Nataša Logar

4 INTERPRETACIJA KORPUSNIH PODATKOV

4.1 Interpretacija korpusnih podatkov – izročki

0.2

a) Korpus = vzorec.

Vzorec je podmnožice celotne populacije, ki je bila na predviden način (i)zbrana za raziskavo. Raziskovalci morajo pri posploševanju podatkov na podlagi vzorca upoštevati način, na katerega je bil ta narejen.

(Peck, Olsen, Delore 2009: 8, 33; nav. po Logar, Dobrovoljc, Arhar Holdt 2015: 467.)

b) Korpus = veliko podatkov

– tudi **številskih podatkov** –

in številske podatke se da **vizualizirati**.

1. Korpus = vzorec

Dejavnikov, ki povzročajo napačno interpretacijo statističnih podatkov, je več:

- Pristranskost raziskovalca
- Nekonsistentnost v definicijah
- Primerjave in povezave med neprimerljivimi oz. nepovezljivimi spremenljivkami
- Napačno sklepanje na celotno populacijo ←

itd.

Primer:

„Število avtomobilskih nesreč, ki so jih v nekem mestu v določenem letu povzročile voznice, je 10. Število avtomobilskih nesreč, ki so jih v istem mestu in istem letu povzročili vozniki, je 40.

Torej: voznice so vozile bolj varno.“

Zakaj je sklep napačen?

(Vir: Jogi 2014: 2.)

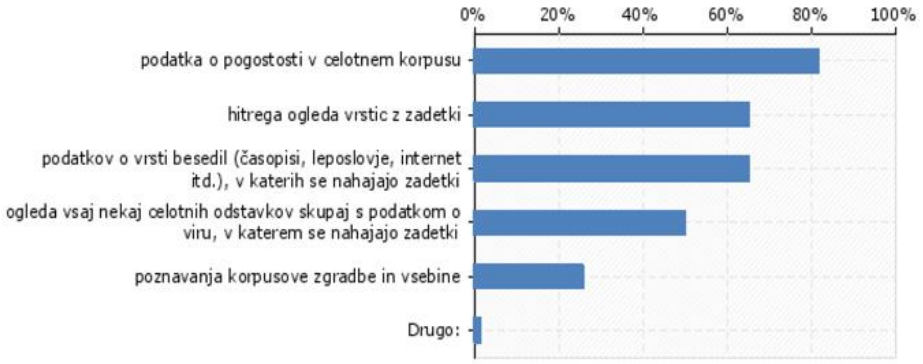
Pri korpusih so za interpretacijo rezultatov poizvedb pomembni dvoji podatki:*

- *predočeni podatki*: podatki, ki so prikazani hkrati z rezultati poizvedb
- *zaledni podatki*: podatki, ki so povezani z izdelavo korpusa

* + poznavanje konkordančnika (v primeru iskanja po spletno dostopnem korpusu)

Anketa med uporabniki korpusa Gigafida (Logar, Dobrovoljc, Arhar Holdt 2015):

Pri interpretaciji podatkov iz korpusa Gigafida večinoma izhajam iz ...



Pogovor ob primerih

1.187.002.502
(ne 1 milijarda)

PRIMER 1: V čem je šibkost naslednje tabele /op.: primer je avtentičen/?

Tabela 1: Primerjava absolutne pogostnosti pojavnic za navedene pare v korpusih FidaPLUS in Gigafida

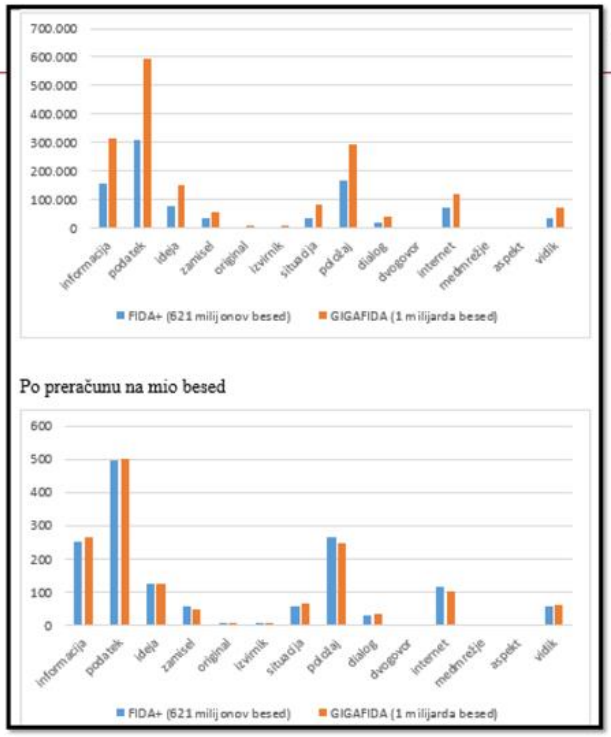
	FIDA+ (621 milijonov besed)	GIGAFIDA (1 milijarda besed)
informacija/podatek	157.306/307.641	317.803/594.487
ideja/zamisel	78.346/35.291	149.996/56.857
original/izvirnik	5.513/6.140	11.009/10.824
situacija/položaj	35.871/165.981	82.306/293.253
dialog/dvogovor	20.813/474	43.155/708
internet/medmrežje	73.222/2.827	121.394/5.265
aspekt/vidik	2.400/34.898	4.699/74.898

>>

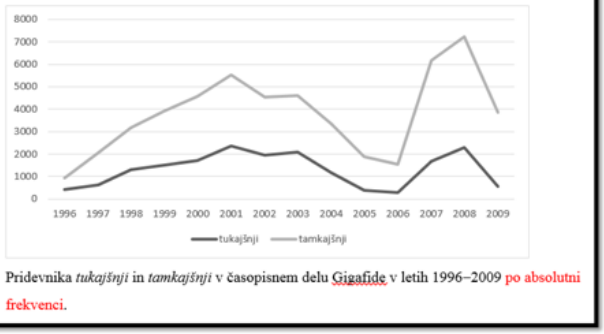
Po preračunu na milijon besed:

	FidaPLUS (621 milijonov besed)	GIGAFIDA (1.187.002.502 besedi)
informacija	253	268
podatek	495	500
ideja	126	126
zamisel	57	48
original	9	9
izvirnik	10	9
situacija	58	69
položaj	267	247
dialog	33	36
dvogovor	1	0,5
internet	118	102
medmrežje	5	4
aspekt	4	4
vidik	56	63

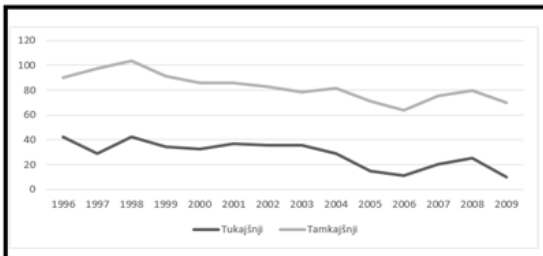
>>
Ali v sliki:



Po preračunu na mio besed



Pridevnika *tukajšnji* in *tamkajšnji* v časopisnem delu Gigafide v letih 1996–2009 po absolutni frekvenci.



Pridevnika *tukajšnji* in *tamkajšnji* v časopisnem delu Gigafide v letih 1996–2009 glede na število vseh besed iz časopisov v teh letih (v številu besed na milijon).

Pozor: absolutna : relativna frekvenca – ČAS

(Vir druge slike:
Kalin Golob, Logar 2014.)

PRIMER 2 /op.: izhodišče je avtentično/

Radijska oddaja Izvidnica (Val 202, 25. 12. 2015), voditelj Damjan Zorc:

„Ali je spletna materinščina vse bolj anglizirana, pravzaprav bi najbrž moral reči amerikanizirana?“

V čem je problematičnost naslednjega pristopa in posledično interpretacije pridobljenih podatkov?

Pristop: primerjava seznama besed v korpusu spletne slovenščine JANES s seznamom besed v korpusu splošne slovenščine KRES > gl. naslednjo sliko.

Seznam besed

Korpus: JANES v0.4

Referenčni korpus: KRES (uravnoteženi)

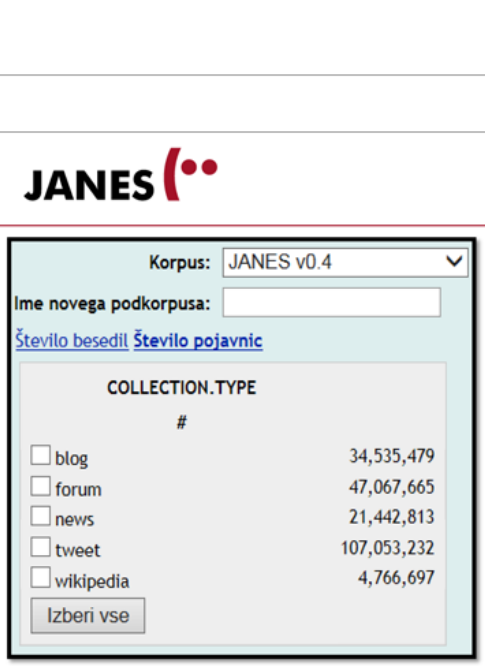
Spremeni cilj in referenčni (pod)korpus

Stran 1 Pojdi Naslednja >

	JANES v0.4		KRES (uravnoteženi)		
lemma	Frekvenca	Frekvenca/mil	Frekvenca	Frekvenca/mil	Rezultat
t.co	2,773,177	12906.5	0	0.0	12907.5
The	66,929	311.5	31	0.3	248.5
IS	137,826	641.5	218	1.8	228.6
via	107,203	498.9	143	1.2	228.6
posted	44,411	206.7	10	0.1	191.8
Slovenc	50,619	235.6	55	0.5	162.4
he	28,713	133.6	1	0.0	133.5
Facebook	59,691	277.8	164	1.4	118.1
photo	38,625	179.8	75	0.6	111.4
kea	32,693	152.2	70	0.6	96.9
automatically	20,238	94.2	3	0.0	92.9
fora	108,790	506.3	552	4.6	90.9
OF	158,740	738.8	918	7.6	85.8
checked	20,555	95.7	19	0.2	83.5
Wikipedija	18,449	85.9	5	0.0	83.4
Tuta	32,385	150.7	100	0.8	82.9
followed	19,606	91.2	15	0.1	82.0
Us	30,663	142.7	100	0.8	78.5
tvit	16,569	77.1	3	0.0	76.2
ne vedeti	16,132	75.1	0	0.0	76.1
po moj	14,572	67.8	0	0.0	68.8
I'm	21,496	100.0	67	0.6	64.9
How	13,441	62.6	3	0.0	62.0
tkati	59,716	277.9	424	3.5	61.7
IT	107,206	498.9	875	7.3	60.5
hehe	23,143	107.7	123	1.0	53.8
Justa	39,725	184.9	299	2.5	53.4
my	61,845	287.8	553	4.6	51.7
unfollowed	10,858	50.5	0	0.0	51.5
liked	11,102	51.7	3	0.0	51.4
Ip	33,428	155.6	269	2.2	48.4
Lp	26,438	123.0	195	1.6	47.4
NSi	9,864	45.9	0	0.0	46.9
What	14,297	66.5	57	0.5	45.8

1. JANES : KRES

>> Problematičnost pristopa je v neupoštevanju zgradbe korpusa JANES. Gl. naslednjo sliko.



JANES: zgradba

JANES

IJS

>> Če upoštevamo zgradbo korpusa JANES, dobimo naslednje slike.

Kakšna je torej pravilna interpretacija podatkov (s previdnostjo, da smo si ogledali le vrhnji del rezultatov)?

Seznam besed

Korpus: JANES v0.4
Podkorpus: Blogi

Referenčni korpus: KRES (uravnoteženi)
Spremeni cilj in referenčni (pod)korpus

Stran 1 Pojdi Naslednja >

	JANES v0.4 : Blogi		KRES (uravnoteženi)		
lemma	Frekvenca	Frekvenca/mil	Frekvenca	Frekvenca/mil	Rezultat
Slovenc	10,602	307.0	55	0.5	211.4
Kouvran	6,073	175.8	20	0.2	151.7
he	4,661	135.0	1	0.0	134.8
androma	3,005	87.0	0	0.0	88.0
DI	2,569	74.4	0	0.0	75.4
Zzzzzz	2,342	67.8	0	0.0	68.8
Androma	2,140	62.0	0	0.0	63.0
The	2,683	77.7	31	0.3	62.6
blog	21,271	615.9	1,155	9.6	58.3
Lp	5,112	148.0	195	1.6	56.9
cinik	3,632	105.2	106	0.9	56.5
branka22	1,936	56.1	5	0.0	54.8
Andro	2,175	63.0	21	0.2	54.5
komsomolec	2,156	62.4	23	0.2	53.3
www.youtube.com	1,786	51.7	2	0.0	51.9
Tebe	2,813	81.5	92	0.8	46.7
hehe	2,825	81.8	123	1.0	41.0
lp	4,338	125.6	269	2.2	39.2
IS	3,678	106.5	218	1.8	38.3
johanbank	1,276	36.9	1	0.0	37.6
politkomisar	1,975	57.2	68	0.6	37.2
OF	10,663	308.8	918	7.6	35.9
dajan	1,514	43.8	39	0.3	33.9
sailor	1,252	36.3	13	0.1	33.6
Zzzzz	1,061	30.7	0	0.0	31.7
Kobilica-poet	950	27.5	0	0.0	28.5
Tupamaross	975	28.2	4	0.0	28.3
www.publishwall.si	921	26.7	0	0.0	27.7
d.o.o.	907	26.3	0	0.0	27.3
sintrati	1,021	29.6	15	0.1	27.2
Us	1,608	46.6	100	0.8	26.0
BiH	857	24.8	0	0.0	25.8
tupamaross	883	25.6	4	0.0	25.7
lea199	823	23.8	0	0.0	24.8
gromovnik	904	26.2	15	0.1	24.2

IJS

1. JANES – blogi : KRES

Seznam besed

Korpus: JANES v0.4
Podkorpus: Forumi

Referenčni korpus: KRES (uravnoteženi)
Spremeni cilj in referenčni (pod)korpus

Stran 1 Pojdi Naslednja >

	JANES v0.4 : Forumi		KRES (uravnoteženi)		
lemma	Frekvenca	Frekvenca/mil	Frekvenca	Frekvenca/mil	Rezultat
ne vedeti	11,252	239.1	0	0.0	240.1
lp	25,564	543.1	269	2.2	168.3
kea	11,977	254.5	70	0.6	161.6
feltna	8,934	189.8	23	0.2	160.2
po moj	7,282	154.7	0	0.0	155.7
Tuta	13,337	283.4	100	0.8	155.4
Lp	18,579	394.7	195	1.6	151.1
tkati	22,811	484.6	424	3.5	107.4
LP	26,291	558.6	563	4.7	98.6
kW	4,233	89.9	0	0.0	90.9
ops	4,786	101.7	27	0.2	83.9
The	3,802	80.8	31	0.3	65.0
he	2,983	63.4	1	0.0	63.8
cca.	12,641	268.6	389	3.2	63.7
v glaven	2,625	55.8	0	0.0	56.8
hehe	5,275	112.1	123	1.0	55.9
IS	7,220	153.4	218	1.8	54.9
Us	4,624	98.2	100	0.8	54.2
Tebe	4,337	92.1	92	0.8	52.8
bencinar	3,482	74.0	56	0.5	51.2
slikica	4,857	103.2	126	1.0	50.9
oglasnik	3,752	79.7	73	0.6	50.3
šarlatanski	2,490	52.9	10	0.1	49.8
xenon	2,465	52.4	11	0.1	48.9
homologirati	2,313	49.1	7	0.1	47.4
kJ	2,129	45.2	0	0.0	46.2
pozdravljen	30,556	649.2	1,614	13.4	45.2
GET	2,446	52.0	23	0.2	44.5
pasati	6,769	143.8	279	2.3	43.7
x4	2,151	45.7	9	0.1	43.5
homologacija	3,903	82.9	114	0.9	43.1
mašina	6,479	137.7	277	2.3	42.0
amortizer	2,557	54.3	43	0.4	40.8

IJS FILOZOFSKA FAKULTETA

2. JANES – forumi : KRES

Seznam besed

Korpus: JANES v0.4
Podkorpus: Novice

Referenčni korpus: KRES (uravnoteženi)
Spremeni cilj in referenčni (pod)korpus

Stran 1 Pojdi Naslednja >

	JANES v0.4 : Novice		KRES (uravnoteženi)		
lemma	Frekvenca	Frekvenca/mil	Frekvenca	Frekvenca/mil	Rezultat
Slovenc	9,527	444.3	55	0.5	305.7
NSi	1,960	91.4	0	0.0	92.4
JJ	4,107	191.5	211	1.8	70.0
bratuškov	1,472	68.6	0	0.0	69.6
Bratušek	1,643	76.6	15	0.1	69.0
The	1,671	77.9	31	0.3	62.8
www.youtube.com	1,147	53.5	2	0.0	53.6
DeSUS	1,115	52.0	0	0.0	53.0
Us	1,963	91.5	100	0.8	50.6
KPK	1,235	57.6	25	0.2	48.5
AB	981	45.7	0	0.0	46.7
he	973	45.4	1	0.0	46.0
Zoki	962	44.9	6	0.0	43.7
ne vedeti	898	41.9	0	0.0	42.9
Cerar	5,325	248.3	615	5.1	40.8
Juncker	983	45.8	24	0.2	39.1
IS	2,303	107.4	218	1.8	38.6
SMC	987	46.0	32	0.3	37.2
bravo	6,409	298.9	1,036	8.6	31.2
MIRNČAN	634	29.6	0	0.0	30.6
po moj	620	28.9	0	0.0	29.9
ZL	1,071	49.9	89	0.7	29.3
Ukrajina	5,298	247.1	940	7.8	28.2
zapatist	617	28.8	8	0.1	27.9
Patria	1,849	86.2	266	2.2	27.2
kea	853	39.8	70	0.6	25.8
haha	4,804	224.0	934	7.8	25.7
K_ris	527	24.6	0	0.0	25.6
law1	521	24.3	0	0.0	25.3
LOL	517	24.1	0	0.0	25.1
Tuta	948	44.2	100	0.8	24.7
neoliberalen	1,240	57.8	170	1.4	24.4
hehe	1,002	46.7	123	1.0	23.6
BiH	473	22.1	0	0.0	23.1
Dandet	463	21.6	0	0.0	22.6

IJS FILOZOFSKA FAKULTETA

3. JANES – novice : KRES

Seznam besed

Korpus: JANES v0.4
Podkorpus: Wikipedija

Referenčni korpus: KRES (uravnoteženi)
Spremeni cilj in referenčni (pod)korpus

Stran 1 Pojdi Naslednja >

	JANES v0.4 : Wikipedija		KRES (uravnoteženi)		
lemma	Frekvenca	Frekvenca/mil	Frekvenca	Frekvenca/mil	Rezultat
Wikipedija	17,766	3727.1	5	0.0	3579.5
CEST	4,047	849.0	0	0.0	850.0
Wikipedist	3,733	783.1	0	0.0	784.1
Welcome	3,781	793.2	13	0.1	716.8
wikipedija	5,050	1059.4	67	0.6	681.4
Дрбоо	3,226	676.8	0	0.0	677.8
пвоалжтоаь	3,225	676.6	0	0.0	677.6
Willkommen	3,225	676.6	0	0.0	677.6
Welk	3,225	676.6	0	0.0	677.6
Bonvenon	3,225	676.6	0	0.0	677.6
Bem-vindo	3,225	676.6	0	0.0	677.6
Benvenuti	3,226	676.8	5	0.0	650.8
enciklopedičen	3,130	656.6	65	0.5	427.1
wikilettremini	1,962	411.6	0	0.0	412.6
peskovnik	6,472	1357.8	305	2.5	384.7
tilda	2,784	584.1	68	0.6	373.9
Wikipedi	1,620	339.9	2	0.0	335.3
vadnica	2,478	519.9	101	0.8	283.3
ČET	1,235	259.1	0	0.0	260.1
orakelj	2,448	513.6	144	1.2	234.4
vandalizem	2,764	579.9	197	1.6	220.4
ping	1,715	359.8	82	0.7	214.7
WP	1,361	285.5	46	0.4	207.3
GFLY	982	206.0	0	0.0	207.0
blokiranje	1,933	405.5	139	1.2	188.7
netname	885	185.7	0	0.0	186.7
wikipedist	901	189.0	4	0.0	183.9
inetno	866	181.7	0	0.0	182.7
nepristranski	3,606	756.5	382	3.2	181.6
Slovenc	1,175	246.5	55	0.5	169.9
IRC-kanal	790	165.7	0	0.0	166.7
Slovene	1,138	238.7	55	0.5	164.6
Ziga	766	160.7	5	0.0	155.3
IS	2,051	430.3	218	1.8	153.5
navajanje	3,776	792.2	524	4.4	148.2

IJS FILOZOFSKA FAKULTETA

4. JANES – Wikipedija : KRES

Seznam besed

Korpus: JANES v0.4
Podkorpus: Tviti

Referenčni korpus: KRES (uravnoteženi)
Spremeni cilj in referenčni (pod)korpus

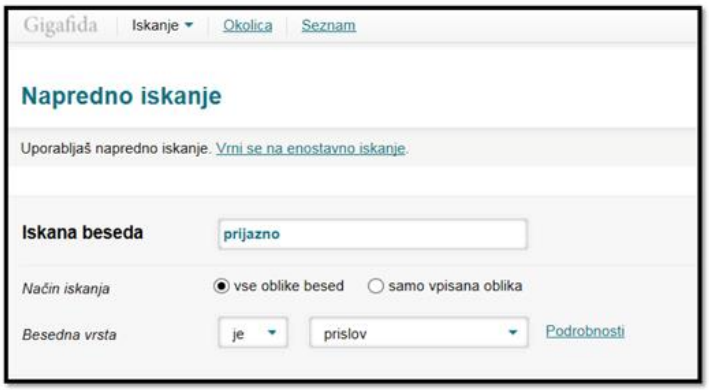
Stran 1 Pojdi Naslednja >

	JANES v0.4 : Tviti		KRES (uravnoteženi)		
lemma	Frekvenca	Frekvenca/mil	Frekvenca	Frekvenca/mil	Rezultat
t.co	2,773,160	25904.5	0	0.0	25905.5
via	106,884	998.4	143	1.2	456.9
The	58,121	542.9	31	0.3	432.6
IS	122,574	1145.0	218	1.8	407.8
posted	44,357	414.3	10	0.1	383.5
Facebook	58,606	547.4	164	1.4	232.2
photo	38,434	359.0	75	0.6	221.9
automatically	20,201	188.7	3	0.0	185.1
he	19,842	185.3	1	0.0	184.8
Slovenc	26,988	252.1	55	0.5	173.8
checked	20,537	191.8	19	0.2	166.6
followed	19,557	182.7	15	0.1	163.3
fora	94,899	886.5	552	4.6	159.0
tvit	16,346	152.7	3	0.0	150.0
OF	129,044	1205.4	918	7.6	139.9
I'm	20,924	195.5	67	0.6	126.2
How	13,071	122.1	3	0.0	120.1
Us	21,945	205.0	100	0.8	112.5
kea	18,576	173.5	70	0.6	110.4
IT	93,191	870.5	875	7.3	105.5
Justa	38,239	357.2	299	2.5	102.9
unfollowed	10,858	101.4	0	0.0	102.4
liked	11,079	103.5	3	0.0	102.0
my	58,862	549.8	553	4.6	98.5
Tuta	17,063	159.4	100	0.8	87.6
What	13,688	127.9	57	0.5	87.5
rt	54,377	507.9	582	4.8	87.3
twitter	12,909	120.6	55	0.5	83.5
It's	12,518	116.9	54	0.4	81.4
photos	9,396	87.8	14	0.1	79.5
day	20,598	192.4	183	1.5	76.8
AB	7,913	73.9	0	0.0	74.9
Haha	7,671	71.7	0	0.0	72.7
good	20,318	189.8	207	1.7	70.2
LOL	7,367	68.8	0	0.0	69.8

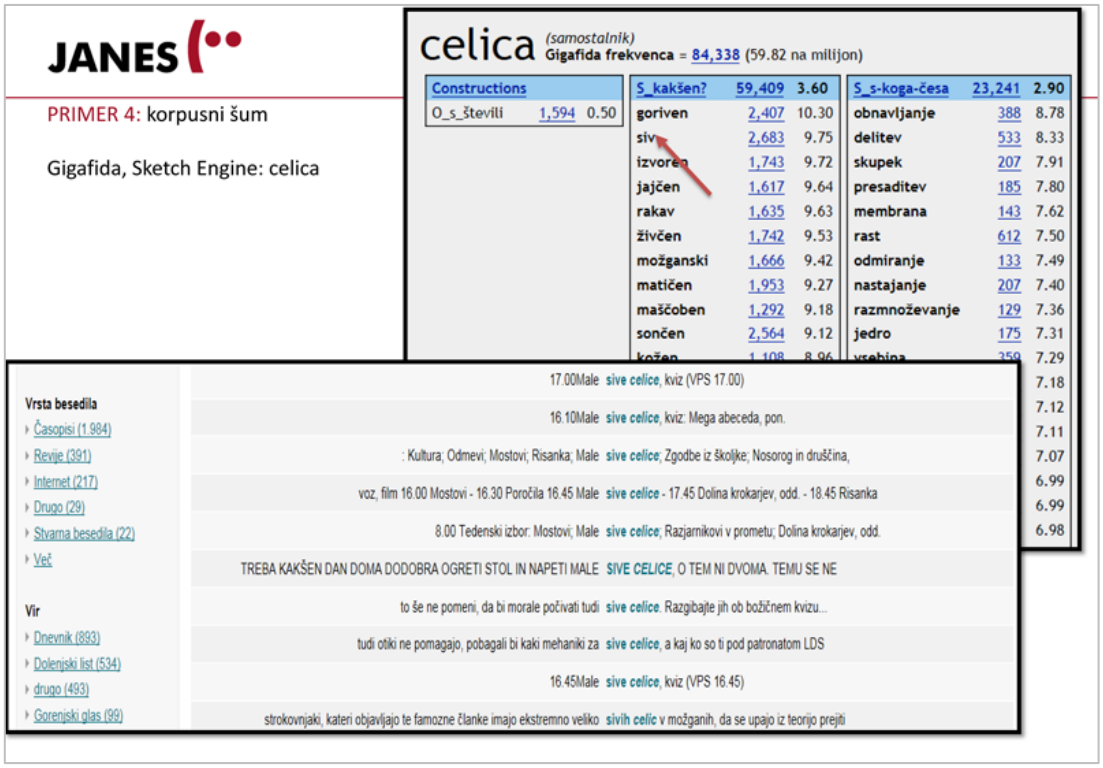
IJS FILOZOFSKA FAKULTETA

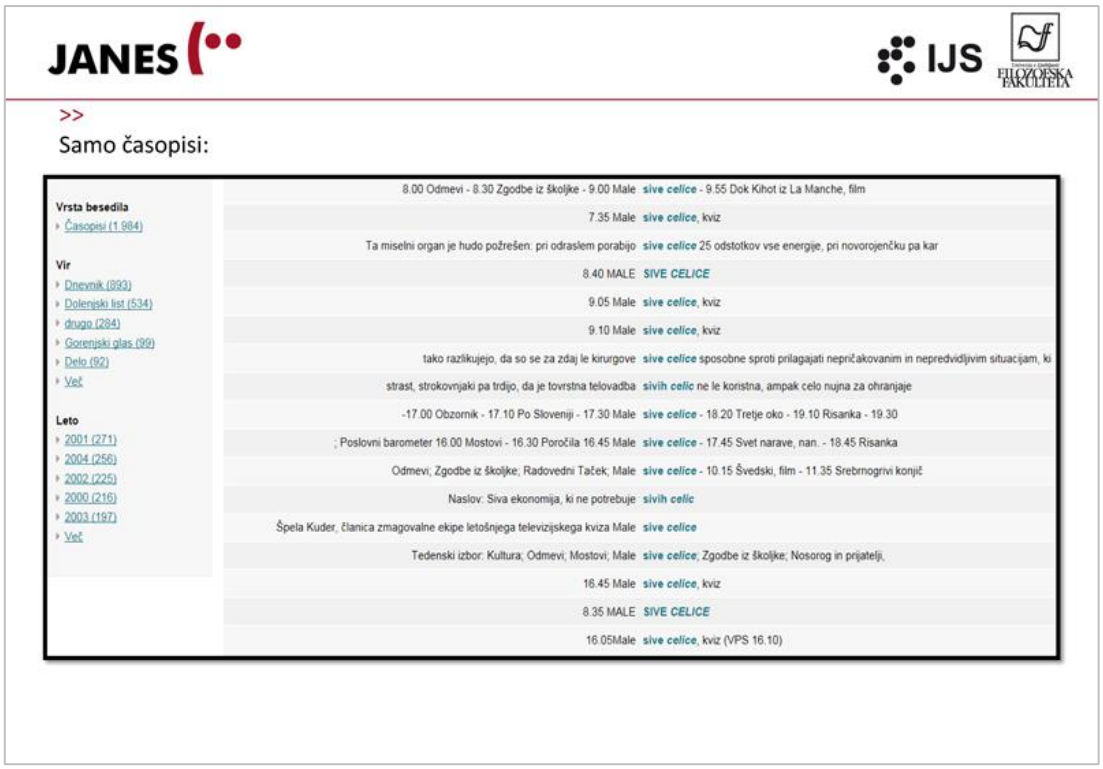
5. JANES – tviti : KRES

PRIMER 3: Kje je razlog, da smo ob pogoju, ki ga kaže prva slika, dobili med rezultati tudi drugo konkordanco na drugi sliki?



Torej pozor: označevanje (različni označevalniki, napake pri označevanju)


JANES
PRIMER 4: korpusni šum
Gigafida, Sketch Engine: celica
celica (samostalnik)
Gigafida frekvenca = 84,338 (59.82 na milijon)
Constructions
O_s_števili 1,594 0.50
S_kakšen? 59,409 3.60
goriven 2,407 10.30
siv 2,683 9.75
izvoren 1,743 9.72
jajčen 1,617 9.64
rakav 1,635 9.63
živčen 1,742 9.53
možganski 1,666 9.42
matičen 1,953 9.27
maščoben 1,292 9.18
sončen 2,564 9.12
S_s-koga-česa 23,241 2.90
obnavljanje 388 8.78
delitev 533 8.33
skupek 207 7.91
presaditev 185 7.80
membrana 143 7.62
rast 612 7.50
odmiranje 133 7.49
nastajanje 207 7.40
razmnoževanje 129 7.36
jedro 175 7.31
7.29
7.18
7.12
7.11
7.07
6.99
6.99
6.98
Vrsta besedila
Časopisi (1.984)
Revije (391)
Internet (217)
Drugo (29)
Stvarna besedila (22)
Več
Vir
Dnevnik (893)
Dolenjski list (534)
drugo (493)
Gorenjski glas (99)
17.00Male sive celice, kviz (VPS 17.00)
16.10Male sive celice, kviz: Mega abeceda, pon.
: Kultura; Odmevi; Mostovi; Risanka; Male sive celice; Zgodbe iz šoljke; Nosorog in družina,
voz, film 16.00 Mostovi - 16.30 Poročila 16.45 Male sive celice - 17.45 Dolina krokarjev, odd. - 18.45 Risanka
8.00 Tedenski izbor: Mostovi; Male sive celice; Razjarnikovi v prometu; Dolina krokarjev, odd.
TREBA KAKŠEN DAN DOMA DODOBRA OGRETI STOL IN NAPETI MALE SIVE CELICE, O TEM NI DVOMA. TEMU SE NE
to še ne pomeni, da bi morale počivati tudi sive celice. Razgibajte jih ob božičnem kvizu...
tudi otiki ne pomagajo, pobagali bi kaki mehaniki za sive celice, a kaj ko so ti pod patronatom LDS
16.45Male sive celice, kviz (VPS 16.45)
strokovnjaki, kateri objavljajo te famozne članke imajo ekstremno veliko sivih celic v možganih, da se upajo iz teorijo prejti


JANES
IJS
>>
Samo časopisi:
Vrsta besedila
Časopisi (1.984)
Vir
Dnevnik (893)
Dolenjski list (534)
drugo (284)
Gorenjski glas (99)
Delo (92)
Več
Leto
2001 (271)
2004 (256)
2002 (225)
2000 (216)
2003 (197)
Več
8.00 Odmevi - 8.30 Zgodbe iz šoljke - 9.00 Male sive celice - 9.55 Dok Kihot iz La Manche, film
7.35 Male sive celice, kviz
Ta miselni organ je hudo požrešen: pri odraslem porabijo sive celice 25 odstotkov vse energije, pri novorojenčku pa kar
8.40 MALE SIVE CELICE
9.05 Male sive celice, kviz
9.10 Male sive celice, kviz
tako razlikujejo, da so se za zdaj le kirurgove sive celice sposobne sproti prilagajati nepričakovanim in nepredvidljivim situacijam, ki
strast, strokovnjaki pa trdijo, da je tovrstna telovadba sivih celic ne le koristna, ampak celo nujna za ohranjaje
-17.00 Obzornik - 17.10 Po Sloveniji - 17.30 Male sive celice - 18.20 Tretje oko - 19.10 Risanka - 19.30
; Poslovni barometer 16.00 Mostovi - 16.30 Poročila 16.45 Male sive celice - 17.45 Svet narave, nan. - 18.45 Risanka
Odmevi, Zgodbe iz šoljke; Radovedni Taček; Male sive celice - 10.15 Švedski, film - 11.35 Srebrnogrivi konjič
Naslov: Siva ekonomija, ki ne potrebuje sivih celic
Špela Kuder, članica zmagovalne ekipe letošnjega televizijskega kviza Male sive celice
Tedenski izbor: Kultura; Odmevi; Mostovi; Male sive celice; Zgodbe iz šoljke; Nosorog in prijatelji,
16.45 Male sive celice, kviz
8.35 MALE SIVE CELICE
16.05Male sive celice, kviz (VPS 16.10)

PRIMER 5: Kaj in kako iščem?

Korošec (1998: 16): poročevalski avtomatizmi, ki so po vsebini sklic na drug vir:

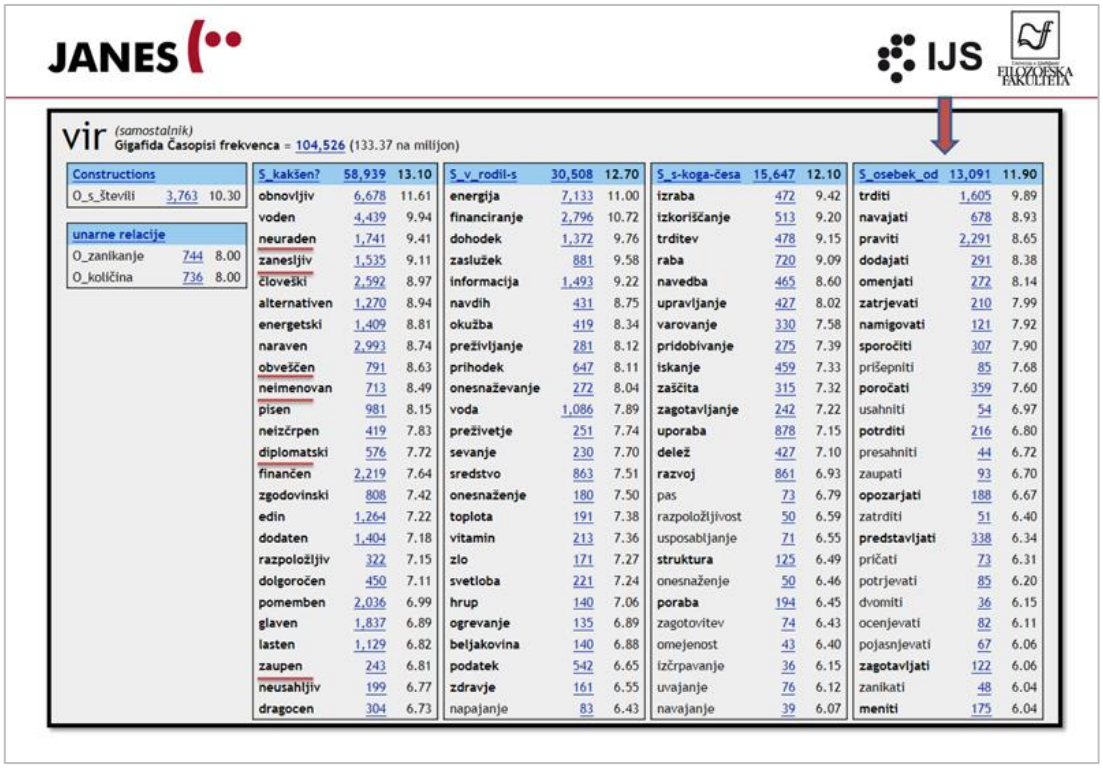
Kot poročajo tuji viri …
Kot poroča agencija …
Iz dobro obveščenih krogov …
Iz krogov, ki so blizu vlade, se je izvedelo …

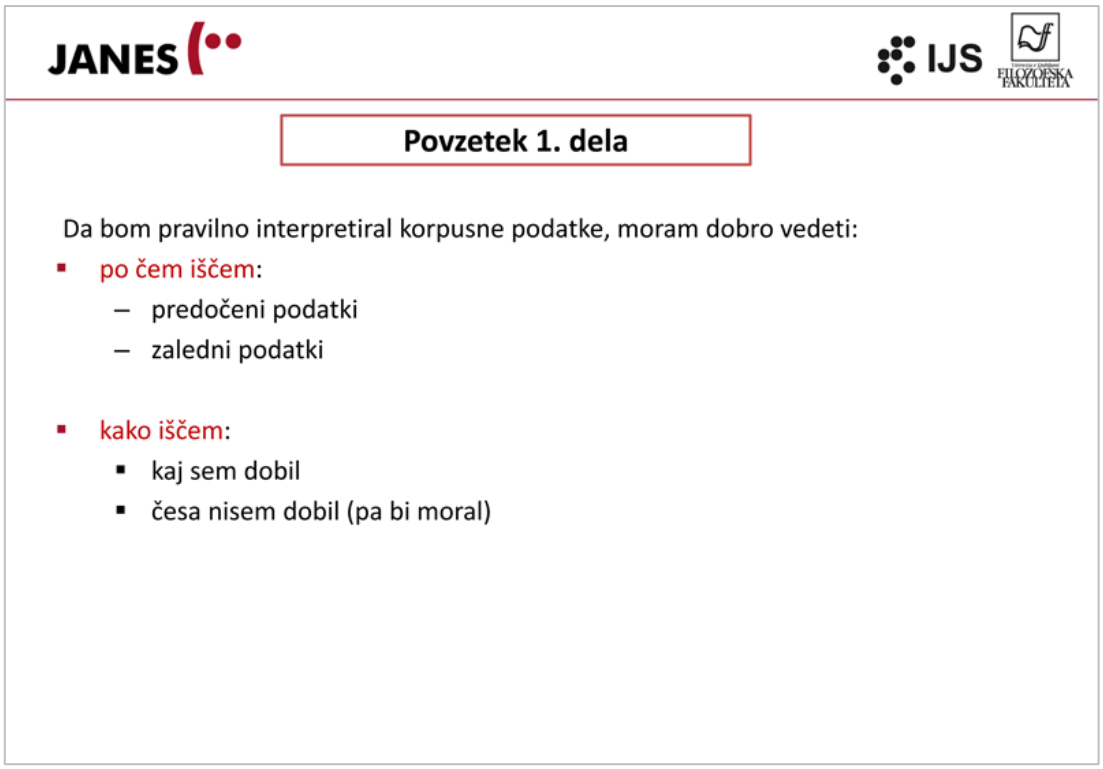
(Več v Logar 2008.)

18 poglobljenih intervjujev z novinarji 11 slovenskih medijev (Logar, Kalin Golob 2015)

> več kot 30 enot, največ z jedrnim samostalnikom *vir:*

viri blizu X	*vir, ki ne želi pred kamero*
naši viri	*vir, ki ne želi biti imenovan*
neuradni viri	*vir, ki mu zaupamo*
uradni viri	*vir, ki ga poznamo*
zanesljivi viri	*iz dobro obveščenih virov*
zelo zanesljivi viri	*od neimenovanega vira*
preverjeni viri	*več virov (nam je potrdilo)*
neimenovani vir	*ime vira je znano v uredništvu*
zaupni vir	*s pomočjo virov*
anonimni vir	


JANES
IJS
FILOZOFSKA FAKULTETA
vir (samostalnik)
Gigafida Časopisi frekvenca = 104,526 (133.37 na milijon)
Constructions
O_s_števili 3,763 10.30
unarne relacije
O_zanikanje 744 8.00
O_količina 736 8.00
S_kakšen? 58,939 13.10
obnovljiv 6,678 11.61
voden 4,439 9.94
neuraden 1,741 9.41
zanesljiv 1,535 9.11
človeški 2,592 8.97
alternativen 1,270 8.94
energetski 1,409 8.81
naraven 2,993 8.74
obveščen 791 8.63
neimenovan 713 8.49
pisen 981 8.15
neizčrpen 419 7.83
diplomatski 576 7.72
finančen 2,219 7.64
zgodovinski 808 7.42
edin 1,264 7.22
dodaten 1,404 7.18
razpoložljiv 322 7.15
dolgoročen 450 7.11
pomemben 2,036 6.99
glaven 1,837 6.89
lasten 1,129 6.82
zaupen 243 6.81
neusahljiv 199 6.77
dragocen 304 6.73
S_v_rodil-s 30,508 12.70
energija 7,133 11.00
financiranje 2,796 10.72
dohodek 1,372 9.76
zaslužek 881 9.58
informacija 1,493 9.22
navdih 431 8.75
okužba 419 8.34
preživljanje 281 8.12
prihodek 647 8.11
onesnaževanje 272 8.04
voda 1,086 7.89
preživetje 251 7.74
sevanje 230 7.70
sredstvo 863 7.51
onesnaženje 180 7.50
toplota 191 7.38
vitamin 213 7.36
zlo 171 7.27
svetloba 221 7.24
hrup 140 7.06
ogrevanje 135 6.89
beljakovina 140 6.88
podatek 542 6.65
zdravje 161 6.55
napajanje 83 6.43
S_s-koga-česa 15,647 12.10
izraba 472 9.42
izkoriščanje 513 9.20
trditev 478 9.15
raba 720 9.09
navedba 465 8.60
upravljanje 427 8.02
varovanje 330 7.58
pridobivanje 275 7.39
iskanje 459 7.33
zaščita 315 7.32
zagotavljanje 242 7.22
uporaba 878 7.15
delež 427 7.10
razvoj 861 6.93
pas 73 6.79
razpoložljivost 50 6.59
usposabljanje 71 6.55
struktura 125 6.49
onesnaženje 50 6.46
poraba 194 6.45
zagotovitev 74 6.43
omejenost 43 6.40
izčrpavanje 36 6.15
uvajanje 76 6.12
navajanje 39 6.07
S_osebek_od 13,091 11.90
trditi 1,605 9.89
navajati 678 8.93
praviti 2,291 8.65
dodajati 291 8.38
omenjati 272 8.14
zatrjevati 210 7.99
namigovati 121 7.92
sporočiti 307 7.90
prišepniti 85 7.68
poročati 359 7.60
usahniti 54 6.97
potrditi 216 6.80
presahniti 44 6.72
zaupati 93 6.70
opozarjati 188 6.67
zatrditi 51 6.40
predstavljati 338 6.34
pričati 73 6.31
potrjevati 85 6.20
dvomiti 36 6.15
ocenjevati 82 6.11
pojasnjevati 67 6.06
zagotavljati 122 6.06
zanikati 48 6.04
meniti 175 6.04


JANES
IJS
FILOZOFSKA FAKULTETA
Povzetek 1. dela
Da bom pravilno interpretiral korpusne podatke, moram dobro vedeti:
po čem iščem:
predočeni podatki
zaledni podatki
kako iščem:
kaj sem dobil
česa nisem dobil (pa bi moral)

Nekaj primerov dobrih sklepanj na celoto

(Vse iz: Slovenščina na spletu in v novih medijih, 2015, ur. D. Fišer.)

1.

Prispevek preverja vrednost novega korpusnega gradiva za normativistične raziskave, in sicer z analizo pogostosti in zapisovanja zvez samostalnika z nesklonljivim levim prilastkom (*solo petje, RTV prispevek*) v korpusih Janes in Kres.

Gradivo korpusa Janes razkrije jasnejše trende zapisovanja tovrstnih zvez narazen, v primerjavi s korpusom Kres, kjer je število zvez znatno nižje, trendi v zapisu pa so bolj heterogeni. /…/ Na drugi strani izbrana metodologija razgali težave, ki jih pri luščenju podatkov povzroči neenotnost avtomatskega označevanja, in s tem pokaže na zadrege s kategorizacijo nesklonljivih levih prilastkov, ki se ob samostalniku pojavljajo.

(Arhar Holdt, Dobrovoljc 2015: 4.)

2.

Iz korpusne analize je razvidno, da je tako pri uradnih kot zasebnih računih močan trend zapisovanja imen industrijskih izdelkov z veliko začetnico. Z izjemo kategorije mobilne telefonije je pri uradnih računih povsod odstotek zapisa z veliko začetnico vsaj 70-odstoten, pri zasebnih pa vsaj 54-odstoten.

Med vsemi dvajsetimi obravnavanimi lemami je izstopala le lema *twitter*, ki jo zasebni uporabniki pogosteje pišejo z malo začetnico (54,9 %), medtem ko v zapisih uradnih računov še vedno prevladuje raba velike začetnice (76,0 %). Glede na to, da sicer splošen trend zapisovanja imen industrijskih izdelkov z veliko začetnico odstotkovno izstopa predvsem pri uradnih računih, lahko sklepamo, da proizvajalci in podjetja stremijo k doslednemu zapisu takih izdelkov z veliko začetnico oziroma z morebitnimi izvirnimi zapisi tipa *iPhone*.

(Goli, Popič, Fišer 2015: 32.)

3.

V preučevanih spletnih besedilih, zlasti v tvitih in komentarjih, se pisci pogosto izražajo na način, ki želi biti drugačen, učinkovit in odmeven. Ena od možnosti, kako to doseči pri hitrem, impulzivnem načinu pisne komunikacije /.../, je uporaba frazemov. /.../ Učinek, kot kažejo analizirani primeri, želi biti predvsem humoren, prenovitveni postopki pa so vezani na izrabljanje pomenskih lastnosti frazemov /.../ in vpletanjem aktualnih družbenih in političnih pa tudi osebnih dogodkov.

Zgledi primerjave med korpusoma Janes in Kres potrjujejo predvidevanje, da je raba frazemov tako v izhodiščnem pomenu kot v prenovitvah za spletna besedila zelo značilna, posledično pa je mogoče sklepati tudi na specifične lastnosti spletne komunikacije (hitra odzivnost, drugačnost, učinkovitost, neformalnost), ki se tudi sicer kažejo na vseh ravneh njenega jezikovnega opisa.

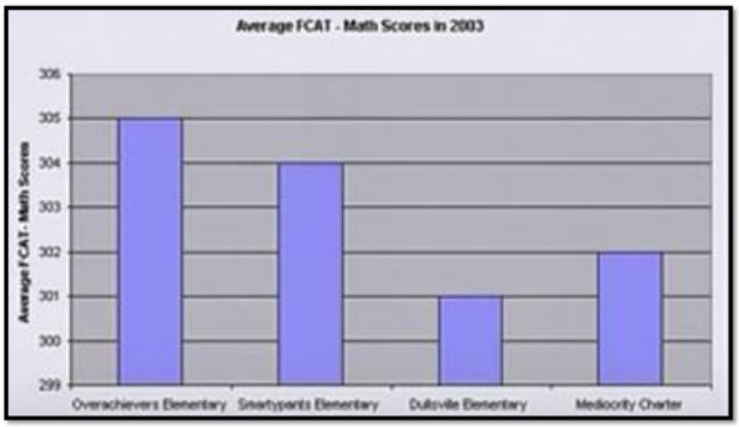
(Justin, Hirci, Gantar 2015: 36–37.)

2. Korpus = veliko številskih podatkov >> vizualizacija

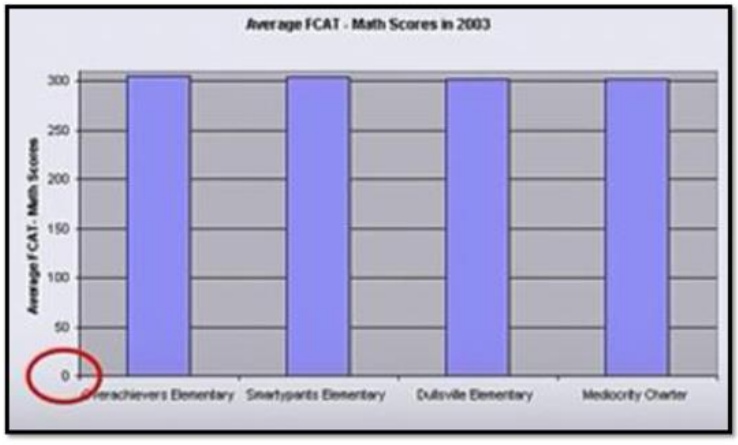
1.

Exposed: How Fox News Lies With Statistics, 29. 11. 2012
https://www.youtube.com/watch?v=w7EvBxRYNME

2. Kaj je narobe z naslednjo predstavitvijo podatkov?
(Vir: Konst 2011.)

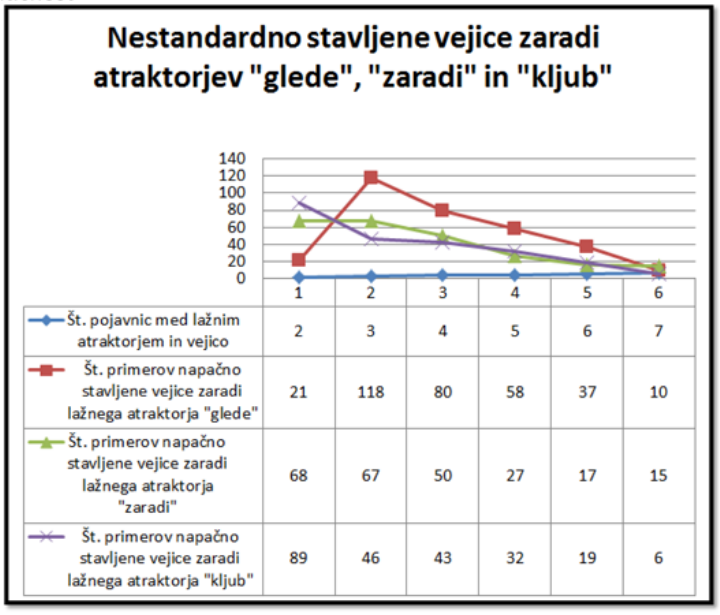


>>



Pogovor ob primerih

PRIMER 1: V čem vse je problematičnost naslednje slike
/op.: primer je avtentičen/?

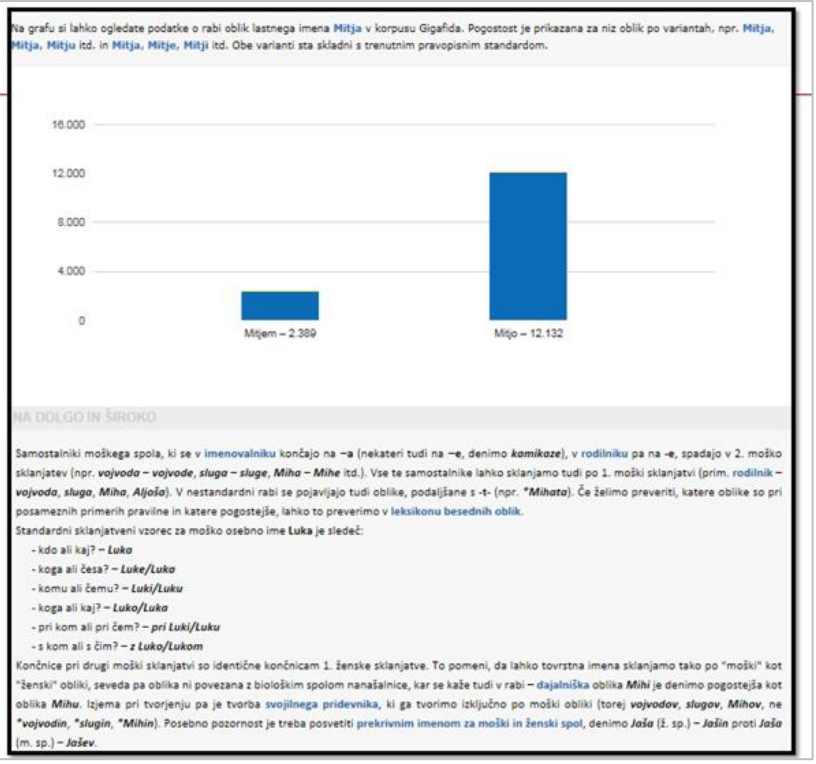


	1	2	3	4	5	6
Št. pojavnic med lažnim atraktorjem in vejico	2	3	4	5	6	7
Št. primerov napačno stavljene vejice zaradi lažnega atraktorja "glede"	21	118	80	58	37	10
Št. primerov napačno stavljene vejice zaradi lažnega atraktorja "zaradi"	68	67	50	27	17	15
Št. primerov napačno stavljene vejice zaradi lažnega atraktorja "kljub"	89	46	43	32	19	6

PRIMER 2: S pomočjo naslednje slike odgovorite na vprašanje: Pišete uradno vabilo. Katero končnico pri imenu Mitja v orodniku ednine bi uporabili?

Kateri podatek na sliki je na odločitev najbolj vplival? Kako hitro ste prišli do odločitve?

(Vir: Slogovni priročnik.)



Povzetek 2. dela

Ko vizualiziram korpusne podatke, moram:

- upoštevati pravila in dobre prakse statistikov
- razumeti moč vizualiziranih podatkov

SKLEP

- Ne samo da se korpusne podatke da interpretirati, ampak ne gre drugače, kot da jih interpretiramo – interpretacija (preprosta ali kompleksna) je sestavni del vsake uspele korpusne poizvedbe.
- Dolžnost sestavljalcev korpusa je priprava opisov korpusne zgradbe in prestavitev odločitev v zvezi z njo; dolžnost uporabnikov korpusa pa je, da te opise in predstavitve poznajo. Samo tako bodo lahko rezultate svojih poizvedb interpretirali na pravilen način.

Literatura

Logar, Nataša (2008): Poročevalstvo kot del jezika v medijih: kot poroča naša dopisnica, v domovini ni rednega dovoda transijade. V: PEZDIRC BARTOL, Mateja (ur.): Slovenski jezik, literatura, kultura in mediji: zbornik predavanj. Ljubljana: Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete. 157–163. Dostopno prek: http://centerslo.si/wp-content/uploads/2015/10/ssjlk_44_zbornik.pdf.

Logar, Nataša (2015): Gradnja referenčnih korpusov na novo: nadgradnja Gigafide. V: GORJANC, Vojko, idr. (ur.): Slovar sodobne slovenščine: problemi in rešitve. Ljubljana: Znanstvena založba Filozofske fakultete. 218–240.

Logar, Nataša, Dobrovoljc, Kaja, Arhar Holdt, Špela (2015): Gigafida: interpretacija korpusnih podatkov. V: SMOLEJ, Mojca (ur.): Slovnica in slovar - aktualni jezikovni opis (Obdobja 34). Ljubljana: Znanstvena založba Filozofske fakultete. 467–477. Dostopno prek: http://www.centerslo.net/files/file/simpozij/simp34/zbornik%202/Logar-Dob-Arh-Hol.pdf.

VIRI:

- Exposed: How Fox News Lies With Statistics (29. 11. 2012). Dostopno prek: https://www.youtube.com/watch?v=w7EvBxRYNME (26. 6. 2016).
- Izvidnica (25. 12. 2015). Val 202. Dostopno prek: http://val202.rtvslo.si/2015/12/izvidnica-37/ (26. 6. 2016).
- Jogi, Usha (2014): Interpretation of data and statistical falacies. Research hub 1/3. Dostopno prek: http://www.slideshare.net/rhimrj/interpretation-of-data-and-statistical-fallacies (25. 6. 2016).
- Kalin Golob, Monika, Logar, Nataša (2014): Prostor v poročevalskem skupnem sporočanjskem krogu. Slavistična revija 62/3. 363–373.
- Korpus Gigafida. Dostopno prek: http://www.gigafida.net (24. 6. 2016).
- Korpus JANES 0.4. Dostopno prek: http://nl.ijs.si/janes/ (27. 6. 2016).
- Konst, Bill (2011): Identifying Misleading Graphs – Konst Math. Dostopno prek: https://www.youtube.com/watch?v=ETbc8GIhfHo (25. 5. 2016).
- Logar, Nataša, Kalin Golob, Monika (2015): Jezikovne izbire pri upovedovanju zaupnih virov informacij: iz zgodovine v sodobnost. Teorija in praksa 52/4. 651–669.
- Slogovni priročnik. Dostopno prek: http://slogovni.slovenscina.eu/ (25. 6. 2016).