

Mestna občina Celje
Komisija Mladi za Celje

**BESEDILNI KORPUSI
SLOVENSKEGA JEZIKA**
RAZISKOVALNA NALOGA

Avtorji:

Luka Dimitrijević

Filip Burnik

Urh Burnik

Mentor:

Črt Močivnik, prof.

Celje, marec 2014



I. osnovna šola Celje

BESEDILNI KORPUSI SLOVENSKEGA JEZIKA

RAZISKOVALNA NALOGA

Avtorji:

Luka Dimitrijević, 9. b

Filip Burnik, 9. b

Urh Burnik, 9. b

Mentor:

Črt Močivnik, prof.

Mestna občina Celje, Mladi za Celje

Celje, marec 2014

POVZETEK

Namen raziskovalne naloge je predstaviti elektronske zbirke besedil slovenskega jezika. Korpusi dajejo celovito podobo jezika in so v današnjem času izhodišče/merilo za sestavljanje slovnice in slovarjev.

Raziskovalna naloga prinaša pregled besedilnih oz. jezikovnih korpusov (Gigafida, KRES, GOS, FidaPLUS, Nova beseda in Šolar). Ukvarjamo se z vprašanjem, koliko so korpusi poznani med učenci in odraslimi, kakšna je njihova vloga v vzgojno-izobraževalnem procesu. Korpuse smo med sabo tudi primerjali in ugotovili, kaj imajo skupnega oz. v čem se razlikujejo. Namen korpusov je tudi ugotavljanje »živosti jezika«, zato nas je zanimalo, katere novejšje besede so npr. v korpusih in jih v slovarjih trenutno še ni.

Z raziskovanjem korpusov smo opozorili na njihov pomen in jih poskušali presoјati in primerjati tudi na podlagi njihovih virov (leposlovje, publicistika, parlamentarni govori idr.). Pregledno smo predstavili tudi njihovo delovanje oz. praktično uporabo v vsakdanjem življenju.

Ključne besede: korpusno jezikoslovje, besedilni korpusi, živost slovenskega jezika

KAZALO VSEBINE

1	Uvod	1
1.1	Opis raziskovalnega problema in namen	1
1.2	Hipoteze	2
1.3	Metodologija, opis raziskovalnih metod	2
2	Teoretični del.....	3
2.1	Teoretična podlaga za raziskavo o besedilnih korpusih	3
2.2	Vrste korpusov.....	3
2.2.1	Gigafida.....	4
2.2.2	KRES	5
2.2.3	GOS	6
2.2.4	FidaPLUS	7
2.2.5	Nova beseda	8
2.2.6	Šolar	9
2.3	Zunanja podoba korpusov	10
3	Primerjava in uporaba korpusov – razprava in ugotovitve	13
4	Rezultati ankete.....	15
5	Primer uporabe korpusov in jezikovnih priročnikov	17
6	Sklepne ugotovitve	18
7	Zaključek.....	20
	Viri in literatura	21
	Priloga.....	22

KAZALO TABEL

<i>Tabela 1: Zajem posnetkov za GOS po letih na dan 6. 11. 2012 (vir: www.korpus-gos.net).....</i>	<i>6</i>
<i>Tabela 2: Sestava korpusa Nova beseda (2000–2008) (vir: bos.zrc-sazu.si)</i>	<i>8</i>
<i>Tabela 3: Pregled korpusov po številu besed in časovnih intervalov.....</i>	<i>13</i>
<i>Tabela 4: Rezultati anket (N = 61)</i>	<i>15</i>
<i>Tabela 5: Rezultati anket (N = 61)</i>	<i>16</i>
<i>Tabela 6: Pregled pojavitve besede »tajkun« v besedilnih korpusih</i>	<i>17</i>

KAZALO GRAFOV

<i>Graf 1: Struktura sestave korpusa Gigafida (vir: www.gigafida.net).....</i>	<i>4</i>
<i>Graf 2: Struktura sestave korpusa KRES (vir: www.korpus-kres.net)</i>	<i>5</i>
<i>Graf 3: Gradivo korpusa FidaPLUS glede na tip (vir: www.fidaplus.net).....</i>	<i>7</i>
<i>Graf 4: Večina gradiva so eseji oz. spisi (65 %) (vir: www.korpus-solar.net)</i>	<i>9</i>

1 Uvod

1.1 Opis raziskovalnega problema in namen

Pred opredelitvijo raziskovalnega problema želimo definirati pojem **računalniško jezikoslovje**. S tem pojmom poimenujemo raziskovalno smer, ki se ukvarja z raziskovanjem jezika s pomočjo statističnih metod obdelave jezikovnega gradiva ter razvojem računalniških orodij, jezikovnih virov ipd. (Arhar Holdt 2011: 16).

Korpusno jezikoslovje prinaša povsem nov pogled na jezik. **Besedilni korpusi** so obsežne zbirke besedil v določenem jeziku, zajete v določenem obdobju iz množičnih medijev. Sestava korpusov je različna – odvisna od virov, iz katerih izhajajo uporabljeni podatki. Slednji so zajeti iz časopisnega in revijalnega tiska, knjižne produkcije, stripov, spleta, reklamnih besedil, prepisov parlamentarnih razprav ipd.

Namen raziskovalne naloge je predstaviti elektronske zbirke besedil slovenskega jezika. Korpusi dajejo celovito podobo jezika in so v današnjem času izhodišče/merilo za sestavljanje slovníc in slovarjev.

Raziskovalna naloga prinaša pregled naslednjih besedilnih oz. jezikovnih korpusov:

- a) **Gigafida** (1,2 milijarde besed) – korpus pisne slovenščine
- b) **KRES** (100 milijonov besed) – podkorpus Gigafide
- c) **GOS** (milijon besed) – korpus govorne slovenščine
- d) **FidaPLUS** (620 milijonov besed) – predhodnica Gigafide
- e) **Nova beseda** (318 milijonov besed) – namenjen za slovaropisne potrebe
- f) **Šolar** (milijon besed) – korpus pisnih izdelkov osnovnošolcev in srednješolcev

Ukvarjamo se z vprašanjem, koliko so korpusi poznani med učenci in odraslimi, kakšna je njihova vloga v vzgojno-izobraževalnem procesu itd. Korpuse smo med sabo tudi primerjali in ugotavljali, kaj imajo skupnega oz. v čem se razlikujejo. Namen korpusov je tudi ugotavljanje »živosti jezika«, zato nas je zanimalo, katere novejšje besede so npr. v korpusih in jih v slovarjih trenutno še ni.

Z raziskovanjem korpusov smo opozorili na njihov pomen in jih presojali in primerjali tudi na podlagi njihovih virov (leposlovje, publicistika, parlamentarni govori idr.). Pregledno bomo predstavili tudi njihovo delovanje oz. praktično uporabo v vsakdanjem življenju.

1.2 Hipoteze

Na osnovi splošnega pregleda besedilnih korpusov slovenskega jezika, ki so dostopni na spletu, in na podlagi našega poznavanja/uporabe, smo si zastavili naslednje hipoteze:

1. *Besedilni korpusi slovenskega jezika so ljudem nepoznani.*
2. *Uporaba korpusov v vsakdanjem življenju je redka.*
3. *Večina ljudi – uporabnikov spleta ne zna uporabljati besedilnih korpusov.*
4. *Glede na uporabo in sestavo so si korpusi med sabo različni.*
5. *V korpusih je delež novejših besed mnogo večji kot v aktualnem SSKJ.*

Z uporabo metodologije oz. raziskovalnih metod, navedenih v nadaljevanju, smo želeli ugotoviti, ali navedene hipoteze držijo.

1.3 Metodologija, opis raziskovalnih metod

Naše raziskovalno delo je temeljilo na pregledu besedilnih korpusov slovenskega jezika in anketi, ki smo jo izvedli v obdobju december 2013–januar 2014. Gradivo smo skrbno pregledali in z mentorjevo pomočjo določili, na kaj moramo biti pozorni. Osredotočeni smo bili zlasti na sestavo in dostopnost korpusov ter njihovo uporabo pri vsakdanjem delu.

Preučili smo več korpusov, najobsežnejši in referenčni korpus za slovenščino je **Gigafida**¹ (1,2 milijarde besed), v katerega je vključen korpus FidaPLUS in KRES. Pri našem raziskovanju smo uporabili zlasti:

- a) metodo dela z viri in literaturo,
- b) metodo obdelave podatkov.

Naše raziskovanje je temeljilo zlasti na podlagi empiričnega preverjanja stanja. Gradivo, ki je dostopno na različnih spletnih straneh, smo pregledovali sistematično – v štirimesečnem obdobju smo pregledali šest besedilnih korpusov in knjižno gradivo, ki smo ga pridobili v Osrednji knjižnici Celje. Vključili smo tudi anketo, s katero smo preverili poznavanje besedilnih korpusov v različnih starostnih skupinah ljudi. Raziskovalna naloga temelji na načelih deskriptivne ali opisne metode, saj smo preučevali gradivo na ravni opisovanja dejstev.

¹ Velja za *referenčni korpus*, v primerjavi z ostalimi korpusi ga odlikujeta predvsem velik obseg ter natančno domišljena metoda gradnje, ki se odraža v uravnoteženih razmerjih med zajetimi besedili (Gorjanc 2005: 8).

2 Teoretični del

2.1 Teoretična podlaga za raziskavo o besedilnih korpusih

Besedilni korpusi so elektronske zbirke avtentičnih besedil, nastale po vnaprej določenih merilih in z določenim ciljem ter opremljena z orodji, ki omogočajo večplastno iskanje jezikovnih podatkov (Krek 2005: 225).

Med označbe, ki jih vsebujejo nekateri korpusi, spadajo **leme** – osnovne oblike besed, oblikoskladenjske označbe, skladenjske označbe in še nekatere druge označbe, ki se uporabljajo tudi za namene prepoznavanje govora in računalniškega prevajanja.

2.2 Vrste korpusov

Glede na jezik besedil delimo korpuse na **enojezične (specializirane)** in na **večjezične**. Za potrebe našega raziskovanja predstavljamo samo enojezične.

Enojezični splošni jezikovni korpusi so:

- a) korpus pisnega jezika Gigafida,
- b) referenčni korpus KRES,
- c) korpus pisnega jezika Nova beseda,
- d) učni korpus ssj500k,
- e) prosto dostopen korpus ccGigafida,
- f) prosto dostopen korpus ccKRES,
- g) korpus starejše slovenščine IMP,
- h) skladenjsko označen korpus SDT,
- i) govorni korpus GOS.

Enojezični specializirani jezikovni korpusi so:

- a) Šolar: korpus pisnih šolskih izdelkov,
- b) iKorpus: korpus besedil s področja informatike in računalništva,
- c) KORP: korpus besedil odnosov z javnostmi.

2.2.1 Gigafida²

Gigafida je najboljšežnejši slovenski korpus (1.187.002.502 besed).

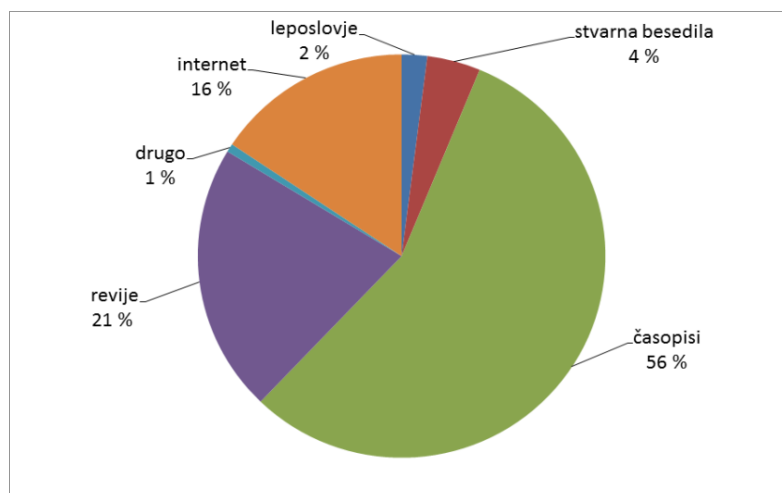
Narejen je bil zato, da bi raziskovali sodoben slovenski jezik na več ravneh. Namenjen je znanstvenikom in raziskovalcem v jezikoslovju, tudi učiteljem slovenščine v osnovnih in srednjih šolah, njihovim učencem, tistim, ki se slovenščino učijo kot tuj jezik, pa tudi vsem, ki gredo namesto na knjižno polico iskat odgovor na svojo jezikovno zadrego na svetovni splet.

Gigafida vsebuje besedila, ki so izšla med letoma 1990 in 2011. Gre za tiskana besedila (besedila, ki so leposlovna ali s stvarno vsebino v obliki revij ali časopisov) in za besedila, pridobljena s spletnih strani (novičarski portali, predstavitvene strani večjih slovenskih podjetij in pomembnejših državnih, pedagoških, raziskovalnih, kulturnih ipd. ustanov).

Gigafida vsebuje skoraj ves korpus **FidaPLUS** in **KRES**. Ta vsebuje tudi druge vrste informacij. Vsak posamezni dokument, ki ga vsebuje, zajema informacijo o viru, letu, nastanku, vrsti besedila, naslovu in avtorju, če je ta znan.

Gigafida je jezikoslovno označen korpus, kar pomeni, da sta k prav vsaki besedi v korpusu pripisana še dva podatka. Prvi je osnovna oblika besede ali lema, drugi je oblikoskladenjska oznaka.

Lastnik **Gigafide** je Ministrstvo za izobraževanje, znanost, kulturo in šport.



Graf 1: Struktura sestave korpusa Gigafida (vir: www.gigafida.net)

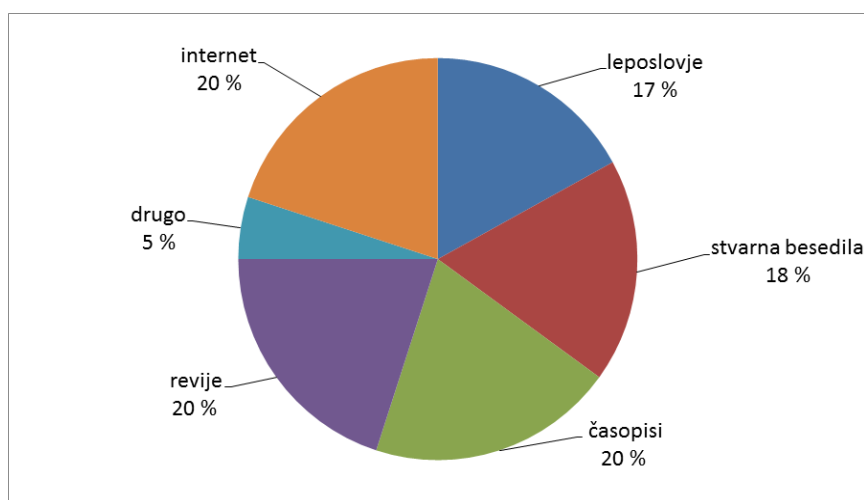
² <<http://www.gigafida.net/>> (Dostop 10. 1. 2014)

2.2.2 KRES³

KRES je podkorpus korpusa **Gigafida**. **KRES** je nastal po pobudi projekta *Sporazumevanje v Slovenskem jeziku 2008 do 2012* (vsebuje 99.831.145 besed). Poleg besed vsebuje tudi druge vrste informacij, saj je jezikoslovno označen korpus. Vsak posamezni dokument, ki je v korpusu **KRES**, skupaj jih je 21.456, vsebuje informacijo o viru (npr. *Mladina*, *Delo*, *Dnevnik*), letu nastanka, vrsti besedila (npr. leposlovje, revija), naslovu in avtorju, če je ta znan. **KRES** je namenjen različnim uporabnikom, kot so znanstveniki in raziskovalci, ki jih zanima sodobni slovenski jezik ter učiteljem v osnovnih in srednjih šolah, njihovim učencem, tistim, ki se slovenščino učijo kot drugi ali tuji jezik, pa tudi vsem drugim, ki se znajdejo pred jezikovno dilemo, povezano s sodobno slovenščino.

KRES je namenjen kakršnim koli jezikoslovnim poizvedovanjem, ki imajo težnjo po merodajnosti, kolikor pač ta izhaja iz vzorca korpusa, ki ima vnaprej premišljeno, znano in utemeljeno ter uravnoteženo zgradbo.

Rezultati poizvedb po tem korpusu so zaradi drugačne besedilnovrstne sestave seveda drugačni kot rezultati iskanja po **Gigafidi**. Za korpus, ki predstavlja celovito podobo izbranega jezika, je pomembno, da so obsežni in **besedilnovrstno pestri**. Že omenjeni korpus **Gigafida** je t. i. *referenčni korpus*, težko pa bi mu pripisali uravnoteženost, saj je v njem 77 % besed iz periodike (časopisi, revije) in npr. le dobrih 6 % besed iz knjig (leposlovje, stvarna besedila). Kot Gigafidin uravnoteženi *podkorpus* so zato ustvarjalci/sodelavci predhodno načrtovali 100-milijonski **KRES**, katerega sestava je razvidna iz grafa:



Graf 2: Struktura sestave korpusa KRES (vir: www.korpus-kres.net)

³ <<http://www.korpus-kres.net/>> (Dostop 25. 1. 2014)

2.2.3 GOS⁴

GOS je korpus govorne slovenščine (okrog 120 ur pogovorov v najrazličnejših situacijah). Narejen je bil zato, ker vemo zelo malo o tem, kako govorimo v vsakdanjem življenju in z namenom, da bi lažje izvedeli, kakšno slovenščino sploh govorimo. Namenjen je tistim, ki želijo raziskovati govorno slovenščino bodisi z jezikoslovnega bodisi s katerega drugega, recimo sociološkega in jezikovnotehnološkega vidika.

Korpus **GOS** je sestavljen tako, da so v njem zbrani posnetki govorov, tako da bi korpus čim natančneje prikazal današnjo slovenščino v najpogostejših vsakdanjih situacijah. Poseben podsklop korpusa **GOS** (10 %) predstavljajo šolske razprave. V korpusu **GOS** so pri snemalnem gradivu upoštevani tudi kriteriji reprezentativnosti govorcev, zato je v tistem delu, ki zajema posnetke zasebnih pogovorov, zajet ustrezen delež govorcev iz različnih regij, obeh spolov, različnih starosti ter različnih izobrazbeni ravni. Korpus **GOS** zajema posnetke iz let 2008–2010. Korpus **GOS** kot **baza podatkov** zajema:

a) posnetke govora,

b) pogovorni zapis govora (*»zapiši, kot slišiš«*),

c) standardizirani zapis govora (*»zapiši, kot pišemo«*),

č) standardiziranemu zapisu avtomatsko dodane podatke o osnovni obliki in oblikoslovnih lastnostih besed,

d) podatke o situaciji, kjer je bil narejen posnetek,

e) podatke o govorniku.

Lastnik korpusa **GOS** je Ministrstvo za izobraževanje, znanost, kulturo in šport.

Leto snemanja	ŠT. BESED	%
2004	5.754	0,56
2008	164.405	15,88
2009	574.252	55,48
2010	290.690	28,08
Skupaj:	1.035.101	100 %

Tabela 1: Zajem posnetkov za GOS po letih na dan 6. 11. 2012 (vir: www.korpus-gos.net)

⁴ <<http://www.korpus-gos.net/>> (Dostop 29. 1. 2014)

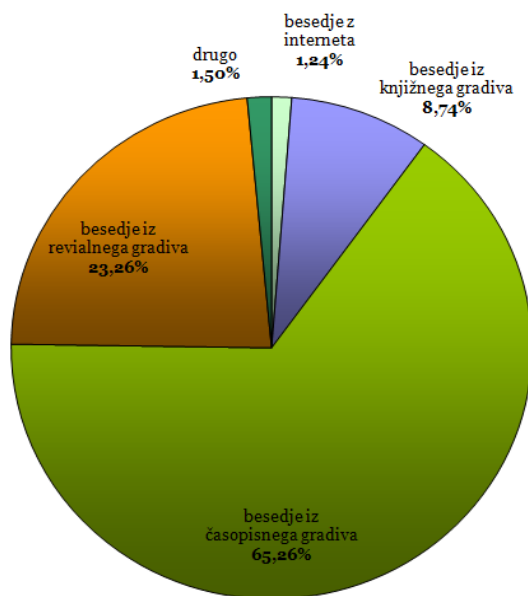
2.2.4 FidaPLUS⁵

Korpus **FidaPLUS** je predhodnik korpusa **Gigafida**. Besedilni korpus **FidaPLUS** je rezultat projekta *Jezikovni viri za slovenščino*. Projekt pomeni nadgradnjo in nadaljevanje projekta Fida, ki je potekal med leti 1997 in 2000. Korpus **FidaPLUS** je referenčni korpus slovenskega jezika. To pomeni, da je obsežna elektronska zbirka, ki obsega velik vzorčni delež besedil slovenskega jezika.

Zgrajen je z namenom, da bi omogočil temeljit vpogled v jezik na najrazličnejših ravneh in področjih. Zato je pomemben vir za uporabno in teoretično jezikoslovje, npr. slovaropisje v vseh oblikah (eno- in večjezikovni slovarji), poučevanje jezika (učbeniki), jezikovne tehnologije (črkovalniki) ter za druge družboslovne in humanistične vede, npr. literarno vedo, psihologijo in sociologijo.

Korpus **FidaPLUS** vsebuje besedila najrazličnejših zvrsti iz večine slovenskih dnevnih časopisov, mnogih revij ter knjižnih publikacij različnih založb (lit. besedila, učbeniki, stvarna lit.). Poleg tega korpus zajema tudi besedila z interneta, prepise parlamentarnih govorov, tako imenovan besedilni drobiž (reklamna besedila, plačilni listki, računi).

Korpus ponuja možnosti natančnega raziskovanja slovenskega jezika v besedilih, objavljenih od leta 1990 do 2006.



Graf 3: Gradivo korpusa FidaPLUS glede na tip (vir: www.fidaplus.net)

⁵ <<http://www.fidaplus.net/>> (Dostop 6. 2. 2014)

2.2.5 Nova beseda⁶

Nova beseda (162 milijonov besed od leta 1998 do 2005) je besedilni korpus, ki je namenjen predvsem za slovaropisne potrebe *Inštituta za slovenski jezik Frana Ramovša ZRC SAZU*, obenem pa tudi vsem drugim, ki se ukvarjajo z izobraževanjem in raziskovanjem slovenskega jezika. Leta 1999 je bil korpus internetno oblikovan. Spomladi leta 2000 je bilo korpusu dodano gradivo časopisa **DELO**, kar je povečalo obseg na 28 milijonov besed. Poleti 2000 se je obseg korpusa z dodatki novih besedil povečal na 48 milijonov besed. V letu 2005 se obseg korpusa poveča na 162 milijonov besed. Maja 2000 je bil korpus prenesen s strežnika **Filozofske fakultete Univerze v Ljubljani** na strežnik **Laboratorija za korpus slovenskega jezika Inštituta za slovenski jezik Frana Ramovša ZRC SAZU**. Nova beseda je po našem mnenju zapletena za uporabo, saj smo imeli pri raziskovanju veliko težav zaradi nepreglednosti.

Lastnik korpusa je **Laboratorij za korpus slovenskega jezika Inštituta za slovenski jezik Frana Ramovša ZRC SAZU**.

Obseg in zvrstnost korpusa se je z leti povečevala, leta 2005 na 162 milijonov besed. Velika večina besedil je iz zadnjih desetih let (razen leposlovja), danes obsega **318 milijonov besed**:

Posamezna področja korpusa Nova beseda	Število besed (v milijonih)
slovensko leposlovje	12
slovensko polleposlovje	2
znanstvene in tehnične publikacija	3
dnevnik DELO	213
zapisi sej DZ RS, Odprti kop TVS	42
periodika – tedniki, mesečniki	44
zakoni Republike Slovenije	2

Tabela 2: Sestava korpusa Nova beseda (2000–2008) (vir: bos.zrc-sazu.si)

⁶ <http://bos.zrc-sazu.si/s_beseda3.html> (Dostop 6. 2. 2014)

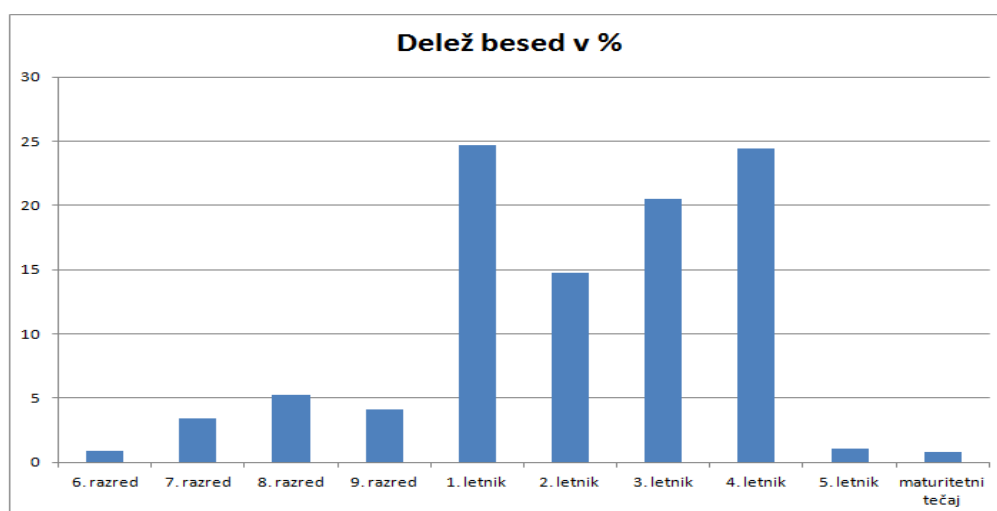
2.2.6 Šolar⁷

Je korpus šolskih pisnih izdelkov (967.477). Šolar je korpus besedil, ki so jih učenci osnovnih (19 %) in srednjih šol (81 %, od tega gimnazija 43 %) samostojno tvorili pri pouku. Nastal je v okviru projekta *Sporazumevanje v slovenskem jeziku v letih 2009 in 2010*. Gre za prvi tovrsten korpus v Sloveniji. Od večine drugih korpusov se razlikuje po tem, da:

a) besedila niso nastala s pobudo projekta, ampak predstavljajo dejansko šolsko produkcijo učencev in dijakov;

b) so jezikovni popravki, ki so označeni v korpusu, realni – naredili so jih učitelji, ne raziskovalci.

Zgrajen je bil z namenom, da bi razširili vpogled v pisanje šolajoče se mladine, populacije, katere jezikovna produkcija je bila doslej s korpusnim pristopom še neraziskana. Korpus Šolar predstavlja velik korak k zapolnitvi tega manka. Korpus je zanimiv zaradi vključenih besedil, ki nudijo vpogled v pisno produkcijo učencev, po drugi strani pa tudi zaradi učiteljskih popravkov, ki nudijo informacije o tem, kaj učitelji največ popravljajo. Šolar je delno regijsko uravnotežen, saj je delež besedil, ki so nastala v šolah na jugozahodnem delu Slovenije, in delež besedil iz šol severozahodnega dela Slovenije, v razmerju 60 : 40 (Kosem, Rozman, Stritar 2012: 29). Šolar je namenjen raziskovanju pisne zmožnosti šolajočih se populacij, hkrati pa je namenjen jezikoslovcem in učiteljem za raziskovanje jezikovne zmožnosti v času šolanja ter pripravo učnih vsebin in gradiv. **Korpus je edinstven ne samo v slovenskem, ampak tudi v evropskem oz. svetovnem merilu.**



Graf 4: Večina gradiva so eseji oz. spisi (65 %) (vir: www.korpus-solar.net)

⁷ <http://www.korpus-solar.net/run.cgi/first_form?corpname=fidapplus-solar> (Dostop 10. 2. 2014)

2.3 Zunanja podoba korpusov

Prikaz zunanje podobe posameznih korpusov pri uporabi iskanja besede »*tajkun*« (splošno, tj. osnovno iskanje, 22. 2. 2014).

a) Gigafida (16.365 zadetkov)

The screenshot shows the Gigafida search results for the word "tajkun". The search bar at the top contains "tajkun" and the "Najdi" button is highlighted. Below the search bar, there are links for "Uporabljaš enostavno iskanje" and "Napredno iskanje". The results are displayed in a table with columns for "Osnovne oblike", "Vrsta besedila", "Vir", and "Leto". The table shows various results, including "Internet (9.668)", "Časopisi (6.224)", "Revije (406)", "Leposlovje (5)", "Škarna besedila (2)", "Več", "Vir", "24ur.com (4.310)", "siol.net (2.921)", "druzo (1.916)", "Dnevnik (1.552)", "Delo (1.410)", "Več", "Leto", "2010 (6.522)", "2008 (3.474)", "2011 (3.263)", "2009 (1.623)". The search results are displayed in a table with columns for "Osnovne oblike", "Vrsta besedila", "Vir", and "Leto". The table shows various results, including "Internet (9.668)", "Časopisi (6.224)", "Revije (406)", "Leposlovje (5)", "Škarna besedila (2)", "Več", "Vir", "24ur.com (4.310)", "siol.net (2.921)", "druzo (1.916)", "Dnevnik (1.552)", "Delo (1.410)", "Več", "Leto", "2010 (6.522)", "2008 (3.474)", "2011 (3.263)", "2009 (1.623)".

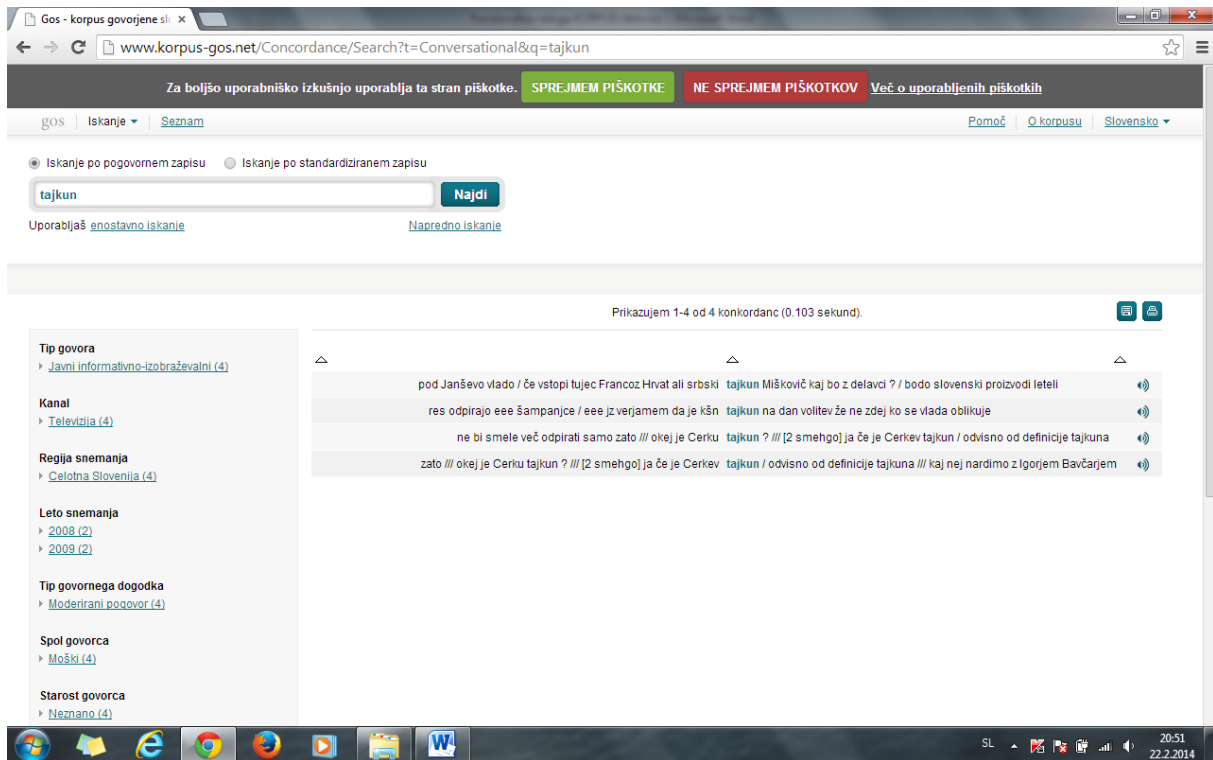
Vir: <http://www.gigafida.net/>

b) KRES (1.003 zadetki)

The screenshot shows the KRES search results for the word "tajkun". The search bar at the top contains "tajkun" and the "Najdi" button is highlighted. Below the search bar, there are links for "Uporabljaš enostavno iskanje" and "Napredno iskanje". The results are displayed in a table with columns for "Osnovne oblike", "Vrsta besedila", "Vir", and "Leto". The table shows various results, including "Internet (793)", "Časopisi (178)", "Revije (29)", "Leposlovje (3)", "Več", "Vir", "24ur.com (567)", "siol.net (141)", "druzo (85)", "Delo (58)", "druzo.si (47)", "Več", "Leto", "2010 (509)", "2011 (287)", "2008 (113)", "2009 (30)", "2007 (28)". The search results are displayed in a table with columns for "Osnovne oblike", "Vrsta besedila", "Vir", and "Leto". The table shows various results, including "Internet (793)", "Časopisi (178)", "Revije (29)", "Leposlovje (3)", "Več", "Vir", "24ur.com (567)", "siol.net (141)", "druzo (85)", "Delo (58)", "druzo.si (47)", "Več", "Leto", "2010 (509)", "2011 (287)", "2008 (113)", "2009 (30)", "2007 (28)".

Vir: <http://www.korpus-kres.net/>

c) GOS (4 zadetki)



Gos - korpus govornih slovarov

www.korpus-gos.net/Concordance/Search?t=Conversational&q=tajkun

Za boljše uporabniško izkušnjo uporabljajte stran piškotke. [SPREJMEM PIŠKOTKE](#) [NE SPREJMEM PIŠKOTKOV](#) [Več o uporabljenih piškotkih](#)

gos | Iskanje | Seznam

Iskanje po pogovornem zapisu | Iskanje po standardiziranem zapisu

tajkun [Najdi](#)

Uporabljaš [enostavno iskanje](#) [Napredno iskanje](#)

Prikazujem 1-4 od 4 konkordanc (0.103 sekund).

Tip govora
Javni informativno-izobraževalni (4)

Kanal
Televizija (4)

Regija snemanja
Celotna Slovenija (4)

Leto snemanja
2008 (2)
2009 (2)

Tip govornega dogodka
Moderirani pogovor (4)

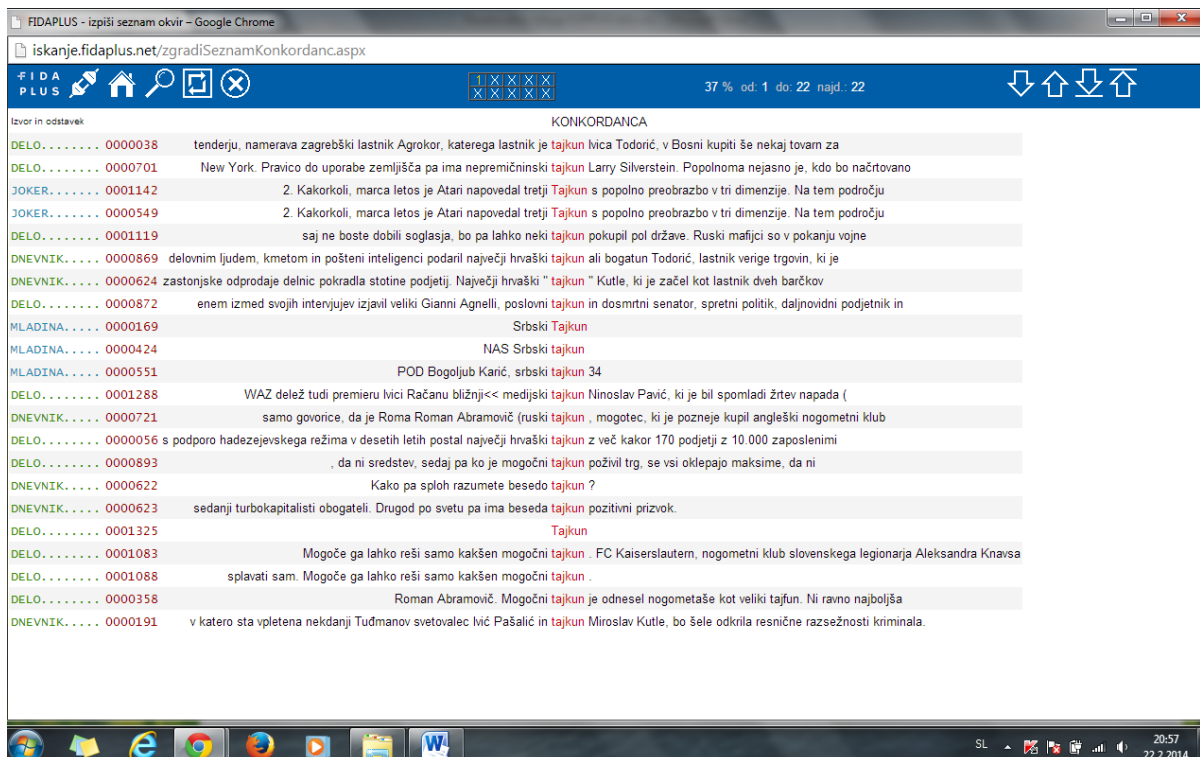
Spol govorca
Moški (4)

Starost govorca
Neznano (4)

pod Janševo vlado / če vstopi tujec Francoz Hrvat ali srbski **tajkun** Miškovič kaj bo z delavci ? / bodo slovenski proizvajalci letali
res odpirajo eee šampanjce / eee jz verjamem da je kšn **tajkun** na dan volitev že ne zdej ko se vlada oblikuje
ne bi smele več odpirati samo zato /// okej je Cerku **tajkun** ? /// [2 smehgo] ja če je Cerkev tajkun / odvisno od definicije tajkuna
zato /// okej je Cerku tajkun ? /// [2 smehgo] ja če je Cerkev **tajkun** / odvisno od definicije tajkuna /// kaj ne nardimo z Igorjem Bavčarjem

Vir: <http://www.korpus-gos.net/>

d) FidaPLUS (88 zadetkov)



FIDAPLUS - izpiši seznam okvir - Google Chrome

iskanje.fidaplus.net/zgradiSeznamKonkordanc.aspx

FIDA PLUS

37 % od: 1 do: 22 najd.: 22

KONKORDANCA

DELO. 0000038 tenderju, namerava zagrebški lastnik Agrokor, katerega lastnik je **tajkun** Ilica Todorčič, v Bosni kupiti še nekaj tovarn za
DELO. 0000701 New York. Pravico do uporabe zemljišča pa ima nepremičninski **tajkun** Larry Silverstein. Popolnoma nejasno je, kdo bo načrtovano
JOKER. 0001142 2. Kakorkoli, marca letos je Atari napovedal tretji **Tajkun** s popolno preobrazbo v tri dimenzije. Na tem področju
JOKER. 0000549 2. Kakorkoli, marca letos je Atari napovedal tretji **Tajkun** s popolno preobrazbo v tri dimenzije. Na tem področju
DELO. 0001119 saj ne boste dobili soglasja, bo pa lahko neki **tajkun** pokupil pol države. Ruski mafijci so v pokanju vojne
DNEVNIK. 0000869 delovnim ljudem, kmetom in poštini inteligenci podaril največji hrvaški **tajkun** ali bogatun Todorčič, lastnik verige trgovin, ki je
DNEVNIK. 0000624 zastojne odprodaje delnic pokradla stotine podjetij. Največji hrvaški " **tajkun** " Kutle, ki je začel kot lastnik dveh barčkov
DELO. 0000872 enem izmed svojih intervjujev izjavil veliki Gianni Agnelli, poslovni **tajkun** in dosmrtni senator, spretni politik, daljnovidni podjetnik in
MLADINA. 0000169 Srbski **Tajkun**
MLADINA. 0000424 NAS Srbski **tajkun**
MLADINA. 0000551 POD Bogoljub Karič, srbski **tajkun** 34
DELO. 0001288 WAZ delež tudi premieru Ilici Račanu bližnji<< medijski **tajkun** Ninoslav Pavič, ki je bil spomladi žrtev napada (
DNEVNIK. 0000721 samo govornice, da je Roma Roman Abramovič (ruski **tajkun** , mogotec, ki je pozneje kupil angleški nogometni klub
DELO. 0000056 s podporo hadezejskega režima v desetih letih postal največji hrvaški **tajkun** z več kakor 170 podjetij z 10.000 zaposlenimi
DELO. 0000893 , da ni sredstev, sedaj pa ko je mogočni **tajkun** poživil trg, se vsi oklepajo maksime, da ni
DNEVNIK. 0000622 Kako pa sploh razumete besedo **tajkun** ?
DNEVNIK. 0000623 sedanjih turbokapitalisti obogateli. Drugod po svetu pa ima beseda **tajkun** pozitivni prizvok.
DELO. 0001325 **Tajkun**
DELO. 0001083 Mogoče ga lahko reši samo kakšen mogočni **tajkun** . FC Kaiserslautern, nogometni klub slovenskega legionarja Aleksandra Knava
DELO. 0001088 splavati sam. Mogoče ga lahko reši samo kakšen mogočni **tajkun** .
DELO. 0000358 Roman Abramovič. Mogočni **tajkun** je odnesel nogometaše kot veliki tajfun. Ni ravno najboljša
DNEVNIK. 0000191 v katero sta vpletena nekdanji Tudmanov svetovalec Ilic Pašalić in **tajkun** Miroslav Kutle, bo šele odkrila resnične razsežnosti kriminala.

Vir: <http://www.fidaplus.net/>

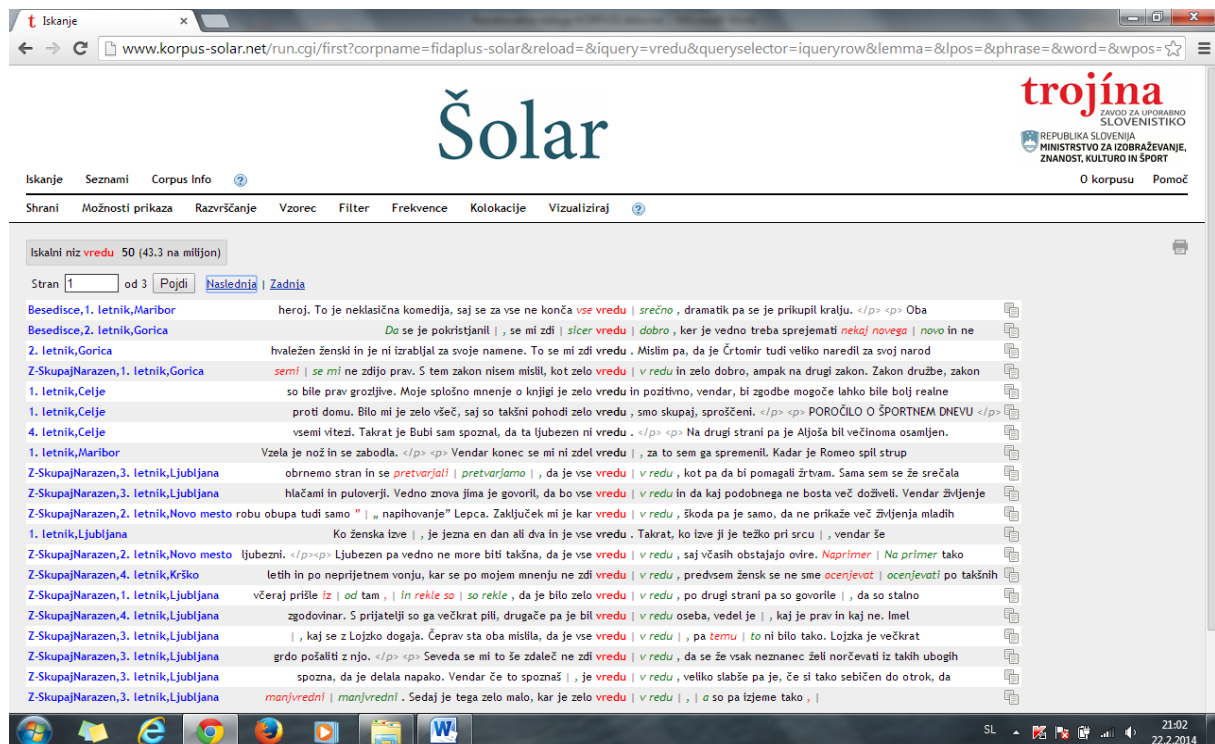
e) Nova beseda (409 zadetkov)



Vir: http://bos.zrc-sazu.si/s_beseda3.html

g) Šolar

Ker gre za korpus šolskih pisnih izdelkov osnovnošolcev in srednješolcev, navajamo primer napačnega zapisa »vred« nam. pravilnega, tj. zapisanega narazen »v redu«.



Vir: <http://www.korpus-solar.net/>

3 Primerjava in uporaba korpusov – razprava in ugotovitve

S primerjavo korpusov želimo ugotoviti, koliko so si med seboj podobni oz. katere so najbolj opazne razlike. Vsem je skupno, da nam dajejo informacije o živosti jezika oz. povedano drugače – korpusi so namenjeni preučevanju jezika vsakdanje rabe. Omogočajo številne možnosti statističnih obdelav velike količine jezikovnih podatkov, ki nam koristijo v vsakdanji rabi, tudi pri pouku jezika in književnosti.

Njihova uporaba je večinoma enostavna, poiščemo jih s pomočjo spletnih brskalnikov, si preberemo kratka navodila za uporabo in že lahko raziskujemo. Izbiramo lahko med osnovnim ali naprednim/podrobnejšim iskanjem, vnesemo želeno besedo in program nam izpiše zadetke.

Učencem nam je zelo blizu korpus **Šolar**, saj si lahko v naboru številnih pravopisnih zgledov poiščemo primere, ki nam koristijo pri pisanju spisa ali tvorbi poustvarjalnega besedila. V veliko pomoč so nam konkretni napačni in pravilni zgledi. Tako lahko že pri pouku (npr. pri analizi preverjanja pred ocenjevanjem) na konkretnih primerih pogledamo, katerim najpogostejšim napakam se lahko izognemo. Najpogosteje gre za pisanje skupaj/narazen ali pa za veliko/malo začetnico, včasih pa tudi za kakšno oblikoslovno (sklonska oblika samostalnika) oz. skladijsko dilemo (podredje in priredje).

V nadaljevanju navajamo nekaj primerjav/razlik med korpusi:

KORPUS	ŠTEVILO BESED	OBDOBJE ZBIRANJA
Gigafida	1,2 milijarde	1990–2011
KRES	100 milijonov	2008–2012
GOS	1 milijon	2008–2010
FidaPLUS	620 milijonov	1990–2006
Nova beseda	318 milijonov	2000–2008
Šolar	1 milijon	2009–2010

Tabela 3: Pregled korpusov po številu besed in časovnih intervalov

Besedilni korpusi slovenskega jezika se med sabo razlikujejo ne samo po številu besed in časovnem intervalu zbiranja, kot je prikazano v tabeli 3, ampak tudi glede na sestavo gradiva oz. virov zajemanja podatkov.

Gigafida kot najobsežnejši slovenski korpus vsebuje besede, zajete iz raznovrstnih besedil v obdobju 1990–2011. Večinoma gre za tiskana besedila, ki so bila zajeta že v predhodnici **FidaPLUS** (1990–2006) in podkorpusu **KRES** (2008–2012). Kot zanimivost naj dodamo, da se je potrebno za uporabo korpusa **FidaPLUS** registrirati oz. skleniti enoletno spletno pogodbo o uporabi besedilnega korpusa med Filozofsko fakulteto UL in končnim uporabnikom. Registracija je enostavna, uporaba in dostopnost podatkov sta brezplačni. Vsi ostali korpusi so uporabnikom prosto dostopni.

Struktura sestave **Gigafide** je sledeča: revije in časopisi obsegajo 77 % gradiva, besedila s spleta (novičarski portali, državne ustanove) obsegajo 16 %, leposlovje zgolj 2 %.

Na tem mestu je zanimiva primerjava s korpusom **Nova beseda** (318 milijonov, 2000–2008), ki ima podobno strukturo: leposlovje obsega 4 % (12 milijonov b.), dnevnik Delo pa kar 67 % vseh besed. Dodane so še transkripcije sej DZ RS in posameznih oddaj Televizije Slovenija.

Najbolj izstopata korpusa **GOS** in **Šolar**. Oba obsegata zgolj milijon besed, vendar sta za uporabnike zelo zanimiva in uporabna.

Korpus **GOS** prikazuje namreč podobo govorjene slovenščine v najpogostejših vsakdanjih situacijah. Za uporabnika/raziskovalca so zanimive **podkategorije**: *tip govora* (zasebni, javni), *kanal* (radio, TV, telefon), *regija snemanja*, *leto snemanja*, *tip govornega dogodka* (pogovor v družini, v trgovini, moderirani pogovor idr.), *spol in starost govorca*, *izobrazba in prvi jezik govorca*.

Korpus **Šolar** je za nas, učence, zelo uporaben. Gre za zbirko šolskih besedil, ki so jih tvorili učenci in dijaki, jezikovne popravke pa opravili učitelji. Ko se znajdemo pri jezikovni oz. pravopisni dilemi, lahko s preprostim vpisom pridemo do takojšnje pravilne rešitve. Korpus nam posreduje tudi naslednje podatke: *šola* (OŠ, gimnazija idr.), *predmet*, *razred*, *regija*, *vrsta besedila* (npr. spis, esej), *tip napake* (skupaj/narazen). S preprosto uporabo pridemo do že znanih splošnih ugotovitev, kot je npr. površna uporaba dvojine v primorski regiji. Zanimiv je tudi zapis besede »**ponavadi**« in »**po navadi**«. Seveda sta prisotna oba zapisa, popravek je nedosleden, priporočljivejši je zapis narazen, tj. »**po navadi**« (med besedi lahko vrinemo dodatno besedo, npr. po *stari* navadi).

4 Rezultati ankete

Spol	Število
moški	30
ženski	29
ne zaupa	2
skupaj	61

Starost	Število + spol
13–18	18 Ž, 24 M, 2 NZ
19–35	1 Ž, 0 M
36–50	3 Ž
več kot 50	6 M, 7 Ž
skupaj	61

Tabela 4: Rezultati anket (N = 61)

Stopnja izobrazbe	Število
nedokončana OŠ	11
dokončana OŠ	26
nižje poklicno	3
srednje poklicno	4
gimnazijsko, srednje poklicno-tehnično ...	6
višje strokovno	6
visoko strokovno, univerzitetno ...	5
doktorat	0
skupaj	61

Ali uporabljate SSKJ?	Število
da	30
ne	31
skupaj	61
Ali veste, kaj so besedilni korpusi?	Število
da	3 (2 sta obkrožila da, vendar delno vesta)
ne	52
delno	6 (eden je vedel, ampak obkrožil delno)
skupaj	61

Tabela 5: Rezultati anket (N = 61)

Komentar rezultatov anket

Iz anket smo ugotovili, da so besedilni korpusi nepoznani. Zanimivo je, da pozna korpuse več starejših ljudi kot mladih (osnovnošolcev in srednješolcev). Seveda tega ne smemo posploševati, navsezadnje gre za manjši vzorec anket (N = 61).

Zaradi teh podatkov se nam zdi ozaveščanje ljudi o besedilnih korpusih in njihovi uporabi nujna, predvsem v šolah (anketa je bila opravljena v obdobju december 2013–januar 2014).

Rezultati anket so bili podobni našim pričakovanjem. Pozitivno nas je presenetila samo uporaba SSKJ (skoraj 50 % anketirancev). Nihče od anketirancev ni vedel, kaj so besedilni korpusi, samo dva sta približno zapisala, za kaj gre.

Kljub splošni uporabi spleta je to področje še dokaj neraziskano in slabo približano uporabniku. Besedilni korpusi so prisotni že vsaj desetletje, vendar ne služijo svojemu namenu (zgolj ozkemu krogu raziskovalcem). Razloge gre iskati predvsem v zasnovi konkordančnih orodij, ki so zaradi svoje specializiranosti uporabnikom neintuitivni ter posledično neprijazni (Kosem 2012).

5 Primer uporabe korpusov in jezikovnih priročnikov

Na podlagi besede »**tajkun**«, ki je zadnja leta pogosto v uporabi, bomo opravili primerjavo pojavitve besede med **besedilnimi korpusi**, **Slovarjem slovenskega knjižnega jezika**⁸ (1997, 2008) in **Slovarjem novejšega besedja slovenskega jezika** (2012).

Pojavitev besede »**tajkun**« v posameznih korpusih:

KORPUS	ŠTEVILO ZADETKOV
Gigafida	16.365
KRES	1.003
GOS	4
FidaPLUS	88
Nova beseda	409

Tabela 6: Pregled pojavitve besede »**tajkun**« v besedilnih korpusih

SSKJ (1997, 2000, 2008) besede »**tajkun**« še nima vključene.

V **Slovarju novejšega besedja slovenskega jezika** (2012) zasledimo sledeči slovarski sestavek (str. 355):

tajkun –a m, člov. (û) zelo bogat in vpliven podjetnik z močnimi političnimi povezavami: *medijski tajkun; tranzicijski tajkun; Menedžerji ali tajkuni so z malverzacijami, izmišljenimi dokapitalizacijami in pritiski prišli do večinskih deležev podjetij* E agl. tycoon, jap. taikun »mogočnej, vladar«, kit. *taijun* iz *tai* »velik« + *jun* »vladar«

Beseda »**tajkun**« se v korpusih prvič pojavi leta 1997, kot zanimivost navajamo podatke za pojavitve besede v največjem korpusu **Gigafida** (glede na vrsto gradiv je največ pojavitve besede »**tajkun**« s spleta in iz časopisov – glede na večanje števil se potrjuje **živost jezika**): 1997 (6), 1998 (6), 1999 (50), 2000 (61), 2001 (26), 2002 (33), 2003 (74), 2004 (63), 2005 (55), 2006 (136), 2007 (973), 2008 (3.474), 2009 (1.623), 2010 (6.522), 2011 (3.263).⁹

⁸ Spletna različica SSKJ <<http://bos.zrc-sazu.si/sskj.html>> (Dostop 21. 2. 2014)

⁹ Podatki pridobljeni 21. 2. 2014.

6 Sklepne ugotovitve

Iz anket smo ugotovili:

A) besedilni korpusi so nepoznani, saj ljudje ne vedo, čemu služijo (samo 5 % jih ve);

B) uporaba SSKJ je solidna, saj ga anketiranci uporabljajo (skoraj 50 %).

Uporaba korpusov:

A) Korpusi so lažji za uporabo kot SSKJ, saj je iskanje besed hitrejše in nam posreduje več podatkov (pri korpusih moramo besedo samo napisati v iskalnik, pri SSKJ pa jo moramo poiskati oz. jo imamo v elektronski različici zapisano samo v slovarski obliki).

B) Korpusi so poučni, pri pouku bi jih lahko uporabljali pogosteje, kar bi pripomoglo k poznavanju besed. To lahko potrdijo tudi učenci naše šole, saj uporabljamo besedilne korpuse tudi pri pouku slovenščine.

V sklepnih ugotovitvah želimo preveriti, ali naše hipoteze držijo ali ne.

H 1: Besedilni korpusi slovenskega jezika so ljudem nepoznani.

Navedena hipoteza v celoti drži, to potrjujejo tudi rezultati anket (*52 od 61 jih je odgovorilo z ne*). Ljudje – uporabniki spleta bi morali biti bolj ozaveščeni o besedilnih korpusih, kajti slednji so uporabni na različnih področjih. Z ozaveščanjem ljudi bi povečali prepoznavnost korpusov in njihovo uporabo, kar bi pripomoglo k razvoju in poznavanju slovenskega jezika.

H 2: Uporaba korpusov v vsakdanjem življenju je redka.

Tudi ta hipoteza v celoti drži, saj jo potrjujejo rezultati ankete (*52 ljudi je odgovorilo, da ne pozna besedilnih korpusov, zaradi tega jih tudi ne morejo uporabljati*). Iz opravljene raziskave, predstavljene v prispevku Simona Kreka in Špele Arhar (2010), lahko hipotezo še dodatno podpremo. Raziskovalca sta prišla do zaključka, da je uporaba korpusov slaba, za reševanje domačih nalog ga uporablja predvsem študentska populacija. Referenčni korpus FidaPLUS se uporablja pri poučevanju na stopnji študija, uporabljajo pa ga tudi nekateri učitelji osnovnih oz. srednjih šol za pripravo učnih gradiv.

H 3: Večina ljudi – uporabnikov spleta ne zna uporabljati besedilnih korpusov.

Iz naših lastnih izkušenj lahko zapišemo, da imamo pri uporabi **Nove besede** veliko težav, saj je nepregledna in oblikovno neprivlačna, za razliko od **korpusa GOS** in **Gigafida**, ki sta za uporabo preprosta (osnovno iskanje). Nekoliko več težav imamo, ko želimo poiskati besedo npr. v določenem sklonu (napredno iskanje). Gre za t. i. **konkordance** (prikaz pojavov določene besede ali niza skupaj v sobesedilu) oz. **iskalne parametre** (npr. oblikoslovne, skladijske idr.). Skratka, če gre za osnovno iskanje, je uporaba pri večini korpusov preprosta, sicer pa imajo uporabniki pri naprednejšem/podrobnejšem iskanju precej težav.

H 4: Glede na uporabo in sestavo so si korpusi med sabo različni.

Hipoteza popolnoma drži, saj se korpusi med sabo razlikujejo glede sestave in namena uporabe, vsebujejo besede iz različnih virov (*npr. korpus **Gigafida** zajema večino besedil s področja časopisov in revij (77 %), korpus **GOS** pa vsebuje besede, zajete iz posnetkov govorov*).

H 5: V korpusih je delež novejših besed mnogo večji kot v aktualnem SSKJ. (X)

Ta hipoteza trenutno še drži, saj večina korpusov, ki smo jih raziskali, vsebuje besede, ki so se uporabljale v obdobju od 1990 do 2012. Naslednje leto je predviden izid novega SSKJ, ki bo vključeval tudi besede iz besedilnih korpusov. Nekaj besed lahko zasledimo že v **Slovarju novejšega besedja slovenskega jezika** (2012).

7 Zaključek

Raziskovanje besedilnih korpusov nam je bilo v izredno veselje, saj smo spoznavali področje, ki je lahko za uporabnike spleta še kako privlačno in zanimivo. Besedilni korpusi so prihodnost raziskovanja **živosti jezika**, saj nam podajo celovito podobo o pogostosti in uporabi določene besede v daljšem/krajšem časovnem obdobju.

Osredotočili smo se na **referenčne besedilne korpuse**, ki so najobsežnejši in zagotavljajo reprezentativen izbor besedil določenega jezika, zato služijo temeljnim raziskavam na področju slovnice in slovarja. V to kategorijo spada naš največji slovenski korpus **Gigafida** (prej **FidaPLUS**) in **Nova beseda**.

V skupini **govornih korpusov** smo predstavili korpus govorne slovenščine – **GOS**. Namenjen je tistim raziskovalcem, ki želijo preučevati govorno slovenščino bodisi z jezikoslovnega bodisi s katerega drugega vidika.

Za našo vsakdanjo uporabo je zelo koristen **korpus šolskih pisnih izdelkov Šolar**. Z njim lahko kadar koli preverimo zapis posamezne besede v konkretni rabi. Dobimo povratno informacijo ne samo o pravilnosti uporabe, ampak tudi popravek v zapisanem primeru iz šolske rabe, zakaj je temu tako.

Žal so besedilni korpusi še vedno slabo poznani, tudi v vzgojno-izobraževalnem procesu se uporabljajo redko. Če bi bodoči potencialni uporabniki vedeli, katere informacije lahko s preprostim vpisom besede dobijo na enem mestu, bi se uporaba zagotovo povečala.

Z nazorno predstavitevijo korpusov želimo prispevati k ozaveščanju uporabe besedilnih korpusov. Pomembno se nam zdi, da si znanje širimo tudi z uporabo izobraževalnih aplikacij, ki nam jih v sodobnem času nudi vse hitreje razvijajoči svetovni splet. Svoje vedenje o maternem jeziku lahko obogatimo tudi z besedilnimi korpusi – uporabljajmo jih.

Viri in literatura

Spletni naslovi besedilnih korpusov:

Gigafida: <<http://www.gigafida.net/>>

KRES: <<http://www.korpus-kres.net/>>

GOS: <<http://www.korpus-gos.net/>>

FidaPLUS: <<http://www.fidaplus.net/>>

Nova beseda: <http://bos.zrc-sazu.si/s_beseda3.html>

Šolar: <http://www.korpus-solar.net/run.cgi/first_form?corpname=fidaplus-solar>

Špela Arhar Holdt, 2011: *Luščenje besednih zvez iz besedilnega korpusa z uporabo dvodelnih in tridelnih oblikoskladenjskih vzorcev*. Ljubljana: Trojina, zavod za uporabno slovenistiko.

Aleksandra Bizjak Končar, Marko Snoj (ur.), 2012: *Slovar novejšega besedja slovenskega jezika*. Ljubljana: Založba ZRC SAZU.

Vojko Gorjanc, 2005: *Uvod v korpusno jezikoslovje*. Domžale: Izolit.

Vojko Gorjanc, Simon Krek (ur.), 2005: *Študije o korpusnem jezikoslovju: zbornik*. Ljubljana: Krtina.

Iztok Kosem (ur.), 2012: *Analiza jezikovnih težav učencev: korpusni pristop*. Ljubljana: Trojina, zavod za uporabno slovenistiko.

Simon Krek in Špela Arhar, 2010: Slovenski besedilni korpusi: kako v razred? *Sodobna pedagogika*, št. 1. 224–241.

Aleš Pogačnik (ur.), 2008: *Slovar slovenskega knjižnega jezika* (ponatis v petnajstih knjigah). Ljubljana: DZS.

Tadeja Rozman (ur.), 2012: *Empirični pogled na pouk slovenskega jezika*. Ljubljana: Trojina, zavod za uporabno slovenistiko.

Marko Stabej, 2010: *V družbi z jezikom*. Ljubljana: Trojina, zavod za uporabno slovenistiko.

Mojca Šorli (ur.), 2012: *Dvojezična korpusna leksikografija: slovenščina v kontrastu: novi izzivi, novi obeti*. Ljubljana: Trojina, zavod za uporabno slovenistiko.

Jože Toporišič, 2000: *Slovenska slovnica*. Maribor: Obzorja.

Jože Toporišič [et al.], 2001: *Slovenski pravopis*. Ljubljana: Založba ZRC, ZRC SAZU.

Darinka Verdonik (ur.), 2011: *Slovenski govorni korpus Gos*. Ljubljana: Trojina, zavod za uporabno slovenistiko.

Priloga

Anketa – Besedilni korpusi (na naslednji strani)

ANKETA – BESEDILNI KORPUSI

Spoštovani anketiranec, spoštovana anketiranka,
smo Filip, Urh in Luka, devetošolci – učenci I. osnovne šole Celje. Za potrebe raziskovalne
naloge te prosimo/vas prosimo, če izpolniš/izpolnite anketo o besedilnih korpusih. Anketa je
anonimna.

1. Spol (obkrožite): a) moški b) ženski c) nočem zaupati

2. Koliko ste stari (obkrožite)? a) 13–18 b) 19–35 c) 36–50 č) več kot 50

3. Kakšen je vaš besedni zaklad?

a) obsežen

b) soliden

c) skop

č) drugo _____

4. Kakšno stopnjo izobrazbe imate?

a) 1. stopnja izobrazbe – nedokončana OŠ

b) 2. stopnja izobrazbe – OŠ

c) 3. stopnja izobrazbe – nižje poklicno izobraževanje

č) 4. stopnja izobrazbe – srednje poklicno izobraževanje

d) 5. stopnja izobrazbe – gimnazijsko, srednje poklicno-tehnično izobraževanje, srednje
tehnično oz. drugo strokovno izobraževanje

e) 6. stopnja izobrazbe – višje strokovno izobraževanje

f) 7. stopnja izobrazbe – visoko strokovno izobraževanje, univerzitetno diplomsko
izobraževanje, magisterij

g) 8. stopnja izobrazbe – doktorat

OBRNI!

5. Ali uporabljate SSKJ (Slovar slovenskega knjižnega jezika)?

a) da

b) ne

6. Ali veste, kaj so (besedilni) korpusi?

a) da

b) ne

c) delno

7. Če ste obkrožili črko a, napišite, kaj so besedilni korpusi. Navedite tudi primer.

Zahvaljujemo se ti/vam za sodelovanje.