

lí slovníkovej práce ďalšiu generáciu. Za veľký pracovný a intelektuálny vklad si tento kolektív spolu so Slovníkom súčasného slovenského jazyka oprávnenne zaslúžili certifikát Slovak Gold.

*Mária Pisárčiková*

## *Slovenský národný korpus v druhej etape*

MÁRIA ŠIMKOVÁ

V roku 2001 bol vo štvrtom čísle časopisu Kultúra slova publikovaný príspevok A. Jarošovej *Národný korpus slovenského jazyka: lingvistické a počítačové aspekty*, v ktorom autorka informovala o jednom z dôležitých fenoménov súčasnej svetovej lingvistiky a rozvoja informačných technológií – o jazykových korpusoch. Upozornila aj na nevyhnutnosť budovať elektronický korpus textov slovenského jazyka v súvisi s potrebou adekvátnej materiálovej bázy na koncipovanie nového výkladového slovníka súčasného jazyka (jeho prvý zväzok A – G vyšiel v r. 2006), na aktualizáciu existujúcich lexikografických a pravopisných príručiek (Krátky slovník slovenského jazyka, Pravidlá slovenského pravopisu a pod.) i gramatických diel (morfologický a syntaktický opis slovenčiny) a, ako autorka zdôraznila v závere, v súvisi „nielen s elektronizáciou základného jazykovedného výskumu, s kodifikačnými príručkami a koncepciou starostlivosti o štátny jazyk, ale... [korpus] je dôležitým predpokladom prežitia „malého“ jazyka v budúcej informačnej spoločnosti“ (Jarošová, 2001, s. 193 – 199). Budovanie Národného korpusu slovenského jazyka v Jazykovednom ústave E. Štúra SAV (ďalej JÚLŠ SAV) a elektronizácia jazykovedného výskumu na Slovensku sa stali skutočnosťou na základe uznesenia vlády Slovenskej republiky č. 137 z 13. 2. 2002 podľa návrhu Slovenskej akadémie vied, Ministerstva školstva Slovenskej republiky a Ministerstva kultúry Slovenskej republiky, ktoré je ústredným orgánom štátnej správy v oblasti

starostlivosti o štátny jazyk. Projekt bol v predkladanej podobe schválený do r. 2006 (podrobnejšie rozpracovaný a publikovaný v štúdiách M. Šimkovej, 2003, 2004), od 1. 1. 2007 pokračuje ***Budovanie Slovenského národného korpusu a elektronizácia jazykovedného výskumu na Slovensku (druhá etapa)*** na základe osobitnej zmluvy medzi MŠ SR, MK SR a SAV. V tomto príspevku zrekapitulujeme výsledky riešenia projektu po piatich rokoch a predstavíme plány na ďalšie obdobie.

Veľkej časti počítačovej verejnosti na Slovensku, ale i viacerým záujemcom zo zahraničia je už dôverne známa internetová stránka Jazykovedného ústavu L. Štúra SAV ([www.juls.savba.sk](http://www.juls.savba.sk)), a to najmä vďaka ponuke lingvistických zdrojov a slovníkov. Všetkým používateľom slovenského jazyka sú na nej online bezplatne prístupné elektronické verzie aktuálnych vydaní ***základných kodifikačných príručiek*** (Krátky slovník slovenského jazyka, Pravidlá slovenského pravopisu), ale i starší slovník väčšieho rozsahu (Slovník slovenského jazyka, 1959 – 1968), ktorý síce nepatrí medzi súčasné kodifikačné príručky, no poskytuje obraz o vývine, dynamike súčasnej slovenčiny a obsahuje aj množstvo dodnes platných spisovných slov a tvarov, ktorými rozširuje možnosti komunikácie v slovenskom jazyku. (Žiaľ, stretávame sa aj s takou [obmedzenou] vydavateľskou praxou, podľa ktorej slová nenachádzajúce sa v Krátkom slovníku slovenského jazyka nepatria do spisovnej slovnej zásoby, čo je prejavom veľkého nepochopenia zamerania a rozsahu slovníka spracúvajúceho jadro slovnej zásoby a okliešťovaním slovenčiny na 60-tisíc heslových slov.) Medzi lexikografické príručky sú z hľadiska obsahu a techniky spracovania zaradené aj onomastické príručky, resp. zdroje – Názvy obcí Slovenskej republiky (Vývin v rokoch 1773 – 1997) a Priezviská na Slovensku (podľa stavu v roku 1995). V uvedených príručkách a zdrojoch je možné vyhľadávať vo všetkých naraz alebo jednotlivito, či vo viacerých podľa výberu. Spôsob práce je síce odlišný od listovania v tlačenej príručke, ale má svoje nesporné výhody: nájdenie hľadaného slova vo všetkých zdrojoch súčasne, možnosti vyhľadávania podľa prefixov, suffixov, časti slova a pod. Po zadaní hľadaného slova a zobrazení heslovej state sa pri základných údajoch o konkrétnej príručke dá prezrieť aj úvod k jej tlačenej verzii alebo iné informácie o nej. Osobitné skupiny v položke Lingvistické zdroje tvoria elektronické verzie vybraných monografií a zborníkov z produkcie JÚLEŠ SAV alebo klasických autorov

(napr. už inak nedostupná akademická Morfológia slovenského jazyka z r. 1966, Rukoväť spisovnej reči slovenskej od S. Cambela z r. 1902), ako aj posledných vyše desať ročníkov všetkých časopisov vydávaných v JÚLŠ SAV. Elektronického spracovania a internetového sprístupnenia sa dostalo aj bibliografii Slovenskí jazykovedci od L. Dvonča, ktorej štyri zväzky (od r. 1925 do r. 2000) predstavujú unikátny súbor bibliografie vednej disciplíny nielen na Slovensku. Malým príspevkom technických možností a zaujímavou školskou pomôckou je automatický prekladač do štúrovskej slovenčiny, ktorý vznikol ako ukážka možnosti automatického spracovania zmien v ortografii. Špecifickou položkou v lingvistických zdrojoch je skúšobná verzia Slovenskej terminologickej databázy, ktorej cieľom je prispieť k zlepšovaniu odbornej komunikácie a zvyšovaniu terminologickej kultúry na Slovensku.

Okrem poslednej položky – Slovenskej terminologickej databázy, ktorá patrí medzi základné úlohy v rámci budovania Slovenského národného korpusu a elektronizácie jazykovedného výskumu na Slovensku, celý súbor lingvistických zdrojov sprístupnených na stránke JÚLŠ SAV je akoby vedľajším produktom tohto projektu. Nemalé investície do digitalizácie a technického spracovania uvedených zdrojov sa však bohato vrátili a vracajú: stránka má priemerne 30-tisícovú dennú návštevnosť, ľudia pracujúci so slovom majú ľahký prístup k základným kodifikačným príručkám slovenského jazyka, čo nám viacerí lingvisti v zahraničí úprimne závidia, mladí ľudia si dávajú linku JÚLŠ SAV do svojich obľúbených položiek (vidíme to v ich blogoch), na internete sa diskutuje o slovenčine a jej používaní. On-line prístup k lingvistickým zdrojom veľmi oceňujú lektori slovenského jazyka a iní záujemcovia o slovenčinu v zahraničí, ktorí inak nemajú veľa možností dostať sa k aktuálnym lexikografickým príručkám. (WWW adresa JÚLŠ SAV sa nachádza napr. aj na takej stránke, ako je [How-to-learn-any-language.com](http://How-to-learn-any-language.com) – The website about teaching yourself languages.) V jazykovej poradni JÚLŠ SAV sa možno o niečo znížil počet bežných otázok, na ktoré si záujemcovia môžu nájsť odpoveď sami jednoduchým vyhľadaním v kodifikačných príručkách na internete, na druhej strane sa niektorým používateľom zložitejšie orientuje v rokoch vydania, a tak sa informujú o možnej platnosti slov zo staršieho slovníka. Postupným pridávaním ďalších zdrojov, ako sú napr. aktuálne sprístupnené Pravidlá slovenského pravopisu z r. 1931 a 1940, akademický Slovník cudzích slov, ukážka z Historického slovníka slovenského jazyka

v podobe 5. zväzku (v rezerve je Synonymický slovník slovenčiny, výber zo Slovníka slovenských nárečí; nový Slovník súčasného slovenského jazyka sa plánuje sprístupňovať cyklicky: pri vydaní ďalšieho zväzku tlačou sa predchádzajúci sprístupní elektronicky), si návštevníci tejto stránky azda zvyknú všímať základné informácie a rozlišovať jednotlivé zdroje, ktorých elektronické spracovanie je neoceniteľnou pomôckou pri výskume jazyka i pri práci s ním.

Základnou úlohou skončeného i nadväzujúceho projektu, avšak nie s takým priamočiarym využitím, ako sme uviedli, je budovanie Slovenského národného korpusu v celej šírke štýlov, žánrov a vecných oblastí, ale aj regiónov, vydavateľstiev, generácií a pod., ktoré bolo v prvej fáze ohraničené na písané texty z obdobia rokov 1955 – 2005, v druhej fáze sa v spracúvaní písaných textov pokračuje tak, aby boli zachytené jednak tie najaktuálnejšie, jednak sa začína vytvárať samostatný celok textov pred roka 1955. Okrem písaných textov sa pripravuje aj korpus hovorenej slovenčiny, budujú sa paralelné korpusy a pod.

Slovenský národný korpus ako elektronický súbor jazykových dát s výkonnými nástrojmi na vyhľadávanie a triedenie skúmaných jazykových prostriedkov je od r. 2003 prístupný cez stránku JÚLŠ SAV (v položke Lingvistické zdroje, ale i v položke O ústave/Oddelenia), samostatne na adrese <http://korpus.juls.savba.sk>, neskôr bol odkaz naň zaradený aj na stránku Ministerstva školstva SR. Začiatkom r. 2007 sa sprístupnila najnovšia, šiesta verzia hlavného, základného korpusu *prim-3.0*, ktorá obsahuje 350 miliónov textových jednotiek (okrem slov je to aj interpunkcia a číslicové či iné neslovné zápisy, ako emotikony a pod.), a spolu s ňou druhá verzia ručne morfológicky anotovaného korpusu v rozsahu takmer 512-tisíc textových jednotiek. Každý text v korpuse je podložený súhlasom autora alebo majiteľa autorských či distribučných práv na jeho spracovanie a zaradenie do celku korpusu podľa licenčnej zmluvy a má podrobnú bibliografickú a štýlovo-žánrovú anotáciu. Celý korpus je automatizovane lematizovaný (každý slovný tvar má pri sebe informáciu o základnom tvare – leme) a automatizovane morfológicky označovaný po natrénovaní softvéru na ručne morfológicky anotovaných textoch. Výbrané texty sa ručne anotujú aj syntakticky. O postupoch pri získavaní textov, ako aj o princípoch ich spracovania od technického čistenia a konvertovania do jednotného formátu cez segmentá-

ciu až po jednotlivé úrovne anotácie nájde používateľ podrobné informácie na stránke Slovenského národného korpusu.

Práca s korpusom je veľmi jednoduchá. Neregistrovaný používateľ môže napísať do vyhľadávacieho okienka na stránke Slovenského národného korpusu slovo v jeho konkrétnom alebo základnom tvare a zobrazia sa mu kontexty jeho výskytov v rozsahu max. 100 textových jednotiek v podobe konkordančného výpisu, t. j. hľadané slovo je v strede riadka, zľava a sprava je príslušný kontext. Pri každom výskyte si možno v ľavom stĺpci zobraziť plnú bibliografickú a štýlovo-žánrovú anotáciu zdrojového textu. Takto má neregistrovaný používateľ k dispozícii jednu, aktuálnu verziu veľkého korpusu *prim* a aktuálnu verziu ručne morfológicky anotovaného korpusu. Voľne môže vyhľadávať aj v paralelných korpusoch a ostatných zdrojoch.

Registrovaní používatelia dostávajú na základe podpísania podmienok používania korpusu osobitný prístup s heslom na prácu s korpusom prostredníctvom korpusového manažéra Manatee s klientom Bonito a majú k dispozícii všetky verzie a podkorpusy (napr. samostatný podkorpus publicistických textov, samostatný podkorpus pôvodnej slovenskej umeleckej tvorby a pod.). V nich môžu nielen jednoducho vyhľadávať, ale uvedený počítačový nástroj im umožňuje vyhľadané kontexty (neraz desiatky tisícov dokladov) triediť, zisťovať ich rôzne štatistické hodnoty a distribúcie, vytvárať kolokácie a pod. Vyselektovaný materiál si môžu uložiť do vlastného počítača a ďalej s ním pracovať. Všetci používatelia sú viazaní používať korpus výlučne na vedeckovýskumné a iné nekomerčné ciele a citovať korpus i jednotlivé zdroje v súlade s autorským zákonom.

Využívanie Slovenského národného korpusu (každoročne vyše 200 registrovaných používateľov, neregistrovaní sú zahrnutí v už spomínanej 30-tisícovej dennej návštevnosti) sa realizuje vo všetkých oblastiach, pre ktoré sa jazykové korpusy budujú:

a) Na Slovensku sa začala systematicky rozvíjať nová vedná disciplína – *korpusová lingvistiká* ako odbor (počítačovej) lingvistiky, ktorej predmetom je skúmanie jazykových javov v prirodzených kontextoch vo veľkom množstve reálnych textov. Na základe analýzy korpusových textov sa overujú doterajšie lingvistické teórie a môžu vzniknúť nové hypotézy a teórie, čoho dôkazom je aj v r. 2006 ukončený grant *Morfosyntaktická analýza Slovenského národného korpusu* (v spolupráci s FF PU Prešov) a zborník

*Sondy do morfosyntaktického výskumu slovenčiny na korpusovom materiáli* (2006) či 2. zväzok *Valenčného slovníka slovenských slovies (na korpusovom základe)* autorky J. Nižníkovej (Prešov 2006), monografia M. Ivanovej *Valencia statických slovies* (Prešov 2006), ako aj viacero jazykovedne zameraných diplomových prác. Najvýznamnejšou aplikačnou zložkou v tejto oblasti je lexikografické využitie: mnoho korpusov sa budovalo a buduje na podporu tvorby slovníkov a lexikografi sú v súčasnosti azda najčastejšími používateľmi korpusov – na báze materiálu Slovenského národného korpusu sa koncipuje nový 8-zväzkový Slovník súčasného slovenského jazyka a aktualizujú sa jednotlivé vydania doterajších lexikografických i pravopisných príručiek.

b) Niektoré výsledky zo spracovania korpusu, ako sú zoznamy slov, spoločné výskyty slov (kolokácie), frekvencia grafém, slabík, slov a spojení atď., sa používajú aj v líniových aplikáciách – v neurológii, logopédii, psychológii, didaktike a pod., no predovšetkým v oblasti **počítačového spracovania prirodzeného jazyka**, kde sa tvoria napr. systémy na spracovanie textov (automatická kontrola pravopisu, gramatiky či štylistiky, strojový preklad textov), systémy na rozpoznávanie reči a pod. Na materiáli Slovenského národného korpusu vzniklo viacero diplomových a doktorských prác na vysokých školách a výskumných pracoviskách technického zamerania, v záverečnom štádiu je vývoj vlastného morfologického analyzátora tvarov slovenského jazyka.

c) Korpus môže byť a je dobrým zdrojom príkladov a viet potrebných **pri výučbe cudzieho, ale aj materinského jazyka**, využívajú ho domáci učitelia i lektori slovenského jazyka v zahraničí na prípravu rôznych cvičení (porov. Ivoríková, 2006; Šimková, 2007). Osobitne vítanou pomôckou v tejto oblasti sú paralelné korpusy, z ktorých sú v rámci Slovenského národného korpusu sprístupnené zatiaľ tri (Parallel Corpus of Computer Terms, Francúzsko-slovenský paralelný korpus, Rusko-slovenský paralelný korpus) a postupne sa plánujú ďalšie (najbližšie by to mal byť slovensko-český, slovensko-chorvátsky a slovensko-anglický paralelný korpus).

Jazykové korpusy, a teda aj Slovenský národný korpus, zo svojej podstaty slúžia ako referenčný zdroj, ako materiál na ďalšie výskumy. Ako súbor veľkého množstva textov (väčšinou už publikovaných knižne alebo v printových médiách) poskytuje korpus možnosť výskumu reálneho fungovania

jazyka v praxi aj v širšom časovom horizonte, nie je to však slovník, podľa ktorého by sa používatelia mali orientovať v spisovnosti/nespisovnosti či správnosti/nesprávosti jazykových prostriedkov. Texty sa do korpusu získavajú v elektronickej podobe z vydavateľstiev alebo od autorov tak, ako vyšli tlačou, alebo sa z tlačenej podoby (najmä pri starších textoch, spreď r. 1995) vyhotovuje technická rozmnoženina skenovaním, rozpoznávaním a rekonštruovaním, opäť však v podobe identickej s pôvodne vydaným textom. V korpuse sa preto nachádzajú aj rôzne chyby – preklepy, ale i pravopisné či gramatické chyby, o štylistických nehovoriac, neraz však aj jazykové hry a rôzne okazionálne tvary, ktoré sa pri tvorbe korpusu nemôžu opravovať. Stámiliónový rozsah textov v korpuse nie je korektorsky reálne zvládnuteľné a ich korigovanie ani nie je úlohou tvorcov korpusov – väčšina textov prešla jazykovou úpravou vo vydavateľstvách. Ak by sa aj taká úloha mala formulovať, často by sa muselo náročne rozhodovať, čo opraviť a čo nie jednak zo štylistického hľadiska a jednak vzhľadom na to, že korpus sa nebuduje na jednorazové či krátkodobé použitie a dynamika jazyka ovplyvňuje aj dynamiku pravopisnej normy. Realnosť a prirodzená nedokonalosť textov v korpuse má navyše svoje využitie pri výskume najčastejších/najtypickejších chýb a spätnom zdokonaľovaní metodiky výučby pravopisu či strojopisu. Programy na kontrolu a opravu pravopisu sa takisto môžu na korpusoch tvoriť, pretože to, čo je naozaj chybné, je pri veľkom objeme dát štatisticky zanedbateľné a počítačové nástroje takéto jednotlivosti neberú do úvahy, resp. „vedia“ rozlíšiť a správne interpretovať výskyt nejakého tvaru vo veľmi odlišnom rozsahu (napr. niekoľkodesaťtisícová frekvencia správneho tvaru a 5 či 6 výskytov nesprávneho tvaru), pričom im pomáha slovník paradigiem z morfológického analyzátoru. Pravda, takto sa môžu zo spracovania stratiť okrajové javy a aj preto sa v súčasnosti budujú korpusy stále väčšieho rozsahu, aby sa zachytili aj periférne jazykové prostriedky.

Jednou z najčastejších otázok, s ktorými sa stretávame, je: Koľko slov má slovenčina a ktoré sú najviac a najmenej frekventované slová. Na prostrednú časť otázky sa odpovedá najjednoduchšie. Zoznam najfrekventovanejších slov a lem sa sprístupňuje spolu s novou verziou korpusu na internete (<http://korpus.juls.savba.sk/stats/>). Na prvých miestach sa okrem bodky a čiarky nachádzajú predložky *v*, *na*, spojka *a*, polyfunkčné slovko *sa* a pomocné sloveso *byť*. Kým pri najčastejších výskytoch sa hodnoty jednot-

livých slov nachádzajú v dost' veľkých odstupoch od seba, v strede tabuľky sa stáva, že rovnakú frekvenciu majú dve, tri i viaceré slová, a na konci sú slová/tvary s výskytom 3, 2 i 1, ktorých je však na týchto posledných miestach aj niekoľko tisíc s rovnakou frekvenciou. Preto sa nedá jednoznačne povedať, ktoré slovo má najnižšiu frekvenciu. Je ich veľa a okrem rôznych nesprávnych zápisov sú to zväčša cudzie alebo archaické slová, úzko špecializované termíny a pod. Ich ojedinelý až unikátny výskyt v korpuse však môže byť spôsobený nedostatkom textov z istej oblasti a po rozšírení korpusu sa môže zmeniť. Pri dobrej koncepcii budovania korpusu a prihliadaní na vyváženosť či reprezentatívnosť (primerané zastúpenie rôznorodých textov) by okrajové javy nikdy nemali mať extrémne vysokú frekvenciu, ale ich jestvovanie by malo byť dostatočne zachytené.

Najnáročnejšia je odpoveď na prvú časť otázky o počte slov v slovenčine, ale týka sa to akéhokoľvek jazyka. V učebniciach či slovníkoch cudzích jazykov sa dočítame, že na základné porozumenie a komunikáciu stačí zvládnuť tisíc až tritisíc slov. Najnovšie vydanie Krátkeho slovníka slovenského jazyka spracúva približne 60-tisíc heslových slov, Slovník slovenského jazyka z r. 1959 – 1968 obsahoval necelých 130-tisíc hesiel a v novom veľkom výkladovom slovníku slovenčiny sa plánuje spracovať okolo 240-tisíc heslových slov. To však ešte stále nie sú všetky slová, ktoré tvoria slovnú zásobu slovenčiny, ešte sú tu nárečia, historická slovná zásoba, termíny... S veľkou pravdepodobnosťou by sme presiahli číslo milión, a to značne. Ak uvádzame, že Slovenský národný korpus obsahuje v poslednej verzii zo začiatku tohto roka 350 miliónov textových jednotiek, ide o súčet výskytov rôznych slov, pričom počet pribúdajúcich nových unikátnych slov klesá priamo úmerne s veľkosťou korpusu. Korpusové štatistiky uvádzajú, že v korpuse, ktorý obsahuje 100 miliónov textových jednotiek, čo je v súčasnosti minimálna veľkosť všeobecného korpusu, sa 8-tisíc jednotiek nachádza v 95 percentách textu a zvyšných 5 percent reprezentuje 500-tisíc jednotiek. S narastaním korpusu sa tento pomer mení – počet jednotiek v 95 percentách sa zväčšuje, počet jednotiek vo zvyšných 5 percentách postupne klesá. Trend v budovaní korpusov je jednoznačný: čím viac dát, tým sú to lepšie dáta.

Slovenský národný korpus by sa mal v druhej etape svojho budovania rozrásť na 600 miliónov textových jednotiek v základnom korpuse písaných



textov od r. 1955 do súčasnosti, t. j. do r. 2011, dokedy bola druhá časť projektu schválená. Tento materiál sa bude využívať predovšetkým pri tvorbe 8-zväzkového Slovníka súčasného slovenského jazyka, ktorého koncipovanie potrvá minimálne do r. 2013, ale plánuje sa na ňom aj príprava ďalších slovníkov, ktoré prispejú k exaktnejšiemu poznaniu jazykového systému slovenčiny a typologickým i iným výskumom: frekvenčný, retrográdny, koločasný slovník. Priebežne sa bude pokračovať v sprístupňovaní lingvistickejých zdrojov a slovníkov na internete, pripravuje sa ich spracovanie a distribúcia aj na CD/DVD nosičoch rovnako ako aj časti hlavného korpusu. Rozširovanie paralelných korpusov sa zameria najmä na česko-slovenský a slovensko-český paralelný korpus, ktorý môže poslúžiť ako materiálová báza na tvorbu prekladového slovníka. Osobitnou súčasťou bude budovanie korpusu hovorenej slovenčiny, ktorej výskum sa stáva jednou z dôležitých úloh viacerých slovakistických pracovísk (Prešov, Banská Bystrica, Nitra, Bratislava), a tvorba Slovenskej terminologickej databázy, ktorá v súčasnosti obsahuje vyše 3000 termínov z rôznych oblastí a mala by sa rozšíriť najmä o terminológiu z oblasti práva a ekonómie.

Oddelenie Slovenského národného korpusu Jazykovedného ústavu E. Štúra SAV v Bratislave spolupracuje pri plnení jednotlivých úloh s pracoviskami na Slovensku i v zahraničí, každý druhý rok organizuje medzinárodnú konferenciu *Slovko* o počítačovom spracovaní prirodzených, predovšetkým slovanských jazykov a v plnom rozsahu naplňa všetky možnosti na elektronizáciu jazykovedného výskumu na Slovensku a stabilné postavenie slovenčiny v medzinárodnom kontexte rozvoja informačných technológií, čo pri hodnotení prvej fázy ocenili aj zahraniční posudzovatelia, keď Slovenský národný korpus zaradili medzi 5 – 8 špičkových jazykových korpusov na svete, aj vedenie SAV, keď tento kolektív odmenilo v r. 2005 Cenou Slovenskej akadémie vied za budovanie infraštruktúry pre vedu.

#### LITERATÚRA A ZDROJE

IVORÍKOVÁ, Helena: Využitie Slovenského národného korpusu (SNK) v školskej praxi. (SNK ako pomôcka pri vyučovaní slovenského jazyka.) <http://korpus.juls.savba.sk/publications/1> 2006.

JAROŠOVÁ, Alexandra: Národný korpus slovenského jazyka: lingvistické a počítačové aspekty. In: Kultúra slova, 2001, roč. 35, č. 4, s. 193 – 199.

ŠIMKOVÁ, Mária: Počítačové spracovanie prirodzeného jazyka a Slovenský národný korpus. In: Počítačová podpora prekladu. Zborník prednášok. Ed. M. Smolík – J. Šoltys – F. Tomášik. Bratislava: Slovenská spoločnosť prekladateľov odbornej literatúry 2003, s. 15 – 19.

ŠIMKOVÁ, Mária: Slovenský národný korpus – východiská a plány. In: Slovenčina na začiatku 21. storočia. Ed. M. Imrichová. Prešov: Fakulta humanitných a prírodných vied Prešovskej univerzity 2004, s. 150 – 158.

ŠIMKOVÁ, Mária: Slovenský národný korpus ako pomôcka pri vyučovaní slovenského jazyka. In: K problematike vyučovania materinského jazyka a literatúry. II. Ed. M. Vojtech. Bratislava: Univerzita Komenského 2007 (v tlači); <http://korpus.juls.savba.sk/publications/>

URL: <http://korpus.juls.savba.sk>

---

**KS**