, spontánní projevy rodilých mluvčích, v praxi se však někdy do mluvených korpusů zařazují i projevy připravené a často stylizované, šířené v médiích.²⁹

České mluvené korpusy řady ORAL i nový korpus ORTOFON se drží přísnějšího vymezení mluveného jazyka a mluvených korpusů – obsahují neformální spontánní projevy rodilých mluvčích. Od mluvených korpusů je účelné odlišovat, i když se to někdy nečiní, korpusy řečové, které jsou zaměřeny na zachycení zvukové stránky jazyka. Současné mluvené korpusy fungují částečně i jako korpusy řečové – nabízejí vedle ortografického přepisu mluveného projevu rovněž jeho přepis fonetický, původní zvukový záznam, někdy k tomu přistupuje i videozáznam, případně přepis gest.

Ke studiu zvukové stránky jazyka ovšem nestačí pouze záznam autentického (neformálního) mluveného projevu. Důležitou složkou korpusů řečových (speech corpora) jsou např. záznamy čtení speciálně pro tyto potřeby připravených textů, čtení izolovaných slov, jejich párů a skupin sestavených s ohledem na některé významné fonetické jevy. Zpravidla jde o záznamy nahrávané ve studiu, v experimentálních podmínkách. Řečové korpusy tohoto typu jsou tedy typické korpusy speciální, nikoli obecné.

29 Pro jejich zařazování se vyslovuje např. J. Sinclair (Sinclair, 1996).

Korpusy žákovské pracují s pojmem autentičnosti poněkud odlišně od toho, jak se chápe v korpusech obecných. Především nevnímají autentičnost jako hodnotu diskrétní, nýbrž skalární. Prosté rozlišení textů vzniklých ze skutečné komunikační potřeby ve standardní situaci a textů ostatních je pro potřeby výzkumu osvojování, vyučování i užívání jazyka žáky příliš hrubé, protože nebere ohled na různé situační faktory, za nichž text vzniká a které jeho podobu i stupeň autenticity ovlivňují a které mohou být z badatelského i didaktického hlediska zajímavé.

Pracujeme-li se škálovým pojetím autenticity, máme na jednom pólu škály projevy autentické, tj. žákovy přirozené, učitelem neřízené a zřetelkém k výuce neovlivňované vyjadřování, vzniklé z komunikační potřeby v autentické komunikační situaci (např. při nákupu, objednávání hotelu, při rozhovoru s kamarády v restauraci apod.). Na druhém pólu jsou projevy žáka zřejmě a výrazně řízené jinou osobou, většinou učitelem, a to s pozorností zaměřenou primárně na jazykovou formu sdělení (projevy experimentálně elicitované). Může jít např. o překlad do druhého jazyka, převádění vět do minulého času nebo množného čísla podle zadání apod.

Mezi oběma póly je velmi rozsáhlá oblast projevů, jejichž vznik byl rovněž podnícen a jeho zpracování ovlivněno jinou osobou než mluvčím (zpravidla učitelem, ale také rodičem, vychovatelem, badatelem, sběračem dat apod.), ale které vznikaly alespoň zčásti se zřetelkém k funkci či obsahu sdělení (např. s cílem vyprávět kamarádovi obsah filmu, popsat obrázek tak, aby ho kamarád rozpoznal apod.), nikoli k jazykové formě. Tento typ projevů bývá označován jako projevy klinicky elicitované (Ellis – Barkhuizen 2005: 23). Za klinicky elicitované můžeme pokládat většinu prací vzniklých ve škole pro školní potřeby, především slohové práce (eseje) různého druhu, které bývají předmětem sběru pro korpusy.

Míra řízenosti u prací vzniklých ve školním kontextu je velmi různá. Typicky experimentální povahu mají různé gramatické transformace, ale také výsledky mluvních cvičení zaměřených na zvukovou stránku řeči. Experimentální povahu mají také korpusy (či části korpusů) obsahující přepisy nahrávek čtených textů, čtených seznamů izolovaných slov apod., které bývají do akvizičních, a zvláště žákovských korpusů mluveného jazyka zařazovány. Korpusy tohoto typu jsou přirozené velmi užitečné pro výzkum i pro pedagogické využití. Tak např. řečový korpus *ISLE (Interactive Spoken Language Education)*³⁰ obsahuje materiál z čtení jednoduchých vět a minimálních párů; řečový korpus angličtiny čínských žáků *ESCCL (English Speech Corpus of Chinese Learners)* zahrnuje čtené dialogy; korpus *LeaP (Learning Prosody in a Foreign Language)*³¹ zahrnuje jak materiál z čtení (izolovaných

30 <http://catalog.elra.info/en-us/repository/browse/ELRA-Soo83/>

31 http://www.homes.uni-bielefeld.de/gibbon/Docs/LeapCorpus_Manual.pdf

slov, krátkého příběhu), tak z převyprávění i volného rozhovoru; čtení (nepřipravené čtení většinou úryvků z beletrie) je základem korpusu *Learners' Corpus of Reading Texts*.³²

U klinicky elicitovaných prací hraje roli řada faktorů: vztah mezi iniciátorem a žákem (učitel, kamarád, neznámá osoba; věkový rozdíl; vzájemná známost apod.), povaha úkolu (součást jazykového testu, úkol zadaný se zřetelem k obsahu a funkci), přípravné aktivity, možnost využívat jazykové pomůcky, téma (blízké, známé, důvěrné vs. vzdálené, neznámé; žánr atd.).

Autentické projevy můžeme zaznamenávat např. v mimoškolním prostředí, v rodině. Pro korpusy žákovské je získání autentické jazykové produkce značně obtížné, zejména pokud sledujeme i nižší úrovně ovládnání jazyka. Autentická produkce žáků na nejnižších úrovních ovládnání jazyka je velmi omezená nejen kvantitativně, ale také funkčně, tematicky a žánrově. Většina známých žákovských korpusů je proto založena na sběrech dat klinicky nebo experimentálně elicitovaných.

Akviziční korpusy zaměřené na první jazyk, tedy korpusy jazyka dětí a mládeže předškolního a školního věku, se od korpusů žákovských v tomto ohledu poněkud odlišují. Pro studium jazyka mládeže je nezbytné získat její autentické neformální a spontánní projevy, vernakulární jazyk. Jeho sběr není snadný, proto se vedle nahrávek neformální komunikace v autentickém prostředí volí většinou různé jiné cesty, jak získat projevy blízké autentickým – např. stimulováním dialogu mezi vrstevníky o tématu, které je jim blízké a emotivní, aranžováním situace, která má typické rysy situace neformální, apod.

I u školní mládeže je ovšem podstatné studovat, jak si žáci osvojují tu varietu rodného jazyka, která je vyžadována a zprostředkovávána školou, jak zvládají písemný projev, jak si osvojují tzv. intelektuální lexikum apod., ani korpusy jazyka mládeže se tedy neomezují jen na projevy vernakulární.

Žákovská data v korpusech řady CZESL i data v akvizičních korpusech češtiny jako prvního jazyka (*SKRIPT2012*) jsou rovněž založena na klinicky elicitované produkci. Jde pravidelně o písemné práce psané ve školním kontextu, ale za různých podmínek, s různou mírou přípravy apod. Autentickým se blíží data v obou akvizičních korpusech češtiny mluvené, tedy *SCHOLA2010*³³ a zčásti *ROMi_1.0*³⁴.

Obecně lze říci, že převážná většina akvizičních a žákovských korpusů je téměř zcela založena na jazykovém materiálu získaném z projevů klinicky

32 Dostupný z http://sldr.ortolang.fr/voir_depot.php?lang=en&id=15.

33 Jde o přepisy autentických vyučovacích hodin, které zahrnují záznam autentické mluvené produkce žáků ve škole.

34 Nahrávky promluv dětí ze sociokulturně znevýhodňujícího prostředí; ani ty ovšem nejsou autentické plně.

elicitovaných, v omezenější míře elicítovaných experimentálně a jen velmi omezeně (především u korpusů zachycujících vernakulární jazyk mládeže) z projevů autentických nebo autentickým projevům se blížících. Aby bylo možné posoudit míru a povahu řízenosti těchto projevů, je nezbytné při budování akvizičních korpusů zaznamenávat všechny relevantní charakteristiky textu, mluvčích, situace produkce a způsobu sběru, které by mohly charakter jazyka v korpusu ovlivnit, a uvádět je v korpusu jako strukturní atributy mnohem podrobněji, než je obvyklé u korpusů obecných.

3.4. Strukturní atributy

Korpusy jsou vnitřně strukturovány – člení se na menší jednotky; u psaného korpusu SYN2015 a psaných korpusů pozdějších to jsou dokumenty, dále texty (pokud je dokumentem např. číslo novin či sbírka fejetonů), ty se dělí na odstavce a odstavce na jednotlivé věty, resp. větné celky (od počátečního do koncového signálu). (Při tokenizaci se věty dále člení na jednotlivé tokeny – výskyty, pozice.) K strukturním jednotkám jsou v korpusu přiřazeny strukturní atributy, které nesou důležité metatextové informace (metadata); podle nich můžeme v korpusu cíleně vyhledávat a vyhodnocovat výsledky.

Příkladem strukturních atributů psaných korpusů pro češtinu nám může být korpus SYN2015. U tohoto korpusu se pracuje kromě jednoznačného identifikátoru dokumentu celkem s 15 strukturními atributy. Většinou slouží:

- a. identifikaci původu dokumentu: název dokumentu nebo periodika, autor, vydání (u periodik), vydavatel, místo vydání, rok vydání, překladatel, zdrojový jazyk;
- b. podání informace o druhu dokumentu: skupina textových typů (beletrie, oborová literatura, publicistika), textový typ (u beletrie např. próza, kratší próza, drama, poezie; u oborové literatury odborná, populárně naučná, profesní, memoáry a autobiografie, administrativa; u publicistiky publicistika tradiční a volnočasová), skupina oborů/témat, žánr/oblast, médium dokumentu (kniha, časopis, noviny, jiná tiskovina, referenční příručka, učební materiál), periodicitu;
- c. podání informace o cílovém adresátovi – jde o odlišení dokumentů určených dětskému čtenáři od ostatních.

Strukturní členění korpusů mluvených je poněkud odlišné a odlišné jsou i strukturní atributy.

Pro akviziční korpusy, prvního i druhého/cizího jazyka, jsou metatextové informace mimořádně důležité. Dovolují totiž sledovat, jak jednotlivé vnější faktory (spojené se vznikem textu, s jeho autorem, se způsobem

sběru atd.) působí na volbu variety jazyka a na jeho užívání, jak ovlivňují řečové chování mluvčích.

Strukturní atributy u akvizičních korpusů jsou proto výrazně bohatší a detailnější. Bohatství této dokumentace je podstatným měřítkem kvality žakovského, resp. širě akvizičního korpusu a zvyšuje či limituje jeho využitelnost.

Zajímavostí akvizičních korpusů je to, že obvykle obsahují metadata spojená s žákem (autorem textu). Např. v korpusu CZESL-SGT najdeme údaje trojího typu: (a) o osobě mluvčího/pisatele, např. o jeho věku, pohlaví, (b) o jeho prvním jazyce a jazykovém zázemí vůbec (první jazyk, skupina jazyků, další jazyky, úroveň ovládnutí češtiny podle SERR), (c) o podmínkách jeho studia češtiny jako cizího jazyka (možnost seznámení s jazykem v rodině, pobyt v ČR, délka, intenzita a povaha kurzu češtiny až po údaj o užívané učebnici).

Strukturní atributy korpusů CZESL-SGT i dalších žakovských a akvizičních korpusů umožňují sledovat vliv velmi širokého spektra faktorů na osvojování a užívání jazyka; dosavadní výzkumy tyto možnosti ještě ani zdaleka nevyužívají.³⁵

3.5. Poziční atributy - lingvistická (a chybová) anotace

Lingvistické či jazykové značkování spočívá v připojování lingvistické interpretace formou tzv. pozičních atributů k existujícímu korpusu psaného nebo mluveného jazyka, přesněji k jednotlivým korpusovým pozicím, tokenům. Korpusy se v rozsahu a povaze této anotace odlišují a přirozeně se tato anotace také vyvíjí.

Vlastnímu značkování předchází rozčlenění textu, a sice: (a) explicitní segmentace textu do vět (značených počáteční značkou <s> a koncovou </s>); jde o větné celky vymezené počátečním a koncovým signálem, tedy věty jednoduché i souvětí; (b) tokenizace, rozdělení textu na tokeny (korpusové pozice).

Značkování probíhá tak, že je každému tokenu přiřazeno lemma (reprezentativní slovníková podoba) a morfologická značka, tag. Slouží k tomu dva kroky: (a) morfologická analýza (přiřazení všech teoreticky možných morfologických a slovnědruhových interpretací každé korpusové pozici bez ohledu na kontext), (b) disambiguace (výběr správné morfologické a slov-

35 Srov. vyjádření S. Grangerové (Granger 2009: 17): *One must admit, however, that this facility is still seldom used and LC researchers (myself included) have had a tendency to base their analysis on the whole corpus or on subcorpora distinguished only on the basis of the learners' mother tongue. In fact, a properly coded learner corpus makes it possible for researchers to study the effect of a much wider range of variables.*

nědruhové interpretace na základě kontextu a vyloučení ostatních, pokud ovšem situace nevyžaduje zachování většího počtu možností).

Například ve větě *Hráli házenou* dospějeme u tokenu *házenou* morfologickou analýzou ke čtyřem alternativním možnostem (s přiřazenými lemmaty, slovnědruhovou a morfologickou charakteristikou; zde uvádíme zkrácené pouze základní určení):

- substantivum v akuzativu singuláru,
- substantivum v instrumentálu singuláru,
- adjektivum v akuzativu singuláru,
- adjektivum v instrumentálu singuláru.

Disambiguace pak tři z těchto možností vyloučí, ponechá pouze správnou variantu (substantivum v akuzativu singuláru). Morfologická analýza i disambiguace se provádí s využitím speciálních programů; jde ale o proces velmi náročný, zejména u syntetických jazyků s bohatou flexí a homonymií, včetně homonymie gramatické, jako je čeština.

Lemmatizace a morfologické značkování představují základní typy lingvistické vybavy korpusů. K nim přistupuje dnes jako třetí anotace syntaktická, která byla pro češtinu vyvinuta v rámci *Pražského závislostního korpusu* s využitím domácí, české syntaktické tradice. Díky tomu se mohou na bázi syntakticky anotovaných korpusů vyvíjet i programy pro potřeby školní výuky. V současné době využívají syntaktickou anotaci rovněž synchronní korpusy budované v Českém národním korpusu; prvním syntakticky anotovaným je korpus SYN2015.

Všechny uvedené typy lingvistické anotace (i další, zde nezmíněné, jako je anotace fonetická, fonologická či sémantická) mohou být velmi užitečné i v akvizičních korpusech. Zejména u žákovských korpusů je ovšem lingvistické značkování velmi svízelné. Žákovský jazyk totiž zahrnuje ve vysoké míře nestandardní jazyková data, nelze tedy na ně bez předběžné úpravy aplikovat nástroje pro automatické zpracování. Před lingvistickou anotací je tudíž potřeba korpusy upravit tak, aby bylo automatické zpracování umožněno, tj. opravit odchylky od obvyklého jazykového úzu (emendovat je). Emendace se provádí manuálně (taková úprava je spolehlivá, ale časově i finančně velmi náročná), automaticky nebo semiautomaticky.

Na emendaci pak někdy navazuje kromě standardní lingvistické anotace rovněž anotace, která byla vyvinuta specificky pro účely akvizičních, zvl. žákovských korpusů, totiž anotace chybová, tedy přiřazení informací o typu chyby.³⁶ Termínem chyba se zde označuje odchylka od obvyklého úzu rodilého mluvčího nebo od jeho jazykové intuice. V literatuře i v praxi se setkáváme s různými typologiemi jazykových chyb, založenými na různých kritériích o různé jemnosti. Jen relativně malý počet žákovských korpusů je

36 Podrobněji k chybové anotaci žákovských korpusů např. Štindlová, 2011.

dnes chybovým značkováním vybaven; volí se přitom různě detailní strategie – někdy jsou chyby pouze vyznačeny, emendovány, jindy též pojmenovány (zařazeny do určité kategorie), případně hodnoceny atd.

Z korpusů řady CZESL jsou emendovány, lemmatizovány a otagovány, včetně tagování chybového, CZESL-SGT a CZESL-MAN. CZESL-SGT byl emendován automaticky s pomocí specifických programů pro kontrolu pravopisu a gramatiky; vznikly tak dvě vrstvy, původní a emendovaná, slovům v obou těchto vrstvách byla přiřazena, opět automaticky, lemmata a tagy (slovnědruhé a morfologické značky), dále pak údaje o chybě. CZESL-MAN prošel emendací manuální a návazně semiautomatickou anotací, rovněž včetně anotace chybové.

Při hledání v korpusu CZESL-SGT můžeme vyjít z původního tvaru (word, k němu je vázáno lemma, tag; k typům vyhledávání viz dále) nebo z tvaru emendovaného (word₁, lemma₁, tag₁), můžeme hledat typy chyb (err) apod. Zadáme-li např. word₁³⁷ slova *miluje*, zobrazí se nám na pozici word jak užití tvary standardní, tak i tvary chybové (*míluje*, *milujút*).

Je potřeba upozornit, že u automatické emendace a anotace musíme počítat s nepřesnostmi a omyly, a to jak ve vlastní emendaci, tak v anotaci. Pokud jsme si toho ale vědomi, může nám korpus CZESL-SGT posloužit jako mimořádně bohatý zdroj dat pro výzkum i výuku.

37 Viz dále.

4. Práce s korpusem

V následující kapitole se seznámíme s několika korpusem češtiny a se základními způsoby vyhledávání v nich.

4.1. Český národní korpus (ČNK)

4.1.1. Materiál/Obsah

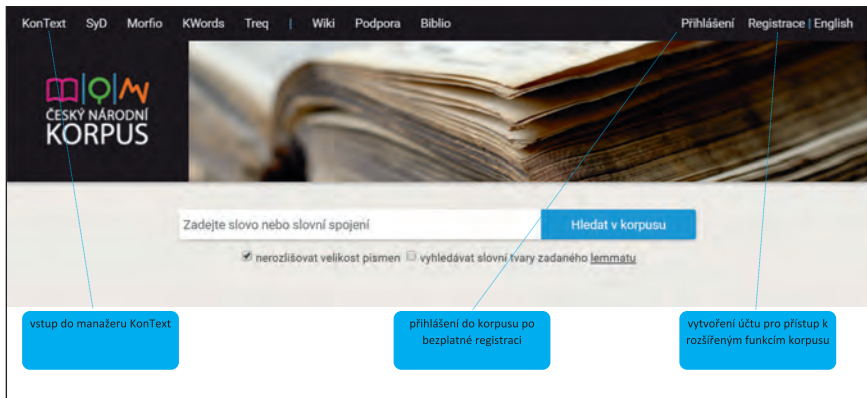
Český národní korpus je projekt, který má za úkol vytváření dílčích korpusem češtiny. Obsahuje především korpusem psaného jazyka, ale také několik korpusem mluvené češtiny. V souhrnu se jedná o nejobsáhlejší jazykový materiál, který je pro češtinu zpracován. Psané korpusem lze dělit na korpusem synchronní, které dokládají současný jazykový úzus, a korpusem diachronní, které zachycují jazyk v jeho vývoji. Nezanedbatelný je také mnohojazyčný paralelní korpus *InterCorp*, který nabízí zarovnané texty (překlady textů) z téměř 40 jazyků. Kompletní přehled všech dostupných korpusem vč. jejich charakteristiky najdete na webu Českého národního korpusem.³⁸

Pro výuku žáků s OMJ jsou relevantní především reprezentativní korpusem současné češtiny, např. SYN2015, nebo akviziční korpusem, např. žákovský korpus CZESL. V závislosti na cíli výuky a po zvážení jazykového původu žáků lze využít také texty paralelního korpusem *InterCorp*.

4.1.2. Vyhledávač neboli korpusový manažer

ČNK využívá korpusový manažer KonText, dostupný z webové stránky ČNK www.korpus.cz. KonText je umístěn v horním menu vlevo. Lze jej využívat i bez registrace, po bezplatném zaregistrování je ale k dispozici více funkcí.

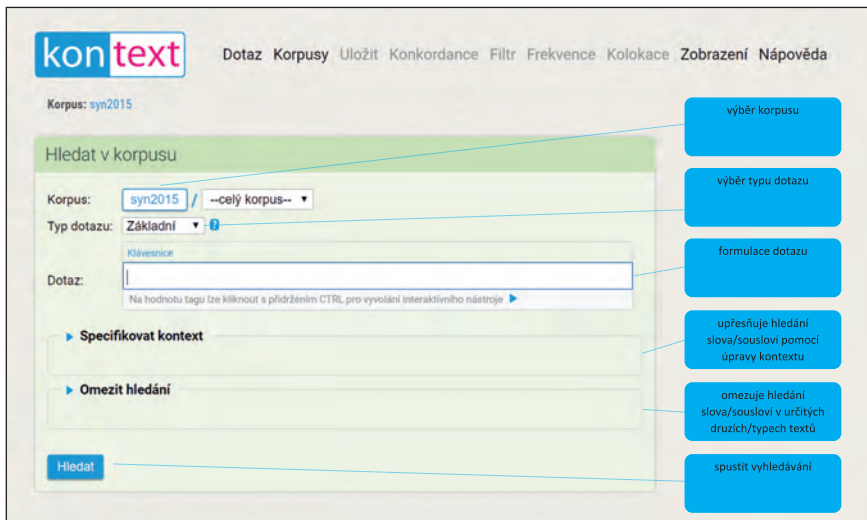
38 <https://wiki.korpus.cz/doku.php/cnk:uvod>



Obrázek 1: Úvodní stránka manažeru KonText

4.1.3. Vyhledávání

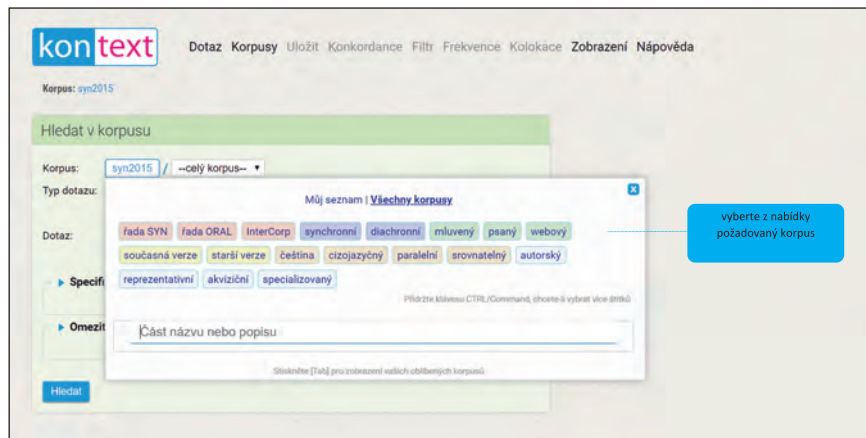
Před vyhledáváním je třeba si ujasnit základní parametry, které ovlivňují jeho výsledek. Jedná se o volbu korpusu, výběr typu dotazu, zadání dotazu, příp. specifikaci kontextu a omezení hledání. První tři jmenované kategorie jsou klíčové a nelze je vynechat. Na volbě těchto tří kategorií také přímo závisí výsledek korpusové analýzy. Poslední dvě kategorie jsou fakultativní a umožňují přesnější zadání dotazu.



Obrázek 2: Zadání dotazu v KonTextu

4.1.4. Volba korpusu

V první řadě je třeba zvolit korpus, v němž bude hledání probíhat. Výběrem pole s přednastaveným korpusem se otevře tabulka se štítky, podle nichž lze v korpusech hledat. Pro lepší přehlednost je možné se orientovat podle seznamu korpusů na encyklopedii wiki.³⁹ Každý ze štítků pod sebou skrývá nabídku všech relevantních korpusů, z nichž si lze následně vybrat ten nejvhodnější.



Obrázek 3: Výběr zdrojového korpusu

4.1.5. Typ dotazu

Před zadáním vlastního dotazu je třeba vybrat jeho typ. Přednastavený je **základní** typ dotazu. Jedná se o intuitivní způsob zadání dotazu. Pokud je do pole dotazu zadáno slovo v základním tvaru (u podstatných jmen první pád jednotného čísla, u sloves infinitiv, u přídavných jmen nestupňovaný tvar mužského rodu apod.), zobrazí se výsledky obsahující všechny tvary hledaného slova (od slova *mladý* doklady obsahující všechny slovní tvary – *mladými*, *mladým*, *mladého* apod.). Pokud v **základním** typu dotazu zadáme konkrétní slovní tvar (např. *mladými*), budou výsledky analýzy zahrnovat pouze texty obsahující tento slovní tvar. V **základním** dotazu lze zadávat i slovní spojení, přičemž platí stejná pravidla jako u jednoslovných zadání. U souloví v konkrétním tvaru (např. *s velkou pravděpodobností*) budou výsledky obsahovat texty s tímto konkrétním tvarem. Pro vyhledání všech tvarů souloví je třeba zadat jeho komponenty v základním tvaru, tj. např. pro spojení *s velkou pravděpodobností* zadáme dotaz ve tvaru *s velkým pravděpodobnost*.

39 <https://wiki.korpus.cz/doku.php/cnk:uvod>

Výsledky pak obsahují sousloví s *velkou pravděpodobností, s větší pravděpodobností, s největší pravděpodobností*. Využití základního typu dotazu je tedy vhodné pro jednodušší dotazy a doporučuje se pro korpusové začátečníky. Jeho nevýhodou je, že nemůže obsahovat tzv. regulární znaky, tj. znaky mající zástupnou funkci (více o regulárních znacích obsahuje wiki⁴⁰).

Druhým typem dotazu je *lemma*. Do pole dotazu se zadává základní tvar slova a ve výsledcích se zobrazují všechny tvary zvoleného slova. Výhodou dotazu typu *lemma* je, že umožňuje doplnit dotaz o specifikaci slovního druhu, a tím u některých slov zamezit vysoké chybovosti z důvodu homonymie. Příkladem může být slovo *stát*, které v češtině existuje jako podstatné jméno i sloveso.

Třetím typem dotazu je *fráze*. Při volbě *fráze* se do pole dotazu zadává slovní spojení v konkrétním tvaru, které je pak také v témže tvaru zobrazeno ve výsledcích analýzy.

Čtvrtým typem dotazu je *slovní tvar*. Do pole dotazu se zadává konkrétní slovní tvar, který je pak zobrazen ve výsledcích. Pokud je při tomto typu zadán neutrální slovní tvar (např. infinitiv), je na rozdíl od *základního* typu dotazu ve výsledcích zobrazen jen infinitiv, nikoli další slovní tvary. Slovní tvar je možné upřesnit volbou slovního druhu (např. *mezi* – předložka vs. podstatné jméno v 6. pádě jednotného čísla).

U typů dotazů *lemma* a *slovní tvar* je na rozdíl od *základního* typu dotazu možné volit slovní druhy, a tím eliminovat tvarovou homonymii.

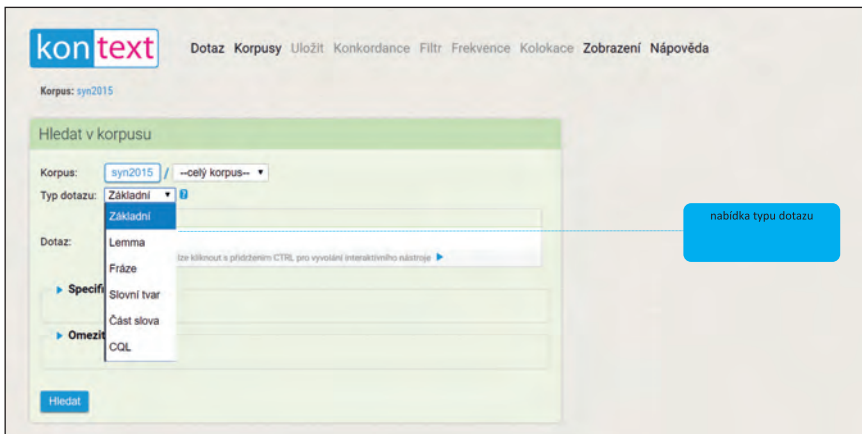
Předposledním nabízeným typem dotazu je *část slova*. Umožňuje vyhledat slovní tvary, které obsahují zadaný řetězec znaků, a zahrnout do dotazu též regulární (tj. zástupné) znaky. Řetězcem znaků je myšlen pouhý sled znaků/písmen bez nároku na morfologické dělení (kořen slova, předpony, přípony atd.). Hledáme-li tedy pomocí tohoto typu dotazu slova s předponou *před-*, bude ve výsledcích analýzy zobrazen též slovesný tvar *přede* od slova *příst*.

Posledním typem dotazu je *CQL* (= *corpus query language*), což je dotazovací jazyk umožňující komplikovanější vyhledávání. Více k práci s *CQL* najdete na wiki.⁴¹

Pro jednoduchou a rychlou orientaci v rozdílech mezi typy dotazů slouží nápověda ve formě modrého pole s otazníkem, která uživateli poskytne rychlou charakteristiku zvoleného typu dotazu.

40 https://wiki.korpus.cz/doku.php/kurz:regularni_vyrazy

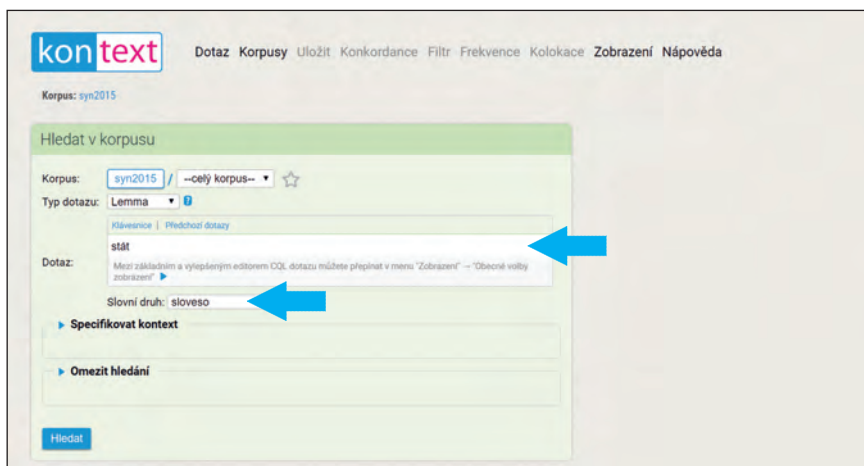
41 https://wiki.korpus.cz/doku.php/kurz:pokrocile_dotazy



Obrázek 4: Volba typu dotazu

4.1.6. Zadání dotazu

Po výběru typu dotazu je třeba dotaz formulovat. V závislosti na vybraném typu dotazu lze zadat základní nebo konkrétní slovní tvar, příp. použít v dotazu regulární znaky. Po výběru některých typů dotazů je zpřístupněno doplňkové okno umožňující definovat slovní druh. U níže uvedeného příkladu je hledáno sloveso *stát* ve všech jeho tvarech (je tedy vyloučena homonymie s podstatným jménem *stát*).



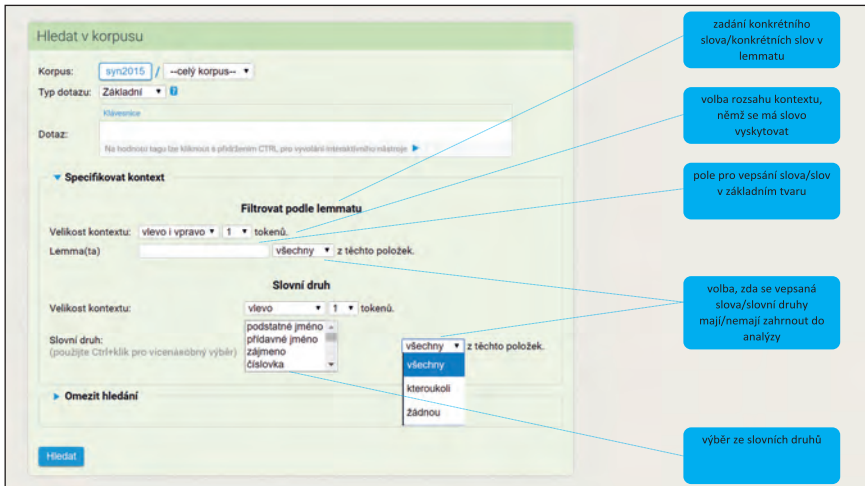
Obrázek 5: Zadání dotazu a specifikace slovního druhu

4.1.7. Specifikace kontextu

Oddíl *specifikace kontextu* se využívá tehdy, pokud uživatel potřebuje ovlivnit kontext vyhledávaného výrazu nad rámec možností, které mu poskytují zadání dotazu. Využívá se tehdy, pokud hledané slovo chceme v jeho nejbližším okolí rozšířit o další slovo/slova, příp. pokud chceme další slovo/slova v okolí z vyhledávání vyloučit. Zadávat lze jak konkrétní slova/slovní tvary, tak kategorie slovních druhů. Specifikace kontextu také slouží tam, kde je třeba vyhledat víceslovné spojení, jehož komponenty nestojí pevně za sebou (k tomu slouží typ dotazu *fráze*).

Po volbě *specifikace kontextu* se zadávací tabulka rozšíří o několik částí. V oddílu *Filtrovat podle lemmatu* je třeba nejprve určit, jaký kontext má být do vyhledávání zahrnut. Kontext se určuje v tokenech (jeden token = zjednodušeně jeden slovní tvar na daném místě), jejichž počet si uživatel volí v tabulce. Před údajem k počtu tokenů uživatel vybírá, zda se počet zahrnutých tokenů týká jen části textu před hledaným slovem, tj. *vlevo*, nebo po hledaném slově, tj. *vpravo*, nebo na obě strany od hledaného slova, tj. *vlevo i vpravo*. Je možné zadávat slova v jejich základních tvarech, tj. *lemmatech*. Pole *Lemma(ta)* umožňuje vepsat jedno nebo více slov. V tabulce na konci řádku je třeba vybrat, zda mají být zadaná slova obsažena *všechna* v zadaném kontextu, nebo zda může být vybráno *kterékoli* ze zadaných slov, příp. zda nemá být vybráno *žádné*. Při zadání jednoho slova znamená výběr *všechny* a *kteroukoli* totéž. U zadání *všechna* se ve výsledcích hledání zobrazí všechny úryvky textů, které obsahují hledaný výraz (v poli *Dotaz*) a zároveň také všechna zadaná slova v rámci nastaveného kontextu. Jako příklad uvedme podstatné jméno *telefon* zadané v poli *Dotaz*, k němuž jsou ve specifikaci kontextu uvedeny atributy *chytrý* a *mobilní*. Výsledky pak budou zobrazovat jen ty úryvky textů, které obsahují všechna tři slova, tj. *chytrý*, *mobilní*, *telefon*. U zadání *kteroukoli* se ve výsledcích hledání zobrazí všechny úryvky textů, které obsahují hledaný výraz (v poli *Dotaz*) a zároveň vždy alespoň jedno ze zadaných slov, ve výše uvedeném příkladu to tedy budou úryvky obsahující slovní spojení *chytrý telefon* a *mobilní telefon*. U zadání *žádnou* se ve výsledcích hledání zobrazí všechny úryvky textů, které obsahují hledaný výraz (v poli *Dotaz*) a zároveň neobsahují ani jedno ze zadaných slov, obsahují tedy *telefon*, ale nikoli *chytrý* ani *mobilní*. Volba *žádnou* má tedy funkci negativního filtru, jímž definujeme, s jakými slovy se slovo zadané v dotazu nemá vyskytovat.

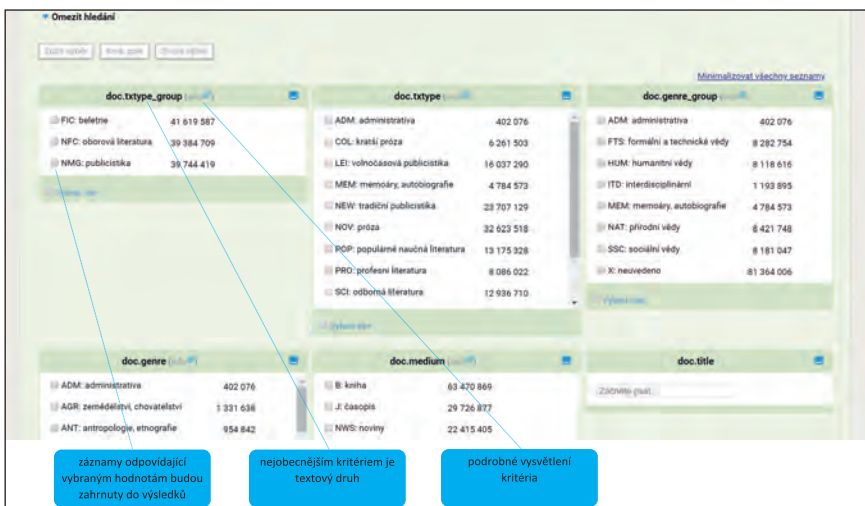
V oddílu *Slovní druh* se postupuje při zadávání stejně jako v oddílu *Filtrovat podle lemmatu*, jen s tím rozdílem, že se nevypisují konkrétní slova v základním tvaru, ale volí se jen slovní druh/druhy jako kategorie.



Obrázek 6: Specifikace kontextu

4.1.8. Omezení hledání

Pokud se uživatel rozhodne, že nechce vyhledávat v celém zvoleném korpusu, může své vyhledávání omezit. Texty jsou v korpusu tříděny podle různých kritérií (podle textového druhu, žánru, typu média, pohlaví autora apod.) a na základě těchto kritérií lze vybrat skupiny textů, které budou z analýzy vyloučeny.



Obrázek 7: Omezení hledání

Po vyplnění úvodní zadávací tabulky (ne všechny oddíly je třeba vždy vyplnit) a zvolení tlačítka *Hledat* se spustí analýza.

4.1.9. Výsledky hledání

V následujícím textu bude popsána výsledková tabulka a možnosti dalšího zpracování zobrazených výsledků. Pro zde použité vzorové vyhledávání byl zvolen výchozí korpus SYN2015, typ dotazu *Lemma*, dotaz *stát* se specifikací slovního druhu *sloveso*.

Ve výsledkové tabulce je zobrazeno shrnutí zadání dotazu a počet nalezených výsledků, tj. počet textů, v nichž se objevil slovní tvar odpovídající zadání. Kromě absolutního počtu výskytů slovního tvaru je uváděn také údaj i.p. m., což je relativní frekvence přepočítaná na celkovou velikost korpusu. Pro kvantifikaci se jedná o relevantnější údaj, než je frekvence absolutní.

Výsledková tabulka obsahuje několik základních částí. Každý řádek (konkordance) je úryvkem textu, v němž se hledané slovo vyskytuje. V každém řádku lze odlišit několik typů údajů, které napříč řádky tvoří sloupce. Zleva se jedná o ikonu umožňující nahlédnout do syntaktické struktury zvolené věty. Následuje bibliografický údaj o zdrojovém textu, po jehož výběru se zobrazí kompletní bibliografické informace o textu. Nejdůležitější je střední část obrazovky, která nabízí úryvky textů s vycentrovaným klíčovým/hledaným slovem (KWIC). Pokud je potřeba zobrazení delšího kontextu, je to umožněno kliknutím na vybrané klíčové slovo. Defaultní nastavení korpusu zobrazuje 40 konkordancí. Pro zobrazení dalších je třeba se přesunout na další stránku výsledků.

The screenshot shows a search results page with the following annotated parts:

- shrnutí zadání dotazu včetně počtu nalezených výskytů**: Points to the search statistics at the top.
- relativní frekvence vztahená k milionu pozic v korpusu**: Points to the 'i.p.m.' column in the search statistics.
- stránkování výsledků**: Points to the page navigation controls at the top right.
- grafika syntaxe**: Points to the syntax tree icon on the left of a result row.
- bibliografický údaj o zdrojovém textu**: Points to the bibliographic information column.
- kontext, ve kterém se hledaný výraz vyskytl**: Points to the text snippet in the KWIC view.
- KWIC (key word in context) hledaný výraz v různých tvarech**: Points to the highlighted search term in the KWIC view.
- jeden výsledkový řádek = konkordance**: Points to a single row of search results.

Obrázek 8: Výsledky hledání

Zobrazením výsledkové tabulky ale práce s korpusem nekončí/nemusí končit. Výsledky lze ještě třídit a filtrovat nebo v nich dále hledat. Všechny nástroje pro zpracování výsledků a ovládání vyhledávání jsou k nalezení v horním menu. V následujícím textu si představíme ty nejdůležitější z nich.

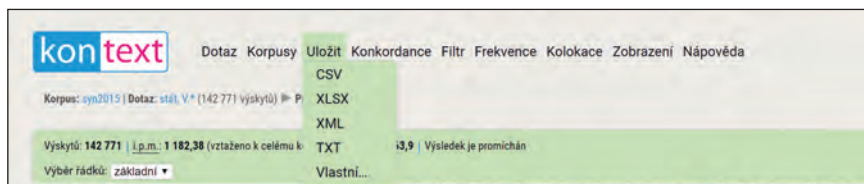
Pomocí záložky *Dotaz* je možné vrátit se do úvodní zadávací tabulky a zahájit tak nové hledání. Nabídka *Předchozí dotazy* zobrazí historii dosavadních dotazů (v rámci jednoho sezení) a umožní se vrátit ke starším vyhledáváním a jejich výsledkům.

V záložce *Korpusy – Dostupné korpusy* je nabízen přehled všech korpusů ČNK, které je možné zvolit jako zdroje pro vyhledávání (obdobně v úvodní zadávací tabulce při volbě korpusu). ČNK umožňuje svým uživatelům vytvoření vlastních subkorpusů, tj. korpusů, které si uživatel definuje z dostupného textového materiálu ČNK sám. Pro vytvoření vlastního subkorpusu musí být relevantní důvod, protože výběrem textů do korpusu jsou významně ovlivněny výsledky hledání a tím i interpretace těchto výsledků. Pro vytvoření a správu vlastních subkorpusů slouží nabídka *Mé subkorpusy*, *Vytvořit vlastní subkorpus* a *Vytvořit veřejný subkorpus*. Vlastní subkorpus bude k dispozici jen danému uživateli, veřejný subkorpus lze nasdílet vybraným uživatelům, např. studentům ve třídě.



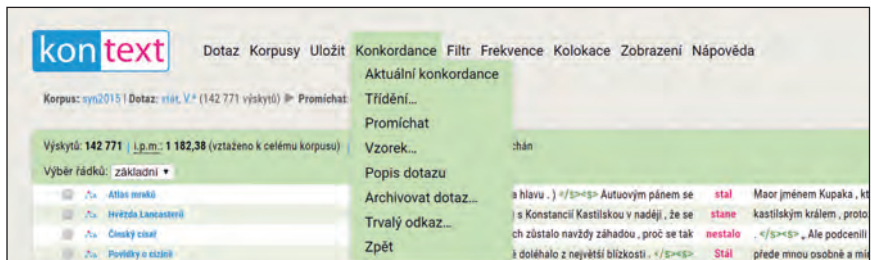
Obrázek 9: Vytvoření vlastního subkorpusu

Výsledky hledání lze uložit na vlastní zařízení, např. na disk počítače. Je tím umožněna práce s daty i v režimu offline. Ukládat lze v různých formátech, např. csv.xlsx.xml či.txt.



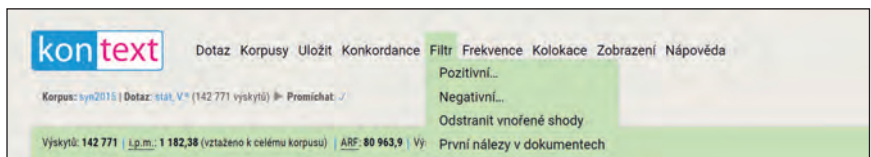
Obrázek 10: Uložení výsledků hledání

S výsledky hledání, resp. jednotlivými konkordancemi, lze dále pracovat. Relevantní může být jejich promíchání (volba *Promíchat*) či výběr náhodného vzorku, jehož velikost určuje uživatel sám (volba *Vzorek*). Tuto volbu lze provést v případech, kdy je počet nalezených konkordancí příliš vysoký na to, aby mohl dojít k jeho manuálnímu zpracování.



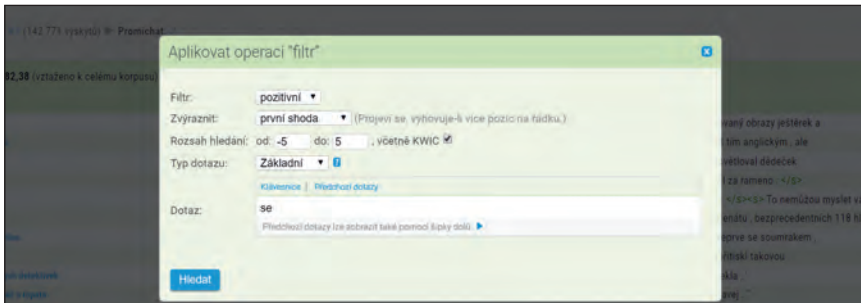
Obrázek 11: Promíchání a výběr vzorku

Výsledky lze různým způsobem také filtrovat. V nabídce je *filtr pozitivní* a *negativní*. V *pozitivním* filtru dochází k přidání další podmínky, např. dalšího slova, se kterým se má hledaný výraz vyskytovat. *Negativní* filtr naopak zaručuje odfiltrování např. nechtěných slovních spojení.

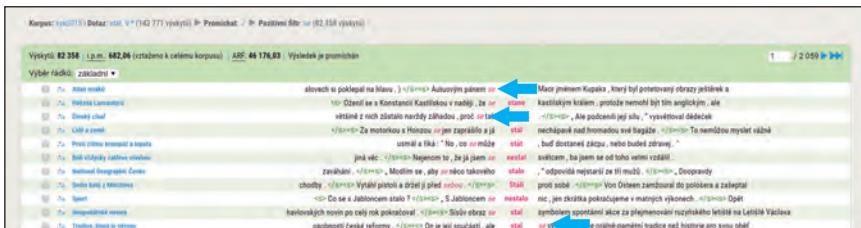


Obrázek 12: Filtrování výsledků

Ukázkou pozitivního filtru v hledání lemmatizovaného slovesa *stát* je přidání podmínky, která zajistí kombinaci slovesa *stát* se zvrtným zájmenem *se*. Zvolený filtr je pozitivní, rozsah hledání -5/5 značí, že zvrtné zájmeno se bude vyskytovat na pozicích pět slovních tvarů vlevo a pět vpravo od klíčového slovesa *stát*. Typ dotazu je zvolen jako základní, čímž je zajištěna flexe zájmena (*se*; *si*). Ve výsledné vyfiltrované tabulce se zvrtná zájmena zobrazují barevně a kurzívou. Obdobným způsobem lze užít i filtr negativní. Příkladem může být hledané přídavné jméno *mobilní*, přičemž za použití negativního filtru jsou ignorovány všechny výskyty s podstatným jménem *telefon* na první pozici vpravo.



Obrázek 13: Využití filtru – výběr užití zvrtného se u slovesa stát



Obrázek 14: Využití filtru – výsledky

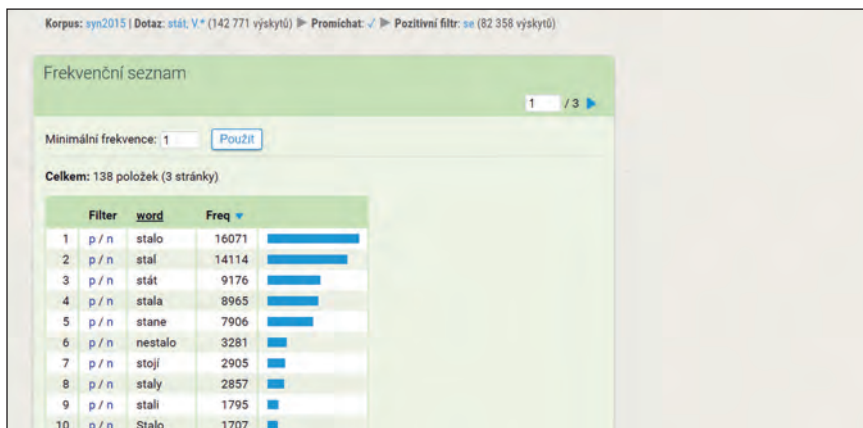
Další z nabídky operací je *Frekvence*. Nabízí statistické údaje o lemmatech, slovních tvarech, rozložení hledaného slova v dokumentech a typech textů a umožňuje provést tutéž statistickou analýzu i na základě vlastních kritérií. Je-li předmětem hledání jedno lemma nebo jeden slovní tvar, nemá smysl zadávat statistickou analýzu lemmatu či slovního tvaru, protože výsledkem je jen zadané lemma či slovní tvar sám. U lemmatu je ovšem možné analyzovat jednotlivé slovní tvary a dozvědět se tak, které slovní tvary z celého paradigmatu převládají.



Obrázek 15: Frekvence – výběr řazení

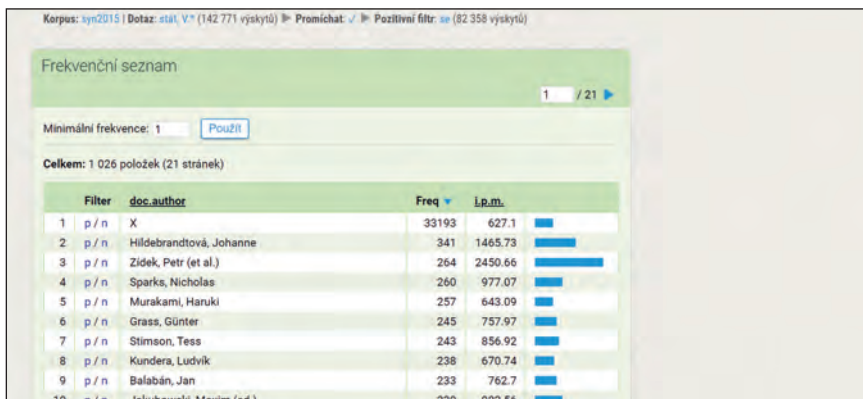
Ve zvoleném příkladu lemmatu *stát* vypadá analýza slovních tvarů takto: nejfrekventovanějším slovním tvarem lemmatu *stát* je *stalo*, následované tvary *stal* a *stát* atd. Frekvenční seznam lze řadit abecedně podle nalezených

tvář (kliknutím na sloupec *word*) anebo u vybraného slovního tvaru použít pozitivní/negativní filtr.



Obrázek 16: Frekvenční seznam - absolutní frekvence

Ve frekvenční analýze dle vlastních kritérií mohou výsledky vypadat následovně (frekvenční seznam byl vytvořen na základě Kritéria *Vlastní - autor*). Seznam je seřazen dle absolutní frekvence. Pro řazení dle relativní frekvence na milion pozic je nutné kliknout na *i.p.m.*



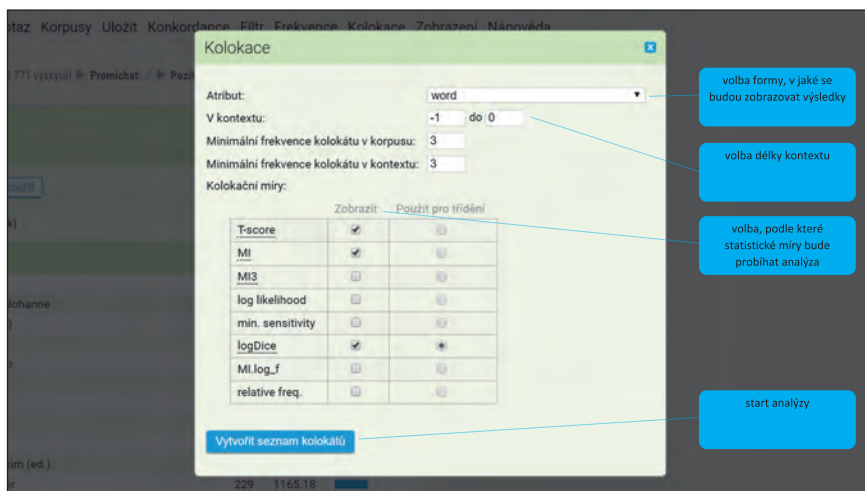
Obrázek 17: Frekvenční seznam

V korpusu lze také nechat analyzovat kontext vyhledaného slova a tím zjišťovat častá slovní spojení (kookurence či kolokace). Slouží k tomu volba *Kolokace*. V tabulce nastavovacích parametrů je třeba zvolit, v jaké formě budou zobrazeny výsledky (položka *Atribut*). Nejčastěji se volí *word* (slovní

tvář) nebo *lemma*. Poté je třeba zvolit prohledávaný kontext. Pozice s minusem označují počet slov vlevo, bez minusu slova vpravo. Výběrem asociačních/statistických měř se určuje, podle kterého algoritmu bude analýza provedena. Každá z asociačních měř bere v úvahu různé parametry ovlivňující spjitost mezi vyhledaným slovem a jeho partnerem/kolokátem (více k různým asociačním mířám najdete na wiki).⁴² Výsledná tabulka nabízí přehled analyzovaných slovních tvarů či lemmat (dle volby), se kterými se hledané slovo nejčastěji spojuje. Výsledky jsou tříděny podle asociační míry zvolené při zadávání, ale lze je nechat přeskupit podle ostatních asociačních měř (jejich zvolením) nebo podle frekvence.



Obrázek 18: Vlastní nastavení kolokace



Obrázek 19: Volba parametrů analýzy kolokací

42 https://wiki.korpus.cz/doku.php/pojmy:asociacni_miry

Korpus: syn2015 | Dotaz: stát: V* (142 771 výskytů) | Promíchat: / | Pozitivní filtr: se (82 358 výskytů)

Kolokace

1

| | Filtr | word | Freq | MI | I-score | logDice |
|-----|-------|---------|-------|-------|---------|---------|
| 1. | p / n | se | 29676 | 4.143 | 162.518 | 8.578 |
| 2. | p / n | může | 1387 | 4.560 | 35.664 | 8.075 |
| 3. | p / n | mohlo | 666 | 6.134 | 25.440 | 7.825 |
| 4. | p / n | nic | 1097 | 4.362 | 31.510 | 7.806 |
| 5. | p / n | něco | 952 | 3.970 | 28.885 | 7.507 |
| 6. | p / n | tak | 1351 | 2.615 | 30.756 | 6.770 |
| 7. | p / n | nikdy | 401 | 3.536 | 18.299 | 6.626 |
| 8. | p / n | to | 2589 | 2.133 | 39.281 | 6.484 |
| 9. | p / n | vlastně | 298 | 3.918 | 16.120 | 6.455 |
| 10. | p / n | něj | 356 | 3.080 | 16.637 | 6.339 |

vyhledané kolokáty
údaj o frekvenci
zvolené asociční míry

Obrázek 20: Kolokace – výsledek hledání

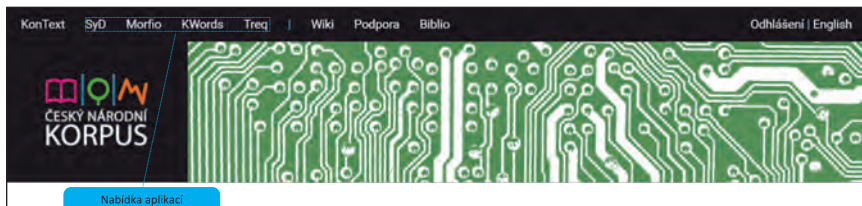
Poslední dvě volby *Zobrazení* a *Nápověda* slouží k nastavení uživatelského prostředí KonTextu a nabízejí odpovědi na nejčastěji kladené otázky. Podrobnější návody k práci s korpusy ČNK a s prohlížečem KonText jsou k nalezení také na wiki.⁴³



Obrázek 21: Výběr zobrazení

4.2. Korpusové aplikace

Vedle vlastní přímé práce s korpusovým vyhledávačem je možné pro práci s korpusovými daty využít také aplikace, které jsou volně dostupné z hlavní webové stránky *korpus.cz*. Jedná se o aplikace SyD, Morfio, KWords a Treq. Všechny jsou dostupné bez registrace a přihlášení, a může je tak využívat kdokoli s přístupem na internet.



Obrázek 22: Náhled menu aplikací

⁴³ <http://wiki.korpus.cz/doku.php>

4.2.1. Aplikace SyD

Tato aplikace umožňuje analýzu a porovnání alespoň dvou jevů, které si v jazyce mohou konkurovat. Může se jednat o varianty pravopisné (*gymnasium* vs. *gymnázium*), morfologické (*mohou* vs. *můžou*), lexikologické (*již* vs. *už*) či syntaktické (*na kurzu* vs. *v kurzu*) a slovosledné (*sebe sama* vs. *sama sebe*). Výsledky je možné nahlížet ze dvou hledisek, synchronního (tedy z hlediska současného jazyka) a diachronního (tedy z hlediska vývoje jazyka v čase). Protože diachronní hledisko není pro výuku žáků s odlišným mateřským jazykem relevantní, věnuje se následující text pouze práci s částí synchronní.

Synchronní náhled nabízí porovnání vybraných variant v textech současného českého jazyka ve formě psané veřejné (analyzován je korpus SYN2010), dále ve formě psané neveřejné (analyzován je korpus KSK-dopisy) a ve formě mluvené (analyzovány jsou korpusy ORAL2006, ORAL2008 a ORAL2013). Výsledky porovnání v této synchronní části nabízejí relativizovaná data o frekvenci vybraných jevů a jejich distribuci v textech (z hlediska textových typů a žánrů) a u mluvených textů data o pohlaví, věku, vzdělání a regionální příslušnosti mluvčích. Zadávat je možné konkrétní slovní tvary nebo lemmata, pro složitější zadání lze využít dotazovací jazyk CQL.

Pro potřeby výuky se jeví využitelnější analýza synchronní. Žáci na ní ověřují rozdíly mezi variantami pravopisnými, morfologickými i lexikálními. Následující obrázky zobrazují a popisují postup práce s aplikací SyD.

Nejprve je třeba najít alespoň dvě slova/slovní tvary, které budou porovnávány. Aplikace umožňuje porovnat i více slovních tvarů (znak +). Volbou *Hledat v současném jazyce* se spustí synchronní analýza.



Obrázek 23: SyD - zadání porovnání tvaru *bychom* a *bysme*

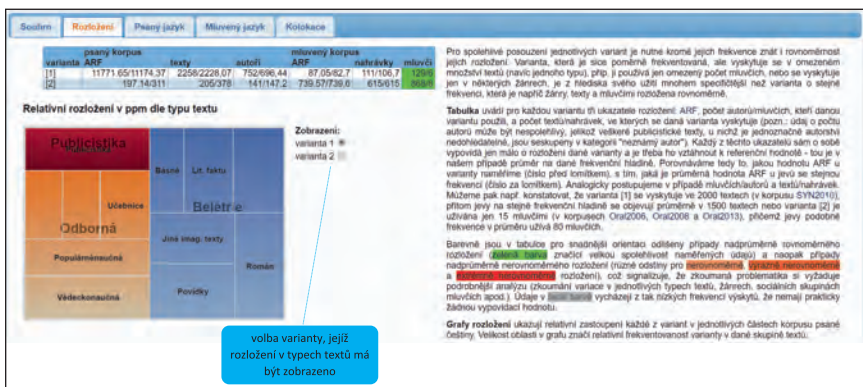
Výsledky se zobrazují v několika různých zpracováních. První výsledky jsou zobrazené v koláčovém grafu a napovídají zastoupení obou zvolených va-

riant v psaném a mluveném jazyce. Ostatní výsledky jsou rozříděné podle dalších kritérií a jsou k nahlédnutí v záložkách pod koláčovými grafy.



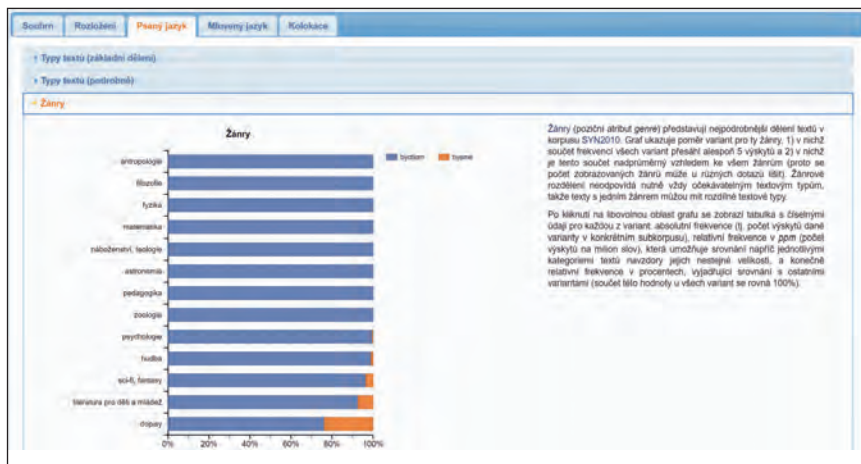
Obrázek 24: SyD – výsledek vyhledávání porovnání tvarů *bychom* a *bysme*

V záložce *Rozložení* je možné nahlédnout do grafu znázorňujícího, v jakých textových typech a žánrech se zvolená varianta nachází. Po zvolení varianty 2 se graf přeskupí. Výsledky jsou opatřeny podrobným komentářem umožňujícím orientaci v problematice interpretace těchto výsledků.



Obrázek 25: SyD – typy textů

V následující záložce jsou graf týkající se psaného jazyka a rozložení zvolených variant v typech textů a textových žánrech zobrazeny podrobněji než v záložce *Rozložení*. Pro ilustraci je v následujícím obrázku znázorněno rozložení podle žánrů.

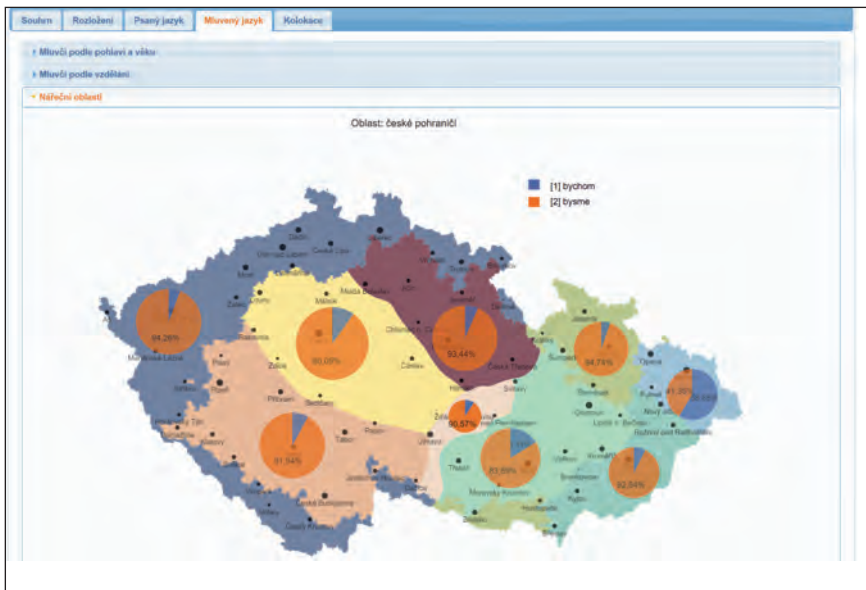


Obrázek 26: SyD - rozložení podle žánrů

Záložka pro mluvený jazyk nabízí jiné kategorie třídění, např. podle pohlaví a věku mluvčího, podle jeho vzdělání či nářeční příslušnosti.



Obrázek 27: SyD - třídění podle pohlaví a věku



Obrázek 28: SyD – třídění podle nářeční příslušnosti

Poslední záložka nabízí nejčastější kolokace k oběma variantám. I zde jsou výsledky opatřeny doprovodným vysvětlujícím textem. Více informací k práci s aplikací SyD je k nalezení ve wiki.⁴⁴



Obrázek 29: SyD – kolokace

44 <https://wiki.korpus.cz/doku.php/manualy:syd>

4.2.2. Aplikace Morfio

Aplikace *Morfio* je navržena tak, aby umožňovala odhalovat a pozorovat slovtvorné pravidelnosti češtiny. Využívá korpusových dat k odhadování rozsahu a produktivity slovtvorných modelů se zaměřením na odvozování. Vychází z poznatků slovtvorby češtiny a vyhledává dvojice (nebo vícečetné kombinace) slov, která se chovají stejně jako zadaný model.

Ve výuce žáků s OMJ lze *Morfio* využít pro odhalování slovtvorných pravidel češtiny, pro uvědomění si jazykového systému a osvojení si nových slovtvorných postupů. Znalosti ze slovtvorby jsou při učení se každému jazyku mocným nástrojem pro rychlé obohacení slovní zásoby o nová slova, a to na základě analogie, kterou slovtvorba disponuje. V případě slovtvorných produktů tak mluvíme o potenciální slovní zásobě, tj. slovní zásobě, kterou žák sice sám nikdy nepoužil ani se s ní nesetkal, ale na základě znalosti slovtvorného pravidla dokáže slovo vytvořit a použít. Práce s *Morfíem* tak představuje induktivní metodu výuky slovtvorby češtiny, která pracuje s jazykovým materiálem zábavnou formou.

Východiskem práce s *Morfíem* je stanovení báze (základu slova) a *formantů* (částí slov, které podléhají změně). Při zadávání do vstupní tabulky lze využít jak kombinace písmen, tak regulárních znaků. Základem pro oba vzory může být tatáž konkrétní báze, např. *škol-*, nebo báze libovolná, tvořená jen regulárními znaky, +, přičemž uživatel volí, zda tato báze, příp. zvolené formanty mají být stejné nebo odlišné pro oba vzory. Oběma vzorům lze přiřknout ještě zvolený slovní druh, je-li tento údaj relevantní (např. slovo končící na *-í* může příslušet k různým slovním druhům, a proto lze slovní druh vybrat). Výběr slovního druhu se doporučuje především proto, aby se eliminovaly nežádoucí výsledky. U některých bází dochází při spojení s různými formanty ke změnám, s nimiž *Morfio* počítá a nazývá je *alternace*. Jedná se o změny typu *knih-a X kniž-ní; psá-t X psa-ní* apod. Důležité je, že ve vstupní tabulce není třeba vždy vyplňovat všechna zadávací pole, ale vždy jen taková, která jsou podstatněná vzhledem k cíli analýzy.

Východiskem je stanovení báze (základu slova) a *formantů* (částí slov, které podléhají změně). Základem slova *škola* (tj. bází) je *škol-*, formantem je *-a*, přičemž i bez korpusu jsme sami schopni vytvořit další slova s jinými formanty od téže báze, např. *školní, školský, školník, školení* apod. Díky *Morfíu* můžeme rychle odhalit, která slova jsou schopna stejné změny jako slovo vyhledávané. Drobnou nevýhodou aplikace *Morfio* je, že nedokáže rozeznat hranice morfémů (např. u předpony *před-* jsou generována slova *předevčirem*, ale také *předobry*).



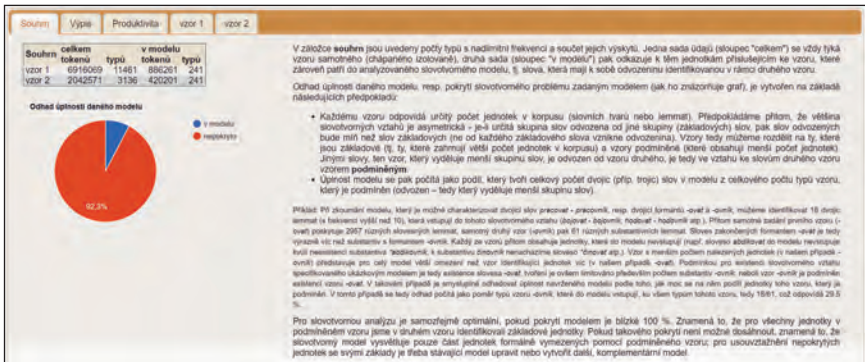
Obrázek 30: Morfio

Následující zadání umožňuje hledat dvojice slov, jejichž báze může být různá, ale která tvoří vždy dvojici, např. substantivum končící na *-a* a tvořící adjektivum končící na *-ní*. Hledají se a ve výsledcích se zobrazují lemmata. Alternace nebyly určeny.

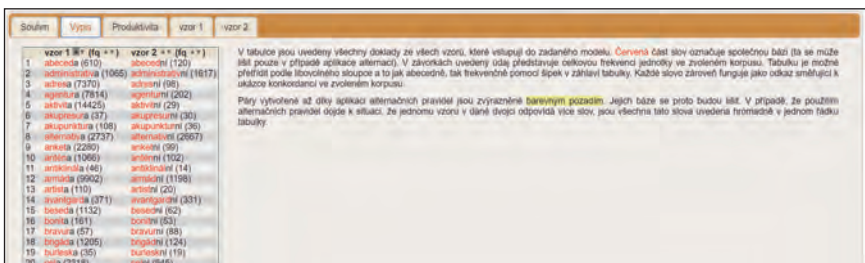


Obrázek 31: Morfio – zadání dotazu

Výsledky analýzy se zobrazí ve formě několika záložek a mohou mít různou míru využití pro různé účely. Každá z výsledkových tabulek je opatřena doprovodným komentářem. Úvodní tabulka a graf představují souhrnné výsledky analýzy. Pro účely výuky se jeví nejvyužitelnější tabulka *Výpis*, nabízející nalezené páry, které jsou kandidáty na hledaný slovtvorný produkt (např. abeceda – abecední). Jednotlivé slovní tvary lze kliknutím aktivovat a dostat se tak do korpusových textů, v nichž lze pozorovat konkrétní užití daného slovního tvaru.



Obrázek 32: Morfio – souhrnné výsledky



Obrázek 33: Morfio – záložka Výpis

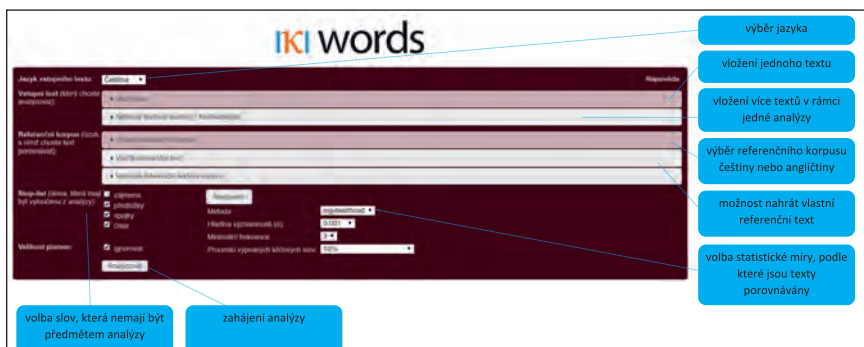
4.2.3. Aplikace KWords

KWords je aplikace umožňující analýzu až 20 vybraných textů, např. stejného textového druhu či žánru, při níž probíhá porovnání těchto vložených textů s referenčním korpusem. Při analýze jsou v textu vyhledávána klíčová slova (měří se relativní frekvence slovoformy každého slovního tvaru a porovnává se s relativní frekvencí téhož slovního tvaru v referenčním korpuse, čímž lze získat seznam slovních tvarů, které jsou ve vybraných textech užity výrazně častěji a lze o nich uvažovat jako o klíčových, resp. majících spojitost s daným textovým druhem, resp. žánrem).

Texty lze v současné době nahrávat v češtině nebo angličtině. Pro relevantní výsledky je však dobré zvolit volbu vhodného referenčního korpusu, s nímž budou vložené texty porovnávány. Nabídka referenčních korpusů lze nalézt na wiki.⁴⁵ Pro výuku žáků s OMJ však není angličtina relevantní a z nabízených českých korpusů doporučujeme využít přednastavený referenční korpus SYN2015.

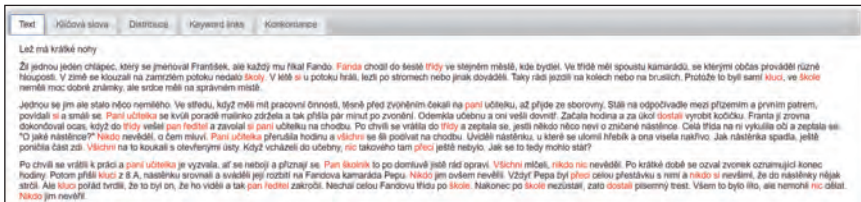
45 <https://wiki.korpus.cz/doku.php/manualy:kwords>

Práce s aplikací je intuitivní. Po otevření aplikace na webové stránce je třeba vyplnit zadávací tabulku. Klíčové je zadání vlastního textu. Analyzovaný text by měl mít od několika set do několika tisíc znaků a je možné ho vložit dvěma způsoby. Text (nebo i více textů) lze přímo kopírovat do připraveného okna, nebo lze v případě vkládání více textů užít tzv. multianalýzy, při níž se do korpusu jednotlivé texty nahrají. Druhým klíčovým krokem je zvolení referenčního korpusu, tedy korpusu, s nímž se vlastní nahrané texty budou porovnávat. Na výběr je několik referenčních korpusů. Pro analýzu současných textů doporučujeme využívat např. SYN2015. Před zahájením analýzy je vhodné zvážit, zda z ní nevyločit některé/všechny nabízené slovní druhy, které jsou většinou vysokofrekvenční a mohou výsledky analýzy znehodnotit, např. předložky, zájmena, spojky apod. Vyloučení z analýzy lze provést v tzv. stop-listu. Vlastní analýza probíhá na základě statistických měř; je možné zvolit jednu ze statistických měř nebo obě dvě zároveň. Analýza začne probíhat po stisku příkazu *Analýzovat* v levé dolní části stránky.



Obrázek 34: KWords – Zahájení analýzy

Výsledky analýzy se zobrazují několika způsoby. V záložce *Text* je zobrazen analyzovaný text a všechna nalezená signifikantní slova jsou zbarvena červeně. Text lze tedy pročitat a soustředit se přitom na vyhledané výrazy.



Obrázek 35: KWords – ukázka analýzy⁴⁶

Výsledky analýzy jsou nabízeny také v jiných formách. V záložce *Klíčová slova* uživatel nalezne souhrnnou tabulku slovních tvarů vyhodnocených jako signifikantní pro zadaný text. Kromě seznamu je zde nabídnuta také hodnota statistické míry, podle které byl výpočet proveden, dále hodnota DIN, která uvádí míru relevance rozdílu mezi výskytem v zadaném textu a v referenčním korpusu (zde platí, že relevantní jsou slovní tvary s hodnotou 75–100, přičemž čím je hodnota vyšší, tím je relevantnější). U vygenerovaných slovních tvarů je také uvedena jejich frekvence v obou textových zdrojích.

| Tvar | LL | DIN | Frekvence | Referenční korpus | TC |
|------------|--------|---------|-----------|-------------------|----|
| 1 šáček | 44.678 | 99.9578 | 3 | 67 | - |
| 2 hrabě | 79.932 | 99.8984 | 6 | 819 | - |
| 3 přeci | 27.993 | 99.2352 | 3 | 1392 | - |
| 4 učitelka | 26.206 | 99.0677 | 3 | 1936 | - |
| 5 kluci | 40.008 | 98.6542 | 5 | 4653 | - |
| 6 třída | 36.811 | 98.0996 | 5 | 6795 | - |
| 7 šafář | 28.433 | 97.2980 | 4 | 7525 | - |
| 8 paní | 50.870 | 96.8812 | 6 | 8952 | - |
| 9 řečnick | 30.632 | 96.6295 | 5 | 11773 | - |
| 10 škola | 24.834 | 96.5393 | 4 | 3701 | - |
| 11 dostal | 17.670 | 96.2569 | 3 | 7850 | - |
| 12 paní | 24.208 | 95.4595 | 5 | 23221 | - |
| 13 nikdo | 32.375 | 92.7107 | 7 | 36372 | - |
| 14 škola | 32.898 | 92.9786 | 6 | 18466 | - |
| 15 všichni | 18.596 | 88.5220 | 5 | 41818 | - |
| 16 it | 11.818 | 82.4893 | 4 | 52725 | - |

Obrázek 36: KWords – výsledek analýzy

Další možností náhledu je distribuce zobrazující rozložení nalezených signifikantních slov v textu a *Keyword links*, jenž zobrazuje vztahy mezi jednotlivými klíčovými slovy. Interpretace těchto dat je určena spíše pro odborníky a zkušené korpusové uživatele, proto se těmto doplňkovým informacím nebudeme dále věnovat. Blížší popis pro zájemce lze nalézt na wiki.

46 Analyzovaný text pochází z: <http://www.cesky-jazyk.cz/slovky/pohadky/lez-ma-kratke-nohy.html#axzz5IU96O5h7>.

Poslední z řady je přehled konkordancí, tj. konkrétních míst v textu v zobrazení KWIC. Po výběru konkordance se soupis rozšíří o všechny konkordance s vybraným slovním tvarem. Tento náhled je výhodný především proto, že shromáždí všechny výskyty vybraného slovního tvaru pod jednu konkordanci a lze na něm, na rozdíl od náhledu lineárního textu, analyzovat chování vybraného slovního tvaru ve všech jeho výskytech najednou.

The screenshot shows the 'KWordy' application interface. At the top, there are tabs for 'Text', 'Klíčová slova', 'Distribuce', 'Klíčový výpis', and 'Konkordance'. The main window displays a concordance table for the keyword 'fanda'. The table is organized into columns: 'Vý POS', 'Text', 'Klíčové slovo', and 'Klíčový kontext'. The 'Text' column contains several lines of text, with the word 'fanda' highlighted in red. The 'Klíčové slovo' column lists the word 'fanda' and its frequency in each line. The 'Klíčový kontext' column shows the surrounding text for each occurrence. Two blue callout boxes are present: one pointing to the word 'fanda' in the 'Klíčové slovo' column, labeled 'klíčové slovo', and another pointing to the 'Klíčový kontext' column, labeled 'rozbalená nabídka konkordancí ke slovnímu tvaru "fanda"'. The table data is as follows:

| Vý POS | Text | Klíčové slovo | Klíčový kontext |
|--------|---|---------------|---|
| 1 | navíc (1) • zato (1) • úkol (1) | dostali (3) | důtku (1) • písničky (1) • vyrobit (1) |
| 2 | (4) • se (1) • ale (1) | fanda (6) | se (2) • nepije (1) • ještě (1) • chodí (1) |
| 3 | 1) LEVÝ KONTEXT | fanda (6) | PROVÝ KONTEXT |
| 4 | trénovat, ale každý měl svou fandu. | fanda | zavazet do země, tedy se stavěním raději |
| 5 | však: jen napoleo. To bylo využívání ke | fanda | ještě dříve v Amer. - protibře maršské např |
| 6 | se ne vyprávěl, ani všechny tyhle | fanda | se zabalit a tak. Je to nejlepší |
| 7 | obslužovali. Ji to tak u státi | fanda | zavazet sám a užít si učitky, jen |
| 8 | řetnu přehrávat z laskavé na něj zavazet. | fanda | se zabalit, zavazet - své věci a |
| 9 | odatny. což byla byla (často), ale | fanda | menší, zavazet, zavazet by se. |
| 10 | ? (1) • přáli (1) • namí (1) • st (1) | kluci (5) | si (1) • se (1) • poříd (1) • (1) |
| 11 | někdo (2) • že (1) • nás (1) • (1) | nic (4) | navyšel (1) • dítat (1) • nemůže (1) • netuší (1) |
| 12 | (3) • a (2) • (1) • (1) | někdo (7) | jim (2) • nic (2) • si (1) • navštívil (1) |
| 13 | (3) • ai (1) • věrní (1) • tak (1) | pan (3) | ředil (2) • školník (3) |
| 14 | (2) • ai (1) • a (1) • se (1) | pan (5) | učebka (2) • učebku (2) |
| 15 | tam (1) • byl (1) • je (1) | přeci (3) | ještě (1) • celou (1) • velký (1) |
| 16 | pan (2) | ředil (5) | a (1) • zakročil (1) • je (1) • pozval (1) |
| 17 | pozdil (1) • kluci (1) • ještě (1) • by (1) | si (4) | namožením (1) • své (1) • a (1) • myslím (1) |

Obrázek 37: KWords – přehled konkordancí

Využití aplikace KWords ve výuce jazyka může být různorodé. Nabízí se možnost analýzy textových druhů a žánrů, na jejímž podkladu žáci samostatně vytvářejí vybraný textový druh či žánr. Zajímavou možností je analýza vlastních vytvořených textů a tím reflexe vlastního idiolektu nebo porovnání textů stejného zadání od žáků s češtinou jako mateřštinou a žáků s OMJ.

4.2.4. Aplikace Treq

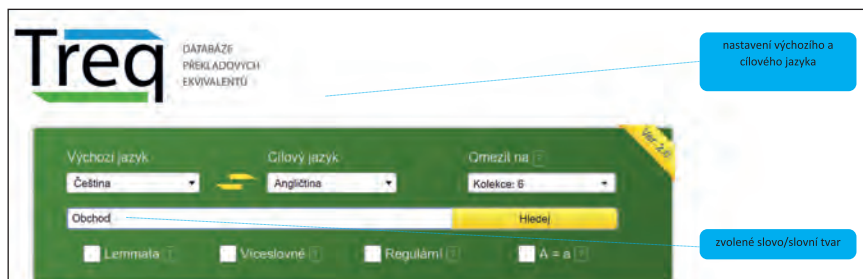
Aplikace Treq je databází překladových ekvivalentů. Protože je založena na textech paralelního korpusu InterCorp, funguje obousměrně pro jazykové páry čeština–cizí jazyk a angličtina–cizí jazyk. Cizí jazyky jsou zde nabízeny v rozsahu odpovídajícím jazykům zastoupeným v InterCorpu.

Samotné vyhledávání probíhá v prvním kroku zvolením výchozího a cílového jazyka. V druhém kroku musí být zvolen hledaný výraz. Vyhledávané slovo musí být slovo z jazyka označeného jako výchozí a lze ho formulovat jako určitý slovní tvar, jako lemma, případně lze vyhledávat víceslovná spojení nebo užít při hledání regulárních výrazů či ignorovat malá a velká písmena (až na vyhledávání konkrétního slovního tvaru je u všech výše jmenovaných potřeba danou kategorii zaškrtnout). Hledání lze provést ve všech textech InterCorpu nebo lze omezit zdrojové texty na beletristické jádro (texty ručně zkontrolované a zarovnané) nebo některou z kolekcí (texty automaticky zarovnané). Výsledkem hledání je seznam překladových ekvivalentů. Proklíkem jednotlivých ekvivalentů se uživatel dostane do korpusu InterCorp, který mu zobrazí zarovnané texty obsahující jak výraz jazyka výchozího, tak nabídnutý ekvivalent jazyka cílového.

Na těchto kontextech lze ověřit, zda zvolený ekvivalent odpovídá nárokům hledaného ekvivalentu.

Své využití najde *Treq* především tam, kde se žáci učí pracovat s cizojazyčnými ekvivalenty a jejich výběrem a také se synonymy. *Treq* sice nelze označit za slovník, protože mu chybí celá řada nezbytných lexikografických informací, ale na rozdíl od běžného slovníku disponuje velkou výhodou, kterou je propojení s autentickými texty ve formě konkordancí obsahujících vyhledaný ekvivalent. Z kontextů tak lze mnohdy vyčíst stejné nebo ještě bohatší informace o významové a pragmatické stránce jazykového ekvivalentu než ze slovníku. Uživatel si užíváním *Trequ* cvičí nejen své jazykové znalosti, ale i svůj jazykový cit a senzibilizuje se na práci s textovým materiálem a jeho výpovědní hodnotou.

Následující obrázky nabízejí náhled úvodní zadávací stránky aplikace *Treq*, stránky s výsledky (nabídkou překladových ekvivalentů) a stránky zobrazující konkordance *InterCorpu* u zvoleného ekvivalentu. Více informací o aplikaci *Treq* je k nalezení na stránkách wiki.⁴⁷



Obrázek 38: Treq – zadávací tabulka

| ▲ Frekvence ▼ | ▲ Procenta ▼ | ▲ Čeština ▼ | ▲ Angličtina ▼ |
|---------------|--------------|-------------|----------------|
| 3364 | 52,5 | obchod | trade |
| 941 | 14,7 | obchod | business |
| 566 | 8,8 | obchod | Trade |
| 366 | 5,7 | obchod | store |
| 358 | 5,6 | obchod | deal |
| 355 | 5,5 | obchod | shop |
| 79 | 1,2 | obchod | industry |
| 47 | 0,7 | obchod | transaction |
| 35 | 0,5 | obchod | bargain |
| 32 | 0,5 | obchod | commerce |

Obrázek 39: Treq – výsledek frekvenční analýzy

47 <https://wiki.korpus.cz/doku.php/manualy:treq>

Korpus: InterCorp v9 - Czech | Databaz obchod | Annotovaný | Download | Settings

Výskytů: 3 000 (p. m. 16,36) (vztaheno k celému korpusu) | AČ: 838,14 | Výsledky je seříděn 1 / 58

Výběr řádků: záměrně

| InterCorp v9 - Czech | | InterCorp v9 - English | |
|----------------------|----------------------|------------------------|--|
| 1 | obchod-vestavaci_m_h | obchod-vestavaci_m_h | Well, in fact there is an awful lot of this, largely because of the total lack of money, trade, banks, art, or anything else that might keep all the non-existent people of the Universe occupied |
| 2 | obchod-hankovic_jar | obchod-hankovic_jar | Here all trade was conducted: |
| 3 | AACH obchod_na_brihu | AACH obchod_na_brihu | than a hamlet, stood on the right bank of the Drina on the slopes of the steep hill below the ruins of the one-time fortress, for then it did not have the size and shape it was to have later when the bridge was built and communications and trade developed. |
| 4 | AACH obchod_na_brihu | AACH obchod_na_brihu | That changed the conditions of life for the whole district and for the town also, influenced trade and communications, and the mutual relations of Turks and Serbs. |
| 5 | AACH obchod_na_brihu | AACH obchod_na_brihu | Then came the time of the occupation and with it livelier trade, easier gain and lower expenses. |

Obrázek 40: Treq – náhled korpusu InterCorp – české slovo *obchod* a anglický ekvivalent *trade*

Využití korpusů ve výuce

5. Náměty pro strategie vyučování a pro tvorbu pracovních listů

5.1. Frekvenční analýza

Korpusová data je možné mj. využít také pro různé strategie vyučování, např. pro výuku slovní zásoby. Pro tento účel je dobře využitelná frekvenční analýza. Ve slovní zásobě jazyka existuje souvislost mezi frekvencí výskytu slova a jeho pořadím. Jde o tzv. Zipfův zákon. Mluvnické současné češtiny (2010) uvádí, že *když seřadíme různá slova v určitém korpusu od nejfrekventovanějšího po nejméně časté a přiřadíme ke každému slovu číslo označující jeho pořadí, můžeme si všimnout, že pořadí slova krát jeho frekvence je víceméně konstantní. Z tohoto vztahu vyplývá, že v jazyce existuje velmi málo hodně frekventovaných slov – to jsou zejména gramatická slova jako předložky a spojky. Většinu lexikonu ale tvoří velké množství slov, která mají velmi nízkou frekvenci. Z uvedeného zákona plynou praktické důsledky pro výuku slovní zásoby ve výuce cizího jazyka. Nejfrekventovanějších 100 slov totiž pokrývá necelých 40 % textu, 1000 slov tvoří 62 % textu a se znalostí 3000 slov jsme schopni rozumět více než 75 % textu (Cvrček a kol 2010: 78).* Jako ukázkou využití frekvenční distribuce slovní zásoby můžeme použít následující tabulku. Data v tabulce pocházejí z korpusu SYN2015. Při práci s korpusovým manažerem KonText jsme vybrali typ dotazu lemma, tedy základní (slovníkový) tvar, který není ovlivněn různou frekvencí výskytů např. různých pádových tvarů. Do vyhledávacího řádku jsme zadali symbol.*, tedy symbol pro vyhledání všech slov o délce alespoň jednoho písmena. Při vyhledávání jsme omezili vyhledávání pouze na beletrii a publicistiku, oborovou literaturu jsme nezahrnuli.

Výsledkem je tabulka zahrnující sto nejvíce frekventovaných slov v daném korpusu. Podle vztahů uvedených výše by mělo platit, že těchto sto slov pokrývá 40 % textů v korpusu SYN2015.

| Pořadí | Vše | Substantiva | Adjektiva | Slovesa | Předložky | Spojky |
|--------|-----|-------------|-----------|---------|-----------|--------|
| 1 | být | rok | velký | být | v | a |
| 2 | se | člověk | celý | mít | na | že |
| 3 | a | den | další | moci | s | ale |

| Pořadí | Vše | Substantiva | Adjektiva | Slovesa | Předložky | Spojky |
|--------|---------|-------------|-----------|------------|-----------------|----------|
| 4 | v | ruka | dobrý | chtít | z | i |
| 5 | ten | dítě | nový | řici | do | jako |
| 6 | na | život | jiný | muset | o | když |
| 7 | on | místo | malý | vědět | k | aby |
| 8 | že | hlava | český | jít | za | nebo |
| 9 | s | doba | poslední | stát | po | než |
| 10 | z | oko | starý | dát | pro | ani |
| 11 | mít | muž | rád | řikat | od | protože |
| 12 | do | žena | vysoký | vidět | u | však |
| 13 | který | město | mladý | začít | před | pokud |
| 14 | já | svět | vlastní | myslet | při | kdyby |
| 15 | ale | čas | jediný | přijít | podle | či |
| 16 | o | dům | dlouhý | dostat | mezi | jestli |
| 17 | i | země | možný | dělat | bez | takže |
| 18 | k | strana | hlavní | udělat | pod | až |
| 19 | jako | chvíle | stejný | nechat | nad | tedy |
| 20 | co | práce | ostatní | najít | přes | ovšem |
| 21 | za | cesta | bílý | dokázat | kolem | jenže |
| 22 | tak | věc | špatný | čekat | proti | zatímco |
| 23 | moci | hodina | plný | hrát | kvůli | zda |
| 24 | svůj | voda | černý | mluvit | během | li |
| 25 | když | koruna | jistý | vypadat | díky | vždyť |
| 26 | po | dveře | evropský | vrátit | kromě | ať |
| 27 | oni | slovo | důležitý | vzít | vedle | dokud |
| 28 | už | Praha | různý | snažit | místo | jakmile |
| 29 | rok | případ | známý | zůstat | mimo | přestože |
| 30 | jak | konec | těžký | cítit | včetně | ačkoli |
| 31 | pro | cena | krásný | zeptat | uprostřed | aniž |
| 32 | všechen | milión | domácí | potřebovat | vůči | neboť |
| 33 | aby | pohled | jasný | znát | vzhledem | anebo |
| 34 | od | peníze | podobný | slyšet | pomocí | proto |
| 35 | jen | tvář | silný | žít | okolo | nicméně |
| 36 | jeho | škola | americký | podívat | navzdory | buď |
| 37 | chtít | problém | bývalý | patřit | podél | jak |
| 38 | řici | část | státní | sedět | skrz | avšak |
| 39 | jeden | týden | pražský | rozhodnout | oproti | byť |
| 40 | člověk | společnost | současný | vést | prostřednictvím | jelikož |
| 41 | muset | firma | příští | chodit | uvnitř | zato |

| Pořadí | Vše | Substantiva | Adjektiva | Slovesa | Předložky | Spojky |
|--------|-------|-------------|------------|------------|-----------|----------|
| 42 | nebo | otec | minulý | dojít | naproti | jestliže |
| 43 | my | auto | místní | získat | nedaleko | jenomže |
| 44 | stát | hlas | mrtvý | otevřít | dle | coby |
| 45 | vědět | tělo | obrovský | pomoci | poblíž | nýbrž |
| 46 | ještě | noha | světový | věřit | ohledně | tak |
| 47 | jít | měsíc | lidský | pracovat | von | tudíž |
| 48 | u | noc | příjemný | ukázat | blízko | jednak |
| 49 | pak | matka | zajímavý | stačit | van | ač |
| 50 | tento | minuta | červený | dívat | zpod | neboli |
| 51 | než | řada | pravý | objevit | | |
| 52 | velký | rodina | společný | jet | | |
| 53 | až | jméno | schopný | zdat | | |
| 54 | ani | zápas | základní | skončit | | |
| 55 | před | ulice | politický | znamenat | | |
| 56 | vy | pravda | veřejný | odpovědět | | |
| 57 | jenž | stůl | letošní | držet | | |
| 58 | ty | pán | šťastný | uvést | | |
| 59 | dva | stát | nízký | dávat | | |
| 60 | můj | procento | zvláštní | pokračovat | | |
| 61 | celý | tým | krátký | umět | | |
| 62 | tam | knih | správný | vyjít | | |
| 63 | hodně | okno | národní | ležet | | |
| 64 | její | pokoj | zelený | lze | | |
| 65 | další | kraj | německý | hledat | | |
| 66 | něco | pan | finanční | brát | | |
| 67 | dát | otázka | otevřený | změnit | | |
| 68 | dobry | situace | blízky | psát | | |
| 69 | den | film | sociální | dodat | | |
| 70 | ne | Kč | vhodný | podařit | | |
| 71 | nový | pocit | samotný | začínat | | |
| 72 | řikat | hra | modrý | odejít | | |
| 73 | nic | projekt | jednoduchý | zastavit | | |
| 74 | také | smrt | skutečný | zjistit | | |
| 75 | vidět | paní | městský | připravit | | |
| 76 | kde | síla | bezpečný | představit | | |
| 77 | první | způsob | zlatý | zvednout | | |
| 78 | teď | vláda | životní | platit | | |
| 79 | ruka | většina | rychlý | sledovat | | |

| Pořadí | Vše | Substantiva | Adjektiva | Slovesa | Předložky | Spojky |
|--------|---------|-------------|------------|------------|-----------|--------|
| 80 | místo | vlas | dětský | trvat | | |
| 81 | při | skupina | osobní | bát | | |
| 82 | začít | důvod | volný | tvrdit | | |
| 83 | protože | světlo | slavný | postavit | | |
| 84 | podle | možnost | skvělý | existovat | | |
| 85 | jiný | hráč | pracovní | vydat | | |
| 86 | každý | kolo | prázdný | líbit | | |
| 87 | druhý | byt | široký | připadat | | |
| 88 | myslet | výsledek | určitý | napsat | | |
| 89 | tenhle | klub | střední | uvidět | | |
| 90 | náš | rameno | nutný | zapomenout | | |
| 91 | takový | Petr | moderní | ptát | | |
| 92 | dítě | syn | oblíbený | poznat | | |
| 93 | mezi | metr | dnešní | napadnout | | |
| 94 | nějaký | srdce | ruský | koupit | | |
| 95 | jejich | system | čistý | působit | | |
| 96 | přijít | přítel | rodinný | otočit | | |
| 97 | život | začátek | úspěšný | milovat | | |
| 98 | však | krok | hezký | číst | | |
| 99 | dostat | akce | jednotlivý | chápat | | |
| 100 | hlava | policie | krajský | pustit | | |

Tabulka 1: Nejfrekventovanější slova v korpusu SYN2015

5.2. Pracovní listy

Následující ukázkové pracovní listy obsahují čtyři části. Na první straně naleznete anotaci a popsaný postup práce s pracovním listem. Druhou stranu tvoří samotný pracovní list, který lze namnožit a rozdat žákům. Po pracovním listu následuje popis práce s korpusem s názornými snímky obrazovky. Na konec je zařazeno řešení úkolů. Pracovní listy obsahují jak cvičení, která vyžadují přímou práci s korpusem během výuky (hands-on), tak cvičení, která jsou založená na jazykových datech z korpusu, ale přímou práci s korpusem během výuky nevyžadují (hands-off). Pracovní listy mají sloužit jako inspirace pro tvorbu vlastních korpusových cvičení.⁴⁸

⁴⁸ Podobné pracovní listy naleznete také pod odkazem <https://www.korpus.cz/proskoly>.

5.2.1 Prostředí KonText

Pracovní list č. 1:

ADJEKTIVA ZDRAVÝ A ZDRAVOTNÍ – Zdravý životní styl = žádné zdravotní problémy?

Anotace

Cvičení tohoto typu je vhodné zařadit do výuky, pokud dělá žákům problém rozlišovat význam podobných adjektiv a užívat je ve správných spojeních. Může jít jak o adjektiva, která mají jiný význam – zde např. *zdravý* a *zdravotní*, tak o adjektiva, která mají v některých kolokacích význam stejný, ale existují různá omezení – např. *šedý* a *šedivý*, kdy se některá substantiva mohou pojít jen s jedním adjektivem (*šedá eminence*). Cvičení také rozšiřují slovní zásobu. Pracovní list vyžaduje samostatnou práci s korpusovým rozhraním KonText a obsahuje tři úkoly, které vedou k tomu, aby si žák vhodné použití adjektiv osvojl sám pomocí vyhledávání typických kolokací, následně formulací pravidel jejich spojování se substantivy, a nakonec jejich použitím v textu.

Postup práce

Před rozdáním pracovního listu je nutné žáky připravit na téma úkolů a vhodně je motivovat k jejich vypracování. V tomto případě může učitel s žáky mluvit o tom, co sami dělají pro své zdraví a jaké existují zdravotní problémy. K diskusi může využít předem připravený text, který bude obsahovat tematizovaná adjektiva, nebo jen otázku v názvu pracovního listu.

Po tomto úvodu učitel obrátí pozornost žáků k adjektivům *zdravý* a *zdravotní* a zeptá se jich, jaký je v nich rozdíl a která substantiva se s nimi mohou pojít. Jako motivační prvek dobře slouží návodné obrázky (zdravé jídlo nebo tělesné aktivity x různá onemocnění, nemocniční personál). Odpovědi žáků lze zapisovat na tabuli formou myšlenkové mapy. Následně učitel rozdá žákům pracovní list. Ti pak splní první úkol dle schopnosti práce s korpusem sami, ve dvojicích nebo společně s učitelem. Před plněním je vhodné projít postup práce s korpusem. Žáci mohou první cvičení také dostat za domácí úkol. V případě, že žáci některým substantivům, která najdou, nerozumí, je třeba jejich význam osvětlit. Vhodné je využít předem připravené obrázky a příkladové věty. Příkladové věty mohou žáci sami najít kliknutím na pozitivní filtr (p/n) u vybraného substantiva v seznamu kolokátů, po kterém se jim objeví všechny věty s daným adjektivem a substantivem. Věty pocházejí z českých textů, je třeba proto počítat s tím, že mohou být pro žáky na nižší jazykové úrovni (asi do B2) nesrozumitelné a naopak matoucí. Důležité je

upozornit žáky také na gramatický rozdíl mezi měkkým adjektivem *zdravotní* a tvrdým adjektivem *zdravý*. Tvary adjektiv si opět mohou ověřit kliknutím na pozitivní filtr u daného substantiva. Nalezené kolokace by během kontroly vždy měli číst se správnou koncovkou adjektiva (tj. *zdravý rozum, zdravá výživa X zdravotní problém, zdravotní sestra*).

Až si žáci najdou substantiva a bude osvětlen jejich význam, měli by se sami nebo ve skupině pokusit zformulovat významovou definici. Mohou vymyslet příklady dalších kolokátů (některé jsou uvedeny v části řešení).

Druhý úkol žáci plní až po kontrole prvního. Opět je nutné, aby před začátkem práce rozuměli zadaným substantivům a aby adjektiva použili se správnou koncovkou.

Třetí úkol mohou žáci plnit samostatně nebo ve dvojicích. Mohou souměřit o to, který text bude nejuvtipnější nebo ve kterém bude použito nejvíce spojení adjektiv a substantiv. Učitel si může připravit začátek textu, na který budou žáci navazovat. Např. *Když byla Klára malá, měla hodně zdravotních problémů. Kvůli nim často chyběla ve škole, a tak ještě v deseti letech neuměla dobře číst a psát...* Texty poté mohou žáci přečíst v plénu nebo je učitel může vybrat a opravit. Texty si také žáci mohou opravit navzájem.

ZDRAVÝ ŽIVOTNÍ STYL = ŽÁDNÉ ZDRAVOTNÍ PROBLÉMY?

1. Najděte v korpusu syn2015 deset nejčastějších substantiv po adjektivech *zdravý* a *zdravotní*. Potom vysvětlíte, co znamenají a kdy se používají.

| | zdravý | zdravotní |
|-----|--------|-----------|
| 1. | | |
| 2. | | |
| 3. | | |
| 4. | | |
| 5. | | |
| 6. | | |
| 7. | | |
| 8. | | |
| 9. | | |
| 10. | | |

2. Co ještě může být *zdravé* a co *zdravotní*? Přiřaďte a napište k substantivům adjektivum se správnou koncovkou.

handicap, holčička, noha, obtíž, prohlídka, zub

zdravý:

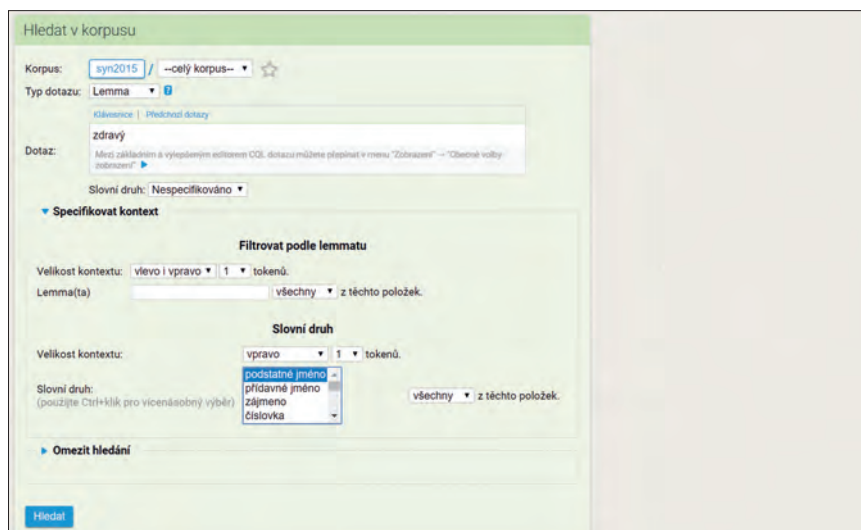
zdravotní:

3. Napište krátký text, ve kterém použijete vždy alespoň třikrát adjektivum *zdravý* a *zdravotní*. Ke každému adjektivu musíte použít jiné substantivum.

Postup práce s korpusem

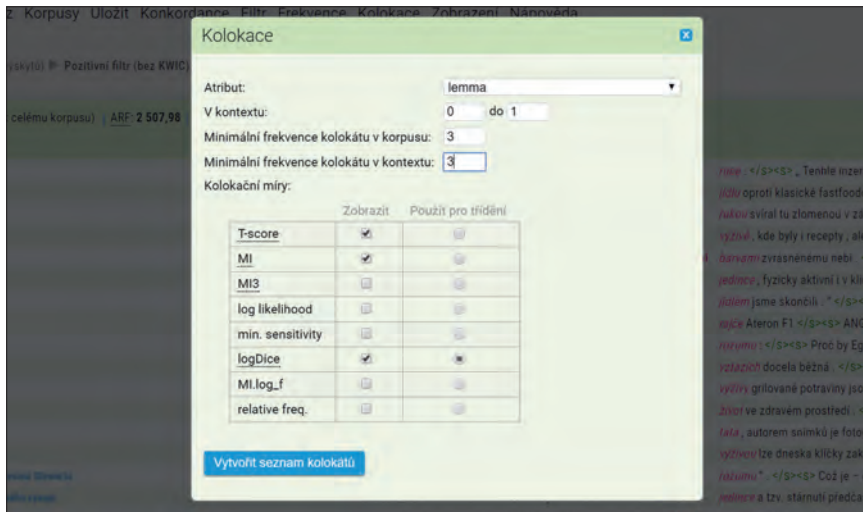
Postup 1

1. V korpusu syn2015 zvolte Typ dotazu lemma a do okénka Dotaz napište hledané adjektivum (*zdravý* nebo *zdravotní*). Dále rozbalte možnost Specifikovat kontext. Zde v části Slovní druh zadejte Velikost kontextu vpravo 1 token. Jako slovní druh zvolte podstatné jméno. Tím najdete všechny věty, ve kterých bude hned vpravo vedle daného adjektiva podstatné jméno.



Obrázek 41

- Až se vám zobrazí konkordance, vyberte možnost Kolokace, dále zvolte Vlastní. Jako Atribut si vyberte lemma, v kontextu od 0 do 1.



Obrázek 42

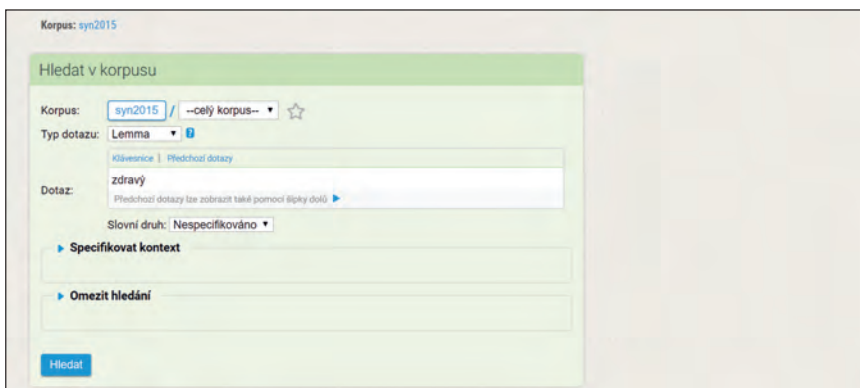
- Zobrazí se frekvenční seznam nejčastějších substantiv, která se pojí s daným adjektivem a stojí ve větě vždy hned vedle něj.

| | Filtr | lemma | Freq | MI | T-score | logDice |
|-----|-------|-------------|------|--------|---------|---------|
| 1. | p/n | rozum | 766 | 11.695 | 27.668 | 11.213 |
| 2. | p/n | výživa | 329 | 11.732 | 18.133 | 10.469 |
| 3. | p/n | strava | 123 | 9.951 | 11.079 | 8.938 |
| 4. | p/n | stravování | 42 | 10.282 | 6.476 | 7.790 |
| 5. | p/n | potravina | 75 | 7.900 | 8.624 | 7.631 |
| 6. | p/n | sebevědomí | 45 | 8.906 | 6.694 | 7.611 |
| 7. | p/n | jídlo | 106 | 7.117 | 10.221 | 7.282 |
| 8. | p/n | jedinec | 64 | 7.445 | 7.954 | 7.271 |
| 9. | p/n | životopráva | 26 | 10.817 | 5.096 | 7.201 |
| 10. | p/n | bydlení | 47 | 7.628 | 6.821 | 7.167 |

Obrázek 43

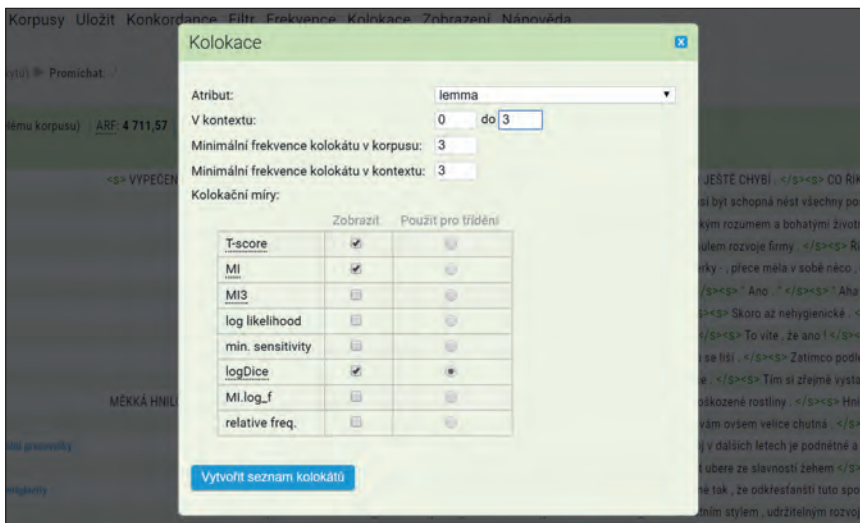
Postup 2

1. V korpusu syn2015 zvolte Typ dotazu lemma a do okénka Dotaz napište hledané adjektivum (zdravý nebo zdravotní).



Obrázek 44

2. Až se vám zobrazí konkordance, vyberte možnost Kolokace, dále zvolte Vlastní. Jako Atribut si vyberte lemma, v kontextu od 0 do 3.



Obrázek 45

3. Zobrazí se vám frekvenční seznam nejčastějších slov následujících po daném adjektivu, a to až do třetí pozice od vlastního adjektiva. Protože jste na začátku nespécifikovali slovní druh, je třeba ze seznamu kolokátů vybrat jen substantiva. Výhodou je, že najdete i kolokáty, které jsou dále od daného adjektiva (např. *zdravý životní styl*).

Korpus: *syn2015* | Dotaz: *zdravý* (10 239 výskytů) ▶ Promíchat ✓

Kolokace

1 ▶

| | Filtr | lemma | Freq | MI | T-score | logDice |
|-----|-------|------------|------|--------|---------|---------|
| 1. | p / n | rozum | 831 | 10.909 | 28.812 | 10.794 |
| 2. | p / n | vyživa | 331 | 10.837 | 18.183 | 9.776 |
| 3. | p / n | styl | 407 | 8.584 | 20.122 | 9.195 |
| 4. | p / n | životní | 438 | 8.117 | 20.853 | 8.959 |
| 5. | p / n | strava | 136 | 9.192 | 11.642 | 8.423 |
| 6. | p / n | potravina | 88 | 7.227 | 9.318 | 7.392 |
| 7. | p / n | jidlo | 127 | 6.474 | 11.143 | 7.264 |
| 8. | p / n | selský | 51 | 9.561 | 7.132 | 7.243 |
| 9. | p / n | stravování | 49 | 9.601 | 6.991 | 7.192 |
| 10. | p / n | jedinec | 79 | 6.845 | 8.811 | 7.141 |
| 11. | n / n | sebevědomí | 48 | 8.095 | 6.903 | 6.998 |

Obrázek 46

4. Pokud chcete využít pozitivní filtr, klikněte na písmeno p ve sloupci filtr. Zde je příklad pozitivního filtru u substantiva *rozum* v seznamu nejčastějších kolokátů adjektiva *zdravý*.

kon text | Dotaz: *Korpusy Uložit Konkordance Filtr Frekvence Kolokace Zobrazení Nápověda*

Korpus: *syn2015* | Dotaz: *zdravý* (10 239 výskytů) ▶ Pozitivní filtr (bez KWIC) *rozum* (831 výskytů) ▶ Promíchat ✓

Výsledek: 831 | 1 p, 0 n, 6,84 (rozděleno k celému korpusu) | 426 427,72 | Výsledek je promíchnut

Výběr řádků: *zdravý rozum* ▶

| | | | | |
|-----|-----|---|---------------------|--|
| 1. | ... | ve městě , kam nechtěl jít dobrovolně sít stádo se | <i>zdravý rozum</i> | ...prošla do novin zpráv , že měla učitel |
| 2. | ... | Hán bylo narušeno na cestě , ...Věštec , že soudečtvo | <i>zdravý rozum</i> | ...a měl rukou vyprávě , ...Jiná mladá a sná |
| 3. | ... | griše ve Vietnamu o v Želeh , ...Kolektivní psychické je | <i>zdravý rozum</i> | ...A Lukáš Kobelov ho sudský moc manažerů |
| 4. | ... | domen v Uh vždy byly a Rychlé rychlosti páni než | <i>zdravý rozum</i> | ...Josef Rychtář o Zelenka Macarovi (, |
| 5. | ... | ... Roman - kteří se uplatí spíše k povírání než ke | <i>zdravý rozum</i> | ...Frundbergio sebetní záchvat přitahá jako zrušení toho , |
| 6. | ... | ... nebo ještě hů , jako by pochýboval o mím | <i>zdravý rozum</i> | ... když j mohu říkat takové věci , ...Poslední |
| 7. | ... | ... záhy vzad letou směřovat nahoru ? ... To mu říká pouhy | <i>zdravý rozum</i> | ... Výzkvě komodo nekurmíje do , ... výky , ... Naneštěstí však soudce |
| 8. | ... | ... stěví vědět jak , ... A navíc , jako se špatně | <i>zdravý rozum</i> | ... by vědět napadlo pokoušet se vstoupit do domu , |
| 9. | ... | ... Max pěstí spány , ... Ažkol se stá odlišit vryky | <i>zdravý rozum</i> | ... andra a o pravci , ... Mladobobal , ... Mladá |
| 10. | ... | ... kvasí rychlob nepropagjí slonk propogene a zábravě baromou Evropu , ... Zdravý | <i>zdravý rozum</i> | ... Aa paděnkřibný hrouk odpórů EKOP a vetpu do Rverende |
| 11. | ... | ... Domovl jsem se , že aleson jeden propví vic | <i>zdravý rozum</i> | ... Ale jich bych sám nepoužil , |

Obrázek 47

Řešení

1. Seznam deseti nejčastějších kolokátů

zdravý

| <u>první postup</u> | <u>druhý postup</u> |
|---------------------|---------------------|
| rozum | rozum |
| výživa | výživa |
| strava | styl |
| stravování | strava |
| potravina | potravina |
| sebevědomí | stravování |
| jídlo | jídlo |
| jedinec | jedinec |
| životospráva | sebevědomí |
| bydlení | bydlení |

zdravotní

| <u>první postup</u> | <u>druhý postup</u> |
|---------------------|---------------------|
| pojišťovna | pojišťovna |
| postižení | postižení |
| pojištění | pojištění |
| stav | péče |
| péče | stav |
| sestra | sestra |
| potíž | potíž |
| problém | problém |
| komplikace | komplikace |
| služba | služba |

Vysvětlením významu může být např. zjednodušená definice ze Slovníku spisovného jazyka českého (<http://ssjc.ujc.cas.cz/>):

zdravý = má bezproblémové zdraví (zdravý jedinec, člověk); přirozený (zdravý rozum, zdravé sebevědomí); dobrý pro zdraví (zdravé jídlo, zdravé bydlení)

zdravotní = týká se zdraví (zdravotní problém, stav); stará se o zdraví (zdravotní sestra, pojišťovna)

2. Spojení adjektiva a substantiva

zdravá holčička, zdravá noha, zdravý zub
zdravotní handicap, zdravotní obtíž, zdravotní prohlídka

Pracovní list č. 2:

ADJEKTIVA ZDRAVÝ, ZDRAVOTNÍ A ZDRAVOTNICKÝ – Kdo vede zdravý životní styl, nemá zdravotní problémy a nepotřebuje zdravotnické služby.

Anotace

Cvičení tohoto typu je vhodné zařadit do výuky, pokud dělá žákům problém rozlišovat význam podobných adjektiv a užívat je ve správných spojeních. Může jít jak o adjektiva, která mají jiný význam – zde např. *zdravý*, *zdravotní* a *zdravotnický*, tak o adjektiva, která mají v některých kolokacích význam stejný, ale existují různá omezení – např. *šedý* a *šedivý*, kdy se některá substantiva mohou pojít jen s jedním adjektivem (*šedá eminence*). Cvičení také rozšiřují slovní zásobu.

Pracovní list obsahuje cvičení založená na jazykovém materiálu z korpusu. Učitel si cvičení připraví předem, žáci ve výuce s korpusem nepracují. Na základě autentického jazykového materiálu si ovšem sami uvědomí, s jakými substantivy se daná adjektiva mohou pojít.

Postup práce

Před rozdělením pracovního listu je nutné žáky připravit na téma úkolů a vhodně je motivovat k jejich vypracování. V tomto případě může učitel s žáky mluvit o tom, co sami dělají pro své zdraví a jaké existují zdravotní problémy. K diskusi může využít předem připravený text, který bude obsahovat tematizovaná adjektiva.

Po tomto úvodu učitel obrátí pozornost žáků k adjektivům *zdravý*, *zdravotní* a *zdravotnický* a zeptá se jich, jaký je v nich rozdíl a která substantiva se s nimi mohou pojít. Jako motivační prvek dobře slouží návodné obrázky (zdravé jídlo nebo tělesné aktivity x různá onemocnění, nemocniční personál). Odpovědi žáků lze zapisovat na tabuli formou myšlenkové mapy. Následně učitel rozdá žákům pracovní list. Ti pak splní první úkol nejlépe ve skupinách. Žáci si v úkolu mají sami všimnout typického okolí daných adjektiv a zformulovat pravidla jejich užití. Až budou žáci hotoví, porovnájí v plénu řešení jednotlivých skupin.

Důležité je upozornit žáky také na gramatický rozdíl mezi měkkým adjektivem *zdravotní* a tvrdými adjektivy *zdravý* a *zdravotnický*.

Druhý úkol žáci plní až po kontrole prvního. Ke každé skupině vět nejprve žáci přiřadí správné adjektivum. Až poté ho doplní do vět ve správném tvaru. Opět mohou spolupracovat ve skupinách.

KDO VEDE ZDRAVÝ ŽIVOTNÍ STYL, NEMÁ ZDRAVOTNÍ PROBLÉMY A NEPOTŘEBUJE ZDRAVOTNICKÉ SLUŽBY.

1. Přečtete si texty a řekněte, kdy používáme *zdravý*, *zdravotní* a *zdravotnický*.

Co ráda vaříte, preferujete **zdravou** výživu?

K vysněné váze vás dovede jedině **zdravý** životní styl.

Zdravá strava by měla začínat už u prvního jídla dne.

Prioritou a hlavním tématem pro letošní rok je zdraví a vše, co s ním souvisí –

zdravá půda, **zdravé** potraviny, zdraví člověka i jeho okolí.

Pokud máte **zdravotní** potíže, vhodné lázně navrhne váš lékař a na platbě se podílí i **zdravotní** pojišťovna.

Čeští sportovci se **zdravotním** postižením zažívají úspěchy i na mezinárodním poli.

Ze **zdravotního** pojištění se hradí **zdravotní** péče poskytnutá pojištěnci s cílem zachovat nebo zlepšit jeho **zdravotní** stav.

Pacient má svobodnou volbu nejenom lékaře, ale i **zdravotnického** zařízení.

Mezi důležité součásti **zdravotnické** dokumentace patří: anamnéza, lékařská zpráva, propouštěcí zpráva, písemný souhlas s poskytováním informací pacientovi a lékařský posudek.

Zdravotníci pracovníci jsou základem k poskytování **zdravotnických** služeb ve všech evropských zemích.

2. Doplňte adjektiva *zdravý*, *zdravotní* a *zdravotnický* ve správném tvaru. V každém okně je stejné adjektivum.

V Praze se uskuteční první mezinárodní veletrh pro lékaře a _____ personál s účastí zástupců německých a rakouských _____ zařízení.

Kvalitní _____ pracovníci jsou základem k poskytování kvalitních

_____ služeb ve všech evropských zemích.

Světová _____ organizace oznámila v roce 2008, že nedostatek kvalifikovaných lékařů v Africe začne do roku 2015 ohrožovat budoucnost tohoto kontinentu.

Národní číslo tísňového volání _____ záchranné služby je 155.

Humanitární organizace Samari pošle _____ materiál do nemocnice na Ukrajině, kterou už několik let pomáhají dobrovolníci Samari opravit.

Sára vaří _____ jídlo.

Důležitou součástí _____ stravování vašich dětí je i pitný režim.

Na počátku onemocnění diabetem mívá nemocný člověk v krvi víc inzulínu, než je běžné u dospělého _____ jedince.

Kvalitní osvětlení je základem _____ bydlení.

Asi nikdo nemá v dnešní době úplně _____ kůži.

Občané mají na základě veřejného pojištění právo na bezplatnou _____ péči.

Na palubě vždy cestuje lékař a _____ sestra.

Velké snížení váhy vede i ke _____ potížím (vypadávání vlasů, ztráta menstruace...).

„V souvislosti s používáním klimatizace roste riziko _____ problémů,“ varuje praktická lékařka MUDr. Gabriela Navrátilová.

Důvodem _____ komplikací lidí bylo hlavně horké počasí.

Postup práce s korpusem

1. V korpusu syn2015 zvolte Typ dotazu lemma a do okénka Dotaz napište hledané adjektivum (*zdravý, zdravotní* nebo *zdravotnický*). Dále rozbalte možnost Specifikovat kontext. Zde v části Slovní druh zadejte Velikost kontextu vpravo 1 token. Jako slovní druh zvolte podstatné jméno. Tím najdete všechny věty, ve kterých bude hned vpravo vedle daného adjektiva podstatné jméno. Kontext nutně specifikovat nemusíte. Pokud to neuděláte, budete ovšem muset ve druhém kroku, tedy hledání typických kolokací, vybrat pouze substantiva (viz postup práce u pracovního listu č. 1).
2. Důležité je vybírat příkladové věty podle frekvence. Proto až se vám zobrazí konkordance, vyberte možnost Kolokace, dále zvolte Vlastní. Jako Atribut si vyberte lemma, v kontextu od 0 do 3. Zobrazí se vám frekvenční seznam nejčastějších slov následujících po daném adjektivu, a to až do třetí pozice od vlastního adjektiva. Pokud jste na začátku nespecifikovali slovní druh, je třeba ze seznamu kolokátů vybrat jen substantiva. Výhodou je, že najdete i kolokáty, které jsou dále od daného adjektiva (např. *zdravý životní styl*).
3. Vyberte si deset nejčastějších kolokátů a zobrazte si příkladové věty kliknutím na pozitivní filtr (p/n) u vybraného substantiva v seznamu kolokátů. Poté se vám objeví všechny věty s daným adjektivem a substantivem.
4. Věty pocházejí z českých textů, je třeba proto počítat s tím, že mohou být pro žáky na nižší jazykové úrovni nesrozumitelné a naopak matoucí. Proto je často třeba věty zjednodušit.
5. Pět příkladových vět ke každému adjektivu použijte pro první cvičení, dalších pět pro druhé s vynechaným adjektivem.

Řešení

1. Řešením mohou být např. zjednodušené definice ze Slovníku spisovného jazyka českého (<http://sjc.ujc.cas.cz/>):
zdravý = má bezproblémové zdraví (zdravý jedinec, člověk); přirozený (zdravý rozum, zdravé sebevědomí); dobrý pro zdraví (zdravé jídlo, zdravé bydlení)
zdravotní = týká se zdraví (zdravotní problém, stav); stará se o zdraví (zdravotní sestra, pojišťovna)
zdravotnický = týká se zdravotnictví: (z. pracovník; z-é orgány; z-á služba; z-á komise; z-é zařízení)
2. zdravotnický - zdravý - zdravotní

Pracovní list č. 3:

PROCVIČENÍ AKUZATIVU - Koupila jsem si čokoládu. Mám radost!

Anotace

Pracovní list obsahuje cvičení založená na jazykovém materiálu z korpusu. Učitel si cvičení připraví předem, žáci ve výuce s korpusem nepracují. Na základě autentického jazykového materiálu si ovšem upevní pravidla pro užití akuzativních koncovek substantiv a adjektiv. Cvičení lze samozřejmě upravit pro další pády.

Postup práce

Cvičení neslouží pro první seznámení s akuzativem. Žáci by s tímto pádem měli být seznámeni v komunikačním kontextu. Vhodná jsou např. témata objednávání v restauraci (*dám si + akuzativ*) nebo představování rodiny (*mám + akuzativ*). Cvičení v tomto pracovním listu jsou zaměřena jen na koncovky a neobsahují jednotlicí komunikační téma. Žáci by si na jejich základě měli upevnit probranou látku aktivním způsobem. Je možné, že skloňování podle některých vzorů (např. *předseda* a *soudce*) pro ně bude nové.

Před řešením úkolů pracovního listu je nutné žáky uvést do tématu. Učitel může žáky upozornit na úvodní věty, zeptat se, jaký je v nich pád, a sbírat další příklady vět s akuzativem. Příkladové věty může shromážďovat na tabuli po skupinách podle rodů. Je vhodné jednotlivé rody barevně odlišit. Pozornost žáků následně obrátí k typickým zakončením. Žáci sami nejspíš nevyjmenují substantiva skloňovaná podle všech vzorů. Učitel jim proto rozdá pracovní list. První větu v prvním cvičení s nimi přečte v plénu a nechá si nadiktovat adjektivum a substantivum v nominativu. Dále žáci postupují samostatně nebo ve skupinách. Vhodné je zkontrolovat řešení hned po femininu, aby si žáci uvědomili rozdíl mezi tvrdým a měkkým adjektivem.

Druhý úkol žáci plní až po kontrole prvního. Opět je vhodné vyřešit první příklad dohromady.

Po splnění druhého úkolu se učitel žáků zeptá, která substantiva a adjektiva se v akuzativu nemění. Pro žáky je motivační, když vědí, že si nemusí zapamatovat takové množství koncovek.

Na tuto fázi formulace pravidel by měla navazovat procvičovací fáze, ve které žáci využijí své znalosti. Materiál k procvičení lze opět najít v korpusu.

KOUPILA JSEM SI ČOKOLÁDU. MÁM RADOST!

1. V jakém pádě jsou podstatná jména v obou větách?
2. Napište první pád tučných přídavných a podstatných jmen.

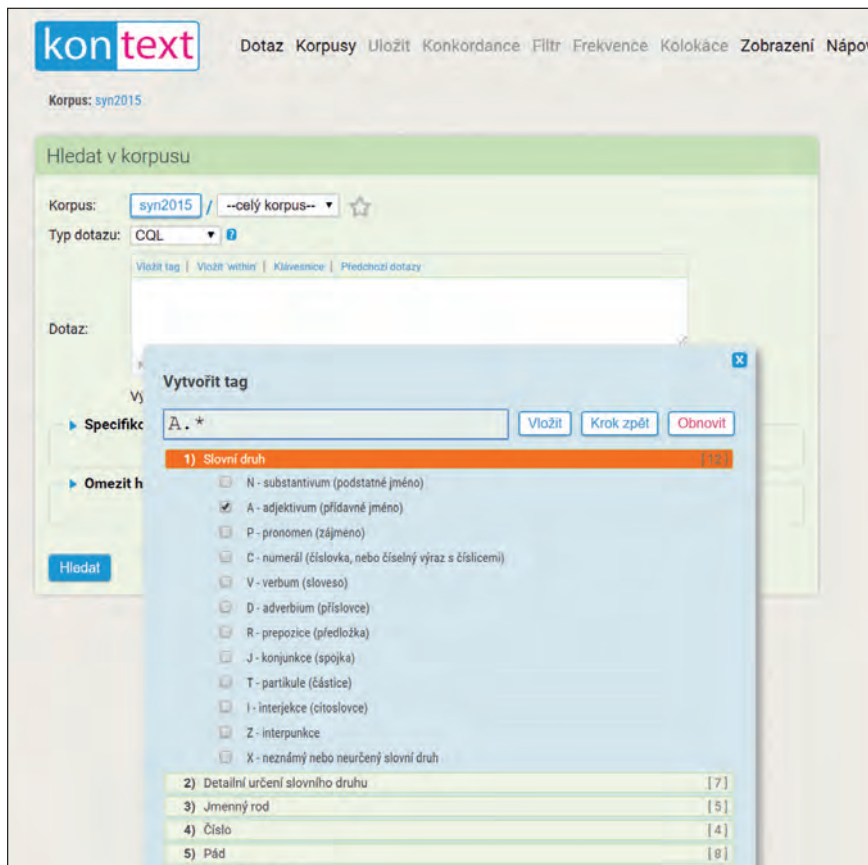
| akuzativ (čtvrtý pád) - ženský rod: | nominativ (první pád) |
|---|------------------------------|
| Koupím si celou láhev . | celá lahev |
| Dělá mi to velkou radost . | |
| Basketbalisté Nymburka mají další výhru : Pohár České pošty. | |
| Dnes máte poslední šanci koupit si levnější vstupenky. | |
| akuzativ (čtvrtý pád) - mužský rod životný: | |
| Mám nového přítele . Už nemám ráda svého muže . | |
| Máme ještě mladšího syna . | |
| Máme v práci staršího kolegu . | |
| Každý král potřebuje dobrého rádce . | |
| akuzativ (čtvrtý pád) - mužský rod neživotný: | |
| Byla jsem celý den doma. | |
| Příští měsíc pojedeme do Argentiny. | |
| akuzativ (čtvrtý pád) - střední rod: | |
| Mám rád dobré jídlo . | |
| Do školy potřebuju čisté oblečení . | |
| Jaké granule jsou nejlepší pro malé štěně ? | |
| Zítرا máme další rande . | |

3. Doplňte tabulku. První řešení je příklad.

| | první pád - koncovky | čtvrtý pád - koncovky |
|-----------------------------|--|------------------------------------|
| ženský rod | | |
| celá láhev | -á (celá), -konsonant (láhev) | -ou (celou) --- (láhev) |
| velká radost | -á (velká), -konsonant (radost) | ___ (___) ___ (___) |
| další výhra | -í (další), -a (výhra) | ___ (___) ___ (___) |
| poslední šance | -í (poslední), -e (šance) | ___ (___) ___ (___) |
| mužský rod životný | | |
| nový přítel / muž | -ý (nový), -tel / měkký konsonant (přítel/muž) | ___ (___) ___ (___) / ___ (___) |
| mladší syn | -í (mladší), -tvrdý konsonant (syn) | ___ (___) ___ (___) |
| starší kolega | -í (starší), -a (kolega) | ___ (___) ___ (___) |
| dobrý rádce | -ý, (dobrý), -e (rádce) | ___ (___) ___ (___) |
| mužský rod neživotný | | |
| celý den | -ý (celý), -konsonant (den) | ___ (___) ___ (___) |
| příští měsíc | -í (příští), -konsonant (měsíc) | ___ (___) ___ (___) |
| střední rod | | |
| dobré jídlo | -é (dobré), -vokál (jídlo) | ___ (___) ___ (___) |
| čisté oblečení | -é (čisté), -vokál (oblečení) | ___ (___) ___ (___) |
| malé štěně | -é (malé), -vokál (štěně) | ___ (___) ___ (___) |
| další rande | -í (další), -vokál (rande) | ___ (___) ___ (___) |

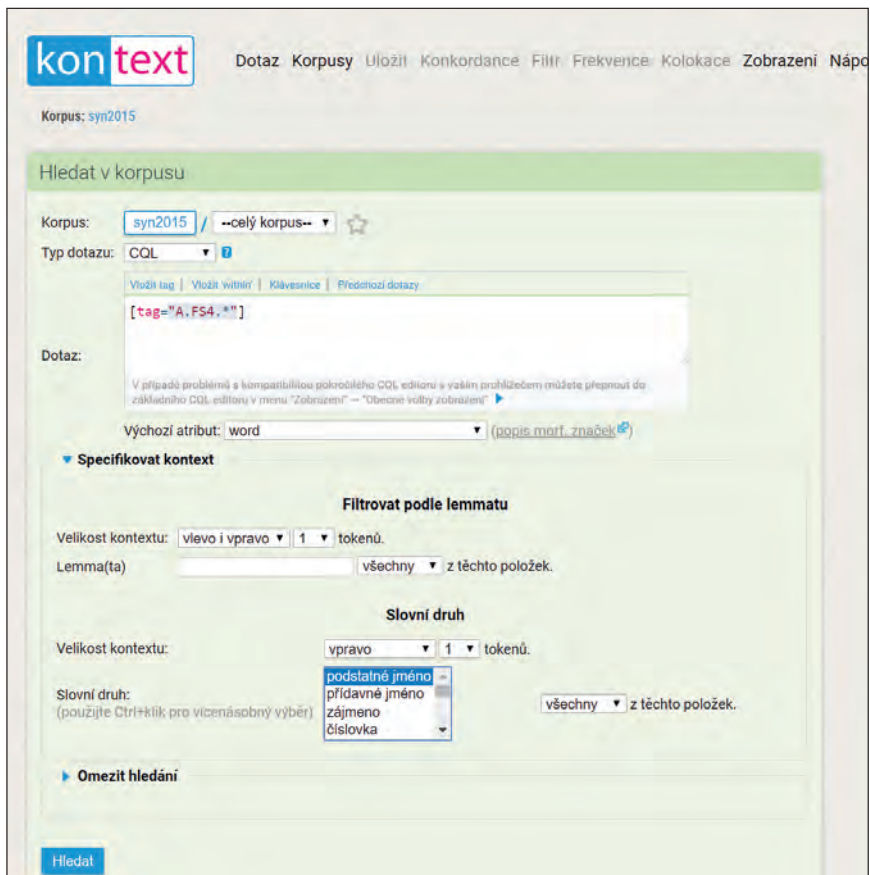
Postup práce s korpusem

1. V rozhraní KonText zvolte korpus syn2015. Jako Typ dotazu zvolte CQL a klikněte na Vložit tag. Jako slovní druh vyberte adjektivum (hledáme věty, ve kterých se budou vyskytovat adjektiva), dále příslušný jmenný rod (femininum, maskulinum animatum, maskulinum inanimatum, neutrum), číslo singular, pád akuzativ. Klikněte na Vložit.



Obrázek 48

2. Výsledný tag se vloží do pole Dotaz. Jako Výchozí atribut ponechte word. Můžete specifikovat kontext tak, aby se přímo za adjektivem objevilo podstatné jméno (viz obrázek). Není to ovšem nutné, vět bez substantiv se i bez specifikace zobrazí jen málo.



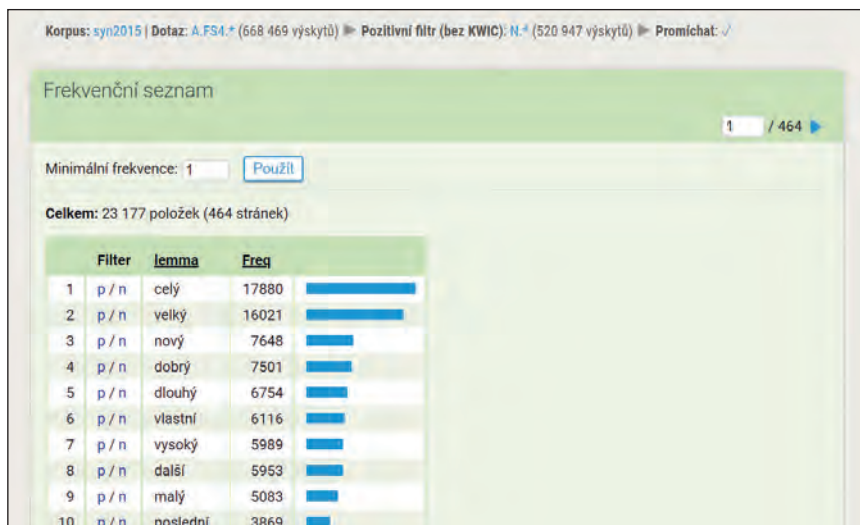
Obrázek 49

3. Dále klikněte na Hledat. Zobrazí se vám konkordance.



Obrázek 50

- Po kliknutí na tučně vyznačené adjektivum se vám zobrazí širší kontext. Z vět je třeba vybrat takové, které budou odpovídat úrovni žáků, případně je třeba věty přizpůsobit. Vhodné je zobrazit si nejprve nejfrekventovanější adjektiva pomocí funkce Frekvence. Po kliknutí na Frekvence a zvolení atributu Lemmata se zobrazí frekvenční seznam.



Obrázek 51

- Příkladové věty zobrazíte funkcí pozitivní filtr (p). Z nich vyberte vyhovující tak, aby byla zastoupena měkká i tvrdá adjektiva i všechny typy skloňování. Věty zjednodušte podle úrovně žáků.

Pracovní list č. 4:

FRAZÉMY TYPU MÍT + PŘÍČESTÍ TRPNÉ - Máte hotovo?

Anotace

Pracovní list obsahuje cvičení založená na jazykovém materiálu z korpusu. Učitel si cvičení připraví předem, žáci ve výuce s korpusem nepracují. Na základě autentického jazykového materiálu se seznámí s frázemi typu *mít + příčestí trpné*. Cvičení lze samozřejmě upravit pro další frazémy.

Postup práce

Cvičení jsou věnována frazémům obsahujícím příčestí trpné. Je proto vhodné, aby byli žáci již dříve seznámeni s pasivem a s konstrukcí *mít + příčestí trpné*, která typicky vyjadřuje výsledný stav (viz název listu).

V prvním úkolu žáci přiřazují k frazémům jejich stručnou významovou definici. Vhodné je, aby pracovali ve dvojicích. Kontrola pak může proběhnout nejprve ve skupinách, poté v plénu.

Druhý úkol žáci plní až po kontrole prvního. Je vhodné vyřešit první příklad dohromady.

Po splnění druhého úkolu mohou žáci vymyslet podobné cvičení. To může být i domácí úkol – žáci mohou příkladové věty najít v korpusu. Zde by ovšem bylo dobré, aby jim učitel pomohl se zadáním (viz část *Postup práce s korpusem*) nebo aby vyhledávali jednoduše přímo pomocí fráze v daném tvaru (např. *má vyhráno*).

MÁTE HOTOVO?

1. Spojte synonyma.

| | |
|-----------------------------------|--|
| 1. mít upито | A. směřovat někam, např. jet někam |
| 2. mít namířeno někam | B. vidět směr své budoucnosti |
| 3. mít nakročeno někam | C. zaměřovat se na něco |
| 4. mít zaostřeno na něco | D. ještě bojovat |
| 5. nemít vyhráno | E. být zabezpečený, zajištěný |
| 6. mít vystaráno | F. mít strach |
| 7. mít zaděláno na problém/úspěch | G. mít v sobě alkohol |
| 8. mít nahnáno z někoho/něčeho | H. být v situaci, která způsobí problém/úspěch |

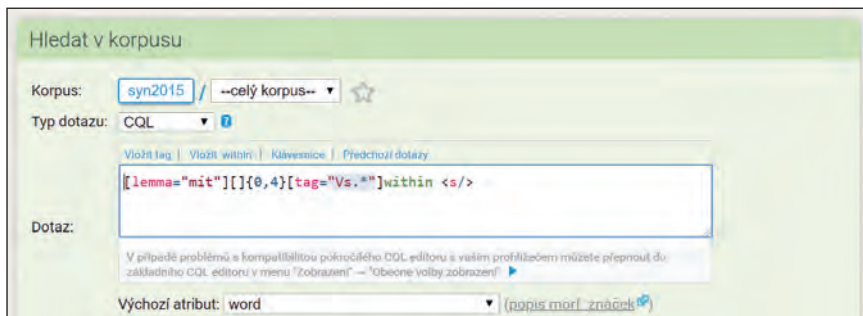
2. Výrazy z levého sloupce tabulky doplňte do vět ve správném tvaru.

1. Muž byl při autonehodě těžce zraněn a nyní leží v kritickém stavu v nemocnici a stále ještě _____ .
2. Neukázněného řidiče, který usedl za volant, ačkoliv _____, zadrželi v úterý po půlnoci policisté v Rožnově pod Radhoštěm.
3. Novináři _____ na politické dění na Slovensku.
4. Recidivista z Hlučína _____ na rekordní již dvaadvacátý pobyt ve vězení. Policisté zjistili, že tentokrát vykradl rodinný dům v Ostravě.
5. Ten muž se vůbec nebojí, ne _____ z policie _____ .
6. V branickém psím útulku se radují. Firma z Poděbrad jim věnovala jeden a půl tuny granulí. „Přinejmenším do konce roku _____,“ odhaduje ošetřovatel Zdeněk Blažek.
7. V rychlíku do Chebu jsem potkala bývalou spolužačku a ptala jsem se jí, kam _____.
8. DiCaprio je pořád tahoun, film s jeho jménem _____ na úspěch.

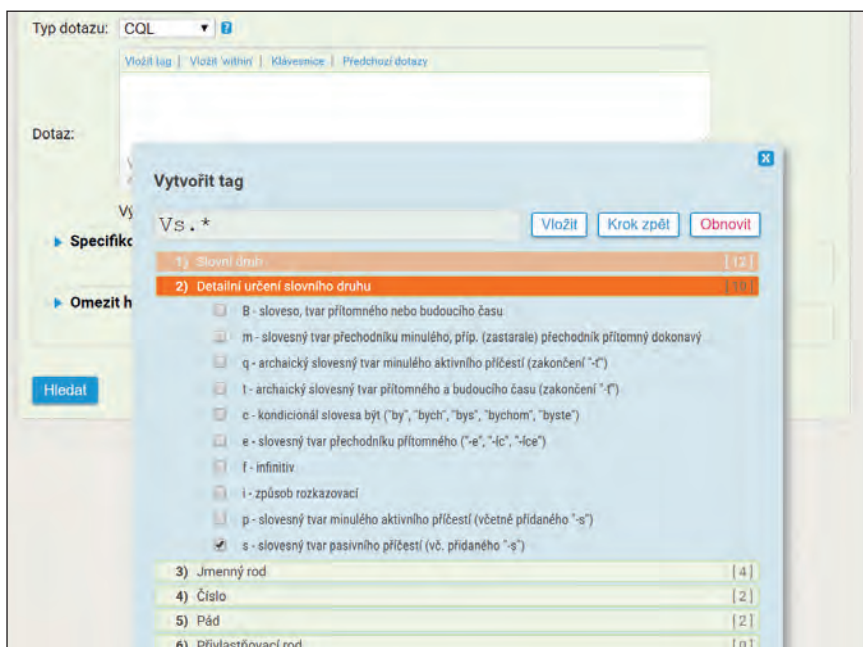
3. Vyberte si pět frází a vymyslete pro spolužáky podobné cvičení.

Postup práce s korpusem

1. V rozhraní KonText si vyberte korpus syn2015 a Typ dotazu CQL. Budete hledat všechny věty se slovesem mít a pasivním přičestím. Kompletní dotaz vidíte na obrázku. Hledáte sloveso mít ve všech tvarech [lemma=„mít“], dále může být ve větě několik libovolných slov – např. 0-4 [] {0,4}, dále pasivní přičestí [tag=„Vs.*“] – tag si můžete vytvořit pomocí tlačítka Vytvořit tag (viz druhý obrázek). Nakonec potřebujete zajistit, aby vše bylo v rámci jedné věty (within <s/>).



Obrázek 52



Obrázek 53

2. Klikněte na Hledat a zobrazí se vám konkordance:

Korpus: **syn2015** | Dotaz: **mít, Vs.+** (15 788 výskytů) ▶ Promíchat

Výsledky: 15 788 (1-16) 100,7% (včetně 4 odlišných tvarů) AŽ: 4 432,00 Výsledky je promíchat

Výběr funkcí: **základní**

| | | | | |
|-------------|---------------------------|--|-------------------------|--|
| 1. mít / n | Mítel a celkové hodnocení | ...mítel jsem dostal. Je to asi čtyři hodiny | mít dokončeno odloženou | ...mít, co mám je. Budeme si muset a |
| 2. mít / n | ...mítel jsem dostal | ...mítel jsem dostal. Je to asi čtyři hodiny | mít být vysvětl | ...mítel jsem dostal. Je to asi čtyři hodiny |
| 3. mít / n | ...mítel jsem dostal | ...mítel jsem dostal. Je to asi čtyři hodiny | mít být vysvětl | ...mítel jsem dostal. Je to asi čtyři hodiny |
| 4. mít / n | ...mítel jsem dostal | ...mítel jsem dostal. Je to asi čtyři hodiny | mít být vysvětl | ...mítel jsem dostal. Je to asi čtyři hodiny |
| 5. mít / n | ...mítel jsem dostal | ...mítel jsem dostal. Je to asi čtyři hodiny | mít být vysvětl | ...mítel jsem dostal. Je to asi čtyři hodiny |
| 6. mít / n | ...mítel jsem dostal | ...mítel jsem dostal. Je to asi čtyři hodiny | mít být vysvětl | ...mítel jsem dostal. Je to asi čtyři hodiny |
| 7. mít / n | ...mítel jsem dostal | ...mítel jsem dostal. Je to asi čtyři hodiny | mít být vysvětl | ...mítel jsem dostal. Je to asi čtyři hodiny |
| 8. mít / n | ...mítel jsem dostal | ...mítel jsem dostal. Je to asi čtyři hodiny | mít být vysvětl | ...mítel jsem dostal. Je to asi čtyři hodiny |
| 9. mít / n | ...mítel jsem dostal | ...mítel jsem dostal. Je to asi čtyři hodiny | mít být vysvětl | ...mítel jsem dostal. Je to asi čtyři hodiny |
| 10. mít / n | ...mítel jsem dostal | ...mítel jsem dostal. Je to asi čtyři hodiny | mít být vysvětl | ...mítel jsem dostal. Je to asi čtyři hodiny |
| 11. mít / n | ...mítel jsem dostal | ...mítel jsem dostal. Je to asi čtyři hodiny | mít být vysvětl | ...mítel jsem dostal. Je to asi čtyři hodiny |
| 12. mít / n | ...mítel jsem dostal | ...mítel jsem dostal. Je to asi čtyři hodiny | mít být vysvětl | ...mítel jsem dostal. Je to asi čtyři hodiny |
| 13. mít / n | ...mítel jsem dostal | ...mítel jsem dostal. Je to asi čtyři hodiny | mít být vysvětl | ...mítel jsem dostal. Je to asi čtyři hodiny |
| 14. mít / n | ...mítel jsem dostal | ...mítel jsem dostal. Je to asi čtyři hodiny | mít být vysvětl | ...mítel jsem dostal. Je to asi čtyři hodiny |
| 15. mít / n | ...mítel jsem dostal | ...mítel jsem dostal. Je to asi čtyři hodiny | mít být vysvětl | ...mítel jsem dostal. Je to asi čtyři hodiny |
| 16. mít / n | ...mítel jsem dostal | ...mítel jsem dostal. Je to asi čtyři hodiny | mít být vysvětl | ...mítel jsem dostal. Je to asi čtyři hodiny |

Obrázek 54

3. Pomocí funkce Frekvence (s atributem slovní tvar) se podívejte na nejčastější spojení:

Korpus: **syn2015** | Dotaz: **mít, Vs.+** (15 788 výskytů) ▶ Promíchat

Frekvenční seznam

Minimální frekvence: 1

Celkem: 11 698 položek (234 stránek)

| Filter | word | Freq | |
|--------|-------|--------------------|----|
| 1 | p / n | má namířeno | 81 |
| 2 | p / n | měla být dokončena | 61 |
| 3 | p / n | měl namířeno | 56 |
| 4 | p / n | máte vyhráno | 42 |
| 5 | p / n | má otevřeno | 39 |
| 6 | p / n | měli namířeno | 38 |
| 7 | p / n | měl být dokončen | 33 |
| 8 | p / n | máte namířeno | 31 |
| 9 | p / n | mít otevřeno | 29 |
| 10 | p / n | má být dosaženo | 26 |
| 11 | p / n | máš namířeno | 26 |
| 12 | p / n | měla být provedena | 26 |
| 13 | p / n | má vyhráno | 25 |
| 14 | p / n | má našlápnuto | 25 |
| 15 | p / n | má být dokončena | 25 |
| 16 | p / n | má nakročeno | 23 |

Obrázek 55

4. Uvidíte, že frazémy jsou velmi frekventované. Pomocí pozitivního filtru (p) si zobrazte požadované věty. Na obrázku vidíte příklad pro *má namířeno*.



Obrázek 56

5. Kliknutím na označenou frázi zobrazíte širší kontext. Nakonec vyberte a upravte vhodné věty pro cvičení.

Řešení

- 1G, 2A, 3B, 4C, 5D, 6E, 7H, 8F
- 1 nemá vyhráno, 2 měl upito, 3 mají zaostřeno, 4 měl nakročeno, 5 nemá nahnáno, 6 máme vysteráno, 7 má namířeno, 8 má zaděláno

Pracovní list č. 5:
IDIOMY – Já snad vyletím z kůže!

Anotace

Cvičení tohoto typu rozvíjejí schopnost rozpoznat a vhodně použít idiomy. Hodí se pro pokročilé žáky, kteří už většinou nemají potíže se základním významem slov. Žáci ve cvičení z tohoto pracovního listu aktivně pomocí korpusu dohledávají chybějící část idiomu, pomocí kontextu by pak měli být schopni vysvětlit jeho význam.

Postup práce

Před samotnou prací s korpusem musejí být žáci seznámeni s existencí idiomů, nejlépe pomocí textu obsahujícího dostatečně názorné idiomy. Pro seznámení lze využít i kvíz v prvním cvičení. Učitel se následně zeptá, zda žáci znají některé další idiomy se slovem *kůže*. Kromě těch, které se objevují v tomto pracovním listu, to mohou být například spojení *kůže líná* nebo *zachránit si kůži*.

Následně žáci samostatně nebo ve skupinách doplní znění idiomu, vypíšou z korpusu jeden příklad a pokusí se vysvětlit význam. Kontrola proběhne v plénu, žáci si případně ještě předtím mohou porovnat řešení. Druhé cvičení mohou žáci také dostat za domácí úkol. Ve třetím cvičení si poté prakticky vyzkouší, zda idiomům správe porozuměli.

JÁ SNAD VYLETÍM Z KŮŽE!

1. Kdy říkáme, že snad vyletíme z kůže?

- a. Když nás svědí celé tělo.
- b. Když nás něco hodně rozčílí.
- c. Když chceme odněkud rychle utéct.

2. Najděte v korpusu syn2015 chybějící části idiomů. Napište příklad a vysvětlete podle kontextu jejich význam.

1) _____ *se zdravou kůží*

příklad:

význam:

2) _____ *v jeho/její kůži*

příklad:

význam:

3) _____ *ve své kůži*

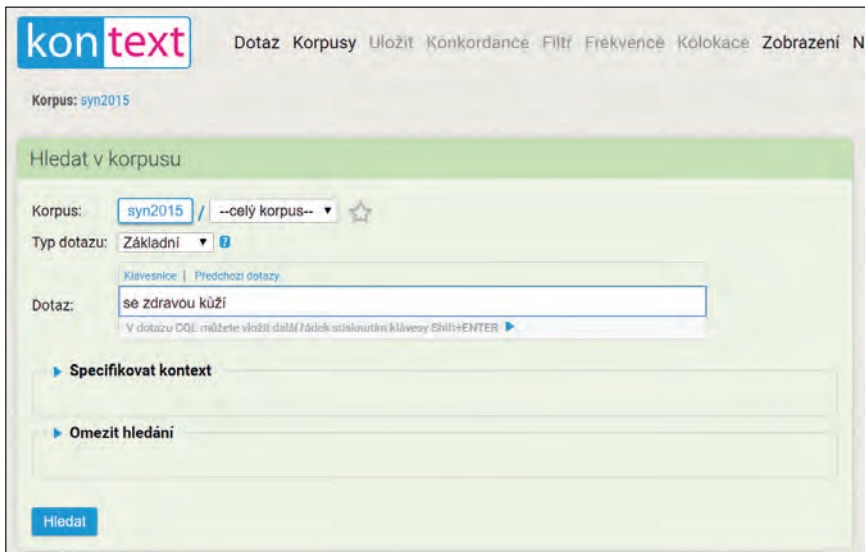
příklad:

význam:

3. Zahrajte se spolužákem dialog, ve kterém jeden z idiomů použijete.

Postup práce s korpusem

1. V rozhraní KonText zadejte postupně do políčka Dotaz vybrané části idiomů. Typ dotazu je základní.



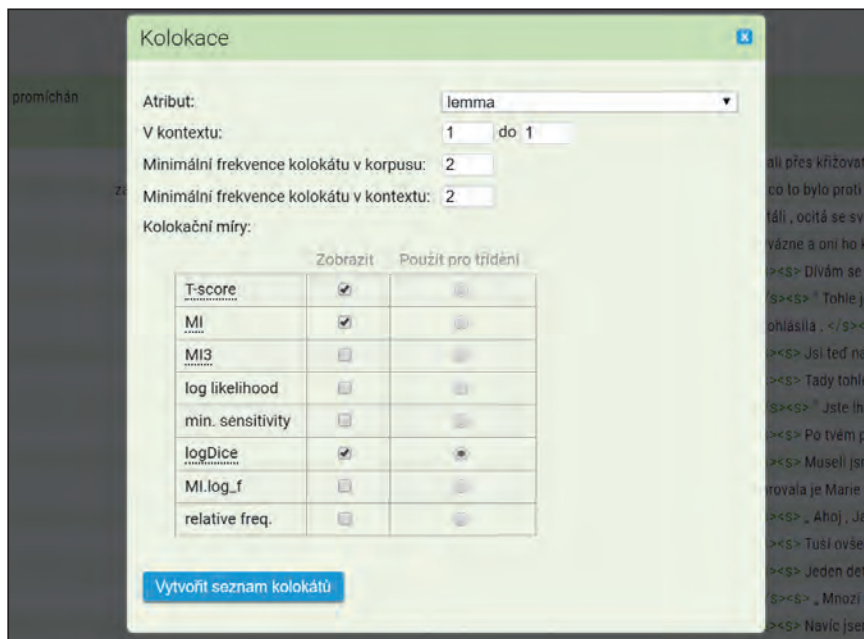
Obrázek 57

2. Po zobrazení konkordancí klikněte na zvýrazněnou část idiomů pro zobrazení širšího kontextu.



Obrázek 58

3. Pokud si chcete zobrazit kolokace, zadejte Atribut lemma, v kontextu od 1 do 1.



Obrázek 59

Korpus: syn2015 | Dotaz: se, zdravou, kůží (51 výskytů) ▶ Promíchat: ✓

Kolokace

1 ▶

| | Filtr | lemma | Freq | MI | T-score | logDice |
|----|-------|-----------|------|--------|---------|---------|
| 1. | p / n | vyváznout | 2 | 12.663 | 1.414 | 6.391 |
| 2. | p / n | . | 31 | 3.377 | 5.032 | -2.798 |
| 3. | p / n | , | 8 | 1.237 | 1.628 | -4.939 |
| 4. | p / n | a | 2 | 0.597 | 0.479 | -5.578 |

Obrázek 60

Řešení

1. b

V konkordancích se budou nejčastěji objevovat idiomy *ne/vyváznout / ne/dostat se z něčeho se zdravou kůží, být v jeho/její kůži, necítit se / nebýt ve své kůži.*

Pracovní list č. 6:

OPRAVA CHYB – Věty se slovesem jít

Anotace

Pracovní list obsahuje cvičení založené na jazykovém materiálu z korpusu. Učitel si cvičení připraví předem, žáci ve výuce s korpusem nepracují. Jejich úkolem je opravit chyby související se slovesem *jít*. Texty pocházejí od nerozličných mluvčích češtiny, učitel je nalezne v korpusu CzeSL-SGT.

Postup práce

Toto cvičení je vhodné zařadit jako shrnující aktivitu, pokud mají žáci problém s užitím slovesa *jít* a jeho prefigovaných derivátů. Je vhodné vybírat do něj věty s chybami typickými pro danou skupinu. Učitel tak může například hledat pouze texty rodilých mluvčích jednoho jazyka odpovídajícího jazyku jeho žáků.

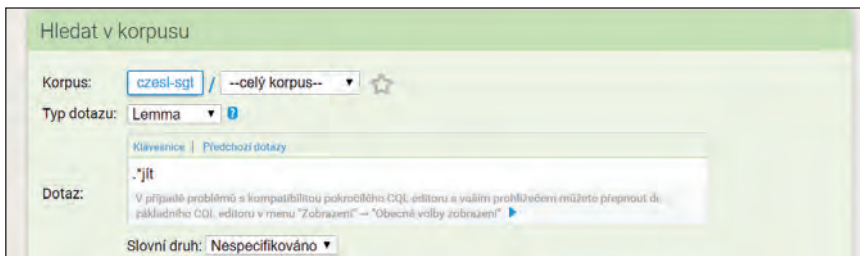
Vhodnou motivací pro úkoly, ve kterých žáci opravují chyby, může být instrukce, že jsou nyní sami učitelé. Učitel by jim měl říct, že jsou texty ve cvičení autentické a že je napsali jiní žáci. První větu může učitel se žáky opravit sám, další poté opravují samostatně nebo ve dvojicích. Kontrolu učitel provede v plénu, měl by nechat žáky přednést různé návrhy oprav a v případě nejasností je nechat opravu vysvětlit.

Opravte chyby ve větách se slovesem „jít“. Chyba může být prefix, adverbium, prepozice nebo „se“.

1. Tancovali jsme, bavili jsme se a pili jsme víno. Potom jsme pošli do klubu a tancovali jsme tam taky.
2. Nechceš půjít se mnou do kina na zajímavý film?
3. Teď jsem poprvé v Čechách. Přišla jsem tady, abych se učila jako Erasmus studentka na Univerzitě Karlově v Praze.
4. Milá Evo, děkuju moc za pozvání. Ale nemůžu přijít k narozeninám, protože jsem špatně napsala matematický test a v neděli se budu učit matematiku.
5. Já jsem studentka už tři roky. Univerzita, na které studuju, je ve stejném městě, kde bydlím. Teprve teď, když jsem přišla v Čechách, jsem měla možnost pochopit, co je život na koleji.
6. Tenkrát jsme se rozhodli, že už nikdy nepřijdeme do té restaurace.
7. Všichni se přišli s dárky a květinami pro Julii, která měla narozeniny.
8. Je mi fakt líto, že k tobě nemůžu přejít. Doufám, že budeš mít nejlepší oslavu narozenin ve svém životě!
9. Ve středu budu psát domácí úkol a ve čtvrtek budu jít do kina s Monikou.
10. Můžeš ke mně přijet na návštěvu vlakem. Bydlím blízko nádraží, tak sejdeme tam.
11. V 14:30 jsem napsal dopis a pak jsem prošel na poštu pro známku.

Postup práce s korpusem

1. Vyberte si korpus CzeSL-SGT, Typ dotazu lemma a napište následující dotaz. Hledáte tak všechny tvary slovesa *jít*, s prefixy i bez prefixu – to zajišťuje regulární výraz.*



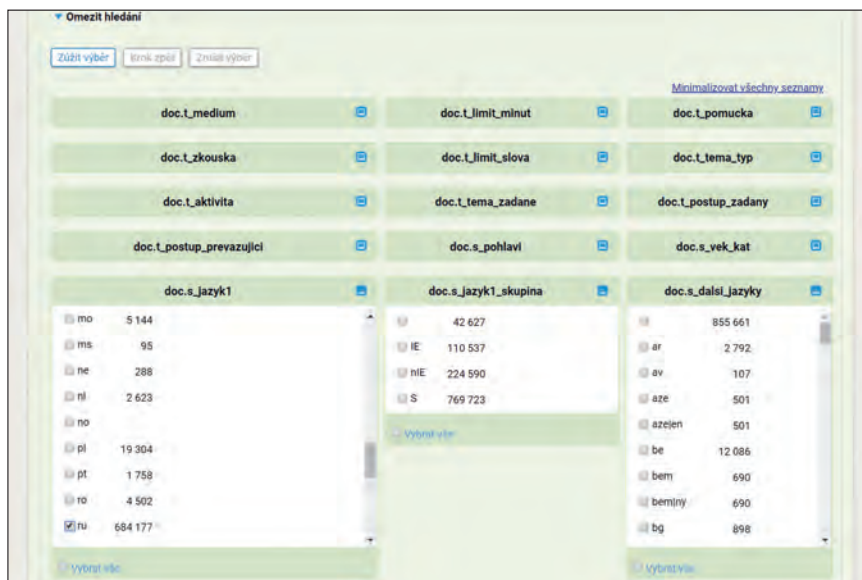
Obrázek 61

2. Klikněte na Hledat a zobrazí se vám konkordance. Vyberte si vhodné věty do cvičení, je vhodné opravit případné další chyby, které vás nezajímají. Kontext zobrazíte kliknutím na zvýrazněné sloveso.



Obrázek 62

3. Pokud chcete hledat pouze v textech určité skupiny mluvčích, je třeba zaškrtnout vybranou kategorii v části Omezit hledání.



Obrázek 63

4. Pomocí funkce Frekvence (s atributem lemmata) se můžete podívat, která slovesa jsou nerodilými mluvčími nejčastěji užívána, a vybrat si příklady podle jejich zastoupení. Některé chyby odhalíte již v tomto seznamu (např. sloveso *půjít*).

Korpus: czech-sgt | Dotaz: "jít,ru (4 579 výskytů) Promíchat ✓

Frekvenční seznam

1 / 1

Minimální frekvence: 1

Celkem: 34 položek (1 stránka)

| | Filter | lemma | Freq |
|----|--------|---------|------|
| 1 | p / n | jít | 3386 |
| 2 | p / n | přijít | 331 |
| 3 | p / n | najít | 302 |
| 4 | p / n | sejít | 243 |
| 5 | p / n | pojit | 76 |
| 6 | p / n | vyjít | 40 |
| 7 | p / n | dojit | 36 |
| 8 | p / n | projit | 36 |
| 9 | p / n | odejít | 35 |
| 10 | p / n | zajít | 18 |
| 11 | p / n | přejít | 14 |
| 12 | p / n | vejít | 11 |
| 13 | p / n | obejít | 7 |
| 14 | p / n | ujít | 7 |
| 15 | p / n | rozejít | 6 |
| 16 | p / n | přijít | 6 |
| 17 | p / n | spojit | 5 |
| 18 | p / n | půjít | 3 |
| 19 | p / n | podejít | 2 |
| 20 | p / n | výjít | 1 |
| 21 | p / n | tajít | 1 |
| 22 | p / n | prijít | 1 |
| 23 | p / n | přjít | 1 |
| 24 | p / n | vjit | 1 |

Obrázek 64

Řešení

1. pošli – šli
2. půjt se mnou – se mnou jít
3. Přišla jsem tady – Přišla jsem sem
4. nemůžu přijít k narozeninám – nemůžu přijít/jít k tobě / na narozeniny
5. když jsem přišla v Čechách – přišla / přijela do Čech
6. nikdy nepřijdeme do té restaurace – nepůjdeme do té restaurace
7. Všichni se přišli s dárky – všichni přišli s dárky
8. že k tobě nemůžu přejít – přijít
9. budu jít do kina – půjdu/jdu
10. tak sejdeme tam – tak se sejdem tam
11. jsem prošel na poštu pro známku – jsem šel / si došel / na poštu pro známku

Pracovní list č. 7:
OPRAVA CHYB - Kvantita

Anotace

Pracovní list obsahuje cvičení založené na jazykovém materiálu z korpusu. Učitel si cvičení připraví předem, žáci ve výuce s korpusem nepracují. Jejich úkolem je opravit chyby v kvantitě. Texty pocházejí od nerodilých mluvčích češtiny, učitel je nalezne v korpusu CzeSL-SGT.

Postup práce

Toto cvičení je vhodné zařadit do lekce věnované pravopisu. Je možné vybírat do něj věty s chybami typickými pro danou skupinu. Učitel tak může například hledat pouze texty rodilých mluvčích jednoho jazyka odpovídajícího jazyku jeho žáků.

Vhodnou motivací pro úkoly, ve kterých žáci opravují chyby, může být instrukce, že jsou nyní sami učitelé. Učitel by jim měl říct, že jsou texty ve cvičení autentické a že je napsali jiní žáci. První větu může učitel se žáky opravit sám, další poté opravují samostatně nebo ve dvojicích. Kontrolu učitel provede v plénu, měl by nechat žáky přednést různé návrhy oprav a v případě nejasností je nechat opravu vysvětlit.

Doplňte nebo škrtněte čárky.

1. A ja jsem chtěl velký byt.
2. Obědvala tam chutné jídlo.
3. Mám češtinu každé rano, ale o vikendu mam volno.
4. Moje město ma různá zajímavá místa.
5. Auta take nepotřebujou jídlo a nemusí odpočívát.
6. Mám se dobře, ale mám malý problem. Musím koupít nový byt.
7. 21. ledna musím byt na univerzitě, protože mám semestrální zkoušku.
8. Studuju češtinu každý den od 9:00 do 13:00.
9. Teď jsem moc spokojený, rad bydlím v Praze.
10. A taký Praha je velmi hezká.

Frekvenční seznam

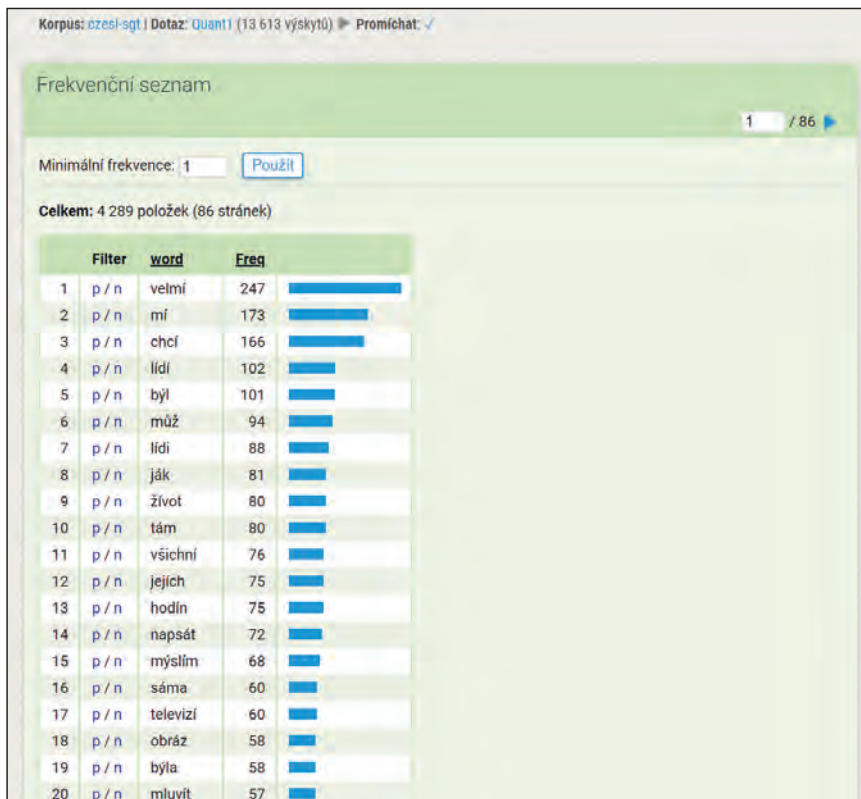
1 / 148 ▶

Minimální frekvence: 1

Celkem: 7 397 položek (148 stránek)

| Filter | word | Freq |
|--------|-----------------|------|
| 1 | p / n ja | 1323 |
| 2 | p / n mam | 618 |
| 3 | p / n take | 493 |
| 4 | p / n každý | 489 |
| 5 | p / n ma | 453 |
| 6 | p / n jídlo | 441 |
| 7 | p / n například | 350 |
| 8 | p / n myslim | 333 |
| 9 | p / n problem | 332 |
| 10 | p / n rano | 329 |
| 11 | p / n máš | 327 |
| 12 | p / n být | 307 |
| 13 | p / n rad | 299 |
| 14 | p / n jit | 290 |
| 15 | p / n krásně | 289 |
| 16 | p / n prochazku | 285 |
| 17 | p / n tyden | 281 |
| 18 | p / n můžeme | 277 |
| 19 | p / n kamarady | 272 |
| 20 | p / n tatinek | 234 |

Obrázek 67



Obrázek 68

Řešení

1. A **já** jsem chtěl velký byt.
2. Obědvála tam chutné **jídlo**.
3. Mám češtinu každé **ráno**, ale o **víkendu** mám volno.
4. Moje město má různá zajímavá **místa**.
- 5.auta také nepotřebujou **jídlo** a nemusí odpočívat.
6. Mám se dobře, ale mám malý **problém**. Musím koupit nový byt.
7. 21. ledna musím **být** na univerzitě, protože mám semestrální zkoušku.
8. Studuju češtinu každý den od 9:00 do 13:00.
9. Teď jsem moc spokojený, **rád** bydlím v Praze.
10. A taky Praha je velmi hezká.

Pracovní list č. 8: **OPRAVA CHYB – Pasivum**

Anotace

Pracovní list obsahuje cvičení založené na jazykovém materiálu z korpusu. Učitel si cvičení připraví předem, žáci ve výuce s korpusem nepracují. Jejich úkolem je opravit chyby v pasivních konstrukcích. Texty pocházejí od nerozdaných mluvčích češtiny, učitel je nalezne v korpusu CzeSL-SGT.

Postup práce

Toto cvičení je vhodné zařadit jako shrnující opakování pasivních konstrukcí. Je možné vybírat do něj věty s chybami typickými pro danou skupinu. Učitel tak může například hledat pouze texty rodilých mluvčích jednoho jazyka odpovídajícího jazyku jeho žáků.

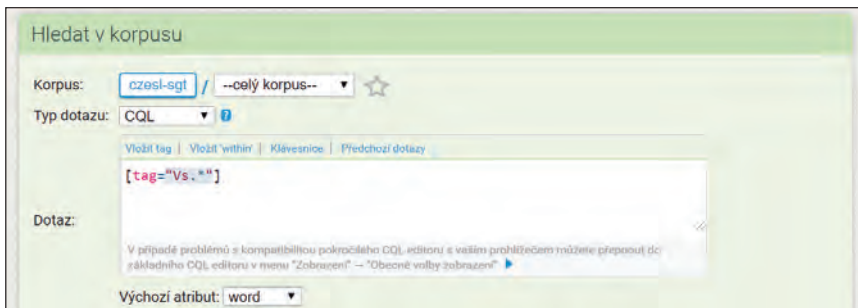
Vhodnou motivací pro úkoly, ve kterých žáci opravují chyby, může být instrukce, že jsou nyní sami učitelé. Učitel by jim měl říct, že jsou texty ve cvičení autentické a že je napsali jiní žáci. První větu může učitel opravit společně se žáky, další poté opravují samostatně nebo ve dvojicích. Kontrolu učitel provede v plénu, měl by nechat žáky přednést různé návrhy oprav a v případě nejasností je nechat opravu vysvětlit.

Opravte chyby v pasivu a pasivních adjektivech.

1. Já jsem měla šťastné dětství, zač velmi vděčena rodičům.
2. Sportovní hala otevřeno denně.
3. Ta učebnice napsána pro cizince, kteří mluví rusky.
4. Na obrázku vidíme manželský pár v letech. Starší pán nabídl paní rámě. Asi je jaro, protože nejsou oblečeny moc teple.
5. Ti lidé nejsou lhostejní, jenom je moc nezajímá, v čem jsi oblečeny a co máš za účes, proto se po tobě v metru nekoukají.
6. Studium v zahraničí je však spojen s mnoha problémy.
7. Ačkoli v Rusku je mnoho náboženských svátků a celkem Rusové jsou pravoslavní, tyto svátky jsou povětšinou neslaveny v kostelích. Můžeme to přirovnat k české oslavě svátků, např. na Vánoce zdaleka ne všichni jdou do kostela, ale zůstávají doma s rodinou.
8. Studujeme proto, abychom mohli být vždy připraveny na všechny životní úkoly.

Postup práce s korpusem

1. Zvolte si korpus CzeSL-SGT a Typ dotazu CQL. Vložte tag pro slovesa, v detailnějším určení slovního druhu zvolte příčestí trpné. Klikněte na Hledat.



Obrázek 69

V konkordancích se vám zobrazí věty s příčestím trpným. Vyberte si v nich věty vhodné do cvičení a případně je upravte. Pomocí funkce Frekvence si můžete zobrazit slovní tvary, které dělají žákům největší potíže.



Obrázek 70

Řešení

1. zač **jsem** velmi věd**ě**č**n**á
2. Sportovní hala **má** otevřeno / **je** otevř**en**á denn**ě**.
3. Ta učebnice **byla** naps**á**na / **je** naps**an**á
4. nejsou obleč**eni**/í
5. jsi obleč**ě**ný
6. je v**š**ak spoj**en**o/é
7. **nejsou** pov**ě**tš**in**ou sl**av**eny
8. abychom mohli b**ý**t v**ž**dy p**ř**iprav**en**i

5.2.2 SyD

Pracovní list č. 9:

KONKURENCE KONCOVEK -UJU A -UJI - Děkuju ti za dárek. Děkuji vám za pomoc.

Anotace

Cvičení tohoto typu je vhodné zařadit do výuky, pokud chceme žákům ozřejmit použití dubletních tvarů, které se liší stylovou platností a/nebo příslušností k jiné vrstvě českého jazyka (nespisovné X spisovné). Tento list je věnován tvarům *děkuju X děkuji*, stejný postup lze využít např. pro tvary *pěct/píct, moci/moct, pší/píšu, mohu/můžu, bychom/bysme* atd. Pracovní list vyžaduje samostatnou práci s korpusovou aplikací SyD a obsahuje tři úkoly, které vedou k tomu, aby si žák význam adjektiv osvojl vlastními silami, a to samostatným objevením pravidel a následně jejich aplikací.

Postup práce

Před obdržením pracovního listu by měli žáci vidět obě morfologické varianty a měli by si jich sami všimnout v textu. V tomto případě se učitel může na začátku zeptat, jak můžeme v češtině děkovat. Je pravděpodobné, že žáci budou znát obě varianty *děkuju* a *děkuji* a možná i další (viz řešení).

Učitel se poté zeptá žáků, kdy by kterou variantu použili. Může je vyzvat, aby si představili, že mluví s kamarádem/ředitelem/učitelem nebo že píše oficiální nebo neoficiální e-mail. Žáci možná budou sami vědět, že varianta *děkuji* je formálnější a častěji se používá v psaných textech. I přesto pro ně může být zajímavé podívat se do korpusu na rozložení v psaném a mluveném jazyku. Předtím, než se do korpusu podívají, můžou hádat, jak výsledek dopadne – tj. kdy se častěji používá která varianta a jaký je poměr. Poté žáci samostatně nebo ve dvojicích vyřeší úkol 2 a 3. Předem musejí znát postup práce s aplikací.

Pokud žáky výsledky zaujmou, mohou se dále podívat na rozložení variant *děkuji* a *děkuju* v psaných textech. Zjistí tak, že varianta *děkuji* se nejčastěji používá v odborných textech, *děkuju* naopak v beletrii.

Učitel se následně žáků může zeptat, která další slovesa mají takové dublety, a žáci se mohou podívat, zda se tyto dublety chovají stejně. Zajímavé je např. sloveso *pracovat*, které se v psaném jazyce v první osobě singuláru častěji objevuje ve formě *pracuji*, v mluveném jazyce se v korpusu najde jen podoba *pracuju*.

Nakonec by žáci měli sami obě varianty vhodně použít. Učitel si předem může připravit různé situace, ve kterých mají žáci někomu ústně nebo pí-

semně poděkovat a použít jednu nebo druhou dubletu. O vhodnosti použití lze samozřejmě diskutovat. Žáci by před vlastní produkcí měli znát nejčastější kolokace, a to především kolokaci *děkovat někomu za něco*. Kolokace najdou po kliknutí na stejnojmennou ikonu, která se v aplikaci zobrazí spolu s výsledky. Zde si můžou kliknutím na jednotlivá slova zobrazit příkladové věty. Kolokace samozřejmě mohou hledat také přes rozhraní KonText.

DĚKUJU TI ZA DÁREK. DĚKUJI VÁM ZA POMOC.

1. Jak můžeme v češtině děkovat?
2. Co je typické pro psaní a co pro mluvení? Podívejte se do korpusu (<https://syd.korpus.cz/>).

Pro psaní je typické:

Pro mluvení je typické:

3. Doplňte do tabulky procenta z grafů.

| varianta | psaní | mluvení |
|----------|-------|---------|
| děkuji | | |
| děkuju | | |

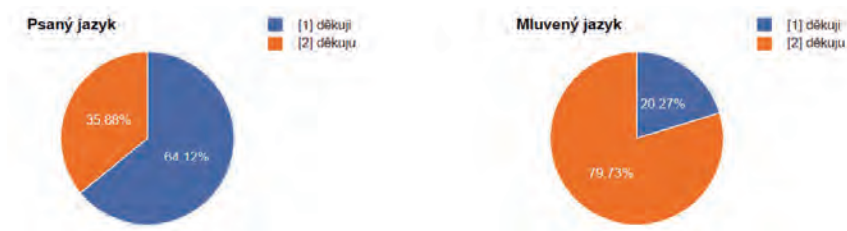
4. Zahrajte se spolužákem minidialog. Spolužákovi musíte v dialogu za něco poděkovat.
5. Napište krátký e-mail učiteli nebo učitelce. V e-mailu musíte za něco poděkovat.

Postup práce s korpusem

Do jednoho okénka aplikace SyD zadejte tvar *děkuji*, do druhého *děkuju*. Nechte zaškrtnuté políčko a=A (nezáleží na tom, jestli je první písmeno malé nebo velké), ale nezaškrťávejte políčko lemma (chcete hledat jen tyto konkrétní tvary slovesa). Klikněte na tlačítko Porovnat varianty.

Řešení

1. děkuji, děkuju (děkuju/i moc, díky, dík, díkec, d')
2. Pro psaní je typické děkuji. Pro mluvení je typické děkuju.



Obrázek 71

| varianta | psaní | mluvení |
|----------|--------|---------|
| děkuji | 64,12% | 20,27% |
| děkuju | 35,88% | 79,73% |

Pracovní list č. 10:

PRAGMATIKA - Rozloučení v oficiálním dopise

Anotace

Tento pracovní list slouží jako zdroj doplňkových cvičení, která lze zařadit do výukové jednotky věnované psaní dopisů. Cvičení jsou zaměřena na pragmatiku, konkrétně pak závěrečný pozdrav v korespondenci. Žákům mohou pomoci uvědomit si, jaký způsob rozloučení je v českých dopisech neutrální.

Postup práce

Žáci by měli nejprve pomocí příkladů probrat formální náležitosti oficiálních dopisů. Poté učitel jejich pozornost obrátí k rozloučení. V některých jazycích je typické loučit se obsáhle. Čeština je naopak relativně strohá. Žáci by měli nejprve sami nebo ve skupinách jmenovat některá typická formální rozloučení. Jako inspirace jim může sloužit například Internetová jazyková příručka (<http://prirucka.ujc.cas.cz/?id=850>). Následně se učitel žáků zeptá, který způsob rozloučení je asi nejčastější. Žáci vysloví své domněnky a poté se samostatně nebo s učitelem podívají do korpusu. Po zobrazení výsledků mohou diskutovat o rozdílech mezi loučením v češtině a v jejich mateřském jazyce. Podobně mohou hledat také rozlučkové formule v neformálních dopisech.

1. Jak se můžeme rozloučit v oficiálním dopise nebo e-mailu? Kde byste hledali inspiraci?
2. Podívejte se do Internetové jazykové příručky (<http://prirucka.ujc.cas.cz/>) a napište varianty rozloučení.

3. Porovnejte varianty. Která je asi nejčastější?

S pozdravem
S přátelským pozdravem
Se srdečným pozdravem
Se srdečnými pozdravy
S přátelskými pozdravy
S přáním pěkného dne
S přáním hezkého dne

4. Podívejte se do korpusu (<https://syd.korpus.cz/>) a řekněte, která varianta je při psaní častější.
5. Podívejte se, které rozloučení se používá nejčastěji v korespondenci.
6. Máte ve vašem jazyce podobné fráze, nebo je rozloučení typicky jiné?

Postup práce s korpusem

Do okének aplikace SyD jednotlivě napište všechny varianty. Nezaškrťávejte políčko a=A (hledáme začátek vět) ani políčko lemma. Klikněte na tlačítko Porovnat varianty.

Pro zobrazení typů textů klikněte na záložku Psaný jazyk a dále Typy textů.

Řešení

1. viz úkol 2 a 3
4. Při psaní se nejčastěji používá varianta S pozdravem



Obrázek 72

5. V korespondenčních textech tato varianta převažuje ještě výrazněji.



Obrázek 73

Pracovní list č. 11: **PRAVOPISNÉ DUBLETY - s/z**

Anotace

Cvičení v tomto pracovním listu jsou zaměřena na pravopisné dublety *gymnázium/gymnasium*, *univerzita/universita* a *filozofie/filosofie*. Pro žáky mohou být oporou při rozhodování o vhodnosti způsobu zápisu. Zadáání lze libovolně variovat, nabízí se zápis s odlišnou kvantitou (*balon/balón*) nebo psaní cizích slov (*management/manažment*).

Postup práce

Učitel žáky nejprve nechá samostatně doplnit s nebo z do slov ve cvičení jedna a řešení konzultovat se spolužákem. Následně svá řešení žáci zkonzultují v plénu. Je pravděpodobné, že nedojde ke shodě na jedné variantě, žáci mohou argumentovat pocitem líbivosti. Učitel zatím nebude řešení komentovat, ale odkáže žáky na aplikaci SyD. Ve dvojicích nebo v plénu do ní žáci postupně zadají všechny dvojice. Ve výsledcích nás tentokrát bude zajímat jen psaný jazyk. V aplikaci žáci zjistí, že výrazně převládá zápis se z (viz řešení). Žáci se mohou pokusit sami přijít na to, v jakých textech se nejčastěji objevuje podoba se s. Jelikož ze zmíněných substantiv se zápis se s nejčastěji vyskytuje v případě *filozofie*, je vhodné zaměřit se především na toto substantivum. Pomocí záložky Typy textů žáci zjistí, že se tato varianta nejčastěji vyskytuje v korespondenci a dále v beletrii. Podrobnější dělení ukáže, že z beletrie je zápis *filozofie* nejčastější v básních a dramatech. Žáci by tedy sami mohli přijít na to, že se tento způsob zápisu používá pro ozvláštňení. Zastaralost zápisu se s by teoreticky žáci mohli odvodit z toho, že se častěji vyskytuje v korespondenci, která dnes není tak často využívána. Z analýzy by si měli odnést především to, že podoba se z je nyní při psaní přejatých jmen variantou neutrální. Nakonec se žáci mohou zamyslet nad tím, která další slova se dříve psala se s a dnes už se píšou se z.

1. Doplňte písmena: gymná__ium, univer__ita, filo__fie. Potom svoje řešení porovnejte se spolužákem. Doplňili jste stejné písmeno?
2. Porovnejte varianty *gymnázium/gymnasium*, *univerzita/universita* a *filozofie/filosofie*. Které varianty se vám více líbí?
3. Podívejte se do korpusu (<https://syd.korpus.cz/>) a řekněte, které varianty jsou při psaní častější.
4. Podívejte se, ve kterých textech se nejčastěji používá varianta *se s*. Proč to tak je?
5. Znáte další slova, která se dříve psala *se s* a nyní se píšou *se z*?

Postup práce s korpusem

Do okének aplikace SyD jednotlivě napište obě varianty. Nechte zaškrtnuté políčko a=A (nezáleží na tom, jestli je první písmeno malé nebo velké), políčko lemma nezaškrtněte. varianty se přiřazují k jednomu lemmatu, proto byste našli jen nové varianty *gymnázium*, *filozofie* a *univerzita*. Klikněte na tlačítko Porovnat varianty.

Pro zobrazení typů textů klikněte na záložku Psaný jazyk a dále Typy textů (nejprve základní dělení, poté podrobně).

Řešení

1. *gymnázium/gymnasium*, *univerzita/universita*, *filozofie/filosofie*
2. dle diskuze
3. Vždy výrazně převládá varianta se z (více než 97% pro gymnázium a univerzita, skoro 67% pro filozofie).
4. Pomocí záložky Typy textů žáci zjistí, že se tato varianta nejčastěji vyskytuje v korespondenci a dále v beletrii. Podrobnější dělení ukáže, že z beletrie je zápis *filosofie* nejčastější v básních a dramatech. Je to kvůli zastaralosti této podoby nebo pro ozvláštnění.
5. Např. dezinfekce, epizoda, fantazie

Pracovní list č. 12:

STYLOVÉ TRIPLETY – Stále mě někdo opravuje. Ještě pořád neumím dobře česky. Proč mi to furt nejde?!

Anotace

Cvičení tohoto typu je vhodné zařadit do výuky, pokud chceme žákům ozřejmit použití dublet (triplet), které se neliší významově, ale pouze svou stylovou platností. Zadání lze libovolně variovat, v tomto pracovním listu se jedná o varianty *stále*, *pořád* a *furt*.

Postup práce

Před obdržení pracovního listu by měli žáci vidět všechny varianty a měli by si jich sami všimnout v textu.

Učitel se poté zeptá žáků, kdy by jakou variantu použili. Může je vyzvat, aby si představili, že mluví s kamarádem / ředitelem / učitelem nebo že píše oficiální nebo neoficiální e-mail. Varianty by se měli pokusit seřadit podle formálnosti. Následně se sami podívají do aplikace SyD, předem musí být obeznámeni s jejím fungováním.

Po zobrazení výsledků si žáci mohou sami ověřit, zda odpovídají jejich očekáváním o formálnosti daných variant. Za tímto účelem se také mohou podívat na typy textů a zjistit, v jakých se nejčastěji používá jaká varianta. Žáci si mohou povšimnout např. relativně častého užití varianty *furt* v korespondenci. Učitel pak může s žáky diskutovat o tom, jak se korespondence blíží mluvenému jazyku a zda jsou tyto tendence silnější např. při psaní sms zpráv. Dále by měl upozornit na to, že s použitím varianty *furt* v mluveném jazyce souvisí také její neformálnost a že se tedy nehodí do oficiálního projevu.

Nakonec si mohou žáci vyzkoušet použití sami. Mohou si například v rollových hrách stěžovat různým lidem a podle toho vhodně měnit lexikum. Paní učitelce tak mohou říkat: *Moje kamarádka pořád mluví a já neslyším, co říkáte*. Spolužákovi mohou říct: *Eva furt kecá a já nic neslyším!* Role se zadanými problémy, situacemi a slovní zásobou si může učitel připravit předem nebo je mohou vymyslet sami žáci.

**STÁLE MĚ NĚKDO OPRAVUJE. JEŠTĚ POŘÁD NEUMÍM
DOBŘE ČESKY. PROČ MI TO FURT NEJDE?!**

1. Porovnejte varianty *stále*, *pořád*, *furt*. Jaká je nejformálnější, jaká je nejméně formální?
2. Podívejte se do korpusu (<https://syd.korpus.cz/>) a řekněte, která varianta se nejčastěji používá při psaní a která při mluvení?
3. V jakých psaných textech se nejčastěji používá *stále*, v jakých *pořád* a v jakých *furt*?

stále:

pořád:

furt:

4. Stěžujte si na to, co často dělá váš kamarád / vaše kamarádka. Použijte vhodnou variantu a vhodná slova podle toho, komu si stěžujete.

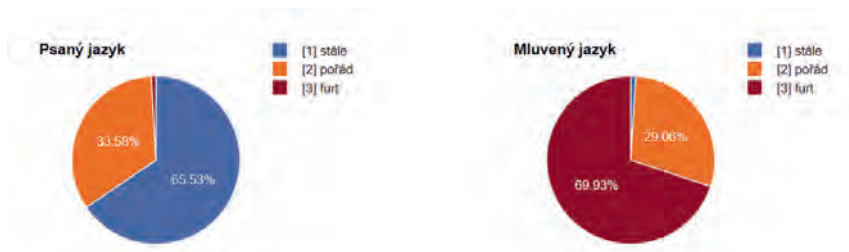
Postup práce s korpusem

Do okének aplikace SyD jednotlivě napište všechny varianty. Nechte zaškrtnuté políčko a=A (nezáleží na tom, jestli je první písmeno malé nebo velké), na políčku lemma nezáleží. Klikněte na tlačítko Porovnat varianty.

Pro zobrazení typů textů klikněte na záložku Psaný jazyk a dále Typy textů.

Řešení

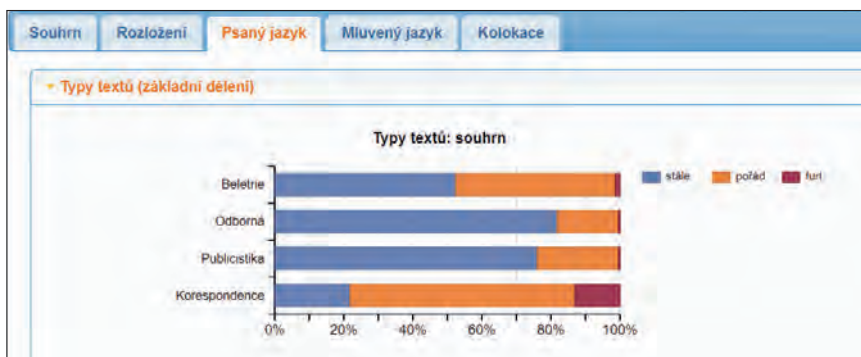
1. stále – pořád – furt



Obrázek 74

2. Při psaní se nejčastěji používá varianta *stále*, při mluvení *furt*.

3. Typy textů



Obrázek 75

Pracovní list č. 13: **OBECNÁ ČEŠTINA - ý>ej, é>í/ý**

Anotace

Cvičení v tomto pracovním listu jsou zaměřena na obecnou češtinu a její použití na území České republiky. Konkrétně jde o konkurenci užití slov s diftongizací ý > ej (dobrej) a bez diftongizace (dobrý) a s úžením é > í (mlíko) a bez úžení (mléko).

Postup práce

Žáci by měli před samotnou prací s pracovním listem vědět, že v České republice existují různé nářeční oblasti. Učitel může žákům pomocí mapy ukázat, kde se nacházejí. Zároveň by měli být seznámeni s existencí obecné češtiny a vhodnosti jejího užití v různých komunikačních situacích. Pracovní list by měl navazovat na práci s textem, ve kterém se objevují obecné české prvky. Těch by si žáci měli povšimnout. Poté obdrží pracovní list.

Učitel se nejprve zeptá žáků, co o obecné češtině vědí. Žáci mohou pracovat nejdříve ve skupinách. Typické charakteristiky poté zapisuje na tabuli učitel nebo žáci. Učitel se následně zaměří na změnu ý>ej a é > í. Vyzve žáky, aby se do korpusu podívali na rozložení jejího výskytu na území ČR pomocí daných příkladů. Příklady lze samozřejmě změnit. Důležité ale je, že v případě sledování změn koncovek je potřeba zadávat celé spojení (např. celý den X celej den), jinak by se vyhledávala i adjektiva v jiném spojení (např. celý auta atd.), což by zkreslilo výsledek. Žáci se mohou před vlastní prací s korpusem pokusit odhadnout, kde bude která varianta převažovat. Při práci je důležité, aby pracovali s výsledky pro mluvený jazyk a nářeční oblasti. Korpus není z hlediska nářečních oblastí vyvážený, proto nemá smysl vyhodnocovat pouze výskyt na celém území ČR zobrazující se v prvních kruhových grafech. Až žáci určí, kde diftongizace a úžení typicky probíhá, vymyslí ve dvojicích příklady dalších slov, kde se ý mění na ej a é na í. Následně zahrají dialog, ve kterém si vyzkoušejí roli „typického Čecha“ a „typického Moravana“. Cvičení lze samozřejmě doplnit o další charakteristiky obecné češtiny, např. převahu koncovky -ý v nominativu plurálu všech rodů (dobrý auta, ženy, muži, hrady).

1. Co víte o obecné češtině? Která obecně česká slova znáte?
2. Pro obecnou češtinu je typická změna *ý>ej* (dobrej kámoš) a *é > í* nebo *ý* (velký auto, mlíko). Podívejte se do korpusu (<https://syd.korpus.cz/>) a řekněte, kde jsou v ČR tyto změny typické a kde ne. Použijte příklady:
výška - vejška, celý den - celej den, mléko - mlíko
3. Která další slova se můžou takto měnit? Napište co nejvíc příkladů.

4. Zahrajte dialog typického Čecha s typickým Moravanem. Jako Čech v něm používejte „ej“ a „ý/í“, jako Moravan mluďte spisovně.

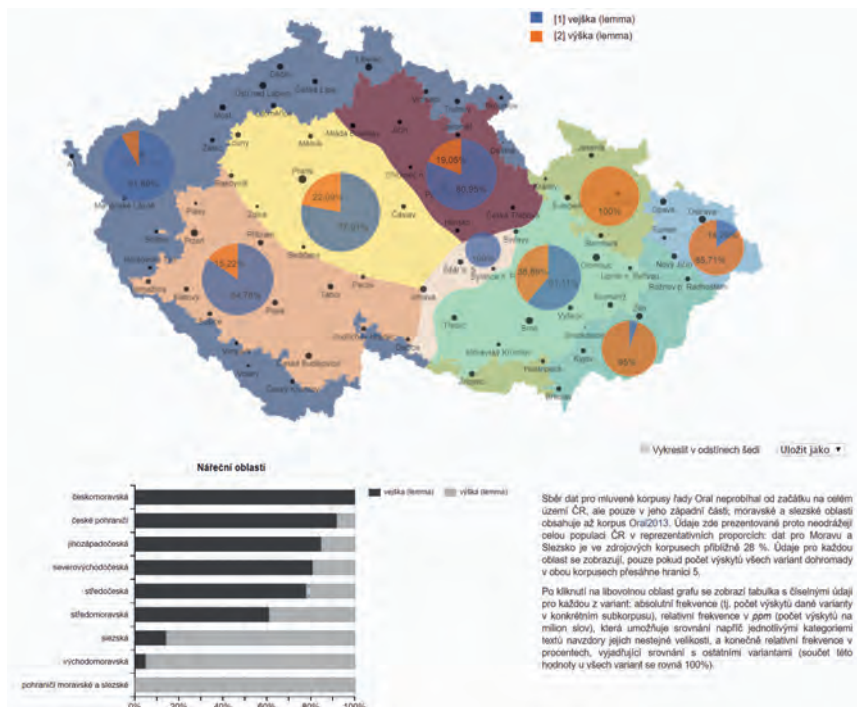
Postup práce s korpusem

Do okének aplikace SyD vždy jednotlivě napište obě varianty. Nechte zaškrtnuté políčko $a=A$ (nezáleží na tom, jestli je první písmeno malé nebo velké), v případě jednoslovných hledání i políčko lemma (hledáme všechny tvary). Klikněte na tlačítko Porovnat varianty. Dále vyberte záložku Mluvený jazyk a Nářeční oblasti. Zobrazí se vám barevná mapa s grafy a zároveň graf pro jednotlivé nářeční oblasti.

Adresy dotazů: <http://syd.korpus.cz/c3RhKzYq.syn> (vejška, výška), <http://syd.korpus.cz/visNjrSk.syn> (celej den, celý den); <http://syd.korpus.cz/h2skGGPz.syn> (mlíko, mléko)

Řešení

viz grafy



Obrázek 76

5.2.3 Morfio

Pracovní list č. 14:

TVORBA ADJEKTIV - Zdravý životní styl = žádné zdravotní problémy?

Anotace

Cvičení tohoto typu je vhodné zařadit pro znázornění různých slovotvorných modelů. Žáci si díky nim lépe uvědomí, že lze od známých slov odvozovat nová slova pomocí přípon. Cvičení lze libovolně variovat, je možné hledat např. slova odvozená předponami (např. *chod* - *vchod*) nebo slova odvozená předponami i příponami (např. *lov* - *úlovek*). V tomto pracovním listu půjde o adjektiva utvořená příponami *-ý* a *-otní* (např. *zdravý* a *zdravotní*).

Postup práce

Tento pracovní list může učitel zařadit po pracovním listu, ve kterém žáci v KonTextu vyhledávají podstatná jména, která se pojí s přídavnými jmény *zdravý* a *zdravotní*. V prvním úkolu si mají na tato podstatná jména vzpomenout a vhodně je spojit s adjektivy, ve druhém pak mohou opakovat definici z předchozího pracovního listu. Odpovědí na třetí otázku má být substantivum *zdraví*, ke kořenu *zdrav* se připojují přípony *-ý* a *-otní*. Pokud žáci neznají názvosloví, může je s ním učitel tímto způsobem seznámit. Následně se učitel žáků zeptá, zda znají nějaká další slova, která vznikla stejným způsobem. Učitel si předem může připravit obrázky nebo věty s vynechanými slovy, která žáci hledají (viz řešení), a žáci se mohou sami pokusit zjistit, o která slova se jedná. Pravděpodobně se jim nepodaří určit všechna slova, a to by mělo sloužit jako motivace pro práci s korpusem. Žáci se sami nebo s učitelem podívají do aplikace Morfio a zjistí, která další slova jsou odvozena stejným způsobem. K některým přídavným jménům by měli být sami schopni utvořit příkladové kolokace. Pokud budou chtít, mohou si nejčastější kolokace najít pomocí rozhraní KontText stejným způsobem, kterým vyhledávali kolokace přídavných jmen *zdravý* a *zdravotní*. Pokud učitel s KonTextem v hodině pracovat nechce, může si nejčastější kolokáty najít předem a žákům je poté nabídnout v pátém úkolu (viz nabídka, zvolit lze samozřejmě i jiná slova). V pátém úkolu jsou úmyslně vynechána přídavná jména *prvý* a *prvotní*, která mají jinou sémantiku.

ZDRAVÝ ŽIVOTNÍ STYL = ŽÁDNÉ ZDRAVOTNÍ PROBLÉMY?

1. Co může být zdravé? Co může být zdravotní?
2. Jak můžete definovat slova zdravý a zdravotní?
3. Od kterého podstatného jména tato přídavná jména vznikla? Jak?
4. Znáte další přídavná jména se stejnými příponami?
5. Použijte aplikaci Morfio a najděte další přídavná jména se stejnými příponami. Napište je a vymyslete k nim příklad.
6. Co může být *husté*, *teplé* a *živé*? Co *hustotní*, *teplotní* a *životní*? Do-
pište přídavná jména k podstatným jménům se správnou koncov-
kou.

| | |
|---|--|
| <i>prostředí, styl, cyklus, úroveň, pojištění, příběh</i> | |
|---|--|

| | |
|---|--|
| <i>plot, organismus, tvor, bytost, sen, diskuse</i> | |
|---|--|

| | |
|--|--|
| <i>koeficient, výkyv, rekord, extrém, stupnice</i> | |
|--|--|

| | |
|---|--|
| <i>voda, počasí, vzduch, klima, jídlo, koupel</i> | |
|---|--|

| | |
|--|--|
| <i>mlha, obočí, vlasy, srst, déšť, polévka</i> | |
|--|--|

| | |
|----------------------------------|--|
| <i>hranice, profil, kontrast</i> | |
|----------------------------------|--|

Postup práce s korpusem:

1. V první kolonce pro společnou bázi ponechte „+“. U prvního vzoru doplňte do kolonky pro odlišný formant *ý*, u druhého *otní*. U obou zvolte morfológickou specifikaci přídavná jména. Vyberte, že se hledají a vyhodnocují lemmata. Velikost písmen můžete ignorovat. Nakonec klikněte na tlačítko Hledat. Pro zobrazení výsledků klikněte na Výpis.

Morfio

→ X společný X odlišný → Morf. specifikace:

vzor 1: + vzor 2: + Morf. specifikace: +

vzor 1: ý vzor 2: otní vzor 1: přídavná jména vzor 2: přídavná jména

vzor 1: A vzor 2: A

Další vzor:

Korpus: SYN2015 Fokusem výřův. rež.: 10 Hledat se: lemmata Vyhodnocují se: lemmata

Velikost písmen: ignorovat Alternace

Hledat Nové zadání Odkaz na tuto zadání http://morfio.korpus.cz/4_DeM621 Nápověda

| Souhrn | Výpis | Produktivita | vzor 1 | vzor 2 |
|--------|----------------|--------------|--|--------|
| | | | vzor 1 [fg **] vzor 2 >> [fg **] | |
| 1 | husty (5417) | | hustý (25) | |
| 2 | prvý (2756) | | první (1696) | |
| 3 | svý (6817) | | svobodní (1583) | |
| 4 | zdravý (10239) | | zdravotní (12610) | |
| 5 | žvý (12739) | | žvotní (19621) | |

V tabulce jsou uvedeny všechny doklady ze všech vzorů, které vstupují do zadávaného modulu. Červená část slov označuje společnou bázi (ta se může lišit pouze v případě aplikace alternací). V závorkách uvedeny jsou představené celkové frekvence jednotky ve zvoleném korpusu. Tabulku je možné přefiltrovat podle lemového sloupce a to jak abecedně, tak frekvence v pomoci šipek v zánovní tabulce. Každé slovo zároveň funguje jako odkaz směřující k aplikaci konkordance ve zvoleném korpusu.

Paty vyřazené až díky aplikaci alternančních pravidel jsou zvýrazněné **barveným pozadím**. Jejich báze se proto lišou jím. V případě, že použijete alternančních pravidel dříve k situaci, že jednemu vzoru v dané dvojici odpovídá více slov, jsou všechna tato slova uvedena horně v jednom řádku tabulky.

Obrázek 79

2. Pro vyhledání kolokátů ve cvičení pět použijte rozhraní KonText. Vyhledejte v něm v korpusu syn2015 příslušná lemmata (*teplý, teplotní, hustý, hustotní, zdravý, zdravotní*). Až se vám zobrazí konkordance, klikněte na Kolokace, zvolte Vlastní a jako Atribut si vyberte lemma, v kontextu od 0 do 3. Ze seznamu kolokátů si poté vyberte podstatná jména dle potřeb a schopností žáků.

Řešení

1. **zdravý:** např. rozum, výživa, strava, stravování, potravina, sebevědomí, jídlo, jedinec, životospráva, bydlení
zdravotní: např. pojišťovna, postižení, pojištění, stav, péče, sestra, potíže, problém, komplikace, služba
2. zjednodušená definice ze Slovníku spisovného jazyka českého (<http://ssjc.ujc.cas.cz/>):
zdravý = má bezproblémové zdraví (zdravý jedinec, člověk); přirozený (zdravý rozum, zdravé sebevědomí); dobrý pro zdraví (zdravé jídlo, zdravé bydlení)
zdravotní = týká se zdraví (zdravotní problém, stav); stará se o zdraví (zdravotní sestra, pojišťovna)
3. Podstatné jméno zní *zdraví*. Vznikla připojením přípon *-ý* a *-otní* ke kořenu *zdrav*.
4. **hustý, hustotní; prvý, prvotní; teplý, teplotní; živý, životní**
5. **životní:** *prostředí, styl, cyklus, úroveň, pojištění, příběh*
živý: *plot, organismus, tvor, bytost, sen, diskuse*
teplotní: *koeficient, výkyv, rekord, extrém, stupnice*
teplý: *voda, počasí, vzduch, klima, jídlo, koupel*
hustý: *mlha, obočí, vlasy, srst, déšť, polévka*
hustotní: *hranice, profil, kontrast*

Pracovní list č. 15:

KONJUGACE – Slovesa čtvrté slovesné třídy I.

Anotace

Pracovní list obsahuje cvičení založená na jazykovém materiálu z korpusu. Učitel si cvičení připraví předem, žáci ve výuce s korpusem nepracují. Na základě autentického jazykového materiálu si ovšem upevní pravidla pro užití osobních koncovek sloves čtvrté slovesné třídy. Cvičení lze samozřejmě upravit pro další slovesné třídy.

Postup práce

Cvičení neslouží pro první seznámení s časováním. Žáci by s časováním sloves měli být seznámeni v komunikačním kontextu. Cvičení v tomto pracovním listu jsou zaměřená jen na koncovky a neobsahují jednotící komunikační téma. Žáci by si na jejich základě měli upevnit dříve probranou látku.

Před rozdáním pracovního listu je nutné žáky uvést do tématu. Žáci by si měli všimnout sloves čtvrté slovesné třídy v textu. Učitel se jich může dále zeptat, která slovesa se časují stejně jako např. *mluvit*. Slovesa může shromažďovat na tabuli po skupinách podle zakončení infinitivu. Učitel poté žákům rozdá pracovní list. První větu v prvním cvičení s nimi přečte v plénu a nechá si nadiktovat infinitiv.

Druhý úkol žáci plní až po kontrole prvního. Učitel nejprve obrátí pozornost žáků k typickým zakončením infinitivu. Dohromady se žáky doplní první informace v tabulce. Dále žáci postupují samostatně. Všechny osobní koncovky najdou v příkladových větách z prvního cvičení.

Po kontrole třetího úkolu žáci aplikují pravidlo v posledním cvičení. Před samostatnou prací by měl učitel ověřit porozumění slovesům. První větu mohou žáci doplnit společně.

1. Napište k tučným slovesům infinitiv.

| | |
|---|--|
| Rozumím ti dobře. | |
| Dám ti pusu, až se vrátíš . | |
| V roli Julie se představí Radka Coufalová. | |
| „Nestíhám, nemám čas,“ slyšíme všude kolem sebe. | |
| Jaké rozdíly vidíte v práci architekta v USA a v Evropě? | |
| Vždycky se říká, že lidé o sobě rádi mluví . | |

2. Doplňte tabulku.

| | |
|---|-----------------------|
| Slovesa s infinitivní koncovkou - ____, - ____, a - ____, mají osobní koncovky: | |
| já | - ____, (např. _____) |
| ty | - ____, (např. _____) |
| on/ona/ono | - ____, (např. _____) |
| my | - ____, (např. _____) |
| vy | - ____, (např. _____) |
| oni | - ____, (např. _____) |
| Všude je písmeno - ____. | |

3. Doplňte vhodné sloveso ve správném tvaru.

koupit, muset, platit, sedět, vidět, zkoušet

(já) _____ si letenku na Floridu.

(ty) _____ jít k lékaři.

Učitelka _____ žáky z matematiky.

My jako organizátoři _____ honorář kapele.

V jaké lavici (vy) _____ ?

Všichni _____, že mu je špatně.

Postup práce s korpusem

1. V aplikaci Morfio vyplňte pole podle obrázku. Pro obě první pole vyberte charakteristiku „společný“, protože kořen slova je v infinitivu i konjugované formě stejný. Ponechte zde výchozí nastavení (.+), protože nezáleží na tom, jak budou hledaná slovesa vypadat. Pro druhé pole – pole pro formant vyberte charakteristiku „odlišný“. V prvním vzoru vyplňujeme zakončení infinitivu. Pro tento typ časování je to vždy vokál+t (až na výjimky, jako je jíst, o které nám nyní nejde). Proto vyberte rozbalením druhého pole pro vzor 1 „samohlásky“ a po zobrazení další nabídky „všechny“ a dopište t. Do políčka pro vzor 2 napište *ím*, tedy osobní koncovku první osoby singuláru. Hledají se tvary, protože nám jde o tvary slovesa, nikoli o lemmata. Dále klikněte na Hledat.



The screenshot shows the Morfio application interface. At the top, the word "Morfio" is displayed in a stylized font. Below it, there are two dropdown menus for selecting characteristics: "společný" (selected) and "odlišný" (selected). To the right, there is a section for "Morf. specifikace:" with two rows of input fields. The first row has "vzor 1:" with a dropdown set to ".+", a text input containing "[:Vok:]t", and a dropdown set to "slovesa" with a "V.*" field. The second row has "vzor 2:" with a dropdown set to ".+", a text input containing "ím", and a dropdown set to "slovesa" with a "V.*" field. Below these are fields for "Další vzor", "Korpus:" (set to "SYN2015"), "Frekvence vyšší než:" (set to "10"), "Hledají se:" (set to "tvary"), and "Vyhodnocují se:" (set to "tvary"). There is also a checkbox for "Velikost písmen:" (checked, "ignorovat") and a text input for "Alternace". At the bottom, there are buttons for "Hledat", "Nové zadání", a URL "Odkaz na toto zadání: http://morfio.korpus.cz/b1gzniGMJ", and "Nápověda".

Obrázek 8o

2. Pro zobrazení frekvenčního seznamu nejčastějších sloves zvolte záložku Výpis a seznam sloves seřaďte podle frekvence. Jak je patrné ze seznamu, musíte výsledky protřídit, jelikož Morfio neumí přiřadit infinitiv k tvaru podle významu. Po vytřídění chybných údajů získáte představu o nejfrekventovanějších slovesech, která poté můžete využít pro tvorbu cvičení.

| Souhrn | | Výpis | Produktivita | vzor 1 | vzor 2 |
|--------|---------------------------|--------------------------|--------------|--------------------------|--|
| | | vzor 1 ▲▼ (fq ▲▼) | | vzor 2 ▲▼ (fq ▲▼) | V tabulce jsou uvedeny lišit pouze v případě přetřídění podle libovolné úkázce konkordancí 'Páry vytvořené až dle alternačních pravidel tabulky. |
| 1 | dělat dělit (23433) | dělím (43) | | vidím (7873) | |
| 2 | vidět (21638) | vidím (7873) | | jím (30) | |
| 3 | jít jít jet ... (20954) | jím (30) | | dím (15) | |
| 4 | dít dát (14221) | dím (15) | | mluvím (1770) | |
| 5 | mluvit (12458) | mluvím (1770) | | najím (45) | |
| 6 | naját najít najet (12418) | najím (45) | | musím (17037) | |
| 7 | musit muset (12060) | musím (17037) | | představím (281) | |
| 8 | představit (9867) | představím (281) | | vrátím (1882) | |
| 9 | vrátit (8850) | vrátím (1882) | | slyším (2601) | |
| 10 | slyšet (7557) | slyším (2601) | | hledím (207) | |
| 11 | hledět hledat (7350) | hledím (207) | | změním (112) | |
| 12 | změnit (6729) | změním (112) | | chodím (1505) | |
| 13 | chodit (6679) | chodím (1505) | | vytvořím (56) | |
| 14 | vytvořit (6585) | vytvořím (56) | | řeším (169) | |
| 15 | řešit (6089) | řeším (169) | | koupím (592) | |
| 16 | koupat koupit (5965) | koupím (592) | | myslím (24971) | |
| 17 | myslet myslit (5912) | myslím (24971) | | platím (260) | |
| 18 | platit (5452) | platím (260) | | vysvětlím (431) | |
| 19 | vysvětlit (5316) | vysvětlím (431) | | spím (462) | |
| 20 | spět spat spát (5307) | spím (462) | | | |

Obrázek 81

Příkladové věty nejlépe najdete přes rozhraní KonText pomocí funkce Slovní tvar, která vám umožní hledat přímo zadaný slovní tvar.

The screenshot shows the KonText search interface. At the top, there are navigation tabs: 'Dotaz', 'Korpusy', 'Uložit', 'Konkordance', 'Filtr', 'Frekvence', 'Kolokače', and 'Zobrazení'. Below this, the current corpus is 'syn2015'. The main search area is titled 'Hledat v korpusu'. It includes a search bar with the text 'rozumím', a dropdown menu for the corpus ('syn2015 / --celý korpus--'), and a search type dropdown set to 'Slovní tvar'. There are also links for 'Klíčesnice' and 'Předchozí dotazy'. Below the search bar, there are options for 'Slovní druh' (set to 'Nespecifikováno') and 'Shoda velikosti písmen' (checked). There are two expandable sections: 'Specifikovat kontext' and 'Omezit hledání'. At the bottom left, there is a blue 'Hledat' button.

Obrázek 82

Nalezené věty je třeba přizpůsobit jazykové úrovni žáků.

Řešení

1. rozumět, vrátit se, představit se, slyšet, vidět, mluvit
2. Osobní koncovky jsou zjednodušené, nepočítá se zde s variantou *rozumějí*.

| Slovesa s infinitivní koncovkou -it , -et a -ět mají osobní koncovky: | |
|--|---|
| já | - <u>ím</u> (např. <u>rozumím</u>) |
| ty | - <u>íš</u> (např. <u>vrátíš se</u>) |
| on/ona/ono | - <u>í</u> (např. <u>představí se</u>) |
| my | - <u>íme</u> (např. <u>slyšíme</u>) |
| vy | - <u>íte</u> (např. <u>vidíte</u>) |
| oni | - <u>í</u> (např. <u>mluví</u>) |
| Všude je písmeno -í . | |

3. koupím - musíš - zkouší - platíme - sedíte - vidí

Pracovní list č. 16:

KONJUGACE – Slovesa čtvrté slovesné třídy II.

Anotace

Cvičení tohoto typu je vhodné zařadit pro názornou ukázkou konjugace sloves. Žáci si díky nim lépe uvědomí, že lze konjugační typ odvodit podle zakončení infinitivu. V tomto pracovním listu bylo jako ilustrační příklad zvoleno sloveso *mluvit*, lze ale samozřejmě použít jakékoliv jiné sloveso s typickým zakončením.

Postup práce

Před obdržením pracovního listu by měli žáci ve vhodném textu najít několik sloves jednoho konjugačního typu, o který poté půjde při plnění úkolů. Sloves by si měli sami všimnout, pokud je to nutné, mohou je vyčasovat. Zde lze využít např. slovesa *mluvit*, *rozumět*, *myslet*. Žáci by měli vědět, že k tomuto typu patří slovesa, která jsou v infinitivu zakončená na *-it*, *-et* nebo *-ět* (viz dotaz v prvním úkolu). Sami by měli být schopní vymyslet další příklady.

Před vlastní prací s korpusem se žáci mohou pokusit odhadnout, která slovesa tohoto typu budou nejfrekventovanější. Učitel si může předem připravit obrázky ilustrující deset nejčastějších sloves (viz řešení) a žáci mohou hádat, o která slovesa jde. Následně dle svých schopností sami, v párech nebo v plénu použijí aplikaci Morfio a ověří si své domněnky. Je možné, že některá slovesa nebudou znát. Učitel by měl po zobrazení výsledků ověřit porozumění např. tím, že žákům opět ukáže obrázky a zeptá se jich, co člověk na obrázku dělá. Žáci jako odpověď musí používat slovesa, která našli.

Je vhodné, aby po vlastní práci s korpusem žáci slovesa použili např. při psaní krátkého příběhu a/nebo aby si je procvičili hrou „aktivity“, během níž jeden žák sloveso namaluje, vysvětlí nebo předvede a ostatní hádají, o které sloveso se jedná.

Pro rychlé a zvědavé žáky vyšší úrovně je možné využít třetí úkol, během něž mají porovnat produktivitu tohoto vzoru se vzorem *pít*. Zde může učitel také upozornit na to, že slovesa s dlouhým *-í* v infinitivní koncovce patří k jinému typu.

Pokud žáky práce zaujme, mohou hledat také zbývající slovesa stejného konjugačního typu, slovesa dalších typů (*dělat*, *kupovat*) nebo slovesa „nepravidelná“ (*plavat* – *plavu...*).

1. Jaký tvar má sloveso „mluvit“ pro „já“? Proč? Jaká další slovesa mají stejnou (identickou) konjugaci?
2. Která slovesa se stejnou konjugací jsou asi nejčastější? Diskutujte ve skupinách. Potom použijte Morfio a najděte deset nejčastějších sloves. Napište si je a vymyslete příběh, ve kterém použijete minimálně pět sloves.

3. Úkol pro rychlé: Kolik sloves má v korpusu stejnou konjugaci jako *mluvit (já mluvím)* a kolik jako *pít (já piju)*?

Postup práce s korpusem.

1. Postup práce je stejný jako u předchozího pracovního listu. V aplikaci Morfio vyplňte pole podle obrázku. Pro obě první pole vyberte charakteristiku „společný“, protože kořen slova je v infinitivu i konjugované formě stejný. Ponechte zde výchozí nastavení (.+), protože nezáleží na tom, jak budou hledaná slovesa vypadat. Pro druhé pole – pole pro formant – vyberte charakteristiku „odlišný“. V prvním vzoru vyplňujeme zakončení infinitivu. Pro tento typ časování je to vždy vokál+t (až na výjimky, jako je jíst, o které nám nyní nejde). Proto vyberte rozbalením druhého pole pro vzor 1 „samohlásky“ a po zobrazení další nabídky „všechny“ a dopište *t*. Do políčka pro vzor 2 napište *ím*, tedy osobní koncovku první osoby singuláru. Hledají se tvary, protože nám jde o tvary slovesa, nikoli o lemmata. Dále klikněte na Hledat.
2. Pro zobrazení sloves tohoto vzoru zvolte Výpis. Zobrazená slovesa následně seřadte podle frekvence kliknutím na šipku směřující dolů u veličiny fq. Tak získáte deset nejčastějších (viz řešení).
3. Pro řešení doplňujícího úkolu (*pít* – *piju*) postupujte analogicky. Do políček pro odlišný formant tentokrát zadejte *ít* a *iju*. Poté porovnejte počet sloves obou typů.

Řešení (na základě korpusu SYN2015)

1. *Mluvím*, protože všechna slovesa, jejichž infinitiv končí na *-et*, *-ět*, *-it* + slovesa *jíst*, *spát*, *stát* mají osobní koncovky *-ím*, *-íš*, *-í*, *-íme*, *-íte*, *-í*.
2. *myslet*, *vědět*, *muset*, *prosít*, *vidět*, *cítit*, *věřit*, *snažit*, *Slyšet*, *vrátit*
3. Stejnou konjugaci jako *mluvit* má asi 800 sloves, jako *pít* 21. Ve výsledcích se objeví 839, respektive 23 sloves, u některých ovšem najdeme i negativní formy. Přesto je zřejmé, že první typ je výrazně produktivnější než druhý. Když se na výsledky druhého hledání podíváte blíže, zjistíte, že se zde objevuje sedm prefigovaných sloves s kořenem *žit* (dle frekvence: *využít*, *použít*, *přežít*, *užít*, *zažít*, *prožít*, *dožít*). Další slovesa jsou v pořadí dle frekvence *zabít*, *pít*, *napít*, *vypít*, *bít*, *rozbít*, *nalít*, *šít*, *opít*, *zašít*, *lít*, *dopít*.

Pracovní list č. 17:

TVORBA SUBSTANTIV SUFIXACÍ – Myslivec dnes ulovil velkého jelena.

Byl to obrovský úlovek!

Anotace

Cvičení tohoto typu je vhodné zařadit pro znázornění různých slovtvorných modelů. Žáci si díky nim lépe uvědomí, že lze od známých slov odvozovat nová slova pomocí předpon a přípon. Cvičení lze libovolně variovat, je možné hledat např. jen slova odvozená předponami (např. *chod* – *vchod*) nebo naopak jen slova odvozená příponami (např. *školit* – *školitel*). V tomto pracovním listu půjde o slova utvořená ze sloves předponou *ú-* a příponou *-ek* (např. *úlovek*).

Postup práce

Učitel by se měl nejdřív ujistit, že žáci rozumí slovu, které bude sloužit jako model. V našem případě je to slovo *úlovek*. Pokud ho žáci neznají, může se učitel na základě příkladové věty v pracovním listu nebo obrázku s myslivcem lovícím jelena zeptat: Co dělal myslivce? Žáci by se měli dostat k odpovědi: Lovil jelena. Poté se učitel opět na základě věty nebo dalšího obrázku zeptá, jaký byl výsledek lovu – ulovil jelena. Výsledek je tedy *úlovek*. Učitel se žáků dále zeptá, na které části lze toto slovo rozdělit, čímž přejde k tematice slovtvorby. Žáci by měli přijít na kořen *lov*, předponu *ú* a příponu *ek*. Pokud neznají názvosloví, může je s ním učitel tímto způsobem seznámit. Následně se učitel žáků zeptá, zda znají další slova, která vznikla stejným způsobem. Učitel si předem může připravit obrázky nebo věty s vynechanými slovy, která žáci hledají (viz řešení), a žáci se mohou sami pokusit zjistit, o která slova se jedná. Pravděpodobně se jim nepodaří určit všechna slova, a to by mělo sloužit jako motivace pro práci s korpusem. Žáci se sami nebo s učitelem podívají do aplikace Morfio a zjistí, která další slova jsou odvozena stejným způsobem. Je vhodné, aby tato slova poté použili např. v krátkém příběhu.

Pokud žáky práce s korpusem zaujme, mohou následně ve skupinách nebo za domácí úkol vyhledávat slova vzniklá na základě jiného slovtvorného modelu (např. jména činitelská, konatelská, obyvatelská atd.).

**MYSLIVEC DNES ULOVIL VELKÉHO JELENA. BYL TO
OBROVSKÝ ÚLOVEK!**

- 1. Co znamená slovo „úlovek“? Jak vzniklo?**
- 2. Které části má toto slovo?**
- 3. Znáte nějaké další slovo se stejnou předponou a příponou?**
- 4. Použijte aplikaci Morfio a najděte další slova se stejnou předponou a příponou. Potom napište krátký text, ve kterém alespoň tři slova použijete.**

Postup práce s korpusem

Do aplikace si přidejte na první místo u obou vzorů ještě jedno pole pro odlišný formant kliknutím na <+. Toto pole ponechte u prvního vzoru prázdné, u druhého do něj vyplňte *ú*. U druhých polí ponechte charakteristiku „společný“ i výchozí nastavení (.+). Do posledního pole pro odlišný morfém napište u prvního vzoru *it*, do druhého *ek*. U prvního vzoru zvolte morfologickou specifikaci slovesa, u druhého podstatná jména. Vyberte, že se hledají a vyhodnocují lemmata. Velikost písmen můžete ignorovat. Nakonec klikněte na tlačítko Hledat.

The screenshot shows the Morfio application interface. At the top, the word "Morfio" is displayed in a colorful font. Below it, there are several input fields and dropdown menus for configuring a search. The search parameters are as follows:

- Left dropdown: <- x odlišný
- Middle dropdown: x společný
- Right dropdown: x odlišný
- Morf. specifikace: >>
- vzor 1: [] + [] e [] slovesa v
- vzor 2: [] + [] ek [] podstatná jména j
- Další vzor: []
- Korpus: SYN2015 Frázeová výřky nel: 10 Hledat se: lemmata Vyhodnocují se: lemmata
- Velikost písmen: ignorovat + Alternace
- Buttons: Hledat, Návě zastavit, Odkaz na toto zadání: <http://morfio.korpus.cz/wf/mrywz>, Nápověda

Below the search parameters, there is a table with the following columns: Souhrn, Výpis, Produktivita, vzor 1, vzor 2.

| vzor 1 | [q **] | vzor 2 | [q **] |
|--------|------------|---------------|--------|
| 1 | ovr (913) | účinek (7532) | |
| 2 | lov (152) | úlovek (996) | |
| 3 | lov (1932) | úlovek (910) | |
| 4 | ovr (2853) | účinek (1179) | |
| 5 | rad (381) | úradek (44) | |
| 6 | lov (431) | úlovek (552) | |
| 7 | lov (187) | úlovek (115) | |

Below the table, there is a text box with the following text:

V tabulce jsou uvedeny všechny doklady ze všech vzorů, které vstupují do zadaného modelu. Červená čára označuje společnou část (a se může lišit podle v případě aplikace alternací). V závorkách uvedeny jsou přídatky se stejnou frekvencí jednotky ve zvoleném korpusu. Tabulku je možné přefiltrovat podle kvantitativního součinu a to jak abecedně, tak frekvencí pomocí šipek v záhlaví tabulky. Každá šipka zároveň funguje jako odkaz směřující k úloze konkordanci ve zvoleném korpusu.

Pro vyřazení až tří aplikací alternativních pravidel jsou zvýrazněné **paronymní spojení**. Jejich báze se proto budou lišit. V případě, že použitím alternativních pravidel dojde k situaci, že jednému vzoru v dané dvojici odpovídá více slov, jsou všechna tato slova uvedena formálně v jednom řádku tabulky.

Obrázek 83

Řešení (na základě korpusu SYN2015)

1. Je to výsledek lovu, zvíře, které jsme ulovili nebo přeneseně cokoliv, co jsme získali např. koupí/výměnou (Ta vzácná známka, kterou jsem koupil od toho starého sběratele, je skvělý úlovek.) Vzniklo od slovesa *ulovit* příponou *-ek*. To zase vzniklo od slovesa *lovit* předponou *-u*.
2. Předponu *ú-*, kořen *lov* a příponu *-ek*.
3. viz 4)
4. činit - účinek, lomit - úlomek, platit - úplatek, radit - úradek, šklebit - úšklebek, tulit - útulek

Jak je z výsledků analýzy patrné, je potřeba provést ještě interpretaci výsledků. Nedá se očekávat, že by slovo *útulek* významově vzniklo ze slovesa *tulit*, formálně ovšem souvislost má.

Pracovní list č. 18:

TVORBA SUBSTANTIV PREFIXACÍ – Vchod, východ, přechod a záchod.

Anotace

Cvičení tohoto typu je vhodné zařadit pro znázornění různých slovtvorných modelů. Žáci si díky nim lépe uvědomí, že lze od známých slov odvozovat nová slova pomocí předpon. Cvičení lze libovolně variovat, je možné hledat např. slova odvozená příponami (např. *cestovat* – *cestovatel*) nebo slova odvozená předponami i příponami (např. *lov* – *úlovek*). V tomto pracovním listu byla zvolena tvorba slov prefixy *v-*, *vý-*, *pře-*, *zá-*, které se všechny mohou pojit s kořenem *chod*.

Postup práce

Tento pracovní list slouží jako inspirace k hledání odpovědi na časté otázky žáků, kteří se ptají, zda se mohou vybrané prefixy pojit s určitými kořeny. Učitel díky němu může také uvést termíny *prefix* a *kořen*.

Na začátku by měli žáci např. pomocí obrázků určit význam slov odvozených od jednoho kořene různými prefixy (zde *vchod*, *východ*, *přechod*, *záchod*). Obrázky si může učitel připravit předem, žáci je také mohou namalovat sami nebo mohou význam slov vysvětlit. Učitel se žáků následně zeptá, co mají slova společného a v čem se liší. Žáci by měli sami přijít na to, že slova mají stejný kořen a liší se svými prefixy. Je vhodné dobrat se také ke slovesu, od kterého byla substantiva odvozena. Jde o sloveso *chodit*, které se stejně jako další slovesa pohybu pojí s více prefixy.

Následně učitel žáky vyzve, aby našli další slova, která se mohou pojit s daným prefixy, případně s prefixy jinými. Právě slovesa pohybu jsou pro tento účel velmi vhodná, jelikož lze změnu významu prefixací názorně ukázat na obrázcích. Stačí, když žáci vyjmenují několik příkladů a uvědomí si princip spojování prefixů s kořeny. Pokud žáci neznají termíny *prefix* nebo *předpona* a *kořen*, může je s nimi učitel tímto způsobem seznámit.

Ve čtvrtém cvičení si pak žáci pomocí aplikace Morfio mohou sami vyzkoušet, jak prefixace funguje. Pokud je práce s aplikací zaujme, mohou si vymyslet další prefixy a zjistit, se kterými slovy se pojí. Učitel by se měl předem ujistit, že si žáci zvolili prefixy, které se všechny alespoň s jedním slovem pojit mohou. Mezi výsledky (viz řešení) se ovšem budou objevovat i méně frekventovaná slova, proto je třeba počítat s tím, že žáci některá slova nebudou znát. Učiteli mohou s vysvětlením významu slov opět pomocí obrázky.

1. Co znamenají tato slova?

vchod, východ, přechod, záchod

2. Co mají všechna slova společného? V čem se liší?

3. Která další slova se mohou podobně měnit? Řekněte příklady.

4. Podívejte se do aplikace Morfio a najděte slova, která mohou mít stejné prefixy jako slova ve cvičení 1. Která znáte? Co znamenají?

Postup práce s korpusem

V aplikaci si navolte místo pro čtyři vzory. Prefixy zadejte do polí v prvním sloupci, která si nastavte pro odlišný formant. U druhých polí zvolte charakteristiku „společný“ a ponechte v nich výchozí nastavení. Jako morfologickou specifikaci u všech vzorů nastavte podstatné jméno. Hledají se lemmata. Můžete ignorovat velikost písmen. Pro zobrazení frekvenčního seznamu klikněte na záložku Výpis.



Obrázek 84

Řešení

2. Společný je kořen *chod*, liší se prefixy.
3. Např. slovesa pohybu (*letět, jet, nést, vézt*). Je třeba upozornit, že u substantiv se v prefixu mění délka (*vyletět-výlet, zaletět-zálet*).
4. východ, vchod, přechod, záchod; výjezd, vjezd, přejezd, zájezd; výklad, vklad, překlad, základ; výlet, vlet, přelet, zálet; výložka, vložka, přeložka, záložka; výstup, vstup, přestup, zástup

Pracovní list č. 19:

TVORBA SUBSTANTIV SUFIXACÍ – Chci být slavný cestovatel!

Anotace

Cvičení tohoto typu je vhodné zařadit pro znázornění různých slovotvorných modelů. Žáci si díky nim lépe uvědomí, že lze od známých slov odvozovat nová slova pomocí přípon. Cvičení lze libovolně variovat, je možné hledat např. slova odvozená předponami (např. *chod* – *vchod*) nebo slova odvozená předponami i příponami (např. *lov* – *úlovek*). V tomto pracovním listu půjde o jména činitelská (např. *cestovatel*).

Postup práce

Učitel by se měl nejdřív ujistit, že žáci rozumí slovu, které bude sloužit jako model. V našem případě je to slovo *cestovatel*. Pokud ho žáci neznají, může si učitel připravit obrázky a krátké příběhy známých českých cestovatelů (např. Zikmund a Hanzelka), pomocí nichž význam žákům osvětlí. Učitel se žáků dále zeptá, jak toto slovo vzniklo. Žáci by měli přijít na to, že je odvozené od slovesa *cestovat* příponou *-tel*. Pokud žáci neznají názvosloví, může je s ním učitel tímto způsobem seznámit. Následně se učitel žáků zeptá, zda znají další slova, která vznikla stejným způsobem. Učitel si předem může připravit obrázky nebo věty s vynechanými slovy, která žáci hledají (viz řešení), a žáci se mohou sami pokusit zjistit, o která slova se jedná. Pravděpodobně se jim nepodaří určit všechna slova, a to by mělo sloužit jako motivace pro práci s korpusem. Žáci se sami nebo s učitelem podívají do aplikace Morfio a zjistí, která další slova jsou odvozena stejným způsobem. Výsledky je vhodné projít předem a označit barevně nebo vybrat jen substantiva, která jsou frekventovaná a pro žáky užitečná (viz řešení). Tato slova pak může učitel promítnout a žáci si z nich vyberou určitý počet a vysvětlí jejich význam. Ostatní žáci hádají, o které slovo se jedná. Je vhodné, aby tato slova poté použili např. v krátkém příběhu.

Pokud žáky práce s korpusem zaujme, mohou následně ve skupinách nebo za domácí úkol vyhledávat slova vzniklá na základě jiného slovotvorného modelu (např. jména konatelská, obyvatelská atd.).

CHCI BÝT SLAVNÝ CESTOVATEL!

1. Znáte nějaké slavné cestovatele?
2. Jak slovo *cestovatel* vzniklo?
3. Které má části?
4. Znáte další slovesa se stejnou příponou?
5. Použijte aplikaci Morfio a najděte další slova se stejnou příponou. Vyberte si pět slov a vysvětlete jejich význam. Ostatní spolužáci musí hádat, která slova to jsou.
6. Potom napište krátký text, kde alespoň tři slova použijete.

Postup práce s korpusem

V aplikaci ponechte v prvním poli pro společnou bázi .+. U prvního vzoru doplňte do pole pro odlišný formant *t* a zvolte morfologickou specifikaci slovesa. U druhého vzoru doplňte formant *tel* a zvolte morfologickou specifikaci podstatná jména. Vyberte, že se hledají a vyhodnocují lemmata. Velikost písmen můžete ignorovat. Nakonec klikněte na tlačítko Hledat. Pro zobrazení seznamu slov zvolte Výpis.



Obrázek 85

Řešení

1. Dle znalostí skupiny a zájmu učitele.
2. Od slovesa *cestovat*.
3. Kořen *cestovat* a příponu *-tel*.
4. např. *cvičitel, chovatel, krotitel, léčitel, luštitel, obdivovatel, objevitel, odesílatel, opatrovatel, osvoboditel, ošetřovatel, pečovatel, pěstitel, podnikatel, pronajímatel, provozovatel, učitel, ředitel, školitel, ubytovatel...*

Jak je patrné z výsledků analýzy, je potřeba provést ještě interpretaci výsledků. Například slova *přítel* a *jetel*, která se objevují mezi výsledky, určitě nevznikla ze sloves *přít* a *jet*, *ředitel* není odvozen od slovesa *ředit*. Slova *činitel* a *dělitel* nejsou jména činitelská.

Pracovní list č. 20:

KLÍČOVÁ SLOVA - Charakteristika

Anotace

Cvičení tohoto typu pomáhají žákům při psaní textů určitého žánru nebo zadání. Pomocí srovnání příslušných textů (zde charakteristik) a vyváženého referenčního korpusu, který obsahuje texty všech žánrů, mohou žáci sami zjistit, která slova jsou pro dané texty typická. Tato slova poté mohou vědomě použít při psaní vlastního textu. Lze vytvořit také offline variantu, kdy si učitel klíčová slova najde předem a žákům je pouze prezentuje. Třetí úkol z tohoto pracovního listu by tak nevyžadoval práci s korpusem ve výuce, ostatní úkoly mohou žáci plnit stejným způsobem.

Postup práce

Pracovní list by měl být zařazen do výukového bloku, během kterého si mají žáci osvojit dovednosti potřebné k psaní textu určitého žánru. Tento pracovní list je věnován charakteristice, lze jej ovšem obměnit pro potřeby jakéhokoliv jiného žánru.

Než žáci začnou s pracovním listem pracovat, měli by se seznámit s tím, jak charakteristika vypadá. Na začátku výukového bloku může učitel pomocí brainstormingu zjistit, co si žáci pod pojmem *charakteristika* vybaví (úkol 1). Sami se mohou pokusit formulovat, co by měla dobrá charakteristika obsahovat a jak by měla být strukturována. Své domněnky by si poté měli ověřit na vhodně zvolené příkladové charakteristice, kterou by si měli přečíst a společně s učitelem okomentovat z hlediska použité slovní zásoby, struktury i obsahu (úkol 2). Okomentované příklady textů různých žánrů, které mají být žáci schopni napsat během maturity z češtiny, naleznete na příklad na stránce <http://www.statnimaturita-cestina.cz>.

Práce s korpusem může být zařazena na toto místo pro zpestření a jako alternativní způsob nacházení klíčových slov typických pro daný žánr. Učitel by se žáků nejprve měl zeptat, která klíčová slova nacházejí v příkladovém textu. Žáci by si je měli vyznačit a poté společně s učitelem diskutovat o tom, zda to jsou opravdu slova typická pro daný žánr. Dále je možné využít vlastní práci s korpusem. Učitel si předem připraví soubor textů daného žánru, které následně s žáky analyzuje pomocí aplikace KWords. Textů musí být několik, aby obsahovaly dostatek materiálu pro analýzu. Ideální je rozsah několika set až několika tisíc slov. Příliš dlouhé texty analyzuje pro-

gram moc dlouho.⁴⁹ Při vkládání je nutné odstranit informaci, odkud byly texty zkopírovány, jinak by se názvy stránek pravděpodobně objevily mezi klíčovými slovy. Žáci mohou samozřejmě najít texty sami během hodiny, v tom případě je ovšem třeba počítat s vyšší časovou náročností. Výhodou je naopak to, že se žáci naučí hledat zdroje inspirace a také budou během práce aktivnější. Alternativně může učitel shromáždit soubor prací žáků s češtinou jako mateřštinou. Na žáky s OMJ může působit motivačně, že analyzují texty svých spolužáků.

Analýzu mohou žáci provádět samostatně, jednodušší je ovšem práce v plénu. Výsledkem analýzy jsou texty s červeně vyznačenými klíčovými slovy. Tato slova lze také zobrazit ve formě frekvenčního seznamu (viz postup práce s korpusem). Žáci si mohou na výsledcích ověřit, zda se nalezená klíčová slova shodují s jejich představou o slovní zásobě typické pro daný žánr. V klíčovém slově se nejspíš objeví slova ze skupiny vzhled, vlastnosti, zájmy, činnosti a jména. Pravděpodobně žáci naleznou také časová určení. Tato slova by měli shromažďovat do skupin a říct, proč se v charakteristice objevují.

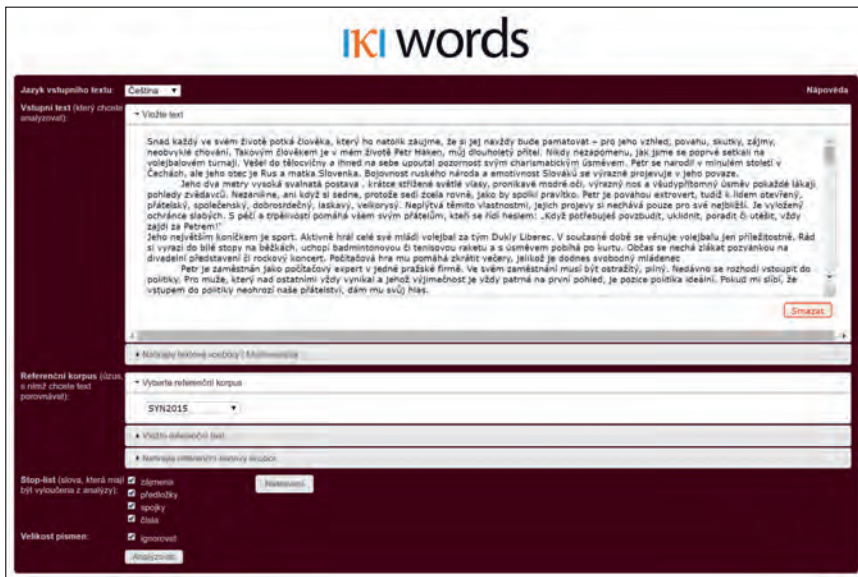
Aplikace umožňuje také zobrazit, jak se jednotlivá klíčová slova propojují (viz postup práce s korpusem).

Na analýzu a interpretaci výsledků by měla navázat vlastní produkce. Žáci by měli napsat vlastní charakteristiku a použít např. deset klíčových slov, která našli.

49 V tomto pracovním listu byly použity texty ze stránek <http://www.statnimaturita-cestina.cz/charakteristika> a <http://www.cesky-jazyk.cz/slohovky/charakteristiky/#axzz5IEY-wRMn2>.

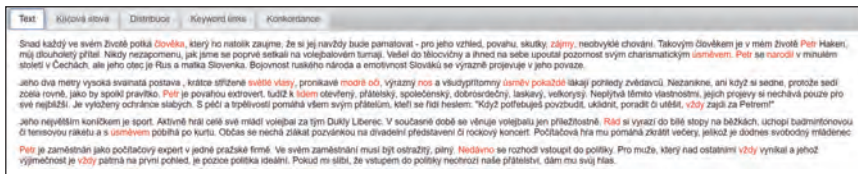
Postup práce s korpusem

1. Vstupní text, který chcete analyzovat (v tomto případě texty charakteristik v rozsahu alespoň 500 slov), zkopírujete do příslušného políčka aplikace KWords, které se rozbalí, když myší najedete na ikonu Vložte text. Jako referenční korpus ponechte syn2015. Pomocí stop-listu můžete z analýzy vyloučit všechna nabízená slova (zájmena, předložky, spojky, čísla). Ponechte zaškrtnuté Ignorovat velikost písma. Nakonec klikněte na Analyzovat.



Obrázek 86

2. Zobrazí se vám analyzovaný text s červeně vyznačenými klíčovými slovy, tedy se slovy, která se v analyzovaném textu vyskytují relativně častěji než v textech z referenčního korpusu.⁵⁰



Obrázek 87

⁵⁰ Příklad je text ze stránky <http://www.statnimatorita-cestina.cz/charakteristika>.

Řešení

Výsledky úkolů 2 a 3 závisí na zvolených vzorových textech. V textech zvolených pro výše uvedenou analýzu byla nalezena tato klíčová slova:

štíra, vážím, zábavný, studuje, Libor, náladu, hraní, světlé, humor, inteligentní, kluk, znamení, nos, zájmy, narodil, modré, školu, úsměvem, úsměv, vlasy, lidem, pokaždé, rád, nedávno, Tomáš, člověka, Petr, vždy, často, opravdu, patří, člověk, oči

V našem případě je na prvním místě slovo *štíra*, které odkazuje na znamení. Z toho lze usuzovat, že se v charakteristikách často mluví o znamení daného člověka. Na prvním místě je proto, že názvy znamení nejsou tak typické pro texty ostatních žánrů. Dalším slovem je *vážím*, jelikož zadání vybraných charakteristik znělo *Charakteristika člověka, kterého si vážím*. Ostatní slova lze shromáždit do skupin, např. *vlastnosti* (zábavný, inteligentní, humor), *vzhled* (světlé vlasy, nos) nebo *časová určení* (pokaždé, nedávno, často), která ukazují na to, že součástí charakteristiky bývá i vyprávění nějakého zážitku, který nás s daným člověkem pojí. Skupiny lze samozřejmě rozšiřovat o další příklady.

5.2.5 Treq

Pracovní list č. 21:

HOMONYMA – Každé ráno slyším, jak kokrhá náš kohoutek a maminka pouští vodu z kohoutku, aby uvařila čaj.

Anotace

Cvičení tohoto typu cílí na rozvoj slovní zásoby v rámci homonymie a polysemie. Žáci pracují s překladovými ekvivalenty ze svého mateřského jazyka nebo z jiného jazyka, který znají. Učitel může materiál z aplikace Treq využít také pro práci offline, a to tak, že nalezené věty prezentuje žákům a ti mají na jejich základě definovat možné významy slova. Následně mohou pomocí aplikace najít další příklady.

Postup práce

Učitel nejprve žáky nechá, aby si přečetli příkladovou větu. Zeptá se jich na význam stejných slov a zda znají další slova, která mají více významů. Poté se zeptá na konkrétní slova prezentovaná ve cvičení. Žáci následně hledají příkladové věty v aplikaci Treq samostatně nebo ve skupinách podle sdíleného dalšího jazyka. Tento pracovní list počítá s překlady do angličtiny. Počet významů se ovšem bude lišit podle cílového jazyka, učitel by proto měl být nejlépe s mateřským jazykem žáků obeznámen. Pokud tomu tak není, neměl by být počet významů stanoven předem. Pracovní list může být také využit pro domácí práci. Ve všech případech je třeba žáky seznámit s postupem práce s korpusem.

1. Přčtete si příkladovou větu. Co znamená slovo „kohoutek“?

Každé ráno slyším, jak kokrhá náš kohoutek a maminka pouští vodu z kohoutku, aby uvařila čaj.

2. Znáte další slova, která mohou mít víc významů?

3. Co mohou znamenat tato slova? Najděte pomocí aplikace Treq nejčastější překlady těchto slov. Napište ke každému překladu příkladovou větu v češtině.

• **los**

1) _____ : _____

2) _____ : _____

• **kolej**

1) _____ : _____

2) _____ : _____

3) _____ : _____

4) _____ : _____

• **topit**

1) _____ : _____

2) _____ : _____

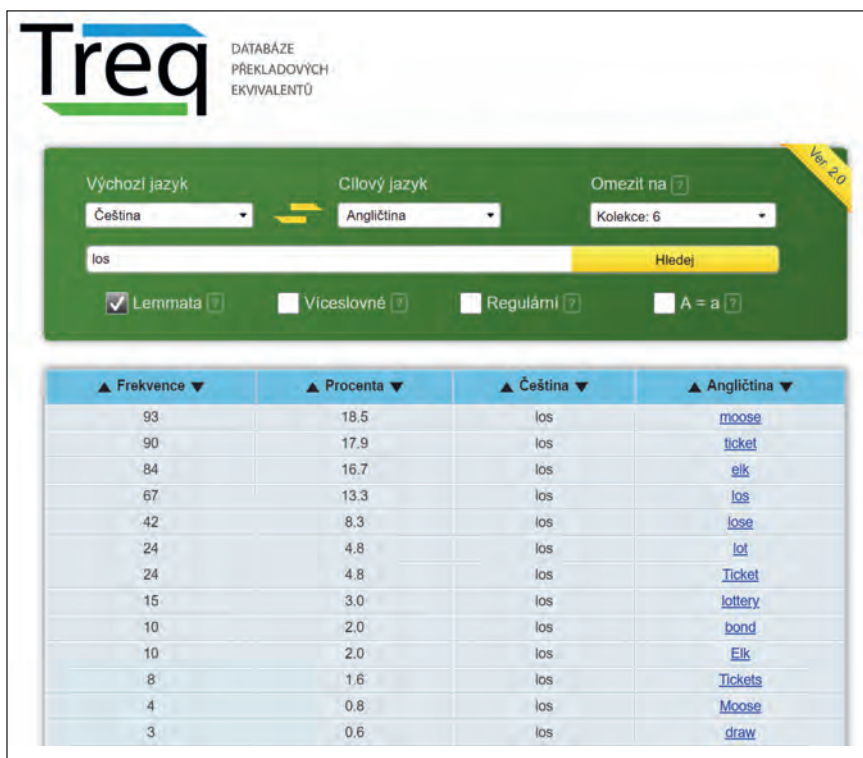
• **pas**

1) _____ : _____

2) _____ : _____

Postup práce s korpusem

Do aplikace Treq zadejte postupně všechna slova, zaškrtněte políčko Lemmata. Můžete ponechat zaškrtnuté a=A. Pro překlad si vyberte cílový jazyk, který se vám hodí. Klikněte na Hledat. Zobrazí se vám seznam překladových ekvivalentů. Vyberte si nejfrekventovanější. Počet překladových ekvivalentů si určete sami podle relevantnosti pro výuku. Kliknutím na překladový ekvivalent se vám zobrazí příkladové věty. Vyberte si ty, které jsou vhodné pro řešení. Ve výsledcích se mohou zobrazovat i nerelevantní výsledky, jelikož některé věty či slovní spojení zůstávají v překladu výjimečně v původním znění. Např. „Was geht **los**?“ otázal se – „WAS GEHT **LOS**, What's going on?“ , he asked.



The screenshot shows the Treq application interface. At the top left is the logo 'Treq' and the text 'DATABÁZE PŘEKLADOVÝCH EKVIVALENTŮ'. Below this is a search form with a green background. The form includes a 'Výchozí jazyk' dropdown set to 'Čeština', a 'Cílový jazyk' dropdown set to 'Angličtina', and an 'Omezit na' dropdown set to 'Kolekce: 6'. A search input field contains the word 'los' and a yellow 'Hledej' button. Below the search bar are four checkboxes: 'Lemmata' (checked), 'Víceslovné', 'Regulární', and 'A = a'. A yellow 'Ver 2.0' badge is in the top right corner of the form.

| ▲ Frekvence ▼ | ▲ Procenta ▼ | ▲ Čeština ▼ | ▲ Angličtina ▼ |
|---------------|--------------|-------------|-------------------------|
| 93 | 18.5 | los | moose |
| 90 | 17.9 | los | ticket |
| 84 | 16.7 | los | elk |
| 67 | 13.3 | los | los |
| 42 | 8.3 | los | lose |
| 24 | 4.8 | los | lot |
| 24 | 4.8 | los | Ticket |
| 15 | 3.0 | los | lottery |
| 10 | 2.0 | los | bond |
| 10 | 2.0 | los | Elk |
| 8 | 1.6 | los | Tickets |
| 4 | 0.8 | los | Moose |
| 3 | 0.6 | los | draw |

Obrázek 90

Řešení (například)

los:

moose: Lovci pronásledovali **losy**, buvoly, mamuty, a divoké polární soby zvané karibu.

ticket: Kupte si **los**, vyhrajte farmu v Africe!

kolej:

track: Na druhou **kolej** přijíždí osobní vlak do Berlína!

college: Kolik studentů navštěvuje zdejší **kolej**?

rail: Ustupte od **kolejí**, vlak právě odjíždí!

dorm: Dnes jsem se přestěhoval na **kolej**.

topit:

drown: **Topí se** a stáhla by vás s sebou na dno.

heat: **Topili** jsme dřevem, bylo naším zdrojem tepla a světla.

pas:

passport: Nemám ani cestovní **pas**.

waist: Popadl ji kolem **pasu**.

Pracovní list č. 22:

IDIOMY – Mám pocit, že snad vyletím z kůže, jaký mám vztek.

Anotace

Cvičení tohoto typu cílí na porozumění idiomům a jejich vhodné užití. Hodí se pro pokročilé žáky, kteří už většinou nemají potíže se základním významem slov. Žáci ve cvičení z tohoto pracovního listu aktivně pomocí korpusu vyhledávají překlad idiomů, pomocí kontextu by pak měli být schopni doplnit idiomy do textu. Jelikož pracují s překladem do mateřského jazyka, mohou v následné diskuzi porovnávat jazykové obrazy světa v různých jazycích, a zlepšovat si tak svoji interkulturní kompetenci.

Postup práce

Před samotnou prací s korpusem musejí být žáci seznámeni s existencí idiomů, nejlépe pomocí textu obsahujícího dostatečně názorné idiomy. Pokud učitel pracovní list nepoužije v návaznosti na text obsahující idiomy se somatismy, obrátí pozornost žáků k metaforickému užití slova *kůže* pomocí příkladových vět v prvním cvičení. Na jejich základě lze také žáky seznámit s pojmem metafora, pokud jim není znám. Učitel se následně zeptá, zda žáci znají další idiomy se slovem *kůže*. Příklady zapisuje na tabuli a nechá žáky uvést kontext užití.

Další idiomy se slovem *kůže* jsou ve třetím cvičení. Žáci samostatně nebo ve skupinách podle mateřského jazyka doplní znění idiomu ve svém jazyce. Tato fáze může proběhnout také jako domácí práce. Následuje diskuze nad idiomy v jednotlivých jazycích. Ve druhém cvičení si žáci poté prakticky vyzkouší, zda idiomům správe porozuměli a zda je dokáží vhodně užít.

1. Přčtete si texty a řekněte, co znamená vyletět z kůže a zachránit si kůži.

Poslyšte, já se naštvu málokdy, ale tohleto je vážně moc. Mám pocit, že snad vyletím z kůže, jaký mám vztek.

Za války někteří lidé dělali hrozné věci, jen aby si zachránili kůži: lhali, kradli nebo i zabíjeli.

2. Znáte další fráze, ve kterých používáme slovo kůže metaforicky?

3. Najděte pomoci Trequ překlady těchto idiomů do svého jazyka. Používá váš jazyk taky slovo kůže?

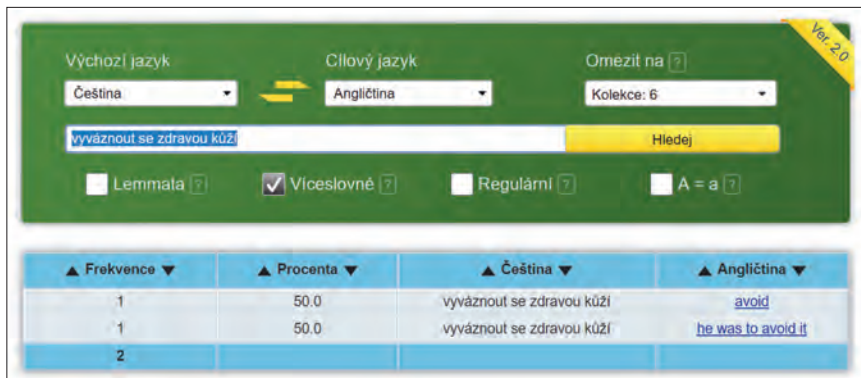
*vyváznout se zdravou kůží
kůže líná
být v něčí kůži
nebýt ve své kůži*

4. Doplňte správný idiom ze cvičení 3 podle kontextu.

1. Včera kolem osmé hodiny ranní u Chotěbuz: Dvě nabouraná osobní auta míjel ve směru od Českého Těšína na Karvinou polský kamion. Řidič kamionu při průjezdu kolem místa nehody snížil rychlost, vzápětí ale do něj zezadu narazila avia. Její řidič _____ (*minulý čas*), smůlu ale měl jeho spolujezdec († 39). „Lékaři i přes veškerou snahu mladšímu muži nedokázali pomoci. Svým zraněním na místě podlehl,“ uvedla mluvčí karvinské policie Zlatuše Viačková.
2. Řeknu vám, nechtěla bych _____ - tátu nemá a máma se o něho nestará! K čemu to povede? Co z něj vyroste?
3. Hned jak se vrátila, poznal, že _____. Ani neodpověděla na pozdrav, jen kývla hlavou.
4. No tak, _____, vstávat!

Postup práce s korpusem

Do aplikace Treq postupně napište všechny idiomy a zaškrtněte políčko Víceslovné. Zvolte si cílový jazyk. Po kliknutí na překladový ekvivalent se vám zobrazí příkladové věty. Vhodné vyberte pro druhé cvičení.



The screenshot shows the Treq application interface. At the top, there are dropdown menus for 'Výchozí jazyk' (set to 'Čeština') and 'Cílový jazyk' (set to 'Angličtina'). A yellow arrow points from the source language to the target language. To the right, there is a 'Omezit na' dropdown set to 'Kolekce: 6'. Below this is a search bar containing the text 'vyváznout se zdravou kůží' and a yellow 'Hledej' button. Under the search bar, there are four checkboxes: 'Lemmata' (unchecked), 'Víceslovné' (checked), 'Regulární' (unchecked), and 'A = a' (unchecked). Below the search bar is a table with the following data:

| ▲ Frekvence ▼ | ▲ Procenta ▼ | ▲ Čeština ▼ | ▲ Angličtina ▼ |
|---------------|--------------|---------------------------|------------------------------------|
| 1 | 50.0 | vyváznout se zdravou kůží | avoid |
| 1 | 50.0 | vyváznout se zdravou kůží | he was to avoid it |
| 2 | | | |

Obrázek 91

Řešení

1. dle mateřského jazyka žáků
2. vyváznul / vyvázl se zdravou kůží – být v jeho kůži – není ve své kůži – kůže líná

Použitá literatura

- Cvrček, V. a kol. (2010). *Mluvnice současné češtiny*. Praha: Karolinum.
- Čermák, F. – Schmiedtová, V. (2004): Český národní korpus – základní charakteristika a širší souvislosti. In: *Národní knihovna, knihovnická revue* 15, 2004, č. 3., str. 152–168.
- Čermák, F. (2011). Korpusy včera, dnes a zítra. In: *Korpusová lingvistika Praha 2011 2 Výzkum a výstavba korpusů* (eds. F. Čermák a kol.). Praha: Nakladatelství Lidové noviny.
- Český národní korpus – základní charakteristika a širší souvislosti. *Národní knihovna, knihovnická revue*, 15, 2004, č. 3., str. 152–168
- Ellis, R. – Barkhuizen, G. (2005). *Analysing Learner Language*. Oxford: Oxford University Press.
- Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In: *Corpora and Language Teaching* (ed. K. Aijmer). Amsterdam and Philadelphia: Benjamins, 2009, str. 13–32.
- Han, N.-R. et al. (2010). Using an Error-Anotating Learner Corpus to Develop an ESL/EFL Error Correction System. Dostupné online 13. 12. 2018: <https://www.cs.rochester.edu/~tetreaul/han-lrec10-final.pdf>
- McEnery, T. – Hardie, A. (2012). *Corpus Linguistics. Method, Theory and Practice*. Cambridge: Cambridge University Press.
- Sinclair, J. (Eagles, 1996). Preliminary recommendations on Corpus Typology. Dostupné z <http://www.ilc.pi.cnr.it/EAGLES96/corpusyp/corpusyp.html>.
- Slovník spisovného jazyka českého [online]. Dostupný z <http://ssjc.ujc.cas.cz/>
- Šebesta, K. – Škodová, S. (2012). *Čeština – cílový jazyk a korpusy*. Liberec: Technická univerzita v Liberci.
- Šebesta, K. (2010). Korpusy češtiny a osvojování jazyka. In: *Studie z aplikované lingvistiky / Studies in Applied Linguistics*, 2, str. 11–33.
- Štindlová, B. (2011). *Žákovský korpus češtiny a evaluace jeho chybové anotace*. Praha: FF UK.
- Šulc, M. (1999). *Korpusová lingvistika. První vstup*. Praha: Karolinum.
- Turton, N.D. and J.B. Heaton (1996). *Longman Dictionary of Common Errors*. Harwich: Longman.
- <http://akces.ff.cuni.cz/>
- <http://wiki.korpus.cz/doku.php/cnk:syn2015>
- <https://wiki.korpus.cz/doku.php/cnk:uvod>

<http://www.webcorp.org.uk/live/>

https://kontext.korpus.cz/first_form

<https://lindat.mff.cuni.cz/cs/>

Learner corpora around the world, dostupné 13. 12. 2018: <https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html>

<https://merlin-platform.eu/>

<http://www.cesky-jazyk.cz/slohovky/charakteristiky/#axzz5IEYwRMn2>

Seznam citovaných korpusů

AKCES 2 verze 2

s. 15

Šebesta, K. – Goláňová, H. et al. *AKCES 2 ver. 2*. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, 2013. Dostupný z WWW: <http://hdl.handle.net/11858/00-097C-0000-0023-3FBB-3>

AKCES 4

s. 15

Šebesta, K. – Bedřichová, Z. – Štindlová, B. et al. *AKCES 4*. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, 2012. Dostupný z WWW: <http://hdl.handle.net/11858/00-097C-0000-000C-2293-0>

AKCES 5

s. 15

Šebesta, K. – Bedřichová, Z. – Šormová, K. et al. *AKCES 5 (CzeSL-SGT)*. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, 2014. Dostupný z WWW: <http://hdl.handle.net/11858/00-097C-0000-0023-95B1-E>

American Heritage Intermediate Corpus

s. 13

M. Paikeday, T. (1973). *The American heritage intermediate corpus*.

Aranea

s. 14

Benko, V. *Srovnatelné webové korpusy Aranea*. Ústav Českého národního korpusu FF UK, Praha 2015. Dostupný z WWW: <http://www.korpus.cz>

BICCEL (Bilingual Corpus of Chinese English Learners)

s. 37

Wen, Q. – Wang, J. *Bilingual Corpus of Chinese English Learners*. National Research Center for Foreign Language Education, Beijing Foreign Studies University.

British Academic Spoken English (BASE)

s. 20

Nesi, H. – Thompson, P. *British Academic Spoken English*. University of Warwick, University of Reading, University of Oxford. Dostupný z WWW: <https://www.sketchengine.eu/british-academic-spoken-english-corpus/#toggle-id-1>

British Academic Written English (BAWE)

s. 20

Nesi, H. – Gardner, S. *British Academic Written English*. University of Warwick, University of Reading, University of Oxford. Dostupný z WWW: <https://www.sketchengine.eu/bawe-corpus/#toggle-id-1>

British National Corpus

s. 19, 23, 27, 39

The British National Corpus, version 3 (BNC XML Edition). 2007. Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. Dostupný z WWW: <http://www.natcorp.ox.ac.uk/>

Brown Corpus

s. 13, 35, 39

A Standard Corpus of Present-Day Edited American English, for use with Digital Computers (Brown). 1964, 1971, 1979. Compiled by W. N. Francis and H. Kučera. Brown University. Providence, Rhode Island. Dostupný z WWW: <https://www.sketchengine.eu/brown-corpus/#toggle-id-1>

Cambridge Learner Corpus

s. 18, 24, 25, 36, 37

Nicholls, D. *The Cambridge Learner Corpus - error coding and analysis for lexicography and ELT*. Diane Nicholls. Dostupný z WWW: <http://ucrel.lancs.ac.uk/publications/CL2003/papers/nicholls.pdf>

Corpus of Contemporary American English

s. 19

Davies, Mark. (2008-) *The Corpus of Contemporary American English (COCA): 560 million words, 1990-present*. Dostupný z WWW: <https://corpus.byu.edu/coca/>

CZESL-MAN

s. 15, 37, 39, 46

Rosen, A. *CzeSL-MAN – a corpus of non-native speakers' Czech with manual annotation*. Praha 2015. Dostupný z WWW: <http://utkl.ff.cuni.cz/~rosen/public/2015-cze-sl-man-en.pdf>

CZESL-PLAIN

s. 15, 39

Šebesta, K. – Bedřichová, Z. – Hana, J. – Hlaváčková, E. – Hnátková, M. – Hrdlička, M. – Janeš, P. – Jelínek, T. – Křen, M. – Lábus, V. – Lundáková, K. – Petkevič, V. – Pierscieniak, P. – Procházka, P. – Rosen, A. – Skoumalová, H. – Škodová, S. – Šormová, K. – Štindlová, B. *CZESL-PLAIN: akviziční korpus psané češtiny, zvl. přepisů písemných projevů nerodilých mluvčích, verze 2 z 22. 1. 2014*. Ústav Českého národního korpusu FF UK, Praha 2012. Dostupný z WWW: <http://www.korpus.cz.seznamcitovanychkorpusu> **175**

CZESL-SGT

s. 15, 37, 39, 46, 107, 109, 119

Šebesta, K. – Bedřichová, Z. – Šormová, K. – Štindlová, B. – Hrdlička, M. – Hrdličková, T. – Hana, J. – Petkevič, V. – Jelínek, T. – Škodová, S. – Poláčková, M. – Janeš, P. –

Lundáková, K. – Skoumalová, H. – Sládek, Š. – Pierscieniak, P. – Toufarová, D. – Richter, M. – Straka, M. – Rosen, A. *CzeSL-SGT: korpus češtiny nerodilých mluvčích s automaticky provedenou anotací, verze 2 z 28. 7. 2014*. Ústav Českého národního korpusu FF UK, Praha 2014. Dostupný z WWW: <http://www.korpus.cz>

CZESL-UD

s. 15

Hana, J. – Hladká, B. (2019). *Czsl - Universal Dependencies Release 0.5*, LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. Dostupný z WWW: <http://hdl.handle.net/11234/1-2927>

DIAKORP

s. 14

Kučera, K. – Řehořková, A. – Stluka, M. *DIAKORP: Diachronní korpus, verze 6 z 18. 12. 2015*. Ústav Českého národního korpusu FF UK, Praha 2015. Dostupný z WWW: <http://www.korpus.cz>

English Profile

s. 18, 25

University of Cambridge. *English Profile*. Dostupný z WWW: <https://www.english-profile.org/>

ESGCL

s. 41

Chen, H. – Wen, Q. – Li, A. *English Speech Corpus of Chinese Learners*. Nantong University, Beijing Foreign Studies University, Chinese Academy of Social Sciences.

HKUST

s. 36, 37

Gregory J., Language Centre, Hong Kong University of Science & Technology, Clear Water Bay, HK. See Milton, John – K.S.T Tong (eds.) (1991). *Text Analysis in Computer-Assisted Language Learning*. Hong Kong: Hong Kong University of Science & Technology.

CHILDES

s. 17

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk*. Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Chung-dahm Corpus

s. 36, 37

Han, N.-R. – Tetreault, J. – Lee, S.-H. – Ha, J.-Y. (b.r.). *Using an Error-Annotated Learner Corpus to Develop an ESL/EFL Error Correction System*. Language Resources and Evaluation, 2010, 8. Dostupný z WWW: <http://cs.rochester.edu/~tetreault/han-lrec10-final.pdf>

InterCorp

s. 14, 47, 70, 71, 72

Rosen, A. – Vavříň, M. – Zasina, A. J. *Korpus InterCorp – čeština, verze 11 z 19. 10. 2018*.

Ústav Českého národního korpusu FF UK, Praha 2018. Dostupný z WWW: <http://www.korpus.cz>

- International Corpus of Learner English (ICLE)** **s. 20**
Graner, S. – Dagneaux, E. – Meunier, F. (2002). *International Corpus of Learner English*. Louvain: UCL.
- Interactive Spoken Language Education (ISLE)** **s. 41**
Atwell, ES. – Howarth, PA. – Souter, DC. (2003). *The ISLE corpus: Italian and German spoken learner's English*. ICAME Journal: International Computer Archive of Modern and Medieval English Journal, 27. 5. 18.
- Japanese Learner English (NICT JLE)** **s. 37**
Izumi, E. – Uchimoto, K. – Hitoshi, I. (2004). The NICT JLE Corpus Exploiting the language learners' speech database for research and education. *International Journal of the Computer, the Internet and Management*. 12.
- KSK-dopisy** **s. 61**
Hladká, Z. *KSK-dopisy (Korpus soukromé korespondence): přepisy ručně psaných dopisů z let 1990–2004*. Ústav Českého národního korpusu FF UK, Praha 2006. Dostupný z WWW: <http://www.korpus.cz>
- KWords** **s. 60, 67, 68, 69, 70, 157, 159, 160**
Cvrček, V. – Vondříčka, P. (2013). *KWords*. FF UK, Praha. Dostupný z WWW: <http://kwords.korpus.cz>
- Lancaster-Oslo/Bergen Corpus** **s. 13**
Lancaster-Oslo-Bergen corpus of modern English: [tagged, horizontal format] Johansson, Stig, 1924–2010. University of Oslo. Institute of British and American Studies, Oslo: s.d
- Learning Prosody in a Foreign Language (LeaP)** **s. 41**
Dostupný z WWW: <https://sourceforge.net/projects/leapcorpus/>
- Learners' Corpus of Reading Texts** **s. 42**
Laboratoire parole et langage – UMR 7309 (LPL) (2010). *A Learners' Corpus of Reading Texts [Corpus]*. ORTOLANG (Open Resources and TOols for LANGuage) - www.ortolang.fr, <https://hdl.handle.net/11403/sldro00015/v2>.
- Longman Spoken and Written Corpus** **s. 30**
Biber, D. – Johansson, S. – Leech, G. – Conrad, S. – Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson, s. 24–35.

LOCNESS**s. 21**

Centre for English Corpus Linguistics. *Louvain Corpus of Native English Essays*. Dostupný z WWW: <https://uclouvain.be/en/research-institutes/ilc/cecl/locness.html>

MERLIN**s. 15**

Štindlová, B. – Čurdová V. (2015): MERLIN: Multilingvální platforma pro evropské referenční úrovně. In: *Časopis pro moderní filologii, dvojčíslo u příležitosti 20. výročí ČNK*. <http://cmf.ff.cuni.cz/?q=en> seznam citovaných korpusů **179**

Michigan Corpus of Academic Spoken English (MICASE)**s. 20**

Simpson, R. C. – Briggs, S. L. – Ovens, J. – Swales, J. M. (2002). *The Michigan Corpus of Academic Spoken English*. Ann Arbor, MI: The Regents of the University of Michigan.

Michigan Corpus of Upper-level Student Papers (MICUSP)**s. 20**

Michigan Corpus of Upper-level Student Papers. (2009). Ann Arbor, MI: The Regents of the University of Michigan.

Morfio**s. 60, 65, 66, 67, 136, 137, 142, 145, 147, 148, 149, 152, 154, 155**

Cvrček, V. – Vondříčka, P. (2013). *Morfio*. FF UK, Praha. Dostupný z WWW: <http://morfio.korpus.cz>

ORAL**s. 14, 38, 40**

Kopřivová, M. – Lukeš, D. – Komrsková, Z. – Poukarová, P. – Waclawičová, M. – Benešová, L. – Křen, M. *ORAL: korpus neformální mluvené češtiny, verze 1 z 2*. 6. 2017. Ústav Českého národního korpusu FF UK, Praha 2017. Dostupný z WWW: <http://www.korpus.cz>

ORAL2006**s. 38, 61**

Kopřivová, M. – Waclawičová, M. *ORAL2006: korpus neformální mluvené češtiny*. Ústav Českého národního korpusu FF UK, Praha 2006. Dostupný z WWW: <http://www.korpus.cz>

ORAL2008**s. 38, 61**

Waclawičová, M. – Kopřivová, M. – Křen, M. – Válková, L. *ORAL2008: sociolinguisticky vyvážený korpus neformální mluvené češtiny*. Ústav Českého národního korpusu FF UK, Praha 2008. Dostupný z WWW: <http://www.korpus.cz>

ORAL2013**s. 38, 61**

Benešová, L. – Křen, M. – Waclawičová, M. *ORAL2013: reprezentativní korpus neformální mluvené češtiny*. Ústav Českého národního korpusu FF UK, Praha 2013. Dostupný z WWW: <http://www.korpus.cz>

ORAL-Z**s. 38**

Kopřivová, M. – Lukeš, D. – Komrsková, Z. – Poukarová, P. – Waclawičová, M. – Benešová, L. – Křen, M.: *ORAL: korpus neformální mluvené češtiny, verze 1 z 2*. 6. 2017. Ústav Českého národního korpusu FF UK, Praha 2017. Dostupný z WWW: <http://www.korpus.cz>

ORTOFON**s. 14, 38, 40**

Kopřivová, M. – Komrsková, Z. – Lukeš, D. – Poukarová, P. – Škarpová, M. *ORTOFON: korpus neformální mluvené češtiny s víceúrovňovým přepisem, verze 1 z 2*. 6. 2017. Ústav Českého národního korpusu FF UK, Praha 2017. Dostupný z WWW: <http://www.korpus.cz>

Pražský závislostní korpus**s. 14, 45**

Bejček, E. – Hajičová, E. – Hajič, J. – Jínová, P. – Kettnerová, V. – Kolářová, V. – Mikulová, M. – Mírovský, J. – Nedoluzhko, A. – Panevová, J. – Poláková, L. – Ševčíková, M. – Štěpánek, J. – Zikánová, Š. *Prague Dependency Treebank 3*. o. Data/software, Univerzita Karlova v Praze, MFF, ÚFAL, Praha 2013. Dostupný z WWW: <http://ufal.mff.cuni.cz/pdt3.o/>

Panevová, J. (2017). PRAŽSKÝ ZÁVISLOSTNÍ KORPUS. In: Petr Karlík, Marek Nekula, Jana Pleskalová (eds.). *CzechEncy - Nový encyklopedický slovník češtiny*. Dostupné online: https://www.czechency.org/slovník/PRAŽSKÝ_ZÁVISLOSTNÍ_KORPUS

ROMi_1.0**s. 15, 42, 186**

Bedřichová, Z. – Šebesta, K. – Šormová, K. *ROMi – první rozsáhlá databanka romského etnolektu češtiny. Lidé města*, 13, 2011. Dostupný z <https://ufal.mff.cuni.cz/romi-10>

SCHOLA 2010**s. 15**

Šebesta, K. – Goláňová, H. – Křen, M. – Procházka, P. *SCHOLA2010: korpus mluvené češtiny ve škole – přepisy nahrávek vyučovacích hodin na českých základních a středních školách*. Ústav Českého národního korpusu FF UK, Praha 2010. Dostupný z WWW: <http://www.korpus.cz>

SKRIPT2012**s. 15, 42**

Šebesta, K. – Goláňová, H. – Jelínek, T. – Jelínková, B. – Křen, M. – Letafková, J. – Procházka, P. – Skoumalová, H.: *SKRIPT2012: akviziční korpus psané češtiny – přepisy písemných prací žáků základních a středních škol v ČR*. Ústav Českého národního korpusu FF UK, Praha 2013. Dostupný z WWW: <http://www.korpus.cz>

SKRIPT2015**s. 15**

Šebesta, K. – Kráčmarová, D. – Rosen, A. – Goláňová, H. – Hana, J. – Hlaváčková, E. – Hnátková M. – Hrdlička, M. – Janeš, P. – Jelínek, T. – Jelínková, B. – Křen, M. –

Lábus, V. – Letafková, J. – Loginová, K. – Petkevič, V. – Pierscieniak, P. – Skoumalová, H. – Škodová, S. – Štindlová, B. – Šormová, K. *SKRIPT2015: srovnávací korpus jazyka žáků z většinové společnosti a ze socio-kulturně znevýhodněného prostředí*. Ústav Českého národního korpusu FF UK, Praha 2013. Dostupný z WWW: <http://www.korpus.cz>

SWECCL (Spoken and Written English Corpus of Chinese Learners) s. 37

Wen, Qiu-Fang – Wang, Li-Fei – Liang, Mao-Cheng (2005): *Spoken and Written English Corpus of Chinese Learners (1.0)*. Peking: Foreign Language Teaching and Research Press.

SyD s. 60, 61, 62, 63, 64, 119, 121, 122, 124, 125, 126, 127, 128, 129, 130, 131, 133, 134

Cvrček, V. – Vondříčka, P. (2011). *SyD – Korpusový průzkum variant*. FF UK, Praha. Dostupný z WWW: <<http://syd.korpus.cz>>

SYN s. 14, 38

Hnátková, M. – Křen, M. – Procházka, P. – Skoumalová, H. (2014). The SYN-series corpora of written Czech. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, s. 160–164. Reykjavík: ELRA. ISBN 978-2-9517408-8-4.

SYN2000 s. 38

Čermák, F. – Blatná, R. – Hlaváčová, J. – Klímová, J. – Kocek, J. – Kopřivová, M. – Křen, M. – Petkevič, V. – Schmiedtová, V. – Šulc, M. *SYN2000: žánrově vyvážený korpus psané češtiny*. Ústav Českého národního korpusu FF UK, Praha 2000. Dostupný z WWW: <http://www.korpus.cz>

SYN2005 s. 38

Čermák, F. – Doležalová-Spoustová, D. – Hlaváčová, J. – Hnátková, M. – Jelínek, T. – Kocek, J. – Kopřivová, M. – Křen, M. – Novotná, R. – Petkevič, V. – Schmiedtová, V. – Skoumalová, H. – Šulc, M. – Velíšek, Z. *SYN2005: žánrově vyvážený korpus psané češtiny*. Ústav Českého národního korpusu FF UK, Praha 2005. Dostupný z WWW: <http://www.korpus.cz>

SYN2006PUB s. 38

Čermák, F. – Doležalová-Spoustová, D. – Hlaváčová, J. – Hnátková, M. – Jelínek, T. – Kocek, J. – Kopřivová, M. – Křen, M. – Novotná, R. – Petkevič, V. – Schmiedtová, V. – Skoumalová, H. – Šulc, M. – Velíšek, Z. *SYN2006PUB: korpus psané publicistiky*. Ústav Českého národního korpusu FF UK, Praha 2006. Dostupný z WWW: <http://www.korpus.cz>

SYN2009PUB**s. 38**

Křen, M. – Bartoň, T. – Hnátková, M. – Jelínek, T. – Petkevič, V. – Procházka, P. – Skoumalová, H. *SYN2009PUB: korpus psané publicistiky*. Ústav Českého národního korpusu FF UK, Praha 2010. Dostupný z WWW: <http://www.korpus.cz>

SYN2010**s. 38, 61**

Křen, M. – Bartoň, T. – Cvrček, V. – Hnátková, M. – Jelínek, T. – Kocek, J. – Novotná, R. – Petkevič, V. – Procházka, P. – Schmiedtová, V. – Skoumalová, H. *SYN2010: žánrově vyvážený korpus psané češtiny*. Ústav Českého národního korpusu FF UK, Praha 2010. Dostupný z WWW: <http://www.korpus.cz>

SYN2013PUB**s. 38**

Křen, M. – Hnátková, M. – Jelínek, T. – Petkevič, V. – Procházka, P. – Skoumalová, H. *SYN2013PUB: korpus psané publicistiky*. Ústav Českého národního korpusu FF UK, Praha 2013. Dostupný z WWW: <http://www.korpus.cz>

SYN2015**s. 38, 43, 45, 47, 54, 67, 68, 75, 78, 81, 83, 85, 91, 95, 100, 104, 139, 148, 151, 161**

Křen, M. – Cvrček, V. – Čapka, T. – Čermáková, A. – Hnátková, M. – Chlumská, L. – Jelínek, T. – Kovářiková, D. – Petkevič, V. – Procházka, P. – Skoumalová, H. – Škrabal, M. – Truneček, P. – Vondříčka, P. – Zasina, A. *SYN2015: reprezentativní korpus psané češtiny*. Ústav Českého národního korpusu FF UK, Praha 2015. Dostupný z WWW: <http://www.korpus.cz> seznam citovaných korpusů **183**

SYN verze 6**s. 35**

Křen, M. – Cvrček, V. – Čapka, T. – Čermáková, A. – Hnátková, M. – Chlumská, L. – Jelínek, T. – Kovářiková, D. – Petkevič, V. – Procházka, P. – Skoumalová, H. – Škrabal, M. – Truneček, P. – Vondříčka, P. – Zasina, A. *Korpus SYN, verze 6 z 18. 12. 2017*. Ústav Českého národního korpusu FF UK, Praha 2017. Dostupný z WWW: <http://www.korpus.cz>

T2K-SWAL Corpus**s. 18**

Biber, D. – Conrad, S. – Reppen, R. – Byrd, P. – Helt, M. 2002. Speaking & writing in the university: A multi-dimensional comparison. *TESOL Quarterly*, 36(1), s. 9–48.

The Academic Corpus**s. 20**

Dostupný z WWW: <https://www.victoria.ac.nz/lals/resources/academicwordlist/information/corpus>

The Longman Learners' Corpus**s. 18**

Dostupný z WWW: <http://global.longmandictionaries.com/longman/corpus#aa>

The PIKUST**s. 36**

Stritar, M. (2009). Slovene as a foreign language: The pilot learner corpus perspective. *Slovenski jezik / Slovene Linguistic Studies* 7, s. 135–152.

The Pilot Arabic Learner Corpus**s. 36**

Alfaifi, A. – Atwell, E. *Arabic Learner Corpus LDC2015S10*. Web Download. Philadelphia: Linguistic Data Consortium, 2015.

Treq**s. 60, 70, 71, 72, 163, 164, 165, 168, 169**

Vavříň, M. – Rosen A. (2015). *Treq*. FF UK, Praha. Dostupný z WWW: <<http://treq.korpus.cz>>

Trinity-Lancaster Corpus**s. 18**

Gablasova, D. – Brezina V. – McEnery, T. (2019): The Trinity Lancaster Corpus: Applications in language teaching and materials development. In: Sandr Götz, S., Liebig, J.: *Learner Corpora and Language Teaching*. John Benjamins, s. 7–28

Varieties of English for Academic Purposes (VESPA)**s. 20**

Granger, S. (2002). A Bird's-eye View of Computer Learner Corpus Research. In Granger S. – Hung, J. – Petch-Tyson, S. (eds). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching. Language Learning and Language Teaching* 6. Amsterdam & Philadelphia: Benjamins (245 pp.), s. 3–33.

WebCorp**s. 35**

Renouf, A. (2003). WebCorp: providing a renewable data source for corpus linguists, in S. Granger & S. Petch-Tyson (eds). *Extending the scope of corpus-based research: new applications, new challenges*. Amsterdam: Rodopi, 39–58. Copyright © 1999–2019 Research and Development Unit for English Studies, Birmingham City University.

Profily autorů

PhDr. Tomáš Gráf, Ph.D. (tomas.graf@ff.cuni.cz) vystudoval anglistiku na FF UK, kde od roku 2010 působí jako odborný asistent v Ústavu anglického jazyka a didaktiky. Ve svém výzkumu se zabývá žakovským jazykem především z pohledu přesnosti a plynulosti. Vyučuje jazykovou akvizici a předměty týkající se didaktiky anglického jazyka a přípravy budoucích učitelů angličtiny.

Mgr. Věra Hejhalová, Ph.D. (vera.hejhalova@ff.cuni.cz) je česká germanistka působící v Ústavu germánských studií FF UK. Vystudovala Pedagogickou fakultu Jihočeské univerzity v Českých Budějovicích (Učitelství pro střední školy, aprobace německý jazyk – latina). Na Filozofické fakultě Univerzity Karlovy se profilovala v doktorském programu germánské jazyky a literatury. Vyučuje korpusovou lingvistiku, didaktiku němčiny jako cizího jazyka, gramatiku a frazeologii němčiny. Centry jejího badatelského zájmu jsou didaktika němčiny jako cizího jazyka, korpusová lingvistika a frazeologie němčiny.

Mgr. Barbora Baštová (barbora.bastova@ff.cuni.cz) je absolventkou magisterského studijního oboru učitelství češtiny jako cizího jazyka v Ústavu českého jazyka a teorie komunikace FF UK (DP „Mlčeti stříbro, mluvití zlato. Rozvíjení mluvních dovedností v současných učebnicích češtiny pro cizince“). Pracuje jako lektorka češtiny pro cizince v Ústavu bohemistických studií se zaměřením na ústní a písemné vyjadřování. Zároveň se jako doktorandka ÚČJTK zabývá využitím jazykových korpusů ve výuce češtiny pro cizince, a to především s ohledem na kolokace sloves pohybu. Vyučuje také v magisterském programu Učitelství češtiny jako cizího jazyka.

Prof. PhDr. Karel Šebesta, CSc. (karel.sebesta@ff.cuni.cz) působí v Ústavu českého jazyka a teorie komunikace FF UK, pedagogicky i badatelsky se zaměřuje na aplikovanou lingvistiku a didaktiku (českého) jazyka jako jazyka prvního i druhého/cizího. Je mj. předsedou oborové rady doktorského studijního oboru Didaktika konkrétního jazyka, iniciátorem a vedoucím dlouhodobého projektu AKCES – Akviziční korpusy českého jazyka, prvního komplexu speciálních korpusů sloužících studiu osvojování a dalšího vývoje českého jazyka i pedagogickým účelům, autorem řady publikací a učebnic v oboru.

Mgr. Kateřina Šormová, Ph.D. (katerina.sormova@ff.cuni.cz) je absolventkou doktorského studia na FF UK, obor český jazyk. V současné době působí jako odborná asistentka v Ústavu českého jazyka a teorie komunikace FF UK, vyučuje zejména v programech navazujícího magisterského studia Učitelství českého jazyka a literatury a Učitelství češtiny jako cizího jazyka, zaměřuje se na osvojování češtiny jako prvního a druhého/cizího jazyka, na čtenářskou gramotnost a na jazykové testování a hodnocení včetně sebehodnocení. Spolupodílela se na vybudování databanky jazykových projevů romských mluvčích češtiny ROMi a korpusu CZESL.

Rejstřík

- akviziční korpusy 7, 12–14, 16, 17, 33, 34, 41–45, 47
akviziční korpusy češtiny 13, 14, 185
angličtina pro pracovní účely 19
angličtina pro akademické účely 19
anotace 17, 22, 23, 26, 29, 31, 34, 37, 44, 45, 46, (pracovní listy: 79, 89, 92, 98, 103, 107, 112, 116, 119, 123, 126, 129, 132, 136, 140, 145, 148, 151, 154, 157, 163, 167
anotační systémy 17
atrice 12
autentičnost 34, 40, 41
- Bilingual Corpus of Chinese English Learners** 37
- Corpus of Learner English** 20
corpus query language 50
CQL 50, 61, 95, 100, 113, 118
- část slova 50
české žákovské korpusy 37
Český národní korpus 35, 47
čítanky 25
- disambiguace** 44, 45
- emendace** 22, 26, 37, 45, 46
English for Academic Purposes 19, 20
English for Occupational Purposes 19
English Grammar Profile 18, 24, 25
English Vocabulary Profile 18, 25, 28
explicitní segmentace textu 44
- falešní přátelé** 26
- faux amis 26, 28
fráze (typ dotazu) 25, 28, 29, 50, 52
fráze (ustálené slovní spojení) 98, 124, 168
frekvence 24, 26, 27, 54, 57, 58, 59, 67, 69, 91, 97, 101, 110, 113, 118, 142, 147
- i. p. m.** 58
Interactive Spoken Language Education 41
- jazyk dětí a mládeže** 15
jazyk dětí předškolního věku 15
jazyk krajanských komunit 15
jazyk nerodilých mluvčích češtiny 15
jazyk sociokulturně znevýhodněných skupin 15
jazyk ve vzdělávacím kontextu 15
Just the Word 27
- kolokace** 27–29, 35, 58–60, 64, 79, 80, 84, 85, 89, 91, 106, 120, 136, 138
konkordance 54, 55, 70, 71, 84, 85, 91, 96, 101, 105, 106, 109, 113, 118, 138
konkordanční 27, 29
KonText 47, 48, 60, 75, 79, 95, 100, 105, 120, 136, 138, 143
kookurence 58
korpus 11, 12, 13, 14
korpusové pozice 44
korpusy abstraktů 31, 32
korpusy atriční 13
korpusy diachronní 14, 47,
korpusy obecné 11, 13, 14, 34, 35–37, 39–41, 43

korpusy osvojování mateřštiny 17
korpusy osvojování cizích jazyků 17
korpusy speciální 11–14, 16, 20, 34, 40
korpusy synchronní 14, 35, 38, 45, 47,
korpusy terapeutické 13, 15
korpusy zaměřené na jazyk krajanů 12,
15
korpusy žakovské 12, 13, 15, 16–26, 28, 31,
34–37, 39, 41–45, 47,
krajanské korpusy 13, 15

learner corpora 16, 17, 34
lemma 44, 45, 46, 50, 52, 54, 57, 59, 61, 66,
70, 75, 83–85, 91, 97, 100, 106, 109, 110,
125, 128, 131, 134, 138, 142, 147, 150, 153,
156, 165,
lemmatizace 37, 45
LINDAT/CLARIN 15
lingvistická anotace 34
Longman Grammar of Spoken and Wri-
tten English 30

metatextové informace 34, 35, 43
mezijazyk 16, 21, 23, 39
morfologická analýza 44, 45
morfologické značkování 22, 45, 46,

nedokončené segmenty 32
nestandardní jazyková data 13, 36,
45

opakované segmenty 32

paralelní korpus 14, 47, 70
pauzy 30, 32
plynulost 27, 30–33
poziční atributy 44
pragmatika 21, 31, 32, 123

projevy vernakulární 42, 43
promíchat 55, 51, 65, 70, 109

regulární znaky/výrazy 50, 51

SERR 17, 18, 40, 44
seznamy slovní zásoby 18, 20, 24, 25, 28,
67
slovní tvar 22, 35, 49, 50, 51, 52, 54, 56–59,
61, 66, 67, 69, 70, 101, 113, 118, 143
specializované korpusy mluveného ja-
zyka 14
specializované korpusy psaného jazyka
14
správná varianta 22
strukturní atributy 34, 43, 44
synchronní korpusy mluveného jazyka
14
synchronní korpusy psaného jazyka 14,
38

tempo mluvy 32
tokenizace 43, 44
tokeny 34, 35, 37, 43–45, 52, 83, 91
typ dotazu 48, 49, 50, 51, 52, 54, 56, 75, 83,
85, 91, 95, 100, 105, 109, 113, 118

velikost 29, 34–40, 54, 55, 83, 91
videokorpusy 21
vyváženost 34, 35, 37, 38–40, 132, 157
vzorek 11, 13, 16, 20, 39, 40, 55, 56

web jako korpus 35
wiki 34, 38, 47, 49, 50, 59, 60, 64, 67, 69, 71

žáci s odlišným mateřským jazykem 23,
26, 61

žáci s OMJ 26, 47, 65, 67, 70, 158

Resumé/Summary

Corpora in language teaching

The present monograph is written in the belief that corpus linguistics offers a wide array of tools which are useful and directly useable by L1 and L2 language teachers. It invites them to explore language using basic exploratory corpus linguistics methods and to use these as tools both for learning about language and for designing materials and activities which will guide students to learn about language and increase their linguistic competence. The specific target group for this publication is those teachers whose classes include students with mother tongues different from the class majority (Czech) and thus often with a lower-than-native command of that language. It is especially these students who need careful language work, as the improvement of their linguistic skills in the language of school instruction has a direct bearing on their success at school.

Acknowledging the fact that corpus linguistics is a field with which Czech teachers may still be largely unfamiliar, the book, first of all, provides a concise introduction to corpus linguistics. Starting, in Chapter 1, by describing and defining what language corpora are and outlining what types of corpora exist and how they have contributed to language teaching, it then explores the area of corpus research and its relevance to language pedagogy. In Chapter 2 it discusses acquisitional corpora and the link between learner corpus research and such fields as second language acquisition, language testing and assessment and language for specific and academic purposes. It further describes how corpus linguistics can contribute to learner-language analysis especially with regard to grammar, vocabulary, phraseology, discourse, pragmatics, fluency and accuracy and discusses its impact on language teaching. Chapter 3 furthers our understanding of the field by focussing on such key parameters of language corpora as their size, the authenticity of the recorded language they contain, their structural and positional attributes, and methods and implications of corpus annotation.

The latter half of the publication aims to introduce teachers to concrete techniques of exploiting language corpora. It does so by introducing the teachers to the Czech National Corpus and the tools which are available through its related web pages, and progresses methodically from the basics of formulating queries and setting various attributes and filters to working

with the concordance listing and accompanying web apps, which are directly exploitable in work with students with different L1s, such as SyD, Morfio, KWords, and Treq.

One of the book's most interesting and useful aspects is the inclusion of sample practical worksheets. The authors have designed these to illustrate how corpus tools can be used for the development of teaching materials whilst bearing in mind the characteristic needs of the target group of students, i.e. the need of many of them to improve their knowledge of Czech in order to be able to succeed in the Czech educational system. The exercises include a variety of didactic approaches, and each worksheet is accompanied by a key which includes both the instructions for carrying out the task using one of the corpus tools and a possible solution. The book thus also offers highly practical concrete solutions for teachers who need to prepare language exercises to aid students' progress in Czech as a second language.

Kateřina Šormová, Karel Šebesta a kol.
Korpusy v jazykovém vyučování II.

Vydala Univerzita Karlova, Filozofická fakulta,
nám. Jana Palacha 2, Praha 1
jako 78. svazek řady Varia

Obrázky Marek Šorm
Typografická osnova František Štorm
Sazba z písma Skolar Dušan Neumahr
Vydala Togga, spol. s r. o., Praha
Vydání první, Praha 2019

Tomáš Gráf

Věra Hejhalová

Barbora Baštová

Karel Šebesta

Kateřina Šormová





FILOZOFICKÁ FAKULTA
Univerzita Karlova

ISBN 978-80-7306-932-0

