



INŠTITUT
ZA NOVEJŠO ZGODOVINO



Digitalna humanistika in zgodovinopisje

Tjaša Konovšek
tjasa.konovsek@inz.si

Katja Meden
katja.meden@inz.si

Problemi zgodovinopisja (v povezavi z DH)

- teme

<http://republicofletters.stanford.edu/>

- postopek raziskovanja

(literatura, vprašanje, raziskava)

- viri (katere in kako)

- DH pri lastnem delu



Arhiv Republike Slovenije

Digitalna humanistika

Definicije zagotavlja ponudnik [Oxford Languages](#) · [Več o tem](#)



digital humanities

noun

an academic field concerned with the application of computational tools and methods to traditional humanities disciplines such as literature, history, and philosophy.

"the unit will advance scholarship in both classical studies and **the digital humanities**"

Raziskovalne infrastrukture

- **DARIAH-EU** (Digital Research Infrastructure for the Arts and Humanities)
 - DARIAH-SI <http://www.dariah.si/>
 - SI-DIH (<https://sidih.si/20.500.12325/menu1>)
- **CLARIN.EU** (Common Language Resources and Technology Infrastructure)
 - CLARIN.SI <https://www.clarin.si/info/o-projektu/>
 - CLARIN.SI repozitorij: <https://www.clarin.si/repository/xmlui/>



Korpusi in konkordančniki

kórpus -a m (ô)

SSKJ² 

1. *vojaška enota iz divizij in posebnih enot*: formirati korpus; poveljnik korpusa / ekspedicijski korpus
2. jezikosl. *zbirka besedil v elektronski obliki s standardizirano obliko zapisa in izbranimi podatki*: analiza korpusa; delo s korpusom; besedila za korpus; korpus in slovar / referenčni korpus *ki je glede na vključena besedila zvrstno kar se da uravnotežen / besedilni korpus*
3. publ., navadno s prilastkom *skupina, zbor*: nastopili so najboljši korpusi; baletni, godalni korpus / diplomatski korpus *diplomatski zbor*; korpus miru zlasti v ameriškem okolju *organizacija iz prostovoljcev, ki skrbi zlasti za gospodarski razvoj in pomoč nerazvitim deželam*
4. knjiž. *zbirka*: korpus srednjeveških fresk
 - ekspr. *ta ženska je pa korpus je zelo debela*

- Extensible Markup Language (XML)
- Text Encoding Initiative
- NoSketch, KonText (konkordančnika)

Attributes **ana** and **msd** both encode the morphosyntactic characteristics of the word in focus (in blue). The **ana** attribute contains Slovene-specific tags according to MULTTEXT-East Specifications, while the **msd** attribute contains the Universal Dependency tagset which greatly simplifies cross-lingual comparison.

The attribute values Q and PART both stand for *particle*. The difference between the two tags is in the tagset used.

The **word form** also known as a token from the running text in the corpus.

The attribute **lemma** indicates the basic word form of the token, that is of the word in focus (in blue).

The corpus also encodes syntactic parses and named entities but since they are not used in this tutorial they were omitted from this illustrative example.

```
<s>
<w ana="mte:Q" msd="UpoSTag=PART" lemma="zlasti">zlasti</w>
<w ana="mte:Sg" msd="UpoSTag=ADP|Case=Gen" lemma="glede">glede</w>
<w ana="mte:Ncmsg" msd="UpoSTag=NOUN|Case=Gen|Gender=Masc|Number=Sing"
lemma="nadzor">nadzora</w>
<w ana="mte:Va-r3s-n"
msd="UpoSTag=AUX|Mood=Ind|Number=Sing|Person=3|Polarity=Pos|Tense=Pres|VerbForm=Fin"
lemma="biti">je</w>
<w ana="mte:Pd-fsn" msd="UpoSTag=DET|Case=Nom|Gender=Fem|Number=Sing|PronType=Dem"
lemma="ta">ta</w>
<w ana="mte:Ncfsn" msd="UpoSTag=NOUN|Case=Nom|Gender=Fem|Number=Sing"
lemma="stvar">stvar</w>
<w ana="mte:Rgp" msd="UpoSTag=ADV|Degree=Pos" lemma="zelo">zelo</w>
<w ana="mte:Agpfsn" msd="UpoSTag=ADJ|Case=Nom|Degree=Pos|Gender=Fem|Number=Sing"
lemma="kočljiv">kočljiva</w>
<pc ana="mte:Z" msd="UpoSTag=PUNCT">.</pc>
</s>
```

One pair of the opening and closing structural tags which, in our case, indicate punctuation (**pc**), word (**w**) or sentence (**s**).

Korpusi parlamentarnih razprav

- Parlamentarne razprave: transkripcije plenarnih zasedanj, zapisniki sej
- Transkripcije govornega jezika
- Dragoceni (sociodemografski) metapodatki:
 - Ime in priimek govornika
 - Letnica rojstva
 - Politične / parlamentarne stranke
 - Spol govornika...

Lastnosti

- Prepisi niso vedno dobesedni
- Izpuščene napake, neprimeren jezik
- Oblikovani v skladu z konvencijami parlamenta:
 - Grški parlament: prekinitve strogo prepovedane, sankcionirane
 - Britanski parlament: prekinitve so popolnoma običajne in niso sankcionirane
 - Poslanci imajo pogosto vnaprej pripravljene govore
- TEI kodiranje in jezikovna anotacija besedila

Official records

"Madam President, I think we all agreed that equal pay is, first of all, a matter of elementary fairness, and the fact that we still have this average pay gap of 16-17% also shows, among other things, the lack of respect for women, which is so deeply rooted in traditions and stereotypes. /.../."

(Věra Jourová, Member of the Commission)

Verbatim transcription

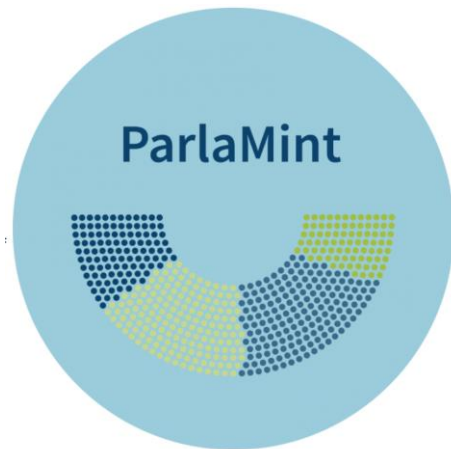
*"**Ladies and Gentlemen, yes gentlemen also are here still. Am...** I think we all agreed that equal pay is, first of all, a matter of elementary fairness, and **that** the fact that still **we** have this **pay gap am...** average 16-17% **this** also shows, among other things, the lack of respect to women, which is so deeply rooted in traditions and stereotypes. /.../."*

(Věra Jourová, Member of the Commission)

Darja Fišer in Kristina Pahor de Maiti. 2021. *Prvič, sem političarka in ne politik, drugič pa ---* : a corpus approach to parliamentary discourse research. Inštitut za novejšo zgodovino, Ljubljana. Dostopno na: <https://sidih.si/cdn/121/ch3.html>

ParlaMint I in II

- ParlaMint I (Julij 2020 – Maj 2021):
 - 17 nacionalnih evropskih parlamentarnih korpusov
 - Vsak korpus vsebuje približno 20 milijonov besed
 - COVID/Reference
- ParlaMint II (December 2021 – Maj 2023)
 - Novi partnerji (25+ korpusov)
 - Novi metapodatki
 - Strojno prevajanje besedila
 - ParlaMint 2.1: <http://hdl.handle.net/11356/1432>
 - ParlaMint-SI (7 in 8. mandat)



SiParl 2.0: Korpus parlamentarnih razprav Državnega zbora Republike Slovenije 1990-2018

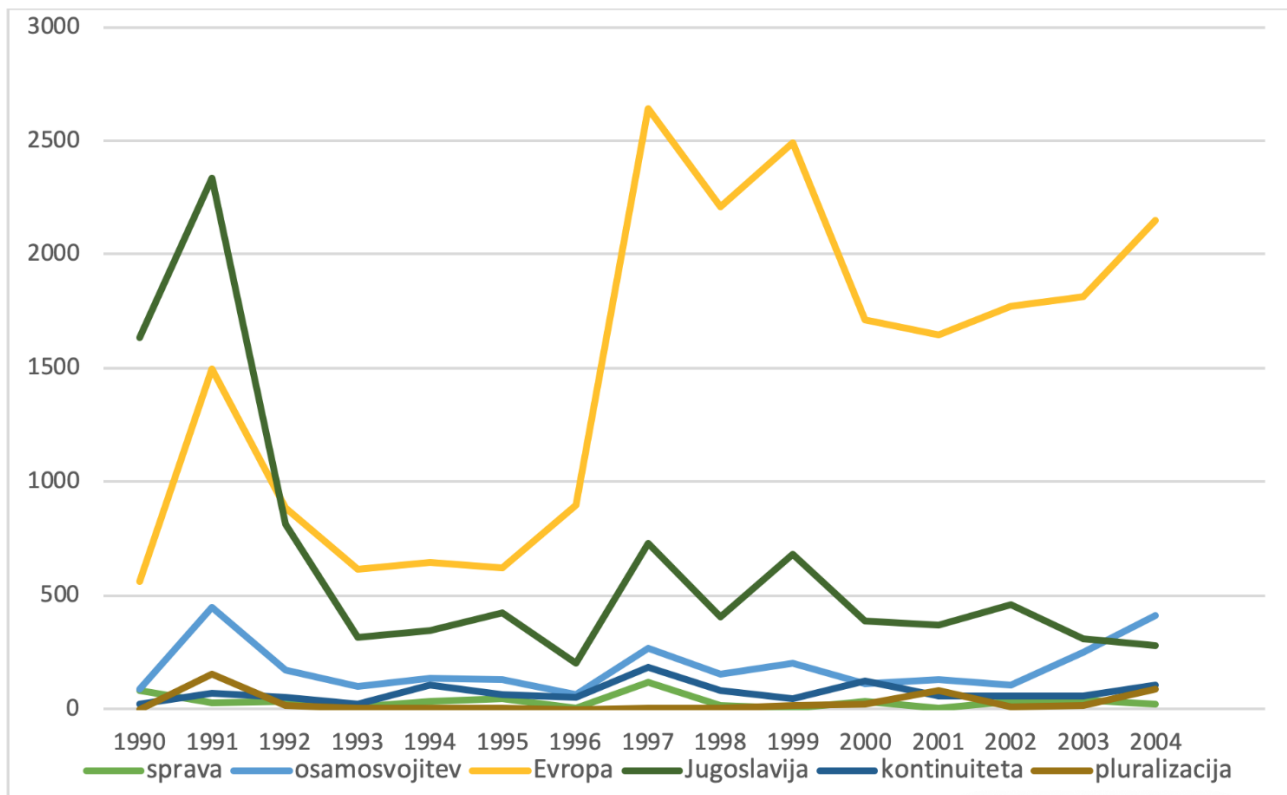
- SiParl 3.0: <http://hdl.handle.net/11356/1748>
- Raziširitev do 13.5.2022 (8. mandat)
- Primerjava s ParlaMint korpusom:
 - 11. Sklic Skupščine v 1991
 - 1. do 8. Mandat
 - Delovna telesa in komisije
 - Prilagojeno kodiranje v TEI
- Konkordančniki - SiParl 2.1 in ParlaMint 2.1

Primer 1: Sprava

<https://sciendo.com/article/10.2478/jnmlp-2021->

0006

- problem: polarizacija političnega prostora, legitimacija države (zgo: empirija in čas)
- vprašanje: odnos do preteklosti in politična moč
- izbrana metodologija: konceptualna/politična zgodovina; digitalna humanistika
- postavljanje časovnice + historični kontekst

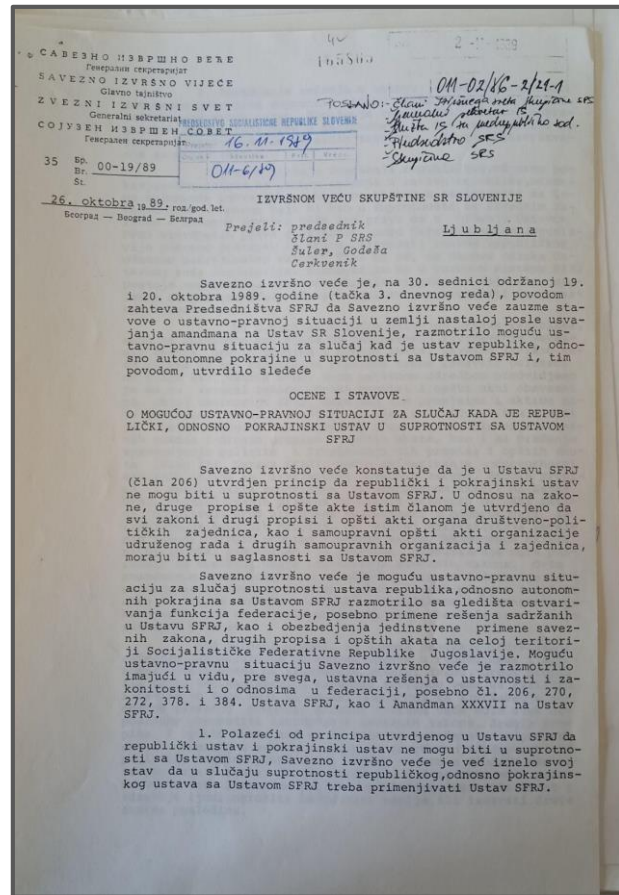
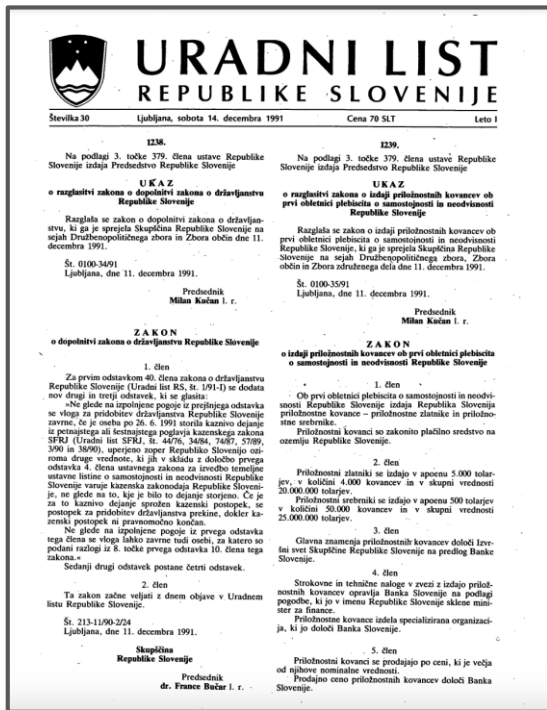


Graf 1: Sprava, osamosvojitev, Evropa, Jugoslavija (1990 – 2004).

Series "kontinuiteta" Legend Entry

Število omemb besed sprava, osamosvojitev, Evropa in Jugoslavija v skupščini oz. državnem zboru med letoma 1990 in 2004. Vir: <https://www.clarin.si/kontext/corpora/corplist> (7. 6. 2021.)

- kombiniranje z drugimi viri
(kvalitativna analiza)

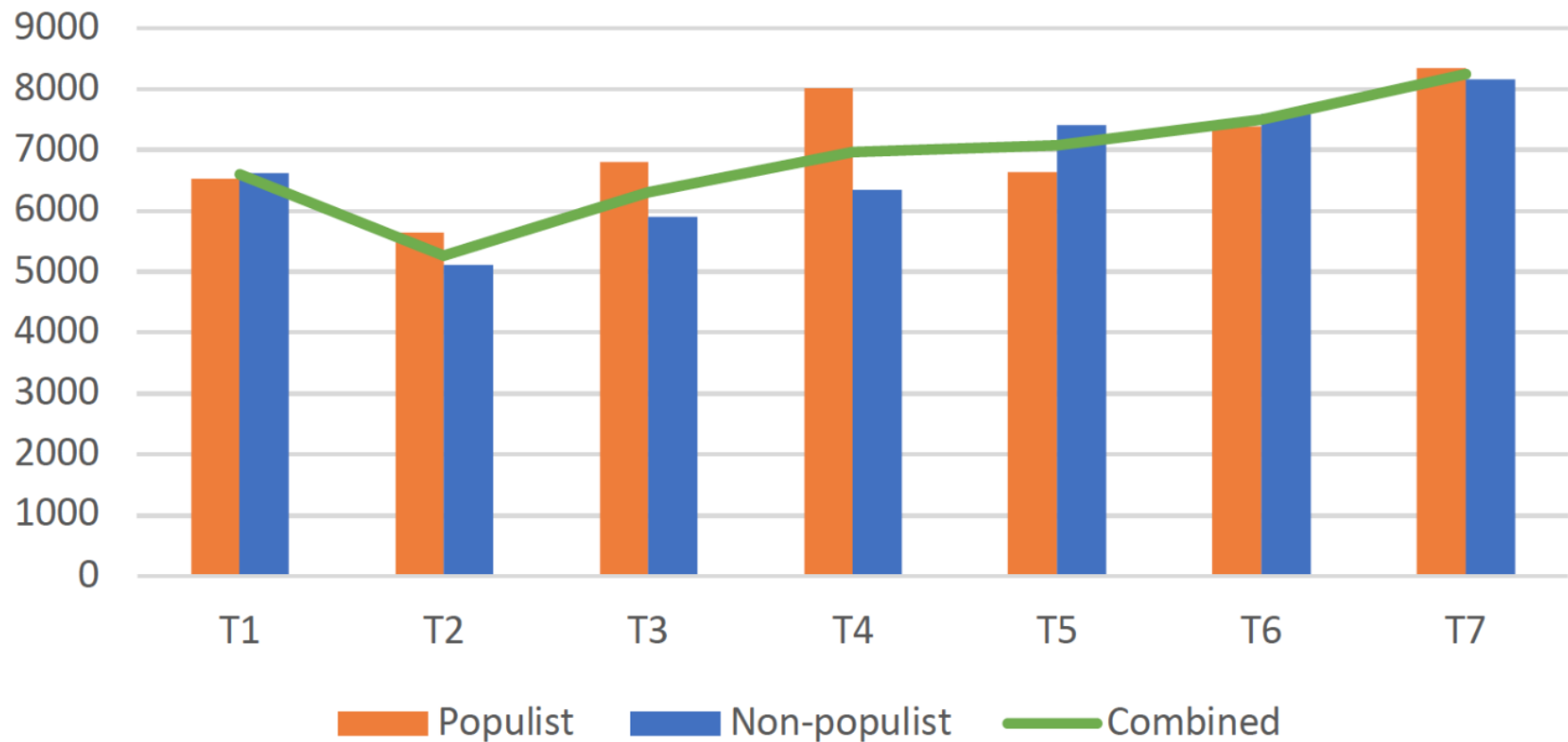


Primer 2: Populizem

- problem: polarizacija političnega prostora, populizem
- vprašanje: kdo in kdaj naslavlja javnost; moč parlamenta
- izbrana metodologija: kulturna/politična zgodovina (*Lebenswelt*); digitalna humanistika
- analiza velike količine podatkov (kvantitativna analiza)

		POPULIST1-7		NON-POPULIST1-7		
	#tokens	49,070,504		77,215,381		
	#lemmas	76		74		
	LEMMA	AF	RF	AF	RF	P:N ratio
P-ONLY	Hrvat	1,341	27.33	0	0.00	/
	žena	397	8.09	0	0.00	/
	Avstrijec	318	6.48	0	0.00	/
	diplomant	300	6.11	0	0.00	/
	storilec	232	4.73	0	0.00	/
	volilec	161	3.28	0	0.00	/
	delojemalec	36	0.73	0	0.00	/
	neslovenec	31	0.63	0	0.00	/
	svojec	27	0.55	0	0.00	/
	delavka	0	0.00	0	0.00	/
N-ONLY	deležnik	0	0.00	1,784	23.10	/
	prejemnik	0	0.00	1,191	15.42	/
	najemnik	0	0.00	983	12.73	/
	dolžnik	0	0.00	752	9.74	/
	vajenec	0	0.00	444	5.75	/
JOINT	kadilec	0	0.00	290	3.76	/
	krajan	0	0.00	172	2.23	/
	oče	929	18.93	329	4.26	4.44
	obrtnik	1,187	24.19	540	6.99	3.46
	davkoplačevalec	4,762	97.04	2,178	28.21	3.44
	migrant	2,627	53.54	1,255	16.25	3.29
	vlagatelj	426	8.68	260	3.37	2.58
	podjetnik	3,880	79.07	2,671	34.59	2.29
	moški	827	16.85	619	8.02	2.10
	ljudstvo	3,089	62.95	2,376	30.77	2.05
	Italijan	272	5.54	216	2.80	1.98
	Slovenka	1,432	29.18	1,143	14.80	1.97
	pacient	1,619	32.99	1,452	18.80	1.75
	zamejstvo	1,067	21.74	966	12.51	1.74
	kmet	6,839	139.37	6,739	87.28	1.60
	prijatelj	1,024	20.87	1,012	13.11	1.59
	naročnik	517	10.54	516	6.68	1.58
	Slovenec	10,103	205.89	11,090	143.62	1.43
	dijak	2,403	48.97	2,670	34.58	1.42
	kupec	1,216	24.78	1,357	17.57	1.41
	državljan	21,570	439.57	24,828	321.54	1.37
	priča	4,061	82.76	4,701	60.88	1.36
	državljanka	6,902	140.65	8,372	108.42	1.30
	narod	4,952	100.92	6,035	78.16	1.29
	žrtev	3,945	80.39	4,810	62.29	1.29
	sosed	738	15.04	928	12.02	1.25
	človek	68,517	1,396.30	86,824	1,124.44	1.24

Rom	627	12.78	808	10.46	1.22
bolnik	1,279	26.06	1,717	22.24	1.17
prošilec	343	6.99	468	6.06	1.15
javnost	16,248	331.12	22,367	289.67	1.14
starš	5,732	116.81	7,893	102.22	1.14
oseba	16,836	343.10	23,762	307.74	1.11
subjekt	3,406	69.41	4,866	63.02	1.10
družina	11,120	226.61	16,298	211.07	1.07
otrok	18,205	371.00	26,762	346.59	1.07
gost	966	19.69	1,438	18.62	1.06
begunec	1,247	25.41	1,879	24.33	1.04
mladina	1,384	28.20	2,101	27.21	1.04
delničar	444	9.05	684	8.86	1.02
tujec	3,169	64.58	4,908	63.56	1.02
zavarovanec	896	18.26	1,394	18.05	1.01
volivec	3,478	70.88	5,544	71.80	0.99
lastnik	8,031	163.66	12,814	165.95	0.99
mati	320	6.52	512	6.63	0.98
družba	23,431	477.50	38,532	499.02	0.96
študent	4,973	101.34	8,202	106.22	0.95
posameznik	7,367	150.13	12,307	159.39	0.94
zavezanec	2,437	49.66	4,096	53.05	0.94
uporabnik	3,441	70.12	5,866	75.97	0.92
nosilec	2,211	45.06	3,812	49.37	0.91
občan	1,558	31.75	2,688	34.81	0.91
prebivalec	5,318	108.37	9,404	121.79	0.89
partner	4,580	93.34	8,312	107.65	0.87
potrošnik	1,657	33.77	3,060	39.63	0.85
generacija	2,279	46.44	4,215	54.59	0.85
delavec	10,768	219.44	20,055	259.73	0.84
invalid	3,032	61.79	5,760	74.60	0.83
prebivalstvo	2,727	55.57	5,452	70.61	0.79
manjšina	2,742	55.88	5,518	71.46	0.78
učenec	1,437	29.28	3,071	39.77	0.74
ženska	2,941	59.93	6,517	84.40	0.71
upokojenec	3,547	72.28	8,097	104.86	0.69
skupnost	16,208	330.30	38,163	494.24	0.67
priladnik	1,375	28.02	3,238	41.93	0.67
upravičenec	1,673	34.09	4,523	58.58	0.58
upnik	566	11.53	1,725	22.34	0.52
podpisnik	465	9.48	1,460	18.91	0.50
udeleženeec	500	10.19	1,685	21.82	0.47
porabnik	129	2.63	540	6.99	0.38
populacija	480	9.78	2,179	28.22	0.35
Total	359,320	7,322.53	524,195	6,788.74	1.08



Primer 3: Konkordančnik in SiParl

- NoSketch Engine: <https://www.clarin.si/noske/>
- KonText: <https://www.clarin.si/kontext/corpora/corplist>
- Primeri:
 - krava
 - mačka
 - ?



institut_za_novejso_zgodovino



Inštitut za novejšo zgodovino



History on the Edge / Zgodovina na Špici



Clarin.si @ClarinSlovenia



INŠTITUT
ZA NOVEJŠO ZGODOVINO

Tjaša Konovšek

tjasa.konovsek@inz.si

Katja Meden

katja.meden@inz.si

katja.meden@ijs.si