

POLETNI RAZISKOVALNI TABOR  
SPLETNE SLOVENŠČINE JANES  
ZA SREDNJEŠOLCE

Oddelek za prevajalstvo  
Filozofska fakulteta, Ljubljana

**4. – 8. julij 2016**

# UČNA GRADIVA

**JANES** (••)

## 2. POLETNI RAZISKOVALNI TABOR SPLETNE SLOVENŠČINE JANES ZA SREDNJEŠOLCE

### Zbornik učnih gradiv

Uredila: Darja Fišer

Univerza v Ljubljani, Filozofska fakulteta, 2016.

To delo je ponujeno pod licenco Creative Commons: Priznanje avtorstva-Deljenje pod enakimi pogoji 4.0 Slovenija.

Založila: Znanstvena založba Filozofske fakultete Univerze v Ljubljani

Izdal: Oddelek za prevajalstvo

Za založbo: dr. Branka Kalenič Ramšak, red. prof. Filozofske fakultete

Ljubljana, 2016

1. izdaja

Publikacija je brezplačno dostopna na spletni strani:

<http://nl.ijs.si/janes/wp-content/uploads/2016/08/Zbornik-gradiv-JANES-tabor-2016.pdf>

CIP - Kataložni zapis o publikaciji

Narodna in univerzitetna knjižnica, Ljubljana

811.163.6(083.97)(0.034.2)

POLETNI raziskovalni tabor spletne slovenščine Janes za srednješolce (2 ; 2016 ; Ljubljana)

Učna gradiva [Elektronski vir] / 2. poletni raziskovalni tabor spletne slovenščine Janes za srednješolce, Ljubljana, Filozofska fakulteta, Ljubljana 4.-8. julij 2016 ; [uredila Darja Fišer]. - 1. izd. - El. knjiga. - Ljubljana : Znanstvena založba Filozofske fakultete Univerze v Ljubljani, 2016

Način dostopa (URL): <http://nl.ijs.si/janes/wp-content/uploads/2016/08/Zbornik-gradiv-JANES-tabor-2016.pdf>

ISBN 978-961-237-849-3 (pdf)

1. Gl. stv. nasl. 2. Fišer, Darja, 1978-  
286114816

## KAZALO VSEBINE

<b>PREDGOVOR</b>	<b>5</b>
<b>1 KORPUSI IN KORPUSNO JEZIKOSLOVJE</b>	<b>7</b>
1.1 Korpusi in korpusno jezikoslovje – izročki	7
1.2 Korpusi in korpusno jezikoslovje – vaje	21
1.2.1 Od konkordance do kolokacije – prvi del	22
1.2.2 Od konkordance do kolokacije – drugi del	24
<b>2 STANDARDNA IN SPLETNA SLOVENŠČINA</b>	<b>26</b>
2.1 Standardna in spletna slovenščina – izročki	26
2.2 Od pravopisa do tvitopisa – vaje	47
2.2.1 Pregibanje osebnih lastnih imen	47
2.2.2 E-tvorjenke	47
2.2.3 Stvarna lastna imena	49
2.2.4 Kratice in simboli	50
2.2.5 Zapis tvorjenk	51
2.3 Od pravopisa do tvitopisa – delavnica	53
<b>3 FRAZEOLOGIJA V SPLETNI SLOVENŠČINI</b>	<b>55</b>
3.1 Frazeologija v spletni slovenščini – izročki	55
3.2 Povzetek	73
3.3 Prepoznavanje različnih tipov večbesednih leksikalnih enot v besedilu – vaje	74
3.3.1 Od kolokacije do frazema	75
3.3.2 Od kolokacije k frazeološkim enotam	76
3.4 Delavnice	77
<b>4 GOVOR IN SPLETNA SLOVENŠČINA</b>	<b>78</b>
4.1 Govor in spletna slovenščina – izročki	78
4.2 Detekstiv – vaje	90
4.2.1 Izhodišče	90
4.2.2 Vaje	90
4.3 Detekstiv – delavnica Odkrivanje govornih elementov na spletnih forumih	92
4.4 Detekstiv – delavnica Odkrivanje govornih elementov na Twitterju	93
<b>5 SLOVENŠČINA V STIKU Z DRUGIMI JEZIKI</b>	<b>94</b>
5.1 Slovenščina v stiku z drugimi jeziki – izročki	94
5.2 Od Novega mesta do Njujorka – vaje	113
5.2.1 Kako zapisujemo tujejezične elemente?	113

---

5.2.2 Koliko je tujejezičnih elementov v JANES-u in iz katerih jezikov so? _____	114
5.2.3 Kdaj in kako se uporablja tujejezične elemente? _____	115
5.2.4 Kdo uporablja tujejezične elemente? _____	116
5.2.5 Stopnja (ne)standardnosti in tujejezični elementi _____	117

## PREDGOVOR

Od 4. do 8. julija 2016 je na Oddelku za prevajalstvo Filozofske fakultete Univerze v Ljubljani potekal 2. raziskovalni tabor spletne slovenščine za srednješolce JANES. Tabor je bil organiziran v okviru temeljnega nacionalnega projekta [JANES](#) – Jezikoslovna analiza nestandardne slovenščine je nacionalni raziskovalni projekt (J6—6842), ki ga od 1. 7. 2014 do 30. 6. 2017 financira Javna agencija za raziskovalno dejavnost Republike Slovenije, sofinancirala pa ga je Slovenska raziskovalna infrastruktura za jezikovne vire in tehnologije [CLARIN.SI](#).

Z raziskovalnim taborom smo želeli srednješolcem približati slovenščino in jim pokazati, da je zanimiv in pomemben predmet proučevanja. Poleg tega je bil naš cilj tudi, da dijake seznanimo z jezikovnimi tehnologijami, s katerimi slovenščina postaja vse bolj opremljena. Tako smo dijake s preglednimi predavanji in vodenimi vajami uvedli v svet korpusnega jezikoslovja in jezikovnih tehnologij, na praktično zasnovanih raziskovalnih delavnicah pod nadzorom mentorjev, vrhunskih strokovnjakov s področja korpusnega jezikoslovja in jezikovnih tehnologij pa so srednješolci dobili še priložnost za raziskovanje izbranega raziskovalnega vprašanja z najsodobnejšimi korpusnimi viri, tehnologijami in metodami, razvitimi v okviru projekta [JANES](#).

Raziskovalni tabor je trajal pet dni, vsak dan je bil posvečen drugi osrednji temi. Tabor je [otvorila](#) dr. Špela Arhar Holdt z uvodom v korpusno jezikoslovje ter predstavitevijo najpomembnejših korpusnih virov in orodij, predstavi osnove korpusnega jezikoslovja in delo z najpomembnejši korpusi za slovenščino, kot so referenčni korpus Gigafida, vzorčni korpus Kres, korpus govorne slovenščine Gos, učni korpus Šolar in korpus spletne slovenščine Janes. [Druga tema](#) je bila namenjena ozaveščanju razlik med standardno in spletno slovenščino na ortografski, leksikalni, skladenjski in slogovni ravni, v okviru katere je dr. Damjan Popič predstavil razvoj slovenske pravopisne norme in osvetlil razlike med standardno in spletno slovenščino, pri čemer je posebno pozornost namenil rabi vejice. [Tretja tema](#), ki jo je pripravila dr. Polona Gantar, je bila posvečena kolokacijskim in frazeološkim posebnostim slovenščine v spletnih skupnostih. V [četrti temi](#) so dr. Ana Zwitter Vitez, dr. Špela Arhar Holdt in dr. Tadeja Rozman jezik družbenih medijev primerjale z govorno slovenščino ter demonstrirale postopke za odkrivanje govornih elementov v spletnih uporabniških vsebinah. Tabor sta s [peto temo](#) sklenila dr. Tamara Mikolič Južnič in Jaka Čibej, ki sta spregovorila o pojavljanju in zapisovanju tujejezičnih prvin ter preklapljanju med jeziki v spletnih žanrih.



---

S pomočjo preglednih predavanj in opravljenih analiz smo dijake seznanili s temeljnimi pojmi s področja korpusnega jezikoslovja, sociolingvistike in analize diskurza ter jim dvignili zavest o izbiri ustreznih jezikovnih različic v različnih govornih položajih. Poleg tega smo jih opremili tudi s kompetencami, ki jim omogočajo, da s pomočjo prosto dostopnih virov za slovenščino samostojno odgovarjajo na najrazličnejša jezikovna vprašanja. Ker menimo, da tovrstne vsebine postajajo vse bolj aktualne, a so še vedno premalo vključene v učne in študijske programe, smo prepričani, da bodo gradiva koristila tudi drugim dijakom, študentom, profesorjem in jezikoslovcem, ki se želijo seznaniti s korpusnim jezikoslovjem in značilnostmi spletne slovenščine, zato smo jih zbrali v skupni publikaciji, ki je pred vami.

doc. dr. Darja Fišer

Ljubljana, avgust 2016

Špela Arhar Holdt in Jaka Čibej

## 1 KORPUSI IN KORPUSNO JEZIKOSLOVJE

### 1.1 Korpusi in korpusno jezikoslovje – izročki

JANES

IJS  
FILOZOFSKA  
FAKULTETA

# Korpusi in korpusno jezikoslovje

Špela Arhar Holdt  
Center za uporabno jezikoslovje ZUS Trojina  
Filozofska fakulteta UL

Ljubljana, 4. 7. 2016

JANES

IJS  
FILOZOFSKA  
FAKULTETA

## Naloga za prvi dan

Spoznati pojme in orodja, ki jih bomo uporabljali ta teden.

- Kaj je korpus?
- Vzorec in generalizacija
- Gradnja korpusov
- Raba korpusov
- Korpus in priročniki
- Slovenski korpusi

JANES IJS FILOZOFSKA  
FAKULTETA

## Želimo raziskovati slovenščino!

časopisi, revije  
pošta, e-pošta  
knjige  
besedilni drobiž  
govorjeni jezik  
spletne strani  
družbena omrežja  
itd.

JANES IJS FILOZOFSKA  
FAKULTETA

# Kako?

Gigafida Iščanje Okolica Seznanj Pomoc Q.korpusu Slovensko

Čofotalnik Najdi

Uporabljaj enostavno iskanje Napredno iskanje

1 2 3 4 naslednja stran

Prikazujem 1-20 od 80 konkordanc (0.343 sekund).

**Osnovne oblike**  
Čofotalnik (80)

**Vrsta besedila**  
Časopisi (53)  
Revije (12)  
Internet (4)  
Svama besedila (1)  
Več

**Vir**  
Dnevnik (28)  
drugo (22)  
Delo (11)  
Večer (6)  
Gorenjski glas (5)  
Več

**Leto**  
2003 (16)  
2004 (15)  
2001 (11)

sta bila od skupno 34 vzorcev dva mikrobiološko neustrezna bazena čofotalnika na kopalšču RC Cizej, Orta vas pri Braslovcah, . To bo kompleks zunanjih bazenov, ki bo v čofotalniku ponujal vodno veselje najmlajim, na grčah in toboganih pa se lahko zabavajo tudi v zunanjem gusarskem bazenu in notranjem čofotalniku . se lahko zabavajo tudi v zunanjem gusarskem bazenu in notranjem čofotalniku . dodatno razveseljuje obnoven bazen s številnimi vodnimi doživetji, otroški čofotalnik in masažni bazen whirlpool. Celovita ponudba Dežele dobrega počutja park z dvema dičama, toboganom in otroškim bazenom - čofotalnikom . Za krepitev zdravja vaši Sawaddee, center tradicionalne tajske . posebne kopeli, za najmlajše pa je tu otroški čofotalnik . Gozdni vodni park ponuja vodne in biseme vrelce, ki se bodo lahko namakali v obeh bazenih ter otroškem čofotalniku . se sprošili v velikem savna centru in telovadli v park z dvema dičama, toboganom in otroškim bazenom - čofotalnikom . Za krepitev zdravja bodo poskrbeli v centru tradicionalne tajske za dezinfekcijo vode. Kopalšču manjkata tudi kompensacijski bazen in čofotalnik za dojenčke ter manjše otroke, ki zdaj pogosto čofotajo tiste, ki želijo toplejšo vodo. Predlagal bi » čofotalnik «. v treh bazenih: zunanjem, v velikem bazenu in čofotalniku . V tamkajšnjem vročem bazenu oziroma savni pa je neustrezen o. Zimsko kopalšče v velikem in malem bazenu ter čofotalniku . V Zdravilišču Laško sta imela kemično neustrezno vodo zunanji Gre za kompleks zunanjih bazenov, ki bo v čofotalniku ponujal vodno veselje najmlajšim obiskovalcem, na grčah in toboganih



## Besedilni korpus



Zberemo veliko količino besedil v elektronski obliki.  
Besedila izberemo tako, da predstavljajo vzorec jezika.  
Besedila pripravimo, da jih je lažje računalniško raziskovati.  
Besedila vstavimo v specializiran program: konkordančnik.

## Besedilni korpus



Zberemo veliko količino besedil v elektronski obliki.  
Besedila izberemo tako, da predstavljajo **vzorec jezika**.  
Besedila pripravimo, da jih je lažje računalniško raziskovati.  
Besedila vstavimo v specializiran program: konkordančnik.

## Vzorec - primer

Recimo, da nas zanima, kakšno glasbo poslušajo slovenski srednješolci. Kako bomo to ugotovili?

Je dovolj, če ocenimo stanje glede na svoje izkušnje? Vprašamo svoje sošolce? Ali vse dijake na svoji šoli? Dijake z različnih srednjih šol? Morda različnih starosti, spolov ali iz različnih regij?



Kakšen **vzorec** moramo zajeti v raziskavi, da lahko na koncu **generaliziramo** rezultate?

## Vzorec - primer

Recimo, da nas zanima, kakšno glasbo poslušajo slovenski srednješolci. Kako bomo to ugotovili?

Je dovolj, če ocenimo stanje glede na svoje izkušnje? Vprašamo svoje sošolce? Ali vse dijake na svoji šoli? Dijake z različnih srednjih šol? Morda različnih starosti, spolov ali iz različnih regij?



Kakšen **vzorec** moramo zajeti v raziskavi, da lahko na koncu **generaliziramo** rezultate?

## Vzorec - korpus

Besedilni korpusi so zgrajeni tako, da predstavljajo **vzorec jezika**. Da lahko pravilno **interpretiramo** in **generaliziramo** ugotovitve, moramo dobro razumeti, kakšna besedila določen korpus vsebuje, kako je bil zgrajen in kakšen je njegov namen.

<b>Korpus Kres</b>	Splošna pisna slovenščina	časopisi, revije, leposlovje, strokovna literatura, spletna besedila, besedilni drobiž
<b>Korpus GOS</b>	Govorjena slovenščina	televizijske in radijske oddaje, javni nastopi, sestanki, zasebna komunikacija
<b>Korpus Janes</b>	Spletna slovenščina	tviti, blogi, uporabniški komentarji, forumi
<b>Korpus Šolar</b>	Jezik šolarjev	šolski eseji in testi + učiteljski popravki

## Besedilni korpus




Zberemo veliko količino besedil v elektronski obliki.  
Besedila izberemo tako, da predstavljajo vzorec jezika.  
Besedila pripravimo, da jih je lažje računalniško raziskovati.  
Besedila vstavimo v specializiran program: konkordančnik.

## Besedilni korpus



Zberemo veliko količino besedil v elektronski obliki.  
Besedila izberemo tako, da predstavljajo vzorec jezika.  
Besedila **pripravimo**, da jih je lažje računalniško raziskovati.  
Besedila vstavimo v specializiran program: konkordančnik.

## Priprava korpusnih besedil

1. K besedilom dodamo vse informacije, ki so na voljo: od kod besedilo izvira, kdaj je nastalo oz. izšlo, kdo je avtor ...
2. Besedila **jezikoslovno označimo**: s posebnim programom pripišemo besedi osnovno obliko, besedno vrsto in druge lastnosti. 

*Vaši malčki lahko varno uživajo v otroškem čofotalniku. (časopis Celjan, 2009)*

Vaši	vaš	svojilni zaimek; 2. oseba, moški spol, množina, imenovalnik, množina svojine
malčki	malček	samostalnik, občno ime; moški spol, množina, imenovalnik
lahko	lahko	splošni prislov; nedoločena stopnja
varno	varno	splošni prislov; nedoločena stopnja
uživajo	uživati	glavni glagol; nedovršni, sedanjik, 3. oseba, množina
v	v	predlog; mestnik
otroškem	otroški	splošni pridevnik; nedoločena stopnja, moški spol, ednina, mestnik
čofotalniku	čofotalnik	samostalnik, občno ime; moški spol, ednina, mestnik

## Besedilni korpus



Zberemo veliko količino besedil v elektronski obliki.  
Besedila izberemo tako, da predstavljajo vzorec jezika.  
Besedila pripravimo, da jih je lažje računalniško raziskovati.  
Besedila vstavimo v specializiran program: konkordančnik.

## Besedilni korpus



Zberemo veliko količino besedil v elektronski obliki.  
Besedila izberemo tako, da predstavljajo vzorec jezika.  
Besedila pripravimo, da jih je lažje računalniško raziskovati.  
Besedila vstavimo v specializiran program: **konkordančnik**.

JANES IJS FILOZOFSKA  
FAKULTETA

Gigafida Iskanje Okočica Seznam Pomoč O korpusu Slovensko

čofotalnik Najdi

Uporabljaš enostavno iskanje Napredno iskanje

1 2 3 4 naslednja stran

Prikazujem 1-20 od 80 konkordanc (0.343 sekund).

**Osnovne oblike**  
Čofotalnik (80)

**Vrsta besedila**  
Časopisi (63)  
Revije (12)  
Internet (4)  
Štvarna besedila (1)  
Več

**Vir**  
Dnevnik (28)  
drugo (22)  
Delo (11)  
Večer (6)  
Gorenjski glas (5)  
Več

**Leto**  
2003 (16)  
2004 (15)  
2001 (11)

sta bila od skupno 34 vzorcev dva mikrobiološko neustrezniz bazena čofotalnika na kopališču RC Cizej, Orta vas pri Braslovčah,  
. To bo kompleks zunanjih bazenov, ki bo v čofotalniku ponujal vodno veselje najmlajšim, na grčah in toboganih pa  
se lahko zabavajo tudi v zunanjem gusarskem bazenu in notranjem čofotalniku .  
se lahko zabavajo tudi v zunanjem gusarskem bazenu in notranjem čofotalniku .  
dodatno razveseljuje obnovljen bazen s številnimi vodnimi doživetji, otroški čofotalnik in masažni bazen whirlpool. Celovita ponudba Dežele dobrega počutja  
park z dvema drčama, toboganom in otroškim bazenom - čofotalnikom . Za krepitev zdravja vabi Sawaddee, center tradicionalne tajske  
, posebne kopeli, za najmlajše pa je tu otroški čofotalnik . Gozdni vodni park ponuja vodne in biserne vrele,  
ki se bodo lahko namakali v dveh bazenih ter otroškem čofotalniku , se sprostiti v velikem savna centru in telovadli v  
park z dvema drčama, toboganom in otroškim bazenom - čofotalnikom . Za krepitev zdravja bodo poskrbeli v centru tradicionalne tajske  
za dezinfekcijo vode. Kopališču manjkata tudi konpenzacijski bazen in čofotalnik za dojenčke ter manjše otroke, ki zdaj pogosto čofotajo  
tiste, ki želijo toplejšo vodo. Predlagal bi » čofotalnik « .  
v treh bazenih: zunanjem, v velikem bazenu in čofotalniku . V tamkajšnjem vročem bazenu oziroma savni pa je neustrezen  
o.. Zimsko kopališče v velikem in malem bazenu ter čofotalniku . V Zdravilišču Laško sta imela kemično neustrezno vodo zunanji  
. Gre za kompleks zunanjih bazenov, ki bo v čofotalniku ponujal vodno veselje najmlajšim obiskovalcem, na grčah in toboganih

JANES IJS FILOZOFSKA  
FAKULTETA

## Besedilni korpus - definicija

Besedilni korpusi so **obsežne zbirke realnih besedil** v elektronski obliki. Besedila so zajeta iz različnih virov na način, da predstavljajo **vzorec jezikovne rabe določene vrste**. Korpusna besedila tipično vsebujejo programsko ali ročno pripisane **oznake**, npr. osnovno obliko besede, besedno vrsto in druge lastnosti besede. Za raziskovanje besedilnih korpusov so besedila umeščena v **konkordančnike** specializirane programe, ki omogočajo napredno iskanje po besedilih, razvrščanje, filtriranje, izvažanje podatkov in podobno.

## Zakaj uporabljati korpusse?

- Ker naša **jezikovna intuicija** ne more natančno predvideti, kako se jezik v širši rabi obnaša (kot ne moremo na pamet predvideti, kakšno glasbo imajo radi slovenski srednješolci).
- Ker s pomočjo računalnika lahko obdelamo večje količine podatkov na naprednejše načine in tako lažje poiščemo relevantne **jezikovne vzorce in trende** (to je posebej pomembno za večje projekte, npr. gradnjo slovarja),
- Ker so (v primerjavi z Googlom, na primer) zgrajeni na transparenten in dokumentiran način, da lahko podatke ustrezno **interpretiramo in generaliziramo**.

## Raba besedilnih korpusov

### Uporabno jezikoslovje

- slovarji, slovnice, pravopis, šolska gradiva ...

### Teoretično jezikoslovje

- nova dognanja o jezikovni rabi in sistemu

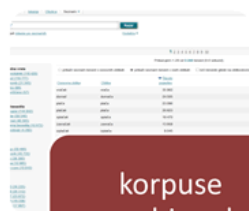
### Jezikovna produkcija

- pisanje, prevajanje, lektoriranje ...

### Ljubiteljsko raziskovanje

- preverjanje jezikovne rabe, zanimivosti ...

## Od korpusa do priročnika



korpusse uporabimo kot vir jezikovnih podatkov



uporabniki lahko preberejo opis jezika v priročnikih

Za slovenščino še ne obstaja veliko priročnikov, ki so narejeni na osnovi korpusnih podatkov (v prihodnosti jih bo več). Trenutno je zato najti razlike med korpusnimi podatki in podatki nekaterih priročnikov.

## Korpusi in jezikovni priročniki



KORPUSI

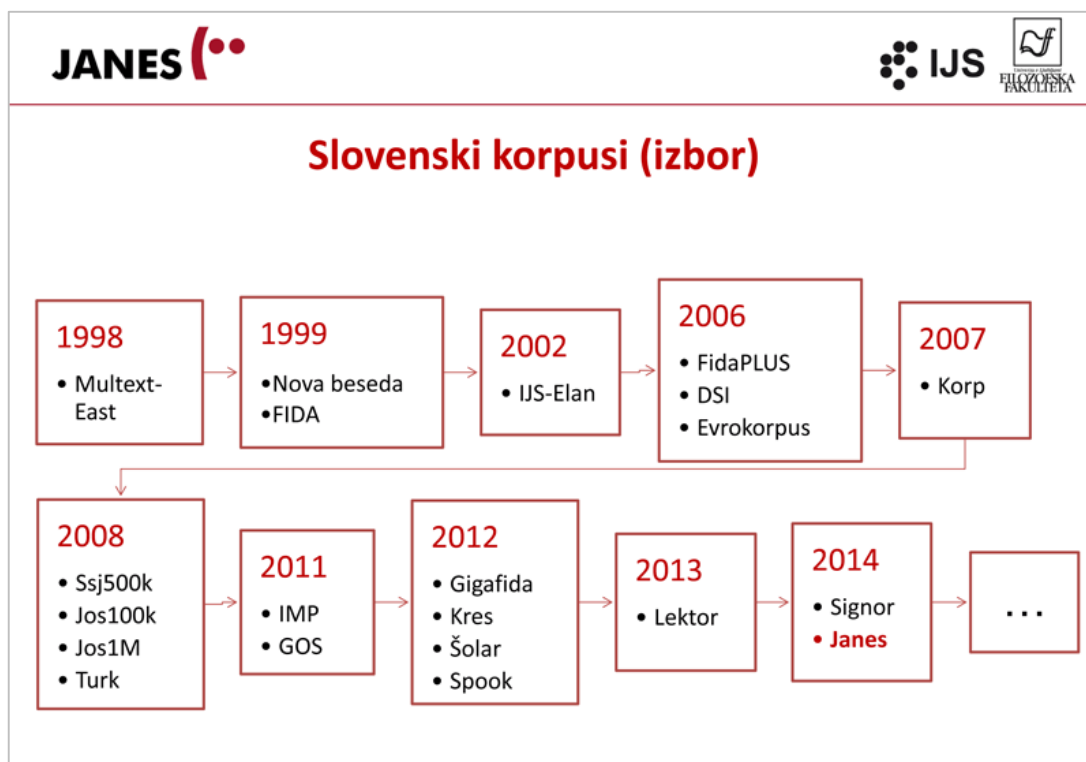
- Vidimo sobesedilo in podatke o besedilu
- Potrebna je sinteza in interpretacija
- Možnost hitrega posodabljanja
- Razen pripisanih oznak so besedila v izvorni obliki



PRIROČNIKI

- Vsak priročnik služi določenemu namenu.
- Jezikovno gradivo je izbrano in urejeno glede na ta namen.
- Priročniki imajo pogosto (tudi) normativno vrednost.



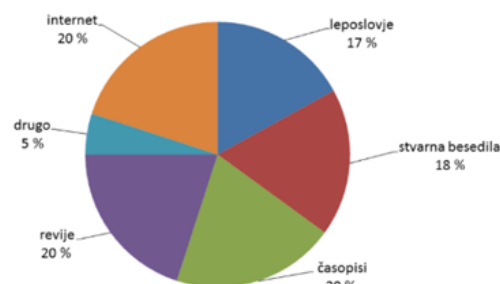
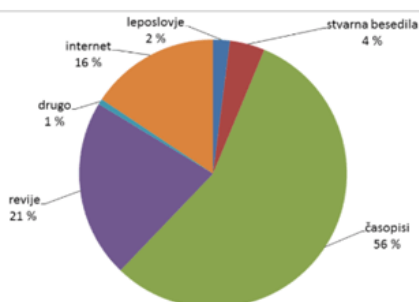


**JANES** IJS FILOZOFSKA  
FAKULTETA

### Gigafida in Kres

- **Gigafida** je obsežna zbirka sodobnih (1990-2011) slovenskih besedil iz časopisov, revij, knjig, s spleta itd. Korpus obsega skoraj 1,2 milijarde besed.
- **Kres** je manjša različica tega korpusa, prinaša cca. 100 milijonov besed.
- Korpusi, ki prinašajo splošni, nespecializirani jezik, imenujemo **referenčni korpusi**. Ti se uporabljajo za izdelavo referenčnih priročnikov, v raziskavah pa jih pogosto uporabljamo tako, da z njimi primerjamo rezultate iz drugih korpusov (npr. rezultate raziskovanja po korpusu Janes primerjamo z referenčnim korpusom, da vidimo, kaj je specifično za spletni jezik, kaj pa se pojavlja tudi v splošni jezikovni rabi).

## Gigafida in Kres



## GOS

- **GOS** je prvi korpus govorne slovenščine. Prinaša posnetke govora v različnih vsakodnevni situacijah.
- Posnetki so **transkribirani** in umeščeni v zmožljiv konkordančnik, s katerim lahko primere govora iščemo, poslušamo in preučujemo.
- Korpus obsega okrog **milijon besed**.
- Namenjen je **raziskovanju govora**.

Transkribcija 1: *ja ne vem po **kermu ključu** se bomo odločali eee koga bomo poslali v samo trgovanje z volno eem ne vem a bo to žrebanje al al glasvanje al kekrkoli*

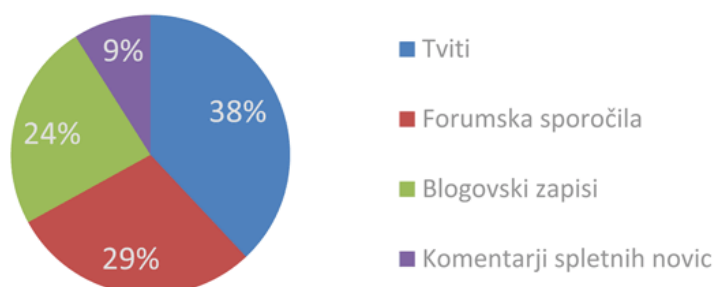
Transkribcija 2: *ja ne vem po **kateremu ključu** se bomo odločali eee koga bomo poslali v samo trgovanje z volno eem ne vem a bo to žrebanje ali ali glasovanje ali kakor*

## Šolar

- **Šolar** vsebuje pisna besedila, ki so jih učenci in dijaki slovenskih šol tvorili pri pouku.
- V precejšnjem delu besedil so posebej označene tudi jezikovne napake, ki so jih v spisih **popravili učitelji**.
- Po jezikovnih napakah oz. učiteljskih popravkih lahko s pomočjo specializiranega konkordančnika tudi iščemo.
- Korpus vsebuje približno **milijon besed**.
- Korpus je primarno namenjen raziskavam šolske pisne produkcije oz. jezikovne zmožnosti šolarjev in pripravi učnih gradiv.

## Janes

- **Janes** je korpus spletne slovenščine. Vsebuje besedila, ki so jih na spletu tvorili uporabniki. Korpus obsega okrog **134 milijonov besed**.



## Janes

### Janes je tudi razlog, da smo tu:

- *JANES – Jezikoslovna analiza nestandardne slovenščine* je nacionalni raziskovalni projekt (J6—6842), ki ga od 1. 7. 2014 do 30. 6. 2017 financira Javna agencija za raziskovalno dejavnost Republike Slovenije.
- Cilj projekta je zgraditi obsežen korpus spletne slovenščine, s pomočjo katerega bomo omogočili empirično podprto jezikoslovno analizo nestandardne spletne slovenščine, izboljšali jezikovnotehnološka orodja za obdelavo besedil, napisanih v nestandardnem jeziku, in izdelali slovarček spletne slovenščine.

## 1.2 Korpusi in korpusno jezikoslovje – vaje

**Besedilni korpusi** so obsežne zbirke realnih besedil v elektronski obliki. Besedila so zajeta iz različnih virov na način, da predstavljajo **vzorec jezikovne rabe** določene vrste. Korpusna besedila tipično vsebujejo programsko ali ročno pripisane **oznake**, npr. osnovno obliko besede, besedno vrsto in druge lastnosti besede. Za raziskovanje besedilnih korpusov so besedila umeščena v **konkordančnike**: specializirane programe, ki omogočajo napredno iskanje po besedilih, razvrščanje, filtriranje, izvažanje podatkov in podobno.

Da lahko pravilno interpretiramo in generaliziramo ugotovitve, moramo dobro razumeti, **kakšna besedila** določen korpus vsebuje, kako je bil **zgrajen** in kakšen je njegov **namen**.

Besedilne korpuse uporabljamo:

- ker naša **jezikovna intuicija** ne more natančno predvideti, kako se jezik v širši rabi obnaša,
- ker s pomočjo računalnika lahko obdelamo večje količine podatkov na naprednejše načine in tako lažje poiščemo relevantne **jezikovne vzorce in trende**,
- ker so zgrajeni na transparenten in dokumentiran način, da lahko podatke ustrezno **interpretiramo in generaliziramo**.

Korpusi se uporabljajo za različne namene v **uporabnem jezikoslovju** (za pripravo slovarjev, slovnice, šolskih gradiv ipd.), **teoretičnem jezikoslovju** (za raziskave, ki lahko vodijo do novih dognanj o jezikovni rabi in sistemu), pri drugih poklicih, ki se posvečajo **pisni produkciji** (pisanje, prevajanje, lektoriranje ipd.) in tudi za **ljubiteljsko raziskovanje** jezika (preverjanje jezikovne rabe, raziskovanje raznih zanimivosti ipd.)

Za slovenščino trenutno še ne obstaja veliko priročnikov, ki so narejeni na osnovi korpusnih podatkov (v prihodnosti jih bo več). Korpusi so tudi sodobnejši od nekaterih obstoječih priročnikov, zato se korpusni podatki in podatki v priročnikih mestoma razlikujejo). V praksi se korpusi zato pogosto uporabljajo kot dopolnilo obstoječim jezikovnim priročnikom.

Za slovenščino je na voljo več različnih korpusov. Na taboru bomo natančneje spoznali naslednje:

IME KORPUSA	VRSTA JEZIKA, POVEZAVA	ZAJETA BESEDILA
<b>Kres</b>	Splošna pisna slovenščina	časopisi, revije, leposlovje, strokovna literatura, spletna besedila, besedilni drobiž
<b>GOS</b>	Govorjena slovenščina	televizijske in radijske oddaje, javni nastopi, sestanki, zasebna komunikacija ...

<b>Janes</b>	Spletna slovenščina	tviti, blogi, uporabniški komentarji, forumi
<b>Šolar</b>	Jezik šolarjev	šolski eseji in testi + učiteljski popravki

**Gigafida** je obsežna zbirka sodobnih (1990-2011) slovenskih besedil iz časopisov, revij, knjig, s spleta itd. Korpus obsega skoraj 1,2 milijarde besed. **Kres** je manjša različica tega korpusa, prinaša cca. 100 milijonov besed. Korpuse, ki prinašajo splošni jezik, imenujemo **referenčni korpusi**. Ti se uporabljajo za izdelavo referenčnih priročnikov, v raziskavah pa jih pogosto uporabljamo tako, da z njimi primerjamo rezultate iz drugih korpusov.

**GOS** je prvi korpus govorne slovenščine. Prinaša posnetke govora v različnih vsakodnevnih situacijah. Posnetki so **transkribirani** in umeščeni v zmogljiv konkordančnik, s katerim lahko primere govora iščemo, poslušamo in preučujemo. Korpus obsega okrog **milijon besed**. Namenjen je raziskovanju govora.

**Šolar** vsebuje pisna besedila, ki so jih učenci in dijaki slovenskih šol tvorili pri pouku. V precejšnjem delu besedil so posebej označene tudi jezikovne napake, ki so jih v spisih **popravili učitelji**. Po slednjih lahko s pomočjo specializiranega konkordančnika tudi iščemo. Korpus vsebuje približno **milijon besed**, namenjen je raziskavam šolske pisne produkcije oz. jezikovne zmožnosti šolarjev in pripravi učnih gradiv.

**Janes** je korpus spletne slovenščine. Vsebuje besedila, ki so jih na spletu tvorili uporabniki, in sicer tvite, forumska sporočila, blogovske zapise in komentarje spletnih novic. Korpus obsega okrog **134 milijonov** besed. Namenjen je raziskovanju nestandardne spletne slovenščine. Korpus je eden od rezultatov nacionalnega raziskovalnega projekta *Jezikoslovna analiza nestandardne slovenščine* (J6—6842), ki poteka med leti 2014 in 2017, v njegovem sklopu pa je organiziran tudi naš poletni tabor.

## 1.2.1 Od konkordance do kolokacije – prvi del

### 1.2.1.1 Korpus KRES

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

1. Odpremo korpus Kres in vtipkamo v iskalno okence besedo *pljuvalnik*. Ogledamo si rezultate v konkordančniku in spoznamo:

- kaj je **konkordanca** oz. **konkordančni niz**, **konkordančno jedro**,
- kje najdemo število konkordanc,
- kako pridemo do širšega **sobesedila**, **metapodatkov o besedilu** in **korpusnih oznak**,
- kaj so **filtri** in kako jih uporabljamo.

2. Raziščemo, kaj pomeni beseda *pljuvalnik*. Kaj pomeni *brbotalnik*? Poznamo sopomenko za brbotalnik? Kako pogosto in v katerih besedilih se pojavlja *brbotalnik*, kako pogosto pa sopomenka?

3. Katere besede na *-nik* še poznamo? Naštujemo tiste, za katere mislimo, da so v korpusu najpogostejše. Odpremo zavihek Seznam in vnesemo iskalni pogoj *\*nik*. Ogledamo si rezultate. Ogledamo si filter Besedna vrsta in komentiramo, kar najdemo pod Neuvrščeno in Prislov.

4. Ogledamo si besedo *lastnik* in razmislimo o besedni zvezi *lastnik + koga ali česa*. Naštujemo nekaj primerov. Odpremo zavihek Okolica in vnesemo besedo *lastnik* + prvo mesto na desni. Ogledamo si rezultate in spoznamo pojem **kolokator** in različne možnosti razvrščanja le-teh. Kolokatorje filtriramo na samostalnike in uredimo glede na pojavitve v okolici.

5. Kdo najde samostalnik, ki se v korpusu najpogosteje pojavlja? Kot zanimivost si ogledamo besedni oblak z najpogostejšimi samostalniki iz korpusa Kres.

6. Kako blizu so bili naši odgovori korpusnim podatkom? Zakaj se je naša **jezikovna intuicija** v nekaterih primerih izkazala za ustrezno, v drugih ne?

### 1.2.1.2 Korpus GOS

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

1. Odpremo korpus GOS in vtikamo v iskalno okence besedo *recimo*. Ogledamo si rezultate v konkordančniku in spoznamo razlike korpusa GOS glede na Kres:

- možnost poslušanja posnetka (poslušamo nekaj primerov),
- podatki o besedilih so drugačni (*kaj pravte recimo na to vse vi ste tudi podjetnica – ogledamo si prekrivanje lastnih imen in oznako za nerazumljivo*),
- razlike v filtrih, poskušamo filtrirati, npr. *Maribor, nejavni nezasebni*. Kakšne vrste komunikacije se najdejo v tej vrsti oznake?

2. Poiščemo v korpusu besedo *ful* in si ogledamo filtre. Kaj lahko ugotovimo glede tipa govora, regije snemanja, spol, starost ... Lahko glede na podatke zaključimo, da ženske uporabljajo besedo *ful* pogosteje kot moški? Za pomoč pri odločitvi si ogledamo podatke O korpusu.

3. Ponovimo, da je korpus GOS transkribiran na dveh nivojih. Odpremo zavihek Seznam in izberemo Iskanje po standardiziranem zapisu, iščemo besedo *lahko*. Ogledamo si rezultate in vsak poišče obliko in poslušča posnetke, ki so najbližje njegovemu narečju ali ki se mu zdijo najbolj zanimivi.

## 1.2.2 Od konkordance do kolokacije – drugi del

### 1.2.2.1 Korpus Janes

[http://nl.ijs.si/noske/janes04.cgi/first\\_form?corpname=janes.04](http://nl.ijs.si/noske/janes04.cgi/first_form?corpname=janes.04)

1. Odpremo korpus Janes in vnesemo v iskalno okence besedo *valjda*. Ogledamo si rezultate v konkordančniku in spoznamo značilnosti vmesnika:

- konkordančni niz je podoben, na levi imamo nekaj osnovnih **metainformacij** o viru besedila, npr. da gre z tvit in kdo je avtor,
- klik na konkordančno jedro odpre sobesedilo, klik na informacije na levi dodatne metapodatke,
- možnost, da si ogledujemo zadetke v obliki povedi (*Možnosti prikaza > Stavek*),
- možnost, da podatke sortiramo – ogledamo si funkcijo *Premešaj*, ki premeša zadetke, da npr. niso na začetku samo tviti,
- omenimo možnost, da vzorčimo in filtriramo konkordančni niz,
- ogledamo si seznam oblik besede (*Frekvenca > Oblike niza*),
- ogledamo si seznam izvornih dokumentov (*Frekvenca -> Dokumenti*)
- ogledamo si možnosti izdelave seznama kolokatorjev: attribute *word*, v razponu 1 do 1. Uredimo zadetke po frekvenci in se pogovorimo o rezultatih.

### 2. Samostojno delo 1: primerjava besed *neumen* in *glup* v korpusu Janes.

- Raziskovanje poteka v parih.
- Eden od dijakov poišče v korpusu Janes besedo *glup*, drugi pa besedo *neumen*.
- Ugotovita, kako pogosto se vsaka od besed pojavlja in postavita hipotezo, zakaj je tako,
- Izdelata seznam kolokatorjev na mestu desno tik ob besedi in si ogledata prvih 100 (dve strani) rezultatov,
- Primerjata oba seznama kolokatorjev in po potrebi posamične konkordance (s klikom na P pred kolokatorjem): kakšno besedišče se pojavlja? Kako pogosti so kolokatorji?
- Katere ugotovitve lahko sklenemo iz podatkov? So podatki potrdili ali ovrgli hipotezo?
- (V sredo bomo spoznali orodje *Primerjalne skice*, ki olajša tovrstne primerjave med besedami.)



### 3. Samostojno delo 2: različni žanri korpusa Janes.

- Delo poteka v štirih skupinah.
- Vsaka od skupin dobi enega od žanrov: *tweet*, *blogi*, *forumi* ali *komentarji* in v konkordančniku za delo izbere ustrezen podkorpus.
- Izberemo besedo ali besedno zvezo, za katero predvidevamo, da se bo pojavljala v vseh žanrih (npr. *itak*).
- Vsaka skupina v svojem podkorpusu naredi naključen vzorec stotih konkordanc.
- Vsaka skupina pregleda svoje konkordance (stavčni pogled) in zabeleži čim več zanimivosti v zvezi z jezikom v njih, npr. ali se pojavljajo posebni znaki in če da, v kakšni funkciji so, kako se uporabljajo ločila, kako se besede zapisujejo, se pojavljajo tujejezične besede ali zveze, sleng, ali katera od besed dobiva nov pomen, kakšen je odnos avtorjev besedila do vsebine, ki jo sporočajo ...
- Skupine primerjajo rezultate in ugotovimo, katere značilnosti se pojavljajo v različnih žanrih, katere pa so značilne za posamezen žanr.

#### 1.2.2.2 Korpus Šolar

[www.korpus-solar.net/](http://www.korpus-solar.net/)

1. Odpremo korpus Šolar in vnesemo v iskalno okence kot napako besedo *otrok*. Ogledamo si rezultate v konkordančniku in spoznamo razlike vmesnika:

- mogoče je iskati po jezikovnih napakah učencev in popravkih učiteljev,
- napake in popravki so izpisani v konkordančnem nizu in v razširjenem kontekstu,
- možnosti urejanja rezultatov so primerljive, ogledamo si podatke *Oblike niza*, *Dokumenti*.

2. Razmislimo in naštejemo nekaj napak, ki se nam zdijo tipične za šolska besedila. Ogledamo si možnosti iskanja s pomočjo *Oznake napake*, npr. napake na ravni besedišča, če je čas, še seznam *Oblike niza* in posamezne primer (npr. *in*, *Hamlet*).

Za obnovitev znanja in dodatne ideje glede uporabe korpusov lahko obiščete spletno stran **Portal jezikovnih virov (viri.trojina.si)**. *Janes* sicer še ni predstavljen, so pa na voljo videopredstavitve korpusov *Gigafida* (ki ima enak konkordančnik kot *Kres*), *GOS* in *Šolar*.

Damjan Popič

## 2 STANDARDNA IN SPLETNA SLOVENŠČINA

### 2.1 Standardna in spletna slovenščina – izročki

JANES

IJS  
FILOZOFSKA  
FAKULTETA

# Standardna in spletna slovenščina – od pravopisa do tvitopisa

Damjan Popič

Ljubljana, 5. 7. 2016

JANES

IJS  
FILOZOFSKA  
FAKULTETA

## #KonecSlovenščine

Radio Salomon @RadioSalomon · 6 hrs

it's time 4 deejaytime...kdor se javi za prvo  
trombeto koketo naj pritisne like... [fb.me  
/25tDTkU7z](https://www.facebook.com/25tDTkU7z)

## #KonecSlovenščine



KOMPAS @Kompas\_1951 · 1 Std.

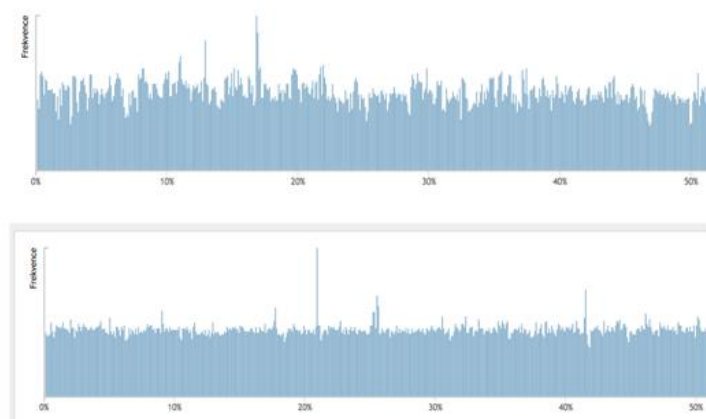
Biseri Mediterana s Costo Fascinoso.  
Neapelj, Savona, Barcelona, Valencia,  
Palma de Mallorca, Palermo [bit.ly/25dRmyy](https://bit.ly/25dRmyy) @costacrueros



## #VejicaJePaNi

	Tviti	Standardna slovenščina
št. vejic	1.866.932	1.474.388
št. vejic/milijon besed	87.998,50	69.495,80
št. pojavnic/stavek	15,88	15,88
št. vejic/stavek	1,40	1,10

## #PosejanostVejic



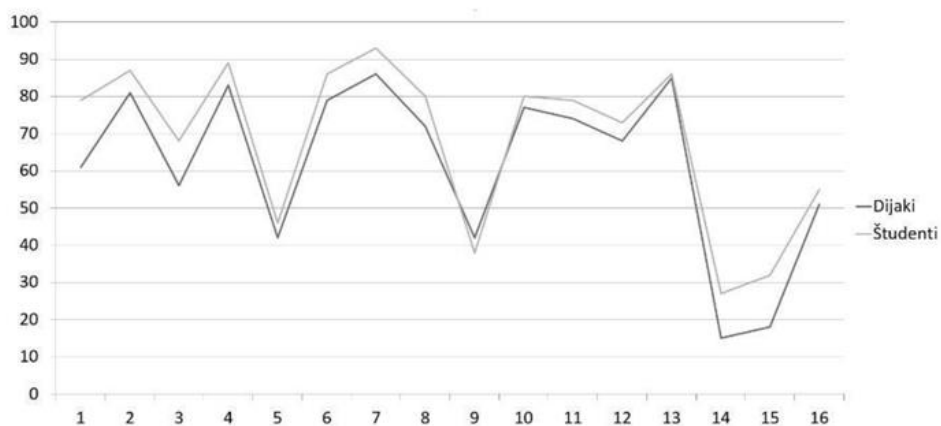
## #KiKoKer

Beseda	Tviti	Beseda	Standard. slov.
da	312.924	da	221.397
ki	160.422	ki	146.763
biti	88.034	biti	78.676
ko	51.247	ko	32.639
če	52.007	če	22.977
ker	42.757	kar	21.441
pa	46.149	kot	20.228
ampak	29.937	se	23.261
ne	26.310	"	61.017
a	22.707	ker	17.402

## #Lektor

Tip	# napak
Slog	8909
Oblika	1397
Pravopis	14.008
Skladnja	4462
Pragmatika	1477

## #Šolar



## (Za)kaj je jezikovni standard?



N. V., 4 leta: MAORICA

(Vir: Stabej 2012a)

## (Za)kaj je jezikovni standard?

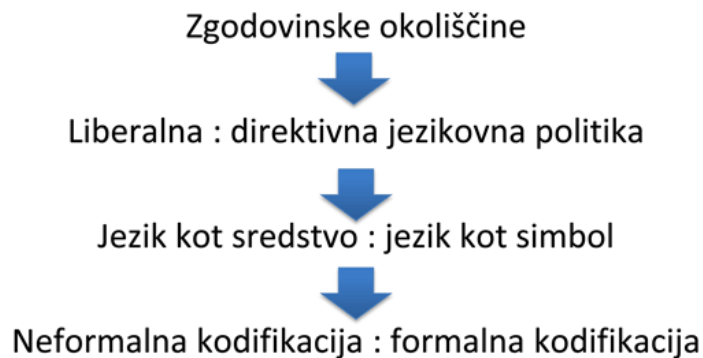
*Gasilec je oseba, ki svoj poklic prinaša tudi domol.*

N. N., 22 let, kandidat za poklicnega gasilca

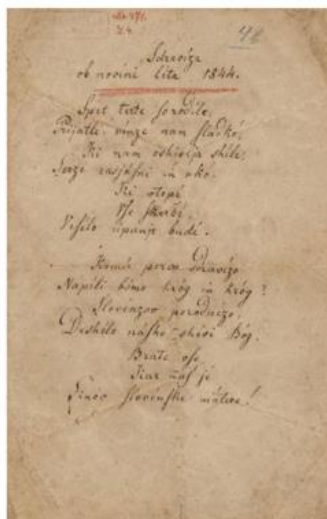
(Vir: Stabej 2012a)



## Kako se rodi jezikovni standard



## Kam torej spada slovenska situacija?



(Vir: Wikipedija)



## Kateri standard?

ABDE € FGH HI I Z JKL L MNNO Ø PRS W 4 3 X TUV 7 4

ABCDEFGHIJKLMNNO PRS 8 ZXTUYV 4

ABC Č DEFGHIJKLMN O PRS Š TUV Z Ž

(Vir: Wikipedija)

## Povest o dobrih ljudeh

Če si hočemo predstavljati vznik svoje narodnosti, si ga moramo nekako takole: ker so Slovenci prišli prvi izmed južnih Slovanov v svoja sedanja bivališča, je gotovo, da so prišli vanja kot prva straža slovanske mase, morda in najbrže kot skupina najsorodnejših in najbližjih rodov. Prišli so v naše sedanje pokrajine kot Sloveni, to se pravi kot nediferencirani člani prvotne velike slovanske skupnosti. Naselili so se v zelo goratem in izoliranem svetu, ki jih je odrezal od Slovanov na jugu in na vzhodu. Živeč v Alpah, v katerih ni živel noben drug slovanski rod, v posebnem podnebjju, na zemlji, ki hrani drugače in ki ima drugačen geološki in oblikovni značaj, in še v drugem političnem in kulturnem svetu kakor bratski rodovi na jugovzhodu, so se polagoma razvili v posebno, zaključeno enoto, ki je ostala v rahlih stikih s sosednimi Hrvati samo na vzhodni periferiji. Vse te naravne, usodne in zgodovinske moči so v stoletjih izpremenile in preoblikovale prvotni značaj in jezik tega ljudstva drugače nego so pretvarjale druge moči njihove slovanske sosedne in brate. Tako je iz njih polagoma nastala zaključena narodna enota Slovencev, ki se je določno zavedela svoje samobitnosti in svojega imena najkeseje že v šestnajstem stoletju, in sicer tudi svoje samobitnosti v razmerju do drugih južnih Slovanov.

Josip Vidmar (1932: 21) cit. po Stabej (2012b: 15)

## Povest o dobrem pastirju

Seveda je slovnica prav dobro sredstvo za ohranitev jezika. Naš tako plemeniti, tako koristni, tako sveti jezik je v naši vojvodini Kranjski v zadnjih letih tako opešal, da se česa takega ne bi mogli spomniti niti najstarejši Kranjci. Kaj je tega krivo? Eden od vzrokov je ta, da ni bilo nobenih pravil, nobene slovnice, kjer bi se dalo poiskati in najti, kaj je prav. Mnogi, ali po pravici povedano, vsi so vzdihovali po slovnici. Jaz tudi. Vidimo, da v vseh drugih jezikih skoraj vsako leto pride na svetlo nova slovnica.

[...]

Končno naj povem, da se glede narečja in izgovarjave nisem ravnal po nobeni določeni pokrajini. V tem sledim učenjakom in jezikoslovcem, ki znajo besedo ceniti in sprejemati. Čvrsta šola je tista, ki ima zanesljiva pravila. Drhal govori po mili volji, kakor se ji zdi, zato kvari jezik. Zaradi tega pa se ni treba bati, da preprost človek ne bo sprejel dobre in pravilne besede ali je ne bo razumel.

Marko Pohlin: Kraynska grammatika, das ist: Die kraynerische Grammatik, oder Kunst die kraynerische Sprache regelrichtig zu reden, und zu schreiben (cit. po Stabej (2012b: 17))

## Povest o dobrem pastirju

Pri svojem delu sem izhajal iz primerjave vseh slovanskih narečij, kar je po mojem prepričanju edino pravo izhodišče. Po eni strani omogoča spoznati resnične značilnosti vsakega izmed njih, po drugi strani pa nas obvaruje pred prezrtjem ali celo občudovanjem nekaterih poganjkov, ki so posledica tujega vpliva in jih je treba iztrebiti, čeprav jih ljudstvo uporablja, saj nasprotujejo slovanskim jezikovnim zakonom (Cigaletov nemško-slovenski slovar (1860).

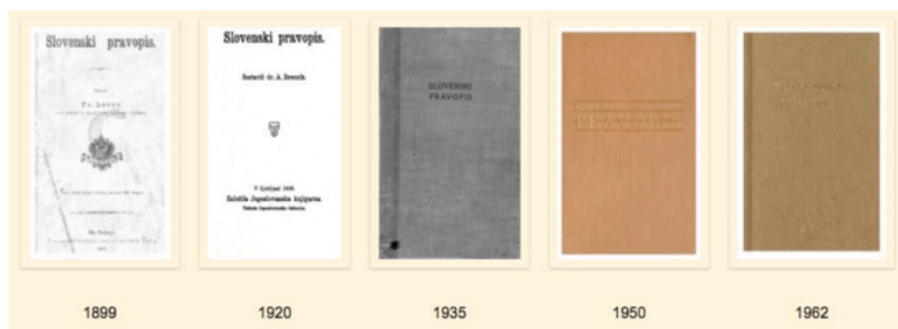
(cit. po Stabej 2012b: 18)

## Povest o dobrem pastirju

Iz preteklosti se učimo za prihodnost. Slovensko jezikoslovje je imelo v vsakem obdobju **jezikoslovca vodnika**, ki je znal in zmozel pozitivno usmerjati razvoj slovenskega jezika, imelo pa je tudi ljudi, ki so tem vodnikom sledili in skrbeli, da se je slovenistično jezikoslovno vedenje **pravilno** in na razumljiv način posredovalo učencem, študentom in vsem zainteresiranim. Ne enih ne drugih med Slovenci nikoli ni bilo v izobilju, pa vendar dovolj, da se je **slovenski jezik razvijal dobro in v pravo smer** (Jesenšek 2012; poudaril D. P.).

(Stabej 2012b: 19; poudaril D. P.)

## Knjiga resnice



(Vir: <http://www.trojina.org/pravopisi/>)

## Knjiga resnice

**\*izglédati -am: \*dobro izgledaš imeniten, zdrav si videti; \*slabo izgledaš slab si videti; \*kako pa izgledaš kakšen pa si; \*izgleda, da bo dež na dež kaže; kaže, podobno je, videti je, zdi se, da bo dež; \*kako bo pa to izgledalo kakšno pa bo to, čému bo pa to podobno, kazno**

(Slovenski pravopis 1962)

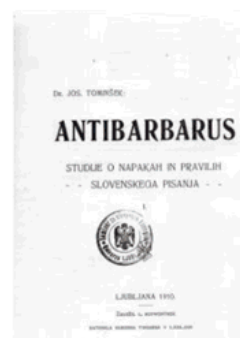
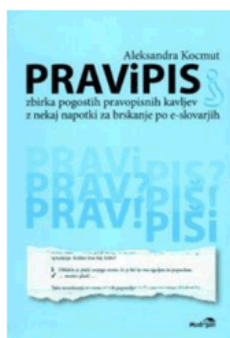
## Tisti, ki bere, je ...

1. *-avec, -ivec, -avka, -ivka*. Iz nedovršnih glagolov ali iz njih glagolskih pridevnikov izpeljani samostalniki, ki zaznamujejo delujoča bitja (osebe in živali), se pišejo s pripono *-avec, -ivec* za moški spol, *-avka, -ivka* za ženski spol:

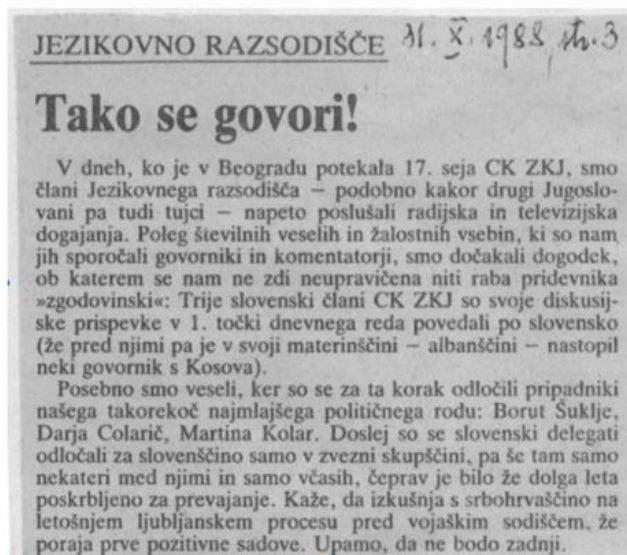
a) *bahavec, bravec, bravka, čakavec, česavka, dajavec, dajavka, darovavka, delavec, delavka, dremavec, glasovavec, glodavec, glodavka, gospodovavka, hiravec, igravec, igravka, izdajavec, izdajavka, jecavec, jecavka, kimavec, kimavka, klavec, kljubovavka, kmetovavec, kopavec, kopavka, kujavec, letavec, majavec, metavec, moledovavka, ocenjevavka, očiščevavec, odlašavec, odlašavka, opazovavka, pohajkovavec, poročevavka, posredovavka, prebivavec, pregledovavka, sejavec, sejavka, sesavec, sestavljavac, skakavec, smejavec, stradavec, svetovavka, škripavec, zajedavec, zalezovavka, žvižgavec;*

(Slovenski pravopis 1962)

## Knjigae resnice



## Jezikovno razsodišče





## Jezikovni standard



(Vir: <http://www.pravopis.si/Zgodovina/tabid/82/Default.aspx>)



## Jezikovni standard



(Vir: Stabej 2012b)

## Jezikovni standard



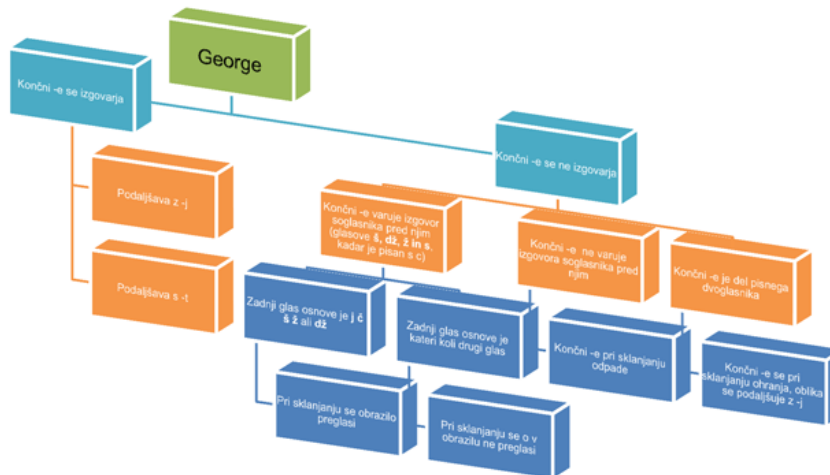
(Vir: Stabej 2012a)

## Jezikovni standard

**izdátek** -tka m (â) To je bil zanj velik ~; mesečni, tekoči ~i;  
zmanjšati ~e za kaj; •~i po osebi *izdatki na/za osebo*

(Slovenski pravopis 2001)

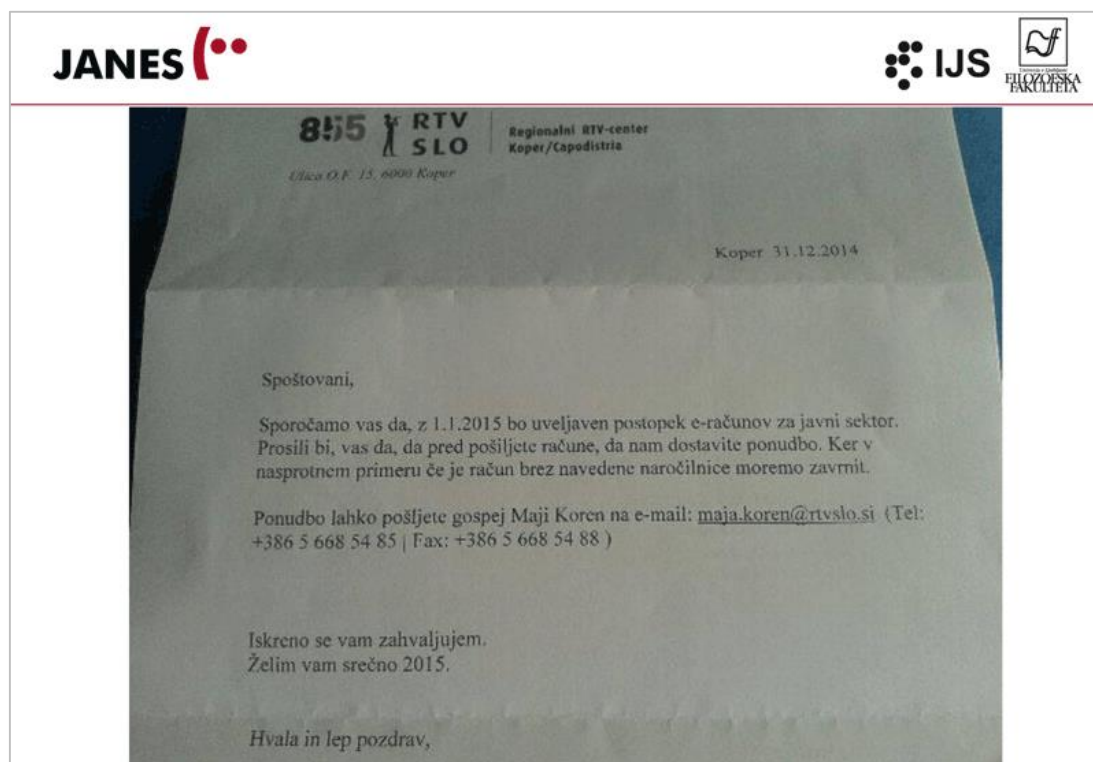
## Jezikovni standard



## Radikalizacija

Vsak državljan Republike Slovenije bi moral biti tudi govorec slovenskega knjižnega jezika, ki je jezikovna zvrst za izražanje najzahtevnejših vsebin v slovenskem jeziku in imajo tudi uradni značaj. Šele na ta način se lahko tudi celovito vključi v na jeziku temelječo slovensko družbo (Tivadar in Tivadar 2015: 43).





**JANES**

**IJS** FILOZOFSKA  
FAKULTETA

### Obvestilo vodstva RTV Slovenija

Odziv vodstva RTV na dopis z obvestilom o prehodu na elektronski račun  
6. januar 2015 ob 10:49  
Ljubljana - RTV SLO

**Radiotelevizija Slovenija - Regionalni RTV-center Koper/Capodistria je dne 31.12.2014 desetim podjetjem posredoval dopis z obvestilom o prehodu na elektronski račun. Ti so od začetka leta 2015 obvezni za vse proračunske uporabnike. Dopis je vseboval več slovničnih napak in ni bil napisan na način, primeren za uradni dopis. Fotografija obvestila, ki je zajela samo njegov del, je od petka, 2. januarja 2015, zaokrožila po socialnih omrežjih in sprožila precejšnje odzive.**

Po seznanitvi z vsemi okoliščinami smo ugotovili, da je obvestilo pripravila, podpisala in poslala javna uslužbenka, ki nima pooblastila za pošiljanje takšnih obvestil v imenu Regionalnega RTV-centra Koper/Capodistria. Z dopisom je želela preprečiti, da bi bili računi po novem letu v sistemu e-računov zavrnjeni. Iz fotografije dela obvestila je mogoče potegniti tudi popolnoma napačen zaključek, da je obvestilo pripravila Maja Koren, kar nikakor ne drži.

Za Radiotelevizijo Slovenija je takšno ravnanje nesprejemljivo, zato bo proti javni uslužbenki primerno ukrepala, prejemnikom obvestila pa posredovala nov dokument in opravičilo.

Radiotelevizija Slovenija, javni zavod  
mag. Marko Filli, generalni direktor

(Vir: <http://www.rtv slo.si/sporocila-za-javnost/obvestilo-vodstva-rtv-slovenija/355138>; podčrtal D. P.)

Spoštovani gospod varuh!

Gledal sem tisto nočno oddajo o slovenskem jeziku in se prepričal, da nimamo več prave jezikovne avtoritete, ki bi opozarjale na napake in imela neki vpliv na razvoj jezika. Stabej, ki enkrat v pogovoru celo dvojine ni uporabil in tista njegova sodelavka gledata po mojem mnenju preveč liberalno, da ne rečem brezbrizno na razvoj slovenščine. Nekdo bi lahko opozarjal javnost, da besede, ki jih pogosto slišimo kot naprimer valda, važi, kul, ful itd. niso slovenske. Napisi po večini gostinskih lokalih so prav po hlapčevsko v tujem jeziku kot coffee to go, take away. Zanimivost pa sem opazil v koprski slaščičarni, last Albanca, ki ima lepo napisano kava za s seboj. Zgled za slovenske hlapce. V Ljubljani sem videl velik napis pred lekarno pri rehabilitacijskim centru Soča DRIVE in lekarna, potem na vogalu Miklošičeve in mislim Dalmatinove ulice je trgovina, ki ima samo angleški napis. O tem sem obvestil inšpektorja na kulturnem ministrstvu, ki je sicer šel tja pogledat, vendar se je izgovarjal na predpise in ni ukrepal ničesar. Škoda, da vaša televizija nima več jezikovne oddaje, ki bi opozarjala na vse jezikovne anomalije in mogoče prebudila speče strokovnjake, da bi spet dobili kakšnega Gradišnika ali Toporišiča.

Lep pozdrav!

R. O.

(Vir: <http://www.rtv slo.si/varuh/pritozba/957>; podčrtal D. P.)

Pozdravljeni!

Malo spremljam oddajo, 6.2.2015, Kam greš Slovenščina, na kateri ste udeleženi tudi vi. Zamerim vam to, da se premalo odločno postavljate za svoje stališče, ki ga popolnoma podpiram. Seveda ne morem verjeti, da voditelj oddaje pusti toliko govoriti onemu Kekcu z očali, ki očitno uveljavlja svoj ego s tem, da govori in govori, kot, da je na pikniku in ima okoli sebe butce. KEKEC, za mene popolni KEKEC, ki nima kaj iskati v taki pogovorni oddaji, je pa neki doktor. No, naj bo doktor, za mene je KEKEC!!!! Toliko! Pa lep pozdrav!

J. M.

(Vir: <http://www.rtv slo.si/varuh/pritozba/957>)

## R.I.P.

Moj žalostni sklep je, da se je razpadanje slovenščine v javni rabi že začelo. Nobene resnične poti ne vidim, po kateri bi tak razvoj preprečili. Slovenščina je že v odhajanju in z njo začenja žal odhajati v preteklost tudi slovenski narod. Slovenskega naroda čez 100–150 let ne bo več in tako bo izginil tudi konstitutivni razlog za obstoj države, v kateri bi bila slovenščina državni jezik. Moja bolečina je velika. Uboga naša materinščina (Orešnik 2010: 147 cit. po Stabej 2012b: 20).



## Kako napisati tole?

<http://www.lektorsko-drustvo.si/lektorsko-brbotanje/pisati-kranjska-ali-kranjska-klobasa>

## Kako naprej?



## Kaj pa je potem spletna slovenščina?



Jan Tomše  
@JTomse11

Follow

Sprememba Facebook statusov pri nadzornikih SDH: "It's complicated!"  
[#Telekom](#) [#privatizacija](#)

View translation

12:38 p.m. - 10 Jun 2015



## Značilnosti "nove" slovenščine

Ugotovitev »večine razprav, ki se skoraj praviloma sklicujejo tudi na ugotovitve tujega jezikoslovja, je, da je jezik, uporabljen v novomedijskih besedilih, slogovno poln posebnosti, kot so **kratkost sporočil, okrajšanost besed in besednih zvez ter povedi, izpustnost in neformalnost**« (Dobrovoljc 2008: 297; poudaril D. P.).

+ Neskladnost s pravopisom ...

## In kaj to pomeni?



Vir: Allianz.com

## Reference

Dobrovoljc, Helena, 2008: Jezik v e-poštnih sporočilih in vprašanja sodobne normativistike. Košuta, Miran (ur.): *Slovenščina med kulturami*. Zbornik Slavističnega društva Slovenije 19. Celovec: Slavistično društvo Slovenije. 295–314.

Jesenšek, Marko (ur.), 2012: *Večno mladi Htinj*. Ob 80-letnici Janka Čara. Maribor: Mednarodna založba Oddelka za slovanske jezike in književnosti, Filozofska fakulteta.

Orešnik, Janez, 2010: Dvoje iz jezikoslovja. Nečak, Dušan (ur.): *Pogledi, Humanistika in družboslovje v prostoru in času*. Ljubljana: Znanstvena založba Filozofske fakultete. 130–147.

Pohlin, Marko, 2003: *Kraynska grammatika. Bibliotheca carnioliae*. Znanstvenokritična izdaja. Ljubljana: Založba ZRC, ZRC SAZU.

Stabej, Marko, 2012a: *Resolucija: renovacija ali revolucija*. Predavanje na konferenci Pravopis, več kot le predpis. 15. 2. 2012. Dostopno na [http://videlectures.net/pravopis2012\\_stabej\\_resolucija/](http://videlectures.net/pravopis2012_stabej_resolucija/) (dostop 22. 8. 2015).

Stabej, Marko, 2012b: Jezik, nazori in nadzor. Bjelčević, Aleksander (ur.): *Ideologije v slovenskem jeziku, literaturi in kulturi* (48. seminar slovenskega jezika, literature in kulture). Ljubljana: Filozofska fakulteta, Oddelek za slovenistiko, Center za slovenščino kot drugi/tuji jezik. 11–20. Dostopno na <http://www.centerslo.net/files/file/ssilk/48%20SSILK/stabej.pdf> (dostop 22. 8. 2015).

Tivadar, Gorazd in Hotimir Tivadar, 2015: Problematika snemanja in zapisovanja govornih besedil na slovenskih sodiščih (na primeru sojenja na okrajnem sodišču v severovzhodni Sloveniji). Tivadar, Hotimir (ur.): *Država in narod v slovenskem jeziku, literaturi in kulturi* (51. seminar slovenskega jezika, literature in kulture). Ljubljana: Znanstvena založba FF UL. 43–50. Dostopno na [http://www.centerslo.net/files/file/ssilk/51\\_SSILK/Tivadar\\_Tivadar.pdf](http://www.centerslo.net/files/file/ssilk/51_SSILK/Tivadar_Tivadar.pdf) (dostop 22. 8. 2015).

Vidmar, Josip, 1932: *Kulturni problem slovenstva*. Ljubljana: Tiskovna zadruga.

## 2.2 Od pravopisa do tvitopisa – vaje

### 2.2.1 Pregibanje osebnih lastnih imen

#### a) Tvorite roditeljski in svojilni pridevnik

George –	Broadway –
Kreisky –	Dumas –
Shakespeare –	Andrew –
Hradetzky –	Crusoe –
Kaye –	Mike –
Chomsky –	Makarios –
Wilde –	Lukács –
García –	Boccaccio –
Wallace –	Nixon –
Lamartine –	Descartes –
Holmes –	Poe –
Jacques –	May –
Rousseau –	Manet –
Zola –	Sydney –
Tesnière –	Bordeaux –
Disney –	

#### b) Lastna imena v oklepajih postavite v ustrezno obliko.

Boji v polfinalu so bili bolj izenačeni, kot so pričakovali. Presenečenje je bil poraz \_\_\_\_\_ (Daniela Hantuchova) proti \_\_\_\_\_ (Francesca Schiavone). Pri moških omenimo poraz najboljšega slovenskega teniškega igralca \_\_\_\_\_ (Marko Tkalec) z Nemcem \_\_\_\_\_ (Waske).

Znan je največji poraženec uvodnega dela tekmovanja Lige prvakov, \_\_\_\_\_ (München – prid.) Bayern. Za uvrstitev v 2. krog so vrata še vedno odprta \_\_\_\_\_ (švicarski Basel), \_\_\_\_\_ (francoski Lens) in \_\_\_\_\_ (grški AEK).

### 2.2.2 E-tvorjenke

iPhone (tudi iP4, iP5 in iP6), iPad, iOS, iTunes, iPod, iMac, iCloud, iStore, iMessage, iWatch, iStyle, iSpot, iHelp, iBooks, iWork, iMovie, iDrive, iPhoto, iP, iPlayer, iLife, iCal, iLoop, iBookstore, iGlove, iStvar, iBand, iCalendar, iSight, iTouch, iWeb, iFlicks, iMaps, iDevice, iUser, iRadar, iSteve

iNaprava, iTelefon, iNapravica, iZadeva, iPriročnik, iTrgovina, iAplikacija pa tudi iCrap, iDiot, iGrača, iFanatik, iFan, iNeki in sklanjanje besede iPhone kot iPhoneta/iPhonetov

#### 10 najpogostejših e-tvorjenk

Beseda / besedna družina	Pojavitve
e-mail, e-mailing, e-majlirati	1699
e-pošta, e-poštni	722
e-knjiga, e-knjigica	420
e-naslov	201
e-novice, e-novičke	248
e-volitve	242
e-račun	181
e-trgovina, e-trgovinica	151
e-uprava, e-upravljanje	140
e-fotograf, e-fotografija	114
e-bančništvo, e-banka	104

#### 10 najpogostejših e-tvorjenk s področja avtomobilizma

Beseda	Pojavitve
dCi	470
S-Max	375
C-Max	197
S-line	184
T-Jet	146
S-Class / s-klasa	98
E-Class / e-klasa	85
X-Trail	84
C-Elysée	73
B-Max	68

#### 10 najpogostejših imen produktov

Beseda	Pojavitve
B-complex / B-kompleks	78
iGO	68
X-treme	57
e-Hdi	34
uTorrent	32



<b>V-lube</b>	32
<b>iRobot</b>	29
<b>nVidia</b>	24
<b>x-box / xBox<sup>1</sup></b>	16
<b>X-TRM</b>	13
<b>X-Reality</b>	12
<b>N-Gage</b>	11

### 2.2.3 Stvarna lastna imena

združeni narodi  
evropska unija  
vlada republike slovenije  
mestno gledališče ljubljansko  
druga svetovna vojna  
lekadol  
silvestrovo  
vlak novomeščan  
fordov novi mondeo  
vozim se s fordov  
liberalna stranka  
mestna hiša  
slovenska ustava

Predsednik slovenske vlade se je sestel s predstavniki ministrstva za notranje zadeve/notranjega ministrstva.

Fotografiram z nikonom.  
Komarje uničujem s pipsom, faraonke pa s faracidom.

dan državnosti, prešernov dan, silvestrovo, pust, novo leto, božič

aspirin, lekadol, apavrin

musliman  
južni slovani  
avto renault clio storia  
večernica (nagrada)

---

<sup>1</sup> Pri slednji besedi gre za napačen zapis imena igralne konzole Xbox.

gostilna pri mraku  
zlato vegovo priznanje  
telefon nokia  
ohmov zakon  
sizifovo delo  
uradno je bila zadnja odprava na luno apollo 17.

## 2.2.4 Kratice in simboli

### a) Zapišite roditeljsko obliko.

SSKJ	FidaPLUS	Nato
NUK	SMS	FBI
BTC	FF	Unesco
Fida	FDV	IBM

### b) Okrajšajte.

tega leta	lastnoročno	literarnozgodovinski
družba z omejeno	rimskokatoliški	in tako dalje
odgovornostjo	na primer	to je
tako imenovani	italijansko-nemški	

### c) Označite presledke.

18kg	ø1,5mm	800kg/m <sup>3</sup>
20°C	5'10"	300€
50%	§290	100N
30x40 cm	45°	

### d) Popravite, če je to potrebno.

<i>GSM-aparat</i>	<i>A-skupina (skupina A)</i>
<i>TV-sprejemnik</i>	<i>7.b razred</i>
<i>90-letnica</i>	<i>DNK test</i>
<i>G-dur</i>	<i>test DNK</i>
<i>C-vitamin (vitamin C)</i>	<i>20 do 30 odstotno povišanje cen</i>

### e) Označite vse pravilne zapise.

TV-hiša	TV Slovenija	HD-televizor	HD TV	TV program	HD-TV
---------	--------------	--------------	-------	------------	-------

## 2.2.5 Zapis tvorjenk

### a) Popravite zapis, kjer je to potrebno.

vitamin C, program TV, legokocke, žalbeseda, pedenjčlovek, temperabarve, dumdumka, dumdumkrogla, tele-objektiv, brutoteža, Josipdol, Slovenija les, rendez-vous, fin de siecle, East River, človek žaba, možbeseda, Cankar-dramatik, Schleswig-Holstein, angloamerišvo, mini golf, mini golfovski, super bencin, nevojak, nekovina, antidelec, nebodigatreba, Sun Jatsen, črnobelost, zlatolaska, sreče-lov, avtogaraža, rdečebela zastava, rjavo ok, vrvohodec, tritisočak, prelep, koreferent, mso-podpisnik, TV Slovenija, jugovzhod, generalpolkovnik, TV hiša, A diplomska naloga, sladkokisel, vitamin C, naravoslovnokulturni spomenik, SMS sporočilo, G-dur, kolovoz, elektro oprema, avto šola, e pošta, polčas, tenis igrišče, zimsko športni, 4. c razred, tempera barva, Ljubljana-Center, frizersko brivski salon, protidelec, črnobel, Korošec Dularjev učbenik, velikonočni, čokoladno rjav, modrikasto bel, temno zelen, bledorjav, častilakomen, novo poročenec, TV Dober dan, Avstro-Ogrska, avstro-ogrskost, italijanskonemški, Anglo-Američan, tule, letu, kdor koli, slej ko prej, sem ter tja, ravno tako, čim bolj, čimveč, natanko, od zunaj, nadrobno, nasvidenje, avto moto zveza, rent-a-car služba

fitnes center / fitnesscenter / center za fitnes

fitnes oprema / fitnessoprema / oprema za fitnes

tenis igrišče / teniško igrišče / igrišče za tenis

golf klub / golfklub

mini golf / minigolf

ananasliker / ananas liker / ananasov liker

diskoklub / disko klub

džezbalet / džez balet

brutoteža / bruto teža

popzvezda / pop zvezda

kavabar / kava bar

Škofjeloški pasijon

državnozbornske volitve

lesnoindustrijski obrat

premoženjskopravni postopki

svetovnonazorski

socialnovarstveni dodatek

vodnogospodarski

državnupravni postopki

triinpolodstotna podražitev

tisočpetstokrat

termoelektrarna

visokofrekvenčni

vojaškoobrambni

psihofizični razvoj otroka

bližnjevzhodni mirovni sporazum  
narodnoosvobodilni boj  
petdesetodstotni delež  
devetdesetletnica  
afroameriška glasba

svetopisemska zgodba  
južnoameriški  
živosrebrn

## 2.3 Od pravopisa do tvitopisa – delavnica


Dijaki delajo samostojno, predavatelj in asistent sta na voljo za pomoč in sugestije. Vsaka skupina si izbere enega od vsebinskih sklopov, ki smo ga spoznali pri vajah:

1. Pregibanje osebnih lastnih imen
2. E-tvorjenke<sup>2</sup>
3. Stvarno lastno ali občno poimenovanje?
4. Kratice in simboli
5. Zapis tvorjenk

Vsaka skupina si v okviru izbrane tematike izbere ustrezno tematiko in pripravi besedilo z naslednjo strukturo:

- I. Opredelitev naloge: raziskovalno vprašanje in teze.
- II. Opis poteka raziskave: utemeljitev izbora vsebinskega sklopa, izbor gradiva oz. korpusov in orodja za njihovo analizo in opis metodologije.
- III. Raziskava in beleženje ugotovitev.
- IV. Zaključki glede na zastavljeno raziskovalno vprašanje.

Nekaj posebnih operatorjev v jeziku CQL<sup>3</sup>:

operator	pomen	primer iskanja	opis iskanja
.	katerikoli znak	[tag="So..d"]	poišči vse občne samostalnike v dajalniku
 (znotraj besede)	katerakoli znak znotraj oglatih oklepajev	[lemma="miš [eo] lovka"]	poišče vse oblike lem <i>mišelovka</i> in <i>mišolovka</i>
*	0 ali več pojavitev (znaka, črke, besede ...)	[word="tvit.*"]	poišči vse besede, ki se začnejo na tvit- (tudi <i>tvit</i> )
+	1 ali več pojavitev	[word="..+ost"]	poišči vse besede, ki se končajo na -ost (brez <i>ost</i> )
?	0 ali 1 pojavitev	[word="Kleme?n.*"]	poišči besede, ki se začnejo na Klemen- ali Klemn-
	disjunkcija ("ali")	[lemma="(blogger blogar)"]	poišči vse oblike lem <i>blogger</i> in <i>blogar</i>
!	negacija ("ne")	[tag="S.*" & word!="@.*"]	poišči vse samostalnike, ki se ne začnejo z znakom @
(?i)	poljubna velikost črk	[word="( ?i)ozn"]	poišči vse besedne oblike ozn, ne glede na zapis

<sup>2</sup> [word="(?) [a-zčšž] -. \*"] (e-tvorjenke tipa *č-podstava* in *Č-podstava*) [word="[a-zčšž] [A-ZČŠŽ] . \*"] (e-tvorjenke tipa *čPodstava*)

<sup>3</sup> Pripravila Kaja Dobrovoljc.

operator	pomen	primer iskanja	opis iskanja
[ ]	katerakoli beseda	[word="tistega" ] [ ] [word="dne"]	poišče vse zveze besed <i>tistega</i> in <i>dne</i> , med kateri je vrinjena še ena beseda
{k,n}	interval (med k in n pojavitev znaka, črke, besede)	[tag="P.*"] {3,5} [tag="S.*"] {0,3} – med nič in tri {3} – točno tri {3,} – vsaj tri	poišči samostalnike, pred katerimi je od tri do pet zaporednih pridevnikov
<s> </s>	začetek ali konec stavka	<s> [lemma="živjo"]	stavki, ki se začnejo s pozdravom <i>živjo</i>
within <s/>	išči znotraj stavka	[word="tistega" ] [*] [word="dne" ] <b>within</b> <s/>	v stavku poišči vse zveze besed <i>tistega</i> in <i>dne</i> , med katerima je vrinjeno poljubno število besed

### Spletne povezave:

- Slovenski pravopis 2001 -- slovar: <http://bos.zrc-sazu.si/sp2001.html>
- Pravila: [http://bos.zrc-sazu.si/c/sp/sp2001\\_pravila.pdf](http://bos.zrc-sazu.si/c/sp/sp2001_pravila.pdf)
- Slogovni priročnik: <http://slogovni.slovenscina.eu>
- Sloleks: <http://www.slovenscina.eu/sloleks>
- Besana: <http://besana.amebis.si/pregibanje/>
- Slovar slovenskega knjižnega jezika: <http://bos.zrc-sazu.si/sskj.html>
- ŠUSS (odgovori na jezikovna vprašanja): <http://www2.arnes.si/~lmarus/suss/>
- Pregled slovenskih pravopisov:  
<http://www.pravopis.si/Zgodovina/tabid/82/Default.aspx>
- Jezikovna svetovalnica pri Inštitutu Frana Ramovša: <http://isjfr.zrc-sazu.si/svetovalnica#v>
- Pravopis za francosko govoreča okolja: <http://orthonet.sdv.fr/ž>
- Dodatne pravopisne vaje:  
[https://dl.dropboxusercontent.com/u/13258131/USSJ\\_Dodatne\\_vaje.pdf](https://dl.dropboxusercontent.com/u/13258131/USSJ_Dodatne_vaje.pdf)
- Korpusi: [http://nl.ijs.si/noske/janes.cgi/first\\_form?corpname=janes](http://nl.ijs.si/noske/janes.cgi/first_form?corpname=janes)

Polona Gantar

## 3 FRAZELOGIJA V SPLETNI SLOVENŠČINI

### 3.1 Frazeologija v spletni slovenščini – izročki

JANES

IJS

FILZOLOŠKA  
FAKULTETA

# Frazeologija v spletni slovenščini

tud drugač se da to povedat

Polona Gantar

2. poletni raziskovalni tabor  
spletne slovenščine za gimnazijce JANES  
Ljubljana, 6. 7. 2016

JANES

IJS

FILZOLOŠKA  
FAKULTETA

## Kar lahko storiš danes, ....

Spoznali bomo osnovne pojme, ki se uporabljajo v frazeologiji, različne tipe večbesednih leksikalnih enot in njihove lastnosti. Spoznali bomo kako nam korpusi lahko pomagajo pri preučevanju besedne povezovalnosti in kako se prek frazeologije kažejo značilnosti slovenske družbe

- Kaj je frazeologija
- Kaj so frazeološke enote
- Kako uporabljamo korpusse pri frazeoloških raziskavah
- Kako in zakaj frazeologijo uporabljamo v jeziku

**JANES** pametni telefon

**Kaj je frazeologija?**

prava čokolada  
gosta megla  
topla greda  
varnostni trikotnik  
frazemi  
jabolko od ...  
drev esa  
volk iz gozda  
maslo ...  
vrabec  
kozol  
iz  
zelja  
črna ovca

kolokacije  
naravna jezikovna raba  
terminologija  
pregovori  
ustvarjalnost drugačnost

**sobese dilo...**

korpusni viri in orodja  
slovarji, priročniki

IJS FILOZOFSKA  
FAKULTETA

**JANES** Kar se Janezek nauči, ... to Janez zna

Da bi bila naša besedila natančnejša ali bolj živa, besede povezujemo v **besedne zveze**. Nekatero tvorimo sproti, medtem ko pišemo ali govorimo. To so **proste besedne zveze**.

Druge besede uporabljamo vedno v enaki zvezi, zato jih imenujemo **stalne besedne zveze**. Te so dveh vrst. Nekatero uporabljamo v prvotnem pomenu, saj pomenijo enako kot besede, iz katerih so sestavljene (npr. *prometna nesreča*). Pomena drugih pa ne moremo razbrati iz pomena besed iz katerih so sestavljene, saj pomenijo nekaj čisto drugega. Te stalne besedne zveze so uporabljene v prenesenem pomenu in jih imenujemo **frazemi** (npr. *iti na nož, priti na svoj račun*). Uporabljamo jih v vsakodnevni pogovori, pogoste pa so tudi v besedilih novinarjev ali oglasih.

Slovenščina 8, Čuden et al. 2005, str. 49



Da bi bila naša besedila natančnejša ali bolj živa, besede povezujemo v **besedne zveze**. Nekatero tvorimo sproti, medtem ko pišemo ali govorimo. To so **proste besedne zveze**.

Druge besede uporabljamo vedno v enaki zvezi, zato jih imenujemo **stalne besedne zveze**. Te so dveh vrst. Nekatero uporabljamo v prvotnem pomenu, saj pomenijo enako kot besede, iz katerih so sestavljene (npr. *prometna nesreča*). Pomena drugih pa ne moremo razbrati iz pomena besed iz katerih so sestavljene, saj pomenijo nekaj čisto drugega. Te stalne besedne zveze so uporabljene v prenesenem pomenu in jih imenujemo **frazemi** (npr. *iti na nož, priti na svoj račun*). Uporabljamo jih v vsakodnevni pogovori, pogoste pa so tudi v besedilih novinarjev ali oglasih.

• ENOTE

**besedne zveze**

**proste besedne zveze**

**stalne besedne zveze**

**frazemi**

• LASTNOSTI

natančnejše bolj živo izražanje

v prvotnem pomenu

tvorimo sproti

vedno v enaki zvezi

pomenijo nekaj čisto drugega

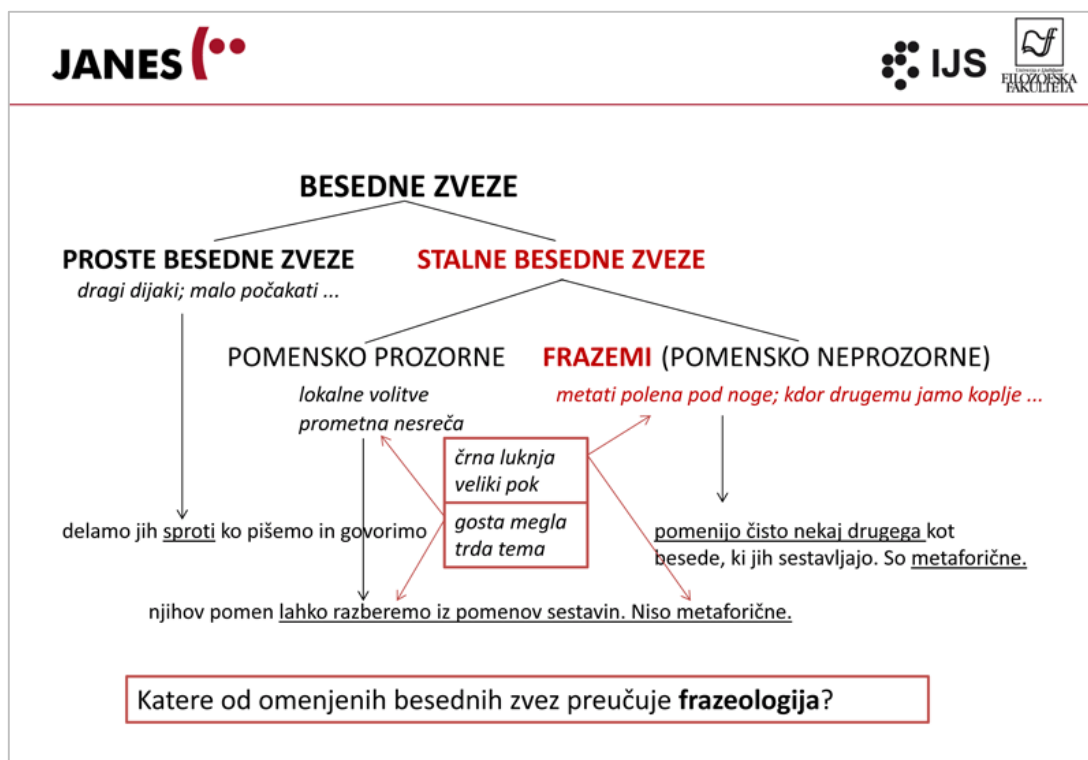
v prenesenem pomenu

v besedilih novinarjev ali oglasih

**Dragi dijaki, lepo pozdravljeni** na **taboru spletne slovenščine**.

**Drage dijake, lepo pozdravljeni** na **tabora spletne slovenščine**.

**Pečeni dijaki, občutno pozdravljeni** na **krožniku spletne slovenščine**.



## Ključne lastnosti

- **večbesednost** besedne zveze
- **prostost : ustaljenost** Nekatero besede uporabljamo vedno v enaki zvezi, zato jih imenujemo stalne besedne zveze za razliko od prostih zvez.
- **pomenska prozornost in neprozornost ter preneseni pomen** Nekatero besedne zveze uporabljamo v prvotnem pomenu, saj pomenijo enako kot besede, iz katerih so sestavljene, pomena drugih pa ne moremo razbrati iz pomena besed, iz katerih so sestavljene. Stalne besedne zveze so uporabljene v prenesenem pomenu.
- **živost, drugačnost, ekspresivnost** Da bi bila naša besedila natančnejša ali bolj živa, besede povezujemo v besedne zveze.

## Vsak začetek je težak

- beseda ali besedna zveza?
- stalna ali spremenljiva besedna zveza?
- pomensko prozorno ali neprozorno?
- preneseni ali dobesedni pomen?

## eno- ali večbesednost?

- Alfa romeo
- al dente, at hoc, a priori
- d. o. o., itd., tj.
- omg!
- narodnozabavni, narodno-zabavni, narodno zabavni
- naprimer – na primer; hokus pokus – hokuspokus
- bati se, smejati se, svitati se ... (predati kaj : predati se)
- popihati jo, cukati ga (: cukati ga za rokav)

**Leksikalna enota** je lahko eno- ali večbesedna. Leksikalno enoto določa dejstvo, da ima samostojni pomen. Ta pomen je lahko predvidljiv iz pomenov sestavin ali pa ne.

**Stalnost oblike in pomena:**  
posameznih sestavin ni mogoče  
zamenjati z drugimi

imeti trdo **kožo** : \*imeti trdo **povrhnjico**  
**soliti** pamet komu : \***poprati** pamet komu  
imeti **mačka** : \*imeti **psa**

Pa kaj vam morem, morem nič. Vi ste **stari mački**, jaz ne. ☞ Sam sem mislil, da je n pravilno mesto - kjer se to da. Tudi **stari mački** ne moremo novincem nič. To je sk Delight, Prospect, Noctiferia (in pa **starih mačkov** kot so Pomaranča, Laibach..), v ega M. Hladnik, ki je na wikipediji že **star maček**, potolažil, da je to absolutno d no res, je pa tudi res, da je Fomalont **stari maček**, če predpostavimo da je Kopejl (/) pa bom verjetno potreboval pomoč **starih mačkov**. ☞ Toplo vabljen k pisanju, le z e prepuščam vam administratorejem, **starim mačkom** na tem področju, za vaše delo s zabiš ali narobe narediš, se zgodi tudi **starim mačkom**. ☞ Članek Ohotsko morje sem t ko vpraša tudi direktno kakšnega od " **starih mačkov**", ki so trenutno online :-). ☞ Sez je zaradi vandalizma, če pa to naredi **star maček** kot si ti, me pa to malo bolj vrž članka pa napiši, kaj se je dogajalo. **Stari mački** namreč hitro najdemo svoje poj starega sodstva in še vedno ga vodijo **stari mački** iz komunizma ☞ Zakone ki sprej ☞ Lukav tale J.C. Juncker. Prekaljeni **stari maček**, ni kaj :) ☞ Malo bo: "Ajm sori, e in poznejši prevzem stranke s strani **starih mačkov** tipa Gregor Golobič) bi na nasle je, ki se ne bi delili na leve in desne, **stari mački** naj takoj izstopijo iz politike. I

- **Kurja koža** ol ouver! Rade #Šerbedžija je Zakon!
- Omojboh... 30 sekund #Sanremo2014 pa že **kurja koža**, solzne oči...
- včeraj sem prešlatala celo zimsko kolekcijo. **kurja koža** v trenutku!
- 3cm veliko **kurjo kožo** dobim že pri trailerju, kaj šele bo.
- omg, meni kar **kurja koža** postane, ko vidim ono fotografijo
- bemu sunac, **kurja koža**, pa napete bradavičke:)
- pa ko čuješ zvok od stare impreze ti kr **kurja koža** rata
- Ja, ko si pa avto kaskiral, ti pa niso upoštevali trenutnih popustov... men gre kr **kurja koža** ...
- Ko slišim za besedo strokovnjak, dobim **kurjo kožo**.
- Na čelu države bo Gospod Janez Janša, rdeči tk kr **kurjo kožo** dobite ko ga omenim

▪ **oblikovne in skladenjske spremembe**

- Flirtanje je lahko nedolžen klepet, včasih pa postane igra z ognjem.
- *Ne bom dovolil, da spet zaidemo v igračkanje z ognjem.*
- *vseeno ne pozabi biti previden: z ognjem se ni dobro igrati!*

? Igrajte se z ognjem!

? kdo se igra z ognjem?

\*Ta ogenj je bil igran

star maček ≠ ta maček je star

▪ **variantnost sestavin**

- *Nizozemci so pravi trn v **peti** Slovencev.*
- *Ja res, ful so koristni ... kot trn v **riti** ...*
- *Me zanima, zakaj jim je ravno Mišič tak trn v **očesu**.*

biti komu trn v peti

- *To sem **ugotovila** tudi na lastni koži*
- *Da je celotna indistrija en velik konglomerat sem **izkusil** na lastni koži.*
- *ker sem to na lastni koži **doživel***
- *js sm to vojno **občutu** na lastni koži*

doživeti kaj na lastni koži

kdo ne more iz svoje kože

▪ **vrivanje besed**

- *Bivši policijski špicelj pač ne more iz **svoje zakompleksane kože**.*
- *Le kdaj boš stopil iz **svoje egocentrične kože** in nehal kritizirati*
- *jelen pederčina ne more iz **svoje pijane kože**.*

## Predvidljiva variantnost

- Čeprav je **stalnost** ena od definicijskih lastnosti stalnih besednih zvez, je **spremenljivost posameznih sestavin** ena od njihovih najbolj tipičnih lastnosti. Stalnost frazeoloških enot je treba obravnavati v smislu **predvidljive variantnosti**, predvidljivost variant pa je mogoče empirično preučevati prav s pomočjo **korpusov**.

## pomensko prozorno ali neprozorno?

**črn** -a -o prid. **1.** ki je take barve kot oglje ali saje

**črna barva, črna obleka, črne hlače**

črn oblak

črni kruh, črna kava, črni čaj

Črno morje, Črna gora, Črni Vrh

črna kronika, črni humor, črni petek

črna gradnja, črno odlagališče

črna vdova

## Dobesedni ali preneseni pomen?



**črna luknja** je v astrofiziki teoretična zgoščitev mase, katere težnostno polje je tolikšno, da ubežna hitrost presega hitrost svetlobe. Zato pridevnik »črna«, saj nič, niti svetloba, ne more uiti njeni težnosti. Čeprav je zveza »črna luknja« zelo razširjena, teorija ne govori o kakršnikoli »luknji« v običajnem pomenu besede.

- Večja kot je **črna luknja**, bolj požira vase zvezde tako da jih je vedno manj.
- *Gravitacijski valovi so valovanje v prostoru-času, ki ga povzroči nastanek **črnih lukenj** ali drugih silovitih dogodkov v vesolju.*
- *Že leta 1789 je Laplace na osnovi Newtonove teorije gravitacije napovedal možen obstoj **črnih lukenj**.*
  
- *Del moje mize je **črna luknja** za SD kartice. Sedaj moram ugotoviti, kateri del mize je to ...*
- *Jap, v **črni luknji** je vse mogoče, tja tud moj denar zgineva :D*
- *... bom rekel, da so ženske torbice kot **črne luknje**. U never know what U'll find..*

## JANES star maček : starejši maček



- *Rjavec ... hja .. on je **star maček** z vsemi "žaubami" namazan.*
- *Ma ga bodo že zdrnali. Saj so sami **stari mački** na kupu.*
  
- *Cesta je bila smrt že tolikim mojim mačkam, nazadnje moji Zorici, **najstarejši mački**.*
- *Zanima me, kako naj poteka privajanje, da ne bo s strani **starejšega mačka** prišlo do markiranja.*



## Tud drugač se da to povedat ...

Iz tega ne more nastati nič dobrega, kajti zaradi **umiritve razmer** raje **prikrijejo dejstva**. In v celi državi ni »pogumnega **politika**«, ki bi dovolj **odločno** problematiziral takšno početje. To je posmeh vsem, ki so zaradi vladnih ukrepov prisiljeni **povsod** varčevati. Tudi sindikati in upokojenci **nočejo priznati resnice**, ne vidijo **daljnosežnosti problema**. Njihova **zmaga bo imela negativne posledice** za celotno Slovenijo. Skoraj neopazno je bil odstavljen prvi nadzornik HSE, eden redkih, ki **je upal opozoriti na nepravilnosti**.

Iz tega ne more nastati nič dobrega, kajti zaradi **umiritve razmer** raje **prikrijejo dejstva**.

In v celi državi ni »pogumnega **politika**«, ki bi dovolj **odločno** problematiziral takšno početje.

To je posmeh vsem, ki so zaradi vladnih ukrepov prisiljeni **povsod** varčevati.

Tudi sindikati in upokojenci **nočejo priznati resnice**, ne vidijo **daljnosežnosti problema**.

Njihova **zmaga bo imela negativne posledice** za celotno Slovenijo.

Skoraj neopazno je bil odstavljen prvi nadzornik HSE, eden redkih, ki **je upal opozoriti na nepravilnosti**.

Iz tega ne more nastati nič dobrega, kajti zaradi **ljubega miru** raje **pometejo smeti pod preprogo**.

In v celi državi ni »politika **z jajci**«, ki bi **udaril po mizi** in problematiziral takšno početje.

To je posmeh vsem, ki so zaradi vladnih ukrepov prisiljeni varčevati **na vsakem koraku**.

Tudi sindikati in upokojenci **tiščijo glavo v pesek**, ne vidijo dlje **od svojega nosu**.

**Ta zmaga bo Pirova**, poražena pa bo cela Slovenija.

Skoraj neopazno je bil odstavljen prvi nadzornik HSE, eden redkih, ki je **dregnil v sršenje gnezdo**.



- Članek Alenke Puhar v Časniku. Vreden branja. Ob ognju iz polen, vrženih pod noge.
- Sem nad princem obupala pa zdej konja čakam.
- samo pazi, da ne narediš iz vina vodo.
- E, v tem tiču grmi zajec.
- Ni vse zlato, kar se Janezek nauči.
- Ljudje z dlako na jeziku, so srečnejši.
- Volk dlako menja, Erjavc pač ne.

## FRAZELOGIJA IN KORPUSNE RAZISKAVE

- odkrivanje **pogostih in tipičnih besednih sopojavitvev** s pomočjo
  - **korpusov** (elektronskih besedilnih zbirk) in
  - **računalniških orodij**
    - statistični izračuni besedne povezovalnosti
      - n-grami
      - kolokacije
      - stavčni vzorci

**bi-grami**

se je  
da je  
ki je  
pa je  
ki so  
da bi  
je bil  
je v  
da se  
naj bi

**tri-grami**

ki ga je  
ki se je  
pa se je  
da bi se  
da se je  
ki jih je  
se je v  
ki jaj bi  
ki jo je  
ki je v

**štiri-grami**

za okolje in prostor  
**ne glede na to**  
**glede na to da**  
je v tem da  
na to da je  
**po drugi strani pa**  
**res pa je da**  
število izvajalcev  
število udeležencev  
pa naj bi se

**BESEDNE ZVEZE**

**PROSTE BESEDNE ZVEZE**

*dragi dijaki; malo počakati ...*

**STALNE BESEDNE ZVEZE**

**POMENSKO PROZORNE**

*lokalne volitve  
prometna nesreča*

**FRAZEMI (POMENSKO NEPROZORNE)**

*metati polena pod noge; kdor drugemu jamo koplje ...*

delamo jih sproti ko pišemo in govorimo

črna luknja  
veliki pok  
gosta megla  
trda tema

pomenijo čisto nekaj drugega kot besede, ki jih sestavljajo. So metaforične.

njihov pomen lahko razberemo iz pomenov sestavin. Niso metaforične.

- Kaj pa tipične sopojavitve besed, ki nimajo samostojnega pomena? (vzorci, n-grami, kolokacije)
- Kaj pa besede s samostojnim pomenom, nastale po metaforični poti, ki niso frazeološke ampak terminološke (*črna luknja, veliki pok, pametni telefon ...*)

## BESEDNE ZVEZE

### proste neleksikalizirane\*

#### stavčni vzorci

SBZ1 vzeti SBZ3 SBZ4 iz SBZ2  
kdo vzame komu kaj iz česa  
Kdo vzame komu kaj od kod

#### kolokacije

trda tema  
gosta megla

### stalne zveze-leksikalizirane enote\*

#### stalne besedne zveze

črna luknja  
pametni telefon  
veliki pok

#### frazeološke enote

delati iz muhe slona  
pustiti koga na cedilu

\***Leksikalna enota** je lahko eno- ali večbesedna. Določa jo dejstvo, da ima samostojni pomen. Ta pomen je lahko predvidljiv iz pomenov sestavin ali pa ne.

Pri njih je pomlad in že **cvetijo** zvončki in trobentice.  
A poljub je poljub. Od tega se **cveti**.  
V politiki in gospodarstvu **cveti** korupcija.

**Stavčni vzorci** so ustaljeni (predvidljivi) vzorci, ki jih sestavlja **beseda** in predvidljivi elementi v njenem besedilnem okolju, ki so potrebni za to, da se „sproži“ določen pomen te besede.

### CVETETI

‘poganjati cvetove’	kaj cveti kako/kdaj/kje cveti	roža, rožica, vijolica ... bogato, bujno, obilno; poleti, zgodaj, na vrtu
‘prinašati dobiček’	kaj cveti kaj cveti komu	posel, trgovina, turizem
‘izgledati zadovoljen in zdrav’	kdo cveti (od česa) kdo cveti (zaradi česa)	sreča, zadovoljstvo

## Kolokacije

Kolokacije so zveze besed, ki se v jeziku pogosto pojavljajo skupaj.

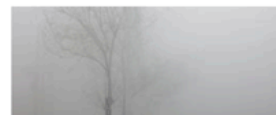
Pogostost sopojavljanja in trdnost zveze je mogoče izračunati.

S kolokacijami se v jeziku izražamo naravno – to je značilno predvsem za matere govorce.

Kolokacije so zato pomembne pri učenju tujega jezika.

Kolokacije so zato tudi pomembna sestavina slovarjev in jezikovnih priročnikov.

<http://www.igra-besed.si/>



## Kolokacije

„you shall know a word by the company it keeps“

(Firth, J. R. 1957)

povej mi s kom se družiš in povem ti ... kaj pomeniš

Kolokacije besede *krona*

Pomeni besede *krona*



„večbesedne besede“

Stalne besedne zveze se obnašajo kot **besede**, le da so sestavljene iz **več besed**, npr. *spalna srajca*. Materni govornici jih zato včasih **težko prepoznamo**.

Njihovega **pomena kot celote** ni mogoče vedno izpeljati iz **vsote pomenov** njegovih sestavin. Včasih se približujejo kolokacijam, včasih pa so blizu FE.

Pogosto nimajo **neposrednega prevoda** v tuj jezik, ali pa v tujem jeziku njihova zgradba **ni večbesedna**.

Njihov pomen je zato treba opisati v **slovarju**.



stara mama =  
grandmother

prava čokolada = cooking  
chocolate

SLO	CZ	DE	ANG
čez čas/ sčasoma	po čase/ začas	im Laufe der Zeit/ über die Zeit	in time
v hipu/ kot bi trenil/ takoj	na první pohled	im Augenblick	in (less than) no time
skrajni čas	nejvyšší čas/ načase	höchste Zeit	high time/ about time

„večbesedne besede“

Stalne besedne zveze se obnašajo kot **besede**, le da so sestavljene iz **več besed**, npr. *bančni avtomat*.

Njihovega **pomena kot celote** ni mogoče vedno izpeljati iz **vsote pomenov** njegovih sestavin.

Stalne besedne zveze navadno **nimajo metaforičnega pomena** – to jih ločuje od frazeoloških enot – lahko pa nastanejo po metaforični poti, npr. *črna luknja*

Zaradi svojega **terminološkega pomena** so navadno izključene iz frazeologije, hkrati pa so v jeziku zelo pogoste, saj se **nova predmetnost** pogosteje opisuje z besednimi zvezami kot besedami.



črna vdova

topla greda



prepovedani položaj



pametni telefon

„drugačnost za vsako ceno“

Katera beseda?

*piš me v uh*

Katera besedna oblika, vrsta?

žebljica, glavica, bes, plentaj,

Slovnična pravilnost?

*odpreti na stežaj - stežaj? priti na kant - kant?*

Pomenska nepričakovanost

*briti (se) norca iz koga* – briti se/briti si noge/  
briti komu dlake/glavo

Opuščanje besed (elipsa)

*Ti meni, jaz tebi.*



## Frazeološke enote

ekspresivnost, drugačnost, opaznost = USTVARJALNOST!

FE so ustaljene metafore, ekspresivni izrazi, zato se z njimi **celotna jezikovna skupnost** izraža **metaforično**.

Njihova raba zato odraža **splošno človeške lastnosti**, pa tudi prepoznavne **družbene pojave** in **narodove značilnosti**.

Pogosto so prav zaradi ekspresivnega naboja povod za ustvarjalno in **duhovito rabo**.



six pack



one pack = [uampek]

*spraviti v red* – urediti  
*mazati si roke z (drekom)* – ukvarjati se z neprijetnimi ali kaznivimi stvarmi

*biti krvav pod kožo*  
*naj sosedu krava crkne*  
*kdor ne skače ni Slovenc*

*kdor drugemu jamljo koplje ...*  
*... ziher ni sebičen*  
*ljubi svojega bližnjega ...*  
*... samo pazi, da te ne dobijo*

*zrno do zrna ... gluten*  
*kamen na kamen ... nepremičninski davek*

*Kdor visoko leta ... je Peter Prevc*

## FRAZELOGIJA V SPLETNI SLOVENŠČINI

**JANES** IJS FILOZOFSKA  
FAKULTETA

## Frazeologija v družbi z jezikom

brezposelni ...	... moški	... ženska	... politik
<b>in upokojenci</b> diplomanti samohranilka <b>mladi</b> prejemniki socialne pomoči	odrasel <b>poročen</b> <b>samski</b> pravi <b>mlad</b>	<b>samska</b> <b>mlada</b> <b>sanjska</b> odrasla sodobna	skorumpiran koruptiven <b>vrl</b> obsojen nesposoben
oseba in iskalci prve zaposlitve <b>mladi</b> invalid <b>in upokojenci</b>	<b>poročen</b> sanjski <b>samski</b> <b>mlad</b> čeden	<b>mlada</b> <b>samska</b> <b>sanjska</b> moški in ženska <b>lepa</b>	vodilen <b>vrl</b> opozicijski vpliven <b>srbski</b>

JANES KRES

**JANES** IJS FILOZOFSKA  
FAKULTETA

## tipično slovensko ...

Kdo je tipičen slovenski	Kaj je tipično slovensko
<b>tpična slovenska</b> družina folklor fora zgodba	<b>tpična slovenska</b> fovšija mentaliteta razmišljanje
<b>tpičen slovenski</b> film volilec voznik	<b>tpična slovenska</b> jed vas pokrajina

JANES KRES

NEGATIVNO  
NEUTRALNO



## 3.2 Povzetek

**Frazeologija** je veda, ki preučuje **večbesedne leksikalne enote** (VLE) in njihovo pomensko in skladijsko obnašanje v besedilu. Obsežni korpusi in računalniška orodja omogočajo prepoznavanje jezikovnih vzorcev in tipičnih sojavitev besed, zato je mogoče VLE preučevati veliko bolj natančno, kot če besede preučujemo izolirano in na podlagi intuitivnih predvidevanj. V korpusu pride do izraza njihova zgradba, pomenske lastnosti in celostna besedilna vloga.

V frazeologiji ne obstaja enotno pojmovanje tega, kaj je **frazeološka enota**. Načeloma velja, da gre za **večbesedno zvezo**, ki ima **samostojen** od sestavin neodvisen navadno **ekspresiven pomen**, je **skladijsko trdna** in jo imamo govorniki v zavesti kot **celoto**.

Za našo raziskavo bomo uporabili delitev, ki obsega 3 tipe

- **KOLOKACIJE** so pogoste sojavitve besed, ki sicer nimajo samostojnega pomena kot celota, ampak izražajo naravno, pristno jezikovno rabo, npr. *gosta megla, krepka juha, dobra družba, najboljša prijateljica, vroče poletje, prva ljubezen* itd.
- **STALNE BESEDNE ZVEZE** (SBZ) so večbesedne enote, ki imajo samostojen, navadno neekspresiven pomen. Najpogosteje označujejo predmete, rastline in živali, npr. *gorsko kolo, rdeči ribez, črna vdova*, ter pojave, npr. *topla greda, sončni zahod*. Načeloma se obnašajo kot besede in so pogosto značilne za določeno strokovno področje, npr. (nogomet) *prepovedani položaj*, (zdravstvo) *kurja slepota*, (državna uprava) *rojstni list*.
- **FRAZEMI** (FE) so večbesedne enote s samostojnim ekspresivnim pomenom, ki imajo lahko zgradbo besedne zveze, stavka ali samostojne povedi, npr. *mamin sinček, izgubiti živce; sreča je opoteča; Ne ga srat*. Za frazeme je značilno, da imajo nepredvidljiv pomen, nepričakovano in spremenljivo zgradbo in ustvarjajo številne priložnosti za ustvarjalno, enkratno rabo.

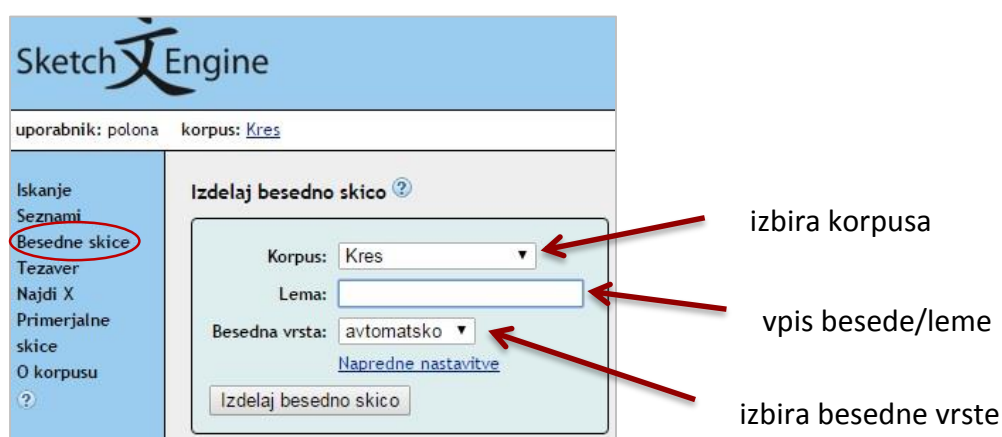
### 3.3 Prepoznavanje različnih tipov večbesednih leksikalnih enot v besedilu – vaje

Priporočljivo je, da si sproti beležite postopek in ugotovitve, saj vam bo to prišlo prav pri izdelavi raziskovalne naloge. Za vnašanje rešitev lahko uporabite kar ta dokument, tako da ga s svojim imenom shranite na računalnik.

Odpremo orodje **Sketch Engine** (SkE):

- NoSkE: [http://nl.ijs.si/noske/janes04.cgi/first\\_form?corpname=janes.04](http://nl.ijs.si/noske/janes04.cgi/first_form?corpname=janes.04)

- FRISke: [https://sketch.cjvt.si/bonito/janes04.cgi/first\\_form](https://sketch.cjvt.si/bonito/janes04.cgi/first_form) - BESEDNE SKICE



Slika 1: Začetno okno v orodju Sketch Engine

Analiza bo potekala v orodju **Sketch Engine** (SkE) z uporabo funkcije »Iskanje« (konkordance) in »Besedne skice« (ang. Word Sketches). Uporabljali bomo **korpus KRES** za analizo splošne slovenščine in **korpus JANES** za analizo spletne slovenščine. Znotraj korpusa JANES lahko izbiramo med različnimi podkorpusi: Janes Blog, janex Comment, Janes Forum, Janes Tweet itd.

V okence »Lema« vtipkamo poljubno besedo v **osnovni obliki** (lemi). To, kar dobimo, so **besedne skice**.

**Besedne skice** so avtomatsko izdelan povzetek slovničnega in kolokacijskega obnašanja določene besede.

uporabnik: polona korpus: Kres

gramatična relacija

koža (samostalnik)  
 Kres frekvenca = 20,475 (170.0 na milijon)

Constructions	S kakšen?	6,085	2.0	S predi-pred	4,540	2.5	S s-koga-kaj	4,079	2.9	S koga-kaj	2,082	3.2
S_biti_s5	237	2.2										
občutljiv	363	9.79		pod	471	8.07	nega	362	9.71	vlažiti	44	9.03
suh	451	9.73		skozi	153	7.62	staranje	153	9.6	dražiti	56	8.91
gol	177	9.1		na	1,882	5.98	čvrstost	36	8.04	pomirjati	45	8.62
masten	130	8.97		brez	83	5.36	tip	191	8.03	negovati	53	8.38
gladek	112	8.68		prek	11	4.86	čiščenje	83	7.99	gladiti	26	8.2
ovčji	84	8.63		iz	191	4.83	plast	116	7.98	zgladiti	26	8.08
pomarančen	79	8.49		čez	23	4.77	vnetje	62	7.76	navlažiti	22	8.03
razdražen	69	8.44		do	115	4.72	luščenje	29	7.75	ščiti	50	7.87
nečist	65	8.37		preko	6	4.26	predelovalec	26	7.62	napeti	33	7.85
rusten	308	8.29		z	423	4.19	prožnost	29	7.54	učvrstiti	18	7.83
napet	68	7.91		po	124	3.82	elastičnost	26	7.54	izsušiti	25	7.79
zrel	70	7.88		v	692	3.77	videz	104	7.52	predreti	16	7.58
trd	102	7.87		za	259	3.57	barva	289	7.51	očistiti	45	7.47
živalski	84	7.81		ob	30	3.11	vlaženje	23	7.45	odreti	12	7.4
razpokan	44	7.79		od	21	2.19	draženje	25	7.43	razdražiti	12	7.35
čvrst	58	7.76		k	7	1.85	obnavljanje	28	7.42	mehčati	12	7.27
nežen	72	7.64		zaradi	6	1.61	povrhnjica	21	7.25	odstraniti	23	7.15
izsušen	39	7.61		med	11	1.59	strojenje	19	7.23	razvajati	17	7.12
kozji	39	7.55		o	7	0.28	površina	125	7.21	odrgniti	9	7.07
debel	83	7.52					ten	20	7.17	osvežiti	16	7.05
mehek	67	7.46					občutljivost	28	7.14	prebosti	10	6.93
zdrav	113	7.42					srbenje	19	7.12	odnesti	37	6.78
srbeč	30	7.29					tonus	18	7.09	poživiti	11	6.74
utrujen	46	7.27					okužba	43	7.03	zaščiti	30	6.73
poškodovan	47	7.23					vlačnost	20	7.02	hladiti	11	6.71

kolokatorji  
 frekvenca  
 št. pojavitev v korpusu  
 statistična jakost

Slika 2: Besedne skice za besedo koža

- **LEMA** je beseda v osnovni obliki.
- **KOLOKATORJI** so besede, ki se tipično sopojavljajo s preučevano besedo in z njo tvorijo KOLOKACIJE (kolokator + koža = kolokacija). Da zvezo štejejo za kolokacijo, mora imeti v korpusu najmanj 3 pojavitve. Če kliknemo na kolokator, se nam odprejo konkordance, ki ta kolokator vsebujejo.
- **FREKVENCA** prikazuje, kolikokrat se kolokator pojavi ob preučevani besedi oz. koliko kolokacij tvori z njo v izbranem korpusu. Pomembno je razlikovati med absolutno in relativno frekvenco
- **STATISTIČNA JAKOST** določa trdnost in sosdvisnost besed, ki vstopajo v kolokacijo.

### 3.3.1 Od kolokacije do frazema

#### 1. naloga

- V izhodiščnem okencu (gl. Slika 1) izberemo korpus KRES, odtipkamo besedo: *jajce* in določimo besedno vrsto: samostalnik.
- Kolikokrat se beseda *jajce* pojavlja v korpusu KRES in kolikokrat v korpusu JANES? Ali lahko na podlagi tega kaj sklepamo?  
 → **frekvenco**.
- V dobljeni besedni skici analiziramo **kolokatorje** v posameznih dveh stolpcih (relacije: S\_kakšen?, S\_kdo-kaj? in S\_kako-kdaj\_p?). Kakšne pomene besede *jajce* izkazujejo kolokatorji v obeh korpusih? Preverimo tudi relacije z glagoli in s predlogi.

- Ali se morda pod katerimi kolokatorji skrivajo frazeološke enote. Pod katerimi kolokatorji? Katere?
- Skušaj ugotoviti, kaj pomenijo te frazeološke enote tako, da jih kot celoto vtipkaj v iskalno okence.

## 2. naloga

- Razdelimo se v skupine oz. v vrste. V prvi vrsti dijaki delajo s korpusom Kres, v drugi pa s korpusom Janes
- Izdelamo besedno skico za pridevnik *bolan*.
- Najprej zabeležimo **frekvenco leme** v obeh korpusih. Kaj ugotovimo?
- Analiziramo **kolokatorje** v relacij S kakšen? v obeh korpusih. Katere pomene nakazujejo kolokatorji v prvem stolpcu? Poskusimo vsak pomen ponazoriti z ustrezno **konkordanco** – zgleodom iz korpusa.
- Naredimo enako raziskavo še za pridevnik *hud* v obeh korpusih

### 3.3.2 Od kolokacije k frazeološkim enotam

#### 1. naloga

V tej nalogi se bomo osredotočili na prepoznavanje različnih tipov večbesednih enot (kolokacije, stalne zveze, frazeološke enote). Obe skupini bosta primerjali rezultate iskanja po korpusu, pri čemer bo ena skupina uporabljala besedne skice, druga pa običajno iskanje v konkordančniku. Obe skupini bosta uporabljali najprej podkorpus Tweet, nato pa še celotni korpus Janes.

- V konkordančnik oz. besedno skico vtipkajte besedo *dlaka*.
  - Katere VLE je mogoče izluščiti na podlagi **besedne skice**. Zapišite kolokacije, stalne zveze in FE:
  - Katere VLE je mogoče izluščiti na podlagi splošnega iskanja.

JANES (Tweet): iskati dlako v jajcu, brez dlake na jeziku, za mišjo dlako, volk dlako menja, narave nikoli, dlaka se ježi, gre pokonci, se dvigne, gre gor ...

- Ali najdemo te zveze tudi v SSKJ. Kaj pomenijo?
- Ali je v SSKJ katera VLE, ki je v korpusu nismo našli
- Če bomo imeli veliko časa na voljo, se gremo lahko pred delavnicami še igro besed: <http://www.igra-besed.si/>

### 3.4 Delavnice

Delavnice so namenjene samostojnemu raziskovanju v skupinah. Pri analizi uporabljamo korpuse in orodja za njihovo analizo (konkordančnik, besedne skice itd.). Za ppt-predstavitev je mogoče uporabiti predlogo, ki jo najdete na spletni strani. Vsaka raziskovalna naloga naj vsebuje:

- raziskovalno vprašanje
- opis korpusa in orodij za analizo
- opis postopka in
- ugotovitve

Raziskovalne teme:

- »To se da tud drgač povedat«: analiza frazeoloških prenovitev v spletni slovenščini
- *Leži kot krava in laže kot pes teče*: živalske prisprodebe v spletni in splošni slovenščini
- *Hudič, vrag, zlodej in satan* na kupu: ali na spletu preklinjamo drugače?
- Slovenec Slovincu – Slovenac: primerjalna analiza besed, ki označujejo regionalno pripadnost: *slovenski, štajerski, dolenski, primorski in gorenjski*
- *Janes, naj te koklja brčne*: primerjalna analiza besed *kokl(j)a, kura* in *kokoš* v splošni in spletni slovenščini
- Vsi drugačni – vsi enakopravni: primerjalna analiza standardno-nestandardnih besednih parov: *hlev-štala, kišta – zaboj, milo – žajfa, obraz – ksiht, knedl – cmok, tla – poden, varžet - žep ...*
- Da si prste oblizneš: večbesedne enote in prehranjevalne navade na spletu (krompir, jajce, juha, vino ...)
- Deli telesa v frazeologiji spletne slovenščine

Ana Zwitter Vitez, Špela Arhar Holdt in Tadeja Rozman

## 4 GOVOR IN SPLETNA SLOVENŠČINA

### 4.1 Govor in spletna slovenščina – izročki

JANES

IJS  
FILOZOFSKA  
FAKULTETA


# Govor in spletna slovenščina

Ana Zwitter Vitez  
FHŠ Univerza na Primorskem, FF Univerza v Ljubljani  
Špela Arhar Holdt  
FF Univerza v Ljubljani, ZUS Trojina  
Tadeja Rozman  
FF Univerza v Ljubljani, ZUS Trojina

Ljubljana, 7. 7. 2016

JANES

IJS  
FILOZOFSKA  
FAKULTETA

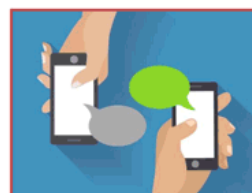


- interdisciplinarno področje
- neraziskani rezultati
- pomembne družbene posledice

## O čem bomo govorili

- specifikje govornjenega jezika
- zakaj in kako raziskovati govor?
- odnos med govorom, pisno in spletno slovenščino

## Od jutra do večera ...



## 1. stereotip: pišemo standardno, govorimo nestandardno



tinka110  
09.02.12  
08:16

Spoštovani, knjižni jezik je eno, pogovorni pa drugo.

Knjižni jezik se uporablja v pisani besedi (vam kakšna ura učenja ne bi škodila), pogovorni pa v izreki.

Odgovori

Citiraj sporočilo

Povezava



[a velikokrat ne ostane le pri tem  
kar smo se lahko prepričali tudi  
med našim snemanjem](#)

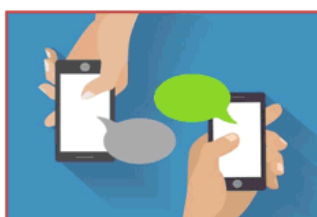
*Antigona se je kljub temu odločila,  
da Polinejka pokoplje.*

[ma tu si lahko duol čez rit  
potjgneš](#)

Igor-za tekmo bo pa jutri pomoje odloču tomažk hahaha +1  
-0  
# 18.08.2015 ob 08:52  
Prijavi neprimerno vsebino



Pišemo standardno, govorimo pa nestandardno?



Pa vendar: specifikke govora

Sprotno tvorjenje besedil

[eem eem eee bom eee hitro povedal eee iz eee znam nardit a masažu](#)

Izgovorjava

[tu je blo ko truebla brez konca](#)

Interakcija s sogovorcem

[ja hvala lepa jst eee eee \[1 delno\] se seveda \*\*strinjam\*\* eee v veliki meri](#)

## 2. stereotip: na spletu pišemo, kot govorimo



### Kako naj se otrok nauči pravilne slovenščine?

Je to danes v poplavi jezikovnih zmazkov sploh mogoče?

Ko smo mi hodili v šolo, je bila večina pisane besede s katero smo se srečevali, napisana pravilno. Knjige, časopisi, televizija, ravno tako plakati, slogani ...

Zdaj pa napak kar mrgoli in to v kao lektoriranih tekstih. Da o slogu ne govorimo. Luštna zbirka knjig za otroke, bralce začetnike. Vsak drugi stavek se začenja z in.

Prijateljica je lektorica pri eni od založb. Pravi, da je obup, da ima čisto premalo časa in da gre lahko samo enkrat skozi knjigo. Tekst dobi direktno od prevajalca in pravi, da je večina prevodov obupnih. /.../

Skratka, kako in kje naj se otrok nauči pravilne slovenščine, če je pa dnevno v stiku z neko spakedranščino.

Pa pustimo internet in SMS sporočila odstrani.

(med.over.net, Al' prav se piše ...?)

## Še več stereotipov ...

Internet slabo vpliva na jezik.  
Internet kvari slovenščino.

Mladi ne znajo več pravilne slovenščine.

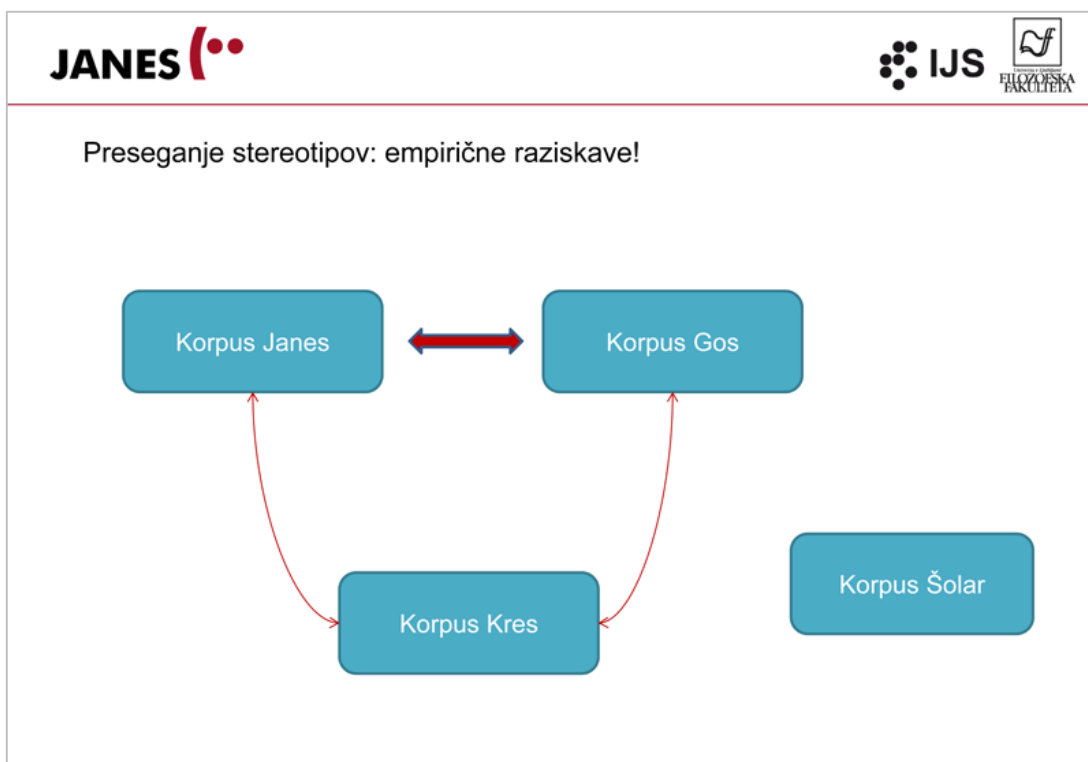
Preseganje stereotipov: empirične raziskave!

Korpus Janes

Korpus Gos

Korpus Kres

Korpus Šolar



**JANES**

IJS FILOZOFSKA  
FAKULTETA

Kako primerjamo korpuse med seboj?

>> najbolj tipične besede enega korpusa

Ključne besede:  
- v opazovanem korpusu so pogoste,  
- v primerjanem korpusu ni veliko.

NoSketch Engine

uporabnik: janes korpus: JANES v0.1.1 Forum podkorpus: Forum L2,3

Iskanje  
Seznami  
O korpusu

NoSketch Engine

uporabnik: defaults korpus: GOS (govorni)

Iskanje  
Seznami  
O korpusu

Shrani  
Spremeni  
nastavitve

Seznam besed

Korpus: GOS (govorni)

Referenčni korpus: KRES (uravnoteženi)

Spremeni cilj in referenčni (pod)korpus

Stran 1 Pojdi Naslednja >

lemma	GOS (govorni)		KRES (uravnoteženi)	
	Frekvenca	Frekvenca/mil	Frekvenca	Frekvenca/mil
eee	23.232	20917.5	63	0.5
eem	2.979	2682.2	3	0.0
mhm	4.480	4033.7	214	1.8
aja	1.318	1186.7	134	1.1
mmm	1.525	1373.1	207	1.7
evo	308	277.3	0	0.0
ne_biti	249	224.2	0	0.0
nnn	245	220.6	7	0.1
aha	2.061	1855.7	1.247	10.4
aaa	290	261.1	86	0.7

## Metapodatki v korpusu Gos

The screenshot shows the search results for the word "rekel" in the Gos corpus. The interface includes a search bar with the word "rekel" entered and a "Najdi" button. Below the search bar, there are navigation options for "Iskanje po pogovornem zapisu" and "Iskanje po standardiziranem zapisu". The results are displayed in a table with columns for "Pogovorni zapis", "Standardizirana oblika", and "Število pojavitev".

Pogovorni zapis	Standardizirana oblika	Število pojavitev
reku	rekel	885
reko	rekel	192
rekel	rekel	83
rek	rekel	27
rjeku	rekel	23
riku	rekel	18

## gos

Iskanje po pogovornem zapisu  Iskanje po standardiziranem zapisu

Uporabljaš [enostavno iskanje](#)

[Napredno iskanje](#)

-obseg:  
milijon besed (112 ur posnetkov)

-vzorec:  
demografsko reprezentativen vzorec govorcev v najpogostejših govornih situacijah

-sestava:  
60% javni diskurz: mediji, predavanja, šolski pouk  
40% nejavni diskurz: zasebni pogovori, delovni sestanki

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

## In še en stereotip: neuporabnost jezikoslovnih raziskav

Finance.si LIVE DANES BRANOST ČASOPIS eFINANCE BORZA MINIBON

Strani: 1

bacekjon sporočil: 735

[#823386] 30.05.10 10:29 [Odgovori](#) + 12

Ja, dokler bo država dušila realni sektor, bodo vsi hoteli biti družboslovci, pravniki, humanisti, in podobni mešalci megle. Ekonomistov pa fax očitno ne usposobi za ekonomijo... saj modeli očitno niso sposobni odpreti lastne firme.

## In še en stereotip: neuporabnost jezikoslovnih raziskav

Finance.si LIVE DANES BRANOST ČASOPIS eFINANCE BORZA MINIBON

plagiatorstvo  
K. T. Guttenberg, K. Markež

anonimne grožnje  
G. Bush, K. Kresal, Z. Jelinčič, J. in U. Janša

literarne vede  
Eva Pacher (*Čudoviti Klon*), David Benjamin (*Sedem*)

kadrovanje  
"Ijudje kot kapital in potencial podjetij" (Jackson 1999)

profiliranje strank in njihovih kupnih navad  
strategije ponudbe in oglaševanja (Shaw idr. 2001)

## Zakaj raziskovati govor?

- proučevanje jezika
- standardizacija jezika
- učenje jezika
- razvojna psihologija
- zdravstvo (npr. prepoznavanje kognitivnih motenj, odpravljanje govornih motenj)
- jezikovnotehnološka orodja (npr. osebe z različnimi oviranostmi)
- sociološke vede
- ...



- interdisciplinarno področje
- neraziskani rezultati
- pomembne družbene posledice



Nekaj odličnejših (v glavnem frazeoloških) ponazoritev za nakazane skladenjske posebnosti neposredno ubesedovalnega govorjenja: 1. popolne ponovitve: *Hodil je hodil, pa ni nikamor prišel. – Beli so beli zobje. – Stopi stopi k sosedu! – Priden priden, psiček.* 2. ponovitve nekako soodnosnega pomena: *Trpim, delam kot konj. – Joka, stoka. – Otroci so mi goli, bosí. – Pojdi, kupi. – Grem, bom videla.* 3. zveze ponovljenih izrazov s členkoma *pa in ne*: *Verjemi, ne verjemi, tako je. – Pameten pa, pameten! – Prijatelj gor, prijatelj dol.* 4. osebek in povedek sta isto: *Služba je služba – Trgovina je trgovina. – Vojska je vojska. – Siromak je siromak bil in bo ostal.* (t. i. istorečje ali tautologija). 5. zveza dovršnika in nedovršnika: *Čaka, ne dočaka. – To čim, ne iztočim. – Lovi, ne ujame. – Gleda, se ne nagleda.* 6. drugi izraz je stopnjevan: *belo prebelo, strašno prestrašno, joj prejoj, bel belcat, sam samcat; lepo, prekrasno; miren, pohleven.* 7. zveza samostalnika in pridevnika iz iste podstave: *trapa trapasta, norec nori, prismođa prismojena, bik bikasti, štor štorasti.* 8. ponovitve vezniki: *Hodili smo in hodili. – Ne razumem pa ne razumem. – On pa ne in ne. – Otrok kot otrok. – Kupil ga bom in ga bom.* 9. ogovorilni glagolski izrazi: *Tebi je tako vidiš, smešno. – On je govoril vse mogoče, veš, jaz pa sem mu verjela. – Lej, sem pridi. 10. členki: Če hoče, pa naj bo. – To ti je bilo veselo na ribniku. – Potem pa še kaj mi verjemi! – Le kje se bo on poboljšal! – Kaj boš zabavljal! 11. poudarni zaimki členki in podobno: *Kakšen čuden si! – Mislím, da je preprosto nor. – Kaj komarj, ampak mrzlica! – Tak povej že, kar hočeš! – Prav treba ti je bilo odpreti usta! – Tse ve, da bom storil, kakor ste želeli.**

vir: Jože Toporišič, Slovenska slovnica

#### Knjižni jezik

Vsak narod ima svoj knjižni jezik, z natančno določenimi pravili, zapisanimi v jezikovnih priročnikih. Je edina socialna zvrst, ki omogoča sporazumevanje po vsej Sloveniji in ima vsenarodno in narodnopovezovalno vlogo. Najstrožja zvrst knjižnega jezika se imenuje zborni jezik.

Uporabljamo ga za sporazumevanje v javnosti oziroma takrat, ko je naše besedilo namenjeno širšemu krogu ljudi. Skoraj vedno ga uporabljamo v zapisanih besedilih, včasih pa zborni tudi govorimo. Zborni jezik je npr. jezik medijev (televizije, radia, časopisov in revij), v njem poteka pedagoški govor v šoli, uporabljajo pa ga tudi drugi nosilci javnih funkcij, npr. politiki, duhovniki idr.

Pedagoški govor in učni pogovor

Kadar na pogovor nismo pripravljeni, vendar se od nas pričakuje knjižna izgovorjava, so v govoru dopustne manjše napake. Takrat nastane **knjižni pogovorni jezik**. To je bolj sproščena različica knjižnega jezika, ki se uporablja pri nepripravljenih pogovorih z ljudmi iz drugih slovenskih pokrajin. Dopusten je v govorjenih radijskih in TV-oddajah, ki potekajo v živo, v anketah, intervjujih ... Knjižni jezik dovoljuje le manjše odmikne od slovenske knjižne norme:

- o izpust končnega l-ja v nedoločniku (moraš odgovoriti, ne morem se učiti ...),
- o izpust neobstoječnih samoglasnikov, zlasti i-ja (so bli).

vir: Slovenščina 1, i-učbenik za slovenščino v 1. letniku gimnazij  
(<https://eucbeniki.sio.si/slo1/index.html>)





## 4.2 Detekstiv – vaje

### 4.2.1 Izhodišče

1. Odpri konkordančnik korpusa Gos.

Najdeš ga na strani <http://www.korpus-gos.net/>.

2. Z orodjem SketchEngine izdelaj seznam ključnih besednih oblik korpusa Gos glede na korpus Kres.

Na spletni strani [http://nl.ijs.si/noske/janes04.cgi/first\\_form?corpname=janes.04](http://nl.ijs.si/noske/janes04.cgi/first_form?corpname=janes.04) izberi možnost **Seznami** in nastavi **korpus Gos**. Pri **Možnostih izpisa** izberi opcijo **Ključne besede**, za referenčni korpus pa izberi **korpus Kres**.

### 4.2.2 Vaje

1. Katere so 3 najbolj tipične besedne oblike korpusa Gos glede na korpus Kres? Zakaj jih tako pogosto najdemo v govoru, v pisnih besedilih pa ne? Na seznamu poišči še tri takšne besede.

2. Kaj je še značilno za besedne oblike na seznamu? Zakaj so redke v korpusu Kres?

3. Na seznamu ključnih besednih oblik poišči oblike *zaj*, *guor*, *fse*, *vejš*, *tk*, *čeprov*, *ge*, *bul*. Zapiši njihovo standardno obliko in ugotovi, iz katerih regij prihajajo govorniki? Koliko izgovornih različic imajo njihove standardne oblike? Izberi eno standardno obliko in poslušaj, kako jo izgovorijo govorniki iz treh različnih regij.

4. Izberi pogovorno besedno obliko, ki je značilna za tvoje govorno področje. Je ta beseda značilna samo za eno regijo? Preveri v korpusu Gos.

5. Na seznamu ključnih besednih oblik poišči besedo *čav*. V katerem tipu diskurza korpusa Gos je najbolj prisotna? Kako stari govorniki jo največ uporabljajo? Koliko besed na ključnem seznamu je takih, ki jih večinoma uporabljamo v neformalnih situacijah?

6. Ali meniš, da bi katero izmed besednih oblik s seznama lahko našli v korpusu Šolar? Preveri v korpusu.

Najdeš ga na [www.korpus-solar.net](http://www.korpus-solar.net).

7. Izberi eno besedo ali besedno obliko, ki jo po tvojem mnenju uporabljate predvsem mladi, in v korpusu Gos preveri, če to drži?

8. Ključne besedne oblike korpusa Gos predstavi v besednem oblaku.

Odpri datoteko [Podatki za besedni oblak.xls] na spletni strani <http://nl.ijs.si/janes/dogodki/tabor-2016/>. V novem zavihku odpri spletno stran <http://www.wordle.net/> in izberi zavihek Advanced. Prilepi podatke iz tabele v okence in izdelaj besedni oblak.

9. Katera beseda se po tvojem mnenju pogosteje uporablja v govoru: *policist* ali *policaj*? Zakaj? Preveri v korpusu Gos in pokomentiraj rezultate. Uporabo obeh variant preveri tudi v korpusu Kres. So rezultati predvidljivi ali presenetljivi?

Korpus Kres najdeš na <http://www.korpus-kres.net/>.

---

### Dodatne vaje

10. Zanima nas, kateri so najpogostejši samostalniki korpusov Gos in Janes.

V orodju SketchEngine izbereš iskanje po korpusu Gos. Pri opciji Vrste iskanj izbereš iskanje z ukazi CQL. Nato v vrstico CQL vpišeš ukaz [tag="So.\*"], s katerim dobiš seznam vseh konkordanc, ki vsebujejo občne samostalnike. Nato v levem meniju izbereš opcijo Frekvence in klikneš Izdelaj frekvenčni seznam.

Zdaj lahko v novem zavihku ponoviš enak postopek na korpusu Janes. Kaj opaziš?

11. Radi bi ugotovili, ali se govor bolj izobraženih razlikuje od govora manj izobraženih.

V orodju SketchEngine izbereš iskanje po korpusu Gos. Pri opciji Vrste iskanj izbereš iskanje z ukazi CQL. Nato v vrstico CQL vpišeš ukaz <s education="srednja šola"> [word=".\*"]. Dobiš seznam konkordanc, nato v levem meniju izbereš opcijo Frekvence in klikneš Izdelaj frekvenčni seznam. Zdaj lahko v novem zavihku ponoviš enak postopek, le da pri iskanju CQL vpišeš ukaz <s education="fakulteta ali več"> [word=".\*"]. Kaj opaziš?

## 4.3 Detektiv – delavnica Odkrivanje govornih elementov na spletnih forumih

Izdelaj seznam ključnih besed korpusa Forum glede na korpus Kres.

Na spletni strani [http://nl.ijs.si/noske/janes04.cgi/first\\_form?corpname=janes.04](http://nl.ijs.si/noske/janes04.cgi/first_form?corpname=janes.04) izberi možnost Sezname in izberi korpus Janes v 0.4 Forum. Pri Možnostih izpisa izberi opcijo Ključne besede, za referenčni korpus pa izberi korpus Kres.

1. Ali na tem seznamu najdemo besede *eee*, *mhm*, *eem*? Zakaj?
2. Oglej si prvih 20 besed na seznamu. Kaj opaziš?
3. Na seznamu poišči besede *drgač*, *misl* in *zih*. Kaj imajo te besede skupnega? Izpiši še tri take besede. Zdaj te besede poišči v korpusu Gos. Kaj lahko sklepaš o profilu avtorja besedila, ki jih uporablja? Ali bi te besede kdaj zapisal/a v šolskem spisu? Odgovor pokomentiraj.
4. Na seznamu poišči besede *vred*, *pomoje* in *nevem*. Kateri pojav zaznamuje te besede? Poskusi te besede poiskati v korpusu Gos: dobiš več rezultatov z zapisom skupaj ali narazen? Kaj pa v korpusu spletnih forumov Janes?
5. Na seznamu ključnih besed poišči besede *feltne*, *oglasnik* in *turbina*. Jih najdeš v korpusu Gos? Zakaj?
6. Seznam ključnih besed korpusa Gos in korpusa spletnih forumov predstavi v besednem oblaku.

Vsaka skupina v okviru izbrane tematike pripravi predstavitev z naslednjo strukturo:

- I. Raziskovalno vprašanje: v kolikšni meri drži, da "na spletnih forumih pišemo, kot govorimo"?
- II. Potek raziskave: utemeljitev izbora vsebinskega sklopa, izbor gradiva oz. korpusov in orodja za njihovo analizo in opis metodologije.
- III. Raziskava in beleženje ugotovitev. Predstavitev besednega oblaka in ugotovitev na podlagi točk 1, 2, 3 in 4. Kateri zaznani pojavi odsevajo jezikovno neznanje, kateri pa kreativnost v jeziku? Kaj na podlagi podatkov iz korpusov lahko povemo o spolu, starosti ali regiji avtorja besedila?
- IV. Zaključki glede na zastavljeno raziskovalno vprašanje.

## 4.4 Detektiv – delavnica Odkrivanje govornih elementov na Twitterju

Izdelaj seznam ključnih besednih oblik korpusa Twitter glede na korpus Kres.

Na spletni strani [http://nl.ijs.si/noske/janes04.cgi/first\\_form?corpname=janes.04](http://nl.ijs.si/noske/janes04.cgi/first_form?corpname=janes.04) izberi možnost Sezname in izberi korpus Janes v 0.4 Tweet. Pri Možnostih filtriranja izberi Izloči besede in izberi datoteko tabor-tuje.txt, ki jo dobiš na spletni strani <http://nl.ijs.si/janes/dogodki/tabor-2016/>. Pri Možnostih izpisa izberi opcijo Ključne besede, za referenčni korpus pa izberi korpus Kres.

1. Ali na tem seznamu najdemo besede *eee*, *mhm*, *eem*? Zakaj?
2. Oglej si prvih 20 besed na seznamu. Kaj opaziš?
3. Na seznamu poišči besede *jst*, *dons*, *jujr*. Po čem so si te besede podobne? Izpiši še tri take besede. Zdaj te besede poišči v korpusu Gos. Kaj lahko sklepaš o profilu avtorja besedila, ki jih uporablja?
4. Na seznamu poišči besede *fora*, *valda* in *ful*. Kaj je njihova skupna lastnost? Izpiši še tri take besede. Zdaj te besede poišči v korpusu Gos. V katerem tipu diskurza prevladujejo? Ali bi te besede kdaj zapisal/a v šolskem spisu? Odgovor pokomentiraj.
5. Na seznamu ključnih besed poišči besede *videoposnetek*, *twit* in *btw*. Jih najdeš v korpusu Gos? Zakaj?
6. Seznam ključnih besed korpusa Gos in podkorpusa tвитov predstavi v besednem oblaku.

Vsaka skupina v okviru izbrane tematike pripravi predstavitev z naslednjo strukturo:

- I. Raziskovalno vprašanje: v kolikšni meri drži, da “na Twitterju pišemo, kot govorimo”?
- II. Potek raziskave: utemeljitev izbora vsebinskega sklopa, izbor gradiva oz. korpusov in orodja za njihovo analizo in opis metodologije.
- III. Raziskava in beleženje ugotovitev. Predstavitev besednega oblaka in ugotovitev na podlagi točk 1, 2, 3 in 4. Ali neformalne besede in izgovoru podoben zapis odsevata jezikovno neznanje avtorjev besedila? Zakaj? Kaj na podlagi podatkov iz korpusov lahko povemo o spolu, starosti ali regiji avtorja besedila?
- IV. Zaključki glede na zastavljeno raziskovalno vprašanje.

Tamara Mikolič Južnič in Jaka Čibej

## 5 SLOVENŠČINA V STIKU Z DRUGIMI JEZIKI

### 5.1 Slovenščina v stiku z drugimi jeziki – izročki

**JANES**

IJS  
FILOZOFSKA  
FAKULTETA

## Slovenščina v stiku z drugimi jeziki

Tamara Mikolič Južnič  
Jaka Čibej

Poletni raziskovalni tabor JANES 2016

Ljubljana, 8. 7. 2016

**JANES**

**O čem bomo govorili**

IJS  
FILOZOFSKA  
FAKULTETA

- Uvodna vprašanja
- Jeziki v prostoru in času
- Pojavljanje "tujega" v slovenščini
- Nekaj statistike o tujejezičnih elementih v korpusu JANES
- Zapisovanje tujejezičnih elementov
- Oblike tujejezičnih elementov
- Preklapljanje koda in načini preklapljanja
- Vzroki za preklapljanje

## Za začetek – mirodvor

- Kaj je to, mirodvor? Kdo bi to besedo kdaj mogel razumeti, ako ne bi znal, da Nemeč veli *Friedhof*? Učeni slovenski novinarji so očetje te strašne prikazni; kajti v nemškem je *Friede* mir, *Hof* dvor, obtorej *Friedhof*, mirodvor, a nič ne de, če *Friedhof* (govori se tudi *Freithof*) nij "dvor miru", nego samo "ograjen dvor" (*eingefriedeter Hof*). Kar nikomur na božjem svetu ne bi na um prišlo, to na um pride bistrournemu Slovencu. [...] Uboga slovenščina! Mirodvorci s toboj delajo, kakor svinja z mehóm; prestóрили so te v gnjusen mirodvorsk jezik.  
(Vertec 1871, Jezikoslovne stvari)

## "Ne mešajte popačenk v slovenski besednjak"

- Radijci na prvem programu - začelo se je uveljavljati angleška beseda vesel ali dober vikend. Dajte to prenehat, ne mešajte angleških popačenk v ta besednjak. Sicer bo ta ful kul generacija v prihodnosti pela Zdravljico v angleščini. Celo doktorji znanosti govorijo in rečejo deset besed, tri so tujke, tri so angleške, tri popačenke, sej komi še kej ostane. Slovenščina je ogrožena, res, na žalost.  
(Marjana z Gorenjske, Jezikovni svetovalni servis na Radio Slovenija 1, 2014)



## Smo res tako posebni?

- Je uporaba tujejezičnih elementov nov fenomen?
- Se to dogaja samo slovenskim govorcem?
- Je to bolj stvar govora/nestandardnega jezika ali knjižnega jezika?
- Prevzete besede vs. tujejezični elementi

## Uporabljate tuje besede v vsakdanji komunikaciji?

- Kratka vaja – v parih zapišite tri primere tujejezičnih prvin, ki bi jih uporabili pri vsakdanji komunikaciji.



## Uporabljate tuje besede v vsakdanji komunikaciji?

- angleščina

aja tvoj twingec je tud **ful kjut**

- italijanščina

En postanek in **ciao!**

Lepo bodi. **Čau.**

- nemščina

menda tut tam velja, da je **muss** samo umret.

- južnoslovanski jeziki

Ma **ajde bračo.**

## Jeziki niso osamljeni

- Noben jezik ni otok.
- Uporaba tujih besed skozi zgodovino
  - stik z drugimi kulturami
  - sprememba življenjskih okoliščin
  - bogatenje jezika
  - sčasoma nekatere besede postanejo del jezika (prevzete besede)
- Kateri jeziki so v preteklosti vplivali na slovenščino?



## "Tuje" v slovenščini

- **Prevzemanje** tujih besed
- Vrste prevzetih besed
  - sposojenke (pica, pice, pici, pico ...)
  - tujke (peugeot, peugeota, peugeotu, ...)
  - citatne besede (first lady)



Trening, birca, lajf iz gud.

Danes sem se naučil, da Betty Ford ni bila Bette Ford (prva je bila First Lady, slednja pa matador).

#WhattaMistakaToMaka

## Preklapljanje koda (code switching)

- uporaba drugega koda (ali drugih kodov) poleg svojega maternega jezika, izmenjavanje različnih jezikov/kodov
- pogoj za preklapljanje:
  - zadostno poznavanje (vsaj dveh) kodov
- Zakaj preklapljamo – jezikovni vidik:
  - premagovanje jezikovnih omejitev (manjkajoči izrazi)
  - učinkovitost (ruter – usmerjevalnik)
  - uspešn(ejš)a komunikacija (hashtag – ključnik)
  - ...

## Govor vs. splet

- Je raba tujejezičnih elementov v vsakdanjem govoru podobna tisti v tvitih, blogih, forumih, komentarjih?
- Korpus GOS
  - 8 zadetkov (7,2 na milijon) za good/gud
  - 10 zadetkov (9 na milijon) za gut [6x Lara Gut]
- Korpus JANES
  - 874 zadetkov (5,4 na milijon) za good/gud
  - 68 zadetkov (0,4 na milijon) za gut

## Kako pogosto uporabljamo tuje besede?

- Tviti:
  - **8,353,515** (78.031,4 na milijon) besed označenih kot tujejezičnih
  - **6,210,610** tujejezičnih besed z malo začetnico (58.014,2 na milijon)
  - Tviteraši: 70 % vsaj 3 tujejezične besede, 40 % vsaj 10
- Blogi:
  - **406,553** (11.772,0 na milijon)
- Forumi:
  - **569,716** (12.104,2 na milijon)
- Komentarji:
  - **120,853** (5,636,1 na milijon)

Lastna imena  
Imena pesmi  
Citati

## Kako zapisujemo tujejezične elemente?

- **citatni zapis**
  - kolk maš bonus score?
- **poslovenjeni zapis** – kakšen?
  - šerati, sherati, sharati, shareati, share-ati ...
- **krajšave**
  - tnx, plz




a pa gre za **gud** zadevo, al je **good old copy / paste**

kdo pa zdaj najbolj **muva** na zabavi? **Ajm 2 gud 2 B** hud.

**damn..** a ne dela **gud**

## Kako zapisujemo tujejezične elemente?

p   N good	767	
p   N gud	107	

p   N gut	65	
p   N guta	2	
p   N gutom	1	

## Kako zapisujemo tujejezične elemente?

Fajn se zapleta!!! Me **liiiiikeee** !!!!

V obrazni knjigi (facebook) bi se izrazil "Aj **lajk**".

nej me en pospravi, da mi nebo treba kopalnice spucat, **plis**

Zato, Volkec, **pliz**, en **šnelkurz** pravnega natakanja goriva.

Kolegu iz pisarne ukradli laptop in projektor. V kolikor kdo prepozna storilca se priporoča za pomoč. **RT plz** .

Raketa na pogon z golf žogicami? Yes, **please** :)

## JANES Kako zapisujemo tujejezične elemente?

word	Frekvenca	
p   N lajk	457	
p   N lajkov	411	
p   N lajkam	193	
p   N lajka	154	
p   N lajkat	112	
p   N lajke	92	
p   N lajkajo	73	
p   N lajki	71	
p   N lajkal	70	
p   N lajkaš	65	
p   N lajkala	52	
p   N lajkaj	46	
p   N lajkanje	29	
p   N lajkate	27	
p   N lajkali	25	
p   N lajkajte	25	
p   N lajkamo	24	
p   N lajkom	23	
p   N lajku	17	
p   N lajkas	17	
p   N lajkanju	12	
p   N lajko	10	
p   N lajkalo	10	
p   N lajkanja	9	
p   N lajkih	8	
p   N lajkanjem	7	
p   N lajkati	6	
p   N lajkano	6	
p   N lajkou	5	
p   N lajkof	5	
p   N lajkfehtanje	5	
p   N lajkfehtanja	5	
p   N lajkane	5	

## JANES

### Kako zapisujemo tujejezične elemente?

p   N cool	485	
p   N kool	40	
p   N coolu	1	

p   N kul	17,794	
p   N kull	84	
p   N kulj	71	
p   N kuuul	29	
p   N kuul	16	
p   N kuuuul	10	
p   N huuuuuu	6	
p   N kuuuuuul	2	
p   N kuuuuul	2	
p   N kuuu	1	
p   N kll	1	

## Kako zapisujemo tujejezične elemente?

- Kdaj in zakaj citatno, kdaj in zakaj pa poslovenjeno?
- Kaj nam način zapisa pove o avtorju?
  - Negotovost pravilnega zapisa?
  - Izražanje identitete?
  - Pogostost rabe?

## Kako se pojavljajo tujejezični elementi?

- Posamezne besede

najbolj **overrated** pilot v zgodovini F1

**Random** otroški pogovor pod oknom.

najboljša scena **ever**

## Kako se pojavljajo tujejezični elementi?

- Besedne zveze

In ja **by the way**, nič vez in poznanstev

Sanjala sem, da sem si postrigla trepalnice na enem ocesu,  
kot **fashion statement**.

Malo mi je na "**short notice**".

## Kako se pojavljajo tujejezični elementi?

- Stavki

Kje, kje, kje? **Please do tell**, lahko tudi na DM ...

ne tega delat.. **think happy thoughts**. Zdej maš nov telefon!

doma mam še eno črno banano, **but inside its good , i think...**



## Kako se pojavljajo tujejezični elementi?

- Povedi

Fantje, kapo dol! **Respect!**

Morda bo pa nekdo drug danes na blogu opisoval pripetljaj pred opičjo kletko – z druge perspektive seveda ... **You know what I mean.**

Najbrž sem edina oseba, ki sovraži petke. Vedno gora dela in za vse se mudi. **Saturdays are more my thing.**

@davidkovic Point je v temu, da ne štekam tega Point je v temu, da ne štekam tega	<b>totally</b> <b>random</b>	random ponujanja. A danes pa čigumije posebej ponujanja. A danes pa čigumije posebej
časopis na Petrolu in ti model ponudi še munchmallow. Ne zastonj, za plačat. Totally random much? ##g Evo, jutranja joga obraza opravljen	<b>munchmallow</b> <b>Totally</b> <b>random</b>	. Ne zastonj, za plačat. Totally random much? ##g Evo, jutranja joga obraza opravljen
D ##g @Razdelilec Sicer je pa bolj djembe je zelena v fris. Center za ravnotezje je zelena v fris. Center za ravnotezje not	<b>trained</b> <b>not</b>	, ne bobni bobni. ##g @mgerencer Okej, to happy. #hrup ##g @multikultivator intervju
napredovala iz 91/54 na 99/64. By doing Lihkar je bil. ##g Odločila sem se, da imam Antisocial personality disorder. #SelfDiagnoseWednesday ##g @multikultivator makes no sense. A glavobol tudi gre, če	<b>happy</b> <b>freaky</b> <b>Behind</b> <b>nothing</b> <b>Antisocial</b> <b>personality</b>	. #hrup ##g @multikultivator intervju je kot North West. ##g Behind the Candelabra the Candelabra smo prevedli Moje življenje . Spi se mi. #pritisk ##g Ah, čudoviti počitniški personality disorder. #SelfDiagnoseWednesday ##g @multikultivator makes no sense. A glavobol tudi gre, če
deziert: D ##g @2sto7 Ah, štekam. A " coffee to , ki " razbija barvo globoko v laseh ". I lol'd . #reklame ##g Za malo manj načičkano poroko	<b>That</b> <b>to</b> <b>go</b> <b>I</b> <b>lol'd</b> <b>No</b>	go " imajo in a je dražji ali cenejši? " imajo in a je dražji ali cenejši? :) ## lol'd. #reklame ##g Za malo manj načičkano poroko . #reklame ##g Za malo manj načičkano poroko
@opica @TomTomanic ##g @photography_A_O @vesnchek Zato je Subway že prvič propadel -	<b>too</b>	shit, tudi na Pošti lahko kupiš vse, pa many fucking questions! : D @davidkovic

## Vrste preklapljanja

- **Besedno preklapljanje**

("tag switching") – hashtag ali druga beseda v tujem jeziku znotraj sicer povsem slovenskega besedila

- prevzemanje ali preklapljanje?

Čakala ob 4: 45 pred avtobusom za Benetke, ko mi je le kapnilo, da grem danes na Dunaj **#facepalm** #prezgodajzamožgane

Kak **response** dobiš?

## Vrste preklapljanja

- **Znotrajstavčno preklapljanje**

iz slovenščine v tuji jezik in nazaj v enem stavku

Silimo pa vas tudi "**out of the box**": sanjali boste o arhitekturi, se učili od Roosevelta in kršili ustaljena pravila.

Bin popolnoma spremeni temo. (Njemu se **nije "baš prepucavalo"** s teboj).

Predvsem je zanimiva nova **drag n drop** funkcija pri tabelah in pa ebooks...

## Vrste preklapljanja

- **Medstavčno preklapljanje** – tujejezični izraz je stavek/poved zase

kot pa da sem sumničava do vaskogar, polna dvomov in frustracij. **If I believe in good there will be good!** Torej, Sosed, krutosti bodo.

Pozdravljen, Matjaž! **Long tajm nou sil!** (Petra s faksa in klepetavih kav)

Prvič sem slišala za Kevina Kampla in prvič sem ga tudi videla. **Is this a joke?** Kevin. Kampl. S tisto frizuro. Resno?

## Zakaj preklapljamo?

- Vzroki za preklapljanje v elektronskih medijih:
  - Izkazovanje solidarnosti oz. pripadnosti skupini, izražanje identitete
  - Naslavljanje izbranega občinstva
  - Izražanje socialnega statusa
  - Pomanjkanje izraza
  - Pomanjkanje terminologije
  - Ustaljene fraze
  - Pisanje o določeni temi (npr. tabuji)
  - Izražanje čustev, odnosa (npr. neodobravanje, jeza, veselje)
  - Retorično sredstvo (prepričevanje bralcev, poudarjanje)

## Naslavljanje izbranega občinstva

**Thank you everybody**, hvala vsem za izjemno vzdušje in podporo! Vedno je lepo biti v MB. Vas nosim s sabo do konca!  
**#badweekend**

NA TARBIŽU POSVET O VEČJEZIČNEM ŠOLSTVU **A TARVISIO  
CONVEGNO SULLE LINGUE STORICHE DELLA VALCANALE**

Če je že konzervativec morda bolj vezan na določene religiozne slogane, pa bi se liberalec kot zgled provokacijam v imenu sloganov zlahka ognil. **No good. Nije dobro.** Ni dobro.

Po pravilu: **Quod licet lovi non licet Bovi**, oziroma: **Nije žvaka za seljaka.**

## Izražanje pripadnosti/identitete

Si predstavljate da bi župan Benetk načrtoval kaj takega? **I Veneziani lo avrebbero già' buttato in Canal grande ... Robe de matti!**

Vse je **pronto** . Se vidimo zvečer v Moby Dicku:)

**Ekola, fatta la xe ;)** Najlepša hvala vsem za izjemo podporo!  
**#mms2014 Viva** Ištra!

## Pomanjkanje izraza

O, lej, ko si javno objavil svoj email naslov. Dodan na  
mojo **spam listo**! ; -)

Tudi **big deal** delat iz njenega appearanca je nesmisel.

Včeraj smo v Lutkovnem gledališču LJ, z  
odličnimi **stand up-isti** in glasbeniki, uspeli zbrati  
sredstva za našo Bosno. Bravo vsi! :)

V Delu smo pred časom pisali o "**ready-made**" spletni  
trgovini, če kaj pomaga

## Ustaljene fraze

najboljs je bit kr doma dons... **just in case**;) )

@bosstjanz čestitke! **great job!** it takes  
effort and dedication. res bravo!

Kje je moj spanec izpred dveh tednov? 5 ur  
na noč? **Come on!**

## "Problematične" teme

Ja, itak. Če bi bilo življenje tako preprosto, bi bilo pa res fajn. **Sex, hot guys and good karma.**

Bejba ostaja fajn, ponosen na njene odločitve....  
**And those boobs , oh, them boobs :P**

Srebrna medalja, osma za Slovenijo! **Goddamit !**

## Izražanje čustev/odnosa

Uni **assholi**, ki se spravijo v prodajo nečesa retardiranega in potem te hočjo pofrendat na LinkedInu, da bi si širli network.

Zdej pa tko: če me teli **effin'** loleki od #big1team ne nehajo dodajat na @LinkedIn, jim bom fajn **lajf** zagrenila. **#promise**

**omg!** a so lahko besedila od bfm še boljša :P **i likeeee**

**OMG**, Amazon je po 10 letih " Not Ales? " spremenil v " Sign Out ". **\*Does a little happy dance\***

## Retorično sredstvo – prepričevanje bralcev

Moram tole napisat ker ... vidim, vidim, da me berete,  
»**but the hell if I know who you are**«! Ne resno, tukaj je zelo malo komentarjev in če ne bi bilo gospoda Google, bi morda včasih že pomislila, da pišem sama sebi!

Chicks dig scars. Zajec me je ravno popraskal po roki. **Still counts, right?**

Točno to. Misliš da lahko obstaja boljši Joker? **No way!** :)

## Retorično sredstvo - poudarjanje

Fantje, kapo dol! **Respect!** #Maribor

I Look To You je klasika, ena najlepših pesmi vseh časov, ki je zdaj dobila tako perfektno preobleko, da nimam pojma kdaj bom izklopil gumb za repeat. **I rest my case**, pravkar sem dobil najboljši, najmočnejši in najlepši komad vseh časov.

## Zakaj vi uporabljate tujejezične elemente?

- Kateri od zgoraj navedenih razlogov je za vas osebno najpogostejši?
- Katere razloge bi še dodali?



## 5.2 Od Novega mesta do Njujorka – vaje

### 5.2.1 Kako zapisujemo tujejezične elemente?

1. Se pogosteje uporablja citatni zapis ali poslovenjene oblike?

a. Odpremo JANES v. 0.4.:  
[http://nl.ijs.si/noske/janes04.cgi/first\\_form?corpname=janes.04](http://nl.ijs.si/noske/janes04.cgi/first_form?corpname=janes.04)

b. V iskalno okence za enostavno iskanje vpišemo naslednje besede v citatnih oblikah in poiščemo njihovo pogostnost.

- like

\_\_\_\_\_

- good

\_\_\_\_\_

- please

\_\_\_\_\_

2. Poleg citatnih oblik se, kot smo videli, uporabljajo tudi različne poslovenjene oblike.

a. Katere so po vašem mnenju poslovenjene oblike zgornjih besed?

b. Napišite te oblike v okence za enostavno iskanje in poiščite njihovo frekvenco. Prepišite najdene vrednosti za citatne oblike (vse skupaj) in poslovenjene oblike (vse skupaj).

• Oblike za *please* in njihova frekvenca:

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

• Oblike za *good* in njihova frekvenca:

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

• Oblike za *like* in njihova frekvenca:

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

3. Katere oblike torej prevladujejo pri izbranih besedah, citatne ali poslovenjene?

\_\_\_\_\_

## 5.2.2 Koliko je tujejezičnih elementov v JANES-u in iz katerih jezikov so?

1. Najprej si pogledjmo, koliko je tujejezičnih elementov v korpusu.

a. V korpusu JANES so besede, ki jih je avtomatski označevalnik prepoznal kot tuje, vendar program ne ločuje med različnimi jeziki. Besede so preprosto označene s posebno, vedno enako kodo. Ko iščemo s CQL, lahko uporabimo to kodo in vse take besede izluščimo naenkrat.

- V polje CQL vpišite iskalni ukaz [tag="Nj"] (Nj je oznaka za tujejezične elemente)
- Koliko je vseh takih konkordanc? \_\_\_\_\_

b. Za primerjavo odprimo tokrat govorni korpus GOS in preverimo, koliko je tujejezičnih elementov v tem korpusu.

- Vtipkajmo enak iskalni ukaz v polje CQL. Rezultat je \_\_\_\_\_.
- Zakaj prihaja do take razlike?

c. Ali se prisotnost tujejezičnih elementov spreminja glede na besedilni tip? Poiščite skupno število tujejezičnih elementov v vsakem izmed štirih podkorpusov JANES-a.

- Na vrhu izberite podkorpus, nato kot vrsto iskanja izberite CQL in v ukazno polje vpišite [tag="Nj"].
- Rezultati:

Blogi	(JANES	v0.4	Blog):	_____
Komentarji	(JANES	v0.4	News):	_____
Forumi	(JANES	v0.4	Forum):	_____
Tviti	(JANES	v0.4	Tweet):	_____

2. Videli smo, da se poleg angleščine, ki nedvomno prevladuje, v spletni slovenščini pojavljajo tudi besede iz drugih jezikov. Kako bi ugotovili, kateri jeziki se poleg angleščine še pojavljajo?

a. Izberite si enega izmed podkorpusov, poiščite vse tujejezične elemente v tem podkorpusu. Napravite frekvenčni seznam besed. Nato na prvih desetih straneh frekvenčnega seznama »ročno« preverite, kateri jeziki se pojavljajo (poleg angleščine). Zapišite te jezike (in primere rabe, ki ste jih našli) spodaj:

---

---

---

3. Kako točni so naši rezultati?

- a. Ker avtomatsko označevanje ni 100 %, se pojavljajo napake: med besedami, označenimi kot neslovenskimi, je tudi nekaj takih, ki to zagotovo niso. Poiščite jih 5.

---

---

---

---

---

- b. Zakaj prihaja do takih napak?

### 5.2.3 Kdaj in kako se uporablja tujejezične elemente?

1. Kako se pojavljajo tujejezični elementi? Kakšne vsebine bodo izražene v tujih jezikih?

- a. Razmislite, na kakšen način bi se lahko pojavljali tujejezični elementi v korpusu JANES (naštejte, kar se spomnite s predavanja).

---

---

---

- b. Ustvarite seznam konkordanc vseh tujih besed v korpusu JANES.

- Izberite korpus JANES v0.4.
- Pri vrstah iskanj izberite CQL in v iskalno polje vpišite [tag="Nj"].

- c. Preglejte dobljene konkordance in ocenite, ali se primeri, ki ste jih našli, ujemajo z vašimi predvidevanji (in spominom). Navedite vsaj 5 različnih načinov in po en primer za vsak način.

---

---

---

---

---

- d. Kateri način vam je najbližji oz. kaj sami uporabljate v podobnih okoliščinah (socialna omrežja, forumi, blogi ipd.)? Napišite tri besede/besedne zveze, ki jih najpogosteje uporabljate, in primerjajte njihovo pogostnost v vseh štirih podkorpusih.

- Besedo/besedno zvezo preprosto vpišite v polje za enostavno iskanje, pri čemer najprej izberite vsak podkorpus posebej. Prepišite, kolikokrat na milijon se pojavlja v posameznem podkorpusu.

Besed/Bes. zveza

Tweet

Forum

Blog

News

---

---

---

e. Kdaj se po vašem mnenju avtorji odločajo za **preklapljanje** – preskok iz slovenščine v drugi jezik in uporabo tujejezičnih elementov (če odmislimo naslove, lastna imena ipd.)?

- Poiščite vse tujejezične besede v podkorpusu Tweet.
- Napišite, kateri so razlogi za preklapljanje v opazovanih primerih:

---

---

---

---

---

- Ali avtorji izrazito preklapljajo na kakem posebnem mestu v svojih tvitih?

---

---

#### 5.2.4 Kdo uporablja tujejezične elemente?

1. Korpus JANES je označen tudi s podatki o tem, kdo so avtorji besedil, ki jih vključuje. Zlasti podkorpus TWEET ponuja veliko informacij v tem smislu. Poglejmo si, pri katerih uporabnikih se tujejezični elementi pojavljajo pogosteje.

a. V podkorpusu TWEET so zasebni uporabniki označeni kot 'private', podjetja, organizacije in drugi javni subjekti pa s 'corporate'. Kateri tip uporabnikov uporablja več tujejezičnih elementov?

- Vključimo lastnosti besedil pri iskanju.
- Izberemo možnost 'corporate' pri izbiri TEXT.SOURCE.
- Z iskanjem CQL poiščemo vse tujejezične elemente v izbranih tvitih. Rezultat za javna besedila: \_\_\_\_\_
- Nato gremo nazaj na iskanje in pri TEXT.SOURCE izberemo možnost 'private'.
- Z iskanjem CQL poiščemo vse tujejezične elemente v teh tvitih. Rezultat za zasebna besedila: \_\_\_\_\_

2. Zabeležen je tudi spol uporabnikov.

a. Ali tujejezične elemente več uporabljajo ženske ali moški?

- Pri lastnostih besedil imamo možnost izbire spola avtorja. Rezultat za ženske: \_\_\_\_\_

Rezultat za moške: \_\_\_\_\_

3. Kdo (kateri posameznik) je v svojih tvitih uporabil absolutno največ tujejezičnih elementov?

- Izdelamo konkordance vseh tujejezičnih elementov v podkorpusu TWEET.
- Izberemo Frekvence in v spodnjem kvadratu (Frekvenčna razporeditev po lastnostnih besedil) Izberemo prvo možnost (text.author).
- Avtor/-ica z največjim absolutnim številom tujih besed v svojih tvitih je:

\_\_\_\_\_

a. Kaj pa če upoštevamo relativno frekvenco, kdo najpogosteje uporablja tujejezične elemente?

- Na zgoraj izdelanem frekvenčnem seznamu kliknemo na napis »Rel. frekvenca (%)«
- Največji odstotek tujejezičnih elementov je v svojih tvitih uporabil/a:

\_\_\_\_\_.

b. Katere tujejezične elemente pa ta oseba uporablja?

- Kliknemo na »p« na začetku vrstice ob imenu avtorja, da odpremo konkordance vseh tvitov s tujejezičnimi elementi tega avtorja.
- Izdelamo frekvenčni seznam tujejezičnih elementov za te konkordance.
- Izberite še kakega drugega avtorja in preverite, ali so besede pri vrhu podobne.

### 5.2.5 Stopnja (ne)standardnosti in tujejezični elementi

1. Podkorpus Tweet je označen tudi glede na stopnjo nestandardnosti besedila (oznake L1, L2 in L3). Oglejmo si, ali višja stopnja nestandardnosti pomeni tudi pogostejšo uporabo tujejezičnih elementov. Kakšna so vaša pričakovanja?

a. Največ tujejezičnih elementov bo verjetno v besedilih z oznako

\_\_\_\_\_.

b. Preverimo, koliko je tujejezičnih elementov v besedilih z najnižjo stopnjo nestandardnosti (torej v besedilih, ki najbolj sledijo pravilom o standardni rabi slovenščine).

- Izberemo podkorpus TWEET.
- Pri Lastnostih besedil izberemo L1.
- V iskalno polje CQL napišemo iskalni ukaz za tujejezične elemente: [tag="Nj"].
- Število tujejezičnih elementov v besedilih, označenih z L1, je

\_\_\_\_\_.

- c. Enako naredimo za besedila, označena z L2 in L3.
- Število tujejezičnih elementov v besedilih, označenih z L2, je \_\_\_\_\_.
  - Število tujejezičnih elementov v besedilih, označenih z L3, je \_\_\_\_\_.
- d. Razmislimo o razlogih za take rezultate. Pobrskajte po konkordancah za posamezne oznake nestandardnosti in preverite, v kakšnem sobesedilu se kje pojavljajo tujejezični elementi.
2. Preverimo lahko tudi, kdo uporablja najbolj nestandarden jezik *in* največ tujejezičnih elementov.
- a. Preverimo, kdo uporablja več tujejezičnih elementov v najbolj nestandardnih tvitih, moški ali ženske.
- Izberemo podkorpus TWEET.
  - Pri Lastnostih besedil izberemo L3.
  - Pri Lastnostih besedil izberemo spol avtorjev (enkrat ženski, enkrat moški).
  - V iskalno polje CQL napišemo iskalni ukaz za tujejezične elemente: [tag="Nj"].
  - Skupno število tujejezičnih elementov pri ženskah v najbolj nestandardnih tvitih je: \_\_\_\_\_  
Skupno število tujejezičnih elementov pri moških v najbolj nestandardnih tvitih je: \_\_\_\_\_
- b. Poiščite tri avtorje moškega in tri ženskega spola, ki uporabljajo najbolj nestandarden jezik in hkrati največ tujejezičnih elementov.
- Ko po zgoraj navedenem postopku dobimo konkordance, zgradimo frekvenčni seznam, pri čemer v oknu *Frekvenčna razporeditev po lastnostih besedil* izberemo »text.author«.
  - V najbolj nestandardnih tvitih največ tujejezičnih elementov uporabljajo:  
moški: \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_
  - ženske: \_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_