

POLETNI RAZISKOVALNI TABOR  
SPLETNE SLOVENŠČINE JANES  
ZA SREDNJEŠOLCE

Oddelek za prevajalstvo  
Filozofska fakulteta, Ljubljana

24. – 28. avgust 2015

# UČNA GRADIVA

**JANES** (••)

# POLETNI RAZISKOVALNI TABOR SPLETNE SLOVENŠČINE JANES ZA SREDNJEŠOLCE

## Zbornik učnih gradiv

Uredila: Darja Fišer

Univerza v Ljubljani, Filozofska fakulteta, 2015.

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, 2015

1. izdaja

Publikacija je brezplačno dostopna na spletni strani:

<http://nl.ijs.si/janes/wp-content/uploads/2015/10/Gradiva.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 (2015 ; Ljubljana)

Učna gradiva / Poletni raziskovalni tabor spletne slovenščine Janes za srednješolce, Ljubljana, 24.-28. avgust 2015 ; [uredila Darja Fišer]. - 1. izd. - El. knjiga. - Ljubljana : Znanstvena založba Filozofske fakultete Univerze v Ljubljani, 2015

ISBN 978-961-237-780-9 (pdf)

1. Gl. stv. nasl. 2. Fišer, Darja, 1978-  
281635584

---

## KAZALO VSEBINE

1	PREDGOVOR .....	5
	Špela Arhar Holdt in Jaka Čibej: .....	7
2	KORPUSI IN KORPUSNO JEZIKOSLOVJE .....	7
2.1	IZROČKI – korpusi in korpusno jezikoslovje .....	9
2.2	Od konkordance do kolokacije – prvi del .....	23
2.2.1	Korpus KRES.....	23
2.2.2	Korpus GOS.....	23
2.3	Od konkordance do kolokacije – drugi del .....	24
2.3.1	Korpus Janes.....	24
2.3.2	Korpus Šolar.....	26
	Damjan Popič in Katja Zupan: .....	27
3	STANDARDNA IN SPLETNA SLOVENŠČINA .....	27
3.1	IZROČKI – standardna in spletna slovenščina .....	27
3.2	Delavnica - Norma: od pravopisa do tvitopisa.....	45
3.2.1	Potek delavnic in vsebina raziskovalnih nalog.....	45
3.2.2	Programski jezik CQL .....	46
3.2.3	Spletne povezave.....	47
	Polona Gantar in Jaka Čibej: .....	48
4	FRAZELOGIJA V SPLETNI SLOVENŠČINI .....	48
4.1	IZROČKI – frazeologija v spletni slovenščini.....	49
4.2	Vaje .....	59
4.2.1	Prepoznavanje različnih tipov večbesednih leksikalnih enot v besedilu .....	59
4.2.2	1. Naloga: KOLOKACIJE .....	60
4.2.3	2. Naloga: OD KOLOKACIJE DO FRAZEMA.....	61
4.3	Delavnice .....	62
4.3.1	RAZISKOVALNE TEME .....	62
	Ana Zwitter Vitez in Špela Arhar Holdt: .....	63
5	GOVOR IN SPLETNA SLOVENŠČINA .....	63

<b>5.1</b>	<b>IZROČKI – govor in spletna slovenščina</b> .....	<b>63</b>
<b>5.2</b>	<b>Detektiv – vaje</b> .....	<b>76</b>
5.2.1	Izhodišče .....	76
5.2.2	Vaje .....	76
<b>5.3</b>	<b>Detektiv - delavnica (skupine Tvit1, Tvit2, Tvit3)</b> .....	<b>77</b>
5.3.1	Odkrivanje govornih elementov na Twitterju .....	77
5.3.2	Naloge.....	77
<b>5.4</b>	<b>Detektiv - delavnica (skupini Forum1 in Forum2)</b> .....	<b>78</b>
5.4.1	Odkrivanje govornih elementov na spletnih forumih .....	78
5.4.2	Naloge.....	78
	<b>Tamara Mikolič Južnič in Katja Zupan:</b> .....	<b>79</b>
<b>6</b>	<b>SLOVENŠČINA V STIKU Z DRUGIMI JEZIKI</b> .....	<b>79</b>
<b>6.1</b>	<b>IZROČKI – slovenščina v stiku z drugimi jeziki</b> .....	<b>79</b>
<b>6.2</b>	<b>IZROČKI – vaje</b> .....	<b>96</b>
<b>6.3</b>	<b>Vaje: od Novega mesta do Njujorka</b> .....	<b>111</b>
6.3.1	Kako zapisujemo tujejezične elemente? .....	111
6.3.2	Koliko je tujejezičnih elementov v JANES-u in iz katerih jezikov so? .....	112
6.3.3	Kdaj in kako se uporablja tujejezične elemente? .....	113
6.3.4	Kdo uporablja tujejezične elemente?.....	115
6.3.5	Stopnja (ne)standardnosti in tujejezični elementi.....	116

## 1 Predgovor

Od 24. do 28. avgusta 2015 je na Oddelku za prevajalstvo Filozofske fakultete Univerze v Ljubljani potekal 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, s sredstvi razpisa za predstavljanje, uveljavljanje in razvoj slovenskega jezika (JPR-UPRS-2015) pa ga je sofinanciralo Ministrstvo za kulturo.

Z raziskovalnim taborom, na katerem so srednješolci dobili priložnost raziskovati slovenščino z najsodobnejšimi viri, tehnologijami in metodami, razvitimi v okviru projekta JANES (<http://nl.ijs.si/janes/>), smo želeli dijakom približati slovenščino in pokazati, da je izjemno zanimiv in pomemben predmet proučevanja. Poleg tega pa je bil naš cilj tudi, da dijake seznanimo z jezikovnimi tehnologijami, s katerimi slovenščina postaja vse bolj opremljena. Tako smo na praktično zasnovanih raziskovalnih delavnicah dijake uvedli v svet korpusnega jezikoslovja in jezikovnih tehnologij ter jih opremili s kompetencami, ki jim omogočajo, da s pomočjo prosto dostopnih virov za slovenščino samostojno odgovarjajo na najrazličnejša jezikovna vprašanja.

Z analizo izbranih jezikovnih prvin smo osvetlili razlike med standardno in spletno slovenščino na ortografski, leksikalni, skladenjski in slogovni ravni ter razlike med jezikom družbenih medijev in govorjeno slovenščino. S pomočjo opravljenih analiz so dijaki spoznali temeljne pojme s področja sociolingvistike in analize diskurza ter dvignili zavest o izbiri ustreznih jezikovnih različic v različnih govornih položajih.

Raziskovalni tabor je trajal pet dni. Vsak dan je bil posvečen drugi osrednji temi, ki je bila najprej predstavljena na uvodnem predavanju, temu pa so sledile vaje in delavnice. Na njih so dijaki v manjših skupinah pod nadzorom mentorjev, vrhunskih strokovnjakov s področja korpusnega jezikoslovja in jezikovnih tehnologij, samostojno raziskovali izbrano raziskovalno vprašanje. Ker menimo, da bodo gradiva koristila tudi drugim dijakom, študentom, profesorjem in jezikoslovcem, ki se želijo seznaniti s korpusnim jezikoslovjem in značilnostmi spletne slovenščine, smo jih zbrali v skupni publikaciji, ki je pred vami.

Poglavja sledijo programu tabora: v prvem poglavju *dr. Špela Arhar Holdt* 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. V drugem poglavju *dr. Damjan Popič* oriše razvoj slovenske pravopisne norme podobnosti in osvetli razlike med standardno in spletno slovenščino. Tretje poglavje, ki ga je pripravila *dr. Polona Gantar*, je posvečeno večbesednim leksikalnim enotam v spletni slovenščini s poudarkom na kolokacijah in frazemih. *Dr. Ana Zwitter Vitez* v četrtem poglavju izpostavi podobnosti in razlike med računalniško posredovano komunikacijo in govornim jezikom ter demonstrira odkrivanje govornih elementov v spletnih uporabniških vsebinah. Gradiva sklenemo s poglavjem *dr. Tamare Mikolič Južnič* o pojavljanju in zapisovanju tujejezičnih prvin ter preklapljanju med jeziki v spletnih žanrih.

doc. dr. Darja Fišer

Ljubljana, oktober 2015

Špela Arhar Holdt in Jaka Čibej:

## 2 KORPUSI IN KORPUSNO JEZIKOSLOVJE

**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 **ljubitelsko 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, 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 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.

**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.





## 2.1 IZROČKI – korpusi in korpusno jezikoslovje

**JANES** (••)  

# Korpusi in korpusno jezikoslovje

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

Ljubljana, 24. 8. 2015

**JANES** (••)  

## 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**   

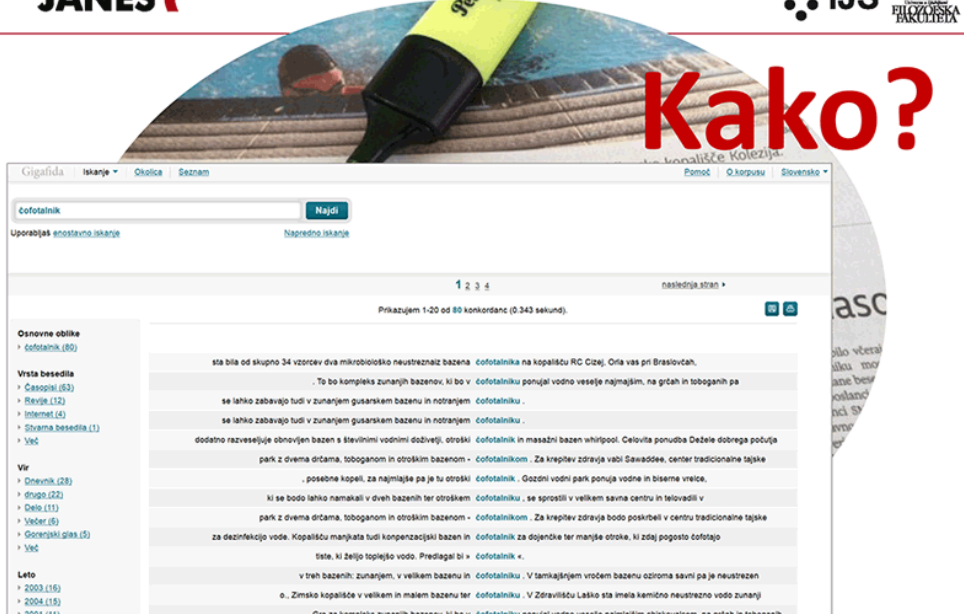
**Želimo raziskovati slovenščino!**



časopisi, revije  
 Matjaž in Gr... življenj  
 primorske novice  
 Predsednik...  
 DELO  
 VEČER  
 knjige  
 govorni jezik  
 besedilni drobiž  
 itd.  
 spletna strani  
 družbena omrežja

**JANES**   

**Kako?**



Gigafida | Iščanje | Okolica | Seznam | Pomot | 0.korpusu | Slovensko

Cofotalnik

Uporabljaš enostavno iskanje | Napredno iskanje

1 2 3 4 | naslednja stran

Prikazujem 1-20 od 60 koncordanc (0.343 sekundi)

Osnovne oblike  
 Cofotalnik (80)

Vrsta besedila  
 Casopisi (53)  
 Revije (12)  
 Internet (4)  
 Svojima besedila (1)  
 Vse

Vir  
 Dnevnik (28)  
 Strupa (22)  
 Delo (11)  
 Večer (8)  
 Gorenjski glas (5)  
 Vse

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

sta bila od skupno 34 vzorcev dva mikrobiološko neustrezna bazena - Cofotalnika na kopališču RC Cizje, Orta vas pri Braslovcah.  
 . To bo kompleks zunanjih bazenov, ki bo v Cofotalniku ponjal vodno vsejše najmljajm, na grtah in toboganih pa  
 se lahko zabavajo tudi v zunanjem guarskem bazenu in notranjem Cofotalniku .  
 se lahko zabavajo tudi v zunanjem guarskem bazenu in notranjem Cofotalniku .  
 dodatno razveseljuje obnoven bazen s strelinimi vodnimi doživljaji, oruški Cofotalnik in masadni bazen whirlpool. Celovita ponudba Dežele dobnega počitja  
 park z dvema dječama, toboganom in otroškim bazenom - Cofotalnikom . Za krepitev zdravja vabi Savadose, center tradicionalne tajske  
 posebne kopeli, za najmljaje pa je tu oruški Cofotalnik. Gozdni vodni park ponuja vodne in biserne vrele,  
 ki se bodo lahko namakali v dveh bazenih ter otroškem Cofotalniku . se sprostiti v velikem savna centru in telovadli v  
 park z dvema dječama, toboganom in otroškim bazenom - Cofotalnikom . Za krepitev zdravja bodo poskrbeli v centru tradicionalne tajske  
 za dezinfekcijo vode. Kopališču manjata tudi konpenzacijski bazen in Cofotalnik za dojenčke ter manjše otroke, ki zdaj pogosto cofotajo  
 tiste, ki željo toplejšo vodo. Predlagal bi » Cofotalnik «.  
 v treh bazenih: zunanjem, v velikem bazenu in Cofotalniku . V tamkajšnjem vročem bazenu oziroma savni pa je neustrezen  
 o, Zmsko kopališče v velikem in malem bazenu ter Cofotalniku . V Zdravilišču Laško sta imela kemično neustrezno vodo zunanji  
 . Gre za kompleks zunanjih bazenov, ki bo v Cofotalniku ponjal vodno vsejše najmljajm obiskovalcem, na grtah 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?

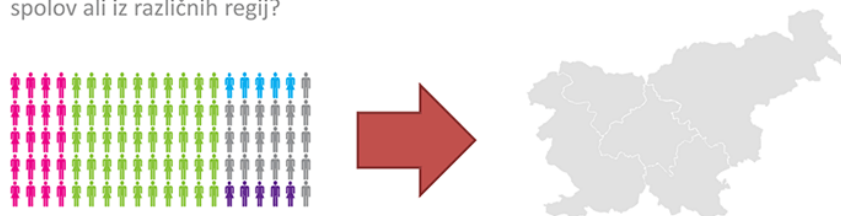


Kakšen **vzorec** moramo zajeti v raziskavi, da lahko na koncu **generaliziramo** rezultate?

### 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)*

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

Gigafida Iskanje ▾ Okolica Seznam Pomoč O.korpusu Slovensko ▾

Najdi

Uporebjaš 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)
- ▶ [Stvarna besedila](#) (1)
- ▶ [Več](#)

**Vir**

- ▶ [Dnevnik](#) (28)
- ▶ [drugo](#) (22)
- ▶ [Delo](#) (11)
- ▶ [Vešer](#) (6)
- ▶ [Gorenjski glas](#) (5)
- ▶ [Več](#)

**Leto**

- ▶ [2003](#) (16)
- ▶ [2004](#) (15)
- ▶ [2001](#) (11)

sta bila od skupno 34 vzorcev dva mikrobiološko neustrezna iz bazena [čofotalnika](#) na kopališču RC Cizej. Oria vas pri Braslovčah,

. To bo kompleks zunanjih bazenov, ki bo v [čofotalniku](#) ponujal vodno veselje najmlajšim, na gričah 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 dirčama, 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 telovadnici v

park z dvema dirčama, 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. Zmiskio kopališče v velikem in malem bazenu ter [čofotalniku](#) . V Zdravilišču Leško sta imela kemično neustrezno vodo zunanji

. Gre za kompleks zunanjih bazenov, ki bo v [čofotalniku](#) ponujal vodno veselje najmlajšim obiskovalcem, na gričah in toboganih

## 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 korpus?

- 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 ...

## Korpusi in jezikovni priročniki



### KORPUSI

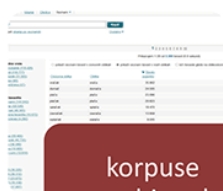
- Vidimo sobesedilo in podatke o besedilu
- Potrebna je sinteza in interpretacija
- Možnost hitrega posodabljanja
- Razen pripisanih oznak so besedila v izvorni obliki



### PRIROČNIKI

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

## Od korpusa do priročnika

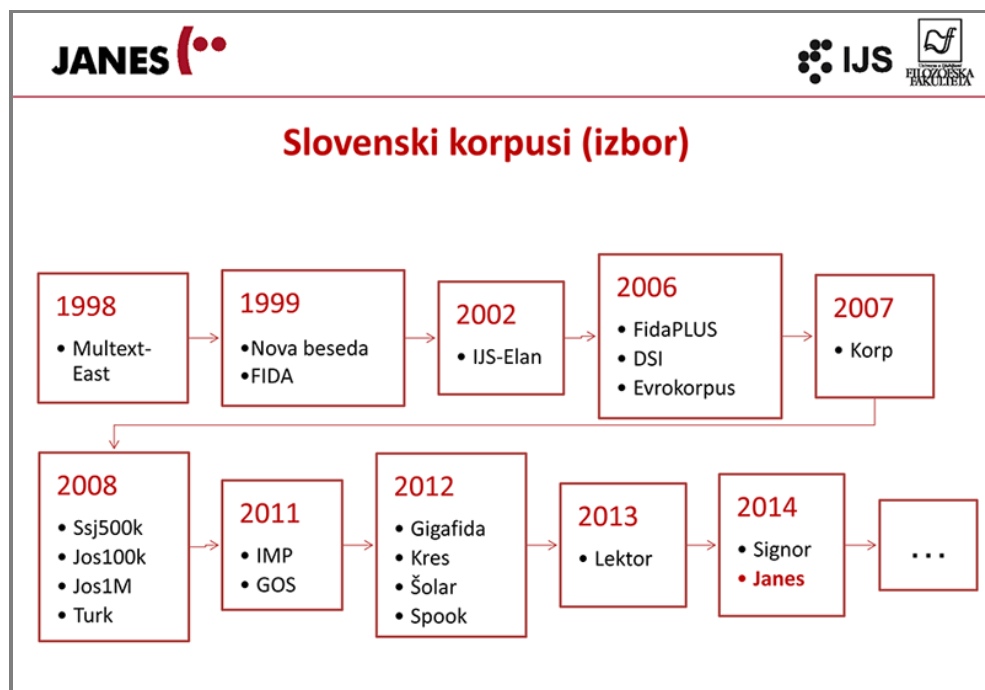




korpuse uporabimo kot vir jezikovnih podatkov



uporabniki lahko preberejo opis jezika v priročnikih

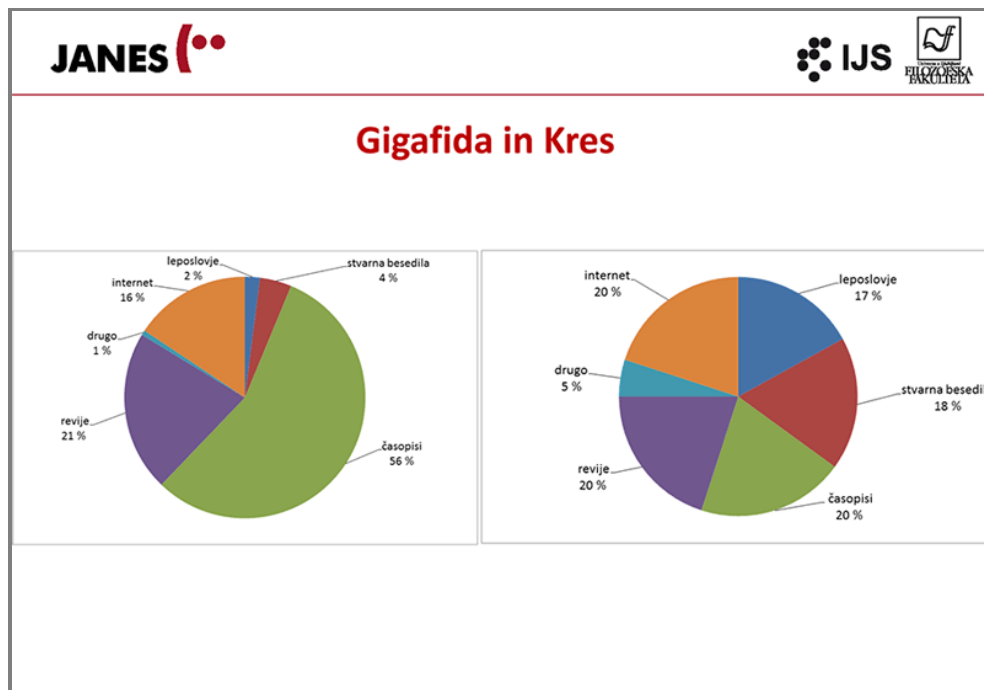
Za slovenščino še ne obstaja veliko priročnikov, ki so narejeni na osnovi korpusnih podatkov (v prihodnosti jih bo več). Trenutno je zato najti razlike med korpusnimi podatki in podatki nekaterih priročnikov.



**JANES** (••)  

### 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).



**JANES** (••)  

### GOS

- **GOS** je prvi korpus govorne slovenščine. Prinaša posnetke govora v različnih vsakodnevni situacijah.
- Posnetki so **transkribirani** in umeščeni v zmojliv 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*

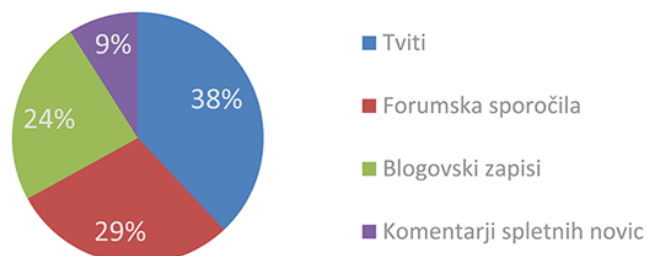
[Transkribcija 2:](#) *ja ne vem po **kateremu** ključu se bomo odločali eee koga bomo poslali v samo trgovanje z volno eem ne vem a bo to žrebanje ali ali glasovanje ali kakor*

## 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 podporo jezikoslovno analizo nestandardne spletne slovenščine, izboljšali jezikovnotehnološka orodja za obdelavo besedil, napisanih v nestandardnem jeziku, in izdelali slovarček spletne slovenščine.

**Hvala za pozornost,  
po odmoru pa nad korpuse!**

Špela Arhar Holdt  
spela.arhar@trojina.si

## 2.2 Od konkordance do kolokacije – prvi del

### 2.2.1 Korpus KRES

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

1. Odpremo korpus Kres in vtipkamo v iskalno okence besedo *pljuvalnik*. Ogleamo 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? Naštejemo tiste, za katere mislimo, da so v korpusu najpogostejše. Odpremo zavihek Seznam in vnesemo iskalni pogoj *\*nik*. Ogleamo si rezultate. Ogleamo si filter Besedna vrsta in komentiramo, kar najdemo pod Neuvrščeno in Prislov.

4. Ogleamo si besedo *lastnik* in razmislimo o besedni zvezi *lastnik + koga ali česa*. Naštejemo nekaj primerov. Odpremo zavihek Okolica in vnesemo besedo *lastnik* + prvo mesto na desni. Ogleamo 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 ogleamo 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?

### 2.2.2 Korpus GOS

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

1. Odpremo korpus GOS in vtipkamo v iskalno okence besedo *recimo*. Ogleamo 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* – ogleamo 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 poslušá posnetke, ki so najbližje njegovemu narečju ali ki se mu zdijo najbolj zanimivi.

## 2.3 Od konkordance do kolokacije – drugi del

### 2.3.1 Korpus Janes

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

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: *tvit*, *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.

### 2.3.2 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**. *Janes*, ki je povsem nov projekt, sicer še ni predstavljen, so pa na voljo videopredstavitve korpusov *Gigafida* (ki ima enak konkordančnik kot *Kres*), *GOS* in *Šolar*.

**Damjan Popič in Katja Zupan:**

## 3 STANDARDNA IN SPLETNA SLOVENŠČINA

### 3.1 IZROČKI – standardna in spletna slovenščina

**JANES** (••)  

## Standardna in spletna slovenščina

Damjan Popič

Ljubljana, 25. 8. 2015

**JANES** (••)  

## (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)

## **(Za)kaj je jezikovni standard?**

tu ste res eni nehvaleni pa poglejmo kolk je slovenija sposobna narest dober film sama pizdarija od slovenskih filmov brez repa in glae.Kolk je vaših filmov gledanih in pohvaljenih to se ne vprašate sam pljuvat znate neki in nič druga sam taki pač ste še srbija z minimalnim vložkom za film naredi 50x boljši film od vaših tak da bodite raje tiho in tistmu ko ni všeč naj ne komentira lepo zbriše in basta ko boste vi delali dobre filme in naredili neki takega pol se javte pizde nehvaležne!!!!!!

—(Rambo[Killer])

sm ze glezu use terminatore pa so bli usi carski upam da je ta zadnji deu še bol carski za gledat1 liga

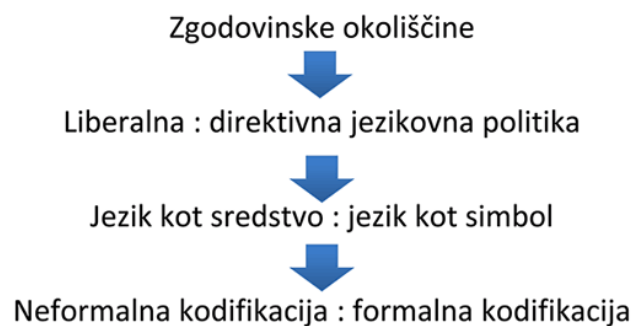
— mimejaga

## Tole je Slovenija

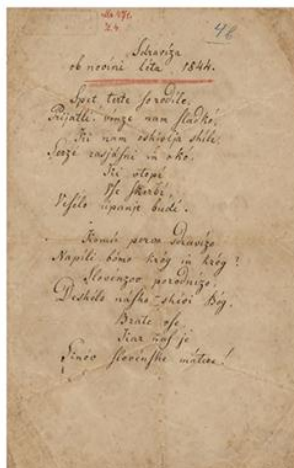


(Vir: [http://www2.arnes.si/~glisentvid10/oseb\\_stran/slovenska\\_narecja\\_v1.html](http://www2.arnes.si/~glisentvid10/oseb_stran/slovenska_narecja_v1.html))

## Kako se rodi jezikovni standard



## Kam torej spada slovenska situacija?



(Vir: Wikipedija)

## Kateri standard?

ABDE©FGHHIİŻJKLŁMNNŃOƆPRŚWϣ3ϠTUVŀϣ

ABCDEFGHIJKLMNŃOPRS8ZXTUYVϣ

ABCČDEFGHIJKLMNOPRSŠTUVZŽ

(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 Slovenci, 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 podnebnju, 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 slovenske sosede in brate. Tako je iz njih polagoma nastala zaključena narodna enota Slovencev, ki se je določno zavedela svoje samobitnosti in svojega imena najklesneje 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 koristen, 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 besede 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 Pohl: Krayska 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 zmožel 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)

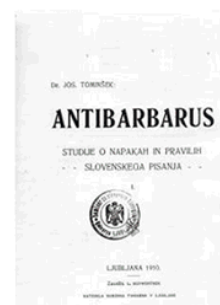
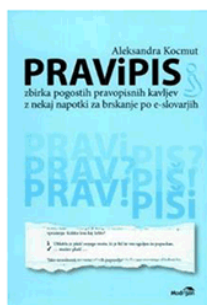
## Tisti, ki bere, je ...

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, sestavlavec, skakavec, smejavec, stradavec, svetovavka, škripavec, zajedavec, zalezovavka, žvižgavec;*

(Slovenski pravopis 1962)

## Knjigae resnice



## Jezikovno razsodišče

JEZIKOVNO RAZSODIŠČE 31. 8. 1988 št. 3

### Tako se govori!

V dneh, ko je v Beogradu potekala 17. seja CK ZKJ, smo člani Jezikovnega razsodišča – podobno kakor drugi Jugoslavlani pa tudi tujci – napeto poslušali radijska in televizijska dogajanja. Poleg številnih veselih in žalostnih vsebin, ki so nam jih sporočali govorniki in komentatorji, smo dočakali dogodek, ob katerem se nam ne zdi neupravičena niti raba pridevnika »zgodovinski«: Trije slovenski člani CK ZKJ so svoje diskusijske prispevke v 1. točki dnevnega reda povedali po slovensko (že pred njimi pa je v svoji materinščini – albanščini – nastopil neki govornik s Kosova).

Posebno smo veseli, ker so se za ta korak odločili pripadniki našega takorekoč najmlajšega političnega rodu: Borut Suklje, Darja Colarič, Martina Kolar. Doslej so se slovenski delegati odločali za slovenščino samo v zvezni skupščini, pa še tam samo nekateri med njimi in samo včasih, čeprav je bilo že dolga leta poskrbljeno za prevajanje. Kaže, da izkušnja s srbohrvaščino na letošnjem ljubljanskem procesu pred vojaškim sodiščem, že poraja prve pozitivne sadove. Upamo, da ne bodo zadnji.

## Jezikovni standard



*In zdaj, Slovenci, čujte blagovest, slovenski pravopis vzemite v pest! — Po zgledu svetega Cirila in Metoda nam dana nova je pisava in metoda.*



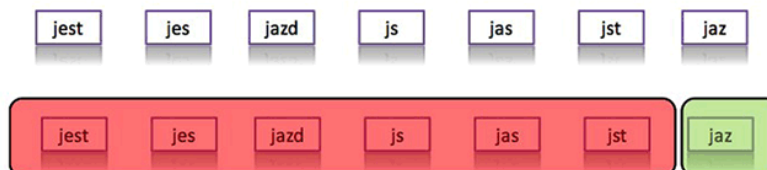
(Vir: <http://www.pravopis.si/Zgodovina/tabid/82/Default.aspx>)

### Jezikovni standard



(Vir: Stabej 2012b)

### Jezikovni standard



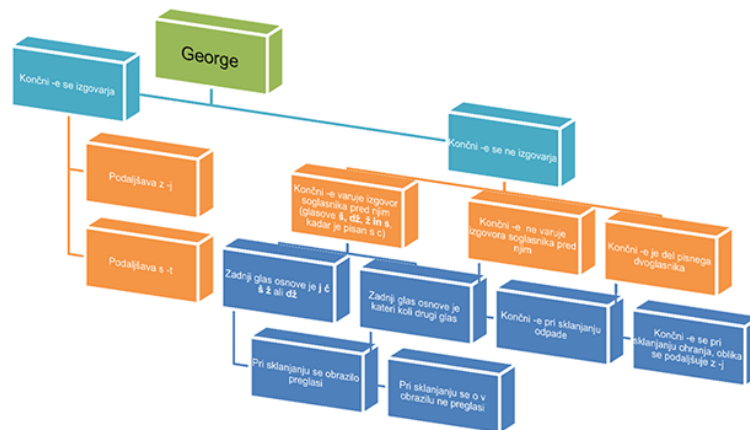
(Vir: Stabej 2012a)

## 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).

Koper 31.12.2014

Spoštovani,

Sporočamo vas da, z 1.1.2015 bo uveljaven postopek e-računov za javni sektor. Prosili bi, vas da, da pred pošljete račune, da nam dostavite ponudbo. Ker v nasprotnem primeru če je račun brez navedene naročilnice moremo zavrniti.

Ponudbo lahko pošljete gospej Maji Koren na e-mail: [maja.koren@rtvslo.si](mailto:maja.koren@rtvslo.si) (Tel: +386 5 668 54 85 | Fax: +386 5 668 54 88 )

Iskreno se vam zahvaljujem.  
Želim vam srečno 2015.

Hvala in lep pozdrav,

### 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 prepričati, da bi bili računi po novem letu v sistemu e-računov zavrtni. 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.rtvlo.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.rtvlo.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.rtvslo.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>

**JANES** (••)

IJS 

### 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

**Hvala za pozornost!**

A red correction marker is shown erasing the word "ERROR" written in pencil on lined paper. The marker is positioned diagonally, and red shavings are visible around the erasing point. The email address "damjan.popic@ff.uni-lj.si" is written in blue ink below the word.

[damjan.popic@ff.uni-lj.si](mailto:damjan.popic@ff.uni-lj.si)

Vir: <http://www.e-bis.com.au/bookkeeping-blog/correcting-gst-mistakes/attachment/correcting-gst-mistakes-3>

## Reference

Dobrovoljc, Helena, 2008: Jezik v e-poštnih sporočilih in vprašanja sodobne normativistike. Kožuta, 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. Bjelčevič, 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.

## 3.2 Delavnica - Norma: od pravopisa do tvitopisa

### 3.2.1 Potek delavnic in vsebina raziskovalnih nalog

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>1</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. **Opredelevanje 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.**

---

<sup>1</sup> [word="(?)i][a-zčšž]-.\*"] (e-tvorjenke tipa č-podstava in Č-podstava)  
[word="[a-zčšž][A-ZČŠŽ].\*"] (e-tvorjenke tipa čPodstava)

### 3.2.2 Programski jezik CQL

Nekaj posebnih operatorjev v jeziku CQL:<sup>2</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>mišalovka</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

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

<sup>2</sup> Pripravila Kaja Dobrovoljc.

### 3.2.3 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/sski.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=janer](http://nl.ijs.si/noske/janes.cgi/first_form?corpname=janer)

Polona Gantar in Jaka Čibej:

## 4 FRAZEOLOGIJA V SPLETNI SLOVENŠČINI

**Frazeologija** je veda, ki preučuje **večbesedne leksikalne enote** (VLE) in njihovo pomensko in skladenjsko obnašanje v besedilu. Obsežni korpusi in računalniška orodja omogočajo prepoznavanje **jezikovnih vzorcev** in **tipičnih sopojavitvev 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 **skladenjsko trdna** in jo imamo govorce v zavesti kot **celoto**.

Za našo raziskavo bomo uporabili delitev, ki obsega 3 tipe:

- **KOLOKACIJE** so pogoste sopojavitve 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.





## 4.1 IZROČKI – frazeologija v spletni slovenščini



# Frazeologija v spletni slovenščini

Polona Gantar

Ljubljana, 26. 8. 2015



## Dober dan za frazeologijo

Spoznati pojme in orodja za analizo večbesednih enot

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

- **Frazeologija** je **jezikoslovna veda**, ki preučuje različne tipe **večbesednih enot**, njihove **izrazne, pomenske, skladske in besedilne lastnosti**. Poleg tega z izrazom frazeologija označujemo tudi **zbir frazeoloških enot**, ki jih uporabljamo v določenem jeziku.

## ključne besede

- večbesednost
- pomenska neprozornost
- slovnična „nepravilnost“

## problemi s težavami

### eno- ali večbesednost?

- Alfa romeo
- d. o. o., omg! espe
- slovensko-italijanski
- bati se, smejati se, svitati se ...
- al dente, at hoc, a priori
- naprimer – na primer; hokus pokus – hokuspokus
- popihati jo, cukati ga

## vsak začetek je težak

### pomensko prozorno ali neprozorno?

- **črn**  
črna barva, črna obleka, črne hlače  
črn oblak  
Črno morje, Črna gora, Črni Vrh  
črni kruh, črna kava, črni čaj, črni ribez  
črna vdova  
črna kronika, črni humor, črni petek  
črna gradnja

**JANES** (••) pametni telefon IJS 

**Kaj je frazeologija?**

gosta megla prava čokolada jabolko od drev esa

topla greda varnostni trikotnik frazemi

maslo ... volk iz gozda

terminologija **sobese** vrabec

pregovori ustvarjalnost drugačnost


kozol iz zelja

črna ovca

kolokacije naravna jezikovna raba jezikovni vzorci korpusni viri in orodja

slovarji, priročniki



**JANES** (••) IJS 

**Večbesedne enote**

večbesednost tipičnost	terminološkost	ustvarjalnost drugačnost
<b>KOLOKACIJE</b>	<b>STALNE BESEDNE ZVEZE</b>	<b>PREGOVORI, FRAZEMI</b>
naravna jezikovna raba	pomenska neprozornost	pomenska neprozornost
učenje jezika	neprevedljivost	ekspresivnost
računalniška obdelava jezika	terminološkost	neprevedljivost
korpusi	računalniška obdelava jezika	ustvarjalnost, drugačnost
orodja za analizo korpusov		računalniška obdelava jezika
<b>ZGRADBA</b>	<b>POMEN, FUNKCIJA</b>	<b>JEZIKOVNA RABA</b>

## Kolokacije

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

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

(Firth, J. R. 1957)

### Kolokacije besede *krona*

### Pomeni besede *krona*



## Kolokacije

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

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

(Firth, J. R. 1957)

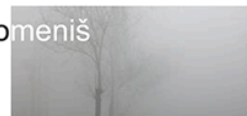
Kolokacije so zveze besed, ki se v jeziku pogosto pojavljajo skupaj.

Pogostost sopojavljanja in trdnost zveze je mogoče izračunati.

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

Kolokacije so zato pomembne pri učenju tujega jezika.

Kolokacije so zato tudi pomembna sestavina slovarjev in jezikovnih učbenikov.



## Stalne besedne zveze

„večbesedne besede“

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

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

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

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

Stalne besedne zveze poimenujejo konkretne predmete, pojave ... navadno na določenem **strokovnem ali terminološkem področju**.



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/ tako	na první pohled	im Überblick	in (less than) no time
skrajni čas	nejvyšší čas/ načase	höchste Zeit	high time/ about time

## Stalne besedne zveze

„večbesedne besede“

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

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

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

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



stara mama ≠ grandmother

prava čokolada ≠ cooking chocolate

## Frazeološke enote

„jezikovne nepravilnosti“

Ena ali dve besedi?

*piš me v uh = pišmevuhovstvo, nebodigatreba*

Katera beseda?

žebljica, glavica, bes, plentaj,

Katera besedna oblika, vrsta?

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

Slovnična pravilnost?

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

Pomenska nepričakovanost

*vlačiti koga po zobeh; nositi srce na dlan; pustiti koga na cedilu; delati iz muhe slona*  
imeti besedo na koncu jezika



Besedilna pravilnost?

*Ti meni, jaz tebi.* – Če ti daš meni nekaj, dam jaz tebi nekaj

## Frazeološke enote

ekspresivnost, drugačnost, opaznost = USTVARJALNOST!

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

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

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

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

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

*kdor drugemu jamljo koplje ...*  
*... ziher ni sebičen*



six pack



one pack = [uampek]

*ljubi svojega bližnjega ...*  
*... samo pazi, da te ne dobijo*

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

*Kdor visoko leta ...*  
*je Peter Prevc*

## Frazeologija v družbi z jezikom

brezposelni ...

in upokojenci  
 diplomanti  
 samohranilka  
**mladi**  
 prejemniki socialne pomoči

oseba  
 in iskalci prve zaposlitve  
**mladi**  
 invalid  
 in upokojenci

... moški

odrasel  
 poročen  
 samski  
 pravi  
**mlad**

poročen  
 sanjski  
 samski  
**mlad**  
 čeden

... ženska

samska  
**mlada**  
 sanjska  
 odrasla  
 sodobna

**mlada**  
 samska  
 sanjska  
 moški in ženska  
**lepa**

... politik

skorumpiran  
 koruptiven  
**vrl**

**neg.**

obsojen  
 nesposoben

vodilen  
**vrl**  
 opozicijski  
 vpliven

**poz.**

**srbski**



## tipično slovensko ...

**Kdo je tipičen slovenski**

**Kaj je tipično slovensko**

**tipična slovenska**

**družina**

folklor  
fora  
zgodba

**tipična slovenska**

**fovšija**

mentaliteta  
razmišljanje

**tipičen slovenski**

**film**

volilec  
voznik

**JANES**

NEGATIVNO

**KRES**

**tipična slovenska**

**jed**

vas  
pokrajina

NEUTRALNO

## Frazeologija in računalniško jezikoslovje

### jezikovni vzorci

@BlogSlovenija mogoče ime in priimek?  @BlogSlovenija **tipična slovenska** družina, fotr pijanc, mama pijanka, samohranilka je ca. 6-8 ur na dan. Živimo skupaj, kot **tipična slovenska** družina, v isti hiši, z ločenim gospodinjstvom kateremu primeru bi se najbolj približala **tipična slovenska** družina? <http://t.co/EpwSaJhU6G>  Zakaj na zapuščinski bodo stopili skupaj? To pa ni **tipična slovenska** družina. :) @SuzanaP24ur  @SustarsicM Ne <http://t.co/H7JwaUC0rV> <http://t.co/qMCWuaL00o>  FOTO: **Tipična slovenska** družina nasmejala Vanečo - <http://t.co/H7JwaUCC> <http://t.co/H7JwaUC0rV> <http://t.co/gnnarpul6Q>  FOTO: **Tipična slovenska** družina nasmejala Vanečo - <http://t.co/H7JwaUCC> največja smet na svetu!  In to so ponavadi **tipične slovenske** družine, kjer očitno preko trpljenja, žaljenja Anja rodila sem se v čisto običajni precej **tipični Slovenski** družini. Saj veš ob nedeljah k maši potem , bi rekli, da je popolnoma enak najbolj **tipični slovenski** družini (ki je seveda globoko disfunkcionalna pripadajočimi disfunkcionalnostmi značilnimi za **tipično slovensko** družino, a vseeno ne razumem, kaj nekatere dvema staršema (če za primer vzamemo le **tipično slovensko** družino, kjer je oče deloholik in cel dan sprejemljivo.  Micl02:  ajaaa, to je un film, ko **tipično slovensko** družino predstavljajo črnici ... WTF, sicer

## frazeologija in računalniško jezikoslovje

IZDELAVA RAČUNALNIŠKO BERLJVIH JEZIKOVNIH OPISOV

=S\_kakšen?/S\_kdo-kaj?  
2:pridevnik brez\_GSD{0,5} 1:samostalnik

# LBS-XX #####  
#/1/ <struktura>PBZ0 sbz2</struktura>  
#/2/ <struktura>pbz0 SBZ2</struktura>

↑ gramatične relacije

PROCESIRANJE NARAVNEGA JEZIKA

$$Dice(f_A, f_B) = \frac{2 \frac{f_{AB}}{N_A} \frac{f_{AB}}{N_B}}{\frac{f_A}{N_A} + \frac{f_B}{N_B}} \approx \frac{2 \frac{f_{AB}}{N}}{\frac{f_A}{N_A} + \frac{f_B}{N_B}} = \frac{2f_{AB}}{f_A + f_B}$$

↑ statistični izračuni

IZDELAVA ORODIJ ZA ANALIZO KORPUSOV

družina (samostalnik)  
JANES v0.3 frekvenca = 24.246 (150.3 na

Constructions			S_kakšen?	7,554	2.9
O_z_lastnim_imenom	241	1.2	enostarševski	210	9.78
O_s_števili	419	1.0	primaren	158	9.14

blazno (prilov)  
JANES v0.3

S_je_pred_p	1,294	31.3
blaten	83	10.27
smešen	43	7.13
kul	40	6.67
seksi	11	6.52
dolgočasen	16	6.34
navdušen	15	6.23
zoprn	8	6.17
ponosen	15	6.13
simpatičen	16	6.12
občutljiv	12	6.08
pameten	48	6.07
fin	8	6.04

kraljev	131	9.01
ogrožen	135	8.74
članski	97	8.48
romski	88	8.42
istospolen	86	8.31
lovski	78	8.24
mlad	377	8.19
difunkcionalen	61	8.02

↑ prikaz podatkov

### 1. analiza korpusa

### 3. zapis jezikovnih vzorcev

- kdo<sub>[1]</sub> zmag
- zmag
- zmag na čem<sub>[2]</sub>
- zmag v čem<sub>[2]</sub>
- zmag v čem<sub>[3]</sub>
- zmag v čem<sub>[4]</sub>
- zmag med kom<sub>[4]</sub> (množina)
- zmag med kom<sub>[5]</sub> (množina)
- zmag pred kom<sub>[5]</sub>
- zmag nad kom<sub>[5]</sub>
- zmag s čim<sub>[6]</sub>

### 2. zapis pomena in udeležencev v „največjem“ stavčnem vzorcu

če ŠPORTNIK ali ŠPORTNA EKIPA<sub>[1]</sub> zmag, doseže na ŠPORTNEM TEKMOMANJU<sub>[2]</sub> v določenem ŠPORTU<sub>[3]</sub> ali glede na določena MERILA<sub>[4]</sub> v odnosu do drugih ŠPORTNIKOV ali ŠPORTNE EKIP<sub>[5]</sub> NAJBOLJŠI REZULTAT<sub>[6]</sub>

### 4. zapis udeležencev

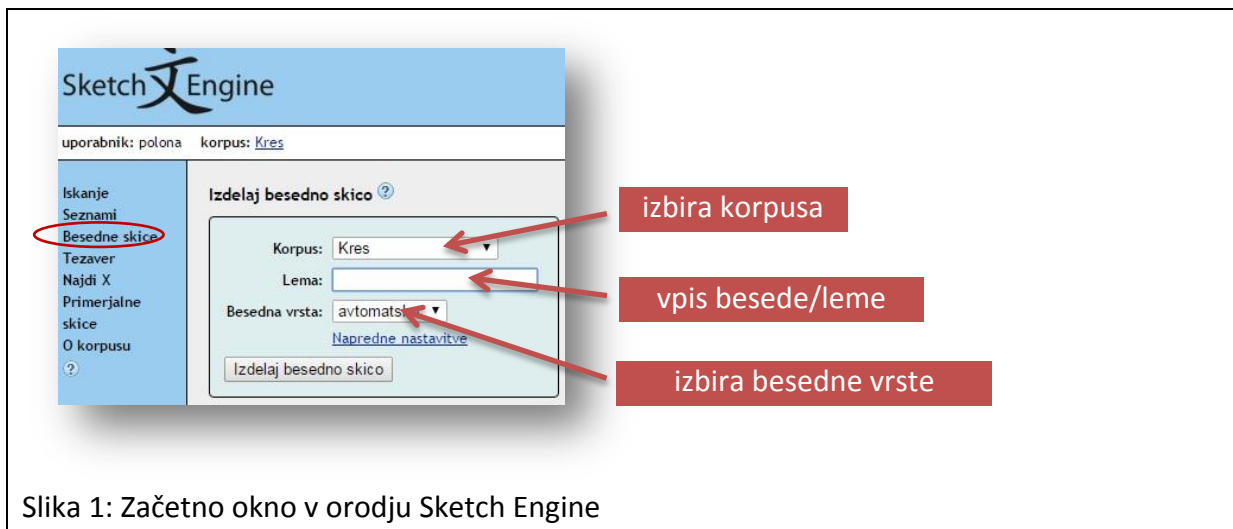
- [1], [5] – ŠPORTNIK, ŠPORTNA EKIPA velemojster, skakalec, smučar, domačin, domačinka, gostitelj, barca, španec, favorit, ekipa ...
- [2] – ŠPORTNO TEKMOMANJE na\_ tekma, na\_turnir, na\_dirka, na\_tekmovanje, na\_miting, v\_tekma
- [3] – ŠPORT
- [4] – MERILA v\_niz, v\_etapa, v\_disciplina, v\_dvoboj, v\_kategorija
- [6] – REZULTAT z\_točka, z\_gol, z\_zadetek

### 5. avtomatsko zgenerirani potencialni vzorci

## 4.2 Vaje

### 4.2.1 Prepoznavanje različnih tipov večbesednih leksikalnih enot v besedilu

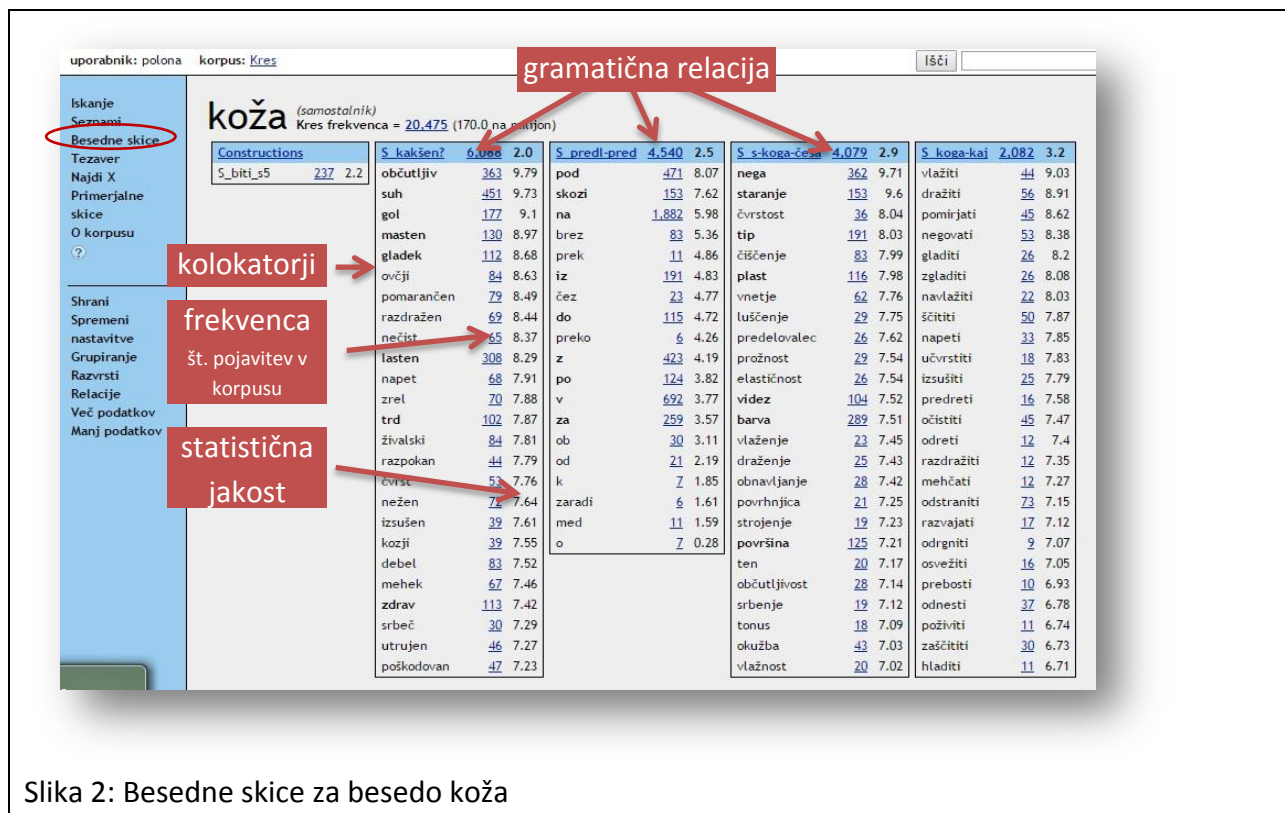
- Priporočljivo je, da si sproti beležite postopek in ugotovitve, saj vam bo to prišlo prav pri izdelavi raziskovalne naloge. Za vnašanje rešitev lahko uporabite kar ta dokument, tako da ga s svojim imenom shranite na računalnik.
- Odpremo orodje **Sketch Engine** ([SkE](#)) oz. [wsketch form](#):



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, janis 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)

kolokatorji

frekvenca  
št. pojavitev v korpusu

statistična jakost

Constructions	S_kakšen?	S_predi-pred	S_s-koga-česa	S_koga-kaj
S_biti_s5	237 2.2	471 8.07	362 9.71	44 9.03
občutljiv	363 9.79	153 7.62	153 9.6	56 8.91
suh	451 9.73	1,882 5.98	36 8.04	45 8.62
gol	177 9.1	83 5.36	191 8.03	53 8.38
masten	130 8.97	11 4.86	83 7.99	26 8.2
gladek	112 8.68	191 4.83	116 7.98	26 8.08
ovčji	84 8.63	čez	23 4.77	62 7.76
pomarančen	79 8.49	do	115 4.72	29 7.75
razdražen	69 8.44	v	6 4.26	26 7.62
nečist	65 8.37	preko	z	29 7.54
lasten	308 8.29	po	423 4.19	26 7.54
napet	68 7.91	v	124 3.82	104 7.52
zrel	70 7.88	za	692 3.77	289 7.51
trd	102 7.87	ob	30 3.11	23 7.45
živalski	84 7.81	od	21 2.19	25 7.43
razpokan	44 7.79	k	7 1.85	28 7.42
čvrst	53 7.76	zaradi	6 1.61	21 7.25
nežen	74 7.64	med	11 1.59	19 7.23
izsušen	39 7.61	o	7 0.28	125 7.21
kozji	39 7.55			20 7.17
debel	83 7.52			28 7.14
mehek	67 7.46			19 7.12
zdrav	113 7.42			18 7.09
srbeč	30 7.29			43 7.03
utrujen	46 7.27			20 7.02
poškodovan	47 7.23			125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02
				125 7.21
				20 7.17
				28 7.14
				19 7.12
				18 7.09
				43 7.03
				20 7.02

- V izhodiščnem okencu (gl. Slika 1) izberemo **korpus KRES**, odtipkamo besedo: *hud* in določimo besedno vrsto: pridevnik.
- Kolikokrat se beseda *hud* pojavlja v korpusu KRES in kolikokrat v korpusu JANES? Ali lahko na podlagi tega kaj sklepamo?
- V dobljeni besedni skici analiziramo **kolokatorje** v prvih dveh stolpcih (relacije: **S\_kdo-kaj?** in **S\_kako-kdaj\_p?**). Kakšen pomen besede *hud* izkazujejo kolokatorji v korpusu KRES? Katera je njihova skupna lastnost?
- Ali v drugem stolpcu kateri kolokator izstopa? Zakaj? Preverimo še konkordance in vir.
- Postopek analize kolokatorjev za besedo *hud* ponovimo še na **korpusu JANES** in primerjajmo rezultate.
- Kateri kolokatorji so prekrivni v obeh korpusih in kateri izstopajo?
- **Zapišimo ugotovitve**

#### 4.2.3 2. Naloga: OD KOLOKACIJE DO FRAZEMA

- Izdelamo besedno skico za samostalnik *tič*.
- Najprej zabeležimo **frekvenco leme** v obeh korpusih. Kaj ugotovimo?
- Analiziramo **kolokatorje** v relacij **S\_kakšen?** v **korpusu KRES**. Katere pomene nakazujejo kolokatorji v prvem stolpcu? Poskusimo vsak pomen ponazoriti z ustrezno **konkordanco** – zgledom iz korpusa.
- Ali so v obeh korpusih prekrivni kolokatorji in pomeni? Kako je s frekvenco kolokatorjev v enem in drugem korpusu?
- Ali pri katerem od kolokatorjev v korpusu KRES najdemo **frazeme**? Kaj pa v JANES-u? Katere? Kaj je zanačilno za njihovo rabo? Izpišimo primere.
- Ali pridejo do izraza še kateri drugi frazemi v obeh korpusih, če analiziramo **konkordance** (V orodju SkE izberemo možnost »Iskanje«)? Pregledamo do 10 kolokacijskih nizov.
- Ali lahko prepoznamo kako **ustvarjalno rabo** frazema (t. i. prenovitev)? Zapišimo si primere in vir (korpus).
- Kako je prenovitev frazema dosežena (npr. združitev dveh frazemov, dobesedna raba ipd.).
- Ali so frazemi, ki jih je mogoče najti v obeh korpusih, opisani v Slovarju slovenskega knjižnega jezika (SSKJ: preverite na <http://www.fran.si/>)?
- Ali najdemo v SSKJ frazeme, ki jih v korpusu KRES in JANES ni? Katere?
- **Zabeležimo ugotovitve**

## 4.3 Delavnice

Delavnice so namenjene samostojnemu raziskovanju v skupinah. Pri analizi uporabljamo korpuse in orodja za njihovo analizo (konkordančnik, besedne skice itd.).

Vsaka raziskovalna naloga naj vsebuje:

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

### 4.3.1 RAZISKOVALNE TEME



- »To se da tud drgač povedat«: analiza frazeoloških prenovitev v spletni slovenščini
- *Leži kot krava in laže kot pes teče*: živalske prispodobe 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 brcne*: 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 ...*

Ana Zwitter Vitez in Špela Arhar Holdt:

## 5 Govor in spletna slovenščina

### 5.1 IZROČKI – govor in spletna slovenščina


**JANES** (••)  

---

## Govor in spletna slovenščina

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

Ljubljana, 27. 8. 2015

**JANES** (••)  

---



- interdisciplinarno področje
- neraziskani rezultati
- pomembne družbene posledice

### O čem bomo govorili

- specifikke govornega jezika
- odnos med govorom, pisno in spletno slovenščino
- kaj lahko besedilo pove o značilnostih avtorja

Od jutra do večera ...





**JANES** (••) IJS 

**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



Na vrh

**JANES** (••) IJS 

gos

Iskanje po pogovornem zapisu  
  Iskanje po standardiziranem zapisu

Najdi


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


<p><b>JANES</b> (••)</p>	<p>IJS </p>
<p><a href="#">a velikokrat ne ostane le pri tem kar smo se lahko prepričali tudi med našim snemanjem</a></p>	<p>Antigona se je kljub temu odločila, da Polinejka pokoplje.</p>
<p><a href="#">ma tu si lahko duol čez rit potjegneš</a></p>	<p>igor-za tekmo bo pa juttr pomoje odloču tomažk hahaha +1 -0      18.08.2015 ob 08:52      Prijavi neprimerno vsebino</p>

**JANES** (••) IJS 

Pišemo standardno, govorimo pa nestandardno?












www.clipartof.com · 1046439





## Pa vendar: specifike 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



**DELO**  
vred. 19.08.2015

Novice Svet Gospodarstvo Mnenja Šport **Kultura** Znanje

Kultura » Knjige » Znanstvena slovenščina (tragedija v dveh dejanjih)

Tweet (2) S+1 (2) Recommend (72)

### Znanstvena slovenščina (tragedija v dveh dejanjih)

Kakšna utegne biti razlika med slovesnostjo ob ustanovitvi slovenske Univerze leta 1919 in ob praznovanju njene stotletnice?

Aljžija Zupan Seod  
pon, 06.01.2014, 21:00

Ključne besede: Univerza, Zakon o visokem šolstvu, slovenščina, angleščina, Aljžija Zupan Seod



 **Neznan** | nedelja, 23. avgust 2015, ob 14:37:01

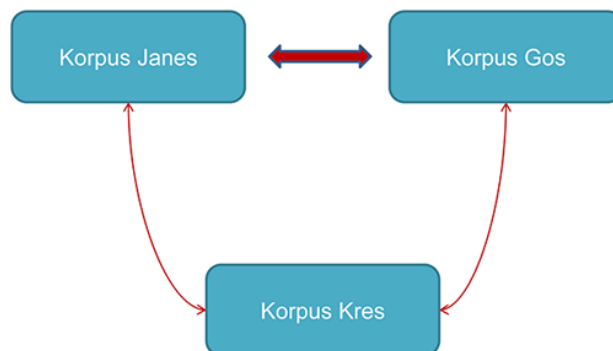
Prava tragedija je v resnici " piši kot govoriš " slovenščina , ki jo piše mladina.

3 K 5 K

Preseganje stereotipov: empirične raziskave



Preseganje stereotipov: empirične raziskave



## 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 v0.3.1 Forum podkorpus: Forum L2,3

Iskanje  
Seznami  
O korpusu  
?

Shrani  
Spremeni  
nastavitve

Seznam besed

Korpus: JANES v0.3.1 Forum  
Podkorpus: Forum L2,3

Referenčni korpus: KRES (uravnoveženi)  
[Spremeni cilj in referenčni \(pod\)korpus](#)

Stran 1 | [Pojdi](#) | [Naslednja >](#)

word	JANES v0.3.1 Forum : Forum L2,3		KRES (uravnoveženi)		Rezultat
	Frekvenca	Frekvenca/mil	Frekvenca	Frekvenca/mil	
lp	18.374	749.3	249	2.1	244.6
feltno	4.603	187.7	8	0.1	177.0
lp	9.247	397.5	195	1.6	152.1
ops	3.554	144.9	14	0.1	130.7
LP	15.595	635.9	563	4.7	112.3
com	4.399	179.4	84	0.7	106.3
felne	1.502	77.6	0	0.0	78.6
xenon	1.941	79.2	6	0.0	76.3
sm	31.049	1266.1	2.033	16.9	70.9
oglasnik	1.770	72.2	12	0.1	66.5
tuning	1.250	79.5	32	0.3	63.6
jst	11.729	480.3	802	6.7	62.8
felge	1.512	61.7	2	0.0	61.6
mam	11.604	473.2	829	6.9	60.2
original	5.093	207.7	310	2.6	58.4

## Metapodatki v korpusu Gos

gos Iskanje Seznam

Iskanje po gos Iskanje Seznam

Iskanje po pogovornem zapisu  Iskanje po standardiziranem zapisu

Uporabljaj

Uporabljaj iskanje po seznamih

Osnovne oblike  1 2

Tip govora

- Nezasebni zapisi
- Javni inform.
- Javni razved.
- Nezasebni razved.
- Več

Kanal

- Osebnost. sbr.
- Tiskovna izd.
- Zasebnost. r.
- Radio (161)
- Več

Standardizirana osnovna oblika

- reči (1.278)

Besedna vrsta

- glasol (1.278)

Tip govora

- Nezasebni (576)
- Javni informativno-izobraževalni (328)
- Javni razvedrilni (213)
- Nezasebni razvedrilni (161)
- Več

loči besede glede na oblikoslovne lastnosti

Prikazujem 1-20 od 30 zadetkov (0.31 sekund)

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

### 3. stereotip: uporabnost 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 ( <i>Čudoviti Klon</i> ), David Benjamin ( <i>Sedem</i> )							
kadrovanje	"ljudje kot kapital in potencial podjetij" (Jackson 1999)							
profiliranje strank in njihovih kupnih navad	strategije ponudbe in oglaševanja (Shaw idr. 2001)							

### Ugotavljanje avtorstva besedil: "Volivci v trenirkah"

uradna spletna stran parlamentarne stranke

psevdonim

vznemirjena javnost:

1. "volivci s tujim naglasom"
2. "volivci v športnih oblačilih (trenirkah)"
3. "s kemičnim svinčnikom napisana številka, ki jo morajo obkrožiti na glasovnici"

### Hipoteza in zajem besedil

Če je avtor besedilo anonimno objavil na uradni spletni strani stranke, je verjetno na isti spletni strani objavil še kakšno besedilo pod drugim/pravim imenom.

#### Kriterij za zajem

- tri mesece pred in tri mesece po objavi spornega besedila

#### Korpus

- 75 besedil 21 avtorjev

- 55.000 pojavnic (650 do 9000 pojavnic na avtorja).

### Priprava besedil

pretvorba v format .txt,  
anonimizacija besedil in tvorjenje glav dokumentov  
oblikoslovno označevanje (Grčar idr. 2012)

Oznaka: Ppnmnid  
Kratka oznaka: Pp  
Lema: relativen

viča sta dvignila umetno ustvarjen strah in slovenska radodarnost z državljanstvi

Relativni zmagovalec predčasnih volitev Zoran Jankovič je na volitvah prejel dobrih 300.000 glasov.

Vsi komentatorji so se veliko ukvarjali s strukturo tega volilnega telesa, ob tem pa so se v celoti namerno izognili analizi strukture njegovih volivcev vidika novih državljanov.

Izračun značilke

Leksikalne značilke:

- raznolikost besedišča (*lexical density*)
- Brunetova formula (Brunet 1988)
- hapax legomena (Holmes 1992)
- Honoréjeva statistika (Honoré 1979)

Berljivostne značilke:

- Flesh-Kincaid
- Coleman-Liau
- Automated Readability Index
- Gunning Fog

<http://www.usingenglish.com/members/text-analysis/help/readability.html>

[http://nl.ijs.si/isit14/proceedings/isit2014\\_23.pdf](http://nl.ijs.si/isit14/proceedings/isit2014_23.pdf)

Rezultati analize

Majer: leksikalne značilke

Avtor	DRB	BI	HS	HL	DRL	BI-L	HS-L	HL-L
G	0.02 ± 0.11	0.14 ± 0.61	70.30 ± 372.30	0.02 ± 0.11	0.02 ± 0.12	0.03 ± 0.95	77.01 ± 193.59	0.01 ± 0.11
A	0.05 ± 0.02	0.11 ± 0.04	56.17 ± 230.80	0.05 ± 0.01	0.06 ± 0.02	0.35 ± 0.02	95.76 ± 254.19	0.05 ± 0.00
H	0.06 ± 0.06	0.22 ± 0.27	15.38 ± 74.07	0.06 ± 0.06	0.06 ± 0.06	0.39 ± 0.44	12.16 ± 83.69	0.05 ± 0.05
O	0.06 ± 0.02	0.26 ± 0.18	420.85 ± 315.92	0.07 ± 0.03	0.06 ± 0.03	0.46 ± 0.33	197.93 ± 154.95	0.06 ± 0.03
N	0.10 ± 0.04	0.48 ± 0.13	1.36 ± 58.86	0.10 ± 0.04	0.09 ± 0.05	0.72 ± 0.29	53.15 ± 134.35	0.08 ± 0.04
D	0.07 ± 0.04	0.34 ± 0.18	196.48 ± 175.76	0.07 ± 0.04	0.10 ± 0.04	0.84 ± 0.30	316.35 ± 83.34	0.09 ± 0.04
P	0.10 ± 0.06	0.44 ± 0.30	183.11 ± 43.80	0.10 ± 0.05	0.10 ± 0.04	0.72 ± 0.32	167.55 ± 109.35	0.09 ± 0.03
M	0.10 ± 0.02	0.47 ± 0.05	242.78 ± 160.22	0.10 ± 0.02	0.12 ± 0.03	1.01 ± 0.13	378.17 ± 135.76	0.12 ± 0.02
J	0.12 ± 0.00	0.48 ± 0.00	38.12 ± 0.00	0.12 ± 0.00	0.15 ± 0.00	1.10 ± 0.00	91.45 ± 0.00	0.13 ± 0.00
B	0.13 ± 0.04	0.57 ± 0.30	443.25 ± 231.55	0.10 ± 0.02	0.16 ± 0.03	1.27 ± 0.34	26.28 ± 367.71	0.14 ± 0.00
E	0.17 ± 0.00	0.78 ± 0.00	434.37 ± 0.00	0.17 ± 0.00	0.17 ± 0.00	1.33 ± 0.00	471.62 ± 0.00	0.18 ± 0.00
F	0.15 ± 0.03	0.74 ± 0.31	279.47 ± 94.69	0.14 ± 0.03	0.17 ± 0.04	1.34 ± 0.48	36.19 ± 117.19	0.15 ± 0.04
V	0.17 ± 0.06	0.95 ± 0.63	68.63 ± 263.70	0.16 ± 0.05	0.17 ± 0.07	1.49 ± 0.86	53.65 ± 303.20	0.15 ± 0.05

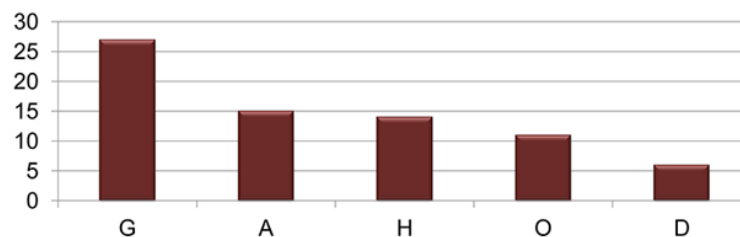


Razlikovalna moč značilk

**Avtor G**

Značilka	Utež
Delež različnih besed	0.79 ± 0.27
Delež različnih lem	0.78 ± 0.26
Hapax legomena	0.77 ± 0.26
Hapax legomena (leme)	0.76 ± 0.26
Brunetov indeks (leme)	0.68 ± 0.30
Brunetov indeks	0.54 ± 0.33
Honoréjeva statistika	0.48 ± 0.27
Honoréjeva statistika (leme)	0.47 ± 0.28
Razmerje med št. besed in št. povedi	0.46 ± 0.27
Razmerje med št. znakov in št. besed	0.45 ± 0.26

Podobnost avtorjev z anonimnim besedilom



Avtorji z najmanjšo povprečno absolutno razliko glede na anonimno besedilo.

**Omejitve raziskave**

križno preverjanje  
prisotnost dejanskega avtorja besedila

**Odprta vprašanja**

dolžnosti raziskovalca  
družbene posledice

**Prihodnost jezikovnih tehnologij**

Povej mi, s kom se družiš, in povem ti, kdo si.  
Povej mi, kaj bereš, in povem ti, kdo si.  
Povej mi, kaj voziš, in povem ti, kdo si.  
Povej mi, kaj ješ, in povem ti, kdo si.  
Povej mi, kako se znojiš, in povem ti, kdo si.

Povej mi **kar koli** in povem ti, kdo si.

**Kaj bomo delali?**

Kako govorimo?

Ali na spletu pišemo, kot govorimo?

Kaj lahko na podlagi podatkov iz korpusov povemo o avtorju besedila?

## 5.2 Detektiv – vaje

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

### 5.2.2 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. Na seznamu ključnih besednih oblik poišči oblike *zaj*, *guor*, *fse*, *vejš*, *tk*, *čeprov*, *ge*, *bul*. 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.

3. 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?

4. 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-2015/>. V novem zavihku odpri spletno stran <http://www.wordle.net/> in izberi zavihek **Advanced**. Prilepi podatke iz tabele v okence in izdelaj besedni oblak.

5. 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?

6. 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?

## 5.3 Detektiv - delavnica (skupine Tvit1, Tvit2, Tvit3)

### 5.3.1 Odkrivanje govornih elementov na Twitterju

**Izdelaj seznam ključnih besednih oblik korpusa Twitter (stopnja L 2, 3) glede na korpus Kres.**

Na spletni strani [http://nl.ijs.si/noske/janes.cgi/first\\_form?corpname=janes](http://nl.ijs.si/noske/janes.cgi/first_form?corpname=janes) izberi možnost **Seznami** in izberi korpus **Janes v 0.3.3 Tweet L2, 3**. Pri **Možnostih izpisa** izberi opcijo **Ključne besede**, za referenčni korpus pa izberi **korpus Kres**.

### 5.3.2 Naloge

1. Ali na tem seznamu najdemo besede *eee*, *mhm*, *eem*? Zakaj?
2. Na seznamu poišči besede *jst*, *dans* (stran 2), *nekej* (stran 3). 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?
3. Na seznamu poišči besede *fajn*, *komot* (stran 2) in *ful*. Kaj je njihova skupna lastnost? Izpiši še tri take besede. Zdaj te besede poišči v korpusu Gos. V katerem tipu diskurza prevladujejo?
4. Na seznamu ključnih besed poišči besede *videoposnetek*, *tvit* in *tnx*. Jih najdeš v korpusu Gos? Zakaj?
5. Seznam ključnih besed korpusa Gos in podkorpusa tvitov 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.

## 5.4 Detektiv - delavnica (skupini Forum1 in Forum2)

### 5.4.1 Odkrivanje govornih elementov na spletnih forumih

**Izdelaj seznam ključnih besed korpusa Forum (stopnja L 2, 3) glede na korpus Kres.**

Na spletni strani [http://nl.ijs.si/noske/janes.cgi/first\\_form?corpname=janes](http://nl.ijs.si/noske/janes.cgi/first_form?corpname=janes) izberi možnost **Seznami**. Nastavi korpus **Janes v 0.3.1 Forum** in podkorpus Forum L2, 3. Pri **Možnostih izpisa** izberi opcijo **Ključne besede**, za referenčni korpus pa izberi **korpus Kres**.

### 5.4.2 Naloge

1. Ali na tem seznamu najdemo besede *eee*, *mhm*, *eem*? Zakaj?
2. Na seznamu poišči besede *lahk*, *jest* in *kak*. 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?
3. Na seznamu poišči besede *vredu*, *pomoje* in *nevem*. Kateri pojav zaznamuje te besede? Poskusi te besede poiskati v korpusu Gos: dobiš več rezultatov z zapisom skupaj ali narazen?
4. Na seznamu ključnih besed poišči besede *feltne*, *oglasnik* in *turbina*. Jih najdeš v korpusu Gos? Zakaj?
5. 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.

**Tamara Mikolič Južnič in Katja Zupan:**

**6 Slovenščina v stiku z drugimi jeziki**

**6.1 IZROČKI – slovenščina v stiku z drugimi jeziki**



**JANES** (••)  

---

## Slovenščina v stiku z drugimi jeziki

Tamara Mikolič Južnič

Ljubljana, 28. 8. 2015

**JANES** (••) **O čem bomo govorili**  

---

- Uvodna vprašanja
- Jeziki v prostoru in času
- Pojavljanje "tujega" v slovenščini
- Primerjava med govorom in pisnim jezikom na spletu
- Nekaj statistike o tujejezičnih elementih v korpusu JANES
- Zapisovanje tujejezičnih elementov
- Oblike tujejezičnih elementov
- Načini preklapljanja
- Vzroki za preklapljanje

## Uporabljate tuje besede v vsakdanji komunikaciji?

- Katere?
- Kateri tuji jezik najpogosteje?
  - angleščina  
aja tvoj twingec je tud **ful kjut**
  - italijanščina  
En postanek in **clao!**      Lepo bodi. **Čau.**
  - nemščina  
menda tut tam velja, da je **muss** samo umret.
  - južnoslovanski jeziki  
Ma **ajde bračo.**

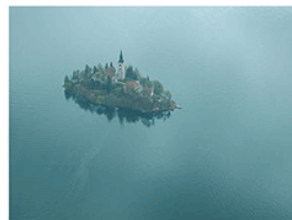
## Smo res tako posebni?

- 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



## 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
  - tujke
  - citatne besede



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

## "Tuje" v slovenščini

- **Preklapljanje** (code switching)
  - uporaba drugega jezika (ali drugih jezikov) poleg svojega maternega jezika, izmenjavanje različnih jezikov
  - pogoj za preklapljanje:
    - (vsaj) dvojezičnost
  - Zakaj preklapljamo – jezikovni vidik:
    - premagovanje jezikovnih omejitev
    - učinkovitost
    - uspešn(ejš)a komunikacija
    - ...

O tem več kasneje!

### 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:
  - **309.554** (4.507,0 na milijon) besed označenih kot tujejezične
  - 142.850 tujejezičnih besed z malo začetnico (2343,1 na milijon)
  - Tviteraši: 70 % vsaj 3 tujejezične besede, 40 % vsaj 10
- Blogi:
  - **173.721** (4.529,7 na milijon)
- Forumi:
  - **57.342** (3.833,9 na milijon)
- Komentarji:
  - **259.425** (5.518,1 na milijon)

*Lastna imena  
Imena pesmi  
Citati*

## Kako zapisujemo tujejezične elemente?



- citatni zapis
- poslovenjeni zapis (kakšen?)
- okrajšave
- ponavljajoče se črke ...




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 **liiliikee** !!!!

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** pravilnega 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** :)

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

## Kako zapisujemo tujejezične elemente?

p   <a href="#">N</a>	cool	485	
p   <a href="#">N</a>	kool	40	
p   <a href="#">N</a>	coolu	1	

p   <a href="#">N</a>	kul	17,794	
p   <a href="#">N</a>	kull	84	
p   <a href="#">N</a>	kulj	71	
p   <a href="#">N</a>	kuuul	29	
p   <a href="#">N</a>	kuul	16	
p   <a href="#">N</a>	kuuuul	10	
p   <a href="#">N</a>	huuuuuu	6	
p   <a href="#">N</a>	kuuuuuul	2	
p   <a href="#">N</a>	kuuuuul	2	
p   <a href="#">N</a>	kuuu	1	
p   <a href="#">N</a>	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	<b>totally</b>	random ponujanja. A danes pa čigumije posebej
Point je v temu, da ne štekam tega totally	<b>random</b>	ponujanja. A danes pa čigumije posebej
časopis na Petrolu in ti model ponudi še	<b>munchmallow</b>	. Ne zastonj, za plačat. Totally random
še munchmallow. Ne zastonj, za plačat.	<b>Totally</b>	random much? ##g Evo, jutranja joga obraza
munchmallow. Ne zastonj, za plačat. Totally	<b>random</b>	much? ##g Evo, jutranja joga obraza opravljena
D ##g @Razdelilec Sicer je pa bolj djembe	<b>trained</b>	, ne bobni bobni. ##g @mgerencer Okej, to
je zelena v fris. Center za ravnotezje	<b>not</b>	happy. #hrup ##g @multikultivator intervju
zelena v fris. Center za ravnotezje not	<b>happy</b>	. #hrup ##g @multikultivator intervju je
priimek, bo North Noth. Kar je še bolj	<b>freaky</b>	kot North West. ##g Behind the Candelabra
Kar je še bolj freaky kot North West. ##g	<b>Behind</b>	the Candelabra smo prevedli Moje življenje
napredovala iz 91/54 na 99/64. By doing	<b>nothing</b>	. Spi se mi. #pritisk ##g Ah, čudoviti počitniški
Lihkar je bil. ##g Odločila sem se, da imam	<b>Antisocial</b>	personality disorder. #SelfDiagnoseWednesday
. ##g Odločila sem se, da imam Antisocial	<b>personality</b>	disorder. #SelfDiagnoseWednesday ##g @multikultivator
menstruacijo? VE-DNO. ##g @There_is_no_I	<b>That</b>	makes no sense. A glavobol tudi gre, če
dezert: D ##g @2sto7 Ah, štekam. A " coffee	<b>to</b>	go " imajo in a je dražji ali cenejši?
dezert: D ##g @2sto7 Ah, štekam. A " coffee to	<b>go</b>	" imajo in a je dražji ali cenejši? :) ##
, ki " razbija barvo globoko v laseh ".	<b>I</b>	lol'd. #reklame ##g Za malo manj načičkano
ki " razbija barvo globoko v laseh ". I	<b>lol'd</b>	. #reklame ##g Za malo manj načičkano poroko
. @opica @TomTomanic ##g @photography_A_O	<b>No</b>	shit, tudi na Pošti lahko kupiš vse, pa
@vesnchek Zato je Subway že prvič propadel -	<b>too</b>	many fucking questions! : D @davidkovic

## Vrste preklapljanja

- **Besedno preklapljanje**

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

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

Kak **response** dobiš?

**Prezemanje ali preklapljanje?**

## 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 s/!** (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?

- 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)

## 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!

## 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.**

## 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 Googla, 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?

## 6.2 IZROČKI – vaje



**JANES** (••)  

---

**Od Novega mesta do Njujorka**  
Vaje

Tamara Mikolič Južnič

Ljubljana, 28. 8. 2015

**JANES** (••)  

---

I.



## I. Kako zapisujemo tujejezične elemente?

1. Se pogosteje uporablja citatni zapis ali poslovenjene oblike?

- a. Odpremo NoSke, izberemo korpus JANES v. 0.3.
- b. V iskalno okence za enostavno iskanje vpišemo naslednje besede v citatnih oblikah in poiščemo njihovo pogostnost.
  - like
  - good
  - please

## I. Kako zapisujemo tujejezične elemente?

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

- a. Kako lahko poiščemo poslovenjene oblike?
- b. Katere so po vašem mnenju poslovenjene oblike zgornjih besed?
- c. 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).
  - *please*
  - *good*
  - *like*

## I. Kako zapisujemo tujejezične elemente?

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

II.

## II. Koliko je tujejezičnih elementov v JANESu 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.

- CQL: [tag="Nj"]
- Koliko je vseh takih konkordanc?

## II. Koliko je tujejezičnih elementov v JANESu in iz katerih jezikov so?

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

- Vtipkajmo enak iskalni ukaz v polje CQL.
- Zakaj prihaja do take razlike?

## II. Koliko je tujejezičnih elementov v JANESu in iz katerih jezikov so?

- c. Ali se prisotnost tujejezičnih elementov spreminja glede na besedilni tip? Poiščite skupno število tujejezičnih elementov v vsakem izmed štirih podkorpusov JANESa.
- Na vrhu izberite podkorpus, nato kot vrsto iskanja izberite CQL in v ukazno polje vpišite [tag="Nj"].
  - Izberite posamezni podkorpus in preverite število konkordanc.  
BLOGI (JANES v0.3 Blog)  
KOMENTARJI (JANES v0.3.1 Comment)  
FORUMI (JANES v 0.3.1 Forum)  
Tвити (JANES v0.3.3 Tweet)

## II. Koliko je tujejezičnih elementov v JANESu in iz katerih jezikov so?

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?
- Izberite si enega izmed podkorpusov, poiščite vse tujejezične elemente v tem podkorpusu.
  - Napravite frekvenčni seznam besed.
  - Nato na prvih 10 straneh frekvenčnega seznama »ročno« preverite, kateri jeziki se pojavljajo (poleg angleščine). Zapišite te jezike (in primere rabe, ki ste jih našli).

## II. Koliko je tujejezičnih elementov v JANESu in iz katerih jezikov so?

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?

III.

### III. Kdaj in kako se uporablja tujejezične elemente?

1. Kako se pojavljajo tujejezični elementi? Kakšne vsebine bodo izražene v tujih jezikih?
  - a. Na kakšen način bi se lahko pojavljali tujejezični elementi v korpusu JANES?

### III. Kdaj in kako se uporablja tujejezične elemente?

- b. Ustvarite seznam konkordanc vseh tujih besed v korpusu JANES.
  - Izberite korpus JANES v0.3.
  - Pri vrstah iskanj izberite CQL in v iskalno polje vpišite [tag="Nj"].
- c. Ali se primeri, ki ste jih našli, ujemajo z vašimi predvidevanji? Navedite vsaj 5 načinov, na katere se pojavljajo tujejezični elementi.

### III. Kdaj in kako se uporablja tujejezične elemente?

- d. Kateri način vam je najbližji oz. kaj sami uporabljate v podobnih okoliščinah?
- Napišite tri besede/besedne zveze, ki jih najpogosteje uporabljate in primerjajte njihovo pogostnost v vseh štirih podkorpusih.
  - Besedo/besedno zvezo vpišite v polje za enostavno iskanje (izberite vsak podkorpus posebej).
  - Prepišite kolikokrat na milijon se pojavlja v posameznem podkorpusu.

### III. Kdaj in kako se uporablja tujejezične elemente?

Beseda/ besedna zveza	Tweet	Forum	Blog	Comment

### III. Kdaj in kako se uporablja tujejezične elemente?

- e. Kdaj se po vašem mnenju avtorji odločajo za preklapljanje?
- Poiščite vse tuje besede v podkorpusu Tweet.
  - Napišite, kateri so razlogi za preklapljanje v opazovanih primerih.
  - Ali avtorji izrazito preklapljajo na kakem posebnem mestu v svojih tvitih?

### IV.



#### IV. 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.

#### IV. Kdo uporablja tujejezične elemente?

- a. zasebni uporabniki = 'private'  
podjetja, organizacije in drugi javni subjekti = '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.
  - Nato gremo nazaj na iskanje in pri TEXT.SOURCE izberemo možnost 'private'.
  - Z iskanjem CQL poiščemo vse tujejezične elemente v teh tvitih.

#### **IV. Kdo uporablja tujejezične elemente?**

2. Zabeležen je tudi spol uporabnikov.
  - a. Ali tujejezične elemente več uporabljajo ženske ali moški?
    - Pri lastnostih besedil izberemo spol avtorja.

#### **IV. Kdo uporablja tujejezične elemente?**

3. Kdo je v svojih tvitih uporabil absolutno največ tujejezičnih elementov?
  - a. Preverimo v podkorporu TWEET
    - Izdelamo konkordance vseh tujejezičnih elementov v podkorporu TWEET.
    - Izberemo Frekvence in v spodnjem kvadratu (Frekvenčna razporeditev po lastnostnih besedil) Izberemo prvo možnost (text.author).

#### IV. Kdo uporablja tujejezične elemente?

- b. Kaj pa če upoštevamo relativno frekvenco, kdo najpogosteje uporablja tujejezične elemente?
- Na zgoraj izdelanem frekvenčnem seznamu kliknemo na napis »Rel. frekvenca (%)«

#### IV. Kdo uporablja tujejezične elemente?

- c. 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.

## V.

### V. Stopnja (ne)standardnosti in tujejezični elementi

1. Podkorporus 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?

## V. Stopnja (ne)standardnosti in tujejezični elementi

- a. Kje je po vašem mnenju največ tujejezičnih elementov?
- 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"].

## V. Stopnja (ne)standardnosti in tujejezični elementi

- c. Enako naredimo za besedila, označena z L2 in L3.
- d. Razmislimo o razlogih za take rezultate. Pobrskajte po konkordancah za posamezne oznake nestandardnosti in preverite, v kakšnem sobesedilu se kje pojavljajo tujejezični elementi.

## V. Stopnja (ne)standardnosti in 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"].
    - Preverimo tujejezične elemente pri moških/ženskah.

## V. Stopnja (ne)standardnosti in tujejezični elementi

- 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«.

## 6.3 Vaje: od Novega mesta do Njujorka

### 6.3.1 Kako zapisujemo tujejezične elemente?

#### 1. Se pogosteje uporablja citatni zapis ali poslovenjene oblike?

- Odpremo JANES v. 0.3.: [http://nl.ijs.si/noske/janes.cgi/first\\_form?corpname=janes](http://nl.ijs.si/noske/janes.cgi/first_form?corpname=janes)
- 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.

- Kako lahko poiščemo poslovenjene oblike?
- Katere so po vašem mnenju poslovenjene oblike zgornjih besed?
- 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?

---

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

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

- 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 ? \_\_\_\_\_
- Za primerjavo odprimo tokrat govorni korpus GOS (<http://www.korpus-gos.net/>) in preverimo, koliko je tujejezičnih elementov v tem korpusu.
  - Vtipkajmo enak iskalni ukaz v polje CQL. Rezultat je \_\_\_\_\_.
  - Zakaj prihaja do take razlike?
- 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 v0.3 Blog): \_\_\_\_\_  
Komentarji (JANES v0.3.1 Comment): \_\_\_\_\_  
Forumi (JANES v0.3.1 Forum): \_\_\_\_\_  
Tviti (JANES v0.3.3 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. Naredite 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?

---

### **6.3.3 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 v0.3.
- 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.

Beseda/Bes. zveza

Tweet Forum

Blog

Comment

---

---

---

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

---

---

---

---

---

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

---

---

### 6.3.4 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:

\_\_\_\_\_

b. 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:

\_\_\_\_\_.

- c. 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.

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

#### 1. Podkorporus 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 podkorporus 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 podkorporus 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: \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_