

UPORABNIŠKA EVALVACIJA KORPUSA FidaPLUS: ZASNOVA VPRAŠALNIKA, PRVI REZULTATI

Špela Arhar

Trojina, zavod za uporabno slovenistiko, Škofja Loka

UDK 811.163.6'322.2

V sklopu priprav na nadgradnjo referenčnega korpusa za slovenski jezik poteka uporabniška evalvacija trenutno aktualnega referenčnega korpusa FidaPLUS. Osnova evalvacije je elektronski vprašalnik, v katerem korpusni uporabnik opredeli svoje izkušnje s posameznimi korpusnimi orodji oz. značilnostmi korpusnega vmesnika. Prispevek opisuje zgradbo vprašalnika ter podaja prvo analizo odgovorov – z osredotočanjem na tisti del vprašanj, ki so zaprtega tipa.

FidaPLUS, uporabniška evalvacija, korpusni vmesnik, Sporazumevanje v slovenskem jeziku, Konkordančnik ASP32

As part of the preparations for the upgrade of the Slovene reference corpus FidaPLUS, a user evaluation of the current corpus is underway. The basis of the evaluation is an electronic questionnaire by means of which corpus users rate their experience with specific corpus tools and corpus interface features. The paper describes the structure of the questionnaire and provides a preliminary response analysis focusing on the closed questions.

FidaPLUS, user evaluation, corpus interface, Communication in Slovene, ASP32 Concordancer

1 Korpus FidaPLUS

Referenčni korpus FidaPLUS je na internetni strani (<http://www.fidaplus.net/>) prosto dostopen za uporabo od začetka leta 2007. Uporabnikom je delo s korpusom omogočeno s pomočjo Konkordančnika ASP32, ki je bil razvit posebej za potrebe slovenskega korpusnega jezikoslovja ob pripravi korpusa FIDA (<http://www.fida.net/>).¹

Konkordančnik ASP32 je zmogljivo programsko orodje, ki omogoča zahtevnejše oblike iskanja po lematiziranem ter oblikoskladenjsko označenem korpusu ter ponuja orodja za selekcijo ter organizacijo v konkordančnem nizu predstavljenih jezikovnih podatkov. Funk-

cionalnosti programa so bile v literaturi že večkrat predstavljene,² prav tako so bili pripravljene priročniki za učenje dela s korpusom, v katerih je uporaba konkordančnika natančneje razložena.³

Kljub temu, da je bila ob gradnji korpusa FidaPLUS že opravljena delna poenostavitev konkordančnega vmesnika (Arhar, Gorjanc 2007: 103–106), pa je – glede na to, da je referenčni korpus namenjen različnim tipom uporabnikov – smotrno razmisliti o dodatnih programskih spremembah, tako na ravni funkcionalnosti kot preglednosti ponujenih orodij. Naslednja nadgradnja referenčnega korpusa se načrtuje v sklopu projekta *Sporazumevanje*

1 Konkordančnik ASP32 se poleg pri omenjenih korpusih uporablja tudi pri korpusu KoRP, specializiranem korpusu besedil odnosov z javnostmi: <http://www.korp.fdv.uni-lj.si/>.

2 Denimo v Arhar, Gorjanc 2007, Gantar 2007, Gorjanc 2005, Gorjanc, Vintar 2000.

3 Arhar 2007, Logar, Arhar 2008.

v slovenskem jeziku,⁴ pri katerem je predvidena povsem nova zasnova konkordančnika, pri kateri bodo med drugim upoštevane tudi povratne informacije zainteresirane uporabniške javnosti, na kar se osredotoča pričujoči prispevek.

2 Vprašalnik za uporabniško evalvacijo

2.1 Priprava in distribucija vprašalnika

Uporabniška evalvacija poteka s pomočjo elektronskega vprašalnika. Začela se je julija 2009 in bo trajala predvidoma šest mesecev, tj. do konca leta 2009. Vprašalnik je bil pripravljen s pomočjo internetnega servisa *Free Online Surveys*,⁵ ki poleg priprave vprašalnika omogoča tudi njegovo distribucijo, npr. enostavno vključitev v že obstoječe internetne strani, pošiljanje po e-pošti ipd. Povezava na obravnavani vprašalnik je bila vključena na internetne strani korpusa FidaPLUS.⁶

2.2 Sestava vprašalnika

Vprašalnik se osredotoča na dve glavni temi v zvezi s posameznimi funkcijami Konkordančnika ASP32: (I) kako pogosto uporabnik določeno programsko funkcijo uporablja, (II) kakšne predloge ima uporabnik za izboljšavo obravnavane programske funkcije. Poleg vprašanj, neposredno vezanih na konkordančnik, vprašalnik prinaša splošna vprašanja o namenu in načinu rabe referenčnega korpusa.

Večina vprašanj je zaprtega tipa – uporabnik izbira med vnaprej podanimi odgovori (nekatera vprašanja dovoljujejo samo en odgovor, pri nekaterih pa je možnih več odgovorov). V primeru, da je med odgovori na voljo kategorija *Drugo*, ima možnost vpisa zelenega

odgovora. Pri tem sklopu vprašanj je uporabnikov odgovor po večini pogoj za nadaljevanje izpolnjevanja vprašalnika. Vprašanja odprtega tipa, na katera odgovor ni obvezen, so na mestih zbiranja mnenj ter predlogov za izboljšavo korpusa, in sicer v povezavi z: (I) načini iskanja po korpusu ter (II) korpusnimi orodji, ki so na voljo (npr. statistike, sito itn.), uporabniki pa lahko podajo tudi splošno mnenje o tem, (III) kaj jih pri korpusu FidaPLUS najbolj moti ter (IV) česa pri nadgradnji ne bi spreminjali.

Vprašalnik sestavljajo tri strani. Prva prinaša vprašanja za opredelitev vzorca vprašanih (glej poglavje 3.1) ter vprašanja o splošni oceni korpusa FidaPLUS (glej poglavje 3.2). Druga stran se osredotoča na zbiranje informacij o pogostnosti rabe posameznih funkcionalnosti Konkordančnika ASP32 (glej poglavje 3.3), zadnja pa je namenjena uporabniškim komentarjem glede možnih korpusnih sprememb. Vprašanja odprtega tipa zahtevajo posebno obravnavo in zato ostajajo izven dome-ta pričujočega prispevka. Prav tako na tem mestu ni prostora za natančnejšo interpretacijo pridobljenih podatkov – slednja bo opravljena po zaključku evalvacije.

3 Prvi rezultati evalvacije

V prispevku so predstavljeni trenutni rezultati: do 24. avgusta 2009 je prvo stran vprašalnika izpolnilo 143 korpusnih uporabnikov, 125 jih je vprašalnik izpolnilo do konca.

3.1 Uvodna vprašanja o načinu in namenu rabe korpusa FidaPLUS

Uvodni del vprašalnika prinaša vprašanja, ki opredeljujejo vzorec vprašanih (opredelitev starosti ter mesta zaposlitve oz. študija) ter

4 <http://www.slovenscina.eu>.

5 Na voljo na strani <http://freeonlinesurveys.com/>. Plačljiva različica programa omogoča izdelavo poljubno dolgega vprašalnika z vprašanji različnega tipa, program pa tudi delno analizira rezultate ter jih grafično predstavi.

6 Uporabniki so bili k reševanju vprašalnika pozvani prek ustreznih poštnih seznamov (*SlovLit*, *SDIT*), forumov (forum prevajalcev, študentov slovenistike v Ljubljani ter Mariboru), na blogu *slovenistika* ipd. Z začetkom novega študijskega leta bodo na vprašalnik dodatno opozorjeni profesorji ter študentje relevantnih študijskih smeri.

sprašujejo po namenu ter pogostnosti uporabe korpusa FidaPLUS. Uporabniki so vprašani, kako so izvedeli za korpus FidaPLUS, kako so se naučili njegove uporabe ter katera druga podobna orodja uporabljajo pri svojem delu.

Skoraj polovica vprašanih je starih med 25 in 34 let (45,5 %), sledita kategoriji 18 do 24 let (22,4 %) ter 35 do 44 let (18,9 %). Vprašalnik je rešilo manj uporabnikov, starejših od 44 let (13,3 %), uporabnikov, mlajših od 18 let, pa v vzorcu ni.

Več kot polovica vprašanih je zaposlenih in se pri delu ukvarja z jezikom (50,3 %), sledijo študentje na dodiplomski (29,2 %) ter podiplomski (12,4 %) ravni, nekaj je tudi zaposlenih, ki jim jezik pri delu ni ključni interes (3,7 %).⁷

Največ vprašanih uporablja korpus nekajkrat mesečno (38,5 %), dosti pa tudi pogosteje: nekajkrat tedensko (20,3 %) ali skoraj vsak dan (16,8 %). Odgovor *nekajkrat letno* je izbralo 17,5 % vprašanih, *korpus sem začel-a uporabljati šele pred kratkim* pa 4,9 %.

Pri vprašanju za kakšen namen večinoma uporabljajo korpus FidaPLUS,⁸ so korpusni uporabniki najpogosteje izbrali odgovore: *kot pomoč pri lektoriranju besedil* (20,3 %), *za izvajanje jezikoslovnih raziskav* (15,9 %), *kot pomoč pri prevajanju* (15,0 %) in *pisanju* (14,1 %) *besedil*. Za ljubiteljsko raziskovanje jezika uporablja korpus 13,4 % vprašanih, za pripravo učnega gradiva 6,9 %, za pripravo seminarških oz. diplomskih nalog 6,7 %, magistrskih oz. doktorskih nalog 4,8 % ter za reševanje domačih nalog 1,6 % vprašanih.

Večina vprašanih je za korpus FidaPLUS izvedela od učitelja oz. profesorja (39,3 %) ali od kolegov (30,3 %), manj pa iz strokovne literature (12,4 %) ali na strokovnem izobraževa-

nju (9,0 %). Po naključju je korpus našlo 1,7 % vprašanih.

Dela s korpusom so se vprašani po večini naučili sami s pomočjo priročnika (45,2 %) ali pa jih je naučil učitelj oz. profesor (20,4 %) ali kolega oz. sodelavec (18,3 %). Manjši delež vprašanih se je uporabe korpusa naučil na strokovnem izobraževanju (6,5 %) oz. se uporabe še ni naučil (2,7 %).

Za raziskovanje jezika večina vprašanih uporablja tudi druga orodja, največ internetne brskalnike (111 vprašanih), korpus Nova beseda (74 vp.), korpuse tujih jezikov (58 vp.) ter specializirane korpuse (29 vp.), poleg tega pa tudi Wordsketches (24 vp.) ter WordSmith (11 vp.).⁹ Odgovor *ne uporabljam* je izbralo 10 vprašanih.

3.2 Splošna evalvacija korpusa FidaPLUS

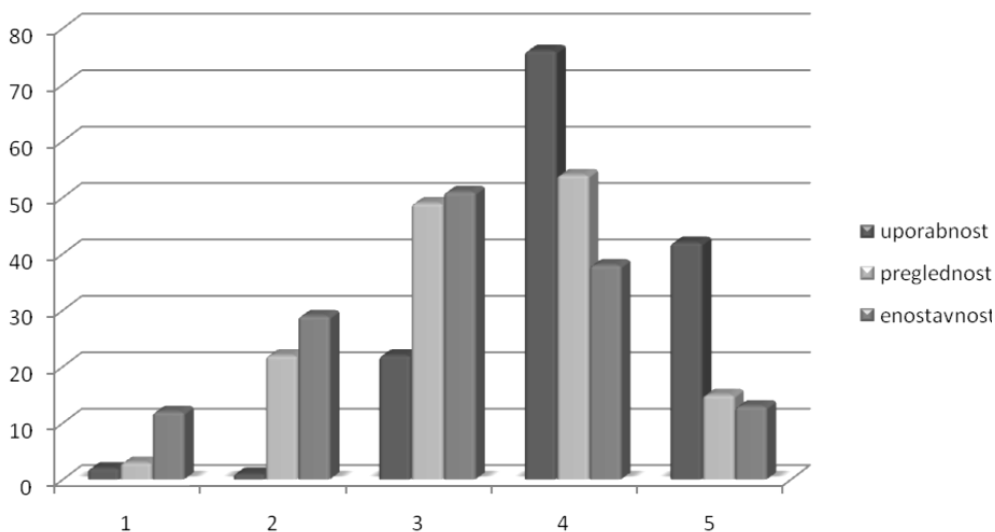
V sklopu splošne evalvacije korpusa so vprašani z oceno od 1 (najslabše) do 5 (najboljše) ocenjevali: (I) uporabnost korpusa za svoje delo, (II) preglednost korpusnega vmesnika (razporeditev ikon, prikaz podatkov itn.) ter (III) enostavnost uporabe korpusa. V drugem koraku, kjer je na voljo tudi odgovor *ne poznam*, so vprašani na enak način ocenjevali (IV) uporabnost priročnika *Kaj početi s korpusom FidaPLUS* ter (V) uporabnost pomoči, ki je na voljo znotraj korpusnega vmesnika: tabele z oblikoskladenjskimi oznakami (t. i. kodami MSD), povzetek načinov iskanja itn.

Rezultate ocenjevanja prikazujeta Grafa 1 ter 2. Vsi grafi v nadaljevanju prispevka na osi x prinašajo nabor odgovorov, med katerimi vprašani uporabnik lahko izbira, na osi y pa število vprašanih, ki so izbrali posamezni odgovor.

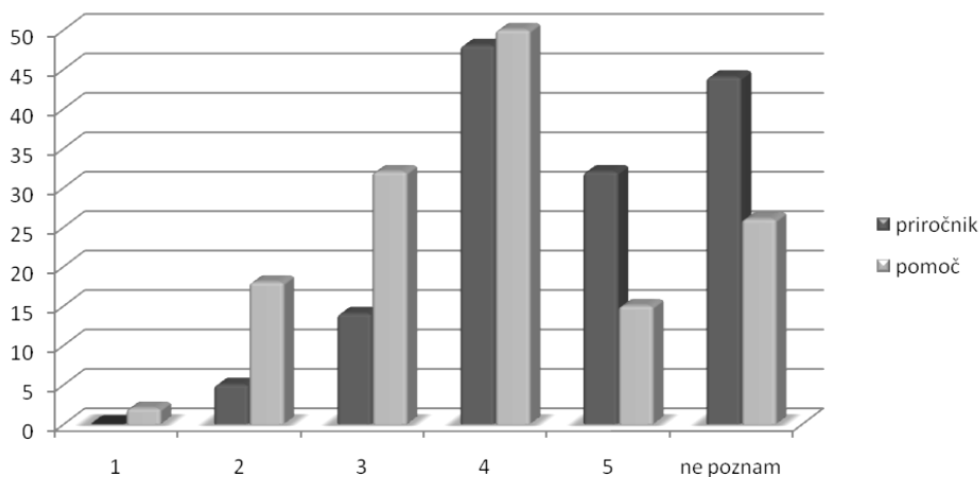
⁷ Obravnavano vprašanje, kot tudi sledeča vprašanja poglavja 3.1, ima med odgovori na voljo kategorijo *Drugo*, ki tipično prinaša okrog 5 % odgovorov. Te odgovore na tem mestu puščamo ob strani, vključeni pa bodo v končno analizo rezultatov.

⁸ Pri obravnavanem vprašanju in sledečih vprašanjih poglavja 3.1 lahko vprašani izbere več odgovorov.

⁹ Nova beseda: http://bos.zrc-sazu.si/s_beseda.html, Wordsketches: <http://www.sketchengine.co.uk/>, WordSmith: <http://www.lexically.net/wordsmith/>.



Graf 1: Uporabniška ocena uporabnosti, preglednosti ter enostavnosti rabe korpusa FidaPLUS.



Graf 2: Uporabniška ocena priročnika za delo s korpusom FidaPLUS ter vmesniške pomoči.

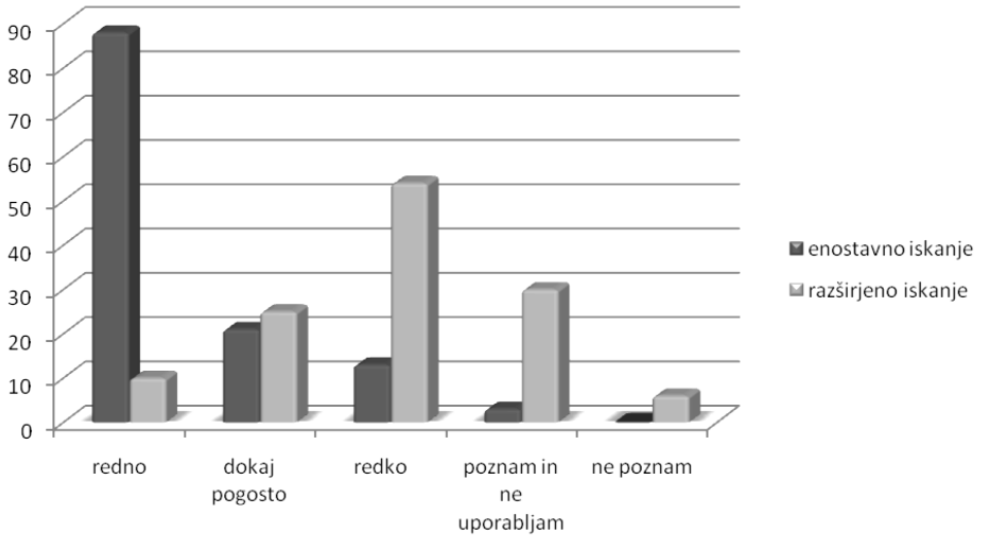
Vprašani korpusni uporabniki v splošnem najvišje ocenjujejo uporabnost korpusa FidaPLUS ter priročnik za delo s korpusom (oboje ima povprečno oceno 4,1). Nekoliko manj so vprašani zadovoljni z vmesniško pomočjo (povp. ocena 3,5) ter splošno preglednostjo

vmesnika (povp. ocena 3,4), najnižjo oceno pa so pripisali enostavnosti rabe korpusa (povp. ocena 3,1). Zgovoren je podatek, da vmesniške pomoči ne pozna 18,2 % vprašanih, priročnika pa kar 30,8 % vprašanih.

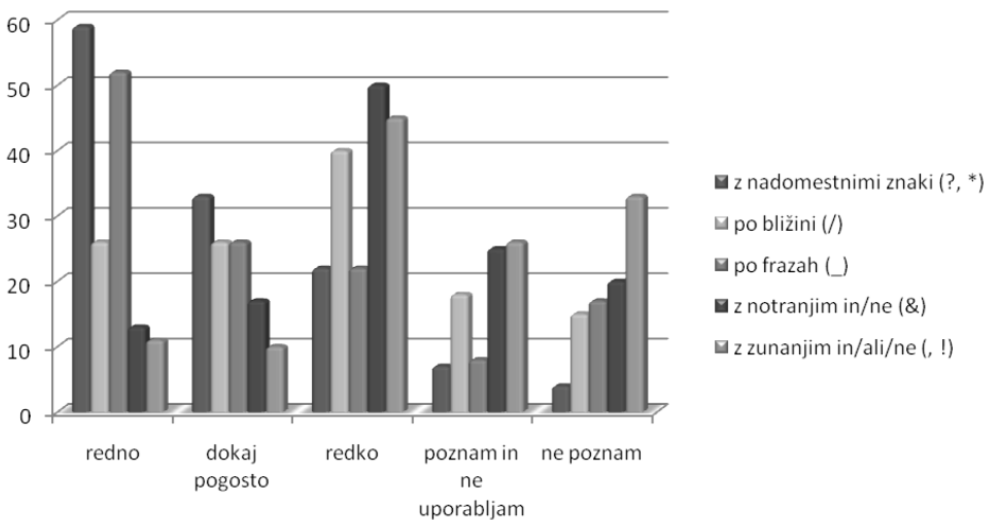
3.3 Pogostnost uporabe posameznih funkcij Konkordančnika ASP32

Vprašani so nadalje ocenjevali, kako pogosto uporabljajo določene funkcije Konkordančnika ASP32 za izdelovanje konkordančnega

niza: najprej glede pogostnosti uporabe enostavnega ter razširjenega iskanja (Graf 3), nato glede pogostnosti uporabe posameznih regularnih izrazov kot pomoč pri sestavi iskalnega pogoja (Graf 4).



Graf 3: Pogostnost rabe enostavnega ter razširjenega iskanja.



Graf 4: Pogostnost rabe regularnih izrazov kot pomoč pri iskanju.

Graf 3 kaže, da vprašani uporabniki bistveno pogosteje uporabljajo enostavno iskanje kot razširjeno. Vprašani, ki razširjeno iskanje uporabljajo, so imeli v nadaljevanju možnost opredelitve, kateri načini omejevanja zadetkov se jim zdijo uporabni za delo: filtriranje po zvrsti je izbralo 58 vprašanih, po lektoriranoosti 52 vprašanih, po letu izida 49 vprašanih, po prenosniku 41 vprašanih, po oznaki Cobiss pa 17 vprašanih.

Zanimive rezultate kaže tudi Graf 4. Od zahtevnejših oblik iskanja vprašani najpogosteje uporabljajo iskanje z nadomestnimi znaki ter iskanje po frazah, manj pogosto iskanje po bližini, najredkeje pa iskanje z notranjim oz. zunanjim in/ali/ne. Slednjega načina iskanja obenem velik delež (26,4 %) vprašanih uporabnikov ne pozna. Kot je razvidno iz grafa, se relativno pomanjkljivo poznavanje kompleksnih načinov izdelave iskalnega pogoja izkazuje pri vseh naštetih kategorijah, razen pri iskanju z nadomestnimi znaki.

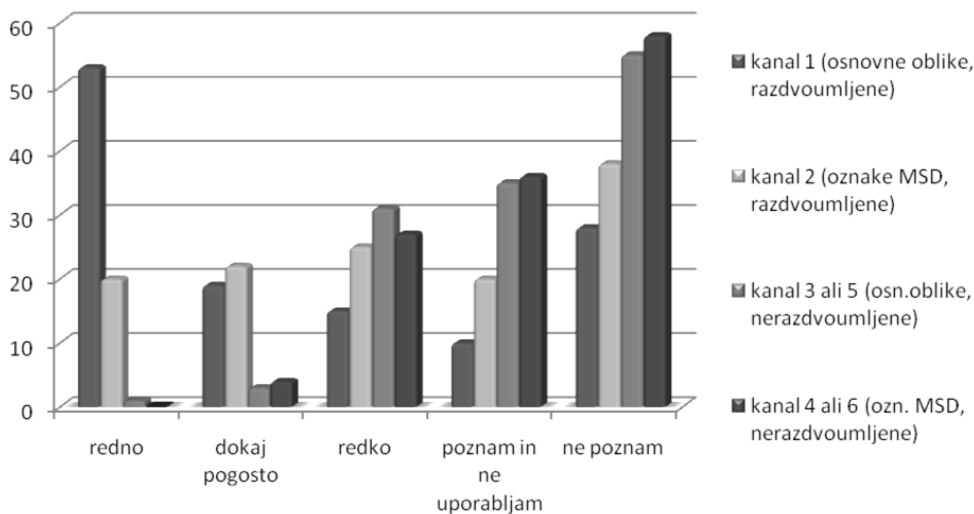
V naslednjem koraku so vprašani ocenjevali, kako pogosto uporabljajo iskanje s pomočjo kanalov, tj. iskanje na osnovi besedne leme oz. s pomočjo oblikoskladenjskih oznak. Prvi ter drugi kanal prinašata razdvoumljene

korpusne oznake, ostali pa nerazdvoumljene (Arhar, Gorjanc 2007: 103).

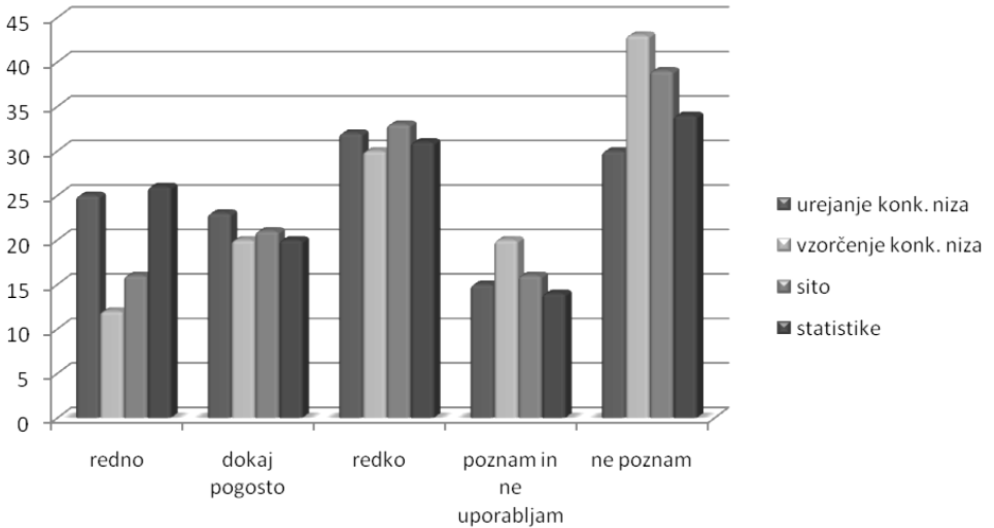
Kot je razvidno iz Grafa 5, vprašani dokaj pogosto uporabljajo prvi, nekoliko manj drugi kanal, ostale štiri kanale pa uporabljajo zelo redko. Po pričakovanjih velik delež vprašanih uporabnikov ne pozna tretjega oz. petega kanala (44,0 %) ter četrtega oz. šestega kanala (46,4 %), preseneča pa visok delež vprašanih, ki ne pozna drugega (30,4 %) oz. celo prvega kanala (22,4 %).

Nazadnje Graf 6 prikazuje še rezultate ocenjevanja, kako pogosto vprašani uporabljajo metode nadaljnje obdelave konkordančnega niza, ki so na voljo v Konkordančniku ASP32, tj. urejanje konkordančnega niza, vzorčenje konkordančnega niza, sito ter statistike za izdelavo seznama kolokatorjev.

Rezultati kažejo, da vprašani dokaj pogosto uporabljajo urejanje konkordančnega niza ter statistike za iskanje kolokatorjev, redkeje pa sito ter vzorčenje konkordančnega niza. Ponovno se izkazuje zelo visok delež vprašanih, ki navedenih funkcij ne pozna – vzorčenja ne pozna 34,4 % vprašanih, sita 31,2 %, statistik 27,2 % ter urejanja niza 24,0 % vprašanih.



Graf 5: Pogostnost rabe kanalov za iskanje po korpusu.



Graf 6: Pogostnost rabe funkcij za nadaljnjo obdelavo konkordančnega niza.

4 Sklep

Odgovori na vprašanja odprtega tipa, ki so na tem mestu puščeni ob strani, pričajo o načelni zadovoljnosti vprašanih uporabnikov s funkcijami, ki jih redno uporabljajo, veliko pa je tudi idej za izboljšave vmesnika v smer večje prijaznosti. Odgovori na vprašanja odprtega tipa bodo v nadaljevanju kategorizirani glede na vsebino predloga oz. komentarja ter v ustrezno zgoščeni obliki zajeti v končno evalvacijsko poročilo.

Uporabniška izkušnja se bo upoštevala pri načrtovanju korpusnih posodobitev, med katerimi je glede na trenutne rezultate predvidena predvsem (I) poenostavitev iskalnih možnosti, (II) priprava preglednejše in dostopnejše pomoči, (III) odstranitev funkcij, ki jih uporabniki ne uporabljajo, ter (IV) nadgradnja pogosto uporabljenih funkcij. Kot eden glavnih ciljev se kaže predvsem zasnova konkordančnika s preglednim ter intuitivnim vmesnikom, ki bo olajšal – trenutno še dokaj zahtevno – delo z referenčnim korpusom za slovenščino.

Literatura in spletni viri

- ARHAR, Špela, 2007: *Kaj početi z referenčnim korpusom FidaPLUS*. Ljubljana: Univerza v Ljubljani, Filozofska fakulteta (Elektronski vir na www.fidaplus.net).
- ARHAR, Špela, GORJANC, Vojko, 2007: Korpus FidaPLUS: Nova generacija slovenskega referenčnega korpusa. *Jezik in slovnost* 52/2. 95–110.
- Free Online Surveys: <<http://freeonlinesurveys.com/>>. (Dostop 25. 8. 2009.)
- GANTAR, Polona, 2007: *Stalne besedne zveze v slovenščini: korpusni pristop*. Ljubljana: Založba ZRC, ZRC SAZU.
- GORJANC, Vojko, 2005: *Uvod v korpusno jezikoslovje*. Domžale: Izolit.
- GORJANC, Vojko, VINTAR, Špela, 2000: Iskanja po Korpusu slovenskega jezika FIDA. Erjavec, Tomaž, Žganec Gros, Jerneja (ur.): *Jezikovne tehnologije 2*. Ljubljana: IJS. 20–26.
- KOSEM, Iztok, 2008: *User-friendly corpus tools for language teaching and learning*. Garcia-Garcia, A. (ur.): *Proceedings of the 8th Teaching and Language Corpora Conference*. Lizbona: Associação de Estudos e de Investigação Científica do ISLA-Lisboa. 183–192.
- LOGAR, Nataša, ARHAR, Špela, 2008: *Kaj početi s korpusom strokovnih besedil KoRP*. Ljubljana: Fakulteta za družbene vede.
- Sporazumevanje v slovenskem jeziku: <<http://www.slovenscina.eu>>. (Dostop 25. 8. 2009.)
- Wordsketches: <<http://www.sketchengine.co.uk/>>. (Dostop 25. 8. 2009.)
- Wordsmith: <<http://www.lexically.net/wordsmith/>>. (Dostop 25. 8. 2009.)

Korpusi

Fida: <<http://www.fida.net/slo/index.html>>. (Dostop 25. 8. 2009.)

FidaPLUS: <<http://www.fidaplus.net>>. (Dostop 25. 8. 2009.)

KoRP: <<http://www.korp.fdv.uni-lj.si/>>. (Dostop 25. 8. 2009.)

Nova beseda: <http://bos.zrc-sazu.si/s_beseda.html>. (Dostop 25. 8. 2009.)