

Polona Gantar

10 KORPUSNA FRAZELOGIJA

10.1 Korpusna frazeologija – izročki

JANES

IJS FILOZOFSKA
FAKULTETA

Korpusna frazeologija

Povej mi s kom se družiš in povem ti ...

Polona Gantar
apolonija.gantar@ff.uni-lj.si

Poletna šola spletne slovenščine za študente
Ljubljana, 4. do 8. julij 2016

JANES

Vsebina

IJS FILOZOFSKA
FAKULTETA

- **Korpusno jezikoslovje**
 - O korpusih in korpusnem jezikoslovju
 - O možnosti izrabe korpusov v leksikologiji in frazeologiji
- **Frazeologija kot jezikoslovna veda**
 - osnovni pojmi frazeološkosti
 - večbesednost
 - ustaljenost
 - pomenska (ne)prozornost
 - jezikovnosistemske »nepravilnosti«
 - ekspresivnost, metaforičnost
 - raziskava: frazeološke prenovitve v spletni slovenščini
 - zamejitev področja
 - frazeologija v ožjem in širšem smislu
- **Korpusna frazeologija**
 - leksikalna gramatika
 - sintagmatsko-paradigmatska razmerja (kolokacije)
 - Projekt Corpus pattern Analysis
 - orodja za analizo večbesednih enot v korpusu
 - Besedne skice
 - skladijski in stavčni vzorci (LBS)
 - Kolokacijski slovar slovenskega jezika
 - Parseme shared task
 - Raziskovalna naloga: Korpusna analiza fraznih glagolov

KORPUSI IN KORPUSNO JEZIKOSLOVJE

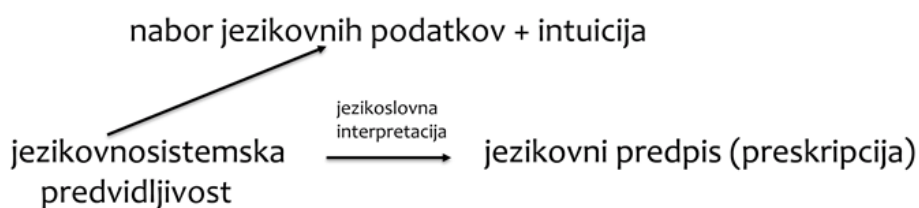
Razvoj korpusov

- **Korpusi v preteklosti**
 - zbirke jezikovnih podatkov
- **Korpusi danes**
 - obsežne elektronske zbirke jezikovnih podatkov
 - računalniška berljivost besedil: možnost procesiranja jezikovnih podatkov
 - **standariziran zapis:** označevanje korpusnih dokumentov; podatki o besedilu ločeni od besedila samega; primerljivost in tipologizacija (TEI – Text Encoding initiative)
 - **označenost:**
 - zapis korpusa: dokumentiranost besedil
 - pripisovanje jezikoslovnih oznak: strukturna in jezikoslovna označenost

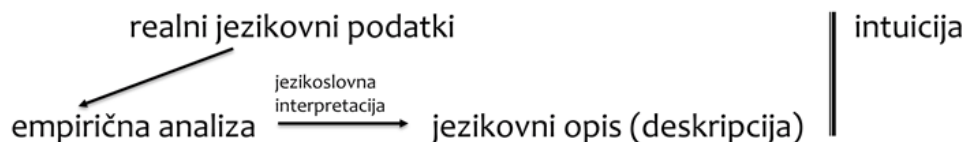
Označenost korpusov

- **strukturno označevanje**
 - tokenizacija osnovne korpusne enote – pojavnice (presledek)
 - segmentacija segmenti, npr. povedi (končno ločilo)
- **jezikoslovno označevanje** jezikoslovna interpretacija
 - **lematizacija**
 - list - sam. m. sp. (list-a, list-u, list-Ø; list-om; list-i, list-ov, list-e, list-om ...)
 - lista – sam. ž. sp. (list-e, list-i, list-o; list-e; list-Ø, list-am)
 - **oblikoskladenjsko označevanje** sobesedilno pogojene morfološke oznake

Pred- (ne)korpusno jezikoslovje



Korpusno jezikoslovje



intuicija

- intuicija pri pridobivanju jezikovnih podatkov pove več o tem, kako deluje intuicija, kot o jeziku samem
 - **samospraševanje** (Kaj pomeni ... ? Ali lahko rečem ...? Tega nikoli ne bi rekel ... Koklikokrat rečem?)
 - **nesprejemljivi : sprejemljivi stavki**
- intuicija rojenih govorcev je nepogrešljiva pri jezikoslovni interpretaciji podatkov

empirična analiza

Empirična analiza jezika, kot se kaže v *dejanski rabi* (empirijska – izkustvo), pomeni registriranje realnega stanja in ne zanašanje na intuitivno vedenje o jeziku.

“ ... če želimo o jezikoslovju govoriti kot o znanosti, potem morajo biti raziskave omejene izključno na to, kar je mogoče v jeziku opazovati ...”

Bloomfield (30. Leta 20. stol.) revija *Language*

Možnosti izrabe korpusnih podatkov

▪ Jezikoslovne raziskave

- leksikologija
- frazeologija
- leksikografija (jezikovni opis; priročniki)

- **oblikoslovje:** ty, tenkju, tenks, tnx, thanks, thnx, tenx ; tviterašiti, tweeterašiti, twitterašiti ...
- **skladnja:** alfa samec; RTV prispevek
- **pomenoslovje:**
 - „I don't believe in word sense“ (Kilgarriff 2009)
 - „, you shall know the word by the company it keeps“ (Firth 1975)

JANES Interdisciplinarne raziskave

▪ Korpusi in jezikovnotehnoške raziskave

- jezikovno rudarjenje
- strojno prevajanje
- analiza sentimenta
- prepoznavanje imenskih enitet
- skladijsko označevanje
- semantično označevanje
- označevanje VLE

- **Sociološke raziskave** brexit – greexit... kdor ne skače ni Slovenc

- **Kulturološke raziskave** kaj je tipično slovensko ...

FRAZELOGIJA KOT JEZIKOSLOVNA VEDA

Frazeološke enote so večbesedne strukturno in pomensko ustaljene enote, ki glede na svoje sestavne elemente izkazujejo samostojni, od sestavnih elementov neodvisni pomen. V besedilih se uporabljajo predvsem za izražanje negativnega ali pozitivnega vrednotenja in so pogosto glede na nefrazeološke ustreznike stilno zaznamovane.

- večbesednost
- strukturna in pomenska ustaljenost
- pomenska neprozornost
- jezikovnosistemske „nepravilnosti“
- ekspresivnost in metaforičnost

JANES ena ali več besed?

tokenizacija in
lematizacija!

- Alfa romeo, alma mater
- al dente, at hoc, a priori
- d. o. o., omg! espe
- narodnozabavni, narodno-zabavni, narodno zabavni
- naprimer – na primer; navsezadnje – na vse zadnje
- bati se, smejati se, svitati se ...

- računati na koga : računati na pamet
- priti do denarja : priti do razglednega mesta
- cukati ga : cukati ga za rokav

JANES ena ali več besed?

- d. o. o., omg! espe
- al dente, at hoc, a priori
- Alfa romeo
- narodnozabavni, narodno-zabavni, narodno zabavni
- naprimer – na primer; navsezadnje – na vse zadnje
- bati se, smejati se, svitati se ...

- **računati na** koga : računati **na pamet**
- **priti do** denarja : priti **do** razglednega mesta
- **cukati ga** : cukati **ga** za rokav

frazni glagol

prosti glagolski morfem

zaimék

predlog: za izražanje meje v
prostoru do katere sega dejanje

JANES ena ali več besed - Janes



▪ frazni glagoli

- up loudati - uploudati, daunoloudati
- up date - updatati, updejtati

trenutno se **uploada** ampak bo še trajalo varno, da si vse slike **up loadal** na FaceBook

▪ izrazi, nastali po frazeološki poti

- piš me v uh - pišmevuh; pišmevuhovstvo
- majke mi - majkemi, jebi ga - jebiga
- je bela cesta - jebela cesta - jebelacesta
- ne bodi ga treba - ne-bodi-ga-treba - nebodigatreba

Majke mi, toti ni čist diht...
@merineseri midve sva v žlahti, **majkemi**

▪ drugo

- omb - o moj bog
- lp - lep pozdrav
- smsjati - sms-jati
- konec koncev - koneckoncev, kratko malo - kratkomalo
- stand up - standup; on line - on-line - online

O MOJ BOG kaki tviti
OMG, kir popizditis dobivam

JANES ena ali več besed - Janes



da danes zopet ne bo nič, odstraniš make	up	, daš v pižamo in skriješ pod kovter.
Najbolje pa je, ce to storis sam. Heads	up	. ¶ V Bauhausu so me gledali kot največjega
Čakam, da požegnajo, da lahko grem. Thumbs	up	za šefa, ki mi je prinesel svež rogljiček
zanimivo, omenja tudi Jacksona oz. njegovo back	up	vokalistko . ¶ La femme qui se poudre - čudovit
za novo leto ". Letni časi so res fucked	up	. ¶ Če imaš vrane v akustičnem dvorišču
če si muzikaličen te zna odpeljat daleč. Up	to you . ¶	. @tjablonsky + Bojan Adamič.
pretiravali @GobaFunk ¶ @strupko sicer pa, shut	up	! :) končno najdem podobnomislečo punco...
letih. Žalostno, da je moral priti push	up	od drugod. ¶ @BKovacec @EvaKobe24ur
je bila totalna čudovitka, hvala za stand	up	comedy šov. #soocenjeLJ ¶ Kdo je stilistka
uporabil samo VIP ¶ Današnji sanjski line	up	bi bil: Dead Boys, Dead Kennedys in kot
pa politični noži v hrbtih... Luka, grow	up	! ¶ @petrasovdat V pričakovanju izjave, da
ne bi vedla:) ¶ Glavno, da imam še perfect	mejkap	, ko v fejk kroksicah nesem smeti ven
http://t.co/SB6VzUFJTt ¶ @Pawla40 ej sam en mejhn	heads up	, kle ni lajkov tko k na FBju! #welcome ¶
odstopil, pa ga žal nima! ¶ DROLC go home and	šadap	! ¶ Hehe tipa ko sem videl na odmevih sem
@BanicGregor ¶ A majo ta tedn vremenarji kšn	Standup	comedy slučajno? Vrjetno bo zvečer bl
http://t.co/UfWN1M2tpT ¶ Si že videl tale delček	lineupa	za Tomorrowland? Nooro kajne? ! : D Kdo

Izrazna in pomenska ustaljenost?

- **Kurja koža** ol ouver! Rade #Šerbedžija je Zakon!
- Omojboh... 30 sekund #Sanremo2014 pa že **kurja koža**, solzne oči...
- včeraj sem prešlatala celo zimsko kolekcijo. **kurja koža** v trenutku!
- 3cm veliko **kurjo kožo** dobim že pri trailerju, kaj šele bo.
- pa ko čuješ zvok od stare impreze ti kr **kurja koža** rata
- omg, meni kar **kurja koža** postane, ko vidim ono fotografijo
- bemu sunac, **kurja koža**, pa napete bradavičke:)
- kadarkoli se ga spomnim, me oblije **kurja polt**
- Zakaj se mi je **kurja polt** naredila, ko sem prvič poslušala Luko Mesca.
- Rusom in Kitajcem se počasi **ježi koža**, dolgo ne bodo več prenašali psihopatov.
- Ko slišim za besedo strokovnjak, dobim **kurjo kožo**.
- Na čelu države bo Gospod Janez Janša, rdeči tk kr **kurjo kožo** dobite ko ga omenim

poz.

neg.

- **variantnost sestavin**
 - Nizozemci so pravi trn v **peti** Slovencev.
 - Ja res, ful so koristni ... kot **trn v riti** ...
 - Me zanima, zakaj jim je ravno Mišič tak trn v **očesu**.
- **vrivanje besed**
 - trn v **poslanski peti**
 - **Kosovski** trn v **črnogorski peti**
 - zasebniki v lekarnstvu so **hud** trn v peti javnim lekarnam
- **oblikovne in skladienske spremembe**
 - Slaba banka je trn v peti »oldboysom«
 - Alenka je trn v **Cerarjevi peti**
- **variantnost v besedilni realizaciji**
 - kadar se ji **zapiči v peto kak trn resnice**, ...
 - Flirtanje je lahko nedolžen klepet, včasih pa postane **igra z ognjem**.
 - Ne bom dovolil, da spet zaidemo v igračkanje z **ognjem**.
 - vseeno ne pozabi biti previden: z **ognjem** se ni dobro igrati!

JANES Jezikovnosistemske „nepravilnosti“ IJS



- **oblikoslovje**
- **besedna oblika**
Je skoda usakga twita. Kr nej bodo na ignor
Čakal sem te ko kreten, ti pa me daš na ignore.
- **besedna vrsta**
KrNeki test 😊
prej sem kr neki fejku z krnekimi copati pa so me vcasih podplati dlje casa bolel
- **skladnja**
- **nepričakovan skladenjski vzorec**
Dvorni norci tudi iz sebe norce brijejo
gre za (kaj)
- **pretvorbena omejenost**
*zdrava pamet *pamet je zdrava*
*star maček *maček je star*
- besedilo**
- **elipsa**
mi o volku, kozel iz zelja
človek človeku volk
- **referenčnost**
delati iz muhe slona
tresla se je gora, rodila se je miš

JANES



Pomenska (ne)prozornost?

• star

- star model avta
- stara mama
- staro mestno jedro
- stari običaji
- star papir
- star prdec
- stara celina
- stara zaveza



▪ črn

- črna barva
- črn oblak
- Črno morje
- črna kava
- črn kruh
- črna kronika
- črna gradnja
- črna vdova



JANES Pomenska (ne)prozornost?

stara mama
dnevna soba
spalna srajca
prekmurska gibanica
pametni telefon

?

SLO	CZ	NEM	ANG
čez čas/sčasoma	po čase/začas	im Laufe der Zeit	in time
v hipu/v trenutku/kot bi trenil/takoj	na první pohled	im Augenblick	in (less than) no time
skrajni čas	nejvyšší čas	höchste Zeit	high time/about time

JANES

Ekspresivnost, metaforičnost, jezikovna ustvarjalnost

FE kot kolektivne metafore – frazeološki pomen

Nenevtralnno izražanje

spraviti v red – urediti
mazati si roke z (drekom) – ukvarjati se s čim neprijetnim ali kaznivim
delati iz muhe slona – pretiravati

Iz tega ne more nastati nič dobrega, kajti zaradi **umiritve razmer** raje **prikrijejo dejstva**. In v celi državi ni **pogumnega politika**, ki bi dovolj **odločno** problematiziral takšno početje. To je posmeh vsem, ki so zaradi vladnih ukrepov prisiljeni **povsod** varčevati. Tudi sindikati in upokojenci **nočejo priznati resnice**, ne vidijo **dalnosežnosti problema**. Njihova **zmaga bo imela negativne posledice** za celotno Slovenijo. /.../ Skoraj neopazno je bil odstavljen prvi nadzornik HSE, eden redkih, ki **je upal opozoriti na nepravilnosti**.

Iz tega ne more nastati nič dobrega, kajti zaradi **ljubega miru** raje **pometejo smeti pod preprogo**. In v celi državi ni politika **z jajci**, ki bi **udaril po mizi** in problematiziral takšno početje. To je posmeh vsem, ki so zaradi vladnih ukrepov prisiljeni varčevati **na vsakem koraku**. Tudi sindikati in upokojenci **tiščijo glavo v pesek**, ne vidijo dlje **od svojega nosu**. **Ta zmaga bo Pirova**, poražena pa bo cela Slovenija. /.../ Skoraj neopazno je bil odstavljen prvi nadzornik HSE, eden redkih, ki je **dregnil v sršenje gnezdo**.

FE kot kolektivne metafore – frazeološki pomen

Nenevtralnno izražanje

Opis splošno človeških lastnosti, narodovih značilnosti, družbenih pojavov (narodovi stereotipi)

Sekundarna metaforizacija – FE kot povod za **ustvarjalno in duhovito rabo**.



six pack



one pack = [vampek]

spraviti v red – urediti
mazati si roke z (drekom) – ukvarjati se s čim neprijetnim ali kaznivim
delati iz muhe slona – pretiravati

krvav pod kožo
naj sosedu krava crkne
kdor ne skače ni Slovenc

ljubi svojega bližnjega ...
... samo pazi, da te ne dobijo

zrno do zrna ...

gluten

kamen na kamen ...

nepremičninski davek

JANES Frazeološke prenovitve



- **Konferenca Slovenščina na spletu in v novih medijih**
- Justin, Hirci in Gantar, 2015: Rana ura, slovenskih fantov grob: analiza frazeoloških prenovitev v spletni slovenščini <http://nl.ijs.si/janes/wp-content/uploads/2015/11/JANES15-06-Rana-ura.pdf>
- **Teoretična izhodišča raziskave** (Kržišnik 1990 idr.)
- **frazeološke enote**
 - raba FE je v vsakem besedilu **opazna**
 - višja pogostnost FE je lahko znak **bogatejšega in pomensko bolj razplastenega jezika** (Kržišnik 1990)
- **frazeološke prenovitve**
 - frazeološke prenovitve kot hotene in namerne inovacijske spremembe; niso systemske in so zato enkratne
 - spremenijo obliko in/ali modificirajo pomen
 - učinkovita prenovitev mora vzpostavljati **razvidno razmerje do pomena izhodiščne FE**, hkrati pa prinašati **nove pomenske razsežnosti**.
- **računalniško posredovana komunikacija**
 - nekorrigirana sporočila
 - specifični žanri
 - elementi nestandardnosti, govornosti; inovativnost, individualnost, učinkovitost, opaznost ...

JANES



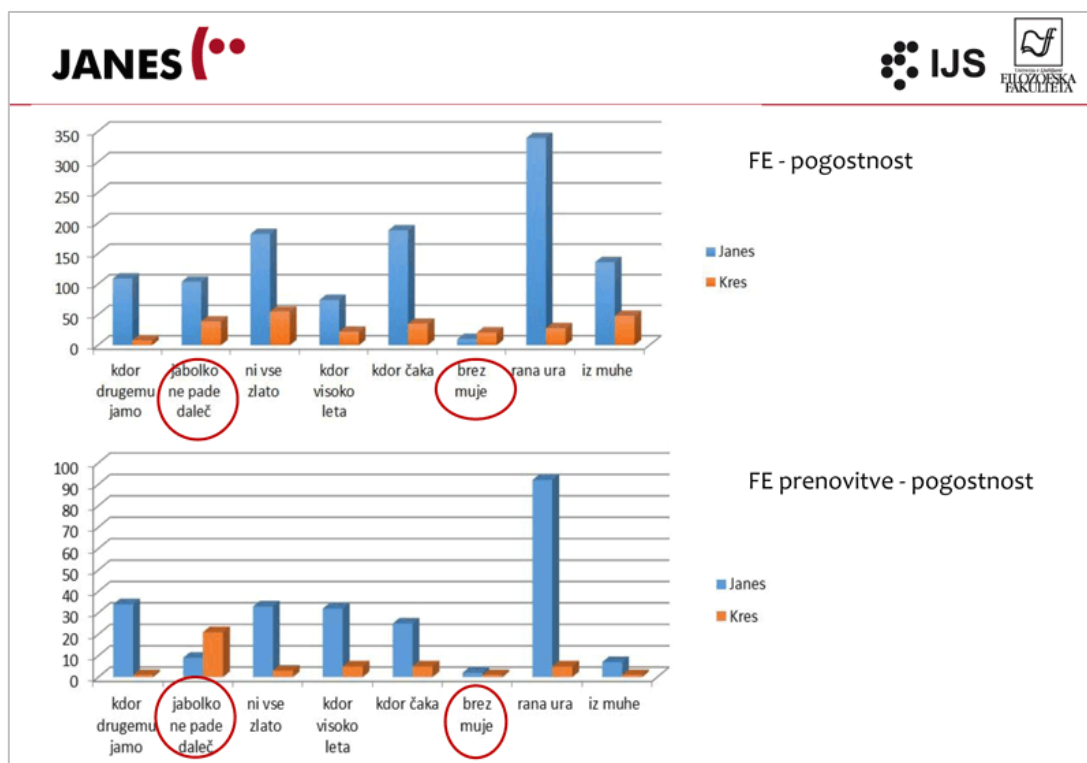
- **Namen raziskave**
 - ugotoviti zastopanost frazeoloških enot v spletno specifičnih žanrih
 - delež frazeoloških prenovitev
 - značilnosti ustvarjalne jezikovne rabe
- **Korpus**
 - izhajali smo iz podkorpusa Janes v0.3 Tweet (Fišer et al. 2014)
 - Za primerjavo smo uporabili korpus Kres (Logar et al. 2012); izločili smo internetna besedila.
- **Orodja in metode**
 - Sketch Engine (Kilgarriff et al. 2004), in sicer funkcije
 - iskanje v konkordančniku
 - urejanje besedilnega okolja in filtriranje besedil
 - besedne skice in
 - frekvenčni prikaz obravnavane FE v določenem besedilnem tipu

Obravnavane FE

- *kdor drugemu jamo koplje, sam vanjo pade*
- *jabolko ne pade daleč od drevesa*
- *ni vse zlato, kar se sveti*
- *kdor visoko leta, nizko pade*
- *kdor čaka, dočaka*
- *brez muje se še čevelj ne obuže*
- *rana ura, zlata ura*
- *iz muhe delati slona*

Kvantitativna analiza

Iskalni pogoj	Korpus Janes (161 mio besed)				Korpus Kres (120,5 mio besed)			
	vse pojavitve	Izho diščne FE	prenovitve	nejasno /dobesedno	vse pojavitve	izhodiščne FE	prenovitve	nejasno/dobesedno
<i>kdor drugemu jamo</i>	109	59	34	12	8 (3)	7	1	0
<i>jabolko ne pade daleč</i>	104	91	9	4	39 (2)	18	21	0
<i>ni vse zlato</i>	182	146	33	3	55 (13)	48	3 (1)	4 (1)
<i>kdor visoko leta</i>	74	38	32	4	22 (3)	17	5	0
<i>kdor čaka</i>	188	155	25	10	35 (5)	27	5	3 (1)
<i>brez muje</i>	10	3	2	5	21 (0)	16	1	4
<i>rana ura</i>	339	215	92	32	28 (5)	17	5 (2)	6
<i>iz muhe</i>	136	128	7	1	48 (11)	45	1	2
skupaj	1142	835	234	71	256 (42)	195	42 (3)	19 (2)



JANES Kvalitativna analiza

- **ponovna aktualizacija (dobesedni pomen)**
 - *kdor drugemu jamo koplje, je prvega že zasul*
 - *kdor drugemu jamo koplje, je grobar*
 - *kdor visoko leta, je pilot/ptič/stevardesa*
 - *kdor visoko leta, daleč vidi*
 - *kdor čaka, ima dobre živce*
 - *kdor čaka, postane star*
 - *rana ura, zaspana ura*
 - *rana ura, rano dan zjebe*

- **demetaforizacija -> ponovna vzpostavitev metaforičnega premika**
 - *iz muhe delati Airbus/Dumbota*
 - *jabolko ne pade daleč od hruške*
 - *brez muje se še papuč ne sezuje*
 - *ni vse zlato, kar je nafta (nafta = črno zlato)*
 - *žena ne pade daleč od drevesa*

- **križanje FE**
 - kdor drugemu jamo koplje, je sam svoje sreče kovač
 - ni vse zlato, kar se Janezek nauči
 - kdor visoko leta, se najslajše smeje
 - kdor visoko leta, ne pade daleč od drevesa
 - rana ura, slovenskih fantov grob

- **družbena aktualizacija**
 - kdor drugemu jamo koplje, vlada vanjo pade
 - kdor drugemu jamo koplje... je jutri vabljen v Rovte
 - ni vse zlato, kar je Žižek
 - kdor visoko leta, je Peter Prevc
 - rana ura, zlati Rolex

DA	NE
<ul style="list-style-type: none">• večbesednost<ul style="list-style-type: none">✓ popihati jo? priti do (česa)	smejati se
<ul style="list-style-type: none">• pomenska neprozornost<ul style="list-style-type: none">✓ star maček	stara mama pametni telefon potni list
<ul style="list-style-type: none">• ekspresivnost, metaforičnost<ul style="list-style-type: none">✓ (imeti) dolg jezik	taščin jezik* črna luknja*

frazeologija v ožjem smislu vključuje večbesedne enote, ki izkazujejo od sestavin neodvisen in hkrati metaforičen pomen.

- **glagoli se/ si** (prosti morfem : povratni zaimek)
 - smejati se* (prosti morfem)
 - umivati koga – umivati se (be)* (povratnost – zaimek)
 - predvidevati kaj – kaj se predvideva* (tvornost/trpnost)
 - delati kaj : delati se* (pomenska razlikovalnost)
- **frazni glagoli**
 - gre za (koga/kaj)*
 - priti do (česa)*
 - računati na (koga/kaj)*
- **zveze s pomensko oslavljenimi glagoli**
 - delati družbo*
 - dati mir*
 - vzeti na znanje*
- **zveze s slovničnimi/funkcijskimi sestavinami**
 - v zvezi z/s, v bistvu, z drugimi besedami ...*
 - ne glede na, za razliko od ...*
 - bolj ali manj, sem in/pa/ter tja ...*
 - po eni strani - po drugi strani ...*

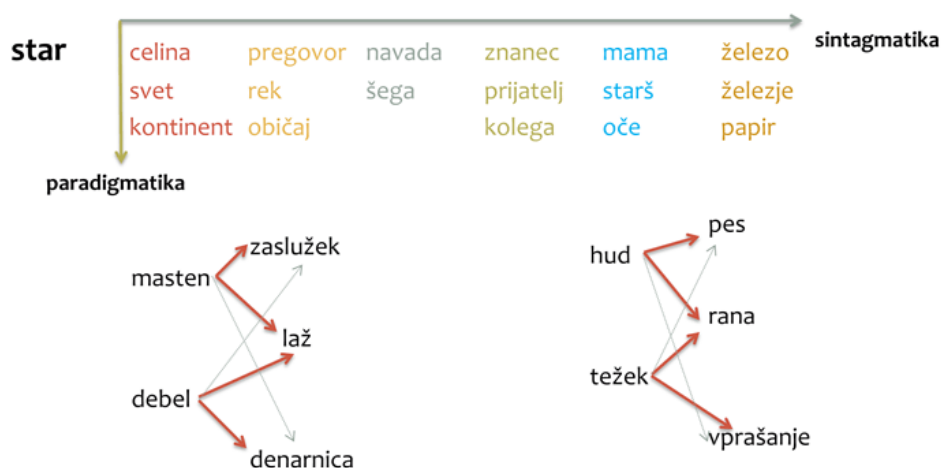
KORPUSNA FRAZELOGIJA

„you shall know a word by the company it keeps“ (Firth 1957: 11)

„I don't believe in word senses“ (Kilgarriff 1997: 91)

- Pomen kot zaključena enota ne obstaja, obstajajo zgolj **pomenske tendence**. (Hanks 2013)
- Pomenske tendence besed vplivajo na **izbire v svojem besedilnem okolju**.
- Vsak pomen se oblikuje **v konkretni govorni situaciji** in je odvisen od vsakokratnega jezikovnega in zunajjezikovnega **konteksta**.
- Pomeni so neposredno povezani z **vzorci, strukturami in frazeološkimi enotami**, v katerih se pojavljajo.
- Izbira pomena, kjerkoli v besedilu, ima odločilen vpliv na okoliške izbire. (J. M. Sinclair, *Corpus Concordance Collocation*, 1991)

„you shall know a word by the company it keeps“ (Firth 1957)



JANES IJS FILOZOFSKA
FAKULTETA

uporabnik: polona korpus: Kres

gramatična relacija

Isči

Iskanje
Seznam
Besedne skice
Tezaver
Najdi X
Primerjalne
skice
O korpusu
Shrani
Spremeni
nastavitve
Grupiranje
Razvrsti
Relacije
Več podatkov
Manj podatkov

koža (samostalnik)
Kres frekvenca = 20.475 (170.0 par milijon)

Constructions	S_kakšen?	S_predi-pred	S_s-kopa-česa	S_kopa-kaj
S_biti_s5 237 2.2	občutljiv 263 9.79	pod 471 8.07	nega 262 9.71	vlažiti 44 9.03
	suh 451 9.73	skozi 153 7.62	staranje 153 9.6	dražiti 56 8.91
	gol 177 9.1	na 1,882 5.98	čvrstost 36 8.04	pomirjati 45 8.62
	masten 130 8.97	brez 83 5.36	tip 191 8.03	negovati 53 8.38
	gladek 112 8.68	prek 11 4.86	čiščenje 83 7.99	gladiti 26 8.2
	ovčji 84 8.63	iz 191 4.83	plast 116 7.98	zgladiti 26 8.08
	pomarančen 79 8.49	čez 23 4.77	vnetje 62 7.76	navlažiti 22 8.03
	razdražen 69 8.44	do 115 4.72	luščenje 29 7.75	ščiti 50 7.87
	nečist 62 8.37	preko 6 4.26	predelovalec 26 7.62	napeti 33 7.85
	lasten 308 8.29	z 423 4.19	prožnost 29 7.54	učvrstiti 18 7.83
	napet 68 7.91	po 124 3.82	elastičnost 26 7.54	izsušiti 25 7.79
	zrel 70 7.88	v 692 3.77	videz 104 7.52	predreti 16 7.58
	trd 102 7.87	za 259 3.57	barva 289 7.51	očistiti 45 7.47
	živalski 84 7.81	ob 30 3.11	vlaženje 23 7.45	odreti 12 7.4
	razpokan 44 7.79	od 21 2.19	draženje 25 7.43	razdražiti 12 7.35
	čvrst 53 7.76	k 2 1.85	obnavljanje 28 7.42	mehčati 12 7.27
	nežen 72 7.64	zaradi 6 1.61	povrhnjica 21 7.25	odstraniti 73 7.15
	izsušen 39 7.61	med 11 1.59	strojenje 19 7.23	razvajati 17 7.12
	kozji 39 7.55	o 2 0.28	površina 125 7.21	odrgniti 9 7.07
	debel 83 7.52		ten 20 7.17	osvežiti 16 7.05
	mehek 67 7.46		občutljivost 28 7.14	prebosti 10 6.93
	zdrav 113 7.42		srbenje 19 7.12	odnesti 37 6.78
	srbeč 30 7.29		tonus 18 7.09	poživiti 11 6.74
	utrujen 46 7.27		okužba 43 7.03	zaščiti 30 6.73
	poškodovan 47 7.23		vlažnost 20 7.02	hladiti 11 6.71

kolokator

frekvenca
št. pojavitev v
korpusu

statistična
jakost
(salinece)

JANES IJS FILOZOFSKA
FAKULTETA

Baza kolokacijskega slovarja slovenskega jezika

http://bkssj.cjvt.si/

KSSJ:

- avtomatsko izluščeni podatki za 2.5000 iztočnic iz LBS + avtomatsko izluščeni podatki za novih 2.500 iztočnic

BKSSJ (trenutno):

- 2.500 iztočnic
- 2.310.100 kolokacij
- 72.117 struktur
- 5 zglედov/kolokacijo

študentski

študentski
študentski in
študentski blok
študentski center
študentski dom

Rezultati iskanja

Število rezultatov: 199

Iztočnice:

študentski

Kolokacije:

[pretežno] pretežno študentski

[študentski] študentski servis

[študentski] študentski dom

[študentski] študentski bon

[blok] študentski blok

[mlad] mlad in študentski

študentski
pridevnik

študentski bon
pridevnik + samostalnik

- študenti se bodo tako s študentskimi boni lahko še naprej prehranjevali tudi med počitnicami.
- Več kot 90 odstotkov anketiranih študentov mariborske univerze se prehranjuje s študentskimi boni, pri čemer večini zadošča 20 bonov na mesec.
- Pa vendar kljub manjšim težavam ostaja prehrana na študentske bone priljubljena in izkoriščena možnost.
- V zakon je bila dodana tudi sankcija prepovedi prodaje študentskih bonov za dobo najmanj enega in največ treh mesecev.
- Po skoraj dva meseca trajajoči spleti in zmedbi pred poslovalnicami za nakup študentskih bonov prodaja bonov sedaj poteka nemoteno.

Corpus Pattern analysis (Hanks)

- Projeciranje pomena iz sobesedila na posamezno besedo v korpusu -> Pattern Dictionary of English Verbs

- Theory of Norms and Exploitations* (Hanks 1994, 2013; Hanks in Pustojevsky 2004, 2005).

- osnovni vzorci (norms)
- skladenjske realizacije (alternations)
- ustvarjalne rabe (exploitations)

- **[[Human 1]] adore [[Human 2 | Deity | {Animal = Pet}]]**
[[Human 1]] has extremely strong positive emotions with regard to [[Human 2 | Deity | {Animal = Pet}]]
- **[[Human]] adore [[Entity | Eventuality]]**
[[Human]] enjoys [[Entity | Eventuality]] very much
- Pivovarna Laško bo vložila pritožbo.
[Human | institution]] vloži [{Entity = Document}]
[Human | institution]] predloži/da [{Entity = Document}] z namenom sprožiti določen postopek

*Teorija jezikovnih konvencij in možnosti njihove izrabe

<http://pdev.org.uk/#browse?q=f;f=C>

stavčni vzorci v leksikalni bazi za slovenščino

PATTERN: **[[Human 1 = Health Professional]] vaccinate [[Human 2 = Patient]]** {(against [[Disease]])}

IMPLICATURE: **[[Human 1 = Health Professional]] injects a vaccine into [[Human 2 = Patient]]** to produce immunity against [[Disease]]

cepiti *glagol*

če se **ČLOVEK** cepi, prejme CEPIVO proti določeni BOLEZNI |ali| če **ZDRAVNIK** cepi **ČLOVEKA**, mu v telo vnese CEPIVO proti določeni BOLEZNI

- kdo cepi
- kdo se cepi
- kdo se cepi proti čemu
- kdo cepi koga/kaj
- kdo cepi koga proti čemu
- kdo cepi koga s čim

avtomatsko pridobivanje stavčnih vzorcev izkorpusa

gramatična relacija	vzorec v LBS	glagol	korpusni primer
V_S_V_O3	<v>kdo/kaj G komu</v>	čestitati, zoperstaviti, dvoriti, gospodinjiti, pridružiti	Bushu je ob <k>izvolitvi</k> <i>čestital</i> tudi palestinski voditelj Jaser Arafat. Ne kar hitro, pošteno so si vzeli čas in mi dvorili, kot se spodobi.
V_S_V_O3_O2	<v>kdo/kaj G komu koga/česa</v>	ne dovoliti	»On mi tega ne dovoli.«
V_S_V_O3_O4	<v>kdo/kaj G komu kaj</v>	predlagati, preprečiti, nakloniti, prekrižati	Bela mačka mi je prekrižala pot, mogoče bi bilo dobro vplačati loto.
V_S_V_O4_predl	<v>kdo/kaj G koga/kaj + predlog</v>	obvarovati, odvrčati, predvidevati	Liz se ni pritoževala, niti ga ni skušala odvrčati od tega
V_lahko-G		privoščiti	»Naslednje logično vprašanje: si lahko privoščiva vzdrževanje dveh gospodinjstev?
V_gl_Cit		pojasniti	A tudi tam so nas zavrnili zaradi istega razloga," nam je pojasnil župan Jazbec.

čestitati (glagol) Gigafida

Constructions		
O_tretja_oseba	4.134	1.10
O_nedoločnik_cs	1.772	4.70
V_lahko_G	909	10.90
O_povratni_si	600	4.70
O_povratni_se	525	0.60
V_S_V_O3	516	47.80

O_tretja_oseba (glagolske zveze z sebkom v dajalniku)

bolí koga kaj
svita se komu kaj
briga koga

O_nedoločnik_cs (zveze z faznimi in modalnimi glagoli)

morati, začeti, želei, hoteti, nameravati ...; treba je [opozarjati, čestitati ...]

V_lahko_G (zveze z modalnimi prislovi)

lahko računate, predvidevate ...

O_povratni_si (zveze s povratnosvojilimi zaimki oz. morfemi)

smejati se; upati si/se

V_S_V_O3 vzorec: Sam-Glag-Sam3

kdo čestita komu

gojiti (glagol)

<u>kom-čem_g5</u>	4165	
<u>S_v-d_g5</u>	2579	5.1
<u>S_na-d_g5</u>	1192	5.9
<u>S_po-d_g5</u>	220	1.8
<u>S_ob-d_g5</u>	88	2.3
<u>S_pri-d_g5</u>	59	1.1
<u>S_o-d_g5</u>	27	0.6

S_v-d_g5

gojiti v [rastlinjaku, lončku nasadu, vrtnariji]
gojiti v [ribogojnici, akvariju, epruveti]
gojiti v [Sredozemlju]

S_na-d_g5

gojiti na [plantažah, gredicah, vrtu, balkonu]
gojiti na [farmi, posestvu]
gojiti na [Kitajskem, Primorskem]

S_po-d_g5

gojiti po [vrtovih]

S_ob-d_g5

gojiti ob [opori]

S_pri-d_g5

gojiti pri temperaturi [x] stopinj

S_o-d_g5

gojiti stereotipe o [harmoniki]
gojiti predsodke o [duševnem zdravju]

PARSEME

PARSING and Multi-word Expressions

<http://typo.uni-konstanz.de/parseme/>

- interdisciplinarna raziskovalna mreža v okviru evropske COST akcije
- računalniško procesiranje večbesednih enot
- standardizacija
- vključevanje v medjezikovno primerljive procesljive jezikovne vire
- PROJEKT: izdelava učnega korpusa z označenimi glagolskimi večbesednimi enotami za 20 jezikov (slovenščina, nizozemščina, grščina, bolgarščina ...)
 - mednarodni standardi in metode označevanja (Guidelines for annotators)
 - označevalnik: SentenceMarkupSystem
 - VLE:
 - glagolski frazemi (in pregovori) – ID (*stopiti na žulj*)
 - zveze s pomensko izpraznjenimi glagoli – LVC (*sprejeti odločitev, biti v dvomih* ...)
 - predložni glagolski frazemi/frazni glagoli – VPC (*priti do - 'doseči; trpeti za - 'oboleti'*)
 - drugi glagolski frazemi - OTH

JANES

označevalnik

IJS

FILOZOFSKA
FAKULTETA

ssj393.2230.7882 - Urejevalnik stavka

Dodaj povezavo O-RESLT SRL Označi zvezo ID ID-p LVC VPC OTH NPU PPU OTH-nv Označi jedra

ena
del
del
dol
dol
dve
štiri
dol
modra
modra
modra

Poročni zvonovi naj bi za zaljubljeni par zapeli še le prihodnje poletje .

ID ID

JANES

IJS

FILOZOFSKA
FAKULTETA

- **analiza predložnih glagolskih zvez z vidika prepoznavanja večbesednih leksikalnih enot**
- izbira korpusa/vzorca
- orodje za analizo: SkE: besedne skice
- izhodišče raziskave: glagol + (konkretni) predlog
- CQL: [tag="Gg.*"] [word="pod"] ali [tag="Gg.*"] [tag="D.*"]
- urejanje konkordančnega niza
- prenos v excelovo datoteko
- analiza ...

- sedeti pod borovcem
- [popustiti, sesesti se, klekniti ...] pod bremenom
- [dogajati, imeti, potekati ...] pod budnim očesom (koga/česa)
- podpisati se pod kaj – 'strinjati, podpirati,
- ...

- ČERMÁK, František, 2001: Substance of idioms: perennial problems, lack of data or theory? *International Journal of Lexicography* 14/1. Oxford: Oxford University Press. 1–20.
- ČERMÁK, František, 1985: Frazeologije a idiomatika. František Čermák, Josef Filipec: *Česká lexikologie*. Praha: Academia. 166–248.
- FIRTH, John, R., 1957: Modes of Meaning. *Papers in Linguistics*. London: Oxford University Press. 1934–51.
- GANTAR, Polona, Iztok KOSEM and Simon KREK, 2016: Discovering Automated Lexicography: The Case of the Slovene Lexical Database. *Int J Lexicography* (2016) 29 (2): 200-225 doi:10.1093/ijl/ecw014
- Gantar, Polona, 2007: Stalne besedne zveze v slovenščini: korpusni pristop. Ljubljana, Inštitut za slovenski jezik Frana Ramovša ZRC SAZU
- GANTAR, Polona, 2008: (Slovenska) leksika med leksikonom in slovnico. *Jezik in slovstvo* 53/5. 19–35.
- HANKS, Patrick, 2013: *Lexical Analysis: Norms and Exploitations*. Cambridge: MIT Press.
- KREK, Simon, 2012: New Slovene sketch grammar for automatic extraction of lexical data. *SKEW3, tretja mednarodna delavnica orodja Sketch Engine, Brno, Češka, 21-22. marec 2012*.
- KREK, Simon, 2015: Leksikografska orodja za slovenščini: slovnica besednih skic. Gorjanc, V., Gantar, P., Kosem, I. in Krek, S. (ur.): *Slovar sodobne slovenščine: problemi in rešitve*. Ljubljana: Znanstvena založba Filozofske fakultete UL. 358-378.
- KREK, Simon in KILGARRIFF, Adam, 2006: Slovene Word Sketches. Erjavec, T., Žganec Gros, J. (ur.): *Proceedings of the 5th Slovenian and 1st International Languages Technology Conference*. Ljubljana, Slovenia. 62–67.
- KRŽIŠNIK, Erika, 1990: Tipologija frazeoloških prenovitev v Cankarjevih prozih besedilih. *Slavistična revija* 38/4. 399–420.
- KILGARRIFF, Adam, 1997: »I Don't Believe in Word Senses.« *Computers in the Humanities* 31(2). Str. 91–113.
- SINCLAIR, J. M., 1991: *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Verdonik, D. (2015): Jezikovnoteoretska načela v korpusnem jezikoslovju. *Slovenščina* 2.0, 3 (1): 1–27.
- ŽELE, Andreja, 2012: *Pomensko-skladenjske lastnosti slovenskega glagola*. Zbirka *Linguistica et philologica*. Ljubljana: Založba ZRC, ZRC SAZU