



Slovenščina 2.0

# Revija Slovenščina 2.0 in smernice RDA

Špela Arhar Holdt, FRI UL

Vojko Gorjanc, FF UL

Tomaž Erjavec, IJS

CC BY-SA



# Slovenščina 2.0

Spletna revija *Slovenščina 2.0* izhaja dvakrat letno od leta 2013 in nagovarja raziskovalno ter pedagoško sfero, ki vključuje poučevanje in raziskovanje jezika ter razvoj orodij za računalniško obdelavo jezika in govora. Področje objav so primarno **empirične raziskave slovenščine** ter interdisciplinarne in aplikativne raziskave z različnimi poststrukturalnimi pristopi. Zlasti pri slednjih velja odprtost revije za objave avtorjev, katerih raziskave niso povezane s slovenščino. Jezika objav sta slovenščina in angleščina.

Revije Filozofske fakultete UL: <https://revije.ff.uni-lj.si/slovenscina2/>.

# Za vtis glede vsebine: tematske številke

- 2013: **Jezikovne tehnologije** (ur. Tomaž Erjavec, Jerneja Žganec Gros)
- 2014: **Leksikografija** (ur. Iztok Kosem, Michael Rundell)
- 2016: **Računalniško posredovana komunikacija** (ur. Darja Fišer)
- 2017: **Slovenščina v dvojezičnih okoliščinah** (ur. Vojko Gorjanc, Iztok Kosem, Matejka Grgič, Sonja Novak Lukanovič)
- 2018: **Slovnične raziskave za jezikovni opis** (ur. Špela Arhar Holdt, Polona Gantar, Vojko Gorjanc, Robert Grošelj)
- 2019: **Jezik in družbeni spol** (ur. Vojko Gorjanc, Marko Stabej)

# Značilnosti revije, ki jih je treba upoštevati

- Nacionalno relevantna, s tem tudi omejena tematika (slovenščina).
- Večina avtorjev je domačih raziskovalcev, mednarodne avtorje pritegne neposreden uredniški angažma, npr. pri tematskih številkah.
- Novi empirični slovenistični pristopi proizvajajo in uporabljajo **jezikovne podatke**, kot so besedilni korpusi, ročno označena besedila in računalniški leksikoni.
- Metodologija zbiranja, objavljanja in citiranja teh podatkov se vzpostavlja v zadnjih letih (npr. po zaslugi CLARIN.SI – o tem kasneje).
- **V uredništvu želimo spremeniti odnos ciljne raziskovalne skupnosti do področnih raziskovalnih podatkov. Ker je potencialnih avtorjev malo, začinjamo z mehкими pristopi (priporočila, možnost izjem).**

# Tipični raziskovalni podatki

- **Scenarij 1:** Članek opisuje nove raziskovalne podatke, npr. določeno ročno označeno učno množico, ki je širše uporabna za razvoj jezikovnih tehnologij, postopkov za strojno procesiranje slovenščine in podobno.
  - Zagotoviti je treba, da so v ustrezno urejeni in dokumentirani obliki podatki dostopni pod ustrezno licenco. Za tovrstne vire skrbi repozitorij CLARIN.SI.
- **Scenarij 2:** Članek uporablja obstoječo podatkovno infrastrukturo, npr. podatke z repozitorija CLARIN.SI, korpusne podatke iz konkordančnikov ali od kod drugod.
  - Zagotoviti je treba, da so uporabljeni raziskovalni podatki ustrezno citirani in da je njihova raba v članku ustrezno razložena. Ključno je zagotavljanje ponovljivosti raziskave (opredelitev verzije vira, točnega iskalnega pogoja itd.).

# CLARIN.SI



# CLARIN.SI



- CLARIN ERIC: Evropska raziskovalna infrastruktura za skupne jezikovne vire in tehnologije (Common Language Resources and Technology Infrastructure).
- Sedež na Nizozemskem, 20 držav članic + 4 opazovalke.
- CLARIN.SI: Slovenska nacionalna infrastruktura CLARIN (od 2014).
- Domuje na Institutu „Jožef Stefan“ (E8, E3, CMI).
- CLARIN.SI je organiziran kot konzorcij 11 partnerjev: 4 univerze, 3 instituti, 2 društvi, 2 podjetji.

# CLARIN.SI storitve

- **Certificirani repozitorij jezikovnih virov in orodij:**

- dolgotrajno hranjenje
- avtentikacija in avtorizacija
- stalni identifikatorji
- eksplicitni pogoji uporabe in licence
- principi FAIR



re3data.org  
REGISTRY OF RESEARCH DATA REPOSITORIES  
<http://doi.org/10.17616/R33M0X>  
CLARIN.SI repository



- **Spletni konkordančnik:**  
**orodje za analizo velikih in označenih besedilnih korpusov.**
- Spletne storitve za jezikoslovno analizo.
- Druga orodja in spletne storitve.
- Center znanja za južnoslovanske jezika (skupaj s CLARIN-BG).
- Financiranje manjših projektov.
- Podpora dogodkom (npr. JT-DH 2020).

# CLARIN.SI repozitorij

- Okoli 160 jezikovnih virov, od tega 123 slovenskih:
  - skoraj vsi pod eno od licenc Creative Commons
- Npr.
  - [Slovenian parliamentary corpus siParl 1.0 \(1990-2018\)](#)
  - [Corpus of Academic Slovene KAS 1.0](#)
  - [Spoken corpus Gos VideoLectures 4.0](#)
  - [Digital library and corpus of historical Slovene IMP 1.1](#)
  - [Semantic lexicon of Slovene sloWNet 3.1](#)
  - [Twitter sentiment for 15 European languages](#)
  - [Bilingual terminology extraction dataset KAS-biterm 1.0](#)
  - [Slovenian-German Dictionary of Maks Pleteršnik \(1894-1895\)](#)
  - [Emoji Sentiment Ranking 1.0](#)
  - [ELMo embeddings model, Slovenian](#)



Za citiranje vnosa uporabite naslednjo referenco ali jo izvozite v prednastavljeno obliko:

BIBTEX

CMDI

Verdonik, Darinka; Potočnik, Tomaž; Sepesy Maučec, Mirjam and Erjavec, Tomaž, 2017, *Spoken corpus Gos VideoLectures 2.0 (transcription)*, Slovenian language resource repository CLARIN.SI, <http://hdl.handle.net/11356/1158>.





# Ubeseiditev smernic

**Revija *Slovenščina 2.0* podpira odprt dostop do raziskovalnih podatkov.**

V sodelovanju z Nacionalnim vozliščem [Research Data Alliance in Slovenia](#) in skladno z načeli odprte znanosti si prizadevamo za odprt dostop do raziskovalnih podatkov. Odprti podatki predstavljajo dragocen potencial za nadaljnjo rabo in omogočajo preverbo ter ponovljivost znanstvenih sklepov. Avtorjem zato priporočamo, da upoštevajo sledeče smernice:

1. Če v svoji raziskavi **uporabljate obstoječe podatkovne vire**, poskrbite za njihovo ustrezno citiranje. Način citiranja tipično opredeljuje repozitorij, kjer so podatki na voljo. Smernice za citiranje virov z repozitorija CLARIN.SI, ki je referenčna točka za jeziko(slo)vne podatkovne vire, so [tukaj](#).
2. Citiranje se pričakuje tudi v primeru, da **uporabljate podatke, ki so na voljo v uporabniških vmesnikih**, npr. v korpusnih konkordančnih ali na slovarskih portalih. V teh primerih se navede v metodologiji spletna stran vira in (če je to smiselno z vidika zagotavljanja ponovljivosti rezultatov) iskalni pogoj ter časovni žig podatkov. Ker so URL-naslovi iskalnih pogojev tipično zelo dolgi, svetujemo uporabo spletne storitve CLARIN [shoftref.org](http://shoftref.org) za dodeljevanje stalnega identifikatorja URL-jem. V literaturi se navede referenca vira, ki vsebuje informacije o avtorstvu in verziji uporabljenega vira. Referenca je običajno na voljo v kolofonu vira.

3. Če raziskava temelji **na vaših lastnih podatkih**, je dobrodošla njihova predhodna objava. Repozitorij, na katerem se zbirajo jezikoslovne podatkovne baze, je [CLARIN.SI](https://www.clarin.si). Manjše podatkovne množice je mogoče dati na voljo tudi v drugi obliki, npr. v splošnih repozitorijih, npr. [Zenodo](https://zenodo.org/). Repozitoriji, kot je CLARIN.SI, imajo oblikovano uredniško politiko, ki opredeljuje pogoje za objavo. Če se boste odločili za objavo rezultatov na repozitoriju, priporočamo, da temu namenite dovolj časa.

Recenzenti revije *Slovenščina 2.0* so pozorni na citiranje in dostopnost raziskovalnih podatkov. Citiranje se pričakuje brez izjeme. Presoja glede objave lastnih podatkov se priporoča, vendar je odločitev na strani avtorja. V primerih, ko objava lastnih podatkov ni smiselna ali mogoča, priporočamo avtorjem, da svojo odločitev ob oddaji prispevka na kratko utemeljijo. V primeru vprašanj lahko pišete na uredništvo (mail).

**[Več o raziskovalnih podatkih \(link na stran RDA\)](#)**

# Implementacija novosti

- Smernice bodo dodane na spletno stran revije po pridobljenih povratnih informacijah (vključno s to konferenco).
- Poleg smernic bomo dodali primere raziskovalnih podatkov v razdelek o citiranju referenc in oblikovanju seznama literature.
- Dobrodošlo bi bilo poljudno napisano oz. uporabniku prijazno spletno mesto, kjer si lahko avtorji in recenzenti preberejo več o iniciativi RDA, razpoložljivih (slovenskih?) repozitorijih ...

# Hvala za pozornost!

[spela.arhar@cjvt.si](mailto:spela.arhar@cjvt.si)

[vojko.gorjanc@ff.uni-lj.si](mailto:vojko.gorjanc@ff.uni-lj.si)

[tomaz.erjavec@ijs.si](mailto:tomaz.erjavec@ijs.si)

Slovenščina 2.0: <https://revije.ff.uni-lj.si/slovenscina2/>

CLARIN.SI: [www.clarin.si](http://www.clarin.si)