

Online SLO-DE-SLO: Spletni slovensko-nemški in nemško-slovenski slovar

Birte Lönneker*, Katarina Rozman†

*Institut für Germanistik II
Von-Melle-Park 6, D-20146 Hamburg
birte.loenneker@uni-hamburg.de

†Univerza v Ljubljani, Filozofska fakulteta
Aškerčeva ulica 2, SI-1000 Ljubljana
katarina.rozman@siol.net

Povzetek

V članku je predstavljen spletni slovar *Online SLO-DE-SLO*, ki postaja eden izmed pomembnejših slovenskih virov z uporabo modernih jezikovnih tehnologij na medmrežju, kar kažejo tudi statistike poizvedovanj, ki jih beleži poseben program na strežniku. Slovensko-nemški oz. nemško-slovenski slovar je prosto dostopen vsem uporabnikom medmrežja. Uvodoma so v članku predstavljeni trenutni način delovanja slovarja, njegova struktura in vsebina ter viri besedišča. Obravnavani so tudi vidiki interaktivnosti v slovarju ter pojem interaktivnosti, ki se pogosto uporablja v povezavi z medmrežnimi aplikacijami. Na koncu je razložena nova struktura slovarja, ki je že sedaj podlaga za poskusno različico z novo oblikoslovno funkcijo, ki je koristna pri poizvedovanju in pri prikazovanju najdenih vpisov slovarja.

Online SLO-DE-SLO: A Slovenian-German and German-Slovenian online dictionary

This paper presents the online dictionary *Online SLO-DE-SLO*, a Slovenian resource using modern language engineering technologies on the internet. Its high importance as a resource is also shown by statistics of user queries, collected by a program on the server. The Slovenian-German-Slovenian dictionary is freely available to everybody who has access to the World Wide Web. The paper first briefly presents the current functioning, structure, and contents of the dictionary as well as the sources of the contained vocabulary. Interactivity aspects of the dictionary are also dealt with, based on some reflections on the concept of interactivity, often used in discussions about web applications. Finally, the new structure of the dictionary is explained. It serves already now as a basis for a test version including a new morphologic function, useful when querying the dictionary and displaying the retrieved entries.

1. Uvod

Spletni slovensko-nemški in nemško-slovenski slovar *Online SLO-DE-SLO*, ki ga v tem prispevku predstavljava, je prvi in edini prosto dostopen slovar te vrste na internetu. Slovar je javno objavljen na svetovnem spletu od aprila 2001. Na začetku je bil slovar primeren predvsem za nemško govoreče ljudi, ki se učijo slovenščino, danes pa je namenjen vsem ljudem, ki jih zanimata slovenščina in nemščina v kombinaciji ter trenutno vsebuje več kot 7.000 vpisov. Statistike na strežniku kažejo, da je *Online SLO-DE-SLO* zelo priljubljen spletni jezikovni vir. S povečanim številom vpisov in z izboljšano strukturo, ki jo že uporablja pri novi poskusni različici, bo slovar lahko postal tudi dragocen del drugih sistemov jezikovnih tehnologij.

V prispevku je najprej na splošno opisana vsebina slovarja (2. poglavje), v nadaljevanju pa še spletni vmesnik in statistika poizvedovanj (3. poglavje), vidiki interaktivnosti pri spletnih slovarjih (4. poglavje), natančna predstavitev današnje strukture slovarja *Online SLO-DE-SLO* in prihodnja izboljšana struktura ter vmesnik (5. poglavje). Sklep (6. poglavje) vsebuje ugotovitve dosedanjega dela s slovarjem *Online SLO-DE-SLO* in predstavlja načrte za prihodnost.

2. Vsebina slovarja

Vsebino slovarja lahko opišemo na dva načina; **a)** glede na to, koliko besed in s katerega področja je mogoče najti v slovarju ter **b)** kako so te besede vpisane v slovar. To poglavje se zato razdeli v dve podpoglavji. Viri in področja besedišča so obravnavani v podpoglavju 2.1.,

podpoglavje 2.2. pa bolj natančno predstavlja enoto "vpis" in vsebino vpisa v slovarju *Online SLO-DE-SLO*.

2.1. Viri in področja besedišča

Slovar *Online SLO-DE-SLO* je bil na začetku namenjen predvsem učečim se slovenščine kot drugega/tujega jezika, po novem pa vsebuje tudi že besedilno gradivo z različnih področj, čeprav največji del slovarja še vedno temelji na enojezičnem učbeniku za slovenščino kot drugi/tuji jezik *Odkrivajmo slovenščino* (Čuk et al., 1996), ki je namenjen pouku odraslih začetnikov slovenskega jezika na tujih univerzah (npr. na Univerzi v Hamburgu) in tečajih slovenščine kot drugega/tujega jezika v Sloveniji. Besedišče tega prvega dela slovarja (več kot 5.500 vpisov) torej lahko opišemo kot večinoma splošno in vsakdanje.

Vir	Področje besedišča	Število vpisov
Učbenik	splošno, vsakdanje	5.544
Časopis DELO	splošno, politika, gospodarstvo, šport	743
Besedni seznam (strokovni)	strojništvo, orodjarstvo	829
<i>Skupno</i>		7.116

Tabela 1: Trenutna vsebina slovarja *Online SLO-DE-SLO*

Leta 2003 je bilo v slovar dodanih več kot 700 vpisov, pri čemer sta bila osrednja kriterija, da besede slovar tedaj še ni vseboval ter pogostost rabe le-te: obravnavane so bile namreč najpogostejše uporabljene besede v časopisu DELO. Ti vpisi so lahko tudi manj splošni, saj besedila v

časopisu pogosto tematizirajo dogodke s področja politike, gospodarstva in športa. Postopek izbora časopisnega besedišča je opisan v Lönneker in Jakopin (2003).

Marca leta 2004 je bilo dodanih še okoli 800 terminoloških vpisov s področja strojništva, izbranih na podlagi strokovnega seznama¹ in pripravljenih s pomočjo slovenskih in nemških strokovnjakov. Trenutno slovar vsebuje več kot 7.000 vpisov. V tabeli 1 najdemo pregled opisane vsebine slovarja; v poglavju 3.1 bo razloženo, kako uporabnik lahko omeji iskanje na izbrano področje.

2.2. Vpis

V prejšnjem poglavju nisva obravnavali enote "beseda", ampak "vpis". Razlog za to je, da lahko predstavlja vpis v *Online SLO-DE-SLO* nekaj drugega kot beseda. Vpis si lahko predstavljamo kot eno vrstico v datoteki, iz katere nastaja slovar. Vsak vpis v slovarju je dvojezičen. Trenutno obstajajo tri vrste vpisov:

- osnovna besedna oblika oziroma slovenska lema (npr. *opisovati*) in natančno en prevod v nemščino – drugi prevod je torej tudi drug vpis;
- oblikoslovna oblika (npr. *opisujem*) in njen prevod v nemščino;
- slovenski stavek (npr. *Telefon ne dela.*), del stavka (npr. *točno v tistem času*) ali kolokacija (npr. *človeška ribica*). Stavki in besedne zveze ilustrirajo uporabo besed v kontekstu; včasih je namreč dobesedni prevod nemogoč ali nepopoln. Kolokacija je posebna večbesedna semantična enota, v kateri se besede pogosto ne prevajajo dobesedno (neposredno). Taki vpisi seveda vključijo tudi prevode teh večbesednih enot; za zgornje primere so to: 'Das Telefon geht nicht', 'genau zu der Zeit' in 'Grottenolm'.

Mogoče je najti tudi vpise, ki so večbesedni v slovenskem jeziku, ampak ustrezajo samostojni besedi v nemškem jeziku, npr. *človeška ribica* 'Grottenolm', *delo v parih* 'Partnerübung' (pri vajah v učbeniku); *telefonski naročnik* 'Fernsprechteilnehmer' itd. Leme in oblike so ponavadi označene z oblikoslovno oznako ali drugačno informacijo o slovnici, npr. tudi z vezljivostjo glagolov, pri obeh jezikih. Vsi slovenski vpisi so opremljeni z naglasnimi znamenji, tako da je mogoče razbrati naglašeni zlog posamezne besede. Na koncu vpis vsebuje še ime vira, na katerem temelji. Vpis v *Online SLO-DE-SLO* ima torej splošno obliko kot je ilustrirano na sliki 1, v kateri oznaka *lt* pomeni tabulator, podpisani številki 1 in 2 pa nemški oz. slovenski jezik.

(lema oblika stavek) ₁ \t(slovnica) ₁ \t
(lema oblika stavek) ₂ \t(slovnica) ₂ \t(vir)

Slika 1. Splošna oblika vpisa v *Online SLO-DE-SLO*

V datotekah slovarja je vpis napisan v eni vrstici, tako da ga lahko preberemo v podatkovno zbirko, iz katere uporabniki poizvedujejo preko spletnega vmesnika. Tretje poglavje predstavlja spletni vmesnik in trenutne možnosti poizvedovanja.

¹ Osnovno zbirko besed s področja strojništva je prvi avtorici pričujočega prispevka ponudil Jani Melik.

3. Spletni vmesnik in poizvedovanje

S svojim prosto dostopnim spletnim vmesnikom² je *Online SLO-DE-SLO* zelo priljubljen spletni jezikovni vir. V tem poglavju je razložena uporaba trenutnega vmesnika in možnosti poizvedovanja (podpoglavje 3.1.) ter podanih nekaj številke o poizvedovanjih, ki jih je zabeležil program na strežniku (podpoglavje 3.2.). Na koncu (podpoglavje 3.3.) je na kratko omenjeno, zakaj spletni vmesnik ni koristen samo za uporabnike, ampak tudi za avtorje slovarja.

3.1. Uporaba vmesnika

Vmesnik slovarja obstaja v dveh različicah, namreč v nemškem in slovenskem³ jeziku. Na osnovni strani obeh različic imajo uporabniki možnost iskanja niza z natančnim ujemanjem, torej natančno s tem, kar so natipkali. Izjeme so naslednje:

- ni pomembno, ali napišemo besedo z velikimi ali z malimi črkami;
- posebne znake slovenske oz. nemške abecede, ki jih morebiti prav vsi uporabniki spleta nimajo na svoji tipkovnici, lahko nadomestimo s črkami *c*, *s*, *z* za slovenske črke *č*, *š* in *ž* ali s kombinacijami *ae*, *oe*, *ue* in *ss* za nemške črke *ä*, *ö*, *ü* in *ß*.

Vmesnik vsebuje dve okenci za vnos, in sicer eno za vnos slovenskih besed, ki jih prevajamo v nemščino, in eno za vnos nemških besed, ki jih prevajamo v slovenščino. Tudi v tem se torej tukaj predstavljen slovar loči od drugih spletnih slovarjev, ki ponujajo eno samo okence in prevajajo v oba jezika sočasno⁴. Takšen postopek je bolj smiseln v primeru, da jezika nimata velikega števila enakopisnic (npr. *je* je enakopisnica v nemščini in v slovenščini, zato bi v tovrstnem primeru takšen slovar izpisal obe možnosti).

Za razširjen način iskanja izberemo eno od možnosti poizvedovanja iz levega navigacijskega stolpca: DE→SLO (če hočemo prevajati iz nemščine v slovenščino) ali SLO→DE (za nasprotno smer). Tako pridemo do razširjenega vmesnika (glej sliko 2). Ta omogoča najprej omejitev iskanja po virih. Tukaj uporabnik lahko izbira področje in obseg besedišča, ki ga zanima: besedišče na podlagi učbenika, časopisnih besedil ali strokovnih seznamov. Seveda jih lahko tudi kombinira. Kombinacija izbire vseh virov je v slovarju privzeta, kar pomeni iskanje po celotnem slovarju. Funkcija "omejitve iskanja" na posamezne zbirke omogoča uporabnikom, da uporabljajo slovar *Online SLO-DE-SLO* kot bi bil sestavljen iz večih knjig: Strokovnjak lahko izbira samo seznam strojniških izrazov, učenec jezika le učbenike, itd. Na tak način *Online SLO-DE-SLO* izkorišča prednost elektronskega in še posebej spletnega medija, ki omogoča visoko stopnjo prilagodljivosti ponujenega gradiva. Po de Schryverju (2003) je ena izmed prednosti elektronskih slovarjev v primerjavi z tiskanimi slovarji ta, da lahko uporabnikom omogoča izbiro natančno *takšnega* slovarja,

² <http://www.rrz.uni-hamburg.de/slowenisch> [29. avgusta 2004].

³ http://www.rrz.uni-hamburg.de/slowenisch/index_si.htm [29. avgusta 2004].

⁴ Npr. osnovna verzija nemško-angleškega slovarja <http://dict.leo.org> [28. avgusta 2004].

ki ga potrebujejo za določen namen: "Many different target user groups [...] are simultaneously catered for in paper dictionaries, making such products highly redundant for any one use(r). [...] In EDs [electronic dictionaries] users could be enabled to select exactly *that* type of dictionary required for the consultation at hand" (de Schryver 2003:188).

Na razširjenem vmesniku slovarja *Online SLO-DE-SLO* lahko izbiramo pod naslovom *Kako najti iskani niz?* še med tremi možnostmi iskanja:

- **natančno ujemanje** (enako kot na osnovni strani);
- **natančno ujemanje ali kot del besedila**; ta način najde tudi nize, ki se pojavljajo znotraj besed; npr. za niz *ogl* dobimo seveda *oglas* in *ogled*, pa tudi *ubogljiv* in *vrtočlavica*, tako kot še veliko drugih besed, oblik in stavkov;
- **besede, ki se začnejo s tem nizom**; ta način najde npr. za niz *ogl* besede *oglas*, *ogled*, in vse druge vpise, ki se začnejo z nizom *ogl*, ne pa tudi besed *ubogljiv* ali *vrtočlavica*.

Kot rezultat dobimo seznam najdenih vpisov, v katerem so le-ti urejeni v tabeli po abecednem redu in sicer v izhodiščnem jeziku, t.j. jeziku, v katerem smo napisali iskano besedo ali niz. To pomeni, da so pri vnosu slovenskega niza glasov (besede ali dela besede), kot npr. *opi*, izpisi (ki jih uporabnik iz slovenščine želi prevesti v nemščino) urejeni tako kot prikazuje slika 3.

3.2. Nekaj statistike o poizvedovanjih

V zimskem obdobju 2002/2003 je bilo mesečno zabeleženih že približno 10.000 poizvedovanj po slovarju. V obdobju od aprila 2003 do aprila 2004 je program na strežniku zabeležil več kot 180.000 poizvedovanj, kar

pomeni povprečno 15.000 poizvedovanj na mesec. V maju 2004 je bilo zabeleženih natančno 23.228 poizvedovanj.

Evalvaciji poizvedovanj (Lönneker in Jakopin, 2003; Jakopin in Lönneker, 2004) sta pokazali, da je v uporabniški vmesnik spletnega slovarja vneseno več poizvedovanj v smeri nemško-slovensko kot v smeri slovensko-nemško. Na začetku je bilo razmerje približno 3 : 1 (Lönneker in Jakopin, 2003); v zadnjem času pa se je nekoliko izravnalo in trenutno znaša približno 3 : 2 (Jakopin in Lönneker, 2004). Na podlagi teh rezultatov pa še ne vemo, ali je slovar bolj priljubljen pri nemško ali slovensko govorečih uporabnikih, saj uporabnik seveda lahko prevaja iz maternega jezika tako kot iz tujega jezika. Statistike uporabe različnih vmesnikov v nemškem oz. slovenskem jeziku pa pokažejo približno isto razmerje (3 : 2 za nemški jezik). Če denimo uporabniki poizvedujejo z vmesnika v svojem maternem jeziku, lahko sklepamo, da je slovar še vedno bolj uspešen pri nemško govorečih uporabnikih.

3.3. Kaj lahko spoznamo iz poizvedovanj

Spletni vmesnik ni koristen samo za uporabnike, pač pa tudi za avtorje slovarja. Avtorji namreč na podlagi statistik poizvedovanj lahko analizirajo potrebe uporabnikov brez njihove vednosti, torej tudi brez vplivanja na njihove vedenje: "The electronic environment makes possible the truly exciting, namely the potential to retrieve feedback without the user being aware of this" (de Schryver, 2003, 188). Iz statistik poizvedovanj lahko prepoznamo veliko zanimivih podatkov, kot npr. kako narašča oz. upada pogostost uporabe slovarja, katera smer prevoda je bolj priljubljena pri uporabnikih, in nenazadnje, katere besede in stavke ter iz katerih področj uporabniki najpogosteje vnašajo oz. iščejo.

The screenshot shows the 'Online SLO-DE-SLO' website interface. At the top right, it says 'Online SLO-DE-SLO Spletni slovensko-nemški in nemško-slovenski slovar'. Below this is a navigation bar with 'Online SLO-DE-SLO | Poizvedovanje' and a German flag icon. On the left, there are search options: 'Poizvedovanje: DE->SLO' and 'SLO->DE'. The main content area is titled 'Slovensko-nemško poizvedovanje' and features a search box with 'opi' entered. Below the search box, there are three radio button options for 'Kako najti iskani niz?': 'Natančno ujemanje', 'Natančno ujemanje ali kot niz v besedilu', and 'Besede, ki se začnejo s tem nizom'. The third option is selected. There is also a 'Nemško' button. At the bottom, there is a footer with 'Copyright ? Birte Lönneker 2001-2004.' and 'Zadnja sprememba strani: 27.07.2004'.

Slika 2. Iskanje za vpise, ki se v slovenščini začnejo z nizom *opi*

Online SLO-DE-SLO
Spletni slovensko-nemški in nemško-slovenski slovar

Online SLO-DE-SLO | Poizvedovanje

Poizvedovanje:
 ▶ DE->SLO
 ▶ SLO->DE

Rezultati

Izbrani viri: - *Učbeniki* - *Časopisna besedila* - *Strojništvo* - *Gradivo v pripravi*

Besede, ki se začnejo z: *opi*

nemško	slovnica	slovensko	izgovorjava	slovnica	korpus
Affe	N nom.sg.m.	opica	opica	N nom.sg.f.	Odkrivajmo slovenščino stran 71
Affen...		opičji	opičji	(opičja, opičjo) ADJ	Odkrivajmo slovenščino stran 71
beschreiben	V	opisati	opisati	V perf.	Odkrivajmo slovenščino stran 59
beschreiben	V	opisovati	opisovati	V impf.	Odkrivajmo slovenščino stran 136
schildern	V	opisovati	opisovati	V impf.	Odkrivajmo slovenščino stran 136
<i>ich</i> beschreibe	V 1.sg.präs.	opisujem	opisujem	V 1.sg.präs.impf.	Odkrivajmo slovenščino stran 136
<i>ich</i> schildere	V 1.sg.präs.	opisujem	opisujem	V 1.sg.präs.impf.	Odkrivajmo slovenščino stran 136
<i>ich</i> beschreibe	V 1.sg.präs.	opišem	opišem	V 1.sg.präs.perf.	Odkrivajmo slovenščino stran 59

Copyright ? Birte Lönneker 2001-2004.

Slika 3. Rezultati iskanja za vpise, ki se v slovenščini začnejo z nizom *opi*

Prva dva podatka, to sta pogostost uporabe in priljubljenost smeri prevoda, sta kvantitativna: lahko jih izračunamo neposredno iz številskih vrednosti, in zato ju je na podlagi podatkov v zbirki relativno enostavno navesti (glej 3.2.). Nasprotno pa je tretji podatek kvalitativen: pove, *kakšno* besedišče najbolj zanima uporabnike. Postopka za evalvacijo tega podatka sta dva, in sicer **a)** primerjava poizvedovanj s trenutno vsebino slovarja (Lönneker in Jakopin, 2003) ali pa **b)** primerjava poizvedovanj z morebitnimi vrstami besedilnih korpusov, iz katerih bi lahko izbrali dodatne vpise slovarja. Točka b) nas je v zadnjem času najbolj zanimala; težave pri pripravljanju slovarjev se namreč tudi po Anžlovarju (2004) začnejo z izborom besedišča.

Bolj podrobno nas je zanimalo, če bi bilo v *Online SLO-DE-SLO* koristno najprej razširiti del besedišča, ki temelji na časopisu DELO. Pri raziskavah, ki so natančno opisane v Jakopin in Lönneker (2004), nas torej niso zanimale najpogostejše besede uravnoveženega korpusa (Gorjanc in Krek, 2001), temveč določanje besedilne vrste, ki bi najbolj ustrezala željam uporabnikov. Narejena je bila primerjava frekvenc vseh lem v zabeleženih poizvedovanjih z relativnimi frekvencami lem treh po besedilni vrsti različnih korpusov: ta so bila leposlovna, časopisna (DELO) ter strokovna besedila.⁵ Frekvence lem iz literarnih virov so se pri evalvaciji pokazali kot najbolj podobne frekvencam lem iz poizvedovanj. Zato trenutno časopis DELO ni najustreznejši vir za izpopolnjevanje slovarja *Online SLO-DE-SLO*, temveč bi potrebe uporabnikov najbolj zadovoljila vključitev izbora novih besed in izrazov iz leposlovnih besedil.

⁵ Korpusi so deli slovenskega besedilnega korpusa *Nova Beseda* (<http://bos.zrc-sazu.si> [4. septembra 2004]), ki vsebuje več kot 120 milijonov besed.

4. Interaktivnost

Je slovar *Online SLO-DE-SLO* interaktiven? Odgovor na to vprašanje ni enostaven, saj obstaja več možnosti razlage pojma "interaktivnost". Izraz izvira s področja sociologije (podpoglavje 4.1.), na medmrežju pa ima drugačen in ožji pomen (podpoglavje 4.2.). Če namreč razširimo prvi pomen interaktivnosti na področje spletnih slovarjev, pridemo do nekaterih idealov leksikografije (podpoglavje 4.3.). V vsakem podpoglavju obravnavava tudi bolj konkretno interaktivnost slovarja *Online SLO-DE-SLO* po vsakokratni razlagi pomena interaktivnosti v določenem kontekstu.

4.1. Interaktivnost na področju sociologije

Izraz izvira s področja sociologije, kjer pomeni: "vzajemno delovanje, medsebojno učinkovanje ljudi, skupin" (*Leksikon Cankarjeve Založbe*, 1998). V splošnem nemškem leksikonu *Der Brockhaus in fünf Bänden* (1993) še izvemo, da se delovanje individuumov v tem medsebojnem učinkovanju prilagaja glede na pričakovanja drugih individuumov: "Wechselbeziehung zwischen den Individuen innerhalb der Gesellschaft, besonders die wechselseitige Orientierung des Handelns an den Erwartungen der anderen in Gruppen und größeren sozialen Systemen." Interaktivnost je torej lastnost družbenega sistema, ki se označuje po meri interakcije ljudi v tem sistemu.

Niti tiskan, niti digitalen slovar ni delujoč individuum in se zato ne more sam udeležiti interakcije v tem pomenu. Če pa pod pojmom interakcije razumemo medsebojno učinkovanje *ljudi*, lahko samo rečemo, da preko slovarja nastaja interakcija med avtorjem in uporabnikom, saj avtor napiše in oblikuje slovar na tak način, kot se mu zdi najbolj primerno za uporabnika. Slednji uporablja slovar na način, ki naj bi ga po uporabnikovem mnenju predvidel

tudi avtor. Taka interakcija med individuimi (med avtorjem in uporabnikom) je zelo posredna. Bolj neposredna bi bila tista, kjer si avtor in uporabnik izmenjata mnenje o slovarju (npr. preko pisma ali elektronske pošte). Zato najdemo na spletnih straneh slovarja *Online SLO-DE-SLO* tudi elektronski naslov avtorice.

4.2. Interaktivnost na področju medmrežja

Na področju računalništva ima pridevnik "interaktiven" po *Leksikonu Cankarjeve Založbe* (1998) še drug splošen pomen, namreč "lastnost sistema, da vsakemu uporabnikovemu ukazu v kratkem času ponudi ustrezen odgovor." Ustrezen odgovor je lahko izpis v ukazovalni vrstici, prikaz vnesenih črk v program za obdelavo besedila, predvajanje glasbe po kliku na gumb itd.

Na medmrežju pa termin "interaktiven" natančneje pomeni, da program na strežniku ustvarja dinamične strani v formatu HTML; dinamične so zato, ker se ne nahajajo vedno pripravljene na strežniku, temveč so zapisane avtomatično, v hipu zahteve oz. poizvedovanja. Vsebina dinamične strani je odvisna od želje uporabnika, ki jo izrazi preko vnesenih podatkov: "an 'interactive' Web application is the one that creates dynamic HTML pages dependent on user input" (Beca et al., 1997, 3). V tem "medmrežnem" smislu npr. odgovor na klikanje na povezavo do statičnih strani HTML ni interaktiven, ker pri tem postopku strežnik ne oblikuje novih strani, temveč samo pošlje že obstoječe strani.

Online SLO-DE-SLO je interaktiven v prvem ("statičnem") in v drugem ("dinamičnem") smislu. Preko poizvedovanja lahko dobimo dinamične strani z ustreznimi rezultati, v nasprotju z "neinteraktivnim" spletnim slovarjem, kjer bi gotovo oblikovane strani vsebovale celotno vsebino slovarja.⁶ Seveda najdemo tudi povezave do statičnih strani v slovarju *Online SLO-DE-SLO*, take povezave pa obstajajo tudi na dinamičnih straneh z rezultati: če dobimo na dinamični strani vsaj en rezultat (vsaj eno ujemanje), najdemo tam tudi povezavo do statične strani z razlagami slovničnih informacij (v nemškem jeziku) in za vsako ujemanje tudi še povezavo do opisa vira, iz katerega je vpis vzet; seveda najdemo povezavo "nazaj" do strani, s katere smo poslali poizvedovanje, tako kot do prve strani slovarja itd. Takšne povezave lahko imenujemo "notranje povezave", ker napotijo do strani, ki so na istem strežniku. "Zunanje povezave", ki napotijo do strani na drugih strežnikih, so zbirane v ločeni sekciji slovarja pod naslovom *Povezave*.

Seveda si lahko predstavljamo primere, v katerih bi bilo koristno imeti zunanje povezave tudi na dinamični strani rezultatov, npr. za vsak vpis bi lahko ponudili povezavo na drug spletni slovar ali na konkordančnik korpusa. Razlog zato, da takih povezav (še) ni, je organizacijske narave. Ne le, da bi bilo potrebno napisati nove funkcije v programu na strežniku, ampak bi jih bilo potrebno spreminjati tudi vsakokrat, ko bi se spremenili programi za druge slovarje oz. konkordančnike, ki jih bi bilo treba torej stalno opazovati.

4.3. Druge interaktivne leksikografske sanje

Interakcijo med avtorjem in uporabnikom slovarja, ki je že bila obravnavana v podpoglavju 4.1., bi bilo mogoče urediti tudi preko drugega vmesnika, na katerem bi uporabnik lahko vnesel manjkajoče besede in izraze skupaj z morebitnimi prevodi.

Kot sta opazila Mangeot-Lerebours in Sérasset (2002), take možnosti ne bi smeli ponuditi brez kontrole, ker bi uporabniki tudi lahko (namenoma ali ne) vnesli napačne podatke. Poleg tega pa bi uporabniki, če bi jim ponudili možnost dodajanja vnosov, pričakovali, da bi avtorji predlagane vnose čimprej obravnavali in ustrezne predloge čimprej vključili v javno različico slovarja. Te zahteve je opazil že Burke (1995, 42–43) v zvezi z popraviljem napak v spletnih slovarjih. Ker pa je *Online SLO-DE-SLO* privaten projekt in ni stalno financiran, trenutno še nima možnosti vzpostavitve takega vmesnika ali pa foruma uporabnikov, kjer bi le-ti lahko sami diskutirali o prevodih besed⁷, ali kjer bi si med seboj pomagali in popravljali kot v izpopolnjenem uporabniškem sistemu *Papillon*, ki ga predstavljata Mangeot-Lerebours in Sérasset (2002). Namesto tega raje iskoriščamo možnost posredne interakcije, kjer avtorji analizirajo poizvedovanja uporabnikov in razširjajo besedišče na podlagi rezultatov iz teh analiz, kot je bilo predstavljeno zgoraj v poglavju 3.3.

De Schryver (2003) je opisal še veliko drugih "leksikografskih sanj", povezanih z interaktivnostjo v ožjem in širšem pomenu, npr. tudi t.i. sanje o dostopnosti (de Schryver, 2003, 173–182). Eden izmed številnih načinov izboljševanja dostopnosti do vsebine slovarja je možnost omejitve iskanja na različna tematska področja slovarja, ki jo *Online SLO-DE-SLO* že ponuja (glej poglavje 3.1. zgoraj). Drug način je iskanje glede na oblikoslovne kategorije, ki je posebej koristno pri oblikoslovno bogatih jezikih kot sta npr. slovenščina in nemščina. Prve korake do tovrstnega iskanja po *Online SLO-DE-SLO* opiševa v naslednjem poglavju.

5. Struktura slovarja: več možnosti

Kot je bilo omejeno pri opisu enote "vpis" (podpoglavje 2.2.), uporablja slovar podatkovno zbirko, ki teče na strežniku s pomočjo programa MySQL. Podatkovna zbirka lahko bere tekstovne datoteke, kjer tabulatorji razmejujejo posamezna podatkovna polja; npr. tabela v podatkovni zbirki, ki bi neposredno brala vpise s strukturo prikazano na sliki 1, bi imela pet polj. Že na sliki 2 zgoraj je bilo mogoče videti, da različni programi za pripravljane leksikografskih datotek ter za prikazovanje rezultatov v slovarju *Online SLO-DE-SLO* dodajo še nekaj slovničnih informacij. V tem poglavju bova obravnavali trenutno funkcijo, ki uporablja takšno znanje o slovnici (podpoglavje 5.1.) in poskusno funkcijo, ki iskorišča povezavo med različnimi oblikoslovnimi oblikami (podpoglavje 5.2.). Na kratko bova nakazali tudi možnost in morebitno prednost XML-zapisa slovarja (podpoglavje 5.3.).

⁶ Statičen dvojezičen slovarček, ki vsebuje tudi slovenski jezik, je npr. Hladnikova *Slovenščina za popotnike*, <http://www.ff.uni-lj.si/sft/> [8. september 2004].

⁷ Take forume ponuja že omenjen nemško-angleški spletni slovar <http://dict.leo.org>.

Št.	DE	SLO	Slovn. DE	Slovn. SLO	Vir
1	Fußgänger	pešec	N nom.sg.m.	N nom.sg.m.	[Učbenik]
2	Lampe	svetilka	N nom.sg.f.	N nom.sg.f.	[Učbenik]
3	Leuchte	svetilka	N nom.sg.f.	N nom.sg.f.	[Učbenik]
4	wohnen	prebivati	V	V impf.	[Učbenik]
5	[...]	[...]	[...]	[...]	[...]

Tabela 2: Struktura glavne tabele (poenostavljeno).

Št.	Slovnicična beseda DE	DE	Slovnicična beseda SLO	SLO	Slovnicična oznaka DE	Slovnicična oznaka SLO	Vir	Št. glavne tabele
1	des	Fußgängers		pešca	N gen.sg.m.	N gen.sg.m.	[Učbenik]	1
2	ich	wohne		prebivam	V 1.sg.präs.	V 1.sg.präs.impf.	[Učbenik]	4
3	[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]

Tabela 3: Struktura tabele za besedne oblike (poenostavljeno).

5.1. Od leksikografskih datotek do podatkovne zbirke

Za pretvorbo leksikografskih datotek v tabele podatkovne zbirke uporabljava poseben program, ki ga je prva avtorica napisala v programskem jeziku Perl. Pri pretvorbi program doda še nekaj informacij, tako da je tabela v podatkovni zbirki bogatejša in ima več polj kot leksikografske datoteke. Na sliki 2 zgoraj je viden npr. izpis dveh različic slovenskega dela slovarja, ene brez in ene z izgovarjavo (*opica* vs. *opica*), kar je pravzaprav enostavno, saj program za pretvorbo samo doda ali odstrani naglas. Bolj zahtevna pri tem postopku je delitev polj z informacijami za nemške vnose, ki jih vidimo tudi v rezultatih na sliki 2; npr. '*ich* schildere', kjer je le zaimek *ich* napisan v poševnem tisku. Program razdeli nemško polje v leksikografski datoteki v dve polji za tabelo v podatkovni zbirki: npr. izraz *ich schildere* 'opisujem' razdeli v *ich* (zaimek in zato nepolnopomenska beseda) in *schildere* (oblika glagola in zato polnopomenska beseda). V podatkovni zbirki imajo nepolnopomenske (slovnicične) in polnopomenske besedne vrste ločena polja. Zato jih prepozna tudi program za prikazovanje rezultatov in jih lahko natisne na dva različna načina.

Tovrstne informacije so zanimive tudi pri interpretaciji poizvedovanj. Program lahko najde vpis *ich schildere* tudi za krajšo poizvedbo *schildere*, in sicer že pri iskanju natančnih ujemanj. Slovar se torej odzove na enak način pri poizvedovanju besede *schildere* kot pri poizvedovanju "frazo" *ich schildere*, kar pomeni večjo fleksibilnost: Uporabniki namreč včasih vpišejo le glagolsko obliko, včasih pa poleg te vpišejo tudi nepolnopomenske (slovnicične) besede.

5.2. Izboljšana struktura in oblikoslovno iskanje

V pripravi je izboljšana struktura slovarja, ki je sestavljena iz dveh tabel. Ena od teh tabel vsebuje le besedne oblike, ki niso v osnovni slovarski obliki (glej tabelo 3). Vsak vpis v tej tabeli kaže preko številke v polju "Št. glavne tabele" na številko ustreznega vpisa v drugi tabeli, ki je glavna tabela in vsebuje osnovne slovarske oblike (glej tabelo 2). Tako je omogočen bolj ustrezen

vrstni red izpisa rezultatov, kjer besedne oblike neposredno sledijo svoji osnovni slovarski obliki (npr. oblika *opisem* sledi besedi *opisati* v seznamu rezultatov).

Na nejavni strani slovarja sva dodali poskusno funkcijo poizvedovanja, ki iskorišča to izboljšano strukturo slovarja. Pri vsaki poizvedbi poišče osnovno obliko in vse zanjo znane besedne oblike, tudi če se ne začnejo z istim nizom, in jih zapiše neposredno po osnovni obliki. Besedne oblike so v tej poskusni različici slovarja dodatno označene s sliko puščice. Slika 4 prikazuje rezultate istega iskanja kot pri sliki 3, urejene in oblikovane na nov način v poskusni verziji. S slike 4 je razvidno, da so v poskusni verziji oblike glagolov *opisati* in *opisovati* napisane neposredno po njegovih osnovnih oblikah, v nasprotju z trenutno verzijo, kjer so vse napisale po abecednem redu in se zato nahajajo na koncu seznama (slika 3). Ker pa nam rezultati poizvedovanj že sedaj izpišejo precej zadetkov in so dinamične strani dolge, posebej pri nenatančnem iskanju, sva se odločili, da se bova odpovedali stolpcu, v katerem so bile do sedaj prikazane slovenske besede brez informacije o naglasu. Uporabniki bodo torej potem dobili slovenske besede le z informacijo o naglasu (glej sliko 4).

5.3. Možnost XML-zapisa

Z zgoraj predstavljeno novo strukturo, ki je prvi korak k oblikoslovni komponenti, bo slovar še ustrežnejši vir za druge jezikovne aplikacije. Nadalje bo mogoče tudi raziskati, kako lahko napišemo strukturo slovarja *Online SLO-DE-SLO* v formatu XML, podobno kot so avtorji Erjavec et al. (2004) naredili za japonsko-slovenski slovar.

Vendar to vprašanje trenutno ni najbolj pereče, saj sva že napisali ustrezne programe za pretvarjenje leksikografskih datotek v format podatkovne baze MySQL in slovar brez težav deluje tudi v tej obliki. Dolgoročno pa bi XML-različica imela prednost, da bi se lažje poslužili sodobnih programov za leksikografsko delo, torej za pripravljanje vpisov, saj je na tem področju XML zelo priljubljen format (Krek, 2003). Poleg tega bo XML-različica omogočala tudi lažjo izmenjavo podatkov, npr. z drugimi slovarji za slovenski jezik, tako kot z drugimi jezikovnimi aplikacijami.

Online SLO-DE-SLO
Spletni slovensko-nemški in nemško-slovenski slovar

Online SLO-DE-SLO | Poizvedovanje

Poizvedovanje:
 ▶ DE→SLO
 ▶ SLO→DE

Rezultati

Izbrani viri: - Učbeniki - - Strojništvo - Gradivo v pripravi

Besede, ki se začnejo z: *opi*

nemško	slovnica	slovensko ↓	slovnica	korpus
Affe	N nom.sg.m.	opica	N nom.sg.f.	Odkrivajmo ... str. 71
Affen...		opičji	(opičja, opičjo) ADJ	Odkrivajmo ... str. 71
beschreiben	V	opisati	V perf.	Odkrivajmo ... str. 59
▶ <i>ich</i> beschreibe	V 1.sg.präs.	▶ opišem	V 1.sg.präs.perf.	Odkrivajmo ... str. 59
beschreiben	V	opisovati	V impf.	Odkrivajmo ... str. 136
▶ <i>ich</i> beschreibe	V 1.sg.präs.	▶ opisujem	V 1.sg.präs.impf.	Odkrivajmo ... str. 136
schildern	V	opisovati	V impf.	Odkrivajmo ... str. 136
▶ <i>ich</i> schildere	V 1.sg.präs.	▶ opisujem	V 1.sg.präs.impf.	Odkrivajmo ... str. 136

Copyright ? Birte Lönneker 2001-2004.

Slika 4: Rezultati iskanja v jasnejem vrstnem redu - poskusna verzija

V ta namen bo potrebno najprej izvedeti, kakšno strukturo XML datotek oz. kateri DTD (*document type description*) te aplikacije uporabljajo. XML-zapis trenutne vsebine slovarja bi bil seveda mogoč že sedaj, saj je najin program sposoben izpisati celotno vsebino slovarja tudi v HTML-obliki, ki pa jo potem lahko brez težav pretvorimo v XML-obliko. Toda če bi pri tem uporabili DTD po standardu TEI⁸, kot so npr. to naredili Erjavec et al. (2004), bi ga zaradi strukture najinega slovarja uporabili na drugačen način. V slovarju *Online SLO-DE-SLO* so namreč npr. stavki in deli stavkov samostojni vpisi, v japonsko-slovenskem slovarju pa so podrejeni izbranemu geslu, kar bi posledično pomenilo uporabo drugih elementov iz istega DTD-ja.

6. Sklep

Predstavili sva namen, delovanje, trenutno in prihodnjo (poskusno) strukturo slovarja *Online SLO-DE-SLO*.

Delo bova nadaljevali na dva načina. Vsebinsko bova razširili, kar pomeni, da bova dodali več vpisov. Ker je ovrednotenje statistike poizvedovanj po slovarju pokazalo, da uporabnike zanimajo predvsem besede in fraze, ki se pojavljajo v vsakdanji rabi jezika, pa tudi pogovorne besede (Jakopin in Lönneker, 2004), bodo nove besede izbrane iz novejših učbenikov in iz slovenske izvirne literature. Tovrstni učbeniki, ki vsebujejo veliko primerne leksike iz vsakdanjih situacij, in temeljijo na komunikativni metodi, so npr. *Slovenska beseda v živo* (Markovič et al., 2002), *A, B, C ... 1, 2, 3, gremo* (Pirih Svetina in Ponikvar, 2003), *Slovenščina na koncu jezika* (Pirih Svetina, 2003) ter *S slovenščino nimam težav* (Markovič et

al., 2003). Kar zadeva slovensko izvirno literaturo, kratke zgodbe in romani vsebujejo dialoge in opise, v katerih lahko najdemo uporabo veliko pogovornih izrazov.

Drugi del načrtovanega dela predstavlja izpopolnjevanje strukture slovarja in objava oblikoslovne funkcije pri poizvedovanju, kot je bilo predstavljeno v 5. poglavju.

7. Zahvala

Avtorici se zahvaljujeva Evropski zvezi za leksikologijo EURALEX za nagrado Laurence Urdang iz leta 2001, s katero je bilo omogočeno delo na slovarju v letih 2002 in 2003, ter Uradu za Slovenski jezik za odobritev sofinanciranja projekta leta 2004.

8. Literatura

- P. Anžlovar. 2004. Slovarji: težave se začnejo že z izborom besed. *Nedelo*, 6. junija 2004:27.
- L. Beca, G. Cheng, G. C. Fox, T. Jurga, K. Olszewski, M. Podgorny, P. Sokolowski, T. Stachowiak in K. Walczak. 1997. *TANGO – a Collaborative Environment for the World Wide Web*. Tehnično poročilo, Northeast Parallel Architectures Center, Syracuse University, New York.
- Brockhaus in fünf Bänden*. 1993. Brockhaus.
- S. M. Burke. 1998. *The Design of Online Lexicons*. Magistrsko delo, Northwestern University, Evanston, IL.
- M. Čuk, M. Mihelič in G. Vuga. 1996. *Odkrivajmo slovenščino*. Filozofska fakulteta, Seminar slovenskega jezika, literature in kulture.
- T. Erjavec, K. Hmeljak Sangawa, I. Srdanović in A. Vahčić. 2004. Making an XML-based Japanese-Slovene Learners' Dictionary. V: *Proceedings of the*

⁸ <http://www.tei-c.org/> [8. septembra 2004]

- Fourth International Conference on Language Resources and Evaluation (LREC-04), Lizbona, Portugalska, 26. – 28. maja 2004, ELRA.*
- V. Gorjanc in S. Krek. 2001. A corpus-based dictionary database as the source for compiling Slovene-X dictionaries. V: *Proceedings of the 6th Conference on Computational Lexicography and Corpus Research (COMPLEX-01), Birmingham, 28. junija – 1. julija 2001*, Centre for Corpus Linguistics, Department of English, University of Birmingham.
- S. Krek. 2003. Sodobna dvojezična leksikografija. *Jezik in slovnstvo*, 1:45–60.
- P. Jakopin in B. Lönneker. 2004. Query-driven dictionary enhancement. V: *Proceedings of the 11th EURALEX International Congress, Lorient, Francija, 6. – 10. julija 2004*, EURALEX.
- Leksikon Cankarjeve Založbe*. 1998. Cankarjeva založba, 3. izd.
- B. Lönneker in P. Jakopin. 2003. Contents and evaluation of the first Slovenian-German online dictionary. V: *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL-03) Conference Companion, Budimpešta, Madžarska, 12. – 17. aprila 2003*, ACL.
- M. Mangeot-Lerebours in G. Sérasset. 2002. Frameworks, Implementation and Open Problems for the Collaborative Building of a Multilingual Lexical Database. V: *Proceedings of the SEMANET Workshop (Post COLING-2002 Workshop), Taipei, Taiwan, 31. avgusta 2002*.
- A. Markovič, V. Halužan, M. Pezdirc Bartol, D. Škapin in G. Vuga. 2003. *S slovenščino nimam težav*. Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete, 2. izd.
- A. Markovič, D. Škapin, K. Rigler Šilc in Š. Kajič Kmetič. 2002. *Slovenska beseda v živo*. Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete, 3. izd.
- N. Pirih Svetina. 2003. *Slovenščina na koncu jezika*. Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete, ponatis 1. izd.
- N. Pirih Svetina in A. Ponikvar. 2003. *A, B, C ... 1, 2, 3, gremo*. Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete.
- G.-M. de Schryver. 2003. Lexicographers' dreams in the electronic-dictionary age. *International Journal of Lexicography*, 16/2:143–199.