



ZRC SAZU

User-focused language resources in Slovenia: towards universal solutions

Janoš Ježovnik

janos.jezovnik@zrc-sazu.si



Introduction

1. Slovenian language resources and user-generated content
2. Slovenian language resources responding to user needs
3. Conclusion



Types of user-generated content

With regard to the proportion of content originating from users (Lew 2014: 10):

- entirely or almost entirely user-generated,
- half-collaborative = combined with professional content,
- supplemental = professional content dominates.



Razvezani jezik

// The Untied Tongue //

- (sort of a) Slovenian version of the Urban Dictionary
- entirely user-generated
- 2004–2019
- ~ 7000 headwords of various types



zaviti se v metliško črnino

Utapljati žalost v dotični rujni kapljici.

--

Ker ni zadel na lotu, se je za cel konec tedna zavil v metliško črnino.

šlinga

a) Vezalka za zavezovanje obuvala. Mn. *šlinge*.

Zavezal si je *šlinge* na čevljih in odhitel na igrišče.

Žakelj je zavezal kar s *šlingo*, ki mu je ostala od njegovih starih čevljev.

b) zanka, pentlja,

Razvezani prosti slovar žive slovenščine

v nulo

V nulo; izg. u nulo. Zelo natančno, brez napake. Primer: To si pa v nulo naredil. V nulo je zadel.



Gotof je

1. Politično odpisana osebnost. Nekdo, ki je zaradi nespotskega oklepanja oblasti ostal le še politični zombi.
2. Opis pre-koitoalno-skanglanega človeka, ki ve kaj hoče. Želja po dokončni konzumaciji demokracije....
3. Konec je z njim.



Language resources of the *Centre for Language Resources and Technologies*

- SloLeks 2.0 – Slovenski oblikoslovni leksikon (viri.cjvt.si/sloleks; 100.802 entries)
// Slovene Morphological Lexicon //
- Slovar sopomenk sodobne slovenščine (viri.cjvt.si/sopomenke; 100.837 entries)
// Thesaurus of Modern Slovene //
- Kolokacijski slovar sodobne slovenščine (viri.cjvt.si/kolokacije; 35.989 entries)
// Collocations Dictionary of Modern Slovene //
- ...



jezik samostalnik, občno ime, moški spol; [jêzik] 227.954 pojavitev | 2019-10-22



Oblike

[jêzik - jezika]

Manjka naglasni vzorec?

EDNINA

Oblika

Posnetek in IPA

jêzik

'je:zik

jezíka

je'zi:ka

jezíku

je'zi:ku

jêzik
neživo

'je:zik

jezíku

je'zi:ku

jezíkom

je'zi:kɔm

DVOJINA

Oblika

Posnetek in IPA

jezíka

je'zi:ka

jezíkov

je'zi:kɔv

jezícoma

je'zi:kɔma

jezíka

je'zi:ka

jezíkih

je'zi:kix

jezícoma

je'zi:kɔma

MNOŽINA

Oblika

Posnetek in IPA

jeziki

je'zi:ki

jezíkov

je'zi:kɔv

jezíkom

je'zi:kɔm

jezike

je'zi:ke

jezíkih

je'zi:kix

jeziki

je'zi:ki



Povezane iztočnice

jezikov
pridevnik, svojilni



hiteti

glagol / pogostost: ●●●●○ / pomensko členjeno geslo / 2022-11-15



Sopomenke Protipomenke



1 pri dejavnosti

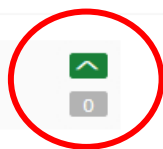
2 o premikanju

3 minevati

Sopomenke

1 pri dejavnosti

brzeti



drveti

:

dirjati

:

pohiteti

:

2 o premikanju

brzeti

:

drveti

:

divjati

:

leteti

:

dirjati

:

ubirati jo



rezati jo



drobencljati

:

drobencati



sukati se



poditi se



hitro hoditi



teči

:

švigniti

:

planiti

:

dreti

:

vrveti

:



1 izdelati risbo

2 ponazoriti

Kolokacije

Pogostost



Splošno besedišče



Skrij iztočnico



1 izdelati risbo

ILUSTRIRATI + ZA + samostalnik v tožilniku

ilustrirati za založbo

ilustrirati za revije

ilustrirati za časopise

ilustrirati za otroke

ILUSTRIRATI + V + samostalnik v mestniku

ilustrirati v knjigi

Kolokacija

ilustrirati v knjigi



Kar nekaj med številnimi fotografijami, ki **ilustrirajo** besedilo v sami **knjigi**, jih je precej boljših.

Ribiška zveza Slovenije, 2003



Omenila bi rada še nekaj, kar bistveno zadeva moj pristop k načelom in tehniki zgodnje analize, ki ju **ilustriram** v svoji **knjigi**.

ISH Fakulteta za podiplomski humanistični študij, 1998





ZRC SAZU

Univerza v Ljubljani
Universitas Labacensis



cjvt

Responsive dictionary

Corpus-based, automated extraction

„[T]he compilation of the dictionary database is immediately followed by providing the language community with access to a large amount of relevant, albeit somewhat noisy language information.“ (source: CJVT website)



ZRC SAZU



ZRC SAZU

Inštitut za slovenski
jezik Frana Ramovša

Growing dictionary

Corpus-based, manually edited

„Parts of the dictionary in the sense of a certain number of finalized dictionary entries are published at regular intervals on the web, even though the dictionary as a whole is not completed. An addition to the dictionary is published once a year, consisting of dictionary entries edited in the last year.“

(source: Jezikovna svetovalnica // Language Counselling website; transl. JJ)



ZRC SAZU

Inštitut za slovenski
jezik Frana Ramovša

Fran language portal:

- online since 2014
- currently 43 dictionaries, 1 linguistic atlas, 2 language counselling services + other resources
- made-for-print/retrodigitized (37) and made-for-web/growing dictionaries (7)



The screenshot shows the homepage of the Fran language portal. At the top, the word "Fran" is displayed in a large, serif font. Below it, the text "Dictionaries of the Fran Ramovš Institute of the Slovenian Language ZRC SAZU" is centered. A search bar with the placeholder text "Search all dictionaries" and a magnifying glass icon is positioned below the text. A navigation menu with the following items is located below the search bar: General, Etymological, Historical, Terminological, Dialectal, Counselling, and Collections. At the bottom of the page, five book covers are displayed in a row. From left to right, they are: a gold cover with "SJS²" and "SJK" in white; a green cover with "//slovar //slovenskega //knjižnega"; a white cover with "Sinonimni slovar slovenskega jezika" and a red and blue graphic; a dark blue cover with a white "S" and "SLOVENSKI"; and a blue cover with "//ePravopis" in white.



ZRC SAZU



ZRC SAZU

Inštitut za slovenski
jezik Frana Ramovša

Sprotni slovar slovenskega jezika

// Growing dictionary of the Slovenian Language //

- covers mainly neologisms = new lexemes and senses
- published as a growing dictionary since 2014 on Fran
- currently 1336 entries, mainly user-suggested

//sprotni
//slovar
//slovenskega
//jezika

sprotni



This word was only recently adopted by Slovenian; you can suggest an adequate Slovenian term or a more naturalized version, or you can vote for other users' suggestions.

sêlfi, selfie

[sêlfi]

samostalnik moškega spola

ê

[Prikaži oblike](#)

1. fotografija samega sebe, navadno narejena s prenosnim telefonom, zlasti za objavo na spletu

sebek

posneti zvezdniški selfi

objava selfija

1.1 kot pridevnik

ki je z zvezi s tako fotografijo, zlasti vedenjem, ki izhaja iz osredotočenosti le nase

Vzgojili smo **selfi** generacijo, ki je obsedena sama s seboj.

Polovica ljudi dela, druga polovica pa kramlja med **selfi** pavzo.

selfi palica, selfie palica

zložljiva, raztegljiva palica, na konec katere se za lažjo izdelavo selfija pritrdi prenosni telefon, manjša kamera

selfiestick

Fleksibilna **selfi palica** z daljinskim upravljalnikom omogoča pametnim telefonom, akcijskim kameram in kompaktnim fotoaparatom varen oprijem.

Kompaktna **selfie palica** je narejena iz aluminijeve zlitine.

[Iztočnica](#)

[Izgovor](#)

[Besedna vrsta](#)

[Tonemski naglas](#)

[Oblike](#)

[Pomeni](#)

[Stalne zveze](#)

//sproti
//slovar
//slovenskega
//jezika

sproti

Growing Dictionary of the
Slovenian Language (SSSJ)



Suggested translation			User	Date	
jazček	1	22	ka	13/10/2014	Prenesi v: Twitter
narcisek	3	19	golob	14/10/2014	Prenesi v: Twitter
sebček	6	9	amater	15/10/2014	Prenesi v: Twitter
sebko	4	12	Anonimen	15/10/2014	Prenesi v: Twitter
samsek	1	16	majci	15/10/2014	Prenesi v: Twitter
sebči	1	15	amater	15/10/2014	Prenesi v: Twitter
svojček	2	15	Lara	15/10/2014	Prenesi v: Twitter
sebek	32	9	dr	15/10/2014	Prenesi v: Twitter
samofotka	1	11	Anonimen	17/10/2014	Prenesi v: Twitter
jazek	4	7	Elen	02/11/2014	Prenesi v: Twitter
samoslik	6	9	April	06/11/2014	Prenesi v: Twitter
selfek	5	10	Anonimen	19/11/2014	Prenesi v: Twitter



sebeslik	1	7	just_me_93	07/02/2015	Prenesi v: Twitter
jazlik	0	9	just_me_93	07/02/2015	Prenesi v: Twitter
sebefot	0	10	just_me_93	07/02/2015	Prenesi v: Twitter
avtoportret	2	6	maja	08/04/2015	Prenesi v: Twitter
samič	1	5	Uroš Ferrari	14/06/2015	Prenesi v: Twitter
sebič	2	5	Uroš Ferrari	14/06/2015	Prenesi v: Twitter
selfe	0	7	nj	21/09/2015	Prenesi v: Twitter
samsebeslik	1	9	mmija	23/09/2015	Prenesi v: Twitter
samovšeček	0	7	Marko	29/10/2015	Prenesi v: Twitter
obnofot	1	8	öüß	25/01/2016	Prenesi v: Twitter
smotko => (SM se fOTKO)	3	5	damir	15/03/2016	Prenesi v: Twitter
osebek	0	4	panda	08/06/2016	Prenesi v: Twitter
všečnik	0	3	JK	27/06/2016	Prenesi v: Twitter



samovšečko	👍 0	👎 4	slovaram	28/10/2016	Prenesi v: Twitter
sébi	👍 0	👎 4	RD	03/12/2016	Prenesi v: Twitter
intskot	👍 0	👎 5	milanadem	28/12/2016	Prenesi v: Twitter
sebik	👍 0	👎 4	milanadem	28/12/2016	Prenesi v: Twitter
sebko	👍 2	👎 1	milanadem	28/12/2016	Prenesi v: Twitter
osebek	👍 0	👎 5	milanadem	28/12/2016	Prenesi v: Twitter
osebko	👍 0	👎 3	Anonimen	12/11/2017	Prenesi v: Twitter
afnač	👍 2	👎 5	jože	12/01/2018	Prenesi v: Twitter
selfač	👍 0	👎 4	Anonimen	22/10/2018	Prenesi v: Twitter
moj-i	👍 0	👎 1	Slovek	18/05/2022	Prenesi v: Twitter
mojček	👍 0	👎 1	Lukec	26/09/2022	Prenesi v: Twitter
fotoportret	👍 0	👎 1	Videx	11/11/2022	Prenesi v: Twitter

[Suggest a new word!](#)



sébek

[sébək], [sêbək]

samostalnik moškega spola

ê; ê

Prikaži oblike

fotografija samega sebe, navadno narejena s prenosnim telefonom, zlasti za objavo na spletu

selfi

Sprednja kamera za **sebke** ponuja možnost zajema fotografij z osmimi milijoni točk.

Deklici, ki ga je prosila, naj se z njo fotografira, je s **sebkom** polepšal dan.

Zavedajte se, da grede nekaterim ljudem **sebki** na živce in lahko zato dobite negativne komentarje.

Iztočnica

Izgovor

Besedna vrsta

Tonemski naglas

Oblike

Pomeni



Franček educational language portal:

- online since 2021
- adapted for the schooling population (primary, secondary school)
- 3 stages: younger pupils (yr. 1–5), older pupils (yr. 6–9), secondary-school students
- instead of separate dictionaries >> thematic modules

FRANČEK



– word senses

– synonyms

– inflection

– pronunciation

– phraseology

– dialect uses

– etymology

– word history



Kaj pomeni beseda **abeceda**?

Abeceda so črke v dogovorjenem zaporedju v latinični pisavi.

slovenska **abeceda**

Slovenska pisava ima 25 črk, ki so urejene po **abecedi**.

Po **abecedi** so razvrščene besede v slovarjih, gesla v leksikonih ali enciklopedijah, priimki v redovalnici ali imeniku telefona.

Ali veš, da imamo v slovenščini 29 glasov in samo 25 črk **abecede**?

Beseda **abeceda** je lahko tudi del stalnih besednih zvez:



atletska abeceda

PRI ŠPORTU



Atletska abeceda so gibalne vaje za tehniko teka, za krepitev in raztezanje mišic.

Najbolj nas je mučila **atletska abeceda**, učili smo se tekaške tehnike.



Morsejeva abeceda

Morsejeva abeceda je nabor črk, števk in ločil, pri katerem je vsak od teh znakov izražen s kombinacijo pik in črt v zapisu, kratkih in daljših zvočnih signalov ali svetlobnih utripov.



Ko se naučite **Morsejeve abecede**, lahko pošljete najrazličnejša skrivna sporočila z udarjanjem dveh žlic, s trkanjem po mizi ali z utripanjem z žepno svetilko.



Kaj pomeni beseda **abeceda**?

1. POMEN

RAZLAGA: **nabor črk v ustaljenem zaporedju v latinični pisavi**

TIPIČNE ZVEZE: črka abecede; konec, začetek abecede; učenje abecede • angleška, latinska, slovenska abeceda; naša abeceda • naučiti, učiti (se) abecedo; poznati, znati abecedo • naučiti, učiti (se) abecede • razvrstiti, urediti po abecedi; slediti si po abecedi

ZGLEDA: Če je v istem letu izšlo več del istega avtorja, jih ločimo z malimi črkami **abecede**. Igralec, ki potegne ploščico, katere črka je najbližja začetku **abecede**, prične prvi.

1. PODPOMEN

RAZLAGA: **nabor črk v ustaljenem zaporedju v kaki pisavi sploh**

TIPIČNE ZVEZE: arabska, grška, hebrejska abeceda • črka abecede • naučiti, učiti (se) abecedo

ZGLEDA: Etruščani so svojo pisavo priredili po grški **abecedi**. Muzej sestavlja daljša glavna stavba, oblikovana v podobi osme črke v hebrejski **abecedi** het, ter dodatna zgradba v obliki črke jod.

2. POMEN

RAZLAGA: **začetno, osnovno znanje določenega področja**

TIPIČNE ZVEZE: atletska, nogometna abeceda; gasilska abeceda; osnovna abeceda • naučiti, učiti (se) KATERE abecede • naučiti, učiti (se) KATERO abecedo; poznati, znati KATERO abecedo

ZGLEDA: Poleg spoznavanja atletske **abecede** in osnov pravilnega teka bodo organizatorji za vse tekače in njihove spremljevalce zagotovili tudi meritve, ki jih bodo opravile medicinske sestre. Strokovno usposobljen predavatelj vas bo naučil računalniške **abecede**.



Kaj pomeni beseda **abeceda**?

1. nabor črk v ustaljenem zaporedju v latinični pisavi

TIPIČNE ZVEZE: črka abecede; konec, začetek abecede; učenje abecede • angleška, latinska, slovenska abeceda; naša abeceda • naučiti, učiti (se) abecedo; poznati, znati abecedo • naučiti, učiti (se) abecede • razvrstiti, urediti po abecedi; slediti si po abecedi

Če je v istem letu izšlo več del istega avtorja, jih ločimo z malimi črkami **abecede**.
Igralec, ki potegne ploščico, katere črka je najbližja začetku **abecede**, prične prvi.

1.1. nabor črk v ustaljenem zaporedju v kaki pisavi sploh

TIPIČNE ZVEZE: arabska, grška, hebrejska abeceda • črka abecede • naučiti, učiti (se) abecedo

Etruščani so svojo pisavo priredili po grški **abecedi**.

Muzej sestavlja daljša glavna stavba, oblikovana v podobi osme črke v hebrejski **abecedi** het, ter dodatna zgradba v obliki črke jod.

2. začetno, osnovno znanje določenega področja

TIPIČNE ZVEZE: atletska, nogometna abeceda; gasilska abeceda; osnovna abeceda • naučiti, učiti (se) KATERE abecede • naučiti, učiti (se) KATERO abecedo; poznati, znati KATERO abecedo

Poleg spoznavanja atletske **abecede** in osnov pravilnega teka bodo organizatorji za vse tekače in njihove spremljevalce zagotovili tudi meritve, ki jih bodo opravile medicinske sestre.

Strokovno usposobljen predavatelj vas bo naučil računalniške **abecede**.



ZRC SAZU

BESSED



BOJ



Terminologišče – terminology portal:

- 17 terminological dictionaries
- divided by professional fields
- Terminološka svetovalnica // **Terminological counselling** // (577 replies)
- linked to Fran & searchable there



Iskalnik

Iztočnica

korenina

Razlaga

Ustreznik (termin v tujem jeziku)

Jezik ustreznika

Angleščina

Išči

Ponastavi

Najdenih: 28

korenína

adventívna korenína

anulózna korenína

dèskasta korenína

dihálna korenína

dôlga korenína

glávna korenína

hodúljasta korenína

korenína -e ž

organ. organ brstnic /Cormophyta/, praviloma v tleh in vedno brez listov, ki omogoča zlasti absorpcijo vode in v njej raztopljenih, predvsem mineralnih snovi, ter pritrjevanje rastline v podlago ali na njo

ang.: root *n*

nem.: Wurzel *f*

lat.: radix *f*



Jezikovna svetovalnica

// (General) language counselling //

- 3571 replies
- linked to Fran & searchable there

ePravopis

// Slovenian Normative Guide/Dictionary //

- 8940 entries
- „problematic“ lexis (from a normative point of view)

Pravopis 8.0

// suggestions for the renovated normative guidelines //



Pravopis 8.0 ^{predlog}

Pravila novega slovenskega pravopisa za javno razpravo





ZAPIS POLGLASNIKA V IZGLASNIH SOGLASNIŠKIH SKLOPIH

{84} Polglasnik v izglasnih soglasniških sklopih ob odsotnosti samoglasnika tvori začasno zlogovno jedro. Ob dodajanju končnic in obrazil se glasovna soseščina spremeni in polglasnik izpade.

[ODDAJTE KOMENTAR](#)

{85} Zapisujemo ga s črko (e) med nezvočnikom in zvočnikom ter med dvema zvočnikoma, in sicer:

1. v imenovalniku ednine samostalnikov in pridevnikov moškega spola (npr. kvader [kvádər], jasen [jásən]);
2. v roditeljski dvojini in množini samostalnikov ženskega in srednjega spola (npr. kamra [kámra], rod. dv./mn. kamer [kámər]; veslo [vêslò], rod. dv./mn. vesel [vêsel]);
3. pri izpeljankah z obrazili na soglasnik, npr. -ček, -ce, -čan; -ski, -ni ... (pismo [písmo], manjš. pisemce [písəmce], vrst. prid. pisemski [písəmski]);
4. v deležniku na -l moškega spola (lezel [lézəu]).

[ODDAJTE KOMENTAR](#)

{86} POSEBNOST

Pri samostalnikih, ki se v imenovalniku ednine (december, meter) ali v roditeljski množini (jedro, Pekre) končajo na r, se ob tvorbi izpeljanke z obrazilom na soglasnik črka r pojavi v medsoglasniškem položaju in označuje glasovni sklop [ər]. Polglasnika pred r ne zapisujemo:

- o december [decêmbər], rod. decembra [decêmbra], vrst. prid. decembrski [decêmbərski];
- o jedro [jêdro], rod. dv./mn. jeder [jêdər], vrst. prid. jedrski [jêdərski], jedrni [jêdərni];
- o Pekre [pêkre], rod. dv./mn. Peker [pêkər], vrst. prid. pekrski [pêkərski], preb. Pekarčan [pêkərčan].

Če je e v končaju -(er) izgovorjen kot [e], črka (e) ne izpada: laser [láser], rod. laserja [láserja], vrst. prid. laserski [láserski].

Priimki, ki se glasijo enako kot (izvirna) občna poimenovanja ali prebivalska imena, lahko imajo uradno veljavne variante brez polglasnika: Turek in Turk (toda Turek |pridpadnik turškega naroda, prebivalec Turčije|), Zajec in Zajc (toda zajec), Kranjec in Kranjc (toda Kranjec |prebivalec Kranjske|).

O vlogi polglasnika pri pisnem in govornem krajšanju osnove gl. poglavje »Kje se pojavljajo neobstojni samoglasniki?«.

[ODDAJTE KOMENTAR](#)



Public debate:

1. Pisna znamenja, Krajšave – dec. 2019 || 213 comments
// Written characters, Abbreviations //
2. Velika in mala začetnica, Prevzete besede in besedne zveze – mar. 2021 || 125 comments
// Upper- and lower-case initial letters, Loanwords and MWEs //
3. Glasoslovni oris za pravopis, O prevzemanju iz posameznih jezikov – jun. 2023 || 273 comments
// An outline of phonetics, Rules for adopting and pronouncing words from foreign languages //

Published as e-books:

- argumentation for each paragraph
- all comments from the public debate (anonymised)



Conclusions

- „Responsive“ vs. „growing“ language resource?
 - Both, each fulfill different purposes!
- User input & user needs:
 - Actual, contemporary language needs
 - Accommodating different types of language user groups
- Language use >> language norm >> codification



ZRC SAZU

• Fran Ramovš Institute of the Slovenian Language • Institute of Archaeology • Milko Kos Historical Institute • France Stele Institute of Art History • Institute of Musicology • Institute of Slovenian Literature and Literary Studies • Institute of Slovenian Ethnology • Slovenian Migration Institute • Institute of Philosophy • Anton Melik Geographical Institute • Karst Research Institute • Ivan Rakovec Institute of Palaeontology • Jovan Hadži Institute of Biology • Institute of Ethnomusicology • Sociomedical Institute • Institute of Cultural History • Institute of Anthropological and Spatial Studies • Institute of Culture and Memory Studies •