

1. SPLOŠNI PODATKI

Naslov zbirke podatkov:

Korpusni vpogled v nestandardno spletno slovenščino: novi samostalniki in semantični premiki v komentarjih družbenih omrežij (JANES v1.0): raziskovalni podatki

Avtor: Maša Rolih

ORCID: <https://orcid.org/0009-0008-5623-1656>

Ustanova: Znanstveno-raziskovalno središče Koper, Inštitut za jezikoslovje

E-naslov: masa.rolih@zrs-kp.si

Trajanje raziskave: januar 2025 – junij 2025

Financiranje: ARIS P5-0409 *Razsežnosti slovenstva med lokalnim in globalnim v začetku tretjega tisočletja*

Jezik podatkov: slovenščina

Ključne besede: slovenščina, spletna slovenščina, nestandardna leksika, novi samostalniki, semantični premiki, korpus JANES v1.0, družbena omrežja, leksikografija, kvalifikatorji

2. DOSTOP IN LICENCA

Licenca: CC BY 4.0 – Creative Commons Priznanje avtorstva 4.0 Mednarodna

3. SEZNAM DATOTEK

- **PreberiMe.pdf** – dokument z opisom metapodatkov, postopkov zbiranja in obdelave podatkov
- **RAZ_Rolih_Masa_2025.txt** – podatkovna zbirka: seznam samostalnikov iz korpusa JANES v1.0 (družbena omrežja), z dodanimi kvalifikatorji, kontekstualno določenimi pomeni in izpisom konkordanc

4. OPIS PODATKOV

4.1 Izvleček

Zbirka vsebuje raziskovalne podatke o novih in releksikaliziranih samostalnikih v nestandardni spletni slovenščini, izluščenih iz korpusa JANES v1.0 (2014–2018). Vključeni so izrazi, ki niso evidentirani v osrednjih slovarskih virih (SSKJ, SP, eSSKJ, SNB, Sprotni slovar), ter samostalniki s semantičnimi premiki v sodobni rabi. Za vsako enoto so navedeni samostalniška oblika,

kvalifikatorji (npr. neform., manj form., mlad., avtomobil., polit., splet., mod., glasb., šport., slabš., evfem., vulg., šalj., ekspr.), pomen v kontekstu in konkordančni primer iz korpusa. Podatki so bili pridobljeni z avtomatskim luščenjem samostalnikov v orodju noSketch Engine in nato filtrirani glede na frekvenco, prisotnost v slovarskih virih ter diskurz. Pri tem je bila upoštevana predvsem raba v spletnih forumih in družbenih omrežjih, kjer je pogosta tvorba novih samostalnikov ali pomenskih premikov. Zbirka omogoča vpogled v dinamiko sodobne nestandardne slovenščine in je osredotočena na tiste leksikalne enote, ki kažejo spremembe v besedotvorju, semantiki in pragmatično-ekspresivni rabi.

4.2 Vir podatkov

Podatki izhajajo iz **korpusa JANES v1.0 (družbena omrežja)**, dostopnega na repozitoriju CLARIN.SI (<https://nl.ijs.si/janes/>).

Korpus vključuje komentarje in objave, ustvarjene med **1. 7. 2014 in 1. 2. 2018**, in vsebuje približno **250 milijonov pojavnic** ter **13 milijonov besedil**.

4.3 Tip podatkov

- računalniško posredovana komunikacija (komentarji, objave na družbenih omrežjih),
- nestandardna slovenščina,
- podatkovna enota: samostalni, izluščen iz korpusnega gradiva.

4.4 Postopek zbiranja

1. **Luščenje samostalnikov:** z orodjem noSketch Engine so bili iz korpusa izpisani samostalniki.
2. **Frekvenčni filter:** odstranitev nizkofrekvenčnih pojavitev.
3. **Diskurzivna relevantnost:** izbor izrazov, ki imajo v komunikaciji jasno pragmatično, identitetno ali stilno funkcijo.

4.5 Merila za vključitev

- **novost v leksiki:** izrazi, ki jih ni v osrednjih slovarskih virih (SSKJ, SP, eSSKJ, SNB, Sprotni slovar slovenskega jezika),
- **semantični premik:** izrazi, ki so sicer v slovarjih, a v sodobni rabi kažejo drugačen pomen,
- **pragmatična označenost:** izrazi, katerih raba je zaznamovana s slogovnim odmikom, ekspresivnostjo ali kontrastom do standardnega jezika.

4.6 Struktura podatkovne zbirke

Za vsak samostalni so zabeleženi:

- **oblika izraza,**
- **konkordančne vrstice** (primeri rabe v korpusu),
- **kontekstualno določen pomen,**
- **kvalifikatorji** (formalnostni, diskurzni, prostorski, pragmatični/ekspresivni).

5. METODOLOGIJA OBDELAVE

5.1 Določitev pomenov

Pomeni so bili rekonstruirani z:

- analizo konkordančnih vrstic (uporaba orodja noSketch Engine),
- primerjavo s slovarskimi viri,
- oceno pragmatične funkcije izraza (npr. ironična raba, ekspresivna obarvanost, slogovni poudarek).

Za vsak izraz je bil dodan tipičen primer rabe, ki ponazarja kontekst in olajša razumevanje pomena.

5.2 Določitev kvalifikatorjev

Kvalifikatorji temeljijo na tipologiji Komac in Race (2022) ter Mikolič (2016).

Uporabljen je bil naslednji vrstni red:

1. **formalnostni/družbenostni** (npr. neform., manj form.),
2. **diskurzni/tematski** (npr. mlad., splet., polit., tehnol.),
3. **prostorski** (npr. primor., osrednjeslov.),
4. **pragmatični in ekspresivni** (npr. ekspr., iron., evfem., vulg., poud.).

Vsi izrazi so bili obravnavani kot **neformalni** (izvor v pogovornem, slengovskem ali regionalnem diskurzu). Tisti z nižjo stopnjo zaznamovanosti so dodatno označeni kot **manj formalni**.

5.3 Posebnosti analize

- **Intenzitetna oznaka (poud.):** pripisana vsem izrazom, saj njihova raba vstopa v kontrast s standardnim jezikom.
- **Ekspresivnost:** določena na podlagi konteksta, saj je čustvena obarvanost pogosto odvisna od diskurzivne situacije.

- **Pragmatična funkcija:** upoštevali smo, ali izraz deluje kot modifikacija intenzitete, identitetni označevalec, slogovno/čustveno zaznamovan.

6. RELEVANTNOST PODATKOV

Zbirka omogoča:

- prepoznavanje novih samostalnikov v spletni slovenščini,
- evidentiranje semantičnih premikov obstoječih izrazov,
- vpogled v pragmatične in slogovne funkcije nestandardne leksike,
- možnost primerjalne analize s slovarskimi viri in rabo v različnih diskurzih.

7. REFERENČNI SEZNAM

Fišer, D., Erjavec, T., Ljubešić, N. (2016). *JANES v0.4: Korpus slovenskih spletnih uporabniških vsebin. Slovenščina 2.0*, 4(2), 67–100. Dostopno na:

http://www.trojina.org/slovenscina2.0/arhiv/2016/2/Slo2.0_2016_2_04.pdf

Korpus besedil s slovenskih družbenih omrežij Janes v1.0 // Corpus of texts from Slovene social media sites Janes V1.0 Dostopno na:

<https://www.clarin.si/ske/#dashboard?corpname=janer>

Gliha Komac, N., & Divjak Race, D. (2022). Tipologija slovarskih oznak za tretjo izdajo Slovarja slovenskega knjižnega jezika (eSSKJ). *Slavistična revija*, 70(4), 612–625.

Mikolič, V. (2016). Slovnični, slovarski in jezikovnodidaktični opis modifikacije moči v procesu upovedovanja. *Toporišičeva obdobja* (pp. 325–333). Znanstvena založba Filozofske fakultete.

<http://centerslo.si/wp-content/uploads/2016/11/Mikolic.pdf>