

JAZYKOVÉ A TEXTOVÉ ZDROJE SLOVENSKÉHO NÁRODNÉHO KORPUSU A ICH VYUŽITIE VO VÝUČBE SLOVENČINY AKO CUDZIEHO JAZYKA

Autorka: Mgr. Katarína Chlpíková

Kontakt: Slovenský národný korpus, Jazykovedný ústav Ľ. Štúra, Slovenská akadémia vied, Bratislava, katarinach@korpus.sk

Autorka: PhDr. Mária Šimková, Ph.D.

Kontakt: Slovenský národný korpus, Jazykovedný ústav Ľ. Štúra, Slovenská akadémia vied, Bratislava, marias@korpus.sk

ABSTRAKT

V príspevku predstavujeme Slovenský národný korpus ako pomôcku pri výučbe slovenčiny ako cudzieho jazyka, najmä pri tvorbe cvičení. Korpusové databázy sú bohatým, dynamickým a interaktívnym zdrojom textového materiálu a jazykových informácií, s ktorým môžu pracovať lektori i frekventanti kurzov Sj ako cudzieho jazyka. Vyzdvihneme niektoré osobitosti korpusu oproti internetovým zdrojom, ako je napr. podrobná bibliografická a štýlovo-žánrová anotácia, vyhľadávanie spojení slov, paralelné a iné špecifické korpusy. Ponúkneme návrhy cvičení, ktoré môže lektor Sj ako cudzieho jazyka využiť na kurzoch.

Kľúčové slová: korpus, jazykové elektronické zdroje, všeobecný korpus, paralelné korpusy, morfológická anotácia, praktické cvičenia, výučba slovenčiny ako cudzieho jazyka

1. Úvod

Učenie sa slovenčiny ako cudzieho jazyka sa teší stále väčšiemu záujmu cudzincov vrátane zahraničných študentov. Didaktika tejto výučby zaznamenala v posledných rokoch výrazný progres a vyvíja sa podľa aktuálnych potrieb spoločnosti a záujemcov o komunikáciu v slovenskom jazyku.

V rámci inovácií v metodických postupoch, ku ktorým patrí napríklad zážitkové učenie či situačné úlohy, sa do popredia dostávajú aj moderné formy výučby s využitím elektronických zdrojov, ktoré zahŕňajú aktuálnu slovnú zásobu, jej používanie v písaných textoch i spontánných prehovoroch, ako aj rôzne jazykové informácie. Veľa možností, ako zefektívniť výučbu, ponúka

aj Internet a na vedeckých i pedagogických pracoviskách sa vyvíjajú stále novšie e-learningové projekty.

Možnosti výučby slovenčiny ako materinského alebo ako cudzieho jazyka pomocou korpusov a iných elektronických zdrojov ešte nie sú na Slovensku dostatočne rozpracované, zmieňujú sa o nich zatiaľ predovšetkým korpusoví lingvisti (napr. Ivoríková, 2006; Šimková, 2007; Gajdošová – Ľos Ivoríková, 2014). V predkladanom príspevku stručne predstavíme aktuálne elektronické zdroje Slovenského národného korpusu (ďalej SNK) vhodné pri výučbe slovenčiny a ponúkneme návrhy cvičení na báze jedného z korpusov písanej slovenčiny, ktorý sa nachádza medzi primárnymi textovými zdrojmi SNK. Pri príprave projektu SNK a tvorbe jeho súčastí sa od začiatku počítalo s vedecko-výskumným a edukačným využitím s cieľom poskytnúť čo najviac informácií o slovenčine bežným používateľom jazyka, lingvistom, prekladateľom, spisovateľom, učiteľom a všetkým, ktorí sa jazykom zaoberajú, čiže aj frekventantom kurzov slovenčiny ako cudzieho jazyka (porov. úvodné informácie na stránke <http://korpus.sk/>, resp. časť o využití korpusov v rôznych oblastiach na <http://korpus.sk/what.html>). Pri tvorbe cvičení sme sa opierali o práce spomínaných autoriek a o relevantné zahraničné zdroje.

2. Aktuálne dostupné jazykové a textové zdroje Slovenského národného korpusu

Čo je elektronický korpus textov a ako môže lektorom slovenčiny ako cudzieho jazyka pomôcť pri výučbe slovenčiny? Autori knihy *From Corpus to Classroom* zdôrazňujú dôležitú vlastnosť korpusu: štruktúrovaný súbor textov slúži na *kvantitatívne* a *kvalitatívne* analýzy (O'Keeffe, A. a kol., 2007, s. 1). Práve táto črta robí korpus jedinečným oproti ostatným internetovým zdrojom. Vďaka bibliografickej a štýlovo-žánrovej anotácii textov a morfolologickej anotácii slov ponúkajú korpusové zdroje široké možnosti na skúmanie jazykových javov prakticky zo všetkých rovín jazyka – morfológie, sémantiky, syntaxe, lexikológie a hovorený korpus aj fonetiku a fonológiu. Na základe bibliografickej a štýlovo-žánrovej anotácie je možné identifikovať pôvod textov a triediť ich podľa žánrov a iných vlastností, čo umožňuje špecifikovať vyhľadávanie a skúmať využiteľnosť určitého javu v konkrétnom žánri, štýle, v tematickej oblasti, v určitom regióne, období a pod.

V čom sa korpusové textové zdroje líšia od textov voľne dostupných na Internete:

1. *Zaradenie slov do širších kontextov.* Pre frekventantov kurzu slovenčiny ako cudzieho jazyka je dôležité od začiatku vnímať a poznávať slová v kontexte. V korpuse majú možnosť vyhľadať akékoľvek slovo alebo výraz vo všetkých kontextoch, v ktorých sa zadané slovo alebo výraz v korpuse vyskytuje. Kontexty sa zobrazia v konkordancii, v ktorej sa nájdené (kľúčové) slová aj kontexty dajú ďalej triediť podľa rôznych vlastností: podľa tvaru či morfológických kategórií hľadaného slova, podľa slov vyskytujúcich sa vľavo alebo vpravo od hľadaného slova a pod. Lektor si môže na základe získaného autentického materiálu vypracovať úlohy na rozpoznanie významu slova, spájateľnosti slov, zaradenie do pádov na základe spájateľnosti s inými slovami a podobne (bližšie v časti 3). Frekventant si nielen precvičuje slovnú zásobu a gramatické javy, ale priamo vidí, ako slovo alebo slovné spojenie funguje v reálnych textoch.

2. *Frekvenčné zoznamy.* Z korpusu ako textového zdroja sa dajú generovať frekvencie slov, s akými sa v danom korpuse vyskytujú. Keďže korpusové databázy v súčasnosti dosahujú miliardové veľkosti (napr. hlavný korpus SNK prim-6.1 obsahuje takmer 830 miliónov textových jednotiek, webový korpus takmer 2,4 miliardy textových jednotiek, t. j. slov, číslíc, značiek, interpunkčných znamienok), frekvencie slov v týchto korpusoch už celkom dobre odrážajú ich reálnu frekvenciu v komunikácii. Frekvencie slov v základných aj všetkých tvaroch, frekvencie podľa slovných druhov, frekvencie dvojíc, trojíc a štvoríc slov hlavného korpusu sú zverejnené na stránke korpusu pri príslušnej verzii (napr. [http://korpus.sk/prim\(2d\)6\(2e\)1\(2f\)frequencies.html](http://korpus.sk/prim(2d)6(2e)1(2f)frequencies.html)). Lektor si vie aj sám vytvoriť frekvenčné zoznamy z konkrétnych textov a zisťovať, aké slová, príp. slovné spojenia sa najčastejšie používajú v príslušných typoch textov a aké ďalšie slová majú vo svojom najbližšom okolí.

3. *Texty, ktoré odrážajú živý, hovorový jazyk.* Korpus písaných textov je zložený z autentických textov odborného, publicistického a umeleckého štýlu a tzv. hovorené korpusy obsahujú aj spontánne prehovory v štandardnej slovenčine (<http://korpus.sk/shk.html>) alebo v nárečí (<http://korpus.sk/dialect.html>) a ich prepisy. Lektor slovenčiny v nich môže nájsť príklady na používanie slov a fráz v bežnom styku a vybrať pre frekventantov príklady z bežnej reči. V prípade, že sú prepisy prelinkované so zvukom, môžu si študenti vypočuť konkrétny prehovor a učiť sa porozumieť aj hovorenému slovu v bežnej komunikácii (v hluku auta, vlaku či reštaurácie). V písaných textoch, ktoré sa do korpusovej databázy zaraďujú v takej podobe, ako boli publikované, sa nachádzajú aj preklepy a iné chyby, ktoré sa pri písaní či korektúrach pred

vydaním nezachytili. Tieto chyby sa pri tvorbe korpusu neopravujú, preto lektor by mal skontrolovať vybraný kontext skôr, ako ho použije v úlohe. Chyby však môže vo vyučovaní aj tvorivo využiť: frekventanti dostanú za úlohu, aby ich v texte našli a opravili. Okrem pravopisných chýb môže ísť aj o nesprávne väzby medzi slovami či nesprávne slovné tvary a študenti si ich aktívnou prácou môžu lepšie uvedomiť a osvojiť si správne podoby.

4. *Spájateľnosť slov.* V korpusových zdrojoch je možné vyhľadať spojenie akýchkoľvek slov a zistiť frekvenciu spojení. Frekventanti majú možnosť tieto spojenia vidieť priamo v textoch a na základe kontextov určovať ich význam alebo zisťovať, aké slová sa spájajú s konkrétnym slovným druhom (bližšie v časti 3).

5. *Špecializované korpusy.* Výhodou korpusu je, že si jeho používateľ môže nastaviť parametre výberu a triedenia materiálu podľa potreby: od výberu konkrétneho korpusu (vyvážený korpus, nárečový korpus, korpus publicistických textov, korpus umeleckých textov a ďalšie korpusy; podrobnejšie na <http://korpus.sk/res.html>) cez výber textov napr. podľa autora, roku vydania, jazyka originálu až po konkrétne atribúty hľadaného výrazu, ako je slovný druh, gramatické kategórie, ďalšie výrazy, s ktorými sa hľadané slovo spája alebo nespája atď. Niektoré parametre sú používateľmi žiadané častejšie a v takých prípadoch sa pripravujú špecializované korpusy. V SNK sa v súvislosti s rôznymi projektmi pripravili osobitné korpusy podľa vecných oblastí: napr. korpus ekonomických textov, korpus náboženských textov, korpus právnych textov (<http://korpus.sk/dom.html>).

6. *Paralelné korpusy.* Všetky doteraz spomínané korpusy a podkorpusy sú jednojazyčné – obsahujú texty písané po slovensky alebo komunikáty prednášané v slovenčine. Pri výučbe slovenčiny ako cudzieho jazyka môžu výraznou mierou pomôcť aj *paralelné korpusy*, ktoré sú zložené zvyčajne z dvoch (alebo troch a viacerých) rovnakých textov v rôznych jazykoch. Ide o vzájomné preklady medzi dvoma jazykmi alebo o preklad toho istého textu z celkom iného jazyka. Slovenský národný korpus v súčasnosti obsahuje 8 dvojajazyčných paralelných korpusov: slovensko-anglický, slovensko-bulharský, slovensko-český, slovensko-francúzsky, slovensko-latinský, slovensko-maďarský, slovensko-nemecký a slovensko-ruský paralelný korpus (podrobnejšie o ich veľkosti a zložení na <http://korpus.sk/par.html>). Texty sú spárované po vetách, takže pri hľadaní slova v jednom z jazykov sa ukážu kontexty tohto slova v dvoch jazykoch a študent hovoriaci iným jazykom ako slovenským sa môže učiť slovenské výrazy na základe reálnych textov a prekladových ekvivalentov, ktoré sa v textoch často nachádzajú v značne

širšom rozsahu ako v prekladových slovníkoch. Na vyšších úrovniach znalosti slovenčiny môže študent hodnotiť správnosť alebo vhodnosť zvoleného prekladu, navrhovať iné ekvivalenty a riešenia prirodzenejšie pre daný jazyk.

Ukážka zo slovensko-ruského paralelného korpusu – v ruskej časti korpusu sme hľadali slovo *бабушка*, ktorého ekvivalentom nie je *babička*, ale pre slovenčinu bežnejšie pomenovanie *stará mama*.

7. *Korpusové nástroje*. Okrem tvorby korpusov ako textových zdrojov sa v oddelení SNK vyvíjajú aj *korpusové nástroje*, ktoré sú vytvorené na báze korpusových dát a zároveň v nich umožňujú rýchlo vyhľadať niektoré informácie o jazykových prostriedkoch a ich fungovaní alebo umožňujú previesť text do inej podoby. Medzi ne patrí napríklad nástroj DIAKRITIK (<http://korpus.sk/diakritik.html>), pomocou ktorého je možné automatizovane prepísať zadaný text bez diakritiky na korektný slovenský text s mäkčeňmi a dĺžkami a naopak, čo je veľmi dobrá pomôcka napr. aj pri korigovaní cudzincom písaného slovenského textu. Nástrojom LUDEVIT (<http://korpus.sk/ludev.it.html>) sa dá previesť text v súčasnej slovenčine na text v štúrovčine a pokročilí frekventanti tak môžu hľadať rozdiely medzi súčasnou a štúrovskou podobou slovenčiny a sledovať zmeny vo vývoji slovenského pravopisu. Pomocou nástroja na identifikáciu kolokácií KOLOKAT (<http://korpus.sk/kolokat.html>) sa dá ľahko zobraziť pevnosť spojení alebo voľnosť až neprípustnosť spoluvýskytov dvoch slov. Najnovším nástrojom v SNK je DEVELOPER (<http://korpus.sk/developer.html>), ktorý zobrazuje frekvenciu slova alebo porovnanie výskytov dvoch slov v textoch spracovaných v korpuse od r. 1955 po súčasnosť.

Osobitnú časť v celku SNK predstavujú *jazykové zdroje*. Záujemcovia o slovenčinu často pracujú s elektronickými slovníkmi, v ktorých sa dá vyhľadávať na stránke <http://slovniky.juls.savba.sk>. Táto podoba elektronického spracovania kodifikačných príručiek a iných, aj starších slovníkov bola vytvorená a spravuje sa v rámci projektov SNK, medzi ktoré patrí aj plná digitalizácia slovníka A. Bernoláka (<http://www.juls.savba.sk/ediela/ber nolak/>). Súčasťou slovníkového rozhrania sú jazykové zdroje SNK vytvorené v procese prípravy a skvalitňovania morfolologickej anotácie korpusu, ako je slovník tvarov podstatných mien (<http://korpus.juls.savba.sk/substantiva.html>) a v pracovnej verzii dostupný slovník slovesných tvarov (<http://slovniky.korpus.sk/?d=verbdb>). Obidve databázy môžu poslúžiť lektorom aj študentom na zistenie alebo overenie správnosti konkrétnych tvarov slov z dvoch hlavných slovných druhov. Pri precvičovaní správnej výslovnosti slovenských slov môžu pomôcť načítané

slová, ktoré sa dajú vypočítať v dvoch formátoch. Nachádzajú sa takmer na konci uvedeného slovníkového rozhrania a za nimi nasledujú prekladové frázy zo slovensko-anglického a slovensko-českého paralelného korpusu. Pri niektorých špecifických úlohách pre pokročilých frekventantov alebo neskôr pre prekladateľov sa dajú dobre využiť aj súpisy termínov spracované v Slovenskej terminologickej databáze (<http://data.juls.savba.sk/std/>).

3. Využitie korpusových databáz na tvorbu cvičení pri výučbe slovenčiny ako cudzieho jazyka

V nasledujúcej časti ponúkame niekoľko cvičení a návrhov na prácu s vyváženým korpusom prim-6.1-public-vyv ako inšpiráciu pre lektorov slovenčiny ako cudzieho jazyka. Cvičenia sme rozdelili do troch okruhov: *sáľateľnosť slov*, *sémentika slov* a *ohýbanie slov*.

Spáľateľnosť slov. Zo základnej slovnej zásoby sme vybrali substantívum *mesto*, s ktorým budeme pracovať prevažne pri frekventantoch od úrovne A2, kedy už ovládajú pády v slovenčine. Študenti slovenčiny ako cudzieho jazyka sa s týmto substantívom prvýkrát stretávajú v lekcii *Mestá a miesta* (učebnice *Krížom krážom A1* a *Hovorme spolu po slovensky! A*) či v lekcii *Čo má naše mesto?* (učebnica *Tri, dva, jeden – slovenčina A1, A2.1*). Vo všetkých spomínaných lekciiach sa autori snažia podať tému bývania v meste pútavým spôsobom prostredníctvom rôznych situácií, s ktorými sa ľudia v meste stretávajú. Túto tému môžeme pripraviť aj pomocou korpusu, s ktorým sa dá on-line pracovať priamo na vyučovacej hodine. Z hľadiska spáľateľnosti nás pri substantíve *mesto* bude zaujímať, aké prídavné mená sa vyskytujú v jeho najbližšom okolí, teda stoja bezprostredne pred ním.

Na prípravu cvičení sme zvolili vyvážený korpus prim-6.1-public-vyv (<http://bonito.korpus.sk>), ktorý je zložený v rovnakom pomere (po jednej tretine) z textov umeleckého, odborného a publicistického štýlu a svojou veľkosťou (takmer 320 miliónov slov a iných textových jednotiek) predstavuje primeraný zdroj materiálu na prácu v škole. Vo vyhľadávači NoSketch Engine, ktorý sa používa v Slovenskom národnom korpuse, zvolíme spôsob vyhľadávania pomocou CQL (Corpus Query Language – formálny jazyk na zostavenie príkazu a zadanie špecifikácií vyhľadávania). V našom prípade sa bude vyhľadávací príkaz skladať z tagu (morfologickej značky), pomocou ktorého vytvoríme príkaz na vyhľadanie konkrétneho slovného druhu – *adjektíva* a lemy *mesto* (lema je slovo v základnom tvare, v ktorom sú zahrnuté všetky jeho ďalšie tvary, takže nemusíme vyhľadávať výskyty slova osobitne

v jednotlivých pádoch, rodoch, číslach a pod., ale jedným príkazom sa zobrazia všetky tvary, v ktorých sa slovo v korpusе vyskytuje). Príkaz vytvoríme v podobe [tag="A.*"] [lemma="mesto"] a zobrazia sa všetky kontexty obsahujúce slovo *mesto* vo všetkých jeho tvaroch spolu s adjektívami, ktoré pred ním stoja. Z ponuky možností práce s vyhľadaným materiálom vyberieme položku „Frekvenčné zoznamy – vlastné“ a vytvoríme frekvenčný zoznam jednotlivých adjektív v základnom tvare.

	<u>lemma</u>	<u>Frekv</u>	
1. p/n	hlavný	15 716	
2. p/n	nový	5 310	
3. p/n	velký	3 288	
4. p/n	starý	3 235	
5. p/n	pracovný	3 131	
6. p/n	celý	2 077	
7. p/n	malý	1 255	
8. p/n	rodný	1 244	
9. p/n	slovenský	1 069	
10. p/n	okresný	1 041	
11. p/n	kráľovský	838	
12. p/n	krajský	592	
13. p/n	d'alší	524	
14. p/n	kúpeľný	513	
15. p/n	kapský	491	
16. p/n	banský	485	
17. p/n	európsky	477	
18. p/n	prístavný	457	
19. p/n	parkovací	415	
20. p/n	svätý	400	
21. p/n	významný	380	
22. p/n	sídelný	359	
23. p/n	pekný	353	

Obrázok 2: Frekvenčný zoznam adjektív z korpusu prim-6.1-public-vyv, ktoré sa spájajú so substantívom mesto. Prvých 23 výskytov

Na obrázku č. 2 vidíme prvých 23 výskytov z existujúcich 2990. Nie všetky slová však vyhovujú zadanej požiadavke – buď nejde o adjektíva, alebo sa tam dostali aj vlastné mená, alebo ide o slová, ktoré v spojení so substantívom *mesto* na prvý pohľad nedávajú zmysel. Sú to slová *d'alší*, *kapský*, *pracovný* a *parkovací*. Adjektíva *pracovný* a *parkovací* sa v zozname objavili preto, lebo v procese automatizovanej lematizácie bol homonymný tvar *miest* nesprávne priradený k

leme *mesto* a nie *miesto*, o ktoré v tomto prípade ide. *Ďalší* je zámeno a *kapský* je vlastné meno, preto nebudeme s týmito slovami pracovať. V zozname ostalo 19 adjektív, ktoré môžeme použiť pri tvorbe nasledujúcich cvičení:

1. Jednoduchými vetami opíšte Bratislavu (alebo akékoľvek mesto) použitím 5 adjektív z vytvoreného frekvenčného zoznamu. (Úroveň A1.)

2. Prečítajte si text o Bratislave na wikipédii a porovnajte ho s frekvenčným zoznamom v korpuse. Ktoré adjektíva z frekvenčného zoznamu sa zhodujú s adjektívami v texte na wikipédii? (Úroveň A2.)

Lektor môže porovnať, či frekvenčný zoznam z vyváženého korpusu odráža frekvenciu týchto adjektív v textoch na internete.

Reálie:

3. Zadajte do internetového vyhľadávača spojenie *banské mesto*. Ktoré dve slovenské banské mestá sa zobrazia ako prvé? (Úroveň A1.)

Webové stránky, na ktorých budú informácie o bankých mestách, ponúknu aj obrázky miest:

4. Podľa obrázkov na webovej stránke bankých miest opíšte každé mesto piatimi vetami s využitím adjektív z frekvenčného zoznamu vytvoreného z korpusu. (Úroveň A1.)

Pokročilí študenti, napr. na úrovni B1 – B2, môžu dostať samostatnú úlohu na porovnanie významov substantív *mesto* a *miesto*. Na základe frekvenčného zoznamu adjektív z vyváženého korpusu by mohla úloha znieť:

5. Zo zoznamu adjektív vyberte tie, ktoré nesúvisia so slovom *mesto*, ale s iným podobným slovom.

a. S ktorým slovom sa tieto adjektíva spájajú?

b. Prečo vyhľadávací nástroj našiel tieto adjektíva v spojení so slovom *mesto*? V ktorom páde majú slová *mesto* a *miesto* rovnaký tvar?

c. Poznáte ďalšie 3 adjektíva, ktoré použijete so substantívom *miesto*, ale nepoužijete ich so substantívom *mesto*?

d. V zozname adjektív sa nachádza jedno slovo, ktoré nie je adjektívom. Ktoré to je?

Podobné cvičenia môžeme vytvoriť aj na základe frekvenčného zoznamu *predložiek*, ktoré sa používajú v spojení so substantívom *mesto*.

Sémantika slov. Ak sa na výučbu slovnej zásoby pozeráme cez jej sémantické vlastnosti a spájateľnosť s inými slovami v určitom kontexte, korpusy budú výbornou pomôckou na vyhľadávanie týchto kontextov a identifikáciu významov. Vo vyváženom korpuse prim-6.1-public-vyv sme dali hľadať lemu slovesa *chodiť*. Výpis (konkordancia) kontextov s týmto výrazom (obr. 3) obsahuje sloveso *chodiť* v rôznych významoch a v rôznych kontextoch. Tieto významy môžeme porovnať s významami, ktoré ponúka Slovník súčasného slovenského jazyka H – L (<http://slovníky.juls.savba.sk/>):

1. miesto, kam sa chodí pravidelne/často, čiže chodiť niekam dlhodobo, s určitým cieľom
2. fráza „Tak to (už) chodí“
3. vo význame kráčať, robiť kroky
4. mať s niekým vzťah, „chodiť s niekým“

i losangeleský butik The Gauntlet , do ktorého	chodia	tí , čo si chcú nechať prepichnúť niektorú
krok je veľkou udalosťou . Všade s nimi	chodí	veľa ochrancov , naokolo sú davy kričiacich fanúšikov
pracovať dlho do noci , ale tak to	chodí	. V prestávkach som pojedla kopec čerešní a
som pocit , že jej zničím topánky a	chodila	som ako hráč rugby v jej úzkych šatách
asi by som veľa skúmala , ako mám	chodiť	a rozprávať , ale keďže som hrala samú
sme sa naučili o tom , ako to	chodí	vo filmárskom biznise . Je to bláznivý život
že áno a ja som im povedal -	chodíte	stále vpred , kým vás to baví ,
Cameron . Padli si do oka , ale	chodit'	začali spolu až po roku , keď jej
filmu bol jediný festival , na ktorý som	chodil	. Nechodil som na žiadnu inú festivalovú prehliadku
raz cítil dobre - akoby do tejto reštaurácie	chodil	prínajmenšom od jej vzniku . Mimochodom , hmlisto
1 . cenová trieda) . Ja som	chodil	do plzenskej , kde čapovali (ako názov
mali ! Do reštaurácie Praha Expo 58 ľudia	chodili	na zmrzlinu . Sedeli na terase , pozorovali
stretnutie s Jurajom išla a dnes už spolu	chodia	dva roky . Juraj , ktorý dovtedy nič
si sympatickí naďalej , tak sme spolu začali	chodiť	. Vlastne mi už počas toho predošlého intenzívneho

Obrázok 3: Konkordancia slovesa *chodiť* v korpuse prim-6.1-public-vyv

Cvičenia môžu byť zamerané napríklad na to, či dokážu frekventanti kurzu na pokročilej úrovni (od B1) identifikovať v textoch tieto štyri významy:

1. V koľkých vetách je sloveso *chodiť* vo význame „mať s niekým vzťah“?
2. V ktorých dvoch vetách je sloveso *chodiť* súčasťou frazémy? (*Tak to chodí*)
3. Kto v daných vetách „kráča“?
4. Napište vlastné vety so slovesom *chodiť* vo všetkých štyroch významoch.

Podobne môžeme pracovať aj s verbom *mať* (obr. 4), ktoré sa vo vete vyskytuje buď v plnom lexikálnom význame *vlastniť*, *držať/nosiť*, *dosiahnuť niečo*, *označovať príbuzenský či spoločenský stav*, *obsahovať*, *považovať* atď., alebo vo funkcii modálneho, limitného či existenciálneho slovesa (*mať obed v kuchyni*) a tiež ako spona v spojení s trpným prídastím, ako sa uvádza napríklad aj v Krátkom slovníku slovenského jazyka (<http://slovníky.juls.savba.sk/>). Pre frekventantov úrovne A1 a A2 postačuje známy lexikálny význam *niečo vlastniť* a modálny význam *mať niečo urobiť*. Na pokročilej úrovni by mali frekventanti rozoznať aj ďalšie významy.

ú do neznámeho prostredia , lebo to môže **mať** negatívny vplyv na ich ďalší výkonnostný r
keď sú tam a keď nie , tak **majú** veľa kritických pripomienok . Do úvahy vš
, kým ich nezarobím . To by sa **mali** naučiť aj kluby . Nie je predsa normálne
z nich , v Trnave či Trenčíne , **má** až s odstupom času cenu zlata . Keď
sa už ťažko dalo reagovať . „ Štrba **mal** situáciu , ktorá predchádzala vyrovnaniu ,
la . “ So skladbou stredovej formácie bude **mať** púchovský kormidelník pred sériou troch s
ajväčší podiel na chýbajúcich informáciách **mali** lekári , ktorí argumentovali lekársnym tajc
Neskôr sa vytvorila odborná skupina , ktorá **mala** vyšetriť okolnosti liečby prezidenta . Zisti
neho starali , nie . Povedal , že **má** záujem na objektívnom vyšetrení pochybe
bilingválnych knížiek vydavateľstva Petrus **má** po doterajšej angličtine premiéru aj nemč
lavicový experiment s reformou , ktorá by **mala** Slovensko priblížiť k fungujúcemu modelu

Obrázok 4: Konkordancia slovesa *mať* v korpuse *prim-6.1-public-vyv*

Na obrázku č. 4 vidíme kontexty z vyváženého korpusu *prim-6.1-public-vyv*, v ktorých sa sloveso *mať* nachádza v rôznych významoch.

1. Z viet vypíšte sloveso *mať* vo funkcii modálneho slovesa. (Úroveň A2.)

S ktorými ďalšími slovesami sa spája?

2. S ktorými substantívami sa spája sloveso *mať* vo význame *vlastniť*?

Pre potreby výučby slovenčiny ako cudzieho jazyka na vyššej úrovni môže lektor skúsiť vyhľadať v korpuse texty, v ktorých bude slovo *mať* vo funkcii slovesa aj podstatného mena. Frekventanti kurzu by mali identifikovať slovný druh na úrovni vety.

Ohýbanie slov. Cudzinci, ktorí nehovoria flektívnym typom jazyka ani taký nikdy neštudovali, mávajú s ohýbaním slov v slovenčine často problém. Slovo v inom ako základnom tvare vnímajú ako úplne nové slovo, a to im bráni porozumieť aj veľmi jednoduchým vetám. Pri precvičovaní ohýbania slov môže poslúžiť korpus, v ktorom vďaka lematizácii a morfolologickej anotácii vieme pracovať s tvarmi jednotlivých slov priamo v texte.

Prostredníctvom základného vyhľadávania vo vyhľadávači NoSketch Engine zadáme v položke „lemma“ substantívum *sused* a kliknutím na tlačidlo „hľadať“ sa zobrazí konkordancia kontextov tohto slova vo všetkých jeho tvaroch, ktoré sa v korpuse nachádzajú. Po kliknutí na „Zobrazenie – vlastné“ na ľavej strane obrazovky môžeme zadať „Atribúty – tag“ a zobrazia sa morfologické značky, ktoré sú vysvetlené na internetovej stránke Slovenského národného korpusu (<http://korpus.sk/morpho.html>). Takýto zápis môže slúžiť napríklad ako kontrola alebo pomôcka pre študentov (obr. 5).

Uložiť	SME94/05	to zároveň aj majstrovstvá Dolného Rakúska, od susedov /SSmp2 príde asi stovka bežcov, očakávame,
Zobrazenie	ACamb1	poobzeráte okolo seba a zachránite jedného zo svojich susedov /SSmp2 pred samotou, potom tieto slová nepíš
<u>Vlastné</u>	VWoo1	hry, spisovať zákony, stretávať sa so susedmi /SSmp7 a zhovárať sa s cudzincami zo zahranič
<u>KWIC</u>	MCel1	. Pozerajú, ako sa správajú spolužiaci alebo susedia /SSmp1 a hovoria si : Som slušnejší ako oni
<u>Veta</u>	JSar1	. Červenkastá ruka nepatrila jemu, lež jeho susedovi /SSms3, chlapovi s čiernymi fúzmi . Ten fuzáň
Triedenie	MPek1	morom a nemusia sa preto báť expanzie mocnejších susedov /SSmp2 . „ Samostatný, zvláštny slovenský nárc
<u>Vlastné</u>	HN05/10	vyvarovalo totálnej občianskej vojny aj vojny so svojimi susedmi /SSmp7 . Neľahká úloha integrovať Rusko zostá
<u>Vľavo Vpravo</u>	BLOG	a museli sme si ho ísť požičať do susedov /SSmp2 . Ujo doniesol veľký vandlík a do neho
<u>KWIC</u>	PODV 02/05	republiky . Perfektné ! Samozrejme, ak zobudíte suseda /SSms4, ktorý vás odvezie do Humenného autc
	LŤaž15	. . Ty mlč, Perun, tvoj sused /SSms1 vie orať, ale nevie rozprávať ? Viem

Obrázok 5: Konkordancia lemy *sused* v korpuse *prim-6.1-public-vyv* s morfologickými značkami. prvých 10 výskytov

Lektori môžu sami vybrať slovo len v určitom páde, čísle alebo rode, a to prostredníctvom vyhľadávania cez CQL, ktoré sme opísali pri prvej úlohe. Po zadaní voľby „vložiť tag“ vyberú atribúty, ktoré má dané slovo alebo celé vyhľadávanie obsahovať, napríklad: adjektívum a substantívum, ktoré ležia bezprostredne vedľa seba a sú v singulári, v ženskom a mužskom rode, životné, v genitíve ([tag="A.[mf]s2.*"] [tag="S.[if]s2.*"]). Na obrázku č. 6 vidíme dáta vygenerované zo všetkých textov vyváženého korpusu *prim-6.1-public-vyv*.

Vyberieme stránku s primeranými, zrozumiteľnými kontextami a môžeme vytvárať úlohy na precvičovanie genitívu.

Uložiť	MThu17	technika ich zrejme trochu mátie . Spomínaní absolventi	jazykovej školy	ľudoopov boli všetko šimpanzy . Prekonáva
Zobrazenie	SMar6	1993) . V roku 1990 absolvoval z	byzantskej hudby	magna cum laude na hudobnej škole F .
Vlastné	PDvo6	- Kričku možno uzavrieť v roku 1947 nálezom	veľkomoravskej osady	v Liptovskom Michale , ktorá nezanikla ani
KWIC	Kor00/12	zamestnancov televízie , ktorí sa nechcú stať figúrami	politckej manipulácie	" opozičnej zmluvy " ČSSD - ODS ,
Veta	HN09/10	Danone dosiahol v 3 . štvrtroku výnosy zo	základnej činnosti	3 , 78 miliardy eur (113 ,
Triedenie	MAT199/1	tretej kritickej amplitúdy a medzou cyklickej elasticity ,	cyklickej citlivosti	a prípadne medze únavy (rovn . 2
Vlastné	HN08/07	americký tajný úrad a jeho šéf z obrovského	diabolského tvora	s rohmí na hlave vychoval bojovníka na str
Vľavo Vpravo	PODV 03/31	trvá ošetrovanie rajčiakov proti fytoftóre , septorióze a	alternatívovej škvrnitosti	, zeleru proti septorióvej škvrnitosti listov
KWIC	DDvor1	- 1411 a v sprostredkovaní mieru - cesta	poľského kráľa	do Uhorska - vojenská výprava proti Benáti
	SME9/07	Od západu sa do strednej Európy rozširuje výbežok	tlakovej výše	, avšak počasie v našom regióne ešte ovply

Obrázok 6: Konkordancia adjektíva a substantíva ženského a mužského rodu, životné, v singulári, v genitíve vo vyváženom korpuse prim-6.1-public-vyv

1. Vyberte z textov 3 spojenia, ktorým rozumiete, a napíšte ich v základnom tvare (úroveň A2).
2. S ktorými predložkami sa viaže genitív v zobrazených vetách? (Úroveň A2.)
3. Z viet vypíšte substantíva, na ktoré sa viaže spojenie v genitíve. (Úroveň A2.)
4. Vyberte jednu vetu, v ktorej zmeníte spojenie v genitíve na spojenie v akuzatíve. Veta by mala ostať významovo podobná. (Úroveň B1.)

Pri výučbe pádov v slovenčine môže lektorom poslúžiť možnosť zvoliť lemu priamo v konkordancii kľúčového slova (obr. 7). V možnosti „Zobrazenie – Vlastné“ na ľavej strane obrazovky zvolíme atribút „lemma“, a tak sa kľúčové slovo zobrazí iba v základnom tvare priamo vo vete. Frekventanti kurzu slovenčiny takvidia zmenu tvaru slova a môžu precvičovať tvorenie jeho správnych tvarov.

Na obrázku č. 7 sú kontexty s vyznačeným hľadaným výrazom v základnom tvare (substantívum + sloveso *kráčať*).

Uložiť	MThu17	technika ich zrejme trochu mátie . Spomínani absolventi	jazykovej školy	ľudoopov boli všetko šimpanzy . Prekonáva
Zobrazenie	SMar6	1993) . V roku 1990 absolvoval z	byzantskej hudby	magna cum laude na hudobnej škole F .
Vlastné	PDvo6	- Kričku možno uzavrieť v roku 1947 nálezom	veľkomoravskej osady	v Liptovskom Michale , ktorá nezanikla ani
KWIC	Kor00/12	zamestnancov televízie , ktorí sa nechcú stať figúrami	politckej manipulácie	" opozičnej zmluvy " ČSSD - ODS ,
Veta	HN09/10	Danone dosiahol v 3 . štvrtroku výnosy zo	základnej činnosti	3 , 78 miliardy eur (113 ,
Triedenie	MAT199/1	tretej kritckej amplitúdy a medzou cyklickej elasticity ,	cyklickej citlivosti	a prípadne medze únavy (rovn . 2
Vlastné	HN08/07	americký tajný úrad a jeho šéf z obrovského	diabolského tvora	s rohmi na hlave vychoval bojovníka na stri
Vľavo Vpravo	PODV 03/31	trvá ošetrovanie rajčiakov proti fytoftóre , septorióze a	alternatívnej škvrnitosti	, zeleru proti septóriovej škvrnitosti listov
KWIC	DDvor1	- 1411 a v sprostredkovaní mieru - cesta	poľského kráľa	do Uhorska - vojenská výprava proti Benáti
	SME9/07	Od západu sa do strednej Európy rozširuje výbežok	tlakovej výše	, avšak počasie v našom regióne ešte ovply

Obrázok 7: Zobrazenie lemy kľúčového výrazu v konkordancii substantívum + verbum v korpuse prim-6.1-public-vyv

1. V textoch nájdite spojenia, v ktorých je podľa kontextu sloveso *kráčať* spojené s 3. osobou singuláru alebo plurálu. Vypíšte tieto spojenia gramaticky správne. (Úroveň A2 – B1.)

2. Nájdite spojenia, ktoré sú v inštrumentáli a správne ich prepíšte – vety, kde je pred vyznačeným spojením predložka *s*. (Úroveň A2, ak by išlo o jednoduchšie vety aj úroveň A1.)

3. Vypíšte spojenie, v ktorom sa sloveso *kráčať* neviaže na vyznačené substantívum (*kráčať párik / kráčať párik detí*). (Úroveň A2 – B1.)

- Spojenie dajte do správneho tvaru a napíšte, v akých pádoch sú substantíva v tomto spojení (v *akuzatíve* a *genitíve*). (Úroveň A2)
- Dajte do správneho tvaru substantívum vo vyznačenom spojení a napíšte:

A. v akom je rode a páde

B. celé slovné spojenie, kde sa toto substantívum nachádza

4. V ktorých vetách nasleduje za slovesom *kráčať* príslovka? Príslovky vypíšte (*šľachta, ďalej, vpred*) (úroveň A2).

Lektor slovenčiny ako cudzieho jazyka si pri učení pádov môže pomôcť aj *paradigmami podstatných mien*, ktoré sa zobrazia po vyhľadani zadaného výrazu na konci stránky <http://slovníky.juls.savba.sk> alebo <http://slovníky.korpus.sk/> spolu s vybranými vetami z korpusu, v ktorých sa dané substantívum nachádza.

ZÁVER

V príspevku sme priblížili niektoré vlastnosti a funkcie korpusových databáz využiteľné predovšetkým pri tvorbe cvičení pre záujemcov o štúdium slovenčiny ako cudzieho jazyka. Prostredníctvom konkrétnych ukážok sme predstavili praktické využitie korpusu pri osvojovaní si slovnej zásoby cez sémantické vlastnosti slov a ich spájateľnosť s inými slovami alebo na precvičovanie takého zložitého javu, ako je deklinácia podstatných mien či časovanie slovíes.

Vzhľadom na obmedzený priestor sme sa o viacerých jazykových a textových zdrojoch SNK mohli zmieniť iba veľmi stručne. Práca s nimi si vyžaduje určitú zručnosť a vedomosti o tom, ako korpus využívať, preto Oddelenie Slovenského národného korpusu Jazykovedného ústavu Ľ. Štúra SAV poskytuje záujemcom bezplatné školenia, o ktorých informuje na stránke <http://korpus.sk>.

ZOZNAM POUŽITEJ LITERATÚRY

ASTON, Guy et al.: *Learning with corpora*. Bologna : CLUEB, 2001, 277 s.

CHLPÍKOVÁ, Katarína: *Využitie korpusovej databázy vo výučbe slovenčiny ako cudzieho jazyka. Záverečná práca odborného kurzu Didaktika slovenčiny ako cudzieho jazyka*. Studia Academica Slovaca Univerzita Komenského. Bratislava 2015, 28 s.

GAJDOŠOVÁ, Katarína – IVORÍKOVÁ, Helena. *Slovenský národný korpus a možnosti jeho využitia v školskej praxi*. In *Slovenčinár*, 2014, roč. 1, č. 1, s. 44–60.

IVORÍKOVÁ, H.: *Využitie Slovenského národného korpusu (SNK) v školskej praxi (SNK ako pomôcka pri vyučovaní slovenského jazyka)*. 2006, dostupné na WWW: <http://korpus.sk/attachments/publications/2006-ivorikova-vyuzitie%20snk.pdf>

KAMENÁROVÁ, Renáta – ŠPANOVÁ, Eva – TICHÁ, Hana – IVORÍKOVÁ, Helena – KLESCHTOVÁ, Zuzana – MOŠAŤOVÁ, Michaela: *Krížom-krážom. Slovenčina A1*. Bratislava: Studia Academica Slovaca Univerzita Komenského, 2007, 187 s. ISBN 978–80–223–2441–0

O'KEEFFE, Anne – MCCARTHY, Michael – RACRET, Ronald: *From Corpus to Classroom. Language use and language teaching*. Cambridge University Press, 2007, 315 s. ISBN 9780511497650

PEKAROVIČOVÁ, Jana – ŽIGOVÁ, Ľudmila – MOŠAŤOVÁ, Michaela: *Vzdelávací program slovenčina ako cudzí jazyk*. Studia Academica Slovaca Univerzita Komenského. Bratislava: STIMUL, 2007, 83 s. ISBN 978-80-89236-28-2

ŠIMKOVÁ, M.: *Slovenský národný korpus ako pomôcka pri výučbe slovenského jazyka*. In K problematike vyučovania materinského jazyka a literatúry 2. Bratislava : Univerzita Komenského, 2007, s. 64-77. ISBN 978-80-223-2389-5

ULIČNÁ, Martina – ANDOROVÁ, Iveta – BÁČKAIOVÁ, Klaudia – GABRÍKOVÁ, Adela: *Tri, dva, jeden – slovenčina. Úroveň A1, A2.1*. Bratislava: Univerzita Komenského, Jazyková a odborná príprava cudzincov a krajanov ČĎV, 2015, 204 s. ISBN 978-80-223-3869-1

ZEBEGNEYOVÁ, Angela – PUZDEROVÁ, Anna – BAKOVÁ, Beáta: *Hovorme spolu po slovensky! Slovenčina ako cudzí jazyka A*. Bratislava: Univerzita Komenského, Ústav jazykovej a odbornej prípravy zahraničných študentov ČĎV, 2010, 166 s. ISBN 978-80-223-2945-3

Internetové zdroje:

<http://korpus.sk>

www.slovník.juls.savba.sk

www.slovník.korpus.sk

<http://bonito.korpus.sk>