

Primož Jakopin

Inštitut za slovenski jezik Frana Ramovša ZRC SAZU
Korpusni laboratorij

Oblikoslovni označevalnik pri korpusu Beseda

JezikovnO Tehnološki Abonma, četrtek, 13. marca 2008

Oblikoslovno označevanje

(Primož Jakopin in Aleksandra Bizjak)

- nabor oblikoslovnih oznak
- ročni označevalnik
- strojni označevalnik

Nabor oblikoslovnih oznak

Besedne vrste

samostalniška beseda

glagolska beseda

pridevniška beseda

zaimki

števnički

prislovi in ostale besedne vrste

Oblikoslovne kategorije

spol	število	sklon	oseba	stopnjevanje	določnost	razmerje
m moški	e ednina	1 imenovalnik	a prva	j primernik	i dol. oblika	pr priredno
ž ženski	d dvojina	2 rodilnik	b druga	jj presežnik		po podredno
s srednji	p množina	3 dajalnik	c tretja			
		4 tožilnik				
		5 mestnik				
		6 orodnik				

samostalniška beseda

	oznaka	spol	število	sklon	primer	
samostalnik	S	{m,ž,s}	{e,d,p}	{1,2,3,4,5,6}	Sme1	(dan)
glagolnik	SG	{m,ž,s}	{e,d,p}	{1,2,3,4,5,6}	SGse2	(spoznanja)
osebna imena	IO	{m,ž,s}	{e,d,p}	{1,2,3,4,5,6}	IOme1	(Martin)
imena prebivalcev	IP	{m,ž,s}	{e,d,p}	{1,2,3,4,5,6}	IPme2	(Čeha)
veroslovna imena	IV	{m,ž,s}	{e,d,p}	{1,2,3,4,5,6}	IVme3	(Bogu)
živalska imena	IŽ	{m,ž,s}	{e,d,p}	{1,2,3,4,5,6}	IŽže3	(Liski)
zemljepisna imena	IZ	{m,ž,s}	{e,d,p}	{1,2,3,4,5,6}	IZme2	(Črnomlja)
mitološka imena	IM	{m,ž,s}	{e,d,p}	{1,2,3,4,5,6}	IMme5	(Hadu)
stvarna imena	IS	{m,ž,s}	{e,d,p}	{1,2,3,4,5,6}	ISže2	(Iliade)

glagolska beseda

	oznaka	oseba	spol	število	sklon	primer
povedni sedanjik	G	{a,b,c}	{e,d,p}		Gce	(plava)
pom. gl. biti sed.	GP	{a,b,c}	{e,d,p}		GPce	(je)
pom. gl. biti prih.	GFP	{a,b,c}	{e,d,p}		GFPcp	(bodo)
pom. gl. biti nik. obl.	GZP	{a,b,c}	{e,d,p}		GZPae	(nisem)
gl. biti v smislu obstajanja	GO	{a,b,c}	{e,d,p}		GOae	(sem)
gl. biti v relacijskem smislu	GR	{a,b,c}	{e,d,p}		GOae	(je)
gl. obst. biti nik. obl.	GZO	{a,b,c}	{e,d,p}		GZOce	(ni)
gl. obst. biti prih.	GFO	{a,b,c}	{e,d,p}		GFObe	(bos)
glagol tipa imeti nik. obl.	GZ	{a,b,c}	{e,d,p}		GZbe	(nimaš)
velelnik	GV	{a,b,c}	{e,d,p}		GVbe	(glej)
opisni deležnik na -I	GL		{m,ž,s}	{e,d,p}	GLže	(obrisala)
opis. delež. gl. biti na -I	GLB		{m,ž,s}	{e,d,p}	GLBme	(bil)
trpni deležnik na -n/-t	GN/GT		{m,ž,s}	{e,d,p}	{1,2} GNme1	(rojen)
deležje na -č/-ši	GČ/GŠI				GČ	(loveč)
nedoločnik	GNE				GNE	(povedati)
namenilnik	GNA				GNA	(gledat)
pogojnik	GBI				GBI	(bi)
prosti glagolski morfem	Gmp				Gmp	(se)

pridevniška beseda

	oznaka	spol	število	sklon	stopnj.	dol.	primer
pridevnik	P	{m,ž,s}	{e,d,p}	{1,2,3,4,5,6}	{/,j,jj}	{/,i}	Pme1i (pomladni)
opisni deležnik na - I	PL	{m,ž,s}	{e,d,p}	{1,2,3,4,5,6}		{/,i}	PLmp4 (uspele)
deležnik stanja na - n/- t	PN/PT	{m,ž,s}	{e,d,p}	{1,2,3,4,5,6}		{/,i}	PNme4i (zgrešeni)
deležnik na - č/- ši	PČ/PŠI	{m,ž,s}	{e,d,p}	{1,2,3,4,5,6}		{/,i}	PČže2 (cvetoče)
povedkovnik	PD	{m,ž,s}	{e,d,p}		{/,j,jj}		PDme (rad)
svoj. prid. iz osebnih imen	PIO	{m,ž,s}	{e,d,p}	{1,2,3,4,5,6}			PIOžp4 (Andrejeve)
svoj. prid. iz imen prebivalcev	PIP	{m,ž,s}	{e,d,p}	{1,2,3,4,5,6}			PIPme1 (Brikin)
svoj. prid. iz veroslovnih imen	PIV	{m,ž,s}	{e,d,p}	{1,2,3,4,5,6}			PIVse4 (Kronovo)
svoj. prid. iz zemljepisnih imen	PIZ	{m,ž,s}	{e,d,p}	{1,2,3,4,5,6}			PIZže5 (Krimski)
svoj. prid. iz mitoloških imen	PIM	{m,ž,s}	{e,d,p}	{1,2,3,4,5,6}			PIMme5 (Hadovem)
svoj. prid. iz stvarnih imen	PIS	{m,ž,s}	{e,d,p}	{1,2,3,4,5,6}			PISže5 (Mohorjevi)

zaimki

		oznaka oseba	spol	število spol	število sklon	primer
osebni zaimek	ZO	{a,b,c}	{m,ž,s,/}{e,d,p}		{1,2,3,4,5,6}	ZOcme5 (njem)
osebni povr. zaimek	ZOP				{2,3,4,5,6}	ZOP2 (sebe)
svojilni zaimek	ZSV	{a,b,c}	{m,ž,s,/}{e,d,p}	{m,ž,s}{e,d,p}	{1,2,3,4,5,6}	ZSVAeme2 (mojega)
povr. svoj. zaimek	ZSVP			{m,ž,s}{e,d,p}	{2,3,4,5,6}	ZSVPme6 (svojim)
vprašalni zaimek	ZV			{m,ž,s}{e,d,p}	{1,2,3,4,5,6}	ZVse1 (kaj)
oziralni zaimek	ZR			{m,ž,s}{e,d,p}	{1,2,3,4,5,6}	ZRme2 (kakršnega)
nikalni zaimek	ZNI			{m,ž,s}{e,d,p}	{1,2,3,4,5,6}	ZNIže2 (nikakršne)
poljubnostni zaimek	ZPO			{m,ž,s}{e,d,p}	{1,2,3,4,5,6}	ZPOme1 (kdo)
ozir. poljub. zaimek	ZRPO			{m,ž,s}{e,d,p}	{1,2,3,4,5,6}	ZRPOme6 (komerkoli)
nedoločni zaimek	ZNE			{m,ž,s}{e,d,p}	{1,2,3,4,5,6}	ZNEse1 (nekaj)
drugostni zaimek	ZD			{m,ž,s}{e,d,p}	{1,2,3,4,5,6}	ZDme5 (drugem)
totalni zaimek	ZT			{m,ž,s}{e,d,p}	{1,2,3,4,5,6}	ZTse1 (vse)
istostni zaimek	ZI			{m,ž,s}{e,d,p}	{1,2,3,4,5,6}	ZIme4 (isti)
mногостни заемник	ZM			{m,ž,s}{e,d,p}	{1,2,3,4,5,6}	ZMse3 (marsičemu)
kazalni zaimek	ZK			{m,ž,s}{e,d,p}	{1,2,3,4,5,6}	ZKmp1 (ti)
poudarni zaimek	ZPU			{m,ž,s}{e,d,p}	{1,2,3,4,5,6}	ZPUže1 (sama)
oz. zaim. kot veznik	ZVR					(ki)

štевniki

	oznaka	spol	štевilo	sklon	primer
glavni števnik	ŠG	{m,ž,s}	{e,d,p}	{1,2,3,4,5,6}	ŠGže3 (petintridesetim)
vrstilni števnik	ŠV	{m,ž,s}	{e,d,p}	{1,2,3,4,5,6}	ŠVme2 (prvega)
ločilni števnik	ŠL	{m,ž,s}	{e,d,p}	{1,2,3,4,5,6}	ŠLse6 (dvojim)
množilni števnik	ŠM	{m,ž,s}	{e,d,p}	{1,2,3,4,5,6}	ŠMme5 (trojnih)
nedoločni števnik	ŠNE				ŠNE (nekaj)
število	Š				Š (77.500,00,-)

prislovi in ostale besedne vrste

	oznaka	podrednost	stopnjevanje	sklon	primer
prislov	A		(/,j,jj)		A (resnično)
členek	Č				Č (kar)
zaničani členek	ČZ				ČZ (ne)
členek v vezniški vlogi	ČV				ČV (ali)
predlog	E			{2,3,4,5,6}	E2 (iz)
veznik	V	{pr,po}			Vpr (in)
medmet	M				M (oh)
	oznaka	spol	število	sklon	primer
kratica z malo začetnico	K				K (št.)
kratica z veliko začetnico	KI	{/,m,ž,s}	{/,e,d,p}	{/,1,2,3,4,5,6}	KI (ŠTUNFF)
naslov spletnne strani	KURL				KURL (http://www.delo.si)



*Inštitut za slovenski jezik Frana Ramovša ZRC SAZU
Korpusni laboratorij*



Slovarske in besedilne zbirke

- **Nova beseda**
- **Besede slovenskega jezika**
- **Slovar slovenskega knjižnega jezika**

- Odzadnji slovar slovenskega jezika
- Besedišče slovenskega jezika
- Oblikoslovno označevanje
- Besedne leme iz zbranih del C. Kosmača
- Slovarček krajšav

- **Beseda**
- **Palindromi**
- **Določevanje osnovnih besednih oblik**
- **Ugotavljanje pogostnosti besednih oblik**

Spletni naslov: bos.zrc-sazu.si

*Inštitut za slovenski jezik Franja Ramovša ZRC SAZU
Laboratorij za korpus slovenskega jezika*

Oblikoslovno označevanje

Spletni viri

- [Iskanje po oblikoslovno označenih besedilih](#)
- [Datoteke s podatkovnimi viri za uporabo označevalnika](#)
- [Program EVA, za pogoje uporabe glej stran \[www.laze.org/eva\]\(http://www.laze.org/eva\)](#)

Bibliografija

- [Oblikoslovni označevalnik korpusa Beseda, JOTA marec 2008](#)
- [Razlaga oblikoslovnih oznak v zbranih delih Cirila Kosmača](#)
- [Part-of-speech tagging of Slovenian text](#)
- [O strojnem podprtju oblikoslovnem označevanju slovenskega besedila, \[Slavistična revija\]\(#\), Vol. 45/3-4, 1997, p. \[513\]-532.](#)



SEVE

V 3.28

Primož Jakopin

Text, Graphics, Data Base,
DeskTop Publishing
and Computer Aided Instruction
on ATARI ST

Statistični oblikoslovni označevalnik

program

besedilo (1)

datoteke s podatkovnimi viri za označevalnik:

- besedne oblike z oblikoslovnimi oznakami in frekvencami (2)
- n-terice besed z oblikoslovnimi oznakami in frekvencami (3)
- n-terice oblikoslovnih oznak s frekvencami (4)
- zbirka zvez besednih vrst z utežmi (5)

Učna zbirka (200.000 besed - 2,3,4)

- Ciril Kosmač: Pomladni dan
- Delo na internetu, 6.5.1997 - 17.6.1997
- George Orwell: 1984

datoteka 1: besedilo

Tisti pomladni dan je bil lep, svetel in zveneč, kakor
iz čistega srebra ulit.

Res je, da so se mrki oblaki grenkih doživetij večkrat
priopodili na jasno nebo mojega spomina, res je, da so stare in
nove bolečine dokaj pogosto zamolklo butale ob steno mojega srca, res
je, da so se včasih vzdignili vrtinci nekdanjih viharnih
mladostnih čustev, res je, da so se na dnu hladnega tolmuna
življenjskega spoznanja vzdihovaje obračali težki kamni potopljenih
hrepenenj – toda vse to ni zasenčilo, ni pomendralo, ni
izpodkopalo in ni raztrgalo široke in globoke njive mojega
miru.

datoteka 2: besedne oblike z oblikoslovnimi oznakami in frekvencami

```
celo-Č,230;Pse1,2;Pse4,2;Pže4,8;Pže6
ceļoletnega-Pme2
ceļoletno-Pže4
ceļonočni-Pže5
ceļoten-Pme1,3;Pme4,2
ceļoti-Sže5,15
ceļotna-Pmd4,2;Pže1,2
ceļotne-Pže2,2
ceļotnega-Pme2,2;Pse2,4
ceļotnem-Pme5,2
ceļotni-Pme4i,2;Pže5,2
ceļotnih-Pmp2;Pžp2
ceļotno-Pse1,6;Pse4;Pže4,7;Pže6,2
ceļovite-Pže2,2
ceļovitih-Pžp2
ceļovito-Pse4;Pže4
```

datoteka 3: nterice (2-5) besed z oblikoslovnimi oznakami in frekvencami

```
jaz ne•Z0ae1 Vpo,1;Z0ae1 ČZ,26
jaz nimam•Z0ae1 GZae,2
jaz nimam lepe•Z0ae1 GZae Pže2
jaz nimam lepe obleke•Z0ae1 GZae Pže2 Sže2
jaz nimam lepe obleke da•Z0ae1 GZae Pže2 Sže2 Vpo
jaz nimam časa•Z0ae1 GZae Sme2
jaz nimam časa da•Z0ae1 GZae Sme2 Vpo
jaz nimam časa da bi•Z0ae1 GZae Sme2 Vpo GBI
jaz nisem•Z0ae1 GZPae,3
jaz nisem strinjal•Z0ae1 GZPae GLme
jaz nisem strinjal ampak•Z0ae1 GZPae GLme Vpr
jaz nisem strinjal ampak velikokrat•Z0ae1 GZPae GLme Vpr A
jaz nisem takoj•Z0ae1 GZPae A
jaz nisem takoj vstal•Z0ae1 GZPae A GLme
```

datoteka 4: nterice (2-5) oblikoslovnih oznak s frekvencami

```
Pže1 Sže1•869
Pže1 Sže1 A•29
Pže1 Sže1 A A•1
Pže1 Sže1 A A GLže•1
Pže1 Sže1 A GLže•8
Pže1 Sže1 A GLže A•2
Pže1 Sže1 A GLže E4•1
Pže1 Sže1 A GLže E5•1
Pže1 Sže1 A GLže Sme4•1
Pže1 Sže1 A GLže Sže4•1
Pže1 Sže1 A GLže Vpr•1
Pže1 Sže1 A GLže ZR•1
Pže1 Sže1 A GLžp•1
Pže1 Sže1 A GLžp E4•1
Pže1 Sže1 A GNE•1
Pže1 Sže1 A GNE Sže4•1
Pže1 Sže1 A GNže1•1
Pže1 Sže1 A GPce•1
Pže1 Sže1 A GPce GLme•1
Pže1 Sže1 A GTže1•1
Pže1 Sže1 A GTže1 E6•1
Pže1 Sže1 A Gce•2
```

datoteka 5: zbirka zvez besednih vrst z utežmi

```
EPPS•Ec Pgnc Pgnc Sgnc•50000
EPS•Ec Pgnc Sgnc•50000
ES•Ec Sgnc•50000
PPPS•P*gnc P*gnc P*gnc Sgnc•10000
PPS•P*gnc P*gnc Sgnc•20000
PS•P*gnc Sgnc•30000
ŠS•ŠVgnc Sgnc•50000
```

«Tisti pomladni dan je bil lep, svetel in zveneč, kakor iz čistega srebra ulit.»

«Res je, da so se mrki oblaki grenkih doživetij večkrat pripodili na jasno nebo mojega spomina, res je, da so stare in nove bolečine dokaj pogosto zamolklo butale ob steno mojega srca, res je, da so se včasih vzdignili vrtinci nekdanjih viharnih mladostnih čustev, res je, da so se na dnu hladnega tolmuna življenjskega spoznanja vzdihovaje obračali težki kamni potopljenih hrepenenj – toda vse to ni zasenčilo, ni pomendralo, ni izpodkopalo in ni raztrgalo široke in globoke njive mojega miru. »O, ne! »Vse, kar se neukročenega in nezadoščenega divja po mojih notranjih temah, ni moglo odplaviti te dragocene, rodovitne zemlje, ki jo je vame naplavila, plast za plastjo, petintrideset let dolga, največkrat kalna in razburkana reka bridkih preizkušenj.»

Disk Block Edit Find Print Print direct Data Tools Help Varia System Lessons End

Tisti pomladni dan je bil lep, sv
iz čistega srebra ulit.
Res je, da so se mrki oblaki gre
pri podili na jasno nebo mojega spomina, res
nove bolečine dokaj pogosto zamolklo butale
je, da so se včasih vzdignili vrtinci nekda
mladostnih čustev, res je, da so se na dnu
življenjskega spoznanja vzdihovaje obračali
hrepenej – toda vse to ni zasenčilo, ni po
izpodkopalo in ni raztrgalo široke in globok
miru. ☺, ne! ☺ Vse, kar se neukročenega in
mojih notranjih temah, ni moglo odplaviti t
rodovitne zemlje, ki jo je vame naplavila,
petintrideset let dolga, največkrat kalna i
preizkušenj. ☺

Character statistics ▶

Line utilities ▶

File procedures ▶

Part-of-speech tagging ▶

Spell checker routines ▶

Translation ▶

Procedures with names ▶

Words to syllables, prefixes ... ▶

Split lines for columns ▶

OCR routines ▶

Raw format utilities ▶

Break file into words

- multiple files into words w/file names

Mark words not in dictionary

Count words in a block

Break file into sentences

Prepare text&tags for output

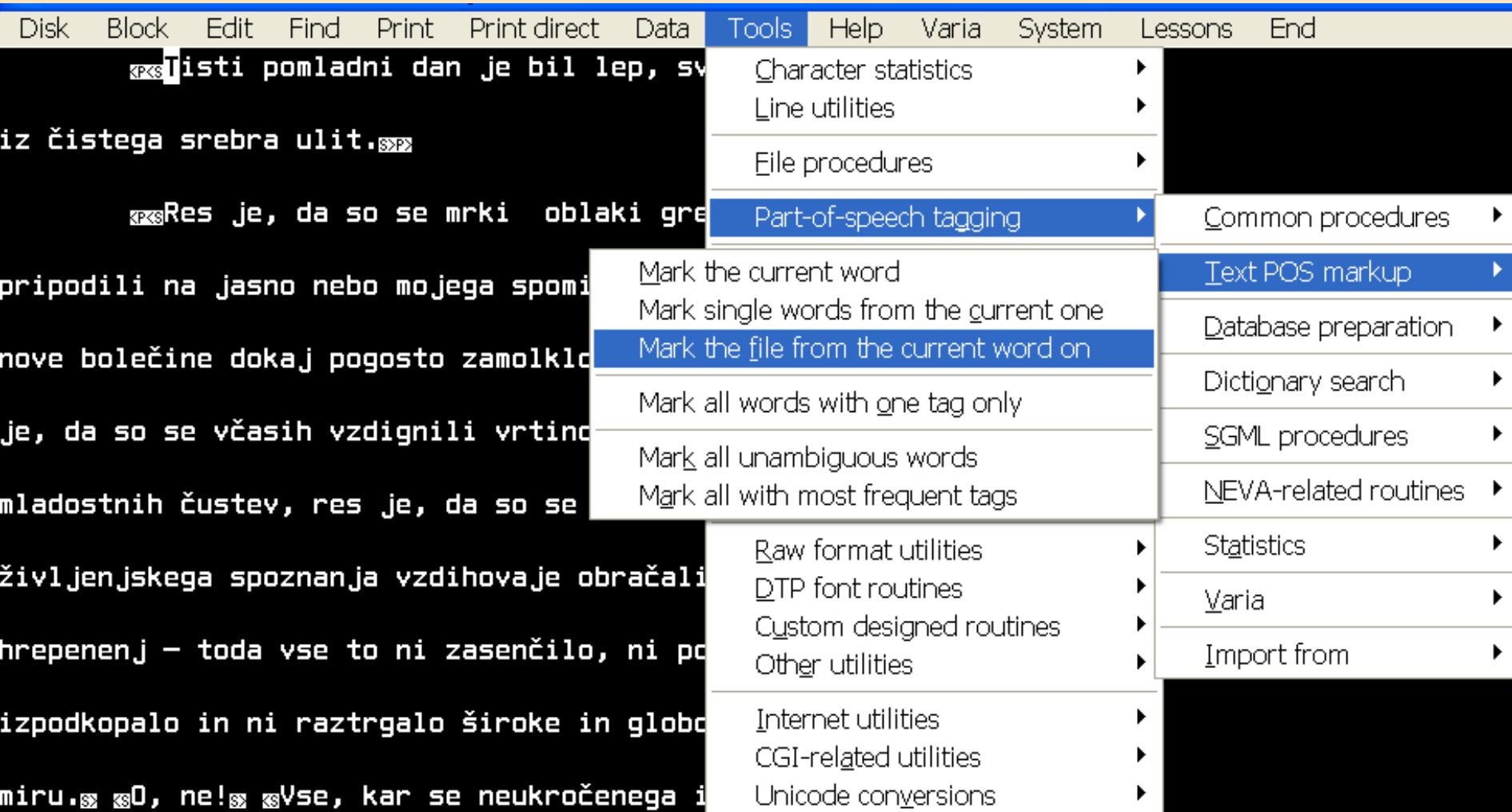
Remove white space

Varia ▶

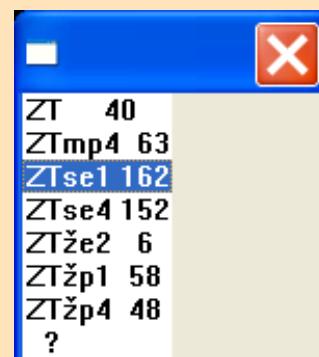
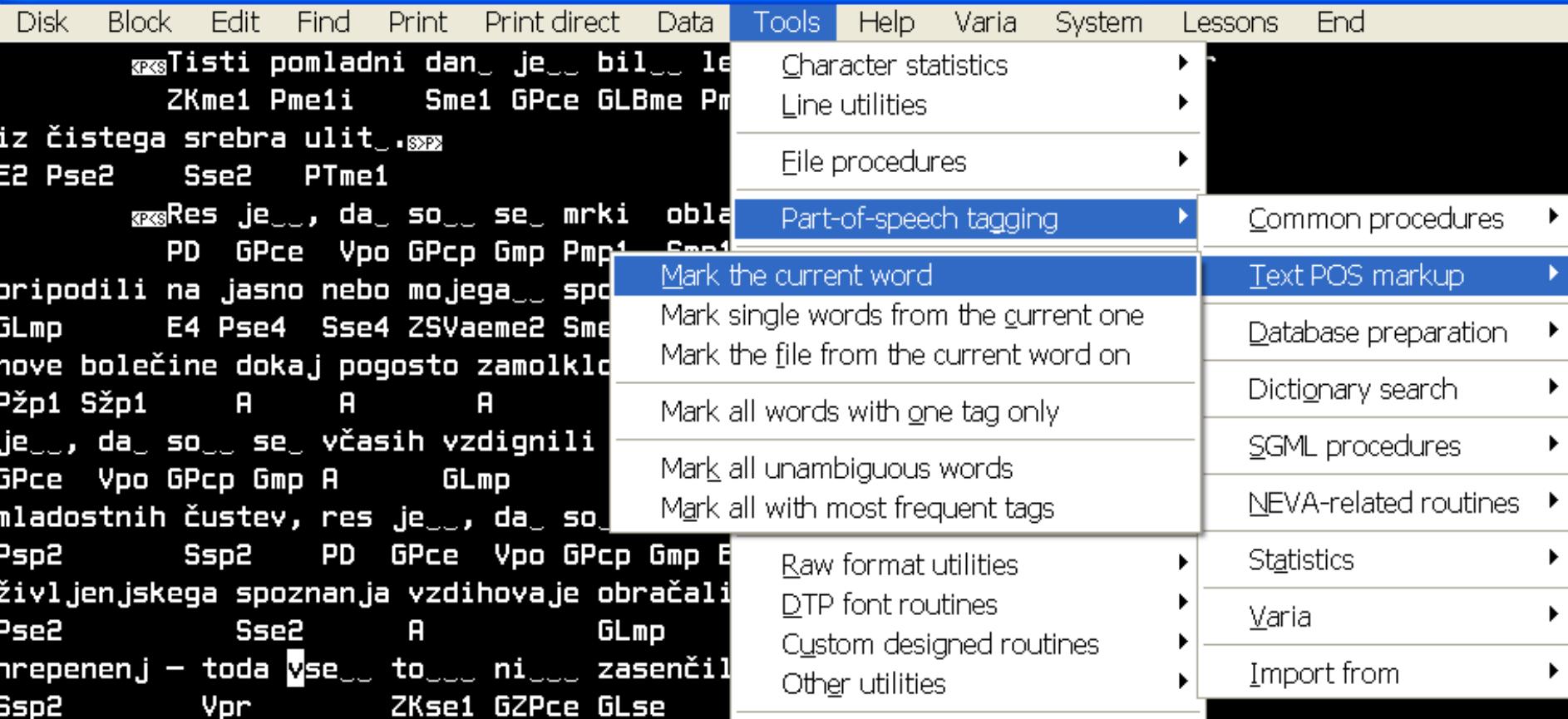
Import from

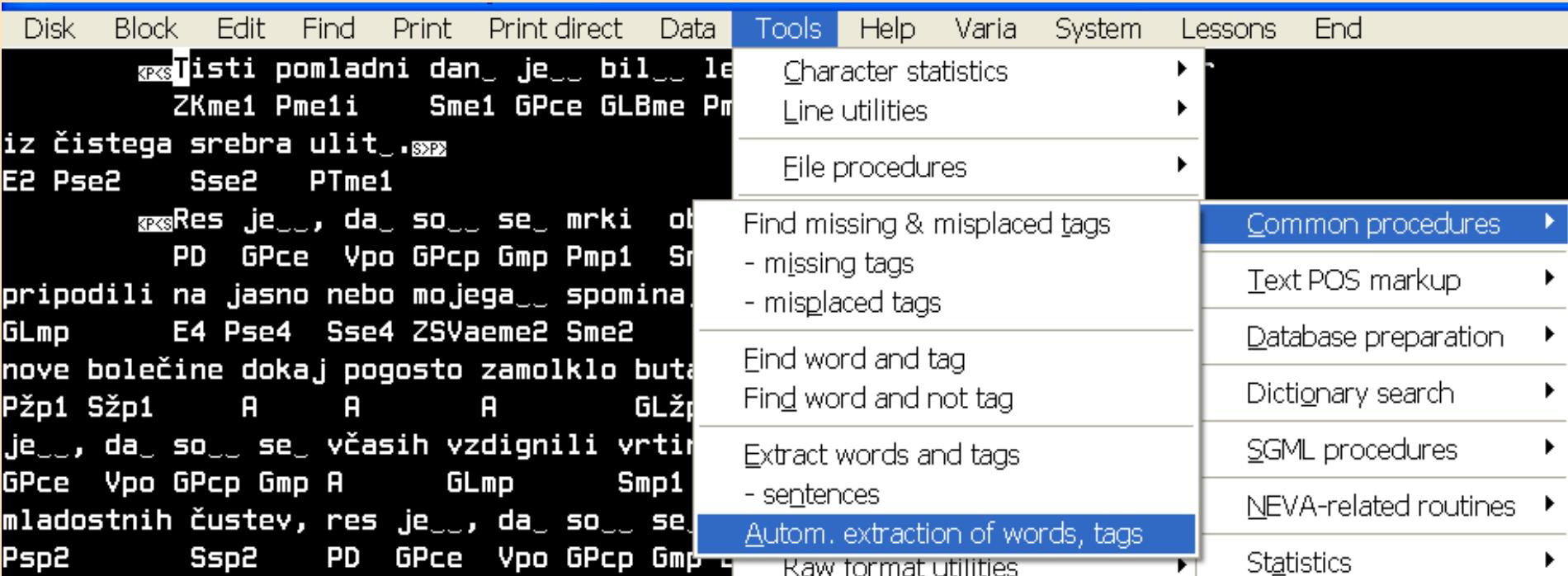
¶ Tisti pomladni dan je bil lep, svetel in zveneč, kakor
iz čistega srebra ulit.¶

¶ Res je, da so se mrki oblaki grenkih doživetij večkrat
priopodili na jasno nebo mojega spomina, res je, da so stare in
nove bolečine dokaj pogosto zamolklo butale ob steno mojega srca, res
je, da so se včasih vzdignili vrtinci nekdanjih viharnih
mladostnih čustev, res je, da so se na dnu hladnega tolmuna
življenskega spoznanja vzdihovaje obračali težki kamni potopljenih
hrepenenj – toda vse to ni zasenčilo, ni pomendralo, ni
izpodkopalo in ni raztrgalо široke in globoke njive mojega
miru.» O, ne! » Vse, kar se neukročenega in nezadoščenega divja po
mojih notranjih temah, ni moglo odplaviti te dragocene,
rodovitne zemlje, ki jo je vame naplavila, plast za plastjo,
petintrideset let dolga, največkrat kalna in razburkana reka bridkih
preizkušenj.¶



Tisti pomladni dan je bil lep, svetel in zveneč, kakor
ZKme1 Pme1 Sme1 GPce GLBme Pme1 Pme1 Vpr PČme1 Vpo
iz čistega srebra ulit.
E2 Pse2 Sse2 PTme1
Res je, da so se mrki oblaki grenkih doživetij večkrat
PD GPce Vpo GPcp Gmp Pmp1 Smp1 Psp2 Ssp2 A
priopili na jasno nebo mojega spomina, res je, da so stare in
GLmp E4 Pse4 Sse4 ZSVaeme2 Sme2 PD GPce Vpo GPcp Pžp1 Vpr
nove bolečine dokaj pogosto zamolklo butale ob steno mojega srca, res
Pžp1 Sžp1 A A A GLžp E4 Sže4 ZSVaese2 Sse2 PD
je, da so se včasih vzdignili vrtinci nekdanjih viharnih
GPce Vpo GPcp Gmp A GLmp Smp1 Psp2 Psp2
mladostnih čustev, res je, da so se na dnu hladnega tolmuna
Psp2 Ssp2 PD GPce Vpo GPcp Gmp E5 Sse5 Pme2 Sme2
življenskega spoznanja vzdihovaje obračali težki kamni potopljenih
Pse2 Sse2 A GLmp Pmp1 Smp1 PNsp2
hrepnenj – toda vse to ni zasenčilo, ni pomendralo, ni
Ssp2 Vpr ZTse1 ZKse1 GZPce GLse GZPce GLse GZPce
izpodkopalo in ni raztrgalо široke in globoke njive mojega
GLse Vpr GZPce GLse Pže2 Vpr Pže2 Sže2 ZSVaeme2
miru. O, ne! Vse, kar se neukročenega in nezadoščenega divja po
Sme2 M ČZ ZTse1 ZRse1 Gmp Pse2 Vpr Pse2 Gce E5
mojih notranjih temah, ni moglo odplaviti te dragocene,
ZSVaežp5 Pžp5 Sžp5 GZPce GLse GNE ZKže2 Pže2
rodovitne zemlje, ki jo je vame naplavila, plast za plastjo,
Pže2 Sže2 ZVR Z0cže4 GPce Z0ae4 GLže Sže1 E6 Sže6
petintrideset let dolga, največkrat kalna in razburkana reka bridkih
ŠG1 Ssp2 Pže1 A Pže1 Vpr PNže1 Sže1 Pžp2
preizkušenj.
Sžp2





bil·GLBme,1	ki·ZVR,1	petintrideset·ŠG1,1	težki·Pmp1,1
bolečine·Sžp1,1	lep·Pme1,1	plast·Sže1,1	tisti·ZKme1,1
bridkih·Pžp2,1	let·Ssp2,1	plastjo·Sže6,1	to·ZKse1,1
butale·GLžp,1	miru·Sme2,1	po·E5,1	toda·Vpr,1
da·Vpo,4	mladostnih·Psp2,1	pogosto·A,1	tolmuna·Sme2,1
dan·Sme1,1	moglo·GLse,1	pomendralo·GLse,1	ulit·PTmei,1
divja·Gce,1	mojega·ZSVaeme2,2;ZSVaeše2,1	pomladni·Pmei1,1	vame·ZOae4,1
dnu·Sse5,1	mojih·ZSVaežp5,1	potopljenih·PNsp2,1	večkrat·A,1
dokaj·A,1	mrki·Pmp1,1	preizkušenj·Sžp2,1	vharnih·Psp2,1
dolga·Pže1,1	na·E4,1;E5,1	priopdili·GLmp,1	vrtinci·Smp1,1
doživetij·Ssp2,1	največkrat·A,1	razburkana·PNže1,1	vse·ZTsei,2
dragocene·Pže2,1	naplavila·GLže,1	raztrgalo·GLse,1	vzdignili·GLmp,1
globoke·Pže2,1	ne·ČZ,1	reka·Sže1,1	vzdihovaje·A,1
grenkih·Psp2,1	nebo·Sse4,1	res·PD,4	včasih·A,1
hladnega·Pme2,1	nekdanjih·Psp2,1	rodovitne·Pže2,1	za·E6,1
hrepenenj·Ssp2,1	neukročenega·Pse2,1	se·Gmp,4	zamolklo·A,1
in·Vpr,6	nezadoščenega·Pse2,1	so·GPcp,4	zasenčilo·GLse,1
iz·E2,1	ni·GZPce,5	spomina·Sme2,1	zemlje·Sže2,1
izpodkopalo·GLse,1	njive·Sže2,1	spoznanja·Sse2,1	zveneč·PČme1,1
jasno·Pse4,1	notranjih·Pžp5,1	srca·Sse2,1	čistega·Pse2,1
je·GPce,6	nove·Pžp1,1	srebra·Sse2,1	čustev·Ssp2,1
jo·ZOče4,1	o·M,1	stare·Pžp1,1	široke·Pže2,1
kakor·Vpo,1	ob·E4,1	steno·Sže4,1	življenskega·Pse2,1
kalna·Pže1,1	oblaki·Smp1,1	svetel·Pme1,1	
kamni·Smp1,1	obračali·GLmp,1	te·ZKže2,1	
kar·ZRsei,1	odplaviti·GNE,1	temah·Sžp5,1	

Disk Block Edit Find Print Print direct Data Tools Help Varia System Lessons End

Tisti pomladni dan je bil le
ZKme1 Pme1 Sme1 GPce GLBme Pm
iz čistega srebra ulit...
E2 Pse2 Sse2 PTme1
Res je, da so se mrki obla
PD GPce Vpo GPcp Gmp Pmp1 Smp1
pri podili na jasno nebo mojega spomina, r
GLmp E4 Pse4 Sse4 ZSvaem
nove bolečine dokaj pogosto za
Pžp1 Sžp1 A A A
je, da so se včasih vzdi
GPce Vpo GPcp Gmp A GLmp
mladostnih čustev, res je, d
Psp2 Ssp2 PD GPce V
življenskega spoznanja vzdiho

Character statistics ►

Line utilities ►

File procedures ►

Part-of-speech tagging ►

Spell checker routines ►

Prepare files 2, 3 and 4

Extract words,tags,word numbers

Pack words,tags,word numbers

Find tags for numbers as words

Make composite tag from lemma,wform,tag

Make lemmas and simple tags from composites

Common procedures ►

Text POS markup ►

Database preparation ►

Dictionary search ►

SGML procedures ►

NEVA-related routines ►

Statistics ►

Extract SGML markup

Extract &...; SGML entities

Extract given SGML tag

Delete SGML markup

SGML setup

Delete the current SGML tag & its end

Delete the current SGML tag

Delete the current end-of-SGML tag

Jump to character outside tags

Mark the SGML markup

Check the SGML markup

Find lemmas & tags from dictionary

Find just POS tags from dictionary

Find lemmas for (word,tag) pairs

Find lemmas for unknown words

- simple lemmas for unknown words

- lemmas for words, unpacked dictionary

Find flexion from dictionary

Dictionary search ►

SGML procedures ►

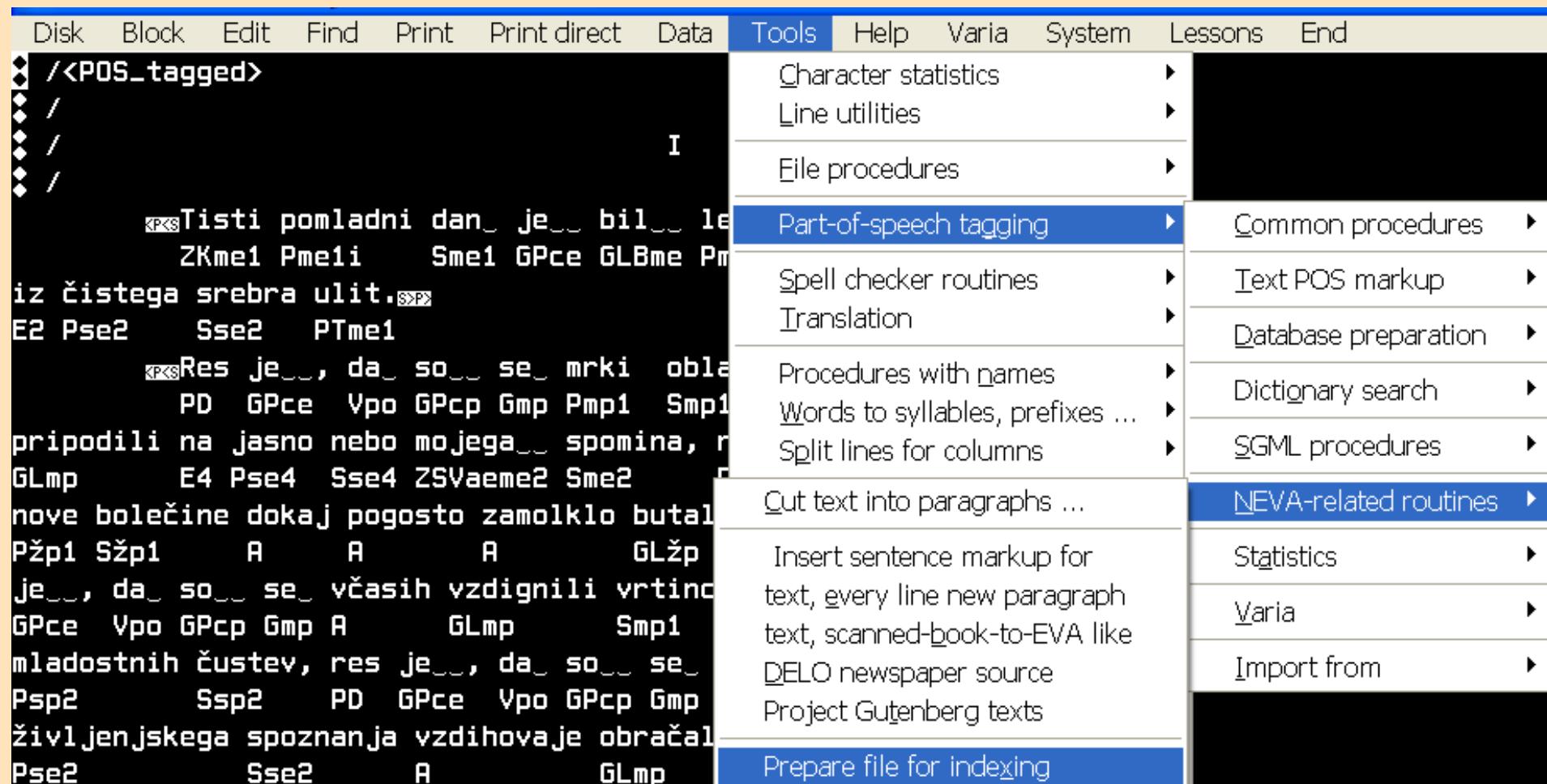
NEVA-related routines ►

Statistics ►

Varia ►

Import from ►

SGML procedures ►



1 /<POS_tagged>/ /<POS_tagged>/
Resnik je GPcep, da Vpor so GPcp se Gmp mrki Pmp1 oblaki Smp1 grenkih Psp2 doživ
OcMo, ne ČZ!»
Vse ZTsei, kar ZRsei se Gmp neukročenega Pse2 in Vpr nezadoščenega Pse2 div ja GPcep po E5
Tisti ZKme1 pomladni Pme1 dan Sme1 je GPcep bil GLBme torek Vpr resnično A lep P
In Vpr prav Č takšna ZKže1 je GPcep bila GLBžep Kadetka IOže1, ki ZVR se Gmp je G
Prvič A sem GPae se Gmp zbudil GLme zdavnaj A pred E6 zoro Sže6, še Č sredi E2
Nič ZNIsei čudnega Sse2, saj Vpr je GPcep bilo GLBse v E5 tisti ZKmp5 napetih PTmp5, a V
Takrat A me ZOae4 je GPcep vsak ZTme4 večer Sme4 sprejel GLme pod E4 streho Sže4 v E5 dr
Spal GLme sem GPae rahlo A in Vpr zelog A nemirno A.»

2 /<POS_tagged>/<POS_tagged>/<POS_tagged>/
Tudi Č zda j A sem GPae sunkovito A vrgel GLme glavo Sže4 kvišku A in Vpr se Gmp
Ležal GLme sem GPae v E5 nizki Pže5 in Vpr ozki Pže5 podstrešni Pže5 čumnati Sže5, polni
Vprašujoče A sem GPae ostrmel GLme, toda Vpr samo Č za E4 hip Sme4, zakaj Vpr še Č pred
Nasmehnili GLme sem GPae se Gmp in Vpr svetla Pže1 misel Sže1, ki ZVR je GPcep posvetila G
Zaprli GLme sem GPae oči Sžp4, srce Sse4 in Vpr vse ZTse4 telo Sse4 pa Vpr mi ZOae3 je G
»Doma A sem G0ae vendar Vpr!»

3 /<POS_tagged>/<POS_tagged>/<POS_tagged>/
Pet najstih ŠG5 letih Ssp5 spet A ležim GLme pod E6 domačo Pže6 streho Sže6!...»
Vrgel GLme sem GPae se Gmp vznak A in Vpr z E6 lačno Pže6 slastjo Sže6 potegnil
Bil GLBme je GPcep mehak Pme1, gost Pme1 in Vpr prijetno A hladen Pme1 kakor Vpor črno Pse
Tako jo A sem GPae začutil GLme v E5 žilah Sžp5 novog Pže4, opojno Pže4 moč Sže4.»

4 /<POS_tagged>/<POS_tagged>/<POS_tagged>/
S E6 prešerno Pže6 radostjo Sže6 sem GPae prekrižal GLme roker Sžp4 pod E6 tilnikom Sme6
Zazibalog GLse me ZOae4 je GPcep ter Vpr me ZOae4 začelo GLse dvigati GNE više Ajo in Vpr
Prav Č kmalu A me ZOae4 je GPcep dvignil GLse v E4 takoz K visoko Pže4 zanesenost Sže4,

5 /<POS_tagged>/<POS_tagged>/<POS_tagged>/
»O M rodni Pme1 dom Sme1, o M hiše Sže2 očetove Pže2 streha Sže1 tip ZObe1!»
Siromaku Sme3 si GPbe grad Sme1 in Vpr popotniku Sme3 v E5 dalji Sže5 uteha Sže1

Besede Sžp1 so GPcp zvenele GLžp polno A in Vpr slovesno A, kakor Vpor bi GBI
Široki Pmp1 zvoki Smp1 so GPcp mi ZOamp1 širili GLmp prsi Sžp4 in Vpr srce Sse4.
A Vpr glej GVbe, kar Č nenadoma A se Gmp je GPcep od E2 srca Sse2 odbil GLme grenak Pme1

Iskanje po zbirkì oblikoslovno označenih besedil

Išči

:pžp* sžp* pmp2 smp2

Primeri:	se je zdelo	- vse pojavitve fraze "se je zdelo".
	svet	- besede, ki vsebujejo "svet".
	:a a a	- vsi primeri treh zaporednih prislovov.
	:pže* sže*	- vse pojavitve parov: pridelnik ženskega spola v ednini z ujemajočim se samostalnikom.

Oblikoslovno označena besedilna zborka, 1.000.000 besed, obsega:

1. Zbrana dela Cirila Kosmača - 408.000 besed
2. Tomo Križnar: O iskanju ljubezni - 132.000 besed
3. George Orwell: 1984 - 91.000 besed
4. Platon: Država - 93.000 besed
5. Sveto pismo Nove zaveze - 150.000 besed
6. Gustave Flaubert: Bouvard in Pécuchet - 86.000 besed
7. Časopis DELO na internetu (vzorec iz leta 1997) - 52.000 besed

Iskalni izraz sme biti dolg do 74 znakov, število prikazanih zadetkov pa je omejeno na 200.

beseda

oblikoslovno označena besedila

:pžp* sžp* pmp2 smp2 (18)

Bilo je zaradi vzdušja hokejskih igrišč,
upravljeni državi²⁷ ter se zato izogiba
revščini, begu in prosjaštvu; bile so tu tudi
oficirske kuhinje, mlečnozobi kadeti, modre,
se smeje: smejejo se ženske, smejejo se
opoldanskem soncu visi gladka senožet, ki jo
travnikih po obeh straneh reke so se vlekle

* * * O

konflikte z okoljem, prepire in socialne drame,
gorah jih je istočasno ščitila tudi pred
Kamor je seglo oko, so ga dražile

Ukazal je skovati
na kolo in gonil z vso silo čez vse tiste
tržnice, škoda je prizadejala tudi režimu

Baje so se

Bilo je vse čmo
kot pri nas, recimo, koder je vse polno

mrzlih kopeli, skupinskih izletov
častnih služb, visokih položajev,
življenske podobe slavnih mož,
otožne oči ruskih ujetnikov,
potne glave starih možakarjev,
dolge sence visokih mecesnov
neskončne vrste temnopolih vojakov
privtnih namerah iranskih mul
Teroristične akcije radikalnih šiitov
podobne vsebinam bombajskih filmov.
velikimi vojskami azijskih osvajalcev.
žive barve razkošnih vencev
velike količine bakrenih novčičev
krmežljave ulice krmežljavih ljudi
zveste sile, hitrih umikov
cele mase trdoživilih poganov,
peklenskih strmin, meglenih prepadow
starih bajt, težkih spominov

in splošne čistosti mišljenja, ki se je
pravnih poslov in podobnih zaposlitev; ljub
izmed katerih so nekateri postali slavni
harmonike, s prtenimi rjuhami pogrnjene
smeje se luč pod stropom, svetle žlice v
režejo v poševne pasove.

v zelenih uniformah.
glede njihove ekspanzije v smislu izvoza
v Libanonu, atentati in ugrabitev so mulam

cvetja, fantastičnih ženskih sarijev, čudoviti
in nanje napisal vrednost, kot če bi bili
trideset kilometrov daleč na letališče.
vajenim gverilcem pa so le redko prišle
ki se jih prej nista dotaknila niti hinduizem
in hudournikov.
na pokopališčih ali na zgodovinskih spomej

Hvala za pozornost in
nasvidenje!