



Priprava učnih gradiv s korpusom šolskih pisnih izdelkov Šolar 3.0

[Filozofska fakulteta, 24. 11. 2023]

Špela Arhar Holdt

Spela.ArharHoldt@ff.uni-lj.si

Predstavitev

- **Center za jezikovne vire in tehnologije:** raziskovalna enota Univerze v Ljubljani, razvoj **jezikovnih virov** in **orodij** za sodobno slovenščino.
- Npr. Slovar sopomenk sodobne slovenščine, Kolokacijski slovar sodobne slovenščine, Sloleks, korpus Gigafida, orodje Vejice ...
- Raziskovalni projekti, mdr. **Empirična podlaga za digitalno podprt razvoj jezikovnih kompetenc** (ARIS J7-3159).



Ogrevanje:

šolsko popravljajanje besedil

Antigona se bori za pravičnost. Tako tudi
sili Ismena dokazati, da ne more nekone ljudi
resiliti in da nekone ^{oni} dolgaji ⁿⁱ bogi. Tudi dokazati, da

Izzivi šolskega popravljanja besedil

- V učnem procesu vrsto in količino popravkov **prilagajamo** razvojni stopnji učenke oz. učenca in drugim relevantnim okoliščinam.
- Določeni popravki so nedvoumni (npr. ločila, črkovanje, velike ali male začetnice, zapis skupaj ali narazen), velik del je pa **interpretativen** in tu so možne raznovrstne odločitve (slogovni, vsebinski popravki).
- Upoštevati je treba jezikovno **variantnost**: ne zanima nas le, kaj popravljati, ampak tudi, česa ne.
- Priprava preišljene, kakovostne povratne informacije zahteva **veliko časa**, učinki pa so vidni šele na srednji in dolgi rok.
- Prakse so relativno **slabo raziskane**, ne obstajajo jasne smernice, tudi ne pri izobraževanju za učiteljski poklic.

Korpus Šolar 3.0

- Zbirka pisnih izdelkov, ki so jih učenci in dijaki (zadnja triada OŠ in različne SŠ) samostojno tvorili pri pouku, predvsem
- 5.485 besedil (1.638.229 besed), predvsem **spisi in eseji**.
- Besedila smo zbrali s pomočjo učiteljic in učiteljev iz **vseh slovenskih regij in različnih vrst šol**.
- Besedila, ki so bila po večini pisana na roko, smo **pretipkali in dodatno uredili**, da jih lahko analiziramo z digitalnimi orodji.
- Cca. tretjina besedil (35,6 %) vsebuje **učiteljske popravke**, ki smo jih transkribirali in vsebinsko kategorizirali: skupno je v korpusu **36.570** avtentičnih jezikovnih popravkov z oznako problema.

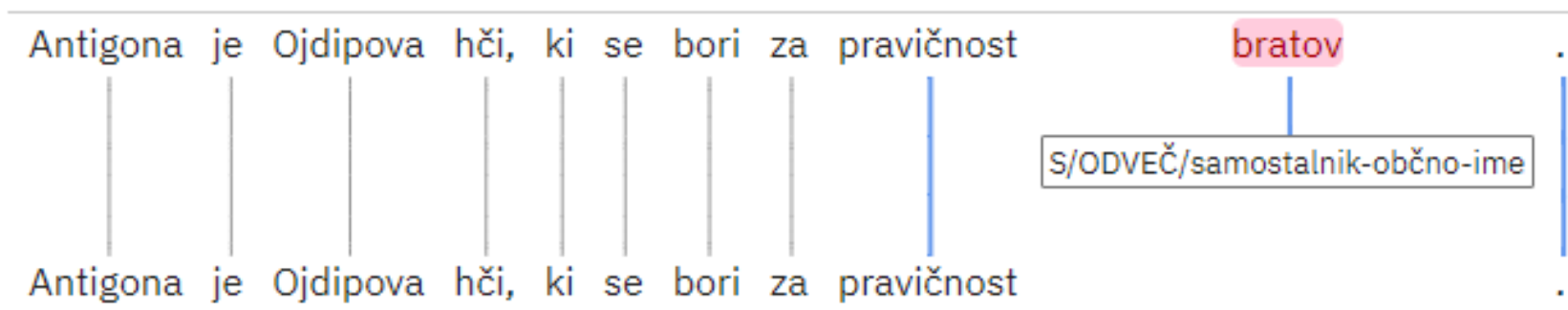
Izvorno besedilo

Antigona je Ojdipova hči, ki se bori za pravičnost bratov. Antigona prekrši Kreonov ukaz in Kreon za to izve in jo pokliče k njemu. Da bi izvedel, zakaj je to naredila.

Popravljeno besedilo

Antigona je Ojdipova hči, ki se bori za pravičnost. Prekrši Kreonov ukaz in ko Kreon za to izve, in jo pokliče k sebi, da bi izvedel, zakaj je to naredila.

Oznake popravkov (orodje CJVT Svala)



Jezikovni popravki v korpusu Šolar

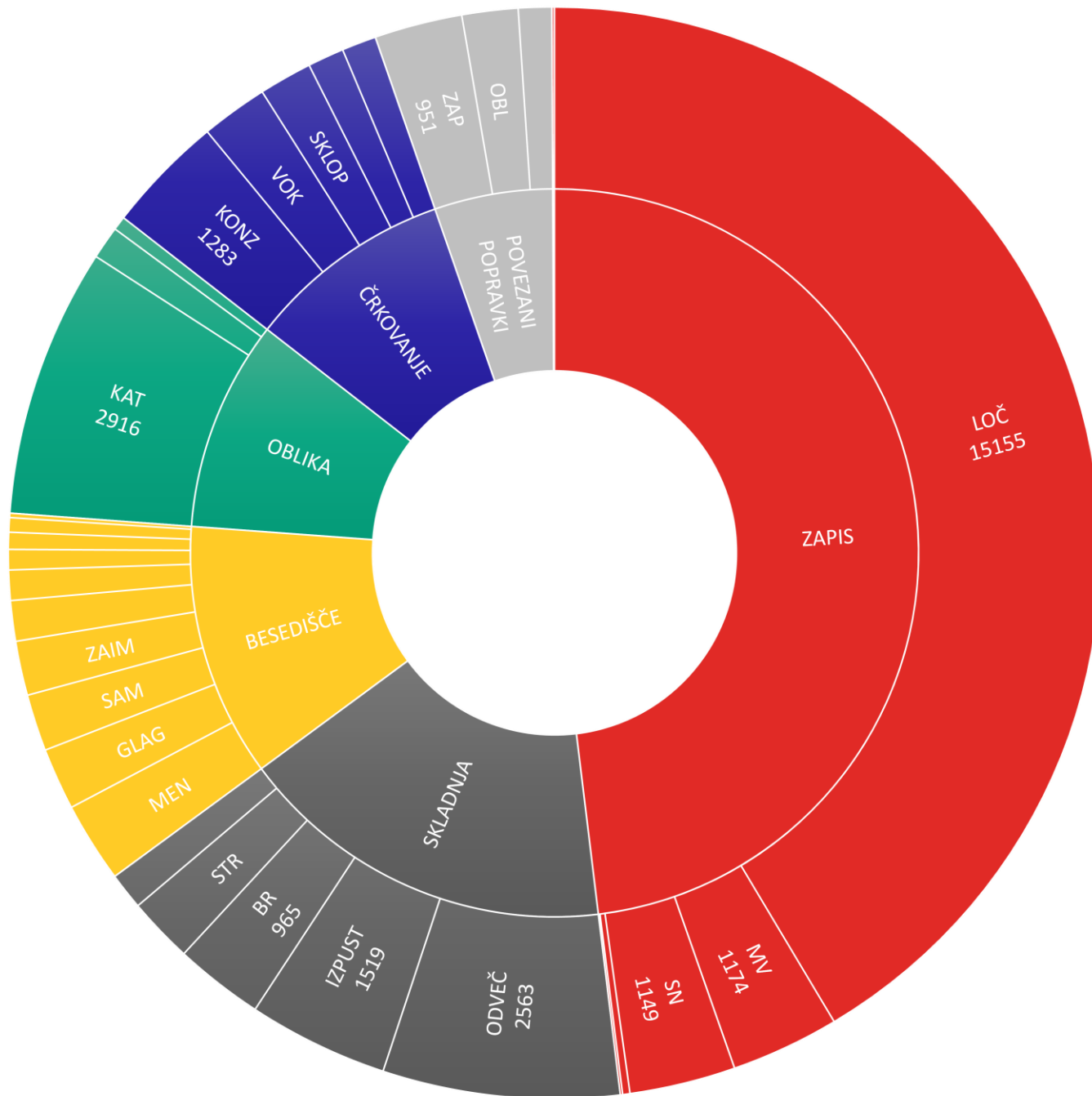
- Jezikovni popravki so razvrščeni v 180 različnih skupin, kar opisujejo smernice [Arhar Holdt et al. 2022](#).
- Oznake, s pomočjo katerih lahko najdemo primere jezikovnih problemov, so sestavljene iz treh delov (**Z/MV/osebna-imena**)
 - **1. nivo = jezikovna ravnina:** zapis, črkovanje, oblika, besedišče, skladnja, povezani popravki + oznake za nečitljivo besedilo.
 - **2. nivo = tip problema,** npr. *zapis* se deli na popravke ločil, male/velike začetnice, pisanja skupaj/narazen, zapisa kratic in zapisa števil.
 - **3. nivo = natančnejši opis problema,** npr. *zapis začetnice* se deli na *osebna-imena*, *pridevnik-ski*, *premi-govor*, *začetek-povedi* ...

Gradivo:

smernice za označevanje

1	ČRKOVANJE	6
1.1	Vokali	6
1.1.1	Odvečni vokal = Č/VOK/odveč	6
1.1.2	Izpuščeni vokal = Č/VOK/izpust	7
1.1.3	Menjava AO = Č/VOK/menjava-ao.....	7
1.1.4	Menjava EI = Č/VOK/menjava-ei	8
1.1.5	Menjava UO = Č/VOK/menjava-uo	8
1.1.6	Druge menjave vokala = Č/VOK/menjava-drugo.....	8

<https://www.cjvt.si/prop/wp-content/uploads/sites/23/2023/11/Smernice-za-oznacevanje-korpusa-Solar-v1.1.pdf>



Gradivo:

frekvenčni seznam problemov

ID besedila	ID odstavka	D stavk	Ime besedila	Leto	Šola	Predm	Razre	Op. besede	Regiji	oj	Koda popravka	1 raven	2 raven	3 raven	Šklic popra	Izvirna poved	Končna po
2	solar1s	solar1s.3	solar1s.3.:KUS-G-slo-1-GO-E-2009-10001	2009	gimnazija slovenščir 1. letnik	esej ali sp Gorica	1 Z/LOČ/nerazvrščeno	Z			1 Z/LOČ/nerazvrščeno	Z	LOČ	nerazvrščeno	#solar1s.3.1.1 "Ne da sovražim, da ljubim sem na svetu", "Ne da sovražim,		
3	solar1s	solar1s.3	solar1s.3.:KUS-G-slo-1-GO-E-2009-10001	2009	gimnazija slovenščir 1. letnik	esej ali sp Gorica	1 Z/LOČ/nerazvrščeno	Z			1 Z/LOČ/nerazvrščeno	Z	LOČ	nerazvrščeno	#solar1s.3.1.1 "Ne da sovražim, da ljubim sem na svetu", "Ne da sovražim,		
4	solar1s	solar1s.3	solar1s.3.:KUS-G-slo-1-GO-E-2009-10001	2009	gimnazija slovenščir 1. letnik	esej ali sp Gorica	1 O/KAT/povratnost	O			1 O/KAT/povratnost	O	KAT	povratnost	#solar1s.3.1.1 "Ne da sovražim, da ljubim sem na svetu", "Ne da sovražim,		
5	solar1s	solar1s.3	solar1s.3.:KUS-G-slo-1-GO-E-2009-10001	2009	gimnazija slovenščir 1. letnik	esej ali sp Gorica	1 B/SAM/občno-besedišče	B			1 B/SAM/občno-besedišče	B	SAM	občno-besedišče	#solar1s.3.2.1 V tem verzju je dala vedeti, da živi iz ljube V tem verzju je da		
6	solar1s	solar1s.3	solar1s.3.:KUS-G-slo-1-GO-E-2009-10001	2009	gimnazija slovenščir 1. letnik	esej ali sp Gorica	1 Z/LOČ/nerazvrščeno	Z			1 Z/LOČ/nerazvrščeno	Z	LOČ	nerazvrščeno	#solar1s.3.2.3 V tem verzju je dala vedeti, da živi iz ljube V tem verzju je da		
7	solar1s	solar1s.3	solar1s.3.:KUS-G-slo-1-GO-E-2009-10001	2009	gimnazija slovenščir 1. letnik	esej ali sp Gorica	1 B/MEN/drugo	B			1 B/MEN/drugo	B	MEN	drugo	#solar1s.3.3.4 Ljubezen poznamo v večjih oblikah; najp Ljubezen poznam		
8	solar1s	solar1s.3	solar1s.3.:KUS-G-slo-1-GO-E-2009-10001	2009	gimnazija slovenščir 1. letnik	esej ali sp Gorica	1 O/KAT/spol	O			1 O/KAT/spol	O	KAT	spol	#solar1s.3.3.7 Ljubezen poznamo v večjih oblikah; najp Ljubezen poznam		
9	solar1s	solar1s.3	solar1s.3.:KUS-G-slo-1-GO-E-2009-10001	2009	gimnazija slovenščir 1. letnik	esej ali sp Gorica	4 Č/SKLOP/nj	Č			4 Č/SKLOP/nj	Č	SKLOP	nj	#solar1s.3.3.1 Ljubezen poznamo v večjih oblikah; najp Ljubezen poznam		

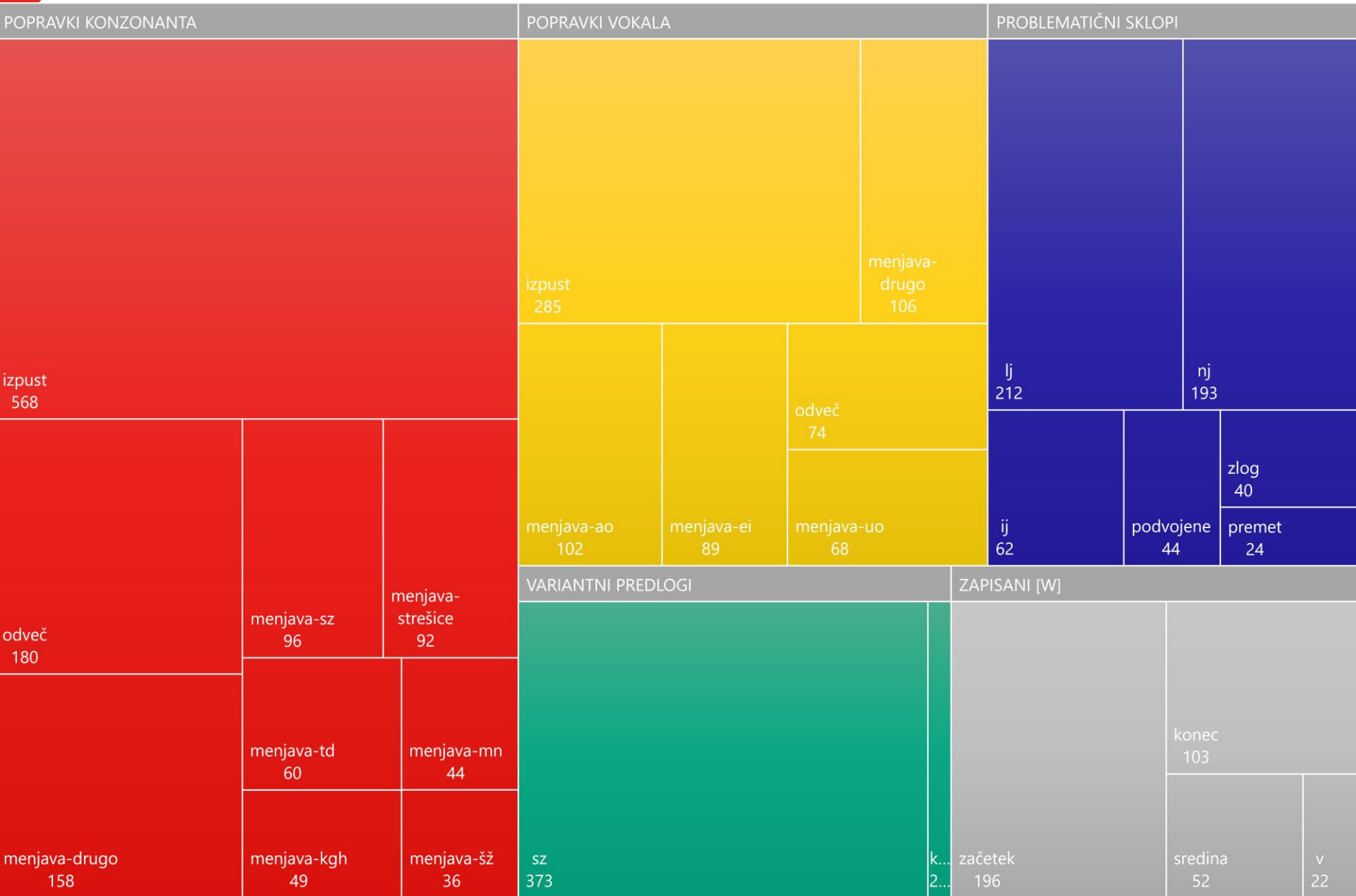
<https://www.cjvt.si/prop/wp-content/uploads/sites/23/2023/11/Izobrazevanje-uciteljev-Frekvencni-seznam-jezikovnih-popravkov-Solar-3.0.xlsx>

Delo s programom: Šolar 3.0 v konkordančniku

<https://solar.cjvt.si/>

https://www.clarin.si/noske/run.cgi/first_form?corpname=solar30_orig;align=

POPRAVKI ČRKOVANJA



■ POPRAVKI KONZONANTA
 ■ POPRAVKI VOKALA
 ■ PROBLEMATIČNI SKLOPI
 ■ VARIANTNI PREDLOGI
 ■ ZAPISANI [W]

POPRAVKI OBLIKE

POPRAVEK KATEGORIJE								POPRAVEK...			
sklon-drugo 490	število-em 317	čas 276	število-dm 246					neglagolska-končnica 106			
								glagolska-končnica 104			
						nedoločnik-namenilnik 99		sklon-dm 80		glagolska-osnova 51	preglas-in-cč 40
sklon-rt 341	spol 206	nedoločnik-kratki 165	oblika-zaimka 73	naklon 57		vid 46		neglagolsk... osnova 32	neobst... vokal 24		
				določnost 41		stopnjevanje 31		DODATNA OZNAKA			
	povratnost 175			število-ed 124	oseba 67	sklon-mo 32	način 26	ned... ose... 24	besede-mati-hči 68		variane 46
									besede-otrok		

■ POPRAVEK KATEGORIJE ■ POPRAVEK PARADIGME ■ DODATNA OZNAKA

Tako tudi želi Kreonu dokazati , da ne more zakone

Tako tudi želi Kreonu dokazati , da ne more zakonov

O/KAT/sklon-rt

Corpus: Šolar: učenci (razvojni)

Simple query:

Query type: simple lemma phrase word CQL

Lemma: PoS: unspecified

Phrase: PoS: unspecified match case

Word form: PoS: unspecified

CQL: [errtype="O_KAT_sklon-rt"]

Default attribute: errtype = Type of error

Zamenjati poševnice s podčrtajem!

NoSketch Engine
Šolar: učenci (razvojni)

- Home
- Search
- Word list
- Corpus info
- My jobs
- User guide

- Save
- Make subcorpus
- View options
- KWIC
- Sentence
- Sort
- Left
- Right
- Node
- References
- Shuffle
- Sample
- Filter
- Sub-hits
- 1st hit in doc
- Frequency
- Node tags
- Node forms
- Doc IDs
- Text types
- Collocations
- Visualize
- Menu position

Query **O_KAT_sklon-rt** 341 (178.76 per million)

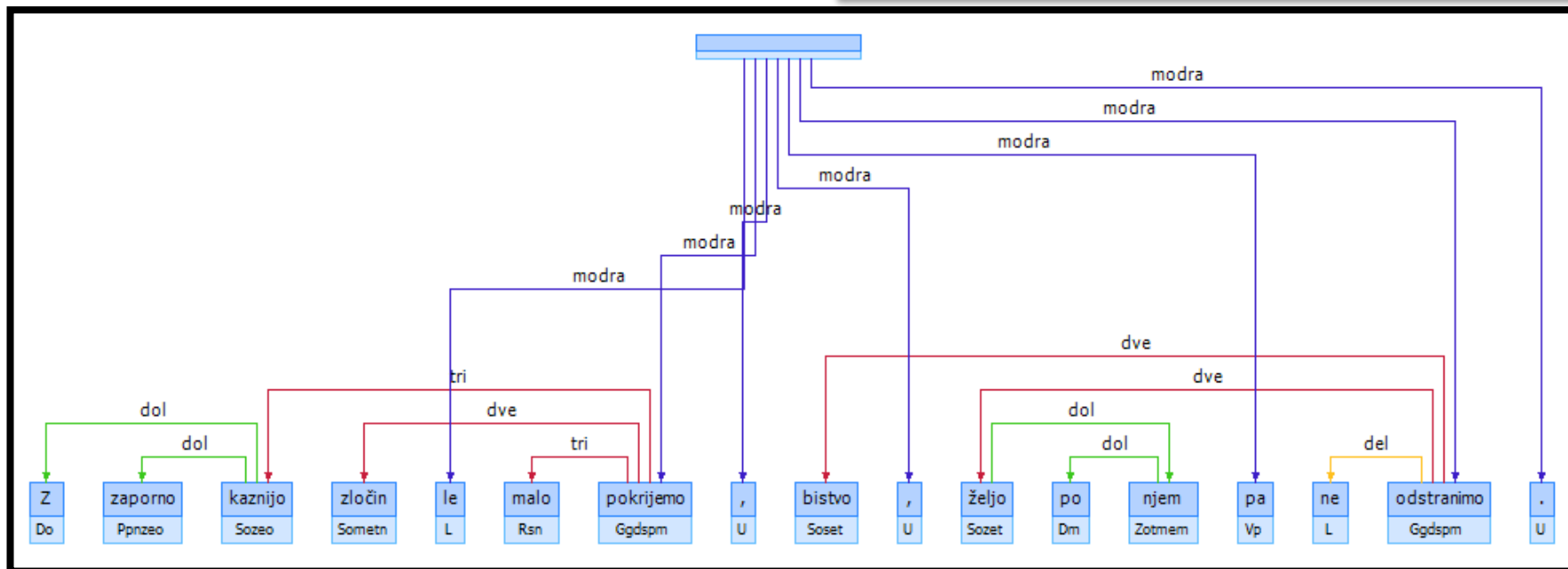
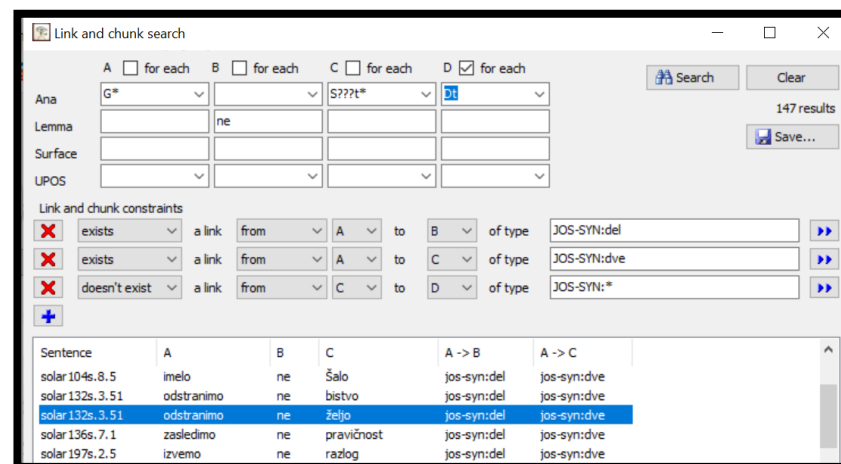
Page 1 of 14

<p>KUS-G-slo-1-GO-E-2009-10002, 1. letnik, gimnazija</p> <p>KUS-G-slo-2-GO-E-2009-10004, 2. letnik, gimnazija</p> <p>KUS-G-slo-2-GO-E-2009-10004, 2. letnik, gimnazija</p> <p>KUS-G-slo-1-GO-E-2009-10006, 1. letnik, gimnazija</p> <p>KUS-G-slo-1-GO-E-2009-10006, 1. letnik, gimnazija</p> <p>KUS-G-slo-2-GO-E-2009-10007, 2. letnik, gimnazija</p> <p>KUS-G-slo-2-GO-E-2009-10008, 2. letnik, gimnazija</p> <p>KUS-G-slo-4-GO-E-2009-10011, 4. letnik, gimnazija</p> <p>KUS-G-slo-4-GO-E-2009-10011, 4. letnik, gimnazija</p> <p>KUS-G-slo-4-GO-E-2009-10021, 4. letnik, gimnazija</p> <p>KUS-G-slo-4-GO-E-2009-10024, 4. letnik, gimnazija</p> <p>KUS-G-slo-4-GO-E-2009-10024, 4. letnik, gimnazija</p> <p>KUS-G-slo-4-GO-E-2009-10025, 4. letnik, gimnazija</p> <p>KUS-G-slo-4-GO-E-2009-10025, 4. letnik, gimnazija</p> <p>KUS-G-slo-4-GO-E-2009-10027, 4. letnik, gimnazija</p> <p>KUS-G-slo-4-GO-E-2009-10027, 4. letnik, gimnazija</p> <p>KUS-G-slo-4-GO-E-2009-10027, 4. letnik, gimnazija</p> <p>KUS-G-slo-4-MB-E-2010-10039, 4. letnik, gimnazija</p> <p>KUS-G-slo-1-MB-E-2008-10040, 1. letnik, gimnazija</p> <p>KUS-G-slo-3-LJ-E-2009-10052, 3. letnik, gimnazija</p> <p>KUS-G-slo-3-LJ-E-2009-10052, 3. letnik, gimnazija</p> <p>KUS-G-slo-3-LJ-E-2009-10052, 3. letnik, gimnazija</p> <p>KUS-G-slo-3-LJ-E-2009-10057, 3. letnik, gimnazija</p> <p>KUS-G-slo-3-LJ-E-2009-10057, 3. letnik, gimnazija</p> <p>KUS-G-slo-4-GO-E-2009-10065, 4. letnik, gimnazija</p>	<p>Njena sestra pa je ravno nasprotnega mišljenja,</p> <p>. Na prvem mestu je bilo sploh to, da bi naučili</p> <p>mestu je bilo sploh to, da bi naučili slovenski</p> <p>nič omajalo, dokler ni moje telo začelo kazati</p> <p>, dokler ni moje telo začelo kazati vidne</p> <p>ženi in ne Nežki pa tudi osvajal je baronico in ne</p> <p>dan ali Matiček se ženi je komedija, katere</p> <p>, čustva. Samega sebe ni doživljal kot</p> <p>. Samega sebe ni doživljal kot samostojno</p> <p>. Na zaslíšanju je sicer zanikala krívdo, ni pa</p> <p>Berk je bolj intelektualca kot zgolj partizan.</p> <p>. Vendar pa gre še vedno za eno življenje in to</p> <p>in notranji boj. To velja za Silvana Kandorja,</p> <p>boj. To velja za Silvana Kandorja, glavnega</p> <p>in zunanjim življenjem. Želijo si notranjega,</p> <p>od njih samih. Vem, da se usode ne da predvideti,</p> <p>tak, da je delal vse po svoje. Bil je za Vesno,</p> <p>za pomoč, takšnih ki se borijo za mir in</p> <p>Argonu povedala, kaj se v njegovi hiši dogaja ji</p> <p>z Bogom ne more nihče pozdraviti in spremeniti</p> <p>Bogom ne more nihče pozdraviti in spremeniti to</p> <p>ko sem le eden od peščice, ki meni tako in ne morem</p> <p>, kako absurdno je življenje. ♀ Mislim, da</p> <p>, kako absurdno je življenje. ♀ Mislim, da ta</p> <p>glavni cilj v katerem vidi smisel. Ne ustavi</p>	<p>kar</p> <p>slovenski</p> <p>jezik</p> <p>vidne</p> <p>znake</p> <p>Nežko</p> <p>bistvo</p> <p>samostojno</p> <p>osebo</p> <p>to</p> <p>Vojno</p> <p>osebe</p> <p>glavnega</p> <p>lika</p> <p>uresničitev</p> <p>jo</p> <p>kar</p> <p>drugacnih</p> <p>tega</p> <p>to</p> <p>miselnost</p> <p>nič</p> <p>ta</p> <p>uboj</p> <p>jo</p>	<p>v začetku navzven ne kaže preveč, vendarle po „</p> <p>jezik. In ko so potem ljudje, kot je bil Linhart</p> <p>. In ko so potem ljudje, kot je bil Linhart pisali</p> <p>znake utrujenosti. V tistem obdobju sem eno</p> <p>utrujenosti. V tistem obdobju sem eno izmed</p> <p>, kar pa je obema ugajalo. Le Matiček ni vedel za</p> <p>ne razumeš, če ne veš tudi zgodovinskega</p> <p>osebo z lastnimi ideali in cilji, ampak kot</p> <p>z lastnimi ideali in cilji, ampak kot partizana</p> <p>hotela utemeljiti, zato so jo obsodili na smrt.</p> <p>ne jemlje tako resno kot Gregor. Berk je</p> <p>, ki ti nekaj pomeni. Zelo težko bi se spopadala s</p> <p>lika romana Senčni ples. Kandor se po smrti</p> <p>romana Senčni ples. Kandor se po smrti svoje</p> <p>svojih sanj, želja, ampak so primorani ostati v</p> <p>spremeniti, a vendar naj bi se vsak za svojo</p> <p>ne bi smel. Zato je bil tudi prisiljen iti v</p> <p>ki želijo uničiti problem razlik med ljudmi,</p> <p>ni vrjel. Zato je njegova žena naredila načrt.</p> <p>miselnost. Sam menim, da je prav cerkev kriva za</p> <p>. Sam menim, da je prav cerkev kriva za številne</p> <p>spremeniti. Omejiti je treba cerkveno oblast a</p> <p>uboj in zakrivil namerno, kajti v nekem</p> <p>in zakrivil namerno, kajti v nekem trenutku, se</p> <p>nobena oblast in moč ki je nad njo. To je njena</p>
--	---	--	---

Page 1 of 14

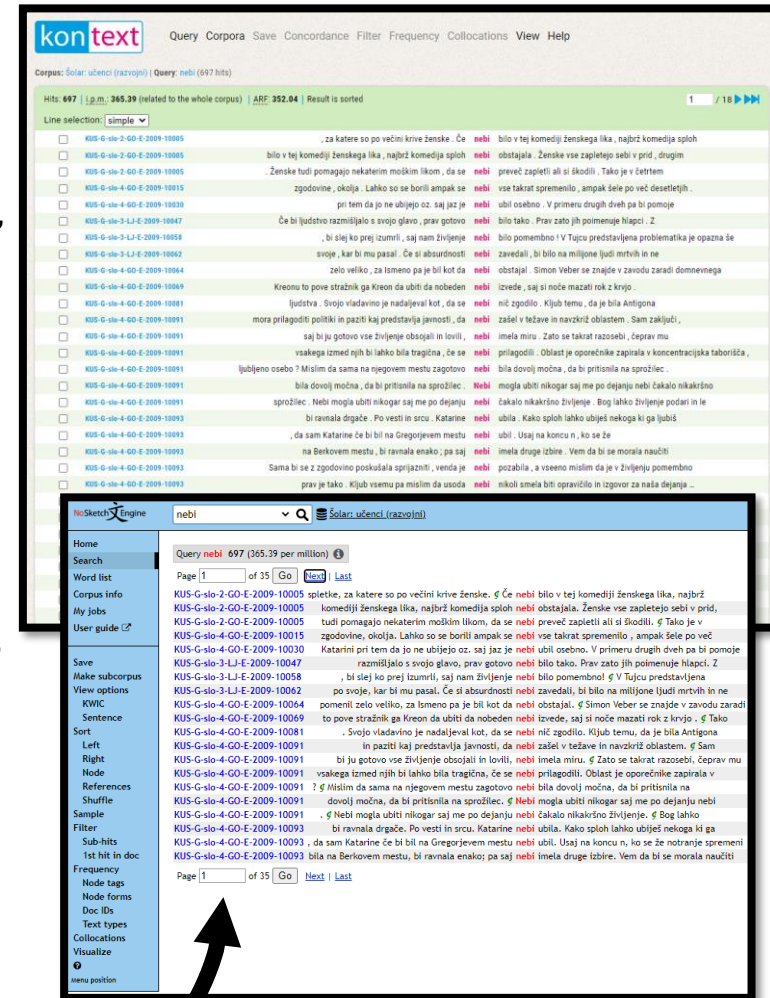
Novosti različice 3.0

- Skladienske oznake sistema JOS-SYN, po katerih je mogoče tudi iskati s programom Q-CAT.
- 93,89 % natančnost (SUK).



Dostopnost korpusa

- Kosem, Iztok; Rozman, Tadeja; Pori, Eva; Arhar Holdt, Špela; Kocjančič, Polonca; Laskowski, Cyprian; Klemenc, Bojan, 2019, [Developmental corpus ccŠolar 1.0](#), Slovenian language resource repository CLARIN.SI.
- Kosem, Iztok; Arhar Holdt, Špela; Stritar Kučuk, Mojca; Krek, Simon; Krapš Vodopivec, Irena; Stabej, Marko; Pori, Eva; Goli, Teja; Lavrič, Polona; Laskowski, Cyprian; Kocjančič, Polonca; Klemenc, Bojan; Rozman, Tadeja, 2019, [Developmental corpus Šolar 2.0](#), Slovenian language resource repository CLARIN.SI.
- Kosem, Iztok; Arhar Holdt, Špela; Stritar Kučuk, Mojca; Krek, Simon; Krapš Vodopivec, Irena; Stabej, Marko; Kocjančič, Polonca; Laskowski, Cyprian; Klemenc, Bojan; Pori, Eva; Rozman, Tadeja, 2019, [Developmental corpus \(without language corrections\) Šolar 2.0 Clear](#), Slovenian language resource repository CLARIN.SI.
- Arhar Holdt, Špela; Goli, Teja; Lavrič, Polona; Laskowski, Cyprian; Klemenc, Bojan; Rozman, Tadeja; Stritar Kučuk, Mojca; Krek, Simon; Krapš Vodopivec, Irena; Stabej, Marko; Kosem, Iztok, 2019, [Error-annotated developmental corpus Šolar 2.0 Error](#), Slovenian language resource repository CLARIN.SI.
- Arhar Holdt, Špela; Rozman, Tadeja; Stritar Kučuk, Mojca; Krek, Simon; Krapš Vodopivec, Irena; Stabej, Marko; Pori, Eva; Goli, Teja; Lavrič, Polona; Laskowski, Cyprian; Kocjančič, Polonca; Klemenc, Bojan; Krsnik, Luka; Kosem, Iztok; 2022, [Developmental corpus Šolar 3.0](#), Slovenian language resource repository CLARIN.SI.



Podobni korpusi za druge jezike

- Angleščina: **The LUCY Corpus** (Sampson 2003): mladostniki (33.000 besed), otroci 9–12 let (30.000 besed); **the LOCNESS corpus** (Granger, 1998): britanski učenci (60.209 besed), britanski in ameriški študenti (264.095 besed); **zbirka esejev Parr** (2010): 20.947 esejev novozelandskih učencev 4.–12. leto šolanja.
- Nemščina: **H1, H2, E2, ERK1** (Berkling 2016, 2018): šolarji 6–11 let (81.385 pojavnic); **the Litkey Corpus** (Laarman-Quante et al. 2019): 2.–4. razred (212.505 pojavnic); **the KoKo Corpus** (Abel et al. 2014): srednješolski eseji (950.000 pojavnic).
- Italijanščina: **CItA** (Barbagli et al. 2016): 1.352 srednješolskih esejev; 5.000 spisov 1.–5. razred (Marconi et al. 1993) in 2.500 spisov iz 1. letnika rimskih SŠ (Borghini 2013).
- Francoščina: **È:CALM** (Ho-Dac et al. 2020), ki združuje cca. 6.700 besedil od OŠ do fakultete iz različnih korpusov (Garcia-Debanco in Bonnemaïson 2014; Doquet et al. 2017; Jacques in Rinck 2017; Wolfarth et al. 2017).
- **LEONIDE** (Glaznieks et al. 2022): 2.512 srednješolskih esejev z večjezikovnega južnotirolskega območja. **DOESTE v0.5** (Martins et al. 2020): učenci in dijaki 10–18 let (244 esejev v portugalsščini in 450 v brazilski portugalsščini). **The Icelandic (Child) Error Corpus** (Arnardóttir et al. 2021, Ingason et al. 2021): 176 srednješolskih esejev in 119 besedil učencev 10.–15. leta. Češki **Chyby** (Pala et al. 2003): cca. 410.000 besed študentskih izdelkov – vključuje avtentične popravke dveh profesorjev.

Raziskave in objave

Opis namena in sestave korpusa Šolar 1.0 in 2.0

- Iztok Kosem, Sara Može (2011): [Rešitve slovnicih zagat na doseg miške: analiza napak v besedilih učencev in dijakov za potrebe elektronskega slovnicega vira](#). V: S. Krajnc (ur.) Meddisciplinarnost v slovenistiki, (Obdobja, Simpozij). Ljubljana: Znanstvena založba Filozofske fakultete, str. 249-257.
- KOSEM, Iztok; ROZMAN, Tadeja; STRITAR, Mojca. [How do Slovenian primary and secondary school students write and what their teachers correct: A corpus of student writing](#). In: *Proceedings of Corpus Linguistics Conference*. 2011. p. 20-22.
- Tadeja Rozman, Irena Krapš Vodopivec, Mojca Stritar Kučuk, Iztok Kosem. 2012. [Empirični pogled na pouk slovenskega jezika](#). Zbirka Sporazumevanje.
- Iztok Kosem, Mojca Stritar Kučuk, Sara Može, Ana Zwitter Vitez, Špela Arhar Holdt, Tadeja Rozman. 2012. [Analiza jezikovnih težav učencev, korpusni pristop](#). Zbirka Sporazumevanje.
- KOSEM, Iztok, ROZMAN, Tadeja, ARHAR HOLDT, Špela, KOCJANČIČ, Polonca, LASKOWSKI, Cyprian Adam. [Šolar 2.0: nadgradnja korpusa šolskih pisnih izdelkov](#). V: ERJAVEC, Tomaž (ur.), FIŠER, Darja (ur.). *Jezikovne tehnologije in digitalna humanistika: zbornik konference*, 2016. V Ljubljani: Znanstvena založba Filozofske fakultete. Str. 95-100.

Specializirani vmesnik za korpusne s popravki

- KOSEM, Iztok, Vit Baisa, Vojtěch Kovář and Adam Kilgarriff. [User-friendly interface of error/correction-annotated corpus for both teachers and researchers](#). *Book of Abstracts LCR 2013*, Norway, September 2013, pp. 82–84.

Raba korpusa za pripravo didaktičnih gradiv

- ARHAR HOLDT, Špela, DOBROVOLJIC, Kaja, POPIČ, Damjan. [Reprezentacija standardnega in nestandardnega v virih SSJ](#). V: ŽELE, Andreja (ur.). *Družbena funkcijskost jezika*: (vidiki, merila, opredelitve). Ljubljana: Znanstvena založba Filozofske

fakultete, 2013. Str. 19-27.

- ARHAR HOLDT, Špela, ROZMAN, Tadeja. *Korpus Šolar: gradivni vir za raziskave pisne produkcije učencev*. *Bilten Bravo*. dec. 2015, letn. 11, št. 23, str. 17-23.
- ARHAR HOLDT, Špela; KOSEM, Iztok; GANTAR, Polona. *Corpus-based resources for L1 teaching: The case of Slovene*. *Handbook on digital learning for K-12 schools*, 2017, 91-113.
- ARHAR HOLDT, Špela, POLLAK, Senja, ROBNIK ŠIKONJA, Marko, KREK, Simon. [Referenčni seznam pogostih splošnih besed za slovenščino](#). V: FIŠER, Darja (ur.), ERJAVEC, Tomaž (ur.). *Jezikovne tehnologije in digitalna humanistika: zbornik konference, 2020*. Str. 10-15.
- ARHAR HOLDT, Špela, KOSEM, Iztok, PORI, Eva. [Jezikovni viri CJVT in njihova raba v izobraževalne namene](#). V: ULČNIK, Natalija (ur.), ANTLOGA, Špela. *Slovenščina na dlani 4*. Maribor: Univerza v Mariboru, Univerzitetna založba, 2021. Str. 19-36.

Raba korpusa za pripravo pedagoškega jezikovnega opisa

- Arhar Holdt, Špela, Rozman, Tadeja, 2015: [Možnosti uporabe podatkov iz korpusa Šolar za pripravo slovarskih priročnikov](#). Smolej, Mojca (ur.): *Obdobja 34: Slovnica in slovar, aktualni jezikovni opis*. Ljubljana: Znanstvena založba Filozofske fakultete. 67–74.
- ROZMAN, Tadeja, ARHAR HOLDT, Špela, POLLAK, Senja, KOSEM, Iztok. [Kolokacije v korpusu Šolar](#). *Jezik in slovnstvo*. 2018, letn. 63, št. 2/3, str. [117]-128.

Metodologija korpusne gradnje

- ARHAR HOLDT, Špela, KOSEM, Iztok, STRITAR KUČUK, Mojca. [Metode in orodja za lažjo pripravo korpusov usvajanja jezika](#). V: PIRIH SVETINA, Nataša (ur.), FERBEŽAR, Ina (ur.). *Na stičišču svetov: slovenščina kot drugi in tuji jezik*. 1. natis. Ljubljana: Založba Univerze, 2022a. Str. 23-30, Zbirka Obdobja, 41.
- ARHAR HOLDT, Špela, LAVRIČ, Polona, ROBLEK, Rebeka, GOLI, Teja, 2022b: [Kategorizacija učiteljskih popravkov: Smernice za označevanje korpusa Šolar](#). Različica 1.1. Rezultat projekta Razvoj slovenščine v digitalnem okolju.

Uporaba za strojno učenje in evalvacijo orodij za popravlanje

- Holozan, Peter (2013). [Uporaba strojnega učenja za postavljanje vejic v slovenščini](#). *Uporabna informatika* (Ljubljana), letnik 21, številka 4, str. 196-209.
- KRAJNC, Anja, ROBNIK ŠIKONJA, Marko. [Postavljanje vejic v slovenščini s pomočjo strojnega učenja in izboljšane korpusa Šolar](#). V: FIŠER, Darja (ur.). *Zbornik konference Slovenščina na spletu in v novih medijih, Ljubljana, 25.-27. november 2015*. Ljubljana: Znanstvena založba Filozofske fakultete, 2015. Str. 38-43.
- ANDOVA, Andrejaana. *Ugotavljanje berljivosti besedil z uporabo statističnih mer in strojnega učenja*. 2017. PhD Thesis. Univerza v Ljubljani.
- ŽAGAR, Aleš. *Readability Model Based on ccGigafida and Šolar Corpora*. In: *MEi: CogSci Conference 2018*. 2018. p. 103.
- BOŽIČ, Martin. [Globoke nevronske mreže za postavljanje vejic v slovenskem jeziku: diplomsko delo: univerzitetni študijski program prve stopnje Računalništvo in informatika](#). Ljubljana: [M. Božič], 2020.
- PETRIČ, Timotej. [Predlogi jezikovnih popravkov v slovenščini z modelom SloBERTa: diplomsko delo: univerzitetni študijski program prve stopnje Računalništvo in informatika](#). Ljubljana: [T. Petrič], 2022.
- MOKOTAR, Rok. [Obvladovanje slovnicih napak v šolskih pisnih izdelkih z metodami za obdelavo naravnega jezika: diplomsko delo: univerzitetni študijski program prve stopnje Računalništvo in informatika](#). Ljubljana: [R. Mokotar], 2023.