

# Korpus [GO]vorjene [S]lovenščine – GOS



**DARINKA VERDONIK – FERI UM**  
**PROGRAMSKA SKUPINA: NAPREDNE METODE INTERAKCIJ V**  
**TELEKOMUNIKACIJAH**

**SODELAVCI: ANA ZWITTER VITEZ – TROJINA**  
**IZTOK KOSEM – TROJINA**  
**SIMON KREK – AMEBIS D.O.O.**  
**MARKO STABEJ – FF UL**  
**JANA ZEMLJARIČ MIKLAVČIČ – FF UL**



MINISTRSTVO ZA ŠOLSTVO IN ŠPORT

MINISTRSTVO ZA VISOKO ŠOLSTVO, ZNANOST IN  
TEHNOLOGIJO



# Kaj je govorni korpus



- Elektronska in v računalniško berljivem zapisu narejena zbirka transkripcij pogovorov.
- Lahko vsebuje tudi povezavo z zvočnimi posnetki, na podlagi katerih so narejene transkripcije.
- Lahko ima uporabniški vmesnik, ki omogoča avtomatizirano iskanje po korpusu na podlagi iskanja besed ali fraz.

# Korpus GOS



Gos - korpus govora x

← → ↻ 🏠 [www.korpus-gos.net](http://www.korpus-gos.net) ☆ ☰

🇪🇺 European Co... ★ Borza projektn... 📁 Web of Knowl... 🇮🇹 International J...

gos | Iskanje ▾ | [Seznam](#) [Pomoč](#) | [O korpusu](#) | [Slovensko ▾](#)

## gos



Iskanje po pogovornem zapisu  Iskanje po standardiziranem zapisu

[Najdi](#)

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

---

[Prijombe in komentarij](#) | ©2010 FERi UM, Trojina, Amebis, FF UL

  **Naložba v vašo prihodnost**  
OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA  
Evropski sklad za regionalni razvoj

Iskanje po pogovornem zapisu
  Iskanje po standardiziranem zapisu

itak

Najdi

Uporabljaš [enostavno iskanje](#)[Napredno iskanje](#)[◀ prejšnja stran](#)

1 2 3 4 5 6 7 8 9 10

[naslednja stran ▶](#)

Prikazujem 151-200 od 489 konkordanc (1.778 sekund).



## Osnovne oblike

▶ [itak \(489\)](#)

## Tip govora

- ▶ [Nejavni zasebni \(363\)](#)
- ▶ [Javni razvedrilni \(49\)](#)
- ▶ [Nejavni nezasebni \(49\)](#)
- ▶ [Javni informativno-izobraževalni \(28\)](#)
- [Več](#)

## Kanal

- ▶ [Osebni stik \(275\)](#)
- ▶ [Telefon \(152\)](#)
- ▶ [Radio \(31\)](#)
- ▶ [Televizija \(31\)](#)
- [Več](#)

## Regija snemanja

- ▶ [Ljubljanska \(221\)](#)
- ▶ [Novomeška \(39\)](#)
- ▶ [Celotna Slovenija \(38\)](#)
- ▶ [Mariborska \(29\)](#)
- ▶ [Celjska \(28\)](#)
- [Več](#)

## Leto snemanja

- ▶ [2009 \(282\)](#)

sem že kupu // eee jez tud tistga ne k tistu <b>itak</b> ne mislim nekej / če grem enkrat na mesec // pa se	
še petkrat po moje nisl šov / ma sej ti maš <b>itak</b> Brlog tam // [2 smehna] Brlog ja // ej ka se da tam kje	
je zdej zihr ful uporabljano ne // niti ne ne / [2 smehna] je <b>itak</b> sam televizija noter / zdej ne vem k bo un komad	
ven vrgl sploh pač ni več legalno notr // aja ja <b>itak</b> ja / ka se še nisi posušil // ne / lej kuk sem	
sej za srednji ja // in pol se je spomnu de <b>itak</b> [priimek] ne ker mata pa [priimek] pa [priimek] to / de	
ni sem z [ime] govoru ja ja sej v obvestilih <b>itak</b> nč ne piše ne / sam pač on je reku razlagov	
ne ubije maš še zmeri mal zdravja // ne ne ne <b>itak</b> da te ne sam včas da te v vodo pomeče	
vprašanje če bi še na iksepjih delal // ja ne misim <b>itak</b> da bi delal // v dosu [1 neraz] // [1 delno] če pr men	
kul jez bi bil [1 neraz] / pa bi // sam jez ga <b>itak</b> nisem nikdar špilov tko da mi / ja ne vš koga	
z [ime] ko sem igral ne // ja on je pa <b>itak</b> nadudlan te stvari // on je pa specializiran veš kake stvari	
pa kako bi reku [1 glas] pač nehov in mi je <b>itak</b> vsi statistike so mi šle dol ne // ja sam kok	
ja no // sam še te eseje pišem / skoz skoz // ja <b>itak</b> // dones sem // pa sej s ga hiter spisu no / mislm	
še naredu // ja // k drugač mhm ta Fdv stari // ja <b>itak</b> // [2 smehgo] sam veš tko je [1 neraz] // [1 premoj] [prazen] // sam skoz s:	
esej seminarska ja // seminarska / ja vem ja / druge ni / pač <b>itak</b> una [2 zvok] [2 zvok] ga j psihirala na polno in eee / ne vem	
pač jst sem tko sklenu kompromis / una j po moje <b>itak</b> hotla že da bi jutv vse poslal sam jst sem	
v četrttek pač neki sfuramo // [1 premoj] [prazen] // je pa pač <b>itak</b> bedna k je miljavžnt stvari še drugih / nimaš kej // ja	
sreč da dobim tisto / pizda uno šestko sedemko // pozitivko / ja <b>itak</b> // res se mi ne da jebat še s tem // ja	

about\_table4\_... .gif

about\_table1\_... .gif

Pokaži vse prenose ...

# Opis končnega korpusnega gradiva



TIP DISKURZA	KANAL	ŠT. BESED	%
JAVNI INFORMATIVNO- IZOBRAŽEVALNI	televizija	104.030	10,1%
	radio	95.117	9,2%
	osebni stik	153.997	14,9%
JAVNI RAZVEDRILNI	televizija	104.955	10,1%
	radio	125.567	12,1%
NEJAVNI NEZASEBNI	osebni stik	122.031	11,8%
	telefon	33.862	3,3%
NEJAVNI ZASEBNI	osebni stik	226.530	21,9%
	telefon	69.012	6,7%
<b>SKUPAJ:</b>		<b>1.035.101</b>	<b>100,00%</b>

# Dostop do GOS-a



- Prosto dostopen, samo za nekomercialno rabo:
  - licenca Creative Commons: »nekomercialno« + »priznanje avtorstva« + »deljenje pod istimi pogoji«
- Preko spletnega konkordančnika:
  - ki je zasnovan uporabnikom prijazno
  - in upošteva potrebe raziskovalcev iz različnih vej
  - ter skuša z enostavnostjo uporabe približati korpus tudi uporabnikom v izobraževanju in različnih poklicih
- Neposredno do tekstovnega dela gradiv za zahtevnejše uporabnike (XML, TEI)

# Zakaj?



- Ker vemo zelo malo o tem, kako govorimo v vsakdanjem življenju – v spontanem, nepripravljenem pogovoru ali javnem diskurzu. To ni niti standardizirani (predpisani) jezik niti čisto (idealizirano) narečje, temveč živa, resnična govorica.
- Kakšno slovenščino torej v resnici govorimo? To lahko izvemo le tako, da jo posnamemo v njeni čim bolj avtentični različici, jo zapišemo ter nato raziščemo in poslušamo. Zato smo zgradili GOS.

# Za koga



- Za vse, ki želijo raziskovati govorjeno slovenščino, bodisi z jezikoslovnega bodisi s kakega drugega, recimo sociološkega ali jezikovnotehnološkega vidika.
- Za razvoj govornih tehnologij, zlasti avtomatske razpoznavne govora.
- Ostali uporabniki: GOS-ov spletni vmesnik je uporabniško nadvse prijazen in enostaven, zato da bi ga uporabljali tudi:
  - učitelji v šoli pri pouku slovenščine ali v jezikovnih tečajih za tujce,
  - lektorji govora na radiu, televiziji ali v gledališču,
  - tolmači, pisci in drugi, ki se tako ali drugače srečajo z vprašanji, povezanimi z govorjeno slovenščino.



# Kdo



- Projekt Sporazumevanje v slovenskem jeziku: [www.slovenscina.eu](http://www.slovenscina.eu) in projekt e-vsebin
- Institucije:
  - FERI UM, programska skupina Napredne metode interakcij v telekomunikacijah
  - TROJINA – zavod za uporabno slovenistiko
  - AMEBIS d.o.o.
  - FF UL
- Ustvarjalci GOS-a in spletnega vmesnika: sodelavci v projektih, študenti FF UL, FDV, FF MB, FERI MB
- Govorci, ki so dobrohotno dovolili vstop v svojo zasebnost: sorodniki, govorci, znanci ... snemalcev in sodelavcev projekta, učitelji idr.
- Institucije, ki so odstopile posnetke ali dovolile snemanje pod svojim okriljem: medijske hiše, osnovne in srednje šole, podjetja



Več informacij: [www.korpus-gos.net](http://www.korpus-gos.net), O korpusu

Hvala za pozornost!