



Entropija v slovenskih leposlovnih besedilih

Primož Jakopin



Foto: Matjaž Rebolj

Primož Jakopin vodi Laboratorij za korpus slovenskega jezika na Inštitutu za slovenski jezik Frana Ramovša ZRC SAZU. Ukvarja se z jezikovnimi tehnologijami in z njimi povezanimi kvantitativnimi raziskavami slovenskega jezika, statističnim opisom besedil in algoritmi za oblikoslovno označevanje besed. Sodeluje pri sestavljanju lematizacijskega slovarja in frekvenčnega slovarja slovenskih besed. Je tudi avtor programske opreme, s katero so bili pripravljene in na svetovni splet postavljene besedilni korpus *Nova beseda* ter več slovarjev Inštituta (<http://bos.zrc-sazu.si>).

Linguistica et philologica

Zbirka Linguistica et philologica

Primož Jakopin

Entropija v slovenskih leposlovnih besedilih

Izdal Inštitut za slovenski jezik Frana Ramovša ZRC SAZU
Za izdajatelja Varja Cvetko Orešnik

Založila ZRC SAZU, Založba ZRC

Za založnika Oto Luthar

Glavni urednik Vojislav Likar

Likovno-grafična urednica Milojka Žalik Huzjan

Tehnična ureditev in prelom Primož Jakopin s programom EVA

Tisk Littera picta, d. o. o., Ljubljana

Izid knjige je podprlo Ministrstvo za šolstvo, znanost in šport Republike Slovenije.

Digitalna različica (pdf) je pod pogoji licence CC BY-NC-ND 4.0 prosto dostopna:
<https://doi.org/10.3986/9616358774>

CIP – Kataložni zapis o publikaciji
Narodna in univerzitetna knjižnica, Ljubljana

004:811.163.6
004:6:821.163.6

JAKOPIN, Primož

Entropija v slovenskih leposlovnih besedilih / Primož Jakopin. –
Ljubljana : Založba ZRC, ZRC SAZU, 2002. – (Zbirka Linguistica et
philologica)

ISBN 961-6358-77-4

121042688

PRIMOŽ JAKOPIN

**ENTROPIJA V SLOVENSKIH
LEPOSLOVNIH BESEDILIH**



Predgovor

Entropija, količina, ki je v premosorazmernem razmerju z informacijo – lahko bi rekli, več kot je o neki stvari znanega, manjša je njena entropija in manj kot o njej vemo, večja je njena entropija – se s časom povečuje, vse od velikega poka pred milijardami let naprej. Vse tisto, kar je bilo še včeraj enostavno in bližje, je danes že malo manj jasno in dlje, ne glede na nenehni in nezaustavljivi napredek našega védenja in znanosti nasploh. Seveda se entropija ne povečuje povsod – globalizacija, s katero izginjajo krajevne posebnosti, jo na marsikaterem območju hitro zmanjšuje: pomislimo samo npr. na hitro hrano ali na uniformnost računalniških programov za široko rabo.

Tu je entropija predstavljena kot mera nedoločenosti slovenskih leposlovnih besedil, ki je bila včasih merjena s poskusi, ko so naključno izbranim bralcem pokazali veliko kratkih iztržkov iz besedil, do največ nekaj deset znakov, in so morali potem uganiti naslednjo črko, ki bi sledila na koncu. Splošno znano je, da vsebuje besedilo manj informacije, kot bi jo naključno nametana zaporedja črk – če v besedilu izpustimo vsako drugo črko, ga še vedno lahko skoraj vedno pravilno prepoznamo. Rezultati, predstavljeni v tej knjigi so bili zbrani v okviru dela na doktorski disertaciji iz teorije informacij, ki jo je avtor zagovarjal na Fakulteti za elektrotehniko Univerze v Ljubljani leta 1999. Leposlovna besedila in njihove sestavne dele, predvsem črke, besede in povedi, osvetljujejo še z drugega zornega kota, kot smo vajeni pri jezikoslovnih ali literarnovednih raziskavah.

V zadnjem času je nastalo že kar nekaj del, ki s kvantitativnega stališča opisujejo jezike, ki so nam blizu, npr. hrvaški (Moguš 1999) ali angleški (Leech, Rayson in Wilson 2001) in več internetnih besedilnih korpusov, tudi za slovenski jezik (Erjavec idr. 1998, Jakopin 2001), ki so podlaga za kvantitativne raziskave jezika (npr. Grzybek 2001). S sodobno računalniško tehnologijo pa bi bilo mogoče, ob ustreznih zakonskih dopolnitvah (prim. Jakac–Bizjak 2001) vsaj načelno napraviti tudi zbirko, ki bi zajela vsa objavljena dela v slovenskem jeziku (Jakopin 2002) in nanjo opreti še veliko popolnejšo raziskavo slovenskega jezika.

Pričujoče knjige ne bi bilo brez pomoči mojih sodelavcev, mentorjev in prijateljev, ki so dali vsak svoj prispevek. Najprej se moram iz srca zahvaliti prof. dr. Nikoli Pavešiću in prof. dr. Janezu Orešniku, za naklonjenost, skrb in spodbudo, s katero sta spremljala nastanek tega dela. Gmotno in tudi drugače me je ves čas podpirala Filozofska fakulteta, za gre iskrena zahvala obema dekanoma, prof. dr.

Ludviku Horvatu in pok. prof. dr. Franetu Jermanu, ter g. Amaliji Šiftar. Prof. dr. Milan Hladnik mi je s svojim večletnim zbiranjem elektronskih verzij slovenskih leposlovnih besedil, ki so javno dostopna na internetu, zelo olajšal zbiranje virov. Matjaž Rebolj mi je pomagal pri prenosu knjige *Devet fantov in eno dekle* v elektronsko obliko, kolegi iz slavistične knjižnice na fakulteti so se vedno hitro in prijazno odzvali, kadar je bilo treba, doc. dr. Igor Grdina pa je prepoznal besedilo knjige *Gorski potoki*, ki v elektronski obliki ni imelo ne avtorja ne naslova. Dr. Marina Zorman je veliko pomagala pri sestavljanju pregibalnih shem (paradigem) za samostalnike, pridevnike in glagole ter razvrščanju le-teh v sheme, kar je bilo nujno za učinkovito čiščenje in klasifikacijo besednih oblik v besedilnih zbirkah. Podarila mi je tudi nekaj Kosmačevih del, nekaj pa so mi jih odstopili v Knjižnici Jožeta Mazovca. Tri dela iz prvega vzorca je prijazno odstopil dr. Klaus Detlef Olof; seznam besed in frekvenc korpusa *Bank of English* je ljubeznivo posredoval Jeremy Clear iz podjetja Cobuild.

Delo takega obsega seveda ne bi bilo mogoče opraviti brez pomoči naklonjenega okolja, povezanega z reševanjem podobnih problemov. Našel sem ga na Inštitutu za slovenski jezik Frana Ramovša. Inštitut je del Znanstvenoraziskovalnega centra Slovenske akademije znanosti in umetnosti, katerega direktorju, dr. Otu Lutharju, gre posebna zahvala. Knjiga bi bila brez pomoči sodelavcev Inštituta bistveno revnejša, na nekatere dele, npr. na ročno preverjeni slovar lem Kosmačevih zbranih del, pa ne bi upal niti pomisliti. Največ zaslug ima predstojnica, dr. Varja Cvetko Orešnik, ki ji ni bila nobena ideja nesprejemljiva, nobena pot predolga in ki bi, roko na srce, težko pomagala še bolj. Aleksandra Bizjak je preverila vsa besedila, ki so bila označena z avtomatičnim oblikoslovnim označevalnikom; njen je tudi nabor oznak, ki sem jih pri tem uporabil. Dr. Peter Weiss je celotno besedilo jezikovno pregledal; njegova je nova beseda *enkratnica* (namesto *hapax legomenon*). Lučka Uršič je prebrala in popravila precejšen del Kosmačevega opusa, pretipkala je tiste najstarejše časopisne vire, ki prostorov Narodne in univerzitetne knjižnice niso smeli zapustiti. V tej knjižnici sem našel tudi računalnik, ki je bil dovolj zmogljiv, da sem lahko entropijo *n-terčkov* pripeljal tako daleč, kot sem želel – zelo mi je stal ob strani Ivan Groznik, pomagala pa sta še Rok Vidmar in Matjaž Kragelj.

Vso zahvalo sem dolžan še kolegoma iz časov, ko je Univerza v Ljubljani šele prišla do prvega lastnega računalnika s terminalsko mrežo; pomagala sta pri projektu, ki še ni imel ne imena ne domovinske pravice. Mag. Melita Ambrožič je pretipkala *Tantadruja* in *Balado o trobenti in oblaku*, dr. Jure Dimec pa znaten del *Pomladnega dneva*.

Naj se na koncu zahvalim še vsem drugim, ki so mi pomagali, pa jih nisem omenil. Predvsem seveda svoji družini, ki ni izgubila upanja, ki se je morala marsičemu odreči in veliko potrpeti, da je lahko nastalo to delo.

Vsebina

Predgovor	i
1 Uvod	1
1.1 Namen dela	1
1.2 Pregled vsebine	2
1.3 Pomembni dosežki	3
2 Viri	5
2.1 Besedila	5
2.1.1 Prvi vzorec	5
2.1.2 Drugi vzorec	7
2.1.3 Vzorca kot del celote	9
2.2 Strojna in sistemska programska oprema	10
2.3 Programska oprema	10
3 Priprava besedil	13
3.1 Format besedil	13
3.1.1 Dodatni simboli	15
3.1.2 Zaglavja besedil	18
3.2 Označevanje gradnikov besedila	20
3.2.1 Odstavki	21
3.2.2 Povedi	22
3.2.3 Premi govor	22
3.2.4 Citati, vzdevki in ostalo	23
3.3 Vprašanje napak	24
3.3.1 Napake pri prepoznavanju	25

3.4 Zbirka besed iz SSKJ	25
3.4.1 Samostalniki	26
3.4.2 Pridevniki	28
3.4.3 Glagoli	30
3.4.4 Preverjanje s slovarjem izpeljank	33
4 Oblikoslovno označevanje	35
4.1 Uvod	35
4.2 Prvi označevalnik	37
4.3 Drugi označevalnik	38
4.4 Oznake drugega vzorca	44
5 Statistični opis	45
5.1 Znaki	47
5.1.1 Nabor	49
5.1.2 Porazdelitve	52
5.1.3 Nterčki znakov	55
5.2 Besede	59
5.2.1 Rast besednega zaklada	59
5.2.2 Dolžine besed	62
5.2.3 Najpogostejše besede	64
5.2.4 Besedni nterčki	70
5.3 Povedi	79
5.3.1 Dolžine povedi	79
5.3.2 Najpogostejše povedi	84
5.3.3 Lipogramske povedi	86
6 Entropija	89
6.1 Definicije	89
6.2 Entropija nterčkov za oba vzorca	91
6.2.1 Algoritem	92
6.2.2 Rezultati	94
6.3 Ocena entropije z modelom	96
6.3.1 Huffmanovo kodiranje	96
6.3.2 Izbira postopka	101
6.3.3 Model z razrezom na enako dolge n-terčke	102

6.3.4 Model z razrezom na besede	104
6.3.5 Model z optimalnim razrezom na n-terčke	106
6.3.6 Komprimiranje vzorcev z drugimi programi	111
6.4 Druge entropije	113
7 Sklep	115
Literatura	117
Priloga A - Seznam del v prvem vzorcu	127
Priloga B - Seznam del v drugem vzorcu	129
Priloga C - Abecedni seznam oblikoslovnih oznak v drugem vzorcu	131
Priloga Č - Abecedni seznam lem, njihovih oblikoslovnih oznak in frekvenc v drugem vzorcu	137
Priloga D - Zanimivi naslovi na internetu	203
Entropy of Slovenian Literary Texts (Abstract)	205

*Niti en Hun ni zamudil trenutka, da bi ne bil potegnil meča,
zadnji jezdec pa so vrgli celo kopja in predrli troje konj.*
(Po Finžgarju, 1978)

1. poglavje

Uvod

1.1 Namen dela

Osnovni namen tega dela je odgovoriti na vprašanje, ki si ga zdaj zastavlja že tretji rod slovenskih računalničarjev, kakšna je informacijska vsebina besedil v našem jeziku, odgovoriti z viri in tehnologijo, ki je dostopna danes. Ob nastanku znanosti o informacijah, teorije informacij, pred točno pol stoletja (Shannon 1948) še ni bilo orodja, s katerim bi lahko izmerili vse parametre sporočil. Zaradi ročne obdelave so morali vzorci biti majhni in izmerjene poti po besedilih kratke. Današnji računalniki, skupaj z veliko večjo dostopnostjo elektronskih besedilnih virov, omogočajo, s primerno programsko opremo, obdelavo tisočkrat večjega gradiva in obdelavo v globino, kjer so bile do sedaj možne le ocene.

Tabela 1: Časovni pregled velikosti obdelanih besedil

leto	besed	objava
1962	4.000	Gyergyék 1962
1973	6.000	Gyergyék 1973
1974	60.000	Gyergyék s sod. 1974
1980	100.000	Vasle 1980
1994	650.000	Kristan idr. 1994
1995	1.600.000	Jakopin 1995a
2002	3.100.000	to delo

Drugi namen tega dela je statistično osvetliti za merjenje entropije zbrano besedilo v luči črk, besed in povedi ter postaviti nekaj mostov na poti do še veliko večjih, s korpusi velikih narodov primerljivih besedilnih zbirk.

Kot je razvidno že iz naslova, je delo omejeno na leposlovna besedila, točneje na prozna leposlovna besedila. Leposlovju pravimo tudi *literatura* ali *umetniška literatura* (SSKJ 1994), ki je razložena kot *umetnost, ki ima za izrazno sredstvo besedo, jezik, književnost* (SSKJ 1994). Podobni opisi leposlovja so še ... *je torej umetnost kot glasba ali slikarstvo, vendar v nasprotju z njima jezikovna vrsta umetnosti.* (Kos 1994) in morda najtočnejši *označuje ... umetnost, upodabljaajočo z uporabo*

dovršene jezikovne oblike: torej leposlovje (Prijetelj 1952, Kmecl 1996).

Razlog za omejitev dela na leposlovna besedila je predvsem v večji homogenosti takih besedil, njihovi kakovosti in večji bližini slovenskemu jeziku. Pesniške zbirke so bile izvzete zaradi težje dostopnosti in strukturne neoprijemljivosti pri nekaterih sodobnejših delih.

1.2 Pregled vsebine

Drugo poglavje je posvečeno virom, obema besedilnima vzorcema, ki skupaj obsegata 3.100.000 besed. Prvi vzorec, 7/8 celote, ki je imel pri preverjanju jezikovnega modela v šestem poglavju tudi vlogo učnega vzorca, vsebuje besedila 60 del 41 avtorjev, 46 izvirnih in 14 prevodov, z letnicami izida od 1858 do 1996. Drugi del vsebuje zbrano delo Cirila Kosmača, 408.000 besed oziroma 52 del, objavljenih med 1931 in 1988. V poglavju je ocenjena tudi skupna velikost celotne dosedanje slovenske leposlovne produkcije, na 12.000 del; oba vzorca skupaj predstavljata med 0,5% in 1% te celote.

Tretje poglavje obravnava pripravo besedil. Vsa so bila, v skladu z mednarodnimi priporočili označena do ravni povedi, drugi vzorec pa tudi do ravni besed. Besedila so tudi tako pripravljena, da je mogoč neposreden prenos vseh v format HTML za objavo na internetu. Drugi del poglavja opisuje odstranjevanje napak in v zvezi s tem postopek za tvorjenje zbirke 3.500.000 besednih oblik iz gesel *Slovarja slovenskega knjižnega jezika*, ki je bila nato uporabljena pri čiščenju besed v gradivu.

Četrto poglavje je namenjeno oblikoslovnemu označevanju besedil. Na ta način je bil označen drugi vzorec, s statističnim oblikoslovnim označevalnikom (angl. *part-of-speech tagger*), sestavljenim v ta namen. Označeno besedilo je bilo ročno pregledano na Inštitutu za slovenski jezik ZRC SAZU in ob tem je bilo ugotovljeno, da znaša natančnost označevalnika 92,4%. Po lematizaciji je bil napravljen slovar besednega zaklada Cirila Kosmača, ki obsega 15.140 lem iz 407.938 besednih oblik in je bistveno večji od doslej edinega lematiziranega opusa del Franceta Prešerna: 2.750 lem iz 13.757 besednih oblik (Suhadolnik 1985).

Peto poglavje navaja statistični opis obeh vzorcev, ki je najtemeljitejši doslej. Najprej je prikazan nabor znakov (168 različnih), nato porazdelitve črk, ki kažejo na prevlado črke *e* (7,92 % vseh znakov) nad *a* (7,74 %). Sledi opis *n*-terčkov znakov za *n* = 1 do *n* = 14 in besednih oblik. Povprečna dolžina le-teh v obeh vzorcih skupaj znaša 4,55 črke. Prvič v tovrstnih raziskavah so obdelane tudi povedi (prvi vzorec 201.000, drugi 37.000) – navedene so njihove dolžine, naštete najpogostejše, na koncu poglavja pa še podatki o lipogramskih povedih.

Šesto poglavje je posvečeno entropiji. Opisan je izvorni algoritem, ki s pomočjo informacije o gibanju novih znakovnih n -terčkov s frekvenco 1 (enkratnic) omogoča neposredno računanje entropije za višje n (do $n = 62$) in pri še veliko večjih besedilnih zbirka, kot je obravnavana zbirka. Za oba vzorca so navedene entropije, normirane vezane entropije in pogojne entropije do $n = 24$; iz teh podatkov in iz gibanja števila novih enkratnic je avtor ocenil zgornjo mejo entropije v slovenskih leposlovnih besedilih na 2,2 bita na znak. V nadaljevanju so opisani trije modeli za ponazoritev slovenskih leposlovnih besedil, ki bi prišli v poštev za konstruiranje komprimirnega mehanizma, posebej prilagojenega slovenskemu jeziku. S prvim vzorcem kot učno bazo in z drugim kot preizkusnim je bila pri zadnjem modelu, modelu z optimalnim razrezom na n -terčke dosežena vrednost 2,7 bita na znak.

V sedmem, sklepnem poglavju je kritično pregledano opravljeno delo, orisane so možnosti in smeri nadaljnjih raziskav na področju kvantitativne analize slovenskih besedil.

Sledi seznam navedenk (110 enot) in pet prilog. V prilogi A je naveden seznam in bibliografski opis vseh del iz prvega vzorca, v prilogi B vseh del drugega vzorca in v prilogi C abecedni seznam oblikoslovnih oznak v drugem vzorcu s frekvencami. Priloga Č vsebuje abecedni seznam lem, njihovih oblikoslovnih oznak in frekvenc v drugem vzorcu za leme, ki so dosegle frekvenco vsaj 10, v prilogi D pa so navedeni še najpomembnejši, z delom povezani naslovi na internetu. Prilogi A in B ter celoten seznam iz priloge Č (za vse frekvence) so objavljeni tudi na internetu. Za pregledovanje seznama lem in oblikoslovnih oznak preko interneta je bil pripravljen poseben iskalni program.

1.3 Prispevki dela

Na enoten način je bil pripravljen in statistično opisan največji besedilni korpus doslej, prispevki k teoriji informacij in h kvantitativnemu jezikoslovju pa bi lahko bili naslednji:

1. Ocena zgornje meje entropije za leposlovnna besedila, 2,2 bita na znak.

Prvič je bila podana konkretna ocena zgornje meje entropije za slovenska leposlovnna besedila, 2,2 bita na znak, ki je bila prej ocenjevana le posredno. Vrednost je nekoliko večja kot pri angleškem jeziku (2,0 bita na znak). Ob tem je bil razvit postopek za neposredno računanje entropije za večje n in za velike besedilne vzorce.

2. Nizkoentropijski jezikovni model za besedila.

Sestavljen je bil model za ponazoritev slovenskih leposlovnih besedil,

ki sloni na pogostostih znakovnih n-terčkov. Model je bil uporabljen kot osnova za komprimirni postopek, obnesel pa se je tudi kot sredstvo za preverjanje, ali je dano besedilo slovensko ali ne in kako daleč je od našega jezika.

3. Korpus slovenskih leposlovnih besedil.

Iz elektronskih besedilnih virov različnega izvora in kvalitete je nastal prečiščen fond nad 3 milijone besed (dva besedilna vzorca), ki je bil označen do ravni povedi, 408.000 besed pa ima tudi oblikoslovne oznake in podatke o lemi. Pri tem je bil izpeljan slovar besednih oblik za osnovne besedne vrste, ki temelji na Slovarju slovenskega knjižnega jezika in obsega 3,5 milijona enot. Vseh 112 besedil (prvič objavljenih od 1858 do 1996) je bilo opremljenih na enak način, za avtomatično konverzijo v format HTML; objavljena so na naslovu: http://bos.zrc-sazu.si/main_si_l2.html.

4. Kvantitativna analiza besedilnih vzorcev.

Na obeh vzorcih je bila opravljena kvantitativna analiza s porazdelitvami črk in znakovnih nizov, ki je na doslej največji zbirki prvič segla čez raven besednih oblik, tudi na besedne n-terčke in na povedi.

5. Statistični oblikoslovni označevalnik.

Narejen je bil statistični oblikoslovni označevalnik z natančnostjo 92 %. Z njim je bil označen celoten korpus slovenskega pisatelja Cirila Kosmača, in po lematizaciji izdelan slovar njegovega besednega zaklada (408.000 besed, 15.000 lem). Z njim je mogoče za nadaljnje raziskovalne namene učinkovito pripravljati večje količine besedil.

Rezultati tega dela osvetljujejo pomembno področje na meji med teorijo informacij in jezikoslovjem, kjer je bilo že dolgo načrtovane raziskave večjega obsega mogoče izvesti šele s tehnologijo, dostopno v zadnjem času.

2. poglavje

Viri

2.1 Besedila

Idealni vir za tako nalogo bi bila vsa doslej objavljena slovenska leposlovna dela. Tak vir bi bil popoln in na njem izmerjene vrednosti bi imele povsem trdno veljavo. Da bi se popolnosti vsaj na neki način približal, se je pisec teh vrstic odločil za dve besedilni zbirki: prvi vzorec, v katerem bi bilo kar največ del, in drugi vzorec, v katerem bi bila vsa objavljena dela priznanega avtorja. Pogoji, potreben za izpolnitev zahteve po popolnosti drugega vzorca, je tudi ta, da izbrani avtor ne bo ničesar več napisal, kar pomeni, da ga ni več med nami. Po daljšem premišljevanju in posvetovanjih je izbira padla na Cirila Kosmača (1910–1981). Sodobnejšega avtorja, ki bi ustrezal zastavljenim pogojem, žal ni bilo mogoče najti.

2.1.1 Prvi vzorec

Vsa leposlovna besedila, ki v današnjih časih pridejo na police v knjigarnah, so že dobro desetletje pripravljena z računalnikom in tako, vsaj do tiska, obstajajo tudi v elektronski obliki. Pot do večje zbirke teoretično torej ne bi smela biti težka, omejena pa je predvsem z bojaznijo avtorjev do zlorabe avtorskih pravic. Tudi zakon o knjižničarstvu, ki bi od tiskarjev zahteval, da oddajo obvezni izvod tudi v elektronski obliki, če ga imajo, še ni bil sprejet in je zato Narodna in univerzitetna knjižnica po tej plati povsem brez virov. Tako je bila izredno dragocena *Zbirka slovenskih leposlovnih besedil*, ki jo od junija 1995 naprej na internetu na naslovu <http://www.ijs.si/lit/leposl.html> gradi Miran Hladnik (npr. Hladnik 1995). Iz nje, tudi iz besedil, ki še niso bila pripravljena za objavo na internetu, izvira več kot polovica prvega vzorca. Večina drugih del je iz sodelovanja med avtorjem in založbo Mihelač, nekaj jih je iz arhiva Inštituta za slovenski jezik Frana Ramovša ZRC SAZU, Varja Cvetko Orešnik je odstopila spomine svojega očeta, dve svoji deli Gitica Jakopin (tretje je prenesel v elektronsko obliko pisec teh vrstic), nekaj jih je avtorju posredoval Klaus Detlef Olof s celovške univerze,

Orwellov 1984 pa je preskrbel Tomaž Erjavec. Natančnejši popis del z bibliografskimi podatki je v prilogi A (str. 105 do 107), kratek pregled pa v spodnji tabeli.

Tabela 2: Dela prvega vzorca z letom prvega izida in številom besed

1. Ivan Cankar	Hiša Marije Pomočnice	1904	29.702
2. Ivan Cankar	Potepuh Marko in Kralj Matjaž	1905	26.547
3. Ivan Cankar	Hlapec Jernej in njegova pravica	1907	18.423
4. Ivan Cankar	Mimo življenja	1920	52.316
5. Janez Cigler	Sreča v nesreči	1836	33.778
6. Dragotin Cvetko	V prostoru in času	1995	97.335
7. Fran S. Finžgar	Sama	1912	49.194
8. Fran S. Finžgar	Strici	1927	15.673
9. Nataša Gale	Fran Gestrin	1927	25.603
10. Miran Hladnik	Kmečka povest (Iz polpret. časa)	1992	17.254
11. Gitica Jakopin	Žarometi	1962	36.808
12. Gitica Jakopin	Devet fantov in eno dekle	1963	29.136
13. Gitica Jakopin	Slovo od dekljštva	1995	57.116
14. Josip Jurčič	Nemški valpet	1867	7.969
15. Josip Jurčič	Sosedov sin	1868	18.680
16. Janko Kersnik	Ponkrčev oča	1882	2.184
17. Janko Kersnik	Mačkova očeta	1886	1.601
18. Janko Kersnik	Jara gospoda	1893	19.015
19. Alojz Kraigher	Peter Drozeg	1916	19.994
20. Tomo Križnar	O iskanju ljubezni	1989	132.098
21. Tomo Križnar	Samotne sledi	1993	81.744
22. Fran Levstik	Martin Krpan z Vrha	1858	5.682
23. Fran Levstik	Popotovanje iz Litije do Čateža	1858	8.214
24. Fran Levstik	Sveti doktor Bežanec v Tožbanji vasi	1870	11.293
25. Milan Lipovec	Čubejska prigoda	1972	7.153
26. Florjan Lipuš	Zmote dijaka Tjaža	1972	51.203
27. Florjan Lipuš	Srčne pege	1991	52.864
28. Miroslav Malovrh	Opatov praporščak	1903	51.420
29. Fran Maselj-Podlimbarski	Gorski potoki	1895	33.272
30. Ivan Pregelj	Mlada Breda	1913	84.350
31. Ivan Pregelj	Thabiti Kumi	1933	9.362
32. Prežihov Voranc	Boj na požiralniku	1935	10.435
33. Prežihov Voranc	Samorastniki	1937	12.949
34. Marijan Pusavec	Zbiralec nasmehov	1991	30.518
35. Ivan Sivec	Kruh ponoči spi	1994	49.047
36. Ivan Sivec	Triglavski kralj	1994	59.604
37. Jakob Sket	Miklova Zala	1884	35.471
38. Josip Stritar	Rosana	1877	18.344
39. Rudi Šeligo	Triptih Agate Schwarzkobler	1968	25.739
40. Gustav Šilih	Beli dvor	1938	131.661
41. Damijan Šinigoj	Neizstreljeni naboj	1994	94.375
42. Ivan Tavčar	Janez Sonce	1885	51.772
43. Ivan Tavčar	Otok in struga	1897	16.118
44. Ivan Tavčar	Cvetje v jeseni	1917	26.349
45. Janez Trdina	Bajke in povesti o Gorjancih	1882-1888	59.241
46. Ivan Zorec	Izgnani menihi	1937	49.676

Tabela 2 (nadaljevanje): Prevodi prvega vzorca z letom izida, številom besed in prevajalcem

47. Christian Andersen	Andersenove pravljice	1992	53.216	Janko Moder
48. Lewis Carroll	Alica v ogledalu	1978	27.410	Gitica Jakopin
49. Lewis Carroll	Aličine prigode v čudežni deželi	1983	23.443	Gitica Jakopin
50. Robert Cringely	Naključni imperiji	1995	100.475	Primož Jakopin
51. Alexandre Dumas	Trije mušketirji	1995	55.734	Tomo Virk
52. Gustave Flaubert	Bouvard in Pécuchet	1995	85.702	Stane Ivanc
53. Peter Handke	Ponovitev	1988	65.259	Silvija Borovnik
54. Jordan Horowitz	Kako vzgojiti očeta	1994	25.188	Mira Hladnik
55. George Mikes	Priročnik za snobe	1994	29.125	Mitja Meršol
56. Libuše Moníková	Fasada	1994	113.764	Štefan Vevar
57. George Orwell	1984	1983	91.022	Alenka Puhar
58. Platon	Država	1976	92.844	Jože Košar
59.	Sveto pismo Nove zaveze	1984	149.338	več prevajalcev
60. Jules Verne	V osemdesetih dneh okoli sveta	1996	54.003	F.Stopar, M.Medvedšek
Skupaj			2.721.416 besed	

Knjige obsegajo časovni razpon od 1858 do 1996, se pravi zadnjih 138 let, s tem da je četrtnina (15) iz 19. stoletja, približno polovica (29) pa je novejša, iz časa po letu 1962. Dela so razvrščena po abecedi avtorjev, znotraj istega avtorja pa kronološko. Pri izvornih delih je navedeno leto prve objave posameznega dela, tudi če je bil kot podlaga za prenos v elektronsko obliko uporabljen kasnejši ponatis.

Knjige so bile v različnih formatih, od formata HTML pri večini del iz Hladnikove zbirke, do formata programa WordStar, s katerim si očitno pomaga še veliko piscev, Microsoftovega izmenjalnega formata RTF, internega formata programov Word in WordPerfect, nekaj datotek je bilo pa celo v formatih avtorjevih programov STEVE in EVA. Konverzija je bila v vseh primerih opravljena s programom EVA.

Vsa dela seveda niso leposlovna v najstrožjem pomenu. Dilema nastane predvsem pri spominski literaturi in pri potopisih, kjer je črto dostikrat težko potegniti. Vključenih je tudi 14 prevodov, tako zaradi količine kot zaradi večje zaokroženosti vzorca. Nekaj del, ki jih je avtor še imel, pa so bila že zelo esejistična, v vzorec ni vključenih. Žal tudi ni prišla v poštev precej obsežna zbirka sicer neleploslovnih besedil, ki jo je zbrala M. Zorman za izdelavo svoje disertacije (Zorman 1997).

2.1.2 Drugi vzorec

Ta vzorec vsebuje vsa objavljena dela Cirila Kosmača; od leta 1931 do 1988. Kot pri prvem vzorcu je popoln popis z vsemi bibliografskimi podatki naveden v prilogi (priloga B, stran 109 in 110), kratek pregled pa v tabeli 3 na naslednji strani. Pri zbiranju Kosmačeve bibliografije sta bila uporabljena dva vira – popisa Helge Glušičeve (Glušič 1975: 22) in Ivana Cesarja (Cesar 1981). Dela so razvrščena približno kronološko po nastanku oz. objavi zadnje različice, kadar je bilo

Tabela 3: Dela drugega vzorca z letom prvega izida in številom besed

1.	Božična noč v ječi	1931	648
2.	Študenta Petra povest o materi	1932/33	2.347
3.	Potepuh Najdu	1932/33	3.170
4.	V znamenju rojstva in prerojenja	1932	1.212
5.	Božično pismo iz celice 589	1932	1.837
6.	Gornje mesto	1933	1.581
7.	Velika nedelja	1933	3.643
8.	Hodil po zemlji sem naši ...	1933	5.567
9.	Kraška simfonija	1933	2.914
10.	Cerkovnik Martin	1933	3.951
11.	Na sveti večer	1934	2.507
12.	Reka	1934	1.179
13.	Hiša št. 14	1934	12.925
14.	Ogorek	1935	3.168
15.	Obisk	1935	2.860
16.	Zločin Bernarda Tula	1935	3.027
17.	Prazna ptičnica	1935 (1988)	26.476
18.	Človek na zemlji	1935	16.058
19.	Sreča	1936	5.348
20.	Gosenica	1936	4.738
21.	Kruh	1936	2.630
22.	Življenje in delo Venca Poviškaja	1937	19.539
23.	15. marec 1937	1937	2.283
24.	Tistega lepega dne	1938	7.669
25.	Zlato	1941	2.910
26.	Dragičeva smrt	1946	2.484
27.	Poletje	1947	1.759
28.	Očka Orel	1947	12.450
29.	Težka nedelja	1948	15.268
30.	Na svoji zemlji	1949	30.885
31.	Beli konj	1956	1.407
32.	V žagi	1950	4.983
33.	Hudo je, človek moj	1950	7.959
34.	Smrt nedolžnega velikana	1952	15.722
35.	Pot v Tolmin	1953	12.579
36.	Pomladni dan	1953	61.567
37.	Mož, ki ni poznal obupa	1953	1.064
38.	Žuželč in njegova krava	1953	1.670
39.	Prvi prizor	1953	2.109
40.	Pomlad s čapljo	1953	4.443
41.	Sredi vasi	1956	1.853
42.	Kovač in hudič	1959	6.296
43.	Medvejke	1959	1.651
44.	Ozimina	1959	4.257
45.	Utrinki iz ječe	1959	1.449
46.	Klopotec	1960	2.375
47.	Tantadruj	1964	13.791
48.	Balada o trobenti in oblaku	1964	45.012
49.	V gaju	1972	2.628
50.	Ringaraja	1972	3.997
51.	Kamen in njiva	1984	5.302
52.	Pravljica o maku	1984	3.451

Skupaj

407.938 besed

različic (npr. pri *Baladi*) več. S prenosom gradiva v elektronsko obliko je avtor

začel že v začetku osemdesetih let, ob pomoči Melite Ambrožič in Jureta Dimca (*Pomladni dan, Balada, Tantadruj*) – s pretipkavanjem na terminalih univerzitetnega računalnika DEC-10, večji del, ki je še ostal, pa je bil prenesen v letu 1997. Vnos je tokrat opravil avtor, v pretežni meri strojno (z bralnikom slike Microtek in optičnim prepoznavanjem znakov s programom EVA), na Inštitutu za slovenski jezik ZRC SAZU. Le tiste redke predvojnne tiske, ki niso smeli iz prostorov NUK, predvsem iz časopisa Istra, je na prenosni računalnik pretipkala Lučka Uršič. Pri tem ji je malo pomagala Marija Jakopin, ki je tudi pretipkala dve črtici iz časopisa Delo. Vneseno besedilo je prebrala in popravila Lučka Uršič.

Izvezeto ni nobeno znano Kosmačevo delo – vključeni sta tudi zadnji dve deli v seznamu (51 in 52), ki sta, dopolnjeni, izšli po smrti in tudi *Prazna ptičnica*, ki je pisec očitno ni nameraval objaviti, pa je bila tako prvič natisnjena šele leta 1988.

2.1.3 Vzorca kot del celote

Pri vrednotenju rezultatov, ki nastanejo pri obdelavi vzorca in ne celote, se vedno pojavi tudi vprašanje, kolikšen delež celote predstavlja vzorec. Z vprašanjem števila in obsega vseh slovenskih leposlovnih del se, kot vse kaže, še nihče ni resneje ukvarjal. Dobro je obdelano obdobje od 1919 do 1935 (Pivec-Stelè 1936), tabe-

Tabela 4: Popis leposlovnih izvirnih knjig in prevodov ter vseh knjig in prevodov

1919	37	16	190	20	1955	97	119	934	138
1920	36	19	196	27	1956	109	107	811	139
1921	38	24	232	35	1957	92	125	874	144
1922	46	28	268	46	1958	96	108	864	152
1923	65	40	295	55	1959	98	165	865	203
1924	29	29	209	45	1960	97	141	981	189
1925	34	39	249	54	1961	120	165	1036	214
1926	44	35	230	40	1962	117	217	1040	272
1927	39	27	266	29	1963	142	220	1007	263
1928	39	25	255	30	1964	131	231	1201	280
1929	56	31	265	48	1965	208	254	1265	322
1930	57	48	238	57	1966	175	263	1098	356
1931	63	69	249	81	1967	156	263	1102	438
1932	57	45	244	53	1968	178	216	1304	358
1933	58	50	232	55	1969	203	231	1194	425
1934	65	54	220	61	1970	186	252	1446	453
1935	54	36	175	45	1971	207	275	1633	491
					1972	270	223	1649	352
1945-47	185	156	1116	346	1973	276	277	1802	471
1948	41	43	471	119	1974	244	252	1855	439
1949	75	40	487	88	1975	333	296	1759	493
1950	71	64	679	109	1976	446	264	1838	497
1951	64	77	854	110	1977	276	224	1784	389
1952	67	83	723	110	1978	128	243	1669	421
1953	70	62	956	87	1979	288	236	1859	408
1954	69	101	1077	116					
					Skupaj	6.132	6.608	43.246	10.173

le za posamezna leta v razdobju 1945–1979 pa je mogoče najti v slovenskih bibliografijah (Slovenska bibliografija 1945 ... 1979), ki jih je izdajal NUK za posamezna koledarska leta. Izvleček je naveden v tabeli 4 na prejšnji strani. Sklepamo lahko, da predstavlja časovni razpon iz tabele 4, 52 let, večji del slovenske leposlovne produkcije, saj tendenca rasti v zadnjih dveh desetletjih ni šla vedno samo navzgor (npr. Kodrič-Dačić 1997: str. 32). Približna ocena, ki jo je za celotno število slovenskih leposlovnih knjižnih del podala Anka Sollner-Perdih iz slavistične knjižnice Filozofske fakultete v Ljubljani, namreč število 12.000 (med 11.500 in 12.500), se po vsem tem ne zdi več optimistično pretirana, ampak blizu resnice. Oba vzorca skupaj potem predstavljata med 0.5 in 1 % celote, po obsegu pa, ker vsebuje celota tudi znaten delež pesniških zbirk, ki so krajše od proznih besedil, verjetno bliže 1 %. Celoto, dosedanji slovenski knjižni leposlovni opus, lahko tako ocenimo na velikostni red 300.000.000 besed. Delež obeh vzorcev skupaj med slovenskimi proznimi deli večje teže, ki jih je približno 500 (Lah 1997), je verjetno še večji.

2.2 Strojna in sistemska programska oprema

Delo je bilo opravljeno predvsem z navadnim prenosnim računalnikom, ki ga ima avtor na uporabo od Inštituta za slovenski jezik ZRC SAZU v Ljubljani (ISJ). Nekaj tabel je bilo izračunanih na namiznem računalniku ISJ, za vnos besedil je bil uporabljen starejši namizni računalnik na ISJ z bralnikom slike, za ročno preverjanje besedil pa isti računalnik. Pri preverjanju oblikoslovnih oznak je bil uporabljen najprej namizni računalnik na ISJ, zadnjih nekaj mesecev pa njegov novejši naslednik na ISJ.

Pri sestavljanju in preizkušanju pregibalnih shem in pri poizvedbah na internetu je bil uporabljen avtorjev računalnik na Filozofski fakulteti v Ljubljani. Za vnos knjige *Devet fantov in eno dekle* je služil namizni računalnik v knjižnici Oddelka za zgodovino na Filozofski fakulteti z bralnikom slike.

Izračun entropij n -terčkov obeh vzorcev je bil opravljen na strežniku Narodne in univerzitetne knjižnice v Ljubljani (Pentium Pro 2 x 400 MHz, 512 MB pomnilnika, Windows NT). Pri tem se je zelo obnesla tudi hitra povezava med NUK in ZRC SAZU (500 KB/sek). Pretipkavanje besedil v NUK je potekalo na starejšem prenosnem računalniku Marije Jakopin.

Vnos prvih treh knjig Cirila Kosmača je potekal na terminalih, priključenih na računalnik DEC-10 v Računskem centru Univerze v Ljubljani.

2.3 Programska oprema

Vsa dela, priprava virov in njihova obdelava so bila skoraj v celoti, vključno s

pripravo rokopisa in besedila knjige za tisk, opravljena z urejevalnikom EVA (<http://www.laze.org/eva>, Jakopin 1995c), ki je bil v ta namen dopolnjen z vsemi potrebnimi orodji. Uporabljeni sta bili najprej verzija za DOS, v letu 1998 in kasneje pa verzija za okolje Windows, obe napisani v jeziku C (Zortechov C oz. Microsoftov Visual C). Iskalni program za pregledovanje priloge Č na internetu je NEVA (strežniška različica programa EVA). Edina druga programska oprema je bil program za prepoznavanje znakov Recognita, s katerim je bil vnesen roman *Devet fantov in eno dekle*.

3. poglavje

Priprava besedil

3.1 Format besedil

Kot je bilo že omenjeno v poglavju o virih, so bile elektronske verzije besedil obeh vzorcev pripravljene na zelo različne načine. Navadno se tudi niso strogo držale vrstičnega reda v tiskani različici dela, ampak so bile vrstice, predvsem kadar so bila dela pretipkana, dolge toliko, kolikor črk je imel urejevalnik v eni vrstici na zaslonu. Odstavki so se seveda ujemali z izvirnikom, meje strani pa so v veliki večini primerov manjkale. Pri odločanju, kako naj bo besedilo v obeh vzorcih interno shranjeno, je bilo upoštevanih več zahtev:

1. berljivost besedila
2. ustrezanje mednarodnim standardom za shranjevanju besedil
3. primernost za nadaljnje obdelave
4. enostaven prenos na internet

Prva zahteva, zahteva po berljivosti, je bila postavljena zaradi čim lažjega ročne-ga branja in preverjanja besedil. Dodatne informacije, vnesene v besedilo – meje odstavkov, povedi, oznake za premi govor in podobno, morajo biti sicer vidne, vendar ne tako vpadljive, da zaradi njih ni jasno videti besedila samega. Da bi bila zahteva izpolnjena, so bile v okviru možnosti, ki jih je avtorju nudila arhitektura urejevalnika EVA (16-bitni nabor znakov, lasten rastrski zapis njihovih slik), vse dodatne oznake kodirane kot posebni znaki.

Druge zahteve ni potrebno posebej utemeljevati. V zadnjih dveh letih so standardi (npr. Erjavec 1997) in priporočila (Baker idr. 1997) za shranjevanje besedil dobili že bolj konkretne in jasne oblike; strukturiranje besedil pri obeh vzorcih je bilo opravljeno do ravni povedi; posebej je označen premi govor. Dodatne oznake, vnešene v besedilo, je mogoče enostavno nadomestiti z oznakami po standardu SGML oz. XML.

Tretja zahteva, predvsem tako oblikovanje besedila, da ga za kasnejšo morebitno oblikoslovno označitev (angl. *part-of-speech tagging*) ne bi bilo treba preurejati, je pomenila, da morajo biti vrstice krajše, kot bi sicer lahko bile, saj so oblikoslovne oznake velikokrat daljše od besed in je potrebno za ujemanje začetkov besed in oznak pod njimi predvideti nekaj dodatnega prostora. Na sl. 1 je najprej naveden primer neoznačenega besedila, pisan z običajnimi zaslonskimi črkami urejevalnika, na sl. 2 pa isti odstavek, tokrat dopolnjen z oznakami za začetke in konce odstavkov in povedi ter z oblikoslovnimi oznakami besed v vsaki drugi vrstici:

Slika 1: Primer neoznačenega besedila (C. Kosmač, Kamen in njiva)

A še ni imel sreče. Prišel je prav blizu.
 Ustavil se je na kolovozu z drugimi kamni. Tisto pomlad
 je stari Pologar postavil nov suhi zid pod kolovozom pa
 je v zid vzdal tudi kamen. Vendar mu je bilo
 zdaj lepše kakor v gozdu. Bil je blizu njive, pa
 jo je lahko gledal, kako se je ogrinjala s snegom,
 kako se je prebujala, kako se je pregrinjala z zlato
 odejo zrele pšenice, z bledomodro odejo cvetočega lanu in
 z belo obleko cvetoče ajde.

Slika 2: Označeno besedilo iz slike 1

ⓂA__ še ni___ imel sreče.Ⓜ ⓂPrišel je__ prav blizu.Ⓜ
 Vpr Č GZPce GLme Sže2 GLme GPce Č A
 ⓂUstavil se_ je__ na kolovozu z_ drugimi kamni.Ⓜ ⓂTisto pomlad
 GLme Gmp GPce E5 Sme5 E6 ZDmp6 Smp6 ZKže4 Sže4
 je__ stari Pologar postavil nov_ suhi_ zid_ pod kolovozom pa_
 GPce Pme1 IOme1 GLme Pme4 Pme4i Sme4 E6 Sme6 Vpr
 je__ v_ zid_ vzdal tudi kamen.Ⓜ ⓂVendar mu_____ je__ bilo_
 GPce E4 Sme4 GLme Vpr Sme1 Vpr ZDcme3 GPce GLRse
 zdaj lepše kakor v_ gozdu.Ⓜ ⓂBil__ je__ blizu njive, pa_
 A Aj Vpo E5 Sme5 GL0me GPce E2 Sže2 Vpr
 jo_____ je__ lahko gledal, kako se_ je__ ogrinjala s_ snegom,
 ZDcže4 GPce A GLme ZV Gmp GPce GLže E6 Sme6
 kako se_ je__ prebujala, kako se_ je__ pregrinjala z_ zlato
 ZV Gmp GPce GLže ZV Gmp GPce GLže E6 Pže6
 odejo zrele pšenice, z_ bledomodro odejo cvetočega lanu in_
 Sže6 Pže2 Sže2 E6 Pže6 Sže6 PČme2 Sme2 Vpr
 z_ belo obleko cvetoče ajde.ⓂⓂ
 E6 Pže6 Sže6 PČže2 Sže2

Za besedami, katerih oblikoslovne oznake so daljše, je še ustrezno število mehkih presledkov.























V skladu s to zahtevo so bila besedila na novo formatirana v vrstice, ki niso bile daljše kot 57 znakov, kar je pomenilo vsaj 21 znakov dodatnega prostora v vsaki; vrstica urejevalnika EVA z običajnimi zaslonskimi črkami ima pri zaslonski ločljivosti 800 x 600 pik namreč 78 enako širokih znakov.

Četrta zahteva, naj bo besedilo shranjeno tako, da bo možna kar najenostavnejša pretvorba v format HTML, je zahtevala ureditev predvsem zaglavja pred vsakim besedilom in uvedbo nekaterih dodatnih oznak, npr. za dele besedila, ki morajo biti napisani z enako širokimi črkami (npr. ladijski vozni red v romanu J. Verna *V osemdesetih dneh okoli sveta.*) Odločiti se je bilo treba tudi, kako interno prikazati meje odstavkov. V knjigah so navadno pisani z zamikom približno 5 presledkov, na internetu pa določilo <p> izpusti eno prazno vrstico in so tako odstavki v knjižnih delih tam navadno upodobljeni brez zamika v desno, zato pa s prazno vrstico. Ker pregledovalnik na internetu običajno uporablja proporcionalno široke črke, je presledek razmeroma ozek, ožji tudi od večine malih črk in pet presledkov na začetku odstavka ni prav opaznih. Zato je v obeh vzorcih na začetku odstavka zamik 8 presledkov, ki se pri pretvorbi v obliko HTML spremenijo v trde presledke in jih pregledovalnik ne združi. Določilo <p> je pri pretvorbi v obliko HTML izpuščeno, namesto </p> pa uporabljen le prelom vrstice –
.

3.1.1 Dodatni simboli

Zaradi zahtev iz prejšnjega razdelka in zaradi omejitve, da se v analizi upošteva jo le slovenski deli besedil (tuji citati pa ne), so bili uvedeni naslednji dodatni simboli iz tabele 5.

Tabela 5: Simboli za označevanje besedila

 	246, 247
 	248, 249
 	250, 251
 	252, 253
 	254, 255
 	256, 257
 	258, 259
 	260, 261
 	262, 263
 	264, 265
 	266, 267

Gre za 22 simbolov, ki nastopajo v parih – prvi stoji pred označenim delom besedila, drugi pa za njim. V tabeli sta za slikama obeh znakov vedno navedeni še njuni kodi. Prvi par simbolov, s kodama 246 in 247, služi za označevanje tistega

je bil že pogostokrat, poljubljal je mnogokatere ustnice in opeval mnogo oči; kadar se je naveličal, se je poslovil ves solzen in z izgovorom, ki je dišal po humaniteti:»

»

»Na mojo pot ne trosi rož dišečih:
temna in strašna pred menoj leži ...«

»

In je šel. Pravili so čudne bajke o njegovem vagabundstvu in bolj češčen od njegovih verzov je bil njegov havelok. On sam je vedel, da je njegovo

Predzadnja para, s kodama 260 in 261 ter 262 in 263, označujeta citate in vzdevke (angl. *citation in so called*):

» »

Oni pa so pokazali tri Pavlova pisma, katerim nobeno sodišče ni moglo utajiti ali izpodbiti pristnosti in veljave. Svoboda se jim je torej potrdila; radi nje so jim dali hrvaški sosedje pridevek 'kmetiški baroni'.

» »

zvesto vlekla na uho. Vse pesmi si nisem zapomnil, tudi ni ne v jezikovnem, ne v drugih ozirih vredna, da bi se uvrstila v sicer široke predale naše književnosti, prvi kitici pa še zdaj znam. Glasita se takole:»

»

Je maček na streh'
kaj dela, kaj prav'
Mijav, mijav, mijav!»

Je mucka pri njem,
ga boža po glav'.
Mijav, mijav, mijav!»

»

»Oh, ta je imenitna, ta pa je imenitna!« je pritrjeval Lazar z glavo. »Ta šči plje!»

Mačka je pesem v živo zadela; pospešil je korake in mahnil proti Gradišču. Obračal je pogled zdaj na levo

Na koncu seznama je najprej par s kodama 264 in 265:

» »

ki je namenjen označevanju delov besedila, za katere označevalec še ne ve, kam bi jih uvrstil – v končni verziji obeh vzorcev se ne pojavi več. Zadnji par, s koda-

ma 266 in 267, pa označuje dele besedila, ki naj jih internetna pregledovalnika prikažeta s črkami enake širine:

Ⓜ Ⓜ

Rothalom in Allahabadom. Ⓜ In tu je izračun, ki ga je določil Ⓜ Morning_Chronicle Ⓜ:

ⓂⓂ

Ⓜ Od Londona do Sueza skozi Mont-Cenis in Brindisi, z železnico in parnikom	7 dni
Od Sueza do Bombaya, s parnikom	13 dni
Od Bombaya do Kalkute, z železnico	3 dni
Od Kalkute do Hong_Konga (Kitajska), s parnikom ..	13 dni
Od Hong_Konga do Jokohame (Japonska), s parnikom .	6 dni
Od Jokohame do San_Francisca, s parnikom	22 dni
Od San_Francisca do New_Yorka, z železnico	7 dni
Od New_Yorka do Londona, s parnikom in železnico .	9 dni

Skupno: 80 dni

ⓂⓂⓂ

3.1.2 Zaglavja besedil

Ker bi bilo opravljanje večjega števila postopkov pri urejanju, popravljanju in označevanju nad besedili, če bi bilo vsako v svoji datoteki, še veliko bolj zamudno, nepregledno in neobvladljivo, se je avtor odločil, da bo zložil vsa besedila vsakega vzorca skupaj. Nastali sta dve datoteki, po vseh postopkih je prva velika slabih 25 megabajtov, druga pa, z oblikoslovnimi oznakami vred malo manj kot 7. Prvo se ravno še da obdelovati na računalniku z 32 megabajti pomnilnika in operacijskim sistemom Windows 95.

Na strežniku interneta mora biti vsako besedilo seveda v svoji datoteki, in če naj bodo vse te datoteke (tipa HTML) strojno generirane iz ene same v formatu urejevalnika EVA, je treba pri pretvorbi od nekod črpati tiste parametre o vsakem besedilu, ki so značilni zanj. Avtor se je odločil, da bodo shranjeni v posebnem zaglavju pred vsakim besedilom, kadar je le-to oblikoslovno označeno, pa še v kratkem dodatku na koncu. Primer iz prvega vzorca je zaglavje romana *Devet fantov in eno dekle*:

```

♦ /<file_name>devet_fa.html</file_name>
♦ /<file_number>12</file_number>
♦ /<file_abbreviation>9F</file_abbreviation>
♦ /<file_size_words>29136</file_size_words>
♦ /<file_creation_date>17. aprila 1998</file_creation_date>
♦ /<file_date_of_change>9. avgusta 1998</file_date_of_change>
♦ /<file_title>Gitica Jakopin: Devet fantov in eno dekle</file_title>
♦ /

```

```

☒

```

Gitica Jakopin

Devet fantov in eno dekle

```

☒

```

I

```

☒

```

☒ V našo hišo doma še danes vodijo močne, kar preveč strme stopnice; na desni strani jih drži staromodna, z železnimi palicami povezana ograja, in na vrhu je veranda. ☒ To je odkrita veranda, skromna, toplejša

Zaradi analogije z imeni oznak v jeziku SGML, ki so v angleškem jeziku, so tudi imena spremenljivk iz zaglavja v tem jeziku. Zaglavja pri drugem vzorcu so zelo podobna, le da imajo eno spremenljivko več:

```

♦ /<file_number>33</file_number>
♦ /<file_abbreviation>HJ</file_abbreviation>
♦ /<file_size_words>7959</file_size_words>
♦ /<file_creation_date>8. januarja 1998</file_creation_date>
♦ /<file_date_of_change>12. maja 1998</file_date_of_change>
♦ /<file_title>Ciril Kosmač: Hudo je, človek moj!</file_title>
♦ /
♦ / hudo.482
♦ \
♦ /<POS_tagged>
♦ /

```

```

☒

```

☒ Pismonoša Alojzij Klemenčič je danes tudi doživel
 Sme1 IOme1 IOme1 GPce A Vpr GLme
 svoj dan. ☒ Ko je Travnikarjevemu stricu srečno obesil
 ZSVPme4 Sme4 Vpo GPce PIOme3 Sme3 A GLme

Zaglavje je napisano v obliki komentarskih vrstic (to je vsaka vrstica v EVI, ki se začne z dvojnimi karom, presledkom in poševnico – levo ali desno). Vsaka nava-

dno vsebuje ime spremenljivke, njeno vrednost in oznako za konec (ime spremenljivke s poševnico spredej).

Določilo *hudo.482* v prejšnjem primeru pomeni številko strani v izvirniku, določilo `<POS_tagged>` pa, da je besedilo oblikoslovno označeno; na koncu datoteke bi seveda sledila še komentarska vrstica z določilom `</POS_tagged>`.

Ukaz v urejevalniku EVA, ki iz datoteke posameznega vzorca naredi desetine datotek tipa HTML, potrebuje le še pomožno datoteko z osnutkom, parametre besedila in njegovo vsebino pa najde v osnovni datoteki. Datoteka z osnutkom, oropana nebitvenih dodatkov in malo bolj na gosto napisana, bi bila lahko videti takole:

```
<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=ISO-8559-2">
<meta name="description" content="$file_title">
<title>Primož Jakopin, doktorska disertacija, priloga I: prvi vzorec,
datoteka $file_number </title> </head>
<body>
Primož Jakopin:<br>
<i>Zgornja meja entropije v slovenskih leposlovnih besedilih</i><br>
Doktorska disertacija. Priloga I - Prvi vzorec - datoteka $file_number
($file_size_words besed)<br><br>
<a href="index.html"><i>Seznam del v Prilogi I</i></a><br>
<hr>
$file_title<br>
<hr><br>
.....
<hr>
<p><font size=2>
<i>Stran je bila postavljena $file_creation_date in nazadnje spremenjena
$file_date_of_change</i></font></p>
</body>
</html>
```

Vsaka datoteka tipa HTML, ki bo nastala iz datoteke z vzorcem, bo imela na začetku prvi del datoteke z osnutkom do vrstice iz dvojnih karov, kjer pa bodo imena spremenljivk pri vsaki datoteki (označena so z znakom \$) nadomeščena z vrednostmi iz zaglavij ustreznega besedila. Na mesto vrstice dvojnih karov (v EVI navadno označuje konec strani) pride besedilo sámo, seveda ustrezno prekoderano, na konec vsake datoteke pa drugi del osnutka, spet z ustreznimi vrednostmi spremenljivk.

3.2 Označevanje gradnikov besedila

Od tega, do katere globine je besedilo v elektronski obliki označeno, je v največji meri odvisno, kaj se bo dalo z njim napraviti. Leposlovna besedila v svetovnih jezikih, predvsem seveda v angleškem jeziku, ki so dostopna preko interneta so navadno shranjena kot običajne besedilne datoteke (z imenskim podaljškom TXT), ki jih lahko urejamo s poljubnim urejevalnikom ali pa so, navadno s pro-

gramom, prevedena v obliko HTML. V prvem primeru so vrstice tudi v datoteki navadno zvesta podoba vrstic v tiskanem besedilu, z izpuščenimi mejami strani, v drugem pa internetna pregledovalnika sama skrbita za dolžine vrstic; označeni so le odstavki, med njimi pa prazne vrstice. Tuji besedilni korpusi, npr. *British National Corpus* (glej prilogo D na str. 201), so seveda označeni vsaj do ravni povedi, velikokrat, še posebej v zadnjem času, pa so označeni tudi oblikoslovno (angl. *part-of-speech*).

V primeru obeh vzorcev iz tega dela, zbranih predvsem za ugotavljanje entropije, se je bilo torej treba odločiti, kako daleč naj bosta označena, kaj naj bo osnovna enota besedila, ki še pride v poštev pri analizi. Za to, da morata biti oba vzorca označena, je poleg tujih zgledov seveda še več razlogov. Po eni strani vsako označevanje namreč prinese tudi tako ali drugačno preverjanje besedila; pri preverjanju pa vedno pride tudi do odpravljanja napak, ki umetno povečujejo entropijo. Po drugi strani pa je označevanje nujno za vsako resnejšo statistično analizo besedila in za uporabnost besedila še pri drugih, kasnejših znanstvenih raziskavah ali obdelavah. Kot osnovni gradnik se glede na vrsto besedila v obeh vzorcih kar sama vsiljuje poved kot najmanjša zaključena enota. Odstavek kot alternativna možnost je veliko ohlapnejši. Posebej v delih besedila, kjer je veliko dialogov s premim govorom, pa so meje odstavkov velikokrat tudi meje povedi.

Glede na razpoložljiv čas in možnosti se je avtor odločil, da bosta oba vzorca označena do ravni povedi in premega govora; kot je bilo že omenjeno v razdelku o dodatnih simbolih, so bili označeni tudi tisti deli besedila, ki ali niso slovenski (tuji citati, popačena slovenščina) ali pa sploh niso besedilo. Število posameznih gradnikov – odstavkov, povedi, premega govora in drugega – je navedeno v razdelku o črkah (str. 48).

3.2.1 Odstavki

Označitev odstavkov v besedilih načelno ni bila problematična. Besedila iz Hladnikove zbirke na internetu so že imela ustrezni oznaki za začetek in konec: <p> in </p>, pri drugih besedilih pa se je dalo orientirati ali s pomočjo praznih vrstic ali pa po zamiku na začetku prve vrstice odstavka. Označitev in reformatiranje odstavkov (na 57 znakov dolge vrstice) sta bila opravljena po predobdelavi besedil z ustreznimi makroprogrami urejevalnika EVA, podobnimi skriptom v jeziku PERL (glej prilogo D na str. 201), ki se navadno uporabljajo v ta namen v drugih okoljih. Pri obeh vzorcih je prišlo do napačnih označitev predvsem pri vrinjenih delih besedila, kot so npr. citirani verzi ali pesmim podoben govor (npr. v *Novi zavezi*). Nekaj se jih je dalo odkriti po mali začetnici na začetku takega vložka, nekaj pa le z ročnim pregledom sumljivih mest.

3.2.2 Povedi

Pri označitvi povedi je treba upoštevati bistveno več elementov kot pri odstavkih. Štiri besedila – Platonova *Država*, Orwellov *1984* ter Kosmačeva *Pomladni dan* in *Prazna ptičnica* – so imela konce povedi (znak s kodo 255) označene že prej (Jakopin in Bizjak 1997); tu je bilo treba dodati le oznake začetkov povedi na ustrezna mesta pred prvo naslednjo veliko začetnico. Pri drugih besedilih so bila kot konec povedi upoštevana tale štiri ločila: pika (.), klicaj (!), vprašaj (?) in dvo-pičje (:), če jim je sledila beseda, pisana z veliko začetnico. Pri pikah so bile upoštewane te kratice:

Tabela 6: Kratice, upoštewane pri strojnem prelomu na povedi

angl.	dol.	gen.	izr.	oz.	st.
arch.	dr.	gor.	l.r.	Ph. D.	str.
c. k.	etc.	hon.	mag.	pl.	Sv.
c. kr.	fr.	ilustr.	Mr.	plem.	sv.
c.kr.	g.	ing.	Mrs.	prof.	št.
dipl.	ga.	itd.	npr.	sl.	t.i.
doc.	gdč.	itn.	oec.	St.	t.j.

Beseda z veliko začetnico za kratico namreč ne pomeni vedno tudi začetka nove povedi. Nekaj kratic je bilo avtorju znanih že od prej, druge pa je poiskal med najpogostejšimi besedami, ki jim sledi pika. Pojavitve števil, za katerimi je pika, so bile pregledane ročno. Upoštevana je bila tudi omejitev, da gnezdenje povedi čez meje odstavkov ni dovoljeno – da se mora poved obvezno končati s koncem odstavka. Omejitev se je izkazala za nestvarno v približno desetih primerih. Večdelne kratice, npr. *c. kr.*, je avtor povezal v celoto tako, da je pri njih povsod nadomestil navaden presledek s trdim presledkom (koda 185), tako da so pri prelomu besedila na besede ostale skupaj kot ena enota. Skupaj je povezal tudi kratice in besede za njimi, kadar tvorijo pojem, ki ga poznamo kot celoto, predvsem pri večbesednih lastnih imenih, npr. *Sv. Lucija* ali *Hong Kong*.

3.2.3 Premi govor

Označitev premege govora je bila med zamudnejšimi opravili pri pripravi besedil. V manjšem številu primerov je bil premi govor označen z znaki, ki jih najdemo v knjigah: « » , ‘ ’ „ ” “ ”. Pri optičnem prepoznavanju znakov so bila nekatera izmed teh ločil napačno interpretirana kot vejice, >> in <<, pike in podobno. Avtor je programsko s posebnim znakom, dvojnim vprašajem, označil mesta v besedilu, kjer je za posamezne pare prihajalo do napačnega gnezdenja, in tista mesta potem ročno popravil.

V večjem delu je bil začetek in konec premega govora označen z istim znakom, dvojnimi narekovajem: " in tu je, spet zaradi pretežno optičnega vnosa, prihajalo do zelo velikega števila neujemanj pri gnezdenju – reda velikosti 1.500 napak za oba vzorca. Napako (npr. manjkajoči zaključek premega govora) je program žal lahko odkril šele pri naslednjem premem govoru in je bilo treba pri ročnem popravljanju iskati včasih daleč nazaj.

V marsikaterih tiskih pa je bil začetek premega govora označen s pomišljajem (–) na začetku odstavka, konec je pa ali bil ali pa manjkal, kot v spodnjem primeru iz *Pomladnega dneva*:

– Tudi nevesta gre, tudi nevesta! – sva se s sestro zagnala k mami, ki je že v sveži ruti in čistem predpasniku stala na pragu.

– Tudi nevesta? – je sklenila roke. – Joj, kje je medaljonček?

– V predalu šivalnega stroja, je povedala sestra.

– Hitro ponj!

A nihče se ni ganil. Vsi smo strmeli samo v gospo v črnini, ki se je bližala s težkimi, trudnimi koraki.

Postopek je bilo v tem primeru mogoče delno avtomatizirati s kontrolo gnezdenja in ročnimi popravki v drugi fazi.

3.2.4 Citati, vzdevki in drugo

Ker je predmet proučevanja le slovenski del leposlovnih besedil, je bilo treba označiti vse dele, ki niso bili v slovenskem jeziku. Pri drugem vzorcu, ki je bil večkrat cel ročno pregledan, so bili ti deli besedila označeni ročno, v prvem vzorcu pa v več fazah. Najprej je avtor med zelo pogostimi besednimi oblikami poiskal tiste, ki jih v slovenskem jeziku sploh ni ali so redke, v drugih jezikih, predvsem angleškem, nemškem, italijanskem, francoskem, srbskem in hrvaškem ter latinskem (veliko romanov je zgodovinskih), pa so zelo pogoste. Primeri so: *an, and, are, as, das, de, der, di, die, du, e, est, et, for, für, I, is, ist, la, non, of, quod, su, sunt, the, u, und in you*. V veliki večini primerov je bila njihova okolica citat v tujem jeziku in potem ročno označena. Delo s sorazmerno največ tujimi citati je *Thabiti Kumi Ivana Preglja* – dolgo je 9.362 besed; v njem je kar 93 latinskih citatov s 521 besedami ali skoraj 6 % celote. Nekaj preostalih citatov v drugih jezikih je avtor odkril pri pregledovanju besedil zaradi kakih razlogov, ki niso bili povezani z iskanjem citatov, še največ pa pri popravljanju napačnih besed.

Citati v slovenskem jeziku in vzdevki so bili v knjižnih verzijah pri delih, kjer je bil premi govor označen s simboloma < in >, navadno označeni z drugačnimi navednicami: , ‘ „ “ ’ “ ”. Kadar je bil označen z dvojnimi narekovajem (") spredaj

in zadaj, so bili citati in vzdevki običajno v enojnih navednicah ('). Odločitev, kdaj je del besedila vzdevek in ne citat, ni bila vedno povsem zanesljiva; strogega kriterija, kdaj je nekaj res vzdevek (angl. *so-called*) in ne citat (angl. *citation*), namreč ni. Primer:

»Po ulicah so hrumela oklepna vozila, redki ljudje so dejali, da so »'naša'«. »Čigava naša?« »Odgovora nisva iskala, čeravno bi bil koristen.«

Poleg citatov, slovenskih in tujih, ter vzdevkov so bili posebej označeni še deli besedila, ki jih ni mogoče šteti za besedilo, so pa vseeno v njem. Sem sodijo predvsem številke opomb in podobne oznake, npr. številke vrstic v poglavjih *Nove zaveze*. Ni jih bilo težko strojno označiti, ker so imeli spredaj zvezdico. Primer:

»*12« »Po preselitvi v Babilon se je Jojahínu rodil Šealtiel, Šealtielu Zerubabél, »*13« Zerubabélu Abihúd, Abihúdu Eljakím, Eljakímu Azór, »*14« Azóru Cadók, Cadóku Ahím, Ahímu Eliúd, »*15« Eliúdu Eleázar, Eleázaru Matán, Matánu Jakob, »*16« Jakobu pa se je rodil Jožef, mož Marije, iz katere je bil rojen Jezus, ki se imenuje Kristus.«

3.3 Vprašanje napak

Pri sestavljanju vsake besedilne zbirke se pojavi tudi vprašanje napak. Napak v besedilih je več vrst, začenši z avtorjevimi (pri znanih imenih so sicer redke, so pa), do tiskarskih, tistih, ki nastanejo pri prenosu dela v elektronsko obliko, s tipkanjem ali optičnim prepoznavanjem (angl. *OCR - Optical Character Recognition*) in do tistih, ki so nastale pri prenosu iz enega računalniškega formata v drug (npr. deljaji na koncu vrstic, ki se pri reformatiranju ohranijo sredi vrstic, namesto da bi izginili).

Stališča do napak v besedilnih zbirkah so različna – poleg klasičnih nazorov, po katerih se je treba proti napakam boriti neusmiljeno in do zadnjega diha, do popolnoma nasprotnih, npr.: »*Errors are integral part of text. If you correct them, you lose information*» (J. M. Sinclair v predavanju na delavnici: *TELR Birmingham Workshop* oktobra 1995 – Jakopin 1996a). Pri slednjem gre za odgovor na vprašanje, zakaj je v besedilnem korpusu *Bank of English* – zdaj ima že 329.000.000 besed – toliko napak (najpogostejša, *hte*, ima frekvenco pribl. 4.000). Tudi druge velike besedilne zbirke so precej nasmetene, kar gre po mnenju avtorja pripisati predvsem dejstvu, da so bili viri za njihovo postavitve zelo omejeni, pritisk in želje pa velike, tako da je v dilemi: ali količina ali kakovost zmagala prva. Tako je zelo zgovorno dejstvo, da avtorji doslej največje raziskave o entropiji angleškega jezika (Brown idr. 1992a), ki je zajela 583 milijonov besed, v svojem

prispevku o napakah niso napisali niti besedice. Pri manj pomembnih besedilih, raznih zapisnikih in podobnem, napake niti niso tako moteče, v leposlovnih besedilih, ki so namenjena branju, tudi v elektronski obliki, pa zelo kvarijo celo podobo. V pripravo leposlovnih del je vloženega vedno zelo veliko truda – več korektur celega besedila že v rokopisu, potem še vsaj dve korekturi iz tiskarne – kar vse izniči površno pripravljena elektronska izdaja. Upoštevati pa je treba konec koncev tudi dejstvo, da napake umetno povečujejo entropije in zmanjšujejo težo vsake statistične raziskave.

Tako se je avtor odločil, da bo besedila kolikor se le da očistil, še posebej drugi vzorec. Pri tem je seveda upošteval vire, ki so mu bili na razpolago, in omejen čas raziskave.

3.3.1 Napake pri prepoznavanju

Besedila, predvsem iz Hladnikove zbirke na internetu in večji del drugega vzorca, so bila v elektronsko obliko prenesena pretežno s pomočjo optičnega prepoznavanja (OCR). Avtorji prenosa so v uvodu vedno poudarili, da so besedilo po prepoznavanju še natančno prebrali in popravili, da pa ne odgovarjajo za morebitne napake, ki bi še ostale. Žal jim je za branje in popravljanje zmanjkalo časa, saj je v skoraj vseh delih napak kar mrgolelo. Poleg napačnih malih črk v besedah, ki se jih ne da odkriti s kakimi preprostimi prijemi, je bilo precej napak še pri ločilih – presledki pred vejicami in pikami, narazen pisane besede, narobe interpretirane navednice in napačne velike začetnice ter števila (npr. 1980, kjer je na zadnjem mestu velika črka O). Te skupine napak je bilo v veliki meri mogoče odkriti programsko, z iskanjem velikih začetnic, obdanih z malimi, ali kombinacij črka–števka in števka–črka. Najpogostejša napaka te vrste je bila beseda *ljudje*, napačno pisana kot *Ijudje*. Njena frekvenca je znašala približno 50.

3.4 Zbirka besed iz SSKJ

Večina napak v besedilih je bila v obliki narobe napisanih besed. Predvsem pogošte besede so za seboj potegnile cele grozde napačnih satelitov. Primeri so besede *samostan*, *tudi* in *zadovoljen*:

Tabela 7: Napake, nastale iz besed *samostan*, *tudi* in *zadovoljen*

samosta	tud	zadooljen
samostamu	tudij	zadovljen
samostanju	tudk	zadovljno
samotan	tui	zadovojen
samotanom	tuid	

Dostikrat se je tudi zgodilo, da so se napake v abecednem seznamu vseh besednih oblik zbrale okoli nenavadne ali nemogoče začetne kombinacije črk (v oklepajih so navedene ali pravilne besedne oblike ali pa kontekst napake):

Tabela 8: Napake na *ps*, *pt* in *vo*

psoebno	(posebno)	voril	(... je go voril ...)
psolvnem	(poslovnem)	vorili	(... nista go vorili radi o ...)
psomladi	(spomladi)	voriva	(Pa go voriva pametno!)
psotala	(postala)	vorčekravnega	(... njegovega vorčekravnega sina ...)
psotavil	(postavil)	vosjki	(... vojski ...)
psoznamo	(spoznamo)	vote	(... sta si ogledala ex vote.)
ptolej	(potlej)	vouz	(... Comment allez vous?)
ptomci	(potomci)		
ptreboval	(potreboval)		
ptrljage	(prtljage)		

Da bi bilo mogoče preveriti obsežna seznama besednih oblik (175.000 različnih besednih oblik v prvem vzorcu in 37.000 v drugem) je bilo treba najti kar največjo zbirko vseh možnih slovenskih besednih oblik. Avtor je imel na razpolago Slovar slovenskega knjižnega jezika (SSKJ 1994); sodeloval je pri prenosu v elektronsko obliko kot avtor programa za urejanje in optično prepoznavanje. Slovar obsega 93.151 gesel, med katerimi prevladujejo samostalniki, pridevniki in glagoli. Te besedne vrste dajo iz osnovnih besednih oblik (lem) tudi največ izpeljank.

Tabela 9: Število besed in besednih oblik v SSKJ po vrstah

	paradigem	osnovnih besednih oblik	izpeljanih oblik	različ. izpelj. oblik
samostalnik	59	51.790	941.788	442.451
pridevnik	18	21.674	1.300.644	265.840
glagol	59	18.124	1.245.244	377.837
Skupaj	136	91.588	3.487.676	1.075.779

Pri oblikovanju paradigem in pri klasifikaciji besednih lem naglasi niso bili upoštevani. Z njimi bi bilo število paradigem (in število besednih izpeljank) še nekoliko večje. Razlog je bil predvsem v tem, da so v pisanih besedilih, kot bo videti tudi v razdelku o statistiki črk, naglasi redki.

3.4.1 Samostalniki

Pri samostalnikih je bilo 24 paradigem za moški spol (drugačen četrti sklon pri samostalnikih za živa bitja število precej poveča), 21 za ženski in 14 za srednji, število lem pa 22.459 za moški, 21.156 za ženski in 8.175 za srednji. Pri gradnji paradigem je bil upoštevan predvsem slovnični del uvoda v SSKJ (SSKJ 1994). Primer samostalniških paradigem je naveden v tabeli 10.

Tabela 10: Tri samostalniške paradigme ženskega spola

ž13⁻, i, i, -, i, jo[•] i, i, im, i, ih, mi[•] i, i, mač[•] ima, i, ih, mač[•] ima
 ž14⁻, e, i, -, i, i, jo[•] e, ač[•] Δev, am, e, ah, ami[•] iče, ač[•] Δev, ama, iče, ah, ama
 ž15⁻, e, i, -, i, i, jo[•] e, Δev, am, e, ah, ami[•] i, Δev, ama, i, ah, ama

Paradigmo ž13 ima npr. samostalnik *podlaket*, ž14 *cerkev*, ž15 pa *breskev*. Tabela paradigem je urejena kot podatkovna zbirka, kjer ima vsak zapis 4 polja. V prvem je ime oz. šifra paradigme, sledijo pa 3 polja z definicijami za vseh 6 sklonov ednine, množine in dvojine. S pomišljajem je navedena uporaba leme, drugače pa le dodatki na roditeljsko osnovo; alternativne možnosti so ločene z znakom &.

Prva vrstica za lemo *podlaket*, ki ima roditeljsko osnovo *podlakt*, konkretno pomeni, da nastane prvi sklon ednine kar z uporabo leme (-), drugi sklon z dodatkom -i (i) na roditeljsko osnovo (*podlakti*), tretji je enak drugemu, četrti prvemu, peti spet drugemu, za šestega pa je treba roditeljski osnovi dodati -jo (s *podlaktjo*). Množina ni problematična, pri dvojini pa sta posebnost le dajalnik in orodnik, pri katerih sta dve možnosti – obrazili -ma in -ima na roditeljsko osnovo (*podlaktma* in *podlaktima*). Izsek iz slovarja lem s podatki o paradigmah in roditeljski osnovi je naveden v tabeli 11.

Tabela 11: 21 samostalniških lem s paradigmo in roditeljsko osnovo

ž13 [•] •žlobudravost	ž13 [•] ΔΔt [•] •podlaket	ž15 [•] ΔΔv [•] •bistritev
ž13 [•] •žolčnost	ž13 [•] ΔΔt [•] •podlaket	ž15 [•] ΔΔv [•] •bit ev
ž13 [•] •žrtvovanost	ž13 [•] ΔΔΔnJ [•] •povodenJ	ž15 [•] ΔΔv [•] •blagoslovitev
ž13 [•] •žuljavost	ž13 [•] ΔΔΔnJ [•] •sovodenJ	ž15 [•] ΔΔv [•] •blažitev
ž13 [•] ΔΔn [•] •raven	ž14 [•] ΔΔv [•] •cerkev	ž15 [•] ΔΔv [•] •bogatitev
ž13 [•] ΔΔt [•] •nadlaket	ž15 [•] ΔΔv [•] •apnitev	ž15 [•] ΔΔv [•] •bratev
ž13 [•] ΔΔt [•] •nadlaket	ž15 [•] ΔΔv [•] •bet ev	ž15 [•] ΔΔv [•] •breskev

Tudi ta zbirka je pripravljena kot podatkovna zbirka, tokrat s tremi polji. V prvem je šifra paradigme, v drugem navodilo za tvorjenje roditeljske osnove, v tretjem pa lema. Navodilo je lahko prazno, kar pomeni, da je roditeljska osnova kar enaka lemi, lahko pa ima najprej n simbolov za odstranitev n končnih črk leme, ki jim sledi še obrazilo, ki ga je treba nataktniti na tako dobljen krn. Pri lemi *breskev* je npr. treba odstraniti zadnji dve črki in dodati v, da dobimo roditeljsko osnovo *breskv*. V tabeli 12 so navedene strojno dobljene izpeljanke iz štirih lem: *žolčnost*, *podlaket*, *cerkev* in *breskev*. Vsaki izpeljanki sledi še navodilo, kako iz nje dobiti lemo, poleg tega pa še oblikoslovna oznaka. Ta je pri samostalnikih štirimestna – najprej velika črka S (za *samostalnik*), potem spol (*m*, *ž* ali *s*), število (*e*, *p* in *d*) ter sklon (1-6). Primer je *breskvami*. Prvi del deskriptorja (do velike črke) – 4ev zahteva, da je treba izpeljanki odvzeti na koncu štiri črke (*bresk*) in dodati *ev*, da dobimo lemo – *breskev*. Drugi del deskriptorja pove, da gre za samostalnik ženskega spola v ednini in orodniku. Veliko izpeljank je tudi enakih – npr. za roditeljski in mestni ednine, zato je število različnih izpeljanih besednih oblik pri samostalnikih za več kot pol manjše od števila vseh izpeljank (442.451 proti 941.788).

Tabela 12: Programsko tvorjene izpeljanke 4 samostalniških lem ženskega spola

žolčnost	Sže1	podlaket	Sže1	cerkev	Sže1	cerkvah	3evSžd5
žolčnosti	1Sže2	podlakti	2etSže2	cerkve	2evSže2	cerkvama	4evSžd6
žolčnosti	1Sže3	podlakti	2etSže3	cerkvi	2evSže3	breskev	Sže1
žolčnost	Sže4	podlaket	Sže4	cerkev	Sže4	breskve	2evSže2
žolčnosti	1Sže5	podlakti	2etSže5	cerkvi	2evSže5	breskvi	2evSže3
žolčnostjo	2Sže6	podlaktjo	3etSže6	cerkvijo	4evSže6	breskev	Sže4
žolčnosti	1Sžp1	podlakti	2etSžp1	cerkve	2evSžp1	breskvi	2evSže5
žolčnosti	1Sžp2	podlakti	2etSžp2	cerkva	2evSžp2	breskvijo	4evSže6
žolčnostim	2Sžp3	podlaktim	3etSžp3	cerkev	Sžp2	breskve	2evSžp1
žolčnosti	1Sžp4	podlakti	2etSžp4	cerkvam	3evSžp3	breskev	Sžp2
žolčnostih	2Sžp5	podlaktih	3etSžp5	cerkve	2evSžp4	breskvam	3evSžp3
žolčnostmi	2Sžp6	podlaktmi	3etSžp6	cerkvah	3evSžp5	breskve	2evSžp4
žolčnosti	1Sžd1	podlakti	2etSžd1	cerkvami	4evSžp6	breskvah	3evSžp5
žolčnosti	1Sžd2	podlakti	2etSžd2	cerkvi	2evSžd1	breskvami	4evSžp6
žolčnostma	2Sžd3	podlaktma	3etSžd3	cerkve	2evSžd1	breskvi	2evSžd1
žolčnostima	3Sžd3	podlaktima	4etSžd3	cerkva	2evSžd2	breskev	Sžd2
žolčnosti	1Sžd4	podlakti	2etSžd4	cerkev	Sžd2	breskvama	4evSžd3
žolčnostih	2Sžd5	podlaktih	3etSžd5	cerkvama	4evSžd3	breskvi	2evSžd4
žolčnostma	2Sžd6	podlaktma	3etSžd6	cerkvi	2evSžd4	breskvah	3evSžd5
žolčnostima	3Sžd6	podlaktima	4etSžd6	cerkve	2evSžd4	breskvama	4evSžd6

3.4.2 Pridevniki

Pridevniki so besedna vrsta z največjim številom pregibnih oblik – iz 21.674 lem je nastalo 1.300.644 izpeljank (265.840 različnih); vseh pridevniških paradigem je bilo 18. Velikemu številu oblik botrujejo predvsem trije spoli pri vsakem pridevniku in pri večini še stopnjevanje. Primer pridevniške paradigme je v tabeli 13.

Tabela 13: Pridevniška paradigma s primerom

```

p4*
,-ega,-emu,&-ega,-em,-im*-i,-ih,-im,-e,-ih,-imi*-a,-ih,-ima,-a,-ih,-ima
>,-e,-i,-o,-i,-o*-e,-ih,-im,-e,-ih,-imi*-i,-ih,-ima,-i,-ih,-ima
<,-ega,-emu,<,-em,-im*-a,-ih,-im,-a,-ih,-imi*-i,-ih,-ima,-i,-ih,-ima
=i,-ega,-emu,=i&=ega,=em,=im*=i,=ih,=im,=e,=ih,=imi*=a,=ih,=ima,=a,=ih,=ima
=a,=e,=i,=o,=i,=o*=e,=ih,=im,=e,=ih,=imi*=i,=ih,=ima,=i,=ih,=ima
=e,-ega,-emu,=e,=em,=im*=a,=ih,=im,=a,=ih,=imi*=i,=ih,=ima,=i,=ih,=ima
naJ=i,naJ=ega,naJ=emu,naJ=i&naJ=ega,naJ=ea,naJ=im*naJ=i,naJ=ih,naJ=i,naJ=e,
naJ=ih,naJ=imi*naJ=a,naJ=ih,naJ=ima,naJ=a,naJ=ih,naJ=ima
naJ=a,naJ=e,naJ=i,naJ=o,naJ=i,naJ=o*naJ=e,naJ=ih,naJ=im,naJ=e,naJ=ih,naJ=imi*
naJ=i,naJ=ih,naJ=ima,naJ=i,naJ=ih,naJ=ima
naJ=e,naJ=ega,naJ=emu,naJ=e,naJ=em,naJ=im*naJ=a,naJ=ih,naJ=im,naJ=a,naJ=ih,
naJ=imi*naJ=i,naJ=ih,naJ=ima,naJ=i,naJ=ih,naJ=ima
-i
-o
+e,=e
naJ+e,naJ=e
p4*hud*a*o*ΔJš1
    
```

Paradigma je precej obsežnejša od samostalniške, opis pridevnika (v našem primeru *hud*) pa kratek. Najprej je šifra njegove paradigme, sledi pridevnik sam, v moškem spolu, za njim navodilo za oblikovanje osnovne izpeljanke ženskega spola (in tvorjenje rodilniške osnove obenem), v četrtem polju za srednji spol, v petem pa za tvorjenje primernika. Pri paradigmi je v prvi vrstici njena šifra, v drugi navodilo za izpeljanke ednine, množine in dvojine moškega spola za vseh 6 sklonov (vmes so vejice), v tretji podobno za ženski spol in v četrti za srednji. Prazno navodilo (za prvi sklon in prvo varianto četrtega) pove, da je treba vzeti kar pridevnik sam, pomen dodatnih simbolov pa je razložen v tabeli 14:

Tabela 14: Simboli v pridevniških paradigmah

- rodilniška osnova
- & sledi še ena varianta za isti sklon
- < osnovna oblika za ženski spol
- > osnovna oblika za srednji spol
- = primerniška osnova (primernik moškega spola brez *i* na koncu)
- + primerniška osnova brez končnega *š*
- ø oblike ni (npr. pri paradigmah, kjer ni določne oblike)

V opisu podatkov o paradigmi sledijo v 5., 6. in 7. vrstici navodila za vse tri spole primerniških oblik, v 8., 9. in 10. pa še za presežnik. Vrstice od 5. do 10. nastopajo samo v tistih paradigmah, kjer obstaja stopnjevanje pridevnikov. Sledi prazna vrstica, za njo pa opis določne oblike, pod njim osnovne oblike prislova ter še, če sta, obeh višjih prislovnih stopenj.

V tabeli 15 na naslednji strani je naveden primer strojnega tvorjenja izpeljank za pridevnik *hud*. Iz leme je nastalo 172 izpeljanih besednih oblik, od tega 36 različnih. Na levi strani vsakega stolpca so, podobno kot pri samostalnikih, navedene izpeljanke, na desni pa deskriptor, ki opisuje pot do leme in navaja oblikoslovno oznako za to izpeljanko. Oblikoslovne oznake so spet štiri, pet ali šestmestne – najprej *P* za vrsto (*pridevnik*), potem koda za spol (*m*, *ž*, *s*), število (*e*, *p*, *d*) in sklon (*1-6*), na koncu pa, po potrebi, še *i* za določno obliko, *j* za primernik in *jj* za presežnik (Jakopin in Bizjak 1997). Tudi opis leme je tokrat včasih razširjen – pri presežniških oblikah je najprej navedeno, kaj naj se zgodi s koncem izpeljanke, da bi dobili lemo, potem, za dvopičjem, pa še, kaj je treba napraviti na začetku izpeljanke. Primer je npr. *najhujšima 5d:3Pžd6jj* – na koncu izpeljanke je treba najprej odrezati 5 črk (5), dodati *d*, potem pa še spredaj odrezati tri črke (3, naj). *Pžd6jj* pove, da gre za pridevnik ženskega spola v dvojini, orodniku (6) in v presežniku (*jj*).

Tabela 15: Izpeljanke iz pridevnika *hud*

hud	Pme1	huda	1Psp1	hujsi	3dPzd1j	najhujsa	3d:3Pze1jj
hudega	3Pme2	hudih	2Psp2	hujsih	4dPzd2j	najhujsse	3d:3Pze2jj
hudemu	3Pme3	hudim	2Psp3	hujsima	5dPzd3j	najhujsi	3d:3Pze3jj
hud	Pme4	huda	1Psp4	hujsi	3dPzd4j	najhujsso	3d:3Pze4jj
hudega	3Pme4	hudih	2Psp5	hujsih	4dPzd5j	najhujsi	3d:3Pze5jj
hudem	2Pme5	hudimi	3Psp6	hujsima	5dPzd6j	najhujsso	3d:3Pze6jj
hudim	2Pme6	hudi	1Psd1	hujsse	3dPselj	najhujsse	3d:3Pzp1jj
hudi	1Pmp1	hudih	2Psd2	hujssega	5dPse2j	najhujsih	4d:3Pzp2jj
hudih	2Pmp2	hudima	3Psd3	hujssemu	5dPse3j	najhujsim	4d:3Pzp3jj
hudim	2Pmp3	hudi	1Psd4	hujsse	3dPse4j	najhujsse	3d:3Pzp4jj
hude	1Pmp4	hudih	2Psd5	hujssem	4dPse5j	najhujsih	4d:3Pzp5jj
hudih	2Pmp5	hudima	3Psd6	hujsim	4dPse6j	najhujsimi	5d:3Pzp6jj
hudimi	3Pmp6	hujsi	3dPme1j	hujsa	3dPsp1j	najhujsi	3d:3Pzd1jj
huda	1Pmd1	hujssega	5dPme2j	hujsih	4dPsp2j	najhujsih	4d:3Pzd2jj
hudih	2Pmd2	hujssemu	5dPme3j	hujsim	4dPsp3j	najhujsima	5d:3Pzd3jj
hudima	3Pmd3	hujsi	3dPme4j	hujsa	3dPsp4j	najhujsi	3d:3Pzd4jj
huda	1Pmd4	hujssega	5dPme4j	hujsih	4dPsp5j	najhujsih	4d:3Pzd5jj
hudih	2Pmd5	hujssem	4dPme5j	hujsimi	5dPsp6j	najhujsima	5d:3Pzd6jj
hudima	3Pmd6	hujsim	4dPme6j	hujsi	3dPsd1j	najhujsse	3d:3Pse1jj
huda	1Pze1	hujsi	3dPmp1j	hujsih	4dPsd2j	najhujssega	5d:3Pse2jj
hude	1Pze2	hujsih	4dPmp2j	hujsima	5dPsd3j	najhujssemu	5d:3Pse3jj
hudi	1Pze3	hujsim	4dPmp3j	hujsi	3dPsd4j	najhujsse	3d:3Pse4jj
hudo	1Pze4	hujsse	3dPmp4j	hujsih	4dPsd5j	najhujssem	4d:3Pse5jj
hudi	1Pze5	hujsih	4dPmp5j	hujsima	5dPsd6j	najhujsim	4d:3Pse6jj
hudo	1Pze6	hujsimi	5dPmp6j	najhujsi	3d:3Pme1jj	najhujsa	3d:3Psp1jj
hude	1Pzp1	hujsa	3dPmd1j	najhujssega	5d:3Pme2jj	najhujsih	4d:3Psp2jj
hudih	2Pzp2	hujsih	4dPmd2j	najhujssemu	5d:3Pme3jj	najhujsim	4d:3Psp3jj
hudim	2Pzp3	hujsima	5dPmd3j	najhujsi	3d:3Pme4jj	najhujsa	3d:3Psp4jj
hude	1Pzp4	hujsa	3dPmd4j	najhujssega	5d:3Pme4jj	najhujsih	4d:3Psp5jj
hudih	2Pzp5	hujsih	4dPmd5j	najhujssem	4d:3Pme5jj	najhujsimi	5d:3Psp6jj
hudimi	3Pzp6	hujsima	5dPmd6j	najhujsim	4d:3Pme6jj	najhujsi	3d:3Psd1jj
hudi	1Pzd1	hujsa	3dPze1j	najhujsi	3d:3Pmp1jj	najhujsih	4d:3Psd2jj
hudih	2Pzd2	hujsse	3dPze2j	najhujsih	4d:3Pmp2jj	najhujsima	5d:3Psd3jj
hudima	3Pzd3	hujsi	3dPze3j	najhujsim	4d:3Pmp3jj	najhujsi	3d:3Psd4jj
hudi	1Pzd4	hujsso	3dPze4j	najhujsse	3d:3Pmp4jj	najhujsih	4d:3Psd5jj
hudih	2Pzd5	hujsi	3dPze5j	najhujsih	4d:3Pmp5jj	najhujsima	5d:3Psd6jj
hudima	3Pzd6	hujsso	3dPze6j	najhujsimi	5d:3Pmp6jj	hudi	1Pme1i
hudo	1Pse1	hujsse	3dPzp1j	najhujsa	3d:3Pmd1jj	hudi	1Pme4i
hudega	3Pse2	hujsih	4dPzp2j	najhujsih	4d:3Pmd2jj	hudo	1A
hudemu	3Pse3	hujsim	4dPzp3j	najhujsima	5d:3Pmd3jj	huje	2doAj
hudo	1Pse4	hujsse	3dPzp4j	najhujsa	3d:3Pmd4jj	hujsse	3doAj
hudem	2Pse5	hujsih	4dPzp5j	najhujsih	4d:3Pmd5jj	najhuje	2do:3Aj
hudim	2Pse6	hujsimi	5dPzp6j	najhujsima	5d:3Pmd6jj	najhujsse	3do:3Aj

3.4.3 Glagoli

Pregibanje pri glagolih je še bolj zapleteno kot pri pridevniki. Poleg običajnih glagolskih oblik, ki nastopajo pri vseh glagolih, najdemo pri mnogih glagolih še deležja, ki se pregibajo po predpisih dveh pridevniških paradigem (8. in 9.).

Primer je glagol *videti* v tabeli 16.

Tabela 16: Glagolska paradigma s primerom

g7•
~eti
-im,-iš,-i--imo,-ite,-i jo--iva,-ite,-ita
-i--imo,-ite--iva,-ita
~el,~ela,~elo~eli,~ele,~ela~ela,~eli,~eli
~et
p8•
,-ega,-emu,&-ega,-em,-im~i,-ih,-im,-e,-ih,-imi~a,-ih,-ima,-a,-ih,-ima
>,-e,-i,-o,-i,-o~e,-ih,-im,-e,-ih,-imi~i,-ih,-ima,-i,-ih,-ima
<,-ega,-emu,<,-em,-im~a,-ih,-im,-a,-ih,-imi~i,-ih,-ima,-i,-ih,-ima
-i
-o
p9•
,-ega,-emu,&-ega,-em,-im~i,-ih,-im,-e,-ih,-imi~a,-ih,-ima,-a,-ih,-ima
>,-e,-i,-o,-i,-o~e,-ih,-im,-e,-ih,-imi~i,-ih,-ima,-i,-ih,-ima
<,-ega,-emu,<,-em,-im~a,-ih,-im,-a,-ih,-imi~i,-ih,-ima,-i,-ih,-ima
-i
-e
g7•videti•ΔΔim•se•A videvši,č videč a e,N viden a o

Primer je glagol *videti* v tabeli 16. Poleg glagolske paradigme (7) sta navedeni še obe pridevniški paradigmi, ki lahko nastopata – 8. za deležnik na *-n* in 9. za deležnik na *-č*. Opis konkretnega glagola ima 5 podatkovnih polj: v prvem je kot po navadi šifra ustrezne paradigme, v drugem glagol sam, v tretjem navodilo za tvorjenje sedanjiške osnove (v tem primeru je treba odvzeti na koncu tri črke in za prvo osebo ednine dodati *im*), v četrtem informacija, ali je glagol lahko tudi prehodan (*se*), v petem pa morebitni podatki o dodatnih oblikah, ki jih je mogoče izpeljati iz glagola. V tem primeru je to deležje *videvši* in dva deležnika, na *-č* in na *-n*, ki imata še podatke o ženskem in srednjem spolu; stopnjevanje pri glagolskih izpeljankah ne pride v poštev. Opis paradigme je šestvrstičen: v prvi vrstici je kot po navadi šifra paradigme, v drugi navodilo za tvorjenje nedoločniške osnove, v tretji navodila za oblike vseh treh števil in oseb (ednina, množina, dvojina, prva, druga in tretja oseba) sedanjega časa, v četrti za velelnike vseh treh števil (v ednini za drugo osebo, v množini in dvojini za drugo in tretjo), v peti za deležnike na *-l*, za vsa tri števila in za vse tri spole, v šesti vrstici pa za namenilnik. Pomišljaj (–) v opisu označuje sedanjiško osnovo, ki jo dobimo v tretjem polju pri lemi, vijuga (~) pa nedoločniško osnovo, ki jo dobimo s pomočjo leme in navodila v drugi vrstici paradigme.

Paradigmi *p8* in *p9* sta taki kot pri pridevnikih in tudi zelo podobni – razlikujeta se le v obliki prislova na koncu. V tabeli 17 so prikazane vse izpeljanke glagola iz prejšnjega primera (*videti*). Izpeljank je 252, od tega 40 različnih. Osebe v oblikoslovnih oznakah so označene z a (prva), b (druga) in c (tretja); oznake so podrobneje razložene v članku (Jakopin in Bizjak 1997).

Tabela 17: Izpeljanke iz glagola *videti* s podatki o lemi in oblikoslovnimi oznakami

videti	GNE	videče	1GČzp1	videča	1PČze1	videna	1GNmd1	videni	1PNmp1
vidim	2etiGae	videčih	2GČzp2	videče	1PČze2	videnih	2GNmd2	videnih	2PNmp2
vidiš	2etiGbe	videčim	2GČzp3	videči	1PČze3	videnima	3GNmd3	videnim	2PNmp3
vidi	1etiGce	videče	1GČzp4	videčo	1PČze4	videna	1GNmd4	videne	1PNmp4
vidimo	3etiGap	videčih	2GČzp5	videči	1PČze5	videnih	2GNmd5	videnih	2PNmp5
vidite	3etiGbp	videčimi	3GČzp6	videčo	1PČze6	videnima	3GNmd6	videnimi	3PNmp6
vidijo	3etiGcp	videči	1GČzd1	videče	1PČzp1	videna	1GNze1	videna	1PNmd1
vidiva	3etiGad	videčih	2GČzd2	videčih	2PČzp2	videne	1GNze2	videnih	2PNmd2
vidita	3etiGbd	videčima	3GČzd3	videčim	2PČzp3	videni	1GNze3	videnima	3PNmd3
vidita	3etiGcd	videči	1GČzd4	videče	1PČzp4	videno	1GNze4	videna	1PNmd4
vidi	1etiGVbe	videčih	2GČzd5	videčih	2PČzp5	videni	1GNze5	videnih	2PNmd5
vidimo	3etiGVap	videčima	3GČzd6	videčimi	3PČzp6	videno	1GNze6	videnima	3PNmd6
vidite	3etiGVbp	videče	1GČse1	videči	1PČzd1	videne	1GNžp1	videna	1PNžel1
vidiva	3etiGVad	videčega	3GČse2	videčih	2PČzd2	videnih	2GNžp2	videne	1PNžel2
vidita	3etiGVbd	videčemu	3GČse3	videčima	3PČzd3	videnim	2GNžp3	videni	1PNžel3
videl	1tiGLme	videče	1GČse4	videči	1PČzd4	videne	1GNžp4	videno	1PNžel4
videla	2tiGLže	videčem	2GČse5	videčih	2PČzd5	videnih	2GNžp5	videni	1PNžel5
videlo	2tiGLse	videčim	2GČse6	videčima	3PČzd6	videnimi	3GNžp6	videno	1PNžel6
videli	2tiGLmp	videča	1GČsp1	videče	1PČse1	videni	1GNžd1	videne	1PNžp1
videle	2tiGLzp	videčih	2GČsp2	videčega	3PČse2	videnih	2GNžd2	videnih	2PNžp2
videla	2tiGLsp	videčim	2GČsp3	videčemu	3PČse3	videnima	3GNžd3	videnim	2PNžp3
videla	2tiGLmd	videča	1GČsp4	videče	1PČse4	videni	1GNžd4	videne	1PNžp4
videli	2tiGLzd	videčih	2GČsp5	videčem	2PČse5	videnih	2GNžd5	videnih	2PNžp5
videli	2tiGLsd	videčimi	3GČsp6	videčim	2PČse6	videnima	3GNžd6	videnimi	3PNžp6
videt	iGNA	videči	1GČsd1	videča	1PČsp1	videno	1GNse1	videni	1PNžd1
videvši	A	videčih	2GČsd2	videčih	2PČsp2	videnega	3GNse2	videnih	2PNžd2
videč	GČme1	videčima	3GČsd3	videčim	2PČsp3	videnemu	3GNse3	videnima	3PNžd3
videčega	3GČme2	videči	1GČsd4	videča	1PČsp4	videno	1GNse4	videni	1PNžd4
videčemu	3GČme3	videčih	2GČsd5	videčih	2PČsp5	videnem	2GNse5	videnih	2PNžd5
videč	GČme4	videčima	3GČsd6	videčimi	3PČsp6	videnim	2GNse6	videnima	3PNžd6
videčega	3GČme4	videči	1GČmeli	videči	1PČsd1	videna	1GNsp1	videno	1PNse1
videčem	2GČme5	videče	A	videčih	2PČsd2	videnih	2GNsp2	videnega	3PNse2
videčim	2GČme6	videč	PČme1	videčima	3PČsd3	videnim	2GNsp3	videnemu	3PNse3
videči	1GČmp1	videčega	3PČme2	videči	1PČsd4	videna	1GNsp4	videno	1PNse4
videčih	2GČmp2	videčemu	3PČme3	videčih	2PČsd5	videnih	2GNsp5	videnem	2PNse5
videčim	2GČmp3	videč	PČme4	videčima	3PČsd6	videnimi	3GNsp6	videnim	2PNse6
videče	1GČmp4	videčega	3PČme4	videči	1PČmeli	videni	1GNsd1	videna	1PNsp1
videčih	2GČmp5	videčem	2PČme5	videče	A	videnih	2GNsd2	videnih	2PNsp2
videčimi	3GČmp6	videčim	2PČme6	viden	GNme1	videnima	3GNsd3	videnim	2PNsp3
videča	1GČmd1	videči	1PČmp1	videnega	3GNme2	videni	1GNsd4	videna	1PNsp4
videčih	2GČmd2	videčih	2PČmp2	videnemu	3GNme3	videnih	2GNsd5	videnih	2PNsp5
videčima	3GČmd3	videčim	2PČmp3	viden	GNme4	videnima	3GNsd6	videnimi	3PNsp6
videča	1GČmd4	videče	1PČmp4	videnega	3GNme4	videni	1GNmeli	videni	1PNsd1
videčih	2GČmd5	videčih	2PČmp5	videnem	2GNme5	videno	A	videnih	2PNsd2
videčima	3GČmd6	videčimi	3PČmp6	videnim	2GNme6	viden	PNme1	videnima	3PNsd3
videča	1GČze1	videča	1PČmd1	videni	1GNmp1	videnega	3PNme2	videni	1PNsd4
videče	1GČze2	videčih	2PČmd2	videnih	2GNmp2	videnemu	3PNme3	videnih	2PNsd5
videči	1GČze3	videčima	3PČmd3	videnim	2GNmp3	viden	PNme4	videnima	3PNsd6
videčo	1GČze4	videča	1PČmd4	videne	1GNmp4	videnega	3PNme4	videni	1PNmeli
videči	1GČze5	videčih	2PČmd5	videnih	2GNmp5	videnem	2PNme5	videno	A
videčo	1GČze6	videčima	3PČmd6	videnimi	3GNmp6	videnim	2PNme6		

3.4.4 Preverjanje s slovarjem izpeljank

V tabeli 9 je bilo navedeno skupno število izpeljanih besednih oblik, 3.487.676, ki nastanejo iz 91.588 lem v Slovarju slovenskega knjižnega jezika. Za preverjanje fonda besed v obeh vzorcih in za ugotavljanje besedam pripadajočih oblikoslovnih oznak v drugem vzorcu je bil slovar besednih oblik SSKJ organiziran v dve datoteki. V prvi so shranjene abecedno urejene besedne oblike in pri vsaki še stisnjene kode s podatki o lemah in oznakah, kakršni so bili navedeni v tabelah 12, 15 in 17 – dolga je 10.470.706 bajtov, v drugi pa dekomprimiranju namenjeni slovar 7.080 stisnjenih kod in ustreznih podatkov – dolga je 142.252 bajtov.

Poleg tega je bil pri preverjanju uporabljen še slovar besed pri postavljanju in preizkušanju prvega oblikoslovnega označevalnika zbranega gradiva (Jakopin in Bizjak 1997, 330.000 besed), ki vsebuje štiri prozna dela (*Pomladni dan* in *Prazno ptičnico* Cirila Kosmača, Platonovo *Državo* in Orwellov *1984*) ter vzorec iz časopisa Delo. Ta slovar obsega 45.304 različne besedne oblike (ki imajo seveda tudi podatke o oblikoslovnih oznakah). Rezultati preverjanja so navedeni v tabeli 18.

Tabela 18: Rezultat preverjanja besednih oblik iz obeh vzorcev

	Prvi vzorec	%	Drugi vzorec	%
Skupaj besednih oblik	174.579	100,00	36.473	100,00
Prepoznanih s slovarjem oblik iz SSKJ	143.933	82,45	31.656	86,79
Prepoznanih s slovarjem iz gradiva	4.335	2,48	2.924	8,02
Ostane	26.311	15,07	1.893	5,19

Žal niso ohranjeni podatki o številu različnih besed v obeh vzorcih na začetku, še pred kakršnim koli popraviljem. Napake, ki so bile odkrite pri različnih delih z vzorcema, že ob njunem sestavljanju samem, so bile popravljane sproti in vodenje evidence ob tem ni bilo izvedljivo. Napak je bilo ponekod zelo veliko – tudi več na vrstico, ponekod pa le ena na nekaj vrstic ali še manj.

Kot je videti iz zgornje tabele, je bilo s pomočjo izpeljank iz SSKJ v prvem vzorcu prepoznanih dobrih 82 % besed, v drugem pa skoraj 87 %. Po upoštevanju še besed iz doslej obdelanega gradiva je na koncu v prvem vzorcu ostalo dobrih 26.000 besed ali 15 % celote, v drugem vzorcu pa veliko manj, le slabih 2.000 ali 5 % vsega; dva romana iz tega vzorca (37 % celote) sta bila namreč že vključena v gradivo. Med neprepoznanimi besedami so predvsem lastna imena, ki jih je bilo v prvem vzorcu (vključuje tudi potopisna in spominska dela ter prevode) razmeroma več kot v drugem. To dejstvo utegne biti eden glavnih razlogov za opazno razliko pri preseku med posameznim vzorcem in SSKJ (82 % proti 87 %).

Na hipotetično vprašanje, koliko je bilo lem, iz katerih izhajajo prepoznane in neprepoznane besedne oblike, je odgovor mogoče podati za drugi vzorec, ki je bil oblikoslovno označen in lematiziran. Prepoznane besedne oblike v njem so izhajale iz 13.524 lem (11.978 iz SSKJ in 1.546 iz že obdelanega gradiva) od vsega skupaj 15.140, se pravi da je bil tu delež prepoznanega 89,33 %, skoraj 6 % manj kot pri besednih oblikah. Za ugotavljanje števila lem, iz katerih izhajajo prepoznane oblike pri prvem vzorcu, bi bilo potrebno prvi vzorec še lematizirati, ocenimo pa lahko, da bi bil delež tudi tam nižji.

Preostale besede iz prvega vzorca so bile ročno pregledane in popravljene – na računalniku z 32 MB pomnilnika sta bili v njem sočasno lahko tako vzorec sam kot tudi slovar neznanih besed. Preverjanje sumljive besede v slovarju (ukaz na funkcijski tipki jo je potegnil v iskalni niz in aktiviral iskanje v vzorcu) je trajalo manj kot 10 sekund. Ob tem so bili sproti označevani še do takrat neodkriti tuji citati.

Postopek pri drugem vzorcu, kjer je bilo besed sicer veliko manj, a jim je bilo treba poleg preverjanja dodati še oblikoslovne oznake, je bil podoben, le da je bil slovar neznanih besed obrnjen (urejen po koncih). Končnice besed vsebujejo veliko informacije o pregibanju, ki je v neposredni zvezi z oblikoslovno oznako, in tako je delo lažje.

4. poglavje

Oblikoslovno označevanje

4.1 Uvod

Kot je bilo že omenjeno v razdelku o označevanju in o vprašanju napak, se pri gradnji vsake besedilne zbirke, tudi take, ki bi bila namenjena statistični raziskavi in merjenju entropije, postavi problem, kako daleč besedila označiti in preveriti. Oblikoslovno označevanje (angl. *part-of-speech tagging* ali *POS tagging*, npr. Dermatas in Kokkinakis 1995) doda vsaki besedi v besedilu še slovnično informacijo; označevanje zahteva, da je beseda ne samo prav napisana (sama zase), ampak mora biti pravilna tudi v kontekstu. Oblikoslovna označitev pa ne prinese samo bolj čistega besedila in novih možnosti za kvantitativno analizo. Je tudi predpogoj za lematizacijo besedila, ki šele da odgovor na osnovno vprašanje o bogastvu besednega zaklada in o entropiji v širšem smislu. V pričujočem delu je bil oblikoslovno označen drugi vzorec; oblikoslovne označitve prvega vzorca, tudi zaradi velikega števila neznanih besed (26.311), ni bilo mogoče dobro izvesti že v tem času.

Oblikoslovno označevanje se je pojavilo v zgodovini kvantitativnega jezikoslovja kmalu za prvimi besedilnimi korpusi in je, podobno kot te zbirke same, vezano predvsem na angleški jezik. Prvi je bil Brownov korpus, imenovan po univerzi Brown (Francis in Kučera 1982) s 1.000 besedili s po 2.000 besedami iz ameriških virov: skupaj 2 milijona besed, vsa iz besedil, napisanih leta 1961. Korpus je bil označen ročno in je kasneje služil kot podlaga za preizkušanje različnih kvantitativnih modelov. Vključen je tudi, kar se oblikoslovnega označevanja tiče, v doslej najbolj uporabljano in proučevano učno bazo, korpus Penn Treebank s 4,5 milijona besed (Marcus 1993). Tu je poleg Brownovega korpusa še največ gradiva iz časopisa *Wall Street Journal*; cela zbirka je bila na novo ročno označena, z različnimi nabori oznak.

Hiter razvoj računalnikov je v začetku devetdesetih let omogočil nastanek in kasneje tudi označitev že za velikostni razred večjih zbirk, od katerih sta najbolj znani *British National Corpus* in *Bank of English*. BNC je zaključen, obsega 100 milijonov besed, nastal je od leta 1992 do 1994 s sodelovanjem več evropskih uni-

verz – domicil ima na oksfordski, označevalnik pa je z univerze v Lancastru. Zanimivo je, da sta bila uporabljena dva nabora oznak, prvi z 61 oznakami za celoten korpus in drugi, razširjeni, s 139 oznakami za jedro korpusa (2 milijona besed), v katerem so bile oznake tudi ročno preverjene in popravljene, s točnostjo 99,7 %. Natančnost strojnega označevanja cenijo na 98,3 %, dvoumnosti pri označevanju pa niso bile povsod razrešene – 4,7 % besed ima več kot eno oznako (Leech idr. 1994). BNC je (skoraj brezplačno) na razpolago vsem raziskovalcem iz držav Evropske unije. Nemški korpus, približno 200 milijonov besed, zbran na Inštitutu za nemški jezik (IDS) v Mannheimu in francoski korpus, FRANTEXT s 3.000 besedili in podobnim obsegom, sta uporabnikom zunaj matičnih držav veliko teže dostopna in zato tudi dosti manj znana.

Bank of English je projekt, ki sta ga v začetku osemdesetih let zastavili univerza v Birminghamu in založba Collins (prim. Jakopin 1996a); ustanovili sta podjetje Cobuild, ki zbirko dopolnjuje in je zdaj dosegla že 330 milijonov besed. Proti plačilu jo je mogoče uporabljati prek interneta, oblikoslovno pa so jo označili v Helsinkih (Järvinen 1994) – točnost so ocenili na 98,5 %. Glavna centra za oblikoslovno označevanje sta raziskovalni center podjetja Rank Xerox v francoskem Grenoblu, kjer se ukvarjajo s približno 15 evropskimi jeziki, med njimi češkim, poljskim in ruskim ter arabščino (Chanod 1997), in podjetje Lingsoft (skupaj z univerzo) iz Helsinkov, predvsem za angleščino (npr. Tapanainen in Voutilainen 1994), v zadnjem času pa tudi še za nekaj drugih jezikov. Ustanovi sta precej povezani – raziskovalci objavljajo skupaj, s tem da je morda nekaj več teže na finski strani. Angleški jezik fleksijsko ni bogat in so bili nabori oblikoslovnih oznak zato lahko majhni – pri korpusu Penn npr. le 48 oznak (36 za besede in 12 za ločila), pri korpusu BNC 61, pri Brownovem 87 in tudi pri francoskem le 88 (Chanod in Tapanainen 1995).

Drugače je pri slovanskih jezikih, ki so znani po svoji pregibnosti. Če upoštevamo sloves praške jezikoslovne šole, ni čudno, da so v slovanski jezikovni družini doslej prišli najdlje na praški univerzi. V sedemdesetih letih so najprej zbrali in ročno označili 600.000 besed dolg korpus, ki so ga potem uporabljali, tudi s pomočjo Xeroxovega centra v Grenoblu, za izpopolnjevanje lastnega strojnega oblikoslovnega označevalnika. Pred petimi leti so ustanovili *Inštitut za češki nacionalni korpus*, s financiranjem iz raziskovalnih virov in z namenom vzpostavitve 100 milijonov besed obsegajočega besedilnega korpusa (Klimová 1996 in Čermák 1997). Trenutno je prek interneta mogoče dobiti konkordance iz 20 milijonov besed predvsem časopisnega jezika. Leta 1996 jim je uspelo s sredstvi ministrstva za izobraževanje (najprej za tri leta) in s sodelovanjem več čeških fakultet in inštitutov ustanoviti *Laboratorij za jezikoslovne podatke* pri Inštitutu za formalno in uporabno jezikoslovje Fakultete za matematiko in fiziko. Njegov namen so raziskave na področju procesiranja naravnega jezika z uporabo kombinacije simboličnih in statističnih metod, tako za pisani kot tudi za govorni jezik (Hajičová 1997).

Nabor oblikoslovnih oznak je seveda pri češkem jeziku bistveno večji, prek 1.000 enot (Hladká in Hajič 1995), in neprimeren za uporabo običajnih statističnih metod, ki so se obnesle pri angleškem in sorodnih jezikih.

Jeziki malih narodov so seveda, če ni kakih izrednih okoliščin (portugalščino na drugi strani oceana uporablja za velikostni razred več govorcev kot v matični domovini), na obrobju zanimanja multinacionalnih družb, ki se ukvarjajo z jezikovnimi tehnologijami. Tako je npr. korpus bosanskega jezika (1,6 milijona besed) nastal na univerzi v Oslu (Santos 1998), oblikoslovni označevalnik za portugalski jezik (in esperanto) pa najdemo na univerzi v Aarhusu na Danskem (Bick 1996) – s sodelovanjem Pasija Tapanainena iz Helsinkov. Pri nas je stanje, kljub skromnim virom, pri raziskavah govorjenega jezika precej blizu svetovni ravni (Mihelič 1998), na področju pisanega jezika pa je zaostanek večji; poleg drugih razlogov se pozna tudi, da skrb za jezik ni institucionalizirana, kot je npr. v skandinavskih državah (Orešnik 1995).

4.2 Prvi označevalnik

Avtor se je seznanil z oblikoslovnim označevanjem, po virih iz korpusa *Penn Treebank*, na delavnici Gregoryja Grefenstetteja v Budimpešti (Grefenstette 1996). Do sredine leta 1997 je nastal prvi oblikoslovni označevalnik (Jakopin in Bizjak 1997), ki je zahteval že pripravljen slovar besed in možnih oznak za vse besede v besedilu in ki je slonel predvsem na upoštevanju n -terčkov besed in oznak do globine 5 iz že obdelanega (učnega) gradiva:

$$(b_{i-2}, o_{i-2}) (b_{i-1}, o_{i-1}) (b_i, o_i) (b_{i+1}, o_{i+1}) (b_{i+2}, o_{i+2}) \quad (1)$$

V izrazu (1) so besede označene z b , oznake z o , nanaša pa se na okolico i -te besede v novem besedilu. Označevalnik je bil dvostopenjski – v prvi fazi je označil samo besede, ki so imele ali

- a) samo eno oznako ali (2)
- b) za katere je obstajala kaka najmanj dve in ne več kot pet besed dolga okolica, pri čemer šteje tudi opazovana beseda, ki je imela natančno en, iz slovarja označenih stavkov razviden niz oznak.

Druga stopnja označevalnika je slonela na predpostavki, da je zaporedje besed v besedilu markovska veriga n -tega reda, kjer je vsak naslednji člen (dogodek) odvisen od prejšnjih $n-1$ členov. Tokrat so bile upoštevane okolice (n -terčki) iz slovarja oznak, pri čemer je bila okolica spet dolga od dve do pet besed – opazovana beseda in od ena do štiri besede pred njo ali od ena do štiri besede za njo. Označevalnik je za vsako okolico poiskal množico vseh možnih nizov oznak in pogledal, kateri od teh nizov se v slovarju n -terčkov (dvojčkov, trojčkov, četverčkov in peterčkov) dejansko pojavijo. Če je bil en sam, ga je proglasil za pravega in iz njega vzel oznako za opazovano besedo. Če je bilo možnih nizov možnih oznak, ki so

se pojavili tudi v učnem slovarju, več, je označevalnik pregledal še vse preostale (širše) okolice. V primeru, da niti prva niti druga faza za opazovano besedo nista našla ustrezne oznake, je ostala beseda neoznačena.

Pri sestavi nabora oblikoslovnih oznak (4.797 enot) so bili upoštevani slovenska slovnica (Toporišič 1984) in tuji zgledi, predvsem že omenjeni češki, obvezne besedne vrste in pregibni vzorci, navedeni v priročniku za ocenjevanje leksike (Underwood in Navaretta 1997), ter nabor, uporabljen pri mednarodnem projektu MULTEXT-East (MULTEXT-East 1997). Iz slovenske slovnice izhajajoči bogato razčlenjeni nabor oznak sicer res zaplete izdelavo avtomatskega označevalnika in omeji število statističnih metod, ki jih je mogoče uporabiti, po drugi strani pa nas približa «idealu različnih oznak za vse razrede besed, ki imajo različne slovnicihne pomene» (Garside idr. 1987). Majhno število oznak tudi dodatno oteži izbiro leme – primer je angleška izpeljanka *lying* (ali izvira iz leme *lie* (lagati) ali iz *lie* (ležati)).

S tem označevalnikom je bila obdelana zadnja tretjina romana *Pomladni dan* in roman *Prazna ptičnica* v drugem vzorcu. Natančnost je bila ocenjena z 80 %, obe deli pa sta bili kasneje še ročno preverjeni in popravljeni.

4.3 Drugi označevalnik

Da bi bilo mogoče označiti cel drugi vzorec in da bi bilo mogoče strojno dodeljene oznake tudi ročno preveriti v kratkem času, je bil potreben boljši označevalnik. Poleg tega, da je bil prvi premalo natančen, je bil zaradi svoje interaktivne narave tudi precej počasen. Pri novem označevalniku, ki ga je avtor napisal v začetku leta 1998, sta oba postopka (upoštevanje zgodovine in ugotavljanje oznak iz podatkov o markovski verigi oznak učnega gradiva) združena in poenostavljena, dodana pa je še redukcija oznak, ki zmanjša nabor pri slovarju n-terčkov oznak na tretjino. Označevalnik se pri delu opira na tri slovarje: slovar besednih oblik z vsemi možnimi oznakami, slovar pogostejših (s frekvenco 2 ali več) besednih n-terčkov ($n = 2-5$) z oznakami in slovar pogostejših n-terčkov oznak. V tabeli 19 je izsek iz prvega slovarja.

Tabela 19: Besedne oblike na *dolgo-* z oznakami

dolgo•A,131;Pse1,7;Pse4,2;Pže4,14;Pže6,3;ŠNE,2
 dolgolascem•Sme6;Smp3,1
 dolgolasci•Smp1,2;Smp6
 dolgoletnega•Pme2;Pme4,1;Pse2
 dolgoletnem•Pme5;Pse5,1
 dolgoletni•Pme11,1;Pme41;Pmp1,2;Psd1;Psd4;Pžd1;Pžd4;Pže3;Pže5
 dolgoletnimi•Pmp6;Psp6;Pžp6,1
 dolgonoga•Pmd1;Pmd4;Psp1;Psp4;Pže1
 dolgoprstna•Pmd1;Pmd4;Psp1;Psp4,1;Pže1
 dolgoropi•Pme11,1;Pme41;Pmp1;Psd1;Psd4;Pžd1;Pžd4;Pže3;Pže5
 dolgorogo•A,1;Pse1;Pse4;Pže4;Pže6
 dolgoroka•Pmd1;Pmd4;Psp1;Psp4;Pže1

V tabeli so v prvem polju vsakega podatkovnega zapisa besedne oblike, v drugem pa možne oblikoslovne oznake, ločene s podpičji. Če oznake izvirajo iz že označenega gradiva, imajo tudi frekvenco, ločeno z vejico, če so bile dobljene iz SSKJ (glej razdelek 3.4) ali določene naknadno, pa frekvence še nimajo. Kratice v oznakah so: *A* prislov, *P* pridevnik, *S* samostalnik, *ŠNE* nedoločni števnik, *m*, *ž*, *s* spol, *e* ednina, *d* dvojina, *p* množina, številke pa označujejo sklon. Kot je videti, predvsem dvojsinske oblike, ki so v besedilih redke, zelo povečajo število možnih oznak. Nasploh je, če upoštevamo vse izpeljanke iz samostalnikov, pridevnikov in glagolov v SSKJ (tabela 9, 3.487.676 izpeljank, od tega 1.075.779 različnih), povprečno 3,24 oznake na besedno obliko. To je bistveno več kot pri angleščini ali portugalščini (pribl. 2) in blizu vrednosti za češki jezik (3,64 – Hladká in Hajič 1995).

Ta slovar uporabi označevalnik za tvorjenje niza vseh možnih oznak za vse besede v opazovani povedi, in kadar ima beseda eno samo oznako (najpogostejši veznik *in* ima samo oznako *V_{pr}* – priredni veznik), je ta tudi že prava in je postopek za to besedo končan.

Nabor oznak, uporabljen pri prvem označevalniku, je bil zaradi predvidene lematizacije označenega besedila dopolnjen z razčlenitvijo glagola *biti* v vseh njegovih oblikah. Razčlenjen je na tri dele:

Tabela 20: Razčlenitev glagola *biti*

lema	razlaga	primer
1. biti_o	biti v smislu obstajanja v času in prostoru	Hiša <i>je</i> tam.
2. biti_r	biti v relacijskem smislu	Ona <i>je</i> lepa.
3. biti_p	pomožni glagol biti	Skoraj <i>sem</i> videl volka.

Lema *biti* je ostala glagolu *biti*, *bijem*. S to razširitvijo in z uvedbo oznake IN za nazive (npr. *Don*) se je število oznak v naboru povzpelo na 4.907. Kot je bilo že navedeno v tabeli 18, je v drugem vzorcu 36.473 različnih besednih oblik. Oznake za 31.656 se je dalo poiskati s pomočjo slovarja oblik iz SSKJ, 2.924 iz že obdelanega gradiva, 1.893 pa jih je ročno klasificiral avtor.

Drugi označevalnikov slovar vsebuje vse besedne n-terice ($n = 2-5$) iz že obdelanega gradiva, ki imajo frekvenco vsaj 2, in njihove oznake. Izsek iz tega slovarja (330.000 upoštevanih besed, za vejico so navedene frekvence), najprej peterčkov, s katerimi algoritem začne, zatem četverčkov, trojčkov in dvojčkov, je naveden v tabeli 21. Če bi program v besedilu naletel na frazo *zdelo se ji je, da*, ki ima, kot je razvidno iz tabele, samo en niz oznak: *GLse Gmp ZOče3 GPce Vpo* (deležnik na -l srednjega spola v ednini, prosti glagolski morfem, osebni zaimek tretja oseba ženski spol ednina dajalnik, pomožni glagol biti tretja oseba ednine in podredni veznik), bi bil postopek dodelitve oznak tem petim besedam že končan. V primeru besednega trojčka *že v prvem*, ki ima dva možna niza oznak: *Č E5 ŠVme5* in *Č E5 ŠVse5*, je razločevanje končano za prvi dve besedi (v obeh primerih sta oznaki *členek* in *predlog, ki se veže z mestnikom*), tretja beseda pa še osta-

ne za nadaljnji postopek (vrstilni števniki ednina mestnik moškega ali srednjega spola).

Tabela 21: Izsek iz slovarja besednih n-terčkov z oznakami

zdelo se ji je da·GLse Gmp ZOče3 GPce Vpo,5
 zdelo se ji je naravno·GLse Gmp ZOče3 GPce Pse1,2
 zdelo se ji je samo·GLse Gmp ZOče3 GPce ZPUse1,2

v tem trenutku se·E5 ZKme5 Sme5 Gmp,5
 v tem trenutku so·E5 ZKme5 Sme5 GPcp,2
 v tem času na·E5 ZKme5 Sme5 E4,1;E5 ZKme5 Sme5 E5,1

že v prvem·Č E5 ŠVme5,1;Č E5 ŠVse5,1
 že v sredo·Č E4 Sže4,2
 že v začetku·Č E5 Sme5,3

druge države·ZDže2 Sže2,1;ZDžp1 Sžp1,1;ZDžp4 Sžp4,1
 druge državljane·ZDmp4 Smp4,2
 druge državne·ZDže2 Pže2,1;ZDžp1 Pžp1,1

Učinkovitost slovarja besednih n-teric z oznakami je sorazmerna velikosti učnega vzorca, že obdelanega in preverjenega gradiva. Po podatkih, dostopnih iz gradiv projekta MULTEXT-East (MULTEXT-East 1997) znaša 50 % pri 1 milijon besed velikem učnem besedilu. Pri označevanju drugega vzorca, ki je potekalo v več fazah in kjer se je spreminjala tudi velikost učnega gradiva (največ 330.000 besed: *Pomladni dan*, *Prazna ptičnica*, Platon: *Država*, G. Orwell: 1984, vzorec časopisa *DELO*), je bil ta delež približno 35 %.

V tretjem slovarju označevalnika so shranjeni podatki o pogostejših (spet s frekvenco 2 ali več) n-terčkih oznak iz učnega besedila, posebej za kratke povedi (1–5 besed) in posebej za splošne primere. Ker je število oznak (4.907) zelo veliko, učno besedilo označevalnika pa še vedno razmeroma skromno, je bilo treba za povečanje učinkovitosti tega slovarja nabor zmanjšati – iz njega izločiti vso informacijo, ki pri razločevanju ni nujna. S transformacijsko tabelo, ki je navedena v tabeli 22, so bile oznake, ki imajo spol, število in sklon, za vse vrste, ki nastopajo v samostalniški vlogi, združene v oznako *S*, vse pridevniške v *P*, vse zaimkovne pa v *Z*. S tem je bilo število oznak v tem slovarju zmanjšano s 1.498 na 514, frekvence oznak pa ustrezno večje.

Tabela 22: Redukcijska tabela pred tvorjenjem n-terčkov oznak

GČ	P	ISP	S	PIM	P	PŠI	P	ZI	Z	ZR	Z	ZSVcd	Z
GN	P	ISS	S	PIO	P	PT	P	ZK	Z	ZRPO	Z	ZSVcme	Z
GT	P	IV	S	PIP	P	S	S	ZM	Z	ZSVad	Z	ZSVcp	Z
IB	S	IZ	S	PIS	P	SG	S	ZNE	Z	ZSVae	Z	ZSVče	Z
IM	S	IŽ	S	PIV	P	ŠG	P	ZNI	Z	ZSVap	Z	ZSVP	Z
IO	S	KI	S	PIZ	P	ŠM	P	ZO	Z	ZSVbd	Z	ZT	Z
IP	S	P	P	PL	P	ŠV	P	ZPO	Z	ZSVbe	Z	ZTS	Z
IS	S	PČ	P	PN	P	ZD	Z	ZPU	Z	ZSVbp	Z	ZV	Z

Algoritem pri razločevanju preostalih dvoumnosti po uporabljenih n-terčkih besed in oznak iz drugega slovarja upošteva najprej dvojčke, potem trojčke, zatem četverčke in nazadnje peterčke. Če uspe najti za vse možne dvojčke oznak opazovane besede in njenih neposrednih sosed v slovarju n-terčkov oznak en sam par, je ta pravi, oznaka iz njega upoštevana in postopek končan. Če ne, se postopek nadaljuje s širšimi okolicami, do $n = 5$. Izsek iz splošnega dela tega slovarja (330.000 besed učnega besedila), za $n = 2 - 5$, je naveden v tabeli 23 (za podatkovnim ločilom, dvignjeno piko, so navedene frekvence).

Tabela 23: Izsek iz slovarja n-terčkov oznak

Pme5 Pme5:58	E4 Pže4 Sse4:2	Pme1 Pme1 Sme1 Pme2:2	Vpr GLse Gmp GPce Vpo:10
Pme5 Sme5:561	E4 Pže4 Sže4:307	Pme1 Pme1 Sme1 Sme1:9	Vpr GLse Gmp ZOae3 GPce:3
Pme5 Sse5:9	E4 Pže4 Vpr:20	Pme1 Pme1 Sme1 Sme2:2	Vpr GLse Gmp ZOcm3 GPce:9
Pme5 Vpo:3	E4 Pže4 Zže4:5	Pme1 Pme1 Sme1 Sže2:2	Vpr GLse Gmp ZOCze3 GPce:2
Pme5 Vpr:19	E4 Pže4jj Sže4:5	Pme1 Pme1 Sme1 Vpr:6	Vpr GLže E6 Pže6 Sže6:2

N-terčki iz obeh slovarjev imajo tudi frekvence, tako da je mogoče med postopkom spremljati, kolikokrat se je vsaka izmed možnih oznak pri opazovani besedi pojavila, in seštevati vmesne frekvence te oznake. Na ta način je mogoče v primeru, da postopek ne pripelje do popolne razrešitve dvoumnosti, med preostalimi oznakami izbrati najpogostejšo. Pri postopku razreševanja dvoumnosti je mogoče vplivati na nastavitve, ki so razvidne s slike 3.

Slika 3: Nastavitve pri oblikoslovnem označevanju

Nastavitve strojnega označevanja in razločevanja

Postopek

- 1. N-terčki oznak iz že obdelanega gradiva
- 2. N-terčki besed in oznak
- 3. 1. postopek, nato pa še 2.
- 4. 2. postopek, nato pa še 1.
- 5. Nič - vse dvoumnosti naj ostanejo

Globina n-terčkov oznak

5, 4, 3, 2
 5, 4, 3 5, 4 5

Gl. n-terčkov besed in oznak

5, 4, 3, 2
 5, 4, 3 5, 4 5

Kaj naj se zgodí s preostalimi dvoumnostmi

- Ostanejo naj
- Te besede označi z dvema vprašajema
- Izberi oznako, ki se je v postopku najpogosteje pojavila
- Izberi oznako, ki je v seznamu dane besede najpogostejša
- Izberi tisto oznako, ki ima največjo frekvenco v celem vzorcu

Največje število oznak na obliko, ki naj se še izpiše

Če po uporabi vseh treh slovarjev še vedno ostane za opazovano besedo v novem besedilu več oznak, se navadno izkaže za najboljšo v zgornji sliki izbrana možnost: Izberi oznako, ki se je v postopku najpogosteje pojavila. Za preverjanje šolskih primerov pride dostikrat prav možnost, da odklonimo vsako razreševanje

dvoumnosti (Nič - vse dvoumnosti naj ostanejo) in pustimo, da program vse možne oznake pri besedah izpiše (Ostanejo naj). V spodnjem primeru je bila ta možnost uporabljena na povedi *Danes je lepo vreme*:

■Danes je_____lepo_____vreme_____■
 A 6Dce,6Pce,6Rce,6ce,2Dče2 A,Pse1,Pse4,Pže4,Pže6 Sse1,Sse4

Pri besedici *je* so najprej vse tri možnosti glagola *biti* – v smislu obstajanja, v relacijskem smislu in kot pomožni glagol, sledi tretja oseba ednine glagola *jesti* v sedanjem času, nazadnje pa je še osebni zaimek ženskega spola v tretji osebi ednine in v roditelju. Kadar je možnih oznak zelo veliko, lahko njihovo število pri izpisu omejimo z zadnjim določilom iz nastavitve na sliki 3.

Na sliki 4 je naveden primer besedila pred označitvijo, iz odstavka v drugem vzorcu – odlomek iz pravljice *Kamen in njiva*. Ima 89 besed v 9 povedih.

Slika 4: Primer besedila iz drugega vzorca pred oblikoslovno označitvijo

■In tako se je kamen zaljubil v njivo.■ ■Zmeraj
 jo je gledal.■ ■Bolj in bolj se mu je zdela lepa
 in privlačna.■ ■Pozimi se je skrila pod belo snežno
 odejo.■ ■Spomladi se je predramila vsa temna in
 zaspana.■ ■A kmalu se je zbudila in oblekla.■ ■In kako
 je menjavala svoje obleke!■ ■Včasih je bila zlatorumena
 v cvetju oljne repice, potem v zlati obleki zoreče
 pšenice, potem spet v bledomodri obleki cvetočega lanu,
 pa spet v beli obleki cvetoče ajde, vsa bela kot nevesta.■
 ■Tokrat je bila kamnu najbolj všeč.■

Po oblikoslovni označitvi z upoštevanjem nastavitvev iz slike 3 nastane besedilo s slike 5. Napačno označene besede so 4 – njihovim oznakam je spredaj dodana zvezdica. Prvi *tako* bi moral biti označen z A, *cvetočega* s PČme2, *cvetoče* s PČže2, *bila* v zadnji vrstici pa z GLRže. Natančnost označevanja je bila v tem primeru 85/89, kar je 95,51 %.

V splošnem je bila točnost označevalnika nekoliko manjša. Oznake celotnega drugega vzorca so bile ročno preverjene in popravljene na Inštitutu za slovenski jezik ZRC SAZU; natančnost je mogoče podati za zadnjih 157.933 besed (39 % drugega vzorca) od 319.936, ki so bile skupaj označene s tem označevalnikom. Napačno označenih besed je bilo 12.089, kar pomeni 92,35 odstotno natančnost. Številka je nekje vmes med natančnostjo prvega označevalnika (80 %) in najboljšimi dosežki za angleški jezik (98,5 % – Järvinen 1994 in 98,6 % – Ratnaparkhi

Slika 5: Primer z označevalnikom obdelanega besedila iz drugega vzorca

In_ tako se_ je_ kamen zaljubil v_ njivo. Zmeraj
 Vpr *Vpo Gmp GPce Sme1 GLme E4 Sže4 A
 jo_ je_ gledal. Bolj in_ bolj se_ mu_ je_ zdela lepa
 Z0cže4 GPce GLme A Vpr A Gmp Z0cme3 GPce GLže Pže1
 in_ privlačna. Pozimi se_ je_ skrila pod belo snežno
 Vpr Pže1 A Gmp GPce GLže E4 Pže4 Pže4
 odejo. Spomladi se_ je_ predramila vsa_ temna in_
 Sže4 A Gmp GPce GLže ZTže1 Pže1 Vpr
 zaspana. A_ kmalu se_ je_ zbudila in_ oblekla. In_ kako
 Pže1 Vpr A Gmp GPce GLže Vpr GLže Vpr ZV
 je_ menjavala svoje_ obleke! Včasih je_ bila_ zlatorumena
 GPce GLže ZSVPžp4 Sžp4 A GPce GLRže Pže1
 v_ cvetju oljne repice, potem v_ zlati obleki zoreče
 E5 Sse5 Pže2 Sže2 A E4 Pže5 Sže5 PČže2
 pšenice, potem spet v_ bledomodri obleki cvetočega lanu,
 Sže2 A A E5 Pže5 Sže5 *GČme2 Sme2
 pa_ spet v_ beli obleki cvetoče ajde, vsa_ bela kot nevesta.
 Vpr A E5 Pže5 Sže5 *A Sže2 ZTže1 Pže1 Vpo Sže1
 Takrat je_ bila_ kamnu najbolj všeč.
 A GPce *GLže Sme3 Ajj PD

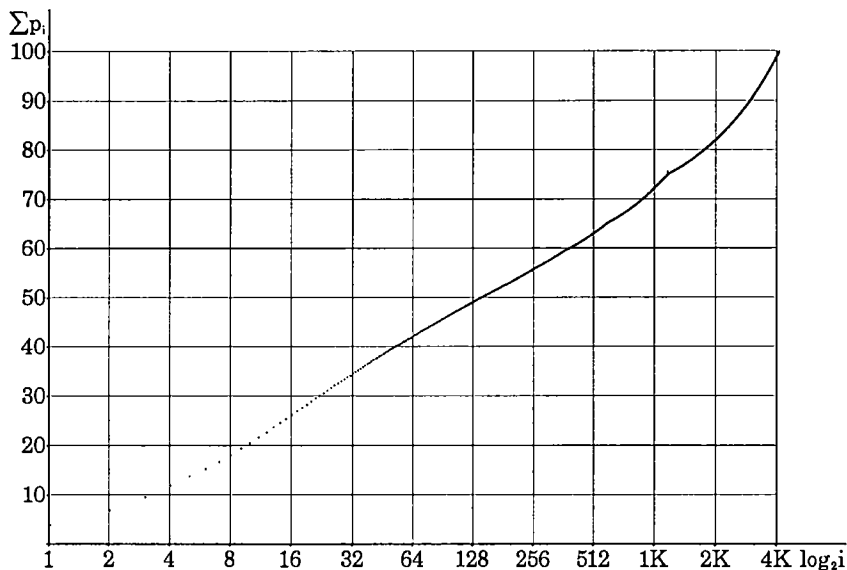
1996); slednji so bili izmerjeni na vzorcih iz časopisa *Wall Street Journal*, ki je bil ročno označen v okviru projekta *Penn Treebank*. Najpogostejše napake so navedene v tabeli 24.

Tabela 24: Najpogostejše napake označevalnika (beseda, prav, narobe, frekvenca)

da	Vpo	Vpr	478	da	Vpo	Č	158
ne	ČZ	Vpo	373	bil	GLRme	GLBme	136
ne	ČZ	Vpr	319	kakor	Vpo	Vpr	134
tako	A	Č	292	zakaj	ZV	Vpr	127
je	GRce	GOce	231	tudi	Č	Vpr	104
je	GRce	GPce	183	tako	A	ZK	101
ali	ČV	Vpr	164	kaj	ZVse1	Č	93
tako	A	Vpr	163	kaj	Č	ZVse4	92

Več o nadaljnjih perspektivah označevalnika je videti s slike 6, ki prikazuje krivuljo rasti vseh trojčkov (beseda, prava oznaka, napačna oznaka) s frekvencami. Različnih napak (trojčkov) je bilo 4.188. Iz krivulje razberemo, da bi bilo za 25-odstotni dvig natančnosti treba popraviti 16 najpogostejših napak, za 50-odstotnega približno 128, za 75-odstotnega pa že več kot 1000.

Slika 6: Krivulja rasti za napake označevalnika



4.4 Oznake drugega vzorca

V drugem vzorcu je bilo označenih 407.938 enot – 407.858 besed in 80 števil. Izmed 4.907 možnih različnih oznak jih je bilo uporabljenih 1.498 ali 30,5 %. Frekvenčni slovar oznak vsebuje priloga C (stran 111), popoln slovar lem in njihovih oznak s frekvencami pa je naveden v prilogi Č (stran 117). V spodnji tabeli so po frekvenci razvrščene besedne vrste.

Tabela 25: Besedne vrste v drugem vzorcu, absolutno in v odstotkih

glagoli	130.208	31,92
samostalniki	83.088	20,37
zaimki	41.259	10,11
vezniki	40.094	9,83
predlogi	35.553	8,72
pridevniki	26.661	6,54
prislovi	26.279	6,44
členki	19.151	4,69
števniki	3.048	0,75
medmeti	1.565	0,38
povedkovniki	1.006	0,25
kratice	26	-
Skupaj	407.938	100,00

Med glagoli je bilo 36.977 ali 28,4 % pomožnih, med samostalniki pa 13.194 ali

15,9 % imen. Podrobnejši pregled najpogostejših oznak je razviden iz tabele 26. Najpogostejša je oznaka za priredni veznik, sledijo prislov, glagolski deležnik na –l moškega spola v ednini, oznaka za pomožni glagol *je*, členek na petem mestu, potem sledijo oznaka za prosti glagolski morfem *se*, podredni veznik, predlog, ki se veže z mestnikom, sedanjik v tretji osebi ednine in glagolski deležnik na –l ženskega spola v ednini.

Tabela 26: Najpogostejših 60 oznak s frekvencami

Vpr	28.529	E4	9.605	GPcp	3.771	Pme1	2.706	E3	1.923	ZOče4	1.396
A	25.756	Sme1	7.830	Sže5	3.742	GPae	2.663	GZPce	1.908	Sse2	1.347
GLme	24.463	E6	7.553	ZOcme4	3.338	GNE	2.640	IOže1	1.902	Pse1	1.343
GPce	22.615	Sže4	7.010	ZVR	3.224	Sme5	2.628	ZV	1.888	Sse5	1.317
Č	15.353	IOme1	6.626	GBI	3.223	Smp1	2.326	Sme6	1.869	Sse1	1.311
Gmp	14.631	Sže1	6.093	ČZ	3.215	GRce	2.303	GLse	1.826	GLRme	1.282
Vpo	11.565	Sme4	5.734	Pže1	3.073	ZOcme3	2.119	Smp4	1.624	ZKse1	1.262
E5	11.018	E2	5.219	Sme2	2.996	Gae	2.073	Sžp1	1.597	Sžp2	1.189
Gce	10.700	GLmp	4.780	Sžp4	2.814	Sže6	2.041	M	1.565	GFPce	1.144
GLže	9.650	Sže2	3.839	Gcp	2.760	Sse4	1.943	Pže4	1.494	Smp2	1.109

5. poglavje

Statistični opis

Pri vsaki kvantitativni obdelavi je zanimivo vprašanje, kako se lotiti ustreznega opisa gradiva. Pri leposlovnih besedilih gre za opis gradnikov hierarhije – delo, stran, odstavek, poved, beseda, znak (oziroma črka v ožjem smislu). Pri elektronskih besedilih navadno ni podatkov o mejah strani, ki tudi pri knjižnih verzijah niso stalne – razlikujejo se od izdaje do izdaje. V tabeli 27 je navedeno število teh gradnikov za oba vzorca.

Tabela 27: Število enot v obeh vzorcih

	Znakov	Besed	Povedi	Odstavkov
Prvi vzorec	16.784.110	2.721.416	201.445	62.493
Drugi vzorec	2.497.308	407.938	37.459	12.710
Skupaj	19.281.418	3.129.354	238.904	75.203

Drugi vzorec, ki vsebuje vsa dela (52) Cirila Kosmača, je po velikosti v znakih približno 15 % prvega vzorca, ki obsega 60 del 41 avtorjev. Oba skupaj sta velika približno stotino največje sedanje besedilne zbirke na svetu – *Bank of English*, kjer pa je leposlovja malo.

Statistični opis je razdeljen na razdelke o znakih, besedah in povedih, ki jim sledi še poglavje o entropiji.

5.1 Znaki

Najmanjši delci pisanega besedila in drugih negovorjenih sporočil, delci, ki se jih ne da še naprej deliti, so znaki. Sem spadajo črke, ločila, med katerimi je najpomembnejši presledek, simboli, kakršen je recimo paragraf, in razne oznake, ki sicer niso del besedila, kot npr. kazalci na opombe. Namesto izraza znaki je v literaturi dostikrat uporabljen tudi izraz simboli.

Tabela 28: Nabor znakov iz obeh vzorcev skupaj, absolutno in v odstotkih

koda	frekv.	%	koda	frekv.	%	koda	frekv.	%			
17	874	-	T	84	30038	0.16	í	161	703	-	
18	874	-	U	85	2907	0.02	ó	162	699	-	
23	966	0.01	V	86	22376	0.12	ú	163	127	-	
24	966	0.01	W	87	1149	0.01	ñ	164	7	-	
25	621	-	X	88	541	-	Č	166	6417	0.03	
26	621	-	Y	89	224	-	č	167	204987	1.06	
32	3462851	17.96	Z	90	14711	0.08	š	168	3913	0.02	
!	33	29454	0.15	[91	8	-	Š	169	138166	0.72
"	34	4	-]	93	8	-	š	170	2699	0.01
#	35	259	-	a	97	1475098	7.65	ž	171	89474	0.46
\$	36	1	-	b	98	261496	1.36	ž	173	86	-
%	37	5	-	c	99	90995	0.47	ž	174	1	-
&	38	23	-	d	100	471153	2.44	ž	179	2	-
'	39	900	-	e	101	1522835	7.90	ž	185	2971	0.02
(40	1623	0.01	f	102	12096	0.06	ž	246	9358	0.05
)	41	1625	0.01	g	103	225488	1.17	ž	247	9358	0.05
*	42	9289	0.05	h	104	144299	0.75	ž	248	1344	0.01
+	43	2	-	i	105	1274003	6.61	ž	249	1344	0.01
,	44	309893	1.61	j	106	655512	3.40	ž	250	75203	0.39
-	45	25632	0.13	k	107	500609	2.60	ž	251	75203	0.39
.	46	183151	0.95	l	108	741056	3.84	ž	252	56240	0.29
/	47	262	-	m	109	449816	2.33	ž	253	56240	0.29
0	48	2618	0.01	n	110	875342	4.54	ž	254	238904	1.24
1	49	6457	0.03	o	111	1280526	6.64	ž	255	238904	1.24
2	50	4310	0.02	p	112	446213	2.31	ž	256	1712	0.01
3	51	3046	0.02	q	113	283	-	ž	257	1712	0.01
4	52	2300	0.01	r	114	705392	3.66	ž	258	3814	0.02
5	53	1977	0.01	s	115	695780	3.61	ž	259	3814	0.02
6	54	1766	0.01	t	116	587227	3.05	ž	260	893	-
7	55	1569	0.01	u	117	264971	1.37	ž	261	893	-
8	56	1969	0.01	v	118	514335	2.67	ž	262	556	-
9	57	2025	0.01	w	119	675	-	ž	263	556	-
:	58	15014	0.08	x	120	628	-	ž	266	3	-
;	59	9802	0.05	y	121	3195	0.02	ž	267	3	-
<	60	1	-	z	122	285122	1.48	ž	372	1	-
=	61	3	-	ž	129	88	-	ž	398	1	-
?	63	18924	0.10	ž	130	2095	0.01	ž	400	7	-
A	65	17178	0.09	ž	131	134	-	ž	484	1	-
B	66	14939	0.08	ž	132	25	-	ž	486	2	-
C	67	3419	0.02	ž	133	154	-	ž	517	1	-
D	68	12200	0.06	ž	135	17	-	ž	518	28	-
E	69	3917	0.02	ž	136	144	-	ž	536	30	-
F	70	3575	0.02	ž	138	551	-	ž	548	2	-
G	71	8018	0.04	ž	139	3	-	ž	571	23	-
H	72	5052	0.03	ž	140	1	-	ž	593	10112	0.05
I	73	15294	0.08	ž	141	9	-	ž	616	2	-
J	74	11087	0.06	ž	144	2	-	«	618	53773	0.28
K	75	27593	0.14	ž	147	115	-	»	619	53773	0.28
L	76	9856	0.05	ž	148	52	-	»	648	2	-
M	77	21442	0.11	ž	149	12	-	-	658	2710	0.01
N	78	26969	0.14	ž	150	1	-	,	678	159	-
O	79	14797	0.08	ž	151	115	-	»	679	405	-
P	80	34869	0.18	ž	153	6	-	»	682	3363	0.02
Q	81	320	-	ž	154	1	-	»	683	3363	0.02
R	82	8934	0.05	ž	158	3	-	«	697	159	-
S	83	24770	0.13	ž	160	1248	0.01	«	698	405	-

5.1.1 Nabor

V tabeli 28 na prejšnji strani je naveden skupni nabor znakov iz obeh vzorcev. Pri tem niso upoštevana zaglavja pred besedili, so pa zajete oznake meja povedi, odstavkov, premege govora in podobnega, vsi dodatni simboli, ki so že bili navedeni v tabeli 5. Vsega skupaj je v obeh vzorcih nastopalo skupaj 168 različnih znakov, v prvem 165, v drugem pa 128. Največ črk, ki ne nastopajo v slovenski abecedi, so prinesli prevodi, predvsem delo Libuše Moníkové *Fasada*. V njem je bilo celo nekaj citatov v cirilici, ki so bili označeni kot neslovensko besedilo in nadomeščeni z zvezdico. Trije dodatni znaki, ki jih je v skupni nabor prispeval drugi vzorec, so bili znak za paragraf ter par dvojnih navednic zgoraj in spodaj – simbola „ in “.

Kode znakov iz tabele 28 so kode, kakršne imajo ti znaki v naboru urejevalnika EVA (16-bitni nabor znakov, njegov naslov na internetu je naveden v prilogi D na strani 293). Pari znakov z nizkimi kodami razmejujejo vrsto pisave – par (17, 18) pomeni začetek in konec poudarjenega tiska (skupaj je bilo takih delov tiska 874), par (23, 24) začetek in konec kurzive (966-krat), par (25, 26) pa začetek in konec potenc (predvsem za opombe, 621-krat). Indeksov v besedilih ni bilo, bil pa je, v besedilu prvega vzorca, dvakrat uporabljen poseben znak za kvadrat (², koda 648).

Znaki od kode 32 do 122 sledijo naboru ASCII, začenši s presledkom, ki ima pogostost 17,96, kar je približno odstotek več kot v primerljivi raziskavi (Kristan idr. 1994); razliko gre pripisati predvsem deležu časopisnega jezika v njej, ki ima opazno daljše besede. Neslovenske črke *q*, *w*, *x* in *y* s frekvencami 283, 675, 628 in 3.195 (0,02 % celote) se po pogostosti s črkami našega jezika, niti s črko *f*, ne morejo primerjati, kar je v skladu s pričakovanji. Izredno redki znaki v tem območju, vsi s frekvenco 5 ali manj, so dvojni narekovaj (4), dolar (1), odstotek (5), znak manjše (1) in enačaj (3). Tako malo dvojnih narekovajev v besedilu je zaradi tega, ker jih v tiskanih leposlovnih delih res ni. Povsod tam, kjer ni bilo eksplicitno razvidno, da je bil res mišljen ta znak, npr. pri premem govoru in podobno, je v rabi par « in ». Posebno pozornost zasluži še pomišljaj (–) s kodo 45. V besedilih je navadno nastopal v obeh vlogah – kot nestični (npr. v povedi *Poglej - žoga!*) in kot stični pomišljaj ali vezaj (npr. v izrazu *Ve-li-ka ne-de-lja*). Da ne bi prihajalo do težav pri prelomu na besede in pri podobnih postopkih, je bil pomišljaj v stični vlogi v obeh vzorcih povsod nadomeščen s posebnim stičnim pomišljajem, ki ima kodo 658. Bilo ga je približno desetkrat manj kot nestičnega (2.711 proti 25.632); nestični pomišljaj je bil zelo pogosto uporabljen za označitev premege govora namesto narekovajev.

V območju kod od 129 do 164 nastopajo črke z diakritičnimi znamenji, ki so v rabi predvsem v zahodni Evropi; med njimi so daleč najpogostejši samoglasniki *é*, *á*, *í* in *ó* z ostrivcem (2.101, 1.247, 704 in 701), sledi pa jim *è* s krativcem (565).

Tabela 29: Nabor znakov iz prvega vzorca, absolutno in v odstotkih

	koda	frekv.	%	koda	frekv.	%	koda	frekv.	%	
	17	867	0.01	S	83	20767	0.12	ß	158	2
	18	867	0.01	T	84	24843	0.15	á	160	1215
	23	929	0.01	U	85	2473	0.01	í	161	698
	24	929	0.01	V	86	19187	0.11	ó	162	672
	25	621	-	W	87	1134	0.01	ú	163	127
	26	621	-	X	88	523	-	ř	164	6
	32	2983591	17.78	Y	89	224	-	č	166	5648
!	33	23097	0.14	Z	90	12482	0.07	č	167	178985
"	34	4	-	[91	8	-	š	168	3340
#	35	259	-]	93	8	-	š	169	120232
\$	36	1	-	a	97	1282355	7.64	š	170	2043
%	37	5	-	b	98	228843	1.36	ž	171	78024
&	38	23	-	c	99	77997	0.46	č	173	66
'	39	851	0.01	d	100	416806	2.48	ø	174	1
(40	1563	0.01	e	101	1329937	7.92	ø	179	2
)	41	1565	0.01	f	102	10925	0.07	(185	2082
*	42	9164	0.05	g	103	196547	1.17	©	246	9352
+	43	2	-	h	104	126261	0.75	©	247	9352
,	44	273760	1.63	i	105	1115738	6.65	©	248	993
-	45	21191	0.13	j	106	574850	3.42	©	249	993
.	46	155567	0.93	k	107	433283	2.58	©	250	62493
/	47	262	-	l	108	644761	3.84	©	251	62493
0	48	2580	0.02	m	109	393578	2.34	©	252	45887
1	49	6400	0.04	n	110	767812	4.57	©	253	45887
2	50	4290	0.03	o	111	1127068	6.72	©	254	201445
3	51	3023	0.02	p	112	389618	2.32	©	255	201445
4	52	2269	0.01	q	113	199	-	©	256	1557
5	53	1953	0.01	r	114	611247	3.64	©	257	1557
6	54	1757	0.01	s	115	608133	3.62	©	258	3432
7	55	1561	0.01	t	116	517774	3.08	©	259	3432
8	56	1960	0.01	u	117	235141	1.40	©	260	659
9	57	1996	0.01	v	118	455821	2.72	©	261	659
:	58	12534	0.07	w	119	646	-	©	262	471
;	59	8894	0.05	x	120	621	-	©	263	471
<	60	1	-	y	121	3179	0.02	©	266	3
=	61	3	-	z	122	247971	1.48	©	267	3
?	63	15811	0.09	ü	129	77	-	À	372	1
A	65	15580	0.09	é	130	2067	0.01	À	398	1
B	66	13310	0.08	à	131	134	-	È	400	7
C	67	3047	0.02	ä	132	24	-	È	484	1
D	68	10424	0.06	à	133	139	-	È	486	2
E	69	3572	0.02	c	135	17	-	È	517	1
F	70	3294	0.02	è	136	144	-	È	518	28
G	71	7353	0.04	è	138	539	-	Ë	536	30
H	72	4206	0.03	ï	139	3	-	Ë	548	2
I	73	12696	0.08	î	140	1	-	Ë	571	23
J	74	9796	0.06	ï	141	8	-	…	593	6967
K	75	24024	0.14	É	144	2	-	»	618	45864
L	76	8835	0.05	ô	147	112	-	»	619	45864
M	77	18331	0.11	ö	148	42	-	»	648	2
N	78	22889	0.14	ò	149	12	-	-	658	2168
O	79	12599	0.08	ù	150	1	-	-	678	144
P	80	29974	0.18	ù	151	3	-	¿	682	3035
Q	81	316	-	Û	153	6	-	¿	683	3035
R	82	7955	0.05	Ü	154	1	-	¿	697	144

Tabela 30: Nabor znakov iz drugega vzorca, absolutno in v odstotkih

	koda	frekv.	%		koda	frekv.	%		koda	frekv.	%	
EN PH NS EM SM HS IT AL IC IL JC	17	7	-	R	82	979	0.04	β	158	1	-	
	18	7	-	S	83	4003	0.16	á	160	33	-	
	23	37	-	T	84	5195	0.21	í	161	5	-	
	24	37	-	U	85	434	0.02	ó	162	27	-	
	32	479261	19.19	V	86	3189	0.13	ň	164	1	-	
	!	33	6357	0.25	W	87	15	-	č	166	769	0.03
	'	39	49	-	X	88	18	-	č	167	26002	1.04
	(40	60	-	Z	90	2229	0.09	š	168	573	0.02
)	41	60	-	a	97	192743	7.72	š	169	17934	0.72
	*	42	125	0.01	b	98	32653	1.31	ž	170	656	0.03
,	44	36133	1.45	c	99	12998	0.52	ž	171	11450	0.46	
-	45	4441	0.18	d	100	54347	2.18	ć	173	20	-	
.	46	27584	1.10	e	101	192898	7.72	˘	185	889	0.04	
0	48	38	-	f	102	1171	0.05	☒	246	6	-	
1	49	57	-	g	103	28941	1.16	☒	247	6	-	
2	50	20	-	h	104	18038	0.72	☒	248	351	0.01	
3	51	23	-	i	105	158265	6.34	☒	249	351	0.01	
4	52	31	-	j	106	80662	3.23	☒	250	12710	0.51	
5	53	24	-	k	107	67326	2.70	☒	251	12710	0.51	
6	54	9	-	l	108	96295	3.86	☒	252	10353	0.41	
7	55	8	-	m	109	56238	2.25	☒	253	10353	0.41	
8	56	9	-	n	110	107530	4.31	☒	254	37459	1.50	
9	57	29	-	o	111	153458	6.14	☒	255	37459	1.50	
:	58	2480	0.10	p	112	56595	2.27	☒	256	155	0.01	
;	59	908	0.04	q	113	84	-	☒	257	155	0.01	
?	63	3113	0.12	r	114	94145	3.77	☒	258	382	0.02	
A	65	1598	0.06	s	115	87647	3.51	☒	259	382	0.02	
B	66	1629	0.07	t	116	69453	2.78	☒	260	234	0.01	
C	67	372	0.01	u	117	29830	1.19	☒	261	234	0.01	
D	68	1776	0.07	v	118	58514	2.34	☒	262	85	-	
E	69	345	0.01	w	119	29	-	☒	263	85	-	
F	70	281	0.01	x	120	7	-	...	593	3145	0.13	
G	71	665	0.03	y	121	16	-	Ⓔ	616	2	-	
H	72	846	0.03	z	122	37151	1.49	<<	618	7909	0.32	
I	73	2598	0.10	ü	129	11	-	>>	619	7909	0.32	
J	74	1291	0.05	é	130	28	-	-	658	542	0.02	
K	75	3569	0.14	ä	132	1	-	,	678	15	-	
L	76	1021	0.04	à	133	15	-	»	679	405	0.02	
M	77	3111	0.12	è	138	12	-	«	682	328	0.01	
N	78	4080	0.16	ì	141	1	-	»	683	328	0.01	
O	79	2198	0.09	ô	147	3	-	«	697	15	-	
P	80	4895	0.20	ö	148	10	-	“	698	405	0.02	
Q	81	4	-	ù	151	112	-					

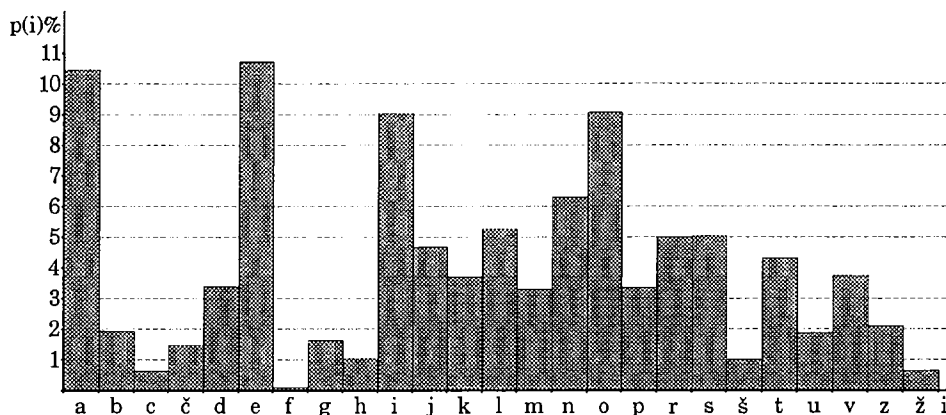
Črke z diakritičnimi znamenji, ki se ne uporabljajo v slovenskem jeziku, npr. samoglasniki s preglasi, so spet zelo redke. Na kodah od 166 do 171 sledijo šumniki Č, č, Š, š, Ž, ž, za njimi pa dve črki od štirih s srbskega oziroma hrvaškega govornega področja, ć (86) in Đ (1). Črka đ ne nastopa niti enkrat, kar pomeni da je bila najverjetneje že v originalnih besedilih transkribirana v *dj*. S kodo 185 malo naprej najdemo trdi presledek, lahko bi mu rekli tudi nedeljivi presledek, s frekvenco 2.057. Z njim je bil nadomeščen navaden presledek v primerih nedeljivih besednih enot, npr. pri krajevnih imenih, kot so *Sv. Lucija, New York* in podobno. Na kodah od 246 do 267 so simboli za označevanje besedila, razloženi v razdelku 3.1.1 (Dodatni simboli). Iz tega dela tabele sledi, da je bilo v besedilu 9.358 primerov nizov, ki niso besedilo (predvsem oznak vrstic v poglavjih *Nove zaveze*), 1.344 citatov v drugih jezikih, 75.203 odstavki, 56.240 primerov premege govora, 238.904 povedi, 1.712 vrstic, ki morajo biti pomaknjene na sredino, 3.814 delov besedila, ki se jih ne sme reformatirati (pesmice, lepo zloženi citati), 893 vzdevkov, 556 citatov v slovenskem jeziku in trije deli, ki morajo biti (na internetu) izpisani s črkami fiksne širine. Na koncu je še nekaj redkejših črk, predvsem z diakritičnimi znamenji, tropičje (s frekvenco 10.112), ki je bilo uvedeno zaradi lažjega določevanja meja povedi, ter razne oblike navednic.

V tabelah 28 in 29 na prejšnjih dveh straneh sta navedena nabora črk za prvi in za drugi vzorec posebej. Odstotne vrednosti v obeh tabelah so dokaj blizu, nekaj razlik je pa vendar opaznih. Tako je v prvem vzorcu presledkov 17,78 %, v drugem pa 19,19 %, kar kaže na krajšo dolžino besed v Kosmačevem besedilu; ta bo potrjena v naslednjem razdelku. V prvem vzorcu pride en premi govor na 4,39 povedi, v drugem pa na 3,62, kar kaže na večji delež dialoga pri Kosmaču. V drugem vzorcu opazimo tudi izredno majhno zastopanost števč – skupaj jih je komaj 0,01 % vseh znakov, v prvem vzorcu pa 0,17 %.

5.1.2 Porazdelitve

Na sliki 7 je prikazana porazdelitev najpogostejših črk v obeh vzorcih, kjer so bile velike in male črke združene, deleži pa se nanašajo na vse črke in ne na vse znake. Opazimo, kot je bilo ugotovljeno že v vseh prejšnjih raziskavah, prevlado črke *e* nad *a*. Odnos je 7,92 % proti 7,74 % glede na vse znake v obeh vzorcih, kar pomeni, da je *e*-jev za 2,33 % več kot *a*-jev. Hipotezi, postavljeni v članku (Kristan idr. 1994), da je odnos med vodilnima samoglasnikoma odvisen od starosti besedila in da je v novejših prevladujoč *a*, je nekoliko bliže odnos v prevodih, novejših besedilih prvega vzorca (6.128.484 znakov). Tam sicer še vedno prevladuje *e*, vendar že manj – 7,73 : 7,66. Po mnenju avtorja je odnos med obema bolj odvisen od žanra besedila – v nekaterih pravljicah je zaradi imen glavnih junakov odnos precej porušen v korist *a*-ja, npr. 7,55 : 8,81 (Jakopin in Musar 1997). Druga razlika je v pogostosti nadaljnjih dveh samoglasnikov, *i* in *o*. V predhodni

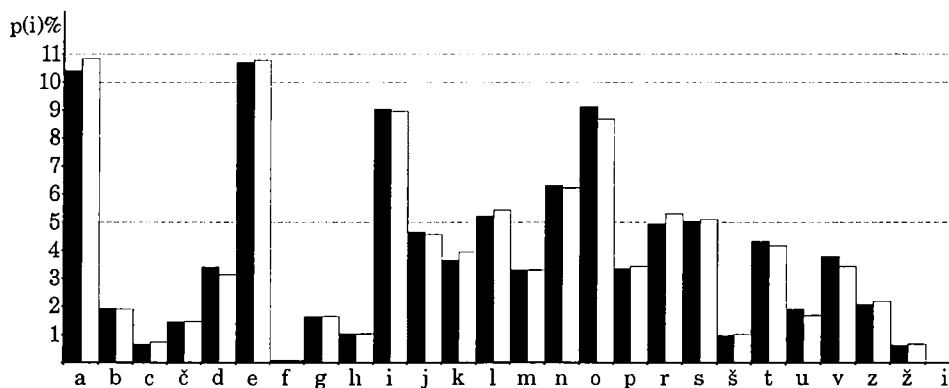
Slika 7: Porazdelitev najpogostejših 25 črk v obeh vzorcih skupaj



raziskavi je črka *i* pogostejša od črke *o*, po rezultatih, dobljenih za oba vzorca skupaj, pa je nekoliko pogostejši *o* – odnos je 6,72 % : 6,69 %. Porazdelitev drugih črk je v skladu s pričakovanji in ne odstopa bistveno od vrednosti, izmerjenih na manjšem vzorcu v predhodni, že omenjeni raziskavi (Kristan idr. 1994).

Primerjava porazdelitev vseh črk v obeh vzorcih je prikazana na sliki 8. Odstotki se tudi tokrat nanašajo na delež posamezne črke med vsemi črkami in ne med vsemi znaki.

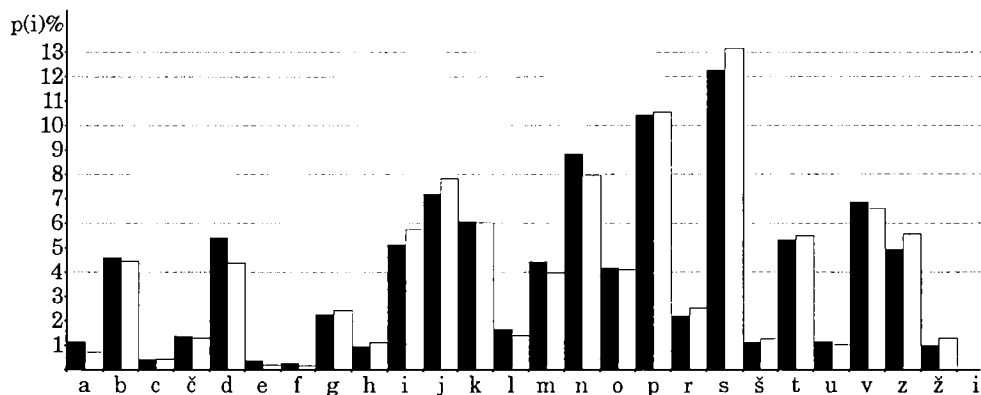
Slika 8: Porazdelitev najpogostejših 25 črk za prvi (črn) in drugi vzorec



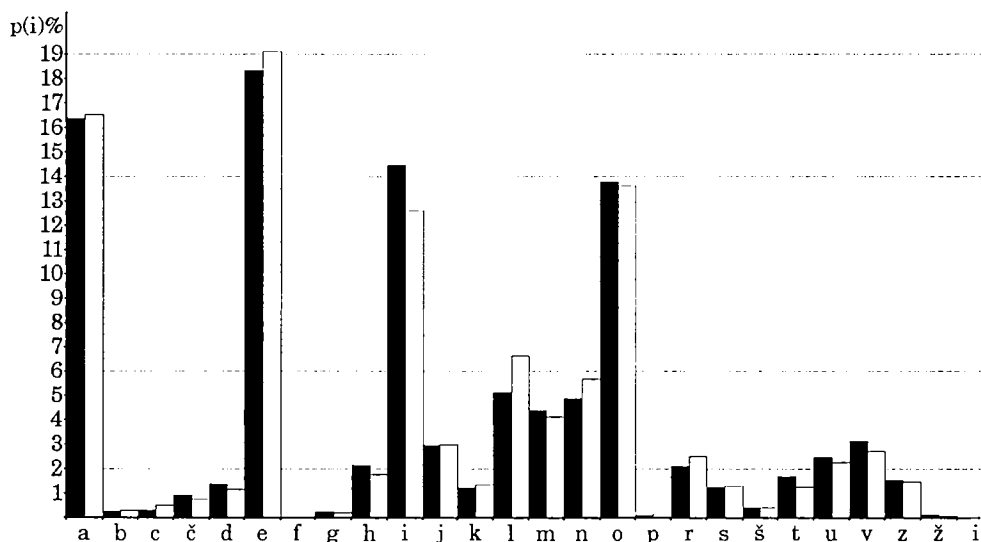
Razlike so majhne, pa zanimive. Takoj je videti, da je v drugem vzorcu delež črke *a* za malenkost večji kot delež črke *e*, obakrat pa spet večji kot v prvem vzorcu, medtem ko so deleži preostalih treh samoglasnikov, črk *i*, *o* in *u*, v drugem vzorcu opazno manjši kot v prvem – lahko bi rekli, da je Kosmačeva govoriča "širša" od govoriče v drugih slovenskih leposlovnih besedilih.

Poleg porazdelitve vseh črk sta najbolj zanimivi še dve porazdelitvi – porazdelitev prvih in zadnjih črk v vseh besedah. Prikazani sta na slikah 9 in 10. Na sliki 9 najprej opazimo, v primerjavi s sliko 8, majhne frekvence samoglasnikov in večjo urejenost cele slike. V nasprotju s prejšnjo raziskavo (kjer je prvi p) je najpogostejša prva črka pri obeh vzorcih s (ki je od vseh črk najlažje izgovorljiva), sledijo p , n , j in v . Črke c , e in f so na začetku besed zelo redko.

Slika 9: Porazdelitev prvih črk vseh besed za prvi (črn) in drugi vzorec



Slika 10: Porazdelitev zadnjih črk vseh besed za prvi (črn) in drugi vzorec



Informacijsko še veliko revnejša je porazdelitev zadnjih črk v vseh besedah. Samoglasniki e , a , i in o in pa l , ki se na koncu besede izgovori kot l , skupaj s

pravim *u-jem* zavzamejo kar 71 % celote. Po drugi strani pa je cela vrsta črk, ki so brez naslednika težko izgovorljive in jih zato na koncih besed najdemo le redko: *b*, *c*, *f*, *g*, *p* in *ž*. V sliko 10 bi po frekvenci sodila tudi črka *y* (0,06 %), ki presega črko *f* (0,03 %); zaradi medsebojne primerljivosti slik 8, 9 in 10 je bila izpuščena.

5.1.3 N-terčki znakov

Znakovni n-terčki, se pravi nizi po 1, 2, 3, 4, 5, ... znakov iz besedila imajo pri kodiranju sporočil veliko uporabno vrednost in tudi sicer o besedilu zelo dosti povedo. Prav pridejo pri krajših mehanizmih za preverjanje pravilnosti besedila, njihovo porazdelitev je mogoče uporabiti tudi pri deljenju besed na koncu vrstic (Jakopin 1995b), na koncu razdelka pa je omenjen njihov pomen za identifikacijo jezika.

Njihovo število z rastočim *n* izredno hitro narašča, še vedno pa seveda dosti počasneje kot število teoretično možnih kombinacij, ki bi jih dobili, če bi bili vsi znaki na vseh mestih enako pogosti. Hitrost naraščanja n-terčkov je razložena v poglavju o entropiji, na naslednjih dveh straneh pa je v tabelah 31 in 32 naštetih 60 najpogostejših n-terčkov s frekvencami, za *n* od 1 do 14 in za oba vzorca. Tabele sta bili napravljeni iz besedil, potem ko so bila že razrezana na povedi, ko so bili izločeni dodatni simboli, tuji citati in razni kazalci (številke opomb in podobno), ki niso besedilo v ožjem smislu; presledek je povsod nadomeščen s podčrtajem (_).

Frekvence najpogostejših n-terčkov v prvem vzorcu za posamezen *n*: *_ / e_ / je_ / _da/ _se_je/ _se_je_ / _je_bil_ / _je_bilo_ / _prijetelj/ ,_da_se_je_ / ,*_je_rekel_ / ?*_je_vprašal* in *,_kakor_da_bi_* padajo z zelo spremenljivim tempom: 5,92 : 1, 2,42 : 1, 1,22 : 1, 4,03 : 1, 1,73 : 1, 1,04 : 1, 2,85 : 1, 1,38 : 1, 3,02 : 1, 1,69 : 1, 1,44 : 1, 1,39 : 1 in 1,57 : 1. Hitrost padanja je, poleg največjega skoka na začetku, očitno povezana z redundantnostjo v besedilu. Ker je iz najpogostejšega trojčka *je_* in iz dejstva, da je *je* najpogostejša beseda, utemeljeno sklepati, da bo najpogostejši četverček *_je_*, je kvocient frekvence med obema le 1,22. Podobno je pri prehodu od šesterčkov na sedmerčke, ko imamo spet niz *_se_je*, ki se skoraj vedno konča z *je* in je potem pričakovani najpogostejši sedmerček *_se_je_* s kvocientom komaj 1,04. Od deseterčkov naprej se hitrost padanja ustali.

Primerjava obeh vzorcev pokaže precejšnje ujemanje pri nizkih *n*: pri enojčkih in dvojčkih je 56 istih nizov v obeh vzorcih, pri trojčkih, četverčkih in peterčkih po 51, pri šesterčkih 45, sedmerčkih 44, osmerčkih 35, deveterčkih 29, deseterčkih in enajsterčkih pa po 22. Istih dvanajsterčkov v obeh vzorcih je bilo le 9: *,_ki_je_bil_ / ,_ki_je_bila/ _ki_je_bila_ / ,*_je_rekel_ / _je_vprašal_ / *_je_vprašal_ / ?*_je_vpraša/ _pripovedova* in *se_je_obrnil*, trinajsterčkov 7: *,_ki_je_bila_ / *_je_vprašal_ / ?*_je_vprašal_ / _se_je_obrnil/ _se_je_oglasil/ se_je_oglasil* in

Tabela 32: Najpogostejši n-terčki v drugem vzorcu s frekvencami; presledek je označen z _

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1.	376.859	e_	67.455	je_	28.622	je_	24.198	e_je_	5.182	se_je_	4.324	se_je_	4.864	_
2.	192.185	a_	53.644	je_	26.185	in_	16.493	_	4.557	se_je_	4.287	je_bil_	2.489	_
3.	191.828	s_	49.487	_	17.599	_	12.944	a_je_	4.478	je_bil_	2.418	kakor_	1.995	a_se_je_
4.	157.335	o_	46.161	in_	17.819	e_je_	5.518	se_je_	4.471	je_bil_	2.416	in_se_	1.562	_
5.	152.863	l_	48.235	in_	16.524	da_	5.956	se_je_	4.482	kakor_	2.851	kakor_	1.161	_
6.	107.898	p_	39.931	po_	16.513	a_je_	4.785	je_p_	3.256	kakor_	1.937	je_bil_	376	da_bil_
7.	95.889	pr_	38.248	se_	14.573	je_	4.676	in_se_	3.245	in_se_	1.781	pa_je_	333	je_rekel_
8.	93.678	r_	34.612	pr_	12.682	na_	4.659	_	2.962	kako_	1.681	da_je_	339	je_bila_
9.	87.342	z_	38.250	_	11.795	da_	4.644	a_in_	2.751	tako_	1.628	in_	929	kakor_b_
10.	88.632	n_	28.248	_	10.989	pri_	4.457	a_b_	2.738	in_	1.606	da_je_	321	je_bila_
11.	68.984	se_	24.141	la_	9.935	bil_	4.281	kako_	2.686	je_za_	1.482	ki_je_	875	akor_b_
12.	67.293	v_	23.768	la_	9.537	pre_	4.254	je_z_	2.571	je_p_	1.458	da_je_	867	je_bilo_
13.	58.293	in_	22.832	_	7.927	_	4.167	e_bil_	2.541	tudi_	1.438	ki_je_	866	kakor_b_
14.	56.467	l_	21.711	bi_	7.868	pa_	4.067	je_p_	2.446	je_p_	1.495	pogled_	824	se_je_z_
15.	56.584	n_	21.353	a_s_	7.738	ila_	4.081	sem_	2.429	_	1.142	da_b_	798	pogleda_
16.	54.124	n_	21.238	e_p_	7.553	je_	3.742	o_in_	2.392	_	1.189	njegov_	739	_
17.	37.676	la_	21.156	da_	7.314	e_po_	3.388	e_in_	2.387	_	1.888	da_se_	733	se_je_p_
18.	35.978	k_	21.828	v_	7.162	in_s_	3.328	o_je_	2.378	_	1.887	ga_je_	724	a_je_bil_
19.	32.573	l_	20.955	em_	7.116	e_je_	3.216	i_je_	2.218	bilo_	1.877	so_se_	723	uprasal_
20.	29.635	z_	20.416	aj_	6.985	_	3.174	svoj_	2.281	e_bil_	1.817	je_bilo_	713	o_se_je_
21.	28.821	ra_	20.318	na_	6.975	sem_	3.158	a_se_	2.198	da_je_	1.888	je_pri_	784	počasi_
22.	27.479	l_	20.867	ne_	6.983	e_pr_	3.222	tako_	2.192	bila_	1.888	je_bila_	698	ki_so_
23.	25.933	po_	19.584	ko_	6.558	hi_	2.931	i_in_	2.180	_	1.881	kako_b_	694	ki_se_
24.	25.193	aj_	19.965	ka_	6.558	ki_	2.978	akor_	2.132	pa_je_	981	e_bil_	691	je_rekel_
25.	17.898	ne_	18.778	is_	6.478	kak_	2.949	je_s_	2.129	in_p_	969	upras_	659	je_je_
26.	12.741	na_	18.629	aj_	6.418	ega_	2.941	kakor_	2.855	a_se_je_	962	e_bilo_	688	se_je_v_
27.	11.443	ko_	17.723	da_	6.406	ala_	2.932	_	2.880	da_je_	946	akor_b_	666	temnikar_
28.	7.547	t_	17.288	ako_	6.389	_	2.914	je_v_	1.966	da_b_	944	kor_b_	664	se_je_s_
29.	7.547	st_	17.144	e_s_	6.374	ne_	2.909	o_se_	1.853	da_je_	935	mu_je_	664	s_koraj_
30.	6.892	ka_	16.891	st_	6.324	ako_	2.984	i_in_p_	1.848	_	931	se_je_z_	652	se_je_p_
31.	5.166	pr_	16.753	ga_	6.222	a_se_	2.889	_	1.883	da_b_	938	pogleda_	654	se_je_
32.	4.861	b_	16.686	ne_	6.206	je_	2.875	in_se_	1.746	zdaj_	926	da_se_	686	sem_se_
33.	4.844	d_	16.623	o_	6.181	je_	2.869	n_se_	1.723	ki_je_	896	se_je_p_	574	priskal_
34.	3.948	ne_	15.524	no_	6.116	e_je_	2.867	je_p_	1.655	svoje_	893	je_rek_	574	živiljen_
35.	3.931	da_	15.182	e_j_	5.947	nil_	2.838	je_p_	1.629	pogled_	888	bi_	568	in_je_
36.	3.562	e_l_	14.927	a_	5.925	ka_	2.838	tako_	1.625	ki_je_	875	_	553	je_reke_
37.	3.147	o_	14.638	nil_	5.828	kako_	2.797	i_se_	1.618	je_je_	861	svoja_	548	pogledal_
38.	3.069	aj_	13.428	pa_	5.641	po_	2.788	je_p_	1.563	samo_	859	kako_	516	in_se_za_
39.	2.494	ri_	13.544	il_	5.553	a_in_	2.752	tudi_	1.494	in_za_	846	rekel_	494	se_je_n_
40.	2.476	em_	13.188	e_b_	5.475	tak_	2.745	bil_	1.467	in_	835	o_se_je_	493	vendar_
41.	2.324	aj_	13.156	a_j_	5.361	sem_	2.719	je_p_	1.454	na_b_	813	o_je_je_	492	je_tako_
42.	2.223	za_	13.124	pre_	5.321	e_je_	2.628	n_	1.458	gledal_	813	uprasal_	488	povedal_
43.	2.178	t_	12.966	so_	5.293	o_po_	2.558	tudi_	1.441	in_za_	812	ki_se_	488	ne_tud_
44.	2.136	n_	12.932	je_	5.279	in_	2.389	je_t_	1.393	ga_je_	886	stopil_	477	rekel_
45.	2.136	d_	12.768	d_	5.224	e_in_	2.394	je_o_	1.387	spet_	888	nekaj_	474	stopil_
46.	1.528	em_	12.749	bil_	5.287	l_je_	2.373	in_	1.371	da_se_	785	počasi_	471	se_je_p_
47.	1.284	te_	12.288	il_	5.131	prav_	2.365	gleda_	1.386	je_pri_	788	pa_se_	463	ker_je_
48.	1.122	no_	12.215	is_	5.066	o_se_	2.347	da_b_	1.255	je_re_	759	lahko_	459	temnika_
49.	1.988	ak_	11.796	ila_	4.982	je_je_	2.238	e_pri_	1.239	upras_	757	ki_so_	456	se_je_za_
50.	968	ve_	11.355	nil_	4.938	_	2.218	i_je_	1.232	so_se_	751	potem_	455	o_je_bil_
51.	965	ar_	11.839	e_n_	4.911	svoj_	2.242	da_s_	1.229	ga_je_	758	ki_so_	452	da_bil_
52.	822	le_	18.933	o_	4.894	svoj_	2.289	bilo_	1.286	e_bilo_	742	ki_se_	451	prilika_
53.	769	bi_	18.323	pr_	4.868	tako_	2.294	i_in_	1.281	njegov_	740	svoja_	448	a_in_se_
54.	655	j_	18.148	e_v_	4.864	e_in_	2.188	nila_	1.198	njegov_	739	rekel_	448	ki_se_
55.	645	od_	9.899	n_	4.824	e_la_	2.184	in_n_	1.168	l_	734	se_mu_	429	ki_se_je_
56.	567	lo_	9.776	a_p_	4.788	kor_	2.169	bila_	1.156	la_in_	729	bil_bil_	428	er_Majce_
57.	532	r_	9.568	ti_	4.662	se_	2.153	pred_	1.131	so_se_	729	je_pre_	428	er_Majce_
58.	413	v_	9.526	so_	4.659	akor_	2.151	se_p_	1.115	e_je_z_	721	sem_se_	428	Peter_M_
59.	378	ga_	9.458	bo_	4.648	za_	2.149	ki_s_	1.118	e_bila_	716	je_reke_	425	r_Majcen_

tem_trenutku_, štirinajsterčkov pa le še 5 (oziroma še manj): *?_je_vprašal_/_se_je_nasmehn/_e_je_nasmehnil/_se_je_nasmehni* in *_se_je_oglasil*. Splošnost prvega vzorca v celoti vzdrži do $n = 10$, ko opazimo prvi niz, ki se ga da umestiti v konkretno delo. Kot po navadi je iz prevoda: niz *_računalni* s frekvenco 818 je iz dela R. Cringelyja *Naključni imperiji* (ali *Kratka zgodovina Silicijeve doline*). Vendar so tudi še pri $n = 14$ le trije "konkretni" nizi med prvimi 60-timi: *Jurij_Ljudevit* iz novele Ivana Tavčarja: *Janez Sonce*, *_Passepartout_* iz prevoda Julesa Verna *V osemdesetih dneh okoli sveta* in *mikroročunalni* iz že omenjenega prevoda R. Cringelyja. Drugi vzorec je manjši in tako najdemo prvi konkretni niz, *Temnikar* že pri osmerčkih, pri štirinajsterčkih pa jih je kar 22, to je 37 %, od tega 16 iz *Balade o trobenti in oblaku*.

N-terčke znakov je mogoče uporabiti tudi za identifikacijo jezika, v katerem je besedilo. O tem je poročal Gregory Grefenstette v članku, kjer so bili primerjalno obdelani trojčki iz prvih milijon znakov enakih besedil v desetih jezikih Evropske unije, ki so izšla na cederomu (Grefenstette 1995). Jeziki so angleški, danski, francoski, nizozemski, italijanski, nemški, norveški, portugalski in švedski. V tabeli 33 je najpogostejših 12 trojčkov za slovenski jezik (oba vzorca skupaj) in omenjenih deset jezikov, kjer je bil kot vir upoštevan omenjeni članek.

Tabela 33: Najpogostejših 12 trojčkov v slovenskem in desetih jezikih EU

	Slo	Ang	Dan	Fra	Niz	Ita	Nem	Nor	Por	Špa	Šve
1.	je_	_th	er_	_de	en_	_di	en_	et_	_de	_de	en_
2.	_je	he_	en_	es_	de_	to_	er_	en_	de_	de_	er_
3.	_po	the	for	de_	_de	_de	_de	er_	os_	os_	et_
4.	_se	nd_	et_	ent	et_	di_	der	_de	do_	_la	tt_
5.	_pr	ed_	ing	nt_	an_	_co	ie_	_ha	que	el_	_de
6.	in_	_an	_fo	_le	n_d	la_	ich	an_	_qu	la_	ar_
7.	_in	and	_af	e_d	_he	re_	sch	de_	_co	que	för
8.	_na	_to	_de	le_	er_	ion	ein	det	as_	as_	om_
9.	se_	ing	nde	ion	_va	ent	che	ar_	ent	ue_	_oc
10.	_za	to_	els	s_d	van	e_d	die	_og	ao_	_qu	ch_
11.	_,k	ng_	lse	e_l	een	le_	ch_	og_	ue_	_co	de_
12.	_bi	er_	ret	_la	ver	o_d	den	te_	_a_	_en	och

Od 120 trojčkov neslovenskih jezikov iz zgornje tabele je kar 76 različnih. Tistih s frekvenco več kot 1 je 22: *_de* (9), *de_* (6), *er_* (6), *en_* (5), *et_* (4), *_co* (3), *ent* (3) in *_la*, *_qu*, *an_*, *ar_*, *as_*, *ch_*, *e_d*, *ing*, *ion*, *la_*, *le_*, *os_*, *que*, *to_*, *ue_* s frekvenco 2. Med vsemi 76 pa ni niti enega, ki bi ga našli tudi med najpogostejšimi slovenskimi. Zelo zanimiva bi bila še primerjava z jeziki, ki so slovenskemu bolj sorodni od zgornjih germanskih in romanskih, kot se je to delno posrečilo pri primerjavi najpogostejših besed.

5.2 Besede

Čeprav so osnovni gradniki pisanega besedila znaki, so besede tisto, kar nekako sami od sebe najprej povežemo z jezikom. Tudi obseg besedilnih zbirk je bil od prvega, Brownovega korpusa naprej vedno merjen v besedah. Pri dosedanjih kvantitativnih raziskavah slovenskega jezika, tudi zaradi velikosti besedilnih vzorcev, pri besedah ni bilo mogoče dlje kot do obdelave besednih oblik kot takih, samih zase. Iz tabele najpogostejših besednih n-terčkov na strani 70 je videti, da frekvence s povezovanjem več besed izredno hitro padajo in tako tudi ta, doslej največja besedilna zbirka nenadoma ni več tako velika, kot bi si človek želel. V tem razdelku so obravnavane besedne oblike (v nadaljnjem imenovane besede), v poglavju o oblikoslovnem označevanju pa še besedne leme, osnovne oblike, ki jih je avtor poiskal za besedila drugega vzorca. V tabeli 34 je navedenih nekaj splošnih podatkov o besedah iz obeh vzorcev. Po stolpcih si sledijo število vseh besed, število različnih besed, število enkratnic, to je besed, ki nastopajo samo enkrat, povprečna frekvenca besede in povprečna dolžina besed.

Tabela 34: Podatki o besedah v obeh vzorcih

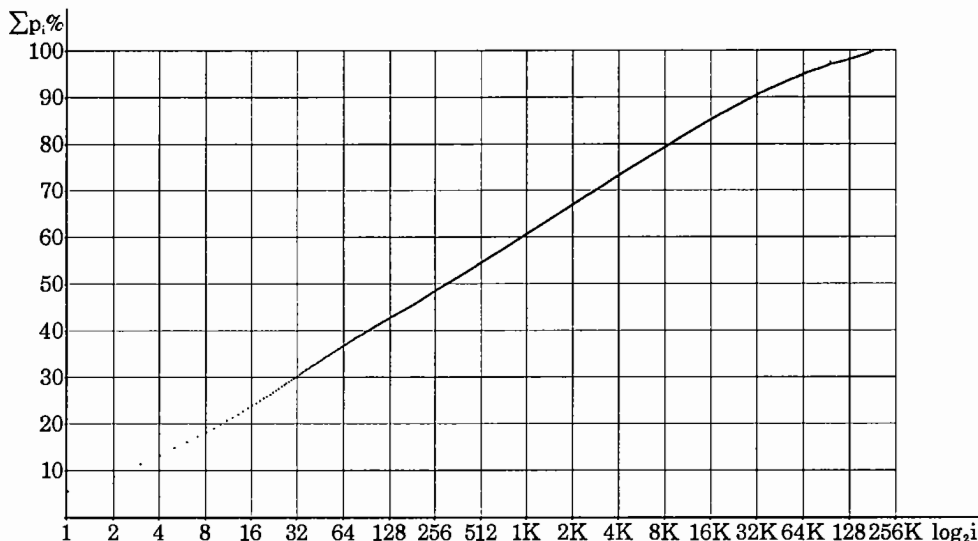
	Vseh besed	Različnih besed	Enkratnic	Povprečna frekvenca	Povprečna dolžina
Prvi vzorec	2.721.416	174.579	84.800	15,59	4,58
Drugi vzorec	407.938	40.475	20.110	10,08	4,38
Skupaj	3.129.354	184.052	87.812	17,01	4,55

Pri štetju različnih besed je bila začetnica besed ohranjena – besede, pisane z veliko začetnico na začetku povedi, so bile štete kot take in niso bile pretvorjene v malo. Ta postopek je bil opravljen kasneje, na besedah drugega vzorca, v poglavju o oblikoslovnem označevanju. Kot je videti iz tabele, se je v prvem vzorcu vsaka beseda pojavila povprečno 16 krat, v drugem pa 10 krat. Glede na to, da je prvi vzorec skoraj osemkrat večji od drugega, kaže ta razmeroma majhna razlika na dokaj veliko pestrost in majhno homogenost prvega. Tudi povprečna dolžina besed kaže na to, da je drugi vzorec, v celoti gledano, bolj gibek od prvega. Skupna povprečna dolžina je za 0.11 manjša od dolžine, ki so jo namerili raziskovalci s fakultete za elektrotehniko (Kristan idr. 1994). Opazen je tudi zelo velik delež besed, ki se pojavijo samo enkrat. V obeh vzorcih je takih med različnimi skoraj polovica.

5.2.1 Rast besednega zaklada

Kako hitro se polni besedni zaklad obeh vzorcev skupaj, se najbolje vidi iz krivulje rasti, ki je prikazana na sliki 11. Krivulja je pollogaritemska – na osi x so

Slika 11: Krivulja rasti za besede v obeh vzorcih: $y_i = \sum_{j=1}^i p_j$



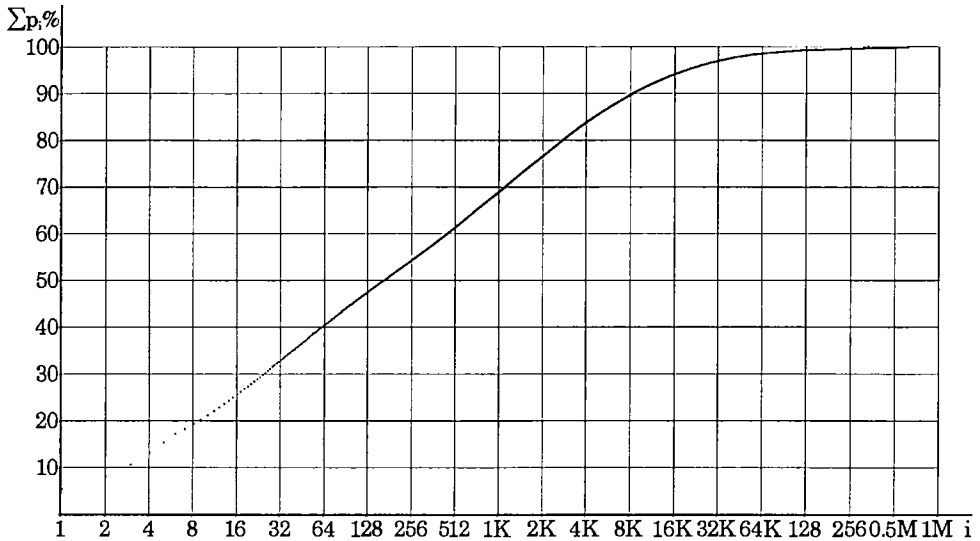
dvojiški logaritmi zaporednih števil besed v seznamu (i), ki je padajoče urejen po pogostostih besed (p_i) – najpogostejša beseda je prva, druga najpogostejša druga . . . , na osi y pa je skupna vsota frekvenc besed v odstotkih od celote. Za boljše ponazoritev so v tabeli 35 navedene relativne pogostosti (delež od celote) za najpo-

Tabela 35: Najpogostejših 64 besed iz obeh vzorcev z relativnimi frekvencami

je	5,63	ne	0,90	s	0,53	to	0,36	kakor	0,30	sta	0,24	me	0,20	nekaj	0,16
in	3,31	ki	0,84	po	0,51	bilo	0,36	ali	0,29	pri	0,23	več	0,18	naj	0,16
se	2,65	bi	0,80	še	0,51	od	0,34	jo	0,28	kar	0,23	samo	0,18	kako	0,16
v	1,86	z	0,70	tako	0,42	že	0,34	bila	0,28	kaj	0,22	ti	0,17	tem	0,16
da	1,56	za	0,68	bil	0,42	bo	0,33	mi	0,27	o	0,22	lahko	0,17	ji	0,15
na	1,24	sem	0,67	tudi	0,41	iz	0,32	In	0,26	če	0,21	ker	0,17	Ko	0,15
so	1,19	ni	0,66	si	0,41	ko	0,32	vse	0,26	do	0,21	pred	0,17	ob	0,15
pa	0,91	ga	0,59	mu	0,41	kot	0,31	jih	0,25	k	0,20	V	0,17	potem	0,14

gostejših 64 besed v seznamu obeh vzorcev. Prva pika na krivulji je delež besede *je* (5,63 %), druga vsota deležev besed *je* in *in* (8,94 %), tretja besed *je*, *in* in *se* (11,58 %) in tako naprej. Iz krivulje je videti, da s prvimi 16 besedami zajamemo 25 % vsega besedila, s 64 besedami (iz tabele 35) malo manj kot 40 % (37,08 %), z 256 pribl. 1/2, s 4.092 3/4, za 90 % pa je potrebnih že 32 K besed. Potek krivulje je dokaj regularen pribl. do $n = 32$, sledi hrbet v območju do $n = 128$ in spet lep potek do pribl. $n = 100.000$, potem pa je jasno viden zlom krivulje na mestu (označen je s kratko pokončno črtico), kjer nastopijo enkratnice, besede s frekvenco 1.

Zanimiva je primerjava te krivulje s krivuljo (Jakopin 1998), ki jo dobimo za besede v korpusu *Bank of English* (stanje iz leta 1996, seznam besed in frekvenc

Slika 12: Krivulja rasti za besede v *Bank of English* (1996)

je odstopil Jeremy Clear). V tem korpusu je bilo 203.648.478 besed, od tega 633.165 različnih in od teh 291.133 enkratnic. Povprečna frekvenca besede je bila 322, krivulja pa je navedena na sliki 12. Na njej opazimo v prvem delu podobno nihanje kot na sliki 11, le da je vzpon seveda hitrejši, 90 % doseže že pri 8 K in potem se asimptoti (100 %) zelo počasi približuje na skoraj tretjini svojega poteka. Nepravilnost na koncu, ki pa ni jasno vidna, spet povzročijo enkratnice.

Podobno kot n-terčke znakov je mogoče tudi najpogostejše besede uporabiti za identifikacijo jezika. V že omenjenem članku (Grefenstette 1995) je avtor za statistično ugotavljanje jezika uporabil najpogostejše kratke besede (do 5 črk). V tabeli 36 so navedene najpogostejše kratke besede za slovenski jezik, jezik Bosanskega korpusa na univerzi v Oslu (Leko 1998) in za 10 jezikov Evropske unije, istih kot v tabeli 33. Slovenske besede so vzete iz obeh vzorcev skupaj in so kar vse –

Tabela 36: Najpogostejših 12 kratkih besed (do 5 črk) v slovenskem in še enajstih jezikih

	Slo	Srh	Ang	Dan	Niz	Fra	Ita	Nem	Nor	Por	Špa	Šve
1.	je	i	the	i	de	de	di	der	og	de	de	och
2.	in	je	and	af	van	la	e	die	det	a	la	i
3.	se	u	to	og	het	le	il	und	han	que	que	att
4.	v	da	of	at	een	et	che	den	i	o	el	som
5.	da	se	a	til	en	des	la	in	er	e	en	en
6.	na	na	in	for	in	les	a	von	på	do	y	är
7.	so	su	was	en	dat	du	in	zu	til	da	a	på
8.	pa	ne	his	om	is	en	per	dem	at	no	los	det
9.	ne	a	that	der	te	un	del	für	som	um	del	av
10.	ki	to	I	er	op	que	un	mit	var	em	se	för
11.	bi	od	he	U	voor	a	non	das	jeg	para	por	med
12.	z	za	as	ikke	met	qui	i	des	med	com	las	den

med prvimi dvanajstimi namreč ni nobena daljša kot 2 črki. Tokrat je med 132 besedicami v drugih jezikih 92 različnih. Takih s frekvenco nad 1 je 24: *a* (6), *en* (5), *i* (5), *de* (4), *in* (4), *la* (3), *que* (3) in *at*, *da*, *del*, *den*, *der*, *des*, *det*, *e*, *er*, *med*, *og*, *på*, *se*, *som*, *til*, *to*, *un* s frekvenco 2. Kot je ugotovil že Grefenstette, so najpogostejši trojčki vseeno boljši indikator od najpogostejših besed. Presek med slovenskim in drugimi jeziki namreč pri besedah ni prazen – v njem najdemo šest pogostih besedic: *da*, *in*, *je*, *na*, *ne* in *se*. In se pojavi še pri angleškem, italijanskem in nemškem jeziku, *da* pri portugalskem in srbskem/hrvaškem, *se* pri srbskem/hrvaškem in španskem, *je*, *na* in *ne* pa še pri srbskem/hrvaškem.

5.2.2 Dolžine besed

Parameter, ki veliko pove o zvrsti besedila, so dolžine besed. V leposlovju so krajše kot v tehničnih besedilih, v delih z dialogom spet krajše kot v opisnih stavkih. V tabeli 37 so navedene povpr. dolžine vseh besed in razl. besed za oba vzorca.

Tabela 37: Povprečne dolžine vseh besed in različnih besed v obeh vzorcih

	Vseh besed	Povprečna dolžina	Standardni odklon	Različnih besed	Povprečna dolžina	Standardni odklon
Prvi vzorec	2.721.416	4,58	2.60	174.579	8,12	2.33
Drugi vzorec	407.938	4,38	2.44	40.475	7,44	2.12
Skupaj	3.129.354	4,55	2.57	184.052	8,00	2.31

Še posebej pri povprečnih dolžinah različnih besed je zelo očitna razlika med prvim in drugim vzorcem. Tega ni več mogoče pripisati večji količini premega govora v Kosmačevem opusu, ampak tudi temu, da je njegova pripoved bolj tekoča (Štih 1958). Dobljene vrednosti za različne besede so manjše od tistih za samostalnike iz Slovarja slovenskega knjižnega jezika (Jakopin 1996b).

V tabelah 38, 39 in 40 so navedene najdaljše besede v obeh vzorcih in v *Bank of English* – pred vsako besedo je izpisana njena dolžina. V tabeli 39 so bila izvzeta večbesedna imena, v tabeli 40 pa je bilo že v viru vse pisano z malo začetnico. Po

Tabela 38: Najdaljše besede v prvem vzorcu

1. 57 prijatelj-sovražnik-prijateljsovražnik-sovražnikprijatelj
2. 45 klik-klik-klik-klik-klik-klik-klik-klik-TRESK
3. 37 ti-hojlarija-hojlarija-hojlarija-drom
4. 28 hodi-hodi-prestopi-hodi-hodi
5. 27 ti-hojlarija-hojlarija-drom
6. 27 petnajststošestinetdesetem
7. 26 pasti-drug-drugemu-v-objem
8. 26 petstošestinsedemdesetkrat
9. 26 petnajstimi-minutami-slave
10. 26 sedemstodevetindvajsetkrat

Tabela 39: Najdaljše besede v drugem vzorcu

1. 30 tisočdevetstodvaintridesetkrat
2. 22 stoštiriinštiridesetih
3. 20 petintridesetletnega
4. 19 literarnokritičnemu
5. 19 petinosemdesetletni
6. 19 dvaindvajsetletnemu
7. 19 kakršnegakolisibodi
8. 18 triinštiridesetega
9. 18 javolheršturmfirer
10. 18 sedemstosedemdeset

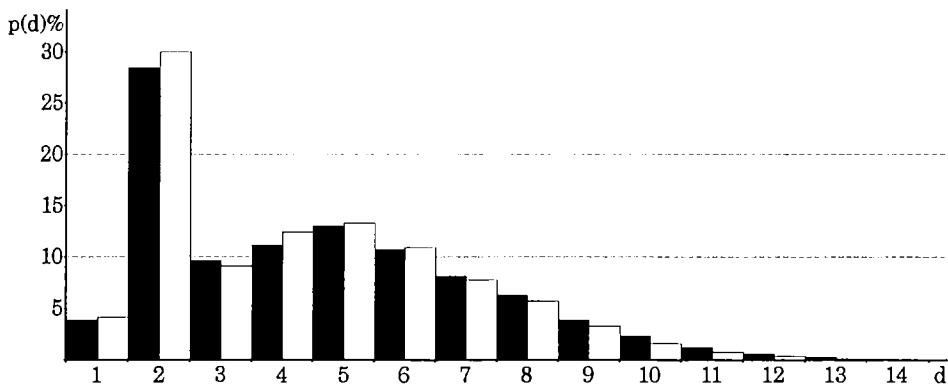
Tabela 40: Najdaljše besede v *Bank of English*

1. 65 confidenceunderstandingcreativitycollaborationexcellenceenjoyment
2. 60 weeeeeeeeeaaaaarrdddrrraaaaaannnnnggggnnnneeeeeeaaarrrrrrwwwww
3. 58 llanfairpwyllgyngyllgogerychwyrndrobwlllantysilogogoch
4. 58 officialpeacehappinesstrudelbeerflagscelebrationshysteria
5. 58 whatchurchfestivalinfebruaryiscelebratedwithlightedcandles
6. 54 fightthepowerlouderthanabombblacksteelintheourofchaos
7. 53 verordnetenfriedefreudestrudelbierfahnenfeierhysterie
8. 52 dandybeanotoppervictorhotspurhornethurricanianabun
9. 52 takethepowerbackbombtrackfistfulofsteelknowyourenemy
10. 43 pneumoultramicroscopicsilicovolcanoconiosis

pričakovanju je povsod najti predvsem večbesedne sestavljenke, če sem štejemo tudi števila. Teh je v tabeli prvega vzorca 3, drugega 7, med najdaljšimi besedami zbirke *Bank of English* pa prav na vrh niso mogla.

Zanimive so še porazdelitve dolžin besed, vseh na sliki 13 in različnih na sliki 14. Stolpci prvega vzorca so spet označeni s črno barvo, drugega pa z belo.

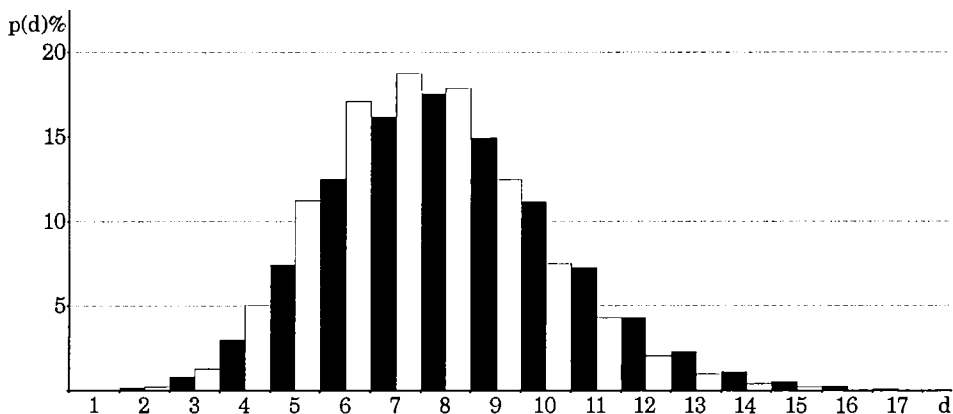
Slika 13: Porazdelitev dolžin vseh besed v obeh vzorcih



Do vključno dolžine 6 so v vseh primerih, razen pri dolžini 3, vrednosti v drugem vzorcu večje kot v prvem, potem pa ves čas do konca nižje. Ker je tudi povprečna

dolžina besed v drugem vzorcu manjša (glej tabelo 31) je tak potek pričakovan. Konica pri dolžini 2 (vezniki, pomožni glagol *je*) je pri obeh vzorcih bistveno višja, kot so jo namerili v predhodni raziskavi (Kristan idr. 1994), sicer pa je potek podoben. Razliko gre pripisati bolj tekočemu jeziku v vzorcih tega dela; predhodni vzorec je vseboval tudi časopisni jezik.

Slika 14: Porazdelitev dolžin različnih besed v obeh vzorcih



Tokrat imamo opravka s precej pravilnima krivuljama, ki pa nimata vrha na istem x . Največ različnih besed pri prvem vzorcu je dolgih 8 črk, pri drugem pa 7.

5.2.3 Najpogostejše besede

V tabelah 42, 43, 44 in 45 na naslednjih straneh so najpogostejše besede iz prvega in drugega vzorca, urejene po frekvencah in po abecedi. Tabeli 42 in 34 odkrivata 250 najpogostejših besednih oblik v vsakem vzorcu. Ker je bil oblikoslovno označen in lematiziran samo drugi vzorec, je treba nekaj sklepov potegniti že iz teh podatkov. Izkaže se, da je prvih 6 besed v obeh seznamih (*je*, *in*, *se*, *v*, *da* in *na*) istih in na istih mestih (brez upoštevanja mest 9 besed), da obsega presek najpogostejših 50 besed v obeh vzorcih 42 besed, presek 100 najpogostejših 86 besed, prvih 250 pa 180 besed. Jezik Cirila Kosmača se je torej razmeroma dobro ujel z jezikom v prvem vzorcu.

Pogled na tabelo 42 tudi razkrije, da je našlo pot v seznam prvih 250 besed eno samo ime (*Bog* na 147. mestu), ki pa tudi ni značilno samo za eno delo. Absolutne frekvence se začnejo s 150.984 in končajo z 953, relativne frekvence (delež celote) se začnejo pa na 5,55 % (*je*) in končajo na 0,0350 % (*kdaj*). Samostalnikov, glagolov in pridevnikov je malo, prevladujejo zainki, prislovi, vezniki in predlogi.

Drugi vzorec je veliko manjši in se zato ne more pohvaliti s tako splošnim seznamom najpogostejših besed. V tabeli 43 zato najdemo imena *Peter, Martin, Matic, Sova, Majcen, Drejc, Temnikar, Tantadruj, Stane* in *Venc*, ki jih z lahkoto umestimo v konkretno črtico ali roman. Absolutne frekvence se začnejo pri 25.798 in končajo pri 161, relativne pa pri 6,32 % in končajo pri 0,0395 %, kar je oboje precej nad vrednostmi v prvem vzorcu. Delež samostalnikov, glagolov in pridevnikov je večji kot pri prvem vzorcu, a še vedno majhen.

V tabeli 41 je navedenih najpogostejših 20 samostalnikov iz obeh vzorcev, pri čemer so bili za prvi vzorec le ocenjeni (iz oblik v tabeli 42), za drugi vzorec pa

Tabela 41: Najpogostejši samostalniki iz vsakega vzorca s frekvencami

1. dan	3.300	1. roka	1.577
2. ljudje	2.939	2. glava	985
3. roka	2.895	3. oči	833
4. čas	2.792	4. otrok	821
5. oči	2.584	5. dan	749
6. leto	2.068	6. hiša	700
7. človek	1.683	7. leto	628
8. oče	1.674	8. vrata	536
9. glava	1.583	9. beseda	513
10. obraz	1.369	10. oče	486
11. gospod	1.354	11. človek	453
12. življenje	1.337	12. glas	438
13. stran	1.298	13. srce	424
14. pot	1.267	14. vas	414
15. glas	1.173	15. obraz	396
16. noč	1.092	16. miza	392
17. beseda	1.071	17. noga	384
18. mož	1.052	18. življenje	373
19. mati	1.033	19. ljudje	369
20. svet	997	20. voda	362

so mesta in frekvence točni, saj izhajajo iz lematiziranega besedila. V obeh seznamih nastopa polovica, to je 10 samostalnikov: *beseda, človek, dan, glava, leto, ljudje, oče, oči, roka*, in *življenje*. Enajsti bi utegnil biti *vas*, a ga v prvem vzorcu ni bilo mogoče enostavno ločiti od enakega zaimka.

V tabelah 44 in 45 so iste besede kot v tabelah 42 in 43, le da so tokrat urejene abecedno in so jim v oklepaju dodani še rangi iz frekvenčno urejenih seznamov. Tako je mogoče za iskano besedo iz prvega ali drugega vzorca takoj ugotoviti, ali je med prvimi 250 in kateri je njen rang po pogostosti.

Tabela 42: Najpogostejše besedne oblike iz prvega vzorca s frekvencami

1. je	150.984	51. lahko	5.318	101. imel	2.334	151. on	1.484	201. glas	1.173
2. in	93.314	52. potem	5.308	102. ste	2.226	152. sicer	1.484	202. moral	1.173
3. se	69.885	53. nekaj	5.211	103. vas	2.200	153. toliko	1.480	203. rad	1.173
4. v	55.564	54. več	5.192	104. dan	2.182	154. boš	1.466	204. vsaj	1.169
5. da	45.181	55. pred	5.043	105. vam	2.180	155. nam	1.460	205. pač	1.152
6. na	36.699	56. ker	5.029	106. bodo	2.125	156. vseh	1.437	206. njegov	1.151
7. so	33.040	57. k	5.010	107. zaradi	2.101	157. dolgo	1.428	207. vsem	1.150
8. ne	27.850	58. naj	4.993	108. kdo	2.092	158. moj	1.420	208. misli	1.143
9. pa	25.643	59. le	4.711	109. njim	2.082	159. dejal	1.419	209. seboj	1.138
10. ki	23.277	60. ob	4.587	110. čez	2.061	160. časa	1.418	210. sedaj	1.131
11. bi	21.973	61. tem	4.495	111. veliko	2.058	161. sebi	1.403	211. tisto	1.128
12. za	20.480	62. zdaj	4.443	112. niti	1.996	162. drugega	1.400	212. njem	1.123
13. z	20.260	63. med	4.320	113. tu	1.971	163. rekla	1.398	213. ravno	1.119
14. ni	19.125	64. ji	4.271	114. treba	1.959	164. čas	1.374	214. dni	1.118
15. sem	18.517	65. ta	4.179	115. zakaj	1.951	165. ljudi	1.373	215. skupaj	1.115
16. ga	15.576	66. prav	4.022	116. nad	1.936	166. obraz	1.369	216. let	1.112
17. še	14.788	67. smo	3.887	117. sva	1.886	167. gospod	1.354	217. teh	1.097
18. po	14.748	68. nič	3.760	118. nas	1.860	168. nazaj	1.350	218. noč	1.092
19. s	14.382	69. te	3.719	119. celo	1.834	169. življenje	1.337	219. svojih	1.084
20. tako	13.818	70. a	3.679	120. zelo	1.824	170. danes	1.332	220. svoji	1.082
21. ko	13.361	71. jaz	3.541	121. ima	1.806	171. šel	1.320	221. takrat	1.078
22. tudi	13.177	72. bili	3.465	122. ves	1.792	172. mogoče	1.313	222. drugo	1.074
23. to	13.054	73. vendar	3.421	123. prišel	1.784	173. vedel	1.309	223. besede	1.071
24. bil	12.332	74. tega	3.279	124. nikoli	1.781	174. kadar	1.302	224. njimi	1.071
25. ali	11.557	75. rekel	3.267	125. drugi	1.762	175. strani	1.298	225. doma	1.063
26. si	11.402	76. svoje	3.245	126. torej	1.707	176. skozi	1.297	226. kmalu	1.056
27. mu	10.940	77. jim	3.083	127. tedaj	1.690	177. roke	1.293	227. dokler	1.053
28. od	10.488	78. toda	3.051	128. človek	1.683	178. no	1.292	228. mož	1.052
29. bilo	10.365	79. zato	3.015	129. videl	1.675	179. njega	1.276	229. more	1.034
30. kot	9.877	80. bolj	3.001	130. oče	1.674	180. vprašal	1.276	230. mati	1.033
31. že	9.537	81. tam	3.000	131. morda	1.643	181. vem	1.271	231. vsa	1.027
32. iz	9.474	82. ter	2.987	132. skoraj	1.637	182. pot	1.267	232. daleč	1.026
33. kaj	9.066	83. saj	2.916	133. vsak	1.623	183. namreč	1.259	233. hitro	1.023
34. bo	8.806	84. pod	2.885	134. sploh	1.612	184. njih	1.247	234. šele	1.016
35. če	8.613	85. brez	2.873	135. roko	1.602	185. dva	1.244	235. imela	1.005
36. vse	8.149	86. sam	2.814	136. glavo	1.583	186. mora	1.236	236. mene	1.005
37. bila	7.957	87. proti	2.774	137. nato	1.575	187. gre	1.227	237. dal	1.003
38. kakor	7.943	88. res	2.746	138. mogel	1.571	188. njo	1.217	238. svet	997
39. mi	7.911	89. niso	2.670	139. ljudje	1.566	189. njegov	1.214	239. ona	993
40. pri	7.530	90. oči	2.584	140. biti	1.559	190. svoj	1.210	240. pride	979
41. jo	7.513	91. vedno	2.550	141. naprej	1.558	191. kje	1.203	241. tri	976
42. kar	7.069	92. bom	2.543	142. takoj	1.557	192. vi	1.199	242. vso	970
43. jih	7.048	93. vsi	2.495	143. tisti	1.556	193. čeprav	1.192	243. eno	969
44. sta	6.720	94. dobro	2.494	144. seveda	1.553	194. čisto	1.192	244. sebe	967
45. o	6.155	95. spet	2.455	145. svojega	1.541	195. okoli	1.189	245. najbolj	964
46. do	6.008	96. kjer	2.448	146. bile	1.531	196. pravi	1.186	246. videti	961
47. ti	5.719	97. nisem	2.440	147. Bog	1.522	197. lepo	1.185	247. prvi	960
48. kako	5.674	98. ampak	2.425	148. prej	1.508	198. sama	1.184	248. moje	958
49. samo	5.620	99. svojo	2.410	149. meni	1.498	199. kajti	1.182	249. leta	956
50. me	5.465	100. malo	2.388	150. hotel	1.493	200. nihče	1.176	250. kdaj	953

Tabela 43: Najpogostejše besedne oblike iz drugega vzorca s frekvenca

1. je	25.798	51. pri	899	101. stopil	361	151. njej	233	201. Drejc	190
2. in	18.471	52. mi	859	102. nad	357	152. svoj	233	202. vem	189
3. se	13.330	53. spet	857	103. naglo	357	153. vedel	231	203. življenje	189
4. v	7.809	54. ter	852	104. bolj	352	154. ozrl	230	204. noge	185
5. da	5.412	55. ti	847	105. Martin	348	155. teta	230	205. Temnikar	185
6. na	5.124	56. kako	845	106. dobro	347	156. dni	228	206. domov	184
7. pa	4.625	57. prav	831	107. nisem	346	157. kadar	228	207. vanj	184
8. so	4.243	58. me	803	108. oče	340	158. ljudje	228	208. bomo	183
9. ne	3.695	59. sta	794	109. tega	339	159. včasih	228	209. glasom	183
10. bi	3.221	60. pred	765	110. vprašal	339	160. povedal	227	210. kam	181
11. z	3.181	61. potem	754	111. seveda	338	161. nemara	226	211. župnik	181
12. ga	3.039	62. res	730	112. videl	332	162. kje	225	212. sama	180
13. ki	2.995	63. do	696	113. bili	329	163. nikdar	225	213. stric	179
14. po	2.948	64. ob	693	114. dan	327	164. zaradi	225	214. potegnil	177
15. sem	2.812	65. več	671	115. šel	323	165. hišo	224	215. sva	177
16. ni	2.636	66. glavo	661	116. otrok	307	166. imel	222	216. smrt	176
17. s	2.441	67. oči	601	117. bodo	303	167. zmeraj	222	217. šele	176
18. še	2.410	68. počasi	595	118. brez	303	168. hitro	219	218. mizo	173
19. za	2.327	69. pod	583	119. zelo	303	169. nas	218	219. bile	172
20. tako	2.155	70. naj	576	120. Matic	301	170. prišel	217	220. glas	172
21. kakor	2.079	71. ji	570	121. skozi	301	171. obraz	216	221. noč	172
22. mu	1.952	72. svoje	567	122. niso	296	172. pravi	216	222. Tantadruj	172
23. že	1.828	73. nekaj	552	123. začel	289	173. vasi	216	223. vrnil	172
24. tudi	1.799	74. roko	539	124. vrata	285	174. sebi	215	224. prikimal	171
25. bo	1.776	75. ta	539	125. dolgo	284	175. dva	214	225. srce	171
26. si	1.735	76. tam	521	126. kdo	283	176. rad	211	226. nam	170
27. bil	1.695	77. zakaj	516	127. okrog	281	177. hiše	210	227. široko	170
28. to	1.528	78. lahko	513	128. človek	278	178. komaj	210	228. kdaj	169
29. iz	1.470	79. nič	504	129. doma	272	179. kot	208	229. očmi	169
30. kaj	1.366	80. no	500	130. Sova	272	180. pogleda	208	230. misli	168
31. k	1.355	81. roke	497	131. Majcen	270	181. čeprav	207	231. njem	168
32. zdaj	1.350	82. bom	490	132. glasno	268	182. obrnil	207	232. njemu	168
33. vse	1.333	83. te	487	133. jaz	266	183. sredi	207	233. svoji	168
34. bilo	1.289	84. smo	479	134. jim	266	184. sebe	204	234. danes	167
35. ko	1.260	85. vsi	475	135. ves	266	185. kjer	203	235. on	167
36. jo	1.246	86. svojo	463	136. celo	265	186. pob	202	236. sonce	167
37. saj	1.151	87. sam	458	137. let	263	187. svojega	202	237. obstal	166
38. samo	1.129	88. takoj	457	138. takrat	261	188. šla	202	238. svojih	166
39. toda	1.127	89. zato	455	139. besede	259	189. enkrat	201	239. dve	165
40. bila	1.106	90. med	448	140. vsak	255	190. mimo	201	240. mati	165
41. a	1.103	91. Peter	438	141. pogledal	252	191. otroci	201	241. sedel	165
42. ker	1.060	92. rekla	437	142. niti	251	192. kmalu	200	242. ponovil	164
43. če	1.052	93. njim	432	143. ima	245	193. tisti	199	243. Stane	164
44. kar	1.027	94. le	429	144. mama	244	194. mirno	197	244. tedaj	164
45. o	1.004	95. tem	429	145. vas	244	195. reče	197	245. tu	163
46. nato	993	96. skoraj	424	146. lepo	243	196. stari	196	246. moj	162
47. od	952	97. boš	389	147. naprej	238	197. tiho	195	247. mora	162
48. ali	947	98. čez	380	148. treba	236	198. tri	193	248. Venc	162
49. rekel	947	99. proti	379	149. sicer	235	199. globoko	192	249. govoril	161
50. jih	924	100. vendar	371	150. vsa	234	200. prišla	192	250. rokami	161

Tabela 44: Abecedno urejene pogoste besedne oblike iz prvega vzorca z rangi in frekvenca

a (70)	3.679	ima (121)	1.806	mož (228)	1.052	pri (40)	7.530	takrat (221)	1.078
ali (25)	11.557	imel (101)	2.334	mu (27)	10.940	pride (240)	979	tam (81)	3.000
ampak (98)	2.425	imela (235)	1.005	na (6)	36.699	prišel (123)	1.784	te (69)	3.719
besede (223)	1.071	in (2)	93.314	nad (116)	1.936	proti (87)	2.774	tedaj (127)	1.690
bi (11)	21.973	iz (32)	9.474	naj (58)	4.993	prvi (247)	960	tega (74)	3.279
bil (24)	12.332	jaz (71)	3.541	najbolj (245)	964	rad (203)	1.173	teh (217)	1.097
bila (37)	7.957	je (1)	150.984	nam (155)	1.460	ravno (213)	1.119	tem (61)	4.495
bile (146)	1.531	ji (64)	4.271	namreč (183)	1.259	rekel (75)	3.267	ter (82)	2.987
bili (72)	3.465	jih (43)	7.048	naprej (141)	1.558	rekla (163)	1.398	ti (47)	5.719
bilo (29)	10.365	jim (77)	3.083	nas (118)	1.860	res (88)	2.746	tisti (143)	1.556
biti (140)	1.559	jo (41)	7.513	nato (137)	1.575	roke (177)	1.293	tisto (211)	1.128
bo (34)	8.806	k (57)	5.010	nazaj (168)	1.350	roko (135)	1.602	to (23)	13.054
bodo (106)	2.125	kadar (174)	1.302	ne (8)	27.850	s (19)	14.382	toda (78)	3.051
Bog (147)	1.522	kaj (33)	9.066	nekaj (53)	5.211	saj (83)	2.916	toliko (153)	1.480
bolj (80)	3.001	kajti (199)	1.182	ni (14)	19.125	sam (86)	2.814	torej (126)	1.707
bom (92)	2.543	kako (48)	5.674	nič (68)	3.760	sama (198)	1.184	treba (114)	1.959
boš (154)	1.466	kakor (38)	7.943	nihče (200)	1.176	samo (49)	5.620	tri (241)	976
brez (85)	2.873	kar (42)	7.069	nikoli (124)	1.781	se (3)	69.885	tu (113)	1.971
celo (119)	1.834	kdaj (250)	953	nisem (97)	2.440	sebe (244)	967	tudi (22)	13.177
čas (164)	1.374	kdo (108)	2.092	niso (89)	2.670	sebi (161)	1.403	v (4)	55.564
časa (160)	1.418	ker (56)	5.029	niti (112)	1.996	seboj (209)	1.138	vam (105)	2.180
če (35)	8.613	ki (10)	23.277	njega (179)	1.276	sedaj (210)	1.131	vas (103)	2.200
čeprav (193)	1.192	kje (191)	1.203	njegov (206)	1.151	sem (15)	18.517	več (54)	5.192
čez (110)	2.061	kjer (96)	2.448	njegovo (189)	1.214	seveda (144)	1.553	vedel (173)	1.309
čisto (194)	1.192	kmalu (226)	1.056	njem (212)	1.123	si (26)	11.402	vedno (91)	2.550
človek (128)	1.683	ko (21)	13.361	njih (184)	1.247	sicer (152)	1.484	veliko (111)	2.058
da (5)	45.181	kot (30)	9.877	njim (109)	2.082	skoraj (132)	1.637	vem (181)	1.271
dal (237)	1.003	lahko (51)	5.318	njimi (224)	1.071	skozi (176)	1.297	vendar (73)	3.421
daleč (232)	1.026	le (59)	4.711	njo (188)	1.217	skupaj (215)	1.115	ves (122)	1.792
dan (104)	2.182	lepo (197)	1.185	no (178)	1.292	smo (67)	3.887	vi (192)	1.199
danes (170)	1.332	let (216)	1.112	noč (218)	1.092	so (7)	33.040	videl (129)	1.675
dejal (159)	1.419	leta (249)	956	o (45)	6.155	spet (95)	2.455	videti (246)	961
dni (214)	1.118	ljudi (165)	1.373	ob (60)	4.587	sploh (134)	1.612	vprašal (180)	1.276
do (46)	6.008	ljudje (139)	1.566	obraz (166)	1.369	sta (44)	6.720	vsa (231)	1.027
dobro (94)	2.494	malo (100)	2.388	oče (130)	1.674	ste (102)	2.226	vsaj (204)	1.169
dokler (227)	1.053	mati (230)	1.033	oči (90)	2.584	strani (175)	1.298	vsak (133)	1.623
dolgo (157)	1.428	me (50)	5.465	od (28)	10.488	sva (117)	1.886	vse (36)	8.149
doma (225)	1.063	med (63)	4.320	okoli (195)	1.189	svet (238)	997	vseh (156)	1.437
drugega (162)	1.400	mene (236)	1.005	on (151)	1.484	svoj (190)	1.210	vsem (207)	1.150
drugi (125)	1.762	meni (149)	1.498	ona (239)	993	svoje (76)	3.245	vsi (93)	2.495
drugo (222)	1.074	mi (39)	7.911	pa (9)	25.643	svojega (145)	1.541	vso (242)	970
dva (185)	1.244	misli (208)	1.143	pač (205)	1.152	svoji (220)	1.082	z (13)	20.260
eno (243)	969	mogel (138)	1.571	po (18)	14.748	svojih (219)	1.084	za (12)	20.480
ga (16)	15.576	mogoče (172)	1.313	pod (84)	2.885	svojo (99)	2.410	zakaj (115)	1.951
glas (201)	1.173	moj (158)	1.420	pot (182)	1.267	še (17)	14.788	zaradi (107)	2.101
glavo (136)	1.583	moje (248)	958	potem (52)	5.308	šel (171)	1.320	zato (79)	3.015
gospod (167)	1.354	mora (186)	1.236	prav (66)	4.022	šele (234)	1.016	zdaj (62)	4.443
gre (187)	1.227	moral (202)	1.173	pravi (196)	1.186	ta (65)	4.179	zelo (120)	1.824
hitro (233)	1.023	morda (131)	1.643	pred (55)	5.043	tako (20)	13.818	že (31)	9.537
hotel (150)	1.493	more (229)	1.034	prej (148)	1.508	takoj (142)	1.557	življenje (169)	1.337

Tabela 45: Abecedno urejene pogoste besedne oblike iz drugega vzorca z rangi in frekvenčami

a (41)	1.103	ji (71)	570	nekaj (73)	552	reče (195)	197	takrat (138)	261
ali (48)	947	jih (50)	924	nemara (161)	226	rekel (49)	947	tam (76)	521
besede (139)	259	jim (134)	266	ni (16)	2.636	rekla (92)	437	Tantadruj (222)	172
bi (10)	3.221	jo (36)	1.246	nič (79)	504	res (62)	730	te (83)	487
bil (27)	1.695	k (31)	1.355	nikdar (163)	225	rokami (250)	161	tedaj (244)	164
bila (40)	1.106	kadar (157)	228	nisem (107)	346	roke (81)	497	tega (109)	339
bile (219)	172	kaj (30)	1.366	niso (122)	296	roko (74)	539	tem (95)	429
bili (113)	329	kako (56)	845	niti (142)	251	s (17)	2.441	Temnikar (205)	185
bilo (34)	1.289	kakor (21)	2.079	njej (151)	233	saj (37)	1.151	ter (54)	852
bo (25)	1.776	kam (210)	181	njem (231)	168	sam (87)	458	teta (155)	230
bodo (117)	303	kar (44)	1.027	njemu (232)	168	sama (212)	180	ti (55)	847
bolj (104)	352	kdaj (228)	169	njim (93)	432	samo (38)	1.129	tih (197)	195
bom (82)	490	kdo (126)	283	no (80)	500	se (3)	13.330	tisti (193)	199
bomo (208)	183	ker (42)	1.060	noč (221)	172	sebe (184)	204	to (28)	1.528
boš (97)	389	ki (13)	2.995	noge (204)	185	sebi (174)	215	toda (39)	1.127
brez (118)	303	kje (162)	225	o (45)	1.004	sedel (241)	165	treba (148)	236
celo (136)	265	kjer (185)	203	ob (64)	693	sem (15)	2.812	tri (198)	193
če (43)	1.052	kmalu (192)	200	obraz (171)	216	seveda (111)	338	tu (245)	163
čeprav (181)	207	ko (35)	1.260	obrnil (182)	207	si (26)	1.735	tudi (24)	1.799
čez (98)	380	komaj (178)	210	obstal (237)	166	sicer (149)	235	v (4)	7.809
človek (128)	278	kot (179)	208	oče (108)	340	skoraj (96)	424	vanj (207)	184
da (5)	5.412	lahko (78)	513	oči (67)	601	skozi (121)	301	vas (145)	244
dan (114)	327	le (94)	429	očmi (229)	169	smo (84)	479	vasi (173)	216
danes (234)	167	lepo (146)	243	od (47)	952	smrt (216)	176	včasih (159)	228
dni (156)	228	let (137)	263	okrog (127)	281	so (8)	4.243	več (65)	671
do (63)	696	ljudje (158)	228	on (235)	167	sonce (236)	167	vedel (153)	231
dobro (106)	347	Majcen (131)	270	otroci (191)	201	Sova (130)	272	vem (202)	189
dolgo (125)	284	mama (144)	244	otrok (116)	307	spet (53)	857	Venc (248)	162
doma (129)	272	Martin (105)	348	ozrl (154)	230	srce (225)	171	vendar (100)	371
domov (206)	184	mati (240)	165	pa (7)	4.625	sredi (183)	207	ves (135)	266
Drejc (201)	190	Matic (120)	301	Peter (91)	438	sta (59)	794	videl (112)	332
dva (175)	214	me (58)	803	po (14)	2.948	Stane (243)	164	vprašal (110)	339
dve (239)	165	med (90)	448	pob (186)	202	stari (196)	196	vrata (124)	285
enkrat (189)	201	mi (52)	859	počasi (68)	595	stopil (101)	361	vrnil (223)	172
ga (12)	3.039	mimo (190)	201	pod (69)	583	stric (213)	179	vsa (150)	234
glas (220)	172	mirno (194)	197	pogleda (180)	208	sva (215)	177	vsak (140)	255
glasno (132)	268	misl (230)	168	pogledal (141)	252	svoj (152)	233	vse (33)	1.333
glasom (209)	183	mizo (218)	173	ponovil (242)	164	svoje (72)	567	vsi (85)	475
glavo (66)	661	moj (246)	162	potegnil (214)	177	svojega (187)	202	z (11)	3.181
globoko (199)	192	mora (247)	162	potem (61)	754	svoji (233)	168	za (19)	2.327
govoril (249)	161	mu (22)	1.952	povedal (160)	227	svojih (238)	166	začel (123)	289
hiše (177)	210	na (6)	5.124	prav (57)	831	svojo (86)	463	zakaj (77)	516
hišo (165)	224	nad (102)	357	pravi (172)	216	še (18)	2.410	zaradi (164)	225
hitro (168)	219	naglo (103)	357	pred (60)	765	šel (115)	323	zato (89)	455
ima (143)	245	naj (70)	576	pri (51)	899	šele (217)	176	zdaj (32)	1.350
imel (166)	222	nam (226)	170	prikimal (224)	171	široko (227)	170	zelo (119)	303
in (2)	18.471	naprej (147)	238	prišel (170)	217	šla (188)	202	zmeraj (167)	222
iz (29)	1.470	nas (169)	218	prišla (200)	192	ta (75)	539	že (23)	1.828
jaz (133)	266	nato (46)	993	proti (99)	379	tako (20)	2.155	življenje (203)	189
je (1)	25.798	ne (9)	3.695	rad (176)	211	takoj (88)	457	župnik (211)	181

5.2.4 Besedni n-terčki

Pestrost jezika se kaže prav v veliki svobodi zlaganja besed med seboj in zato raziskave besednih n-terčkov oziroma fraz, kot bi jim rekli jezikoslovci, zahtevajo zelo velike besedilne vzorce. Če je bilo že pri najpogostejših znakovnih n-terčkih z rastočim n opaziti skokovito upadanje frekvenc, to razumljivo še veliko bolj velja za besedne n-terčke. Predvsem prvi vzorec iz tega dela pa je že tako velik, da je smiselnih nekaj korakov tudi v tej smeri.

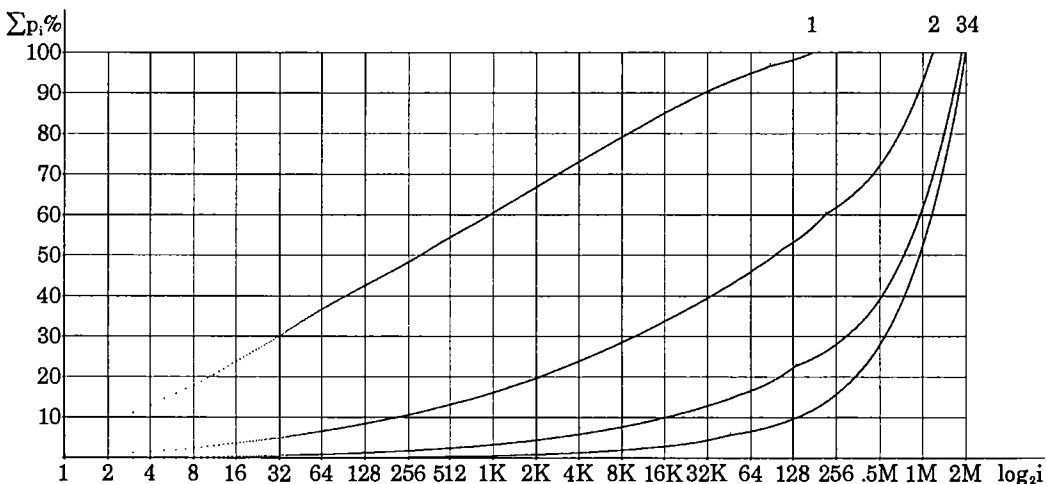
V tabeli 46 je naveden pregled števila besednih n-terčkov prvega vzorca za n od 1 do 5, pri čemer je bila tabela pri znakovnih n-terčkih, tvorjena iz besedil, že razrezanih na povedi in očiščenih oznak in tujih citatov.

Tabela 46: Število besednih n-terčkov (1-5) v prvem vzorcu

n	Vseh n-terčkov	Različnih n-terčkov	Enkratnic	Delež enkr. med različnimi v %	Delež enkr. med vsemi n-terčki v %
1	2.721.416	174.579	84.800	48,57	3,12
2	2.523.050	1.218.506	999.475	82,02	39,61
3	2.326.469	1.930.363	1.793.579	92,91	77,09
4	2.136.579	2.057.833	2.010.939	97,72	94,12
5	1.955.526	1.937.820	1.923.628	99,27	98,37

Opaziti je predvsem zelo hitro rast deleža n-terčkov s frekvenco 1 (enkratnic) z večanjem števila n. Pri peterčkih je delež enkratnic med vsemi že 98,37 %, kar pomeni, da je fraz z dolžino pet besed že zelo malo. Podobo dopolnjuje še krivulja rasti za besede, dvojčke, trojčke in četverčke besed na sliki 15.

Slika 15: Krivulje rasti za besedne n-terčke (1-4) v prvem vzorcu



Prva krivulja z leve, za besede, ima do približno 97 %, kjer nastopijo enkratnice, razmeroma regularen potek. Z najpogostejšimi 16 besedami pokrijemo 25 % celega vzorca, z 256 slabo polovico, s 4.096 približno 75 % in z 32.768 dobrih 90 %. Pri dvojčkih, naslednji krivulji, potrebujemo za pokritje 25 % besedila že 4 K najpogostejših besednih parov, za 50 % besedila nekaj manj kot 100.000, za 75 % več kot 512 K, za 90 % pa že skoraj milijon. Pri trojčkih je treba za 25 % besedila vse besedne trojčke s frekvenco nad 1 (več kot 128.000), za 50 % že približno 700.000 in tako naprej. Pri četverčkih je še precej slabše – z vsemi, ki imajo frekvenco nad 1, pridemo komaj do 6 % celotnega besedila (mesta, kjer se začnejo enkratnice, so na krivulji označena s kratko pokončno črtico nad ustrežno točko).

V tabelah 47 in 48 na naslednjih straneh so navedeni najpogostejši besedni n-terčki ($n = 2-5$) za oba vzorca. V njih so močno udeležene najpogostejše besede: pri prvem vzorcu nastopa pomožni glagol *je* v 54 % dvojčkov, 68 % trojčkov, 74 % četverčkov in 78 % peterčkov. Predvsem pri dvojčkih in trojčkih gre bolj ali manj za vezivo, za povezave veznikov, pomožnih glagolov in prostega glagolskega morfema *se*, sem in tja pa najdemo še kakšen osební zaimék. V tabelah prvega vzorca je pri dvojčkih en sam pravi glagol, samostalnik ali pridevnik, v *je rekel*, pri trojčkih tudi, v *se je zgodilo*, pri četverčkih že 22, a od tega v kar 15 primerih nastopa glagol *zdeti se* (ostanejo od časa do časa/ se je zgodilo, da/ Bouvard in Pécuchet sta/ kar se je zgodilo/ iz dneva v dan/ je bil prepričan, da in prišlo na misel, da), pri peterčkih pa 38 (od tega v 18 primerih nastopata *zdeti se* ali *zazdeti se*). Konkretnih n-terčkov, za katere lahko takoj ugotovimo izvor, je pri prvem vzorcu malo – v celi tabeli 47 sta samo dva: četverček *Bouvard in Pécuchet sta* iz Flaubertovega romana z naslovom iz prvih treh besed in peterček *in mir od Boga, našega iz Nove zaveze*. Frekvence najpogostejših n-terčkov z rastočim n padajo še dosti hitreje kot pri n-terčkih znakov: razmerja so 7,3 : 1 (*je : se je*), 17,2 : 1 (*se je : da bi se*), 7,4 : 1 (*da bi se : ne da bi se*) in 3,5 : 1 (*ne da bi se : se mi je zdelo, da*). Zanimivo je tudi, da ima zadnji, 50. peterček v tabeli 47 frekvenco že komaj 9.

Drugi vzorec je skromnejši in najdemo že med dvojčki 3 take, ki imajo kakšen glagol, samostalnik ali pridevnik: *je rekel*, *je rekla* in *Peter Majcen*. Med trojčki jih je 10: *zmajal z glavo, zamahnil z roko, se je popraskal, se je oglasil, se je obrnil, Peter Majcen je, je popraskal kmet, z obema rokama, se je vrnil* in *in se obrnil*. Konkretnih n-terčkov, ki jim lahko zlahka najdemo izvor, je 34 in so vsi iz del *Balada o trobenti in oblaku, Potepuh Najdù* in *Tantadruj*. Prekrivanje vzorcev je znatno do $n = 3$, sicer pa je pri enojčkih med najpogostejšimi 50 42 enakih, pri dvojčkih 37, pri trojčkih 26, pri četverčkih 7 (*da se mu je/ ne da bi se/ potem pa se je/ se je obrnil k/ se mi je, da/ se mu je zdelo in se mu je, da*) in pri peterčkih samo 2 (*se mu je zazdelo, da* in *se mu je zdelo, da*).

Tabela 47: Najpogostejši besedni n-terčki (2–4) v prvem vzorcu, s frekvenca

	2	3	4	5			
1. se je	20.657	da bi se	1.202	ne da bi se	163	se mi je zdelo, da	47
2. je bil	7.297	da se je	1.169	se mu je, da	108	Zdelo se mi je, da	42
3. da bi	7.077	se mu je	1.145	se mi je zdelo	93	Zdelo se mu je, da	41
4. da je	6.903	ki se je	972	se je zdelo, da	92	se mu je zdelo, da	39
5. je bilo	5.643	pa se je	809	kaj se je zgodilo	79	zdelo se mu je, da	35
6. so se	5.609	ki ga je	797	se mu je zdelo	78	se je zgodilo, da je	26
7. ki je	5.136	ne da bi	668	se mi je, da	73	in zdelo se mu je	21
8. da se	5.123	ko se je	652	ki se mu je	71	ki mu je bilo ime	20
9. in se	4.894	se mi je	643	da se ne bi	70	prišlo na misel, da bi	20
10. je bila	4.618	ki je bil	524	kakor da bi se	65	se mu je, da je	20
11. ga je	4.484	ki jo je	520	se je, da je	64	se mu je, da se	20
12. mu je	4.336	ki so se	495	da se mu je	63	Zdelo se ji je, da	19
13. pa je	3.993	se ji je	465	Zdelo se mi je	62	je bila v tem, da	18
14. ki so	3.712	ki jih je	443	od časa do časa	58	se je zdelo, da je	17
15. bi se	3.003	se je v	431	Zdi se mi, da	57	in zdelo se je, da	15
16. mi je	2.690	in se je	411	se je zgodilo, da	55	je bilo vse, kar je	15
17. ko je	2.528	Ko se je	359	tako da se je	55	je od časa do časa	15
18. je v	2.501	kakor da bi	355	se je obrnil k	54	ne glede na to, ali	15
19. jo je	2.474	da bi ga	350	Zdelo se mu je	53	se je in se je	15
20. ki se	2.452	da je bil	348	mi je zdelo, da	47	se je zdelo, da se	15
21. se mu	2.377	ki je bila	335	Zdelo se je, da	47	se ji je zdelo, da	15
22. sem se	2.269	da so se	315	in ko se je	46	so se odprla vrata in	15
23. da so	2.254	da bi bil	311	se mi zdi, da	46	se mi je, da sem	14
24. ni bilo	2.246	se je zgodilo	299	ne da bi bil	45	tistega, ki me je poslal	14
25. ne bi	2.043	ki so ga	299	zdi se mi, da	45	zdelo se mi je, da	14
26. pa se	1.959	je, da je	296	zdelo se mu je	44	Zdelo se je, da je	13
27. in je	1.880	so se mu	295	Bouvard in Pécuchet sta	43	je bilo jasno, da je	12
28. me je	1.783	ki mu je	287	je bilo, da je	43	je prišlo na misel, da	12
29. so bili	1.778	pa je bil	278	ki mu je bilo	43	ne da bi se bil	12
30. jih je	1.763	se je na	273	potem pa se je	42	je bil prepričan, da je	11
31. ji je	1.742	da ne bi	272	je, da se je	41	je zdelo, da se je	11
32. je rekel	1.740	da ga je	272	se je in je	41	ki so bili z njim	11
33. si je	1.685	da se ne	268	da bi se bil	40	se ji je zazdelo, da	11
34. je, da	1.659	se je, da	258	pa se mu je	40	se ji je, da je	11
35. ne bo	1.639	mu je bilo	254	mu je zdelo, da	39	se mu je zazdelo, da	11
36. je že	1.575	ki so bili	244	da se mi je	38	tri dni in tri noči	11
37. se ne	1.567	da se bo	239	se ji je, da	37	Zazdelo se mu je, da	11
38. se mi	1.528	je bil v	239	kar se je zgodilo	36	da je vse v redu	10
39. ki ga	1.462	ne bi bilo	233	ki ga je bil	36	je prijel za roko in	10
40. je na	1.447	se je bil	233	iz dneva v dan	35	je zgodilo, da se je	10
41. so ga	1.402	se je že	232	je bilo jasno, da	35	milost vam in mir od	10
42. sta se	1.387	je bilo treba	228	je bil prepričan, da	34	ne da bi pri tem	10
43. je še	1.386	da je to	226	ki si ga je	34	ne da bi se bila	10
44. je to	1.336	pa so se	224	prišlo na misel, da	34	okna v zidovih in stenah	10
45. in ga	1.334	je bilo v	221	se ji je zdelo	34	se je odločil, da bo	10
46. in da	1.297	je, da se	216	je v tem, da	33	se je zgodilo, da se	10
47. pa so	1.292	da mu je	216	je, da bi se	33	zdelo se ji je, da	10
48. To je	1.269	pa je bilo	211	ne glede na to	33	Čez nekaj časa se je	9
49. da sem	1.225	in da je	207	s tem, da je	33	ga je obudil od mrtvih	9
50. kar je	1.214	ki so jih	207	vse, kar je bilo	33	in mir od Boga, našega	9

Tabela 48: Najpogostejši besedni n-terčki (2-4) v drugem vzorcu, s frekvenca

	2	3	4	5		
1. se je	4.327	ki se je	265	se je popraskal kmet	46 kajpak,« se je popraskal kmet	36
2. in se	1.563	pa se je	261	kajpak,« se je popraskal	45 kaj hočemo, je že tako	22
3. je bil	1.004	da bi se	211	ne da bi se	38 ne da bi se ozrl	19
4. pa je	977	se mu je	184	je rekel Peter Majcen	32 in rdečega vina, temnih gozdov	14
5. da je	935	da se je	176	Peter Majcen se je	30 rdečega vina, temnih gozdov in	14
6. da bi	929	kakor bi se	139	se je obrnil k	28 temnih gozdov in zelenega morja	14
7. ki je	876	ki je bil	122	nato pa se je	27 vina, temnih gozdov in zelenega	14
8. ga je	759	se je v	105	hočemo, je že tako	26 sonca in rdečega vina, temnih	13
9. so se	732	ki ga je	93	zmajal z glavo in	26 zlatega sonca in rdečega vina	13
10. je bilo	712	ker se je	91	je zmajal z glavo	24 je zmajal z glavo in	12
11. mu je	701	ki so se	84	kaj hočemo, je že	22 je rekel Peter Majcen in	11
12. je bila	696	ki je bila	80	in se zamaknil v	21 ker se je bal, da	10
13. kakor bi	655	ko se je	80	je ponovil Matic in	21 smrt, moj hramček bo zaprt	10
14. da se	609	ne da bi	73	to se pravi, da	21 bo bridka smrt, moj hramček	9
15. bi se	574	se je na	71	da bi se ozrl	20 bridka smrt, moj hramček bo	9
16. pa se	464	zmajal z glavo	69	da se mu je	19 se mu je zdelo, da	9
17. ki so	456	se mi je	69	zamahnil z roko in	19 s svojimi velikimi, nejasnimi	
18. je v	453	da bi ga	66	je zamahnil z roko	17 očmi	8
19. ki se	453	nato pa je	64	se je zdrznil in	16 se mu je zazdelo, da	8
20. sem se	451	z glavo in	63	ki se je v	15 dvignil kazalec, da bi si	7
21. se mu	437	zamahnil z roko	63	od hiše do hiše	15 in vse je kazalo, da	7
22. je rekel	423	se je spet	63	rekel Peter Majcen in	15 je zamahnil z roko in	7
23. in ga	395	so se mu	63	gozdov in zelenega morja	14 je zamajala z glavo in	7
24. je že	353	da ne bi	62	in rdečega vina, temnih	14 je-e sonce, na zemljici mrz	7
25. ne bi	349	se je popraskal	61	in zamahnil z roko	14 kakor bi smrt švignila mimo	7
26. si je	321	se je in	58	kako bi se reklo	14 nebu je-e sonce, na zemljici	7
27. jo je	310	ki jo je	54	Peter Majcen in se	14 zmajal z glavo in se	7
28. ker je	303	da mu je	53	potem pa se je	14 in spomnila se je, kako	6
29. ne bo	292	da se bo	53	rdečega vina, temnih	in vsi so za vas	6
30. mi je	279	da je bil	52	gozdov	14 ja-az zvonce in vsi so	6
31. je še	272	se je oglasil	52	se je ozrl v	14 je ponovil Matic in dvignil	6
32. so ga	272	ki jih je	50	se mu je zdelo	14 ne da bi se ozrla	6
33. Peter		je, da je	49	temnih gozdov in zelenega	14 se je bal, da bo	6
Majcen	270	se je obrnil	49	vina, temnih gozdov in	14 udaril s pestjo po mizi	6
34. je z	249	to se pravi	48	ker se mu je	13 zaradi tega, ker se je	6
35. ter se	246	ker je bil	48	sonca in rdečega vina	13 zvonce in vsi so za	6
36. je rekla	245	Peter Majcen je	47	zlatega sonca in rdečega	13 bi si zapomnil ta ukaz	5
37. me je	244	kajpak,« se je	47	k njemu in ga	12 če je moje okno še	5
38. ko je	243	je popraskal	47	k njemu in mu	12 da bi se ozrl v	5
39. ni bilo	241	kmet	46	kakor bi se mu	12 da bi si zapomnil ta	5
40. je na	241	z obema rokama	46	kakor šiba na vodi	12 imam in ta bo prava	5
41. da so	229	potem pa je	46	se je bal, da	12 in sam ni vedel, kdaj	5
42. in jo	224	se je še	45	se je nasmehnil in	12 in se ozrl okrog sebe	5
43. in si	221	se je vrnil	45	se mi je, da	12 in še ena, pa je	5
44. je s	214	da se mu	44	se mu je, da	12 in zdelo se mi je	5
45. jih je	212	mu je bilo	44	z glavo in se	12 je lase s čela in	5
46. je tako	207	se je bil	43	in se ozrl po	11 je moje okno še zaprto	5
47. je spet	206	da bi si	43	in se ozrl v	11 kakor bi se bala, da	5
48. se ni	206	in se obrnil	42	je prikimal Peter Majcen	11 kakor bi se bal, da	5
49. da bo	206	kakor bi ga	42	lase s čela in	11 kazalec, da bi si zapomnil	5
50. se ne	205	ki mu je	42	zmajala z glavo in	11 ker se mu je zdelo	5

Tabela 49: Najpogostejši besedni n-terčki na začetkih povedi v prvem vzorcu

	1	2	3	4			
1. In	5.034	Ko je	1.157	Ko se je	356	Zdelo se mi je,	48
2. V	4.019	To je	850	Potem se je	130	Zdelo se je, da	47
3. Ko	3.844	Ko se	508	To je bilo	111	Zdelo se mu je,	43
4. Na	2.279	Bil je	485	Tedaj se je	105	Bouvard in Pécuchet sta	39
5. Ali	2.077	Ko so	417	Ko so se	90	Zdi se mi, da	35
6. Če	1.999	Tako je	417	Tako se je	90	In tako se je	28
7. To	1.989	»To je	402	To je bil	87	Jezus jim je odgovoril:	24
8. Tako	1.639	Ko sem	384	In ko je	85	Po teh besedah je	24
9. Toda	1.512	»Kaj pa	305	Ko pa je	85	Videti je bilo, da	22
10. Tudi	1.463	Potem je	294	Ko sem se	74	Zdelo se ji je,	22
11. »Kaj	1.455	Če bi	287	Ko je bil	67	»Zdi se mi, da	22
12. Po	1.361	Bilo je	281	Zdelo se je,	63	Tedaj pa se je	18
13. Z	1.219	Bila je	272	»Kaj pa je	61	»Kaj pa je to?«	18
14. Kaj	1.216	Tedaj je	241	Nato se je	58	Ne da bi se	17
15. Potem	1.214	Zdelo se	217	Zdelo se mi	56	Čez nekaj časa se	16
16. Ne	1.163	In ko	214	To je bila	55	Ko pa se je	16
17. Za	1.113	Nato je	214	Veronika se je	53	On pa jim je	16
18. Pa	1.056	Potem se	211	Zdelo se mu	53	Tu in tam je	16
19. »In	1.055	V tem	210	Ne da bi	51	In ko se je	15
20. »Ne	1.040	Če je	208	In tako je	49	Od časa do časa	15
21. »Ali	949	In če	194	Potem pa je	45	Če ne bi bilo	14
22. S	911	Ko pa	190	Bouvard in Pécuchet	44	In tedaj se je	14
23. Ta	904	Če pa	187	Phileas Fogg je	44	Pri tem se je	14
24. Še	891	Ker je	185	Tu in tam	44	Jezus pa jim je	13
25. Ni	871	»Tako je.«	184	Vse je bilo	44	Njen obraz je bil	13
26. »To	863	Da bi	183	Obrnil se je	43	Obrnil se je k	13
27. A	847	In tako	180	Po teh besedah	42	Res pa je, da	13
28. Vse	843	Vsi so	173	Čez nekaj časa	41	V tem trenutku se	13
29. Pri	824	Pri tem	169	Medtem ko je	41	Njegov glas je bil	12
30. Zdaj	771	Če se	168	V tem trenutku	41	Obrnil se je in	12
31. »Če	731	Vse je	167	Zdi se mi,	41	Po drugi strani pa	12
32. Zato	704	Jaz sem	165	Jezus jim je	40	Potem pa se je	12
33. Tedaj	702	Ali je	163	Ker je bil	40	Pri teh besedah je	12
34. Saj	689	Kaj pa	160	Če bi se	39	V tistem trenutku se	12
35. Od	679	Tako se	159	Ko je bilo	39	Vendar pa se je	12
36. Kako	665	Ko sta	156	Kakor da bi	38	Vrnil se je v	12
37. Da	643	Ko bi	155	Zato se je	38	Bal se je, da	11
38. Ker	632	Tam je	151	Ko sta se	37	Bilo je, kakor da	11
39. Bil	621	Zdaj je	151	Nenadoma se je	37	Bilo je, kot bi	11
40. Jaz	613	Ali pa	149	Tako je bilo	37	Čez nekaj časa je	11
41. Iz	604	Sicer pa	136	»Nič se ne	37	Jezus mu je rekel:	11
42. »Saj	587	Bouvard je	132	Alena se je	36	Kakor da bi se	11
43. Vendar	577	Ta je	131	Ko je prišel	36	Kdor ima ušesa, naj	11
44. Ob	575	Pécuchet je	128	Pécuchet se je	36	Med njimi je bil	11
45. Samo	572	Zdaj pa	128	Zdaj se je	36	Njegov obraz je bil	11
46. Nato	571	Tedaj se	125	In tako se	35	Očitno je bilo, da	11
47. Tam	564	Zato je	125	Ko ga je	35	Zazdelo se mu je,	11
48. Kakor	563	Kaj je	123	On pa je	35	Bilo je, kot da	10
49. »Tako	562	S tem	121	V tistem trenutku	35	Glede na to, da	10
50. »Kako	557	Potem pa	117	Vedel je, da	35	Jezus mu je odgovoril:	10

Tabela 50: Najpogostejši besedni n-terčki na začetkih povedi v drugem vzorcu

	1	2	3	4			
1. In	1546	Ko je	181	Ko se je	88	»Hm, kajpak,« se je	47
2. V	543	To je	136	Nato se je	54	Peter Majcen se je	30
3. Ko	520	Bil je	120	Peter Majcen je	47	Prekleta strešnica se je	11
4. Toda	475	Nato je	110	»Hm, kajpak,« se	47	Raus e patacis, repa	10
5. Nato	367	Ko se	102	In tako je	31	Spomnila se je, kako	10
6. Saj	349	Tako je	97	Peter Majcen se	30	V tem trenutku se	9
7. Na	333	Peter Majcen	87	Potem se je	26	»Prišla bo bridka smrt,	9
8. »In	304	Nato se	85	V tem trenutku	25	Nanca in Kati sta	8
9. To	298	»Kaj pa	85	Vse je bilo	25	Tedaj pa se je	8
10. Pa	263	Potem je	81	Ozrl se je	24	Ozrl se je v	7
11. Zdaj	258	In tako	73	Martin se je	23	»Na-a nebu je-e sonce,	7
12. A	256	Ko so	66	Venc se je	23	»Pa kaj hočemo, je	7
13. Tako	246	Vse je	63	To je bil	18	»Raus e patacis, repa	7
14. Potem	233	Zdaj je	56	Matic se je	17	In tako se je	6
15. Tudi	233	Saj je	55	Nato si je	17	Ko se je Žef	6
16. Vse	227	Bila je	54	Te besede so	17	Naslonil se je na	6
17. »Saj	226	Če bi	51	Kaj pa je	16	Nato pa se je	6
18. »Ali	208	Očka Orel	50	Tantadruj se je	16	Na-a-biram ja-az zvonce in	6
19. Samo	197	Pa tudi	50	Zdaj pa je	16	Nenadoma pa se je	6
20. Ali	195	Bilo je	47	»Kaj pa je	16	Pa kaj hočemo, je	6
21. »Kaj	185	In tudi	47	Nato pa je	15	Potem pa se je	6
22. Če	181	Zdaj pa	47	Oče se je	15	Segel je v žep	6
23. »No,	179	»Hm, kajpak,«	47	Ozrl sem se	15	To pa je bilo	6
24. Kaj	169	Kaj pa	46	Pob se je	15	To se pravi, da	6
25. Še	164	Martin je	46	Počasi se je	15	Travnikarjev stric se je	6
26. Vsi	162	Matic je	46	Potem pa je	15	Zamaknil se je v	6
27. Bil	144	Potem pa	46	To je bilo	15	Zdaj pa se je	6
28. Peter	143	In če	44	Venček se je	15	»In kaj je bilo	6
29. Z	142	In še	44	Bilo mu je	14	Bilo je, kakor bi	5
30. Po	141	In ker	42	Peter se je	14	In potem se je	5
31. No,	139	Temnikar je	41	Zdaj se je	14	In tako je tudi	5
32. Za	139	In zdaj	40	In ker je	13	Ko pa se je	5
33. Ta	136	Sicer pa	40	In res je	13	Ko se je vrnil	5
34. Sova	127	In prav	38	In to je	13	Ko so prišli do	5
35. S	125	Vsi so	38	»V imenu postave,	13	Obrnil se je k	5
36. Tam	121	In potem	37	In da bi	12	Ozrl sem se v	5
37. Kar	118	Tam je	36	Ko pa je	12	Tako je bilo tudi	5
38. »Tantadruj,	107	Venček je	36	Stopil je k	12	Vedel je, da je	5
39. »Pa	104	Moj Jezus	35	Tu pa tam	12	Vse je kazalo, da	5
40. Ne	101	V tem	35	Z obema rokama	12	Zamaknil sem se v	5
41. Martin	100	In to	34	In ko je	11	Zdelo se mi je,	5
42. Počasi	95	Potem se	34	Ko sem se	11	Zravnal se je in	5
43. Pri	94	Žef je	34	Ljudje so se	11	»Javolheršturmfirer!«	
44. »To	93	»To je	34	Naslonil se je	11	je zagrrgal Prekleta	5
45. Iz	92	Ko sem	33	Nemška smrt je	11	»Kaj pa je?« je	5
46. »Ne,	92	Nato pa	33	Otroci so se	11	»Pha!« je zaničljivo prhnil	5
47. Ob	90	Pa saj	33	Prekleta strešnica se	11	»Ti si nor!« je	5
48. Prav	87	Stopil je	33	Res je, da	11	Bilo mu je tako	4
49. Tedaj	87	»Ali je	33	Tedaj pa se	11	In ker je bil	4
50. »Ti	87	Otroci so	32	Temnikarica se je	11	In ko se je	4

Tabela 51: Najpogostejši besedni n-terčki na koncih povedi v prvem vzorcu

1	2	3	4		
1. rekel:	685	je rekel:	290	66 Jezus jim je odgovoril:	24
2. je.	587	»Tako je.«	184	62 kaj se je zgodilo.	21
3. oči.	500	in rekel:	165	40 pa jim je rekel:	21
4. več.	370	je rekel.	142	38 pa mu je rekel:	19
5. njim.	369	z njim.	139	38 je odvrnil Phileas Fogg.	15
6. ne.	327	ni bilo.	137	36 ki me je poslal.	11
7. je.«	314	na tla.	124	35 Jezus mu je rekel:	11
8. glavo.	298	za njim.	114	34 Jezus mu je odgovoril:	10
9. rekel.	296	je odgovoril:	107	34 našega Gospoda Jezusa Kristusa.	9
10. naprej.	277	na svetu.	94	29 je skomignil z rameni.	9
11. roko.	256	s seboj.	94	27 – je vzkliknil Passepartout.	9
12. je:	243	se je.	94	27 Jezus jim je rekel:	9
13. obraz.	239	je rekla.	90	26 ga ni bilo več.	8
14. bilo.	236	v oči.	85	25 je dejal Mr. Fogg.	8
15. domov.	235	k sebi.	80	24 iz dneva v dan.	8
16. seboj.	235	v obraz.	76	23 »Kaj pa je to?«	8
17. sebi.	233	za njo.	74	23 svojo smrt in vstajenje	7
18. njo.	231	v hišo.	73	22 in Gospoda Jezusa Kristusa.	7
19. dejal:	223	z njo.	72	21 je vprašal Mr. Fogg.	7
20. reče:	215	je bilo.	68	21 ki je v nebesih.	7
21. nazaj.	213	k njemu.	68	20 kar se je zgodilo.	7
22. dan.	206	in dejal:	67	20 ni bilo od nikoder.	7
23. sebe.	205	reče Orten.	67	20 pa jim je odgovoril:	7
24. tla.	202	sem rekel.	66	20 je rekel svojim učencem:	7
25. ni.	200	z njimi.	64	19 »V tem imaš prav.«	7
26. se.	197	in rekla:	60	19 »Kaj misliš s tem?«	7
27. življenje.	194	je bil.	60	19 vam, pismouki in farizeji!	6
28. svetu.	192	z glavo.	60	19 pri cerkvi sv. Jakoba.	6
29. »Da.«	191	ali ne?«	58	18 je dejala Rdeča kraljica.	6
30. ljudi.	188	reče Podol.	58	18 ne duha ne sluha.	6
31. rekla:	188	jim je:	57	17 od dne do dne.	6
32. Alica.	187	je vprašal.	56	17 je bilo vse tiho.	6
33. drugega.	186	ali ne.	54	17 je vse v redu.	6
34. strani.	185	je dejal:	53	17 je rekel Krilati lev.	6
35. njega.	184	po sobi.	52	16 na področju osebnih računalnikov.	6
36. let.	183	je rekla:	51	16 – je odvrnil Fix.	6
37. ne?«	183	po tleh.	51	16 je rekel Mali Miklavž.	6
38. sam.	181	v roki.	51	16 in jim je rekel:	6
39. nič.	179	rekla Alica.	49	16 pa mu je odgovoril:	6
40. srce.	173	v njem.	49	16 junij '91, Novo mesto	5
41. pot.	170	za seboj.	49	16 si je rekla Veronika.	5
42. glas.	169	iz sobe.	48	15 Vrata so se odprla.	5
43. roke.	169	pri sebi.	48	15 Veronika,« je rekla Martina.	5
44. doma.	165	z menoj.	48	15 jih ni bilo več.	5
45. besede.	164	za to.	47	15 in se ne gane.	5
46. hišo.	159	na dan.	46	14 to je bilo vse.	5
47. meni.	159	pred njim.	46	14 od nog do glave.	5
48. njem.	156	v sobo.	46	14 je dejal Phileas Fogg.	5
49. noč.	155	z rameni.	45	13 je odvrnil Mr. Fogg.	5
50. vprašal.	154	je zgodilo.	44	13 od glave do nog.	5

Tabela 52: Najpogostejši besedni n-terčki na koncih povedi v drugem vzorcu

	1	2	3	4			
1. oči.	170	Peter Majcen.	79	je popraskal kmet.	40	se je popraskal kmet.	40
2. je.	155	za njim.	58	zamahnil z roko.	33	hočemo, je že tako!	18
3. glavo.	147	z glavo.	55	zmajal z glavo.	26	je rekel Peter Majcen.	17
4. njim.	123	z roko.	45	je že tako!	19	in zamahnil z roko.	11
5. rekel.	114	je rekel.	44	rekel Peter Majcen.	19	patacis, repa in krompir!	11
6. rekel:	112	popraskal kmet.	40	in se obrnil.	15	gozdov in zelenega morja.	10
7. roko.	106	sem rekel.	38	in tako dalje.	14	kakor šiba na vodi.	10
8. naprej.	90	k sebi.	35	in se zamislil.	13	kajpak, se je popraskal.	9
9. več.	84	na tla.	35	in zaprl oči.	13	moj hramček bo zaprt ...	9
10. ni.	84	je rekla.	33	kakor se reče.	12	da bi se ozrl.	8
11. Majcen.	79	k njemu.	33	se je popraskal.	12	je prikimal Peter Majcen.	8
12. vprašal.	73	in rekel:	32	repa in krompir!	11	je zamahnil z roko.	7
13. roke.	71	je vprašal.	29	se je vprašal.	11	je vprašal Peter Majcen.	6
14. vanj.	68	okrog sebe.	28	in zelenega morja.	10	je zmajal z glavo.	6
15. vrata.	67	se je.	28	ozrl okrog sebe.	10	kakor bi hotel reči:	6
16. očmi.	63	z njim.	28	šiba na vodi.	10	kakor griva divjega konja.	6
17. ne.	63	iz hiše.	27	vrata za sabo.	10	sonce, na zemljici mrz.	6
18. teta.	63	v hišo.	26	ga je prešinilo.	9	svojimi velikimi, nejasnimi očmi.	6
19. hišo.	62	v izbo.	25	hramček bo zaprt ...	9	bi smrt švignila mimo.	5
20. otrok.	61	v oči.	25	in se nasmehnil.	9	da bi se ozrla.	5
21. dan.	59	v dolino.	24	je rekel kamen.	9	in zmajal z glavo.	5
22. oče.	59	na noge.	23	je rekla teta.	9	in zmaje z glavo.	5
23. sebe.	59	sem vprašal.	23	Tako je to!	9	iz kota v kot.	5
24. rekla:	58	v kuhinjo.	23	zmignil z rameni.	9	je pomislil Peter Majcen.	5
25. obraz.	58	in rekla:	22	bi se ozrl.	8	kakor oči divje živali.	5
26. hiše.	57	na cesto.	22	je rekel macesen.	8	moje okno še zaprto.	5
27. kmet.	56	ni več.	21	Matic Enaka Palica.	8	od hiše do hiše.	5
28. mama.	55	in šel.	20	prikimal Peter Majcen.	8	ponovil Matic Enaka Palica.	5
29. domov.	53	na mizo.	20	velikimi, nejasnimi očmi.	8	s pestjo po mizi.	5
30. doma.	53	ni bilo.	20	vprašal Peter Majcen.	8	si zapomnil ta ukaz.	5
31. sebi.	51	z rameni.	20	zmaje z glavo.	8	svoje debele, solzave oči.	5
32. besede.	50	k njej.	19	in zaprla oči.	7	vsi so za vas.	5
33. tla.	50	pred njim.	19	je zagrmel Luka.	7	zaprl vrata za sabo.	5
34. pob.	49	proti vratom.	19	naravnost v oči.	7	ga je prekinil pob.	4
35. vprašal:	49	v njem.	19	pomislil Peter Majcen.	7	in glasno požrl slino.	4
36. vasi.	49	v obraz.	19	z obema rokama.	7	in ta bo prava!	4
37. njej.	48	in vstal.	18	zapomnil ta ukaz.	7	je zaničljivo prhnil Furlan.	4
38. glas.	48	očka Orel.	18	zmajala z glavo.	7	ne bele ne črne.	4
39. mizo.	47	v vodo.	18	»Kaj pa je?«	7	opotekel, a padel ni.	4
40. stric.	44	k mizi.	17	bi hotel reči:	6	otrpnil na poraščeni bradi.	4
41. glasom:	44	po dolini.	17	griva divjega konja.	6	pritajeno drsi skozi veje.	4
42. Martin.	44	»Hm, kajpak ...	17	je ponovil Matic.	6	se bo Krist rodil.	4
43. noge.	43	in dejal:	16	je prikimal Hotejec.	6	se ozrl okrog sebe.	4
44. njem.	43	in vprašal:	16	je rekel oče.	6	žalostno zmajal z glavo.	4
45. je!	43	pri srcu.	16	je rekla jelka.	6	in čakali, kaj bo.	3
46. Luka.	42	se ozrl.	16	je rekla mama.	6	iz čistega srebra ulit.	3
47. rekla.	42	Travnikarjev stric.	16	je vprašal pob.	6	je bilo že prepozno.	3
48. hiši.	41	za njo.	16	Mar ni res?	6	je obdržal na nogah.	3
49. njemu.	41	o tem.	15	na vse grlo:	6	je pribil Peter Majcen.	3
50. življenje.	40	se nasmehnil.	15	na zemljici mrz.	6	je prikimal stari mornar.	3

Poleg najpogostejših n-terčkov besed kjer koli v povedih je zanimivo še, s kakšnimi besednimi n-terčki oziroma frazami se povedi najpogosteje začnejo in končajo. V tabelah 49 in 50 so navedeni najpogostejši n-terčki na začetkih povedi v prvem in v drugem vzorcu, v tabelah 51 in 52, vse na prejšnjih straneh, pa na koncih povedi za oba vzorca.

Ker je povedi veliko manj kot besed, so manjše tudi frekvence v tabelah od 49 do 52. V tabeli 49, ki se nanaša na začetke povedi in na prvi vzorec, je videti zelo zanimivo strukturo pri enojčkih. Na začetkih povedi najpogosteje najdemo kratke besede, predvsem veznike, predloge in členke. Skoraj polovica (24) jih je dolgih 2 črki ali manj, tri četrtine 3 črke ali manj in le ena beseda, *vendar*, je dolga 6 črk. Morda malo presenetljivo je odločno na prvem mestu, v drugem vzorcu še veliko bolj, veznik *in*, ki smo ga sicer vajeni sredi povedi. V celi tabeli 49 najdemo le 12 konkretnih n-terčkov med 200, od katerih jih 11 izvira iz prevodov in le 1 iz izvirnega leposlovja (*Veronika se je*). Na začetkih povedi se oba vzorca tudi precej dobro ujemata – pri enojčkih je istih 78 %, pri dvojčkih 52 %, pri trojčkih 28 %, pri trojčkih pa le še 14 % ali 7 (*In ko se je*, *In tako se je*, *Ko pa se je*, *Obrnil se je k*, *Potem pa se je*, *Tedaj pa se je*, *V tem trenutku se in* *Zdelo se mi je*).

Morda še bolj zanimivi so besedni n-terčki na koncih povedi, ki jih najdemo v tabelah 51 in 52. V nasprotju z začetki povedi tu pri enojčkih najdemo skoraj izključno samostalnike, glagole in zaimke. Če odštejemo ločilo na koncu povedi in navednice premega govora, je v obeh vzorcih na prvem mestu glagolski deležnik *rekel*, značilen predhodnik povedi s premim govorom, na drugem mestu glagol *je*, na tretjem mestu pa prvi samostalnik, *oči*. Pri četrtem mestu se vzorca razideta. Ujemanje je pri enojčkih 52 %, pri dvojčkih 44 %, pri trojčkih in četvorčkih pa ni niti enega, ki bi nastopal med najpogostejšimi 50 v obeh vzorcih. Če pogledamo trojčke prvega vzorca, hitro tudi postane jasno, zakaj. Med njimi jih je namreč kar 14 ali več kot četrtina tipa *je rekel Ime: je dejala Alica. je uprašala Alica. je pomislila Alica. je rekla Alica. je uprašala Anica. je rekla Veronika. je rekla Martina. je rekel Bouvard. odvrnil Phileas Fogg. je rekel O'Brien. je rekel Winston. je rekel Pécuchet. je rekel Ray. in je rekel Timmy*. Vsi ti n-terčki so seveda konkretni, takoj jih lahko postavimo v konkretno knjigo in v veliki večini spet izvirajo iz prevodov.

Besedni n-terčki na koncih povedi imajo tudi večjo entropijo od tistih na začetku – v prvem vzorcu ima najpogostejši enojček na začetkih (*In*) relativno frekvenco 3,02 %, na koncih (*rekel*) pa le 0,49 %. Razporeditev n-terčkov besed na začetkih, še bolj pa na koncih povedi bi utegnila biti zanimiva za ugotavljanje avtorstva.

5.3 Povedi

Povedi so, v nasprotju s črkami in besedami, že zaključene enote besedila. Njihova struktura in raznovrstnost ločita besedila med seboj po kvaliteti in tudi drugače. Zato bi bila kvantitativna raziskava povedi, še posebno v oblikoslovno označenem besedilu, zelo zanimivo in hvaležno opravilo. Žal pa je preskok od besed k povedim še večji od tistega s črk na besede in bi bil za tako raziskavo, kakršna je bila še možna pri besedah, potreben bistveno, vsaj za velikostni razred, večji vzorec. Nekaj o povedih je bilo že navedenega v razdelku o besednih n-terčkih, sledi pa še nekaj podatkov o dolžinah povedi in o najpogostejših povedih.

5.3.1 Dolžine povedi

Iz dolžin povedi se da razbrati marsikaj o zvrsti besedila: besedila s kratkimi povedmi so bolj tekoča in lažje berljiva od tistih z dolgimi, dolžine povedi pa veliko povedo tudi o avtorjevem stilu. V tabeli 53 so navedene dolžine povedi za oba vzorca, v besedah.

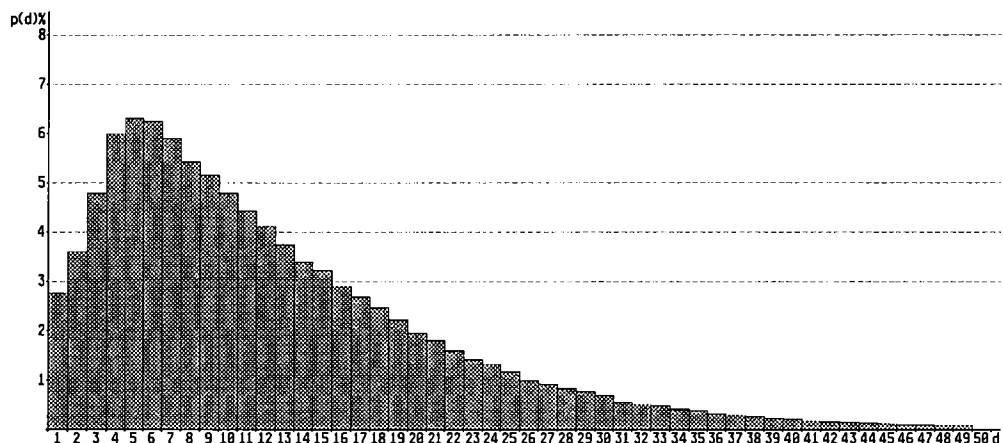
Tabela 53: Dolžine povedi za oba vzorca v besedah

	povedi	najmanjša dolžina	največja dolžina	povprečna dolžina	standardni odklon
Prvi vzorec	201.445	1	1.002	13,51	11,86
Drugi vzorec	37.459	1	284	10,91	9,51
Skupaj	238.904	1	1.002	13,10	11,56

Povprečne dolžine v znakih so 77 za prvi vzorec, 60 za drugi in 75 za oba skupaj. Najdaljša poved v prvem je dolga 6.035 znakov, v drugem pa 1.836. Standardni odklon je bil dodan zaradi zelo velikih razlik med povprečnimi in največjimi dolžinami. Ker je ta odklon v vseh treh primerih skoraj tako velik kot aritmetična sredina, so porazdelitve daleč od normalne. Porazdelitev dolžin za oba vzorca skupaj je na sliki 16. Najbolj pade v oči, da je močno asimetrična v levo. Največ povedi, po približno 6 % celote, je dolgih od 4 do 7 besed, z vrhom pri 5. Porazdelitev se strmo dvigne od skoraj 3 % pri eno besedo dolgih povedih do vrha pri dolžini 5, potem pa vse počasneje pada proti koncu. Aritmetična sredina pri 13 besedah je že precej naprej od vrha.

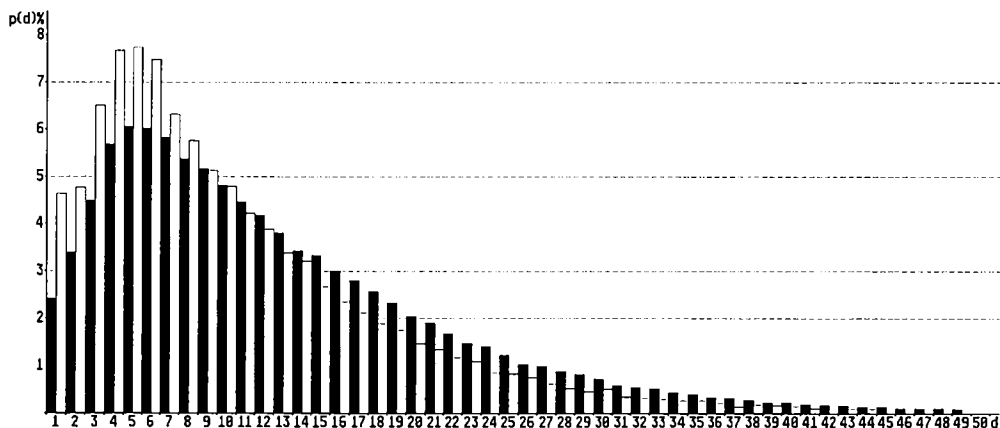
Na sliki 17 sta prikazani porazdelitvi za vsak vzorec posebej. Stolpci prvega vzorca so označeni s črno barvo, drugega pa z belo. Vrh je pri obeh vzorcih ostal, kjer je bil tudi pri skupnem besedilu (5), opaziti pa je znatno prevlado kratkih povedi v drugem vzorcu, do vključno dolžine 8. Največ povedi, po več kot 7 %, je dolgih

Slika 16: Porazdelitev dolžin povedi v besedah za oba vzorca skupaj



4, 5 in 6 besed, zelo veliko, bistveno več kot pri prvem vzorcu pa je tudi zelo kratkih povedi, z dolžino 1, 2 in 3. Očitno je v drugem vzorcu veliko več premega govora, kar lahko preverimo s pomočjo podatkov o premem govoru v tabelah 29 in 30. V prvem vzorcu je 45.887 pojavitev oznak za začetek in konec premega govora, v drugem vzorcu pa 10.353. V prvem pride torej 4,39 povedi na en premi govor, v drugem vzorcu pa le 3,62.

Slika 17: Porazdelitev dolžin povedi v besedah za vsak vzorec posebej



Najkrajše povedi so lepo razvidne iz tabel 57 in 58 v naslednjem razdelku, po 5 najdaljših povedi za oba vzorca pa je navedenih v tabelah 54 in 55. V prvem vzorcu je vseh 5 iz istega dela – *Zmot dijaka Tjaža* Florjana Lipuša, v drugem vzorcu pa so 1., 3. in 5. med najdaljšimi povedmi iz *Hiše št. 14*, 2. iz *Balade o trobenti in oblaku*, 4. pa iz *Življenja in dela Venca Poviškaja*.

Tabela 54: Najdaljših 5 povedi iz prvega vzorca

1. (1002 besedi) Brez rjuh se odeje doma ne smejo uporabljati, med spanjem mora biti blazina obrnjena, da se le ne umaže ali preveč izrabi, bi moral vsakdo imeti med blazino in rjuho, lastno odejo ali kaj podobnega, prvi torej v mesecu mora vsak pustiti rjuhe na postelji, prevleke pa mora vsak sneti z zglavnika zastran menjave, če so rjuhe popackane zaradi nočnih izlivov semena, je treba taka mesta skriti na ljubo nunam, ki pobirajo rjuhe, v kopalnicah je treba že pred slačenjem vsakikrat previdno uporabiti plastično zaveso, poleg tega prha ne sme biti preveč in predolgo odprta, da se voda ne tlači, če se je kaj vode razlilo, jo mora vsak takoj obrisati s čunjo, po kopanju je treba pustiti kopalnico tako snažno, kakor bi jo hotel najti vsak, zato jo je treba osnažiti z gobo ali metlico in paziti, da ne ostane odtok zamašen z lasmi, isto pravilo mora veljati tudi v straniščih, da se leta ne zamašijo, se smejo riti brisati le s toaletnim papirjem, za odpiranje hlačnega razporka je treba uporabljati levo roko, tisto reč je treba izvleči in držati s tremi prsti levic, curek je treba usmeriti v školjko in pri tem paziti, da se ne razpršuje, zadnje kaplje je treba otresti z roko, na tleh ne sme biti ne koščka papirja in ne kapljice vode, med scanjem naj bo pogled uprt naprej, torej pokončno, naravno, zavodar je naraven in pokončen človek, še vedno smo pri dobesednem navajanju, obiskovanje stranišč in kopalnic je treba vključiti v vsakodnevno izpraševanje vesti, ker so dijaki tu najbolj izpostavljeni nečistoti, izpraševanje vesti mora potekati po določenem izpraševalnem pravilniku: se nečisto slačil, dotikal, opazoval, gledal, kupoval, daroval v nesramne namene, nesramno pisal ali risal ali hotel nesramno pisati in risati, nesramno govoril, pel, pripovedoval, se smejal, mežikal, namigoval, poslušal, vohal, k nesramnosti spodbujal, dajal povod, svojega bližnjega ne odvrtač od nje, se ji ne javno uprl, nečisto mislil, želel, hrepenel, si predstavljal, delal ali tudi samo hotel nečisto misliti, govoriti, peti, pripovedovati, se smejati, želiti, hrepeneti, si predstavljal, pogledovati, mežikati, namigovali, poslušati, vohati, k nesramnosti spodbujati, dajati povod, svojega bližnjega ne odvrtač od nje, se ji ne javno upreti, nečistovati, sam ali z drugimi, s kom, s fantom ali deklico, s kakim starim fantom in kako mlado deklico, fantom ali deklico kakšnega rodu, premožnega ali revnega stanu, je uspelo, ni uspelo, kolikokrat in kako pogosto, kako dolgo vsega skupaj in kako dolgo posamič, in tako naprej, zapomni si: kar je za čistočo in telesno zdravje potrebno, to je tudi dovoljeno, vse, kar se s tvojim telesom dogodi, ne da bi bil to sam hotel ali sam zakrivil, vse to ni greh, temveč božja volja, vidiš taka je božja volja, bog te je iz svoje neskončne milosti ustvaril takega, izkaži se vrednega tega, posnemaj dobre zglede, ugašaj luč, luč ne sme nikjer goreti brez potrebe, kaditi je prepovedano, tudi odraslim ni dovoljeno kaditi, ker gre dim v kapelo in Jezus ne prenese kajenja, ako je v kapeli služba božja, naj je nihče ne moti z glasnim govorjenjem, če se sliši na hodnike zborovsko petje, petje korala, odpevov, ljudsko petje in tako dalje, ga je treba spoštljivo poslušati, služba božja je božja služba, v dvorani je televizor vključen navadno od pol osmih do desetih, pomembne oddaje izven tega časa bodo oznanjene na glasni deski, tam, kjer je nabit tudi oznanilo tedenskih božjih služb, torej red svetih maš, televizorja se ne sme dotikati nihče razen tistega, ki je za dotikanje pooblaščen, časa za gledanje televizije ne bo na pretek, vsak stanovalec doma je dolžan prispevati nekaj svojega prostega časa za skupne potrebe, polti na primer nabiranje črnic po faroških gozdovih v sosesčini, doma je treba obirati sadje, ki ga stiskamo in sok prodamo, da se z izkupičkom more znižati zavodnina, vsak je dolžan nabrati določeno količino teh sadežev za skupnost, kdor bo nabral več, bo javno pohvaljen in nabit na glasni deski, pri nabiranju mora vladati red, pred očmi je treba imeti občo blaginjo, nabirati je treba vzorno in marljivo, več kot sto metrov se nihče ne sme oddaljiti od zbiralnega tovarnjaka, kajti zgodilo se je že, da so se nekateri preveč oddaljili, izginiili v gmovju in podraščeni, se tam slekli ali pa tudi ne in kar tako občevali v črničju, ali pa so drug drugemu segali v hlačne žepce in se šlatali, dokler ni prišlo, to je seve prepovedano in zdravju škodljivo, zato bo upoštevanje teh navodil strogo nadzorovano, kajti marsikatera ženska zaman čaka na izliv semena, tu pa igračkanje in potrata, to ne gre več, vodstvo doma tega ne bo trpelo več, zato tisto določilo o sto metrih, to mora zadostovati, poltici črnice, kot smo rekli, jescni pa pobiramo krompir pri večjih okoliških kmetih, če imajo kmetje hčere ali druge nastavljenke, pobirajo te na drugi njivi ali ob drugem času, zato kmetje dajo domu zastonj nekaj stotov krompirja, za kmete je treba moliti, da jim bog da dobro letino, kadar pobiramo krompir na grofovskih poljih, je treba pobirati urno in natančno, misliti je treba na to, da se pri gospodu grofu zapusti dober vtis, grof je ugleden gospod in ima vpliv, počasi prihajamo h koncu dobesednega navajanja hišnega reda, kdor se pregreši proti hišnemu redu, pa četudi samo proti eni sami določbi, dobi za kazensko nalogo spis pod naslovom Krščanske vrlin hišnega reda, kdor je tak spis že pisal in se vnovič pregreši ter mora to nalogo pisati vdrguč, bo izključen iz doma, dva taka spisa pomcnita torej isto kot odpust iz doma, to je Tjaž dobro vedel, da to imamo dokaz in to je Tjažev svojeročni spis o krščanskih vrlinah hišnega reda, ki ga hranimo kot dragocen spomenik med njegovimi dokumenti v trezorju, kljub temu se je znova pregrešil in bi bil moral spis pisati vdrguč, tega seveda ni več storil in ga tudi nismo silili v to, sam si je izrekel sodbo, njegova izključitev iz doma je bila torej vseskozi postavna, torej v skladu z določili hišnega reda in smo imeli potemtakem vso pravico do nje, sam nam je potisnil v roke pravne prijeme.

2. (659 besed) Da bo čez nekaj časa tako končal, tega ni nihče mogel predvidevati in tudi ne slutiti, tega tudi nismo namerali, Tjaž pač ni bil človek za zavod in je bolj zašel kot prišel sem, predolgo se je zadrževal pri nas, zaradi njegovoga socialnega položaja smo ga vzeli, iz krščanskega usmiljenja tako rekoč, nobene prave vzgoje, nobenih dobrih zgledov, nobenega ugleda, brez idealov, in to, starši na slabem glasu, oče odstavljen drvar in zapit cestni pometač, mati kacetovka, podpirala je partizane, tega bi ji ne bilo treba, takih in podobnih svinjarij si je nakopala, vso okolico je okužila, sama je bila kriva in prav se ji je godilo, zločince je treba odstraniti iz človeške družbe in jo zavarovati pred njimi, izdajalce je treba umoriti, iz teh vzrokov smo ga vzeli v dom, samomor na nima z domom sploh nič opraviti, o samomoru ni nikoli govoril, samomorilskih misli ni nikoli razlagal, tudi njegovo vedenje ni dopuščalo tega sklepa, čeprav je bilo nekoliko čudaško, predaleč bi šel, kdor bi spraval njegov samomor kakor koli v zvezo z zavodom, človek njegove sorte pač ni mogel končati drugače kot tako, kot je končal, kaznoval se je sam, ker je zašel na kriva pota in se neprimravljen znašel pred nesmiselnostjo svojega početja, njegova vest je bila tankočutno vzgojena, to se nam je v teh kratkih letih z božjo pomočjo posrečilo, s takšno vestjo je do zadnjih odtenkov spregledal svojo zanikrnost, presedlo mu je dolgočasje tako vtollega životarjenja iz dneva v dan in si je zato sodil sam, to je treba jasno pribiti, ne da bi odobraval njegovega koraka, nikakor, čeprav odpade zdaj neka skrb in smo se z njegovim koncem znebili spet neke odgovornosti, skrb in odgovornost manj, tudi to je nekaj, kar je treba videti, nko breme nam je bilo odvzeto, ljudje delamo in bog nardci, ljudje pripravljamo in bog izvrši, ljudje stopamo, bog stopi, breme nosimo in omahujemo pod njim, pa nas razbremeni on, ki je neskončno dobrtljiv, odvahi od bremen breme, katero najmanj pričakujemo, da ga bo, pa ga, težkim bremenom sami ne moremo do živega, on pa jih odpihne, slava in čast mu za

Tabela 54 (nadaljevanje): Najdaljših 5 povedi iz prvega vzorca

to, naše ljudstvo ni zastonj verno ljudstvo, božja volja je bila, da je umrl, preden je mogel storiti še večje hudobije, prav gotovo bi jih bil storil in je bilo tako za vse najboljše, da se je tako rešil, bog mu bodi dober sodnik, ni imel kaj prida na tem svetu, molili bomo za zveličanje njegove duše, to bo potrebno, jutri bomo s tein začeli, kajti njegova krivda vpije v nebo, res žal za črnilo in papir, saj je vsak lahko vnaprej vedel, da se bo postavil proti nam, vrgel je senco na zavod, ponižal nas je, ki smo bili njegovi prijatelji in smo mu želeli samo dobro, slekel nam je kožo, da bi nas očrnil, vendar mu odpustimo, nismo se umaknili v stolp samozadovoljnosti in se oddali z užaljenostjo, nismo ga samo vrgli iz doma, kakor je po pravici zaslužil, temveč smo mu tudi odpustili, Malharju je tako rekoč odsekal uho, mi pa smo ga ozdravili, brez ušesa ne more biti, Malhar ne more biti brez njega in uhelj ne more biti brez Malharja, kajti uho brez Malharja ni uho in Malhar brez uhlja ni Malhar, vsaka stvar se mora družiti s pravo stvarjo, vsaka stvar mora biti na pravem mestu, vedno smo bili pripravljeni, da ga ozdravimo, ako hi se vrnil, čut in ljubezen do bližnjega sta nam branila, da bi ga dokončno odstranili, kot bi po pravici zaslužil, ne da bi se nam ne bil maral, kdo to pravi, nismo zaprli vseh poti do njega, vedno bi se bil lahko vrnil iz noči v nerazbiten, svetel dan, če bi bil hotel, pa ni, čudovite izraze premore zavod, nerazbiten svetel dan, besede se ti stopijo na jeziku, ni se vrnil in se je vrnil, vidite, to je tisto, vrnil se je samo toliko, da registriramo njegovo smrt, naredimo križec pri njegovem imenu, niti pozdravil ni več.

3. (591 besed) Zavod je storil vse, kar je bilo v Tjaževem primeru mogoče storiti, vzgajal ga je in mu zaradi težje vzgojivosti in trdosrednega vztrajanja v slabem posvetil neznansko veliko skrb in časa, lahko trdimo, da več kot vsem drugim, ki jih je moral zagledelj zancemariti, dajal mu je nasvete, delil koristne nauke za življenje, poskušal zlepa in poskušal zgrda, skratka, storil je vse, ne samo da bi Tjaž živel, temveč da bi živel, kot se za krščenega človeka spodobi, in če bi zdaj še živel, bi ga lahko sami vprašali, ali se je teh naukov držal, zakaj se jih ni, če se jih ni, in tako dalje, od svojih krivih poti ni obogatel, zavod zaradi njih ni obubožal, Tjaž je ostal takšen, kot se je mislil spremeniti, samo da se je prej že osramotil, za isto mero je kleenil v kolenih, za katero se je hotel povzpeti, ljudje so nerodni in se zunaj sami le redkokdaj znajdejo, posebno še, če dom ni imel dovolj prilike, da bi človeka za to pripravljaj, zavod njegove smrti ni zakrivil, pri nas se tega ni naučil, take stvari preprečujemo, kolikor je v naši moči, nismo pogrebni klub ali pokopališko društvo, če se tako vzame, vsoga je kriv sam, ker si ni znal naložiti prave količine svobode, pri nas jo je prejemal po malem, vsakič prav toliko, da si je poželjš za kancev več, ko jo umaknejo, zadostila te je in te spet ni zadostila, ker si je prav zaradi tistega pridržanega kance znova zaželiš, dozirano tako rekoč, po toliko, kolikor je človek prenese, ne da bi pretrpel škodo, v pravi količini torej, tako kot hoče bog, Tjaž pa se je svobode nažrl, drugače njegovih pustolovščin ni mogoče imenovati kot s to grdo besedo nažrl, to so bili tudi vzroki, da je šel rakom žvižgat, zdaj žvižga rakom, kakor je zaslužil, zaradi žrtja se je pretegnil, vedel se je kakor svinja z mehonom, ni bil kos življenju v svobodi, ni našel mere in sredine, zato je videl rešitev v samomoru, mislil je, da se samo tako more za svoje presedlanje opravičiti pred lastno vestjo in pred javnim mnenjem, verjetno so vplivali na tako vednje tudi določeni kompleksi, ki so bili povezani s poklicno neuravnovešenostjo, kar je treba upoštevati kot olajševalne okoliščine, svobodo besede, mnenja, mišljenja, odločanja in ravnanja smo mu dajali tudi mi, seveda v smislu hišnega reda, red mora biti, v njegovih mejah so bile možne vse svoboščine, v tem okviru si je lahko dovolil najbolj si nasprotujoče razlage določil in predpisov, ki vsi izvirajo iz krščanske skrbi za našo mladino, kar je obče znano, in ki so za mlačne Kristjane prav zato precej zahtevni, to uvidimo in upoštevamo, kolikor se da, sicer pa je treba vedno več zahtevati, da se malo doseže, sami se lahko prepričate o krščanskem duhu, ki veje iz vsch določil in predpisov hišnega reda, temu v dokaz citiramo dobesedno od škofijstva potrjena sveta besedila, tako dobesedno citiranje se nam zdi umestno, da tako razodeneo hišni red iz prve roke, torej v nedotaknjeni in neomadeževani, tako rekoč v deviški obliki, kolikor široki javnosti še ni znan, in nam nihče ne more podtakniti, da smo ga po svoje zasakuli, narobe, takšen tekst bo poznavalec snovi napolil do sklepa, kako lahko bi se ga bilo držati in kako lahkomišlno so ravnali tisti, ki se ga niso držali, težko je priti do drugih sklepov, sploh ne moreš priti do njih vpricho nevarnosti, da bodo dobesedni citati mogoče spališlog in pokvarili barvitost našega stališča, vendar je kaj takega pač treba stvari na ljubo vzeti nase.

4. (587 besed) Hišni red torej zelo nazorno zahteva, da se morajo vsi v domu stanujoči počutiti kot družina, zdaj že začenja mo z dobesednim citiranjem, zato mora vsak skrbeti za dobre medsebojne odnose, takšne, kakršne hoče bog, od vsakega stanovalca se zahteva pošteno obnašanje v domu in zunaj njega ter vestno upoštevanje hišnega reda ob vseh prilikah, kakor tudi slučajnih predpisov, stanovska dolžnost vsakogar je, da se pridno uči in da pravočasno opravlja izpite, poleg tega pa skrbi za dušni blagor, grobo zanemarjanje te dolžnosti bi imelo za posledico tudi izključitev iz doma, zaradi nevarnosti izgubljanja časa ni dovoljeno, da se kdo mudi v sobi koga drugega dalj časa kot deset minut, za daljši razgovor so na voljo številni skupni prostori, za skupno učenje v privatni sobi je potrebno izrecno dovoljenje, obiski so dovoljeni le v prtiličnih prostorih in le do pol desetih zvečer, brez posebnega dovoljenja ne sme nihče nikogar pripeljati v stanovanjske prostore ali ga sprejeti v sobi razen lastnih staršev, vendar je treba tudi v tem primeru obvestiti vodstvo doma, služkinje, ki stanujejo v domu, imajo prost vhod le za opravke v uradu doma, vstop v prostore zunaj lastne sobe jim je strogo prepovedan, enako je prepovedan tudi fantom vstop v prostore uslužbencev, v sobah in v vsem domu morata vladati red in snaga in je treba paziti, da se nič ne poškodeje, v čitalnicah in skupnih prostorih je treba pustiti vsako stvar po uporabi na svojem mestu, kdor povzroči kako škodo v domu, jo mora poravnati, če krivec ostane neznan, bodo stroški popravila naloženi vsem, pribijati ali lepiti slike na zidove ali opremo je strogo prepovedano, kdor želi obesiti na steno kak okrasek, potrebuje za to posebno dovoljenje, v vsaki sobi mora viseti razpelo, v razpelo se je treba večkrat ozreti, paziti je treba tudi, da se tok, plin in voda ne uporabljajo po nepotrebnem, za sušenje perila se lahko napelje trak nad umivalnikom med vijakoma, prilepljene ali z zrakom pritrjene plastične kljuke povzročajo škodo in so zato prepovedane, bivanje v domu bo toliko bolj prijetno, kolikor bolj se bo vsak zavzel za dobro obnašanje, zato naj vsak te predpise in hišni red nosi pri sebi ali hrani na priročnem kraju in jih prebira, tudi prenišlje naj o njih, vsak naj skrbi za dober glas med tovariši v hiši in zunaj nje, obrekovanje zavoda in njegovega vodstva ni dovoljeno, ponoči mora imeti vsak največjo skrb, da nikakor ne moti nočne tišine, vrata je treba zapirati zelo počasi, da ne treskajo, kljuke ni dovoljeno nenadoma izpustiti, ker drugače v železju zaškrtja, tudi ključje je treba obračati počasi, da se prepreči ropot, tekanje po stopnicah in hodnikih je prepovedano, hoditi je treba na moč potiho,

Tabela 54 (nadaljevanje): Najdaljših 5 povedi iz prvega vzorca

prav tako ni dovoljeno glasno govorjenje, vsak naj večkrat pomete svojo sobo in jo obriše z mokro cunjjo, ki naj meri 20 x 15 cm, po navoženem podu je treba hoditi v copatah, vsak kos obleke ali papirja mora biti v sobi na pravem mestu, vsako soboto je treba umiti okna z mokro cunjjo in obrisati s časopisnim papirjem, za zračenje sobe navadno zadostuje, da je odprta vrhinja oknica, vsak je odgovoren za morebitno poškodbo šip ali zaves, da se pozimi prepreči nepotrebno izgubljanje toplote, je treba sobo zračiti le kratak čas, oknice je treba premikati zelo previdno, da se ne poškodujejo, raba hladilnika je omejena na določeno plastično škatlo, ki stane 39 šilingov, in na polliterko steklenico, v hladilniku ne sme biti navadno nič drugega, zdaj smo sredi dobesednega navajanja, vsak naj ima svojo odejo, ki jo sme uporabljati, da na njej leži ali sedi.

5. (564 besed) Nunam je najino poznanstvo ostalo še prikrito, potlej sva tvegala zmeraj več, doživel je prve vrhunce z menoj, ostajal do jutra, zdaj ko je mrtev, to lahko povem, pustila sem mu, ker mi je ugajalo, za šalo in iz radovednosti, pomagala sem mu, da je vse zamujeno in zapravljeno spravil pod streho, skraja je odhajal zelo pobit in zaskrbljen, pekla ga je vest in obujal je kes, ob misli na zavod se ga je polaščal živalski nomir, toda počasi se je navadil, mirila sem ga, mu razlagala in ga prepričevala, skupno sva krotila božjast in privide in šlo je, nekaj v njem se je prelomilo na dvoje, nekaj se je obrnilo, kar čutila sem, kako se v njem obrača in presnavlja, kot v kvasu je delovalo v njem, veržilo me je iz odpora, nalajše in iz kljubovanja, v meni je našel sredstvo, da je krenil iz brazde, nikoli ni grdo ravnal z menoj, surovosti ni poznal, to moram priznati in bi priznala tudi v primeru, ko bi še živel, ne govorim samo zaradi njegove smrti tako o njem, bil je dober z menoj, v dokaz navajam, da mi je, čeprav je bil brez denarja, kupil uro, verižico, knjige in druge stvari, ki sem jih potrebovala ali si jih želela, znal je uganiti, vselej je bil nežen in obziren, poniževal me je samo občutek, da mu nisem bila namen, temveč sredstvo, mogoče sem si to samo domišljala, opravičuje pa me dejstvo, da nisva mislila nič resnega, nazadnje me to ni več motilo, sprjaznala sem se, nisem tega storila zastonj, sama sem skrbela, da nisem ostajala prikrajšana, ni bil prvi in ne zadnji možki, bil je eden od tistih, ki so bili po svoji usodi edinstveni in zato nekako vendarle prvi in hkrati zadnji, hočem reči, ta mladi možki je z mojo pomočjo uresničil samega sebe, postal je Tjaž, kakor še nikoli pred menoj in nikoli za menoj, najbolj resnični, prvobitni in izvirni Tjaž, kot ga poleg njega ni bilo drugega, prepričana sem, da nikjer na svetu, kot ga ni bilo drugače in ob drugem času in iz drugih razlogov in z drugih vidikov in v drugi nakladi, pri tem dvomim, da ponovim domnevo, da je ta čudaški človek kdaj prej prekopal svojo naravo res za vsemi duhovnimi in telesnimi dobrinami in z njimi tvegal strast, in če jih je bil izkopal, komu naj bi jih bil izkazal, vse to je imel za preveliko nedostojnost in greh, rajc se jim je odpovedal in jih s silo zatri, junak Tjaž ni bil, manjkalo mu je poguma in samozavesti, bal se je, da bi ga strast ljubljenja ne razvnela tako, da bi zraven od naporov in razburjenja crknil, ugasnil kot sveča, pobodlo bi, mogoče bi v dolžini četrtr kolesa zapelko in najbrž bi zamudil, da je mimo, da je život dotrajal in se enkrat za vselej zavozljal ravno v trenutku, ko bi se pričel orgazem, tako da njegovega konca ne bi več pričakal, pred to možnostjo je vselej trepetal in se zgrozil, čeprav bi bil to lep in lahek konec, kar nasladna, pravičajna smrt bi ga doletela, ne, junak Tjaž nikakor ni bil, kvečjemu je bil junak tisto uro, ko se je odločil za skok z najvišje stolpnice celega mesta, a to je bilo pozneje in čisto ob koncu, toliko da se je še tik pred svojim koncem nalezl junaškega slovesa, junak ali nejunak, naj bo, kakor hoče.

Tabela 55: Najdaljših 5 povedi iz drugega vzorca

1. (284 besed) Na vse zadnje so se socialno čuteči, po sami rimski in božji previdnosti postavljeni tržaški občinski svetniki do grla nasitili, gledati in poslušati brezposelne, na cesto pognane družinske očete, zapuščene vdove in onemogle starce, ki so slabokrvni in jetični od lakote, garjavi, nadušljivi in polni revmatizma od prenočevanja po zatohlih kavernah in zasmrajanih mestnih podzemskih kanalih, skrhanji in sključeni, razjedeni in oglodani od trdega življenja, vsak dan v večjih trumah prihajali na magistrat ter krevljali po dolgih svetlih hodnikih, puščali na preprogah rjave odtise svojih blatnih, pošvedranih obutev, pobirali čike izza pljuvalnikov, se vsekavali kar s prsti, zamolklo kašljali, grkali, smrkali in metali široke pljunke kamor je naneslo, stresali bolhe, uši in stenice, zaudarjali po kistem smradu raztrganih in gnilih cunj, prijemali s svojimi umazanimi lopatastimi rokami za svetle medene kljuge na vratih, vstopali v sobe in se raztezavali po njih kakor smradljiv, dušeč dim, se obesali za suknje gospodom uradnikom, ki so tako lepo dišali po milu, se odkrivali, držali raztrgana pokrivala z obema rokama nizko na kolennih ter dokazovali, da niso divjaki, temveč ljudje, ustvarjeni po božji podobi in volji, trdili, da so Italijani od pamtiveka, fašisti od vsega začetka, državljani, udani Mussoliniju, kralju in papežu, ter prosili košček strehe svojim prezeblim otrokom in bolnim ženam: saj morajo vendar živeti, dokler jim je usojeno; ali naj se mar lepo zleknejo ob cesti in mirno počakajo na ljubo smrt; ali naj se pobesijo na hlačne jermene; ali naj svoje otroke pobijejo, pokoljejo, pomečejo v morje; ali naj noseče žene usmajajo ponoči po cestah in ponujajo po dve liri svoje kosti - o, saj bi jih, toda nihče se ne zmeni zanje, ker je dovolj bolje ohranjenih žensk; ali naj začnejo krasti, pobijati - ali kaj?

2. (240 besed) In zakaj je ne bi dokazali, če se nudi prilika? - mrzlo in uradno dokazujejo, da je ločila postavljati smiselno, čeprav jih ni postavjal po splošno veljavnih pravilih; da je njegov jezik živ, lep in tekoč, čeprav ne bi mogli trditi, da bi bil kremenit; da je bil dober stilist, nemara celo predober, saj ga je včasih zaneslo v drugo skrajnost, v artizem; da je bil mojster verističnega, realističnega dialoga, čeprav ga ne moremo prištevati k čistim realistom; da je v njegovih delih precej pristnih epičnih momentov, a da se nikdar ni povzpел do pravega epičnega pisanja; da je bil zakrknjen lirik s precejšnjim romantičnim navdihom; da je bila njegova satirična žila tako močna, da je včasih brizgnilo iz nje tudi tam, kjer ne bi bilo treba; da so njegove misli kar lepo zaokrožene, večkrat celo izvorne, čeprav niso nenavadno globoke; da je bil miselno napreden, čeprav z lahkoto zasledimo, da se ni ukvarjal s proučevanjem družboslovnih del; da so njegovi značajni doživeti in živi, ljudje iz mesa in krvi, čeprav niso monumentalne, enkratne figure; da je gledal na svet nekoliko zviška in z rahlim posmehom, čeprav je bil v resnici prežet s humanizmom; da je napisal zelo malo, a da je njegovo pisanje kulturno, dokaj svojsko in sugestivno, čeprav mu naposled vendarle nekaj manjka; da so da je skratka, vsemu navzlic navkljub je bil mož človek in umetnik in zato je prav res škoda, ker je legel v prerani grob

Tabela 55 (nadaljevanje): Najdaljših 5 povedi iz drugega vzorca

3. (164 besed) Odprla je svoja težka okovana hrastova vrata in začela požirati stare črvice omare z neštevilnimi predali; škripajoče postelje s trhlili nogami; polomljene stole; raztrgane divane, iz katerih je silila vlažna in umazana morska trava in ki so bili v tej gneči tako podobni ranjencem s preparanimi trebuhu; kolesa otroškega vozička, pribita k navadnemu zaboju s širokim napisom Cirio zmeroni in povsod ali pa Vzemite po kosilu pomarančo Gaddi, ki Vas osveži in Vam pospešuje prebavo in v njem jokajoč, koščen otrok; razklane kovčge, pisane cule, raztrgane vreče in podobno slikovito ljudsko bogastvo; požirala je hiša starec z redkimi, ščetinastimi, sivimi bradami; upognjene ženice s pridruženim stokanjem na ustih; trde, zarjavele može, ki so preklinjali pod težkimi bremenji; žene s širokimi nosečimi trebuhu in z dojenčki na mlahavih, cunjastih prsih; otroke, ki so venomer vreščali, hlipali in smrkali; hihitajoče se blodnice po ljudskih cenah in žvižgajoče potepuhe, ki so mirno prihajali s kapo postrani, s pisano kriččo ruto za vratom, in kadili svoje večne čike.

4. (156 besed) Poslovil se je od nadaljnjih radosti in tegob življenja na zemlji, od norčavih južnih vetrov, ki so razigrano pihali od vseh štirih strani na vse štiri strani sveta, veselo oznanjali pomlad in naravnost z otroško razigranostjo vneto tlačili dim iz dimnikov nazaj v kuhinje, da so se gospodinjce enerile okrog ognjišč in s solzami zalivale jedi, ki so se kuhale po piskrih; poslovil se je od zemlje, ki se je dramila iz zimskega spanja, od teloha, ki je že odcvetal in se vdano pripravljaj na smrt, od cvetočega resja, od čebel, od vrh in njihovih mačic, od mladik, ki so se vjugale k mlademu soncu in naglo poganjale očesca, od vonja po pravkar nakidanem gnoju, od bližajočih se velikonočnih praznikov in od otrok, ki so se podili po lužah, se lasali, se obdelovali s pestmi in proti ter se vselej, kadar koli so zagledali Venca Poviškaja, z novimi močmi grabili za trebuche in kričali za njim.

5. (138 besed) O, tudi njim od skrbi sivijo lasje: delnice N. G. I. rapidno padajo; delnice Cosulich Line so že brez vsake vrednosti in jih lahko mirne duše vzamejo s sabo na stranišče; je v obupnem položaju - treba bo napovedati prisilno poravnavo, da se nekoliko opomorejo; se ne obrestuje; žena ima kamne v nehurju in hoče na vsak način v Viareggio; služkinja je razbila dragoceno japonsko vazo in milostljiva bo prav gotovo dobila živčni napad, ko bo to zvedela; hči je zbežala s centurionom nilice in zdaj pošilja pisma ter grozi z družabnim škandalom, če ji ne odštejejo mastne dot; v gimnaziji je afera s sinom, ki je pokazal več talenta za zvodništvo mladih deklet kakor za Risorgimento; Lola, angel plavi, zahteva sto lir na večer, in če jih ne dobi, odpotuje - težko je dati stotak, toda še težje je biti brez Lole.

5.3.2 Najpogostejše povedi

Povedi je za velikostni red manj kot besed. Iz tega in iz dejstva, da je največ vsaj nekajbesednih, bi bilo pričakovati, da bo najpogostejša poved imela že precej nizko frekvenco. V tabelah 57 in 58 je navedenih približno 100 (do zaključene meje neke frekvence) najpogostejših povedi iz prvega in iz drugega vzorca. Po pričakovanju prevladujejo kratke povedi, vzete iz premega govora ali povezane z njim. V prvem vzorcu je približno desetina (13) konkretnih, od tega 9 iz *Nove zaveze* (*Jezus je odgovoril: Jezus ji reče: Jezus jim je odgovoril: Jezus jim je rekel: Jezus mu je odgovoril: Jezus mu je rekel: Jezus mu odvrne: Jezus pa jim je rekel: Jezus pa mu je rekel.*), v drugem pa še dosti več, predvsem iz *Balade o trobenti in oblaku* in iz *Tantadruja*.

Po zvrsti je v prvem vzorcu med prvimi 109 32 % pritrilnih (npr. *Da. Tako je. Vsekakor.*), 22 % vprašalnih (npr. *Kako? Ali ne? Kako to misliš?*), 18 % takih, ki uvajajo premi govor (npr. *Rekel jim je: Krpan odgovori: Odgovorili so mu.*), 13 % nikalnih (npr. *Ne. Ne vem. Mislim, da ne.*), 9 % velelnih (npr. *Lahko noč! Ha, ha! Ne boj se!*) in samo 6 % povednih (npr. *Nič. Zasmejal se je.*).

Pri drugem vzorcu je podoba precej drugačna: pritrilnih povedi je 13 %, vprašalnih je 22 %, takih, ki uvajajo premi govor sploh ni, nikalnih je 6 %, velelnih 39 %, povednih pa 20 % (npr. *Tišina. «Hm, kajpak,» se je popraskal kmet. Tišina. «Na nebu je sonce, na zemljici mraz.*).

V preseku med obema vzorcema (naveden je v tabeli 56) je le 16 povedi, ki so vse zelo kratke in stereotipne.

Tabela 56: Presek najpogostejših povedi obeh vzorcev

»Da.	Kam?	»Ne!	»Seveda.
Kaj?	»Kje?«	»Ne, ne!	»Tako!
»Kaj?	Ne!	Nič!	»Tako je!
»Kako?	»Ne.	Nič.	Zakaj?

Tabela 57: Najpogostejše povedi v prvem vzorcu s frekvencami

1. »Da.«	191	38. »Ne vem.«	14	75. »Da, tako je.«	8
2. »Tako je.«	184	39. »Strinjam se.«	14	76. Odgovorili so:	8
3. »Vsekakor.«	86	40. Rekel je:	14	77. »Kaj pa je to?«	8
4. »Seveda.«	71	41. »Jasno.«	13	78. »Kje?«	8
5. »Gotovo.«	54	42. »Pravilno.«	13	79. »Ne boj se!	8
6. »Ne.«	47	43. »Zakaj?«	13	80. »Ne, ne!	8
7. »Kako?«	46	44. Ne!	12	81. »Nič.	8
8. Doktor Pipec.	38	45. »Popolnoma pravilno.«	12	82. »Pridi!«	8
9. Doktorand Koleneč.	35	46. »Se strinjam.«	12	83. »Tako je!	8
10. Ali ne?«	34	47. Dejal je:	11	84. Jezus jim je rekel:	7
11. »Prav gotovo.«	33	48. Jezus mu je rekel:	11	85. Jezus ji reče:	7
12. »Nikakor ne.«	32	49. Kaj?	11	86. Jezus pa mu je rekel:	7
13. »Nujno.«	32	50. Lahko noč!«	11	87. Je tako?«	7
14. Ne.	31	51. Odgovoril je:	11	88. Kajne?«	7
15. »Kaj?«	30	52. Jezus je odgovoril:	10	89. Kaj je to?	7
16. »Kaj?«	28	53. Jezus mu je odgovoril:	10	90. Nikakor ne!	7
17. »Zakaj?«	26	54. »Ja.	10	91. Odgovorili so mu:	7
18. Rekel jim je:	26	55. »Nikakor ne!«	10	92. »Dobro.«	7
19. Jezus jim je odgovoril:	24	56. »Prav.«	10	93. »Kaj misliš s tem?«	7
20. »Očitno.«	24	57. »Veronika!«	10	94. »Kako misliš to?«	7
21. »Dobro.	23	58. »Zakaj ne?«	10	95. »Kateri?«	7
22. Odgovoril jim je:	21	59. Da.	9	96. »Kdo?«	7
23. »Ne.	21	60. Ha, ha!«	9	97. »Mislim, da ne.«	7
24. »Ne!	20	61. Jezus pa jim je rekel:	9	98. »Ne!«	7
25. »Kako to misliš?«	19	62. »Da!	9	99. »Res?«	7
26. Amen.	18	63. »Gotovo.	9	100. »Sem.«	7
27. »Tako se zdi.«	18	64. »Jaz?«	9	101. »Seveda!	7
28. »Kako?«	17	65. »Nič!«	9	102. »Tako!	7
29. Zakaj?	17	66. On pa jim je rekel:	9	103. »Vsekakor.	7
30. »Brez dvoma.«	16	67. »Štiri!«	9	104. »V tem imaš prav.«	7
31. »Seveda.	16	68. »Tako?«	9	105. Prav.	7
32. »Da.	15	69. Zasmeljal se je.	9	106. Rekli so mu:	7
33. »Dobro!	15	70. Jezus mu odvrne:	8	107. Seveda!	7
34. »Res je.«	15	71. Kam?«	8	108. Tako!	7
35. Nič.	14	72. Krpan odgovori:	8	109. Vprašali so ga:	7
36. »Ja.«	14	73. Ne vem.	8		
37. »Ne vem.	14	74. Nič!	8		

Tabela 58: Najpogostejše povedi v drugem vzorcu s frekvenca

1. »Hm, kajpak,« se je popraskal kmet.	32	42. »Pa kaj hočemo, je že tako! ...	5
2. Tišina.	24	43. »Stane!	5
3. »Hm, kajpak ...	22	44. Aha!	4
4. »Kaj?	22	45. Ali razumesš?«	4
5. »Ne!	18	46. Hahaha!«	4
6. Tak!	14	47. Halo!	4
7. »Aha!	14	48. Halo!«	4
8. Hm?	12	49. In kajne, gospod?«	4
9. Ne!	12	50. Kaj še!	4
10. Mhm ...	11	51. Kaj?	4
11. »Pha!	11	52. Kje pa!	4
12. A?	10	53. Nič!	4
13. »Seveda.	10	54. Nič.	4
14. Tako je to!	9	55. Pa kaj hočemo, je že tako! ...«	4
15. »Hm, kajpak,« se je popraskal.	9	56. Peter je molčal.	4
16. »Prišla bo bridka smrt, moj hramček bo zaprt ...«	9	57. Pijte! ...«	4
17. »Tak!	9	58. Raus e patacis, repa in krompir!«	4
18. Zakaj?	8	59. Stane!	4
19. »Drejc!	8	60. Ta-a-ta-an ta-a-ta-an ta-a-an-ta-druj.	4
20. »Tako je!	8	61. Top!	4
21. Kam?	7	62. Trenutek tišine.	4
22. No!	7	63. V Obrekarjevi izbi.	4
23. »Halo!	7	64. Vidiš,«	4
24. »Raus e patacis, repa in krompir!«	7	65. »Ana!	4
25. A?«	6	66. »A?!	4
26. Ju-u-hu-u, ju-u-hu-u ju-hu-hu-hu-huj ...«	6	67. »Bom!	4
27. Mar ni res?	6	68. »Da.	4
28. Uuu!	6	69. »Ha!	4
29. »Boga mi!	6	70. »Kaj pa je?«	4
30. »Na!	6	71. »Kje?«	4
31. »Na-a nebu je-e sonce, na zemljici mraz.	6	72. »Ne, ne!	4
32. »Saplater!	6	73. »Ne.	4
33. »Stric!	6	74. »No!	4
34. Ali me poslušáš?	5	75. »Pha!« je zaničljivo prhnil Furlan.	4
35. Molk.	5	76. »Pijte!« je rekel.	4
36. Na!	5	77. »Poslušam ...	4
37. Na-a-biram ja-az zvonce in vsi so za vas.	5	78. »Saj!	4
38. Seveda.	5	79. »Saplater?	4
39. »A!	5	80. »Tako!	4
40. »Kako?	5	81. »Tako.	4
41. »Ooo-hooo ...«	5	82. »Zakaj pa?	4

5.3.3 Lipogramske povedi

Pri povedih so zanimivi še lipogramski pojavi, to je deli besedila brez ene ali več črk. Primera sta npr. angleška poved

Youth, throughout all history, had had a champion to stand up for it; to show a doubting world that a child can think; and, possibly, do it practically;

you would constantly run across folks today who claim that 'a child don't know anything'. (Wright 1939) in francoska

. . . *Ton bouton, ou tout va mourir, ton bouton, bastion final ou j'irai m'annulant, ou j'irai m'absorbant, m'abolissant dans un amour toujours a accomplir, dans l'absolu sursaut ou nous vivrons un jour, confondus a jamais, dans la passion ou dans l'oubli, dans la nuit ou tout disparaît, dans l'infini instant ou nous n'aurons qu'un corps!* (Perec 1969). Oba primera sta vzeta iz del, ki sta bili napisani brez uporabe črke *e*. Niti v prvem niti v drugem vzorcu ni kakega takega dela, je pa precej povedi, ki so zanimive s tega vidika. V tabelah 59 in 60 je navedenih po pet najdaljših povedi iz prvega in iz drugega vzorca, ki ne vsebujejo črke *e*. Vsega skupaj je v prvem vzorcu 9.102 takih povedi ali 4,52%, v drugem pa 2.958 ali 7,90%.

Tabela 59: Najdaljše povedi iz prvega vzorca brez črke *e* z dolžinami v znakih

- 121 In Augusta Comta: duhovniki morajo vzgajati mladino, voditi vsa duhovna opravila, pozivati oblast, naj uravnava rojstva.◀
 97 Tri noči sta ga skrivaj nosila od Jaromila, ga nasipala v listnjaku in ga pokrivala kakor zaklad.
 91 Tistim pa, ki so poklicani, Judom in Grkom, oznanjamo Kristusa, božjo moč in božjo modrost.
 89 Nikomur nista zaupala, zato sta kar sama zdravila živali, jim dajala čistila in odvajala.
 85 Mornarji so ga dvignili, nato so ladjo zavarovali tako, da so jo vso opasali z vrvmi.

Tabela 60: Najdaljše povedi iz drugega vzorca brez črke *e* z dolžinami v znakih

- 90 »Zastrupil,◀ so v zboru ponovili tuji župniki in mu požugali z rožnatimi zalitimi kazalci.
 69 S trudnim korakom drsiš po pločniku in pod pločnikom kar votlo bobni.
 65 Z mračno sivino zro iz polmraka, kakor s trudnimi, plašnimi očmi.
 61 Na gmajnah so rasli bori, podobni jatam črnih žalostnih ptic.
 60 Sčasoma so postali vsi tako domači, kakor bi zrasli na vasi.

V tabeli 59 sta 1. in 4. poved iz prevoda Flaubertovega romana *Bouvard in Pécuchet*, 3. in 5. iz *Nove zaveze*, druga pa iz novele *Boj na požiralniku* Prežihovega Voranca. Zanimive so še povedi, ki vsebujejo največji delež najpogostejše črke. V tabelah 61 in 62 je navedenih po prvih pet takih povedi v prvem in drugem vzorcu, ki imajo vsaj štiri besede (največje deleže imajo medmeti v zelo kratkih povedih).

Tabela 61: Vsaj 4 besede dolge povedi iz prvega vzorca z deležem črke *e* v odstotkih

- | | |
|------------------------------|--|
| 32 Je ne reče ene in je ne!◀ | 31 Če berete, pač berete; če ne berete, pač ne berete. |
| 32 Sesedel se je vase. | 30 Dekletce se ne gane. |
| 32 Še eden in še eden. | |

Tabela 62: Vsaj 4 besede dolge povedi iz drugega vzorca z deležem črke *e* v odstotkih

- | | |
|--------------------------|----------------------------|
| 33 Deset let se že tepe. | 30 »Kje je šele jelševje?◀ |
| 32 Le jejte, le jejte! | 30 Vse stene se reže vame. |
| 31 Ne, ne vleče je. | |

Popolnoma drug zven kot povedi brez črke *e* imajo povedi brez črke *a*. Najdaljše za oba vzorca so navedene v tabelah 63 in 64. 1., 3. in 5. poved v tabeli 63 sta iz *Nove zaveze*, 2. iz Platonove *Države*, 4. pa iz Cankarjeve *Hiše Marije Pomočnice*.

Tabela 63: Najdaljše povedi iz prvega vzorca brez črke *a* z dolžinami v znakih

- 135 Kot božji izvoljenci, sveti in ljubljani, si torej oblecite čim globlje usmiljenje, dobrotljivost, ponižnost, krotkost, potrpežljivost.
108 Če si to prisvojimo pod videzom kreposti, se bomo dobro počutili pred bogovi in ljudmi v življenju in smrti.
106 Tvoji potomci bodo priseljenci v tuji deželi, kjer bodo živeli v suženjstvu in trpeli stisko štiristo let.
101 Stene so bile polne svetih podob; rdeče luči so gorele pred njimi; velo cvetje je viselo od okvirjev.
97 »Pojdi domov k svojim ljudem in jim sporoči, kolikšne reči ti je storil Gospod in se te usmilil.«

Tabela 64: Najdaljše povedi iz drugega vzorca brez črke *a* z dolžinami v znakih

- 76 Hudič bi mu bil verjetno srečno ušel, če bi se bil spustil v grmovje k reki.
73 Bori, bori v tihi grozi, bori, bori v nemi grozi, bori, bori, bori, bori!
67 Še je vse polno ledenih studencev, deročih potokov, modrobelih rek.
66 Brezdušne biče in človeško trpljenje je spremenil pesnik v lepoto.
64 Vse je bilo tiho, vse se je pogreznilo v sivo novembrsko tišino.

Po pet povedi, daljših kot tri besede in z največjim deležem črke *a* za oba vzorca najdemo v tabelah 65 in 66. Zanimivo je, da je v tabeli 65 le ena taka, ki nima črke *e*, v tabeli 66 so pa take kar 4.

Tabela 65: Vsaj 4 besede dolge povedi iz prvega vzorca z deležem črke *a* v odstotkih

- 30 Onadva sta ga vprašala:
30 Teta Agata je zajokala.
29 Ampak aroma je ta prava!
29 Čakala ga je vsa.
29 Česa pa nama tam manjka?

Tabela 66: Vsaj 4 besede dolge povedi iz drugega vzorca z deležem črke *a* v odstotkih

- 32 je vprašala vaša mama.
31 Nanca, ta pa ta!
29 Hahaha, ja, prav imaš, hahaha,«
29 Kaj pa jaz, a?
28 Ja, kaj pa sanjaš?

6. poglavje

Entropija

Če je informacija obvestilo tistemu, ki jo je prejel, da se je nekaj zgodilo, je entropija mera za nedoločenost nekega sistema sporočil. Večja kot je entropija sistema, več informacije potrebujemo, da bi ga opisali, če pa je o njem vnaprej že vse znano, je njegova entropija nič. V zvezi z besedili je pojem entropije navadno povezan s pojmom nedoločenosti naslednje, še neprebrane črke besedila. Večja kot je entropija besedila, manj možnosti ima naključni bralec, da bo po n že prebranih črkah uganil črko $n+1$, ne da bi jo videl. Entropija besedila bi bila seveda največja, če bi bila verjetnost pojavitve vseh črk enaka ($1/m$, če je m število črk) in če bi bile nepovezane – če bi bila verjetnost pojavitve vsake naslednje črke neodvisna od vseh prejšnjih. Ker pa ni tako, predvsem samoglasniki nastopajo veliko večkrat kot soglasniki in nekateri znakovni n -terčki so tudi zelo pogosti, velika večina od teoretično možnih pa se sploh nikoli ne pojavi, je entropija besedila manjša in besedilo redundantno – lahko ga rekonstruiramo nazaj, tudi če v njem manjka tri četrtine črk, seveda ne naključno izbranih (Pavešić 1997, str. 64). Spodnja meja entropije za besedila v nekem jeziku je zelo težko določljiva, se pa da, s čedalje večjimi obdelanimi vzorci, počasi zniževati oceno za njeno zgornjo mejo.

6.1 Definicije

Po očetu teorije informacije Claudu Eugenu Shannonu (Shannon 1948) je mogoče entropijo nekega sistema X izraziti kot

$$H(X) \equiv H = - \sum_{i=1}^m p_i \log_2 p_i \quad (3)$$

kjer pomeni m število vseh možnih stanj sistema X , p_i pa verjetnost pojavitve i -tega stanja. Vse verjetnosti p_i so večje ali enake 0, njihova vsota pa mora biti 1. Če vzamemo za sistem množico vseh znakovnih n -terčkov za dan m in z N označimo število vseh n -terčkov, z m število različnih in če privzamemo, da je verjetnost pojavitve i -tega n -terčka enaka

$$p_i = \frac{f_i}{N} \quad (4)$$

kjer je f_i frekvenca i -tega n -terčka, lahko izraz (3) poenostavimo (npr. Jakopin 1981) v

$$H = \log_2 N - \frac{1}{N} \sum_{i=1}^m f_i \log_2 f_i \quad (5)$$

ki je bil tudi dejansko uporabljen pri računanju. Enota entropije, če smo jo pisali kot v (3), je en bit (od angl. *binary digit*) – količina, ki opisuje sistem z dvema možnima stanjema. Entropijo H , ki opisuje množico n -terčkov nekega besedila, imenujemo tudi vezana entropija, normirana vezana entropija pa je entropija na črko besedila, ki jo označimo s

$$H_n = \frac{1}{n} H \quad (6)$$

Entropija H_n pove, koliko bitov je povprečno treba za najkrajše zakodiranje ene črke besedila, če bi ga kodirali z n -terčki. Na koncu ostane še najzanimivejša entropija F_n , to je pogojna entropija n -te črke, če prejšnjih $n-1$ že poznamo. Po Shannonu lahko hitro izpeljemo, da velja zveza

$$F_n = H_n - H_{n-1}, \quad n = 2, 3, \dots \quad (7)$$

in privzamemo, da je $F_1 = H_1$. V prejšnjih raziskavah angleškega (Brown idr. 1992a) in slovenskega jezika (Kristan idr. 1994) vezane in seveda tudi pogojne entropije za višje n niso bile izračunane; v prvem primeru zaradi zelo velikega vzorca (583 milijonov besed), ki ni dopuščal neposrednega ugotavljanja n -terčkov in njihovih frekvenc, v drugem pa si raziskovalci zaradi premajhnega vzorca (650.000 besed) niso upali iti čez $n = 8$. Raziskovalci zgornje meje entropije za angleški jezik so si pri ugotavljanju svoje ocene pomagali z modelom, kjer so naslednji trojček v besedilu napovedovali iz prejšnjih dveh. Besedilo so tako aproksimirali z markovsko verigo približno osmega reda – da je vsak naslednji znak odvisen od prejšnjih osmih. Da bi to lahko dosegli, so morali izračunati n -terčke do $n = 9$.

Entropijo slovenskih leposlovnih besedil bi bilo načelno mogoče izračunati, če bi jih imeli v celoti na razpolago, v elektronski obliki. Ker tega še ni, jo moramo oceniti s pomočjo entropij obeh vzorcev, ki skupaj merita (glej razdelek 2.1.3) približno 1% celote. Zmogljivosti računalnikov, predvsem glede pomnilnika, so se v zadnjih dveh letih toliko dvignile, da je postalo možno tako neposredno merjenje entropije obeh vzorcev, kot izvedba modela, ki pri napovedovanju pogostnosti n -terčkov preizkusnega (drugega) vzorca upošteva vse n -terčke učnega (prvega) vzorca, ki so se pojavili vsaj dvakrat. V nadaljevanju je najprej naveden izračun entropij za oba vzorca, potem pa še preizkus z modelom, ki temelji na razširjenem Huffmanovem kodiranju.

6.2 Entropija n -terčkov za oba vzorca

V tabeli 68 so navedene entropije za prvi vzorec, v tabeli 67 pa za drugi vzorec. Ker je prvi precej raznovrsten, drugi pa homogen in kompleten, jih v tem primeru ni bilo smiselno dajati skupaj. Navedene so celotna vezana entropija sistema n -terčkov H za posamezen n , normirana vezana entropija H_n (entropija na znak) in pogojna entropija H_n n -tega znaka, če je predhodnih $n-1$ že znanih. Poleg tega so navedeni še: število vseh n -terčkov, ki so se pojavili, število različnih, povprečna frekvenca n -terčkov, število enkratnic (tistih, ki se pojavijo samo enkrat) in število novih enkratnic: tistih n -terčkov s frekvenco 1 za dan n , ki ne izvirajo iz kake krajše enkratnice – take, ki se je pojavila že za nižji n . Razumljivo je nam-

Tabela 67: Entropije znakovnih n -terčkov v drugem vzorcu

n	H	H_n	F_n	Vseh n -terčkov	Različnih n -terčkov	Povprečna frekvenca	Enkratnic	Novih enkratnic
1	4,459	4,459	4,459	2.296.775	101	22.740,35	2	2
2	7,963	3,982	3,504	2.296.700	1.841	1.247,53	324	322
3	10,911	3,637	2,948	2.296.625	13.121	175,03	2.712	2.388
4	13,315	3,329	2,404	2.296.550	62.078	36,99	16.815	14.103
5	15,283	3,057	1,968	2.296.475	196.532	11,68	74.095	57.287
6	16,887	2,815	1,604	2.296.400	449.426	5,11	228.307	154.222
7	18,123	2,589	1,236	2.296.325	773.038	2,97	485.939	257.654
8	19,030	2,379	0,907	2.296.250	1.091.856	2,10	784.557	298.645
9	19,677	2,186	0,647	2.296.175	1.371.101	1,67	1.074.334	289.811
10	20,130	2,013	0,453	2.296.100	1.600.442	1,43	1.332.741	258.450
11	20,440	1,858	0,310	2.296.025	1.779.061	1,29	1.548.459	215.769
12	20,649	1,721	0,209	2.295.950	1.912.514	1,20	1.719.968	171.564
13	20,788	1,599	0,139	2.295.875	2.009.353	1,14	1.850.505	130.597
14	20,882	1,492	0,094	2.295.800	2.079.296	1,10	1.948.967	98.525
15	20,946	1,396	0,064	2.295.725	2.129.397	1,08	2.022.425	73.525
16	20,989	1,312	0,043	2.295.650	2.165.325	1,06	2.076.840	54.483
17	21,019	1,236	0,030	2.295.575	2.191.073	1,05	2.117.018	40.250
18	21,041	1,169	0,022	2.295.500	2.209.902	1,04	2.146.991	30.046
19	21,056	1,108	0,015	2.295.425	2.223.711	1,03	2.169.345	22.428
20	21,068	1,053	0,012	2.295.350	2.233.987	1,03	2.186.180	16.909
21	21,076	1,004	0,008	2.295.275	2.241.786	1,02	2.199.090	12.984
22	21,083	0,958	0,007	2.295.200	2.247.697	1,02	2.208.987	9.971
23	21,088	0,917	0,005	2.295.125	2.252.252	1,02	2.216.714	7.801
24	21,092	0,879	0,004	2.295.050	2.255.838	1,02	2.222.841	6.201

reč, da če ima recimo neki trojček v besedilu frekvenco 1, pa ni na koncu enote, do koder merimo, bodo vsi višji n -terčki, ki ga imajo na začetku (ki izvirajo iz njega), tudi imeli frekvenco 1. N -terčki so bili tvorjeni iz besedila, očiščenega vseh dodatnih simbolov (oznak) in tujih citatov ter do meja povedi (nobena poved ni bila deljena preko vrstic) zloženega skupaj v 31.000 znakov dolge vrstice, kolikor je znašala zgornja meja urejevalnika. Pri ugotavljanju entropij, v nasprotju s

prejšnjimi raziskavami (Gyergyék 1973, Kristan idr. 1994), tudi niso bile več enačene velike in male črke in vsa ločila združena v en znak. Entropije so bile izračunane do $n = 62$, v tabelah pa so navedene do $n = 24$, kjer pogojna entropija pri obeh vzorcih pade pod 0,005 bita.

Tabela 68: Entropije znakovnih n -terčkov v prvem vzorcu

n	H	H_n	F_n	Vseh n -terčkov	Različnih n -terčkov	Povprečna frekvenca	Enkratnic	Novih enkratnic
1	4,456	4,456	4,456	15.802.362	133	118.814,75	10	10
2	7,994	3,997	3,538	15.801.851	3.683	4.290,48	571	561
3	11,020	3,673	3,026	15.801.340	30.023	526,31	7.270	6.699
4	13,565	3,391	2,545	15.800.829	149.094	105,98	45.446	38.179
5	15,739	3,148	2,174	15.800.318	528.051	29,92	195.247	149.811
6	17,643	2,941	1,904	15.799.807	1.441.169	10,96	655.505	460.288
7	19,272	2,753	1,629	15.799.296	3.066.203	5,15	1.716.853	1.061.411
8	20,587	2,573	1,315	15.798.785	5.177.804	3,05	3.416.698	1.699.956
9	21,594	2,399	1,007	15.798.274	7.375.385	2,14	5.462.046	2.045.508
10	22,334	2,233	0,740	15.797.763	9.385.084	1,68	7.533.769	2.071.930
11	22,861	2,078	0,527	15.797.252	11.078.807	1,43	9.424.877	1.891.364
12	23,221	1,935	0,360	15.796.741	12.420.307	1,27	11.030.525	1.605.966
13	23,460	1,805	0,239	15.796.230	13.422.587	1,18	12.307.935	1.277.779
14	23,615	1,687	0,155	15.795.719	14.139.003	1,12	13.270.707	963.177
15	23,715	1,581	0,100	15.795.208	14.638.044	1,08	13.972.904	702.621
16	23,779	1,486	0,064	15.794.697	14.980.642	1,05	14.474.510	502.046
17	23,821	1,401	0,042	15.794.186	15.213.792	1,04	14.827.769	353.712
18	23,848	1,325	0,027	15.793.675	15.372.308	1,03	15.075.733	248.432
19	23,866	1,256	0,018	15.793.164	15.479.668	1,02	15.248.294	173.043
20	23,878	1,194	0,012	15.792.653	15.553.149	1,02	15.368.825	121.018
21	23,886	1,137	0,008	15.792.142	15.604.148	1,01	15.453.885	85.550
22	23,891	1,086	0,005	15.791.631	15.640.078	1,01	15.514.768	61.376
23	23,895	1,039	0,004	15.791.120	15.665.784	1,01	15.559.113	44.840
24	23,898	0,996	0,003	15.790.609	15.684.179	1,01	15.591.459	32.847

6.2.1 Algoritem

Algoritem, ki je bil uporabljen pri računanju podatkov v tabelah 68 in 67, je skušal kar se najbolj da varčevati s pomnilniškim prostorom in obenem, na koraku n , izkoristiti vso informacijo koraka $n-1$, da tudi čas računanja ne bi bil pretirano dolg. Računanje rezultatov v tabeli 68 je na za takratne pojme (1998) zelo sodobnem računalniku (2 x Pentium II 400 MHz, 512 MB pomnilnika, Windows NT) trajalo nekaj manj kot 4 ure. Postopek bi bil izvedljiv že tudi s 96 MB pomnilnika, česar pa avtor seveda ni vedel vnaprej. Prvi poskus, s klasičnim (8-bitnim) pesimističnim algoritmom, da lahko namreč iz enega n -terčka nastane toliko $(n+1)$ -terčkov, kolikor je minimum dveh števil:

- a) števila znakov v naboru
- b) njegove frekvence

in ki je seveda tudi rezerviral toliko prostora zanje, je namreč naletel na premajhen pomnilnik (512 MB) že pri računanju deveterčkov. Da bi bilo lažje nadaljevati, je bilo besedilo najprej prekodirano iz 16-bitnega nabora (tabela 28) v 8-bitnega, v katerem so imeli vsi znaki kode, manjše od 192. Zatem pa sta bili uporabljena dva premisleka:

1. da lahko iz n -terčka s frekvenco 1 nastane $(n+1)$ -terček s frekvenco kvečjemu 1;
2. da je pričakovano število $(n+1)$ -terčkov, ki nastanejo iz n -terčka i s frekvenco $f_i > 1$, mogoče dobro aproksimirati s kvadratnim korenem iz f_i , kadar je ta vrednost manjša od števila znakov v naboru;

Pomnilnik je bil nato, po opravljenem koraku n in pred korakom $n+1$, razdeljen na 4 dele:

- (1) del, kjer so bile shranjene tiste enkratnice, ki so bile odkrite na enem prejšnjih korakov od 1 do n ("matere" enkratnic s koraka n) in ki so se na koraku n tudi še dejansko pojavile na začetku kakega n -terčka;
- (2) del, kjer so bili shranjeni n -terčki s pričakovano frekvenco $f_{i,n+1}$ (na koraku $(n+1)$) 3 ali manj – zanje je bil rezervirano $f_{i,n+1}$ enot prostora z dodatnim, še praznim znakom;
- (3) del, predviden za $(n+1)$ -terčke, katerih pričakovana frekvenca je bila večja kot 3;
- (4) prostor za $(n+1)$ -terčke, ki bi nastali iz n -terčkov in je bila njihova dejanska frekvenca večja od predvidene (angl. *overflow space*).

Pomembno pri tem je, da sta bila dela (1) in (2) organizirana kot abecedno urejeni tabeli (abeceda kodov), zgrajeni po načelu podobnih tabel pri zbirkah črkovalnikov – vsak nov niz je imel pred seboj en znak, ki je povedal, koliko znakov v tem nizu je enakih kot v prejšnjem. To število je imelo prišteto 192, da ga je bilo mogoče odkriti. Tako je bilo torej mogoče shranjevati nize do dolžine 62 (znak s kodo 63 oz. 255 je pomenil konec zapisa) in od tod izvira potreba po znižanju kod v naboru na 191 ali manj. S tem so bili seveda shranjeni samo različni konci enkratnic in tabela, ki bi sicer lahko bila zelo velika, ohranjena v zmernih okvirih. Frekvence vseh nizov v tabelah od 1 do 3 do bile pred korakom $n+1$ postavljene na 0. Med postopkom na koraku $n+1$ je za bilo za vsak nov niz najprej preverjeno, ali je naslednik kake enkratnice iz dela 1. Če je bil, mu je bila frekvenca povečana na 1 in postopek končan. Če ni bil, je bilo potrebno preveriti, ali je zanj predviden prostor v delu 2 ali v delu 3. Če je bil predviden, je bila frekvenca spet povečana (v delu 2 do 3, v delu 3 brez omejitve) in naloga opravljena. Če je nov niz še vedno ostal, je bil dodan na konec dela 4. Iskanje po delih 1, 2 in 3 je potekalo z bisekcijo. Po pregledu celotnega besedila (ki je bilo medtem tudi v pomnilniku), je bil najprej sortiran del 4, nato pa opravljeno zlivanje delov 2, 3 in 4 (angl. *merge*) s sprotnim odmetavanjem tistih (pričakovanih) nizov, ki so imeli frekvenco 0. Na koncu je bila iz dela 1 in zlitih 2, 3 in 4 tvorjena izpisna datote-

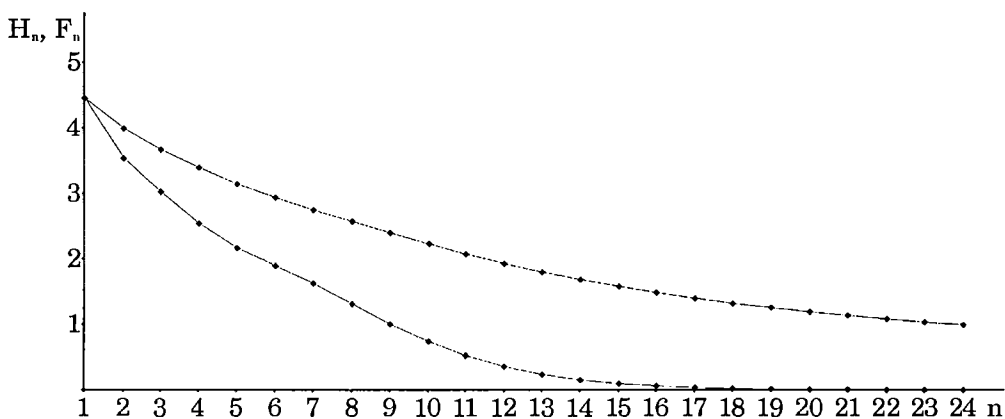
ka, ki je bila shranjena na disk (od $n = 5$ v črkovalnikovem formatu) in odstranjena iz pomnilnika, nato pa še novi deli 1, 2 in 3 za korak $n+2$.

S to metodo bi bilo mogoče s sedanjimi računalniki obdelovati tudi za velikostni red večja besedila.

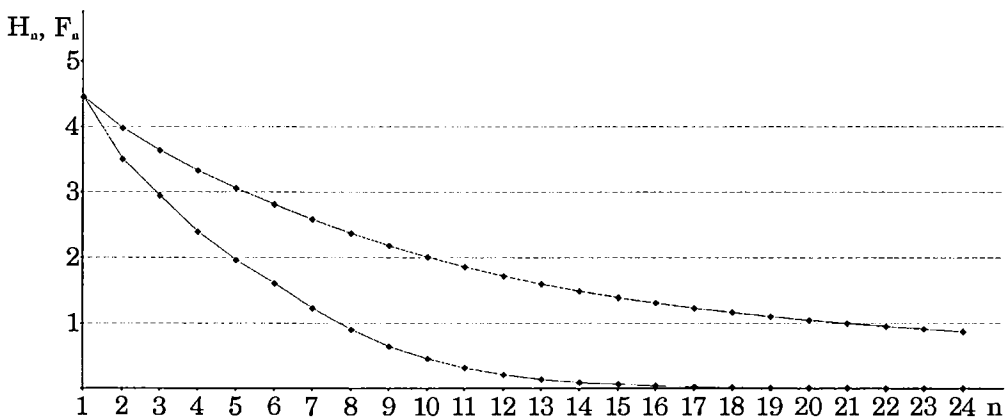
6.2.2 Rezultati

Kot je videti iz tabel 68 in 67, se entropija začne pri 4,459 v prvem oziroma 4,456 v drugem vzorcu. Ugotovitev, da ima večji vzorec večjo entropijo kot manjši (Kristan idr. 1994), seveda vseeno drži – entropija sistema n -terčkov se pri prvem vzorcu dviguje do 23,898 bita, pri drugem vzorcu pa do 21,092 bita. Še bolj zanimivo je opazovati padanje normiranih vezanih entropij (entropij na znak) in pogojnih entropij. Grafično je prikazano na sl. 18 (za prvi vzorec) in 19 (za drugega).

Slika 18: Normirana vezana entropija in pogojna entropija v prvem vzorcu

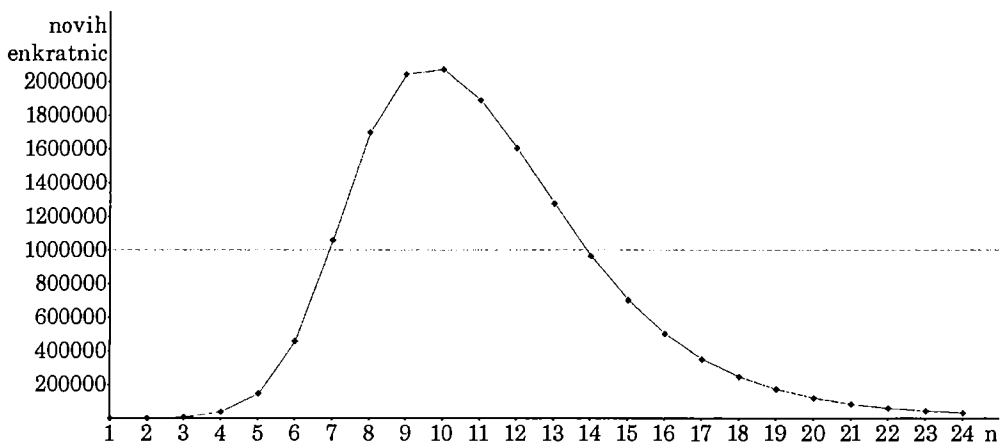


Slika 19: Normirana vezana entropija in pogojna entropija v drugem vzorcu



Entropija na črko, H_n , monotono pada v obeh vzorcih, pri prvem od 4,456 do 0,996 bita na črko, v drugem pa od 4,459 do 0,879. Pogojna entropija F_n (spodnji krivulji na obeh slikah) v drugem vzorcu tudi monotono pada do 0, le precej hitreje kot v prvem. Pri prvem vzorcu potek gibanja pogojne entropije ni povsem enakomeren – opaziti je upočasnitev padanja od $n = 5$ do $n = 7$, kjer je rahla stopnica, potem pa se krivulja, kot na smučarski skakalnici, spusti v nižino do ničle. Pojavi se vprašanje, kje potegniti črto, do kod vzorca še sledita hipotetičnim vrednostim za celoto vseh slovenskih leposlovnih besedil in od kod naprej ju zaradi njune omejenosti ne moremo več upoštevati. Odgovor na to vprašanje se po mnenju avtorja skriva v zadnjih stolpcih tabel 68 in 67. V tabeli 68 opazimo, da se število novih enkratnic (zadnji stolpec) povečuje do $n = 10$, do koder hitro raste tudi še število različnih n -terčkov (šesti stolpec); povprečna frekvenca deseterčkov je prav tako še razmeroma blizu 2 (1,68). Pri enajsterčkah vzorcu poide sapa in zmanjka mu življenjske moči – število novih enkratnic znatno pade, povprečna frekvenca se zniža na vrednost, ki je že bližje 1 kot 2 (1,43) in hitrejša pot navzdol je odprta. Gibanje novih enkratnic je upodobljeno na sliki 20.

Slika 20: Število novih enkratnic v prvem vzorcu



Iz vsega naštetega, iz navedenih podatkov obravnavanega besedilnega vzorca je mogoče torej sklepati, da zgornje meje entropije ni možno postaviti pod 2,233 bita na znak (vrednosti pri deseterčkah), ki je blizu polovice entropije H_1 (2,228 bita na znak), znane ocene za limito entropije H_∞ , ko se n bliža neskončnosti. Ocena za zgornjo mejo entropije v slovenskih leposlovnih besedilih je torej 2,233 oziroma

2,2 bita na znak

Pogojna entropija pri deseterčkah znaša 0,740 bita na znak, kar je v mejah, dobljenih pri ocenjevanju pogojne entropije s poskusi za angleški jezik (0,6 – 1,3 bita

na znak, Shannon 1951). Ocenjena vrednost, 2,2 bita na znak, je (0,6 – 1,3 bita na znak, Shannon 1951). Ocenjena vrednost, 2,2 bita na znak, je za desetino višja od vrednosti, ki so jo izmerili raziskovalci IBM-ovih laboratorijev na dvestokrat večjem vzorcu (Brown idr. 1992a), 2,0 bita/znak. Osnovna dva razloga sta večja prebignost slovenskega jezika in pa to, da vzorca vsebinsko nista povsem primerljiva. V ameriškem vzorcu leposlovja praktično ni bilo, le roman Arthurja Conana Doylea *Prigode Sherlocka Holmesa*, večina gradiva pa je bila poslovna korespondenca (angl. *office correspondence*) velikih multinacionalnih družb. Pričakovati bi bilo, da je entropija takega gradiva nižja od entropije leposlovnih besedil.

6.3 Ocena entropije z modelom

Ocena zgornje meje entropije, dobljena v prejšnjem razdelku iz entropije n -terčkov prvega vzorca, bi bila seveda prepričljivejša, če bi jo bilo mogoče podpreti še z jezikovnim modelom. V ta namen smatramo prvi vzorec za učni vzorec, podatki katerega bodo služili za konstrukcijo modela, s katerim bomo potem besedilo preizkusnega (drugega) vzorca lahko komprimirali, kodirali tako, da bo skupna dolžina kod najkrajša in s tem dobili še eno oceno za entropijo.

Če je teorija informacij pogosto definirana kot proučevanje učinkovitega kodiranja, kar se tiče hitrosti in zanesljivosti prenosa (Ingels 1971), je komprimiranje podatkov veja teorije informacij, ki se ukvarja z minimiziranjem količine podatkov, ki jih je treba prenesti ali shraniti. Težišče reševanja problemov pri komprimiranju podatkov se je v zadnjem desetletju sicer preneslo na metode za učinkovito shranjevanje slik, kakršna sta standard JPEG (npr. Pennebaker in Mitchell 1993, Wallace 1991) in fraktalno komprimiranje (npr. Fisher 1995, Jacquin 1993), vendar je ostalo komprimiranje besedil v sodobnih raziskovalnih prizadevanjih slej ko prej aktualno.

V nadaljevanju je razložen kodirni algoritem, ki je bil nato uporabljen pri konstruiranju modela za komprimiranje besedil.

6.3.1 Huffmanovo kodiranje

Način kodiranja z minimalno redundanco (da se skupno število bitov kod simbolov kar najmanj razlikuje od entropije simbolov), če se število simbolov na izvoru približuje neskončnosti, je Shannon–Fanojevo kodiranje (Shannon in Weaver 1949, Fano 1949), kodiranje z zajamčeno minimalno redundanco tudi že brez zahteve po neskončnem številu simbolov pa je Huffmanovo kodiranje (Huffman 1952, Sayood 1997: tretje poglavje, program v Press 1992: razdelek 20.4). Pri obeh načinih za določitev kod, prirejenih vhodnim simbolom, pri besedilih so to navad-

no znaki, lahko pa tudi nizi znakov, dobimo kot rezultat binarne kode (kode iz binarnih števk), najprej pa potrebujemo tabelo simbolov z njihovimi pogostnostmi, urejeno po nenaraščajočih pogostnostih. Pri Shannon–Fanojevi metodi na vsakem koraku seznam razdelimo v dve skupini, tako da sta vsoti pogostnosti simbolov v obeh skupinah kar najbolj enaki. Vsem simbolom zgornje skupine potem dodelimo binarno števk 0 kot prvo števk njihovega koda, vsem simbolom spodnje skupine pa 1. Vsako skupino nato po istem načelu razdelimo naprej in njenim simbolom dodelimo dodatno binarno števk. Postopek ponavljamo, dokler v vsaki skupini ne ostane samo po en simbol. V spodnji tabeli so prikazane ustrezne kode za znake povedi *Gori_na_gori_gori*.

Tabela 69: Znaki povedi *Gori_na_gori_gori*. s frekvenca in Shannon–Fanojevimi kodami

_	3	00	a	1	1100
i	3	010	G	1	1101
o	3	011	n	1	1110
r	3	100	.	1	1111
g	2	101			

Poved ima 18 znakov, torej bi morali prvi skupini imeti skupno frekvenco kar najbližje $18/2$ ali 9. V prvi zgornji skupini se torej znajdejo znaki *_*, *i* in *o*, ki imajo skupno frekvenco ravno 9, v spodnji pa vsi ostali. Prvi dobijo začetno števk 0, drugi 1. Ponovna delitev skupine *_*, *i* in *o* se ustavi pri *i*, ker je $9/2 = 4.5$ in je $3+3$ ravno tako daleč od 4.5 kot 3. *_* dobi torej še eno 0, *i* in *o* pa 1. Deljenje skupine *i* in *o* je trivialno, ker sta podskupini samo dve in imata enako frekvenco – *i* dobi še eno ničlo, *o* pa 1. Skupine s simbolom *_* ni treba deliti še naprej, ker ima samo en element. Prvo spodnjo skupino *r*, *g*, *a*, *G*, *n* in *.* delimo analogno. Skupno število bitov za kodiranje povedi *Gori_na_gori_gori* je tako $55 (3 \times 2 + 3 \times 3 + 3 \times 3 + 3 \times 3 + 2 \times 3 + 1 \times 4 + 1 \times 4 + 1 \times 4 + 1 \times 4)$ ali 3,06 bita na znak, kar je malenkost slabše od entropije sistema znakov iz tabele 69, ki znaša 3,00 bita na znak (po zvezi (3) iz razdelka 6.1).

Algoritem Huffmanovega kodiranja lahko na kratko opišemo s štirimi koraki:

1. Razvrstimo *n* izvornih simbolov padajoče po pogostnostih (normiranih med 0 in 1)
2. Združimo zadnja dva simbola v nov simbol, katerega pogostnost je enaka vsoti obeh in ga uvrstimo na ustrezno mesto v seznam, tako da le–ta obdrži svojo urejenost. Opazovana simbola izločimo iz seznama.
3. Korak 2 ponavljamo tako dolgo (*n*–1 krat), dokler ne dobimo samo enega simbola (celote) s pogostnostjo 1. Imenujemo ga koren, nastalo strukturo pa Huffmanovo drevo. Izvorni simboli so njegovi listi.
4. Za vsak izvorni simbol sledimo pot od korena do listov, s tem da na vsakem razvejišču zapišemo 0 za prehod na levo in 1 za prehod na desno. Dobljene števk zložimo skupaj in tako dobimo Huffmanovo kodo za vsak izvorni simbol.

V tabeli 70 je upodobljeno Huffmanovo drevo za simbole povedi *Pravemu dedcu solze ne tečejo ven, temveč mu tečejo noter.* iz drugega vzorca. Poved spada med lipogramске povedi (glej razdelek 5.3.3), saj nima niti ene črke *a*. Vsebuje 58 znakov, od tega 20 različnih. V tabeli so koraki algoritma ponazorjeni z zaporednimi stolpci. Za vsakim simbolom, izvornim in sestavljenim, je za dvignjeno piko navedena njegova frekvenca. Trikotnik v tabeli je razdeljen na dva dela, da ga je bilo mogoče spraviti na list papirja.

Tabela 70: Primer Huffmanovega drevesa z znaki povedi iz drugega vzorca

e·12	e·12	e·12	e·12	e·12	e·12	e·12	e·12	e·12	e·12	e·12	e·12	e·12	e·12
_·9	_·9	_·9	_·9	_·9	_·9	_·9	_·9	_·9	_·9	_·9	_·9	_·9	_·9
o·4	o·4	o·4	o·4	o·4	o·4	o·4	o·4	vd·5	nu·6	nu·6	„szjr·8	„szjr·8	vdo·9
t·4	t·4	t·4	t·4	t·4	t·4	t·4	t·4	o·4	vd·5	čm·6	nu·6	tlPac·8	„szjr·8
č·3	č·3	č·3	č·3	č·3	lPac·4	lPac·4	lPac·4	t·4	o·4	vd·5	čm·6	nu·6	tlPac·8
m·3	m·3	m·3	m·3	m·3	č·3	„sz·4	„sz·4	lPac·4	t·4	o·4	vd·5	čm·6	nu·6
n·3	n·3	n·3	n·3	n·3	m·3	č·3	jr·4	„sz·4	lPac·4	t·4	o·4	vd·5	čm·6
u·3	u·3	u·3	u·3	u·3	n·3	m·3	č·3	jr·4	„sz·4	lPac·4	t·4	o·4	
v·3	v·3	v·3	v·3	v·3	u·3	n·3	m·3	č·3	jr·4	„sz·4	lPac·4		
d·2	d·2	d·2	d·2	d·2	v·3	u·3	n·3	m·3	č·3	jr·4			
j·2	j·2	j·2	j·2	j·2	d·2	v·3	u·3	n·3	m·3				
r·2	r·2	r·2	r·2	r·2	j·2	d·2	v·3	u·3					
a·1	„·2	„·2	„·2	„·2	r·2	j·2	d·2						
c·1	a·1	sz·2	sz·2	sz·2	„·2	r·2							
l·1	c·1	a·1	lP·2	lP·2	sz·2								
P·1	l·1	c·1	a·1	ac·2									
s·1	P·1	l·1	c·1										
z·1	s·1	P·1											
„·1	z·1												
·1													
e·12	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8
nučm·12	e·12	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8	tlPac·8
_·9	nučm·12	e·12	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8
vdo·9	_·9	nučm·12	e·12	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8	tlPac·8
„szjr·8	vdo·9	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8	tlPac·8
tlPac·8	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8	tlPac·8	„szjr·8

Na prvem koraku sta se v nov simbol „ s frekvenco 2 združila simbola , in . in se pomaknila 6 mest navzgor, za zadnji simbol s frekvenco 2, za r. Na drugem koraku sta bila to l in P, na tretjem a in c, na četrtem pa je nov simbol, sz., nastal že iz dveh sestavljenih. Na devetnajstem koraku smo prišli do korena, sestavljene-ga iz vseh izvornih simbolov in s skupno frekvenco 58. Huffmanovo kodo najpo-gostejšega simbola e zdaj dobimo tako, da mu sledimo po vseh razvejiščih, kjer še ni bil samostojen. Najprej, pri razcepu simbola _vdo„szjr·8 in enučm v _vdo„szjr·8 in enučm je bil v drugem delu in torej dobi spredaj 1. Pri nadalj-nji cepitvi skupine enučm v e in nučm pa je bil v prvem delu, dobil še 0 in je to že tudi vse, ker je postal samostojen. Njegova Huffmanova koda je torej 10.

V tabeli 71 je naveden seznam vseh znakov te povedi s frekvencami, Shannon–Fanojevimi in Huffmanovimi kodami. Shannon–Fanojeve kode hitro prepoznamo po strogem naraščanju vodilnih enojk proti dnu tabele. Skupno število bitov kod za vse znake povedi je 225 pri Shannon–Fanojevem in 224 pri Huffmanovem kodiranju, kar je 3,88 oziroma 3,86 bita na črko. Entropija sistema znakov povedi znaša, spet po izrazu (3) iz razdelka 6.3.1, 222,49 bitov ali 3,84 bita na znak.

Tabela 71: Znaki iz tabele 70 s frekvencami, Shannon–Fanojevimi in Huffmanovimi kodami

e	12	00	10
–	9	010	000
o	4	0110	0011
t	4	0111	0110
č	3	1000	1110
m	3	1001	1111
n	3	1010	1100
u	3	10110	1101
v	3	10111	00100
d	2	11000	00101
j	2	11001	01010
r	2	11010	01011
a	1	11011	011110
c	1	11100	011111
l	1	111010	011100
P	1	111011	011101
s	1	111100	010010
z	1	111101	010011
,	1	111110	010000
.	1	111111	010001

Huffmanovo kodiranje besedil na nivoju znakov prinese sicer že znaten prihranek v primerjavi s kodiranjem znakov s fiksno dolžino osmih bitov, ki je zdaj običajno pri shranjevanju in prenašanju (npr. standard ASCII in izpeljanke), skoraj polovico, a vendar manj kot bi bilo možno še z upoštevanjem povezanosti črk v besedilu in upoštevanjem krajevnih posebnosti v besedilu. Klasični Huffmanov algoritem, opisan v tem razdelku, zato navadno označujejo kot statično Huffmanovo kodiranje, ker so kode posameznih simbolov med vsem kodirnim postopkom nespremenjene. Ta način kodiranja tudi zahteva dva prehoda preko izvornih simbolov – prvega za izračun njihovih pogostnosti in drugega za kodiranje v ožjem smislu. Dinamično Huffmanovo kodiranje je postopek, pri katerem se Huffmanovo drevo gradi in spreminja sproti, brez predhodnega ugotavljanja pogostnosti simbolov, z vsakim novim prebranim simbolom glede na njegovo dotedanjo frekvenco (Faller 1973, Gallager 1978, Knuth 1985). Po teh treh raziskovalcih se izpopolnjen postopek, uporabljen v nekaterih komprimirnih programih v okolju UNIX, imenuje tudi algoritem FGK. Kasneje je k zmanjšanju dela pri sprotnem uravnoteževanju Huffmanovega drevesa in k minimiziranju tudi največje dolžine kode za posamezen simbol prispeval še J. Vitter (Vitter 1987).

Tabela 72: Huffmanov kod za črke, števke in ločila slovenskih leposlovnih besedil

A	0,000928550	1001010100	o	0,069218303	0101
B	0,000807521	1001011010	ó	0,000037784	000011110100100
C	0,000184813	100101111111	ò	0,000000649	00001010100011100110
Č	0,000346868	1011111011	ô	0,000006216	10010100000111101
D	0,000659466	00001010010	D	0,024119859	10110
E	0,000211732	100101010111	q	0,000015297	1001010000001011
F	0,000193245	100101111101	r	0,038129672	00111
G	0,000433410	10010101010	s	0,037610100	01000
H	0,000273084	000011110101	š	0,007468506	1000001
I	0,000826711	1001010110	t	0,031742312	01111
J	0,000599303	00001010101	u	0,014322898	100001
K	0,001491528	101111000	ú	0,000006865	10010100000111100
L	0,000532762	00001111100	ù	0,000006216	000010101000111000
M	0,001159038	0000111100	v	0,027802165	10001
N	0,001457798	101111001	w	0,000036487	000011110100101
O	0,000799846	1001011011	x	0,000033946	100101000000100
P	0,001884829	100101001	y	0,000172704	101111110101
Q	0,000017297	0000101010001111	z	0,015412152	011101
R	0,000482924	10010100010	ž	0,004836480	00001110
S	0,001338932	0000101000	0	0,000141515	1001010000000
Š	0,000211516	100101111100	1	0,000349030	10111111001
T	0,001623692	100101100	2	0,000232975	100101000110
U	0,000157137	0000101010010	3	0,000164650	0000101010000
V	0,001209525	0000101011	4	0,000124326	1001010000010
W	0,000062109	10010100000011	5	0,000106866	1001010101100
X	0,000029244	100101000001101	6	0,000095460	1001011111100
Y	0,000012108	1001010000011111	7	0,000084812	00001010100010
Z	0,000795197	1001011110	8	0,000106433	1001010101101
Ž	0,000145893	0000111101000	9	0,000109461	1001010001111
a	0,079735811	0010		0,187343573	11
á	0,000067460	00001111010011	!	0,001592124	100101110
à	0,000008324	10010100000010101	'	0,000048649	10010111111010
â	0,000007243	10010100000111001	(0,000087731	101111101001
b	0,014135058	100100)	0,000087839	1011111101000
c	0,004918697	00001011	*	0,000502113	10010100001
č	0,011080487	0000100	,	0,016751138	011100
ć	0,000004649	0000101010001110010	-	0,001385527	101111111
d	0,025467980	10100	-	0,000146488	0000101010011
e	0,082316215	0001	.	0,009900151	0000110
é	0,000113244	1001010001110	/	0,000014162	1001010000011101
è	0,000029784	100101000001100	:	0,000811576	1001010111
ê	0,000007784	10010100000111000	;	0,000529843	00001111101
f	0,000653844	00001010011	?	0,001022929	0000111111
g	0,012188660	101110	...	0,000546600	00001111011
h	0,007800023	1000000	»	0,002906677	10111101
i	0,068865705	0110	«	0,002906677	10111110
í	0,000038000	000010101000110	‘	0,000181786	101111110000
ì	0,000000486	00001010100011100111	’	0,000181786	101111110001
j	0,035433430	01001	,	0,000008595	00001010100011101
k	0,027060212	10011	‘	0,000008595	10010100000010100
l	0,040057475	00110	,,	0,000021892	100101111110110
m	0,024314618	10101	“	0,000021892	100101111110111
n	0,047316249	00000			

Dinamično Huffmanovo kodiranje sicer navadno ne prinese prihrankov pri prostoru (Vitter 1987), pač pa, zaradi le enega prehoda preko podatkov namesto dveh pri statičnem, pri hitrosti komprimiranja.

Poleg dinamičnih algoritmov je bilo Huffmanovo kodiranje deležno še raziskav v smeri razširitev (Bshouty in Falk 1992) in novih teoretičnih študij v zadnjem času (Calude in Tomescu 1997). Huffmanovo kodiranje je služilo za osnovo s strani agencije NASA leta 1995 nagrajenemu algoritmu za komprimiranje pri prenosu slik z vesoljskih postaj na zemljo (Rice idr. 1993, Venbrux idr. 1992). Poleg tega je bil ta način kodiranja uporabljen še pri algoritmih, ki niso povezani s komprimiranjem, npr. pri konstruiranju optimalnih iskalnih dreves (Hu in Tucker 1971, Itai 1976) zlivanju seznamov (Brent in Kung 1978) in generiranju optimalnih evaluacijskih dreves pri prevajanju izrazov (Parker 1980).

V tabeli 72 na prejšnji strani so navedene pogostnosti in Huffmanovi kodi za 105 znakov v besedilih obeh vzorcev, ki so dosegli pogostnost vsaj 0,0000005. Najkrajšo kodo, 2 bita, ima presledek, kode samoglasnikov *a*, *e*, *i* in *o* so dolge po 4 bite, najdaljša pa je koda črke *ì*, 20 bitov.

6.3.2 Izbira postopka

Poleg Huffinanovega kodiranja in njegovih adaptacij so še drugi kodirni postopki, pri katerih se vsota dolžin kod za izvirne simbole asimptotično približuje entropiji izvirnega sistema. Sem sodijo aritmetično kodiranje (Abramson 1963, Rubin 1979, Witten idr. 1987), algoritem LZW (Ziv in Lempel 1977, Welch 1984), algoritem BSTW (Bentley idr. 1986) in algoritem BWT (Burrows–Wheelerjeva transformacija: Burrows in Wheeler 1994, Fenwick 1996). Vsak ima svoje prednosti in pomanjkljivosti. Ker je osnovni namen tega dela predvsem iskanje mehanizma, s katerim bi bilo mogoče identificirati izvirne simbole kot tiste nize znakov, katerih pogostnosti se najbolj splača izkoristiti pri konstrukciji modela z minimalno entropijo, je v nadaljnjem kot konkretni algoritem preverjanja še naprej uporabljeno Huffmanovo kodiranje.

Pri načrtovanju konkretnih postopkov, namenjenih posebej komprimiranju besedil, so raziskovalci, kot sicer predvsem za angleški jezik, doslej ubirali različne poti. Predvsem so skušali izpostaviti pogoste elemente v besedilu. B. Hahn (Hahn 1974) je predlagal kodiranje besed v datoteki z besedilom, pri katerem so njihove kode kazalci v slovar besed. Kombiniran postopek, pri katerem bi bil rezultat kodiranja sestavljen iz kazalcev v slovar pogostih besed in pogostih besednih končnic, ostalo pa kodirano na klasičen način, je predlagal R. Tropper (Tropper 1982). Precej pozornosti je bilo posvečeno kodiranju pogostih znakovnih dvojčkov v besedilu (Snyderman in Hunt 1970, Cortesi 1982, Knuth 1985). R. A. Wagner (Wagner 1973) je poskusil najpogostejše besedne skupine (frazе), ki zaslužijo posebno obravnavo, določiti ročno, druga raziskava (Fraenkel idr. 1983) pa je obravna-

vala strojno identifikacijo pogostih predpon in pripon. Doslej že že večkrat ome-njena raziskava (Brown idr. 1992a, Brown idr. 1992b), ki je temeljila na statistiki deveterčkov, dobljeni iz 583 milijonov besed dolgega vzorca, je postregla z doslej najnižjo oceno entropije za angleška besedila, 2,0 bita na znak. Pogostnosti deve-terčkov, dobljene iz vzorca, so bile potem uporabljene na 2 milijona besed dolgem preizkusnem vzorcu, znanem Brownovem korpusu z besedili iz leta 1961 (Francis in Kučera 1982). Na kratko bi lahko rekli, da je bila entropija Brownovega korpu-sa ocenjena z očmi in védenjem nekoga, ki je prebral učni vzorec in si ga zapom-nil. *Jezik je v smislu svoje funkcije določen kot komunikacijski sistem, ki ekološko odseva lastnosti svojih uporabnikov - človeških bitij* (Tobin 1997, str. 145) in je zato sistem, ki mora z najmanj energije posredovati čimveč sporočila. Ta zahteva ima seveda svoj antipod – dodatno zahtevo po zanesljivosti prenosa, ki zahteva redundantnost, da smo sposobni tako ali drugače poškodovano oz. moteno sporoči-lo še vedno prav prebrati ali razumeti. Kako velik je vzorec, ki ga pozna in se ga spomni tipičen srednje izobražen pripadnik nekega jezikovnega področja, je to nekaj milijonov besed ali manj, ali pa več, nekaj deset milijonov ali celo nekaj sto milijonov besed kot v primeru vzorca IBM–ovih raziskovalcev, je težko oceniti.

Avtor se je odločil, da bo pri oceni zgornje meje entropije z jezikovnim modelom prvi vzorec, 2.700.000 besed, uporabil kot učni vzorec, drugega, 400.000 besed, pa kot preizkusni. Oba sta seveda veliko manjša od ameriškega zgleda, prvi dvesto-krat, drugi pa petkrat. Zato je seveda pričakovati manjšo učinkovitost podatkov učnega vzorca na preizkusnem in posledično šibkejšo (višjo) oceno za zgornjo mejo entropije.

6.3.3 Model z razrezom na enako dolge n-terčke

Najprej je bila uporabljena najpreprostejša možnost, razrez prvega in drugega vzorca na enako dolge n-terčke in določitev ustreznih Huffmanovih kod. Pri tem sta bili besedili obeh vzorcev smatrani za eno samo vrstico in zaporedno raz-rezani, zadnji n-terček, če je bil prekratek, pa odvržen. V nasprotju z n-terčki v razdelku 6.2 (tabeli 68 in 67), kjer so bili generirani vsi možni, gre tu le za zapo-redne n-terčke. Primer je npr. razrez povedi *To_je_ vse.* na peterčke – pri razre-zu, ki sledi, dobimo samo 2: *To_je* in *_vse.*, v razrezu iz razdelka 6.2 pa 6: *To_je*, *o_je_*, *_je_v*, *je_vs*, *e_vse* in *_vse.* V tabelah 73 in 74 so navedeni: n, število vseh n-terčkov po razrezu, število različnih n-terčkov, skupna dolžina Huffmanovih kod za vse n-terčke (v bitih), dolžina Huffmanovih kod na znak besedila, dolži-na najkrajše in najdaljše Huffmanove kode. Tabeli segata od $n = 1$ do 20, kjer pade dolžina Huffmanovih kod na znak besedila pod 1 bit. Opazimo nižje vredno-sti na znak besedila kot v tabelah 68 in 67, kar je razumljivo zaradi manjšega števila n-terčkov; povprečna frekvenca n-terčka pade pri prvem vzorcu pod 2 že pri osmerčkih, pri drugem vzorcu pa pri sedmerčkih. Vidimo tudi, da se z rasto-čim n zmanjšuje razlika med najkrajšo in najdaljšo Huffmanovo kodo – v tabeli

Tabela 73: Dolžine Huffmanovih kod pri razrezu prvega vzorca na n-terčke

n	vseh n-terčkov	različnih n-terčkov	skupno bitov	bitov na črko	najkrajša Huff. koda	najdaljša Huff. koda
1	15.802.872	133	71.160.133	4.5030	3	24
2	7.901.436	3.280	63.368.664	4.0099	5	23
3	5.267.624	22.677	58.191.686	3.6823	7	22
4	3.950.718	96.318	53.621.781	3.3932	7	22
5	3.160.574	280.525	49.537.784	3.1347	9	22
6	2.633.812	589.240	45.756.764	2.8955	10	21
7	2.257.553	919.393	41.992.115	2.6572	10	21
8	1.975.359	1.156.054	38.323.126	2.4251	11	21
9	1.755.874	1.275.809	34.938.247	2.2109	11	21
10	1.580.287	1.310.359	31.859.619	2.0161	13	21
11	1.436.624	1.289.850	29.104.778	1.8417	14	20
12	1.316.906	1.238.465	26.673.762	1.6879	15	21
13	1.215.605	1.173.826	24.544.717	1.5532	15	21
14	1.128.776	1.106.523	22.684.277	1.4355	16	21
15	1.053.524	1.041.361	21.053.062	1.3322	17	20
16	987.679	980.920	19.677.795	1.2452	16	19
17	929.580	925.682	18.464.196	1.1684	16	20
18	877.937	875.628	17.383.179	1.1000	16	20
19	831.730	830.217	16.414.586	1.0387	17	19
20	790.143	789.239	15.542.545	0.9835	17	20

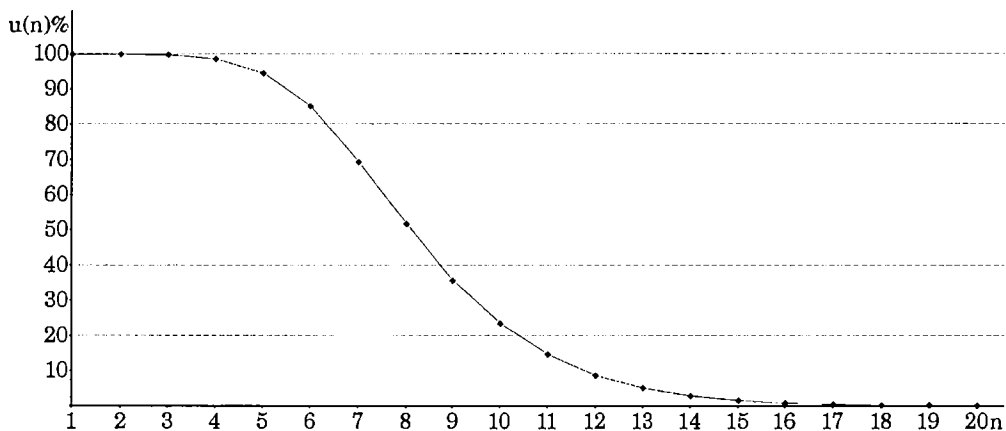
Tabela 74: Dolžine Huffmanovih kod pri razrezu drugega vzorca na n-terčke

n	vseh n-terčkov	različnih n-terčkov	skupno bitov	bitov na črko	najkrajša Huff. koda	najdaljša Huff. koda
1	2.296.849	101	10.345.994	4.5044	2	21
2	1.148.424	1.611	9.177.023	3.9955	5	20
3	765.456	10.292	8.366.728	3.6427	6	20
4	574.212	41.174	7.626.828	3.3206	7	19
5	459.369	98.603	6.919.121	3.0124	9	19
6	382.808	159.760	6.242.933	2.7180	9	19
7	328.121	197.771	5.586.563	2.4323	9	19
8	287.106	213.629	4.991.709	2.1733	11	18
9	255.205	214.456	4.473.365	1.9476	11	18
10	229.684	207.291	4.043.712	1.7606	11	18
11	208.804	196.467	3.674.764	1.5999	12	18
12	191.404	184.618	3.358.406	1.4622	13	18
13	176.680	172.688	3.085.576	1.3434	13	18
14	164.060	161.665	2.849.628	1.2407	14	18
15	153.123	151.710	2.644.098	1.1512	14	17
16	143.553	142.613	2.463.311	1.0725	15	17
17	135.108	134.468	2.303.519	1.0029	14	18
18	127.602	127.126	2.164.754	0.9425	14	17
19	120.886	120.204	2.043.466	0.8897	14	17
20	114.842	114.594	1.935.556	0.8427	14	16

73 imajo vsi dvajseterčki razen najpogostejših 1153 dolžino 20 bitov, v tabeli 74 pa vsi razen prvih 22 dolžino 16 bitov. Ob nizkih vrednostih na znak pri višjih n

je pri konstrukciji praktičnega komprimirnega mehanizma seveda treba upoštevati še strošek, ki ga potegne za seboj slovar različnih n -terčkov s Huffmanovimi kodami. Tudi če bi bil abecedno urejen in shranjen v črkovalnikovem načinu (kjer je pred vsakim naslednjim n -terčkom najprej shranjen podatek o številu začetnih črk, ki se ujema s predhodnikom, potem pa le rep n -terčka, ki se razlikuje od prejšnjega, za njim pa Huff. koda), nikakor ni majhen. Drugače je seveda, če ta strošek vzamemo v zakup, na račun modela – recimo, da je tak slovar že vgrajen v sprejemnik in oddajnik, ki sta specializirana za besedila v slovenskem jeziku. V tem primeru potrebujemo, pred odločitvijo o morebitnem preizkusu modela, dobljenega s podatki o n -terčkih prvega vzorca, na drugem vzorcu, še podatek o pokrivanju, kolikšen delež n -terčkov drugega vzorca se ujema z n -terčki prvega. Ta podatek je razviden iz slike 21.

Slika 21: Pokritje n -terčkov drugega vzorca z n -terčki prvega vzorca



Žal je prekrivanje zelo dobro samo za zelo nizke n . Pri šesterčkih je le 85%, potem pa je padanje še hitrejše in že pri dvanajsterčkih je vrednost pod 10% (8.6%). Ker je število bitov na črko pri peterčkih v prvem vzorcu še precej nad 3, preizkus v tem primeru opustimo.

6.3.4 Model z razrezom na besede

Pri tem se odločimo, da besedilo obeh vzorcev razrežemo na robovih besed – ločilno črto potegnemo pred vsako besedo in za njo. Primer razreza je odlomek iz drugega vzorca – iz romana *Pomladni dan*:

... |pričil_lizginilal_>|Kaj_llel_rogoviliml_lsl_ltol_langleščino!<_lsem_lsel_lozmerjall_>|Tennysona
|_recitirašl_lmel_ljel_lzaničljivol_lopomnilal_lpametl_>|Saj_liveml_<_lsem_lzamrmrall_...

Tako dobimo po prelomu seznam enot, ki so ali besede ali pa ločilni nizi. Najpogostejša enota je že na prvi pogled najpogostejši ločilni niz, presledek. V tabeli 75 je navedeno število enot in dolžine Huffmanovih kod za oba vzorca:

Tabela 75: Število enot in dolžine Huffmanovih kod pri razrezu na besede

	vseh znakov	besednih enot	različnih enot	skupaj bitov	bitov na znak
Prvi vzorec	15.802.362	5.446.180	176.792	41.323.662	2.62
Drugi vzorec	2.296.775	818.362	40.747	5.905.093	2.57

Vrednostim v zadnjem stolpcu manjka sicer še kar nekaj do zelene meje, vendar so obetavne. V tabeli 76 so navedene najpogostejše enote iz besednega razreza za oba vzorca. Do desetega mesta so na vrhu isti nizi; opazimo tudi, da ima samostojni presledek Huffmanovo kodo dolgo le en bit, da pa so le eno ali dve črki dolgi razmeroma slabo kodirani – z vsaj tremi biti na znak.

Tabela 76: Najpogostejše besedne enote, njihove frekvence in Huffmanove kode

prvi vzorec			drugi vzorec		
enota	frekvenca	Huffmanova koda	enota	frekvenca	Huffmanova koda
–	2.214.919	1	–	329.333	1
,–	258.876	00010	,–	33.989	00100
je	150.558	01110	je	25.768	01011
..	119.505	001000	..	20.586	000001
in	87.126	010100	in	16.555	001010
se	69.576	0000001	se	13.306	010101
v	51.034	0011010	v	7.168	0100011
da	43.590	0101010	da	5.159	00000011
na	34.025	00000110	na	4.674	00010001
so	32.915	00001010	so	4.242	00011100
ne	24.867	00111001	pa	4.224	00011101
pa	24.114	01000001	._→	3.941	00111011
ki	23.238	01000110	bi	3.219	01100001
bi	21.862	01001111	ne	3.178	01100100
za	19.092	01101001	ga	3.035	01101111
z	18.912	01101100	z	3.017	01110000
sem	18.300	01111011	ki	2.993	01110010
ni	18.048	01111101	sem	2.793	01111011
._→	17.766	000000000	po	2.771	01111101
ga	15.524	000111011	ni	2.532	000011000

Poskusimo zdaj uporabiti enote prvega vzorca za kodiranje drugega. Izkaže se, da 20.950 besednih enot nima svojega ustreznika v prvem vzorcu. Med njimi je 9.416 različnih in od tega kar 3.264 (35%) takih z veliko začetnico, ki so predvsem imena. Če med nize prvega vzorca uvedemo še poseben niz, ki ga bomo uporabili za označitev, da sledi neznana besedna enota in mu damo pogostnost 20.950/818.362, ugotovimo, da je dolžina njegove Huffmanove kode 6 bitov. Izkaže se tudi, da bi za Huffmanove kode znanih besednih enot porabili skupaj 5.814.874 bitov, da je v prvemu vzorcu neznanih besednih enotah skupaj 154.870 znakov in pa, da bi za kodiranje dolžin teh 20.950 enot s Huffmanovimi kodami porabili skupaj 70.916 bitov (3,39 bita na dolžino). Če neznane znake kodiramo z navadno Huffmanovo kodo (vsak znak posebej), ki zasede povprečno 4,50 bita,

porabimo za kodiranje enot drugega vzorca s kodami enot prvega z omenjenim dopolnilom za neznane enote, skupaj 6.707.685 bitov ($5.814.874 + 20.950 \cdot 6 + 70.196 + 154.870 \cdot 4,5$). Če to delimo s številom znakov v drugem vzorcu, 2.296.775, dobimo

2,92 bita na znak

našo prvo z modelom besed prvega vzorca na drugem vzorcu dobljeno oceno za entropijo. Poglejmo še, kdo so glavni krivci, da ocena ni nižja. Najpogostejših 16 besed (s frekvencami) iz drugega vzorca, ki se v prvem niso nikoli pojavile, je: *Majcen* (270), *Drejc* (189), *Temnikar* (185), *Tantadruj* (172), *Venc* (162), *Venček* (151), *Žef* (150), *Kadetka* (125), *Tildica* (116), *Nanca* (106), *Hotejec* (105), *Najdù* (104), *Temnikarica* (102), *Modrijan* (97), *Okrogličar* (83) in *Nančika* (73). Skoraj same osebe, ki so tipične za Cirila Kosmača in jih v drugih literarnih delih ne najdemo.

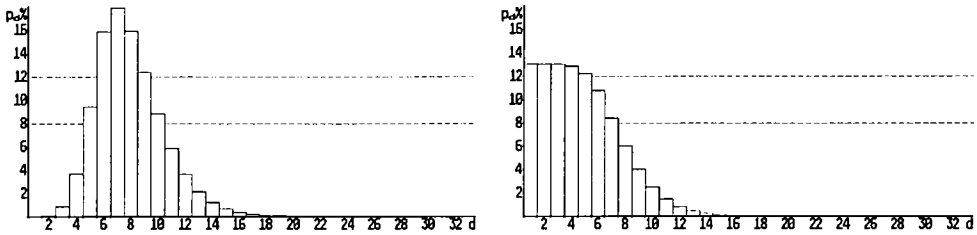
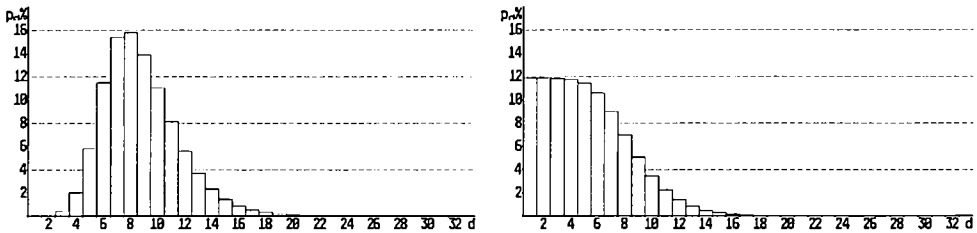
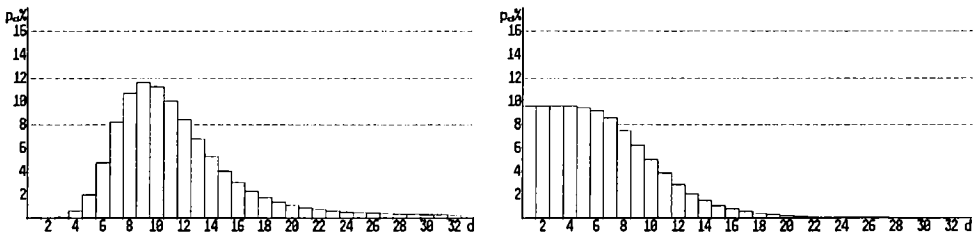
6.3.5 Model z optimalnim razrezom na n-terčke

Glede na vse v prejšnjih razdelkih omenjene in obravnavane pristope ostane na koncu še možnost oblikovanja modela, v katerem bi dali priložnost vsem nizom znakov prvega vzorca, ki dosežejo določeno frekvenco. S tem bi se izognili tako težavam s slabim pokrivanjem (slika 21 v razdelku 6.3.3) enako dolgih n-terčkov: algoritem bi na vsakem koraku uporabil najdaljši možen n-terček kot tudi šibkostim, ki jih prinaša razrez na besede: veliko kratkih besed v vrha frekvenčnega seznama, npr. *je*, *se*, *bil* se z znatno frekvenco pojavlja tudi skupaj. Pot do takega seznama, ki utegne biti zelo obsežen, pa je prav tako znana – seznam vseh n-terčkov s frekvencami je osnovni produkt algoritma za ugotavljanje entropije, opisanega v razdelku 6.2.1. Pojavi se še vprašanje, kako daleč iti, do katere dolžine. Iz slik 22, 23, 24, 25, 26 in 27 na naslednji strani je razvidno, da začne število različnih n-terčkov od $n = 20$ naprej močno upadati in se je avtor zato odločil postaviti mejo pri 32. V tabeli 77 je navedeno število različnih in vseh n-terčkov z dolžinami od 1 do 32, ki imajo frekvenco vsaj 10, vsaj 5 in vsaj 2.

Tabela 77: Število n-terčkov ($n=1-32$) prvega vzorca za minimalne frekvence

minimalna frekvenca	različnih n-terčkov	vseh n-terčkov
10	1.454.567	120.517.900
5	3.427.898	133.038.634
2	16.401.724	164.019.971

Kot vidimo, je različnih n-terčkov z dolžinami od 1 do 32, ki imajo frekvenco vsaj 10, slab poldrugi milijon, s frekvenco 5 že skoraj 3,5 milijona, s frekvenco 2 pa že več kot 16 milijonov. Njihove skupne frekvence se gibljejo od 120 milijonov pri minimalni frekvenci 10 do 164 milijonov pri minimalni frekvenci 2. Skratka, jezikovni model za slovenski jezik, ki bi poznal pogoste nize in njihove frekvence, bi potreboval precej prostora, a še vedno v okvirih zmogljivosti sedanjih računal-

Sliki 22 in 23: Število različnih(22) in vseh(23) n -terčkov s frekvenco vsaj 10 po dolžinahSliki 24 in 25: Število različnih(24) in vseh(25) n -terčkov s frekvenco vsaj 5 po dolžinahSliki 26 in 27: Število različnih(26) in vseh(27) n -terčkov s frekvenco vsaj 2 po dolžinah

nikov. Iz slik 22, 24 in 26 razberemo, da se porazdelitve, ko povečujemo število n -terčkov (z nižanjem minimalne frekvence) širijo, vrh pa se pomika od 7 znakov (min. frekv. 10) na 8 (min. frekv. 5) in 9 znakov pri minimalni frekvenci 2. Ker želimo čimveč čimdaljših n -terčkov v modelu, da bo ocena entropije na znak lahko kar najmanjša, pričakujemo, da bo najboljšežnejši model tudi najboljši. Porazdelitve na slikah 23, 25 in 27 so nepravilne in monotono padajoče, saj skupne frekvence za manjši n seveda ne morejo biti manjše od tistih pri večjem n – povprečna frekvenca z rastočim n hitro pada. Vidimo tudi, da je celo pri minimalni frekvenci 2 skupna frekvenca n -terčkov od $n = 20$ naprej praktično zanemarljiva in da torej odločitev o omejitvi na dolžine do 32 ni bila prestroga.

Ostane še konstrukcija Huffmanovih kod za n -terčke vseh treh skupin in preizkus modela na besedilu drugega vzorca. Drevo s 16 milijoni listov in prav toliko razvejišči je bilo še v mejah računalniških virov, dostopnih avtorju. Ko so bile Huffmanove kode za vse n -terčke posameznih variant izračunane, je bilo treba z upoštevanjem n -terčkov in njihovih kod razrezati drugi vzorec tako, da bi bila

vsota kod za vse dobljene nize najkrajša. Algoritem razreza je sestavljen iz treh korakov:

1. na tekočem mestu poišči tisti n -terček iz učnega vzorca, ki se bo z besedilom ujema najdlje in postavi začasno delilno mesto na njegov konec;
2. pomikaj dobljeno delilno mesto za mesto nazaj toliko časa, dokler skupno število bitov na znak dveh nizov – od tekočega mesta do delilnega mesta in najdaljšega niza od delilnega mesta naprej ne bo slabše (večje);
3. pomakni tekoče mesto za delilno mesto in nadaljuj s korakom 1 dokler ne prideš na konec besedila;

Za ilustracijo algoritma je naveden razrez morda najznamenitejše povedi iz drugega vzorca, iz romana Pomladni dan.

Tabela 78: Primer entropijsko najugodnejšega razreza povedi iz drugega vzorca

10. Tistil_pomladni_dan_je_bil_lep_!svetel_lin_zvelneč_!kakor_iz_!čistegal_srebrla_ulitl_Rel
 5. Tistil_pomladni_dan_je_bill_lep_!svetel_in_lzvneč_!kakor_iz_!čistegal_srebra_lulitl_Res_je_da_l
 2. Tistil_pomladni_dan_je_bil_lep_!svetel_in_lzvneč_!kakor_liz_čistega_srebral_ulitl_Res_je_da_sl

V prvi vrstici tabele je naveden razrez, pri katerem so bili uporabljeni vsi n -terčki 1. vzorca z dolžino do 32 znakov, ki so imeli frekvenco vsaj 10, v drugi vrstici vsi s frekvenco vsaj 5, v tretji pa s frekvenco vsaj 2. Prva vrstica ima 12 delilnih mest, druga 11, tretja pa 10. Prva delitev je zajela 81 znakov, druga je prišla 9 znakov dlje, tretja pa še eno mesto naprej, 91 znakov daleč. Zaradi enostavnosti je v nadaljevanju naveden premislek začetka prve delitve.

Tabela 79: N -terčki prvega vzorca s frekvenco vsaj 10, z dolžinami Huffmanovih kod

T·12	o·7	p·8	_·5	i·7	
Ti·15	om·11	po·10	_p·9	i_·9	
Tis·18	oml·18	pom·14	_po·10	i_p·12	
Tist·18	omla·19	poml·19	_pom·14	i_po·13	
Tisti·19	omlad·19	pomla·19	_poml·19	i_pom·17	
Tisti_·20	omladn·22	pomlad·19	_pomla·19	i_poml·23	
Tisti_p·23	omladni·23	pomladn·23	_pomlad·20	i_pomla·23	
		pomladni·23	_pomladn·23	i_pomlad·23	
			_pomladni·23		
_·5	i·7	n·7	l·8	_·5	l·8
d·10	i·9	ni·10	le·11	_l·12	l_·10
_da·11	i_d·13	ni_·12	lep·14	_le·13	l_l·17
_dan·15	i_da·16	ni_d·16	lep_·20	_lep·15	l_le·18
dan·16	i_dan·17	ni_da·18	lep_·20	_lep_·22	l_lep·20
_dan_j·20	i_dan_j·18	ni_dan·20			
_dan_je·20	i_dan_j·21	ni_dan_·21			
_dan_je_·20	i_dan_je·21				
_dan_je_b·22	i_dan_je_·21				
_dan_je_bi·22	i_dan_je_b·23				
_dan_je_bil·22	i_dan_je_bi·23				
_dan_je_bil_·22	i_dan_je_bil·23				

Pri tem so bili uporabljeni podatki iz tabele 79. Na začetku povedi je bil najdaljši ugotovljeni niz dolg 7 znakov (*Tisti_p*), katerega Huffmanov kod je dolg 23 bitov. Niz *Tisti_po* se v prvem vzorcu ni pojavil 10-krat ali več. Začasno delilno mesto je torej postavljeno za sedmim znakom. Najdaljši niz, ki sledi, je *omladni* z dolžino Huffm. kode tudi 23 bitov. Skupaj 46 bitov za 14 znakov ali 3,29 bita na znak. Algoritem zdaj zahteva preverjanje toliko znakov nazaj, da dobimo slabši rezultat. En znak nazaj dobimo niza *Tisti_* in *pomladni* (niz *pomladni_* ni imel frekvenca 10 ali več), ki imata skupno dolžino Huffmanovih kod 20+23, se pravi 43 bitov, spet za 14 znakov, ali 3,07 bita na znak. Dva znaka nazaj opazujemo par *Tisti* in *_pomladni* (niz *_pomladni_* se očitno tudi ni pojavil v naboru), ki imata skupno dolžino Huffm. kod 19+23 = 42 bitov ali 3,00 bita na znak. Še eno mesto nazaj dobimo par *Tist* in *i_pomlad*, ki imata skupno dolžino Huffmanovih kod 18+23 = 41 bitov za 12 znakov, kar je 3,42 bita na znak ali slabše kot prej. Prvo delilno mesto torej pade za *Tisti*, s prirastkom 19 bitov (3,80 bita na znak).

Pri določanju drugega delilnega mesta sta prvi par *_pomladni* in *_dan_je_bil_*, ki imata 23 in 22 bitov dolgi Huffmanovi kodi, skupaj torej 45 bitov za 21 znakov ali 2,14 bita na znak. Mesto nazaj dobimo par *_pomladn* in *i_dan_je_bil* s skupno dolžino 46 bitov (23+23) za 20 znakov ali 2,30 bita na znak, kar je že slabše. Naslednje delilno mesto je torej za *_pomladni*. Pri določanju tretjega delilnega mesta imamo najprej niza *_dan_je_bil_* in *lep_*, s skupno dolžino Huffmanovih kod 22+20 = 42 bitov za 17 znakov ali 2,47 bita na znak. Mesto nazaj je par *_dan_je_bil* in *lep_*, skupaj 22+22 = 44 bitov ali 2,59 bita na znak, kar je slabše in delilno mesto postavljeno za *_dan_je_bil_*. Do tja se je torej skupaj nabralo 19+23+22 = 64 bitov za 26 znakov ali 2,46 bita na znak.

Rezultati razreza drugega vzorca za vse tri variante (min. frekvenca 10, 5 in 2) so navedeni v tabeli 80.

Tabela 80: Rezultati optimalnega razreza drugega vzorca z n-terčki prvega

minimalna frekvenca	znakov	nizov po razrezu	povprečna dolžina nizov	skupaj bitov	bitov na znak
10	2.296.775	331.153	6.94	6.647.122	2,89
5	2.296.775	305.344	7.52	6.469.638	2,82
2	2.296.775	273.301	8.40	6.245.660	2,72

Najkrajše dolžine Huffmanovih kod so bile v vseh treh primerih 6 bitov, najdaljše pa, spet v vseh treh primerih, 27 bitov. Ocena za entropijo, dolžina kod na znak besedila, pade od 2,89 bita pri najskromnejšem naboru n-terčkov do 2,72 bita na znak, če upoštevamo vse n-terčke iz prvega vzorca, dolge od 1–32 znakov, ki so se pojavili vsaj dvakrat. Rezultat je opazno večji od ocene iz razdelka 6.2.2, 2,2 bita na znak in je na prvi pogled videti slab. Poskusimo ga osvetliti še z drugega zornega kota. V ta namen razdelimo drugi vzorec na dva, praktično enako velika dela (49,96% : 50,04%) – prvi obsega vsa dela od začetka do vključno sce-

narija *Na svoji zemlji*, drugi pa vse ostalo. Če uporabimo prvi del kot učno bazo in generiramo iz njega vse n -terčke (1–32), izračunamo njihove Huffmanove kode in jih potem po v tem razdelku opisanem algoritmu uporabimo za generiranje najkrajših kod za drugi del (1.149.161 znakov), dobimo rezultate iz tabele 81:

Tabela 81: Rezultati optimalnega razreza druge polovice 2. vzorca z n -terčki prve polovice

minimalna frekvenca	različnih n -terčkov	vseh n -terčkov	bitov na znak
10	104.040	6.607.609	3,08
5	252.415	7.548.271	2,96
2	1.119.263	9.640.829	2,81
1	2.375.354	11.041.782	2,78

V tabeli smo šli še dlje kot pri prejšnji raziskavi, saj smo v zadnji vrstici uporabili prav vse n -terčke, tudi tiste, ki so se pojavili samo enkrat. Iz vrednosti zadnjega stolpca sklepamo, da je entropija Kosmačevega besedila, če druge polovice s prvo ni mogoče napovedati bolje kot do 2,78 bita na znak, res bližje oceni 2,72 bita na znak, dobljeni v tabeli 80, kot pa kaki nižji. Drug premislek, ki kaže na višjo entropijo slovenskega leposlovnega besedila v primerjavi z vzorcem iz ameriške raziskave (Brown idr. 1992a), najdemo v opisu besednjakov IBM-ovega vzorca (Jelinek 1997, tabela 4.2, str. 75). V njem je navedeno, da je bilo za 15.000 novih besed (lem, ne besednih oblik) potrebno obdelati 640.000 besed obsegajoče besedilo. Drugi vzorec pričujočega dela je bil oblikoslovno označen in lematiziran in zato lahko v prejšnjem stavku navedeni številki primerjamo s Kosmačevim besedilom (glej tabelo 82 v naslednjem razdelku). V drugem vzorcu je bilo za 15.140 lem potrebnih le 408.000 besed besedila. Ta podatek spet kaže na višjo entropijo leposlovnega besedila.

Iz vsega navedenega avtor sklepa, da je z modelom, dobljenim iz besedil prvega vzorca in uporabljenim na besedilih drugega vzorca ugotovljena ocena zgornje meje entropije slovenskega leposlovnega besedila manjša od 2,72 oziroma

2,7 bita na znak

Poleg komprimiranja je model z optimalnim razrezom na n -terčke, ker vsebuje veliko zbirko nizov iz slovenskih besedil s pogostnostmi, mogoče uporabiti še za ugotavljanje, ali je neko besedilo slovensko oziroma, kako daleč je od slovenskega jezika. V tabeli 82 so navedeni rezultati uporabe modela na istem besedilu, prevodu Platonove države, v 16 evropskih jezikih (Projekt TELRI: Erjavec, Lawson in Romary 1998). Tabela je urejena padajoče po povprečnem številu bitov na znak, ki jih je pri kodiranju posameznega besedila dosegel model. V vsaki vrstici je najprej navedena zaporedna številka jezika, nato njegovo ime, prevajalec, leto izdaje dela, ki je služilo za prenos v elektronsko obliko (navadno se ujema z letom prevoda), vodjo prenosa v elektronsko obliko (le prvega, kadar jih je bilo več), šte-

vilo besed v delu, število znakov in na koncu še najpomembnejše, povprečno število bitov na znak, ki jih je potreboval model. Pomišljaj pomeni, da podatek ni bil znan. Besedila, dostopna na TELRI-jevem cedeju je avtor po najboljših močeh poenotil oz. transkribiral – bolgarsko besedilo je bilo že transkribirano v latinično, le šumniki so bili označeni s *ch*, *sh*, *zh*, ruski izvod pa je bil še v cirilici. Besedilom v češkem, slovaškem in poljskem jeziku so bila odvzeta diakritična znamenja (rezultati z njimi vred bi bili povprečno 5.88, 5.37 in 5.34 bita na znak).

Tabela 82: Preizkus modela na besedilu Platonove Države v 16 jezikih

jezik	prevod	leto	prenos v el. obliko	besed	znakov	b/z
- slovenski	Jože Košar	1976	Primož Jakopin	92.741	565.604	2,37
1. srbohrvaški	A. Vilhar, B. Pavlović	1983	Duško Vitas	107.506	613.082	3,77
2. hrvaški	Damir Salopek	1976	Marko Tadić	92.870	532.497	3,84
3. bolgarski	-	-	Patrice Bonhomme	112.676	678.131	3,96
4. češki	Radislav Hošek	1993	František Čermák	110.466	636.201	4,10
5. poljski	Wladyslaw Witwicki	1991	-	107.559	645.532	4,32
6. ruski	-	-	-	99.503	649.102	4,46
7. slovaški	Július Španár	1990	Alexandra Jarošová	99.661	622.463	4,46
8. latvijski	Gustavs Lukstīņš	1982	Andrejs Spektors	45.238	290.508	4,74
9. litvanski	Jonas Dumčius	1981	-	85.144	584.318	4,94
10. angleški	Paul Shorey	-	-	129.331	692.058	5,40
11. francoski	-	1993	-	142.624	817.658	5,67
12. nemški	Karl Vretska	1982	Joachim Hohwieler	104.876	641.333	5,69
13. romunski	Andrei Cornea	1986	Dan Tufis	131.064	658.804	5,76
14. finski	Marja Itkonen-Kaila	-	Anna Mauranen	75.800	582.522	6,11
15. madžarski	Szabó Miklós	1984	Tamás Váradi	105.538	728.501	6,47

Po pričakovanju sta, tudi če upoštevamo rezultate modela, slovenščini najbližja srbski in hrvaški jezik, sledijo pa bolgarski, češki, poljski, ruski in slovaški. Na koncu seznama sta, prav tako ne preveč presenetljivo, oba ugrofinska jezika, le da je finski našemu jeziku po modelu precej bližji kot madžarščina.

6.3.6 Komprimiranje vzorcev z drugimi programi

Za dodatno osvetlitev so navedeni še rezultati komprimiranja besedil obeh vzorcev z najboljšimi programi za komprimiranje datotek, ki jih je mogoče najti na internetu. Zanimanje za tovrstne programe, predvsem za shranjevanje slik, ki delajo brez izgube informacije (angl. *lossless compression*), se je z razmahom digitalne fotografije v zadnjih dveh letih zelo povečalo.

Avtor se je pri odločanju, katere izmed velike množice programov izbrati, oprl predvsem na rezultate preizkusov, ki sta jih opravila Jeff Gilchrist (internetni naslov strani <http://web.act.by.net/~act/act.html>) in Aleksander Ratušin (<http://www.geocities.com/SiliconValley/5577/artest10.html>). Rezultati za prvi in za drugi vzorec so v tabelah 83 in 84 na naslednji strani; v rubriki tip je navedeno, ali je

program brezplačen (fw, od angl. freeware) ali ga je treba kupiti po navadno enomesečnem preizkusnem obdobju (sh, od angl. shareware). Programi v obeh tabelah so urejeni naraščajoče po zmogljivosti, ugotovljeni na prvem vzorcu – se pravi po številu bitov, ki jih je za kodiranje znaka obeh datotek (z besedili prvega oz. drugega vzorca) povprečno potreboval ustrezeni program. Prvih pet programov, z izjemo programa ESP, je delo renomiranih in uveljavljenih proizvajalcev – PKZIP je dejanski standard za izmenjavo komprimiranih datotek na internetu. Podrobnosti, predvsem o implementaciji in navadno tudi o uporabljenih algoritmih, seveda niso znane; zadnjih pet programov, v največji meri najboljši trije, Boa Constrictor, RKIVE in Ufa, uporabljajo tehniko zaporednega komprimiranja podatkov z različnimi postopki, ki se utegnejo po predhodni oceni najboljše obnesti – npr. najprej LZW, potem aritmetično kodiranje. Časi izvajanja so v sekundah, na običajnem namiznem računalniku tipa PC, stikala pa so bila izbrana za največje možno komprimiranje besedil, kadar je program to izbiro imel, sicer pa na največje možno komprimiranje.

Tabela 83: Komprimiranje prvega vzorca s tujimi komprimirnimi programi

ime programa	verzija	dobavitelj	država	leto	tip	dolžina programa	stikala	čas izv. v sek.	bitov na znak
ARJ	2.60	ARJ Software	US	1997	sh	203.199	-m1 -jm	84	3,27
ESP	1.92	GyikSoft	HU	1996	fw	12.644	-	60	3,22
PKZIP	2.04g	PKWARE	US	1993	sh	42.166	-ex	74	3,20
RAR	2.50b3	RarSoft	FI	1999	sh	105.224	-m5	138	3,09
JAR	1.02	ARJ Software	US	1997	sh	343.030	-m4	195	3,00
UHARC	0.2	Uwe Harklotz	GE	1997	sh	99.444	-m3	751	2,73
Arhangel	1.38	Jurij Ljapko	UA	1999	sh	76.416	-	188	2,52
Ufa	0.04b1	Igor Pavlov	RU	1998	sh	133.632	-m5 -mu32	703	2,24
RKIVE	1.92b1	Malcolm Taylor	NZ	1998	sh	120.740	-mt3	2.148	2,23
Boa	0.58b	Ian Sutton	CA	1998	sh	105.984	-m15	756	2,16

Tabela 84: Komprimiranje drugega vzorca s tujimi komprimirnimi programi

ime programa	verzija	dobavitelj	država	leto	tip	dolžina programa	stikala	čas izv. v sek.	bitov na znak
ARJ	2.60	ARJ Software	US	1997	sh	203.199	-m1 -jm	13	3,19
ESP	1.92	GyikSoft	HU	1996	fw	12.644	-	9	3,14
PKZIP	2.04g	PKWARE	US	1993	sh	42.166	-ex	10	3,12
RAR	2.50b3	RarSoft	FI	1999	sh	105.224	-m5	19	3,01
JAR	1.02	ARJ Software	US	1997	sh	343.030	-m4	31	2,89
UHARC	0.2	Uwe Harklotz	GE	1997	sh	99.444	-m3	84	2,63
Arhangel	1.38	Jurij Ljapko	UA	1999	sh	76.416	-	28	2,48
Ufa	0.04b1	Igor Pavlov	RU	1998	sh	133.632	-m5 -mu32	105	2,25
RKIVE	1.92b1	Malcolm Taylor	NZ	1998	sh	120.740	-mt3	270	2,10
Boa	0.58b	Ian Sutton	CA	1998	sh	105.984	-m15	100	2,12

Rezultati sicer niso neposredno primerljivi s postopki iz prejšnjih razdelkov, kjer se modeli opirajo na zelo velike zbirke s podatki, ki so tipični za slovenski jezik,

pa vseeno – postopek z optimalnim razrezom na n -terčke je potreboval na prvem vzorcu 255 sekund in kodiral znake s povprečno 2,36 bita na znak, na drugem pa 44 sekund z že omenjenimi 2,72 bita na znak.

Komprimirni časi programov iz obeh tabel gredo pri prvem vzorcu sicer v desetih minut (RKIVE je potreboval celo dobre pol ure), vendar opazimo, da se mehanizmi, razviti predvsem za komprimiranje slik, tudi na besedilih zelo dobro obnesejo. Opazimo tudi, da se, z eno samo izjemo (program Ufa) povprečno število bitov na črko kodiranih datotek z rastjo vzorca povečuje. Pri najboljšem, programu Boa Constrictor, se dvigne z 2,12 bita na črko pri drugem vzorcu na 2,16 bita na črko pri osemkrat večjem prvem vzorcu. Ocena 2,2 bita na črko za zgornjo mejo entropije slovenskih leposlovnih besedil (pribl. stokrat večja celota od obeh vzorcev) dobi s tem nov argument.

6.4 Druge entropije

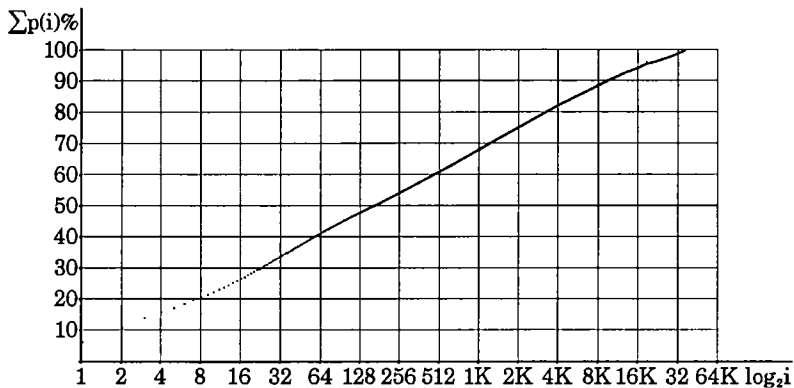
Ker je bil drugi vzorec, kot že rečeno, oblikoslovno označen, je mogoče podati še nekaj vrednosti za entropije, povezane s tem. Iz tabele 85 na naslednji strani vidimo, da znaša slovnična informacija, ki jo vsebuje povprečna beseda v besedilu, slabih 7 bitov (en dodaten znak) in da je ta informacija z besedno obliko že precej dobro določena – razlika med entropijo leme in označenega para je samo 1,618 bita. Na slikah 28, 29 in 30 so navedene krivulje rasti za besedne oblike, za leme in za oblikoslovne oznake v drugem vzorcu.

Tabela 85: Entropije besednih oblik in lem iz drugega vzorca

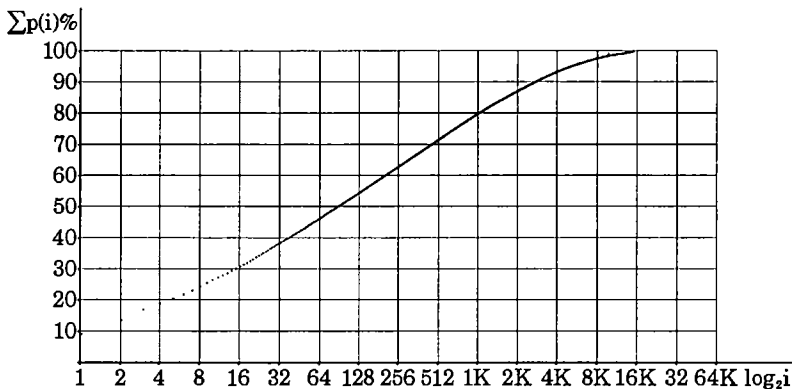
	n	entropija na enoto
različnih parov (besedna oblika, oblikoslovna oznaka)	44.858	11,201
različnih lem (osnovnih besednih oblik)	15.140	9,582
različnih oblikoslovnih oznak	1.498	6,607

Primerjava slik pokaže veliko lepši potek krivulje (končni asimptoti se prilega kot tangenta in ne kot sekanta) pri lemah in seveda še lepši pri oblikoslovnih oznakah.

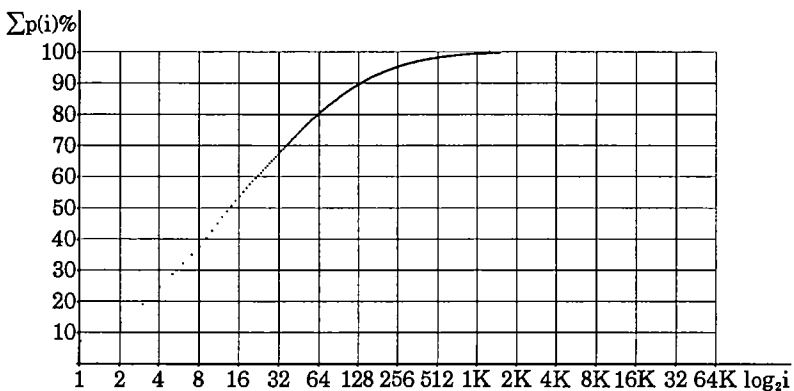
Slika 28: Krivulja rasti za besedne oblike v drugem vzorcu



Slika 29: Krivulja rasti za leme v drugem vzorcu



Slika 30: Krivulja rasti za oblikoslovne oznake v drugem vzorcu



7. poglavje

Sklep

Pot, prehojena do teh vrstic, je bila precej dolga in dostikrat tudi bolj ovinkasta kot ne. Od prvih raziskav entropije v našem prostoru (Matković 1957 in Gyergyék 1962) in pionirskih kvantitativnih analiz slovenskega jezika (Gyergyék s sod. 1974 in Poniž 1974) so minila skoraj tri desetletja. Šele pospešeni razvoj in široka dostopnost prave računalniške moči v zadnjih letih sta omogočila analizo besedil v obsegu pričujočega dela. Pomen besedilnih korpusov za znanost se v zadnjih letih, tudi zaradi naštetega, povečuje (npr. Sinclair 1991, Habert 1997, Oakes 1997, McEnery in Wilson 2001) in vzbuja nove, tokrat ne več tako preuranjene upe v uporabnost kvantitativnih raziskav.

Če se ozremo nazaj in pogledamo, kaj manjka in kaj bi se dalo napraviti bolje, naletimo predvsem na dvojce.

Najprej se pojavi vprašanje velikosti in vsebine vzorca slovenskih leposlovnih besedil. Če upoštevamo, da je povprečen pripadnik našega jezikovnega prostora po zaključku srednje šole v obliki obveznega šolskega čtiva in morda še kakšne druge knjige prebral približno 90 leposlovnih del (4 krat 8 in 4 krat 15), dobimo obseg, primerljiv z velikostjo obeh vzorcev tega dela. Da pa bi lahko napravili model, ki bi služil za podlago res učinkovitemu mehanizmu za komprimiranje slovenskih leposlovnih besedil, bi bilo potrebno dodati predvsem imena, ki so zviševala oceno entropije v modelih iz tega dela (glej npr. konec razdelka 6.3.4, str. 106). Virov je več, najkrajši pa bi bila nedvomno publikacija s kratkimi vsebinami slovenskih proznih del (npr. Janež 1992). Boljša pot bi bila seveda, če bi vsa ta dela vseboval. Drug nujni dodatek bi bila dopolnitev vzorca z besedili, ki smo jim poleg leposlovnih še izpostavljeni, predvsem časopisnimi in televizijskimi.

Drugi premislek se tiče predvsem mehanizma za komprimiranje slovenskih besedil in je delno povezan s prvim. Gre za to, da bi, če naj se spusti niže v bližino ocene 2,2 bita na znak, moral biti hibriden, to je tak, ki bi poleg pogostih znakovnih nizov, ki se pojavljajo v opazovanem besedilu in ki jih algoritem že pozna v učnem vzorcu, upošteval še lokalne pogostosti, kakršne so recimo imena glavnih

junakov ali pa pogostosti, ki spadajo k slogu pisanja in ki jih s pridom izkoriščajo algoritmi v najučinkovitejših komprimirnih programih, kakršni so bili npr. omenjeni v razdelku 6.3.6.

Poleg načelnih premislekov je vedno zanimiva še bolj bližnja bodočnost. Priprave na oblikoslovno označitev prvega vzorca so pri koncu, zbrana in prečiščena pa je tudi že znatna zbirka časopisnega jezika (nad 80 milijonov besed, elektronska različica časopisa DELO od januarja 1998 do decembra 2002 ter revija Monitor, letniki 1999, 2000 in 2001, spletni naslov <http://bos.zrc-sazu.si>), ki bo omogočila kvantitetni preskok s 3 milijonov besed na blizu 100 milijonov besed veliko besedilno zbirko. Z vsem navedenim bi se dalo učinkovito podpreti graditev besedilnega korpusa slovenskega jezika, t. i. Slovenskega nacionalnega korpusa, katerega nastanek že nekaj časa visi v zraku (Hladnik 1995, Balažic idr. 1998, Dobrišek idr. 1998, Erjavec idr. 1998, Kranjc 1998, Jakopin 2000).

Literatura

- [1] Abramson, N. (1963). *Information Theory and Coding*. McGraw-Hill, New York.
- [2] Baker, P., Burnard, L., McEnery, A. in Wilson, A. (1997). *An analytic framework for the validation of language corpora*. ELRA, Pariz.
- [3] Balažic, D., Dimec, Z., Kanič, A., Žumer, M., Gazvoda, J. in Jakac-Bizjak, V. (1998). *Elektronske publikacije : kako jih bomo ohranili za prihodnost?*. Narodna in univerzitetna knjižnica, Ljubljana.
- [4] Bentley, J. L., Sleator, D. D., Tarjan, R. E. in Wei, V. K. (1986). A Locally Adaptive Data Compression Scheme. *Commun. ACM*, 29/4: 320–330.
- [5] Bick, E. (1996). Automatic Parsing of Portuguese. *Proceedings of the Second Workshop on Computational Processing of Written Portuguese*, Curitiba.
- [6] Brent, R. in Kung, H. T. (1978). Fast Algorithms for Manipulating Formal Power Series. *J. ACM*, 25/4: 581–595.
- [7] Brown, P. F., Della Pietra, V. J., Della Pietra, S. A., Lai, J.C., in Mercer, R. L. (1992a). An Estimate of an Upper Bound for the Entropy in English. *Computational Linguistics*, 18/1: 31–40.
- [8] Brown, P. F., deSouza, P. V., Mercer, R. L., Pietra, V. J. D., in Lai, J. C. (1992b). Class-based n-gram models of natural language. *Computational Linguistics*, 18/4: 467–479.
- [9] Bshouty, N. H. in Falk G., (1992). Compression of dictionaries via extended front coding and Huffman coding. *Proceeding of the 4th International Conference Computing and Information*, 361–364.
- [10] Burrows M. in Wheeler D. J. (1994). *A Block-sorting Lossless Data Compression Algorithm*. SRC Research Report 124. Digital Systems Research Center, Palo Alto.

- [11] Calude, C. S. in Tomescu, I. (1997). Optimum Extendible Prefix Codes. *Journal of Universal Computer Science*, 3/11: 1167–1179
- [12] Cesar, I. (1981). *Poetika pripovedne proze Cirila Kosmača*. Prev. Mojca Mihelič, Mladinska knjiga, Ljubljana.
- [13] Chanod, J. P., in Tapanainen, P. (1995). Creating a tagset, lexicon and guesser for a French tagger. *ACL SIGDAT workshop*. University College Dublin. 58–64.
- [14] Chanod, J. P. (1997). Multilingual tools at the Xerox Research Centre. *TELRI Newsletter*, 7: 4–12.
- [15] Cortesi, D. (1982). An Effective Text–Compression Algorithm. *BYTE*, 7/1: 397–403.
- [16] Čermák, F. (1997). The Czech National Corpus: A Brief Survey of the Current State. *TELRI Newsletter*, 4: 28–29.
- [17] Dermatas, E., in Kokkinakis, G. (1995). Automatic stochastic tagging of natural language texts. *Computational Linguistics*, 21/2: 137–163.
- [18] Dobrišek, S., Gros, J., Ipšič, I., Pepelnjak, K., Mihelič, F., in Pavešič, N. (1998). GOPOLIS: slovenska podatkovna zbirka govornjenih poizvedovanj. *Zbornik konference: Jezikovne tehnologije za slovenski jezik*, Ljubljana.
- [19] Erjavec, T. (1997). Računalniške zbirke besedil. *Jezik in slovstvo*, XLII/2–3: 81–96.
- [20] Erjavec, T., Gorjanc, V. in Stabej, M. (1998). Korpus FIDA. *Zbornik konference: Jezikovne tehnologije za slovenski jezik*, Ljubljana.
- [21] Erjavec, T., Lawson, A., in Romary, L. (ur.) (1998). *East meets West - A Compendium of Multilingual Resources*. TELRI, Institut für deutsche Sprache, Mannheim.
- [22] Faller, N. (1973). An Adaptive System for Data Compression. *Record of the 7th Asilomar Conference on Circuits, Systems and Computers*, Pacific Grove, 593–597.
- [23] Fano, R. M. (1949). *Transmission of Information*. M.I.T. Press, Cambridge.
- [24] Fenwick, P. (1996). Block Sorting Text Compression. *Proceedings of the 19th Australasian Computer Science Conference*, Melbourne.

- [25] Finžgar, F. S. (1978). *Pod svobodnim soncem*. Mladinska knjiga, Ljubljana.
- [26] Fisher, Y. (ur.) (1995). *Fractal Image Compression: Theory and Application*. Springer-Verlag, New York.
- [27] Fraenkel, A. S., Mor, M. in Perl, Y. (1983). Is Text Compression by Prefixes and Suffixes Practical? *Acta Inf.*, 20/4: 371–375.
- [28] Francis, N. W., in Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Houghton Mifflin, Boston.
- [29] Gallager, R. G. (1978). Variations on a theme by Huffman. *IEEE Transactions in Information Theory*, 24/6: 668–674.
- [30] Garside, R., idr. (1987). *The computational analysis of English: A corpus-based approach*. Longman, London.
- [31] Glušič, H. (1975). *Pripovedna proza Cirila Kosmača*. Slovenska matica, Ljubljana.
- [32] Grefenstette, G. (1995). Comparing Two Language Identification Schemes. *Proceedings of 3rd International Conference on Statistical Analysis of Textual Data*, Rim.
- [33] Grefenstette, G. (1996). Approximate Linguistics. *Proceedings of the 4th International Conference on Computational Lexicography*. Budimpešta. 83–96
- [34] Grzybek, P. (2001). Pogostnostna analiza besed iz elektronskega korpusa slovenskih besedil. *Slavistična revija* 48/2: 141–157.
- [35] Gyergyék, L. (1962). Nekateri stavki iz teorije o informacijah in srednja vrednost informacije na črko slovenske abecede. *Avtomatika*, III/april: 74–80.
- [36] Gyergyék, L. (1973). Prispevek k statistični obdelavi slovenskega pisanega besedila. *Elektrotehniški vestnik*, 40/11–12: 247–252.
- [37] Gyergyék, L., s sod. (1974). *Prispevek k statistični obdelavi slovenskega jezika*. Raziskovalna naloga 122, Fakulteta za elektrotehniko, Ljubljana.
- [38] Habert, B., Nazarenko, A., in Salem, A. (1997). *Les linguistiques de corpus*. Armand Colin/Masson, Paris.

- [39] Hahn, B. (1974). A New Technique for Compression and Storage of Data. *Commun. ACM*, 17/8: 434–436.
- [40] Hajičová, E. (1997). A New Impetus For The Study Of Computerized Corpora In Prague. *TELRI Newsletter*, 5: 36–37.
- [41] Hladka, B., in Hajič, J. (1995). A Simple Czech and English Probabilistic Tagger: a Comparison. *TELRI: Proceedings of the First European Seminar*, Tihany. 191–196.
- [42] Hladnik, M. (1995). Elektronski literarnovedni viri in računalniško pisanje. *Jezik in slovnstvo*, XL/7: 243–254.
- [43] Hu, T. C. in Tucker, A. C. (1971). Optimal Computer Search Trees and Variable–Length Alphabetic Codes. *SIAM J. Appl. Math.*, 21/4: 514–532.
- [44] Huffman, D. A. (1952). A Method for the Construction of Minimum–Redundancy Codes. *Proc. IRE*, 40/9: 1098–1101.
- [45] Ingels, F. M. (1971). *Information and Coding Theory*. Intext, Scranton.
- [46] Itai, A. (1976). Optimal Alphabetic Trees. *SIAM J. Comput.*, 5/1: 9–18.
- [47] Jacquin, A. E. (1993). Fractal image coding: A review. *Proceedings of the IEEE*, 81/10: 1451–1465.
- [48] Jakac–Bizjak, V. (ur.) (2001). *Elektronske publikacije, Kodeks prakse prostovoljnega depozita*. Ljubljana: Narodna in univerzitetna knjižnica.
- [49] Jakopin, P. (1981). *Entropija imen in priimkov v Sloveniji*. Magistrsko delo. Sveučilište u Zagrebu, Zagreb.
- [50] Jakopin, P. (1995a). Nekaj številčk iz Slovarja slovenskega knjižnega jezika. *Slavistična revija*, 43/3: 341–375.
- [51] Jakopin, P. (1995b). O n-terčkih in o novem postopku za deljenje besed. *TIP*, 1/6: 38–40.
- [52] Jakopin, P. (1995c). EVA – A Textual Data Processing Tool. *Proceedings of the First TELRI Seminar*, Tihany. 169–171.
- [53] Jakopin, P. (1996a). TELRI Birmingham Workshop Report. *TELRI Newsletter*, 3: 17–23.

- [54] Jakopin, P. (1996b). Ali so rojstna imena krajša od drugih samostalnikov? *Slavistična revija*, 44/2: 193–200.
- [55] Jakopin, P. (1998). Plato's Republic: Part-of-speech comparison of English, German and Slovenian translation. *Proceedings of the Third TELRI seminar*, Montecatini. 91–102.
- [56] Jakopin, P., in Bizjak, A. (1997). O strojno podprtem oblikoslovnem označevanju slovenskega besedila. *Slavistična revija*, 45/3–4: 513–532.
- [57] Jakopin, P., in Musar, A. (1997). Kvantitativni prikaz prevoda zbirke zgodb Alica v čudežni deželi. *Slava*, 1996/97, 2: 135–142.
- [58] Jakopin, P. (2000). Slovenian texts on the internet. *Zapiski*, 7: 4–7
- [59] Jakopin, P. (2001). Beseda : a Slovenian text corpus. V: Fraser, M., Williamson, N. & Deegan, M. (eds.), *Digital Evidence : selected papers from DRH2000, Digital Resources for the Humanities Conference*, University of Sheffield, September 2000, (pp. 229–241). London: Office for Humanities Communication.
- [60] Jakopin, P. (2002). The feasibility of a complete text corpus. In: *Proceedings of the Third international conference on language resources and evaluation*, May 2002, Vol. II, pp. 437–440. Las Palmas: ELRA.
- [61] Janež, S. (1992). *Poglavitna dela slovenske književnosti : vsebine in označitve. 1, Pripovedna in dramska dela : 1789-1986*. Obzorja, Maribor.
- [62] Järvinen, T. (1994). Annotating 200 Million Words: The Bank of English Project. *Proceedings of the 15th International Conference on Computational Linguistics*, Kyoto.
- [63] Jelinek, F. (1997). *Statistical Methods for Speech Recognition*. MIT Press, Cambridge.
- [64] Klimová, J. (1996). Czech National Corpus. *Proceedings of the Tübingen East-West Computational Linguistics Meeting*. Tübingen.
- [65] Kmecl, M. (1996). *Mala literarna teorija*. Mihelač in Nešović, Ljubljana.
- [66] Knuth, D. E. (1985). Dynamic Huffman Coding. *Algorithms Journal*, 6/2: 163–180.
- [67] Kodrič-Dačić, E. (ur.) (1997). *Zapisi za prihodnost: 190 let obveznega izvoda tiska na Slovenskem*. Narodna in univerzitetna knjižnica, Ljubljana.

- [68] Kos, J. (1994). *Očrt literarne teorije*. DZS, Ljubljana.
- [69] Kranjc, S. (1998). Govorjena besedila in korpus slovenskega jezika. *Zbornik konference: Jezikovne tehnologije za slovenski jezik*, Ljubljana.
- [70] Kristan, B., Jug, M., Kovačič, S., in Gyergyék, L. (1994). Entropija slovenskih besedil. *Elektrotehniški vestnik*, 61/4: 171–179.
- [71] Lah, A. (1997). *Pregled svetovne in slovenske književnosti*. Rokus, Ljubljana.
- [72] Leech, G., Garside, R., in Bryant, M. (1994). CLAWS4: The tagging of the British National Corpus. *Proceedings of the 15th International Conference on Computational Linguistics*. Kyoto. 622–628.
- [73] Leech, G., Rayson, P. & Wilson, A. (2001). *Word Frequencies in Written and Spoken English: based on the British National Corpus*. London: Longman.
- [74] Leko, N., (1998). *Some problems in compiling a frequency dictionary from the Oslo Corpus of Bosnian Texts*. Poročilo. Univerza u Sarajevu, University of Oslo.
- [75] Marcus, M. P. (1993). Building a Large Annotated Corpus of English: the Penn Treebank. *Computational Linguistics*, 19/2: 313–330.
- [76] Matković, V. (1957). *Primjena teorije komunikacije na određivanje entropije hrvatskog jezika*. Doktorska disertacija. Elektrotehnički fakultet, Zagreb.
- [77] McEnery, T. & Wilson, A. (2001). *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- [78] Mihelič, F. (1998). Klepet z računalnikom v slovenščini, slovaščini, češčini ali nemščini?: samodejno razpoznavanje, razumevanje in tvorjenje govora. *Delo*, 138: 10.
- [79] Moguš, M. (ur.) (1999). *Hrvatski čestotni rječnik*. Zagreb : Školska knjiga.
- [80] MULTEXT-East (1997). *East meets West: A Compendium of Multilingual Language Resources*, Vol. II. CD-ROM. TELRI, Mannheim.
- [81] Oakes, M. (1997). *Statistics for Corpus Linguistics*. Edinburgh University Press, Edinburgh.

- [82] Orešnik, J. (1995). *Uradi za jezik v Skandinaviji*. Slovenska akademija znanosti in umetnosti, Ljubljana.
- [83] Pavešić, N. (1997). *Informacija in kodi*. Fakulteta za elektrotehniko, Ljubljana.
- [84] Perec, G. (1969). *La disparition*. Denoël, Paris.
- [85] Pennebaker, W. B. in Mitchell, J. L. (1993). *JPEG Still Image Data Compression Standard*. Van Nostrand Reinhold, New York.
- [86] Pivec-Stelè, M. (1936). Slovenska knjižna produkcija v letih 1919–1935. *Čas, Revija Leonove družbe*, XXX/19: 241–264.
- [87] Parker, D. S. (1980). Conditions for the Optimality of the Huffman Algorithm. *SIAM J. Comput.*, 9/3: 470–489.
- [88] Poniž, D. (1974). *Slovenski jezik - literatura - računalniki*. Obzorja, Maribor.
- [89] Press, W. H., Flannery, B. P., Teukolsky, S. A., in Vetterling, W. T. (1992). Huffman Coding and Compression of Data. V *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd ed.. Cambridge University Press, Cambridge, 896–901.
- [90] Prijatelj, I. (1952). (Slodnjak, A. (ur.). Literarna zgodovina. *Izbrani eseji in razprave Ivana Prijatelja*, I. zv., Slovenska matica, Ljubljana.
- [91] Ratnaparkhi, A. (1996). A Maximum Entropy Model for Part-Of-Speech Tagging. *Proceedings of the Empirical Methods in Natural Language Processing Conference*, University of Pennsylvania.
- [92] Rice, R. F., idr., (1993). Algorithms for High-Speed Universal Noiseless Coding. *Proc. of the 9th Conference AIAA Computing in Aerospace*, San Diego.
- [93] Rubin, F. (1979). Arithmetic Stream Coding Using Fixed Precision Registers. *IEEE Trans. Inform. Theory*, 25/6: 672–675.
- [94] Santos, D. (1998). Providing access to language resources through the World Wide Web: the Oslo Corpus of Bosnian Texts. *Proceedings of The First International Conference on Language Resources and Evaluation*, (Granada, 28–30 May 1998). Granada. 475–481.
- [95] Sayood, K. (1997). *Introduction to Data Compression*. Morgan Kaufmann, San Francisco.

- [96] Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27: 379–423. 623–656.
- [97] Shannon, C. E. (1951). Prediction and Entropy of Printed English. *Bell System Technical Journal*, 30, 50–64.
- [98] Shannon, C. E. in Weaver, W. (1949). *The Mathematical Theory of Communication*. University of Illinois Press, Urbana.
- [99] Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.
- [100] Snyderman, M. in Hunt, B. (1970). The Myriad Virtues of Text Compaction. *Datamation*, 16/12: 36–40.
- [101] Slovenska bibliografija (1945 ... 1979). *Slovenska bibliografija*. Narodna in univerzitetna knjižnica. Ljubljana.
- [102] SSKJ (1994). *Slovar slovenskega knjižnega jezika*. DZS, Ljubljana.
- [103] Suhadolnik, S. (1985). *Gradivo za Prešernov slovar*. Znanstvenoraziskovalni center Slovenske akademije znanosti in umetnosti: Inštitut za slovenski jezik, Ljubljana.
- [104] Štih, B. (1958). Spremna beseda in opombe. *Iz moje doline*. Mladinska knjiga, Ljubljana.
- [105] Tapanainen, P., in Voutilainen, A. (1994). Tagging accurately – Don't guess if you know. *Proceedings of the 4th Conference on Applied Natural Language Processing*. Stuttgart. 47–52.
- [106] Tobin Y., (1997). *Phonology as Human Behavior*. Duke University Press, Durham.
- [107] Toporišič, J. (1984). *Slovenska slovnica*. Obzorja, Maribor.
- [108] Tropper, R. 1982. Binary–Coded Text, A Text–Compression Method. *BYTE*, 7/4: 398–413.
- [109] Underwood, N., in Navaretta, C. (1997). *A Draft Manual for the Validation of Lexica: Preliminary Report*. ELRA, Pariz.
- [110] Vasle, T. (1980). *Statistična obdelava slovenskega besedila*. Diplomaska naloga. Fakulteta za elektrotehniko, Ljubljana.

-
- [111] Venbrux, J., idr., (1992). A VLSI Chip Set for High-Speed Lossless Data Compression. *IEEE Trans. on Circuits and Systems for Video Technology*, 2/4.
- [112] Vitter, J. S. (1987). Design and Analysis of Dynamic Huffman Codes. *Journal ACM*, 34/4: 825–845.
- [113] Wagner, R. A. (1973). Common Phrases and Minimum-Space Text Storage. *Commun. ACM*, 16/3: 148–152.
- [114] Wallace, G. K. (1991). The JPEG Still Picture Compression Standard, *Communications of the ACM*, 34/4: 30–44.
- [115] Welch, T. A. (1984). A Technique for High-Performance Data Compression. *Computer*, 17/6: 8–19.
- [116] Witten, I. H., idr. (1987). Arithmetic Coding for Data Compression. *Commun. ACM*, 30/6: 520–540.
- [117] Wright, E. V. (1939). *Gadsby: A Story of Over 50,000 Words Without Using the Letter 'E'*. Wetzell, Los Angeles.
- [118] Ziv, J. in Lempel, A. (1977). A Universal Algorithm for Sequential Data Compression. *IEEE Trans. Inform. Theory*, 23/3: 337–343.
- [119] Zorman, M. (1997). *Nekateri strukturalno-funkcionalni odnosi med členi sinonimnih vrst v slovenščini*. Doktorska disertacija. Filozofska fakulteta, Ljubljana.

Priloga A

Seznam del v prvem vzorcu

A.1 Izvirna dela

- Ivan Cankar: *Hiša Marije Pomočnice*. L. Schwentner, Ljubljana 1904.
- Ivan Cankar: Potepuh Marko in Kralj Matjaž. v: Anton Knezova knjižnica, zv. 12, Slovenska Matica, Ljubljana 1905.
- Ivan Cankar: *Hlapec Jernej in njegova pravica*. L. Schwentner, Ljubljana 1907.
- Ivan Cankar: *Mimo življenja*. Ljubljana 1920.
- Janez Cigler: *Sreča v nesreči*. Jožef Blaznik, Ljubljana 1836.
- Dragotin Cvetko: *V prostoru in času: spomini*. Slov. matica, Ljubljana 1995.
- Fran S. Finžgar: *Sama. Dom in svet*, 1912.
- Fran S. Finžgar: *Strici*. v: *Vest in zakon. Strici* (ps. F. Detela), Družba sv. Mohorja, Prevalje 1927.
- Nataša Gale: Fran Gestrin. v: *Vest in zakon. Strici* (ps. F. Detela), Družba sv. Mohorja, Prevalje 1927.
- Miran Hladnik: *Kmečka povest* (Iz polpreteklega časa). <http://www.ijs.si/lit/bizar-na.html-12>, Ljubljana, 1992.
- Gitica Jakopin: *Žarometi*. Mladinska knjiga, Ljubljana 1962.
- Gitica Jakopin: *Devet fantov in eno dekle*. Borec, Ljubljana 1963.
- Gitica Jakopin: *Slovo od deklišтва*. Samozaložba, Ljubljana 1995.
- Josip Jurčič: Nemški valpet: povest. *Slovenski glasnik*, Celovec 1867.
- Josip Jurčič: Sosedov sin: povest. v: *Mladika*, Ljubljana 1868.
- Janko Kersnik: Ponkrčev oča. *Ljubljanski zvon*, 1882.
- Janko Kersnik: Mačkova očeta. *Ljubljanski zvon*, 1886.
- Janko Kersnik: Jara gospoda. *Ljubljanski zvon*, 1893.
- Alojz Kraigher: Peter Drozeg. *Slovenski narod*, 1916.
- Tomo Križnar: *O iskanju ljubezni*. Samozaložba, Ljubljana 1989.
- Tomo Križnar: *Samotne sledi*. Samozaložba, Ljubljana 1993.
- Fran Levstik: Martin Krpan z Vrha. *Slovenski glasnik*, 1858.
- Fran Levstik: Popotovanje iz Litije do Čateža. *Slovenski glasnik*, 1858.
- Fran Levstik: Sveti doktor Bežanec v Tožbanji vasi. *Dunajski zvon*, 1870.
- Milan Lipovec: Čubejska prigoda. *Zaliv*, 1972.
- Florjan Lipuš: *Zmote dijaka Tjaža*. Obzorja, Maribor 1972.
- Florjan Lipuš: *Srčne pege*. Wieser, Celovec 1991.
- Miroslav Malovrh: Opatov praporščak. *Slovenski narod*, 1903.

- Fran Maselj-Podlimbarski: *Gorski potoki*. Anton Knezova knjižnica, Slovenska matica, Ljubljana 1895.
- Ivan Pregelj: *Mlada Breda*. Družba Sv. Mohorja, Slovenske večernice 67, Celovec 1913.
- Ivan Pregelj: Thabiti Kumi. *Sodobnost*, Ljubljana 1933.
- Prežihov Voranc: Boj na požiralniku. *Sodobnost*, Ljubljana 1935.
- Prežihov Voranc: Samorastniki. *Sodobnost*, Ljubljana 1937.
- Marijan Pusavec: *Zbiralec nasmehov*. Samozaložba, Ljubljana 1991.
- Ivan Sivec: *Kruh ponoči spi*. Mohorjeva družba, Celje 1994.
- Ivan Sivec: Triglavski kralj. *Družina*, Ljubljana 1994.
- Jakob Sket: *Miklova Zala*. Družba sv. Mohorja, Slovenske večernice 38, Ljubljana 1884.
- Josip Stritar: Rosana. *Dunajski zvon*, 1877.
- Rudi Šeligo: *Triptih Agate Schwarzkobler*. Obzorja, Maribor 1968.
- Gustav Šilih: *Beli dvor*. Tiskovna založba, Maribor 1938.
- Damijan Šinigoj: *Neizstreljeni naboj*. Slon, zbirka Odmevi časa, Ljubljana 1994.
- Ivan Tavčar: Janez Sonce. *Slovenski narod*, 1885.
- Ivan Tavčar: Otok in struga. v: *Povesti*, zv. 2, samozaložba, Ljubljana 1897.
- Ivan Tavčar: Cvetje v jeseni. *Ljubljanski zvon*, 1917.
- Janez Trdina: Bajke in povesti o Gorjancih. *Ljubljanski zvon*, 1882–1888.
- Ivan Zorec: *Izgnani menihi*. Družba sv. Mohorja, Slov. večernice 90, Celje 1937.

A.2 Prevodi

- Christian Andersen: *Andersenove pravljice*. Prev. in prir. Janko Moder. Mihelač, Ljubljana 1992.
- Lewis Carroll: *Alica v ogledalu*. Prev. Gitica Jakopin. MK, Ljubljana 1978.
- Lewis Carroll: *Aličine prigode v čudežni deželi*. Prev. Gitica Jakopin. MK, Ljubljana 1983.
- Robert Cringely: *Naključni imperiji: kratka zgodovina Silicijeve doline*. Prev. Primož Jakopin. Pasadena, Ljubljana 1995.
- Alexandre Dumas: *Trije mušketirji*. Prev. Tomo Virk po prir. Isabelle Jeannot. Devin, Trst 1995.
- Gustave Flaubert: *Bouvard in Pécuchet*. Prev. Stane Ivanc. Mihelač, Ljubljana 1995.
- Peter Handke: *Ponovitev*. Prev. Silvija Borovnik. Wieser, Celovec 1988.
- Jordan Horowitz: *Kako vzgojiti očeta*. Prev. Mira Hladnik. <http://www.ijs.si/lit/kakovzgo.html-12>, Ljubljana 1994.
- George Mikes: *Priročnik za snobe*. Prev. Mitja Meršol, Mihelač, Ljubljana 1994.
- Libuše Moníková: *Fasada*. Prev. Štefan Vevar. Društvo slovenskih pisateljev – Mihelač, Ljubljana 1994
- George Orwell: *1984*. Prev. Alenka Puhar. Knjižnica Kondor, MK, Ljubljana 1983.
- Platon: *Država*. Prev. Jože Košar. DZS, Ljubljana 1976.
- Sveto pismo Nove zaveze*. Nadškofijski ordinariat iz Ljubljane in Britanska in inozemska biblična družba v Beogradu, Ljubljana 1984.
- Jules Verne: *V osemdesetih dneh okoli sveta*. Prir. Pino Foresti, prev. France Stopar in Mojca Medvedšek. Mihelač, Ljubljana 1996.

Priloga B

Seznam del v drugem vzorcu: zbrana dela Cirila Kosmača

- Božična noč v ječi. *Naš rod* III, 1931, str. 84–85.
- Študenta Petra povest o materi. *Val* I, 1932/33, str. 4–9.
- Potepuh Najdu. *Val* I, 1932/33, str. 29–36.
- V znamenju rojstva in prerojenja. *Istra* IV, 1932, 51/52, str. 2.
- Božično pismo iz celice 589. *Istra* IV, 1932, 51/52, str. 4.
- Gornje mesto. Par besedi ob novi knjigi B. Magajne, *Istra* IV, 1933, 19, str. 2.
- Velika nedelja. *Istra* V, 1933, 14, str. 22.
- Hodil po zemlji sem naši ... *Istra* V, 1933, 14, str. 32.
- Kraška simfonija. Pravljica za božične dni, *Istra* V, 1933, 50, str. 9.
- Cerkovnik Martin. *Ljubljanski zvon* 1933, str. 201–211.
- Na sveti večer. *Istra* VI, 1934, 49/51, str. 18.
- Reka. *Istra* VI, 1934, 21, str. 12–13.
- Hiša št. 14. *Sodobnost* 1934, str. 465–496.
- Ogorek. *Jadranski kalendar* 1935, str. 54–60.
- Obisk. *Jadranski kalendar* 1935, str. 109–113.
- Zločin Bernarda Tula. *Sodobnost* 1935, str. 104–111.
- Prazna ptičnica*. Roman, 1935, Ljubljana, Cankarjeva založba 1988, 139 str.
- Človek na zemlji. Novela, 1935, v: *Sreča in kruh*, Ljubljana, Državna založba Slovenije 1946, 153–208.
- Sreča. Novela, 1936, v: *Sreča in kruh*, Ljubljana, Državna založba Slovenije 1946, 5–23.
- Gosenica. Novela, 1936, v: *Sreča in kruh*, Ljubljana, Državna založba Slovenije 1946, 25–42.
- Kruh. Novela, 1936, v: *Sreča in kruh*, Ljubljana, Državna založba Slovenije 1946, 43–52.
- Življenje in delo Venca Poviškaja. Novela, 1937, v: *Sreča in kruh*, Ljubljana, Državna založba Slovenije 1946, str. 81–151.
15. marec 1937. *Istra* IX, 1937, 18, str. 3.

- Tistega lepega dne. Novela, 1938, v: *Sreča in kruh*, Ljubljana, Državna založba Slovenije 1946, 53–80.
- Zlato. *Sodobnost* 1941, str. 49–54.
- Dragičeva smrt. *Novi svet* 1946, str. 785–791.
- Poletje. *Razgledi* 1947, str. 246–249.
- Očka Orel. Novela, 1947, v: *Iz moje doline*, Ljubljana, Mladinska knjiga 1958, 207–247.
- Težka nedelja. *Novi svet* 1948, str. 253–268.
- Na svoji zemlji*. Scenarij, Ljubljana, Slovenski knjižni zavod 1949, 132 str.
- Beli konj. Odlomek, *Naši razgledi* 1956, 6, str. 144.
- V žagi. *Novi svet* 1950, str. 396–406.
- Hudo je, človek moj. *Novi svet* 1950, str. 482–499.
- Smrt nedolžnega velikana. Novela, 1952, v: *Iz moje doline*, Ljubljana, Mladinska knjiga 1958, 148–206.
- Pot v Tolmin. Novela, 1953, v: *Iz moje doline*, Ljubljana, Mlad. knjiga 1958, 40–79.
- Pomladni dan*. Roman, Ljubljana, Prešernova družba 1953, 175 str.
- Mož, ki ni poznal obupa. *Tovariš IX*, 1953, 16, str. 421.
- Žuželč in njegova krava. *Tovariš IX*, 1953, 34, str. 944–945.
- Prvi prizor. Odlomek iz romana, *Tovariš IX*, 1953, 42, str. 1204–1205.
- Pomlad s čaplo. Literarna priloga *Ljudska pravica - Borba*, 27.6.1953.
- Sredi vasi. Odlomek iz romana, *Ljudska pravica* 1956, 5 (8.1.), 16 (21.1.).
- Kovač in hudič. *Koledar Prešernove družbe za leto 1959*, str. 31–39.
- Medvejke. *Delo* 1959, 1, str. 21.
- Ozimina. *Delo* 1959, 64, str. 6–7.
- Utrinki iz ječe. *Slovenski Jadran* 1959, 50, str. 7.
- Klopotec. *Slovenski izseljenski koledar* 1960, str. 76–80.
- Tantadruj*. Novela, Ljubljana, Cankarjeva založba 1964, 74 str. (IV).
- Balada o trobenti in oblaku*. Novela, Ljubljana, Cankarjeva založba 1964, 144 str. (IV).
- V gaju. Novela, v: *V gaju življenja*, Ljubljana, Mladinska knjiga 1972, str. 5–11.
- Ringaraja. Novela, v: *V gaju življenja*, Ljubljana, Mladinska knjiga 1972, str. 162–173.
- Kamen in njiva. V: *Kamen in njiva, Pravljica o maku*, Trst, Založništvo tržaškega tiska 1984, str. 6–43.
- Pravljica o maku. V: *Kamen in njiva, Pravljica o maku*, Trst, Založništvo tržaškega tiska 1984, str. 44–67.

Priloga C

**Abecedni seznam oblikoslovnih
oznak v drugem vzorcu**

A	25756	GFRap	7	GNme4i	1	GTme1	80	IBme2	3	IPže2	1	IZmp2	7
Aj	370	GFRbd	1	GNme6	1	GTme4	1	IBže5	1	IPže4	2	IZmp3	1
Ajj	153	GFRbe	30	GNmp1	77	GTmp1	8	IMme2	1	IPžp1	1	IZmp4	5
Č	15353	GFRbp	1	GNmp2	3	GTmp2	1	IMme4	1	IS	27	IZmp5	7
ČV	583	GFRcd	1	GNmp4	3	GTmp4	1	IMmp1	1	ISme1	46	IZse1	5
ČZ	3215	GFRce	476	GNmp6	1	GTse1	12	IMse4	1	ISme2	2	IZse2	16
E	235	GFRcp	4	GNse1	120	GTse6	1	IMže1	5	ISme3	2	IZse3	1
E2	5219	GLBmd	40	GNse2	1	GTsp1	13	IMže4	1	ISme4	6	IZse4	35
E3	1923	GLBme	275	GNse4	1	GTže1	32	IMže5	2	ISme5	6	IZse5	53
E4	9605	GLBmp	150	GNse6	2	GTže4	3	IMže6	1	ISmp1	4	IZse6	1
E5	11018	GLBse	29	GNsp	1	GTžp1	15	IMžp1	1	ISmp2	1	IZsp1	8
E6	7553	GLBsp	20	GNsp1	17	GTžp4	1	IMžp2	1	ISmp4	7	IZsp2	7
Gad	89	GLBžd	4	GNsp4	1	GTžp6	1	IMžp4	1	ISmp5	3	IZsp4	21
Gae	2073	GLBže	98	GNžd1	3	GVad	30	INme1	6	ISse1	5	IZsp5	5
Gap	412	GLBžp	86	GNže1	167	GVap	56	INme3	1	ISse2	2	IZže1	133
Gbd	12	GLmd	1045	GNže4	9	GVbd	37	INme4	1	ISse3	1	IZže2	118
Gbe	1074	GLme	24463	GNže5	3	GVbe	987	IOmd1	1	ISse4	1	IZže3	26
GBI	3223	GLmp	4780	GNže6	1	GVbp	293	IOme16626		ISse5	1	IZže4	157
Gbp	281	GLOmd	5	GNžp1	69	GZad	2	IOme2	259	ISže1	40	IZže5	180
Gcd	340	GLOme	138	GNžp4	5	GZae	42	IOme3	540	ISže2	17	IZže6	14
Gce	10700	GLOmp	20	GNžp5	1	GZap	13	IOme4	681	ISže3	1	IZžp1	4
Gcp	2760	GLOsd	1	GNžp6	1	GZbd	1	IOme5	113	ISže4	18	IZžp2	7
GČ	38	GLOse	330	GOad	8	GZbe	18	IOme6	109	ISže5	7	IZžp3	3
GČme1	18	GLOsp	1	GOae	38	GZbp	5	IOmp1	20	ISže6	1	IZžp4	20
GČme2	1	GLOžd	1	GOap	31	GZce	84	IOmp4	13	ISžp1	2	IZžp5	29
GČmp1	1	GLOže	126	GObc	35	GZcp	15	IOmp5	3	ISžp2	2	IZžp6	6
GČmp2	2	GLOžp	9	GObp	1	GZOad	1	IOmp6	1	ISžp3	2	IŽme1	31
GČmp6	1	GLRmd	23	GOcd	8	GZOae	7	IOse6	4	ISžp5	2	IŽme3	1
GČse1	2	GLRme	1282	GOce	749	GZOap	3	IOsp2	1	IVme1	119	IŽme5	1
GČse6	3	GLRmp	140	GOcp	92	GZObe	3	IOže1	1902	IVme2	18	IŽže1	26
GČže4	4	GLRsd	2	GONE	8	GZObp	1	IOže2	101	IVme3	9	IŽže2	1
GČže5	2	GLRse	928	GPad	166	GZOce	274	IOže3	149	IVme4	40	IŽže3	3
GFOad	2	GLRsp	21	GPae	2663	GZOcp	10	IOže4	265	IVme5	4	IŽže4	15
GFOae	13	GLRžd	6	GPap	382	GZPad	10	IOže5	62	IVme6	4	IŽže6	1
GFOap	3	GLRže	768	GPbd	28	GZPae	304	IOže6	94	IVmp1	1	K	3
GFObe	5	GLRžp	77	GPbe	199	GZPap	52	IOžp1	8	IVmp4	1	KI	3
GFObp	1	GLsd	5	GPbp	94	GZPbd	5	IPmd1	3	IVže1	11	KIP	8
GFOce	158	GLse	1826	GPcd	753	GZPbe	49	IPmd4	1	IVže2	1	KP	12
GFOcp	4	GLsp	174	GPce	22615	GZPbp	17	IPme1	53	IVže3	2	M	1565
GFPad	49	GLžd	109	GPcp	3771	GZPcd	57	IPme2	5	IVže4	3	P	1
GFPae	440	GLže	9650	GRad	3	GZPce	1908	IPme3	1	IVže5	2	PČme1	9
GFPap	173	GLžp	981	GRae	64	GZPcp	244	IPme4	10	IZ	1	PČme1i	16
GFPbd	8	Gmp	14631	GRap	66	GZRae	35	IPme6	5	IZmd5	1	PČme2	14
GFPbe	354	GNA	97	GRbe	178	GZRap	9	IPmp1	104	IZme1	83	PČme3	4
GFPbp	73	GNE	2640	GRbp	26	GZRbe	21	IPmp2	9	IZme2	113	PČme4	5
GFPcd	33	GNmd1	7	GRcd	5	GZRbp	8	IPmp3	8	IZme3	21	PČme4i	19
GFPce	1144	GNme1	325	GRce	2303	GZRcd	7	IPmp4	17	IZme4	175	PČme5	15
GFPcp	295	GNme2	1	GRcp	380	GZRce	454	IPmp5	4	IZme5	148	PČme6	43
GFRad	4	GNme3	1	GRNE	109	GZRcp	42	IPmp6	12	IZme6	32	PČmp1	13
GFRae	37	GNme4	3	GTmd1	2	IBme1	8	IPže1	26	IZmp1	1	PČmp2	18

PČmp3	4	PIOmp6	6	PLse2	3	Pmp1j	34	PNže5j	2	PTmp5	2	Pže6	504	Ssd5	8
PČmp4	17	PIOse1	17	PLse3	1	Pmp1jj	8	PNže6	69	PTmp6	4	Pže6j	4	Ssd6	3
PČmp5	4	PIOse2	7	PLse4	4	Pmp2	315	PNžp1	66	PTsd1	1	Pže6jj	3	Sse1	1311
PČmp6	8	PIOse4	13	PLse6	3	Pmp2jj	11	PNžp2	33	PTsd4	1	Pžp1	780	Sse2	1347
PČse1	8	PIOse5	3	PLsp1	2	Pmp3	47	PNžp3	4	PTse1	46	Pžp1j	11	Sse3	171
PČse2	17	PIOse6	3	PLsp4	2	Pmp3j	3	PNžp4	69	PTse2	7	Pžp1jj	4	Sse4	1943
PČse4	6	PIOsp2	2	PLsp5	1	Pmp4	435	PNžp5	30	PTse3	2	Pžp2	450	Sse5	1317
PČse5	22	PIOsp3	1	PLže1	20	Pmp4i	1	PNžp6	76	PTse4	13	Pžp2j	2	Sse6	474
PČse6	6	PIOsp4	2	PLže2	7	Pmp4j	1	Psd1	9	PTse5	7	Pžp2jj	8	Ssp1	357
PČsp1	1	PIOžd1	2	PLže3	1	Pmp4jj	4	Psd4	4	PTse6	7	Pžp3	30	Ssp2	567
PČsp2	5	PIOžd4	1	PLže4	15	Pmp5	220	Psd5	2	PTsp1	10	Pžp4	700	Ssp3	78
PČsp4	2	PIOže1118	PLže5	5	Pmp5j	2	Pse1	1343	PTsp2	1	Pžp4j	6	Ssp4	909	
PČsp5	2	PIOže2	58	PLže6	8	Pmp6	257	Pse1j	27	PTsp4	10	Pžp5	310	Ssp5	361
PČsp6	3	PIOže3	14	PLžp1	6	Pmp6j	4	Pse1jj	10	PTsp5	1	Pžp5j	1	Ssp6	273
PČzd1	1	PIOže4	73	PLžp2	4	Pmp6jj	1	Pse2	645	PTsp6	9	Pžp5jj	5	Szd1	77
PČze1	15	PIOže5	31	PLžp3	1	PNmd1	14	Pse2j	12	PTzd1	2	Pžp6	407	Szd2	4
PČze2	23	PIOže6	28	PLžp4	11	PNmd2	1	Pse2jj	9	PTze1	20	Pžp6j	1	Szd3	2
PČze3	2	PIOžp1	21	PLžp5	3	PNmd6	1	Pse3	38	PTze2	7	S	2	Szd4	101
PČze4	25	PIOžp2	8	PLžp6	5	PNme1181	Pse4	478	PTze3	3	SGse1	141	Szd5	13	
PČze5	21	PIOžp3	2	Pmd1	120	PNme1i	74	Pse4j	3	PTze4	13	SGse2	121	Szd6	88
PČze6	15	PIOžp4	20	Pmd1j	3	PNme1j	2	Pse5	291	PTze5	4	SGse3	23	Sze1	6093
PČžp1	4	PIOžp5	5	Pmd1jj	2	PNme2	59	Pse5j	1	PTze6	6	SGse4	128	Sze2	3839
PČžp2	12	PIOžp6	3	Pmd2	5	PNme3	10	Pse6	201	PTžp1	10	SGse5	76	Sze3	694
PČžp4	15	PISže1	1	Pmd3	1	PNme4	47	Pse6j	3	PTžp2	7	SGse6	60	Sze4	7010
PČžp5	8	PISže2	1	Pmd3j	1	PNme4i	36	Pse6jj	1	PTžp3	1	SGsp1	5	Sze5	3742
PČžp6	16	PISže5	1	Pmd4	7	PNme5	47	Psp1	131	PTžp4	6	SGsp2	8	Sze6	2041
PD	603	PIVme1	1	Pmd4jj	1	PNme6	89	Psp1j	1	PTžp5	7	SGsp4	11	Szp1	1597
PDj	37	PIVse6	1	Pmd5	1	PNmp1139	Psp2	120	PTžp6	12	SGsp5	1	Szp2	1189	
PDjj	23	PIZme1	2	Pmd6	2	PNmp2	50	Psp2j	1	Pzd1	41	SGsp6	4	Szp2jj	1
PDmd	8	PIZme6	1	Pme12706	PNmp3	1	Psp3	11	Pzd1j	1	Smd1	228	Szp3	107	
PDme	211	PIZse1	3	Pmeli	694	PNmp4	51	Psp4	174	Pzd1jj	1	Smd2	15	Szp4	2814
PDmp	38	PIZže1	2	Pmelj	87	PNmp5	22	Psp4j	2	Pzd3	2	Smd3	6	Szp5	1078
PDže	81	PIZže4	3	Pmeljj	43	PNmp6	33	Psp4jj	2	Pzd4	27	Smd4	93	Szp6	921
PDžp	4	PLmd1	2	Pme2	664	PNse1	59	Psp5	84	Pzd4j	1	Smd5	9	Š	80
PDžpj	1	PLmd4	1	Pme2j	8	PNse2	26	Psp5jj	2	Pzd5	3	Smd6	14	ŠG	7
PIBže1	1	PLme1	4	Pme2jj	10	PNse3	4	Psp6	72	Pzd6	10	Sme17830	ŠGmd1109		
PIMže2	1	PLmeli	10	Pme3	122	PNse4	51	Psp6j	1	Pže13073	Sme22996	ŠGmd2	6		
PIOmd1	4	PLme2	5	Pme3jj	2	PNse5	18	PTmd1	1	Pže1j	65	Sme3	756	ŠGmd3	3
PIOmd4	1	PLme3	1	Pme4	684	PNse6	22	PTmd3	1	Pže1jj	37	Sme45734	ŠGmd4	93	
PIOme1307	PLme4	2	Pme4i	345	PNsp1	16	PTme1	22	Pže2	863	Sme52628	ŠGmd5	4		
PIOme2	44	PLme4i	5	Pme4j	15	PNsp2	14	PTme1i10	Pže2j	6	Sme61869	ŠGmd6	12		
PIOme3	38	PLme5	13	Pme4jj	17	PNsp4	13	PTme2	2	Pže2jj	10	Smp12326	ŠGme1	31	
PIOme4	61	PLme6	4	Pme5	620	PNsp5	7	PTme3	1	Pže3	95	Smp21109	ŠGme2	7	
PIOme5	46	PLmp1	7	Pme5j	4	PNsp6	8	PTme4	5	Pže3jj	1	Smp3	273	ŠGme3	1
PIOme6	19	PLmp2	3	Pme5jj	7	PNžd1	6	PTme4i	5	Pže41494	Smp41624	ŠGme4	30		
PIOmp1	23	PLmp3	1	Pme6	707	PNže1	255	PTme5	9	Pže4j	7	Smp5	651	ŠGme5	11
PIOmp2	13	PLmp4	10	Pme6j	6	PNže2	53	PTme6	13	Pže4jj	10	Smp6	575	ŠGme6	11
PIOmp3	9	PLmp5	3	Pme6jj	4	PNže3	11	PTmp1	16	Pže5	845	Ssd1	13	ŠGmp1122	
PIOmp4	7	PLmp6	3	Pmp1	996	PNže4	122	PTmp2	9	Pže5j	1	Ssd2	4	ŠGmp2	8
PIOmp5	23	PLse1	2	Pmp1i	1	PNže5	71	PTmp4	3	Pže5jj	10	Ssd4	31	ŠGmp3	10

ŠGmp4117	ŠVme1142	ZDmp6 10	ZKmp3 17	ZNEme1 47	ZOad3 8	ZOcmp4 581
ŠGmp5 19	ŠVme2 27	ZDse1 4	ZKmp4 71	ZNEme2 33	ZOad4 12	ZOcmp5 26
ŠGmp6 34	ŠVme3 4	ZDse2 36	ZKmp5 41	ZNEme3 2	ZOad6 9	ZOcmp6 140
ŠGsd1 9	ŠVme4 76	ZDse3 1	ZKmp6 12	ZNEme4 17	ZOae1 277	ZOcp4 12
ŠGsd2 3	ŠVme5 33	ZDse4 22	ZKsd5 1	ZNEme5 8	ZOae2 112	ZOcsd4 2
ŠGsd4 30	ŠVme6 8	ZDse5 4	ZKse11262	ZNEme6 23	ZOae3 890	ZOcse2 2
ŠGsd5 7	ŠVmp1 22	ZDse6 4	ZKse2 311	ZNEmp1 34	ZOae4 943	ZOcse4 27
ŠGsd6 2	ŠVmp2 2	ZDsp1 1	ZKse3 53	ZNEmp2 1	ZOae5 45	ZOcse5 2
ŠGse1 9	ŠVmp3 5	ZDsp2 1	ZKse4 385	ZNEmp3 4	ZOae6 108	ZOcse6 4
ŠGse2 3	ŠVmp4 6	ZDsp3 1	ZKse5 191	ZNEmp4 4	ZOamd1 28	ZOcsp2 4
ŠGse4 2	ŠVmp5 3	ZDsp5 1	ZKse6 104	ZNEmp5 2	ZOamp 2	ZOcsp3 4
ŠGse6 1	ŠVmp6 3	ZDsp6 1	ZKsp1 11	ZNEmp6 1	ZOamp1 63	ZOcsp4 39
ŠGsp1 8	ŠVsd1 1	ZDže1 11	ZKsp2 9	ZNEse1 163	ZOamp2 17	ZOcsp5 5
ŠGsp2 5	ŠVsd4 1	ZDže2 10	ZKsp4 4	ZNEse2 17	ZOamp3 170	ZOcsp6 3
ŠGsp4 24	ŠVse1 6	ZDže3 3	ZKsp5 9	ZNEse4 159	ZOamp4 124	ZOcžd3 3
ŠGsp5 25	ŠVse2 29	ZDže4 19	ZKsp6 1	ZNEse5 6	ZOamp5 75	ZOcžd4 5
ŠGsp6 34	ŠVse3 2	ZDže5 5	ZKžd1 9	ZNEse6 11	ZOap1 1	ZOcžd6 3
ŠGžd1 57	ŠVse4 30	ZDže6 2	ZKžd4 4	ZNEsp1 2	ZOap6 34	ZOcže1 25
ŠGžd2 2	ŠVse5 17	ZDžp1 5	ZKžd5 1	ZNEsp4 1	ZOažp1 1	ZOcže2 234
ŠGžd4 69	ŠVse6 2	ZDžp2 12	ZKžd6 2	ZNEsp6 2	ZObd3 2	ZOcže3 647
ŠGžd5 8	ŠVsp2 3	ZDžp4 14	ZKže1 375	ZNEže1 25	ZObd4 1	ZOcže4 1396
ŠGžd6 7	ŠVsp4 4	ZDžp5 5	ZKže2 143	ZNEže2 25	ZObd6 1	ZOcže5 158
ŠGže1 49	ŠVsp5 6	ZDžp6 4	ZKže3 14	ZNEže4 14	ZObe1 501	ZOcže6 132
ŠGže2 6	ŠVžd1 1	ZIme1 8	ZKže4 185	ZNEže5 9	ZObe2 56	ZOcžp2 27
ŠGže3 1	ŠVžd4 1	ZIme2 5	ZKže5 159	ZNEže6 10	ZObe3 296	ZOcžp3 9
ŠGže4 32	ŠVže1 69	ZIme4 30	ZKže6 58	ZNEžp1 6	ZObe4 246	ZOcžp4 285
ŠGže5 8	ŠVže2 23	ZIme5 6	ZKžp1 103	ZNEžp2 1	ZObe5 9	ZOcžp5 16
ŠGže6 4	ŠVže3 4	ZIme6 7	ZKžp2 41	ZNEžp3 1	ZObe6 37	ZOcžp6 8
ŠGžp1 40	ŠVže4 55	ZIse2 2	ZKžp3 7	ZNEžp4 5	ZObmd 14	ZOP2 174
ŠGžp2 12	ŠVže5 28	ZIse4 3	ZKžp4 97	ZNEžp5 1	ZObp1 64	ZOP3 143
ŠGžp3 2	ŠVže6 12	ZIse5 6	ZKžp5 48	ZNEžp6 1	ZObp2 7	ZOP4 224
ŠGžp4 73	ŠVžp1 12	ZIse6 1	ZKžp6 23	ZNI 780	ZObp3 104	ZOP5 78
ŠGžp5 29	ŠVžp2 5	ZIsp6 1	ZM 6	ZNIme1 146	ZObp4 60	ZOP6 142
ŠGžp6 17	ŠVžp4 9	ZIže1 5	ZMme1 9	ZNIme2 60	ZObp5 8	ZPO 31
ŠG1 96	ŠVžp5 3	ZIže2 3	ZMme2 1	ZNIme3 21	ZObp6 8	ZPOme1132
ŠG2 1	ŠVžp6 1	ZIže4 6	ZMme3 3	ZNIme4 8	ZObžp1 1	ZPOme2 4
ŠG4 432	Vpo 11565	ZIže5 5	ZMme4 1	ZNIme5 1	ZOcmd1 2	ZPOme3 2
ŠLmp5 1	Vpr 28529	ZIžp4 4	ZMme5 1	ZNIme6 5	ZOcmd2 9	ZPOme4 17
ŠLmp6 1	ZDmd1 2	ZK 986	ZMme6 1	ZNImp1 1	ZOcmd3 42	ZPOme5 4
ŠLsp2 1	ZDme1 66	ZKmd1 17	ZMmp2 1	ZNIse1 141	ZOcmd4 39	ZPOme6 4
ŠLžp1 1	ZDme2 14	ZKmd4 4	ZMse1 10	ZNIse2 48	ZOcmd6 27	ZPOmp1 3
ŠLžp3 1	ZDme3 27	ZKmd5 1	ZMse2 2	ZNIse4 69	ZOcme1 167	ZPOmp2 6
ŠL1 9	ZDme4 35	ZKme1430	ZMse3 1	ZNIse5 2	ZOcme2 223	ZPOmp4 7
ŠL4 13	ZDme5 11	ZKme2119	ZMse4 21	ZNIse6 1	ZOcme32119	ZPOse1 61
ŠMme4 1	ZDme6 12	ZKme3 28	ZMse5 2	ZNIže1 13	ZOcme43338	ZPOse2 15
ŠMse6 1	ZDmp1 45	ZKme4184	ZMsp4 1	ZNIže2 80	ZOcme5 166	ZPOse4 149
ŠMsp6 2	ZDmp2 9	ZKme5162	ZMže1 4	ZNIže4 4	ZOcme6 408	ZPOse6 2
ŠNE 501	ZDmp3 16	ZKme6 63	ZMže2 1	ZNIže5 1	ZOcmp1 7	ZPOsp2 3
ŠR 37	ZDmp4 22	ZKmp1136	ZMže4 1	ZNIžp2 5	ZOcmp2 93	ZPOže1 30
ŠVmd1 2	ZDmp5 3	ZKmp2 52	ZNE 343	ZOad2 5	ZOcmp3 273	ZPOže2 5

ZPOže3	2	ZRPOme5	1	ZSVaeže5	15	ZSVbese1	6	ZSVcmese6	7	ZSVčzemp117	
ZPOže4	11	ZRPOme6	2	ZSVaeže6	1	ZSVbese2	2	ZSVcmesp1	6	ZSVčzemp2	3
ZPOžp1	4	ZRPOse1	3	ZSVaežp1	11	ZSVbese4	4	ZSVcmesp2	6	ZSVčzemp4	7
ZPOžp2	2	ZRPOse4	5	ZSVaežp2	8	ZSVbesp1	7	ZSVcmesp4	3	ZSVčzemp5	4
ZPOžp3	1	ZRPOse5	1	ZSVaežp3	1	ZSVbesp4	3	ZSVcmesp5	3	ZSVčzemp6	1
ZPOžp4	4	ZRse2	21	ZSVaežp4	12	ZSVbesp5	2	ZSVcmeže1133		ZSVčze1	26
ZPUmd1	12	ZRse3	5	ZSVaežp5	9	ZSVbeže1	42	ZSVcmeže2	38	ZSVčze2	7
ZPUme1443		ZRse4	4	ZSVaežp6	3	ZSVbeže2	5	ZSVcmeže3	2	ZSVčze3	1
ZPUme2	15	ZRse5	7	ZSVapmd1	1	ZSVbeže3	1	ZSVcmeže4	51	ZSVčze4	13
ZPUme3	11	ZRse6	6	ZSVapmd3	1	ZSVbeže4	6	ZSVcmeže5	34	ZSVčze5	1
ZPUme4	37	ZRsp5	2	ZSVapme1	60	ZSVbeže5	2	ZSVcmeže6	12	ZSVčze6	2
ZPUme5	13	ZRže1	8	ZSVapme2	11	ZSVbeže6	1	ZSVcmežp1	35	ZSVčzesp1	3
ZPUme6	17	ZRže2	11	ZSVapme3	5	ZSVbežp1	7	ZSVcmežp2	16	ZSVčzesp2	4
ZPUmp1	51	ZRže4	11	ZSVapme4	21	ZSVbpmd3	1	ZSVcmežp3	3	ZSVčzesp5	2
ZPUmp2	2	ZRže5	17	ZSVapme5	7	ZSVbpme1	5	ZSVcmežp4	34	ZSVčžežd4	1
ZPUmp4	3	ZRže6	4	ZSVapme6	5	ZSVbpme5	1	ZSVcmežp5	25	ZSVčžež1	36
ZPUse1	19	ZRžp1	3	ZSVapmp1	22	ZSVbpmp1	1	ZSVcmežp6	4	ZSVčžež2	18
ZPUse2	10	ZRžp2	4	ZSVapmp2	8	ZSVbpse1	3	ZSVcpme1	6	ZSVčžež3	5
ZPUse4	11	ZRžp4	6	ZSVapmp3	5	ZSVbpse5	1	ZSVcpme2	1	ZSVčžež4	16
ZPUse6	1	ZRžp5	11	ZSVapmp4	10	ZSVbpže1	17	ZSVcpme6	3	ZSVčžež5	12
ZPUsp1	3	ZRžp6	3	ZSVapmp5	7	ZSVbpže2	2	ZSVcpmp1	9	ZSVčžež6	2
ZPUsp2	1	ZSVadmp4	1	ZSVapmp6	3	ZSVbpže3	1	ZSVcpmp4	5	ZSVčžežp1	20
ZPUžd1	2	ZSVadse1	1	ZSVapse1	6	ZSVbpže5	1	ZSVcpmp5	1	ZSVčžežp2	14
ZPUže1	165	ZSVadže1	1	ZSVapse2	9	ZSVbpžp1	1	ZSVcpmp6	3	ZSVčžežp3	2
ZPUže2	21	ZSVAemd1	5	ZSVapse3	2	ZSVcdmd5	1	ZSVcpse1	3	ZSVčžežp4	10
ZPUže3	1	ZSVAeme1	155	ZSVapse4	3	ZSVcdme1	1	ZSVcpse2	2	ZSVčžežp5	10
ZPUže4	13	ZSVAeme2	17	ZSVapse5	7	ZSVcdme4	1	ZSVcpse3	1	ZSVčžežp6	7
ZPUže5	14	ZSVAeme3	7	ZSVapsp1	5	ZSVcdsd1	1	ZSVcpse4	5	ZSVPmd4	3
ZPUžp1	10	ZSVAeme4	10	ZSVapsp2	2	ZSVcdsd4	1	ZSVcpse5	1	ZSVPme1	10
ZPUžp2	4	ZSVAeme5	19	ZSVapsp4	2	ZSVcdse4	1	ZSVcsp1	1	ZSVPme2	76
ZPUžp4	6	ZSVAeme6	9	ZSVapsp5	5	ZSVcdže2	1	ZSVcsp2	3	ZSVPme3	30
ZPUžp5	3	ZSVAemp1	21	ZSVapže1	76	ZSVcdžp5	1	ZSVcsp4	3	ZSVPme4	284
ZR	444	ZSVAemp2	7	ZSVapže2	27	ZSVcmemd1	2	ZSVcsp5	3	ZSVPme5	60
ZRmd1	1	ZSVAemp4	7	ZSVapže3	5	ZSVcmemd4	1	ZSVcpže1	11	ZSVPme6	82
ZRme1	49	ZSVAemp5	2	ZSVapže4	22	ZSVcmeme1	109	ZSVcpže2	2	ZSVPmp1	2
ZRme2	18	ZSVAese1	48	ZSVapže5	32	ZSVcmeme2	21	ZSVcpže3	2	ZSVPmp2	50
ZRme3	1	ZSVAese2	13	ZSVapže6	5	ZSVcmeme3	4	ZSVcpže5	1	ZSVPmp3	25
ZRme4	8	ZSVAese3	1	ZSVapžp1	10	ZSVcmeme4	28	ZSVcpže6	1	ZSVPmp4	98
ZRme5	22	ZSVAese4	17	ZSVapžp2	2	ZSVcmeme5	29	ZSVcpžp1	4	ZSVPmp5	23
ZRme6	17	ZSVAese5	8	ZSVapžp3	1	ZSVcmeme6	14	ZSVcpžp2	2	ZSVPmp6	31
ZRmp1	3	ZSVAese6	7	ZSVapžp4	13	ZSVcmemp1	41	ZSVcpžp4	4	ZSVPse1	2
ZRmp2	5	ZSVAesp1	5	ZSVapžp5	2	ZSVcmemp2	5	ZSVcpžp5	1	ZSVPse2	63
ZRmp3	2	ZSVAesp2	4	ZSVbdže1	2	ZSVcmemp3	2	ZSVcpžp6	1	ZSVPse3	8
ZRmp4	10	ZSVAesp4	2	ZSVbeme1	26	ZSVcmemp4	13	ZSVčzemd1	3	ZSVPse4	145
ZRmp5	12	ZSVAesp5	5	ZSVbeme4	5	ZSVcmemp6	5	ZSVčzeme1	63	ZSVPse5	37
ZRmp6	6	ZSVAesp6	1	ZSVbeme5	2	ZSVcmese1	40	ZSVčzeme2	23	ZSVPse6	35
ZRPO	14	ZSVaeže1	86	ZSVbeme6	2	ZSVcmese2	23	ZSVčzeme3	3	ZSVPsp2	10
ZRPOme1	3	ZSVaeže2	27	ZSVbemp1	6	ZSVcmese3	5	ZSVčzeme4	30	ZSVPsp3	1
ZRPOme3	1	ZSVaeže3	2	ZSVbemp2	2	ZSVcmese4	39	ZSVčzeme5	21	ZSVPsp4	17
ZRPOme4	2	ZSVaeže4	30	ZSVbemp4	2	ZSVcmese5	20	ZSVčzeme6	8	ZSVPsp5	14

ZSVPsp6	9	ZSVPžp673	ZTmp2	40	ZTsp2	10	ZTže4	151	ZVme4	16	ZVse5	13	
ZSVPžd3	2	ZT	59	ZTmp3	41	ZTsp3	1	ZTže5	38	ZVme5	5	ZVse6	9
ZSVPžd4	7	ZTmd1	75	ZTmp4	125	ZTsp4	27	ZTže6	49	ZVme6	11	ZVsp1	1
ZSVPžd6	1	ZTmd2	8	ZTmp5	10	ZTsp5	3	ZTžp1	92	ZVmp1	6	ZVsp4	1
ZSVPže1	1	ZTmd3	6	ZTmp6	17	ZTsp6	3	ZTžp2	57	ZVmp2	1	ZVsp6	1
ZSVPže2	158	ZTmd4	12	ZTsd4	1	ZTžd1	6	ZTžp3	3	ZVmp3	1	ZVše1	40
ZSVPže3	26	ZTmd5	9	ZTsd5	2	ZTžd2	3	ZTžp4	134	ZVmp4	1	ZVže2	2
ZSVPže4	394	ZTme1	254	ZTse1	556	ZTžd3	3	ZTžp5	31	ZVmp5	1	ZVže3	1
ZSVPže5	134	ZTme2	25	ZTse2	87	ZTžd4	13	ZTžp6	20	ZVmp6	1	ZVže4	17
ZSVPže6	72	ZTme3	22	ZTse3	17	ZTžd5	5	ZV	1888	ZVR	3224	ZVže5	6
ZSVPžp3	34	ZTme4	263	ZTse4	432	ZTžd6	61	ZVmd1	1	ZVse1	1314	ZVže6	4
ZSVPžp3	9	ZTme5	32	ZTse5	53	ZTže1	1217	ZVme1	1222	ZVse2	5	ZVžp1	6
ZSVPžp4	163	ZTme6	10	ZTse6	32	ZTže2	28	ZVme2	3	ZVse3	8	ZVžp3	1
ZSVPžp5	35	ZTmp1	475	ZTsp1	23	ZTže3	12	ZVme3	9	ZVse4	268	ZVžp4	6

Priloga Č

Abecedni seznam pogostejših lem, njihovih oblikoslovnih oznak in frekvenc v drugem vzorcu.

Navedenih je 3.554 lem s frekvenco vsaj 10; popoln seznam je na voljo na internetu, na naslovu: <http://bos.zrc-sazu.si/ckb.html>. Z *biti p*, *biti r* in *biti o* so označene tri komponente glagola *biti* – pomožni glagol (npr. Janez *je* videl volka.), glagol *biti* v relacijskem smislu (npr. Jaz *sem* lepa.) in glagol *biti* v ožjem smislu obstajanja (npr. Hiša *je* tam.)

- a · Č,128; M,16; Vpr,959
aaa · M,11
ah · M,56
aha · M,66
ajda · Šze1,6; Šze2,17; Šze4,3; Šze5,4
Albin · IOme1,19; IOme3,3; IOme6
ali · ČV,582; Vpr,365
amen · M,11
Amerika · IZže1,2; IZže2,4; IZže4,3; IZže5,2
ampak · Vpr,134
Ana · IOže1,9; IOže2,2; IOže4,2; IOže6
Anca · IOže1,16; IOže3,4; IOže4; IOže5; IOže6
Andreich · IOme1,13; IOme2; IOmp4
Andrej · IOme1; IOme1,16; IOme3,2; IOme4,2;
IOme5
Andrejč · IOme1; IOme1,15; IOme4,2
angel · Sme1,6; Smp1,4
Angela · IOže1,14; IOže3,5; IOže4; IOže5,3
Angelca · IOže1,40; IOže3,4; IOže4,8; IOže6,3
Angelčin · PIOme5; PIOme6; PIOže3,3;
PIOže6,4; PIOžp4
angleški · Pme6; Pmp3; Pže4,6; Pže6; Pžp6
aretirati · Gcp; GLme,4; GLmp,5
Arnac · IOme1,9; IOme3,2; IOme4
ata · Sme1,22
Augusta · IOže1,9; IOže4
avstrijski · Pme1; Pme2; Pme4; Pme5; Pme6;
Pžd4; Pže1; Pže4; Pže5; Pže6,2
avtomobil · Smd1,2; Sme1,16; Sme2,12; Sme3,4;
Sme4,7; Sme5,3; Sme6,5; Smp1,6; Smp2;
Smp4; Smp6,3
baba · Šzd1,2; Šze1,52; Šze2,2; Šze3; Šze4,5;
Šzp1,29; Šzp2,2; Šzp4,7; Šzp5
babica · Šzd1; Šze1,5; Šze2,2; Šze4,2
babji · Pme1,3; Pse1; Pse2; Pse4; Pžel; Pžp4,3
babnica · Šze1,8; Šze2; Šze4; Šzp1,2
Bača · IZže1,7; IZže2,3; IZže4; IZže5,3
baje · Č,23
bajta · Šzd1,2; Šze1,18; Šze2,23; Šze3,14;
Šze4,25; Šze5,26; Šze6,22; Šzp1; Šzp2,3; Šzp4;
Šzp5,2
bajtar · Smd1; Sme1,20; Sme3,3; Sme4,3;
Smp1,8; Smp2; Smp3,2; Smp6
bajtarski · Pme1,3; Pmp3; Pmp4; Pže1,2; Pže4;
Pžp4; Pžp5
baraba · Šze1,7; Šze4; Šzp1; Šzp6
baraka · Šze1; Šze2,2; Šze3; Šze4; Šze5; Šzp1,5;
Šzp2,2; Šzp4,2; Šzp6,3
Barbara · IOže1,22; IOže2,2; IOže3,3; IOže4,5;
IOže5,2; IOže6
Barbaron · IOme1,11; IOme2,2; IOme4; IOme4;
IOmp6
barva · Šze1,2; Šze2,11; Šze4,5; Šze5; Šze6,3;
Šzp1,2; Šzp2; Šzp4,3; Šzp5,2
bas · Smd1; Sme1,7; Sme5,2; Sme6,2
bati se · Gae,25; Gbe,14; Gce,5; Gcp,5; Gcp,6;
GLmd; GLme,50; GLmp,9; GLže,15; GLžp;
GNE; GVbe,11; GVbp
bedak · Smd1; Smd6; Sme1,7; Sme4,2; Smp1,2;
Smp4
beden · Pme1i; Pme1,2; Pmp1,2; Pse1; Pse2,2;
Pže1; Pže2; Pžp2
beg · Sme1; Sme4,3; Sme5,6
bel · Pme1i,4; Pme1,26; Pme2,12; Pme4;
Pme4i,7; Pme4,9; Pme5,27; Pme6,4;
Pmp1,16; Pmp2,3; Pmp4,11; Pmp5,2;
Pmp6,10; Pse1,7; Pse2,5; Pse4,3; Pse5,3;
Pse6; Psp1; Psp4,2; Psp5; Psp6; Pžd1; Pžd6,2;
Pže1,53; Pže2,16; Pže3,2; Pže4,24; Pže5,27;
Pže6,9; Pžp1,29; Pžp2,6; Pžp3,2; Pžp4,18;
Pžp5,9; Pžp6,4
beliti · Gce,3; GLme,2; GLmp,2; GLže,6
belo · A,8; Sse5,2
belogardist · Smd2; Sme1,6; Sme4; Sme6;
Smp1,3; Smp2,2; Smp3; Smp4,5; Smp6,3
berač · Sme1,13; Sme4; Sme4; Sme6; Smp1,5;
Smp2,3; Smp4,6; Smp5,2
Bernard · IOme1,63; IOme2,3; IOme3,2;
IOme4,8; IOme5,3; IOme6,2
bes · Sme1; Sme2,5; Sme5; Sme6,4
beseda · Šzd1,2; Šzd4,2; Šze1,31; Šze2,121;
Šze3,7; Šze4,77; Šze5,11; Šze6,18; Šzp1,64;
Šzp2,10; Šzp2,40; Šzp3,5; Šzp4,75; Šzp5,20;
Šzp6,30
besno · A,29
bevskniti · Gce,3; GLme,10; GLže,8
bežati · Gae; Gap; Gbe; Gce,5; Gcp; GLme,6;
GLmp,2; GLse; GLsp,3; GLže,2; GNE,3;
GVbd; GVbe,13
Bichi · IOme1,72; IOme2,2; IOme3,8; IOme4,15;
IOme6
bič · Sme1,20; Sme2,2; Sme3,6; Sme4,24;
Sme5,2; Sme6,7; Smp1,5; Smp2,2; Smp4,9;
Smp5; Smp6
bik · Sme1,6; Sme3,3; Sme4; Smp6
bilka · Šze2,3; Šze4,4; Šze5; Šze6; Šzp1; Šzp4;
Šzp5
birič · Sme1,2; Sme4,2; Smp1,3; Smp2,2; Smp4
biser · Sme5; Smp1,2; Smp2,5; Smp4; Smp5;
Smp6
bister · Pme1,2; Pme2,3; Pme3; Pmp1; Pže1,7;
Pže2,7; Pže4; Pže5,2; Pžp1,2; Pžp2; Pžp4;
Pžp6

- biti · Gce,11; Gcp; GLme; GLmp,5; GLse; GLže;
GLže,4; GNE,5
- biti o · GFOad,2; GFOae,13; GFOap,3;
GFObe,5; GFObp; GFOce; GFOce,157;
GFOcp,4; GLOmd,5; GLOme,138; GLOmp,20;
GLOsd; GLOse,330; GLOsp; GLOzd;
GLOže,126; GLOžp,9; GOad,8; GOae,38;
GOap,31; GObe,35; GObp; GOcd,8; GOce,7;
GOce,742; GOcp,92; GONE,8
- biti p · GBI,3223; GFPad,49; GFPae,440;
GFPap,173; GFPbd,8; GFPbe,354; GFPbp;
GFPbp,72; GFPcd,33; GFPce,1144;
GFPcp,295; GLBmd,40; GLBme,275;
GLBmp,150; GLBse,29; GLBsp,20; GLBzd,4;
GLBže,98; GLBžp,86; GPad,166; GPae,2663;
GPap,382; GPbd,28; GPbe,199; GPbp,94;
GPcd,753; GPce,22615; GPCp,3771
- biti r · GFRad,4; GFRae,37; GFRap,7; GFRbd;
GFRbe,30; GFRbp; GFRcd; GFRce,476;
GFRcp,4; GLRmd,23; GLRme; GLRme,1281;
GLRmp,140; GLRsd,2; GLRse,928;
GLRsp,21; GLRzd,6; GLRže,768; GLRžp,77;
GRad,3; GRae,64; GRap,66; GRbe,178;
GRbp,26; GRcd,5; GRce,2303; GRcp,380;
GRNE,109; GVbd,3; GVbe,35; GVbp,6
- bitje · Ssd1,2; Sse1,13; Sse2,3; Sse3,2; Sse4,4;
Ssp1,9; Ssp2,3; Ssp4
- bitka · Sže1,8; Sže2; Sže4,2; Sže5,4; Sže6,3;
Sžp1,2; Sžp2; Sžp4,2
- bivši · Pmd2; Pme1,5; Pme4,2; Pmp4; Pže1
- blag · Pme1,2; Pme2,3; Pme3; Pme6,4; Pmp1;
Pmp5,2; Pse1; Pže1; Pže2; Pžp2
- blago · A,7; Sse2,9; Sse4,2; Sse6,2
- blagoslov · Sme1,7; Sme2,2; Sme4,2; Smp6
- blaten · Pmd1; Pme1,2; Pme2; Pme5,4; Pmp5;
Pse4; Psp2; Pže1,2; Pže2; Pže4,4; Pžp2
- blato · Sme2,2; Sse1,5; Sse4,8; Sse5,4; Sse6,6
- blazina · Sže1; Sže2,3; Sže4,15; Sže5; Sže6;
Sžp5
- blažič · IOme1,24; IOme2,2; IOme3; IOme4;
IOme5,2; IOme6
- blažičev · PIOme1; PIOmp4,4; PIOže1,2;
PIOže4,3
- bled · Pme1i; Pme1,24; Pme2; Pme4; Pme4;
Pme4i; Pme5,3; Pmp1; Pmp4; Pse1,4; Pse5;
Psp1; Psp2,2; Pže1,11; Pže5; Pžp1,2; Pžp2,3;
Pžp3; Pžp6,2
- bliskovit · Pže1; Pže6,12
- bliskovito · A,27
- blizu · Aj,54; A,70; E2,17
- bližati se · Gad; Gcd; Gce,24; Gcp,14; GLmd,2;
GLme,16; GLmp,7; GLse; GLže,15; GLžp
- bližina · Sže4,5; Sže5,14
- bluza · Sže1,3; Sže4,6; Sže5,6
- bodrilno · A,15
- Bog · IVme1,46; IVme2,6; IVme3,6; IVme4,29;
IVme5; IVme6,2
- bog · Sme1,70; Sme2,9; Sme3,21; Sme4,28;
Sme5; Sme6,2; Smp1
- bogastvo · Sse1,3; Sse2,3; Sse3; Sse4,7; Sse5,2;
Sse6,3
- bogat · Pme1jj; Pme1,5; Pme6; Pmp1j; Pmp4,2;
Pse1; Pse4; Pse6; Psp1; Pže1; Pže2,2; Pže4,2
- bogve · M,26
- boj · Sme1,15; Sme2,6; Sme3; Sme4,13; Sme5,8;
Smp2; Smp4,2; Smp5
- bojen · Pme1i,3; Pse1; Pse2; Pže1,2; Pže4;
Pže5,3
- bok · Sme2; Sme4,6; Sme5; Smp2,2; Smp4,10;
Smp5,2; Smp6
- bolan · Pme1,19; Pse4; Pže1; Pže1,7; Pže3;
Pže4; Pžp3
- bolečina · Sže1,26; Sže2,10; Sže4,13; Sže5,5;
Sže6,4; Sžp1,2; Sžp2; Sžp4,5; Sžp5; Sžp6
- boleti · Gce,10; Gcp; GLme,2; GLmp; GLse,10;
GNE
- bolezen · Sže1,7; Sže2,3; Sže4; Sže5,3; Sže6;
Sžp1; Sžp2,5; Sžp4; Sžp6
- bolj · Ajj,68; A,352
- bolničar · Smd1,8; Sme1; Smp1,2; Smp2
- bomba · Sžd4,3; Sže1; Sže2,2; Sže4,6; Sžp1;
Sžp4,3; Sžp6,2
- bor · Sme2; Smp1,14; Smp4; Smp6,7
- borba · Sže1,3; Sže2,4; Sže4; Sže5,2
- borec · Sme1,2; Sme2,4; Smp1; Smp2; Smp4;
Smp6,2
- Boris · IOme1,145; IOme2,3; IOme3,16;
IOme4,15; IOme5
- boriti se · Gap; Gcp; Gcp,2; GLme,5; GLmp,6;
GLže,2; GNE,2
- borovec · Smp1,3; Smp2,2; Smp4,2; Smp5;
Smp6,3
- bos · Pme1,8; Pme2; Pme4; Pmp1,6; Psp4,2;
Psp6; Pže1,3; Pže4; Pže6,7; Pžp2,5; Pžp4,3;
Pžp5; Pžp6,8
- Bosanec · IPmd1,2; IPme2; IPmp1,8; IPmp3;
IPmp4; IPmp6
- Bosna · IZže2,4; IZže4,4; IZže5,8
- botrček · Sme1,10; Sme4,2
- Božena · IOže1,30; IOže3; IOže4,3; IOže5,2;
IOže6,4
- božič · Sme1,4; Sme2; Sme4,5; Sme5; Sme6,3
- božičen · Pme4i,4; Pmp6; Pse1,2; Pse4,4;
Pse5,2; Psp2; Pže4,3

- božji · Pme1,26; Pme2,3; Pme3; Pme4,5;
Pme5,8; Pme6,2; Pmp1,10; Pmp2,4; Pmp4;
Pmp6; Pse1,4; Pse2,2; Pse4; Pse5,2; Pže1,28;
Pže2,10; Pže3; Pže4,26; Pže5,6; Pže6,3; Pžp1;
Pžp2; Pžp5
- božorno-boserna · M,27
- brada · Sže1,13; Sže2,4; Sže4,24; Sže5,20;
Sže6,3; Sžp6
- brajda · Sže1,3; Sže4; Sže5,4; Sžp1; Sžp2,2;
Sžp4,3; Sžp5,3; Sžp6,3
- braniti · Gap,5; Gbe; Gce,5; Gcp; GLme,7;
GLmp,5; GLže,8; GNA; GNE,3
- brat · Smd1,2; Sme1,18; Sme2; Sme3,3; Sme4,7;
Smp1; Smp1,12; Smp2,6; Smp3,4; Smp4,5;
Smp5,2
- bratec · Sme1,5; Sme2; Sme4,3; Smp1,3; Smp2;
Smp4,2
- brati · Gae,3; Gap,4; Gce,20; Gcp,5; GLme,36;
GLmp,3; GLže,16; GNE,6; GVbe; GVbp
- bravo · M,11
- brazda · Sže1,2; Sže4,2; Sžp1,3; Sžp2,3; Sžp3;
Sžp4,6; Sžp5,5; Sžp6,2
- brdo · Sse2,6; Sse4,2; Sse5,4; Ssp1,3; Ssp2;
Ssp4,4; Ssp5
- breg · Smd2; Sme2,13; Sme3,7; Sme4,5;
Sme5,37; Sme6; Smp4; Smp5,4
- breme · Sse1,4; Sse4,11; Sse6,6; Ssp1,2; Ssp2;
Ssp4,4; Ssp6
- brenčati · Gce,3; Gcp; GLmd,2; GLme; GLmp,3;
GLse; GLže,2; GLžp,3
- brenkati · Gce; Gcp; GLmd; GLme,2; GLmp,2;
GNE,3
- brez · E2,302
- brezmejen · Pme1,3; Pme2; Pme6; Pse4; Pse6,2;
Pže4,2; Pžp2,2
- brezposelen · Smp1,10; Smp2
- bridek · Pme1,3; Pme1,2; Pme2j; Pme2,3;
Pme4; Pme4i; Pme5; Pme6; Pmp1,2; Pmp2;
Pmp5; Pse1,4; Pse2; Pse6; Psp2; Pžd6;
Pže1,18; Pže2; Pže4,3; Pže5,3; Pže6,3;
Pžp1,4; Pžp2,2; Pžp4; Pžp5; Pžp6
- bridko · A,33
- bridkost · Sže1,21; Sže2,12; Sže4,3; Sže5,5;
Sže6,10; Sžp1; Sžp2,2; Sžp4,5; Sžp5; Sžp6
- brigada · Sže1,14; Sže2; Sže4; Sže6; Sžp1,3
- brigadir · Sme1,28; Sme2,2; Sme3; Sme4,5;
Sme5; Sme6; Smp1
- brigati · Gae; Gce,8; Gcp,2; GLme,2; GLže
Brika · IPže1,21
- brinje · Sse1,9; Sse2,2; Sse4; Sse5; Sse6,2
- brisati · Gae; Gbe; Gce,15; Gcp,3; GLmd,2;
GLme,10; GLmp,4; GLže,10; GLžp; GNE,7
- briti · Gae; Gce,5; Gcp; GLme,11; GLže,2; GNE
- brk · Smp1,11; Smp4,22; Smp5; Smp6,9
- brka · Sžp4,11
- brnenje · SGse1,3; SGse2; SGse4,5; SGse5
- brneti · Gce,4; Gcp,3; GLme; GLmp,2; GLže,9;
GLžp
- bronast · Pme1; Psp4; Pže1; Pže2; Pže4; Pže5,2;
Pže6,2; Pžp1; Pžp2
- brskati · Gce; Gcp; GLme,4; GLmp,3; GLže;
GNE,3
- bruhniti · Gce,3; GLme,3; GLme,4
- brv · Sže2,3; Sže3,2; Sže4; Sže5,4; Sže6,2;
Sžp4,6
- brzostrelka · Sžd4; Sže1,9; Sže2,2; Sže4,27;
Sže6,7; Sžp1,5; Sžp2,2; Sžp4,3; Sžp6,4
- buča · Sže1,6; Sže4,2; Sžp4; Sžp5
- bučati · Gce,3; GLse; GLže,3; GLžp,2; GVbe,4
- bujen · Pmp4,2; Pmp5; Pse1; Pže1,2; Pže2;
Pže4; Pže5; Pžp1
- buljiti · Gce,5; Gcp; GLme,6; GLmp,2; GLže,3
- bunker · Sme1,5; Sme2,5; Sme4; Sme5,3;
Sme6,2
- burja · Sže1,32; Sže2,2; Sže3; Sže4,2; Sže5,3
- burno · Aj,2; A,12
- butati · Gce,4; Gcp; GLme; GLmp,4; GLse;
GLže,3; GLžp,2; GNE
- butniti · Gce,3; GLme,16; GLmp,2; GLse; GLže;
GLžp
- bzikniti · Gce,2; GLme,11
- capa · Sžp4,8; Sžp5,2; Sžp6
- caposquadra · Sme1,61; Sme2; Sme3; Sme4
- Carlo · IOme1,9; IOme2,3
- Carniglia · IOme1,10; IOme2,2
- cel · Pmd1; Pme1,7; Pme4; Pme5,2; Pme6,2;
Pmp4,4; Psd4; Pse1; Pse2; Psp2; Psp4,2;
Pže1,19; Pže2; Pže4,20; Pže6; Pžp4
- celica · Sže1,5; Sže2,23; Sže4,19; Sže5,28;
Sže6,2; Sžp2; Sžp4; Sžp5,4
- celo · A,135; Č,108; Sse5
- cement · Sme1,2; Sme2,2; Sme4; Sme5,3;
Sme6,2
- cena · Sže2; Sže4,7; Sže5; Sžp5
- Cene · IOme1,39; IOme2,2; IOme3,2; IOme4,2;
IOme6
- cerkev · Sže1,13; Sže2,14; Sže3,17; Sže4,19;
Sže5,32; Sže6,16; Sžp2; Sžp2,2; Sžp4,2; Sžp5;
Sžp6
- cerkovnik · Pme1,2; Sme1,10; Sme4; Sme6,2;
Smp2
- cerkven · Pme4; Pmp2; Pse2; Pse5,4; Pse6;
Psp3; Psp4; Pže1,2; Pže4,3; Pžp2; Pžp4,2;
Pžp5,2

- cesar · Sme1,13; Sme2; Sme3,5; Sme4,5; Sme5,2
 Cesare · IOme1,16
 cesta · Šzd1; Šzd2; Šze1,44; Šze2,37; Šze3,4;
 Šze4,85; Šze5,139; Šze6,14; Šzp1,3; Šzp2,3;
 Šzp3,2; Šzp4,3; Šzp5,22; Šzp6
 Cestar · IOme1,35; IOme2; IOme3,7; IOme4,4;
 IOme5,3; IOme6
 cesten · Pme2,2; Pme4i,4; Pme5,3; Pmp5,2;
 Pse5; Pže1; Pže2,2; Pže4
 cev · Šze1,2; Šze2,4; Šze4,6; Šze5,2; Šze6,3;
 Šzp2; Šzp4,2
 Chriznich · IOme1,13; IOme2; IOme3; IOmp4
 cigara · Šze1; Šze2,2; Šze4,6; Šze5; Šzp4
 cigareta · Šze1; Šze2,8; Šze4,5i; Šze5,3; Šze6,2;
 Šzp2,4; Šzp4,13
 cigaretan · Pme1i,7; Pme2,2; Pme4i,4; Pme5,2;
 Pme6; Pmp1,3; Pmp4; Pmp6
 Cika · IŽze1,20; IŽze3,3; IŽze4,4
 cilj · Sme1,3; Sme2; Sme3; Sme4,3; Sme6,2;
 Smp1; Smp2,2; Smp3,3; Smp5
 cipresa · Šze1; Šzp1,6; Šzp2; Šzp4,4; Šzp6,6
 cula · Šzd4; Šze1,4; Šze2; Šze3; Šze4,15; Šze5;
 Šze6; Šzp4,2
 cunjia · Šze4; Šzp1,3; Šzp2,8; Šzp4,10; Šzp5,4;
 Šzp6,2
 curek · Smd3; Sme1,6; Sme2; Sme4,5; Sme5,2;
 Smp1,3; Smp5,3
 cvesti · Gce,5; Gcp; Gcp,4
 cvet · Sme1,6; Sme2,2; Sme4,3; Sme5; Smp1,2;
 Smp2,3; Smp4; Smp6
 cveteti · Gcp; GLme,3; GLmp; GLže,6; GLžp,3;
 GNE
 cvetje · Sse1,6; Sse2,9; Sse3; Sse4,7; Sse5,2;
 Sse6,10
 cvetoč · GČse1; GČze4; PČme1i; PČme2,3;
 PČme5,3; PČse2; PČse5,3; PČze1,2;
 PČze2,15; PČze4,3; PČze5,6; PČze6,2;
 PČžp4,4; Pme4; Pmp2; Pžp2,3
 čakati · Gae,8; Gap; Gbe,4; Gce,54; Gcp,21;
 GLmd,3; GLme,88; GLmp,26; GLse,4; GLzd;
 GLže,37; GLžp,5; GNE,7; GVbe,3
 čaplja · Šze1,21; Šze4,6
 Čar · IOme1,40; IOme2,4; IOme3,5; IOme4,5
 Čargo · IOme1,12
 čas · Sme1,86; Sme2,115; Sme3; Sme4,48;
 Sme5,38; Sme6,4; Smp1,18; Smp2,6; Smp4,17;
 Smp5,19; Smp6
 časopis · Sme4,7; Sme6; Smp2,2; Smp4,5;
 Smp5,3
 čast · Šze1,2; Šze2,2; Šze4,5; Šze5; Šzp5; Šzp6,2
 častnik · Sme1,17; Sme2; Sme3; Sme4,5;
 Smp1,2; Smp3
 če · Vpo,1052
 čebela · Šze1; Šzp1,8; Šzp2,5; Šzp3; Šzp4
 čecha · Šzd1; Šze1,20; Šze2,2; Šze3; Šze4; Šzp1,2
 čednost · Šzd1; Šze1,8; Šze4,2; Šze5,2; Šze6
 čelada · Šze1,5; Šze2; Šze4,5; Šze5; Šze6,3;
 Šzp1,3; Šzp2,2; Šzp4; Šzp5,6; Šzp6,3
 Čelarjev · PIOme1,3; PIOme3,2; PIOme4,2;
 PIOme5; PIOmp1,2; PIOže1,2; PIOže6
 čeljust · Šze4,4; Šzp1,6; Šzp2; Šzp4,16; Šzp5,3;
 Šzp6,11
 čelo · Sme2,31; Sme5,24; Sme6,3; Smp5; Sse1,9;
 Sse2,32; Sse4,72; Sse5,13; Sse6; Ssp4
 čenča · Šze1; Šzp1,6; Šzp2; Šzp3,2; Šzp4,8
 čepeti · GLme,11; GLmp; GLže,3; GLžp,2
 čepica · Šze1,2; Šze4,12; Šze5; Šze6; Šzp5,2
 čeprav · Vpo,207
 češ · Č,117
 češnja · Šze1; Šze2,2; Šze3,2; Šze4,4; Šze6,2;
 Šzp1,12; Šzp2,6; Šzp4,9; Šzp6,2
 četa · Šze1,11; Šze2,4; Šze4,2; Šze5,5; Šze6;
 Šzp1; Šzp2
 četrti · ŠVme1,6; ŠVme4,3; ŠVme6; ŠVmp1;
 ŠVze1,5; ŠVze4,2; ŠVze5; ŠVze6
 čevelj · Sme2; Sme4,3; Sme6,8; Smp1,20;
 Smp2,17; Smp4,24; Smp5,6; Smp6,7
 čez · A,36; E4,344
 čim · A,12
 čimprej · A,15
 čist · Pme1,16; Pme2; Pme4; Pme4j; Pme5,6;
 Pme6,3; Pmp1,4; Pmp4; Pse1,7; Pse2,5;
 Pse6,2; Psp4; Pže1,18; Pže2,6; Pže4,4; Pže5,3;
 Pže6,2; Pžp1,2; Pžp2; Pžp5
 čisto · A,42
 članek · Sme1,2; Sme4; Smp1; Smp4,5; Smp6
 človek · Smd1,6; Smd4,3; Sme1,278; Sme2,32;
 Sme3,35; Sme4,54; Sme5,19; Sme6,9;
 Smp6,17
 človeški · Pme1,2; Pme2; Pmp4; Pse2,3; Pse4,2;
 Pse5; Pse6,2; Pže1,3; Pže2,3; Pže4; Pže5,2;
 Pžp1; Pžp2
 čmrlj · Smd1,2; Smd2; Sme1; Sme4,2; Smp1,3;
 Smp4
 čokat · Pme1i,3; Pme1,7; Pme2,2; Pme4;
 Pme4i,2; Pme4,3; Pme5; Pse1; Pže1,4; Pže3;
 Pže4,2; Pže6; Pžp1; Pžp2
 čop · M,3; Sme2,2; Sme4,3; Sme6,3; Smp5;
 Smp6,2
 čreda · Šzd1; Šze1,3; Šze2; Šze4; Šzp1; Šzp5,3
 črhniiti · Gce,2; GLme,16; GLmp; GNE,3; GVbe
 črka · Šze1; Šze4,3; Šze6,2; Šzp1,4; Šzp2,2;
 Šzp4,6; Šzp6,4

- črn · Pmd4; Pmeli,5; Pme1,20; Pme2,18;
Pme3,3; Pme4; Pme4i,5; Pme4,4; Pme5,12;
Pme6,8; Pmp1,23; Pmp2,4; Pmp4,11;
Pmp5,6; Pmp6,5; Pse1,7; Pse2,4; Pse4,8;
Pse5,4; Pse6; Psp1,2; Psp2,3; Psp4,4; Psp6,4;
Pzd1,2; Pzd4,2; Pže1,42; Pže2,21; Pže4,33;
Pže5,14; Pže6,5; Pžp1,24; Pžp2,8; Pžp4,21;
Pžp5,10; Pžp6,19
- Črnilogar · IOme1,12; IOme2; IOme4,2;
IOme6,2
- Črnilogarica · IOže1,9; IOže5
- črnina · Sže1,2; Sže4; Sže5,9; Sže6
- črta · Sže1,4; Sže2,4; Sže3; Sže4; Sže5,9; Sže6,2
- črviv · Pme1; Pme4i,2; Pmp1; Pmp2; Psd1;
Pse4,2; Psp4; Pže1; Pže3; Pže4,4; Pže5,2;
Pžp2; Pžp4,2; Pžp5
- čudak · Smel,14; Sme4,2; Smp1,2
- čuden · Pmeli; Pme1,26; Pme4i; Pme5,3;
Pme6,11; Pmp1,3; Pmp4,2; Pse1,50; Pse2,14;
Pse3,2; Pse4,4; Pse5; Pse6,3; Psp1; Pže1,18;
Pže2,3; Pže4,5; Pže5,2; Pže6,6; Pžp2; Pžp4,2;
Pžp5
- čuditi se · Gae,5; Gbp; Gce,2; GLme,10;
GLmp,2; GLže,4; GNE
- čudno · A,80
- čudo · Sse1,3; Sse2,6; Ssp1; Ssp2,2
- čudovit · Pme1,6; Pme4,2; Pmp1; Pse1,10; Pse2;
Psp6; Pže1,3; Pže4,3; Pže6; Pžp1,4; Pžp2;
Pžp4
- čudovito · A,13
- čumnata · Sže2,2; Sže4,10; Sže5,8
- čustvo · Sse1,3; Sse2,2; Sse4,2; Ssp1,14; Ssp2,6;
Ssp3; Ssp4,4; Ssp6
- čutara · Sže4,8; Sže6,2
- čutiti · Gae,9; Gap; Gbe; Gcd; Gce,25; Gcp,3;
GLme,68; GLmp; GLže,17; GNE
- da · Č,224; Vpo,5106; Vpr,18
- dahnniti · Gce; GLme,7; GLže,5
- dajati · Gap; Gbe,3; Gce,30; Gcp,5; Gcp,9;
GLmd,2; GLme,11; GLmp,11; GLse,4; GLže,5;
GLžp; GNE,3; GVbe
- daleč · Ajj; A,123
- daljava · Sže2,10; Sže4,11; Sže5,13
- dalje · A,57
- daljen · Pmeli,2; Pme5; Pmp2; Pmp5; Pse2,2;
Pse5,3; Psp1; Pže1; Pže2,2; Pžp3
- dan · GNsel,2; GNžel; PNse1; Smd4,12; Smd5;
Smd6,2; Sme1,94; Sme2; Sme2,12; Sme2,78;
Sme3; Sme4,233; Sme5,15; Sme6; Sme6,9;
Smp1,14; Smp2,128; Smp4,5; Smp4,51;
Smp5,6; Smp5,39; Smp6,10; Sžp4,37
- današnji · Pme1; Pme2,3; Pme3; Pme4;
Pme5,2; Pme6; Pmp2,2; Pse1; Pse2; Psp2,2;
Pže2; Pže4; Pže5
- dandanašnji · A,8; Pmeli,2; Pse1
- dandanes · A,13
- danes · A,167
- dar · Smel,8; Sme2,3; Sme4,5; Sme6,2; Smp1;
Smp2,3; Smp4,3
- dati · Gae,19; Gap,5; Gbe,5; Gcd; Gce,64;
GLmd,2; GLme,118; GLmp,44; GLse,24;
GLže,66; GLžp,3; GNE,18; GNsp1; GVap,11;
GVbd,3; GVbd,5; GVbe,59; GVbp,13
- daven · Pmp6; Psp2; Psp5; Psp6,4; Pže2; Pže4;
Pžp4
- daviti · Gce,5; GLmp; GLse; GLže; GNE,2
- debel · Pmd1,2; Pmd6; Pmeli,4; Pme1j;
Pme1,5; Pme2,3; Pme3; Pme4i,5; Pme4,5;
Pme6,15; Pmp1,4; Pmp2,3; Pmp4,3; Pmp5,4;
Pmp6,4; Psd1; Pse1,2; Pse2,3; Pse4,3; Pse5;
Pse6,6; Psp5,2; Pzd1,4; Pzd4,2; Pže1j,3;
Pže1,17; Pže2,4; Pže3; Pže4,28; Pže5,3;
Pže6,5; Pžp1jj; Pžp1,18; Pžp2jj; Pžp2,10;
Pžp4,24; Pžp6j; Pžp6,5; Ssp2
- debelo · A,22
- deblo · Ssd4; Sse1,4; Sse2,12; Sse3,7; Sse4,17;
Sse5,11; Sse6,3; Ssp1,2; Ssp4,3; Ssp5; Ssp6,8
- dečko · Smel,10; Sme4
- ded · Smel,132; Smp1,2
- deda · Sme2,5; Sme3,15; Sme4,20; Sme5,2;
Sme6,5; Smp2; Smp4
- dedec · Smel,41; Sme2; Sme3; Sme4,3; Sme6;
Smp1,13; Smp3,2; Smp4,3
- dejanje · Sse1,4; Sse2,3; Sse3; Sse4,6; Sse5,3;
Ssp2; Ssp4,2; Ssp5; Ssp6
- dejati · GLmd; GLme,137; GLmp; GLže,13
- dekla · Sže1,17; Sže2,2; Sže3,2; Sže4,15; Sže6;
Sžp1,4
- dekle · Sse1,32; Sse2,7; Sse3,3; Sse4,14; Sse5,2;
Sse6,4; Ssp1,33; Ssp2,8; Ssp3,4; Ssp4,7; Ssp5;
Ssp6,3
- dekletce · Sse1,10; Sse2; Sse4,12; Sse5,2; Sse6
- deklica · Sže1,13; Sže2,3; Sže3; Sže4,6; Sžp1,3;
Sžp3; Sžp4
- del · Smel,4; Sme2,11; Sme3; Sme4; Sme5,9;
Sme6,9; Smp1
- delati · Gae,2; Gap; Gbe,10; Gbp,2; Gcd,4;
Gce,32; Gcp,15; GLmd,5; GLme,52; GLmp,19;
GLsp; GLže; GLže,5; GNA; GNE,30; GVbe,2
- delavec · Smd1; Sme1,5; Sme2; Sme4,2; Smp1,7;
Smp2,3; Smp3,3; Smp4,2; Smp5; Smp6

- delavski · Pme2; Pme4; Pme5,2; Pmp4;
Pmp5,5; Pmp6; Pse4; Pse5; Pže4; Pžp1;
Pžp4,2; Pžp5; Pžp6
- deležen · Pme1,5; Pmp1,7; Pse1; Pže1,3
- deliti · Gce,5; Gcp,3; GLmd; GLme; GLmp,11;
GLže,7; GNE,2
- delo · Sse1,16; Sse2,22; Sse3,4; Sse4,57; Sse5,13;
Sse6; Ssp1,2; Ssp2; Ssp4,10; Ssp5,2
- denar · Sme1,19; Sme2,47; Sme4,56; Sme5;
Sme6,6; Smp4
- desen · Pme1i; Pme2; Pse4,2; Pse5,3; Pže3;
Pže5; Pže6
- deset · ŠGmp6,2; ŠGsp5; ŠGžp2; ŠGžp5,3;
ŠGžp6; ŠG1,7; ŠG4,29
- deska · Šzd4; Šze1; Šze2,2; Šze4,4; Šze5; Šze6;
Šzp1; Šzp2,10; Šzp4,7; Šzp5,5; Šzp6,3
- desna · Šze4,13; Šze5,8
- desni · Pme3; Pme5,2; Pse1,2; Pse2,3; Pže1,3;
Pže2,5; Pže4,19
- desnica · Šze1; Šze2,5; Šze4,12; Šze5,13;
Šze6,10; Šzp4
- detektiv · Smd1,3; Sme1,16; Sme2; Sme3;
Sme4,4; Smp1,3; Smp2; Smp5
- devet · ŠGmp5; ŠGmp6; ŠGžp5; ŠGžp6; ŠG1,3;
ŠG4,16
- deveti · ŠVme1,3; ŠVme4,2; ŠVmp3; ŠVze1,2;
ŠVže4,2; ŠVže6
- dež · Sme1,10; Sme2,4; Sme4; Sme5,5; Sme6
- Dežela · IZže1,2; IZže2,5; IZže4,2; IZže5,4
- dežela · Šze1,9; Šze2,6; Šze4,19; Šze5,17; Šzp2
- dih · Sme1,4; Sme2,4; Sme4,11; Sme6
- dihati · Gae; Gce,8; Gcp; GLme,7; GLmp,2;
GLse,3; GLže,4; GNE
- dim · Sme1,23; Sme2,13; Sme4,16; Sme5,2;
Sme6
- dimnik · Sme1,2; Sme2,7; Sme5; Smp2,3;
Smp4,2; Smp5
- dišati · Gce,7; GLme; GLmp,2; GLse,3; GLže;
GLžp
- divizija · Šze1,6; Šze2,2; Šze5; Šze6
- divje · A,31
- divji · Pme1,6; Pme2,10; Pme3,3; Pme4;
Pme5,3; Pme6; Pmp1,3; Pmp2; Pmp4; Pmp6;
Pse1; Pse2; Pse3; Pže1,5; Pže2,7; Pže4; Pže5;
Pže6; Pžp1; Pžp2; Pžp4
- dlaka · Šze1,5; Šzp3; Šzp4; Šzp6,3
- dlan · Šze2,4; Šze4,20; Šze5,12; Šze6,37; Šzp1,2;
Šzp2; Šzp4,9; Šzp5,5; Šzp6,4
- dnina · Šze4,14; Šzp4; Šzp6
- dno · Sse1,2; Sse2,26; Sse4,15; Sse5,28
- do · E2,672; E,24
- doba · Šze1,2; Šze2,4; Šze4; Šze5,3; Šze6; Šzp2,2
- dober · Pmd1; Pme1i,5; Pme1jj,2; Pme1,58;
Pme2,8; Pme3,2; Pme4jj; Pme4,4; Pme4,21;
Pme5,3; Pme6jj; Pme6,11; Pmp1j; Pmp1jj;
Pmp1,10; Pmp2jj,2; Pmp2,7; Pmp4,4;
Pmp5j,2; Pmp5,2; Pmp6,2; Psd4,2; Pse1jj,2;
Pse1,4; Pse1,43; Pse2,3; Pse2j; Pse2,12; Pse4,8;
Pse6,5; Psp2,3; Psp4; Psp6,4; Pže1jj,2;
Pže1j,5; Pže1,22; Pže2j; Pže2,23; Pže4j;
Pže4jj; Pže4,17; Pže5jj; Pže5,6; Pže6,2;
Pžp1,4; Pžp2,7; Pžp4j; Pžp4,4; Sse2
- dobiti · Gae,6; Gap; Gbe,3; Gce,11; Gcp; GLmd;
GLme,23; GLmp,8; GLže,13; GNE
- dobro · Aj; Ajj; Ajj,11; Aj,24; A,271; Sse1;
Sse4,5; Sse5,2
- dobrodušen · Pme1i,4; Pme4,2; Pme6; Pse6;
Pže1,3; Pže6
- dobrota · Šze1,3; Šze2,4; Šze5,2; Šze6; Šzp1,2;
Šzp2; Šzp4,2; Šzp5
- dočakati · GLme,11; GLmp,2; GLže
- dati · Gce,16; GLme,41; GLmp; GLže,22
- dogajati se · Gce,14; Gcp,4; GLse,5; GLžp,2
- dogodek · Sme1,3; Sme2,3; Sme4,2; Sme5,3;
Sme6; Smp1,3; Smp2,4; Smp4; Smp5,2; Smp6
- dokaj · A,28
- dokazovati · Gce,15; Gcp; GLmd; GLme,2;
GLmp,2; GLse; GLžp,2; GNE
- dokler · ZR,118
- dokument · Sme4,2; Smp1; Smp2,2; Smp4,5
- doleteti · Gce; GLse; GLžd; GLže,21
- dolg · Pmd1; Pme1i; Pme1j,2; Pme1,17;
Pme2,3; Pme3; Pme4; Pme4; Pme4i,7;
Pme4j; Pme4,12; Pme5j; Pme5,18; Pme6jj;
Pme6,10; Pmp1j; Pmp1,16; Pmp2; Pmp4jj;
Pmp4,12; Pmp5,7; Pmp6j; Pmp6,9; Pse1j,2;
Pse1,11; Pse2,2; Pse3; Pse4j,2; Pse4,5;
Pse5,10; Pse6; Psp2,6; Psp4,17; Psp5,9; Psp6;
Pžd1; Pžd4; Pžd6; Pže1j,5; Pže1,49; Pže2,5;
Pže3; Pže4,27; Pže5jj; Pže5,14; Pže6,10;
Pžp1j,3; Pžp1,14; Pžp2,3; Pžp4,15; Pžp5,14;
Pžp6,16; Sme2,2; Smp4,4; Smp5
- dolgčas · Sme1,10; Sme4,2
- dolgo · A,228; Sse5,3; ŠNE,2
- dolina · Šzd2; Šze1,31; Šze2,43; Šze4,63;
Šze5,86; Šze6,5; Šzp1,3; Šzp2,4; Šzp4; Šzp5,4
- dolinica · Šze1,6; Šze2,2; Šze5,2
- dolžnost · Šze1,11; Šze2,2; Šze4; Šze4,8; Šze5,2
- dom · Sme1,15; Sme2,35; Sme3,16; Sme4,28;
Sme5,9; Sme6,2; Smp2,2; Smp4,2; Smp5,2;
Smp6
- doma · A,237

- domač · Pmd1; Pme1i,6; Pme1,3; Pme2,7;
Pme4; Pme4,5; Pme5,3; Pme6; Pmp1,8;
Pmp2; Pmp3; Pmp4,4; Pmp5,2; Pse1,4;
Pse2,5; Pse5; Pse6,2; Psp1; Psp4,2; Psp5;
Pže1,12; Pže2,5; Pže4,12; Pže5,5; Pže6,2;
Pžp1,4; Pžp4,3; Pžp5,3
- domače · A,12
- domačija · Šze1,7; Šze2,6; Šze4,6; Šze5
- Dominik · IOme1,17; IOme4
- Dominov rob · IZme2,3; IZme3; IZme4,12;
IZme5,7; IZme6,3
- domišljija · Šze1,6; Šze2,5; Šze3; Šze4,3; Šze5,8
- domobranec · Smd1,10; Sme1,14; Sme3,2; Sme4;
Smp1,10; Smp2; Smp3; Smp4,2; Smp5
- domobranski · Pme1; Pžd1; Pžd6; Pže1,2;
Pže2,2; Pže5,4
- domov · A,182
- domovina · Šze1,20; Šze2,8; Šze3,2; Šze4,10;
Šze5,5
- doneti · Gce,8; Gcp; GLmp; GLse; GLže,12;
GLžp
- dopoldne · A,17; Sse4
- dopovedati · Gce; GLme,3; GLmp,2; GLže,3;
GNE,5; GVbe
- dopovedovati · Gce; GLmd,3; GLme,2; GLmp;
GLže,2; GNE,7; GVbp
- doseči · Gae,8; Gce,2; GLme,17; GLmp,3;
GLže,7; GLžp; GNE,9
- dotakniti se · Gbe; Gcd; Gce,5; Gcp; GLme,22;
GLmp; GLse,2; GLžd; GLže; GNE,4; GVbe,2
- dotikati se · God; Gce,5; GLme,2; GLmp; GLse;
GLže,4; GLžp; GVbe
- dovolj · A,109; ŠNE,17
- dozoreti · GLmd; GLme,3; GLmp; GLse,4;
GLže,3; GLžp; GNE,3
- doživeti · Gce,2; Gcp; GLme,9; GLmp,7; GLže,3;
GNE,2
- drag · Pme1i,17; Pme1j; Pme1jj; Pme1,3;
Pme4jj; Pmp1,6; Pmp2; Pmp2jj; Pmp4;
Pse1,3; Pže1,7; Pže2jj; Pže4; Pže5; Smp4
- Draga · IZže1; IZže2,3; IZže3,5; IZže4; IZže5,9;
IZže6,4
- Dragarica · IOže1,59; IOže2; IOže3,3; IOže4,2;
IOže5; IOže6
- Dragarjev · PIOme1,2; PIOme3,2; PIOme5;
PIOmp1,2; PIOse4; PIOse5; PIOse6; PIOže4;
PIOže5; PIOže6,2
- Dragič · IOme1; IOme1,47; IOme2,6; IOme3,9;
IOme4,9; IOme6
- dragocen · Pme1i; Pme1jj; Pme2,2; Pmp2;
Pmp4,2; Psp1; Pže2; Pže4; Pžp2
- dražiti · Gae; Gap,2; Gbe; Gbp; Gce,3; GLme,6;
GLmp,5; GLse,3; GLsp
- Drejc · IOme1,190; IOme2,9; IOme3,26;
IOme4,22; IOme5,3; IOme6,4
- dreti · Gbe,2; Gbp; Gce,14; Gcp,6; GLmd;
GLme,15; GLmp,14; GLse,2; GLžd; GLže,4;
GLžp,4; GNE,4; GVbe
- drevesce · Sse1,3; Sse2; Sse4,4; Sse5,2; Sse6;
Ssp2,2
- drevje · Sse1,8; Sse2,5; Sse3; Sse4,5; Sse5;
Sse6,10
- drevo · Sse1,15; Sse2,6; Sse4,9; Sse5,4; Sse6,2;
Ssp1,3; Ssp2,14; Ssp3; Ssp4,2; Ssp5; Ssp6
- drevored · Sme1,3; Sme2,2; Sme4; Sme5,14
- drhteti · Gce,2; GLme,11; GLžp,2; GNE
- drob · Sme1,2; Sme2; Sme4,2; Sme5,14
- droban · Pme1i,6; Pme1,11; Pme2,5; Pme3;
Pme6,2; Pmp1,8; Pmp2,2; Pmp4,8; Pmp6,3;
Pse1,5; Pse4,2; Pžd1,2; Pžd4; Pže1,19; Pže2,3;
Pže6,4; Pžp1,10; Pžp4,11; Pžp6,9
- drobec · Sme1,2; Sme2; Sme4,3; Sme6,3; Smp1;
Smp2,3; Smp4,2
- droben · Pme1; Pme4; Pme4i,2; Pme5; Pže4,12;
Pžp2,7; Pžp5,2
- drobiti · Gae; Gce,3; Gcp; GLmd; GLme,2;
GLmp,7; GLse; GLže,5; GLžp,4
- drseti · Gbe; Gce,10; GLme,3; GLmp,2; GLse,2;
GLžd,7; GLže,6; GLžp,2; GNE,2
- drug · Pme1i; Pžp4,4; ZDme1,5; ZDme1,59;
ZDme2,14; ZDme3,27; ZDme4,10; ZDme4,25;
ZDme5,11; ZDme6,12
- druga · ZDmd1,2; ZDsp1; ZDsp2; ZDsp3;
ZDsp5; ZDsp6; ZDže1,11; ZDže2,8; ZDže3,3;
ZDže4,19; ZDže5,5; ZDže6,2
- drugače · A,66
- drugačen · Pme1,8; Pme2; Pme6; Pmp1,5;
Pse1,3; Pse6; Pže1,6; Pže4; Pžp1,2; Pžp6;
ZDže2,2
- drugam · A,18
- druge · ZDžp1,5; ZDžp2,12; ZDžp4,14; ZDžp5,5;
ZDžp6,4
- drugi · ŠVme1,39; ŠVme2,3; ŠVme4,7;
ŠVme4,14; ŠVme5,21; ŠVme6; ŠVmp1,3;
ŠVmp3; ŠVse1,2; ŠVse2,6; ŠVse4,9; ŠVse5;
ŠVže1,16; ŠVže2,9; ŠVže4,19; ŠVže5,6;
ŠVže6,2; ŠVžp1,2; ZDmp1,45; ZDmp2,9;
ZDmp3,16; ZDmp4,22; ZDmp5,3; ZDmp6,10
- drugič · A,31
- drugo · ZDse1,4; ZDse2,36; ZDse3; ZDse4,22;
ZDse5,4; ZDse6,4
- družba · Šze1,4; Šze2,5; Šze4,2; Šze5,2; Šze6,3

- družina · Šze1,16; Šze2,3; Šze3,2; Šze4,6;
Šze5,5; Šze6,3; Šzp1,2; Šzp4
- drveti · Gcp; GLme,7; GLmp,3; GLse; GLže,2;
GLžp; GNE,2
- drviti · Gcd,2; Gce,7; Gcp,7
- drvo · Sse2,3; Ssp2,4; Ssp4,22
- drzen · Pme1,3; Pme4,2; Pme6; Pmp1,3;
Pmp4,2; Pže1; Pže2; Pže4; Pže5; Pžp1; Pžp6
- držati · Gae,5; Gbe; Gbp; Gcd; Gce,65; Gcp,2;
Gcp,6; GLmd,3; GLme,60; GLmp,16; GLse,5;
GLžd,2; GLže,42; GLžp,3; GNE,5; GVbe,4;
GVbp
- država · Šze1,19; Šze2,7; Šze3,8; Šze4,8; Šze5
- državen · Pmeli; Pme2; Pme4; Pmp2; Pmp4,4;
Pmp5; Pse2,3; Pse4,3; Pže1; Pže2; Pže4;
Pže5,3
- duh · Sme1,7; Sme2,10; Sme4,2; Sme4,7;
Sme5,14; Smp2; Smp6
- duša · Šze1,47; Šze2,46; Šze3; Šze4,49; Šze5,46;
Šze6,7; Šzp1,5; Šzp2,9; Šzp4,3; Šzp5
- duševen · Pmp2; Pmp3; Pmp5; Pse1; Pse2;
Pse3; Pse4,2; Pse5; Pže2,2; Pže4,2; Pžp2;
Pžp5
- dva · ŠGmd1,109; ŠGmd2,6; ŠGmd3,3;
ŠGmd4,93; ŠGmd5,4; ŠGmd6,12; ŠG4,12
- dvajset · ŠGmp6; ŠGsp2; ŠGsp5; ŠGsp6,4;
ŠG1,2; ŠG4,34
- dvakrat · A,63
- dvanajst · ŠGmp6; ŠGsp2; ŠGsp6; ŠGžp5;
ŠG1,3; ŠG4,18
- dve · ŠGsd1,9; ŠGsd2,3; ŠGsd4,30; ŠGsd5,7;
ŠGsd6,2; ŠGžd1,57; ŠGžd2,2; ŠGžd4,69;
ŠGžd5,8; ŠGžd6,7; ŠGžp2
- dvigati · Gae; Gcd; Gce,17; Gcp,7; GLme,12;
GLmp,2; GLse,3; GLže,11; GLžp,3; GNE,4
- dvigniti · Gae; Gbe; God,8; Gce,74; Gcp,7;
GLmd,3; GLme,145; GLmp,15; GLse,5;
GLže,47; GLžp; GNE,9; GVbe
- dvignjen · PNme6,2; PNmp6; PNže5,2;
PNže6,3; PNžp4; PNžp6,5
- dvoje · ŠL1,8; ŠL4,9
- dvom · Sme1,5; Sme2,3; Sme4; Sme5; Sme6;
Smp1; Smp2,2; Smp4
- dvorana · Šze1,3; Šze2,4; Šze4,6; Šze5,4; Šzp5
- dvorišče · Sse1,5; Sse2,9; Sse3; Sse4,35;
Sse5,30; Ssp2; Ssp5
- eden · ŠGme1,20
- edini · Pmd4; Pme1,14; Pme4,2; Pme4,2;
Pse1,4; Pse4,6; Pžd1; Pže1,13; Pže2; Pže4;
Pže4,6; Pže5; Pže6,2
- eh · M,41
- en · ŠGme1,9; ŠGme2,7; ŠGme3; ŠGme4,11;
ŠGme4,16; ŠGme5,11; ŠGme6,11; ŠGmp1;
ŠGmp3; ŠGse1,9; ŠGse2,3; ŠGse4,2; ŠGse6;
ŠGže1,49; ŠGže2,6; ŠGže3; ŠGže4,32;
ŠGže5,8; ŠGže6,4; ŠGžp4
- enak · Pmd1,2; Pme1,5; Pme4; Pme6,3;
Pmp1,3; Pmp4,2; Pmp6,2; Pse2; Pžd1;
Pže1,6; Pže6; Pžp1; Pžp4
- Enaka Palica · IOme1,28; IOže4
- enakomeren · Pme4i,2; Pme5; Pmp5; Pse1;
Pse2; Pse5; Pse6,3; Pžp4; Pžp5,4
- enakomerno · A,18
- enkrat · A,201
- Ernest · IOme1,82; IOme2,2; IOme3,11;
IOme4,7; IOme6,2
- fakin · Smd1; Sme1,3; Sme2; Smp1,4; Smp2,2;
Smp3
- fant · Smd1,18; Smd4; Sme1,59; Sme2; Sme3,2;
Sme4,7; Smp1,60; Smp2,4; Smp3,3; Smp4,12;
Smp6,3
- fara · Šze1,2; Šze2,4; Šze4; Šze5,6
- fašist · Smd1,2; Sme1,14; Sme3,2; Sme4;
Smp1,46; Smp2,7; Smp3,10; Smp4,3; Smp5;
Smp6,2
- fašističen · Pmeli,3; Pme4i,7; Pme5,2; Pmp1;
Pže1; Pže2,5; Pže5,2; Pžp3; Pžp5
- Felix · IOme1,6; IOme2,2; IOme3; IOme4
- figa · Šze1,3; Šze2; Šze4,11; Šzp4
- fižol · Sme1; Sme2,6; Sme4,6
- fotografija · Šzd1; Šze1,7; Šze2,2; Šze3; Šze4,19;
Šze5,2; Šze6,2; Šzp1,3; Šzp2; Šzp4; Šzp5
- frakelj · Sme1; Sme2,2; Sme4,10; Sme6,4
- Franc · IOme1,11; IOme2; IOme3; IOme4,2
- France · IOme1,13
- Francija · IZže2,2; IZže4,8
- Francovich · IOme1,11; IOmp4
- Fratnik · IOme1,18; IOme2; IOme3,4
- fronta · Šzd6; Šze2,4; Šze4,8; Šze5,3; Šzp5,2
- Furlan · IOme1,25; IOme3; IPme1; IPmp1,4;
IPmp6
- gaber · Sme1,13; Sme2; Sme4,3; Sme5,3; Sme6;
Smp2; Smp4,3
- gabrov · Pme1; Pme2; Pme3; Pme5; Pme6,2;
Pže2; Pže4,4
- gaj · Sme1,5; Sme4,3; Sme5,2; Smp1,2
- ganiti · Gbe; Gce,12; Gcp; GLme,14; GLse,3;
GLže,5; GLžp,2; GNE,6
- garati · Gae; Gap,4; Gce,3; Gcp,2; GLme,13;
GLmp,7; GLže,5; GNE,3; GVbe
- gaziti · Gce,4; GLme,4; GLmp; GLže; GNE
- general · Sme1,10; Sme3; Sme4,4; Sme6,2;
Smp1; Smp4

- gibati se · Gce,6; Gcp,2; GLže; GNE,3
 Gilda · IOže1,28; IOže2,2; IOže3,2; IOže4,3
 Gino · IOME1,23; IOme2; IOme4,2
 Giusto · IOme1,11; IOme2; IOme3; IOme4,3
 gladek · Pme1,3; Pme5,5; Pmp1,2; Pmp4;
 Pmp5,2; Pse1,5; Pse2; Psp4; Pže1,8; Pže2,2;
 Pže4,2; Pže5,3; Pže6; Pžp1,3; Pžp2
 gladina · Sže2; Sže4,6; Sže5,12; Sže6
 gladiti · Gce,3; GLme,5; GNE,4
 glagol · Sme1,3; Sme2,2; Smp2,4; Smp4; Smp5
 glas · Sme1,112; Sme2; Sme2,26; Sme3;
 Sme4,60; Sme5,38; Sme6,183; Smp1,5; Smp2;
 Smp4; Smp6,10
 glasan · Pme1; Pme6,3; Pmp2,2; Psel1; Psel,2;
 Pse2; Pse4,2; Pse6,6; Pžel1,2; Pže1,4; Pže5;
 Pžp4,2
 glasneje · A,36
 glasno · Aj,2; A,263
 glava · Sžd1,8; Sžd4,3; Sžd6; Sže1,75; Sže2,78;
 Sže3,7; Sže4,437; Sže5,90; Sže6,224; Sžp1,11;
 Sžp2,8; Sžp4,29; Sžp5,4; Sžp6,10
 glavica · Sže1,2; Sže4,7; Sže5; Sžp6,2
 gledati · Gad,2; Gae,13; Gap; Gbe,6; Gbp,2;
 Gcd,8; Gce,97; Gcp,36; GLmd,10; GLme,158;
 GLmp,34; GLse,4; GLže,46; GLžp,8; GNA,8;
 GNE,9; GVbd; GVbe,3; GVbe,80; GVbp,12
 globel · Sže2,3; Sže4,2; Sže5,3; Sže6; Sžp2,2;
 Sžp4; Sžp5,6
 globina · Sže1; Sže2,5; Sže4,4; Sže5; Sžp2,3
 globok · Pmd6; Pme1,8; Pme2; Pme4i; Pme4,3;
 Pme5,2; Pme6j,3; Pme6,16; Pmp4; Pmp5;
 Pmp6; Psel,7; Pse2; Pse5; Pse6,2; Pžel1;
 Pžel1j; Pžel1,11; Pže2,4; Pže4j; Pže4jj; Pže4,6;
 Pže5,12; Pžp1,4; Pžp2,6; Pžp4j; Pžp4,4;
 Pžp5,6; Pžp6,2
 globoko · Aj; Ajj; Aj,24; A,179
 gluh · Pmeli,2; Pme1,17; Pme6,2; Pmp1; Psel1;
 Psp1; Pžel,12; Pže5,2; Pžp1,3; Pžp4; Pžp6
 gmajna · Sže1,3; Sže2,4; Sže4,7; Sže5,16;
 Sžp1,4; Sžp2,4; Sžp4,6; Sžp5,9
 gnati · Gae,2; Gbe; Gbp; Gcd; Gce,12; Gcp,9;
 GLmd; GLme,12; GLmp,10; GLse; GLže,6;
 GNE,2
 gnezdo · Sse1,2; Sse2,9; Sse3,2; Sse4,5; Sse5;
 Ssp6
 gnih · Pme5; Pse2,2; Pse4; Pse5; Pže1; Pže2;
 Pže4; Pže5; Pžp1; Pžp2
 gnoj · Sme1,2; Sme2,11; Sme4,9; Sme5,4; Sme6
 godba · Sže1,5; Sže2,3; Sže4,2; Sže6
 goditi se · Gce,15; Gcp; GLme; GLse,7; GLže,2;
 GLžp,4
 godrnjati · Gce; Gcp; GLme,3; GLmp; GLže,3;
 GLžp
 gol · Pme1,6; Pme2,2; Pme5,5; Pme6,2;
 Pmp1,2; Pmp4; Pmp6; Psel,3; Pse2,2; Pse3;
 Pse4,3; Pse5,4; Pse6,2; Psp1,2; Psp4; Psp6;
 Pžd4; Pžd5; Pže1,12; Pže2; Pže4,4; Pže5,5;
 Pže6,4; Pžp1,4; Pžp2,5; Pžp4,4; Pžp5; Pžp6,5
 golob · Pse4; Sme1; Smp1,3; Smp1,6; Smp2,5;
 Smp4
 goniti · Gae,2; Gbe; Gce; GLme,4; GLmp,4;
 GLže; GNE; GVbe
 gora · Sže1,16; Sže2,12; Sže3; Sže4,8; Sže6,3;
 Sžp1,15; Sžp2; Sžp2,20; Sžp3,3; Sžp4,14;
 Sžp5,10; Sžp6,6
 gorak · Pme1,2; Pme2; Pmp1,2; Pmp4; Pse1,2;
 Pse4; Pse5; Pže1; Pže4; Pže5,4; Pže6
 goreti · Gce,20; Gcp,2; Gcp,4; GLme,2; GLse;
 GLže,5; GLžp,3
 gori · A,12
 Gorica · IZže2,4; IZže3; IZže4,21; IZže5,16;
 IZže6
 goriški · Pme1i; Pme5,2; Pmp2; Pmp4; Pse5;
 Psp2; Pže2,3; Pže3; Pže5,2; Pžp6
 gorjača · Sže2; Sže4,5; Sže6,4
 gosnica · Sže1,13; Sže2,5; Sže4,13; Sže5,2;
 Sže6,4; Sžp1; Sžp2,2; Sžp4; Sžp5
 gospa · Sže1,8; Sže2; Sže4,4; Sže6
 Gospod · IVme1,14
 gospod · Sme1,77; Sme2,8; Sme3,9; Sme4,18;
 Sme5,4; Sme6,3; Smp1; Smp3,2; Smp4,2
 gospodar · Smd1; Sme1,34; Sme2,5; Sme3,3;
 Sme4,7; Sme6,3; Smp1,5; Smp2; Smp3; Smp6
 gosposinja · Sže1,14; Sžp1,3
 gost · Pme1,6; Pme4i; Pmp1,5; Pmp4,4;
 Pmp5,3; Pmp6; Pmp6; Pse1,10; Pse2,2;
 Pse4,2; Pse5,2; Psp6; Pžel1,17; Pže2,4; Pže4,5;
 Pže5,7; Pže6; Smd1; Sme4; Sme6; Smp1,2;
 Smp3; Smp4,2
 gostilna · Sžd4; Sže1,4; Sže2,17; Sže4,21;
 Sže5,19; Sže6; Sžp1,4; Sžp2,2; Sžp5,4
 gosto · Aj,6; A,12
 gotovo · A,61
 govor · Sme1,2; Sme2,2; Sme4,4; Sme5; Sme6;
 Smp4,6; Smp6,2
 govorica · Sže1,2; Sže2,2; Sže4,2; Sže5,3; Sže6;
 Sžp1,3; Sžp4
 govoriti · Gad; Gae,15; Gap; Gbe,20; Gbp,2;
 Gcd,2; Gce,55; Gcp,5; Gcp,13; GLmd,3;
 GLme,161; GLmp,38; GLse,3; GLže,40;
 GLžp,4; GNE,28; GVbe; GVbe,8; GVbp

- gozd · Smd2; Sme1,12; Sme2,23; Sme3,3;
Sme4,27; Sme5,32; Sme6,6; Smp1,8; Smp2,15;
Smp4,6; Smp5,6; Smp6,2
- grabiti · Gae; Gcd; Gce,2; Gcp,5; GLme,9;
GLmp,4; GLse; GLžd; GLže,3; GLžp
- grablje · Šzp1,3; Šzp4,12; Šzp6,2
- grad · Sme1,4; Sme2; Sme4,8; Sme5,3; Smp1;
Smp4,2; Smp5
- graditi · Gce,3; Gcp; GLme,2; GLmp,4; GNE,3
- grajski · Pme1i,2; Pme2; Pme3,6; Pme4i;
Pme5; Pmp1,2; Pse4; Pže5; Pžp4
- granata · Šze1,10; Šze4; Šzp1,4; Šzp4; Šzp5;
Šzp6,2
- grapa · Šze1,4; Šze2,7; Šze3,2; Šze4,17; Šze5,17;
Šzp1; Šzp2; Šzp2,4; Šzp4,4; Šzp5,9
- Graparjev · PIOme1,5; PIOme3,4; PIOme4;
PIOmp1; PIOže4; PIOže5
- graščak · Sme1,27; Sme3,2; Sme4,2; Smp1,3;
Smp4
- grd · Pme1i,3; Pme1j; Pme1,8; Pme4; Pme4;
Pme4jj; Pme5,3; Pme6,2; Pse1,11; Pse2,4;
Pse4; Pse6; Pžel1j,2; Pžel,13; Pže2,3; Pže4,2;
Pže5; Pžp1,4; Pžp2,2; Pžp5
- grdo · A,27
- grebsti · Gae,2; Gbe; Gce,3; Gcp; GLme,3;
GLže,4; GNE,2
- Gregorčič · IOme1,10; IOme3,3; IOme4,3
- greh · Sme1,22; Sme2,6; Sme4,16; Sme5,2;
Smp1; Smp2,4; Smp4,7; Smp5
- grenek · Pme1; Pme1i; Pme1,9; Pme2,3;
Pme6,3; Pmp1,4; Pmp2; Pmp4; Pmp6j; Pselj;
Pse1,7; Pse2,6; Pse4; Pse5; Pse6; Psp1,5;
Psp2; Pže1j; Pže1,8; Pže4,6; Pže5; Pžp1,3;
Pžp4,2
- grenko · A,16
- grenkoba · Šze1,4; Šze2,4; Šze4,2; Šze5; Šze6,7;
Šzp2
- grič · Sme1; Sme2,7; Sme3; Sme4,7; Sme5,4;
Sme6; Smp1; Smp2; Smp5,2
- gristi · Gce,12; Gcp,2; GLme,8; GLmp; GLže,3;
GNE,2
- griva · Šze1,12; Šze2,3; Šze4,5; Šze5,6
- grlo · Sse2,10; Sse4,31; Sse5,24; Sse6,2; Ssp2,6;
Ssp4
- grm · Smd1; Sme1,9; Sme2,15; Sme3,6;
Sme4,10; Sme5,10; Sme6,7; Smp2; Smp4,6;
Smp6,4
- grmeti · Gbe; Gce,4; GLme,5; GLmp,2; GNE,2
- grmovje · Sse1; Sse2,8; Sse4,24; Sse5,9; Sse6,4
- grob · Pme1i,2; Pme2,2; Pme4; Pme6,2; Pse2,2;
Pse5; Psp4; Pže1; Pže6; Pžp2; Pžp4; Pžp6,2;
Sme1,9; Sme2,11; Sme3; Sme4,15; Sme5;
Sme5,7; Sme6,4; Smp1; Smp2; Smp5
- grobar · Sme1,15; Sme2; Sme3,2; Sme4,2
- grofica · Šze1,5; Šze2,2; Šze4; Šze5; Šzp2
- grom · Sme1,4; Sme2,2; Sme4,3; Sme5,2; Sme6;
Smp1
- gromek · Pme1i,6; Pme2,2; Pme5; Pme6,4
- groza · Šze1,5; Šze2,9; Šze5,7; Šze6,11; Šzp4;
Šzp5
- grozeče · A,16
- grozen · Pme5; Pse2,3; Pse5; Pže1; Pže5,2;
Pžp1,2; Pžp2; Pžp5
- groziti · Gce,3; Gcp,2; GLme; GLže,4; GNE
- grozljiv · Pme1; Pme6; Pse1; Pse2,2; Pžel,3;
Pže5; Pžp1; Pžp4
- gruča · Šze1,13; Šze2,4; Šze3,2; Šze4,3; Šze5,5;
Šze6; Šzp5
- gumb · Smd4; Sme2,5; Sme4,2; Smp4,5
- h · E3,101; M
- ha · M,37
- hahaha · M,22
- halo · M,20
- harmonij · Sme1,8; Sme3; Sme4,6
- harmonika · Šze1,7; Šze2,2; Šze4,4; Šze5;
Šzp1,2
- hčerka · Šzd4; Šze1,3; Šze4,2; Šzp1,5
- hči · Šzd1,2; Šzd4; Šze1,21; Šze2,2; Šze3,2;
Šze4,7; Šze6,3; Šzp1; Šzp2,2; Šzp4; Šzp5,2
- he · M,18
- hebniti · Gce,2; GLme,2; GLže,5; GVbd
- Hedvika · IOžel,8; IOže3; IOže6
- hej · M,20
- he-he · M,15
- hip · Sme1,6; Sme2; Sme4,140; Sme5,20
- hiša · Šzd4,2; Šze1,71; Šze2,181; Šze3,24;
Šze4,154; Šze5,124; Šze6,70; Šzp1,14;
Šzp2,21; Šzp4,15; Šzp5,16; Šzp6,8
- Hiša št. 14 · ISže1,5; ISže2,2; ISže4,4; ISže5;
ISže6
- hišen · Pme2; Pme4i,6; Pme5,2; Pme6; Pmp3;
Pmp4; Psp2; Pže1; Pže2,2; Pže5; Pžp2
- hiter · Pmd1j; Pme1j; Pme5; Pme6; Pmp6;
Pmp6j; Psel,2; Pže1; Pže1j
- hiteti · Gae,2; Gbe; Gcd,3; Gce,31; Gcp,3;
Gcp,3; GLme,11; GLmp,4; GLže,9; GLžp,2;
GVbe; GVbp
- hitro · Aj,43; A,217
- hkrati · A,121
- hlače · Šzp1,13; Šzp2,6; Šzp4,48; Šzp5,15
- hlačen · Pme2,5; Pme6; Pmp2; Pmp4,7; Pmp5,3

hladen · Pmd1; Pmeli; Pme1,7; Pme2,2; Pme4;
Pme5; Pmp1; Pse1,3; Pse2; Pže1,9; Pže4,2;
Pže5,5; Pžp1; Pžp5
hlapec · Sme1,46; Sme2; Sme3,2; Sme4,10;
Sme6,3; Smp1,13; Smp2,3; Smp3; Smp4
hlastno · A,23
hlev · Sme1,5; Sme2,25; Sme3,4; Sme4,22;
Sme5,12; Sme6,21; Smp1; Smp5,2; Smp6,3
Hlip · IOme1,8; IOme3,2; IOme4,2; IOme5
Hlipov · PIOme1,11; PIOme2,4; PIOme6,2
hlod · Sme2,2; Sme4,5; Smp2,2; Smp4,2
hm · M,124
hoditi · Gad; Gae,19; Gap,3; Gbe,6; Gbp;
Gce,45; Gcp,11; GLmd,10; GLme,113;
GLmp,19; GLse,2; GLžd; GLže,30; GLžp,3;
GNE,19; GVbe,6
hodnik · Sme1; Sme2,4; Sme3; Sme4,10;
Sme5,18; Sme6; Smp1,3; Smp2,2; Smp4;
Smp5,10
hoja · Sže1,4; Sže2,2; Sže4,2; Sže5,4
Hotejc · IOme2,5; IOme3,16; IOme4,4; IOme6
Hotejčev · PIOme1,13; PIOme2; PIOme4;
PIOme4,2; PIOže1; PIOže2; PIOže4; PIOžp2
Hotejec · IOme1,105
hoteti · Gad; Gae,25; Gap,42; Gbe,4; Gbe,18;
Gbp,7; Gcd; Gce,9; Gce,24; Gcp; Gcp,14;
GLmd,3; GLme,138; GLmp,12; GLse;
GLže,45; GLžp,6; GVbe,2
hram · Sme2,6; Sme4,20; Sme5,6; Smp4; Smp5
hramček · Sme1,10; Sme5
hrana · Sže1,4; Sže2,3; Sže4,7
hrast · Sme1,11; Sme2,8; Sme3; Sme4,4;
Sme6,4; Smp1; Smp2; Smp4,2
hrastov · Pme1; Pme3; Pme4,4; Pme5; Pmp4;
Psp4,2; Pžd6; Pže1,2; Pže4,3; Pže5; Pže6
lrabet · Sme1,5; Sme2,6; Sme4,67; Sme5,65;
Sme6,53; Smp1,3; Smp4,2; Smp5; Smp6,5
hrepenenje · SGse1,9; SGse2,9; SGse4,4;
SGse5,4; SGse6,5; SGsp1,2; SGsp2,3; SGsp4
hrepeneti · Gae; Gap,2; Gce; GLme,3; GLmp,2;
GLse; GLže,2; GLžp; GNE
hreščati · Gce,3; GLme; GLmp; GLse,2; GLže,3;
GLžp
hreščeč · GČže5; PČme1i,4; PČme4i; PČme5;
PČme6,3
hrib · Sme1; Sme2,7; Sme4,14; Sme5,6; Sme6;
Smp1,6; Smp2,2; Smp4,7; Smp5,3
lribovski · Pme1; Pmp1; Pže1,2; Pže2,3;
Pže4,4; Pžp5
hrumeti · Gce,5; GLme,2; GLmp,3; GLse;
GLže,4; GNE
hrup · Sme1,16; Sme2,4; Sme4,3; Sme6,3

hrupno · A,10
hruška · Sže1; Sže2,3; Sže6,2; Sžp2,4; Sžp4,2
hud · Pme1i,4; Pme1jj,2; Pme1j,3; Pme1,34;
Pme2,2; Pme4; Pme4jj; Pme4jj; Pme5,2;
Pme6; Pmp1,9; Pmp2jj; Pmp4; Pmp5,4;
Pmp6,2; Psel,1j,5; Pse1j,3; Pse1,35; Pse2jj;
Pse2j,7; Pse2,33; Pse4,3; Pse6jj; Pže1jj,3;
Pže1j,3; Pže1,30; Pže2jj,2; Pže2,4; Pže4jj;
Pže4,8; Pže5jj,3; Pže5,4; Pžp1j; Pžp1jj;
Pžp1,5; Pžp2,4; Pžp4,5; Pžp5jj,2; Pžp5,2
hudič · Sme1,107; Sme2,12; Sme3,13; Sme4,31;
Sme5,2; Sme6,6; Smp1,8; Smp2,2; Smp4,2
hudičev · Pme5,2; Pme6; Pmp1; Pse1,4; Pse5,3;
Pže1,7; Pže2; Pže4,2; Pžp1,6; Pžp4; Pžp6,2
hudo · A; Aj,3; Aj,11; A,59
hudomušen · Pme1; Pmeli; Pme2; Pme4i,2;
Pme6; Pže1,3; Pže2
hudomušno · A,28
hvala · M,9; Sže1,16; Sže2,2; Sže4
hvaležen · Pme1,2; Pme6,2; Pmp1,3; Pmp6;
Pse6; Pžp6
hvaliti · Gbp; Gce,3; Gcp; GLme,5; GLmp;
GLže
Idrija · IZže1,31; IZže2,9; IZže3,2; IZže4,14;
IZže5,6
igla · Sže4,3; Sže6,3; Sžp1; Sžp5,2; Sžp6
igrati · Gce,16; Gcp,5; GLmd,4; GLme,23;
GLmp,7; GLže,3; GLžp; GNE,3; GVbe;
GVbp,4
ime · Ssd1; Sse1,39; Sse2,13; Sse4,20; Sse5,26;
Sse6,8; Ssp1; Ssp2; Ssp4,6; Ssp6,3
imeniten · Pmd1; Pme1,12; Pme2; Pmp4;
Pse1,2; Psp6j; Pže1,2; Pžp1
imenitno · A,11
imenovati · Gae; Gap,2; Gce,2; GLme,10;
GLmp,7; GLse; GLže,3; GNE; GVbp
imeti · Gad,2; Gae,67; Gap,18; Gbe,74; Gbp,25;
Gcd,9; Gce,246; Gcp,53; GLmd,20; GLme,222;
GLmp,53; GLse,6; GLsp; GLžd,2; GLže,90;
GLžp,4; GNE,20; GVbe,37; GVbp; GVbp
in · Vpr,18470
iskati · Gae,6; Gbp; Gcd; Gce,14; Gcp,8;
GLmd,2; GLme,25; GLmp,4; GLže,11;
GLžp,3; GNA,6; GNE,12
iskra · Sže1,2; Sžp1,2; Sžp2; Sžp4,4; Sžp6
isti · ZIme1,8; ZIme2,5; ZIme4,30; ZIme5,6;
ZIme6,7; ZIse2,2; ZIse4,3; ZIse5,6; ZIse6;
ZIsp6; ZIže1,5; ZIže2,3; ZIže4,6; ZIže5,5;
ZIžp4,4
Italija · IZže1,5; IZže2,6; IZže3,2; IZže4,4;
IZže5,8; IZže6

- Italijan · IPme1,6; IPme4; IPmp1,2; IPmp2;
IPmp4,2; IPmp6,2
- italijanski · Pme1; Pme2; Pme3; Pme4,3;
Pme6,2; Pmp1,2; Pmp2,3; Pmp3; Pse2; Pse5;
Pse6; Pzd4; Pže1,4; Pže2; Pže4; Pžp2; Pžp4,2
- iti · Gad,6; Gad,16; Gae,16; Gae,126; Gap,4;
Gbe,8; Gbe,29; Gbp; Gbp,8; Gcd,2; Gcd,7;
Gce,26; Gce,154; Gcp,2; Gcp,4; Gcp,22; Gcp,41;
GLmd,42; GLme, 324; GLmp,71; GLse,42;
GLžd,4; GLže,160; GLžp,7; GNE,9; GVad,20;
GVap,17; GVbd,7; GVbe,49; GVbp,9
- Ivana · IOže1,18; IOže3; IOže4
- iz · E2,1470
- izba · Sže1,14; Sže2,41; Sže3; Sže4,75; Sže5,67;
Sže6,2; Sžp1; Sžp5,2
- izbirati · Gcd; Gce; Gcp; GLmd; GLme,3;
GLmp; GLse; GLže,4; GNE,3
- izbrati · Gbe; Gce,2; GLmd; GLme,12; GLmp,3;
GLže,4; GLžp; GNE
- izbruhniti · Gce,5; GLmd; GLme,10; GLže,6
- izbuljiti · Gce,2; GLme,12; GLže,8; GLžp
- izdati · Gcd; Gce; GLme,9; GLmp,4; GLže;
GNE; GVbp,3; Sže4
- izdaviti · Gae,4; Gce; GLme,11
- izdihniti · Gce; GLme,13; GLmp; GLže,5; GNE
- izginiti · Gcd,2; Gce,24; Gcp,2; GLmd,10;
GLme,59; GLmp,7; GLse,2; GLžd,5; GLže,51;
GLžp,10; GVbe,4; GVbp
- izginjati · Gce,4; GLme; GLmp,3; GLže,3; GLžp
- izgovoriti · Gae; Gce; GLme,9
- izgubiti · Gae,2; Gbe,3; Gce,8; Gcp; GLme,33;
GLmp,7; GLse; GLže,6; GLžp,3; GNžp1
- izgublјati · Gae; Gce,6; GLme,3; GLmp,3;
GLže,2; GNE,2
- izgublјen · GNže1; PNmd1,2; PNme1,2;
PNme1,8; PNme2; PNme3; PNme6,2;
PNmp2; PNže1,2; PNže2; PNžp1; PNžp2
- izgublјeno · A,21
- Izidor · IOme1,82; IOme2,3; IOme3; IOme4,5;
IOme6
- izjaviti · GLme,6; GLže,4; GLžp
- izkazati se · Gce; GLmp; GLmp,2; GLse,6
- izkopati · Gap,2; GLmd; GLme,3; GLže,3; GNE
- izmučen · GNme1,5; PNme1,1; PNme2,2;
PNmp4; PNmp6; PNse1; PNže1; PNže6;
PNžp4
- izpit · PNme5,2; PTmp1; PTmp2; PTsp2; Sme1;
Sme2; Sme4; Smp1; Smp2; Smp3,4; Smp4,4;
Smp5,4
- izpiti · Gae; Gce,3; GLme,20; GLmp,2; GNE;
GTme4; GVbe,2; GVbp
- izpod · E2,146
- izprazniti · Gce,3; GLmd; GLme,7; GLmp;
GLže; GNE,2
- izpred · E2,17
- izpustiti · Gae; Gbp,2; Gcp,2; GLme,5; GLmp,7;
GLže,4; GNE
- izraz · Smd1,2; Sme1,7; Sme2; Sme4; Sme5;
Sme6,3; Smp4
- izreči · GLme,9; GNE,3
- izročiti · Gce,3; Gcp; GLme,3; GLmp; GLže,2
- iztegniti · Gce,2; GLme,8; GLže
- iztrgati · Gce,2; Gcp,2; GLmd; GLme,13;
GLmp; GLse; GLže,8; GNE,3
- izvedeti · Gce; Gcp; GLme,6; GLmp,2; GLže,2;
GLžp,2
- izza · E2,82
- ja · Č,29
- jablana · Sže2,4; Sže4,2; Sže5,3; Sžp1,4; Sžp2;
Sžp6
- jabolko · Ssd4; Sse1,4; Sse2,2; Sse4,6; Ssp1,4;
Ssp4,7; Ssp6
- Jaconcic · IOme1,14
- jajce · Ssd2,2; Ssd4; Sse6; Ssp2,6; Ssp4,12; Ssp5;
Ssp6,5
- Jakob · IOme1,53; IOme2; IOme3,10; IOme4;
IOme6
- Jakončič · IOme1,40; IOme2,2; IOme3,2;
IOme4,9; IOme5; IOme6,2
- jama · Sže1,10; Sže2,21; Sže3; Sže4,22; Sže5,11;
Sže6,2; Sžp4; Sžp5
- Janez · IOme1,34; IOme2,2; IOme3,5; IOme4,5;
IOme5; IOme6
- Jankec · IOme1,48; IOme2; IOme3; IOme4,2
- jarek · Sme2,2; Sme4,4; Sme5,5; Smp4,4; Smp5
- jasa · Sže1,4; Sže2,5; Sže4,8; Sže5,5; Sžp4; Sžp5
- jasen · Pme1,j,2; Pme1,10; Pme2,2; Pme4;
Pme5,j; Pme5,2; Pme6,2; Pmp1,4; Pmp4,2;
Pse1,30; Pse2,3; Pse4,6; Pse5; Pse6; Pže1,14;
Pže2; Pže4; Pže5; Pžp1,5; Pžp2; Pžp4; Pžp6
- jasno · Aj,4; A,87
- javiti · Gce,4; GLme; GNA; GNE,4
- javolheršturmfiren · M,18
- javor · Sme1,3; Sme3,2; Sme4; Sme5,7; Smp2
- Javorka · IOže1,30; IOže2,2; IOže4,2; IOže5;
IOže6

- jaz · Sže4,2; ZOad2,5; ZOad3,8; ZOad4,12;
 ZOad6,9; ZOae1,4; ZOae1,6; ZOae1,266;
 ZOae2,46; ZOae2,66; ZOae3,95; ZOae3,795;
 ZOae4; ZOae4; ZOae4,2; ZOae4,4; ZOae4,15;
 ZOae4,30; ZOae4,74; ZOae4,756; ZOae5,45;
 ZOae6,6; ZOae6,102; ZOamd1,28;
 ZOamp1,63; ZOamp2,17; ZOamp3,170;
 ZOamp4,124; ZOamp5,75; ZOamp,2;
 ZOap6,34; ZOažp1
- jecljati · Gce,5; GLme,5; GNE,5
- ječa · Sže1,4; Sže2,27; Sže4,13; Sže5,20; Sže6;
 Sžp1; Sžp2; Sžp4,2; Sžp5,5
- ječar · Smd1,4; Sme1,41; Sme2,2; Sme3,2;
 Sme4,10; Smp1,3; Smp2; Smp4,3
- jed · Sže1,5; Sže2,6; Sže3; Sže4; Sže5,6; Sže6,4;
 Sžp2,2; Sžp4,2
- jeklen · Pme1i; Pme4; Pme4i; Pme5; Pme6;
 Pmp1,2; Pmp2,2; Pmp4; Pmp5; Pmp6,2;
 Pžel; Pže6; Pžp5
- jekniti · Gce; GLme,5; GLmp; GLže,13; GLžp,2
- Jelčica · IOže1,8; IOže2,2; IOže3
- Jelka · IOže1,5; IOže2,2; IOže4,3
- jelka · Sže1,14; Sže2; Sže3; Sže4,3; Sže5; Sže6;
 Sžp3
- jelševje · Sse1,2; Sse4,8; Sse5,5; Sse6,2
- jemati · Gae,2; Gbe; Gce,6; Gcp,4; GLmd,3;
 GLme,3; GLmp,10; GLse,2; GNE,4; GVbp
- Jera · IOže1,64; IOže2,5; IOže3,4; IOže4,7;
 IOže5; IOže6
- jera · Sže3,5; Sže4,4; Sže5,4; Sže6,2
- Jerica · IOže1,13; IOže3,2; IOže4,2
- jermen · Smd4; Sme1; Sme2; Sme4,8; Sme5;
 Sme6; Smp1,4; Smp2; Smp4,5
- Jernej · IOme1,44; IOme2,3; IOme3,3; IOme4,2;
 IOme6
- Jernejček · IOme1,2; IOme1,9
- jesen · Sme1,2; Sme4,4; Sže1,16; Sže2,7; Sže4,9;
 Sže5,19; Sže6; Sžp4
- jesenski · Pme1i; Pme1,4; Pme2; Pme5,2;
 Pmp1; Pse1,6; Pse2,3; Pse4; Pse5; Psp5;
 Pžel,4; Pže2,2; Pže4; Pžp1,2
- jesti · Gae; Gap,3; Gce,2; Gce,6; Gcp; GLmd,3;
 GLme,35; GLmp,8; GLže,3; GNA,2; GNE,11;
 GVbd; GVbe,5; GVbp,4
- jetnik · Sme1,24; Sme2; Sme3,4; Sme5; Sme6;
 Smp1,10; Smp2,6; Smp3,4; Smp4,2; Smp6
- jetniški · Pme1,3; Pme2; Pme4; Pmp1; Pmp2;
 Psp2; Pže2; Pže4,2; Pže5,3; Pžp5
- jez · Sme1,4; Sme2; Sme2,2; Sme4,11; Sme5;
 Sme6,11; Smp4; Smp4
- jeza · Sže1,21; Sže2,27; Sže4,7; Sže5,6; Sže6,19
- jezen · Pme1,8; Pme6,4; Pmp1; Pmp6; Pse2;
 Pse4; Pžel,2; Pže6,4; Pžp2; Pžp5
- jezero · Ssd1; Sse1,4; Sse2,2; Sse4,3; Sse5
- jeziki · Sme1,21; Sme2,19; Sme4,36; Sme5,9;
 Sme6,25; Smp1,3; Smp3; Smp4,3; Smp5;
 Smp6,2
- jeziti se · Gce,6; Gcp,2; GLme,9; GLse,9;
 GLže,4
- jezno · A,90
- Jezus · IVme1,22; IVme4
- joj · M,44
- jok · Sme1,8; Sme2,2; Sme4,22; Sme5,3; Sme6
- jokati · Gae; Gbe,3; Gce,7; GLme,20; GLmp,4;
 GLže,13; GLžp
- Jugoslavija · IZže1,3; IZže2,4; IZže3; IZže4;
 IZže5,3
- Julček · IOme1,10; IOme3; IOme4
- junak · Sme1,7; Sme2,3; Sme4,4; Smp1,3;
 Smp2,2; Smp4,2; Smp5; Smp6,2
- junaški · Pme4; Pmp1,2; Pse1,2; Psp4; Pžel;
 Pže2; Pže4; Pžp4; Pžp5,2
- junaško · A,13
- Jur · IOme1,11; IOme6
- Justin · IOme1,20; IOme3,3
- Justina · IOme4,2; IOže1,44; IOže2; IOže3,2;
 IOže4,16; IOže6,2
- jutranji · Pme1; Pme4,2; Pme5; Pse1,4; Pse5,7;
 Pse6; Pžel,4; Pže2,3; Pže4,3; Pže5,8; Pžp5
- jutri · A,49
- jutro · Sse1,26; Sse2,17; Sse3,7; Sse4,45; Sse5,8;
 Sse6; Ssp1; Ssp5,2
- južen · Pmp2; Pmp5; Pse1; Pse2,2; Psp4; Pžel;
 Pže2; Pže5,2
- ju-u-hu-u · M,14
- k · E3,1356
- kača · Sže1,16; Sže2; Sže4,5; Sžp1
- kadar · Vpo,187; ZR,35
- kadar koli · ZRPO; ZR,3; ZR,6
- kadet · Sme1,35; Sme2; Sme3; Sme4,6; Sme5,2;
 Sme6,2; Smp1; Smp2
- Kadetka · IOže1,125; IOže2,4; IOže3,5;
 IOže4,19; IOže5,2; IOže6,10
- kadetov · Pme1,2; Pme4,4; Pme5,2; Pme6;
 Pmp4,2; Pžel,11; Pže2; Pže3; Pže4
- kaditi · Gae,2; Gbe,2; Gce,9; Gcp,3; GLmd;
 GLme,13; GLmp,4; GLse,5; GNE
- kaj · Č,623; Sme4; ZPOse1,54; ZPOse2,14;
 ZPOse4,136; ZPOse6; ZVse1,290; ZVse2,3;
 ZVse3,8; ZVse4,262; ZVse5,3; ZVse5,10;
 ZVse6,5; ZV,30
- kajne · Č,13
- kajpak · Č,144

- kajti · Vpr,152
kak · ZNEme6,3; ZNEsp6,2; ZNEžp6;
ZPOme1,3; ZPOme2,3; ZPOme3,2;
ZPOme4,3; ZPOme4,3; ZPOmp2,6;
ZPOmp4,3; ZPOse1; ZPOse2; ZPOsp2,2;
ZPOže1; ZPOže1,6; ZPOže2,3; ZPOžp2;
ZPOžp4; ZV
kakah · ŠNE,19
kako · ZPOse4,12; ZPO,14; ZV,809
kakopak · A,17
kakor · Vpo,2038; Vpr,40
kakrašen · ZRmd1; ZRme1,17; ZRme2,6;
ZRme4,2; ZRme6,2; ZRmp1,3; ZRmp4,7;
ZRse2; ZRže1,8; ZRže2,4; ZRže4,4; ZRže5;
ZRžp1,3; ZRžp2,3; ZRžp4,4; ZRžp5
kakšen · ZPOme1,27; ZPOme4,4; ZPOme5,4;
ZPOme6; ZPOmp1,3; ZPOmp4,4; ZPOse1,6;
ZPOse4; ZPOse6; ZPOsp2; ZPOže1,23;
ZPOže2,2; ZPOže3; ZPOže4,3; ZPOžp1,4;
ZPOžp2; ZPOžp3; ZPOžp4,3; ZVmd1;
ZVme1,33; ZVme4,2; ZVme4,3; ZVme5,2;
ZVmp1,5; ZVmp2; ZVmp3; ZVmp4; ZVmp5;
ZVse1,21; ZVse4,3; ZVsp1; ZVže1,34;
ZVže2,2; ZVže4,11; ZVže5; ZVže6; ZVžp1,6;
ZVžp4,5
kalen · Pmp1; Pmp2,2; Pmp3; Pmp4; Pmp6;
Pže1,4; Pže2; Pže4,2; Pže5; Pžp1,3; Pžp4,2;
Pžp5,2; Pžp6,2
kam · ZV,181
kamen · Pže4; Smd1,2; Smd4,2; Sme1,109;
Sme2,22; Sme3,6; Sme4,50; Sme5,15; Sme6,5;
Smp1,11; Smp2,3; Smp4,11; Smp5,3; Smp6,9
kamenit · Pme2; Pme4,2; Pme5,3; Pme6; Pse4;
Pže2; Pže4,2; Pže5,2; Pžp2,4; Pžp4; Pžp5,2
kamenje · Sse1,5; Sse2,4; Sse4,19; Sse5,3;
Sse6,9
kamnit · Pme2; Pme4i; Pme4,2; Pme5; Pse5;
Psp1; Psp4; Pže1; Pže4; Pže5,2; Pže6
kamniten · Pme2; Pme4i,4; Pme5,2; Pse4;
Pže1; Pže4,4; Pže5,4; Pžp4
kamor · ZR,34
kap · Šze1,8; Šze4,4
kapa · Šzd1; Šzd6; Šze1,2; Šze2; Šze4,13;
Šze5,2; Šze6,5; Šžp1; Šžp4,2; Šžp5,2
kaplan · Sme1,8; Sme2; Sme3; Sme4,2; Smp1
kaplja · Šzd1; Šze1,17; Šze2,2; Šze4,5; Šze6,3;
Šžp1,15; Šžp2,3; Šžp4,11; Šžp6
kapljati · Gce,4; Gcp,3; GLmp,2; GLse; GLže,2;
GLžp,3; GNE,2
kar · Č,702; Vpo,92; ZRse2,16; ZRse3,3; ZRse5;
ZRse6,4; ZVR,229
karabinjer · Smd1,24; Smd2,5; Smd3,2; Smd4,2;
Smd6; Sme1,10; Sme4,2; Sme6; Smp1,30;
Smp2,8; Smp3,5; Smp4,4
karajoče · A,23
kasarna · Šzd1; Šze2,5; Šze3,3; Šze4,9; Šze5,3;
Šže6; Šžp1,2; Šžp4
kasneje · A,86
kašča · Šze1; Šze2; Šze4,22; Šze5,8
Kata · IOže1,10; IOže4
kateri · ZRme2,10; ZRme3; ZRme4,6;
ZRme5,22; ZRme6,2; ZRme6,13; ZRmp2,5;
ZRmp3,2; ZRmp4,3; ZRmp5,12; ZRmp6,6;
ZRse2,4; ZRse3,2; ZRse4; ZRse4,3; ZRse5,6;
ZRse6,2; ZRsp5,2; ZRže2,7; ZRže4,7;
ZRže5,16; ZRže6,4; ZRžp2; ZRžp4,2;
ZRžp5,10; ZRžp6,3; ZVme1,4; ZVme2,3;
ZVme4,3; ZVmp6; ZVse2,2; ZVse4,3; ZVse6,4;
ZVsp4; ZVže1,6; ZVže3; ZVže4,4; ZVže5,5;
ZVže6,3; ZVžp3; ZVžp4
Kati · IOže1,14; IOže2; IOže3; IOže4,4
Katra · IOže1,44; IOže2,2; IOže3,4; IOže4,6;
IOže6,6
kava · Šze1,4; Šze2,34; Šze4,24; Šze5,3; Šze6
kavarna · Šze2,2; Šze4,9; Šze5,2; Šžp5,3
kazalec · Sme2,2; Sme4,29; Sme6,39; Smp5;
Smp6
kazati · Gap; Gbp; Gce,34; Gcp,5; GLme,18;
GLmp,2; GLse,25; GLže,9
kazen · Šze1,4; Šze2; Šze4,6; Šze5,2; Šze6; Šžp1;
Šžp2; Šžp4,2; Šžp6
kdaj · ZV,167
kdo · ZPOme1,102; ZPOme2; ZPOme4,7;
ZPOme6,3; ZVme1,181; ZVme3,9; ZVme4,7;
ZVme5,3; ZVme6,11
kdor · ZRme1,32
ker · Vpo,1060
ki · ZVR,2995
kihiniti · GLme,4; GLže,7
kij · Sme4,2; Sme6,8; Smp4; Smp6,2
kila · Šze1,3; Šze4,8; Šze6
kimati · GLme,8; GLže,2
kisel · Pme1,2; Pme5; Pse1,4; Pse4,3; Pse5;
Pse6; Pže1,2; Pže2
kita · Šzd1; Šzd4,11; Šzd6,2; Šze4,2; Šze5; Šžp2;
Šžp4,6
kitara · Šze1,3; Šze4,6; Šžp1; Šžp4,3
kje · ZPO,4; ZV,221
kjer · ZR,201
klada · Šze1,6; Šze4,5
kladivo · Sse1; Sse2; Sse4,5; Sse6,4; Ssp2; Ssp4
klanec · Sme1,2; Sme2,13; Sme3; Sme4,5;
Sme5,33; Sme6; Smp1; Smp2,2; Smp5

- klasje · Sse1; Sse4,6; Sse5,3; Sse6,3
 klečati · Gce,3; GLmd,4; GLme,7; GLže; GLžp;
 GNE
 klet · Šze2,6; Šze3; Šze4,12; Šze5,9; Šzp4;
 Šzp5,4
 klic · Sme1,6; Sme4,3; Sme5; Smp1,2; Smp3
 klicati · Gae,2; Gap; Gbe,2; Gbp; Gce,13; Gcp,5;
 GLme,14; GLmp,15; GLse; GLže,7; GNE,5;
 GVbe,3
 klikniti · Gce,15; GLmd; GLme,4; GLže,6
 klin · Sme2,2; Sme3; Sme4,2; Sme5,2; Smp4;
 Smp5,2; Smp6
 kljub · E3,87
 ključ · Sme1,5; Sme2; Sme4,6; Smp6,2
 kljuka · Šze1; Šze2,2; Šze4,7; Šze5; Šzp4
 kljun · Sme1,4; Sme2; Sme4,4; Smp4,2; Smp6,2
 klobuk · Smd4; Sme1,10; Sme2; Sme4,3,9;
 Sme5,2; Sme6,5; Smp2; Smp4,9; Smp5;
 Smp6,3
 klop · M,29; Šze1,26; Šze2,7; Šze3; Šze4,39;
 Šze5,21; Šze6,3; Šzp1; Šzp4,2; Šzp5
 klopotati · Gce; Gce,17; GLme,4; GLmp,3
 klopotec · Sme1,5,4; Sme2,12; Sme4,30; Sme5;
 Sme6,4
 kmalu · A,200
 kmečki · Pme1,5; Pme6; Pmp1,2; Pmp3; Pmp5;
 Pse1; Pse2,2; Pse5; Psp3; Pže1; Pže5,4; Pže6;
 Pžp2; Pžp3; Pžp4; Pžp5,3
 kmet · Smd1,3; Smd2; Sme1,15,4; Sme2,6;
 Sme3,11; Sme4,20; Sme5,8; Sme6,3; Smp1,65;
 Smp2,4; Smp3,7; Smp4,8; Smp5,14; Smp6,5
 kmetica · Pže2; Šzd4; Šze1,4; Šzp1,5; Šzp2,2;
 Šzp4,2; Šzp6
 kmetija · Šze1,10; Šze2,7; Šze4,8; Šze5,5; Šze6;
 Šzp1; Šzp2; Šzp5,3
 knjiga · Šze1,5; Šze2,9; Šze3; Šze4,18; Šze5,6;
 Šze6,2; Šzp1,5; Šzp2,9; Šzp4,3,9; Šzp5,9; Šzp6
 ko · Vpo,1262
 kobila · Šze1,9; Šze2,2; Šze4,2; Šze5
 Kobilčer · IOme1,12; IOme2; IOme4,4
 kocka · Šze3; Šze4,3; Šzp4,5; Šzp5; Šzp6
 Kočarica · IOže1,12; IOže3
 kočija · Šze1,4; Šze2,3; Šze4; Šze5; Šzp4
 kod · ZV,31
 kodrast · Pmp1; Pmp2,4; Pmp5; Pmp6; Pže4,3;
 Pže5,2
 kokoš · Šze1,4; Šze4,3; Šze5; Šzp1,3; Šzp2,3;
 Šzp3; Šzp4,6; Šzp6,2
 kol · Sme4,9; Sme6,9; Smp1; Smp6,3
 koleno · Ssd1; Sse1; Sse3,2; Sse4,5; Sse5,3;
 Sse6,3; Ssp1,12; Ssp2,9; Ssp4,5,2; Ssp5,19;
 Ssp6,14
 kolesej · Sme2; Sme3,2; Sme4; Sme5; Sme6,6
 koliko · ZV,82
 kolikor · Vpo,27; ZR,5
 kolnik · Sme1; Sme2,7; Sme4,7; Sme5,30; Smp4;
 Smp5,3
 kolo · Ssd1; Ssd5; Sse1,3; Sse2,2; Sse2,4; Sse4,2;
 Sse5,2; Sse5,3; Ssp1,6; Ssp2; Ssp4,5; Ssp5,2
 kolovoz · Sme1,9; Sme2,5; Sme3,2; Sme4,8;
 Sme5,2,6; Sme6,2; Smp5,3
 kolovratiti · Gap; GLme,6; GLmp,2; GLže
 komaj · A,210
 komandant · Sme1,31; Sme2; Sme3,9; Sme4,6;
 Sme6
 komis · Sme1,4; Sme2,5; Sme4,6; Sme6
 komisar · Sme1,32; Sme2; Sme3; Sme4,3;
 Sme5,3; Sme6,3
 komolec · Sme2; Sme4,2; Sme6,13; Smp4,8;
 Smp5; Smp6,15
 končan · GNme1,4; GNse1,5; GNže1,3; PNže5,2
 končati · Gce,4; GLme,24; GLmp; GLse,4;
 GLže,3; GNE,3
 končno · A,13
 konec · Smd2; Smd4,3; Smd5,4; Sme1,5,4;
 Sme2,3,3; Sme3,14; Sme4,20; Sme5,6,7; Smp4;
 Smp5; Smp6,2
 konica · Šze4,4; Šze6,2; Šzp4,4; Šzp5
 konj · Smd1; Sme1,8; Sme2,9; Sme3,2; Sme4,3;
 Sme5,7; Smp1,9; Smp2,12; Smp4,20; Smp5,6;
 Smp6,8
 konjski · Pme4; Pmp1; Pmp6,3; Psp1; Pže2,2;
 Pže4; Pže6,2; Pžp2
 kopati · Gad; Gap; Gcp,2; Gcp,2; GLmd;
 GLme,9; GLmp,2; GLsp,2; GLžp; GNA;
 GNE,3; GVbe
 kopito · Sse1; Sse4; Sse6,10; Ssp1,3; Ssp2;
 Ssp4,4; Ssp6
 korak · Smd4,20; Sme1,19; Sme4,41; Sme5,5;
 Sme6,14; Smp1,30; Smp2,5,4; Smp4,21;
 Smp5,9; Smp6,2,9
 korakati · Gae,2; Gap; Gcd; Gce,6; GLmd,2;
 GLme,13; GLmp,3; GLse; GLže,3; GNE,3
 korenina · Šzd6; Šze3; Šze4,4; Šze5; Šze6,2;
 Šzp1,9; Šzp2,3; Šzp3,2; Šzp4,6; Šzp5,2; Šzp6,6
 koristen · Pme1,2; Pme4; Pmp1; Pmp2; Pse1,3;
 Pže1; Pže4; Pžp4
 korito · Ssd6,2; Sse1,4; Sse2,8; Sse3,9; Sse4,9;
 Sse5,12; Sse6,3; Ssp6
 korpus · Sme1,4; Sme2,7; Sme6
 koruza · Šze1,4; Šze2,5; Šze4,8; Šze5,3
 koruzen · Pme1; Pme4; Pmp1; Pmp4; Pse2;
 Pse4; Psp2; Pže1; Pže2,3; Pžp1; Pžp4

- kos · A; PD,6; Smd4,2; Sme1,28; Sme2,8;
 Sme4,56; Sme5,2; Sme6,3; Smp1,10; Smp2,3;
 Smp3; Smp4,14; Smp6,3
 kosa · Sze1,8; Sze2,13; Sze3; Sze4,22; Sze5,19;
 Sze6,16; Sžp1,4; Sžp4,12; Sžp6
 kosec · Smd3; Sme1,3; Sme4; Smp1,5; Smp2,3;
 Smp3,4; Smp4,2
 kosilo · Sse1,2; Sse2,2; Sse3,2; Sse4,8; Sse5,12;
 Sse6
 kositi · Gae,2; Gce,6; Gcp,5; GLme,4; GLmp,4;
 GNA,5; GNE,6
 kosmat · Pme1,4; Pmp1; Pmp2; Pse1; Pse4;
 Pže1,2; Pže4,2; Pžp2; Pžp5; Pžp6
 kost · Sze1,5; Sze2,6; Sze4,10; Sze6,3; Sžp1,11;
 Sžp2,6; Sžp4,17; Sžp6,4
 kostanj · Sme1,3; Sme2; Sme4,4; Sme6,2; Smp2
 koš · Sme1,5; Sme2; Sme4,5; Sme5,9
 košara · Sžd4; Sžd6,2; Sze1,2; Sze2; Sze3,2;
 Sze4,19; Sze6; Sžp4,3; Sžp6,2
 košat · Pme1i; Pme1,5; Pme2; Pme3; Pme4i;
 Pme5; Pme6; Pmp1,5; Pmp4; Pmp5; Pmp6,3;
 Pse6; Pže1,3; Pže2,3; Pže4,6; Pže5; Pže6,3;
 Pžp1,2; Pžp2,5; Pžp3; Pžp4; Pžp6,2
 košček · Sme1,2; Sme2; Sme4,10; Sme5,2; Sme6;
 Smp1,3; Smp2; Smp4
 koščen · Pme1i,3; Pme1,13; Pme2; Pme4;
 Pme4i,4; Pme6,2; Pmp1,3; Pmp4,4; Pmp6,3;
 Pse4,2; Psp2; Psp6; Pžd4; Pže1,13; Pže3;
 Pže4,12; Pže5,5; Pže6,10; Pžp1,2; Pžp4,2;
 Pžp5,2; Pžp6
 košnja · Sze1,3; Sze2,2; Sze4,3; Sze5,6; Sze6
 kot · Sme1,5; Sme2,15; Sme4,37; Sme5,31;
 Smp2,4; Smp4,6; Smp5,5; Vpo,166
 kotel · Smdl; Sme1; Sme2,4; Sme3; Sme4,5;
 Sme5,6; Sme6,2; Smp2; Smp5,2
 Kovač · IOme1,37; IOme3; IOme5,2; IOme6
 kovač · Sme1,86; Sme2,2; Sme3,3; Sme4,6;
 Sme6
 kovček · Sme2; Sme3,4; Sme4,8; Sme6,3; Smp4;
 Smp6
 kozarček · Smd4,2; Sme4,9; Sme6; Smp4,2
 kozarec · Sme1,4; Sme4,51; Sme5,2; Sme6,2;
 Smp1,3; Smp2,4; Smp4,9; Smp5; Smp6,3
 kozel · Sme1; Sme2,12; Sme4,6; Sme5,3; Smp1
 kozolec · Sme1; Sme2,3; Sme3,3; Sme4,7;
 Sme5,6; Sme6,3; Smp1; Smp2; Smp4,2
 koža · Sze1,18; Sze2,5; Sze3; Sze4,17; Sze5,4;
 Sze6,7; Sžp2
 kraj · E2; Sme1,8; Sme2,42; Sme3; Sme4,11;
 Sme5,16; Smp1; Smp2,6; Smp4,10; Smp5,15
 kralj · Sme1,9; Sme2,2; Sme3,3; Sme4; Smp2
 kraljestvo · Sse1; Sse2,2; Sse4,6; Sse5
 kramar · Smd1,6; Smd2; Smp1,2; Smp2
 kramp · Sme1,10; Sme4,24; Sme5,3; Sme6,7;
 Smp1,2; Smp2; Smp4,2; Smp6
 Kras · IZme1; IZme2,5; IZme3,2; IZme4,10;
 IZme5,6
 krasan · Pmp1,3; Pse1; Psp1; Pže1; Pže1,7
 krasti · Gae; Gap; Gce,6; Gcp,4; GLme,7;
 GLmp; GLse; GLžp; GNE,2
 Kraševc · IOme1,2; IPmp1,5; IPmp2; IPmp3;
 IPmp6
 kraški · Pme1,3; Pme2,2; Pme4; Pme5,2;
 Pmp1; Pmp2; Pse1; Pse4; Pže1,15; Pže2;
 Pže4; Pže5,3; Pžp1; Pžp2,5; Pžp5
 kratek · Pmeli; Pme1j; Pme1,4; Pme2; Pme4;
 Pme4i,2; Pme4,11; Pme5jj; Pme5,12; Pme6,4;
 Pmp1j; Pmp5,4; Pmp6,2; Pse1,7; Pse4,3;
 Pse5,2; Pse6; Psp1j; Pže1j; Pže1,9; Pže2;
 Pže4,5; Pže5; Pže5jj,2; Pže6,4; Pžp1,4;
 Pžp2,3; Pžp4,4; Pžp5,2; Pžp6,8
 kratko · A,29; Sse5,8
 krava · Sžd4; Sze1,29; Sze2,7; Sze3,7; Sze4,35;
 Sze5,3; Sze6,4; Sžp1,9; Sžp2,4; Sžp3,3;
 Sžp4,20; Sžp5; Sžp6,3
 kravji · Pme4,5; Pmp1,3; Pmp2; Pmp4,3;
 Pmp6; Pže1,4; Pže2; Pže4,3
 krčevito · A,28
 krčma · Sze1,7; Sze2; Sze3,2; Sze4,2; Sze5,5;
 Sze6,2
 krepelj · Smp2,4; Smp4,4; Smp5; Smp6
 kreniti · Gae; Gap; Gcd,3; Gce,11; Gcp,4;
 GLmd,2; GLme,37; GLmp,2; GLžd; GLže,12;
 GVbe; GVbp,3
 krepak · Pme1,2; Pme2; Pme6; Pmp2; Pse2,2;
 Psp2; Pže1; Pžp5
 krepko · A,65
 kretnja · Sze1,3; Sze2; Sze4,2; Sze5,2; Sze6,42;
 Sžp1; Sžp3; Sžp5,3; Sžp6,12
 kri · Sze1,90; Sze2,44; Sze3,4; Sze4,41; Sze5,10;
 Sze6,7
 kričati · Gae; Gbd; Gbe; Gbp; Gcd; Gce,24;
 Gcp,6; GLme,22; GLmp,14; GLže,11; GNE,3;
 GVbe,3
 krik · Sme1,16; Sme2; Sme4,4; Sme6,5; Smp1,5;
 Smp4,2; Smp6
 krikniti · Gce,6; Gcp; GLme,9; GLmp,2;
 GLže,10; GNE
 kriliti · Gcp,3; GLme,6; GLže,4
 krilo · Sse1,5; Sse2,8; Sse4,20; Sse5,4; Sse6,4;
 Ssp1; Ssp2,5; Ssp3; Ssp4,9; Ssp5,4; Ssp6,13
 Krist · IVme1,15; IVme2,3; IVme4,3; IVme6,2
 Kristus · IBme1,2; IBme2,2; IOme6; IVme1,8;
 IVme4; IVme4,2; IVmp1; IVmp4

kriv · Pmd1; Pme1,39; Pme6,4; Pmp1,2; Pmp2;
Pmp6; Pse1,2; Psp4; Pže1,20; Pže4; Pžp1,3;
Pžp2,3; Pžp5; Pžp6,2
krivica · Šze1,19; Šze2,5; Šze4,3; Šze5,5; Šzp1,2;
Šzp2,2; Šzp4,3; Šzp5,3
križ · Sme1,16; Sme2,3; Sme4,17; Sme5,3;
Sme6; Smp1,3; Smp2,2; Smp3; Smp4,9;
Smp5,4; Smp6,3
križati · Gcp,4; GLme,2; GLmp,3; GLže,2;
GLžp,5; GNE,2
križem · A,15; E2
križišče · Sse1; Sse5,10; Ssp5
Krn · IZme1,17; IZme2,7; IZme3,3; IZme4,4;
IZme5,3; IZme6,6
krog · A,2; E2,28; Sme1,3; Sme2,2; Sme4,4;
Sme5,9; Smp1,4; Smp4,2; Smp5,3
krogla · Šzd1; Šze1,8; Šze2,2; Šze4,6; Šze6,2;
Šzp1,15; Šzp2,2; Šzp6
krompir · Sme1,29; Sme2,10; Sme4,14; Smp2
krop · Sme1; Sme2,4; Sme5; Sme6,4
krota · Šze1; Šzp1,6; Šzp3; Šzp4,2; Šzp6,2
krožiti · Gcd,3; Gce; Gcp,3; GLmd; GLme,7;
GLmp; GLžd,2; GLže,3; GNE
krožnik · Smd6; Sme1,6; Sme2; Sme4,9;
Sme5,3; Sme6,2; Smp1; Smp4,5; Smp6
krst · Sme1,8; Sme2,2; Sme3; Sme5,3
krsta · Šze1,10; Šze2,2; Šze3,2; Šze4,16; Šze5,7;
Šze6,6; Šzp4,2
krstiti · Gcp,2; GLme,4; GLmp,9; GLžd; GLže
kruh · Sme1,32; Sme2,82; Sme4,40; Sme5,10;
Sme6,4
krvav · Pme1,6; Pme4; Pme4i,5; Pme5; Pmp1;
Pmp4; Pse1; Pže1,4; Pže2; Pže5,2; Pžp1;
Pžp4,2
kučma · Šze4,10; Šze5; Šze6,3
kuhar · Sme1,24; Sme2; Sme3; Sme4,3; Sme6
kuharica · Šze1,3; Šze2,3; Šze3; Šze4,4; Šze5;
Šzp4
kuhati · Gap; Gbe,2; Gce,8; Gcp,4; GLme,9;
GLmp,7; GLže,10; GLžp; GNE; GVbe
kuhinja · Šze1,13; Šze2,26; Šze3,5; Šze4,70;
Šze5,46; Šzp1; Šzp4; Šzp5,2
kuhinjski · Pme1; Pme3; Pme4; Pme5; Pmp1;
Pmp2; Psp2,4; Psp4,2; Psp5,3; Psp6
kup · Sme1,6; Sme2; Sme4,15; Sme5,10; Sme6;
Smp1,2; Smp4,2
kupiti · Gae,2; Gap,2; Gce,5; Gcp; GLmd;
GLme,30; GLmp,5; GLže,10; GNA,2; GNE,3;
GVbe,2
kura · Šze1; Šze3; Šze4; Šzp1,4; Šzp2,2; Šzp3;
Šzp4,4

kurir · Smd1; Sme1,7; Sme3; Sme6; Smp1,3;
Smp6
kurnik · Sme2,3; Sme4,3; Sme5,5
Kuščar · IOme1,13; IOme3; IOme4,2
Kutschera · IOme1,58; IOme2; IOme3,3;
IOme4,9
kvišku · A,105
lačen · Pmd1,3; Pme1i; Pme1,17; Pme4,2;
Pme6; Pmp1,8; Pmp4,2; Pmp6; Pse4; Pže1,2;
Pže6; Pžp2,2
ladja · Šze1,3; Šze2,2; Šze3; Šze4,2; Šze5; Šze6;
Šzp4; Šzp6
lagati · Gae,3; Gbe,4; Gce,5; Gcp; GLme,6;
GLže,2; GNE
Lah · IPme1,7; IPme3; IPme4,2; IPme6,3;
IPmp1,15; IPmp3; IPmp4,5; IPmp6,3
lahek · Pmd1,2; Pme1,4; Pme2jj; Pme6,2;
Pmp5; Pse1j; Pse1,5; Pse2; Pse2j; Pse6;
Pse6j; Psp4; Pže1j; Pže1,7; Pže3; Pže4,6;
Pže5; Pže6; Pžp4
lahko · Ajj,2; Aj,5; Aj,17; A,494; PD,7
lahkota · Šze2,2; Šze4; Šze6,7
lahkotno · A,13
lajati · Gae,3; Gce,7; Gcp; GLme,5; GLmp,3;
GLže; GNE
lakota · Šze1,12; Šze2,11; Šze4,6; Šze5,2; Šze6,5
las · Sme2,4; Sme4,6; Smp1,54; Smp2,31;
Smp5,25; Smp6,7; Šzp6,3
laso · Sme3; Smp4,67
lasten · Pme1i,5; Pme2,2; Pme4; Pme4; Pme4i;
Pme5,6; Pmp6,2; Pse1; Pse2,3; Pse4; Pse5,2;
Psp3; Psp4; Psp5; Pže1,3; Pže2,12; Pže4,9;
Pže5,8; Pže6,2; Pžp1; Pžp2,2; Pžp4,10;
Pžp5,2; Pžp6,3
laški · Pme1,10; Pme2; Pme3; Pme4; Pme6;
Pse2; Pse4; Pže1,3; Pže4,2; Pžp2
laz · Sme1,4; Sme2,2; Sme3,2; Sme4,8; Sme5,7;
Sme6,3; Smp1; Smp1; Smp1,2; Smp2,4; Smp3;
Smp4,6; Smp5,2; Smp5,2
laž · Šze1,4; Šze2; Šze4,2; Šze6; Šzp1; Šzp3;
Šzp4; Šzp5; Šzp6
le · Č,342; Vpo,8; Vpr,79
leči · Gae,4; Gbe,2; Gcd; Gce,6; GLmd;
GLme,33; GLmp,2; GLžd,2; GLže,14; GNE;
GVbe,7
led · Sme1,9; Sme2,3; Sme4,4; Sme5,4; Smp1
leden · Pme2,2; Pme4i; Pme6,2; Pmp1,2; Pmp2;
Psd1; Pse1,3; Pse4; Pže1,3; Pže5,3; Pže6;
Pžp1,4; Pžp2; Pžp5; Pžp6
legati · Gap; Gce; Gcp; GLme,2; GLmp,3; GLse;
GLže; GLžp

- lep · Pmd1; Pme1i,3; Pme1jj,7; Pme1j,2;
Pme1,66; Pme2j; Pme2jj; Pme2,21; Pme4,6;
Pme5jj,3; Pme5,7; Pme6; Pmp1jj; Pmp1,8;
Pmp2jj,2; Pmp2,2; Pmp4,5; Pmp5,2; Pse1j,2;
Pse1,80; Pse2j; Pse2jj,2; Pse2,13; Pse3;
Pse4,2; Pse5,3; Pse6; Psp1,4; Psp2,2; Psp5;
Psp5jj; Pzd1; Pže1jj,7; Pže1j,5; Pže1,106;
Pže2,6; Pže3; Pže4jj,2; Pže4,14; Pže5,8; Pže6;
Pžp1j; Pžp1jj; Pžp1,12; Pžp2jj; Pžp2,10;
Pžp4,6; Pžp5,3; Pžp6,3
- lepo · Ajj,3; Aj,11; A,146; Sse2; Sse5,14
- lepota · Šze1,5; Šze2,6; Šze4,5; Šze5,2; Šze6;
Šzp5
- les · Sme1,6; Sme2,12; Sme3; Sme4,6; Sme5,2;
Sme6,2; Šzp2
- lesen · Pme1,7; Pme2; Pme3; Pme4; Pme4;
Pme4i,3; Pme5,5; Pme6; Pme6; Pmp4; Pmp5;
Pse1,5; Pse5; Pse6; Pzd4; Pže1,4; Pže2,3;
Pže3; Pže4,10; Pže5; Pže6,2; Pžp1,3; Pžp2,2;
Pžp5
- lesketati se · Gap; Gce; Gcp; GLme; GLmp,2;
GLže; GLžp,5
- lesti · Gcd; Gce,9; Gcp,4; GLmd; GLme,6; GLsp;
GLže,4
- lestva · Šze4; Šze5,6; Šzp1,2; Šzp5
- letaló · Sse1,5; Sse2,4; Sse4; Sse5; Ssp1; Ssp2,3
- leteti · Gae,5; Gbe,2; Gce,6; Gcp,2; GLmd,2;
GLme; GLmp,4; GLse,3; GLže,3; GLžp,2
- leto · Ssd1,3; Ssd2; Ssd4,26; Ssd5,4; Ssd6;
Sse1,7; Sse2,27; Sse3,2; Sse4,74; Sse5,11;
Sse6,6; Ssp1,18; Ssp2,262; Ssp4,72; Ssp5;
Ssp5,66; Ssp6,47
- leva · Šze4,3; Šze5,9
- levi · Pme2,5; Pme5,2; Pmp4; Pse1; Pse2,2;
Pse4,13; Pse5,2; Pže1,3; Pže2,7; Pže4,25;
Pže5,6; Pže6
- levica · Šze1,2; Šze2,2; Šze4,5; Šze5,8; Šze6,13
- ležati · Gae,3; Gbe; Gcd,2; Gce,69; Gcp; Gcp;
Gcp,12; GLmd,4; GLme,3; GLme,67; GLmp,7;
GLse,7; GLsp; GLzd; GLže,52; GLžp,4;
GNE,3; GVbe,3
- ležišče · Sse1; Sse2,2; Sse4,8; Sse5,4
- Libero · IOme1,14; IOme3,2; IOme4
- lice · Ssd5; Sse2,5; Sse4,11; Sse5,15; Ssp1,16;
Ssp2,11; Ssp4,9; Ssp5,25; Ssp6
- lina · Šze1,3; Šze3; Šze4,16; Šze5; Šzp4
- Linika · IOže1,7; IOže2,2; IOže3,2; IOže4,2;
IOže5
- lipa · Šze1,2; Šze2; Šze4,2; Šze6,5; Šzp1; Šzp2;
Šzp4,3; Šzp6
- lira · Šzd4,4; Šze2,3; Šze4,4; Šzp2,31; Šzp4,3
- lisa · Šzd1; Šze1,5; Šze2; Šze5,2; Šze6; Šzp1,3;
Šzp4,3; Šzp5; Šzp6
- lisica · Šze4; Šzp3,4; Šzp4,12; Šzp6
- list · Smd1; Smd4; Sme1,15; Sme2,4; Sme3;
Sme4,25; Sme5,2; Sme6,2; Smp1,8; Smp2,2;
Smp4,4; Smp5; Smp6,4
- listič · Sme4; Smp1,11; Smp2; Smp4,4; Smp6,4
- listje · Sse1,15; Sse2,6; Sse4,18; Sse5,6; Sse6,2
- liter · Smd1; Smd4; Sme1; Sme2,5; Sme4,7;
Sme5; Sme6; Smp4,2
- liti · Gce; Gcp; GLme,3; GLse; GLže,8
- ljub · Pme1i,5; Pme4; Pme4i; Pmp1; Pse2;
Pže1jj; Pže4; Pže6jj
- ljubezen · Šze1,27; Šze2,25; Šze3,2; Šze4,8;
Šze5,11; Šze6,15; Šzp4,2
- ljubeznivo · A,17
- ljubiti · Gae; Gap,4; Gce,7; Gcp; GLme,5;
GLže,5; GNE,2
- ljudje · Smd5; Smp1,228; Smp2,46; Smp3,24;
Smp4,49; Smp5,21
- ljudski · Pme1; Pmp1; Pmp4; Pse2,4; Pse4;
Pse6; Psp4; Pže1,2; Pže2,2; Pže4,3; Pže5;
Pžp2,2; Pžp5,2
- ljudstvo · Sse1,4; Sse2,5; Sse4; Sse5,2
- lobanja · Šze1,2; Šze2; Šze4,6; Šze5; Šzp2
- ločiti · Gap; Gcd; Gce; GLmd,2; GLme,2; GLmp;
GLsp; GNE
- Lojz · IOme1,31; IOme6,2
- lok · M,3; Sme1; Sme5,21; Smp5,2
- Lokar · IOme1,15; IOme4,3
- lomiti · Gcd; Gce,14; GLme,4; GLmp,2; GLže,2;
GNE,6
- lonec · Smd4; Sme2,4; Sme4,30; Sme5,2;
Sme6,4; Smp4; Smp5,4; Smp6,4
- lopata · Šze4,7; Šze6; Šzp1; Šzp6
- lopatast · Pže6,3; Pžp4,4; Pžp5; Pžp6,2
- Loputnik · IOme1,17; IOme2; IOme3,10;
IOme4,3
- lotiti se · Gce,2; Gcp,2; GLmd; GLme,11;
GLmp,3; GLže,2; GNE
- loviti · Gae,2; Gap; Gbe; Gce,12; Gcp,3;
GLme,13; GLmp,6; GLže,7; GNE,2
- lubje · Sse1; Sse2,4; Sse4,4; Sse6
- luč · Šze1,38; Šze2,30; Šze3,5; Šze4,30; Šze5,11;
Šze6,10; Šzp1,6; Šzp2,2; Šzp6
- Ludvik · IOme1,35; IOme2; IOme4,8; IOme6,2
- Luka · IOme1,92; IOme3,6; IOme4,5; IOme5;
IOme6
- luknja · Šze1; Šze4,7; Šze5,4; Šzp2; Šzp4,2;
Šzp5
- Lužnica · IOže1,26; IOže3,2; IOže4; IOže6

- Lužnikov · PIOMe1,5; PIOMe2,2; PIOMe6;
PIOŽe1; PIOŽp4,2
- macesen · Sme1,15; Sme3; Sme4,3; Sme5; Sme6;
Smp1; Smp4; Smp6,2
- mačica · Šzp1,4; Šzp2,3; Šzp4; Šzp5; Šzp6,4
- Magajna · IOme1,9; IOme2,2
- mah · Sme1,3; Sme2; Sme4,17; Sme6,7
- mahati · Gae; Gce,5; Gcp,2; GLme,13; GLmp;
GLže,4; GLžp; GNE,4
- mahniti · Gad; Gae; Gce; Gcp; GLmd,3;
GLme,27; GLmp; GLže,3; GVap; GVbe,2
- maj · Sme1,4; Sme2,7; Sme4,5; Sme5,2
- majati · Gce,10; Gcp; Gcp; GLme,11; GLmp,5;
GLže,8; GLžp,2; GNE,2
- Majcen · IOme1,270; IOme2,7; IOme3,17;
IOme4,4
- majhen · Pme1j,2; Pme1j,11; Pme1,24;
Pme2jj,5; Pme4,4; Pme5,3; Pme6,2; Pmp1j;
Pmp1,6; Pmp2; Pmp2jj; Pmp3; Pmp4;
Pmp4jj; Pse1,6; Pse2; Pse2jj,2; Pse4,2; Psp1;
Psp2; Psp4; Psp5,2; Psp6; Pže1j; Pže1jj,2;
Pže1,27; Pže2; Pže3; Pže3jj; Pže4,5; Pže5,2;
Pžp1,5; Pžp2; Pžp3; Pžp4
- majolika · Šze1,4; Šze2; Šze4,10; Šze5; Šze6,3
- mak · Sme1,10; Sme4
- malce · A,74
- malenkost · Šze1,5; Šze2; Šze4,2; Šze5,2;
Šzp1,2; Šzp4; Šzp6
- Mali · IOme1,26; IOme2,2; IOme3,3; IOme4,2;
IOme5,2
- mali · Pme1i,5; Pme2,4; Pme4,4; Pme5,2;
Pme6,2; Pmp1,4; Pmp2,2; Pmp3; Pmp4,5;
Pmp6,2; Psd1; Pse1; Pse2,5; Pse4,2; Pse5;
Psp2; Pže1; Pže2; Pže3,2; Pže4,13; Pže5,4;
Pže6,2; Pžp1,2; Pžp2,3; Pžp4; Pžp6,2; Sme2,8;
Sme6,2; Smp1,2; Smp2,2; Smp3
- malo · A,116; Sse6; ŠNE,15
- mama · Šze1,244; Šze2,14; Šze3,41; Šze4,19;
Šze5,5; Šze6,5
- mamin · Pme1; Pse2; Pse3; Pže1; Pže2,3; Pže3;
Pže4,4; Pže5,3; Pže6
- manj · Ajj,30; A,27; ŠNE,4
- manjkati · Gce,21; Gcp,3; GLme,2; GLse,17;
GLže,2
- mar · A,62; Č,72
- marati · Gae,9; Gbe,2; Gce,11; Gcp,3; GLme,25;
GLmp,9; GLže,12; GVbe,3
- Marija · IOže1,8; IOže2,4; IOže3,3; IOže4,3;
IOže5; IVže1,3
- Marjanca · IOže1,51; IOže3,2; IOže4,4; IOže5;
IOže6
- Marjeta · IOže1,11; IOže3,2; IOže5; IOže6
- marsikaj · ZMse1,10; ZMse2,2; ZMse3;
ZMse4,21; ZMse5,2
- marsikateri · ZMme1; ZMme3; ZMme5;
ZMme6; ZMmp2; ZMsp4; ZMže1,3; ZMže2;
ZMže4
- marsikdo · ZMme1,6; ZMme2; ZMme3,2;
ZMme4
- Martin · IOme1,348; IOme2,8; IOme3,20;
IOme4,41; IOme5,2; IOme6,6
- Martinov · PIOMe1,4; PIOMe2; PIOMe4;
PIOMe4; PIOMe5,2; PIOMe6; PIOMP1,2;
PIOse1; PIOse4; PIOže1; PIOže4; PIOžp1;
PIOžp2; PIOžp4,2; PIOžp5,2
- maslo · Sse1,3; Sse2,3; Sse4,7; Sse5; Sse6,2
- masten · Pmp5; Psp4; Pžel,2; Pže2,3; Pže4,2;
Pže5; Pžp1; Pžp2; Pžp4,5; Pžp5; Pžp6
- maša · Šze1,3; Šze2,5; Šze3,15; Šze4,8; Šze5,9;
Šze6,9; Šzp2,2; Šzp4,4
- materin · Pme1,5; Pme4,2; Pse1; Pse2,2; Pse4;
Pse5,2; Pse6; Pže1; Pže2,3; Pže3; Pže4,3;
Pže5,2; Pže6; Pžp1,2; Pžp2,3; Pžp3; Pžp4;
Pžp5
- mati · Šze1,165; Šze2,35; Šze3,23; Šze4,48;
Šze5,8; Šze6,6; Šzp1,2; Šzp2,2; Šzp3,2; Šzp4
- Matic · IOme1,301; IOme2,2; IOme3,19;
IOme4,30; IOme6,3
- Matiček · IOme1,11; IOme3
- Matičev · PIOMe5; PIOMP1; PIOse1; PIOse4,3;
PIOsp2; PIOže1,2; PIOže2,3; PIOže5; PIOže6;
PIOžp2; PIOžp5
- Matija · IOme1,9; IOme4,2
- mazati · Gae; Gap; Gbe; Gce; GLme; GLmp;
GLže; GNE,2; GVbp
- med · E4,85; E6,359; Sme1,2; Sme2,2; Sme4,2;
Sme6
- medtem · A,50; Vpo,8
- medvejka · Šzp1,2; Šzp2,3; Šzp4,5; Šzp6,3
- megla · Šze1,9; Šze3; Šze4,3; Šze5,3; Šzp1
- meglica · Šze1,6; Šze4,5; Šze5; Šze6; Šzp5
- meh · Sme1,2; Sme4,9; Sme6
- mehak · Pme1,9; Pme4,2
- mehek · Pme1i,2; Pme2; Pme3; Pme4i; Pme5,2;
Pme6,10; Pmp1,5; Pmp4,3; Pmp5,2; Pse1;
Pse2,5; Pse4,4; Pse5; Pse6; Psp4; Pže1,14;
Pže2,2; Pže4,10; Pže5,6; Pže6,3; Pžp1,2;
Pžp2; Pžp5
- meja · Šze1,2; Šze2,9; Šze4,24; Šze5,11; Šzp1,2;
Šzp2; Šzp2,6; Šzp4,3; Šzp5,3; Šzp6,2
- meniti · Gbe; Gce,6; Gcp; GLmd,3; GLme,4;
GLmp,2; GLže
- mera · Šze1,2; Šze2,5; Šze4,12; Šze5,7

- meriti · Gae,2; Gce,2; GLme,5; GLmp; GLže;
GNE
- mesec · Smd1; Smd4,4; Smd5; Sme1,18; Sme2,4;
Sme3; Sme4,18; Sme5; Sme6,7; Smp1,4;
Smp2,10; Smp4,9; Smp5,4; Smp6
- mesečina · Šze1,8; Šze2,5; Šze4,3; Šze5,13;
Šze6,5
- meso · Sse1,10; Sse2,17; Sse4,3
- mesten · Pme1i,6; Pme2; Pme3; Pme5,2;
Pmp1,7; Pmp2,2; Pmp5,2; Pže1,2; Pže3,2;
Pže4,2; Pže5; Pže6; Pžp2
- mesto · Ssd5; Sse1,11; Sse2,24; Sse4,26;
Sse5,53; Sse6,3; Ssp2; Ssp4,2; Ssp5,2
- mešati · Gae; God; Gce,7; Gcp,2; GLmd,2;
GLme,2; GLse; GLžp,2; GNE
- metati · Gbe; Gce,6; Gcp,6; GLmd,2; GLme,10;
GLmp,5; GLže,2; GLžp; GNA; GNE
- meter · Sme1; Sme4,9
- metrum · Smd1; Smd2; Smd4,5; Sme2;
Smp2,10; Smp4,4
- mežikati · Gce,2; Gcp,2; GLme,17; GLžp;
GNE,3
- mhm · M,36
- Mica · IOže1,5; IOže4,6; IOže5,3; IOže6
- mil · Pme1; Pme1i; Pme1j; Pmp2; Pse6; Pže5,6;
Pžp1; Pžp4
- Milan · IOme1,6; IOme2,3; IOme4,2; IZme2;
IZme5,8
- Milica · IOže1,39; IOže2,3; IOže4,7; IOže5;
IOže6,3
- miličnik · Sme1,6; Sme3,2; Sme6; Smp1,6;
Smp2,4; Smp3,2; Smp4,4; Smp6,2
- milijon · Sme1,5; Sme2; Sme4; Smp1,3; Smp4,5
- milo · A,9; Sse2,3; Sse4; Sse5
- milost · Šze1,8; Šze2,2; Šze4,2
- mimo · A,75; E2,126
- mimogrede · A,23
- mina · Šze1,8; Šžp1,2; Šžp2
- miniti · Gce,10; GLme,5; GLmp,2; GLse,24;
GLsp; GLže,19
- minuta · Šzd4,2; Šze1,2; Šze2; Šze4,3; Šze5,3;
Šžp2,21; Šžp5,4; Šžp6
- mir · Sme1,49; Sme2,24; Sme3; Sme4,20;
Sme5,46; Sme6,9
- miren · Pmd1; Pme1j; Pme1,20; Pme2,3;
Pme6j; Pme6,9; Pmp1,7; Pmp5; Pse1,7;
Pse2,3; Pse4,2; Pže1,16; Pže2,6; Pže4,2;
Pže5,2; Pže6; Pžp1,3; Pžp2,2; Pžp4
- mirno · Aj,3; A,185
- misel · Šze1,65; Šze2,17; Šze3,5; Šze4,51;
Šze5,17; Šze6,10; Šžp1,29; Šžp2,26; Šžp3,2;
Šžp4,23; Šžp5,45; Šžp6,7
- misli · Gae,57; Gap,2; Gbd; Gbe,57; Gbp,21;
God; Gce,46; Gcp,8; GLmd,4; GLme,106;
GLmp,14; GLže,34; GLžp; GNE,29; GVbe,5
- mišica · Šžp1,7; Šžp4,2; Šžp6
- miza · Šzd4,4; Šzd5,2; Šze1,12; Šze2,46;
Šze3,61; Šze4,146; Šze5,83; Šze6,26; Šžp1,2;
Šžp4,6; Šžp5; Šžp6,3
- mlad · Pmd1,10; Pmd3j; Pmd4jj; Pme1i,12;
Pme1jj,4; Pme1j,9; Pme1,39; Pme2j; Pme2jj;
Pme2,8; Pme3,2; Pme4i; Pme4jj,4; Pme4j,3;
Pme4,6; Pme5jj; Pme6jj; Pme6,2; Pmp1jj,3;
Pmp1j,2; Pmp1,13; Pmp2jj; Pmp2,7; Pmp3j;
Pmp3,3; Pmp4jj; Pmp4,8; Pmp5,2; Pmp6jj;
Pse1,14; Pse2; Pse2,4; Pse3,3; Pse4,2; Pse5,2;
Pse6,2; Psp1; Psp2,4; Psp4; Psp5,6; Pžd3,2;
Pžd4j; Pže1jj,5; Pže1j,4; Pže1,38; Pže2,2;
Pže4,12; Pže5,3; Pže6,2; Pžp1,6; Pžp2,4;
Pžp4,3; Pžp5,3; Smp4
- mladika · Šze1,9; Šze4,7; Šze5,2; Šžp1,3;
Šžp2,2; Šžp4,2; Šžp5,2; Šžp6,3
- mladina · Šze1,5; Šze2,3; Šze3; Šze4,2; Šze5
- mladost · Šze1,7; Šze2,16; Šze4,3; Šze5,11
- mladosten · Pme1; Pme2; Pme6,2; Pmp1;
Pmp2; Pmp6; Psp1; Psp2; Pže1,2; Pže2,4;
Pže6
- mlatiti · Gcp; GLmd; GLme; GLmp,6; GLse;
GNE,3
- mleko · Sse1,9; Sse2,16; Sse4,14; Sse5,10;
Sse6,9
- mleti · Gce,4; Gcp; GLmd; GLme; GNE,3
- mlin · Sme1,24; Sme2,9; Sme3,3; Sme4,24;
Sme5,11; Smp2
- mnenje · Ssd4; Sse1,5; Sse2,2; Sse4,3; Sse5,10;
Ssp1; Ssp5,2
- množica · Šze1,23; Šze2,4; Šze3,8; Šze4,7;
Šze5,3; Šze6,2; Šžp2; Šžp4
- moč · Šze1,17; Šze2,19; Šze4,25; Šze5,6;
Šze6,12; Šžp1,8; Šžp2,5; Šžp4,8; Šžp5,7;
Šžp6,12
- močan · Pmd1; Pmeli,2; Pme1jj,3; Pme1j,2;
Pme1,33; Pme2j; Pme2,2; Pme4; Pme4,2;
Pme5j; Pme5,4; Pme6; Pmp1j,2; Pmp1,6;
Pmp4,4; Pmp5; Pse1j; Pse1,3; Pse2,4; Pse4,2;
Pse6; Psp1; Pžd4; Pže1j,4; Pže1,12; Pže2,2;
Pže4,2; Pže5; Pže6; Pžp1,5; Pžp2; Pžp2j;
Pžp4; Pžp5; Pžp6,5
- moči · Gad; Gae,65; Gap,10; Gbe,11; Gbp,3;
God; Gce,76; Gcp,15; GLmd,4; GLme,121;
GLmp,11; GLse,3; GLže,41; GLžp,2; PD
- močno · Aj,6; A,13

- moder · Pmeli,2; Pme1,8; Pme2; Pme4i,3;
Pme5; Pme6,7; Pmp1; Pmp2; Pmp4; Pse1,9;
Pse2,5; Pse3; Pse4,4; Pse5; Pže1,2; Pže4,5;
Pže5,2; Pže6; Pžp1,4; Pžp2,4; Pžp4,13;
Pžp5,2; Pžp6,13
- Modrijan · IOme1,97; IOme2,7; IOme3,15;
IOme4,11; IOme5,8; IOme6,5; IPmp1
- Modrijanov · IOmp1,2; PIOme1,12; PIOme2,3;
PIOme4; PIOme4; PIOme5,8; IOmp2;
PIOmp4; IOmp6; PIOse4,3; PIOse5; PIOsp3;
PIOzd1; PIOzd4; PIOže1,4; PIOže2,4; PIOže3;
PIOže4,5; PIOže5; PIOže6,5
- modrikast · Pme1,4; Pme4i; Pme4,2; Pme6;
Pmp1; Pse3; Pse5; Pže1,2; Pžp4,3
- mogoče · A,16
- mogočen · Pme1i,4; Pme1j; Pme1jj; Pme1,8;
Pme2,7; Pme4i; Pme4j; Pme5,3; Pme6;
Pmp2; Pse1,2; Pse2; Pse4; Psp6; Pže1,10;
Pže2; Pže4,3; Pže6,3; Pžp1; Pžp4; Pžp5
- moj · ZSVaemd1,5; ZSVaeme1,154;
ZSVaeme2,17; ZSVaeme3,7; ZSVaeme4,2;
ZSVaeme4,8; ZSVaeme5,19; ZSVaeme6,9;
ZSVaemp1,21; ZSVaemp2,7; ZSVaemp4,7;
ZSVaemp5,2; ZSVaes1,48; ZSVaes2,13;
ZSVaes3; ZSVaes4,17; ZSVaes5,8;
ZSVaes6,7; ZSVaespl,5; ZSVaespl,2,4;
ZSVaespl,2; ZSVaespl,5; ZSVaespl,6;
ZSVaeze1,86; ZSVaeze2,27; ZSVaeze3,2;
ZSVaeze4,30; ZSVaeze5,15; ZSVaeze6;
ZSVaezp1,11; ZSVaezp2,8; ZSVaezp3;
ZSVaezp4,12; ZSVaezp5,9; ZSVaezp6,3
- Moj Jezus · IOme1,152; IOme2,12; IOme3,20;
IOme4,10; IOme5,12; IOme6,2
- moka · Šze1,4; Šze2,19; Šze4,9; Šze5,2
- moker · Pmd1,2; Pme1,2; Pme4i,2; Pme5,4;
Pme6; Pmp1,3; Pmp4,2; Pse1,4; Pse4,3; Psp4;
Psp4,2; Psp6; Pže1,4; Pže2,2; Pže4,3; Pže5;
Pže6,3; Pžp1,7; Pžp2,3; Pžp4,14; Pžp5,3;
Pžp6,7
- molčati · Gae,4; Gae,7; Gap; Gbe; Gcd; Gce,20;
Gcp,4; GLmd,10; GLme,57; GLmp,22;
GLže,18; GLžp,3; GNE; GVbe; GVbp,12
- molče · A,84
- molitev · Smp4; Šze1; Šze3; Šze4,3; Šze5,3;
Šze6; Šžp1; Šžp4,2
- moliti · Gae; Gap; Gce,6; Gcp,2; GLme,8;
GLmp,10; GLže,9; GNA; GNE,2; GVap,2;
GVbe
- molk · Sme1,14; Sme2,4; Sme4,14; Sme5,9;
Sme6
- Mongol · IPme1,2; IPme2; IPmp1,9; IPmp3
- morati · Gad,7; Gae,59; Gap,31; Gbe,19; Gbp,4;
Gcd; Gce,160; Gcp,31; GLmd,5; GLme,108;
GLmp,29; GLse,9; GLže,38; GLžp,3
- morda · A,78
- moriti · Gce,6; GLme,2; GLže,2; GNE
- morje · Sse1,24; Sse2,25; Sse3,5; Sse4,24;
Sse5,21; Sse6,3; Ssp4; Ssp5
- mornar · Sme1,43; Sme2; Sme3; Sme6,2; Smp2;
Smp5
- most · Sme1,2; Sme2; Sme2,5; Sme3,2; Sme4,11;
Sme5,33; Sme6,6; Smp1; Smp2; Smp4,3;
Smp5
- moški · Pme1,3; Pme6,2; Pse1,2; Psp4; Pžd1;
Pže1,4; Pže2; Sme1; Smp1,12; Smp2,2; Smp4;
Smp5,2
- moten · Pmd1; Pme6; Pmp1; Pse1,3; Pse2,3;
Pse4,2; Pže1,2; Pže2; Pže4,2; Pže5,3; Pžp1,7;
Pžp2; Pžp4,3; Pžp6,3
- motiti · Gae,4; Gbe,4; Gce,7; Gcp,4; GLme,3;
GLmp; GLse,4; GLže,2; GNE
- motriti · Gcd; Gce,21; Gcp; GLme,8; GLže,6
- mož · Smd1,4; Sme1,76; Sme2,7; Sme3,19;
Sme4,27; Sme5,2; Sme6,5; Smp1,51; Smp2,3;
Smp3,5; Smp4,9; Smp5,3; Smp6
- možakar · Smd1; Sme1,15; Sme2,3; Sme3;
Sme4,2; Smp1,5; Smp2; Smp4
- možgani · Smp4,9; Smp5,4; Smp6
- mračen · Pmeli; Pme1,6; Pme2; Pme4,2;
Pme4,2; Pme5; Pme6,2; Pmp1,2; Pmp5; Pse1;
Pse4; Pse5; Pže1,5; Pže2; Pže4,6; Pže5,2;
Pže6; Pžp1; Pžp2; Pžp4
- mračno · A,23
- mrak · Sme1,9; Sme2,19; Sme4,5; Sme5,22;
Sme6,3
- mraz · Sme1,34; Sme2,6; Sme3; Sme5,2
- mreža · Šze1,2; Šze3,2; Šze4,3; Šze5,2; Šze6;
Šžp2
- mrha · Šze1,20; Šze2; Šze4
- mrk · Pmd1; Pme1i,2; Pme1,9; Pme2,2;
Pme4,2; Pmp1,11; Pmp2; Pmp5; Pse4; Psp6;
Pže1,2; Pže3; Pžp1,2; Pžp4; Pžp5
- mrko · A,35
- mrlič · Sme1,5; Sme4,2; Sme5,4; Smp2; Smp3;
Smp4,3
- mrmrati · Gce,3; GLme,4; GLmp; GLse; GLže,6
- mrtve · Pmd1,3; Pme1i,5; Pme1,36; Pme2,3;
Pme3; Pme4; Pme4i,2; Pme4,7; Pme5,2;
Pme6; Pmp1,7; Pmp2,4; Pmp5; Pse1,10;
Pse2; Pse5,2; Pže1,14; Pže3; Pže4,4; Pže6;
Pžp1,7; Pžp4,3; Pžp5; Pžp6,2; Sme4; Smp1,2
- mrtvaški · Pme1,2; Pme2,3; Pme4,6; Pme5,5;
Pme6; Pmp6; Pže1; Pže3; Pže4; Pžp1

- mrtvi · Smp2,8; Smp3; Smp4,2
 mrzel · Pmd1; Pme1i,5; Pme1,1,2; Pme2,2;
 Pme4; Pme4i,2; Pme5,6; Pme6,4; Pmp1,2;
 Pmp2; Pmp5; Pmp6,2; Pse1,4; Pse2,3; Pse4,3;
 Pse5,3; Psp4,2; Psp5; Pže1,18; Pže2,4; Pže3;
 Pže4,8; Pže5,6; Pže6,4; Pžp1,3; Pžp2,2;
 Pžp4,5; Pžp5,3; Pžp6,2
 mrzlo · A,16
 mučen · Pme1i,2; Pme1j; Pme1,3; Pme2;
 Pmp1,2; Pmp4; Pmp5; PNse4; Pse1,4; Pse2,2;
 Pse5; Pže1j; Pže1,3; Pže4,2; Pže5; Pžp2,3
 mučenik · Sme4,3; Smp1,7; Smp2,3; Smp3,2;
 Smp4,2
 mučiti · Gce,2; Gcp,3; GLme,10; GLmp,5; GLse;
 GLže; GLžp; GNE
 muditi se · Gce,2,2; GLmp; GLse,5
 muha · Šze1,7; Šze2; Šze4,6; Šzp1,7; Šzp2,2;
 Šzp4,3
 muka · Šze1; Šze4,2; Šze6,11; Šzp1,2; Šzp2;
 Šzp4,4; Šzp5,3
 Mussolini · IOme1,6; IOme2,2; IOme3,3; IOme4
 muuu · M,17
 na · E4,2847; E5,2152; E,65; M,23
 nabirati · Gae; Gae,6; Gce,6; Gcp,4; GLmd,5;
 GLme; GLmp,4; GLže; GLžp,5; GNA; GNE,2
 nabrati · Gae,2; Gce,3; Gcp,2; GLme,7;
 GLmp,6; GLse,5; GLže,15; GLžp,3
 nabrekel · Pme2; Pse1; Pže3; Pžp1,4; Pžp2;
 Pžp5; Pžp6
 Nace · IOme1,20; IOme3,2; IOme4,2; IOme4,3;
 IOme5
 načelnik · Sme1,8; Sme3,2
 način · Sme4,23; Sme5,2; Smp4,2
 načrt · Sme1,2; Smp2,2; Smp4,5; Smp5
 nad · E4,61; E6,296
 nadaljevati · Gae; Gce,8; GLme,37; GLmp;
 GLže,17; GNE
 nadstropje · Sse2,4; Sse4,2; Sse5,5
 nadzornik · Sme1,10; Sme3; Sme4
 naenkrat · A,20
 nag · Pme1,7; Pme2,4; Pme4; Pmp1,5; Pse5;
 Pžp1
 nagajati · Gce,2; Gcp; GLme,4; GLmp,3; GLsp;
 GLže; GLžp; GNE
 nagel · Pme1; Pme2; Pme4i; Pme5; Pmp1;
 Pmp6,2; Pže1,2; Pže2,3; Pže4; Pže6
 nagelj · Sme1,2; Sme2; Sme4,5; Sme5; Sme6,3;
 Smp1,5; Smp2,5; Smp5
 nagibati · Gae,2; Gce,2; Gcp; GLme,11; GLže,5;
 GLžp,2; GNE
 naglas · A,10; Sme6,4
 naglica · Šze5; Šze6,28
 naglo · A,355
 nagnati · Gae; Gap; Gce; Gce,2; Gcp,2; GLme,14;
 GLmp,8; GLže,4
 nagniti · Gae,3; Gce,19; Gcp; GLme,45;
 GLmp,2; GLse; GLže,16; GLžp; GNE,2; GVbe
 nagonsko · A,30
 nahrbtnik · Sme1; Sme2; Sme4,15; Sme5,2;
 Sme6,3; Smp5
 naj · Č,576
 najbrž · A,40
 najbrže · A,20
 Najdù · IOme1,104; IOme3,2; IOme4,2
 najesti · Gce; GLmd; GLme,4; GLmp,6; GLže,4
 najti · Gae,3; Gbe,2; Gcd,2; Gce,15; Gcp,3;
 GLmd,8; GLme,79; GLmp,37; GLse,3; GLžd;
 GLže,18; GLžp,2; GNE,20
 naključje · Ssd1; Sse1,11; Sse3; Sse5,4; Sse6;
 Ssp1,4; Ssp3
 nalašč · A,19
 naleteti · Gap; GLmd,3; GLme,17; GLmp,2;
 GLže,3
 naliti · Gbe; Gcd; Gce,3; GLme,16; GLmp,2;
 GLže,3
 nalivati · Gce,2; GLme,4; GLmp; GLže; GNE,4
 naloga · Šze1; Šze3; Šze4,3; Šze6; Šzp1; Šzp4,3
 naložiti · Gce; GLmd; GLme,7; GLmp,2; GLže
 namen · Sme1; Sme2,2; Sme4,8; Sme6,3;
 Smp1,3; Smp4,4; Smp5
 namenjen · GNme1,10; GNmp1,2; GNžp1,2;
 PNže1
 nameravati · Gce,2; GLme,9; GLže,2
 nameriti · Gce; GLme,9; GLže
 namesto · A,4; Č,2; E2,8
 namrdniti se · Gce,3; GLme,6; GLže,2
 namreč · A,19; Vpr,11
 namršiti · Gce,7; Gcp; GLme,8; GLže,3
 Nanca · IOže1,106; IOže2,4; IOže3,9; IOže4,12;
 IOže5; IOže6,2
 Nančika · IOže1,74; IOže2; IOže3,10; IOže4,23;
 IOže5,3
 Nande · IOme1,22; IOme2; IOme5; IOme6,2
 nanesti · Gce,2; GLme,3; GLmp; GLse,5; GLže;
 GLžp,3
 napasti · Gae; Gce; GLme; GLme,16; GLmp,4
 napenjati · Gae; Gce,8; Gcp; GLme,5; GLže
 naperjen · GNme1; PNme6,3; PNmp6; PNže6;
 PNžp4; PNžp6,6
 napet · GTme1; GTse1,2; Pmp5,3; PNse2;
 PNže5,2; PNžp1; PNžp4; PNžp5; Pse1,3;
 Psp1; PTme6; PTmp6,2; PTse6; Pže1,5;
 Pže4,2

napeti · Gce,4; Gcp; Gcp,4; GLme,11; GLmp,2;
GLse,3; GLže,5; GLžp,2; GVbe
napeto · A,26
napis · Sme1,7; Sme2,2; Sme4,6; Sme6,5;
Smp1,3; Smp4,3; Smp6
napisan · GNmp1; GNse1,7; GNže1,2; GNže4;
GNžp1,3; PNmp1; PNmp2; Pže4
napisati · Gae,6; Gbp,3; Gcp,4; GLme,42;
GLmp,6; GLže,4; GNE,13; GVbe; GVbp,2
napiti se · Gbe,2; Gce; Gcp,2; GLmd,4;
GLme,11; GLmp,4
napol · A,14
napolniti · Gce,2; GLmp,4; GLse; GLže,3;
GLžp,2
napolnjen · GNme1; GNže1,5; GNže5; GNžp1;
PNse1,4
napor · Sme1; Sme2,2; Sme5; Sme6,11
naposled · A,124
napota · Šze4,17
napotiti · Gap; Gcp; GLme,30; GLmp,5; GLže,3;
GNE
napraviti · Gce,2; Gcp; GLme,9; GLmp; GNE;
GVbe,2
naprej · A,238
naproti · A,26
narava · Šze1,7; Šze2,3; Šze4; Šze5,5; Šze6
naraven · Pmp1; Pse1,7; Pžel; Pže2; Pžp1;
Pžp2
naravnost · A,148
narediti · Gae; Gap; Gbp,3; Gce,5; GLmd,2;
GLme,33; GLmp; GLže,4; GNE,2
narejen · GNme1,2; GNse6; PNmp1; PNže6,10
narobe · A,21
naročiti · GLme,17; GLmp; GLže,3
naroče · Sse2,8; Sse4,41; Sse5,28; Sse6; Ssp5
narod · Sme1,6; Sme2,5; Sme3,4; Sme4,4; Sme5;
Smp1,2; Smp2,2; Smp3,3
naroden · Pmd4; Pme1; Pmp5; Pse4; Pže1,3;
Pže2,6; Pže3; Pže4; Pže5; Pžp2; Pžp4
Narte · IOme1,25; IOme3; IOme4
naseliti se · Gcp; GLmd; GLme,4; GLmp;
GLže,3
naslanjati se · Gce,6; Gcp,2; GLme,4; GLmp;
GLže
naslednji · Pme4,16; Pme5; Pme6; Pmp1;
Pmp4; Pse1; Pse4,7; Pže4; Pžp6,2
nasloniti · Gce,8; Gcp; GLmd; GLme,42; GLmp;
GLže,19
naslov · Sme1,3; Sme2,2; Sme4,3; Sme6,5
nasmeh · Sme1,14; Sme2,2; Sme4,10; Sme5,4;
Sme6,14; Smp1

nasmehniti se · Gad; Gae,3; Gce,41; Gcp,3;
GLmd,2; GLme,135; GLmp,6; GLže,47;
GNE,10
nasmihati se · Gce,2; Gcp,4; GLme,3; GLmp,2;
GLže,2
nasproten · Pme4i; Pme5; Pse2,2; Pse4; Pse5,2;
Pže1; Pže4,2; Pže5,2
nastati · Gce; GLme,3; GLse,2; GLže,9
nastlan · GNme1,2; GNse1,2; GNsp1; GNže1;
PNmp1; PNse1; PNže1,2; Pse4
naš · ZSVapmd1; ZSVapmd3; ZSVapme1,60;
ZSVapme2,11; ZSVapme3,5; ZSVapme4,9;
ZSVapme4,12; ZSVapme5,7; ZSVapme6,5;
ZSVapmp1,22; ZSVapmp2,8; ZSVapmp3,5;
ZSVapmp4,10; ZSVapmp5,7; ZSVapmp6,3;
ZSVapse1,6; ZSVapse2,9; ZSVapse3,2;
ZSVapse4,3; ZSVapse5,7; ZSVapsp1,5;
ZSVapsp2,2; ZSVapsp4,2; ZSVapsp5,5;
ZSVapže1,75; ZSVapže2,27; ZSVapže3,5;
ZSVapže4,22; ZSVapže5,32; ZSVapže6,5;
ZSVapžp1,10; ZSVapžp2,2; ZSVapžp3;
ZSVapžp4,13; ZSVapžp5,2
našobiti · Gce,3; GLme,3; GLže,8
natančno · Aj; A,24
nato · A,993
natočiti · GLme,8; GLmp,2; GLže; GNE
naučiti se · Gae; Gbe; Gce,2; GLme,21; GLmp;
GLse; GLže,2; GNE,2; GVbe
nauk · Sme2,2; Sme4,2; Sme5; Sme6,2; Smp2;
Smp4,4; Smp6,3
navada · Šze1,15; Šze2,11; Šze4,7; Šze5,30;
Šzp1; Šzp2; Šzp3; Šzp4,4; Šzp5
navaden · Pme1,3; Pme2; Pme3,2; Pme4;
Pme6,3; Pmp1,6; Pmp2; Pmp4; Pmp5,2;
Pse1,2; Pse2; Pže1,4; Pže4,2; Pžp1
navdušeno · A,28
navdušiti · Gae; Gce,2; GLme,4; GLse; GLže;
GLžp; GNE
naveličan · PNme6,6; Pse1,2; Pže6; Pžp1
naveličano · A,19
navkreber · A,38
navpičen · Pse1; Pse5; Psp1,2; Psp6; Pže1,3;
Pže4; Pže6,2; Pžp1; Pžp2; Pžp6,2
navsezadnje · A,11
navzdol · A,19
navzgor · A,25
nazaj · A,71
ne · ČZ,3215; Vpo,150; Vpr,331
nebesa · Ssp1,6; Ssp2,2; Ssp4,9; Ssp5,14
nebeški · Pme1,6; Pme2; Pme4,2; Pse1; Pse4,3;
Pse5; Pžel,6; Pžp2

- nebo · Sse1,40; Sse2,55; Sse3,9; Sse4,66;
Sse5,42; Sse6,5
- nebogljen · Pme1j; Pme1,11; Pmp1,2; Pže1,2
- nečloveški · Pme1,6; Pme6,3; Pmp6,2; Pse1;
Pse2; Pže1,4; Pžp1
- Neda · IOže1,21; IOže2; IOže4,2; IOže6
- nedelja · Šzd5; Šže1,15; Šže2,4; Šže4,40; Šže6;
Šžp1,2; Šžp2; Šžp4,2; Šžp5,14
- nedolžen · Pme1i,3; Pme1,2; Pme2; Pme4,3;
Pme6,2; Pmp1; Pse1; Pže1,6; Pže2; Pžp1
- negiben · Pme1i; Pme1,10; Pme4i; Pmp1,5;
Pse1,2; Pže1,6; Pže3; Pžp1; Pžp4
- negibno · A,14
- nehati · Gbe,2; Gce,4; GLmd; GLme,8; GLse,2;
GLže,3; GNE; GVad,2; GVbe,22; GVbp,2
- nehote · A,58
- nejasen · Pmeli; Pme6; Pse1; Pže3; Pže4;
Pže6,2; Pžp2; Pžp4,3; Pžp6,13
- nejeverno · A,12
- nejevoljno · A,15
- nek · ZNEme1,8; ZNEme2,31; ZNEme4;
ZNEme4,9; ZNEme5,4; ZNEme6,9; ZNEmp1;
ZNEmp4; ZNEse1,5; ZNEse2,4; ZNEse4,7;
ZNEse5; ZNEse6,5; ZNEže1,22; ZNEže2,22;
ZNEže4,9; ZNEže5,5; ZNEže6,3; ZNEžp3
- nekaj · ŠNE,240; ZNEse1,140; ZNEse2,13;
ZNEse4,152; ZNEse5,5; ZNEse6; ZNEse6,3;
ZNE,20
- nekajkrat · A,32
- nekak · ZNEme1; ZNEme5; ZNEse1,15
- nekako · ZNE,19
- nekakšen · ZNEme1,6; ZNEme2,2; ZNEme3;
ZNEme4; ZNEme5,3; ZNEme6,8; ZNEmp1;
ZNEmp4; ZNEmp6; ZNEse1,3; ZNEse6,2;
ZNEsp1; ZNEsp4; ZNEže1,3; ZNEže2,3;
ZNEže4,5; ZNEže5,3; ZNEže6,7; ZNEžp1;
ZNEžp2; ZNEžp4,2
- nekam · ZNE,90
- nekateri · ZNEmp1,32; ZNEmp2; ZNEmp3,4;
ZNEmp4,2; ZNEmp5,2; ZNEsp1; ZNEže5;
ZNEžp1,5; ZNEžp4,3; ZNEžp5
- nekdaj · A,13
- nekdanji · Pme1,2; Pme2,4; Pme4; Pmp2,3;
Pmp4; Psp2; Pže1; Pže2,3; Pže3; Pže5
- nekdo · ZNEme1,31; ZNEme3; ZNEme4,6;
ZNEme6,3
- neki · Č,20
- nekje · ZNE,82
- nekoč · A,63
- nekoliko · ZNE,129
- nem · Pme1,2; Pmp1; Pse6; Psp1,2; Pže1,3;
Pže2; Pže5; Pžp1,2; Pžp5
- nemara · Č,226
- Nemec · IPme1,3; IPme6; IPmp1,16; IPmp2,2;
IPmp3; IPmp4,3; IPmp5; IPmp6
- nemir · Sme1,10; Sme2,6; Sme3; Sme4,3;
Sme5,2
- nemiren · Pme1,9; Pmp5; Pse4; Psp2; Pže1,4;
Pže2,2; Pžp1; Pžp2
- nemirno · A,22
- nemo · A,10
- nemogoč · Pme6; Pse1,10; Pže1,2; Pžp4,2
- nemški · Pmd1; Pme1,15; Pme2,4; Pme3;
Pme4; Pme5; Pme6; Pmp1,15; Pmp2,2;
Pmp6,2; Pže1,48; Pže4,5; Pže5; Pžp1,2; Pžp6
- nenadoma · A,153
- nenavaden · Pme1i,2; Pme1,5; Pme4; Pme6;
Pmp1,3; Pse1,6; Pse2,14; Pse4; Pže1,4; Pže2;
Pže4; Pže5; Pže6,4; Pžp1
- nenavadno · A,57
- nenehno · A,12
- nepojmljiv · Pme1; Pme6; Pse1; Pse4; Pse5;
Pse6; Pže1; Pže2; Pže4; Pže6
- nepremičen · Pme1,3; Pmp1; Pmp2; Pse1;
Pže1,2; Pže6; Pžp1,3
- nepremično · A,118
- neprestano · A,22
- neprijeten · Pmeli; Pme1,5; Pme4; Pmp5;
Pse1,12; Pse2,2; Pže1,5; Pže2; Pže4,2; Pžp4,2
- neprijetno · A,14
- nerazumljiv · Pme1; Pme5; Pme6; Pse1,3;
Pse2,4; Pse6; Pže1,5; Pže4; Pže6; Pžp1,2;
Pžp4,2
- neroden · Pme1i,2; Pme1,2; Pme4i; Pmp6;
Pse1,21; Pse4,2; Pže1; Pže5; Pže6; Pžp4;
Pžp6
- nerodno · A,27
- neskončen · Pme1; Pme3,2; Pme4; Pme4i;
Pme4,2; Pmp1; Pse1,2; Pse2; Pse5; Psp4;
Pže1,7; Pže3; Pže4,3; Pžp1,2; Pžp4; Pžp5
- neskončno · A,10
- neslišno · A,28
- nesreča · Šze1,22; Šze2,10; Šze4,6; Šze5,5; Šžp1;
Šžp2; Šžp4,2
- nesrečen · Pmeli,17; Pme1j; Pme1,10; Pme2,3;
Pme3; Pme4i; Pme4,5; Pme5; Pme6; Pmp1,3;
Pmp3; Pmp5; Pse1; Pse2,2; Psp5jj; Pže1,10;
Pže3; Pže4,3; Pžp1
- nesrečnik · Sme1,15; Sme3; Sme4,2; Smp1,4;
Smp2; Smp3; Smp4,4; Smp6
- nesti · Gad; Gae,2; Gap; Gbe; Gcd,4; Gce,10;
Gcp,2; Gcp,5; GLmd,3; GLme,26; GLmp,12;
GLže,13; GLžp,8; GNE,8; GVbe; GVbp
- nestrpen · Pme1,6; Pme4,2; Psp2; Pžel

nestrpno · A,40
 neumen · Pme1; Pme1,3; Pmp1,3; Pmp3;
 Pse1,6; Pže1,5; Pže2,2; Pže3; Pžp1,3; Pžp4,2
 neumnost · Šze1,4; Šze2,2; Šze3; Šze4,2; Šzp1,3;
 Šzp2; Šzp4,12
 neusmiljeno · A,21
 nevaren · Pme1,10; Pme2; Pmp1; Pmp4;
 Pse1,3; Pse2; Psp1; Pžel,3; Pže3; Pže4
 nevarnost · Šze2,6; Šze3,2; Šze4,5; Šze5,4; Šzp2;
 Šzp4
 neverjetno · A,17
 nevesta · Šze1,13; Šze2,3; Šze4,7; Šze6,3; Šzp1
 neviden · Pme1; Pme1,3; Pme2; Pme6,2;
 Pmp4; Psp5,2; Pžel,2; Pže4; Pže6,4; Pžp1,2;
 Pžp4,3
 nevihta · Šze1,3; Šze2; Šze4,2; Šze5,3; Šze6
 nezakonski · Pme1,3; Pme2; Pme3; Pme4,3;
 Pme5; Pmp2,2; Pmp3,2; Pse4,2; Psp1; Pže2;
 Pže4
 nezaslišan · Pme4; Pmp1; Pse2,5; Pže2,2; Pže5;
 Pžp4,2
 neznan · Pmd1; Pme1; Pme1,3; Pme2,2; Pme4;
 Pme4i; Pme6; Pmp2; Pmp4; Pse2,4; Pse4,2;
 Pse6; Pžel,6; Pže2,3; Pže4,3; Pžp2,2; Pžp4,2
 neznanško · A,26
 Neža · IOže1,13; IOže3; IOže4
 ne biti o · GZOae,7; GZOap,3; GZObe,3; GZObp;
 GZOce,274; GZOcp,10
 ne biti p · GZOad; GZPad,10; GZPae,304;
 GZPap,52; GZPbd,5; GZPbe,49; GZPbp,17;
 GZPcd,57; GZPce,1908; GZPcp,244
 ne biti r · GZRae,35; GZRap,9; GZRbe,21;
 GZRbp,8; GZRcd,7; GZRce,454; GZRcp,42
 ne hoteti · Gae,11; Gcp,4
 ne imeti · GZad,2; GZae,42; GZap,13; GZbd;
 GZbe,18; GZbp,5; GZce,84; GZcp,15
 nič · Sme4,2; ZNIse1,138; ZNIse2,31;
 ZNIse4,67; ZNIse5,2; ZNIse6; ZNI,297
 nihalo · Sse1,3; Sse2; Sse4,7
 nihče · ZNIme1,136; ZNIme2; ZNIme2,4;
 ZNIme3,18; ZNIme6,4
 nikakor · ZNI,32
 nikamor · ZNI,36
 nikar · ZNI,91
 nikdar · ZNI,225
 nikjer · ZNI,56
 nikoli · ZNI,43
 nit · Šze1,4; Šze2; Šze4; Šze6,2; Šzp1; Šzp4
 niti · Č,155; Vpr,93
 nizdol · A,21

nizek · Pmd1; Pme1i,2; Pme1,11; Pme4i,2;
 Pme4j,2; Pme5,3; Pme6,7; Pmp1; Pmp4;
 Pmp5,2; Psd5; Pse1; Pse2,2; Pse4,2; Pse6;
 Psp4,2; Pžel,4; Pže4j; Pže5; Pže5,3; Pže6;
 Pžp2; Pžp4; Pžp5
 nizko · Aj; Aj,19; A,23
 njegov · ZSVcmemd1,2; ZSVcmemd4;
 ZSVcmeme1,109; ZSVcmeme2,21;
 ZSVcmeme3,4; ZSVcmeme4,4;
 ZSVcmeme4,24; ZSVcmeme5,29;
 ZSVcmeme6,14; ZSVcmemp1,41;
 ZSVcmemp2,5; ZSVcmemp3,2;
 ZSVcmemp4,13; ZSVcmemp6,5;
 ZSVcmese1,40; ZSVcmese2,23; ZSVcmese3,5;
 ZSVcmese4,39; ZSVcmese5,20; ZSVcmese6,7;
 ZSVcmesp1,6; ZSVcmesp2,6; ZSVcmesp4,3;
 ZSVcmesp5,3; ZSVcmeže1,133;
 ZSVcmeže2,38; ZSVcmeže3,2; ZSVcmeže4,51;
 ZSVcmeže5,34; ZSVcmeže6,12;
 ZSVcmežp1,35; ZSVcmežp2,16; ZSVcmežp3,3;
 ZSVcmežp4,34; ZSVcmežp5,25; ZSVcmežp6,4
 njen · ZSVčzemd1,3; ZSVčzeme1,63;
 ZSVčzeme2,23; ZSVčzeme3,3; ZSVčzeme4,5;
 ZSVčzeme4,25; ZSVčzeme5,21; ZSVčzeme6,8;
 ZSVčzemp1,17; ZSVčzemp2,3; ZSVčzemp4,7;
 ZSVčzemp5,4; ZSVčzemp6; ZSVčzese1,26;
 ZSVčzese2,7; ZSVčzese3; ZSVčzese4,13;
 ZSVčzese5; ZSVčzese6,2; ZSVčzesp1,3;
 ZSVčzesp2,4; ZSVčzesp5,2; ZSVčžežd4;
 ZSVčžežel,36; ZSVčžeže2,18; ZSVčžeže3,5;
 ZSVčžeže4,16; ZSVčžeže5,12; ZSVčžeže6,2;
 ZSVčžežp1,20; ZSVčžežp2,14; ZSVčžežp3,2;
 ZSVčžežp4,10; ZSVčžežp5,10; ZSVčžežp6,7
 njihov · ZSVcpme1,6; ZSVcpme2; ZSVcpme6,3;
 ZSVcpmp1,9; ZSVcpmp4,5; ZSVcpmp5;
 ZSVcpmp6,3; ZSVcppe1,3; ZSVcppe2,2;
 ZSVcppe3; ZSVcppe4,5; ZSVcppe5; ZSVcppe6;
 ZSVcppe7; ZSVcppe8; ZSVcppe9; ZSVcppe10;
 ZSVcppe11; ZSVcppe12; ZSVcppe13; ZSVcppe14;
 ZSVcppe15; ZSVcppe16; ZSVcppe17; ZSVcppe18;
 ZSVcppe19; ZSVcppe20; ZSVcppe21; ZSVcppe22;
 ZSVcppe23; ZSVcppe24; ZSVcppe25; ZSVcppe26;
 ZSVcppe27; ZSVcppe28; ZSVcppe29; ZSVcppe30;
 ZSVcppe31; ZSVcppe32; ZSVcppe33; ZSVcppe34;
 ZSVcppe35; ZSVcppe36; ZSVcppe37; ZSVcppe38;
 ZSVcppe39; ZSVcppe40; ZSVcppe41; ZSVcppe42;
 ZSVcppe43; ZSVcppe44; ZSVcppe45; ZSVcppe46;
 ZSVcppe47; ZSVcppe48; ZSVcppe49; ZSVcppe50;
 ZSVcppe51; ZSVcppe52; ZSVcppe53; ZSVcppe54;
 ZSVcppe55; ZSVcppe56; ZSVcppe57; ZSVcppe58;
 ZSVcppe59; ZSVcppe60; ZSVcppe61; ZSVcppe62;
 ZSVcppe63; ZSVcppe64; ZSVcppe65; ZSVcppe66;
 ZSVcppe67; ZSVcppe68; ZSVcppe69; ZSVcppe70;
 ZSVcppe71; ZSVcppe72; ZSVcppe73; ZSVcppe74;
 ZSVcppe75; ZSVcppe76; ZSVcppe77; ZSVcppe78;
 ZSVcppe79; ZSVcppe80; ZSVcppe81; ZSVcppe82;
 ZSVcppe83; ZSVcppe84; ZSVcppe85; ZSVcppe86;
 ZSVcppe87; ZSVcppe88; ZSVcppe89; ZSVcppe90;
 ZSVcppe91; ZSVcppe92; ZSVcppe93; ZSVcppe94;
 ZSVcppe95; ZSVcppe96; ZSVcppe97; ZSVcppe98;
 ZSVcppe99; ZSVcppe100
 njiva · Šzd1; Šze1,18; Šze2,31; Šze3,9; Šze4,15;
 Šze5,30; Šze6,3; Šzp1,16; Šzp2,23; Šzp4,15;
 Šzp5,18; Šzp6,5
 no · Č,500
 noben · ZNIme1,2; ZNIme2,54; ZNIme3,3;
 ZNIme4,2; ZNIme4,6; ZNIme5; ZNIme6;
 ZNImp1; ZNIse1,3; ZNIse2,17; ZNIse4,2;
 ZNIže1,13; ZNIže2,80; ZNIže4,4; ZNIže5;
 ZNIžp2,5
 nocoj · A,69

- noč · Sžd4; Sže1,63; Sže2,44; Sže4,109; Sže5,26;
Sže6,7; Sžp1,3; Sžp2; Sžp4,18; Sžp5,6
- nočen · Pme2,3; Pme5,2; Pmp2; Pmp4; Pse2;
Psp4,2; Pže1,3; Pže2; Pže3; Pže4; Pžp1; Pžp5
- noga · Smp5,3; Sžd1; Sžd3; Sžd5; Sžd6,7;
Sže1,14; Sže2,8; Sže3; Sže4,29; Sže5,3;
Sže6,21; Sžp1,55; Sžp2,35; Sžp3,5; Sžp4,122;
Sžp5,23; Sžp6,55
- nor · Pme1,16; Pme6; Pmp1; Pse1,2; Pže1
- norček · Sme1,63; Sme2; Sme3,2; Sme4,2; Smp1
- norec · Sme1,12; Sme3,2; Sme4,3; Smp1,9;
Smp2,5; Smp3; Smp4,6; Smp6,2
- noreti · Gae; Gbe,3; Gce,9; GLže; GVbe
- nos · Sme1,19; Sme2,18; Sme2,24; Sme3;
Sme4,73; Sme5,24; Sme6,48; Smp2; Smp4,11
- nosilo · Sse6; Ssp2,2; Ssp3; Ssp4,2; Ssp5,5; Ssp6
- nositi · Gap; Gbe,2; Gbp,4; Gce,17; Gcp,14;
GLmd,3; GLme,34; GLmp,8; GLže,6; GLžp,3;
GNE,6
- nosnica · Sže4; Sžp1,2; Sžp4,9; Sžp5; Sžp6,2
- notranji · Pme1; Pme2,3; Pme3; Pme4,3; Pse1;
Pse2; Pse4,3; Pse5; Pže1,2; Pže2; Pže5,2;
Pžp5
- nov · Pme1i,12; Pme1,14; Pme2,4; Pme3;
Pme4,7; Pme4,12; Pme5; Pme6jj; Pme6,2;
Pmp1,3; Pmp2jj; Pmp2,5; Pmp4,2; Pmp5;
Pmp6; Pse1,5; Pse2,12; Pse3; Pse4,8; Pse5,5;
Pse6,2; Psp2; Psp4,4; Psp6; Pže1,11; Pže2;
Pže2jj; Pže3; Pže4,27; Pže5,3; Pže6,7; Pžp1,5;
Pžp2,9; Pžp4,11; Pžp5,2; Pžp6,4
- novela · Sže1,3; Sže2,4; Sže4,3; Sže5; Sžp2;
Sžp4,2; Sžp5
- novica · Sže1,3; Sže4,8; Sže5; Sže6,2
- nož · Sme2; Sme4,20; Sme5; Sme6,19; Smp1;
Smp4; Smp6,2
- nuditi · Gce,4; Gcp,3; GLse,2; GLže
- o · E5,772; M,232
- ob · E4,131; E5,562
- oba · ZTmd1,68; ZTmd2,8; ZTmd3,6;
ZTmd4,12; ZTmd5,9; ZTsd4; ZTsd5,2;
ZTžd1,6; ZTžd2,3; ZTžd3,3; ZTžd4,13;
ZTžd5,5; ZTžd6,61
- občina · Sže1,5; Sže2; Sže3,2; Sže4,4
- občinski · Pme1,4; Pme4,2; Pmp1; Pmp3;
Pmp4,2; Pmp6; Pže1,4; Pže4,3; Pžp1
- občutek · Sme1,16; Sme2,2; Sme3; Sme4,4;
Sme5,3; Sme6,8; Smp1,4; Smp2,3; Smp3,2;
Smp4,2; Smp5,2; Smp6,2
- občutiti · Gae,2; Gap; Gce,6; GLme,10; GLže
- obdržati · Gce,4; GLmd; GLme,8; GLse; GLže
- obesiti · Gce,13; GLme,41; GLmp,9; GLže,4
- obilan · Pme3; Pme4j; Pmp2,2; Pse3; Pse4;
Pže1,6; Pže2; Pže4; Pže4j; Pže6; Pžp2; Pžp4,5
- obirati · Gae; Gbe; Gce,2; Gcp; GLmd,2; GLme;
GLmp,2; GLžd
- obisk · Sme1,3; Sme4,5; Sme5,2; Smp1; Smp2,2;
Smp4,3; Smp6
- obiskati · Gbe; Gcp; GLmd; GLme,10; GLmp,2;
GLže,5; GLžp; GNA,3
- obiti · GLme,4; GLmp; GLse,3; GLže,4; GLžp
- objem · Sme2,4; Sme4,6; Sme5,6; Smp4; Smp5
- objemati · Gce; Gce,3; Gcp; GLme,3; GLmp,2;
GLse; GLže,4; GLžp; GNE
- objeti · Gcd; Gce,13; GLmd,2; GLme,18;
GLmp,2; GLže,13; GNE,2
- oblaček · Smd1; Sme1,3; Sme2,3; Smp1,2; Smp6;
Sse3
- oblačiti · Gce; Gcp,2; GLme,3; GNE,4
- oblak · Smd1; Smd4; Sme1,19; Sme2,9; Sme3,2;
Sme4,8; Sme6,3; Smp1,23; Smp2,6; Smp3;
Smp4,7; Smp5,3; Smp6,5
- oblast · Sže1,6; Sže2,4; Sže3,2; Sže4,5; Sžp1;
Sžp3
- oblečen · GNme1,11; GNmp1; GNže1,7;
GNžp1,3; PNme1,3; PNmp1; PNmp4
- obleči · Gce,2; GLme,10; GLmp; GLže,4; GNE;
GVbd; GVbp
- obledel · Pmp6; Pse4; Pže1; Pže2; Pže4; Pže5,3;
Pže6; Pžp1,2; Pžp4,2; Pžp5
- obleka · Sže1,3; Sže2,9; Sže3,2; Sže4,23;
Sže5,13; Sže6; Sžp2,3; Sžp4,3; Sžp5,8
- obležati · Gae; Gce,9; Gcp,2; GLme,18; GLse,2;
GLže,7; GNE,2
- oblika · Sže1; Sže2; Sže4,2; Sže5,2; Sžp1;
Sžp2,4; Sžp4,4; Sžp5,2
- obljubiti · Gae,2; Gbe; Gce,3; Gcp; GLmd;
GLme,15; GLmp,4; GLže,6
- obok · Sme1,3; Sme2,5; Sme3,2; Sme4; Sme6,11
- obokan · Pme1; Pme2,2; PNme4i; PNme6;
PNsp1; PNsp2; PNže1; Pže4,3
- oborožen · GNmp1,3; PNme1,6; PNmp2,2;
PNže1,2; PNže4; PNže6; Pže2,2
- obračati · Gbe; Gcd; Gce,9; Gcp,2; GLme,9;
GLmp,4; GLže,2; GNE,3
- obraz · Smd1,3; Smd5; Sme1,92; Sme2,49;
Sme3,2; Sme4,124; Sme5,59; Sme6,20;
Smp1,14; Smp2,4; Smp4,15; Smp5,8; Smp6,5
- obregnuti se · Gce,7; GLme,9; GLmp,2; GLže,5
- Obrekar · IOme1,64; IOme2,3; IOme3,13;
IOme4,21; IOme5,5; IOme6,3

Obrekarica · IOže1,39; IOže2,2; IOže4; IOže5,3;
IOže6,2
Obrekarjev · IOmp4,4; IOmp5; PIOmd1;
PIOme1,8; PIOme3; PIOme4,3; PIOme4,3;
PIOme5,3; PIOmp1; PIOmp3; PIOmp5,4;
PIOmp6; PIOse2; PIOže1,2; PIOže2,3;
PIOže5,4; PIOže6,4; PIOžp1
Obrekarjev dob · IZme2,2; IZme4,6; IZme5,6
obrisati · Gae; Gcd; Gce,14; Gcp; GLmd;
GLme,62; GLmp,3; GLže; GLže,21; GNE,3;
GVad; GVbe
obriti · Gce,6; Gcp; GLme,4
obrniti · Gae,6; Gbe,2; Gcd; Gce,13; Gcp,2;
GLmd,3; GLme,207; GLmp,10; GLse,2;
GLže,56; GLžp; GNE; GVbe,2
obronek · Sme2,5; Sme3,3; Sme4,6; Sme5,6;
Sme6,5; Smp4,3; Smp5,2
obrt · Sže1,4; Sže2; Sže3; Sže4,2; Sže5,2
obrv · Sže4; Sže6; Sžp1,3; Sžp2,9; Sžp4,20;
Sžp6,4
obseden · GNže1; PNme1,5; PNmp1,2; PNže1,2
obsedeti · Gce,3; Gcp; Gcp; GLmd; GLme,11;
GLže,3; GNE,2
obsojen · GNme1,4; GNmp1; GNmp2; PNme1;
PNme1; PNme2; PNmp2; PNse3; PNžp1
obstati · Gae; Gcd,7; Gce,2; Gce,53; Gcp,13;
GLmd,15; GLme,166; GLmp,24; GLse,3;
GLže,41; GLžp,4; GNE
obup · Sme1,12; Sme2,17; Sme3; Sme4,5;
Sme5,4; Sme6,19
obupan · GNme1,5; PNme6,3; PNmp1; PNmp5;
PNže1,3; Pže6
obupno · A,15
obviseti · God; Gce,2; GLme,4; GLmp; GLse;
GLže,3; GLžp
obzidje · Sse2; Sse4,3; Sse5,4; Sse6,3; Ssp5
očala · Ssp1,2; Ssp2; Ssp4,14; Ssp6,2
oče · Sme1,340; Sme2,29; Sme3,31; Sme4,50;
Sme5,9; Sme6,15; Smp1,10; Smp2; Smp4
očetov · Pme1,4; Pme2,2; Pme4; Pme5,3;
Pmp4,2; Pse1; Pse4; Pže2; Pže4,2; Pže5,3;
Pže6,2; Pžp1; Pžp2,4; Pžp4
očetovsko · A,11
oči · Sžp1,139; Sžp2,62; Sžp3,8; Sžp4,400;
Sžp5,55; Sžp6,169
očitajoče · A,24
očitati · Gbe; GLme,9; GLmp,6; GLže,4; GLžp;
GNE,3
očitek · Sme1,2; Sme2,2; Sme6,10; Smp1;
Smp2,2
očitno · A,38
očividno · A,21

očka · Sme1,122; Sme2,4; Sme3,8; Sme4,12;
Sme5; Sme6,4; Sže4; Sžp4
od · E2,881; E,71
odbijati · Gce,3; Gcp; GLme,4; GLmp,6; GLse;
GLže
odbiti · Gae; Gce,13; GLme,32; GLse,2;
GLže,13; GNE
oddahnuti se · Gae,5; Gce,11; Gcp,4; GLmd,2;
GLme,61; GLmp,6; GLže,20; GNE
oddaljen · GNže1,3; Pmp1; PNme1,2; PNme2;
PNme4; PNme5; PNže1
oddaljivati se · Gce,4; Gcp; GLme; GLmp,5;
GLže,3
odeja · Sže1,6; Sže2,3; Sže4,23; Sže5; Sže6,9;
Sžp2,2; Sžp4,3; Sžp6
oder · Sme1; Sme2,4; Sme4,8; Sme5,6
odganjati · Gce,2; GLme,12; GLmp,2; GLže;
GNE,3
odgnati · Gce,3; Gcp; GLme,10; GLmp,7;
GLže,4; GNE,9; GVbe
odgovarjati · Gae; Gce,8; Gcp,3; GLme,12;
GLmp,3; GLže,2; GNE; GVbe,2
odgovor · Sme1,5; Sme2,35; Sme4,16; Sme6;
Smp2,2
odgovoriti · Gae,3; Gbe,2; Gcd; Gce,26; Gcp;
GLme,87; GLmp,5; GLsp,2; GLže,33; GNE,2;
GVbe,2; GVbp
odhajati · Gae,4; Gbp; Gce,2; Gcp,3; GLmd;
GLme,15; GLmp,5; GLžd; GLže,4
oditi · Gce,2; Gcp; GLmd,5; GLme,28;
GLmp,16; GLžd; GLže,7; GLžp; GNE,5
odkar · ZR,30
odkašljati se · Gce,2; GLme,8; GLže
odkimatei · Gce,7; Gcp; GLme,57; GLse;
GLže,10; GNE
odkleniti · GLme,7; GLmp,2; GLse
odkod · ZK,10
odkorakati · Gce,3; GLmd; GLme,7
odkrehniti se · Gce,2; GLme,24; GLmp,2
odkrit · Pme1; Pme5; Pmp1; Pse1,5; Pse2;
PTme1; PTmp1; PTže6; Pže1; Pže2
odkriti · Gce; GLmd; GLme,11; GLmp,3; GLže;
GNE,3
odkrito · A,12
odleteti · Gcd; Gce; GLmd,2; GLme,9; GLmp,3;
GLže,6; GLžp; GNE,2
odločen · Pme1,5; Pme2; Pme5,2; Pme6,6;
Pmp1; Pmp6; Pse5; Pže1; Pže5; Pže6; Pžp1;
Pžp2
odločiti · Gce,2; GLme,22; GLmp,2; GLže,6;
GNE,7
odločno · Aj,2; A,119

- odložiti · Gce,4; Gcp; GLmd; GLme,18; GLmp; GLže,6; GNE; GVbe
- odmakniti · Gce,4; GLme,19; GLmp,2; GLse; GLže,4
- odmev · Sme1,15; Sme2,3; Sme5; Sme6; Smp1,6; Smp5
- odmevati · Gce,10; Gcp,2; GLmp,5; GLse,4
- odnehati · Gce,2; Gcp; GLme,8; GLže,3; GLžp; GNE,2; GVbe,2
- odnesti · Gbe; Gce,5; Gcp; GLme,14; GLmp,15; GLse; GLže,22; GNE,5
- odpahniti · Gae; Gce,14; GLmd; GLme,13; GLže
- odpeljati · Gce,4; Gcp,2; GLme,17; GLmp,10; GLže,5; GNE; GVbe
- odpeti · Gce; GLme,2; GLmp,4; GLže,4
- odpirati · Gce,9; Gcp,5; GLmd; GLme,7; GLmp,3; GLsp; GLžd; GLže,2; GLžp,4; GNE,2
- odpovedati · Gae,4; Gbe,3; Gce; GLme,2; GLžp,3; GVbe
- odpraviti se · Gae; Gcp,2; GLmd; GLme,20; GLmp,2; GLže,7; GLžp; GNE,3
- odpravljati se · Gce,4; Gcp; GLme,2; GLmp; GNE,3
- odpreti · Gae,2; Gbe,2; Gce,25; Gcp,11; GLmd,2; GLme,114; GLmp,15; GLse,3; GLsp,38; GLže,56; GLžp,2; GNE,7; GTžp4; GVbe,2; GVbp
- odprt · GTme1; GTmp2; GTsp1,9; GTže1,3; GTžp1,3; PNse6; Pse1,3; Pse2; Psp6,3; PTme1; PTme5,2; PTse1,2; PTse3,2; PTse4,7; PTse5,2; PTsp1; PTsp4,3; PTže1,5; PTže2,2; PTžp1,2; PTžp3; PTžp4; PTžp5,3; Pže6,2; Pžp6,12
- odpustiti · Gap; Gce,5; GLme,6
- odrasel · Pme1,6; Pme3; Pme4,2; Pmp1; Pmp2; Pmp3; Smd1; Smp1,6; Smp6
- odrešilen · Pme1; Pme4i; Pse2; Pse6; Pže1,4; Pže4,3; Pžp2
- odreveneti · Gce,3; Gcp,3; GLme,3; GLmp; GLže; GLžp
- odrezati · Gap; Gce,6; GLmd; GLme,22; GLmp,4; GLže,6; GNE,3
- odsekan · PNme6,2; PNmp1,2; PNmp6,2; PNžd1,2; PNže6,5; PNžp1
- odsekati · Gae; Gbe; Gce; GLme,17; GLmp,2; GLže,3; GNE,2
- odskočiti · GLme,15; GLmp,5; GLže,5
- odсотno · A,15
- odtrgati · Gce,4; Gce,10; Gcp; GLmd,3; GLme,22; GLmp,4; GLse; GLže,12; GNE,6
- odvezati · Gce; GLme,6; GLmp; GLže,6
- odvleči · Gcd; Gce,4; Gcp,3; GLmd,2; GLme,5; GLmp,6; GLže,3
- odvrniti · Gce,11; GLme,29; GLže,14
- oficir · Sme1,12; Sme4,2; Sme6; Smp1,6; Smp3; Smp5; Smp6,3
- ogelj · Sme1,19; Sme2,8; Sme3,2; Sme4,13; Sme5,11; Sme6,4; Smp1; Smp2; Smp4,3; Sse6,3
- ogibati se · Gae; Gce; GLme,4; GLmp; GLže,3; GNE
- oglasiti se · Gap; Gbe; Gcd; Gce,49; Gcp; Gcp,4; GLme,120; GLmp,10; GLse,24; GLže,82; GLžp
- oglašati se · Gce,6; Gcp,3; GLme,2; GLmp,2; GLse,4; GLže,6; GLžp,2; GNE,2
- ogledati · Gae,2; Gbe,2; Gce,2; Gcp,2; GLmd; GLme,16; GLmp,5; GLže,4; GLžp; GNE,3; GVap; GVbe,2; GVbp,3
- ogledovati · Gbe; Gcd; Gce,5; Gcp,4; GLme,14; GLže,2; GLžp; GNE,4
- oglje · Sse1,2; Sse2,2; Sse4,5; Sse6
- oglodan · PNme4i,2; PNmp1; PNmp4; PNže4,3; PNžp1,2; PNžp2; PNžp4; Pže5
- ognjen · Pme2,2; Pme4,2; Pmp4; Pmp6; Pse1; Pže1,3; Pže4; Pžp1,2; Pžp2
- ognjišče · Sme2,7; Sme4,2; Sse1,2; Sse2,5; Sse3,6; Sse4,2; Sse5,14; Sse6,15; Ssp2,2; Ssp5
- ogorek · Sme1,8; Sme2,4; Sme4,15; Sme5,2; Smp1,3; Smp4,3; Smp6
- ograja · Sže1,2; Sže2,2; Sže4,21; Sže5,4; Sže6,2
- ogreti · Gap,2; Gce,5; Gcp; GLme,4; GLže,3; GVap; GVbe
- ogromen · Pme1,3; Pme2; Pme4; Pme4i,2; Pme4,3; Pmp1; Pmp2; Pmp4,2; Pmp5; Pmp6; Psp1,2; Pže1,6; Pže4,3; Pže6; Pžp1,2
- oh · M,70
- oho · M,12
- okajen · PNme1i,7; PNme1,5; PNmp1; PNže1
- okamenel · PLme1; PLme1i,4; PLme5; PLmp4; PLse6; PLsp5; PLže2
- okameneti · GLme,9; GLmp,4; GLse,6; GLže,7; GLžp
- okence · Sse1; Sse3,3; Sse4,13; Sse5; Ssp1
- okno · Sse1,19; Sse2,33; Sse3,30; Sse4,73; Sse5,37; Sse6,13; Ssp1,22; Ssp2,5; Ssp4,19; Ssp5,9; Ssp6,9
- oko · Sse1,18; Sse3; Sse4,26; Sse6,6; Ssp2
- okoli · A,14; E2,2
- okoren · Pme1,2; Pmp1; Pmp6; Pse1,3; Pže1; Pže6,2; Pžp6
- Okouinokou · IOme1,13; IOme2; IOme4,2
- okov · Smp1,4; Smp2,3; Smp4,8; Smp5

- okovan · PNme4,2; PNme6,2; PNmp1,3;
PNmp2,2; PNmp6,3; PNse6; PNsp1; PNsp2;
PNsp4,3; PNže1; PNže4; PNže6; PNžp1
- okreniti · Gcd,5; Gce,79; Gcp,3; GLme,36;
GLmp; GLžd; GLže,7; GNE
- okrog · A,8; E2,271; E,2
- okrogel · Pme1i; Pme1,7; Pme2,3; Pme3;
Pme4i,2; Pme4,2; Pme4,2; Pme5,3; Pmp1,5;
Pmp2; Pmp5,3; Pmp6,3; Pse1,4; Pse4,3;
Pse6,7; Psp1; Psp5; Pžd1; Pže1,11; Pže4,5;
Pže5,2; Pže6,3; Pžp1,3; Pžp4; Pžp6,2
- Okrogličar · IOme1,84; IOme2,2; IOme3,2;
IOme4,11; IOme6
- Okroglo · IZse1,2; IZse2,2; IZse4,4; IZse5,2
- okus · Sme1; Sme2,5; Sme3; Sme4,2; Sme5;
Sme6
- okvir · Sme1,2; Sme2,3; Sme4,4; Sme5,10;
Smp5; Sže5
- olje · Sse1,7; Sse2,6; Sse6
- oltar · Sme1; Sme2,2; Sme3; Sme4,4; Sme5,2;
Sme6,4
- omahniti · Gae,2; Gce,7; Gcp; GLme,7; GLse;
GLže,14; GLžp,2
- omara · Sže1,3; Sže2,6; Sže3,2; Sže4,9; Sže5;
Sže6,3; Sžp1,2; Sžp4,3
- omehčati · Gce; Gcp; GLme,4; GLmp; GLse;
GLže,2
- omeniti · Gce; GLme,12; GLmp,2; GLže,3;
GNE,3
- on · ZOcmd1,2; ZOcmd2,2; ZOcmd2,7;
ZOcmd3,5; ZOcmd3,37; ZOcmd4,36;
ZOcmd6,27; ZOcm1,167; ZOcm2,2;
ZOcm2,109; ZOcm2,112; ZOcm3,168;
ZOcm3,1951; ZOcm4; ZOcm4; ZOcm4,6;
ZOcm4,10; ZOcm4,13; ZOcm4,37;
ZOcm4,39; ZOcm4,61; ZOcm4,83;
ZOcm4,184; ZOcm4,2903; ZOcm5,166;
ZOcm6,408; ZOcm2; ZOcm5,2; ZOčzd3,3
- ona · ZKžd1; ZKže1,44; ZKže3; ZKže4;
ZKže5,5; ZOcmd4; ZOcp2; ZOcp2,3;
ZOcp3,4; ZOcp4,37; ZOcp5,5; ZOcp6,3;
ZOče1,25; ZOče2,100; ZOče2,134;
ZOče3,77; ZOče3,570; ZOče4; ZOče4;
ZOče4; ZOče4,3; ZOče4,3; ZOče4,8;
ZOče4,20; ZOče4,22; ZOče4,41; ZOče4,51;
ZOče4,1245; ZOče5,2; ZOče5,156;
ZOče6,132
- one · ZOčzp2,4; ZOčzp2,23; ZOčzp3,8; ZOčzp4;
ZOčzp4,6; ZOčzp4,6; ZOčzp4,270; ZOčzp5,16;
ZOčzp6,8
- onemeti · GLme,10; GLmp,3; GLže,3; GLžp
- oni · ZKme1,5; ZKme3; ZKme4; ZKme4;
ZKme5; ZKmp1,6; ZKmp3; ZKse2; ZKse3;
ZKse5,5; ZOcmp1,7; ZOcmp2,37; ZOcmp2,56;
ZOcmp3,19; ZOcmp3,254; ZOcmp4; ZOcmp4;
ZOcmp4,2; ZOcmp4,2; ZOcmp4,4; ZOcmp4,5;
ZOcmp4,7; ZOcmp4,12; ZOcmp4,12;
ZOcmp4,535; ZOcmp5,26; ZOcmp6,140;
ZOcp4,12; ZOcp4; ZOcp4; ZOčzp3; ZOčzp4,2
- ono · ZKse1,2; ZKse4,2; ZOce2; ZOce4,27;
ZOce6,4
- onstran · A,5; E2,23
- ooo · M,13
- opaziti · Gce; GLmd; GLme,18; GLmp,2; GLse;
GLže,6
- opazovati · Gce,4; GLme,8; GLže,2
- opaž · Sme2,4; Sme4,5; Sme5,3; Smp4; Sžp5,3
- opeka · Sže2,4; Sže4,5; Sže6
- oplaziti · Gce,5; GLme,7; GLse; GLže,5; GLžp,2
- opletati · Gbe; Gbp; Gce,7; Gcp,3; GLme,10;
GLže,10; GLžp,3; GNE,3
- opoteči se · Gce,4; GLme,13
- opotekati se · Gbe; Gce,3; Gcp; GLme,6; GLmp
- oprati · Gce,4; GLme,3; GLme,4; GLmp,2;
GLže,6; GLžp
- opravek · Sme2,11; Sme3; Sme4,2; Sme5;
Smp3,2; Smp5,2; Smp6
- opravičevati · Gae; Gce; GLme,10; GLže
- opravilo · Sse1,2; Sse2; Sse5,7
- opraviti · Gae,2; Gcd; Gce,3; GLmd,4; GLme,29;
GLmp,15; GLse; GLže,5; GLžp; GNE,10
- opravljati · Gae; Gce,3; Gcp; GLme,9; GLmp,3;
GLžp; GNE
- opravljen · GNse1,6; GNže1,5; PNme5; PNmp5;
PNse1; PNse4; PNse5
- oprijeti se · GLme,6; GLže,3; GLžp; GNE
- oprostiti · Gbp,6; Gce,2; GVbd,5
- oprtnik · Sme1,6; Sme2,4; Sme3; Sme4,10;
Sme6,4; Smp1; Smp2; Smp4,4
- orati · Gce; Gcp,4; GLme,8; GLmp; GNE,3
- oreh · Sme1,3; Sme2,7; Sme3,4; Sme4,4;
Sme6,4; Smp1; Smp2,3; Smp4; Smp6
- Orel · IOme1,106; IOme2,7; IOme3,10;
IOme4,13; IOme5; IOme6,4
- orel · Sme1,22; Sme3,3; Sme4,2; Sme6; Smp1,2
- orgle · Sžp1,10; Sžp2; Sžp4,2; Sžp5
- orlica · Sže1,18; Sže3,3; Sže4; Sžp1
- orlovski · Pme1,3; Pme2,6; Pme4; Pme6,2;
Pžp1; Pžp6
- orodje · Sse1,6; Sse2,2; Sse4,9; Sse6,2
- orožje · Sse1,3; Sse4,6; Sse5,2; Sse6
- orožnik · Smd1,8; Sme1,3; Smp1,16; Smp2,3;
Smp3,6; Smp4,3; Smp5

- osem · ŠGmp5; ŠGsp6,4; ŠGžp2; ŠG1,5; ŠG4,15
 osliniti · Gce,2; GLmd; GLme,8
 ostali · Pme2,2; Pmp1,14; Pmp2; Pmp3,2;
 Pmp6,2; Pse4; Pžp4; Smp4,2
 ostanek · Sme1,3; Sme4; Smp1; Smp4,4; Smp5;
 Smp6
 ostati · Gae,2; Gap; Gbe,3; Gcd; Gce,30; Gcp,4;
 GLmd,4; GLme,73; GLmp,8; GLse,17; GLsp;
 GLžd,2; GLže,47; GLžp,4; GNE,10; GVbe,3;
 GVbp,5
 oster · Pme1i; Pme1,13; Pme2,5; Pme4i;
 Pme4,4; Pme5,2; Pme6,7; Pmp1,2; Pmp2;
 Pmp4; Pmp6,3; Pse1,2; Pse2,2; Pse4; Pse6;
 Psp4j; Pže1,6; Pže2; Pže4,4; Pže5; Pže6,2;
 Pžp1j; Pžp1,4; Pžp2,2; Pžp3,2; Pžp4,3; Pžp5;
 Pžp6,2
 ostrgati · Gce,2; GLme,17; GLže; GNE; GVbe
 ostrmeti · Gce,7; GLmd,3; GLme,14; GLmp,2
 ostro · Aj,3; A,55
 ošiniti · Gbe; Gce,6; GLme,6; GLže,3
 otava · Sže1,4; Sže2,2; Sže4,11; Sže5,13; Sže6,2
 otepati se · Gae; Gce,4; Gcp,3; GLmd; GLme,20;
 GLmp; GLže; GLžp,2; GNE
 Oti · IOže1,14; IOže3; IOže4
 otipavati · Gcd; Gce; Gcp; GLme,4; GLže;
 GNE,3
 otirati · Gce; GLme,7; GLže,2; GNE,11
 otožen · Pme1i; Pme1,6; Pme2; Pmp4i; Pse1,2;
 Pse2; Pže1,2; Pže2,3; Pže6; Pžp1,2; Pžp4
 otožno · A,27
 otožnost · Sže1; Sže2,2; Sže4,2; Sže6,5
 otresati · Gbp; Gce,4; Gcp,2; GLse; GLže,2
 otresti · Gce,4; GLme,22; GLmp; GLže,3;
 GNE,4
 otreti · GLme,3; GLže,7
 otročaj · Smp1,11; Smp2; Smp3,2; Smp4
 otročji · Pme1,9; Pmp1,4; Pse2; Pse6; Pže1;
 Pže6
 otrok · Smd1,8; Smd3; Smd4,3; Sme1,240;
 Sme2,12; Sme3,32; Sme4,90; Sme5,6;
 Sme6,34; Smp1,198; Smp2; Smp2,67;
 Smp3,31; Smp4,80; Smp6,3; Smp6,15
 otroški · Pme1,3; Pme2,3; Pme5; Pme6,3;
 Pmp1; Pse1,2; Pse2,2; Pse4,2; Pse6; Psp1,2;
 Psp2,5; Psp4; Psp5,11; Pže1,5; Pže2; Pže4,4;
 Pže5,2; Pže6,7; Pžp1,2; Pžp2; Pžp3; Pžp4;
 Pžp5,4
 otroško · A,12
 otrpniti · GLme,4; GLmp,5; GLse
 ovinek · Sme1; Sme2,9; Sme3; Sme4,3; Sme5,19;
 Sme6,8; Smp1; Smp2; Smp4,2; Smp5,3
 ovratnik · Sme1,2; Sme2,5; Sme4,9; Sme5
- ozek · A; Pme1i; Pme1j,2; Pme1,4; Pme4;
 Pme4; Pme4i; Pme5,11; Pme6; Pmp1,2;
 Pmp2; Pmp4; Pmp5; Pse1,8; Pse4,3; Psp1,2;
 Psp2,2; Psp4,2; Pželj; Pže1,10; Pže4,3;
 Pže5,12; Pžp2,4; Pžp4; Pžp5,5
 ozirati se · Gae; Gce,9; Gcp,4; GLme,25;
 GLmp,6; GLže,6; GNE,6
 ozmerjati · Gce; GLmd; GLme,8; GLmp;
 GLže,7; GNE
 oznaniti · Gce; GLme,12; GLmp; GLže
 ozračje · Sse1,8; Sse4,4; Sse5,5
 ozreti se · Gae,8; Gae; Gbe; Gcd,2; Gce,40;
 Gcp,5; Gcp,10; GLmd,8; GLme,230; GLmp,15;
 GLže,60; GNE,3; GVbe; GVbp
 oženiti · Gce; GLme,18; GNE,2
 ožgan · GNžel; PNme5,2; PNmp1,4; PNmp2;
 PNse2; PNžp2; PNžp6
 oživeti · GLme,6; GLmp; GLse,2; GLže,5;
 GLžp,2
 pa · Č,1733; Vpr,2892
 pač · Č,135
 padati · Gae,12; Gce,15; Gcp,10; GLme,16;
 GLmp,16; GLse,4; GLže,6; GLžp,3; GNE,6
 paglavец · Sme1,4; Sme6; Smp1,5
 pahnti · Gae; Gce,19; Gcp; GLmd,6; GLme,22;
 GLmp,13; GLže,2; GNE
 pajčevina · Sže2,2; Sže4,3; Sže6,3; Sžp1,4; Sžp2;
 Sžp4
 palec · Smd1,2; Smd6; Sme1; Sme2,2; Sme3;
 Sme4,10; Sme5,4; Sme6,12; Smp2,2; Smp4,2;
 Smp6
 palica · Sže1,10; Sže2,6; Sže4,55; Sže5,5;
 Sže6,32; Sžp4,4; Sžp6,4
 pamet · Sže1,29; Sže2,17; Sže3,5; Sže4,37;
 Sže5,10; Sže6,3
 pameten · Pme1,32; Pme2,2; Pme3; Pme4;
 Pme4j; Pmp1,12; Pmp2; Pmp3; Pmp4,2;
 Pmp6; Pse1,7; Pse2,2; Psp2; Pže1,8; Pže2;
 Pže4; Pžp1,2
 pametno · Aj; Ajj,4; A,15; Sse2
 pankrt · Smd1; Smd6; Sme1,3; Sme4,3; Sme6;
 Smp1; Smp4,2
 papir · Sme1,7; Sme2,14; Sme4,12; Sme6,4;
 Smp2,3; Smp4,14; Smp5,3; Smp6
 par · Smd4; Sme1,7; Sme2,4; Sme4,10; Sme6;
 Smp1,6; Smp2; Smp4; ŠNE
 paragraf · Sme1,6; Sme2; Sme4; Smp1,2; Smp5;
 Smp6
 parobek · Sme2,3; Sme3; Sme4,3; Sme5,2; Smp4
 partizan · Smd1,3; Sme1,31; Sme2,5; Sme3;
 Sme4,2; Sme5; Smp1,59; Smp2,10; Smp3,14;
 Smp4,11; Smp5,8; Smp6,8

- partizanski · Pme1,4; Pmp1; Pmp4; Pmp5;
Pse1,2; Pse4; Pse6; Pžel,5; Pže4; Pže5,4;
Pže6; Pžp1,2
- pas · Sme1,9; Sme2,2; Sme2,19; Sme4,19;
Sme5,5; Sme6,9; Smp1; Smp4,2; Smp6
- pasti · Gae,4; Gap; Gcp; Gbe,3; Gcd; Gce,26;
Gcp,7; GLmd,6; GLme,8; GLme,95; GLmp,3;
GLmp,16; GLse,17; GLsp; GLžd,2; GLže;
GLže,25; GLžp; GLžp,6; GNE,4
- pastir · Smdl; Sme1,14; Sme2; Sme3; Sme4,2;
Sme5; Smp1,8; Smp2,2; Smp3; Smp4,2; Smp6
- paša · Sže4,9; Sže5,2
- patron · Sme1,3; Sme2; Sme3; Sme4; Smp4,5
- Pavel · IOme1; IOme2,11; IOme4,2
- Pavla · IOže1,14; IOže2; IOže3,3; IOže4,3; IOže5
- pazduha · Sže2,5; Sže4,11; Sže6,20; Sžp6
- Pazi · IOme2,2; IOme3; IOme4,3; Ižme1,5;
Ižme3; Ižme5
- paziti · Gbe; Gce; Gcp; GLme,4; GLmp,3;
GNE,2; GVbe,3
- peč · Sže1,5; Sže2,16; Sže4,32; Sže6,2; Sžp4,3;
Sžp5,2
- peča · Sže3,19; Sže5,54
- Pečan · IOme1,11; IOme3,2
- Pečanov · IOmp5; PIOmd1; PIOme1,5; PIOme3;
PIOme4; PIOmp5; PIOže1; PIOže4; PIOžp5
- pečen · GNže1,2; Pmp4,2; PNme1; PNmp2;
PNmp5; PNže1; PNže4; Pžp1; Pžp4
- peči · Gce,8; Gcp; GLme; GLmp,2; GLse;
GLže,10; GNE,2; GNme1,2; GNžp4; GVbe
- pehati se · Gae; Gce,6; Gcp; GLme; GLmp,2;
GNE,4
- pekel · Sme2,8; Sme4,3; Sme5,2
- peljati · Gae; Gap,4; Gbp; Gcd; Gce,8; Gcp,5;
GLme,12; GLmp,16; GLže,12; GNE,2;
GVbp,2
- peniti se · Gce,6; Gcp; GLse; GLže,8; GNE
- Pepa · IOže1,23; IOže2; IOže4,2
- pepel · Sme1,2; Sme2,3; Sme4,4; Sme5,4; Sme6
- Peregrin · IOme1,24; IOme3,2; IOme4,4
- perica · Sže1,11; Sže2
- perilnik · Sme1,9; Sme5,4
- pero · Sse1; Sse4,6; Sse6,3; Ssp6
- pes · Smd4; Sme1,33; Sme2,3; Sme3,3; Sme4,9;
Sme6,2; Smp1,6; Smp2,4; Smp3; Smp4; Smp5;
Smp6
- pesek · Sme1,2; Sme2,3; Sme4,6; Sme5,3;
Sme6,2
- pesem · Sže1,61; Sže2,15; Sže3,3; Sže4,36;
Sže5,3; Sže6,6; Sžp1,2; Sžp2,3; Sžp4,24;
Sžp5,3
- Peskarjev · PIOme1,6; PIOme2; PIOme4;
PIOme6; PIOmp1,2; PIOmp6,3; PIOže1,2;
PIOže2; PIOže4; PIOžp1
- pesnik · Sme1,12; Sme2,4; Sme4; Smp1,5; Smp4
- pest · Smd4; Sžd4; Sže1,4; Sže2; Sže4,28; Sže5;
Sže6,21; Sžp2,2; Sžp4,18; Sžp5; Sžp6,10
- peščen · Pmeli; Pme5; Pmp2; Psp4; Pže1; Pže2;
Pže4,8; Pže5,2; Pžp2; Pžp4,2; Pžp5
- pet · GNžp1; PTže6; ŠGmp2; ŠGmp5,3;
ŠGmp6,4; ŠGsp2; ŠGsp5; ŠGsp6,2; ŠGžp2;
ŠGžp5,4; ŠGžp6,3; ŠG1,11; ŠG4,43
- peta · Sže1; Sže4; Sže5,7; Sže6,5; Sžp1,4;
Sžp2,3; Sžp4; Sžp4,3; Sžp5,5; Sžp6,10
- petdeset · ŠGsp5,2; ŠG1,2; ŠG4,19
- petelin · Sme1,20; Sme2,3; Sme3,2; Sme4,7;
Sme6; Smp1,5; Smp2,21; Smp6
- Peter · IOme1,438; IOme2,5; IOme3,39;
IOme4,63; IOme5,3; IOme6,4
- peti · Gae,2; Gbe,3; Gcd; Gce; Gce,2; Gce,32;
Gcp,9; Gcp,14; GLmd; GLme,33; GLmp,11;
GLse; GLže,31; GLžp,6; GNE,10; GVbe;
GVbp; ŠGmp6; ŠVme1,3; ŠVme2; ŠVme4;
ŠVse5; ŠVže2,2; ŠVže4,2; ŠVže5; ŠVže6
- petindvajset · ŠGsp5; ŠG4,11
- petje · SGse1,9; SGse2,3; SGse3; SGse4,3;
SGse5,2; SGse6,2
- petnajst · ŠGsp5,4; ŠGsp6,7; ŠG1,10; ŠG4,25
- petrolejka · Sže1,5; Sže2,2; Sže4,2; Sžp1
- Pettaros · IOme1,21; IOme3,2; IOmp4
- pha · M,22
- pihati · Gce,9; Gcp,2; GLme,14; GLmp,3; GLse;
GNE,3; GVbe; GVbp
- pihniti · Gce,4; Gcp; GLme,12; GLmp; GLže,3;
GVbe,2
- pijača · Sže1,4; Sže2,11; Sže4,7; Sže5,2
- pijan · Pmeli,8; Pme1,8; Pme3; Pme4; Pme5;
Pmp1,2; Pmp4; Pže4,2; Pže5,2; Pžp5
- pika · Sžd1; Sže1,2; Sže2; Sže4,6; Sže5; Sžp2;
Sžp4,2
- pipa · Sže1,4; Sže2,7; Sže4,23; Sže6,3; Sžp4,4
- pisalen · Pme41,9; Pže1,2; Pže2,2; Pže3; Pže4,2;
Pže5,2; Pže6,5
- pisan · GNse1,4; GNže1; Pme1; Pmeli; Pme2;
Pme4; Pme4i; Pme4,4; Pmp1,2; Pmp3; Pmp4;
PNmp5; PNse1; PNse4; PNsp5; PNže4,7;
PNže6; PNžp5,2; Pse1,2; Pse2; Psp6,2;
Pže1,2; Pže3; Pže4; Pže5,2; Pžp1,5; Pžp4,2;
Pžp6
- pisanje · SGse1,6; SGse2,3; SGse4; SGse5,4;
Sse4; Sse5,2; Sse6
- pisar · Sme1,7; Sme3,3; Sme4; Sme6
- pisarna · Sže2,2; Sže4,8; Sže5,5; Sžp5

pisatelj · Sme1,8; Sme2,4; Smp1,2; Smp4
 pisati · Gae,20; Gbe,2; Gbp,11; Gce,39; Gcp,10;
 GLme,62; GLmp,7; GLse; GLže,21; GNE,23
 pisker · Sme1; Sme4,7; Sme5,2; Sme6,2; Smp1;
 Smp4,5; Smp5; Smp6
 pismo · Sse1,8; Sse2,7; Sse4,28; Sse5; Sse6,3;
 Ssp1,3; Ssp2,4; Ssp4,2; Ssp6
 pismonoša · Sme1,77; Sme3,9; Sme4,2; Sme4,18
 piti · Gae,5; Gbe; Gce,13; Gcp,2; Gcp,9;
 GLmd,3; GLme,43; GLmp,8; GLže,5; GNA,2;
 GNE,5; GVad; GVbe,17; GVbp,14
 plačati · Gae,4; Gce,3; Gcp; GLmd; GLme,3;
 GLmp; GLže,2; GNE,5; GVbe,2
 plačevati · Gce,6; Gcp,2; GLme,2; GNE
 plačilo · Sse1,2; Sse4,8; Sse6
 plamen · Sme1,6; Sme4,2; Sme5,3; Sme6,2;
 Smp1,6; Smp2,2; Smp4,2; Smp5,4; Smp6
 planina · Šze2; Šze4; Šzp1,2; Šzp2,6; Šzp4,4;
 Šzp5,2
 planiti · Gcd,3; Gce,73; Gcp,9; GLmd,3;
 GLme,84; GLmp,7; GLse; GLže,39; GNE,4
 planjava · Šze1,4; Šze2; Šze5,2; Šzp4,3
 plast · Šze1,4; Šze2,2; Šze4; Šze6; Šzp2,2;
 Šzp4,2
 plašč · Sme1,2; Sme2,2; Sme4,8; Sme5,3;
 Sme6,3; Smp4; Smp5
 plašno · A,40
 plav · Pme1i; Pme5; Pmp1,4; Pmp2; Pmp4,3;
 Pzd4,5; Pzd6,2; Pže1,2; Pže4,2; Pžp1; Pžp2;
 Pžp4,2; Pžp5; Pžp6,2
 plavati · Gce,9; Gcp,2; GLme,9; GLmp,4;
 GLse,7; GLže,5; GLžp,4; GVbe
 plaz · Sme1,2; Sme4,5; Sme5,2; Sme6,2
 plaziti se · Gcd; Gce,6; Gcp,2; GLmd,2; GLme;
 GLže; GNE
 plemenit · Pme1; Pme1i; Pme2,4; Pme5;
 Pmp1,2; Pmp2; Pmp5; Pže1; Pže4; Pžp1
 ples · Sme1,7; Sme4,6; Sme5
 plesati · Gcd; Gce,4; Gcp,6; GLmd; GLme,9;
 GLmp; GLse,2; GLsp; GLžd; GLže,5; GVbe;
 GVbp
 pleša · Šze1,4; Šze4,11; Šze5,3; Šze6
 plešast · Pme1,2; Pme4; Pže1; Pže4,7; Pže6;
 Pžp2
 pletenka · Šze4,10
 plevel · Sme2; Sme4,7; Sme6,2
 plezati · Gae,8; Gcd; Gce,2; Gcp; GLme,3;
 GNE,3
 pljuniti · Gce,9; Gcp; GLme,31; GLmp; GLže;
 GNE; GVbe,3
 pločevinast · Pme1i,2; Pme3; Pme4; Pme4i;
 Pme6; Pmp6,2; Pže1; Pže2; Pže4; Pže5

pločnik · Sme2; Sme4,2; Sme5,10; Sme6; Smp5
 plosk · Pme5; Pse1; Pzd6; Pže6,5; Sme1,2
 ploskati · Gce; Gcp,4; GLme; GLmp,3; GLže,2
 ploščad · Šze2; Šze3; Šze4,3; Šze5,9; Šze6
 plot · Sme2; Sme4,2; Sme5,6; Sme6; Smp1;
 Smp6
 plug · Sme1,2; Sme2; Sme4,9; Sme6,2; Smp1,3;
 Smp4,2; Smp6
 po · Č,39; E4,242; E5,2633; E,34
 pob · Sme1,202; Smp1,15
 poba · Smd1,10; Sme2,6; Sme3,22; Sme4,25;
 Sme5,2; Smp2,3; Smp4,2
 pobarati · Gce,2; GLme,7; GLže; GNE
 pobesiti · Gcp,2; GLme,7; GLmp; GLže; GNE,3
 pobirati · Gce; Gcp; GLme,6; GLmp; GLže,3;
 GNA,2; GNE,6
 pobiti · Gcp; GLme,3; GLmp,7
 pobočje · Sse1,9; Sse2,24; Sse3; Sse4,20;
 Sse5,55; Sse6; Ssp1,5; Ssp2,4; Ssp4,3; Ssp5,6
 pobožati · Gce,3; GLme,13; GLse; GLže,8
 pobožen · Pme1; Pme4; Pse3; Pse4; Pže1,5;
 Pže4; Pžp1,3
 pobrati · Gap,2; Gce; Gce,11; Gce,30; Gcp,5;
 GLmd,2; GLme,63; GLmp,23; GLse,2;
 GLže,36; GLžp,3; GNE,4; GVbe; GVbp
 pobrisati · Gce,8; Gcp; GLme,15; GLmp,4;
 GLže,9
 počakati · Gad; Gae,4; Gap; Gce,8; Gcp,2;
 GLmd,5; GLme,31; GLmp,3; GLže,6; GLžp;
 GNE,6; GVbd; GVbe,35; GVbp,15
 počasen · Pme1; Pme1i,2; Pme3; Pme4; Pme4i;
 Pmp6,2; Pže1; Pže2; Pže4; Pže6
 počasi · A,595
 počen · PNme2; PNže1,2; PNže4; PNže5,9
 počenjati · Gae,2; Gcd; Gce,4; Gcp,4; GLme,5;
 GLže,2; GNE
 počepniti · Gce,2; GLme,7; GLmp; GLže,4
 počez · A,10
 počitek · Sme1,2; Sme2,11; Sme3,7; Sme4,4;
 Sme5
 počiti · Gce,8; GLme,3; GLmp; GLse,8; GLžp
 počivati · Gae; Gce,12; Gcp; Gcp,3; GLme,6;
 GLmp,2; GLsp; GLže,4; GLžp,3; GNE,4;
 GVbe,3
 počohati · GLme,8; GLže,2
 počrnel · PLme4i,2; PLme5,2; PLmp6; PLse1;
 PLse3; PLže1; PLže4,4; PLžp4; PLžp6
 pod · E4,166; E6,406; Sme1,3; Sme2,3; Sme4,9;
 Sme5,29; Smp1; Smp2
 podajati · Gce,6; GLmd; GLme,2; GLmp; GNE,2
 podkovan · GNme1; PNme2; PNme6,2;
 PNmp1,12; PNmp6; PNže1

- podnevi · A,11
 podoba · Sžd4; Sžd6; Sže1,31; Sže2,5; Sže4,31;
 Sže5,16; Sže6,3; Sžp1,4; Sžp2,3; Sžp4,10; Sžp5
 podoben · Pmd1; Pme1,25; Pme2,2; Pme4;
 Pme6; Pmp1,14; Pmp2; Pmp3; Pmp4,4;
 Pmp5; Pmp6; Pse1,5; Pse2,10; Pse4,3; Pse5,4;
 Psp1; Psp2; Psp4,2; Psp6; Pžd1; Pže1,44;
 Pže4; Pže6; Pžp1,4; Pžp2,3; Pžp4,3; Pžp5,8;
 Pžp6,3; Sme6
 podolgovat · Pme1,3; Pme2,3; Pme5,2; Pmp4;
 Pse2; Pse3; Psp1; Psp2; Pže2; Pže4,2; Pže5;
 Pžp1
 podpirati · Gae,2; Gcd,2; Gce,4; Gcp; GLmd;
 GLme,4; GLmp; GLže; GNE
 podpisati · Gae; Gbe; GLme,6; GLmp,4; GLže;
 GNE; GVbe; GVbp,2
 podplat · Sme4,2; Sme6; Smp1; Smp2; Smp4,2;
 Smp5,3
 podreti · Gae; Gbe; Gce; Gce,2; GLme,6;
 GLmp,2; GLse; GLžd; GLže,2; GNE
 podvzitati se · Gce,2; Gcp; GLme,6; GNE;
 GVbe,2
 podzemlje · Sse1; Sse2,3; Sse4,3; Sse5,3
 Podzemljšč · IOme1,30; IOme2,2; IOme3,2;
 IOme4; IOme6
 poganjati · Gce,11; Gcp,3; GLme,8; GLmp,3;
 GLse; GLže,8; GLžp,2; GNE,3
 pogladiti · Gce,2; GLme,13; GLže,3
 poglaviten · Pme1i; Pme2; Pmp2; Pmp4;
 Pse1,10; Pse2; Pse4; Pže1jj; Pže1,4; Pže4;
 Pže5; Pžp1; Pžp2; Pžp4
 pogled · Sme1,41; Sme2,21; Sme3,3; Sme4,60;
 Sme5,15; Sme6,108; Smp1,3; Smp2; Smp4,3;
 Smp6,11
 pogledati · Gae,8; Gap; Gap,4; Gbe; Gcd,3;
 Gce,187; Gcp,11; GLmd,2; GLme,252;
 GLmp,15; GLže,97; GLžp; GNA,7; GNE,14;
 GVbd; GVbe,34; GVbp,14
 pognati · Gae,3; Gbe; Gcd,2; Gce,21; Gcp,2;
 GLmd; GLme,84; GLmp,15; GLse,5; GLže,32;
 GLžp,3; GNE,9; GVbe,2
 pogoltniti · Gce,2; GLme,6; GLže; GNE
 pogosto · Aj,2; A,8
 pogovarjati se · Gae,7; Gbe,3; Gce,13; Gcp,3;
 GLmd,8; GLme,22; GLmp,9; GLže,2; GLže,8;
 GLžp; GNE,12
 pogovor · Sme1,5; Sme2,5; Sme3,2; Sme4,9;
 Sme5,2; Smp2; Smp4,3
 pogovoriti se · Gad,2; Gcp; GLmd; GLme,5;
 GLmp,2; GLže; GNE,2
 pograbit · Gcd,3; Gce,24; Gcp,3; GLmd,6;
 GLme,47; GLmp; GLže,17; GLžp; GNE
 pogreb · Sme1,3; Sme2,2; Sme3,6; Sme4,6;
 Sme5,2; Smp5
 pogreti · Gad; Gap,2; Gbe; Gcp,2; GLme,2;
 GLmp,3; GLse; GLžp
 pogrezati se · Gae; Gce,4; Gcp; GLme,3; GLmp;
 GLse; GNE
 pogrezniti se · Gbe; Gce,4; GLme,29; GLmp,2;
 GLse; GLže,12; GLžp
 pogrnuti · Gap; Gce; GLmd; GLme,9; GLže,5
 pogum · Sme1,6; Sme2,8; Sme4,5; Sme6,2
 pohiteti · Gad,2; Gap; Gce,6; Gcp; GLme,4;
 GLže; GVbe
 pohleven · Pme1,3; Pme2; Pmp4; Pse1; Pže1;
 Pže6; Pžp2,2
 pohvaliti · GLme,6; GLže,4
 poiskati · Gae; Gcd; Gce,9; Gcp,4; GLmd,5;
 GLme,30; GLmp,6; GLže,8; GNE,2; GVbe,3
 pojaviti se · Gcd; Gce,2; Gcp; GLme,13; GLžd;
 GLže,2; GNE
 pojesti · GLme,9; GLmp,4; GLže,5
 pokaditi · Gae; Gap; Gce; GLme,5; GLmp,2;
 GLže,2; GNE
 pokarati · Gce,2; GLme,6; GLmp,2; GLže,6
 pokazati · Gae,3; Gbe; Gcd; Gce,57; Gcp,2;
 GLmd; GLme,143; GLmp,8; GLse,14; GLžd;
 GLže,58; GLžp,2; GNE,11; GVbe,7
 pokimati · Gce,16; Gcp,2; GLme,20; GLže,5
 poklekni · Gce,6; GLmd; GLme,9; GLmp,3;
 GLže,6; GVap
 poklic · Sme1,6; Sme2,3; Sme4,4; Sme5,4
 poklicati · Gae,4; Gce,13; GLme,38; GLmp,9;
 GLse; GLže,24; GNE,3; GVap,4; GVbe,4;
 GVbp,2
 pokojen · Pme1i,4; Pme2; Pme4; Pme5,2; Pmp1;
 Pmp6; Pže1; Pže2; Pže4
 pokonci · A,27
 pokončati · Gcp; GLme,2; GLmp,3; GLže; GNE;
 GVbp,3
 pokopališče · Sse1; Sse2,5; Sse3,3; Sse4,7;
 Sse5,10
 pokopati · Gap; Gbp; Gce; Gcp,4; GLme,4;
 GLmp,20; GLže; GVap
 pokositi · Gbe; Gce,3; Gcp; Gcp; GLmd;
 GLmp,2; GNE
 pokrajina · Sže1,23; Sže2,5; Sže3; Sže4,10;
 Sže5,4; Sže6; Sžp1; Sžp4; Sžp5,4
 pokrit · GTme1,4; GTmp1,2; GTse1; GTže1,7;
 GTžp1; PTmp1,2; PTže1; PTže3; PTžp1;
 PTžp2
 pokriti · Gce,8; Gcp; GLmd; GLme,10; GLže,9;
 GVbe
 pokrivati · Gce,2; Gcp,2; GLme,3; GLže,3

- pokrižati · Gce,4; GLme,3; GLže,4; GNE
 pokrov · Sme1,6; Sme2,2; Sme4,8; Sme5,2;
 Sme6,7
 pol · A,51; ŠNE,43
 polastiti se · Gce; GLme,6; GLmp; GLse; GLže
 poldan · Sme1; Sme2,2; Sme3; Sme4,6; Sme5;
 Sse6
 Polde · IOme1,15
 poleči · Gce,5; Gcp,2; GLme,9; GLmp,10;
 GLse,2; GLže,3; GLžp; GNE
 poleg · E2,46
 polenta · Šze1,2; Šze2,8; Šze3; Šze4,5; Šze5
 poleten · Pme1i,4; Pme2,3; Pme4; Pme5; Pmp1;
 Pse1,4; Pse2,3; Pse4; Pse5; Pže1,3; Pže2,2;
 Pže4; Pže5,3; Pžp4; Pžp5
 poleti · A,11
 poletje · Sse1,20; Sse2,10; Sse4,3; Sse5,4; Ssp4
 polglasno · A,18
 polica · Šze1,8; Šze2,7; Šze3,4; Šze4,14; Šze5,21;
 Šžp1; Šžp5,8
 policaj · Sme1,6; Sme2; Sme3; Sme4; Smp1;
 Smp2
 policija · Šze1,8; Šze3,4; Šze4,2; Šze5,5
 policijski · Pme1,5; Pme2,4; Pme4; Pme4;
 Pme5; Pmp1; Pmp6; Pse6,2; Pže1,4; Pže2,4;
 Pže5
 politika · Šze1,6; Šze2,9; Šze4,11; Šze6
 polje · Sse1,6; Sse2,9; Sse3,2; Sse4,18; Sse5,31;
 Sse6,7; Ssp1,10; Ssp2,4; Ssp4,6; Ssp5,2
 poljub · Sme1,5; Sme2; Sme4,4; Smp4; Smp5
 poljubiti · Gce; Gcp; GLmd,2; GLme,9; GLmp;
 GLže,7; GLžp; GNE,3
 polmrak · Sme1,3; Sme2,9; Sme4,5; Sme5,8;
 Sme6,2
 poln · Pme1i; Pme1,18; Pme2; Pme2; Pme4;
 Pme5,3; Pme6,3; Pmp1,8; Pmp4; Pmp6,4;
 Pse1,17; Pse2,5; Pse5; Psp1; Psp2,4; Psp6,2;
 Pžd1; Pže1,30; Pže2,4; Pže4,7; Pže5,2;
 Pže6,2; Pžp1,10; Pžp2; Pžp4,7; Pžp5; Pžp6,2
 polniti · Gce,2; Gcp; GLmp; GLse,2; GLže,3;
 GLžp
 polno · A,12
 polnoč · A,6; Šze4; Šze6,3
 polnočnica · Šze2; Šze3,7; Šze5,2
 Polog · IZme4,11; IZme5,4
 Pologar · IOme1,21; IOme3; IOme4
 Pologarjev · PIOme1,6; PIOme6,2; PIOže1;
 PIOže4
 polomiti · Gce,3; GLme,8; GLmp; GLže,3
 Polona · IOže1,5; IOže3; IOže4,4
 polovica · Šzd4; Šze1,6; Šze2; Šze4,14; Šze6
 položaj · Sme1,5; Sme2,3; Sme3; Sme4,2;
 Sme5,3; Sme6; Smp4,3
 položen · GNžp1; Pme5,3; PNme1; Pse1; Pse5;
 Pse5j; Pže1,3; Pže2; Pžp1; Pžp2; Pžp5,2
 položiti · Gap; Gbe; Gcd,5; Gce,30; Gcp,2;
 GLmd,3; GLme,104; GLmp,11; GLže,28;
 GLžp,2; GNE,7
 polzeti · Gcd; Gce; GLme; GLže,4; GLžp,3
 pomagati · Gae,2; Gce,17; Gcp,2; GLme,10;
 GLmp,2; GLse,12; GLže,3; GLžp; GNA,2;
 GNE,5; GVbe,3; GVbp,7
 pomahati · Gce,2; Gcp; GLme,11; GLmp;
 GLže,2
 pomakniti · Gae; Gcd; Gce,17; GLmd,2;
 GLme,10; GLmp,3; GLže,4; GNE
 pomemben · Pme1; Pme1i; Pme4; Pme5; Pmp2;
 Pse1,2; Pse4; Pže1,3
 pomeniti · Gap; Gcd; Gce,14; Gcp,7; GLme;
 GLmp,2; GLse,3; GLže,4; GNE
 pomesti · Gcp; GLme,2; GLme,7; GLže,2
 pometati · Gae,2; Gbe; Gce; Gce; Gcp; Gcp,3;
 GLme,4; GLmp,2; GLže,2; GLžp; GVbe
 pomežikniti · Gcd; Gce,3; GLme,7; GLže,3
 pomikati · Gae,2; Gcd,2; Gce,5; Gcp,5; GLme,6;
 GLmp,2; GLse,2; GLže,6; GLžp,2; GNE,4
 pomiriti · Gae,2; Gbe,2; Gbp; Gce,18; Gcp;
 GLme,33; GLmp,2; GLse; GLsp; GLže,11;
 GNE,4
 pomirljivo · Aj,2; A,8
 pomisliti · Gae,4; Gap,3; Gce,12; GLme,74;
 GLže,15; GLžp; GNE; GVbe,6; GVbp,3
 pomlad · Šze1,46; Šze2,11; Šze3; Šze4,36;
 Šze5,6; Šze6; Šžp2,2; Šžp4
 pomladanski · Pme1; Pme5,4; Pmp5; Pse4,2;
 Pse5; Pže1,5; Pže4,2
 pomladen · Pme1i,9; Pme2,2; Pme4i,4; Pme5,2;
 Pse1,3; Pse2,5; Pse4,2; Pse5,2; Psp4; Pže1,2;
 Pže2,2; Pže5; Pžp1,2
 pomleti · Gcp; GLme,9
 pomoč · Šze1,2; Šze2,12; Šze4,11; Šze6,5
 pomolčati · Gce,3; GLmd; GLme,12; GLže,13
 pomoliti · Gce,4; Gcp; GLme,5; GLže,2
 pomračiti se · Gce,8; Gcp; Gcp; GLme,3; GLže,2
 ponavljati · Gae,4; Gap; Gce,10; Gcp; GLme,37;
 GLmp,5; GLže,6; GLžp; GNE,2
 ponižno · A,10
 ponoči · A,64
 ponos · Sme1,16; Sme2,15; Sme4,3; Sme6,23
 ponosen · Pmd1; Pme1,20; Pme6,3; Pmp1,6;
 Pmp2; Pžd1; Pže1,2
 ponosno · A,63

- ponoviti · Gae,2; Gcd; Gce,26; Gcp,2;
GLme,164; GLmp,2; GLse; GLže,27; GLžp,2;
GNE
- ponovno · A,11
- ponuditi · Gbe; Gce,3; Gcp; GLme,12; GLže,5
- ponujati · Gae; Gce,3; Gcp; GLme,4; GLmp;
GLže,3; GNA,2; GNE
- popiti · Gcd; Gce; GLmd,2; GLme,8; GLmp,4;
GNE; GVbp
- popoldan · Sme1; Sme2,8; Sme5,2; Smp4,2;
Smp5
- popoldne · A,26; Sse1,3; Sse4,16
- popoln · Pse1,2; Pže1,4; Pže2,4; Pže3; Pže4;
Pže5,3
- popolnoma · A,155
- popraskati se · Gce; GLme,80; GLmp; GLže;
GVbp
- popraviti · Gae; Gce,4; GLme,9; GLmp,3;
GLže,10; GNE,3
- popravljati · Gcd; Gce,5; Gcp; GLmd; GLme,3;
GLmp,2; GLže,2; GNE,4
- poraščen · GNme1,2; GNže1,8; PNme1,2;
PNme5; PNmp2; PNmp6,2; PNse1; PNse2,2;
PNse4; PNse5; PNsp1; PNsp2; PNsp5; PNže1;
PNže4,2; PNže5,7; PNžp1; PNžp5,2
- poreči · Gad; Gae,5; Gbe,3; Gce,11; Gcp,5
- poredkoma · A,14
- poribati · Gae; GLmd,2; GLme,12
- poročiti se · Gae; Gbe; Gce,2; GLme,9; GLže,4;
GLžp; GNE,4
- poroditi · GLme; GLse,2; GLžd; GLže,9; GLžp,2
- poroka · Sže1,3; Sže2; Sže4,2; Sže5,4; Sže6
- posaditi · Gce,2; GLme,10; GLže,7
- posebej · A,23
- poseben · Pme1; Pme2; Pme4,3; Pme6,2;
Pmp1,2; Pmp2; Pmp4; Pse1,2; Pse2,8; Psp1,2;
Psp4; Pže1,2; Pže2,2; Pže4; Pže5,4; Pžp1,2;
Pžp2; Pžp6
- posebno · A,78
- poseči · Gce,3; GLme,7; GLmp,2; GLže,4; GNE
- posedati · Gbe; Gce,2; Gcp; GLme,16; GLmp,3
- posejan · GNse1; GNže1,6; PNme1; PNmp2,2;
PNže4
- posekati · Gce; Gcp; Gcp; GLme,5; GLmp,7;
GNE,2
- posel · Sme1,2; Sme2,2; Sme4,6; Sme5,4;
Smp4,3
- posesti · GLmp,10; GLžp,4
- posijati · Gce,3; GLmp; GLse,7
- poskakovati · Gce,7; GLmd; GLme,2; GLmp,4;
GLže,2; GLžp
- poskočiti · Gcd; Gce,10; GLmd; GLme,45;
GLmp,3; GLse; GLže,11; GVbe
- poslati · Gae; Gap; Gce; Gcp,3; GLme,18;
GLmp,27; GLže,28; GNE,2; GVbe; GVbp
- poslednji · Pme1,4; Pme2; Pme4,3; Pme5;
Pme6,5; Pmp2; Pmp4; Pmp5; Pmp6; Pse1,3;
Pse2; Pse4; Psp1; Pžd4; Pže1; Pže4,5; Pžp1,2;
Pžp4,2; Pžp6,3
- posloviti se · Gae; Gce; GLmd,2; GLme,18;
GLmp,5; GLže,3; GNE,3
- poslušati · Gae,15; Gbe,11; Gce,13; Gcp,4;
GLme,52; GLmp,10; GLže,6; GNE,5;
GVbe,13; GVbp,2
- posmeh · Sme1,4; Sme2; Sme6,6
- posmehovati se · Gce,2; Gcp,4; GLmp,4
- posoda · Sže1,3; Sže2,2; Sže3; Sže4,6; Sže5,2;
Sže6
- pospraviti · Gae; Gce; Gcp; GLme,3; GLmp,2;
GLže,5; GNE
- posrečiti se · Gce,4; Gcp,2; GLse,5; GLže,4
- postaja · Sže1; Sže2; Sže3; Sže4,5; Sže5,3; Sžp5
- postajati · Gce,4; Gcp,3; GLme,15; GLmp,3;
GLse,2; GLže,5; GLžp,3; GNE
- postati · Gbe,3; Gce,6; Gcp,6; GLmd,4;
GLme,47; GLmp,4; GLse,5; GLsp,2; GLže,20;
GLžp,4; GNE; GVbe
- postava · Sžd4; Sže1,20; Sže2,26; Sže4,4; Sže6;
Sžp1,3; Sžp2; Sžp4,3; Sžp5,2; Sžp6
- postaviti · Gae; Gcd; Gce,12; Gcp,2; GLmd,3;
GLme,69; GLmp,29; GLse; GLsp; GLže,25;
GNE,8
- postavljati · Gce; Gcp,5; GLme,4; GLmp,2;
GLže
- postelja · Sže1,15; Sže2,31; Sže3,10; Sže4,46;
Sže5,32; Sže6,2; Sžp1,3; Sžp2,3; Sžp4,4;
Sžp5,2
- postojanka · Sže1,2; Sže2,7; Sže3,2; Sže4,9;
Sže5; Sže6,2; Sžp1; Sžp2; Sžp3; Sžp4; Sžp5,2;
Sžp6
- posušiti · Gce,2; GLme,4; GLmp,4; GLže,6;
GLžp
- posvetiti · Gcp; GLme,9; GLmp; GLse,6; GLže,3
- poševen · Pme5,2; Pmp1,2; Pmp2; Pmp4,2;
Psp4; Pže1; Pže2; Pže5,4; Pžp1; Pžp6
- posevno · A,10
- pošiljati · Gae; Gbe,2; Gce,7; Gcp,2; GLme,6;
GLžd; GLže,2; GNE
- poštar · Sme1,37; Sme2,3; Sme3,2; Sme4,4;
Sme5; Smp1; Smp4; Smp6
- pošten · Pme1,9; Pme2; Pme3; Pme4i,3; Pme6;
Pmp1,2; Pmp2,2; Pmp4,4; Pse1,2; Pse2;
Pse4,2; Pže1,3; Pže2; Pže3; Pžp1

- pošteno · A,20
 pot · Sme1,7; Sme2; Sme4,10; Sme5,4; Sme6;
 Smp1; Smp5; Sže1,34; Sže2,28; Sže4,90;
 Sže5,38; Sže6,13; Sžp1,2; Sžp2,2; Sžp4,2;
 Sžp5,8
 potegniti · Gae,2; Gcd; Gce,80; Gcp,6; GLmd,5;
 GLme,177; GLmp,21; GLse,3; GLže,49;
 GNE,5; GVbe
 potem · A,754
 poten · Pme1; Pme1; Pmeli; Pme2; Pme4i,10;
 Pme5; Pmp6; Pse1; Pse4,11; Pže1,3; Pže4,2;
 Pže5; Pžp1,4; Pžp4,3; Pžp5; Pžp6
 potepati se · Gae,2; Gce,4; Gcp,2; GLmd,2;
 GLme,6; GLmp,2; GNE,4
 potepuh · Sme1,18; Sme2; Sme4; Sme5,2; Sme6;
 Smp1,4; Smp3; Smp4
 poteza · Sže2; Sžp1,12; Sžp2,2; Sžp4,2; Sžp5
 potipati · Gae; Gcp; GLme,9; GLže
 potiskati · Gae; Gce,11; Gcp,2; GLme,3;
 GLmp,2; GLže,3; GNE,6
 potisniti · Gae,2; Gce,26; GLmd; GLme,57;
 GLmp,9; GLse; GLže,16
 potok · Sme1,8; Sme2,3; Sme3,8; Sme4,6;
 Sme5,17; Sme6,5; Smp1,6; Smp2,2; Smp4;
 Smp5; Smp6,2
 potolažiti · Gae; Gce,3; Gcp; GLme,12; GLse;
 GNE,3
 potolči · Gcd; Gce,2; Gcp,3; GLme,10
 potovati · Gae,3; Gap; Gce,4; GLme,2; GNE
 potovka · Sže1,8; Sže2; Sže3,2; Sže4; Sže5; Sžp2;
 Sžp4
 potrditi · Gce; GLmd; GLme,12; GLmp,2; GLse;
 GLže,2; GNE; GVbe,4
 potreba · PD,9; Sže2,12; Sže4; Sže5; Sžp2; Sžp4;
 Sžp5
 potreben · Pme1,16; Pme4; Pmp1,4; Pse1,10;
 Pse4; Pse6; Pže1,11; Pže4; Pžp1; Pžp2
 potrebovati · Gae,3; Gce,4; Gcp; GLme;
 GLmp,6; GLže
 potrepljati · Gbe; Gce,10; GLmd,2; GLme,19;
 GLmp; GLže,4; GNE
 potrkati · Gce,2; GLme,13; GLse,3; GLže,4;
 GNE,3
 potrpeti · Gbp,2; Gce,15; GLme,2; GLže;
 GNE,3; GVbe,2; GVbp,4
 potuhniti se · Gcp,2; GLme,4; GLmp,4; GLže,7
 poudarek · Sme2; Sme6,12
 povabiti · Gce; GLmd; GLme,11; GLmp,2
 povedati · Gae,40; Gap,4; Gbp; Gcd; Gce,22;
 Gcp,3; GLmd; GLme,227; GLmp,30; GLse,2;
 GLže,79; GNA,2; GNE,64; GVbe,48; GVbp,16
 povesiti · Gce,5; GLme,5; GLmp; GLže,6
 povest · Sže1,16; Sže2; Sže4,14; Sže5; Sže6,2;
 Sžp1; Sžp2; Sžp4,17; Sžp5
 Poviškaj · IOme1,30; IOme2,10; IOme3,3;
 IOme4,10; IOme5,2; IOme6,4
 poviškati · Gce; Gcp; GLme,5; GLže; GVbe,9
 povodenj · Sže1,15; Sže2; Sžp1; Sžp3; Sžp5
 povodnja · Sže5,10
 povrniti · Gce,3; GLme,14; GLmp,2; GLse;
 GLže
 povsem · A,11
 povesod · A,33
 pozabiti · Gad; Gae,2; Gbp; Gce,13; Gcp,2;
 GLmd,3; GLme,71; GLmp,8; GLže,10; GLžp;
 GNE,5; GVbe,6
 pozdrav · Sme1,14; Sme2; Sme4,15; Sme5;
 Smp4,2
 pozdraviti · Gbp; Gce,21; Gcp,2; GLmd;
 GLme,30; GLmp; GLže,5; GNE; GVbe
 pozdravljati · Gap; Gce,3; Gcp,5; GLme,4;
 GLmp; GNE,4
 pozen · Pme2,2; Pme5; Pse2,2; Pse5; Pže1,2;
 Pže2,3; Pže5,10
 pozibavati · Gce; GLmd; GLme,2; GLmp;
 GLže,3; GLžp; GNE
 pozimi · A,29
 poznati · Gae,16; Gap,8; Gbe,11; Gbp,4; Gce,15;
 Gcp,6; GLmd,2; GLme,32; GLmp,16; GLse,3;
 GLsp; GLžd; GLže,11; GLžp,4; GNE
 pozno · Aj; A,14
 požgan · PNmp2; PNmp4; PNže1,5; PNže2,2;
 PNže4; PNže5,2; PNže6,2; PNžp2; PNžp6,2
 požgati · GLmp,14
 požirati · Gad; Gae; Gbe; Gce,3; GLme,5;
 GLmp; GLse; GLže,4; GLžp; GNE,2
 požirek · Sme2; Sme4,17; Sme5,2; Sme6,6;
 Smp2; Smp6
 požreti · Gbe; Gce,4; GLmd; GLme,15; GLmp,2;
 GLže,4; GLžp; GNE; GVbe
 požugati · Gce,3; GLme,14; GLmp,3; GLže,7
 požvižgavati · Gcd; Gce,2; Gcp,3; GLme,6;
 GNE,2
 prag · Sme1,4; Sme2,26; Sme3; Sme4,75;
 Sme5,66; Sme6,7; Smp1; Smp4; Smp5;
 Smp5,2
 prah · Sme1,10; Sme2,10; Sme4,23; Sme5,9;
 Sme6,5
 pramen · Smd5; Sme1,5; Sme2,2; Sme4,10;
 Sme5; Sme6
 praskati · Gbp,3; Gce,5; GLme,2; GLže,3; GLžp;
 GNE,2
 prašen · Pmp1; Pse1,2; Pže1,2; Pže2,3; Pže4,2;
 Pže5,4; Pžp1; Pžp4

prašič · Smd4; Sme4,4; Smp1,2; Smp2,4;
Smp3,3; Smp4,2; Smp5
prati · Gce,4; Gcp,3; GLme,2; GLmp; GLže,2;
GNA; GNE,2
pratika · Šze1,4; Šze4,4; Šze6,2
prav · A,11; Č,737; PD,82; Sme4
pravi · Pmd1,5; Pmd3; Pme1i; Pme1,30;
Pme2,3; Pme3; Pme4i,5; Pme4,5; Pme5,17;
Pme6,2; Pmp1,8; Pmp2; Pmp4; Pse1,7;
Pse2,4; Pse4,2; Pse5,2; Pse6,2; Pžd1; Pže1,63;
Pže2,18; Pže4,18; Pže5,3; Pže6,7; Pžp1,5;
Pžp4
pravica · Šze1,4; Šze2,8; Šze4,13; Šze5,14; Šžp1;
Šžp2,2; Šžp3; Šžp4,2; Šžp5
pravičen · Pme1i,4; Pme1,4; Pme2; Pse1; Pže1;
Pže5; Pžp1; Pžp4; Sme4
pravilno · A,10
praviti · Gae,61; Gbe,15; Gbp,5; Gce,168;
Gcp,24; Gcp,59; GLme,14
pravkar · A,36
pravljica · Šze1,12; Šze2,6; Šze4,17; Šze5,2;
Šze6; Šžp1,4; Šžp2,3; Šžp4,6
pravzaprav · A,86
prazen · Pmd1; Pme1,15; Pme2,4; Pme4i,2;
Pme4,4; Pme5,3; Pme6,3; Pmp1,2; Pmp5;
Pmp6,2; Pse1,13; Pse2; Pse4; Pse5,3; Pse6,2;
Psp6; Pže1,36; Pže2,6; Pže4,11; Pže5,5;
Pže6,3; Pžp1,9; Pžp2,7; Pžp4,4; Pžp6,4
prazničen · Pme1i; Pme4i,3; Pme5; Pmp1;
Pmp5; Pže2; Pže4,2; Pžp5
praznik · Sme1,10; Sme2,3; Sme4,3; Sme5;
Sme6,2; Smp1,3; Smp2; Smp4,7; Smp5,5;
Smp6
praznina · Šze1,4; Šze2; Šze3; Šze4,6; Šze5
prebirati · Gae; Gce,6; Gcp,4; GLmd,2; GLme,9;
GLže,2; GNE,3
prebiti · Gap; Gce,2; GLme,2; GLmp; GLže,3;
GNE,3
prebledeti · Gce,4; GLme,9; GLže,3
prebrati · Gce,2; GLmd; GLme,15; GLmp,2;
GLže,2; GNE,3
prebrestiti · Gce; Gcp; GLme,6; GLmp; GNE
precej · A,109; ŠNE,38
precejšnji · Pme6,2; Pse4; Pže1,5; Pže2; Pže4,3;
Pže5,2
pred · E; E4,96; E5; E6,667
predati · Gce,2; Gcp; GLme,3; GLmp,2; GLže,2
predebel · Pme1; Pže1,12; Pže4
preden · A,134
predmestje · Sse2,3; Sse4; Sse5,3; Ssp4; Ssp5,3
predor · Smd6; Sme2,2; Sme4,3; Sme5,2;
Sme6,2; Smp6,3

predpasnik · Sme2,3; Sme4,32; Sme5,2;
Sme6,23; Smp4,2; Smp6
preganjati · Gcd; Gce,2; Gcp,3; GLme,7;
GLmp,7; GLže; GNE,4
pregledati · GLme,9; GLmp,2; GLže; GNE;
GVad
pregnati · Gae,2; Gbe; Gce; GLmd; GLme,2;
GLmp; GLže; GNE
pregreha · Šze1; Šze2; Šze4,4; Šze5; Šze6;
Šžp2,2; Šžp4,4; Šžp6
pregrešen · Pmp2; Pse2,3; Pse4; Pse6; Pže1,2;
Pže4; Pžp1; Pžp2
pregrešiti se · Gce; GLme,8; GLmp
pregrniti · GLme; GLže,9
pregrnjen · GNmd1; GNmp1; GNse1; GNže1,3;
GNžp1,2; PNme1,2; PNme4i; PNže4; PNže6
prehiteti · Gce,3; Gcp,4; GLmd; GLme,5;
GLmp,3
prehladiti · Gae; Gce,2; GLme,2; GLmp,2;
GLže,4
prej · A,126; A,158
prejeti · Gbe; Gbp; Gce; GLme,10; GLmp,2;
GLsp; GLže,4
prejšnji · Pme1i,2; Pme2; Pme4,10; Pme6,4;
Pmp1; Pmp2,2; Pmp5; Pse1; Pže1; Pže2,4;
Pže4; Pže5; Pže6,3; Pžp1; Pžp2
prekiniti · Gce,8; GLmd; GLme,23; GLmp;
GLže,11
prekladati · Gae; Gce,3; Gcp,4; GLme,12;
GLmp,5; GLže,9; GNE,2; GVbe
preklet · PNmp2; PNže1,14; PNže2; PNžp1,2;
PTme1i,5; PTme4; PTmp1,2; PTse1,2;
PTžd1; PTže5,2
Prekleta strešnica · IOme1,64; IOže2; IOže3,2;
IOže4,5
preklinjati · Gae; Gce,3; Gcp,2; GLme,10;
GLmp,8; GLže,2; GLžp; GNE,2; GVbe
preko · E2,27
prekrižati · Gae,2; Gce,7; GLme,15; GLmp,3;
GLže,22; GNE
prelivati · Gce,6; Gcp,2; GLme; GLmp,4;
GLse,5; GLžd; GLže,13
premagati · Gcd; Gce; GLmd; GLme,10;
GLmp,2; GLse; GLže,2; GLžp; GNE,15
premakniti · Gap; Gcd; Gce,9; Gcp,4; GLmd;
GLme,25; GLmp,2; GLse,3; GLže,5; GLžp,2;
GNE,8
premeriti · Gcd; Gce,23; Gcp,2; GLme,42;
GLmp; GLže,15
premetavati · Gae; Gbe; Gbp; Gce,7; Gcp,2;
GLme,12; GLmp,6; GLse; GLže,4; GLžp,2

- premikati · Gcp,3; GLse,2; GLže,3; GLžp;
GNE,4
- premisliiti · Gce,2; Gcp; GLmd; GLme,27;
GLmp,2; GLže,7; GNE,3; GVbe,3
- premišljati · Gae,4; Gce,4; GLme,17; GNE
- premišljavati · GLmd; GLme,5; GLmp; GLže,4;
GNE
- premolki · Sme5,41
- premrli · PLme1; PLmp4; PLže6; PLžp3;
PLžp4,4; PLžp6,2
- prenašati · Gae; Gap; Gce,2; Gcp,3; GLme,9;
GLmp,2; GLže,3; GNE,2
- prenehati · Gae; Gce,3; GLme,2; GLmp; GLse;
GLže; GLžp
- prenesti · Gae,3; Gce; GLmd,2; GLme,5;
GLže,2; GNE,5
- preorati · GLmd; GLme,7; GLže; GLžp
- prepad · Sme1,6; Sme2,5; Sme4,10; Sme5;
Sme6,3; Smp1; Smp2; Smp4
- prepevati · Gcp,3; GLme,9; GLmp,5; GLse;
GLže,3; GLžp,2; GNE,3
- prepirati se · Gae; Gap; Gbe; Gcp,3; GLmd,3;
GLme; GLmp,4; GLže,3; GLžp,4; GNE,2
- preplašen · GNmp1,2; PNme1,4; PNme1,4;
PNme3; PNme4; PNme4; PNmp1; PNmp2;
PNmp4; PNse1; PNse6; PNže1,5; PNže4;
PNžp1,2
- preplašeno · A,58
- prepozno · A,10
- prepričan · GNme1,28; GNmp1,5; GNže1,3
- prepričanje · Sse1,3; Sse2,2; Sse3; Sse4,3; Sse5;
Sse6,2; Ssp4
- prepričati · Gce; GLme,12; GNE,3
- preprost · Pme1,2; Pme2; Pme4,2; Pme5;
Pmp1,3; Pmp3; Pse1,8; Pse2; Pse5; Pže1,5;
Pže2,2; Pže4; Pžp5
- preriniti se · Gce,6; GLmd; GLme,4; GLmp;
GLse
- Preseka · IZže2,3; IZže4; IZže5,8
- preseliti se · GLmd; GLme,5; GLmp,3; GLže,6;
GNE
- presledki · Sme2; Sme5,2; Smp5,4; Sžp5,5
- presneto · A,14
- presoditi · Gce,2; GLme,4; GLže; GNE,5
- presojati · Gce,2; GLme,5; GLse; GLže; GNE
- prestopati se · Gce,3; Gcp; GLmd,2; GLme,6;
GLže,3; GLžp; GVbe
- prestopiti · Gce; GLme,15; GLmp,2; GLže,2
- prestrašeno · A,19
- prestrašiti · GLme,9; GLže,4
- presunljivo · A,13
- Presveta Moštaca · IOme1,8; IOme4,2
- prešerno · A,18
- prešiniti · Gce,6; GLme,2; GLse,23; GLže,11
- prešteti · Gce; Gcp,2; GLme,5; GLže,3; GLžp;
GVbe
- pretegniti · Gce; GLme,13; GLže
- pretep · Sme1,4; Sme2; Smp2; Smp4; Smp5;
Smp6,2
- pretepati · Gap,2; Gce,7; GLmd; GLme,7;
GLmp,4; GLse; GLže; GNE
- pretepsti · Gce,2; Gcp; GLmd; GLme,7;
GLmp,8; GLže,3
- pretikati · Gce,4; GLmd,3; GLme,5
- pretresljiv · Pme1; Pme1; Pmp1; Pmp4; Pse1;
Pse2; Pže1,2; Pže4,4; Pžp2,2
- pretresti · Gce,3; Gcp; GLme,6; GLmp; GLse,6;
GLže,6; GLžp
- pretrgati · Gce,2; GLme,5; GLmp,2; GLže,6
- preveč · A,129; ŠNE,2
- prevelik · Pme4,2; Pmp1; Pmp4; Pse6; Pže1;
Pže2; Pže4; Pže5; Pže6; Pžp4
- previdno · A,71
- prevreči se · Gce,5; GLme,4; GLsp; GLže,4
- prevzeti · Gce; Gcp; GLme,3; GLse,3; GLže,8
- prezgodaj · A,12
- preživeti · Gcd,2; Gce,2; Gcp; GLmd,3;
GLme,12; GLmp; GLže; GNE
- prgišče · Sse2; Sse4,10; Sse6; Ssp5,2
- prhniti · Gce,3; GLme,38; GLže,8
- pri · E5,899
- pribiti · Gae; Gce,14; GLme,60; GLže,7; GVbe
- približati se · Gcd,2; Gce,11; Gcp,3; GLmd,2;
GLme,5; GLmp,3; GLže,6; GNE
- pribočnik · Sme1,25; Sme2; Sme3,2; Sme4;
Sme6; Smp3
- priča · Sže1,4; Sže4,2; Sže5,39; Sžp1; Sžp4
- pričakovanje · SGse1; SGse2,9; SGse4,3;
SGse5,4; Sse6,3; Ssp5
- pričakovati · Gae,2; Gap,2; Gbe; Gbp; Gce,4;
Gcp,4; GLmd; GLme,15; GLmp,7; GLže,11;
GLžp,4; GNE,2
- prida · A,9; PD,8
- pridejati · Gae,14; Gcp,24
- priden · Pmd1; Pme1,8; Pmp1,2; Pže1,8
- pridiga · Sže1,4; Sže2,3; Sže4,10; Sže5; Sže6,2;
Sžp1,2; Sžp2,2; Sžp4; Sžp5
- pridigati · Gbe; Gce,2; Gcp; GLme,9; GLmp,2;
GLže,3; GNE,2; GVbe
- pridno · A,23
- pridrvti · Gcd; Gce,6; Gcp,2; GLmd,2; GLme,5;
GLmp; GLže,2
- pridušeno · A,14

priganjati · Gce,4; GLmd; GLme; GLže,3;
GNE,5
prigaziti · GLme,9; GLže
prignati · Gae; Gbp; GLmd,4; GLme,5; GLmp,6;
GLse
prigovarjati · Gce; GLme,8; GLmp; GLse;
GLže,4; GNE,5
prihajati · Gae; Gbe; Gcd; Gce,30; Gcp,15;
GLme,13; GLmp,22; GLse,5; GLsp; GLžd;
GLže,18; GLžp,8
prijatelj · Smd1,7; Sme1,10; Sme3,3; Sme5;
Sme6; Smp1,5; Smp2,2; Smp3; Smp4,2; Smp6
prijazen · Pme1,7; Pmp1; Pse6; Pže1,9; Pže2;
Pže5; Pžp1
prijazno · Aj; A,57
prijemati · Gce; GLme,4; GLmp,3; GLže; GNE
prijeten · Pme1i; Pme1,7; Pme4,3; Pme6,2;
Pse1,15; Pse2; Pse3; Pse6; Pže1,5; Pže4;
Pže5; Pžp2; Pžp4,2
prijeti · Gae,3; Gap,2; Gce; Gce,19; Gcp,5;
GLmd,5; GLme,67; GLmp,29; GLse,3;
GLže,23; GNE,8; GVbe,2
prijetno · Aj; Aj; A,21
prikazati se · Gce,13; Gcp,8; GLmd; GLme,17;
GLmp,8; GLse,3; GLžd; GLže,45; GLžp,3;
GVbe
prikimati · Gae,2; Gcd; Gce,45; Gcp,2; GLmd;
GLme,171; GLmp,9; GLse,4; GLže,40; GLžp
prikimavati · Gce,5; Gcp; GLmd; GLme,6;
GLmp,2; GLže,2; GNE,2
priklicati · Gce; GLme,5; GLmp,2; GLžd; GNE
prikrit · PNme2,2; PTme1,3; PTme6; PTse6,4
prikupen · Pme1i; Pme1,5; Pme3; Pže1,7
prileten · Pmd1,2; Pme1i; Pmp1i; Pmp2;
Pže1,3; Pže2; Pžp2
prileteti · Gcd; Gce,11; Gcp; GLmd,2; GLme,16;
GLmp; GLže,11; GNE
prilika · Šze1; Šze4,3; Šze5,8; Šžp1,2; Šžp4;
Šžp5,4; Šžp6
priložnost · Šze1,3; Šze2; Šze4,7; Šze5,3
primer · Sme1,3; Sme4,15; Sme5,4; Smp2,2;
Smp5,4
primera · Šze1,3; Šze2,3; Šze3; Šze4; Šze5,6;
Šžp4
primeren · Pme1; Pme2; Pme4; Pme6; Pmp1,2;
Pmp2; Pse1,2; Pse2,3; Pse4,2; Pže1; Pže2,3;
Pže4,4; Pže5,2; Pžp1; Pžp2; Pžp4
Primorska · IZže1,5; IZže2,3; IZže4,5
prinašati · Gap,2; Gce,6; Gcp,3; GLme,2;
GLmp,2; GLže,4; GNE

prinesti · Gae,4; Gcd,2; Gce,15; Gcp,3; GLmd;
GLme,29; GLmp,12; GLse; GLže,34; GLžp,4;
GNE,3; GVbe,3; GVbp
pripeljati · Gae; Gce,2; GLmd,3; GLme,24;
GLmp,11; GLže,2; GNE,3
pripetiti se · Gce,3; Gcp; GLse,11; GLže;
GLžp,2; GNE
priplavati · Gce; GLme,2; GLse,3; GLžd; GLže,3
pripoditi · Gcd,2; Gce,5; Gcp; GLmp,2
pripomniti · GLme,7; GLmp; GLže,4; GNE,4
pripovedovanje · SGse1; SGse4,4; SGse6,2;
Sse2,4; Sse5,7
pripovedovati · Gae,10; Gap; Gce,4; Gcp,4;
GLmd,3; GLme,27; GLmp; GLže,10; GNE,12;
GVbe
pripraviti · Gae; Gce,4; Gcp; GLme,10; GLmp,4;
GLže,11; GLžp; GNE,3; GVbe
pripravljati · Gce,10; Gcp,2; GLmd; GLme,14;
GLmp,2; GLže,5; GNE,7; GVbe
pripravljen · GNme1,5; GNmp1,3; GNse1,6;
GNže1,4; GNžp1; GNžp6; PNme1; PNse1;
PNse4; PNže1; PNže4,2
priprt · GTžp1,2; PTse6; PTsp1; PTsp4,4;
PTsp5; PTžp2,3; PTžp4; PTžp5
priskočiti · Gce,3; GLme,9; GLmp,2; GLže,2
prisoniti · Gce,2; Gcp,2; GLme,6; GNsp1
prisluhni · Gae; Gcd; Gce,5; Gcp; Gcp,2;
GLme,24; GLmp,3; GLsp; GLže,4
prisluškovati · Gae,3; Gcp; GLmd; GLme,4;
GLmp,3; GLže
prisluziti · Gae; Gce; GLme,5; GLže,5
prismoda · Šze1,17; Šze6
prisrčno · A,10
pristanišče · Sse1,2; Sse2; Sse4,6; Sse5,3
pristopiti · Gce,2; GLmd,2; GLme,7; GLmp,2;
GLse
pritajen · PNme1i; PNme6,5; PNmp1; PNmp4;
PNse4,2; PNse6,2
pritajeno · A,34
priteči · Gcd; Gce,6; GLme,3; GLmp,3; GLže,11;
GLžp
priti · Gad; Gap,8; Gbe,6; Gcd,4; Gce,109;
GLmd,37; GLme,217; GLmp,90;
GLse,59; GLsp; GLžd,3; GLže,154; GLžp,13
GNE,14; GVbd,2; GVbe,15; GVbp,2
pritisniti · Gce,6; Gcp,4; GLmd; GLme,4;
GNE,2
pritisniti · Gae; Gcd; Gce,23; Gcp,2; GLme,37;
GLžd,2; GLže,10
pritlikav · Pme1i; Pme1,2; Pme2; Pme3;
Pme4,2; Pme6; Pmp1,3; Pmp4,2; Pse1; Pže4

- pritrđiti · Gae; Gce,14; Gcp; GLme,37; GLmp,5;
GLže,9; GNE
- pritrjevati · Gce,6; GLmd; GLme,5; GLže,2
- privaditi se · Gbe,2; Gce; GLme,7; GLmp,2;
GLže,2; GNE
- privezati · Gad; GLmd; GLme,17; GLmp; GLže;
GNE; GVbd,2
- privid · Smd4; Sme1,13; Sme2; Sme3,2; Sme4,3;
Sme5,3; Smp2,3; Smp4
- privleči · Gce,4; GLmd,2; GLme,4; GLmp,7;
GLže,5
- privoščiti · Gbe,2; Gce,3; GLme,6; GLmp; GLže;
GNE,2
- prizadeti · Gce; Gcp,2; GLme; GLse,2; GLže;
GLžp,5; Pmp1
- prizibati se · Gce,2; GLme,14; GLmp; GLže;
GLžp
- priznati · Gae,7; Gbe; Gbp; Gce,11; GLme,38;
GLmp,3; GLže,11; GNE,10; GVbe; GVbp
- prizor · Sme1,7; Sme2,2; Sme4,5; Sme5; Sme6,3;
Smp1,6; Smp4; Smp5,2
- prizgati · Gae; Gce,6; GLme,27; GLmp; GLže,6;
GLžp; GNE,3; GVbe,2
- priznica · Sže1; Sže2,7; Sže4,3; Sže5,4
- prod · Sme2,2; Sme4,8; Sme5,12; Sme6
- prodajati · Gae; Gbe,2; Gcp,5; GLme,5; GLmp;
GLže,2; GNE; GVbp
- prodati · Gae; Gce,7; GLme,11; GLmp,5;
GLse,2; GLže,8; GNE,5
- profesor · Sme1,8; Sme3; Sme6; Smp1
- proga · Sže1,2; Sže2; Sže4,10; Sže5,7; Sže6;
Sžp4,2; Sžp5
- propasti · Gcp; GLmd; GLme,4; GLmp; GLmp;
GLže; GLžp
- proseče · A,35
- prositi · Gae,13; Gce,7; Gcp,2; GLme,33;
GLmp,9; GLže,13; GNE,5; GVbe,4
- prosojen · Pmd1; Pme1,3; Pme5; Pmp1; Pmp4;
Pse2; Psp6; Pže1,4; Pže5; Pže6,2; Pžp1
- prostor · Sme1,10; Sme2,9; Sme4,9; Sme5,20;
Smp1; Smp4,3
- prostoren · Pme2; Pse1; Pže1,6; Pže4; Pže5
- prostran · Pme1; Pme2; Pme6; Pse4,2; Pse5;
Pže1; Pže4,2; Pže5; Pžp4; Pžp6
- proti · E3,377; E,2
- prozoren · Pme1,2; Pse1; Psp5,2; Pže1,2;
Pže4,2; Pže6
- prožen · Pme1,5; Pme2; Pme4i; Pme5; Pme6;
Pmp1; Pmp4
- prsa · Ssp1,2; Ssp4,8
- prsen · Pme1i,2; Pme2; Pme5,7
- prsi · Ssp6,2; Sžp1,4; Sžp2,6; Sžp3; Sžp4,34;
Sžp5,65
- prst · Smd4; Smd6,4; Sme1,2; Sme2,2; Sme4,31;
Sme6,46; Smp1,22; Smp2,12; Smp4,45;
Smp5,32; Smp6,35; Sže2,4; Sžp4,2; Sžp6,3
- prvi · Smp6; ŠVmd1,2; ŠVme1,71; ŠVme2,11;
ŠVme3,2; ŠVme4,8; ŠVme4,23; ŠVme5,10;
ŠVme6,6; ŠVmp1,15; ŠVmp2,2; ŠVmp3,2;
ŠVmp4,6; ŠVmp5,3; ŠVmp6,3; ŠVsd1; ŠVsd4;
ŠVse1,4; ŠVse2,4; ŠVse4,11; ŠVse5,11;
ŠVsp4,4; ŠVsp5,3; ŠVzd1; ŠVzd4; ŠVže1,36;
ŠVže2,11; ŠVže3,2; ŠVže4,23; ŠVže5,19;
ŠVže6,6; ŠVžp1,9; ŠVžp2,5; ŠVžp4,9;
ŠVžp5,3; ŠVžp6
- prvič · A,68
- psovka · Sže1,2; Sže4; Sžp2,3; Sžp4,2; Sžp6,5
- pšenica · Sže1,2; Sže2,7; Sže4,6
- pšeničen · Pmp1,4; Pmp2; Pmp4; Pmp5; Pse4;
Pse6; Psp2; Pže2; Pže4; Pžp4
- ptica · Sžd1,3; Sže1,20; Sže2,2; Sže3,2; Sže4,4;
Sžp1,14; Sžp2,4; Sžp3; Sžp4
- ptič · Sme1,5; Sme2,2; Smp1,3; Smp2,2; Smp4,4;
Smp5,2; Smp6
- ptičnica · Sže1,2; Sže2; Sže4,8; Sže6,5
- puhniti · GLme,10; GLže,6
- pust · Pme1; Pme1i; Pme5,2; Pmp4; Pse5,2;
Psp5; Pže1,11; Pže2; Pže5; Pže6,3; Sme1;
Sme4; Sme5,2
- putiti · Gad,8; Gae,7; Gap,4; Gbp,3; Gce,23;
Gcp; Gcp,5; GLmd,2; GLme,27; GLmp,14;
GLže,8; GNE,4; GVbe,7; GVbp,15; Pme6
- puška · Sžd3; Sžd4; Sže1,3; Sže2,4; Sže3;
Sže4,26; Sže5; Sže6,15; Sžp1,3; Sžp2; Sžp4,9;
Sžp5,4; Sžp6,12
- puškin · Pse4; Pse6,8; Psp1; Psp2; Psp6; Pžp2;
Pžp6
- putrih · Sme1,3; Sme2; Sme4,8
- raba · Sže4,12
- račun · Sme1; Sme4,11; Sme5; Sme6; Smp1,2;
Smp2,2; Smp4,4; Smp6
- računati · Gae; Gce; GLme,8; GLmp; GLže,3;
GNE,2
- rad · PDjj,23; PDj,37; PDmd,8; PDme,211;
PDmp,38; PDže,80; PDžp; PDžp,4
- rado · Ajj; Ajj,10
- radost · Sže1,7; Sže2,5; Sže5; Sže6,4; Sžp2
- radosten · Pme1; Pme2; Pme5; Pme6,2; Pmp2;
Pmp5; Pse1,4; Pse5; Pže1
- radostno · A,14
- radovedno · A,11

rahel · Pme1,3; Pme2jj,2; Pme3jj; Pme4;
 Pme4i,2; Pme4jj; Pme5,3; Pme6,10; Pmp1,2;
 Pmp5; Pmp6; Pse1; Pse2,2; Pse4,3; Pse6,3;
 Psp4; Pže1,9; Pže2; Pže2jj; Pže4,2; Pže6,4;
 Pžp2; Pžp4,3; Pžp5,3; Pžp6

rahlo · A,79

raje · A,10; PD,2

rajnki · Pme1,5; Pme4,2; Pme5; Pže1,2; Pže2;
 Pže5; Sme1,3

rama · Sže4,30; Sže5,21; Sže6,3; Sžp4,9; Sžp5,6

rame · Ssd5; Sse2,6; Sse4,45; Sse5,14; Ssp1;
 Ssp2,4; Ssp3; Ssp4,27; Ssp5,4; Ssp6,59

rana · Sže1,6; Sže2,4; Sže4,10; Sže5; Sže6;
 Sžp1,4; Sžp2,3; Sžp4,6; Sžp5,2

ranjen · GNme1,3; GNže1; PNme1; PNmp2,2;
 PNse2; PNže1,3; PNže2; PNže4

ranjenec · Sme2; Sme4,3; Smp1,6; Smp2,5;
 Smp3; Smp4,4; Smp6,2

raskav · Pme1; Pme6,2; Pse2; Pse3,2; Pse4,4;
 Pse5,3; Pže1; Pže3; Pže4,2; Pže6; Pžp2

rast · Sže2,6; Sže4,3; Sže5

rasti · Gae,2; Gce,10; Gce,24; Gcp; Gcp; Gcp,2;
 Gcp,4; GLmd; GLme,23; GLmp,10; GLse,7;
 GLže,18; GLžp,10; GNE,2

Raus e patacis · IS,20

ravan · Sže2; Sže4,3; Sže5,7

raven · Pme1,4; Pme2,2; Pme6; Pmp1; Pmp4;
 Pse1,4; Psp1,2; Pžd4; Pže1,7; Pže5,3; Pže6;
 Pžp1,2; Pžp5,4; Sže1

ravnatelj · Sme1,16; Sme3; Sme6; Smp5

ravnati · Gce,4; Gcp,4; GLme; GLmp; GLže,3;
 GNE,4

Ravnica · IZže1,5; IZže2,2; IZže3; IZže4,3;
 IZže5,7

Ravničar · IOmd1; IOme1,57; IOme2,4;
 IOme3,6; IOme4,17; IOme5; IOme6,5; IPmp1;
 IPmp5,2

Ravničarica · IOže1,10; IOže6

Ravničarjev · PIOme3; PIOme5; PIOmp1,3;
 PIOmp5; PIOže1; PIOže5; PIOže6; PIOžp1;
 PIOžp3; PIOžp4

ravno · A,28; Č,11; Sse5,16

ravnotežje · Sse1; Sse4,23

razbeljen · GNže4; PNme1,2; PNme2; PNmp5;
 PNse2; PNse4; PNse5; PNže1,2; PNžp1,2;
 PNžp2; PNžp6

razbijati · Gce,4; GLme,5; GLmp,2; GLse,2;
 GLže; GNE

razbit · GTže1; PNmp4; Pse5,2; PTme1;
 PTme5,2; PTme6,2; PTmp2,4; PTmp5;
 PTmp6; PTse2,4; PTse6; PTže2,3; PTžp2;
 PTžp6

razbiti · Gae; Gce,2; GLme,6; GLmp,4; GLže,3;
 GLžp; GNE,2

razbojnik · Smd1; Sme1,14; Sme4,3; Sme5;
 Smp1,2; Smp2,2; Smp5,2; Smp6

razburiti se · Gce,2; GLme,6; GLže,2; GLžp

razburkan · PNme1i; PNmp1; PNmp5;
 PNse2,2; PNsp4; PNže1,3; PNže5,3

razcapan · GNmp1; Pmd1; Pme1,5; Pmp1,3;
 PNmp1,3; Pže1; Pže4,3; Pžp1

razdeliti · GLmd; GLme,3; GLmp,5; GLže

razen · A,4; E2,8; Pmp1,2; Pmp4,4; Pmp6,2;
 Pse4,2; Psp4,5; Psp6; Pže2; Pžp1,2; Pžp2;
 Pžp4,5; Pžp6,2

razgovor · Sme1,4; Sme2,4; Sme3,2; Sme4,6;
 Smp4

razgrniti · Gae; Gcd; Gce,4; GLme,18; GLmp;
 GLsp; GLže,12

razgrnjen · GNme1,3; GNmp1; GNse1;
 GNže1,4; GNžp1; PNže1,3; PNžp4

raziti · GLmd,3; GLmp,7

razjeden · GNžp1; PNmp1,2; PNmp5,3; PNmp6;
 PNse1; PNže1; PNže3; PNže4; PNžp6

razjokati se · Gcp; GLmd; GLme,11; GLmp,2;
 GLže

razkoračiti se · Gce,5; GLme,16; GLmp; GLže,3

razlagati · Gbe; Gcd; Gce; GLme,32; GLmp;
 GLže,2; GNE,10

razlesti se · Gce,4; GLme,3; GLse; GLže,2;
 GLžp

različen · Pmp1jj; Pmp1,2; Pmp2; Pmp4jj;
 Pmp4,2; Pmp6; Pse2; Psp6; Pžp1; Pžp2jj,5;
 Pžp2,2; Pžp4; Pžp6

razliti · Gce,3; GLme,4; GLse,4; GLsp; GLže,9;
 GLžp; GNE,2

razlivati · Gce,7; Gcp; GLme,3; GLže,3; GLžp;
 GNE,2

razločno · A,22

razložiti · Gcp; GLme,21; GLže; GNE,4; GVbe,2

razmajan · GNže1; PNmp4; PNže4,3; PNže5,2;
 PNžp5,3

razmisliti · Gce,2; GLme,6; GLže,2

razmišljati · Gae,6; Gce,12; Gcp,2; GLme,51;
 GLmp,2; GLže,2; GNE,2; GVbe,2

razočarano · A,10

razoran · GNse1; GNsp1,2; PNme1i; PNme2;
 PNse1; PNse2; PNse4; PNse5; PNže4

razpet · GTme1,2; GTže1,2; GTžp1; PTsp6,4;
 PTže3,2

razpoka · Sže1,2; Sže4,5; Sžp4,3

razred · Sme2,3; Sme4,4; Sme5,7; Smp5

razsajati · Gce; Gcp,5; GLme,5

razsoditi · GLme,7; GLže; GLžp,2

- razsuti · Gce,2; GLme,4; GLmp,3; GLže,2
 razširiti · Gae; Gce,14; GLme,32; GLmp,4;
 GLse; GLsp; GLže,15; GLžp; GNE
 raztrgan · GNme1,3; GNmp4; GNže1,4; GNžp1;
 GNžp5; PNmd1; PNme1,2; PNme4; PNme6;
 PNmp1,4; PNmp4,3; PNmp6; PNse4; PNsp2;
 PNsp4; PNže1; PNže4,3; PNžp1; PNžp2,2;
 PNžp4,4; PNžp6
 raztrgati · Gbe; Gce; Gce,2; Gcp,2; GLme,5;
 GLmp; GLse,3; GLže,4
 razumeti · Gae,35; Gbe,41; Gbp,6; Gcd,3;
 Gce,15; Gcp,6; GLmd,4; GLme,44; GLmp,19;
 GLže,15; GNE,15
 razumljiv · Pme1,3; Pmp2; Pse1,7; Pže1,2;
 Pže5
 razvalina · Šze1; Šzp1,4; Šzp2; Šzp5,3; Šzp6
 razveseliti · Gcd; Gce,5; GLme,23; GLmp;
 GLse; GLže,6; GLžp
 razvneti · GLme,8; GLmp; GLže,3
 rdeč · Pmeli,3; Pme1,7; Pme2; Pme4i; Pme4,4;
 Pme5,2; Pme6,4; Pmp1,7; Pmp2,3; Pmp4,5;
 Pmp6; Pse1,15; Pse2,22; Pse4,5; Pse5,3;
 Pse6,5; Psp1,4; Psp2,2; Psp4,2; Pže1,10; Pže2;
 Pže3; Pže4,13; Pže5,4; Pže6,8; Pžp1,11;
 Pžp4,6; Pžp6,4
 rebro · Ssp1; Ssp4,18; Ssp5
 rebula · Šze1; Šze2,9; Šze5
 reč · Šze1,7; Šze3; Šzp1; Šzp2,3; Šzp4,4; Šzp6,3
 reči · Gad,2; Gae,67; Gap,4; Gap,24; Gbe,6;
 Gcd; Gce,197; Gcp,8; GLmd,12; GLme;
 GLme,947; GLmp,100; GLse,21; GLže,425;
 GLžp,3; GNE,83; GVad; GVbe,19; GVbp,3
 red · Sme1,3; Sme2; Sme4,13; Sme5,21; Sme6
 redok · Pme5; Pme6; Pmp1,10; Pmp2,2;
 Pmp4,4; Pmp5,3; Pmp6,3; Pse1,3; Pse6; Pže1;
 Pže5; Pžp1,4; Pžp2; Pžp4,3; Pžp5; Pžp6,8
 rediti · Gae; Gap; Gbp; Gce,7; Gcp,2; GLmp,5;
 GLže,4; GLže,5; GLžp; GNE,4
 regljati · Gbp; Gce,6; Gcp,4; GLmd,2; GLže,2
 Rejc · IOme1,16; IOme3,3; IOme4,4
 Rejcov · PIOme2,2; PIOme5; PIOme6,2;
 PIOmp2; PIOmp3; PIOže1; PIOže2; PIOže4
 Rejčev · IOže3; PIOme1,4; PIOme2,3; PIOme4,6;
 PIOme5; PIOže1; PIOže2; PIOže4
 reka · Šze1,24; Šze2,36; Šze3,20; Šze4,40;
 Šze5,30; Šze6,4; Šzp1,4; Šzp2,4; Šzp4,2; Šzp6
 rekoč · GČ,32
 rep · Sme1; Sme2; Sme4,16; Sme5,2; Sme6,11
 repa · Šze1,22; Šze2,4; Šze4,3
 res · Č,580; PD,150
 resen · Pmd1; Pme1,8; Pme2; Pme4; Pme4j;
 Pme6,5; Pmp1,4; Pmp4; Pmp5; Pmp6,4;
 Pse1; Psp1; Pže1,2; Pžp1; Pžp4,2
 resnica · Šze1,27; Šze2,7; Šze3,3; Šze4,5;
 Šze5,23; Šze6; Šzp1; Šzp2
 resničen · Pmd1; Pme1; Pmeli,2; Pme2,2;
 Pmp1; Pmp4; Pse1,2; Pse6; Pže1,6; Pže2;
 Pže4,3; Pžp1,2
 resnično · A,28
 resno · A,86
 rešitev · Šze1,6; Šze2,12; Šze3,3; Šze4,11;
 Šze5,2; Šzp4
 rešiti · Gbe; Gce,6; Gcp; GLme,21; GLmp,2;
 GLže,6; GLžp; GNE,9
 reva · Šze1,13; Šze3; Šzp1
 reven · Pme1,3; Pmp1; Pmp2; Pže1,5
 revež · Smd1; Sme1,8; Sme2; Sme4; Sme6,2;
 Smp3,3; Smp4
 revolucija · Šze1,14; Šze2,2; Šze4,3; Šze5,2; Šze6
 revolver · Sme1,2; Sme2,4; Sme3; Sme4,9;
 Sme6,12; Smp4,2; Smp6,2
 rezati · Gce,7; Gcp,7; GLme,12; GLmp,7; GLse;
 GLže,5; GLžp,3; GNE,3
 rezek · Pme1; Pme2; Pme6; Pmp1; Pse1,2;
 Pse4,2; Pže1,4; Pže4
 rezko · A,16
 režati se · Gbe,2; Gbp,4; Gce,3; Gcp; Gcp;
 GLme,5; GLmp; GLže,2; GLžp
 riba · Šze1,4; Šzp1,2; Šzp2,2; Šzp3; Šzp4,5;
 Šzp5,2
 ribati · GLme,2; GLse; GNA,2; GNE,5
 Rim · IZme1,2; IZme2,2; IZme4,4; IZme5,7
 rimski · Pme4; Pmp4; Pse2,2; Pže1; Pže2,3;
 Pže3,2; Pže4; Pže5,5; Pžp5
 risati · Gce,2; Gcp,2; GLme,4; GLmp; GNE
 rjav · Pmd2; Pme2; Pme4,2; Pmp1; Pmp4;
 Pmp5; Pse4; Psp1; Psp6,2; Pže1,4; Pže2,3;
 Pže4,9; Pže5; Pžp1,2; Pžp2,2; Pžp4,2; Pžp5,2
 rjuha · Šze1,2; Šze2,2; Šze3; Šze4,11; Šze6,4;
 Šzp1,3; Šzp2; Šzp4,6; Šzp5,4; Šzp6,2
 rob · Sme1,2; Sme2,6; Sme3; Sme4,45; Sme5,40;
 Sme6; Smp5,6
 Robe · IZžp3; IZžp4,8; IZžp5,11; IZžp6,2
 robec · Sme2,2; Sme4,31; Sme6,10; Smp1;
 Smp6,2; Sse6,3
 ročica · Šze4,2; Šze6; Šzp3; Šzp4,5; Šzp6,6
 rod · Smd1; Sme1,11; Sme2; Sme2,17; Sme4,5;
 Sme5,2; Smp1; Smp2
 Rodica · IOže1,6; IZže2,9; IZže3,2
 roditi · Gbe; Gbp; Gce,15; Gcp,4; GLme,50;
 GLmp,3; GLse; GLže,23; GLžp; GNE,6

- rog · Smd1,2; Sme1,3; Sme4,9; Sme5,2; Sme6,4;
Smp1,2; Smp2; Smp4,2; Smp6; Smp6,5
- rogovilast · Pme1i,11; Pme2; Pme6; Pse1; Pse6;
Pžel
- rojstvo · Sse1,2; Sse2,11; Sse4,2; Sse6; Ssp1
- roka · Šzd1,4; Šzd2; Šzd4,12; Šzd6,68; Šze1,59;
Šze2,61; Šze3; Šze4,300; Šze5,43; Šze6,239;
Šzp1,61; Šzp2,112; Šzp4,2; Šzp4,3; Šzp4,376;
Šzp5,74; Šzp6,161
- rokav · Sme2,3; Sme4,20; Sme5,3; Sme6,12;
Smp1; Smp2,3; Smp4,9; Smp6
- roman · Sme1,6; Sme4,3; Sme5; Sme6; Smp1;
Smp2; Smp4,3
- romanje · SGse1; SGse2,2; SGse5,2; SGse6;
Sse1,2; Sse4,2
- romati · Gae; Gce,4; GLme,10; GLmp,3;
GLže,3; GNA
- ropotati · Gae; Gce,2; GLme; GLmp,3; GLse;
GLsp,2; GLže
- rosa · Šze1,4; Šze2,10; Šze4,2; Šze5
- rosen · Pme5; Pmp1; Pmp2; Pmp4; Pmp5; Pse1;
Pse5; Pžel,6; Pže2; Pže4; Pže5,7; Pžp1,4;
Pžp2,2; Pžp6
- Roza · IOže1,8; IOže2; IOže4,2
- roža · Šze1,3; Šze4,6; Šze5; Šzp1,13; Šzp2,4;
Šzp3; Šzp4,9; Šzp5; Šzp6,2
- rožica · Šze1,2; Šzp1,4; Šzp4,6
- rožnat · Pmd1; Pme2; Pme4i; Pme5; Pme6,2;
Pmp1; Pmp2; Pmp6; Pse1; Pse5; Psp1; Pžel;
Pže5; Pže6; Pžp4
- Rudi · IOme1,10; IOme4; IOme4
- rumen · Pme1; Pme1i; Pme6; Pmp1; Pmp4;
Pmp6; Pse1,3; Pse4; Pže1; Pže4; Pže6,2;
Pžp1; Pžp4,2
- Rupar · IOme1,10; IOme3
- Ruparjev · PIOme1,12; PIOme3,2; PIOme5;
PIOže1,2; PIOže2
- Rusepatacis · IOme1,20; IOme3; IOme4,3
- ruski · Pmp1,3; Pmp2,3; Pse2,2; Pse5,3; Pže5;
Pžp4
- ruta · Šze1; Šze2,6; Šze3; Šze4,35; Šze5,2;
Šze6,6; Šzp4,2; Šzp6,2
- s · E2,285; E6,2155; S
- sad · Sme1,4; Sme2,5; Sme4,8; Sme5,2; Smp1,2;
Smp2; Smp5
- saden · Pme2,3; Pme4i,2; Pme5; Pmp2; Pse1;
Pse2; Pse6,2; Psp2
- sadje · Sse1,3; Sse2,2; Sse4,4; Sse5,2; Sse6
- saj · Č,575; Vpr,570
- sajast · Pmd1,2; Pme1; Pme2; Pme3,2; Pme4i;
Pme5,2; Pmp1; Pmp4; Psp4; Pžel,2; Pže2,3;
Pže4,3; Pže5,2
- sam · ZPUmd1,12; ZPUme1,442; ZPUme2,15;
ZPUme3,11; ZPUme4,16; ZPUme4,21;
ZPUme5,13; ZPUme6,17; ZPUmp1,51;
ZPUmp2,2; ZPUmp4,3; ZPUse1,19;
ZPUse2,10; ZPUse4; ZPUse4; ZPUse4,9;
ZPUse6; ZPUsp1,3; ZPUsp2; ZPUžd1,2;
ZPUže1,165; ZPUže2,21; ZPUže3;
ZPUže4,13; ZPUže5,14; ZPUžp1,10;
ZPUžp2,4; ZPUžp4,6; ZPUžp5,3
- samo · Č,692; Sse5,7; Vpo,51; Vpr,345
- samokres · Sme4,13; Sme6,3
- samota · Šze1,13; Šze2,4; Šze4; Šze5,11; Šze6;
Šzp2; Šzp5,2
- samoten · Pme1i; Pme1,4; Pme2,3; Pme3;
Pme4; Pme4i,2; Pme5,2; Pmp1; Pmp2; Pmp5;
Pse1; Pžd1; Pžel,9; Pže2,3; Pže3; Pže4;
Pže5,3; Pžp1,2; Pžp5,2
- samozavestno · A,16
- sanja · Šzp2,29; Šzp3,3; Šzp5,32; Šzp6,5
- sanjati · Gbe,3; Gce,11; GLmd; GLme,48;
GLmp; GLse,4; GLže,2; GNE,9
- sanje · Šzp1,36; Šzp4,24
- sapa · Šze1,18; Šze2,19; Šze4,80; Šze5,8; Šze6,3
- Saplater · IOme1,27
- se · Gmp,13313
- sebe · ZOP2,174; ZOP3,6; ZOP3,137; ZOP4,16;
ZOP4,30; ZOP4,32; ZOP4,37; ZOP4,54;
ZOP4,55; ZOP5,78; ZOP6,12; ZOP6,130
- seči · Gae; Gbe; Gce,22; Gcp; GLme,45;
GLmp,4; GLže,12; GLžp; GNE,2
- sedaj · A,17
- sede · A,10
- sedem · ŠGmp5; ŠGžp5; ŠG1,10; ŠG4,26
- sedemdeset · ŠG1,2; ŠG4,9
- sedeti · Gad,3; Gae,7; Gbe,5; Gbp,4; Gcd,9;
Gce,56; Gcp,16; GLmd,13; GLme,165;
GLmp,24; GLse; GLže,19; GNE,9; GVap;
GVbe; GVbp,2
- sedlo · Sse1,2; Sse2,3; Sse3,4; Sse4,4; Sse5,2;
Sse6
- segati · Gce,3; Gcp,6; GLme,6; GLmp,4; GLsp,2;
GLžp,2; GNE
- sejati · Gce,6; Gcp,6; GLme,4; GLmp,4; GLže;
GNE
- sekati · Gce,4; Gcp,3; GLmd; GLme,2; GLmp,3;
GLse; GLžp; GNE
- sekira · Šze1,12; Šze2,9; Šze4,47; Šze6,17; Šzp1
- sekunda · Šze2,3; Šze4,4; Šze5,2; Šzp1; Šzp2;
Šzp4
- sem · ZK,49
- seme · Sse1,7; Sse5,2; Sse6; Ssp2
- semenj · Sme1,2; Sme2,3; Sme4,10; Sme5,6

- senca · Ssp4,3; Ssp5,2; Ssp5,7; Sžd1; Sže1,37;
Sže2,4; Sže4,19; Sže5,15; Sže6,3; Sžp1,10;
Sžp2; Sžp4,5; Sžp6,2
- senčen · Pse1,2; Pse2; Pže1,3; Pže2,3; Pže3;
Pže4,2; Pže5,2; Pže6
- senik · Sme2,2; Sme4,6; Sme5,3; Sme6
- seno · Sse2,9; Sse4,3; Sse5,6
- senožet · Sže1,4; Sže2,12; Sže3; Sže4; Sže5,7;
Sže6; Sžp1,4; Sžp2,2; Sžp4,4; Sžp5,3; Sžp6
- sesesti se · Gce,7; GLme,14; GLže; GLže,9
- sesti · Gae,8; Gce,12; Gcp,3; GLmd,17; GLmp,8;
GLsp; GLžd,2; GLže,39; GNE
- sestra · Sžd4; Sže1,22; Sže2,3; Sže3; Sže4,3;
Sže6,10; Sžp1,10; Sžp2; Sžp2; Sžp2; Sžp3;
Sžp4,2
- sesuti · Gce,4; GLme,5; GLmp,2; GLže,7; GLžp;
GNE
- seveda · Č,338
- sezidati · Gcd; GLme,7; GLmp; GLže
- sezuti · Gce; Gcp; GLme,6; GLže,4; GTmp1;
GVbe,2
- shod · Sme1,2; Sme4,2; Smp4; Smp5,5
- si · Gmp,1317
- sicer · Č,99; Vpr,136
- sijati · Gce,14; GLme; GLse,13; GLsp; GLže,2;
GLžp
- sikati · Gce,2; Gcp; GLmd,2; GLme,2; GLmp,4;
GLse; GLže,2; GLžp
- sikniti · Gce,4; GLme,12; GLže,3
- sila · A,4; Sže1,20; Sže2,11; Sže4,12; Sže5,3;
Sže6,42; Sžp1,9; Sžp2,9; Sžp4,9; Sžp5;
Sžp6,10
- silen · Pme1; Pme1j; Pme2j,2; Pme4jj; Pmp1,2;
Pmp2; Pmp4; Pmp5; Pmp6,2; Pse1j; Pse2;
Pže1,4; Pže2; Pže2jj; Pže4,3
- siliti · Gae,2; Gce,14; Gcp,6; GLme,15;
GLmp,13; GLse,2; GLže,11; GLžp; GNE,2
- silno · Aj; A,23
- Silvija · IOže1,12; IOže4,11
- sin · Smd1,4; Sme1,76; Sme2,5; Sme3,14; Sme4;
Sme4,40; Sme5; Sme6,3; Smp1,9; Smp2,3;
Smp3; Smp4,4; Smp5,2
- sinji · Pme2; Pme4; Pme6,2; Pse1,3; Pse2,4;
Pse5,3; Pžp1; Pžp2; Pžp4,2; Pžp5; Pžp6
- sinjina · Sže1,6; Sže2,3; Sže4; Sže5,6; Sže6
- sirena · Sže1,10; Sže2,2
- siromak · Smd1; Sme1,10; Sme3,3; Sme4,3;
Smp1,4; Smp2,2; Smp3,4; Smp4
- sirota · Sže1,21; Sže4; Sžp1,7; Sžp5
- sit · Pme1,3; Pmp1,5; Pže1,4
- siten · Pme1i; Pme1,2; Pme2; Pmp4; Pmp5;
Pmp6; Pse1; Pže1,5; Pže4; Pžp1,3
- sitno · A,15
- siv · Pme1i,2; Pme1,11; Pme2,4; Pme3,2;
Pme4i,5; Pme5,6; Pme6,3; Pmp1,11;
Pmp2,10; Pmp4,6; Pmp5; Pmp6,5; Pse1,7;
Pse2,3; Pse3; Pse4,8; Pse5,2; Pse6; Psp4;
Psp6; Pžd1; Pže1,18; Pže2,3; Pže3,2; Pže4,11;
Pže5,12; Pže6; Pžp1,6; Pžp2,9; Pžp4,13;
Pžp6,16
- skakati · Gap; Gce,5; Gcp,6; GLme,12; GLmp,3;
GLže,6; GNE,4; GVbe
- skala · Sže1,9; Sže2,22; Sže3,4; Sže4,23;
Sže5,19; Sže6,4; Sžp1; Sžp2,6; Sžp4,19;
Sžp5,12; Sžp6,14
- skalnat · Pme1i; Pme2; Pme4; Pme4i,2;
Pme6,2; Pmp1; Pmp5; Pse1; Pse2,2; Psp4;
Pže1,2; Pžp2,4; Pžp4
- skedenj · Sme1,2; Sme2,8; Sme4,10; Sme5,3;
Sme6,3; Smp5
- skesano · A,11
- sklanjati · Gce,8; Gcp,6; GLme,6; GLže,3;
GLžp,4
- skleda · Sžd4; Sže1,3; Sže2,2; Sže3,3; Sže4,16;
Sže5,8; Sže6,6; Sžp1; Sžp4,2; Sžp6,2
- sklednik · Sme1,2; Sme2,2; Sme3; Sme4;
Sme5,3; Sme6
- skleniti · Gae; Gce,10; Gcp,3; GLme,48;
GLmp,6; GLže,26; GNse1
- sklenjen · PNmp1; PNžp4,3; PNžp6,7
- sklep · Smd4; Sme2,2; Sme4,5; Sme6,2; Smp5,3
- sklepati · Gae; Gce; GLme,4; GLže,3; GNE,3
- skloniti · Gae; Gce,30; Gcp,5; GLme,51;
GLmp,4; GLže,20; GLžp; GNE; GVbp
- sklonjen · GNme1; PNže1; PNže3; PNže6,10;
PNžp1; PNžp2; PNžp4; PNžp6
- skobacati se · Gce; GLmd; GLme,9
- skočiti · Gae; Gbe,4; Gcd,4; Gce,25; Gcp,2;
GLmd; GLme,81; GLmp,11; GLže,30; GNE,8
- skop · Pme1,2; Pme2; Pme5,2; Pmp1; Pže1,4;
Pžp2
- Skopičnik · IZme1,3; IZme2; IZme3,7; IZme4,2;
IZme5,5; IZme6,2
- skopneti · Gce,2; GLme,2; GLmp; GLže,7
- skoraj · A,424
- skoro · A,15
- skozi · A,12; E4,289
- skratka · A,20
- skrb · Sže1,6; Sže2,26; Sže6,4; Sžp1,4; Sžp2,4;
Sžp4,4; Sžp5,3
- skrbeti · Gae,2; Gap; Gce,8; GLmd; GLme,7;
GLmp,2; GLže,14; GNE,4
- skrbno · A,10

skrinja · Sže2,5; Sže3,2; Sže4,19; Sže5,12;
 Šzp4,4; Šzp5
 skrit · GTme1,2; GTse1; PTme4i; PTme6;
 PTmp2; PTse1,4; PTse2; PTse4; PTsp1;
 PTže1,2; PTže4,2; PTžp2; PTžp4; PTžp5,2
 skriti · Gae; Gap; Gbe; Gbp; Gce,5; GLme,14;
 GLmp,2; GLže,9; GNE,7; GVbd; GVbe
 skrivaj · A,24
 skrivati · Gae; Gap; Gbe; Gce,16; Gcp,6;
 GLmd,2; GLme,15; GLse,3; GLže,5; GLžp,3;
 GNE,3
 skrivnost · Pmd1; Pme1; Pmeli; Pmelj;
 Pme2; Pme4i; Pme6,3; Pse1,2; Pse2,4; Pse4;
 Pse5; Pže1j; Pže1,5; Pže2j
 skrivnostno · A,13
 skrušeno · A,11
 skuhati · Gae,4; Gce,4; GLme,10; GLmp;
 GLže,10; GNE,2; GVbe; GVbp,2
 skupaj · A,44
 skušati · Gae; Gcd; Gce,14; GLmd; GLme,49;
 GLmp,3; GLže,5
 slab · Pme1j,2; Pme1,20; Pme2; Pme4; Pme5;
 Pme5jj; Pmp1,5; Pmp2jj; Pmp4,2; Pmp5;
 Pse1j,2; Pse1,7; Pse2,7; Pse6; Pže1,10;
 Pže2,5; Pže4,3; Pže6; Pžp1,2; Pžp4; Pžp5,2
 slabo · Aj,4; A,30; Sse3
 slabost · Šze1,8; Šze2; Šze4,2; Šze5; Šzp2;
 Šzp4,2
 slaboten · Pme1,5; Pme2; Pme4i; Pmp1; Pse1;
 Pse2; Pse4; Pže1,5; Pžp2
 sladek · Pme1i,2; Pme1,2; Pme2; Pme4; Pme4i;
 Pme6; Pmp1,2; Pmp3; Pmp4; Pmp5; Pse1,2;
 Pse6; Pžd1; Pže1j; Pže1,6; Pže2,2; Pže4,3;
 Pže5,3; Pže6; Pžp2,3; Pžp6; Šzp4,5
 sladkor · Sme1,5; Sme2,9; Sme4,10; Sme6
 slama · Šze1; Šze2,4; Šze4,7; Šze5,2; Šze6,6
 slamnat · Pže1,2; Pže2; Pže4,3; Pže6,2; Pžp1;
 Pžp5
 slap · Sme1,3; Sme2; Sme3; Sme4,2; Sme5;
 Sme6,2; Smp1; Smp2; Smp6
 sleči · Gae; Gce,3; Gcp,2; GLme,11; GLmp;
 GLže,3; GNE; GVbe
 sled · Sme2,4; Smp1,4; Šze1,4; Šze2; Šze4,2;
 Šze5,2; Šzp4; Šzp5
 slediti · Gce; GLme,6; GLže,2; GLžp
 sleheren · Pme1i,3; Pme4i,8; Pse2
 sleherni · Pme2,4; Pme5; Pse4; Pže2,8; Pže4,2
 slej · A,19
 slika · Šze1,18; Šze2,2; Šze4,12; Šze5,4; Šze6;
 Šzp1,4; Šzp2,2; Šzp4,5; Šzp5,4
 slina · Šze1,7; Šze2,4; Šze4,33; Šze6; Šzp4,3

slišati · Gae,20; Gap; Gbe,16; Gbp,8; Gce,26;
 Gcp,8; GLmd,3; GLme,132; GLmp,40; GLse,2;
 GLže,31; GNE,35
 slišno · A,17
 sloneti · Gae; Gcd; Gce,6; GLmd; GLme,8;
 GLžd; GLže,5
 Slovenec · IPme1,8; IPmp1,5; IPmp2,2; IPmp4;
 IPmp6
 slovenski · Pme1,3; Pme2,4; Pme3; Pme5;
 Pme6; Pmp2,4; Pmp4,2; Pže1,2; Pže2,7;
 Pže4,7; Pže5; Pžp4,5; Pžp5,2
 slovesen · Pme1,5; Pme3; Pme4i; Pme6,2;
 Pse1,4; Pse2; Pže1,3; Pže2; Pže4,2; Pžp1
 slovesno · A,24
 slutiti · Gcd; Gce; Gcp; GLme,6; GLže,4
 slutnja · Šze1,5; Šze2; Šze4; Šze5,2; Šze6,3;
 Šzp2; Šzp4
 služba · Šze1,5; Šze2,5; Šze4,9; Šze5,4
 služiti · Gbe; Gce,17; Gcp,2; GLme,7; GLse;
 GLže,11; GNA; GNE,5
 smeh · Sme1,18; Sme2,13; Sme3; Sme4,24;
 Sme5,7; Sme6,23
 smehljaj · Sme1,10; Sme2; Sme4,3; Sme6,8
 smehljati se · Gcd; Gce,2; Gcp; GLme,12;
 GLže,7
 smejati se · Gae,5; Gap; Gap,2; Gbe,3; Gbp,2;
 Gcd; Gce,15; Gcp,9; GLmd,2; GLme,12;
 GLmp,15; GLse,2; GLže,19; GLžp; GVbe;
 GVbp
 smer · Šze4,9; Šze5,3
 smešen · Pme1; Pmp1,2; Pse1,4; Pse2,2; Pže2;
 Pžp1
 smet · Šze1,2; Šze4; Šzp1,2; Šzp2,2; Šzp4,4;
 Šzp5
 smeti · Gae,19; Gap,5; Gbe,7; Gcd; Gce,28;
 Gcp,12; GLme,7; GLmp,5; GLse; GLže,2
 smiliti se · Gbe,2; Gce,3; GLme,11; GLže,2;
 GNE
 smisel · Sme2,2; Sme4,5; Sme5,4
 Smodin · IOme1,7; IOme4,3
 smola · Šze1,3; Šze2; Šze4,6; Šze5,5
 smolnat · Pme6; Pmp1; Pmp4; Pmp5; Pse5;
 Pže2; Pže3; Pže4; Pžp1,2
 smreka · Šze1,3; Šze2,5; Šze4,4; Šzp1,2; Šzp2,7;
 Šzp4,2; Šzp6,3
 smrkavec · Smd1; Sme1,5; Sme3; Sme4,2; Sme5;
 Smp1,2; Smp2; Smp3,2; Smp6
 smrt · Šze1,126; Šze2,42; Šze3,10; Šze4,50;
 Šze5,34; Šze6,16; Šzp2,2; Šzp5
 smrten · Pme5,2; Pmp5; Pže1,5; Pže2; Pže5,5;
 Pže6,2

- Smukač · IOme1,2; IOme2,2; IOme4; IOme5;
IŽme1,6
- snež · Sme1,36; Sme2,12; Sme4,25; Sme5,19;
Sme6
- snežiti · GLme,2; GLme,31
- sneti · Gae,11; Gbe,2; Gce,9; Gcp; GLmd;
GLmp,4; GLže,8; GNE,2
- snežiti · Gce,8; GLse,4
- snop · Sme1,2; Sme4,3; Sme5,3; Sme6; Smp1;
Smp2; Smp4,4; Smp5,2
- soba · Sže1,13; Sže2,35; Sže3; Sže4,62; Sže5,62;
Sžp1; Sžp2; Sžp4,5; Sžp5; Sžp6
- soča · ISže1,6; IZže1,5; IZže2,6; IZže3,2;
IZže4,3; IZže5,11; IZže6,2
- sočen · Pme1,2; Pmp4; Pmp5; Pse1,2; Pse2,4;
Pse4,2; Psp1,2; Pže4,2; Pže6; Pžp1,7; Pžp2
- sod · Sme1,5; Sme2,3; Sme3; Sme4; Sme5,2;
Smp3,2; Smp4,3
- soden · Pme1,3; Pme2,4; Pme4,5; Pme6
- soditi · Gae; Gce,5; Gcp,3; GLme,3; GLmp;
GLse; GLže,2; GLžp
- sodnik · Sme1,10; Sme2,3; Sme3,6; Sme4,3;
Sme6; Smp1,5; Smp3; Smp4; Smp6
- sodrga · Sže1,12; Sže2,3; Sže4; Sže6
- soha · Sžd4; Sže1,4; Sže3; Sže4,2; Sže5; Sžp4
- sok · Sme1,5; Sme2,7; Sme4,3; Sme5,2
- solza · Sžd1,13; Sžd4; Sže1,5; Sže2,2; Sže4,4;
Sže6; Sžp1,29; Sžp2,2; Sžp2,4; Sžp2,5; Sžp3;
Sžp4,32; Sžp5,3; Sžp6,17
- solzan · Pme4; Pže2; Pžp3; Pžp4,3; Pžp6,4
- sonce · Sse1,143; Sse2,44; Sse3,9; Sse4,24;
Sse5,76; Sse6,6; Ssp4,2
- sončen · A; Pme1; Pme1,3; Pme2,6; Pme3;
Pme4,2; Pme5,4; Pme6,5; Pmp1,7; Pmp2,4;
Pmp4; Pmp5; Pse2; Psp4; Pže1,4; Pže2,12;
Pže4,4; Pže5,4; Pže6,2; Pžp2; Pžp4,2; Pžp5
- sončnica · Sže1,5; Sže2,4; Sže4,22; Sže6,2; Sžp1;
Sžp2
- sosed · Smd1,2; Sme1,7; Sme2,3; Sme3; Sme4;
Sme5; Smp1,4; Smp3; Sže3,2
- soseda · Sže1,6; Sže4; Sže5; Sžp1,2; Sžp6
- soteska · Sže1,2; Sže2,2; Sže4,4; Sže5,3; Sžp5;
Sžp6
- Sova · IOme1,272; IOme2,8; IOme3,24;
IOme4,17; IOme5,2; IOme6,5
- sovrašstvo · Sse1; Sse2,3; Sse4,2; Sse5,2; Sse6,5;
Ssp4
- sovražen · Pme1,3; Pme6,4; Pmp1,2; Pmp6,2;
Pse5; Pže1,3
- sovražiti · Gae; Gce,5; Gcp,2; GLme,7; GLmp;
GLsp
- sovražnik · Sme1,5; Sme2,2; Sme4,2; Sme6;
Smp2
- spanje · SGse4,8; Sse2,5; Sse5,5
- spati · Gae,2; Gap; Gbe,5; Gcd; Gce,19; Gcp;
Gcp,2; GLme,30; GLmp,11; GLže,4; GNA,29;
GNE,6; GVbe,3
- spenjati · Gce; Gcp; GLme,7; GLže,2; GNE,2;
GVbp
- spet · A,857
- splaziti se · Gae,4; Gce; GLme,7
- splezati · Gae,2; GLme,10; GLmp; GLže
- sploh · A,135
- spočetka · A,40
- spodnji · Pme2; Pme4; Pme5,2; Pse4,2; Pse5;
Pže2; Pže4,3; Pže5,2; Pže6; Pžp1,2; Pžp5,4
- spodobiti se · Gce,25; GLse,3
- spogledati se · Gbp; Gcd,9; Gce,5; Gcp,8;
GLmd,15; GLmp,12; GLžp,2
- spomenik · Sme1,7; Sme2; Sme4,7; Sme5,2;
Smp1,3
- spomin · Sme1,29; Sme2,12; Sme4,22; Sme5,19;
Sme6,2; Smp1,12; Smp2,6; Smp3; Smp4,8;
Smp5,3
- spominjati se · Gae,20; Gap; Gbe,11; Gbp;
Gce,4; Gcp,2; GLmd; GLme,13; GLmp; GLse;
GLže,6; GLžp; GNE
- spomladi · A,17
- spomniti se · Gae,10; Gap; Gbe; Gce,6; GLmd,2;
GLme,100; GLmp,4; GLse,3; GLže,37; GLžp;
GNE,9; GVbp
- spoštovanje · Sse2,3; Sse4,3; Sse6,6
- spotakniti · Gce,4; GLme,8; GLže
- spoved · Sže2,3; Sže3,5; Sže5,4
- spoznanje · SGse1,7; SGse2,7; SGse4,2;
SGse5,2; SGsp1,3; SGsp2; SGsp4,2; Sse1;
Sse4,10
- spoznati · Gae,2; Gbe; Gce,3; Gcp; GLmd;
GLme,30; GLmp,3; GLse; GLže,13; GVap
- spraševati · Gae,2; Gbp; Gce; Gcp; GLme,17;
GLmp,7; GLže,5; GNE,2
- spraviti · Gae,6; Gbe,3; Gce,18; Gcp,4; GLmd,2;
GLme,50; GLmp,21; GLse,2; GLžd; GLže,20;
GLžp; GNE,12
- spravlјati · Gae,2; Gbe; Gce,8; Gcp,3; GLme,2;
GLmp,2
- spregovoriti · Gad; Gce,4; GLmd; GLme,10;
GLže,2; GNE,6
- sprehajati se · Gce,5; Gcp,2; GLme,4; GLže,2;
GNE
- sprehod · Sme3; Sme4,8; Smp4
- sprejeti · Gce,7; GLme,12; GLmp,5; GLse,2;
GLže,3; GLžp

spreleteti se · Gce; GLme,16; GLse,16; GLže
 spremeniti · Gce,8; Gcp,2; GLme,11; GLmp;
 GLse,3; GLže,14; GNE
 spremenjen · GNme1,6; GNse1,2; GNže1,2;
 PNme1,2; PNme5; PNme6,4; PNže1
 spremljati · Gcd; Gce,2; Gcp; GLmd,2; GLme,6;
 GLže,4
 sprevideti · GLmd,3; GLme,13; GLmp,2; GVbe
 spricho · A,7; E2,11
 sproščeno · A,11
 sprožiti · Gce,8; GLme,9; GLmp; GLže,2
 spustiti · Gae,3; Gbe; Gcd; Gce,39; Gcp,5;
 GLmd,5; GLme,100; GLmp,17; GLže,2;
 GLže,28; GLžp; GVbp
 spuščati · Gae; Gcd; Gce,9; Gcp,4; GLmd,3;
 GLme,12; GLmp,2; GLse,2; GLže,2; GNE,9
 srajca · Sže1,14; Sže2,12; Sže3,2; Sže4,35;
 Sže5,13; Sžp2; Sžp4,2; Sžp5,2; Sžp6
 sram · PD,42; Sme1,2; Sme2,4
 sramežljivo · A,11
 sramota · Sže1,4; Sže2,6; Sže4,6; Sže6,3
 srbeti · Gce,5; Gcp; GLme; GLse,2; GLžp,2;
 GNE
 srce · Sse1,100; Sse2,97; Sse3,11; Sse4,71;
 Sse5,107; Sse6,15; Ssp1,4; Ssp2,6; Ssp4,4;
 Ssp5,9
 srebrn · Pme1,2; Pme4i,2; Pme5,3; Pme6;
 Pmp1,2; Pmp5; Pzd1; Pže1,4; Pže2; Pže5;
 Pžp2; Pžp4; Pžp6,2
 sreča · IOže1,5; IOže2,2; IOže3; IOže4,5;
 IOže6,3; ISže1
 sreča · Sže1,52; Sže2,35; Sže3,2; Sže4; Sže4,23;
 Sže5,13; Sže6,6; Sžp4
 srečanje · SGse1,5; SGse2,3; SGse4; SGse5,3;
 SGsp2; SGsp6
 srečati · Gae,4; Gap,4; Gbe; Gcd; Gce,4; Gcp,2;
 Gcp,3; GLmd,11; GLme,27; GLmp,9; GLže,8;
 GLžp; GNE,3
 srečen · Pmd1,2; Pme1jj; Pme1,19; Pme2,2;
 Pme4; Pme4,3; Pme5,2; Pme6,2; Pmp1,6;
 Pmp2jj; Pmp5; Pse1,3; Pse4,2; Pse5; Pse6;
 Psp1; Pže1,12; Pže4,3; Pže5,2; Pže6; Pžp1,2
 Srečko · IOme1,15; IOme2,2; IOme3
 srečno · A,57
 sreda · Sže2,5; Sže4,10; Sže5,21
 sredi · E2,186
 srednji · Pme1i; Pme2,3; Pme4; Pme4i; Pme5,2;
 Psp2,6; Psp5,2; Pže1,3; Pže2,5; Pže4; Pže5,2
 sredstvo · Sse1; Sse5; Ssp1; Ssp2,2; Ssp4,4;
 Ssp6
 srh · Sme1,12; Sme4
 srkniti · Gbe; Gce; Gcp; GLme,5; GLže; GVad

Stane · IOme1,164; IOme2,4; IOme3,14;
 IOme4,21; IOme6,2
 stanovati · Gae; Gce,5; Gcp; GLme,4; GLmp,8;
 GLse; GLže,2; GNE
 star · Pmd1jj; Pmd1j,2; Pmd1,8; Pmeli,117;
 Pmel1jj,6; Pmel1j,17; Pmel1,90; Pme2,36;
 Pme3jj; Pme3,13; Pme4i,3; Pme4jj; Pme4,12;
 Pme4,19; Pme5,13; Pme6,18; Pmp1j,11;
 Pmp1,25; Pmp2,10; Pmp3; Pmp4,12; Pmp5,7;
 Pmp6,2; Psd1; Pse1,8; Pse2,2; Pse3; Pse4,8;
 Pse5; Pse6,4; Psp1,2; Psp2; Psp2j; Psp4,13;
 Pzd1j; Pzd1,3; Pzd4,2; Pže1jj,2; Pže1j,4;
 Pže1,108; Pže2,39; Pže3; Pže3,5; Pže4,44;
 Pže5,30; Pže6,14; Pžp1,25; Pžp2j; Pžp2,10;
 Pžp3; Pžp4,28; Pžp5,4; Pžp6,3; Sže5; Sžp2jj
 stara · Sže1,11; Sže2,3; Sže3
 starec · Smd1; Sme1,95; Sme2,7; Sme3,6;
 Sme4,7; Sme5; Sme6,5; Smp1,14; Smp2,3;
 Smp4,2; Smp5
 starejši · Pme3,2; Pmp2,2; Pmp4; Pže6,3; Sme1;
 Smp1
 starinski · Pme1,3; Pme6,2; Pmp6; Pse6;
 Pže1,3; Pže2,2; Pže3; Pže4; Pže5,3; Pže6,2;
 Pžp1; Pžp4
 starka · Sže1,19; Sže3,3
 stati · Gae,17; Gap; Gbe,7; Gcd,4; Gce; Gce,105;
 Gcp,9; Gcp,10; GLmd,8; GLme,141; GLmp,38;
 GLse; GLsp; GLzd,2; GLže,67; GLžp,12;
 GNE,2; GVbe,14; GVbp,5
 stavek · Sme1,3; Sme2,7; Sme4,2; Smp2; Smp4;
 Smp5
 steber · Sme1,3; Sme2,4; Sme3,2; Sme4,8; Sme6;
 Smp1,4; Smp2; Smp4,4
 steči · Gcd,2; Gce,5; GLme,27; GLmp,2; GLse;
 GLže,17; GLžp
 stegniti · Gae,4; Gbe; Gce,24; Gcp; GLme,53;
 GLže,16; GLžp; GNE,2
 stegno · Sse2,2; Sse5,6; Ssp5,3
 stegovati · Gae,4; Gcd; Gcp,2; GLme; GLže;
 GNE
 steklen · Pme1i; Pme2; Pme6; Pmp4,3; Pse1;
 Psp2,3; Psp4; Psp6; Pžp1,7; Pžp2; Pžp6,3
 steklenica · Sže1,2; Sže4,19; Sže5,2; Sže6;
 Sžp2,2; Sžp4,3; Sžp6,2
 steklo · Sse1; Sse2,16; Sse4,4; Sse5,3; Sse6,2;
 Ssp1,3; Ssp4,3
 stena · Szd6; Sže1,9; Sže2,31; Sže3,19; Sže4,49;
 Sže5,41; Sže6,3; Sžp1,15; Sžp2,12; Sžp4,14;
 Sžp5,13; Sžp6,7
 stenski · Pme3; Pže1,8; Pže2,3; Pže4,3; Pže5;
 Pžp4

- steza · Sže1,8; Sže2,7; Sže4,10; Sže5; Sže5,44;
Sže6,2; Sžp1; Sžp2; Sžp2,3; Sžp4; Sžp5,9
- stikati · Gce,2; Gcp,2; GLme,6; GLžp; GNA
- stiskati · Gap,2; Gcd; Gce,22; Gcp,10; GLmd;
GLme,13; GLmp,5; GLse,4; GLže,2; GLžp;
GNE,2
- stisniti · Gad; Gae,2; Gbe; Gcd; Gce,59; Gcp,6;
GLmd,6; GLme,80; GLmp,16; GLse,12;
GLže,34; GLžp,3; GNE,2; GVbe,2
- stisnjen · GNme1; GNžp1,4; PNme1; PNmp4,3;
PNse6; PNsp1; PNže4; PNže6; PNžp1;
PNžp4; PNžp6,3
- stlačiti · Gce; Gcp; GLmd,2; GLme,9; GLmp;
GLže; GNE
- sto · ŠG1,6; ŠG4,27
- stoje · A,12
- stokati · Gbe; Gce,4; Gcp,2; GLmd; GLme,14;
GLmp; GLžd; GLže,5; GNE,2; GVbe,6
- stol · Smd4; Sme1,2; Sme2,4; Sme4,15; Sme5,5;
Sme6; Smp1; Smp2,2; Smp4,4; Smp5; Smp6
- stoletje · Sse2,6; Sse4; Ssp1; Ssp4,4; Ssp5
- stopati · Gae,2; Gce,6; Gcp,10; GLmd,3;
GLme,10; GLmp,3; GLse; GLsp,2; GLže,3;
GLžp
- stopiti · Gad,2; Gae,8; Gap,4; Gbe,3; Gbp,5;
Gcd,6; Gce,111; Gcp,5; GLmd,26; GLme,360;
GLmp,43; GLse; GLžd; GLže,120; GLžp,2;
GNE,13; GVbe,11; GVbp
- stopnica · Sže1; Sže2,2; Sže4,4; Sže5; Sžp1,6;
Sžp2,12; Sžp3; Sžp4,6; Sžp5,50; Sžp6,4
- stopnišče · Sse1; Sse2,3; Sse3,2; Sse4,3; Sse5,12;
Sse6; Ssp1
- storiti · Gae,5; Gce,5; Gcp,4; GLmd; GLme,22;
GLmp; GLže,6; GNE,4
- stradati · Gce; Gcp,3; GLme,2; GLmp,3; GLžp;
GNE
- strah · Sme1,34; Sme2,30; Sme4,6; Sme5,17;
Sme6,17
- strahoten · Pme2; Pme4; Pme5,2; Pme6; Pmp1;
Pmp5; Pse1,2; Pse4; Pže1,4; Pže4; Pžp5
- stran · A,4; Sžd2; Sžd4,5; Sžd5,4; Sžel,3;
Sže2,29; Sže4,32; Sže5,59; Sžp2,10; Sžp4,10;
Sžp5,6
- strast · Sžel,2; Sže4,3; Sže5; Sže6,4; Sžp2; Sžp4
- strastno · A,11
- strašen · Pme1i; Pme1,2; Pme2; Pme4i,4;
Pme5; Pmp1,2; Pmp5,2; Psd1; Pse1,6; Pse2,5;
Psp2,2; Pže1j; Pže1,20; Pže2,2; Pže4; Pže5;
Pžp1,2; Pžp2; Pžp4
- strašiti · Gae; Gbe,2; GLme,9; GLmp; GNE
- strašno · A,25
- straža · Sžel,5; Sže2,4; Sže5,4; Sžp1
- stražmojster · Sme1,43; Sme2; Sme3; Sme4,3
- stražnik · Smd1; Sme1,12; Sme6,2; Smp1,6;
Smp3,2; Smp6
- streči · Gbe; Gce,3; Gcp; GLme,7; GLmp;
GLže,2
- streha · Sže1,17; Sže2,22; Sže3; Sže4,29;
Sže5,10; Sže6,17; Sžp1,4; Sžp2,2; Sžp4,4;
Sžp5,9; Sžp6,2
- strel · Sme1,10; Sme2,4; Sme4,5; Smp1,7;
Smp2,3
- strela · Sže1,14; Sže2,3; Sže4; Sže5; Sže6;
Sžp1,2; Sžp4
- streljati · Gcd; Gce,10; Gcp,5; GLme,7; GLmp,4;
GNE,6
- stresati · Gcd; Gce,10; GLmd,2; GLme,14;
GLmp,2; GLsp,2; GLže,4; GNE,11
- stresti · Gae,2; Gcd; Gce,26; Gcp,2; GLme,69;
GLže,19
- Strežek · IOme1,20; IOme3,2; IOme4,2; IOme5
- Strežkov · PIOme1; PIOmp3; PIOmp5;
PIOže1,5; PIOže2,2; PIOže4
- strgati · Gbe; GLme,21; GLže; GNE,7; GVbe,2
- stric · Sme1,179; Sme2,5; Sme3,16; Sme4,20;
Sme5; Sme6
- strm · Pme2; Pme5,8; Pmp2,3; Pmp4; Pmp5;
Pse1,4; Pse2; Pse3; Pse4; Pse5,12; Psp1,2;
Pže1,8; Pže4; Pže5,10; Pže6; Pžp1; Pžp2,4;
Pžp5,4
- Strmar · IOme1,10; IOme3; IOme4
- strmeti · Gae,3; Gap; Gbe,2; Gcd,3; Gce,44;
Gcp,14; GLme,66; GLmp,9; GLže,12; GLžp,3
- strmina · Sže1; Sže4,2; Sže5,9; Sžp4; Sžp5,3
- strog · Pme1,3; Pme4i; Pme6,2; Pmp1; Pse1;
Pže1,2
- strogo · Aj; Aj,4; A,49
- stroj · Sme1,3; Sme2; Sme4,10; Sme5; Smp4;
Smp5
- strojnica · Sžel,14; Sže2,4; Sže3,6; Sže4,16;
Sže5,3; Sže6,9; Sžp1,10; Sžp2; Sžp4,2; Sžp6
- strop · Sme1,6; Sme2,8; Sme3,3; Sme4,18;
Sme5,7; Sme6,11; Smp1; Smp2,2
- struna · Sžel,5; Sžp1,3; Sžp4,3; Sžp5
- strup · Sme1,7; Sme4,3
- strupen · Pme1; Pme2; Pme4i; Pme5; Pme6,3;
Pse1; Pže1,2; Pže6,2; Pžp1; Pžp4
- strupeno · A,15
- stuliti se · Gce; GLme,8; GLmp; GLse,2; GLže
- stvar · Sžd1; Sžd5; Sžel,66; Sže2,2; Sže4,12;
Sže5,13; Sže6,2; Sžp1,24; Sžp2,15; Sžp4,56;
Sžp5,19; Sžp6,8

- suh · Pmd1,4; Pmeli,2; Pme1,19; Pme2,6;
Pme3; Pme4i,5; Pme4,2; Pme5,7; Pme6,3;
Pmp1,3; Pmp2,2; Pmp4,4; Pse1,4; Pse4,3;
Pse6,3; Psp2; Psp5; Psp6; Pže1,8; Pže2,3;
Pže3,2; Pže4,9; Pže5,3; Pže6; Pžp1,2; Pžp2,4;
Pžp4,8; Pžp5,2; Pžp6
- sukati · Gcd; Gce,8; Gcp,3; GLme,9; GLžp
- suknja · Šze2,4; Šze4,3; Šze5,2; Šze6,3; Šzp4
- suknjič · Sme2,9; Sme4,25; Sme5,2; Smp4,2;
Smp5
- sunek · Sme1,3; Sme5,2; Sme6,9; Smp1; Smp5,6
- suniti · Gce,13; Gcp; GLme,17; GLmp; GLse;
GLže,4; GNE,2
- sunkovito · A,84
- surov · Pme1i; Pme1,2; Pme2,2; Pme6,2; Pmp1;
Pmp4; Pse1; Pse6; Pžp2,2
- sveča · Šze1,5; Šze4,2; Šze5; Šzp1,2; Šzp4,2;
Šzp6,3
- svet · Pmeli; Pmeli,21; Pme2,3; Pme3;
Pme4i,24; Pme4,4; Pme5,4; Pmp5; Pse1,2;
Pse2,2; Pse4,2; Pse5; Psp1,3; Pže1,13; Pže2,4;
Pže4,2; Pže5,4; Pže6,4; Pžp1,2; Pžp4,2;
Pžp5,2; Sme1,50; Sme2,56; Sme3,5; Sme4,61;
Sme5; Sme5,52; Sme6,5; Smp4; Smp5; Smp5;
Smp6
- svetel · Pme1,17; Pme2,4; Pme4; Pme4i,2;
Pme4,3; Pme5,4; Pme6,7; Pmp1j; Pmp1,9;
Pmp2,4; Pmp4,9; Pmp5,2; Pmp6,2; Pse1,20;
Pse2,4; Pse4; Pse5,2; Pse6; Psp1; Psp4;
Pžd1,2; Pže1,29; Pže2,5; Pže4,8; Pže5,4;
Pže6,2; Pžp1,12; Pžp2,4; Pžp4,16; Pžp5,3;
Pžp6,7
- svetiti · Gce,6; Gcp,6; GLmd; GLme,9; GLmp;
GLse,3; GLže,9; GLžp,5
- svetlikati se · Gce,8; Gcp; GLme,3; GLmp,2;
GLse; GLže,3; GLžp
- svetlo · A,15
- svetloba · Šze1; Šze2,4; Šze3,2; Šze5; Šze6,4
- svetnik · Sme1; Sme2,2; Sme4,3; Smp1; Smp2;
Smp3,2; Smp4; Smp6
- svetoven · Pme1,2; Pme1,2; Pme2; Pse2; Pžd1;
Pže1,3; Pže2,5; Pže5; Pže6,4
- svez · Pme1; Pme2,3; Pme4i; Pmp5; Pse2;
Pže2,3; Pže4; Pže5,3; Pžp2,4
- sveže · A,10
- svinčen · Pme2; Pme6; Pse1,2; Pse2; Pse4,2;
Pse6; Psp1; Pžp1,2
- svinčnik · Sme1; Sme4,7; Sme6,2; Smp4
- svinja · Šze1,16
- svišli · Šzp4,5; Šzp5,5
- svoboda · Šze1,6; Šze2,3; Šze4,14; Šze5,5; Šze6
- svoboden · Pme1; Pme2; Pme3; Pse2; Pže1,3;
Pže3; Pže5,2; Pžp2,2
- svoj · ZSVPmd4,3; ZSVPme1,10; ZSVPme2,76;
ZSVPme3,30; ZSVPme4,62; ZSVPme4,222;
ZSVPme5,60; ZSVPme6,82; ZSVPmp1,2;
ZSVPmp2,50; ZSVPmp3,25; ZSVPmp4,98;
ZSVPmp5,23; ZSVPmp6,31; ZSVPme1,2;
ZSVPse2,63; ZSVPse3,8; ZSVPse4,145;
ZSVPse5,37; ZSVPse6,35; ZSVPsp2,10;
ZSVPsp3; ZSVPsp4,17; ZSVPsp5,14;
ZSVPsp6,9; ZSVPžd3,2; ZSVPžd4,7;
ZSVPžd6; ZSVPže1; ZSVPže2; ZSVPže2,157;
ZSVPže3,26; ZSVPže4,3; ZSVPže4,391;
ZSVPže5,134; ZSVPže6,72; ZSVPžp2,34;
ZSVPžp3,9; ZSVPžp4; ZSVPžp4,162;
ZSVPžp5,35; ZSVPžp6,73
- šala · Šze1,12; Šze2,5; Šze4,5; Šze5; Šzp4,5
- ščepec · Sme4,10; Smp4
- še · Č,2411
- šele · Č,176
- šepetati · Gce,2; Gce,3; Gcp; Gcp; GLme,4;
GLmp; GLže,3
- šest · ŠGmp3; ŠGmp5,2; ŠGmp6,2; ŠGžp2;
ŠGžp5,4; ŠG1,8; ŠG4,27
- šesti · ŠVme2; ŠVme3; ŠVže1; ŠVže4,4; ŠVže5
- šiba · Šze1,14; Šze2; Šze4; Šze6; Šzp4
- šibek · Pme1; Pme4j,2; Pme6,2; Pmp1,2; Pmp2;
Pse6; Psp2; Pže1,6; Pže2; Pže4; Pžp6
- šipa · Šze1; Šze4; Šze5; Šzp1,5; Šzp4,4; Šzp6
- širiti · Gce,7; Gcp,6; GLme,6; GLmp; GLse,2;
GLže,3; GLžp,2
- širok · Pmeli,2; Pme1j; Pme1,13; Pme2,14;
Pme3; Pme4; Pme4i,3; Pme4,6; Pme5,14;
Pme6,7; Pmp1,4; Pmp2,2; Pmp4,5; Pmp5,2;
Pmp6,7; Pse1,3; Pse2,5; Pse4,6; Pse5,4;
Psp1,6; Psp2,4; Psp3; Psp4,7; Psp5,2; Psp6,3;
Pže1j,2; Pže1,31; Pže2j; Pže2,2; Pže3,3;
Pže4,15; Pže5,17; Pže6,12; Pžp1,8; Pžp2,5;
Pžp4,6; Pžp5,5; Pžp6,6
- široko · Ajj; A,134; Smp4
- šivati · Gce,3; Gcp; GLme,2; GLže,4; GNE
- škaf · Sme1,2; Sme2,3; Sme3; Sme4,22; Sme5;
Sme6,8; Smp4,2
- škatlá · Šze2,2; Šze4,8; Šze5; Šze6,2; Šzp4,3
- škoda · Šze1,20; Šze2,2
- škorenj · Sme2; Sme4,5; Sme5; Sme6,7; Smp1,4;
Smp2,3; Smp5,6; Smp6,2
- škripati · Gce,2; Gcp,5; GLme,2; GLmp,4;
GLsp,2; GLžp,2; GNE
- šmarnica · Šzp1,5; Šzp2,8; Šzp4,11; Šzp6,2

- Šmonov · PIOME1,14; PIOME2,2; PIOME4;
PIOmp6
- šola · Sze1,4; Sze2,14; Sze3; Sze4,45; Sze5,20;
Sze6,4; Szp2,3; Szp4,9
- šolski · Pme2; Pme4; Pmp6; Pse2; Pse3; Pse4;
Psp4; Psp5; Pze1; Pze2,2; Pze3; Pze4; Pze5;
Pzp2; Pzp4
- šop · Sme1; Sme4,16; Sme6; Smp5
- šopek · Smd4; Sme1,5; Sme4,5; Sme6,4; Smp4,2
- štab · Sme1,4; Sme2,2; Sme4,2; Sme5; Sme6,2
- štefe · IOme1,28; IOme2; IOme3,3; IOme4,2
- štefuc · IOme1,39; IOme3,5; IOme4
- šteti · Gae; Gap,2; Gce,3; GLmd; GLme,3;
GLmp,5; GLze,3; GNE,7; GVbe,2
- številka · Sze1,2; Sze2,12; Sze4,4; Sze5; Sze6;
Szp2,2; Szp4; Szp5
- štiri · ŠGsp1,4; ŠGsp4,15; ŠGsp5,5; ŠGsp6,3;
ŠGžp1,13; ŠGžp2; ŠGžp4,16; ŠGžp5,8;
ŠGžp6,5; ŠG1; ŠG4
- štirideset · ŠGsp5,3; ŠGsp6,5; ŠG1,3; ŠG4,13
- štirinajst · ŠGmp5; ŠG1,3; ŠG4,14
- štirje · Smp6; ŠGmp1,33; ŠGmp2,2; ŠGmp3,3;
ŠGmp4,26; ŠGmp5; ŠGmp6,4
- štor · Sme1,3; Sme3; Sme4,3; Sme6; Smp5;
Smp6
- štrleti · Gce; Gcp; GLmd,3; GLmp,2; GLžd;
GLze,4; GLžp
- štrukelj · Sme1; Smp1,3; Smp2,2; Smp4,7;
Smp5,2
- študent · Smd4; Sme1,17; Sme2,2; Sme3,2;
Sme4,2; Smp1
- šum · Sme1,5; Sme2,12; Sme4,8; Sme5; Sme6,3;
Smp4
- šumeti · Gce,14; Gcp,2; Gcp,2; GLme,2; GLmp;
GLse,9; GLsp; GLze,12; GLžp
- švigati · Gce,5; Gcp,7; GLmd,2; GLme,7;
GLmp,6; GLsp; GLze,2; GLžp,3; GNE,2
- švigniti · Gae,2; Gcd,2; Gce,5; Gcp,3; GLmd;
GLmd; GLme,26; GLmp,3; GLse,2; GLze,28;
GLžp; GVbe,2
- ta · ZKmd1,11; ZKmd4; ZKmd5; ZKme1,261;
ZKme2,59; ZKme3,15; ZKme4,24; ZKme4,69;
ZKmp2,22; ZKmp5,19; ZKsd5; ZKse1,1062;
ZKse2,256; ZKse3,50; ZKse4,335; ZKsp1,7;
ZKsp2,7; ZKsp4,3; ZKsp5,5; ZKze1,185;
ZKze2,81; ZKze3,5; ZKze4; ZKze4,2;
ZKze4,110; ZKze5,117; ZKze6,21; ZKžp2,18;
ZKžp5,35; ZTžp2,15
- tak · M,28; ZKmd1; ZKme1,50; ZKme2,11;
ZKme3; ZKme4,2; ZKme4,5; ZKme5,3;
ZKme6,11; ZKmp1,20; ZKmp2,10; ZKmp3,3;
ZKmp4,2; ZKmp5,9; ZKmp6,2; ZKse2,39;
ZKse4,15; ZKse6,4; ZKsp1,3; ZKsp6;
ZKze1,63; ZKze2,17; ZKze3,3; ZKze4,3;
ZKze5,2; ZKžp1,14; ZKžp2,12; ZKžp3,2;
ZKžp4,28; ZKžp5,9; ZKžp6,6; ZK,11
- taka · ZKze4,24; ZKze6,28
- takle · ZKme1,6; ZKme4; ZKmp1,2; ZKmp2;
ZKmp4; ZKze1,2
- tako · A,1580; Vpo,115; Vpr,222; ZKse1,166;
ZK,5
- takoj · A,457
- takole · ZKse1,4; ZK,35
- takrat · A,261
- takšen · ZKme1,22; ZKme2,5; ZKme4; ZKme4;
ZKme6,3; ZKmp1,7; ZKmp2,3; ZKmp4,3;
ZKmp5,2; ZKse1,4; ZKse2; ZKse5,2; ZKse6;
ZKsp5; ZKze1,18; ZKze2,4; ZKze4; ZKžp1,5;
ZKžp2; ZKžp4,4; ZKžp6
- tale · ZKme1,7; ZKme4; ZKme4,2; ZKme4,3;
ZKmp3; ZKmp5; ZKse1; ZKse4; ZKse5,2;
ZKse6,3; ZKze5; ZKžp4
- tam · ZK,520
- tanek · Pme1,2; Pmp1j; Pmp1,3; Pmp4;
Pse2jj,2; Pse2,2; Pžd4; Pžp4
- tank · Sme1,7; Sme2,4; Sme3; Sme4; Smp1;
Smp2,2; Smp4; Smp6,5
- Tantadruj · IOme1,172; IOme3,5; IOme4,3;
IOme5,4
- tat · Sme1,13; Sme2; Sme4,3; Smp1; Smp1,3
- tatvina · Sze2,8; Sze4,5; Sze6,3; Szp2; Szp4;
Szp5
- ta-a-ta-an · M,12
- te · M,21; ZKžp1,64; ZKžp3,3; ZKžp4,49;
ZKžp6,12
- teči · Gap; Gbe,4; Gcd,2; Gce,4; Gce,46; Gcp,5;
GLme,16; GLmp,4; GLse,7; GLsp,7; GLze,36;
GLžp,2; GNE,3; GVbe,20
- tedaj · A,164
- teden · Sme1,5; Sme2; Sme4,49; Sme5,2;
Sme6,5; Smp2,2; Smp4,6; Smp5,3; Smp6
- tele · Sse1,2; Sse4,6; Ssp1,3; Ssp4,2
- telo · Sse1,46; Sse2,23; Sse3,2; Sse4,55; Sse5,35;
Sse6,21; Ssp1,3; Ssp2,2; Ssp4; Ssp5
- tem · Vpo,12; ZK,7
- tema · Sze1,14; Sze2; Sze2,13; Sze3; Sze4;
Sze4,17; Sze5,24; Sze6,2; Szp5
- temačen · Pme2,2; Pme4i; Pme5,2; Pse1; Pze1;
Pze4,5; Pze5,3

- teman · Pme1j; Pme1,8; Pme2,2; Pme4i;
Pme5,2; Pme6j; Pmp1,7; Pmp2,14; Pmp4,2;
Pmp5,3; Pmp6,3; Pse2; Pse4,2; Pse5; Pse6;
Psp4; Psp5,3; Pže1j; Pže1,12; Pže2,5; Pže5,5;
Pžp1,2; Pžp2; Pžp4,4; Pžp5,3; Pžp6,2
- temeljito · A,12
- temen · Pme4; Pme6; Pse1,5; Pže4,7; Pže6
- Temnik · IZme1; IZme2; IZme4,6; IZme5,5;
IZme6
- Temnikar · IOme1,185; IOme2,2; IOme3,5;
IOme4,21; IOme5,13; IOme6,10
- Temnikarica · IOže1,102; IOže2; IOže3,10;
IOže5,3; IOže6,2; IZže4,16
- Temnikarjev · PIOme1,6; PIOme2; PIOme3;
PIOme4,2; PIOme6,2; PIOmp1,3; PIOse2;
PIOže1; PIOže2,2; PIOže3,2; PIOže4,5;
PIOže6; PIOžp1,3; PIOžp4; PIOžp5
- temno · A,32
- temnozelen · Pme1,2; Pme4i; Pme6; Pmp5;
Pže1; Pže4; Pže5,3
- temveč · Vpr,101
- tenak · Pme1,9; Pme4
- tenek · Pme1,4; Pme4i; Pme5; Pme6,2; Pmp1;
Pmp2,2; Pmp4; Pse4; Pse6; Psp1; Psp6; Pžd1;
Pže1,5; Pže2; Pže4; Pže5,2; Pže6,2; Pžp1,7;
Pžp2,2; Pžp4,5; Pžp5; Pžp6
- tepka · Šze1,13; Šze2,12; Šze4,24; Šze5,9;
Šze6,11; Šžp1
- tepsti · Gae; Gap; Gcd; Gce,4; GLmd; GLme,8;
GLmp,11; GLse; GLže,2; GNE,4; GVbe
- ter · Vpr,852
- teran · Sme1,3; Sme2,11; Sme4,10; Sme5
- tesar · Sme1,17; Sme3; Sme4
- tesno · Aj,12; A,14
- tesnoba · Šze1,10; Šze2,4; Šze3,2; Šze4,6; Šze5;
Šze6,6
- Testen · IOme1,29; IOme2,2; IOme3,2; IOme4,5;
IOme5; IOme6
- teta · Šze1,230; Šze2,4; Šze3,8; Šze4,12; Šze5,3;
Šze6,10; Šžp1; Šžp2; Šžp3
- teža · Šze1,2; Šze2,5; Šze4,4; Šze6,10
- težak · Pme1,16; Pme4,4; Smp1,4
- težava · Šze1; Šze6,23; Šžp2; Šžp4,3; Šžp5,3;
Šžp6
- težek · Pme1,2; Pme1j; Pme1j,4; Pme1,3;
Pme2,2; Pme4; Pme4,3; Pme5,4; Pme6,10;
Pmp1j; Pmp1,12; Pmp2,4; Pmp4,9; Pmp5,7;
Pmp6,14; Psd5; Pse1j,3; Pse1,14; Pse2,4;
Pse4,14; Pse5,3; Pse6,7; Psp1,7; Psp2,2;
Psp4j; Psp4,8; Psp6,4; Pžd4; Pže1j,3;
Pže1,20; Pže2,6; Pže4,27; Pže5; Pže5,3;
Pže6,4; Pžp1,9; Pžp2,5; Pžp4j,2; Pžp4,15;
Pžp5j; Pžp5,3; Pžp6,7
- težiti · Gce,8; GLme,2; GLse; GLže,2; GNE,3
- težko · Aj,3; A,32; Č,6
- ti · ZKmp1,78; ZKmp3,5; ZKmp4,42; ZKmp6,4;
ZKžd1,7; ZKžd4,2; ZKžd4,2; ZKžd6,2;
ZObd3,2; ZObd4; ZObd6; ZObe1,501;
ZObe2,20; ZObe2,36; ZObe3,37; ZObe3,259;
ZObe4; ZObe4; ZObe4; ZObe4,2; ZObe4,7;
ZObe4,14; ZObe4,29; ZObe4,191; ZObe5,9;
ZObe6; ZObe6,36; ZObsd,14; ZObp1,64;
ZObp3,104; ZObp6,8
- tičati · Gce,4; Gcp; GLme,7; GLže; GNE,2
- tih · Pme1j; Pme1,12; Pme4; Pme5,3; Pme6,7;
Pmp1,2; Pse1,21; Pse2; Pse4; Psp1; Pže1j;
Pže1,18; Pže2; Pže4,3; Pže5,4; Pže6
- Tihi dol · IZme1,10; IZme2,3; IZme4,11;
IZme5,5; IZme6
- tihó · Aj; Aj,10; A,169; Sse5,18
- tik · A,36; E2
- tikati se · Gce,35; GLse
- tiktakati · Gce,7; GLže,5; GNE
- Tilčka · IOže1,56; IOže2,2; IOže3,6; IOže4;
IOže5,3; IOže6,6
- Tildica · IOže1,117; IOže2,5; IOže3,7; IOže4,24;
IOže5,6; IOže6,2
- Tilka · IZže1,11; IZže6,2
- tilnik · Sme1; Sme2; Sme4,24; Sme5,5; Sme6,2;
Smp1; Smp4
- Tina · IOme1,3; IOže1,8; IOže2,3; IOže4
- Tinč · IOme1,15; IOme2; IOme3; IOme4; IOme6
- Tinka · IOže1,42; IOže2,4; IOže3,3; IOže4,8;
IOže5,2; IOže6,6
- tisoč · Smp1; Smp2,2; Smp4,6; ŠG; ŠGme1,2;
ŠGme4,2; ŠG1,3; ŠG2; ŠG4,13; ŠNE
- tisoče · A,5; ŠNE,10

- tisti · ZKmd1,2; ZKmd4,3; ZKme1,78;
 ZKme2,43; ZKme3,11; ZKme4,11; ZKme4,62;
 ZKme5,47; ZKme6,20; ZKmp1,22;
 ZKmp2,16; ZKmp3,6; ZKmp4,23; ZKmp5,10;
 ZKmp6,5; ZKse1,23; ZKse3,2; ZKse4,30;
 ZKse5,11; ZKse6,3; ZKsp1; ZKsp2,2; ZKsp4;
 ZKsp5,3; ZKžd1; ZKžd5; ZKže1,63;
 ZKže2,40; ZKže3,5; ZKže4,41; ZKže5,31;
 ZKže6,8; ZKžp1,20; ZKžp2,10; ZKžp3,2;
 ZKžp4,15; ZKžp5,4; ZKžp6,3
- tisto · ZKse2,13
- tiščati · Gae; Gbe; Gbp; Gcd; Gce,9; Gcp,6;
 GLme,27; GLmp,2; GLse,4; GLže,6; GLžp;
 GNE; GVbe
- tišina · Šze1,92; Šze2,19; Šze3; Šze4,33;
 Šze5,13; Šze6,2
- titovka · Šze4,9; Šžp4
- tja · ZK,73
- tla · Ssp2,30; Ssp5,37; Šze3,26
- tlačan · Sme4,2; Smp1,7; Smp2,4; Smp3,2;
 Smp4,4
- tlačiti · GLme,5; GLmp; GLse; GLže,4
- tlak · Sme1; Sme2; Sme4,5; Sme5,8
- tlakovan · GNžp1; PNme5; PNse4; PNže1;
 PNže2,6; PNže4; PNže5,2; PNže6; Pže1
- tlo · Ssp1,8; Ssp3; Ssp4,141
- tnalo · Sse1; Sse2,3; Sse4,22; Sse5,23; Sse6
- to · ZKme5,111; ZKme6,28; ZKse5,170;
 ZKse6,93
- tobak · Sme1,5; Sme2,30; Sme4,14; Sme5,3;
 Sme6
- tod · ZK,14
- toda · Vpr,1127
- tok · Sme1,4; Sme3,2; Sme5,4
- tolažiti · Gae; Gbe; Gce,4; Gcp; GLme,7; GLmp;
 GLže
- tolči · Gce,23; GLme,7; GLmp; GLse; GLže,5;
 GNE,6
- toliko · ŠNE,47; Vpo,16; ZK,58
- Tolmin · IZme2,8; IZme4,19; IZme5,15
- Tolminec · IPme1,4; IPmp1,6; IPmp2; IPmp5
- tolminski · Pme1,9; Pme2,2; Pme4,3; Pmp1;
 Pmp2; Pmp3; Pmp4,3; Pmp5,2; Pse1; Pže1,5;
 Pže2,5; Pže4,2; Pže5,4; Pžp1,3; Pžp2,3; Pžp4;
 Pžp5,2
- Tolminsko · IZse2; IZse4,3; IZse5,10
- tolmun · Sme1,3; Sme2,12; Sme3,2; Sme4,24;
 Sme5,17; Sme6,3; Smp4
- Tomaž · IOme1,20; IOme2; IOme5
- Tone · IOme1,93; IOme3,7; IOme4,7; IOme5;
 IOme6,3
- top · M,26; Pme1,4; Pme2; Pme4; Pme6,3;
 Pmp1; Smd1; Sme1; Sme2; Smp1,6
- tóp · M,12
- topel · Pmd1,2; Pme1i; Pme1,5; Pme2; Pme5,2;
 Pme6,8; Pmp1,2; Pmp5; Pmp6; Pse1,8;
 Pse4,3; Pže1,13; Pže2,3; Pže3; Pže4,6; Pže5,2;
 Pže6,2; Pžp1; Pžp2; Pžp6
- toplo · Aj,2; A,25; Sse5,2
- topol · Sme1,3; Sme3; Sme4; Sme5,2; Smp1,9;
 Smp2,6; Smp3; Smp4,11; Smp6,2; Šze4; Šžp5
- torba · Šze2,2; Šze4,7; Šze6
- torej · Vpr,78
- tovariš · Sme1,22; Sme4,2; Smp1,8; Smp2;
 Smp3,2; Smp4; Smp5; Smp6,3
- tovornjak · Smd1; Sme1,2; Sme2; Sme3,3;
 Sme4,3; Sme5,3; Sme6; Smp2
- tožiti · Gce; Gcp,2; GLme,4; GLže,2; GLžp;
 GNA
- trajati · Gce,2; GLme,2; GLse,7; GLžp; GNE
- trak · Sme1; Sme4,3; Sme5,3; Sme6; Smp1;
 Smp6
- tram · Sme1,5; Sme2,5; Sme3,2; Sme4,8;
 Sme5,2; Sme6,3; Smp1,3; Smp2; Smp4,3;
 Smp5
- trapa · Šze1,26; Šžp1
- trapest · Pme1; Pme4; Pse6; Pže1,6; Pžp1
- trava · Šze1,34; Šze2,19; Šze4,24; Šze5,33;
 Šze6,3; Šžp2
- travnik · Sme1,16; Sme2,10; Sme3,2; Sme4,9;
 Sme5,27; Sme6; Smp1,3; Smp2; Smp4,15;
 Smp5,14; Smp6,3; Šžp2,3
- Travnikarica · IOže1,17; IOže3,3; IOže4,3;
 IOže6
- Travnikarjev · PIOme1,64; PIOme2,2;
 PIOme3,11; PIOme4,10; PIOme5; PIOme6;
 PIOže5
- trd · Pme1,24; Pme2,4; Pme3; Pme4; Pme4i,2;
 Pme5,3; Pme6,6; Pmp1j; Pmp1,11; Pmp2,3;
 Pmp4; Pmp6,7; Pse1,4; Pse2,6; Pse3,2;
 Pse4,3; Pse5,2; Psp4,2; Pže1,13; Pže2j;
 Pže2jj; Pže2,8; Pže3; Pže4,7; Pže5,12; Pže6;
 Pžp1,8; Pžp4,3; Pžp6,2
- trden · Pme1,17; Pme4; Pme6,3; Pmp4; Pmp6;
 Pse1,2; Pse4; Psp2; Psp5,2; Pže1,5; Pže2j;
 Pže4,2; Pže6,2; Pžp1j; Pžp4; Pžp6; Sse2
- trditi · Gae,3; Gce,2; GLme,11; GLmp,6;
 GLže,7; GLžp,3; GNE,7
- trdno · Aj; A,32
- trdo · A,112
- treba · PD,236
- trebuh · Sme1,5; Sme2,6; Sme4,18; Sme5,6;
 Sme6,5; Smp4,5; Smp5,2; Smp6,4

- treniti · Gce,3; GLme,8; GLže,2
 trenutek · Sme1,19; Sme2,15; Sme3,3; Sme4,26;
 Sme5,67; Smp1,4; Smp2,25; Smp4; Smp5,8;
 Smp6
 trepetajoč · GČse6; PČme3; PČme5; PČmp4,2;
 PČze1; PČze4; PČze5,2; PČzp2; PČzp6
 trepetati · Gce,3; Gce,6; Gcp; GLme,17; GLmp;
 GLse,3; GLže,9; GLžp,3; GNE
 treskati · Gcp,4; GLmp; GLse; GLže; GLžp,2;
 GNE
 tresoč · GČme1; PČme6,3; PČmp6; PČze5;
 PČze6,6; PČzp4,2; PČzp5; PČzp6
 tresti · Gae,7; Gbe,9; Gce,10; Gcp,7; GLme,18;
 GLmp,3; GLse; GLsp,4; GLže,10; GLžp,10;
 GNE,3; GVbe,2
 treščiti · Gce,9; GLme,14; GLmp,2; GLse,8;
 GLže,3
 tretji · ŠVme1,13; ŠVme2,3; ŠVme3; ŠVme4;
 ŠVme4,7; ŠVme5,2; ŠVmp1,3; ŠVse2;
 ŠVse4,3; ŠVze1,6; ŠVze2; ŠVze3,2; ŠVze4,2;
 ŠVze6; ŠVzp1
 tretjič · A,14
 trg · Sme1; Sme2,8; Sme4,11; Sme5,17; Smp6
 trgati · Gad; Gce,2; Gcp; Gcp,2; GLmd,4;
 GLme,6; GLmp,3; GLse,2; GLže; GLžp,2;
 GNE
 trgovec · Sme1,4; Sme2,2; Sme4; Sme6; Smp1,3;
 Smp2
 trgovina · Šzd4; Šze1,4; Šze3; Šze4,8; Šze5,6;
 Šzp1,2; Šzp5
 trhel · Pme2,2; Pme5,2; Psp4,2; Pže1,3; Pže4,2;
 Pže5; Pžp1,2; Pžp2,2; Pžp4,2; Pžp5; Pžp6
 tri · ŠGsp1,4; ŠGsp4,9; ŠGsp5,4; ŠGsp6,4;
 ŠGžp1,27; ŠGžp2,6; ŠGžp3,2; ŠGžp4,56;
 ŠGžp5,6; ŠGžp6,7; ŠG1,2; ŠG4,5
 trideset · ŠGsp5,2; ŠGsp6,2; ŠG1,3; ŠG4,18
 trije · ŠGmp1,88; ŠGmp2,5; ŠGmp3,5;
 ŠGmp4,90; ŠGmp5,8; ŠGmp6,18
 trikrat · A,52
 trma · Šze1,5; Šze2,3; Šze4; Šze6,6
 trmast · Pme1,2; Pme1,6; Pme2; Pme4; Pme6;
 Pmp1,3; Pse1; Pse4; Pže1,3; Pže4
 trmasto · A,23
 Trnar · IOme1,33; IOme3; IOme4,3; IOme6
 trobenta · Šze1,58; Šze2,17; Šze4,50; Šze5,3;
 Šze6,2; Šzp4; Šzp5,3
 trobiti · Gad; Gae; Gbd; Gbe; Gcd; Gce,5;
 GLmd; GLme,4; GLmp,2; GVbp
 Trohovka · IOže1,36; IOže2,4; IOže3,3; IOže5
 trpek · Pme1,3; Pme4; Pme5; Pse6; Pže1; Pže4;
 Pže5; Pže6,2; Pžp1,2; Pžp4; Pžp5
 trpeti · Gap; Gce,8; Gcp; GLmd,2; GLme,8;
 GLmp; GLse; GLže,9; GNE,10
 trpljenje · SGse1,7; SGse2,13; SGse3,2;
 SGse4,9; SGse5,8; SGse6,3; SGsp4
 Trst · IZme1,10; IZme2,12; IZme3; IZme4,30;
 IZme5,28; IZme6
 tršat · Pme1,3; Pme1,5; Pme4i; Pme5; Pse4;
 Pse6,2; Pže2
 trta · Šzd4; Šze1,7; Šze2,2; Šze3,2; Šze5; Šzp2;
 Šzp4,2
 trud · Sme1,8; Sme2,3; Sme6
 truden · Pmd1; Pme1,5; Pme6; Pmp1; Pmp6,2;
 Pže1,6; Pže4,2; Pžp1,3; Pžp4; Pžp5; Pžp6,2
 truditi se · Gae,2; Gce,3; GLme,6
 trzati · Gce,4; Gcp,3; GLme,4; GNE,2
 trzniti · Gce,12; Gcp; GLmd,2; GLme,36;
 GLmp; GLže,9
 tu · ZK,163
 tudi · Č,1201; Vpr,598
 tuj · Pme1,6; Pme2; Pme4,2; Pme5,5; Pme6,3;
 Pmp1,6; Pmp2,3; Pmp4,2; Pmp5; Pmp6,2;
 Pse1,2; Pse2,3; Pse5; Psp4,2; Psp6; Pžd1,2;
 Pže1,6; Pže2; Pže4,3; Pže5; Pže6; Pžp1,3;
 Pžp2,3; Pžp3,2; Pžp4,2; Pžp5
 tujec · Sme1,5; Sme2,3; Sme4,4; Sme6,2; Smp1;
 Smp3; Smp4; Smp6
 tukaj · A,96; ZK,28
 Tul · IOme1,18; IOme4
 tuliti · Gcd; Gcp,2; GLme,6; GLmp; GLže,2;
 GNE; GVbe,6
 tvoj · ZSVbeme1,26; ZSVbeme4; ZSVbeme4,4;
 ZSVbeme5,2; ZSVbeme6,2; ZSVbemp1,6;
 ZSVbemp2,2; ZSVbemp4,2; ZSVbese1,6;
 ZSVbese2,2; ZSVbese4,4; ZSVbesp1,7;
 ZSVbesp4,3; ZSVbesp5,2; ZSVbeže1,40;
 ZSVbeže2,5; ZSVbeže3; ZSVbeže4,6;
 ZSVbeže5,2; ZSVbeže6; ZSVbežp1,7
 ubiti · Gae,2; Gbe; Gce,2; Gcp,2; GLme,21;
 GLmp,19; GLse,2; GLže; GLžp; GVbe; Smp2
 ubog · Pme1; Pme1,3; Pme4; Pmp1,3; Pmp3;
 Pmp4; Pse4; Pže1,7; Pže2; Pže3; Pžp4
 ubogati · Gce; GLme,7; GLmp; GLže,3; GNE,3;
 GVbe
 ubrati · Gcd; Gce; GLme,10; GLmp,3; GLže,2;
 GNE,2; GVbe
 učitelj · Sme1,18; Sme2,2; Sme3,2; Sme4,3;
 Sme5; Sme6,2; Smp4
 učiteljica · Šze1,29; Šze2; Šze3,3; Šze5; Šze6
 učiti · Gae; Gbe,2; Gbp; Gce,6; GLme,29;
 GLmp,13; GLže,4; GNE,5; GVbe,5

- udarec · Sme1,18; Sme2,4; Sme3; Sme4,2;
Sme5,6; Sme6,4; Smp1,13; Smp2,11; Smp4,11;
Smp5; Smp6,6
- udariti · Gad; Gae,2; Gce,43; Gcp; GLmd,2;
GLme,91; GLmp,9; GLse,10; GLže,39; GLžp;
GNE,2; GVap; GVbp
- udrihati · Gcd; Gce,12; Gcp,3; GLme,9;
GLmp,2; GLsp; GNE,5; GVbe,2
- uganjati · Gae; Gbp; Gce,2; Gcp; GLme,5;
GLmp; GLže,2; GNE,2; GVbe
- ugasniti · Gce,3; GLme,12; GLse,3; GLže,6;
GLžp
- ugibati · Gce; GLme,5; GLmp,2; GLžp; GNE
- ugnati · GLmd; GLme,3; GLmp,7; GNE,5
- ugodje · Sse1,3; Sse2,7; Sse5; Sse6
- ugotoviti · Gce; GLme,8; GLmp; GLže,2;
GLžp,2
- ugovarjati · GLme,13; GLže,7
- uh · M,18
- uho · Sse1,7; Sse2,2; Sse4,27; Sse5; Sse6,3;
Ssp1,6; Ssp2,6; Ssp3; Ssp4; Ssp4,30; Ssp5,11;
Ssp6,4
- uiti · GLme,7; GLmp,2; GLse,3; GLže,10;
GLžp
- ujeti · Gce,9; Gcp,2; GLmd; GLme,27; GLmp,3;
GLse,2; GLže,11; GLžp
- ujetnik · Sme5; Smp1,5; Smp2,6; Smp4,2; Smp6
- ukaz · Sme1,5; Sme2; Sme3; Sme4,21; Smp2,2;
Smp4
- ukazati · Gae,2; Gce,20; Gcp,2; GLme,38;
GLmp,6; GLže,12; GVbe
- ukazujoče · A,15
- ukovan · GNmd1; GNmp1; GNžp1; PNme1,4;
PNme4; PNžp5; PNžp6
- ukrasti · Gce,2; Gcp; GLmd,2; GLme,3; GLmp;
GLže
- ukvarjati se · Gae,5; Gbe; Gce,4; Gcp; GLme,9;
GLmp; GNE
- ulica · Sže1,2; Sže4,10; Sže5,10; Sžp1; Sžp2,2;
Sžp4; Sžp5,14
- um · Sme2,2; Sme4,14
- umakniti se · Gce,15; Gcp,3; GLmd,2;
GLme,32; GLmp,8; GLže,19; GLžp,2; GNE,8;
GVbe; GVbp,2
- umazan · GNmd1; GNme1,6; GNse1,2;
GNže1,5; GNžp1,2; Pme1; PNmd1; PNmeli;
PNme2,2; PNme4; PNme4,2; PNme5,6;
PNme6; PNmp1,3; PNmp2,2; PNmp4,2;
PNmp5; PNmp6; PNse1; PNse2,2; PNse4,3;
PNse5,2; PNsp5; PNsp6; PNže1,11; PNže2,4;
PNže4,5; PNže5,4; PNže6,2; PNžp1,3;
PNžp2,4; PNžp4,5; PNžp5,2; PNžp6; Pžp1
- umetnik · Sme1,2; Sme2; Sme3; Smp1,3; Smp2;
Smp3; Smp5; Smp6
- umetnost · Sže1,4; Sže2,5; Sže3; Sže5
- umikati se · Gcd; Gce,10; Gcp,5; GLme,5;
GLmp; GLse; GLže,2; GLžp; GNE,4; GVbp
- umirati · Gap; Gce,7; Gcp,3; GLme,8; GLmp,5;
GLse,4; GLsp; GLže,3; GNE,3
- umit · PTme5; PTmp1; PTmp4,2; PTse1,3;
PTsp1; PTžel1; PTžp6
- umiti · Gce,3; Gcp; GLmd,2; GLme; GLme,16;
GLmp; GLže,2; GNE,3; GVbe
- umolkniti · Gcp,2; GLmd,2; GLme,13; GLmp,4;
GLže,11
- umreti · Gae; Gae,2; Gap; Gbe,7; Gcd; Gce;
Gce,13; Gcp; GLme,134; GLmp,4; GLse,5;
GLže,89; GLžp,3; GNE,17
- uničiti · Gce,2; GLme,8; GLmp; GLže; GLžp;
GNE
- uniforma · Smp5; Sže1,2; Sže2; Sže4,3; Sže5,14;
Sžp5,2
- upanje · Sse1,18; Sse2,7; Sse3; Sse4,4; Sse5,2;
Sse6,12
- upati · Gae,14; Gap,2; Gbe,5; Gce,5; Gcp,3;
GLme,19; GLmp,2; GLže,8; GNE
- upirati se · Gae,2; Gce,6; Gcp,2; GLme,16;
GLmp,3; GLse; GLže; GLže,3; GNE,4;
GVbe,2; GVbp
- upognjen · GNme1,3; GNže1,2; GNžp1; PNme1;
PNme2,2; PNme4; PNme4i; PNme5;
PNme6,2; PNmp1; PNmp6; PNse1; PNsp6;
PNže2; PNže4,2; PNže5; PNže6; PNžp1;
PNžp4
- upor · Sme1,6; Sme2; Sme3; Sme4,3; Sme5;
Sme6; Smp5,2
- upornik · Sme1,7; Smp1,2; Smp3; Smp4
- upreti · Gce,11; Gcp; GLme,40; GLmp,3;
GLže,15; GNE,5; GTžp1; GTžp6
- ura · Sžd1,2; Sžd4,4; Sže1,69; Sže2,35; Sže3,3;
Sže4,52; Sže5,34; Sže6,2; Sžp1,4; Sžp2,10;
Sžp4,44; Sžp5,14
- uraden · Pme6,3; Pže1; Pže4,2; Pže6; Pžp2;
Pžp4; Pžp5
- uradno · A,17
- urediti · Gap; Gce; Gcp; GLmd; GLme,8;
GLmp; GLse; GLže,2; GNE
- urejen · GNse1,8; GNže1; PNme1; PNsp1;
PNže1
- urezati · Gce,2; GLmd; GLme,5; GVbe,5
- Usadarica · IOže1,11
- Usadarjev · PIOme1,10; PIOme3; PIOme5,2;
PIOže1,2; PIOže2,3; PIOže3; PIOže5

- usekniti se · Gce; GLme,9; GLže; GLže,8; GNE;
GVbe
- usmajati · Gce,2; Gcp; GLme,8; GLmp; GNE
- usmiliti se · Gce,2; GLme,3; GLmp; GLže,2;
GVbp,6
- usmiljenje · Sse1,2; Sse2,11; Sse4; Sse5; Sže5
usoda · Šzd6; Sže1,2,3; Sže2,7; Sže3,5; Sže4,16;
Sže5,9; Sže6,6; Sžp1
- usoden · Pme2; Pmp5,2; Pse2; Pse4; Pže1,3;
Pže2; Pže5,2; Pžp1
- usojen · GNse1,10
- usta · Ssp1,15; Ssp2,5,3; Ssp3,5; Ssp4,10,4;
Ssp5,18; Ssp6,9
- ustaviti · Gae,4; Gcd,7; Gce,35; Gcp,4;
GLmd,10; GLme,9,4; GLmp,14; GLse,2;
GLže,3,2; GLžp,2; GNE
- Ustinar · IOme1,19; IOme3,2
- ustnica · Sže1,2; Sže4,3; Sže5,3; Sže6,2;
Sžp1,16; Sžp2,15; Sžp3,3; Sžp4,5,1; Sžp5,10;
Sžp6,14
- ustrašiti se · Gce; Gcp; GLme,13; GLmp,2;
GLže,3; GNE
- ustreliti · Gcp,2; GLme,25; GLmp,15; GLže,2;
GNE,2; GVbe
- ustvariti · Gce,2; GLme,7; GLmp,2; GLže;
GNE,2
- usuti · Gce,2; Gcp,3; GLmp,6; GLse; GLže,4
- uš · Sže1,2; Sžp1,2; Sžp2,4; Sžp4,5; Sžp6
- utegniti · Gae,7; Gbe; Gce,8; GLmd; GLme,17;
GLmp,6; GLse,2; GLže,10
- utihniti · Gce,11; Gcp,3; GLme,13; GLmp,5;
GLse; GLžd; GLže,14; GLžp,2; GNE
- utoniti · Gbe; Gbp; Gce; Gcp; GLme,6; GLmp;
GLse; GLže,5
- utripati · Gce,9; Gcp; GLmd; GLme; GLmp;
GLse,14; GLže,4; GNE,6
- utrujen · Pme1,7; PNme1; PNme1i,2; PNme2;
PNme4i; PNme6; PNmp1,3; PNse1; PNže1,5;
PNže4; PNže6; PNžp1,3; PNžp4
- utrujenost · Pže4; Sže1,6; Sže2,4; Sže4,2
- užaljen · GNmp1; GNže1; PNme1i; PNme1,6;
PNme6,4; PNmp2; PNse5; PNže2
- užaljeno · A,46
- užitek · Sme1,2; Sme2,2; Sme4; Sme5,2;
Sme6,18
- uživati · Gae; Gbe; Gce,2; Gcp,2; GLmd;
GLme,8; GLmp,2; GLže
- v · E4,3808; E5,3999; E,3
- vabiti · Gae; Gbe; Gce,4; Gcp; GLme,3; GLmp;
GNE,2
- vajen · Pmd1; Pme1,4; Pmp1,2; Pse1; Pže1,2;
Pžp1
- val · Sme1,7; Sme5; Sme6; Smp1,6; Smp2,3;
Smp3; Smp4,11; Smp5,9; Smp6,2
- valiti · Gce,16; Gcp,4; GLme,3; GLmp,4;
GLže,8; GLžp,2; GNE,2
- valjati · Gce; GLme,7; GLmp,2; GLže; GLžp
- valpet · Sme1,4; Sme3,2; Sme4; Smp1,3; Smp2;
Smp2; Smp3
- vame · ZOae4,60
- varati · Gbe,3; Gce,7; GLme,2; GLže,2; GLžp
- varen · Pme4i; Pmp1; Pse2,2; Pse5,3; Pse6;
Pže1; Pžp5
- vas · Sže1,66; Sže2,66; Sže3,15; Sže4,10,3;
Sže5,12,5; Sže6,24; Sžp1,5; Sžp2,3; Sžp4,2;
Sžp5,2; Sžp6,3
- vaš · ZSVbpmd3; ZSVbpme1,5; ZSVbpme5;
ZSVbpmp1; ZSVbpse1,3; ZSVbpse5;
ZSVbpže1,1,7; ZSVbpže2,2; ZSVbpže3;
ZSVbpže5; ZSVbpžp1
- vaščan · Sme1,2; Smp1,3,4; Smp2,3; Smp3,2;
Smp4,5; Smp5,5; Smp6
- vaški · Pme1,3; Pme2,2; Pme4; Pme5,2;
Pmp1,9; Pmp2,6; Pmp3; Pmp4,5; Pže1,4;
Pže2,2; Pže4,3; Pžp1,4; Pžp2,3; Pžp3; Pžp4,3;
Pžp5
- važen · Pme1,1,2; Pme1,2; Pme6,3; Pmp2;
Pmp6; Pse1,1; Pse1,1,2; Pse2; Pse5,3; Pže1;
Pže4; Pžp1
- včasih · A,228
- včeraj · A,23
- vdan · GNme2; GNmp1; PNme1,2; PNme2;
PNme4i,2; PNme6,2; PNmp1; PNžp6; Pže4
- vdano · A,36
- vdati se · Gae,5; Gce; GLmd,2; GLme,13; GLsp;
GLže,7; GLžp; GNE
- vdreti · Gcp; GLme,2; GLmp,5; GLse,2; GLže;
GNE,3
- več · Ajj,9; Č,627; ŠNE,30
- večen · Pme1i; Pme1,2; Pme2,2; Pme4i,2;
Pme5; Pme6,2; Pmp1; Pmp4,5; Pse2; Pse4,2;
Pse6,2; Pže1,5; Pže4; Pže5,3
- večer · Smd4; Sme1,28; Sme2,15; Sme3,11;
Sme4,8,5; Sme5,4; Sme6,5; Smp1,4; Smp2;
Smp3; Smp4; Smp5,10; Sse2,5
- večeren · Pme2; Pme4i,4; Pme5,2; Pmp1;
Pmp4; Pse1,3; Pse2; Pse4; Pse5,11; Pže1,2;
Pže2,2; Pže5,3; Pžp5,2
- večerja · Sže1,6; Sže2,7; Sže3,3; Sže4,12; Sže5,6;
Sže6; Sžp2; Sžp3
- večkrat · Ajj,7; A,120
- večno · A,14
- večnost · Sže1,7; Sže2,2; Sže3; Sže4,13

- vedeti · Gad,2; Gae,189; Gbd,6; Gbe,137;
Gbp,39; Gcd; Gce,9; Gce,137; Gcp,15; Gcp,34;
GLmd,8; GLme,231; GLmp,42; GLse,2; GLžd;
GLže,65; GLžp,3; GNE,23
- vedno · A,128
- veja · Šze1,19; Šze2,9; Šze4,20; Šze5,4; Šze5,5;
Šze6,4; Šzp1,17; Šzp2,11; Šzp4,23; Šzp6,5
- vek · Sme2; Sme4,2; Sme5,2; Smp2; Smp4,5
- veka · Šzp1,3; Šzp2,2; Šzp4,7; Šzp5,2; Šzp6,5
- velik · Pmd1jj; Pmd2; Pmd4; Pmeli; Pmeli,17;
Pme1jj,2; Pme1j,8; Pme1,51; Pme2j,2;
Pme2,8; Pme3,3; Pme4i,11; Pme4j,2; Pme4,6;
Pme4,6; Pme5jj; Pme5,19; Pme6,32; Pmp1j;
Pmp1jj,2; Pmp1,13; Pmp2,4; Pmp3; Pmp3j;
Pmp4,5; Pmp5,3; Pmp6,9; Pse1jj; Pse1j,2;
Pse1,12; Pse2jj; Pse2,10; Pse3; Pse4; Pse4j;
Pse4,11; Pse5,2; Pse6j,2; Pse6,5; Psp2,2;
Psp3; Psp5,2; Psp6,2; Pže1jj,3; Pže1j,5; Pže2j;
Pže2jj,2; Pže2,13; Pže3,3; Pže4jj,3; Pže4j,3;
Pže4,50; Pže5j; Pže5jj,2; Pže5,22; Pže6jj;
Pže6j,4; Pže6,21; Pžp1j; Pžp1,19; Pžp2,6;
Pžp4j,2; Pžp4,35; Pžp5,8; Pžp6,36
- velikan · Sme1,12; Sme4,4; Sme6,2; Smp1,5
- velikanoč · Šze1,9; Šze4,7
- velikanski · Pme1i,2; Pme1,2; Pme3; Pme4,2;
Pme5; Pme6,2; Pmp1; Pmp5,2; Pse1; Pse4,4;
Pse6; Psp2; Pže1,6; Pže2,4; Pže4,6; Pže5;
Pžp4; Pžp6,2
- veliki · Pmd1,4; Psp1,4; Psp4,3; Pže1,82
- veliko · Aj,14; A,53; ŠNE,13
- veljati · Gce,9; Gcp,2; GLme,5; GLse; GNE,2
- ven · A,30
- Venc · IOme1,162; IOme2,8; IOme3,20;
IOme4,29; IOme5,3; IOme6
- Venček · IOme1,151; IOme3,21; IOme4,12
- vendar · Vpr,371
- vendarle · Vpr,26
- venec · Sme1,2; Sme4,6; Sme6,6; Smp4
- venomer · A,56
- vera · Šze1; Šze2,5; Šze4,3; Šze5,6; Šze6,2
- veriga · Šze1,3; Šze2,5; Šze4,10; Šze5,2; Šze6,5;
Šzp1,2; Šzp2,4; Šzp4,2; Šzp5
- verjeti · Gae,21; Gap,2; Gbe,13; Gbp,3; Gcd;
Gce,9; Gcp,4; GLme,29; GLmp,11; GLže,5;
GNE,16; GVbe,2
- verjetno · A,27
- Veronika · IOme1,5; IOže1; IOže3,3; IOže4,3;
IOže6
- ves · ZTmd1,4; ZTme1,121; ZTme2,15;
ZTme3,6; ZTme4,3; ZTme4,119; ZTme5,8;
ZTme6,6; ZTmp1,475; ZTmp2,39; ZTmp3,41;
ZTmp5,10; ZTmp6,17; ZTse2,82; ZTse3,15;
ZTse5,46; ZTse6,32; ZTsp1,23; ZTsp2,9;
ZTsp3; ZTsp4,26; ZTsp5,3; ZTsp6,3;
ZTže1,181; ZTže3,8; ZTže4,112; ZTže5,28;
ZTže6,46; ZTžp2,39; ZTžp3,3; ZTžp5,31;
ZTžp6,20; ZT,26
- vesel · Pmd1,4; Pmeli; Pme1,44; Pme2,3;
Pme4,2; Pme4,2; Pme6,3; Pmp1,6; Pmp2;
Pmp3; Pmp4; Pmp6,2; Pse1,5; Pse2; Pse6;
Pže1j; Pže1,12; Pže2,3; Pže5; Pže6; Pžp1,3;
Pžp2; Pžp4,4
- veseliti se · Gbe; Gce; Gce,4; GLme,11; GLmp;
GLse,2; GLže,5
- veselje · Sse1,9; Sse2,18; Sse4,9; Sse5; Sse6,15
- veselo · A,143
- vest · Šze1,28; Šze2,3; Šze4,6; Šze5,5; Šze6,2;
Šzp1
- Veter · ISme1,10
- veter · Sme1,49; Sme2,8; Sme3; Sme4,7;
Sme5,46; Sme6,3; Smp1; Smp1,4; Smp2,3
- veža · Šze1,2; Šze2,8; Šze4,44; Šze5,44
- vežen · Pme4i; Psp1,6; Psp2,8; Psp3,2; Psp4,6;
Psp5,5; Psp6,8
- vhod · Smd4; Sme1; Sme2,2; Sme5,2; Sme6,4;
Smp4; Smp6
- vi · ZObp2,7; ZObp4,60; ZObp5,8
- viden · PNme1,2; PNme4; PNme5; PNme6,6;
PNmp1; PNse1,3; PNže1; PNže2; PNže5;
PNže5j,2; PNže6; PNžp4; PNžp6; Pžp4
- videti · Gad; Gae,61; Gap,11; Gbe; Gbe,140;
Gbp,15; Gce,39; Gcp,5; GLmd,10; GLme,332;
GLmp,58; GLse,4; GLže,66; GLžp,4; GNE,68
- videz · Sme4,12
- vidno · A,30
- vihar · Sme1,4; Sme2; Sme5,4; Smp1,2; Smp2;
Smp4
- vihrati · Gce,3; Gcp,10; GLme,4; GLmp,3;
GLže,3; GLžp,2
- vijuga · Šze1,5; Šze4,4; Šzp4
- vijugati se · Gce,6; GLme,6; GLmp,2; GLže,11;
GLžp,3
- vilice · Šzp1; Šzp4,7; Šzp6,4
- vino · Sse1,34; Sse2,52; Sse3; Sse4,9; Sse5,9;
Sse6,2; Ssp4
- vinograd · Sme2,4; Sme4,3; Sme5,6; Sme6,3;
Smp1,4; Smp2; Smp4,3; Smp6,4
- viseč · GČmp6; PČme2; PČme4i; PČme6;
PČmp4,6; PČmp6; PČsp2; PČsp5; PČže6;
PČžp2; PČžp6,2

viseti · Gcd; Gce,22; Gcp,2; Gcp,8; GLme,22;
GLmp,8; GLse,5; GLžd; GLže,28; GLžp,9
visok · Pme1; Pmeli,3; Pme1jj; Pmelj,6;
Pme1,26; Pme2,3; Pme3,2; Pme4i,2; Pme4,3;
Pme5j; Pme5,5; Pme6j; Pme6,11; Pmp1j,8;
Pmp1,3; Pmp2,2; Pmp3; Pmp3j; Pmp4,3;
Pmp5; Pmp6,7; Pse1j; Pse1,12; Pse2jj;
Pse2,3; Pse4,4; Pse5,4; Pse6; Psp4j; Psp5;
Pzd1jj; Pzd6; Pže1jj,3; Pželj,4; Pže1,31;
Pže2j; Pže2,12; Pže3; Pže4jj,2; Pže4,15;
Pže5,9; Pže6jj; Pže6,5; Pžp1jj; Pžp1j,2;
Pžp1,3; Pžp2,4; Pžp4,4; Pžp5,7; Pžp6,3
visoko · Aj; Aj,51; A,48
višava · Šze5,2; Šzp2; Šzp4,5; Šzp5,2
višina · Šze2,3; Šze4,4; Šze5,3; Šzp2; Šzp5
vitek · Pme1; Pme4i; Pme5,2; Pmp1,2; Psel,3;
Pse4; Psp1; Pžel,4; Pže2; Pže4; Pžp1,2
viti · Gce,12; Gcp; GLme; GLže,7
vlačiti · Gae; Gbe,2; Gcd; Gce,6; Gcp,6;
GLme,7; GLmp,7; GLže,6; GLžp
Vladimir · IOme1,46; IOme2; IOme3,3;
IOme4,8; IOme5,2; IOme6
vlak · Sme1,11; Sme2,9; Sme4,6; Sme5,4;
Smp1,2
vlažen · Pme4i; Pme5,2; Pmp1; Psel; Pžel,3;
Pže2,2; Pže5,2; Pžp1
vleči se · Gae; Gbe,4; Gcd; Gce,42; Gcp; Gcp,10;
GLmd,5; GLme,2; GLme,37; GLmp,9; GLse,6;
GLže,18; GLžp,4; GNE,8
vmes · A,20
vneto · A,21
vnuk · Sme1,7; Sme3,2; Sme4,2; Sme6,2
voda · Šze1,80; Šze2,77; Šze3,6; Šze4,107;
Šze5,66; Šze6,12; Šzp1,9; Šzp2; Šzp5; Šzp6,3
voden · PNse4; PNse6; PNže4; PNžp1; PNžp2;
PNžp5; Pse5; Pže1,2; Pže2,2; Pže4,3; Pže5,3;
Pže6; Pžp1; Pžp1,3; Pžp6,3
voditi · Gce,2; Gcp,3; GLme,3; GLmp,2; GLže,5;
GLžp,2; GNE
vogal · Smd2; Sme2,19; Sme4,7; Sme5,3;
Sme6,2; Smp2,3; Smp5
vojak · Smd1,12; Sme1,30; Sme2,4; Sme3;
Sme4,5; Sme6; Smp1,54; Smp2,15; Smp3,11;
Smp4,10; Smp5,2; Smp6,3
vojaški · Pme1,9; Pme3; Pme4; Pme4; Pme5;
Pme6,2; Pmp1,2; Pmp2,4; Pmp5; Psp6; Pžel;
Pže2,2; Pže4,15; Pže5,2; Pže6,2; Pžp2; Pžp5
vojaško · A,10
vojen · Pme1i,4; Pme2,3; Pme4; Pme5; Pmp4,2;
Pmp6; Pže6; Pžp5
vojna · Šzd1,2; Šze1,23; Šze2,18; Šze4,10;
Šze5,30; Šze6,38; Šzp1,3; Šzp4

Vojnac · IOme1,14; IOme3; IOme5,2; IOme6
Vojnačev · IOmp1; IOsp2; PIOmd4; PIOme4;
PIOmp3; PIOmp5,3; PIOže1,3; PIOže4,3;
PIOžp1,4
Vojnačka · IOže1,19; IOže3,4; IOže4,2; IOžp1,8
vojska · Šze1,19; Šze2,7; Šze3,2; Šze4,8; Šze5,6;
Šzp2
vol · Sme1,6; Smp2,2; Smp4,2; Smp6
volja · Šze1,17; Šze2,31; Šze4,26; Šze5,12;
Šze6,2
vonj · Sme1,9; Sme2; Sme4,11; Sme6,4; Smp2
votel · Pme1,4; Pme6,10; Pmp1,2; Psel,7;
Pse4,2; Pse6; Pžel,3; Pže4; Pžp1,3; Pžp6,2
votlo · A,21
voz · Smd4; Sme2,12; Sme3,2; Sme4,25; Sme5,3;
Sme6,8; Smp1,4; Smp2; Smp4; Smp5; Smp6;
Smp6
votel · Sme1,7; Sme4,2; Sme5,3; Smp2
voziti · Gap; Gce,10; Gcp; GLmd; GLme,13;
GLmp,5; GLme,3; GNE
vožnja · Šze1,5; Šze2,4; Šze4,4; Šze5; Šze6,3
vpiti · Gbe; Gcd; Gce,10; Gcp,4; GLme,6;
GLmp,4; GLse; GLže,2; GLžp; GNE,3
vpite · SGse1,6; Sse1; Sse2,3; Sse6
vplivati · Gce,2; Gcp; GLme,3; GLse,2; GLže,2
vprašanje · Sme6; Ssd1; Sse1,33; Sse2,6;
Sse4,22; Sse5,6; Sse6,4; Ssp1; Ssp2,4; Ssp4,4;
Ssp5; Ssp6,2
vprašati · Gae,16; Gbe,2; Gbp; Gcd,2; Gce,81;
Gcp,6; GLmd,5; GLme,339; GLmp,22;
GLže,122; GLžp,4; GNE,12; GVbe,9; GVbp,3
vpraševati · Gae,3; Gbe,3; Gcd,2; Gce,6; Gcp;
GLme; GLmp
vprašujoče · A,47; GČ
vračati · Gae,6; Gbe; Gce,17; Gcp,11; GLmd,2;
GLme,26; GLmp,14; GLse,2; GLže,8; GLžp,4
vrag · Sme1,4; Sme2; Sme3,3; Sme4,5
Vranjek · IZme1,4; IZme2,3; IZme4,13; IZme6,8
vrat · Sme1,6; Sme2,19; Sme4,38; Sme5,23;
Sme6,11; Smp4,4; Smp5; Smp5; Smp5,5
Vrata · IZsp1,8; IZsp2; IZsp4,11
vrata · Ssp1,86; Ssp2,53; Ssp3,61; Ssp4,206;
Ssp5,69; Ssp6,61
vrba · Šze1,2; Šze2,8; Šze4,11; Šze5,7; Šzp1,15;
Šzp2,8; Šzp3,3; Šzp4,5; Šzp5,6; Šzp6,7
vrhje · Sse2,2; Sse3; Sse4,10; Sse5,2; Sse6,2
vreča · Šzd4; Šze1,3; Šze3; Šze4,16; Šze6,4;
Šzp1,2; Šzp2; Šzp4,13
vreči · Gae,4; Gbe,2; Gce,63; Gcp,6; GLmd,2;
GLme,114; GLmp,20; GLse,8; GLže,62;
GLžp,2; GNE,2; GVbe,3
vred · A,22

- vreden · Pme1,10; Pmp1; Pse1,13; Pse2,2; Pse6;
Psp1; Pže1,3; Pže4; Pže6; Pžp1,2
- vreme · Sse1,12; Sse4,6; Sse5,2
- vresje · Sse1; Sse2,5; Sse4; Sse5,4; Ssp1,4; Ssp4
- vreščati · Gce,3; Gcp,3; GLmd; GLme; GLmp,2;
GLže,5; GNE
- vreti · Gce,17; Gcp; GLmp,2; GLse,7; GLsp;
GLže,8; GNE
- vrh · E2,16; Sme1,9; Sme2,24; Sme4,22;
Sme5,30; Sme6,2; Smp1,9; Smp2; Smp3;
Smp4,5; Smp5,9; Smp6,4
- vrisk · Sme1,10; Sme4,2; Sme5; Sme6,5;
Smp1,2
- vriskati · Gbe; Gce; Gcp,6; GLme; GLmp,4;
GLže; GNE; GVbe,2
- vrniti · Gae,10; Gap,2; Gap,2; Gbe,3; Gbp;
Gcd,2; Gce,30; Gcp,6; GLmd,6; GLme,172;
GLmp,19; GLse; GLže,58; GLžp,5; GNE,11;
GVad; GVbp,4
- vroč · Pme2; Pmp4,2; Pse1,4; Pse2,2; Pse4;
Pse5; Psp1; Psp2; Psp4,2; Pže2,2; Pže4,3;
Pže5; Pžp2
- vročina · Šze1,7; Šze2; Šze5,4
- vrsta · Šzd4; Šzd5; Šzd6; Šze1,10; Šze2,34; Šze3;
Šze4,12; Šze5,30; Šze6; Šzp1; Šzp2,5; Šzp4;
Šzp5,7
- vrstica · Šzd1,2; Šzd4,3; Šze1; Šze4; Šzp2,2;
Šzp4
- vrt · Sme1,5; Sme2,5; Sme3,2; Sme4,12;
Sme5,38; Sme6,4; Smp2; Smp5,2; Smp6
- vrtati · Gae; Gbp; Gce,2; Gcp; GLme,2; GLžd;
GNE,2
- vrteti · Gcp,4; GLmd; GLme,17; GLmp,3;
GLse,2; GLsp,2; GLžd,2; GLže,12; GLžp;
GNE,3
- vrtiček · Sme2,3; Sme3,4; Sme4; Sme5; Smp6
- vrtinec · Sme1,2; Sme4,2; Sme5,8; Smp1,3;
Smp2
- vrtiti · Gce,21; Gcp,4
- vrv · Šze1,5; Šze2,2; Šze4,23; Šze5,3; Šze6,4;
Šzp1; Šzp2; Šzp4,5; Šzp5
- vrzel · Šze1; Šze2,2; Šze4,4; Šzp4,2; Šzp5
- vsa · ZTže2,17
- vsaj · Č,108
- vsak · ZTme1,131; ZTme2,10; ZTme3,13;
ZTme4,15; ZTme4,124; ZTme5,24; ZTme6,4;
ZTse1,6; ZTse2,5; ZTse3,2; ZTse4,42; ZTse5,7;
ZTsp2; ZTsp4; ZTze1,36; ZTze2,11; ZTze3,4;
ZTze4; ZTze4,38; ZTze5,10; ZTze6,3;
ZTžp2,3; ZTžp4
- vsakdanji · Pme1,2; Pme2,2; Pme4,6; Pme5,3;
Pme6,3; Pmp2; Pse1,3; Pse2; Pže1,2; Pže4,3;
Pžp1
- vse · ZTse1,546; ZTse4,387; ZTžp1,92;
ZTžp4,133; ZT,33
- vseeno · A,48
- vsekakor · Č,23
- vselej · A,84
- vsepovsod · A,13
- vsesti · Gce; GLme,7; GLže,2
- vsi · ZTmp4,125
- vstajati · Gae; Gap; Gbp; Gce,6; Gcp,5; GLmd;
GLme,5; GLmp,2; GLžd; GLže
- vstati · Gae,8; Gap; Gce,16; Gcp,3; GLmd,4;
GLme,93; GLmp,14; GLse,2; GLže,38; GLžp;
GNE,9; GVbe,12; GVbp
- vstopiti · Gae; Gap; Gce,10; Gcp,2; GLmd;
GLme,16; GLmp,4; GLse; GLže,10; GVbd;
GVbe,2
- vstran · A,12
- vsuti · Gce,3; Gcp,2; GLmp,4; GLse; GLžp
- všeč · PD,36
- vštric · A,10; E2,3
- vtakniti · Gbe; Gce,4; Gcp,2; GLme,13;
GLmp,2; GLže,8; GNE,3; GVbe
- vtikati · Gap; Gce; GLme,4; GLmp; GLže;
GNE,2; GVbe
- vzdigniti · Gae,3; Gce,6; GLmd,2; GLme,71;
GLmp,8; GLse; GLže,24; GLžp; GNE
- vzdih · Sme1,8; Sme3,2; Sme4; Sme5; Sme6,8;
Smp4,2; Smp5
- vzdihniti · Gce,20; GLme,71; GLmp; GLže,51
- vzdrgetati · Gce; GLme,16; GLse; GLže
- vzdržati · Gce; GLme,5; GLse,3; GNE
- vzeti · Gad; Gad,2; Gae,6; Gap,2; Gcd,2;
Gce,31; Gcp,3; GLmd,10; GLme,125;
GLmp,13; GLse,6; GLže,42; GNE,12; GVap,3;
GVbd; GVbe,15; GVbp,3
- vzklik · Sme1,3; Sme2,2; Sme4; Sme5,2;
Sme6,5; Smp6
- vzklkniti · Gcd; Gce,27; GLmd; GLme,56;
GLmp; GLse; GLže,46; GNE
- vzmet · Šze1,5; Šze5; Šzp1,3; Šzp2,2; Šzp4; Šzp5
- vznak · A,26
- vznemiriti · Gcp; GLme,4; GLse; GLsp; GLže,3;
GLžp,2
- vznemirjati · Gce,3; Gcp,2; GLme,5; GLmp;
GLse; GLže; GLžp,2; GNE,2
- vznemirjen · GNme1,5; PNme1; PNse1; PNse2;
PNže1,3
- vzpenjati se · Gae,3; Gcd; Gce,3; Gcp,3;
GLme,4; GLse; GLže,3; GNE,9

- vzravnano · A,15
vzrok · Sme1,3; Sme2,4; Sme4,2; Smp4
vztrajati · GLmd; GLme,10; GLže,2
vztrajno · Aj; A,14
vztrepetati · Gce,2; Gce,5; GLme,30; GLmp;
GLse; GLsp; GLže,12; GLžp
vztrepetavati · Gce,4; Gcp; GLme,2; GLmp,6;
GLse; GLže,3; GLžp,2; GNE
z · E; E2,269; E6,2911
za · E2,21; E4,1526; E6,748; E,32
zabliskati · Gce,2; GLme,2; GLmp; GLse,4;
GLže,4; GLžp,2
zaboleti · Gce,2; GLmp; GLse,6; GLžp
zabosti · Gce,7; GLme,6; GLže
zabresti · Gce,2; GLme,6; GLmp,4; GLže,3
zabrundati · Gce; GLme,7; GLmp; GLže,2
zabrusiti · Gce,3; Gcp,2; GLme,4; GLmp;
GLže,2
zacviliti · Gce; Gcp,2; GLmd; GLme,3; GLse,2;
GLže
začetek · Sme1,6; Sme2,10; Sme4,2; Sme5,7
začeti · Gad; Gae,7; Gap; Gcd,4; Gce,121;
Gcp,3; Gcp,31; GLmd,8; GLme,289; GLmp,41;
GLse,36; GLžd,2; GLže,123; GLžp,10; GNE,3;
GVap,5; GVbe,4; GVbp
začuden · PNme1; PNme1i,2; PNme6,2;
PNmp1,2; PNmp4; PNmp6,3; PNže1,2;
PNže3,2; PNžp4; PNžp6,2; Pse1,2; Pže6
začudenje · Sse1,3; Sse2,8; Sse4,2; Sse5,3;
Sse6,10
začudeno · A,124
začuditi se · Gae,2; Gce,14; GLmd; GLme,61;
GLmp,4; GLže,23; GNE,4
začititi · Gae,6; Gce,3; Gcp; GLme,61; GLmp,3;
GLže,24; GLžp
zadeti · Gae; Gce,6; Gcp; GLmd,3; GLme,31;
GLmp,5; GLse,3; GLže,15; GLžp,5; GNE,3;
GVbe
zadevati · Gcp,4; GLme,5; GLmp; GLse; GLže
zadni · Pme1,24; Pme2,6; Pme3,2; Pme4,3;
Pme4,8; Pme5,4; Pmp1,3; Pmp2,3; Pmp4,12;
Psd4; Pse1,6; Pse2; Pse4,4; Pse5,2; Psp1,3;
Psp2; Psp4,8; Psp5; Pže1,15; Pže2,9; Pže4,21;
Pže5,9; Pže6; Pžp1,10; Pžp2,4; Pžp4,19;
Pžp5,2; Pžp6,4
zadnjič · A,23
zadoneti · Gce,3; Gcp; GLme,4; GLmp,2;
GLže,20
zadovoljen · Pme1; Pme2; Pmp1,3; PNme1,7;
Pže1,2
zadovoljno · A,47
zadovoljstvo · Sse1,3; Sse2,2; Sse5; Sse6,6
zadrega · Sže1; Sže2,4; Sže4,5; Sže5,56; Sže6,5
zadreti se · Gce,27; Gcp; GLmd; GLme,29;
GLmp,9; GLže,8; GLžp; GNE
zadrgalica · Sže1,4; Sže2,2; Sže4,11; Sžp4,2
zadržati · GLmd,3; GLme,10; GLmp,4; GLžd;
GLže,6; GLžp; GNE,4
zadrževati · Gcd; Gce,3; GLmd; GLme,5;
GLmp,3; GLže,4
zadušiti se · Gce,5; GLmp; GLže,2; GNE,3
zaganjati se · Gce,2; GLme,3; GLmp,4; GLsp;
GLže,3
zagledati · Gae,5; Gcd; Gce,14; Gcp,2; GLmd,3;
GLme,90; GLmp,15; GLse,2; GLže,28; GLžp;
GNE
zagnati · Gce,6; Gcp; Gcp,4; GLmd,2; GLme,17;
GLmp,5; GLže,10; GLžp,3
zagodrnjati · Gce,10; GLme,14; GLmp; GLže,4
zagolčati · GLme,14; GLse
zagorel · PLme1,2; PLme2; PLme5,3; PLmp4;
PLmp5; PLsp1; PLže4,2; PLžp4; PLžp5;
PLžp6
zagoreti · Gcp; GLme,2; GLmp; GLse,2; GLže,3;
GLžp
zagrebsti · Gce,2; GLmd; GLme,11; GLže;
GNE,4
zagrešiti · GLme,14; GLmp,2
zagristi · Gce,2; GLme,5; GLmp; GLže; GVbe
zagrmeti · Gce,32; Gcp; Gcp; GLme,61;
GLmp,2; GLse,7; GLže,3
zagroziti · Gce,2; GLme,6; GLmp,2; GLže,8
zahajajoč · PČse4,2; PČse5,8; PČsp2
zahajati · Gce,2; Gcp; GLme,16; GLmp; GLse,2;
GLže,2; GLžp; GNE,3; GVbe
zahlipati · Gce,7; GLme,7; GLse; GLže,15
zahreščati · Gce,5; GLme,8; GLse; GLže,2
zahropeti · Gcp,11; GLme,40; GLse
zahrumeti · Gce,4; GLme; GLmp; GLse,3; GLžp
zahtevati · Gce,9; GLme,5; GLmp,5; GLse;
GLže,5; GNE,2
zahvaliti se · Gap; Gce,3; GLmd,3; GLme,3;
GNE; GVbp
zaigrati · Gcd; Gce,5; GLme,5; GLmp; GLže;
GVbe,2; GVbp,8
zaiti · GLmd; GLme,5; GLmp; GLse,2;
GLže,7
zajahati · Gce; Gce,2; GLmd; GLme,6; GLmp
zajavkati · Gce,3; GLme,9; GLže
zajecljati · Gce,5; GLme,19; GLže
zajemati · Gce; Gcp,6; GLmd; GLme,2; GLmp;
GLsp; GLže,3
zajeti · Gae,4; Gce,10; GLme,33; GLmp,4;
GLže,12; GNE,2; GVbe,2

- zajokati · Gae; Gce,2; Gce,2; GLme,7; GLže,9;
GNE
- zakaditi se · Gbe; Gce,3; Gcp,2; GLme,12;
GLmp,3; GLže,7
- zakaj · Smp2; Vpr,164; ZV,352
- zakajen · PNme1; PNme1,4; PNme2; PNme4,2;
PNme5; PNme6,2; PNmp2; PNmp4; PNse1;
PNse4; PNsp2,3; PNže1,2; PNže2; PNže4,2;
PNže5; PNže6,2; PNžp2; PNžp6; Pžp5
- zaklati · Gcp; GLme,2; GLmp,6; GNE; GVbe
- zakleniti · Gce; GLme,6; GLmp; GLže,3; GVbe
- zakleti · Gce,2; GLmd,2; GLme,7; GNE
- zaklicati · Gce,2; GLme,6; GLže; GLže,5; GNE
- zaključiti · Gce; GLme,23; GLže,4
- zaklon · Sme2,4; Sme4,2; Sme5,4; Smp2
- zakon · Sme1; Sme2; Sme4,4; Sme5,2; Sme6;
Smp1; Smp2,2; Smp4,4; Smp5,2
- zakonski · Pme4; Pse2,3; Pže1; Pže4,3; Pžp2;
Pžp4; Pžp6
- zakopati · Gce; GLme,6; GLmp,4
- zakričati · Gcd; Gce,7; GLme,73; GLmp,6;
GLže,25
- zakriliti · GLme,12; GLže,10
- zakristija · Sže2,2; Sže4,5; Sže5,5
- zakriti · Gce; GLme,3; GLmp; GLžd; GLže,5;
GLžp,3; GNE
- zakrižan · PNse1,2; PNse4,6; PNsp1; PNsp2;
PNsp5; PNsp6; PNže1
- zakrohotati se · Gcd; Gce; Gce,7; Gcp; GLmd,2;
GLme,22; GLmp,10; GLže
- zaleči · Gce,7; GLme; GLse,2; GLže
- zaletavati se · Gcp,2; GLme,2; GLmp,5; GLse,2;
GLže
- zaleteti se · Gce,4; Gcp,2; GLme,9; GLmp;
GLse; GLže
- zalit · GTže1,7; GTžp1,2; Pmp4; PTme1;
PTme1,3; PTme5; PTme6,2; PTmp1; PTmp4;
PTmp6; PTse4,2; PTže1; PTže4,4; PTže5;
PTže6,2; PTžp4,2; PTžp6,3
- zaliti · Gce,3; Gcp; GLme,2; GLže,6; GLžp,4;
GTse1; GTse6
- zaljubljen · GNme1,7; PNme1; PNmp2; PNžp6
- zaloputniti · Gce; Gcp; GLme,11; GLmp; GLsp;
GLže,2
- zamah · Sme2; Sme5,2; Sme6,5; Smp1; Smp2;
Smp6,7
- zamahniti · Gae; Gce,25; GLme,108; GLmp;
GLse; GLže,24; GNE
- zamajati · Gce,11; Gcp; Gcp; GLme,39;
GLmp,2; GLse; GLže,13
- zamakniti se · Gae; Gcd,2; Gce,21; Gcp,5;
GLme,93; GLmp,3; GLže,19; GNE
- zamaknjen · GNme1,6; PNmp1; PNže5;
PNžp1,2; PNžp6; Pse1
- zaman · A,51
- zamaščen · GNme1; PNme2,2; PNme4,4;
PNme5; PNže4; Pže4
- zamegliti se · Gce,2; GLmp; GLse,5; GLžp,2
- zameriti · Gae; Gbp,8; Gce,7; GLme,4; GNE;
GVbp,4
- zamežikati · GLme,26
- zamisliti · Gae; Gcd; Gce,7; GLme,45; GLmp,2;
GLže,8; GNE; GVbp
- zamišljen · PNme1,7; PNmp1; PNže1,4
- zamišljeno · A,28
- zamižati · Gce,9; GLme,25; GLže,7
- zamolkel · Pme1; Pme6,12; Pse1,2; Pse6,2;
Pže1; Pže6; Pžp4
- zamolκλο · A,13
- zamomljati · Gce; GLme,13
- Zamorka · IOže1,12; IOže2,2; IOže3,2; IOže6,2
- zamrmrati · Gae,2; Gce,19; Gcp; GLme,111;
GLmp,2; GLse; GLže,19
- zamukati · Gbe; Gce,3; Gcp,2; GLme; GLže,7;
GLžp
- zanergati · Gce; GLme,5; GLže,4
- zanesti · Gbe; Gce; Gcp; Gcp; GLme,4; GLse,4;
GLže,3
- zaničljiv · Pme1; Pmeli; Pme6,3; Pmp4,2;
Pse6,2; Pže6
- zaničljivo · A,72
- zanimanje · Sse2,5; Sse6,5
- zanimati · Gae; Gce,13; Gcp,2; GLmd,2;
GLme,10; GLmp,6; GLse,4; GLže,9; GLžp;
GNE
- zanimiv · Pmd1; Pme1,2; Pse1,3; Pse2,3;
Pže1,3; Pže4; Pžp1; Pžp4
- zanos · Sme2,3; Sme5,3; Sme6,5
- zapeči · Gce,2; GLse,2; GLže,8; GLžp
- zapestje · Sse1; Sse2,2; Sse4,4; Sse5,2; Sse6;
Ssp2; Ssp5
- zapeti · Gae,2; Gce,10; Gcp,2; Gcp,2; GLmd,3;
GLme,31; GLmp,11; GLže,4; GLžp,3; GNE,6;
GVap; GVbe,2
- zapirati · Gap; Gbp; Gce,3; Gcp,4; GLme;
GLmp,3; GLse; GLsp; GLže; GNE
- zapisan · GNme1,3; GNmp1,2; GNse1,3; GNže1;
PNže4
- zapisati · Gae; Gce; Gcp; GLme,4; GLže; GNE;
GVbe
- zapiskati · Gce,2; GLme,9; GLse; GLže; GVbe
- zaplavati · Gce; GLme,9; GLžd; GLže; GLžp

- zaplesati · Gae; GLme,3; GLmp,3; GLže,2;
 GLžp
 zapoditi · Gce,3; Gcp; GLme,2; GLmp,4; GNE
 zapomniti · Gce,6; GLme,18; GLže,2; GNE,2;
 GVbe,4; GVbp,2
 zaprašen · GNže4; GNže5; GNžp1; PNmp2;
 PNmp5; PNse5; PNse6; PNsp4; PNže1;
 PNžp1
 zapresti · Gce; GLmd,3; GLme,8; GLmp,25;
 GLse,3; GLsp,6; GLže,36
 zapreti · Gae; Gce,18; Gcp; Gcp,2; GLme,65;
 GNA; GNE,2; GTmd1; GTmp1; GVbe; GVbp
 zaprositi · GLme,9; GLmp; GLže,5
 zaprt · GTme1,31; GTsp1,3; PTme4; PTme6;
 PTmp1,2; PTse1,27; PTse5; PTsp1,3; PTsp4;
 PTsp6,3; PTže1; PTžp1; PTžp6
 zapustiti · Gcp,2; GLmd; GLme,12; GLmp;
 GLže,10; GLžp,2; GNE,2
 zapuščen · GNme1,5; GNme3; GNse1; GNže1,4;
 PNme1; PNme2,2; PNme4; PNme4i,2;
 PNmp1; PNže1; PNže4; PNže5,3; PNže6;
 PNžp1; PNžp4
 zaradi · E2,225
 zarasti · Gce,9; Gcp,4; GLže,2
 zaregljati · Gce,5; Gcp,4; GLže,5
 zares · A,33
 zarevsati · Gce,2; GLme,7; GLže
 zarezati · Gce,5; Gcp,2; GLme,8; GLmp,2;
 GLse; GLže
 zarežati se · Gce,4; GLme,15; GLmp; GLže,6;
 GLžp
 zarisati · Gce; Gce,2; Gcp; GLme,5; GLmp;
 GLse; GLže,3; GLžp,3; GNE
 zariti · Gce,3; GLme,7; GLže,3; GLžp
 zarjavel · PLme2,2; PLme3; PLme5,2; PLmp1;
 PLmp4,3; PLmp5; PLse4; PLsp4; PLže1,3;
 PLže2; PLže4; PLže5; PLže6; PLžp1
 zarohneti · Gce; GLme,14
 zaropotati · Gbp; Gce,4; Gcp; GLme,9; GLse,2;
 GLže,2
 zasikati · Gce,4; GLme,3; GLže,4
 zaskrbljeno · A,23
 zaslīšati · Gce,11; Gcp,3; GLme,35; GLmp,7;
 GLse,5; GLže,2
 zaslužiti · Gbe; Gce,5; Gcp,3; GLme,10; GLže,3;
 GNE
 zasmejati se · Gae; Gcd; Gce,11; Gcp,3;
 GLmd,5; GLme,17; GLmp,6; GLže,15; GLžp;
 GNE
 zasmiliti se · Gcd; Gce,2; GLme,9; GLže,3
 zasnežen · GNže1; Pme1i; PNme1i; PNme2;
 PNme5; PNse1; PNse5,2; PNže1,5; PNže2,2;
 PNže4; PNže5,2; PNžp2,2; PNžp4; PNžp5;
 Pse1,3; Pže4
 zaspan · Pmd4; Pme1i,8; Pme4; Pme6; Pmp5;
 Pže1,2; Pžp2; Pžp4
 zaspati · Gap; Gbe,2; Gcd,2; GLmd,4; GLme,26;
 GLmp,2; GLže,4; GNE,8; GVbe,8
 zastati · Gce,7; GLme,4; GLže,6; GLžp
 zastava · Šže1,11; Šže4,9; Šže6; Šžp1; Šžp2;
 Šžp4,5; Šžp5; Šžp6
 zastokati · Gce; Gce,7; GLme,54; GLse,3;
 GLže,21
 zastonj · A,11
 zastrmeti se · Gcd; Gce,12; GLme,18; GLmp;
 GLže,9
 zastrupiti · GLme,8; GLže,2; GNE
 zasukati · Gbd; Gce,10; Gcp,3; GLme,18;
 GLmp,2; GLse,3; GLže,2; GLžp; GVbe; GVbe
 zasvetiti · Gce; Gcp,6; GLse; GLžd; GLže,4;
 GLžp
 zašepetati · Gae; Gap; Gce,8; GLme,32; GLse,3;
 GLže,11
 zaškripati · Gce,5; Gcp,10; GLme,12; GLmp,7;
 GLse; GLsp,8; GLže,6; GLžp,6; GNE
 zašumeti · Gce; GLmp,2; GLse,13; GLžp,2
 zatajiti · Gae; Gap; GLme,5; GLmp,2; GLže;
 GNE
 zatakiniti · Gae,2; Gce,4; GLme; GLme,4;
 GLže,2
 zateči se · Gce; GLmd; GLme,9; GLmp,3;
 GLže,4; GLžp
 zateglo · A,10
 zategniti · Gce,8; Gcp; GLme,40; GLmp,5;
 GLže,7
 zatisniti · Gae; Gce; GLme,8; GLmp; GLže,2;
 GNE
 zato · A,81; Vpr,374
 zatohel · A; Pme1i; Pme5; Pmp2; Pže1; Pže4;
 Pžp1; Pžp5,3
 zatuliti · Gce,8; GLme,21; GLse; GLže,8
 zaupati · Gbp,2; Gce,2; GLme,7; GLmp,3; GVbe
 zavaliti · Gcp; GLmd; GLme,4; GLmp; GLže,4
 zavaljen · Pme4; PNmd2; PNme1i,5; PNme1,5;
 PNme3; PNme4,3; PNmp1,2; PNse4,2;
 PNse6; PNže1,6; PNže3; PNže4; PNžp2
 zavedati se · Gce,4; Gcp; GLme,6; GLmp; GLže
 zavedeti se · Gae; Gce; Gcp; GLme,28; GLmp,2
 zavest · Šže1,8; Šže5; Šže6,2
 zavezati · Gap; Gce,2; GLmd,3; GLme,4; GLmp;
 GLže
 zavijati · Gap; Gce,4; Gcp,2; GLme,6; GLže;
 GNE
 zavitek · Smd4; Sme1,2; Sme4,12; Sme5; Smp1;
 Smp4,2

zaviti · Gae,4; Gap,2; Gbe; Gcd; Gce,5; Gcp,3;
 GLmd,6; GLme,52; GLmp,11; GLže,10;
 GLžp,2; GNE,2; GTme1; GTmp1; GTže,4,2;
 GVbe,2; PTme4i
 zavod · Sme2; Sme4,4; Sme5,5; Sme6
 Zavoglar · IOme1,34; IOme3,3; IOme4,4; IOme6
 zavpiti · Gce,27; GLme,15; GLmp,3; GLse;
 GLže,12; GNE
 zavreščati · Gce,5; GLme,14; GLže,8; GLžp
 zavreti · Gce,4; GLse,3; GLsp,2; GLže,6
 zavriskati · Gcd; Gce; Gcp; GLme,9; GLže;
 GNE
 zavrniti · Gce,2; GLme,6; GLže,6
 zavrteti · Gad; Gcd; Gce,6; Gcp; Gcp,2; GLmd;
 GLme,15; GLmp,2; GLse,3; GLže,14; GVbp
 zavzeti · Gce,2; GLme,6; GLže,5
 zazdeti se · Gce; GLmp; GLse,30; GLže,6;
 GLžp,3
 zazebsti · Gce; GLse,12
 zazibati · Gce,4; GLme,15; GLmp; GLse,3;
 GLže,4
 zazijati · Gce,7; GLme,6; GLmp,2; GLže,10;
 GLžp
 zazveneti · Gce; GLme,3; GLse; GLsp; GLže,3;
 GLžp
 zazvoniti · GLme,3; GLmp,3; GLse,7
 zažariti · Gce,6; Gcp,4
 zažgati · Gcp; Gcp; GLme,3; GLmp,6
 zaživeti · Gce,6; GLmd; GLme,6; GLmp,2;
 GLse; GLže,10; GLžp,3; GNE,2
 zažvižgati · Gce,3; Gcp; GLmd; GLme,13;
 GLmp,2; GLse,2; GLžd; GLže,4
 zbegano · A,12
 zbežati · Gce; Gcp; GLme,18; GLžd; GLže,6;
 GNE,2
 zbirati · Gce,7; Gcp,6; GLme,7; GLmp,7;
 GLže,4; GLžp; GNE,4
 zbogom · M,18
 zbosti · Gce,2; GLme,2; GLmp; GLse; GLže;
 GLžp,4; GNE
 zbran · GNme1; GNmp1,5; GNže1,5; PNžp2
 zbrati · Gae; Gap,2; Gce,5; Gcp,2; GLmd;
 GLme,12; GLmp,18; GLse,3; GLže,13; GNE,5;
 GVbe
 zbuditi · Gce; Gce,6; GLme,53; GLmp,6;
 GLse,3; GLže,11; GLžp
 zdaj · A,1350
 zdaljšati · Gce,2; GLme,5; GLmp,3; GLsp; GLžp
 zdavnaj · A,61
 zdeti se · Gae; Gbe; Gce,51; GLme,14; GLmp;
 GLse,73; GLže,19; GLžp,7; GNE

zdič · Sme2,3; Sme4,9; Sme5,8
 zdramiti · Gae; Gcd,3; Gce,7; GLme,25;
 GLmp,3; GLse,2; GLže,14
 zdrav · Pmd1; Pme1,31; Pme2,2; Pme3; Pme4;
 Pme4; Pme6; Pmp1,6; Pmp2; Pse1,3; Pse2,4;
 Pse4; Pse5,2; Pže1,8; Pže3; Pže4,2; Pže5;
 Pžp1,2; Pžp2; Pžp4; Pžp6,2
 zdravje · Sse1,6; Sse2,6; Sse4,5; Sse6
 zdravnik · Sme1,14; Sme2,3; Sme3,6; Sme4,3;
 Smp1,5
 zdrkniti · Gce,3; GLme,13; GLmp,4; GLse,3;
 GLže,6; GLžp
 zdrobiti · Gae; Gce,2; GLme,2; GLmp; GLse;
 GLže,2; GLžp
 zdrsni · Gce,2; GLme,5; GLse; GLže,3
 zdrveti · Gcd; Gce,3; Gcp; GLme,3; GLmp;
 GLse
 zdrzniti se · Gae,2; Gcd; Gce,15; Gcp,5;
 GLme,52; GLmp; GLže,13
 zdržati · Gae; Gce,2; Gcp; GLme,14; GLse;
 GLže,2; GNE,2
 zebsti · Gce,8; GLse,3
 zelen · Pme1i,2; Pme1,5; Pme2,2; Pme3;
 Pme4i,3; Pme5,6; Pme6,3; Pmp1,3; Pmp4;
 Pse1,9; Pse2,19; Pse4,3; Pse5,12; Pse6,3;
 Psp2; Pže1,20; Pže2,10; Pže4,12; Pže5,23;
 Pže6,7; Pžp1,5; Pžp2,4; Pžp4; Pžp5,3; Pžp6
 zelenika · Sže2,11; Sže5,2
 zelenje · Sse1; Sse2; Sse4; Sse6,7; Sže2,2
 zelo · A,303
 zemlja · Sže1,43; Sže2,55; Sže3,6; Sže4,93;
 Sže5,59; Sže6,6
 zganiti se · Gce,16; Gcp,6; GLmd; GLme,29;
 GLmp,5; GLse,5; GLže,15; GLžp,2; GNE
 zgodaj · A,13
 zgodba · Sže1,13; Sže2,8; Sže4,25; Sže6,2;
 Sžp1,3; Sžp2,6; Sžp3; Sžp4,10; Sžp5
 zgoditi se · Gce,20; GLme,2; GLse,59; GLže,2;
 GNE
 zgodnji · Pme1; Pse1; Pže1,2; Pže2; Pže5,5;
 Pžp5
 zgodovina · Sže2,4; Sže4,4; Sže5,3
 zgoreti · Gae; Gbe; Gce,3; GLmp,5; GLže,7
 zgornji · Pme1,2; Pme5,5; Psp2; Psp5; Pže1,2;
 Pže4,2; Pže5,3; Pžp1
 zgrabiti · Gcd,2; Gce,59; Gcp,4; GLmd,2;
 GLme,64; GLmp,9; GLse,2; GLže,23; GLžp,3;
 GVbe,2; GVbp,2
 zgroziti se · Gce,2; GLme,10; GLže,5
 zgruditi se · Gce,4; GLme,13; GLse; GLže,6
 zgubiti · Gce; GLme,7; GLže,2

- zibati · Gcd; Gce,17; Gcp,6; GLme,22; GLmp,2;
GLse,4; GLže,17; GLžp,6; GNE,4
- zibelka · Sže1,4; Sže4,3; Sže5,2; Sže6
- zid · Sme1,3; Sme2,3; Sme2,9; Sme3,3; Sme4,23;
Sme5,21; Sme6,4; Smp1,12; Smp2,3; Smp4,2;
Smp5; Smp6,3
- zidanica · Sže1,4; Sže2,3; Sže4,2; Sže5; Sže6,2;
Sžp1,2
- zidar · Sme1,7; Sme4; Smp3; Smp4,2
- zidati · Gap; Gcp,3; GLme,3; GLmp,4; GNE,3
- zijati · Gbe,3; Gbp; Gcd; Gce; Gcp; GLme,4;
GLmp,2; GLsp
- zima · Sže1,20; Sže2,4; Sže4,18; Sže5,6; Sžp1,2;
Sžp2; Sžp4,2
- zimski · Pme1,2; Pme4; Pme5,3; Pmp1;
Pmp5,3; Pse1,4; Pse2; Pse4,3; Pse5,2; Pže1,2;
Pže2; Pže4,3; Pžp4
- ziniti · Gce,5; GLme,6; GLže; GVbe
- Zinka · IOže1,66; IOže2,5; IOže3,2; IOže4,3;
IOže5; IOže6,9
- zjokati se · Gae,2; Gce; GLmd; GLme,4; GLmp;
GLže,3
- zjutraj · A,41
- zlagati · Gae; Gce; GLme,7; GLže,2; GNE,4
- zlat · Pme1i; Pme1,6; Pme2; Pme4i; Pme4,2;
Pme5,3; Pme6; Pmp1,8; Pmp2; Pmp4,11;
Pmp6,4; Pse1,11; Pse2,20; Pse4,7; Pse5,2;
Pse6,5; Psp1; Psp2; Psp4; Pzd4; Pže1,14;
Pže2,8; Pže4,6; Pže5,3; Pže6,9; Pžp1,6;
Pžp2,2; Pžp3; Pžp4; Pžp6,3
- zlato · A,4; Sse1,6; Sse2,5; Sse4; Sse6,2
- zlecni se · Gce,2; GLme,20; GLse; GLže,3
- zlekniti se · Gae; Gce,3; Gcp,5; GLme,19;
GLmp; GLse,2; GLže,5; GNE
- zlesti · Gce,3; Gcp; GLme,14; GLmp; GLse,3;
GLže,2; GLžp,2
- zliti · Gce,2; GLme,4; GLmp,2; GLse; GLže
- zlobno · Aj; A,21
- zlomiti · Gae; Gce,2; GLme,7; GLmp,2; GLže,6;
GNE,3
- Zmaga · IOže1,8; IOže4,5; IOže5; IOže6; ISže1,5
- zmaga · Sže1,3; Sže2,4; Sže4; Sže5; Sžp2; Sžp3
- zmagati · GLme,4; GLmp,6; GNE
- zmagoslavno · A,12
- zmajati · Gae,2; Gcd; Gce,25; GLme,80;
GLmp,5; GLže,27; GNE
- zmanjkati · GLmp; GLse,10
- zmazati · Gae,2; GLme,13; GNE
- zmeniti se · Gbe; Gce,9; Gcp,2; GLmd,2;
GLme,17; GLmp,5; GLže,7; GNE
- zmeraj · A,222
- zmerjati · Gce; Gcp,3; GLmd,2; GLme,8;
GLmp,5; GLže,11; GLžp; GNE,5
- zmerom · A,61
- zmešati · Gce; GLse,9; GLže,2
- zmigniti · Gce,6; Gcp; GLme,29; GLže,7; GNE
- zmpačiti se · Gce,2; GLme; GLse,7
- znajti se · Gce,5; Gcp; GLmd,2; GLme,16;
GLmp; GLse; GLže,3; GNE,2
- znak · Sme1,2; Sme2,2; Sme4,9; Smp1
- znamenje · Sse1,6; Sse2,3; Sse4,4; Ssp1,2;
Ssp4,4
- znan · GNse1; Pmd1,3; Pmeli,4; Pme1,7;
Pme4i; Pme6,2; Pmp1,2; Pmp4; PNsp2;
Pse1,6; Pse4; Pže1,16; Pže2; Pže4; Pže6;
Pžp1,2
- znanec · Smd1,2; Sme1,2; Sme2,3; Sme3; Sme4;
Sme6; Smp1,3; Smp4,3; Smp6
- znati · Gae,6; Gbe,3; Gbp; Gcd; Gce,16; Gcp,7;
GLmd,3; GLme,23; GLmp,3; GLse; GLže,8;
GNE,4
- znoj · Sme1,9; Sme2,5; Sme4,4; Sme5,2; Sme6
- znojen · Pme1i; Pmp4; Pse1; Pse4,3; Pse5; Psp4;
Pže1; Pže4; Pžp1,4; Pžp4,7
- znova · A,38
- zob · Sme4; Smp1,4; Smp2,7; Smp4,39; Smp5;
Smp6,13
- zoper · E4,10
- zopet · A,14
- zoprn · Pme1i; Pme1,4; Pme4; Pme6,4; Pse1;
Pže1,4; Pže4,3; Pže5; Pže6
- Zora · IOže1,8; IOže3; IOže4; IOže6,2
- zora · Sže1,4; Sže2,15; Sže3; Sže4,3; Sže5,12;
Sže6,5
- zrak · Sme1,19; Sme2,9; Sme4,24; Sme5,54;
Sme6,3
- zrasti · Gce,4; Gcp; GLme,2; GLme,24; GLmp,5;
GLse; GLsp; GLže,12; GLžp,4
- zraven · A,19; E2,29
- zravnati · Gae; Gce,31; GLme,72; GLmp,4;
GLse,2; GLže,27; GNE; GVbe,2; GVbp
- zrcalo · Sse1,4; Sse2,4; Sse5,9; Sse6,3
- zrel · Pmd1; Pmp1; Pse1,3; Pse2; Pže1,3;
Pže2,3; Pže3; Pžp1,2; Pžp2
- zresniti · Gce,11; Gcp,2; GLme,8; GLže,2
- zreti · Gae; Gap,5; Gce; Gcp; Gcp,6; GLme;
GLmp; GLžp
- zrnje · Sse1,2; Sse2,5; Sse4,2; Sse6
- zrno · Sse1,3; Sse2; Sse4,2; Sse6; Ssp1; Ssp2,2
- zrogoviliti · Gce,3; Gcp; GLme,18; GLže
- zrušiti · Gce,3; Gcp,3; GLme,19; GLmp; GLse,4;
GLže,13; GLžp; GNE,2

- zunaj · A,51; E2
 zunanji · Pme4,2; Pme5; Pme6; Pse2; Pse4;
 Pže5,3; Pžp1; Pžp2
 zvečer · A,52
 zvenec · PČme1,5; PČme4i,2; PČme6; PČmp2;
 PČsp6
 zveneti · Gce,3; GLme,7; GLse,3; GLže,6;
 GLžp,2
 zver · Sže1,7; Sže2; Sže3; Sže4; Sže5,5
 zverina · Sže1,11; Sže4; Sže6; Sžp1,9; Sžp4
 zverizen · PNme1; PNme2,3; PNme5,2; PNsp1;
 PNže1; PNže4; PNžp4
 zveza · Sže1,4; Sže2,2; Sže4,2; Sže5,5
 zvezda · Sže1,2; Sže2,3; Sže4,7; Sže6; Sžp1,2;
 Sžp2,2; Sžp3; Sžp5,4; Sžp6,2
 zvijski · Gce,10; Gcp; GLme,9; GLmp,2; GLže,4;
 GNE,3
 zvišati · Gae; GLme,9; GLmp,2
 zviška · A,47
 zviti · Gce,2; GLme,20; GLmp; GLse; GLžp;
 GNE
 zvok · Sme1,13; Sme2,3; Sme4,3; Sme5,3;
 Smp1,9; Smp2; Smp4; Smp5; Smp6
 zvon · Sme1,22; Sme2; Sme4,3; Sme6; Smp1,21;
 Smp4,2; Smp5,2
 zvonec · Sme1,4; Sme2; Sme4,5; Sme5; Smp1,15;
 Smp2,3; Smp4,26; Smp6,3
 zvonik · Sme1,3; Sme2,7; Sme4; Sme5,8; Sme6;
 Smp5; Smp6,2
 zvoniti · Gae,3; Gce,23; Gcp; GLme,18;
 GLmp,10; GLse,4; GLže; GNA; GNE,7;
 GVbe,2
 zvrniti · Gad; Gce,2; GLmd; GLme,13; GLmp,2;
 GLže,5; GVad
 ž · E6,10
 žaga · Sže1,15; Sže2,3; Sže3,2; Sže4,15; Sže5;
 Sže6,2
 žagar · Smd1,6; Smd4; Sme1,2; Sme3
 žakaj · Sme2; ZV,12
 žal · P; PD,24
 žalost · Sže1,18; Sže2,20; Sže3; Sže4,6; Sže5,6;
 Sže6,4
 žalosten · Pme1,15; Pme2,3; Pme3; Pme4;
 Pme4i; Pme5; Pme6,2; Pmp1,3; Pmp3; Pmp5;
 Pmp6; Pse1,8; Psp2; Pzd1; Pže1,17; Pže2,6;
 Pže4; Pže5; Pže6,2; Pžp1,4; Pžp2,4; Pžp5
 žalostno · A,40
 žandar · Sme1,2; Sme3; Sme4,3; Sme6; Smp1,3;
 Smp4
 žareč · GČme1,3; PČme3; PČme4i,3; PČme6,2;
 PČse1,2; PČse2; PČže6; PČžp2; PČžp6,4
 žarek · Pse5; Sme1,6; Sme4,3; Sme5; Smp1,9;
 Smp2,9; Smp4,3; Smp5; Smp6
 žareti · GLme,8; GLse; GLsp,2; GLže,6; GLžp
 žariti · Gce,11; Gcp,12
 žditi · Gce,3; Gcp; GLme,4; GLmp,2
 že · Č,1828
 žebelj · Sme2; Sme4,5; Sme5; Sme6,2; Smp1,2;
 Smp6,2
 Žef · IOme1,150; IOme2,3; IOme3,8; IOme4,15;
 IOme5,6; IOme6
 žejen · Pmd1,3; Pme1,7; Pmp1,3; Pse1; Pže1,3
 želeti · Gae; Gap; Gbp; Gce; GLme,9; GLže,2;
 GVbe
 železen · Pmeli; Pme1,2; Pme2,3; Pme4i,2;
 Pme5; Pmp2; Pmp4,2; Pse1; Psp1; Psp4;
 Pže1,2; Pže2,2; Pže4,5; Pže6; Pžp1; Pžp2,3;
 Pžp4,4
 železo · Sse1,2; Sse2,5; Sse4,6
 želja · Sže1,10; Sže2; Sže4,4; Sžp1,3; Sžp2;
 Sžp2,2; Sžp4,4; Sžp5,2
 želodec · Sme1; Sme2; Sme4; Sme5,8; Smp4
 žena · Sže1,131; Sže2,11; Sže3,20; Sže4,41;
 Sže6,3; Sžp1,10; Sžp2,3; Sžp3; Sžp4,3; Sžp5;
 Sžp6,2
 ženica · Sžd1; Sže1,12; Sže3; Sžp1,6; Sžp3; Sžp4
 ženin · Pme1; Pže5; Pžp1; Pžp3; Sme1,8; Sme4;
 Sme6,2; Smp1; Smp2
 ženiti · Gbe; Gce; GLme,8; GLmp; GVbe,3;
 GVbp,2
 ženska · Sžd1,6; Sžd4; Sže1,45; Sže2,16; Sže3;
 Sže4,14; Sže6,4; Sžp1,59; Sžp2,10; Sžp3,6;
 Sžp4,13; Sžp5; Sžp6,7
 ženski · Pme1,9; Pme2,2; Pme4,3; Pme6,3;
 Pmp6,2; Pse1; Pse2; Pse5; Psp2; Pže1,2;
 Pže2,2; Pže4,2; Pžp2; Pžp6
 žep · Sme2,35; Sme4,46; Sme5,13; Smp1;
 Smp2,2; Smp4,21; Smp5,12
 žepar · Sme1,7; Sme2,2; Sme3; Sme4; Smp4;
 Smp5,2; Smp6
 žganje · Sse1,8; Sse2,32; Sse4,18; Sse5; Sse6,2
 žgati · Gce,8; Gcp; Gcp; GLme,3; GLse,2; GLže
 žgoč · GČže4; GČže5; PČme4i,2; PČme5,4;
 PČse2,2; PČse5,2; PČze1; PČže4; PČžp4,2;
 Psp1
 žila · Sže1,4; Sžp1,6; Sžp2; Sžp4; Sžp5,10; Sžp6
 žito · Sse1,6; Sse4; Sse5; Ssp2; Ssp4
 živ · Pmd1; Pme1i,4; Pme1,70; Pme2,7;
 Pme4i,3; Pme4,3; Pme4,4; Pme5,2; Pme6,5;
 Pmp1,15; Pmp2,3; Pmp4,2; Pmp5; Pmp6;
 Psd1,2; Pse1,18; Pse2,7; Pse3; Pse4; Pse5,7;
 Psp1; Pžd1; Pžd4; Pže1,62; Pže2,17; Pže3;
 Pže4,11; Pže5,6; Pžp1,19; Pžp3; Pžp4,8; Pžp5;
 Pžp6,8

živahno · Aj; A,10
 žival · Sže1,21; Sže2,7; Sže4,2; Sžp1,10; Sžp2,2;
 Sžp3,2; Sžp4,3; Sžp5,2
 živeti · Gcp,2; GLmd,9; GLme,63; GLmp,20;
 GLse,3; GLže,34; GLžp,4; GNE,41
 živina · Sže1,6; Sže2,2; Sže3,2; Sže4,6; Sže5;
 Sže6,3
 živiti · Gae,7; Gap,7; Gbe,5; Gbp; Gcd; Gce,43;
 Gcp,12
 življenje · Sse1,87; Sse2,93; Sse3,8; Sse4,102;
 Sse5,66; Sse6,16; Ssp2
 življenjski · Pme4; Pse1; Pse2; Pže1,3; Pže2,2;
 Pže4,2; Pže5; Pžp4; Pžp4
 živo · A,43; Sse2,8; Sse4; Sse4,9
 žleb · Sme2; Sme3,3; Sme4,5; Sme5,9; Smp4;
 Smp5
 žlica · Szd4,2; Sže2,3; Sže4,15; Sže5,2; Sže6,8;
 Sžp1,2; Sžp4,3; Sžp6
 žreti · Gbp; Gcd; Gce,8; Gcp,2; GLme; GLmp,4;
 GLže

žrtev · Sže1; Sže2,2; Sže4,3; Sžp1; Sžp2,2; Sžp4;
 Sžp6
 žulj · Smp1,4; Smp2,3; Smp3; Smp4; Smp6,2
 žuljav · Pmp2; Pmp4; Pmp6; Pže1,2; Pže4,3;
 Pže5; Pže6,2; Pžp1,2; Pžp2; Pžp4,4; Pžp6
 župan · Sme1,10; Sme3; Sme4
 župnik · Sme1,181; Sme2,11; Sme3,28;
 Sme4,23; Sme5,2; Sme6,6; Smp1,18; Smp3,3;
 Smp4; Smp5,2; Smp6
 župnikov · Pme2; Pme4; Pmp6; Pse1; Pse2;
 Pže1,3; Pže2; Pže4,3; Pže5; Pžp1; Pžp4; Pžp6
 župnišče · Sse2,12; Sse4,9; Sse5,6; Sse6,3; Ssp5
 Žuželjč · IOme1,24; IOme2,3; IOme3; IOme4,2;
 IOme6
 Žuželjčevka · IOže1,25; IOže3,3; IOže4,4
 žvečiti · Gce,2; Gcp; GLme,13; GNE,2
 žvižg · Sme1,8; Sme4; Sme6
 žvižgati · Gce,7; Gcp,9; GLme,10; GLmp,6;
 GLse; GLže,5; GLžp,2; GNE

Priloga D

Zanimivi naslovi na internetu

<http://www.ff.uni-lj.si/hp/pj/disertacija>

Doktorska disertacija: *Zgornja meja entropije pri leposlovnih besedilih v slovenskem jeziku*

<http://bos.zrc-sazu.si>

Besedilni korpus *Nova beseda*, tudi z besedili prvega in drugega vzorca

<http://bos.zrc-sazu.si/ckb.html>

Priloga Č brez frekvenčne omejitve: seznam vseh lem drugega vzorca z oblikoslovnimi oznakami in frekvencami

<http://www.ijs.si/lit/leposl.html>

Zbirka slovenskih leposlovnih besedil (M. Hladnik, od junija 1995)

<http://www.omnibus.se/beseda>

Beseda, zbirka slovenskih elektronskih knjig v PDF formatu, ki jo v Stockholmu pripravlja Franko Luin

<http://www.fida.net/>

Predstavitvena stran projekta FIDA (Filozofska fakulteta Univ. v Ljubljani, Inštitut Jožef Štefan, založba DZS, podjetje Amebis, od septembra 1998).

<http://www.ids-mannheim.de/ldv/cosmas/corpora-ges.html>

Zbirka nemških besedil na Inštitutu za nemški jezik v Mannheimu (IDS)

<http://frantext.inalf.fr/frantext.htm>

Zbirka francoskih besedil FRANTEXT

<http://visl.hum.ou.dk/Linguistics.html>

Avtomatični označevalnik za portugalščino in esperanto na Univerzi v Aarhusu

<http://www.laze.org/eva>

Urejevalnik EVA

<http://nl.ijs.si/ME/>

Predstavitvena stran projekta MULTEXT-East

<http://www.icp.inpg.fr/ELRA/services/validat.php3>

Evropska agencija za jezikovne vire – ELRA – projekt za vrednotenje jezikovnih virov

<http://www.tekstlab.uio.no/Bosnian/Corpus.html>

Zbirka bosanskih besedil na Univerzi v Oslu

<http://language.perl.com/>

Predstavitvena stran jezika PERL

<http://info.ox.ac.uk/bnc/>

Projekt British National Corpus

<http://americannationalcorpus.org>

Projekt American National Corpus

<http://ucnk.ff.cuni.cz/verejny.html>

Iskalna stran Češkega nacionalnega korpusa

<http://www.rasip.fer.hr/research/compress>

Stran referenčnega centra za komprimiranje podatkov na Univerzi v Zagrebu

<http://www.ross.net/compression/introduction.html>

Zelo zanimiva, komprimiranju podatkov posvečena stran, ki jo je pripravil dr. Ross N. Williams

Internetni naslovi javno dostopnih komprimiranih programov (gl. 6.3.6):

Boa	ftp://ftp.elf.stuba.sk/pub/pc/pack/boa058.zip
RKIVE	http://www.geocities.com/SiliconValley/Peaks/9463/rkv192b1.zip
Ufa	ftp://ftp.simtel.net/pub/simtelnet/win95/compress/ufa004b1.zip
Arhangel	http://www.geocities.com/SiliconValley/Lab/6606/arh138.zip
UHARC	ftp://ftp.elf.stuba.sk/pub/pc/pack/uaharc02.zip
ARJ in JAR	http://www.arjsoft.com
RAR	http://www.rarsoft.com
PKZIP	ftp://ftp.simtel.net/pub/simtelnet/msdos/arcers/pkz204g.exe
ESP	ftp://ftp.elf.stuba.sk/pub/pc/pack/esp_v192.exe

Abstract

ENTROPY OF SLOVENIAN LITERARY TEXTS

Quantitative description of two samples of literary texts in Slovenian is presented. From the most important parameters of character n-grams, from single characters to 24-grams, the upper bound of entropy is estimated.

The second chapter, which follows the introduction with description of the main contributions of this work, is dedicated to the sources of both samples. The first sample contains 60 texts of 41 authors, from Ivan Cankar to Ivan Zorec, 46 original works and 14 translations, published between 1858 and 1996, 17 million characters, 2,7 million words and 200.000 sentences in all. The second sample includes the complete opus of the great Slovenian stylist from the middle of the century, Ciril Kosmač (1910–1980), 52 works published between 1931 and 1988, 2,5 million characters, 408.000 words and 37.000 sentences. Both samples together are estimated at 0,5 % – 1 % of the complete Slovenian literary production.

The third chapter is devoted to preparation of texts. It begins with text headers and additional symbols, introduced to facilitate the task of tagging the text components to the level of sentences, including direct speech and quotes, and to enable an automatic conversion of texts from a single file to a set of HTML files for Internet. The second part of the chapter deals with the question of errors. It describes the procedure which generated a collection of 3,5 million wordforms from 92.000 verb, noun and adjective lemmata from the *Dictionary of the Slovenian Literary Language*. The collection has been used during the cleanup of texts from both samples.

Part-of-speech tagging is described in chapter 4. The second sample has been POS tagged, with a stochastic tagger built for the purpose. Tagged texts were verified and corrected at the Institute for the Slovenian Language (part of Scientific Research Centre of the Slovenian Academy of Sciences and Arts); so the precision of the tagger could be measured directly and it turned out to be 92 %. The most frequent lemmata for the 5 main word classes of the second sample are given in table 1. Percentage values which follow the words were computed

with regard to the total number of words (407.938): 9,11 % equals the frequency of 37.163 and 0,06 % the frequency of 245. The verb *biti* (*to be*) has been divided into 3 components. *Biti p* denotes an auxiliary verb (as in *John is moving the table.*), *biti r* the verb *to be* in relational sense (as in *You are beautiful.*) and *biti o* in the most narrow sense of existence in time or space (as in *There is a house.*)

Table 1: The most common lemmata for the 5 word classes in sample 2 (frequencies in %)

verb	noun	adjective	pronoun	adverb	
1. <i>biti p</i>	9,11	<i>roka</i> 0,39	<i>star</i> 0,19	<i>on</i> 1,60	<i>tako</i> 0,39
2. <i>biti r</i>	1,71	<i>glava</i> 0,24	<i>velik</i> 0,12	<i>ki</i> 0,73	<i>zdaj</i> 0,33
3. <i>ne biti p</i>	0,65	<i>oči</i> 0,20	<i>lep</i> 0,11	<i>jaz</i> 0,70	<i>nato</i> 0,24
4. <i>reči</i>	0,47	<i>otrok</i> 0,20	<i>dolg</i> 0,09	<i>ta</i> 0,69	<i>spet</i> 0,21
5. <i>biti o</i>	0,44	<i>dan</i> 0,18	<i>črn</i> 0,09	<i>ona</i> 0,66	<i>potem</i> 0,18
6. <i>imeti</i>	0,23	<i>hiša</i> 0,17	<i>bel</i> 0,09	<i>svoj</i> 0,55	<i>počasi</i> 0,15
7. <i>vedeti</i>	0,23	<i>leto</i> 0,15	<i>dober</i> 0,08	<i>ves</i> 0,39	<i>lahko</i> 0,13
8. <i>videti</i>	0,20	<i>vrata</i> 0,13	<i>živ</i> 0,07	<i>ti</i> 0,36	<i>takoj</i> 0,11
9. <i>iti</i>	0,19	<i>beseda</i> 0,13	<i>mlad</i> 0,07	<i>vse</i> 0,29	<i>skoraj</i> 0,10
10. <i>stopiti</i>	0,18	<i>oče</i> 0,12	<i>težek</i> 0,06	<i>oni</i> 0,28	<i>bolj</i> 0,10
11. <i>začeti</i>	0,17	<i>človek</i> 0,11	<i>širok</i> 0,06	<i>sam</i> 0,22	<i>naglo</i> 0,09
12. <i>pogledati</i>	0,16	<i>glas</i> 0,11	<i>hud</i> 0,06	<i>kako</i> 0,20	<i>dobro</i> 0,08

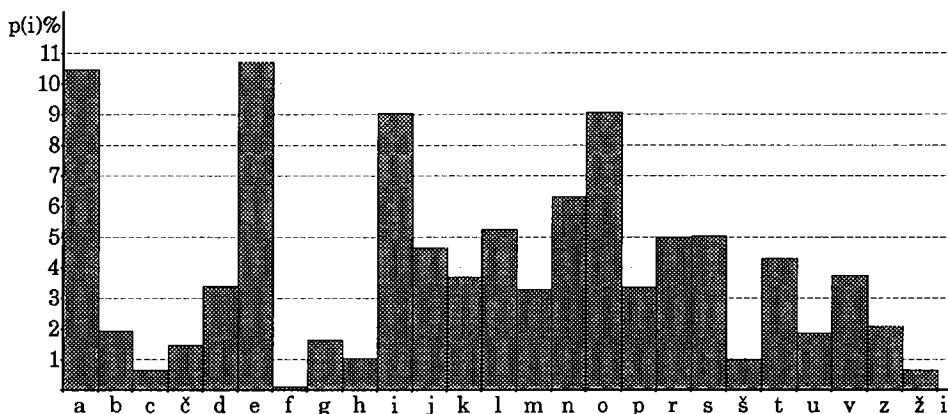
Very approximate English translations are given in table 2. The pronoun *svoj*, marked by an asterisk, is impossible to translate into one word: it means *pertaining to oneself*.

Table 2: Approximate English translations of the words in table 1

verb	noun	adjective	pronoun	adverb	
1. <i>biti p</i>	9,11	<i>hand</i> 0,39	<i>old</i> 0,19	<i>he</i> 1,60	<i>so</i> 0,39
2. <i>biti r</i>	1,71	<i>head</i> 0,24	<i>big</i> 0,12	<i>which</i> 0,73	<i>now</i> 0,33
3. <i>ne biti p</i>	0,65	<i>eyes</i> 0,20	<i>beautiful</i> 0,11	<i>I</i> 0,70	<i>then</i> 0,24
4. <i>to say</i>	0,47	<i>child</i> 0,20	<i>long</i> 0,09	<i>this</i> 0,69	<i>again</i> 0,21
5. <i>biti o</i>	0,44	<i>day</i> 0,18	<i>black</i> 0,09	<i>she</i> 0,66	<i>afterwards</i> 0,18
6. <i>to have</i>	0,23	<i>house</i> 0,17	<i>white</i> 0,09	<i>svoj*</i> 0,55	<i>slowly</i> 0,15
7. <i>to know</i>	0,23	<i>year</i> 0,15	<i>good</i> 0,08	<i>total</i> 0,39	<i>easily</i> 0,13
8. <i>to see</i>	0,20	<i>door</i> 0,13	<i>alive</i> 0,07	<i>you</i> 0,36	<i>at once</i> 0,11
9. <i>to go</i>	0,19	<i>word</i> 0,13	<i>young</i> 0,07	<i>all</i> 0,29	<i>nearly</i> 0,10
10. <i>to step</i>	0,18	<i>father</i> 0,12	<i>heavy</i> 0,06	<i>they</i> 0,28	<i>more</i> 0,10
11. <i>to begin</i>	0,17	<i>man</i> 0,11	<i>wide</i> 0,06	<i>alone</i> 0,22	<i>fast</i> 0,09
12. <i>to look at</i>	0,16	<i>voice</i> 0,11	<i>angry</i> 0,06	<i>how</i> 0,20	<i>well</i> 0,08

Very interesting are the distributions of gender, number and case in the POS tags of the relevant words. The ratios for gender are 55 % (masculine), 33 % (feminine) and 12 % (neuter). For the number the figures are 65 % for singular, 33 % for plural and 2 % for dual. The distribution of cases is 28 % for nominative, 13 % for genitive, 6 % for dative, 27 % for accusative, 15 % for locative and 11 % for instrumental.

Statistical description of both samples is the subject of chapter 5. The common character set is given (168 characters), as well as character sets and letter distributions for both samples. The distribution of the most common 25 letters is shown in figure 1. The vowels are most common with *e* (7,92 % of all characters) on top, followed by *a* (7,74 % of the total). Character n-grams follow and the most common for $n = 1$ to $n = 14$ are given for both samples. Word statistics comes next: the average word length for both samples is 4,55 letters and the 12



most common wordforms are: *je* 'is', *in* 'and', *se* reflexive personal pronoun and free morpheme of reflexive verbs, *v* 'in', *da* subordinator 'that', *na* 'on', *so* 'are' (3rd person), *ne* the particles 'not' and 'no', *pa* 'but', *ki* relative particle, *bi* particle of the conditional and *z* 'with'. The longest words in both samples as well as in the *Bank of English* are given, followed by the most common word n-grams. Description of sentences with sentence lengths and the most common sentences (top 7 in sample 1: *Yes. So it is. By all means. Of course. For sure. No. How?*) come at the end of the chapter.

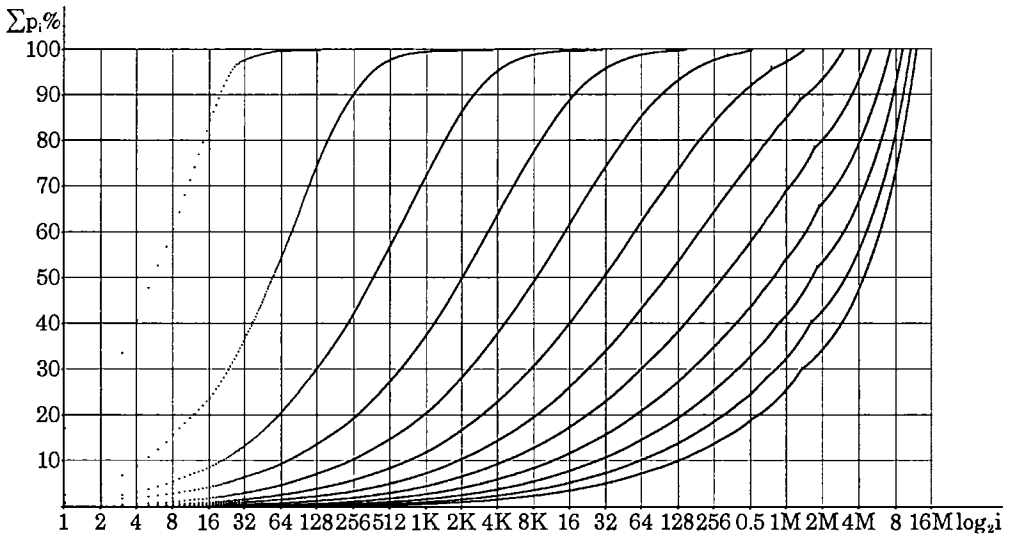
Table 3: Entropy of character n-grams in the first sample

n	H	H _n	F _n	n	H	H _n	F _n
1	4,456	4,456	4,456	13	23,460	1,805	0,239
2	7,994	3,997	3,538	14	23,615	1,687	0,155
3	11,020	3,673	3,026	15	23,715	1,581	0,100
4	13,565	3,391	2,545	16	23,779	1,486	0,064
5	15,739	3,148	2,174	17	23,821	1,401	0,042
6	17,643	2,941	1,904	18	23,848	1,325	0,027
7	19,272	2,753	1,629	19	23,866	1,256	0,018
8	20,587	2,573	1,315	20	23,878	1,194	0,012
9	21,594	2,399	1,007	21	23,886	1,137	0,008
10	22,334	2,233	0,740	22	23,891	1,086	0,005
11	22,861	2,078	0,527	23	23,895	1,039	0,004
12	23,221	1,935	0,360	24	23,898	0,996	0,003

Entropy is the main topic of chapter 6. An algorithm is given, which computes entropies up to $n = 62$ and for sample sizes which can be an order of magnitude bigger than the samples of the thesis. Its achievements are based on special management of the quickly expanding database of Hapax Legomena. Total entropy H , entropy per character H_n , and conditional entropies F_n , are given for both samples, to $n = 24$ (table 3). At that stage the conditional entropy, i. e. the entropy of n -th character if the previous $(n-1)$ are known, falls below 0,005 bits per character in both samples. From the values of H_n and from the flow of new Hapax Legomena from n to $(n+1)$ the upper bound of entropy in Slovenian literary texts is estimated at 2,2 bits per character. A language model, based on probabilities of n -grams from the first sample which achieved the frequency at least 2, is described. The mapping of the text of the second sample, based on the model, achieved an average length of 2,7 bits per character.

How fast the number of different n -grams, if sorted by frequency in descending order (the most frequent first) fills up the entire set in sample 1 is shown in figure 2. The growth curves for n -grams from 1 to 12 follow one another from the left to the right. From the figure it can be seen that to attain half of the entire text sample it is necessary to have 6 most frequent characters (*Space, e, a, i, o, n*), 50 bigrams, 300 trigrams 2.000 4-grams, 8.000 5-grams, 32.000 6-grams, 80.000 7-grams, 300.000 8-grams, 800.000 9-grams and 1.500.000 10-grams.

Figure 2: Growth curves for character n -grams (1-12) in the first sample



Conclusions as well as the possible follow-up research in the field of computational analysis of Slovenian literary corpora are discussed in the final, seventh chapter.

Linguistica et philologica

ISBN 961-6358-77-4



9 789616 358774

Linguistica et philologica

2001

Andreja Žele

Vezljivost v slovenskem jeziku

2000

Jerneja Gros

Samodejno tvorjenje govora iz besedil.

*Postopek za izdelavo sintetizatorja
slovenskega govora*

1999

Alenka Šivic Dular

*Besedna družina iz korena *god- v
slovanskih jezikih. Pomenoslovna
razčlemba v kulturološkem kontekstu*

1998

Varja Cvetko Orešnik

*K metodologiji preučevanja baltoslovansko-
indoiranskih jezikovnih odnosov. Prvi del.*

ISBN 961-6358-77-4



9 789616 358774