

---

## JEZIKOVNI PRIROČNIKI IN NOVI MEDIJI

---

Članek opisuje slovenske jezikovne priročnike in druge vire podatkov o slovenskem jeziku, ki se nahajajo v neknjižnem digitalnem okolju. Začenja s slovarji kot klasičnimi priročniki v različnih računalniških oblikah, preko leksikonov besednih oblik in jezikovnih korpusov kot novih oblik jezikovnih priročnikov do popolnoma računalniških jezikovnih orodij, ki jih najdemo v urejevalnikih besedil, avtomatskih prevajalnih sistemov in govornih tehnologij. Zaključuje s kratko oceno stanja in predvidevanji glede nadaljnjega razvoja.

### 1 Uvod

V naslovu razprave nastopata dve kategoriji, ki ju je nujno natančneje opredeliti, vendar izključno za namene te razprave brez zahtev po terminološki občeveljavnosti. Z izrazom jezikovni priročnik bomo opredelili vse priročniške vire, ki so namenjeni javnosti kot vir podatkov o sodobnem slovenskem jeziku ali pri katerih nastopa slovenščina v paru s tujim jezikom. Nove medije pa bi najlažje opredelili ne s pozitivnim, temveč z negativnim izborom kot vse medije, ki niso tradicionalni knjižni medij jezikovnih priročnikov. Jezikovni priročniki oz. informacije o slovenščini v njih nastopajo v digitalizirani obliki, bodisi na računalniku ali v drugih digitalnih okoljih. Deliti jih je mogoče po različnih kriterijih, eden od njih je denimo razlikovanje med jezikovnimi priročniki, ki jih v taki ali drugačni obliki že poznamo v knjižni obliki in se z novo funkcionalnostjo pojavljajo v digitalnem okolju, do popolnoma programskih virov, ki izven digitalnega okolja ne obstajajo. Drugi možni kriterij je digitalno okolje, v katerem se pojavlja določen priročnik, bodisi da gre za računalniško okolje zaprtega tipa – aplikacija za osebni računalnik, strežnik, dlančnik, specializirana naprava itd. ali komunikacijski medij – internet, mobilna telefonija. Naša predstavitev poteka po tipih priročnikov, ki v grobem sledi prehodu od knjižnih priročnikov v digitalnem okolju do prej neznanih jezikovnih virov ali priročnikov, s posameznimi skoki v druga področja, kjer se je to zdelo smiselno.

## 2 Slovarji

Najbolj klasični jezikovni priročniki v elektronski obliki so obstoječi knjižni slovarji, bodisi enojezični ali dvojezični, neposredno preneseni v računalniško okolje. Preskok v digitalno okolje pa je obenem tudi kvalitativni preskok. Uporabnost slovarja v elektronski obliki je ob ustrezni programski opremi, ki omogoča kompleksna iskanja po celotni slovarski bazi podatkov, vključno z vsebino gesel ali njegovimi posamičnimi deli, neprimerno večja kot v knjižni obliki. Slovarji se z digitalizacijo vsaj deloma znebijo tudi ene od osnovnih zadreg, ki so jih pri jezikih z abecednim zapisom spremljale že od samih začetkov – tiranije abecedne razvrstitve slovarskih iztočnic, ki je le najboljša med slabimi rešitvami in povzroča težave pri iskanju vseh možnih sestavin jezika, ki niso neposredno in brez težav najdljive preko osnovne oblike – večbesednih leksemov, nepravilnih morfoloških oblik, iztočnic z neabecednim zapisom itd. (Landau 2001: 107–109). Slovarji se v digitalni obliki pojavljajo v večini računalniških okolij, kot aplikacije v vseh vrstah računalnikov, na internetu, v obliki SMS sporočil v mobilni telefoniji ali v specializiranih napravah, kot je denimo kombinacija optičnega čitalca in slovarja – Quicktionary™.

### 2.1 Enojezični slovarji

Splošnih enojezičnih slovarjev v knjižni obliki, ki opisujejo sodobni slovenski jezik, ni veliko. S to oznako bi lahko opredelili tri slovarje, med katerimi sta dva izšla v zadnjem času: Slovar slovenskega knjižnega jezika, ki je izšel v petih zvezkih od 1. 1970 do 1991 (SSKJ),<sup>1</sup> Slovenski pravopis iz leta 2001 (SP 2001) in Veliki slovar tujk iz leta 2002 (VST 2002). Tem bi pogojno lahko dodali tudi Besedišče slovenskega jezika z oblikoslovnimi podatki (BSJ), tj. spisek besed, ki niso bile vključene v SSKJ, ter Odzadnji slovar slovenskega jezika po Slovarju slovenskega knjižnega jezika (OSSJ), spisek iztočnic iz SSKJ, razvrščenih po abecedi od zadnje črke do prve. V času pisanja te razprave od prvih treh v elektronski obliki obstaja le Slovar slovenskega knjižnega jezika (skupaj z BSJ in OSSJ), ki se je v elektronski obliki prvič pojavil leta 1997 na enajstih disketah v programu ASP za operacijski sistem DOS in Windows podjetja Amebis. Trenutno je SSKJ dostopen v paketu s programom ASP32, ki deluje v obliki enouporabniške ali mrežne aplikacije v različicah operacijskega sistema Windows za osebne računalnike in strežnike. Program ASP32 ne deluje v operacijskih sistemih Unix, Linux, MacOS in drugih.

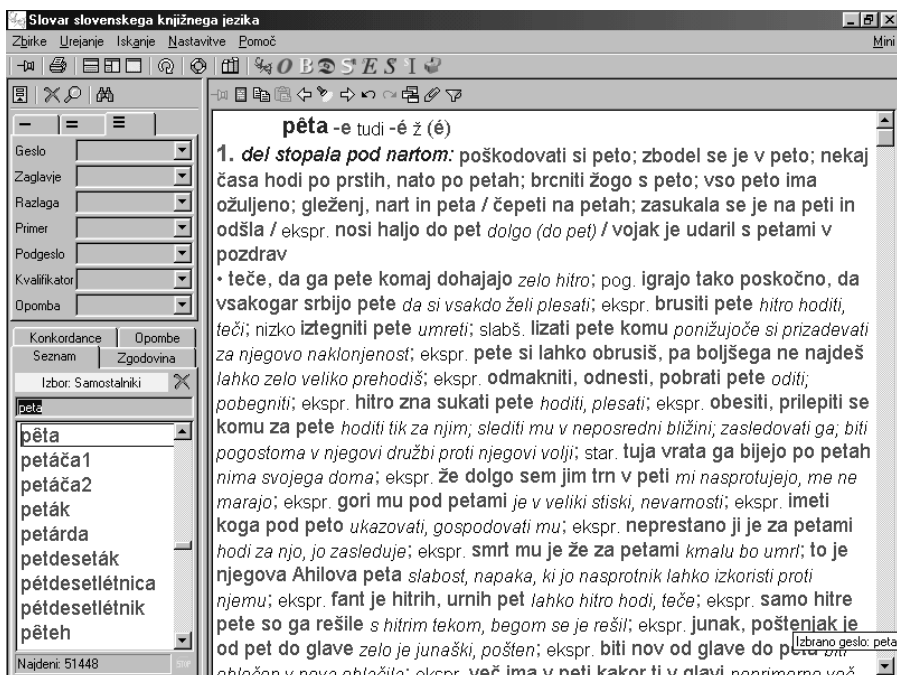
Druga elektronska oblika istega slovarja je javnosti dostopna v programu NEVA, razvitem v Laboratoriju za korpus slovenskega jezika Znanstvenoraziskovalnega centra SAZU, ki od leta 2000 omogoča dostop do SSKJ preko svetovnega spleta na spletnih straneh ZRC SAZU <<http://bos.zrc-sazu.si/sskj.html>>. Dostop do slovarja SSKJ preko svetovnega spleta je brezplačen.<sup>2</sup>

---

<sup>1</sup> Bibliografski podatki vseh jezikovnih virov ter spletnih strani, ki so omenjeni v razpravi, se nahajajo na koncu med navedeno literaturo.

<sup>2</sup> Jeseni 2003 lahko pričakujemo tudi Slovenski pravopis (SP 2001) v elektronski obliki, in sicer v programu ASP32, kasneje po analogiji s SSKJ najbrž tudi v prosto dostopni spletni varianti.

Program ASP32 je »klasični« slovarski vmesnik s funkcijami, ki so postale standard pri podobnih programih: različne možnosti tiskanja gesel in geslovnika, možnost dodajanja opomb, prilagajanja velikosti ali barv črk na ekranu, »sito«, ki s skrivanjem posameznih delov gesla omogoča boljši pregled vsebine pri daljših geslih, odpiranje več oken z več gesli itd. Predvsem pa je močno orodje za kompleksna iskanja – z nadomestnimi znaki za en (?) ali niz znakov (\*) ter operatorji IN, ALI, NE, BLIZU ter t. i. NOTRANJI NE in NOTRANJI IN, ki delujeta znotraj posameznega niza iskanih znakov. Celotno slovarsko bazo lahko preiskujemo po posameznih delih gesel: geslo, zaglavje, razlaga, primer, podgeslo in kvalifikator, iskanja po posameznih segmentih lahko tudi poljubno kombiniramo. Iščemo lahko po vseh »posebnih« znakih, ki se nahajajo v SSKJ, od naglasnih znamenj za izgovorjavo besede do grških črk in matematičnih znakov. Ena od uporabnih lastnosti programa ASP32 je tudi možnost shranjevanja in iskanja po izboru gesel, kar nam olajša pogosta iskanja po geslih, ki smo jih iz celote že izločili s kompleksnejšim iskalnim pogojem.



Možnosti pri spletnem vmesniku so v primerjavi z ASP32 skromnejše, vendar tudi ta omogoča iskanje po celotni vsebini SSKJ, kar je dovolj za osnovno uporabo. Od nadomestnih znakov pozna znak za niz (\*), operatorji pa so naslednji:

: vsebuje	!= ni enako	%< dolžina polja manjša kot
!: ne vsebuje	< manj kot	%= dolžina polja enaka
= je enako	> več kot	%> dolžina polja več kot



dostop do množice slovarjev zelo različnega izvora, navadno pa so to zbirke, ki so rezultat akademskih projektov, posameznih navdušencev, starejše verzije sodobnih slovarjev in podobno. Sčasoma so se poleg splošnih brskalnikov na svetovnem spletu pojavili tudi specializirani iskalniki, ki iskano besedo ali zvezo iščejo po vseh brezplačno dostopnih slovarjih, npr. <<http://www.dictionary.com>>, <<http://www.yourdictionary.com>>, <<http://www.onelook.com>> itd. V večini primerov je tem slovarjem skupna bodisi nedokončanost, zastarelost, relativno majhen obseg ali fragmentarnost, vendar imajo seveda tudi takšni viri svojo vrednost. Angleščina je pri tem edina izjema, saj so nekatere založbe najbrž zaradi hude konkurence začele nuditi brezplačen dostop tudi do svojih najnovejših in najboljših slovarjev.<sup>3</sup> Večina tujih založnikov pa svoje slovarske paradne konje ponuja v obliki verificiranega dostopa za denarno nadomestilo, pogosto v integrirani obliki z vsemi viri, ki jih imajo na voljo, tipičen primer je denimo stran založbe Oxford University Press <<http://www.oxfordreference.com>>.

Odlična knjižnica povezav do spletnih virov za slovenski jezik (in za tuje jezike) se nahaja na strani <<http://www.sigov.si/slovar.htm>>. Večji splošni dvojezični slovarji s slovenščino kot izhodiščnim ali ciljnim jezikom v plačljivi obliki za zdaj še niso dostopni na svetovnem spletu. Od slovarjev, ki obstajajo tudi v knjižni obliki in na CD-ROM-u, velja morda izpostaviti osemjezični Evropski slovar, ki za vsak jezik po navedbah založbe ponuja 10.000 iztočnic<sup>4</sup> in je na voljo brezplačno na spletnem portalu <<http://www.pinkponk.com>>, preko povezave Vedež.

Zanimivo je, da prav dvojezični slovarji navadno prednjačijo pri prehodu v druge tehnologije, saj jih najdemo v okoljih, ki presegajo aplikacije za osebne računalnike ali internet. Do dvojezičnih podatkov iz osemjezičnega Evropskega slovarja je poleg svetovnega spleta mogoče priti tudi preko digitalne telefonije, s pomočjo sporočil SMS. Na navedeni spletni strani je opisan postopek pošiljanja SMS-jev s slovarsko vsebino. Manjši obojesmerni angleški, italijanski, francoski in hrvaški dvojezični slovar z nekaj nad šest tisoč gesli (po navedbah proizvajalca) je mogoče dobiti tudi za dlančnike z operacijskim sistemom PalmOS <<http://slovenian.palm-dictionaries.com>>.

Povsem drugačen dostop do slovarjev pa ponuja zanimiva kombinacija optičnega čitalca, programa za razpoznavo črk in slovarja v pripomočku z imenom Quicktionary <<http://www.wizcomtech.com>>, <<http://www.iskraprins.si>>. Slovenščina je v njem zastopana v dvojezičnem angleško-slovenskem paru. S konico – optičnim čitalcem – te debelejšemu svinčniku podobne naprave z zaslonom iz tekočih kristalov lahko potegnemo po besedilu, natisnjenem na papir, in na zaslonu preberemo prevod besede ali besedne zveze, če je čitalec pravilno prepoznal črke in če beseda ali zveza obstaja v slovarju. Optično branje in prepoznavo črk delujeta zanesljivo,

<sup>3</sup> To sicer ni tema razprave, a naj omenimo le nekaj slovarjev: Longman Dictionary of Contemporary English <<http://www.longmanwebdict.com>>, Cambridge Advanced Learner's Dictionary <<http://dictionary.cambridge.org>>, Merriam-Webster's Collegiate Dictionary <<http://www.m-w.com/netdict.htm>> itd.

<sup>4</sup> To število je najbrž potrebno vzeti z malce rezerve, kajti sam iskalnik ASP32 izkazuje, da je npr. število slovenskih iztočnic 7.828, vsaka dvo- ali večbesedna iztočnica tipa »dobro jutro« ali »kurilno olje« pa je obravnavana kot samostojno geslo.

čeprav na začetku morda potrebujemo nekaj časa, da se privadimo na pravilno lego pri potegu naprave preko besedila. Sam slovar je očitno nastal na podlagi ene od elektronskih verzij Velikega angleško-slovenskega slovarja (Grad 1978), čeprav predelane, najbrž manj vsebinsko kot predvsem z računalniškimi postopki. V njem so tako ostali mnogi Gradovi kuriozumi,<sup>5</sup> število samih gesel pa je najbrž precej manjše, kot ga navajajo proizvajalci (300.000). Listanje po osnovnem geslovníku pokaže grobo oceno okrog 40.000 gesel brez podgesel in frazeoloških enot, kar je približno polovica Velikega angleško-slovenskega slovarja, vendar pri takšni napravi kar spoštljiva številka.

### 2.3 Tezavri ali slovarji sinonimov

Kot je znano, za slovenski jezik slovar sinonimov ali tezaver v knjižni obliki ne obstaja. Na tej točki je elektronska različica torej prehitela bolj logično predhodno knjižno obliko, saj podjetje Microsoft v svojem programskem paketu Microsoft Office od lokalizirane slovenske različice MS Office 97 (zadnja je slovenski MS Office XP), ponuja tudi orodje, ki uporabniku predlaga sopomenke. Slovar, ki ga orodje uporablja, je razmeroma nedodelan, saj je za besedo *fant* denimo zmožen sproducirati sopomenski niz *mladenič, deček, fantič, pob, moški, mož, mladinec, paglavec*, ki seveda niso sopomenke v strogem pomenu, temveč označujejo osebe moškega spola različnih starosti z različnimi pomeni na različnih stilnih ravlinah. Da bi zanesljivo opravljal obljubljeni funkciji, bi bila potrebna (predvsem jezikoslovna) dodelava, saj je za uspešno uporabo potrebno dodobra poznati slovenski jezik, da se pri izbiri ustrezne sopomenke ne opečemo. Kljub vsemu pa je ob omenjenih omejitvah tudi tak vir seveda dobrodošel.

### 2.4 Terminološki enojezični ali dvojezični slovarji

Kot smo že omenili, so terminološki slovarji takorekoč idealni priročnik za medij, kakršen je internet. Če iščemo prevod ali razlago nekega termina, nas pravzaprav ne zanima, ali je v slovarju deset ali deset tisoč iztočnic. Vsaka ad hoc sestavljena lista terminov s prevodi ali razlagami, ki jih je sestavil bodisi navdušeni amater ali akademsko znanstveno telo, ki lahko na svetovnem spletu sproti dopolnjuje terminologijo nekega področja brez dragega tiskanja, nam bo dala (bolj ali manj zanesljivo) informacijo.

Na svetovnem spletu kar mrgoli tovrstnih virov, ki jih lahko najdemo s pomočjo splošnih iskalnih orodij, kakršni so Google, Altavista, Najdi.si itd. ali specializiranih za slovarje, ki so omenjeni v poglavju o dvojezičnih slovarjih. V spletni slovarski zbirki povezav <<http://www.sigov.si/slovar.htm>> je mogoče najti terminološke

<sup>5</sup> Eden od slavnejših je denimo *paramedic* – »zdravnik padalec«. Pri slovenskem prevodu gre za izjemno redek pomen angleške iztočnice, ki označuje zdravnike v padalskih enotah ameriške vojske ali civilne zdravnike, ki se v primeru potrebe s padalom spustijo na nedostopna območja. Manjka pa izjemno pogosti in praktično edini smiselni prevod – »bolničar/-ka«.

slovarje z zelo različnih področij, od zootehnike do besedja na temo slapov, bogato knjižnico povezav najdemo tudi na strani <[http://17slon.com/kbg/prevajanje/slovarji\\_in\\_glosarji.htm](http://17slon.com/kbg/prevajanje/slovarji_in_glosarji.htm)>, precej manjšo tudi na portalu Matkurja <<http://www.matkurja.com/slo/resources/entertainment/dictionaries>>.

Med vsemi terminološkimi slovarji gotovo izstopa Evroterm <<http://www.sigov.si/evroterm>>, terminološka baza s spletnim iskalnikom Sektorja za prevajanje Službe Vlade RS za evropske zadeve, ki vsebuje več kot 30.000 dvojezičnih (slovensko-angleških) vpisov in je nastala pri prevajanju evropske zakonodaje v slovenski jezik (Željko 2002). Geslovník je štirijezični, zato lahko iščemo angleške, francoske, nemške in slovenske iztočnice. Iskalnik omogoča tudi iskanje z nadomestnimi znaki, vendar je pri kompleksnejših iskanjih bolje prebrati navodila, kajti namesto (morda običajnejšega) vprašaja kot znaka za poljubno črko uporabljamo piko, vprašaj pa pomeni, da se predhodna črka v besedi lahko pojavi ali ne, znak \$ na koncu besede pomeni, da se zveza konča na to besedo, v običajnejših vlogah pa lahko uporabimo oglate in zavite oklepaje in druge funkcije t. i. regularnih izrazov.

### 3 Leksikoni (besednih oblik)

Možnosti hranjenja in hitrega dostopanja do velikih količin podatkov ter potrebe povsem računalniških jezikovnih orodij, kot so črkovalniki, delilniki, besedilni korpusi, avtomatski prevajalniki itd. so s seboj prinesli tudi potrebo in možnost organiziranja t. i. leksikonov besednih oblik pri pregibnih besednih vrstah, ki so nujni zlasti pri jezikih z velikim številom oblikoslovnih paradigem. Te informacije so bile do računalniške dobe v precej skrčeni obliki navadno dostopne v zaglavju enojezičnih (ali dvojezičnih) slovarjev kot – denimo v SSKJ – roditeljska oblika ednine pri samostalniki in pridevnikih ali prvoosebna oblika ednine pri glagolih, z občasnimi opozorili pri izjemah, ter v oblikoslovnih razdelkih slovnice slovenskega jezika. Informacija tako v slovnici (nekaj primerov paradigme) kot v slovarju (pri posamezni besedi le ena ali nekaj oblik) je bila torej dokaj rudimentarna in je za uporabo v resnici predpostavljala poznavanje slovnice oz. oblikoslovnih paradigem. Uporabniško gledano je sklepanje iz osnovne oblike na vse oblike besede za domače govorce slovenščine predvsem pri oblikah, ki so oblikoslovno težavnejše in morda manj v rabi, sorazmerno naporno, za tujce, ki se učijo slovenščino, pa takorekoč neobvladljivo. Računalniški medij je za tovrstni priročnik precej bolj primeren kot knjižna oblika, najbrž celo edini možni, kajti izpisovanje vseh oblik določene besede je prostorsko precej potratno, predvidljivost oblik glede na paradigmo, v katero spada, pa vendarle sorazmerno velika.

Leksikoni besednih oblik so javno dostopni v dveh medijih. Na spletu najdemo orodja, ki imajo funkcijo iskalnika po računalniškem leksikonu besednih oblik, na dveh mestih. Zanimivo je, da je logika iskanja pri obeh ravno obratna. Z orodjem, ki ga najdemo na spletnih straneh ZRC SAZU <<http://bos.zrc-sazu.si>>, povezava »Določanje osnovnih besednih oblik«, lahko ugotovljamo osnovno obliko določene besede. Dolžina seznama, ki ga lahko damo hkrati v preverjanje, ni omejena, število osnovnih oblik v geslovníku, ki ga uporabljamo iskalnik, pa ni znano.

Orodje, ki ga lahko najdemo na spletnih straneh podjetja Amebis <<http://www.amebis.si/sklanjanje>>, omogoča obratno pot – od osnovne oblike iskane besede do razpredelnice z vsemi njenimi oblikami, iščemo pa lahko le samostalnike in pridevnike. Iskanje po posameznih oblikah besede ne da rezultata, iskalnik prepozna le obliko za imenovalnik ednine. Isto podjetje ponuja tudi aplikacijo z imenom Amebisov besedni zaklad <<http://www.amebis.si/izdelki/abz>>, ki deluje v že omenjenem programu ASP32 za okolje Windows, s katero lahko preiskujemo leksikon besednih oblik z geslovníkom, ki obsega nad 250.000 iztočnic, tudi lastnoimenskih. S tem orodjem lahko iščemo po osnovnih in po vseh ostalih oblikah, v kombinaciji s slovnično kategorijo, ki je pripisana osnovni obliki.<sup>6</sup>

Leksikoni besednih oblik so izjemno pomembni za vse mogoče aplikacije na področju jezikovnih tehnologij in obe omenjeni aplikaciji sta pravzaprav stranski produkt drugih orodij (Romih in dr. 2002a), pri katerih nujno potrebujemo informacijo o osnovni obliki besede in njeni oblikoslovni paradigmi – črkovalnika in oblikoslovno-skladenjskega označevalnika, potrebnega pri t. i. lematizaciji velikih zbirk besedil.

#### 4 Besedilni korpusi

Korpusi so obsežne zbirke besedil, po katerih lahko iščemo z računalniškim orodjem, ki se imenuje konkordančnik (Erjavec 96/97, Gorjanc 1999b, Gorjanc 2000). Računalniški korpusi so nekako pred dvajsetimi leti nadomestili zbirke listkovnega gradiva, na podlagi katerih so v predračunalniški dobi navadno nastajali slovarji – med njimi tudi SSKJ – v zadnjih desetih letih pa je napredek tehnologije z množično uporabo interneta in skokovitim porastom računalniških kapacitet omogočil, da je korpus skupaj s konkordančnikom postal standardni vir podatkov o jeziku ali kar jezikovni priročnik, na katerega se lahko obrnemo ob jezikovnih zadregah (Hirci 1998, Pisanski 1998). Med različnimi tipi korpusov vlogo jezikovnega priročnika navadno najbolje opravlja t. i. referenčni korpus – čimbolj obsežna zbirka sodobnih besedil, ki odraža stanje v določenem jeziku glede na različne jezikovne zvrsti. Z iskanjem po referenčnem korpusu je mogoče spremljati jezikovno delovanje določene jezikovne skupnosti in na podlagi tega sprejemati lastne jezikovne odločitve. Če ima jezikovna skupnost jezik dobro opisan v drugih jezikovnih priročnikih, jim lahko služi kot dopolnilo. V slovenskem primeru, kjer slovarji temeljijo na že zastarelem gradivu, pa jih lahko do neke mere celo nadomesti.

Korpus je pravzaprav vsaka zbirka besedil, ki je zbrana za določen namen – za opis sodobnega jezika bomo potrebovali denimo veliko zbirko sodobnih besedil, za vzporedni korpus bomo potrebovali besedila, ki obstajajo prevedena v dveh (ali več) jezikih, za korpus strokovnih besedil pa nabor besedil z določenega strokovnega področja. Skupaj z besedili za iskanje po korpusu potrebujemo tudi konkordančnik, računalniški program, ki omogoča hitra iskanja po celoti zbranih besedil

<sup>6</sup> Podrobnejšo analizo tega elektronskega pripomočka lahko najdemo v razdelku Ocene in poročila v tej publikaciji.



po poljubnih pogojih. Da bi izboljšali uporabnost korpusov, lahko v besedilo korpusa z različnimi postopki vnesemo tudi dodatne informacije, najbolj običajno so to t. i. lematizacija in oblikoslovno-skladenjsko označevanje (Jakopin in Bizjak 1997, Zupan 1999), kjer vsaki besedi pripišemo njeno osnovno obliko ali skladenjsko vlogo. V nadaljevanju opisujemo korpus slovenskega jezika, ki so javno dostopni na svetovnem spletu.

#### 4.1 Enojezični

Trenutno največji enojezični korpus slovenskega jezika je korpus FIDA <<http://www.fida.net>>, projekt dveh akademskih in dveh industrijskih partnerjev, ki je dostopen preko interneta za plačilo, razen za raziskovalce na Filozofski fakulteti Univerze v Ljubljani ter na Institutu Jožef Stefan (Erjavec 1998, Erjavec in dr. 1998, Gorjanc 1999a, Romih 1998b, Stabej 1998). Brezplačen je testni dostop, pri katerem je število izpisanih konkordančnih vrstic omejeno na deset. Konkordančnik ASP32, s katerim iščemo po korpusu (Gorjanc in Vintar 2000), ima poleg možnosti kompleksnih iskanj, ki so opisana zgoraj v slovarski varianti, dodana tudi orodja, ki omogočajo statistične izračune in različna filtriranja glede na okolico iskalnega pogoja. Korpus sestavljajo sodobna besedila v obsegu 100 milijonov besed, nastala v 90-ih letih prejšnjega stoletja in so uravnotežena po standardnih korpusnojeziških kriterijih, kar pomeni, da so v njem zastopana v določenih odstotkih glede na vrst (umetnostna, neumetnostna, strokovna, humanistična in družboslovna, naravoslovna in tehnična itd.) in prenosnika (knjižna, periodična, časopisna, revialna, dnevna, tedenska itd.). Korpus je v sorazmerno visokem odstotku lematiziran (besedam je pripisana osnovna oblika, po kateri je mogoče iskati), vendar so v primeru, ko je pri eni od oblik besede možnih več osnovnih oblik, pripisane vse možne oblike.

Drugi največji korpus slovenskega jezika je korpus Nova beseda <<http://bos.zrc-sazu.si>>, povezava »Nova beseda«, ki nastaja v Laboratoriju za korpus slovenskega jezika na Inštitutu za slovenski jezik Frana Ramovša ZRC SAZU. Korpus je brezplačno dostopen preko interneta in obsega 95 milijonov besed, od tega nekaj nad 84 milijonov iz časopisa DELO, ostalo pa sestavljajo računalniška revija Monitor, starejše in novejše slovensko leposlovje ter prevodna literatura. Konkordančnik z imenom NEVA, s katerim iščemo po korpusu, omogoča enostavno iskanje po besedah (tudi z zvezdico kot nadomestnim znakom za poljubni niz črk) in frazah. Izpisujemo lahko poljubno število konkordanc, okolice iskalnega pogoja ni mogoče statistično obdelati, filtrirati ali razvrščati. Korpus ni lematiziran, kematiziran je le njegov enomilijonski del, dostopen preko povezave »Oblikoslovno označena besedila«. Konkordančnik omogoča tudi iskanje po listi vseh besed, ki se pojavljajo v korpusu, pri čemer je mogoče iskati po frekvenci besed, dolžini glede na število črk ter z operatorjema IN ter IN NE.

Tretji korpus, dostopen preko interneta, je kronološko prvi, vendar bistveno manjši enomilijonski korpus, brezplačno dostopen na straneh Instituta Jožef Stefan <<http://nl2.ijs.si/index-mono.html>>. Uporabnost korpusa je glede na količino besedil v

primerjavi z gornjima dvema sorazmerno manjša, funkcija in vrednost korpusa pa je pravzaprav v pionirski vlogi, ki je postavila osnove za oblikovanje drugih korpusov, predvsem korpusa FIDA, glede formata, lematizacije in drugih postopkov gradnje korpusov. Konkordančnik Corpus Workbench, razvit v Stuttgartu (Institut für Machinelle Sprachverarbeitung) omogoča preprosto iskanje po besedah (z zvezdico kot nadomestnim znakom za poljubni niz črk) in frazah. Število izpisov je omejeno na 500 konkordančnih vrstic.

Med omenjenimi korpusi sta prva dva uporabna kot korpusna jezikovna priročnika, vsak od njiju pa ima svoje prednosti in slabosti. Prednosti korpusa Nova beseda sta predvsem prosta dostopnost ter sprotno dopolnjevanje, slabosti pa izrazita neuravnoteženost in nelematiziranost. Dokončno sklepanje o jeziku večinoma na podlagi vsebine enega časopisa je jezikoslovno sporno in lahko privede do napačnih jezikovnih odločitev.

Prednosti in slabosti korpusa FIDA so temu ravno nasprotni. Korpus je relativno dobro uravnotežen, kaže torej sorazmerno ustrezno podobo slovenskega pisnega jezika, in je tudi lematiziran. Kot problem pa se kaže nerazdvoumljenost lem pri oblikah, kjer jih je možnih več – pri obliki »levi« je navedena lema »lev« kot žival, »levi« kot pridevnik ter glagol »leviti« – ter seveda plačljivi dostop.

## 4.2 Vzporedni

S konkordančnikom IMS Corpus Workbench lahko iščemo tudi po dveh paralelnih angleško-slovenskih korpusih <<http://nl2.ijs.si/index-bi.html>>. Prvi je korpus ELAN (Erjavec 1999a, Erjavec 1999b, Vintar 1999a), enomilijonska zbirka prevedenih večinoma pravnih in drugih besedil ter roman 1984 Georgea Orwella, zbranih v projektu z istim imenom na Institutu Jožef Stefan <<http://nl.ijs.si/elan>>. Druga je nekaj manj kot enomilijonska zbirka besedil z imenom TRANS s področij medicine, strojništva, zakonodaje/prava, geologije in turizma, zbranih na Oddelku za prevajalstvo Filozofske fakultete Univerze v Ljubljani v okviru študentskega seminarkega dela <<http://www-ai.ijs.si/~spela/trans-index.html>>.

Vzporedni korpusi so oz. bi lahko bili izjemno pomembni viri predvsem za prevajalce, slovaropisce in za celo področje jezikovnih tehnologij, ki se ukvarja z avtomatskimi prevajalnimi sistemi. Zdi pa se, da je še vedno zelo težko priti do vzporednega korpusa splošnega jezika (Čermák 2002: 274), saj je za razliko od enojezičnih korpusov splošnih paralelnih korpusov izjemno malo, veliko bolje pa so zastopani na posameznih strokovnih področjih kot podlaga in vir za terminološke slovarje, kar navsezadnje kaže tudi sestava obeh omenjenih paralelnih korpusov (Vintar 1999b, Vintar 2002).

### 4.3 Korpusi strokovnih besedil

V okviru Sektorja za prevajanje Službe Vlade RS za evropske zadeve je od leta 2002 preko interneta prosto dostopen tudi celoten približno 1,8 milijonski korpus prevodov zakonodaje Evropske unije <<http://www.sigov.si/evrokorpus>>. Korpus uporablja isto iskalno orodje – v tem primeru v funkciji konkordančnika – kot slovar Evroterm, z istimi iskalnimi možnostmi. Vpisani izraz je lahko v angleščini ali slovenščini, izpis pa je lahko enojezičen ali dvojezičen in od te izbire je tudi odvisno ali ga uporabljamo kot enojezični ali vzporedni korpus. Iskanje lahko omejimo tudi na različna specializirana področja zakonodaje: finance, kmetijstvo, okolje, primarno zakonodajo, promet ter zunanjo in varnostno politiko.

### 4.4 Svetovni splet

Svetovni splet (world wide web) je s širitvijo v preteklih desetih letih pravzaprav sam postal neke vrste megakorpus, po katerem lahko iščemo podobno kot po drugih velikih zbirkah besedil (Kilgarriff 2001). Razvit je bil tudi konkordančnik <<http://www.webcorp.org.uk>>, ki uporablja poljubni splošni iskalnik (Google, Altavista, Metacrawler, Alltheweb ali Northern Light) in izpisuje konkordančne vrstice ter opravlja nekaj osnovnih funkcij konkordančnikov, kot so iskanje kolokacij, filtriranje in urejanje po abecedi levo in desno od iskalnega pogoja. Če iskalnik dovoljuje omejitev iskanja po domeni (npr. iskanje po končnici .si), lahko iščemo samo po straneh, ki so napisane v slovenščini. Spletni konkordančnik deluje sorazmerno počasi, tako da je za enostavna iskanja pravzaprav lažje uporabljati kar splošne iskalnike, konkordančnik pa za raziskovanje okolice iskane besede. Pri pridobivanju jezikoslovnih podatkov s svetovnega spleta se znajdemo tudi pred znanimi zadregami: besede niso lematizirane, še vedno se pojavljajo težave s krilatimi znaki, predvsem pa je medij preveč kaotičen za dokončna sklepanja o frekvenci, rabi ali pomenu besede ali besedne zveze. Služi lahko le kot dodatna sekundarna informacija.

## 5 Jezikovna orodja v pisarniških programskih paketih

### 5.1 Črkovalniki

Črkovalnike – programske orodje, ki preverja pravilnost zapisa besed v različnih programih, od urejevalnikov besedil, programov za razpoznavo črk, za namizno založništvo, poštnih programov itd. – smo omenjali že večkrat (Košir in dr. 1998, Romih 1998a). Tudi črkovalniki so eno od orodij, ki so bili v računalniškem okolju takoj aktualni in so se pojavili med prvimi. Prvi slovenski črkovalnik je bil Mspell avtorja Miha Mazzinija z začetka devetdesetih let, razvoj črkovalnikov pa je po tem času šel v dve smeri oz. se je dogajal v dveh okoljih. En (večinski) del slovensko pišočih uporablja črkovalnik, ki je del Microsoftovega pisarniškega paketa slovenske različice urejevalnika besedil Word 7 do MS Office XP in je delo podjetja

Amebis. Različne generacije istega črkovalnika lahko najdemo tudi v drugih programih, denimo v malce starejšem urejevalniku besedil WordStar, WordPerfectu (do različice Corel WordPerfect 8), programu za razpoznavo črk OmniPage in drugih programih, lahko pa ga uporabljamo tudi kot samostojno aplikacijo MikroBesAna za operacijski sistem Windows in DOS <<http://www.amebis.si/izdelki/crkovalnik>>. Program vsebuje skoraj 1,2 milijona besednih oblik.

Drugi del slovensko pišoče publike uporablja prosto dostopne črkovalnike aspell, ispell in myspell, ki so del projekta GNUsl <<http://nl.ijs.si/GNUsl>>. Aplikacija je nastala v okviru skupine za slovenjenje pri društvu Lugos, v sodelovanju s podjetjem Amebis, in uporablja približno enak slovar kot črkovalnik za MS Office (Košir 2002). Za prosto dostopni brezplačni pisarniški program OpenOffice.org, ki obstaja v slovenski lokalizirani različici za operacijska sistema Windows ter Linux <<http://openoffice.lugos.si>>, je slovenski črkovalnik in na voljo na spletni strani <[http://whiteboard.openoffice.org/lingucomponent/download\\_dictionary.html](http://whiteboard.openoffice.org/lingucomponent/download_dictionary.html)>.

## 5.2 Delilniki

Delilnik, orodje za delitev besed po pravilih slovenskega pravopisa, lahko v obliki modula za vgradnjo najdemo na istih spletnih naslovih kot oba črkovalnika, za razliko od njiju pa je slovarska baza delilnikov različna. Delilnik podjetja Amebis je podobno kot črkovalnik že vgrajen v pisarniški paket Microsoft Office od slovenske različice urejevalnika besedil Word 7 naprej.

## 5.3 Preverjanje slovnice

Poleg črkovalnika in delilnika besed, najstarejših standardnih komponent v urejevalnikih besedil, v njih najdemo tudi orodja za preverjanje skladenjske pravilnosti stavkov. Niti Microsoft Office niti drugi urejevalniki takega orodja za slovenski jezik še nimajo. Edino samostojno orodje za preverjanje slovnice, vendar le za operacijski sistem DOS, ponuja podjetje Amebis pod imenom BesAna <<http://www.amebis.si/izdelki/besana>>.

## 5.4 Seznami samopopravkov

Seznam samopopravkov je orodje, ki je na voljo v paketu Microsoft Office kot integrirana aplikacija in ni na voljo denimo kot vgradni modul. Gre za spisek pogostih tipkarskih napak, ki jih delajo uporabniki pri uporabi urejevalnika besedil. Če imamo orodje vključeno, urejevalnik sam nadomesti besedo iz spiska »napačnih« besed s »pravo« besedo. Dodati je potrebno, da se pri spisku napačnih besed skriva tudi kakšna jezikovna past, saj med njimi najdemo tudi pare, kjer gre za besede z različnimi pomeni (*tehnica – tehnica*), na različni stilni ravni (*košarkaš – košarkar*), neustaljeno rabo, morda celo z manj ustrezno izbiro (*nadobudnež – nadebud-*

*než*), ne le za klasične napake, kjer orodje nadomešča lektorja in korektorja (*precednik – predsednik, povdarek – poudarek; skald – sklad, ail – ali*).

Za prosto dostopne urejevalnike besedil je na spletni strani <<http://nl.ijs.si/gnus/download>> na voljo spisek t. i. napačnih besednih oblik, ki sicer skriva še več omenjenih jezikovnih pasti, saj gre precej daleč v jezikovni normativizem. Tako med napačnimi besedami najdemo tudi naslednje: *bojler, brzina, budžet, candra, cedejka, cimra, citadela, cukrarna, displej, duhamoren* itd.

## 6 Avtomatski prevajalni sistemi

Avtomatski prevajalni sistemi so pri slovenskem jeziku na začetku razvoja. To področje, ki je takorekoč sveti gral računalniškega jezikoslovja in je od petdesetih let prejšnjega stoletja doživljalo vzpone in padce, je v zadnjih letih doživelo nesluhten razmah in pri »velikih« jeziki rezultati že zdavnaj ne zbuja več le prizanesljivih nasmeškov. Zavedati se je treba, da avtomatski prevajalniki verjetno nikoli ne bodo v celoti nadomestili prevajalcev, lahko pa spravijo dokument ali npr. spletno stran v obliko, ko uporabnik brez znanja izhodiščnega jezika razbere osnovno sporočilo ali temo.

Za slovenščino kot jezik z majhnim številom govorcev, razmeroma šibko finančno podporo in sorazmerno majhnim številom ljudi, ki se ukvarjajo z jezikovnimi tehnologijami, je sedanje stanje na tem področju dokaj logično. Vendarle pa lahko, kot rečeno, spremljamo prve poskuse v smer programske opreme, ki ponuja avtomatsko prevajanje. Prvi je domač proizvod, drugi prihaja iz tujine. Podjetje Amebis ponuja prevajalnik, ki zna prevajati iz slovenščine v angleščino, na voljo pa je kot programski paket v treh različnih izvedbah <<http://presis.amebis.si>>, preizkusiti ga je mogoče tudi preko svetovnega spleta <<http://presis.amebis.si/prevajanje.asp>>, kjer omogoča prevajanje besedila z največ petsto besedami. Za zgodnjo razvojno fazo prevajalnik sorazmerno dobro opravi svoje delo, prevod omogoča osnovno razumevanje originalnega besedila, do ravni, na kateri se nahajajo najboljše prevajalniki med angleščino, francoščino, nemščino itd. pa mu manjka še kar precej.

Drugi sistem je prevajalnik med celo množico jezikov in obstaja v več različicah za različne operacijske sisteme oz. medije, tudi preko svetovnega spleta <<http://www.tranexp.com>>. Med šestintridesetimi jeziki je tudi slovenščina (in kot zanimivost, hrvaščina, bosanščina in srbščina). Testno prevajanje preko spleta <<http://www.tranexp.com:2000>>, če je na podlagi tega mogoče o čem sklepati, pokaže, da slovenščina v tem prevajalnem sistemu še ni dosegla ravni, ki bi omogočala razumevanje besedila brez znanja izhodiščnega jezika, preizkusiti pa je bilo mogoče le angleško-slovenski in slovensko-angleški par.<sup>7</sup>

<sup>7</sup> Morda le hudomušni utrinek – pri prevajanju osnovne strani portala [www.CNN.com](http://www.CNN.com) je prevajalni sistem rubriko Politics prevedel s »politična mahinacija«.

## 7 Govorne tehnologije

Sinteza in analiza govora je zadnje področje te razprave in v nekem smislu sega že preko njenega osnovnega zastavka, ki ostaja pri priročnikih za slovenski jezik v digitalni obliki. Najbolj tipični produkt na tem področju, ki bi še vključeval osnovno temo, bi bil fond zapisov izgovarjave besed v slovenskem jeziku, ki bi bil na voljo javnosti, tipično v obliki elektronskega slovarja z izgovarjavo besed, vendar takega priročnika za slovenščino ni. Še najbliže temu je morda zbirka izgovarjav pogostih izrazov, števil in podobnih uporabnih zvez, namenjenih turistom na spletni strani <<http://www.ff.uni-lj.si/sft>> v okviru spletnega priročnika Slovenščina za popotnike.

Od orodij za sintezo in analizo govora, ki so na voljo javnosti, so za zdaj vsa na strani sinteze, ki je – kot kaže – lažja polovica dveh sicer dokaj različnih tehnologij (Gros 2000, Gros 2001, Kačič in dr. 1998, Šef in dr. 1998). Za slovenščino obstaja več sistemov za sintezo govora: prvi je bil razvit na Institutu Jožef Stefan, imenuje se Govorec in se nahaja na strani <<http://ai.ijs.si/govorec>>. Primarno je namenjen slepim in slabovidnim za delo z računalnikom, sistem je za njih brezplačno dostopen, mogoče ga je v obliki datoteke shraniti z interneta in namestiti na računalnik z operacijskim sistemom Windows. Podjetje Masterpoint pa ponuja program eBralnik, ki omogoča testno uporabo preko svetovnega spleta na strani <<http://mpx.masterpoint.si/voice>>.

## 8 Prihodnost

Prihodnost slovenskih računalniških jezikovnih priročnikov najbrž lahko vidimo v sedanjem stanju podobnih virov pri drugih jezikih z večjimi finančnimi in človeškimi možnostmi. Povprečni uporabnik računalnika in drugih tehnologij, ki je govorec jezika, kjer poteka najbolj dinamičen razvoj, lahko tvori besedila v materinem jeziku s pripomočki, ki mu svetujejo glede besedišča in struktur, ki jih uporablja, sorazmerno uspešno lahko razume besedila v tujem jeziku brez poznavanja jezika in lahko govorno komunicira z računalnikom ter drugimi digitalnimi napravami. Govorec slovenščine bo verjetno še nekaj časa prikrajšan pri več segmentih. Najbolj pomembna za nadaljnji razvoj so pravzaprav področja, kjer nihče drug ne more opraviti dela za neko jezikovno skupnost, pri segmentih, ki so vezani na jezik, ne pa na samo tehnologijo. Prva zadrega torej nastane že na ravni, ki še ni računalniška, je pa nujen pogoj za kvalitetne jezikovne pripomočke v digitalnem mediju – spodobna dokumentiranost stanja sodobnega jezika v obliki, ki je primerna za prenos v digitalne aplikacije, od vseh vrst slovarjev do uspešnega oblikoslovno-skladenskega označevalnika. Brez teh elementov je težko graditi naprej.

Drug pomemben element je racionalnost pri uporabi razpoložljivih finančnih sredstev in človeških potencialov, kajti jasno je, da če ena in ista računalniška aplikacija lahko v primerljivih ekonomskih okoliščinah zadovolji potrebe petdeset milijonov uporabnikov ali dveh milijonov, bodo prvi imeli petindvajsetkrat več možnosti, da do nje pridejo. Ta premisa v tej jezikovni skupnosti do sedaj ni bila najbolj upoštevana.

## Literatura

Čermák, František, 2002: Today's corpus linguistics: Some open questions, *International Journal of Corpus Linguistics* 7/2. 265–282.

Erjavec, Tomaž, 1996/97: Računalniške zbirke besedil. *Jezik in slovstvo* 2/3. 81–95.

Erjavec, Tomaž, 1998a: Oznake korpusa FIDA. Kačič, Zdravko, (ur.): *Uporabno jezikoslovje* 6. Tematska številka »Jezikovne tehnologije«. 85–95.

Erjavec, Tomaž, 1999a: Making the ELAN Slovene/English Corpora. Language Technologies – Multilingual Aspects. Špela Vintar (ur.): *Proceedings of the workshop within the framework of the 32<sup>nd</sup> Annual Meeting of the Societa Linguistica Europea, 8–11 July 1999. Ljubljana*: Filozofska fakulteta, Oddelek za prevajanje in tolmačenje. 23–30.

Erjavec, Tomaž, 1999b: Slovensko-angleški korpus ELAN. *Slavistična revija* 47/4. 515–522.

Erjavec, Tomaž, Gorjanc, Vojko in Stabej, Marko, 1998: Korpus FIDA. Tomaž Erjavec in Jerneja Gros (ur.): *Jezikovne tehnologije za slovenski jezik/Language Technologies for the Slovene Language*. Zbornik konference/Proceedings of the Conference. 6. do 7. oktober 1998. Ljubljana: Institut Jožef Stefan. 124–127.

Gorjanc, Vojko, 1999a: Kaj in kako v korpus FIDA. *Razgledi* 13, 23. jun. 1999. 7–8.

Gorjanc, Vojko, 1999b: Korpusi v jezikoslovju in korpus slovenskega jezika FIDA. Kržišnik, Erika in Lokar, Meta (ur.): *35. seminar slovenskega jezika, literature in kulture*. Ljubljana: Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovanske jezike in književnost Filozofske fakultete Univerze v Ljubljani. 47–59.

Gorjanc, Vojko, 2000: Nekatere možnosti jezikoslovne izrabe enojezikovnih korpusov. Orel, Irena in Lokar, Meta (ur.): *36. seminar slovenskega jezika, literature in kulture*. Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovanske jezike in književnosti Filozofske fakultete Univerze v Ljubljani. 335–348.

Gorjanc, Vojko in Vintar, Špela, 2000: Iskanja po Korpusu slovenskega jezika FIDA. Erjavec, Tomaž in Gros Jerneja (ur.): *Informacijska družba, Jezikovne tehnologije*. Ljubljana 17.–19. okt. 2000. 20–26.

Gros, Jerneja, 2000: *Samodejno tvorjenje govora iz besedil: postopek za izdelavo sintetizatorja slovenskega govora*. Ljubljana: Založba ZRC SAZU.

Gros, Jerneja, 2001: Sinteza slovenskega govora – jezikoslovni vidiki. Orel, Irena in Lokar, Meta (ur.): *37. seminar slovenskega jezika, literature in kulture*. Ljubljana: Filozofska fakulteta, Oddelek za slovanske jezike in književnost, Center za slovenščino kot drugi/tuji jezik. 21–31

Hirci, Nataša, 1998: Korpusi v prevodoslovju. Tomaž Erjavec in Gros Jerneja (ur.): *Jezikovne tehnologije za slovenski jezik/Language Technologies for the Slovene Language*. Zbornik konference/Proceedings of the Conference. 6. do 7. oktober 1998. Ljubljana: Institut Jožef Stefan. 113–118.

Jakopin, Primož in Bizjak, Aleksandra, 1997: O strojno podprtem oblikoslovnem označevanju slovenskega besedila. *Slavistična revija* 45/3-4. 513–532.

Kačič, Zdravko in Horvat, Bogomir, 1998: Izgradnja infrastrukture potrebne za razvoj govorne tehnologije za slovenski jezik. Erjavec, Tomaž in Gros, Jerneja (ur.): *Jezikovne tehnologije za slovenski jezik/Language Technologies for the Slovene Language*. Zbornik konfe-

rence/Proceedings of the Conference. 6. do 7. oktober 1998. Ljubljana: Institut Jožef Stefan. 100–104.

Kilgarriff, Adam, 2001: Web as corpus. *Proceedings of Corpus Linguistics 2001*. Lancaster: University of Lancaster.

Košir, Aleš, 2002. Sodobni prosti črkovalniki in baze pravih besednih oblik. Erjavec, Tomaž in Gros, Jerneja (ur.): *Jezikovne tehnologije/Language Technologies*. Zbornik konference/Proceedings of the Conference. 14. do 15. oktober 2002. Ljubljana: Institut Jožef Stefan. 171–172.

Košir, Aleš, Peterlin, Primož in Erjavec, Tomaž, 1998. GNUSSL: Prosto programje in slovensčina. Erjavec, Tomaž in Gros, Jerneja (ur.): *Jezikovne tehnologije za slovenski jezik/Language Technologies for the Slovene Language*. Zbornik konference/Proceedings of the Conference. 6. do 7. oktober 1998. Ljubljana: Institut Jožef Stefan. 35–41.

Landau, Sidney I., 2001: *Dictionaries, The Art and Craft of Lexicography, 2<sup>nd</sup> Edition*. Cambridge: Cambridge University Press.

Pisanski, Agnes, 1998. Virtualna učilnica: uporaba interneta pri poučevanju tujih jezikov. Erjavec, Tomaž in Gros, Jerneja (ur.): *Jezikovne tehnologije za slovenski jezik/Language Technologies for the Slovene Language*. Zbornik konference/Proceedings of the Conference. 6. do 7. oktober 1998. Ljubljana: Institut Jožef Stefan. 96–99.

Romih, Miro, 1998a: Amebis in jezikovne tehnologije. Erjavec, Tomaž in Gros, Jerneja (ur.): *Jezikovne tehnologije za slovenski jezik/Language Technologies for the Slovene Language*. Zbornik konference/Proceedings of the Conference. 6. do 7. oktober 1998. Ljubljana: Institut Jožef Stefan. 29–34.

Romih, Miro, 1998b: Direktorijška struktura korpusa FIDA. Kačič, Zdravko (ur.): *Uporabno jezikoslovje* 6. Tematska številka »Jezikovne tehnologije«. 79–84.

Romih, Miro in Holozan, Peter, 2002a. Infrastruktura za razvoj jezikovnih tehnologij – korpus FIDA in sistem ASES. Erjavec, Tomaž in Gros, Jerneja (ur.): *Jezikovne tehnologije/Language Technologies*. Zbornik konference/Proceedings of the Conference. 14. do 15. oktober 2002. Ljubljana: Institut Jožef Stefan. 166.

Romih, Miro in Holozan, Peter, 2002b. Elektronski slovarji. Erjavec, Tomaž in Gros, Jerneja (ur.): *Jezikovne tehnologije/Language Technologies*. Zbornik konference/Proceedings of the Conference. 14. do 15. oktober 2002. Ljubljana: Institut Jožef Stefan. 165.

Stabej, Marko, 1998: Besedilnovrstna sestava korpusa FIDA. Kačič, Zdravko (ur.): *Uporabno jezikoslovje* 6. Tematska številka »Jezikovne tehnologije«. 96–106.

Šef, Tomaž, Dobnikar, Aleš, Gams, Matjaž, Grobelnik, Marko, 1998. Slovenski govor na internetu. Erjavec, Tomaž in Gros, Jerneja (ur.): *Jezikovne tehnologije za slovenski jezik/Language Technologies for the Slovene Language*. Zbornik konference/Proceedings of the Conference. 6. do 7. oktober 1998. Ljubljana: Institut Jožef Stefan. 60–64.

Vintar, Špela, 1999a: A Lexical Analysis of the ELAN Slovene-English Parallel Corpus. Špela Vintar (ur.): *Language Technologies – Multilingual Aspects. Proceedings of the workshop within the framework of the 32<sup>th</sup> Annual Meeting of the Societa Linguistica europea, 8–11 July 1999*. Ljubljana: Filozofska fakulteta, Oddelek za prevajanje in tolmačenje. 63–69.

Vintar, Špela, 1999b: Računalniško podprto iskanje terminologije v slovensko-angleškem vzporednem korpusu. *Uporabno jezikoslovje* 7-8. 156–169.

Vintar, Špela, 2002. Avtomatsko luščenje izraza iz slovensko-angleških vzporednih besedil. Erjavec, Tomaž in Gros, Jerneja (ur.): *Jezikovne tehnologije/Language Technologies*. Zbornik



konference/Proceedings of the Conference. 14. do 15. oktober 2002. Ljubljana: Institut Jožef Stefan. 78–85.

Zupan, Jure, 1999: Problemi in nekaj rešitev računalniških obdelav slovenskih besedil. *Slavistična revija* 47/3. 277–296.

Željko, Miran, 2002. Pripomočki na spletu za prevajalce zakonodaje EU. Erjavec, Tomaž in Gros, Jerneja (ur.): *Jezikovne tehnologije/Language Technologies*. Zbornik konference/Proceedings of the Conference. 14. do 15. oktober 2002. Ljubljana: Institut Jožef Stefan. 33–38.

## Slovarji

BSJ – *Slovar slovenskega knjižnega jezika z odzadnjim slovarjem slovenskega jezika in besediščem slovenskega jezika z oblikoslovnimi podatki*. Elektronska izd. na plošči CD-ROM, 1998. Ljubljana: DZS, SAZU in ZRC SAZU.

Debenjak, Božidar, Doris in Primož, 1992. *Veliki nemško-slovenski slovar*. Ljubljana: DZS. ES – *Evropski slovar*, 1995. Ljubljana: Mladinska knjiga.

Grad, Anton, Škerlj, Ružena in Vitorovič, Nada, 1978. *Veliki angleško-slovenski slovar*. Ljubljana: Državna založba Slovenije.

OSSJ – *Slovar slovenskega knjižnega jezika z odzadnjim slovarjem slovenskega jezika in besediščem slovenskega jezika z oblikoslovnimi podatki*. Elektronska izd. na plošči CD-ROM, 1998. Ljubljana: DZS, SAZU in ZRC SAZU.

SP – *Slovenski pravopis*, 2001. Ljubljana: SAZU in Inštitut za slovenski jezik Frana Ramovša ZRC SAZU.

SSKJ – *Slovar slovenskega knjižnega jezika z odzadnjim slovarjem slovenskega jezika in besediščem slovenskega jezika z oblikoslovnimi podatki*. Elektronska izd. na plošči CD-ROM, 1998. Ljubljana: DZS, SAZU in ZRC SAZU.

VST – *Veliki slovar tujk*, 2002. Ljubljana: Cankarjeva založba.

## Spletne strani

*Slovarji:*

a) enojezični

<<http://bos.zrc-sazu.si/sskj>>

b) dvojezični

<[http://www.amebis.si/izdelki/elektronsko\\_zaloznistvo.htm](http://www.amebis.si/izdelki/elektronsko_zaloznistvo.htm)>

<<http://www.dictionary.com>>

<<http://www.yourdictionary.com>>

<<http://www.onelook.com>>

<<http://www.oxfordreference.com>>

<<http://www.sigov.si/slovar.htm>>

<<http://www.pinkponk.com>>

<<http://slovenian.palm-dictionaries.com>>

<<http://www.wizcomtech.com>>

<<http://www.iskraprins.si>>

*c) terminološki*

<<http://www.sigov.si/slovar.htm>>  
<[http://17slon.com/kgb/prevajanje/slovarji\\_in\\_glosarji.htm](http://17slon.com/kgb/prevajanje/slovarji_in_glosarji.htm)>  
<<http://www.matkurja.com/slo/resources/entertainment/dictionaries>>  
<<http://www.sigov.si/evroterm>>

*Leksikoni:*

<<http://bos.zrc-sazu.si>>  
<<http://www.amebis.si/sklanjanje>>  
<<http://www.amebis.si/izdelki/abz>>

*Korpusi:*

*a) enojezični*

<<http://www.fida.net>>  
<<http://bos.zrc-sazu.si>>  
<<http://nl2.ijs.si/index-mono.html>>

*b) dvojezični*

<<http://nl2.ijs.si/index-bi.html>>  
<<http://nl.ijs.si/elan>>  
<<http://www-ai.ijs.si/čspela/trans-index.html>>

*c) terminološki*

<<http://www.sigov.si/evrokorpus>>

*d) spletni*

<<http://www.webcorp.org.uk>>  
*Jezikovna orodja v pisarniških programskih paketih:*  
<<http://www.amebis.si/izdelki/crkovalnik/>>  
<<http://nl.ijs.si/GNUsl>>  
<<http://openoffice.lugos.si>>  
<[http://whiteboard.openoffice.org/lingucomponent/download\\_dictionary.html](http://whiteboard.openoffice.org/lingucomponent/download_dictionary.html)>  
<<http://www.amebis.si/izdelki/besana>>

*Avtomatski prevajalni sistemi:*

<<http://presis.amebis.si>>  
<<http://presis.amebis.si/prevajanje.asp>>  
<<http://www.tranexp.com>>  
<<http://www.tranexp.com:2000>>

*Govorne tehnologije:*

<<http://www.ff.uni-lj.si/sft>>  
<<http://ai.ijs.si/govorec>>  
<<http://mpx.masterpoint.si/voice>>