

Plnenie úloh pri tvorbe a rozvoji Slovenského národného korpusu za rok 2021

V poslednom 5. roku riešenia úloh podľa Zmluvy o združení prostriedkov na tvorbu a rozvoj Slovenského národného korpusu (č. 0323/2017) sa členovia riešiteľského kolektívu Slovenského národného korpusu Jazykovedného ústavu Ľ. Štúra SAV sústreďovali najmä na finalizáciu dát do nových verzií korpusov podľa Harmonogramu na rok 2021 a na dokončenie a vydanie plánovanej publikácie. Z hľadiska personálneho zloženia došlo v riešiteľskom kolektíve k niekoľkým zmenám: 1. 1. 2021 nastúpili dvaja odborní pracovníci Mgr. M. Debnár, PhD. na 50 % úväzok a Mgr. L. Gibľáková, PhD. na 75 % úväzok, ktorá však z osobných dôvodov zotrvala v pracovnom kolektíve len do konca roka 2021. Okrem nej odišli z oddelenia ďalšie dve pracovníčky: Mgr. A. Veselá k 31. 7. 2021 a Mgr. K. Rausová k 1. 9. 2021, ktorá pokračuje v doktorandskom štúdiu a bude ďalej na projekte spolupracovať. Na začiatku roka 2021 bolo v oddelení SNK 13 pracovníkov, z toho 7 z nich malo čiastočný úväzok, čo celkovo predstavovalo deväť 100 % pracovných úväzkov a jeden 85 %, čím sa dodržalo ustanovenie čl. III bod 3. Zmluvy o 10 pracovných miestach.

Doplnenie realizácie úloh z roku 2019 podľa Dodatku č. 1

Úloha č. 4: Dokončenie a knižné vydanie publikácie *Časovanie slovíes v slovenčine s korpusovými príkladmi*.

Plnená priebežne, odovzdaná do recenzného konania: december 2021

Publikácia *Časovanie slovíes v slovenčine s korpusovými príkladmi*, ktorá mala byť podľa Dodatku č. 1 vydaná v roku 2019 a ktorej tvorba bola z personálnych dôvodov v roku 2020 prerušená, predstavuje nadviazanie na tradíciu počítačovo podporovaných spracovaní a výskumov jazyka či jednotlivých slovných druhov a javov, akým bolo napríklad metodologicky príbuzné počítačové spracovanie skloňovania slovenských substantív (Garabík et al., 2016: Skloňovanie podstatných mien v slovenčine s korpusovými príkladmi), ktoré bolo vydané v rámci predchádzajúceho projektového obdobia. Dielo trojice autorov M. Zúmríka, D. Majchrákovéj a L. Mihálikovej predstavuje analýzu výsledkov porovnávania morfolologickej databázy, využívanej pri morfolologickej anotácii korpusov, a ďalších databáz obsahujúcich slovesné časovanie, ktoré pracovníci Jazykovedného ústavu Ľ. Štúra SAV vytvorili v priebehu niekoľkých desaťročí. Predkladaná práca, pozostávajúca z predhovoru, metodologického a historického úvodu, analytickej a štatistickej časti, záveru a abecedného kľúča slovíes, je adresovaná širokému okruhu odborníkov i bežných používateľov slovenčiny s cieľom dôkladne opísať mimoriadnu tvarovú rôznorodosť konjugačného systému slovenských slovíes.

Publikácia, ktorá má v rukopise 350 normostrán, prešla recenzným konaním, v súčasnosti sa zapracúvajú pripomienky oboch recenzentov, pričom vydanie publikácie sa plánuje v marci 2022.

Doplnenie realizácie úloh z roku 2020

Úloha č. 11: Vytvorenie a sprístupnenie novej, rozšírenej verzie Slovenského hovoreného korpusu – s-hovor-7.0 a jeho podkorpusov.

Splnená, sprístupnenie: marec 2022

Nová verzia hovoreného korpusu bola zhotovená v máji 2021. Ako bolo vykázané v správe

za rok 2020, obsahuje 869 nahrávok, čo predstavuje necelých 852 hodín zvukových záznamov. Oproti predchádzajúcej verzii s-hovor-6.0 ide o nárast o vyše 100 nahrávok a viac ako 100 hodín zvukových záznamov. Celkovo má táto verzia 7,8 milióna tokenov, čo znamená, že v porovnaní s predchádzajúcou verziou bol zaznamenaný nárast o 1,2 milióna tokenov (k tomuto nárastu došlo v podkorpuse s-hovor-7.0-sane na celkových 4,9 milióna tokenov, kým rozsah podkorpusu s-hovor-6.0-upn zostal nezmenený).

Harmonogram riešenia úloh roku 2021

Harmonogram riešenia úloh (Príloha č. 2 Zmluvy) zahŕňa na rok 2021 deväť úloh, z toho jedna sa delí na dve časti.

Úloha č. 1 Sprístupnenie novej, skvalitnenej a rozšírenej verzie Slovenského národného korpusu – hlavného korpusu prim-10.0 a jeho podkorpusov

Plnená priebežne, sprístupnenie: marec 2022

V záujme skvalitnenia vnútornej anotácie všetkých spracovaných textov sa riešiteľský kolektív rozhodol implementovať viacero zmien, čo však na druhej strane spôsobilo posunutie termínu sprístupnenia nového korpusu všetkým používateľom. Zmeny sú spôsobené prechodom na dokonalejšie nástroje, vylepšením konfigurácie a zavedením nástrojov a modelov vytvorených na pôde SNK ako aj na iných pracoviskách JÚLEŠ SAV. K vylepšeniu morfolologickej anotácie dochádza aj vďaka rozšíreniu morfolologickej databázy.

Hlavný korpus je tvorený priebežne dopĺňanými textami a ich štýlovo-žánrovou anotáciou. Všetky jednotky v texte sú lematizované a morfologicky anotované. Vo **verzii prim-10.0** dochádza oproti prim-9.0 k výraznému zlepšeniu spracovania textu, menovite segmentácie, tokenizácie, lematizácie a morfolologickej anotácie.

K oneskoreniu finalizácie prác prispelo aj veľké množstvo zmien v spracovaní textov. Príprava dát hlavného korpusu sa uzavrela koncom februára 2022, rozdelenie na podkorpusy a zverejnenie korpusov sa uskutoční v priebehu marca 2022. Zatiaľ preto nie je možné poskytnúť štatistické ukazovatele hlavného korpusu a porovnať ho s predchádzajúcou verziou.

Úloha č. 2 Sprístupnenie novej, rozšírenej verzie webového korpusu slovenčiny – web-6.0

Plnená priebežne, zverejnenie korpusov: marec 2022

Webový korpus slovenčiny sa zakladá na dátach z korpusu Araneum Slovacom VI Maximum (22.01) vytvoreného v rámci projektu Aranea. Korpus obsahuje webové texty v slovenčine priebežne zbierané medzi rokmi 2013 – 2021. Segmentácia viet a tokenizácia sú prebraté z pôvodného korpusu Araneum Slovacom, no dáta sú lematizované a morfologicky anotované spôsobom kompatibilným s hlavným korpusom prim-10.0. Očakávaná veľkosť korpusu je 4.39 mld. tokenov. V porovnaní s predchádzajúcou piatou verziou web-5.0 z roku 2020 pôjde o nárast o približne pol miliardy tokenov.

Úloha č. 3 Vytvorenie a sprístupnenie nových verzií minimálne 2 paralelných korpusov

Plnená priebežne, sprístupnenie: február 2022

Z existujúcich 11 paralelných korpusov, ktoré sa od roku 2006 budujú v rámci projektu SNK, boli v roku 2021 zvolené na rozšírenie 2 korpusy: slovensko-nemecký a slovensko-španielsky paralelný korpus.

Prvý z nich – **slovensko-nemecký paralelný korpus** vznikol v roku 2014 a o dva roky neskôr bola sprístupnená jeho zväčšená a aktualizovaná druhá verzia par-skde-2.0 s rozsahom 446 miliónov tokenov. S odstupom piatich rokov bol tento korpus, respektíve jeho podkorpus beletrie, obohatený o vyše 170 beletristických titulov na výsledných 468 miliónov tokenov (v slovenskej časti sa nachádza 229 miliónov tokenov a v nemeckej časti 238 miliónov tokenov). Aktuálne obsahuje celkovo 147 kníh – z toho takmer 90 % tvoria preklady z nemčiny, 3,3 % preklady zo slovenčiny (celkovo 14 kníh) a zvyšných 10 % predstavuje preklady z tretích jazykov do nemčiny. Nové tituly v tretej verzii tohto paralelného korpusu zahŕňajú diela F. Kafku, H. Hesseho, P. Süskinda, R. Steinera či V. E. Frankla, čo nepochybne prispeje k väčším možnostiam komparatívneho alebo translatického výskumu.

Druhým rozšíreným korpusom je **slovensko-španielsky paralelný korpus**, ktorý vznikol v roku 2019 aj v spolupráci s Katedrou romanistiky z Filozofickej fakulty Univerzity Mateja Bela v Banskej Bystrici. Ide o jeden z menších paralelných korpusov, ktorého prvá verzia mala 11,5 milióna tokenov a pozostávala len z beletrie. Aj druhá verzia, vytvorená v roku 2021 sa skladá z textov krásnej literatúry, z ktorých takmer 57 % bolo napísaných v španielčine a 20 % vo francúzštine. Pôvodná slovenská tvorba so španielskym prekladovým náprotivkom predstavuje len 0,02 %. Zvyšných vyše 20 % bolo napísaných v iných, celkovo desiatich jazykoch a následne bolo preložených do španielčiny. Celkový rozsah takmer 36 miliónov tokenov predstavuje viac ako trojnásobok prvej verzie (18,9 milióna tokenov v španielskej časti a 16,7 milióna tokenov v slovenskej časti). Z hľadiska počtu titulov obsahuje nová verzia par-skes-2.0 celkovo 225 kníh, z čoho viac ako polovica (148 kníh) bola pridaná práve pri jeho rozširovaní v roku 2021 vrátane diel J. Amada, J. Cortázara, M. Ángela Asturiasa, M. de Cervantesa Saavedru, M. O. Silvu či P. Coelhoa.

Úloha č. 8 Spracovanie a sprístupnenie minimálne 500 nových termínov v Slovenskej terminologickej databáze: kategória právo

Plnená priebežne

V podprojekte Slovenskej terminologickej databázy sa v prvom rade spracúvala **terminológia práva** v súlade s aktuálnou úlohou zakotvenou v Harmonograme – celkovo bolo vypracovaných a sprístupnených vyše **550 terminologických záznamov**. Spracované termíny možno zaradiť do oblasti ústavného, trestného, správneho, obchodného či rodinného práva, ale aj z teórie práva. Celý súbor termínov bude prechádzať odbornou revíziou a bude ďalej rozširovaný. Riešiteľský tím sa opieral o viacero odborných encyklopedických a terminografických zdrojov (*Právnický slovník* 2009, *Trestné právo procesné* 2006, *Trestné právo hmotné* 2010, *Slovník slovenského práva* 2000, ako aj o odborné tezaury sprístupnené na portáli <https://www.slov-lex.sk>.

Paralelne s prácou na terminológii práva prebiehala odborná revízia a korektúra terminologickej zbierky epidemiológia a imunológia, na základe ktorej sa rozsah zbierky navýšil o desiatky termínov na výsledných 1005 terminologických záznamov. Navyše sa v priebehu roka zrevidovali a opravili stovky starších terminologických záznamov – v kategóriách korpusová lingvistika (271 terminologických záznamov), v kategórii liturgické textílie (217 terminologických záznamov), filozofia (287 terminologických záznamov), požiarna ochrana (125 terminologických záznamov), marketing (103 terminologických záznamov), civilná ochrana (74 terminologických záznamov), bezpečnostnoprávna oblasť (62 terminologických záznamov), sociálna ochrana (19 terminologických záznamov), šach (4 terminologické

záznamy). V súvislosti s aktualizáciami a opravami bolo v existujúcich kategóriách vytvorených 29 nových terminologických záznamov.

K februáru 2022 obsahovala Slovenská terminologická databáza vo svojej verejne prístupnej časti 16 324 terminologických záznamov.

Úloha č. 12 Vytvorenie a sprístupnenie novej verzie Historického korpusu slovenčiny – *hist-6.0* Plnená priebežne, sprístupnenie: marec 2022

Nová verzia **Historického korpusu slovenčiny** sa od predchádzajúcej líši kvantitatívne aj kvalitatívne. V dátach do tejto verzie *hist-6.0* boli uskutočnené úpravy na troch úrovniach: najskôr išlo o revíziu doterajších textov, aby všetky spĺňali podmienku transliterácie (nie transkripcie) s tým, že zásahy do textu sa oproti originálu mohli týkať len používania veľkých písmen a interpunkcie. Ďalšou úpravou bolo zjednotenie a zjednodušenie označenia cudzojazyčných textov a skratiek. Na úrovni vyhľadávania bola doplnená možnosť výberu rozlišovať alebo nerozlišovať ypsilon od ě. Do aktuálnej verzie pribudlo 5 manuálne prepisovaných pamiatok z predpisovného obdobia – 3 administratívno-správne a 2 z oblasti liečiteľstva. Navyše sa do tejto verzie historického korpusu podarilo získať texty z dvoch publikácií APVV projektu č. APVV-16-0374: *Slovaciká z bývalého Uhorského kráľovstva na príklade Horného Uhorska* (1500 – 1780).

Z personálnych dôvodov došlo k oneskoreniu skompilovania a zverejnenia korpusu, a preto zatiaľ nie je možné poskytnúť jeho štatistické ukazovatele a porovnať ho s predchádzajúcou verziou.

Úloha č. 13 Sprístupnenie novej, rozšírenej verzie textového Korpusu nárečí – *dialekt-5.0* Plnená priebežne, sprístupnenie: marec 2022

V roku 2021 sa pripravovali dáta do piatej verzie nárečového korpusu. Z naskenovaných a rekonštruovaných zdrojov sa do novej verzie spracovalo 51 zdrojových textov rozličného rozsahu, s dátami z predchádzajúcej verzie korpusu tak nová verzia obsahuje cez 100 zdrojových textov. Zväčša ide o ukážky nárečových textov z popularizačných textov alebo vlastivedných monografií, ale aj nárečové texty viacerých dialektológov, napr. J. Dudásovej, G. Horáka, E. Jónu, V. Kováčovej.

V rámci zhromažďovania dát sa získali texty Štefana Tóbika z publikácie *Gemerské nárečia. 2. časť* a *Gemerská nárečová čítanka* a texty z publikácie J. Dudásovej s názvom *Goralské nárečia z pohľadu súčasnej slovenskej jazykovedy*, ktoré však pre komplikovanosť spracovania vo fáze konverzie textu s ohľadom na špecifické znaky používané v texte neboli do súboru pripravených dát do piatej verzie korpusu zaradené.

Z personálnych dôvodov došlo k oneskoreniu skompilovania a zverejnenia korpusu, a preto zatiaľ nie je možné poskytnúť jeho štatistické ukazovatele a porovnať ho s predchádzajúcou verziou.

Úloha č. 14a Anotácia pomenúvajúcich entít a viacslovných spojení Plnená priebežne

V súlade s koncepciou vypracovanou v roku 2018 pokračovali práce na ručnej anotácii pomenovaných entít textov z korpusu slovenskej Wikipédie. Rovnako ako minulý rok sa na anotáciu vyberali texty primeraného rozsahu s cieľom dosiahnuť rovnomerné zastúpenie tém z hľadiska obsahu. Do súčasného stavu bolo zanotovaných 299 textov, v ktorých bolo označených

viac ako 232 tisíc pomenovaných entít. Na podmnožine súborov textov s anotáciami bol natrénovaný model na automatické rozpoznávanie pomenovaných entít s použitím nástroja NameTag 1. Model má ku dňu dokumentácie úspešnosť 84.32 % priemernej miery mikro-F1 pri určovaní jednoduchých kategórií pomenovaných entít a 78.2 % pri určovaní detailných kategórií.

V ďalšom projektovom období je v pláne nadstavba na vytvorených dátach a rozšírenie anotácie o linky na pomenované entity a koreferencie.

Úloha č. 15a Pracovné semináre o používaní korpusových zdrojov podľa požiadaviek záujemcov

Plnená priebežne

Na 7 praktických seminároch pre začínajúcich a pokročilých používateľov SNK (uskutočnených prezenčne i online vzhľadom na pandemickú situáciu a obmedzenia) sa zúčastnilo celkovo 32 záujemcov z vysokoškolských pracovísk (Katedra slovenského jazyka Filozofickej fakulty Univerzity Komenského v Bratislave a Katedra romanistiky Filozofickej fakulty Univerzity Mateja Bela v Banskej Bystrici) ako aj z radov prekladateľského združenia DoSlov. Oddelenie zároveň zorganizovalo 3 interné semináre používania korpusových databáz pre doktorandov a záujemcov z JÚLŠ SAV. Ďalší záujemcovia využili mailové a telefonické konzultácie, príp. mohli využiť inštruktážne návody na používanie korpusov na facebookovej a instagramovej stránke SNK, a takisto na YouTube kanáli SNK.

Úloha č. 15b Organizácia 11. ročníka medzinárodnej konferencie Slovko a publikovanie príspevkov

Splnená, konanie konferencie: 13. – 15. október 2021, vydanie zborníka: október 2021

Príprava jedenásteho ročníka bienálnej konferencie SLOVKO sa začala na jeseň 2020 zostavením vedeckého a organizačného výboru, aktualizáciou témy a rozoslaním pozvánok. Na začiatku roku 2021 sa uzatvárali prihlášky záujemcov o účasť na konferencii. V ďalšom kroku sa registrovali prihlásené príspevky v angličtine a zabezpečovalo sa ich recenzné posúdenie, ktoré skomplikoval veľký záujem o publikovanie. Vyše 80 prihlásených príspevkov bolo potrebné podľa pravidiel konferencie prideliť dvom členom vedeckého výboru na anonymné posúdenie, prípadne tretiemu členovi, ak sa prví dvaja pridelení výrazne nezhodli. Do výsledného súboru príspevkov sa na základe tejto procedúry dostala len takmer tretina – 34 príspevkov. V priebehu leta 2021 nasledovala formálna úprava a zjednocovanie jednotlivých rukopisov do finálnej podoby publikácie, ktorá vyšla začiatkom októbra ako samostatné tematické číslo Jazykovedného časopisu (dostupný online tu: <https://korpus.sk/slovko.html>) a ktorá je indexovaná v Scopuse a ďalších databázach, čo zvyšuje potenciálny impakt publikovaných štúdií (2 z nich autorsky pripravili členovia riešiteľského tímu).

Vzhľadom na protipandemické opatrenia sa napokon samotná konferencia konala v hybridnej forme: online i prezenčne. Celkovo vyše 60 účastníkov z Česka, Ruska, Slovinska, Rakúska, Nemecka, Švédska, Ukrajiny, Chorvátska a Slovenska (vyše 40 prezenčne) tak mohlo prezentovať svoje výskumné výsledky, projekty, prístupy a tiež vymeniť si skúsenosti pri budovaní a využívaní jazykových zdrojov a digitálnych infraštruktúr. Hybridná forma podujatia predstavovala pre jeho organizátorov nielen zvýšenú finančnú ale aj organizačnú náročnosť.

Úloha č. 16 Dokončenie a vydanie monografie *Lexikálno-gramatický opis adjektív v písanej a hovorenej slovenčine na báze korpusového materiálu*

Plnená priebežne

Čiastočne opatrenia súvisiace s pandemiou COVID-19, ale predovšetkým personálne zmeny v oddelení v posledných dvoch rokoch neumožnili rozpracovanie tejto úlohy v tej podobe, ako bola pôvodne zamýšľaná. V roku 2021 bol uskutočnený čiastkový výskum adjektívnych sufixov v odborných textoch zaradených do celku korpusov SNK, z ktorého vzišli 2 nadväzujúce štúdie v anglickom jazyku: 1) *Usage and Empirical Productivity of International Adjectival Suffixes in Slovak Based on General and Specialised Corpora* publikovaná v špeciálnom čísle Jazykovedného časopisu a 2) *Usage and empirical productivity of international adjectival suffixes in Slovak revisited* v litovskom terminologickom časopise Terminologija.

Okrem týchto dvoch štúdií bolo v rámci výskumného zámeru v rokoch 2018 – 2019 vypracovaných a publikovaných 6 ďalších príspevkov:

Levická, J.: Konkurenčnosť synonymných adjektív v lekárskej terminológii. 2018

Gajdošová, K.: Vzťahy medzi vybranými augmentatívnymi adjektívami. 2019

Gajdošová, K.: Sufixálne augmentatívne adjektíva v súčasnej slovenčine. 2019

Krolčíková, R.: Fungovanie adjektív vo vybraných dielach Dušana Mitanu (Analýza na báze Slovenského národného korpusu). 2019

Šimková, M. Deminutívne adjektíva v súčasnej slovenčine (na báze Slovenského národného korpusu). 2019

Šimková, M. Minulé a súčasné kontexty prídavného mena *politický*. (2019)

Uvedených 8 štúdií však nevykazuje ani dostatočný tematický záber, ani náležitú súvzťažnosť, ktorá by umožňovala koncipovanie samostatnej publikácie.

Priebežné úlohy vyplývajúce zo Zmluvy

Väčšina vyššie uvedených výsledkov bola založená na priebežnom plnení úloh, ktoré zakotvuje aj Harmonogram a z ktorých sa v roku 2021 realizovali predovšetkým tieto:

Úloha č. 1: Sťahovanie a spracúvanie textov povinných výtlačkov

Sťahovanie a spracúvanie textov povinných výtlačkov a zo zoznamu www adries, z ktorých sa pravidelne sťahujú diela na základe licenčnej zmluvy podľa interného harmonogramu SNK (napr. noveslovo.sk, pcrevue.sk, tyzden.sk, operaslovakia.sk, babetko.rodinka.sk, birdz.sk).

Úlohy č. 1, 2, 10, 11 – 13: Dopĺňanie a) všeobecného korpusu, b) paralelných korpusov a c) ďalších korpusov SNK aktuálnymi textami

Naskenovalo sa takmer 12 000 strán textov z knižných a časopiseckých titulov, rekonštruovalo, opravilo a skontrolovalo sa vyše 10 450 strán textov. Do archívu SNK bolo vložených 413 nových dokumentov s príslušnou anotáciou.

Úlohy č. 1, 2, 10, 11 – 13: a) Oslovovanie poskytovateľov textov, b) spracúvanie licenčných zmlúv, c) správa databázy poskytnutých textov

Pokračovalo sa v revízii a aktualizácii údajov v databáze poskytovateľov, sumarizovali sa získané a chýbajúce texty, uzatvorilo sa 9 nových zmlúv (napríklad s Filozofickým ústavom SAV).

Osobitne sa vyhľadávali a sťahovali texty z internetu s otvorenými licenciami – časopisy Slovenské divadlo, Ekologické štúdie, Životné prostredie, Studia Politica Slovaca, Geografický časopis, Vojenské reflexie a niektoré publikácie UMB, Pedagogika.sk.

Úlohy č. 1, 2, 10, 11 – 13: Konverzie textov do jednotného formátu

V roku 2021 boli v rámci automatických i jednotlivých konverzií z archívu SNK skonvertované tisíce dokumentov vrátane 1 604 891 správ získaných na základe licenčnej zmluvy z Tlačovej agentúry Slovenskej republiky.

Úlohy č. 1, 2, 10, 11 – 13: Bibliografická a štýlovo-žánrová anotácia textov

Do banky SNK pribudlo 243 nových bibliografických a štýlovo-žánrových anotácií.

Úlohy č. 8, 10: a) Zhromažďovanie a b) spracúvanie odborných textov pre databázu termínov vybraných vedných odborov

Snaha o navýšenie odborných textov v korpusových databázach sa v roku 2021 premietla do spracovania tisícok nových dokumentov predstavujúcich odborné texty v celkovom rozsahu vyše 50 miliónov tokenov.

Úloha č. 11: a) Zhromažďovanie a b) spracúvanie zvukových záznamov do hovoreného korpusu

V rámci priebežných prác bolo v roku 2021 získaných 360 nahrávok, ktoré boli uložené do archívu Slovenského hovoreného korpusu s potenciálom využitia v ďalšej verzii tohto korpusu.

Úloha č. 12: a) Získavanie a b) spracúvanie pramenných materiálov do historického korpusu

Do novej verzie Historického korpusu slovenčiny, ktorá bola naplánovaná na rok 2021, sa na základe konzultácií s pracovníkmi Oddelenia dejín slovenčiny, onomastiky a etymológie JÚLŠ SAV vybralo a spracovalo niekoľko textov z archívu tohto oddelenia (*Liptovský herbár*, *Turčiansky receptár* a niekoľko administratívno-právnych pamiatok). Prepis pôvodných textov realizovali traja anotátori.

osobitné: Správa a aktualizácia prezentačnej stránky SNK na Facebooku

Prezentácia výstupov a práce oddelenia SNK sa v roku 2021, rovnako ako v predchádzajúcom roku z dôvodu protiepidemických opatrení realizovala predovšetkým v online priestore, a to najmä prostredníctvom korpusovej stránky na sociálnej sieti Facebook a YouTube kanáli. Zároveň sa uskutočnila jedna online prednáška o využití korpusov v preklade pre študentov filologických odborov na Filozofickej fakulte UK. Okrem toho boli v rádiu Devín odvysielané dve eseje M. Zumríka na jazykovedné témy.

Na korpusovej FB stránke bolo celkovo publikovaných 73 príspevkov, pričom 2 z nich si pozrelo takmer 5 000 záujemcov. Korpusová FB stránka má aktuálne 1 666 sledovateľov z takmer 40 krajín sveta.

Registrovaných aktívnych používateľov korpusových zdrojov SNK bolo v roku 2021 celkovo 673.

osobitné Správa a aktualizácia a) počítačovej siete, b) pracovných staníc, c) úložísk, d) záložných zdrojov

Realizovala sa priebežná údržba, oprava, montáž, výmena a aktualizácia počítačov

(fyzických a virtuálnych) a iných hardvérových i softvérových nástrojov. V rámci prevencie výpadkov internetovej komunikácie bol zavedený sekundárny DNS server prevádzkovaný Centrom spoločných činností SAV. Prístup k prehľadávaniu korpusov bol rozšírený z pôvodne šiestich na jedenásť serverov a zefektívnili sme využitie hardvérových zdrojov na zrýchlenie prístupu.

osobitné Terminologická poradenská služba

Realizuje sa v terminologických komisiách, kde má SNK JÚLEŠ SAV zastúpenie, ako aj formou mailov, odpovedí na otázky prichádzajúce do jazykovej poradne JÚLEŠ SAV či priebežnými osobnými konzultáciami.

osobitné Korpusovolingvistická poradenská služba

Bolo poskytnutých 22 osobných a telefonických konzultácií, ďalšie otázky boli zodpovedané pri pridelovaní nových alebo obnovovaných registráciách na používanie SNK.

Práce nad rámec harmonogramu 2021

Popri úlohách vyplývajúcich zo Zmluvy sa členovia oddelenia zapájajú aj do grantových projektov: v roku 2021 to boli dva projekty APVV – **Digitálna zbierka slovenskej prózy (APVV-20-0414)** a **Jazykové chyby v slovenčine ako cudzom jazyku na báze akvizičného korpusu, (APVV-19-0155)**. V rámci druhého z nich vznikla pilotná verzia akvizičného korpusu s rozsahom 137 704 tokenov, ktorá je zatiaľ prístupná len členkám riešiteľského kolektívu.

Oddelenie Slovenského národného korpusu participuje prostredníctvom svojho zástupcu na medzinárodnej iniciatíve ELRC (European Language Resource Coordination). V roku 2021 oddelenie SNK podľa požiadaviek koordinátora ELRC usporiadalo **3. slovenský workshop ELRC**. Cieľom online workshopu, ktorý sa konal v slovenčine i angličtine, bolo poukázať na súčasný stav a budúce výzvy v oblasti rozvoja umelej inteligencie a počítačového spracovania prirodzeného jazyka na Slovensku a zvýšenie povedomia o možnostiach služby automatického prekladu *eTranslation*. Workshopu sa zúčastnili zástupcovia konzorcia ELRC, Európskej komisie, slovenských verejných inštitúcií (ministerstiev), odborníci v danej oblasti a prekladatelia.

Úlohy stanovené v zmluve na rok 2021 boli splnené, resp. budú splnené v krátkodobom horizonte s výnimkou 1 publikácie, ktorá nebola dokončená najmä z personálnych dôvodov.

Vypracovali: Jana Levická
 Ivor Uhliarik
 Katarína Gajdošová
 Katarína Rausová