

Oblikoslovno označevanje in besediloslovna analiza

Aleksandra Bizjak

IZVLEČEK: V prispevku avtorica predstavlja uporabo oblikoslovnega označevalnika za besediloslovno raziskavo, ki je bil izdelan na Inštitutu za slovenski jezik ZRC SAZU. Z oblikoslovnim označevalnikom je mogoče hitro in zanesljivo odkrivati in izpisovati oblikoslovne vzorce, nanizane v linearnem poteku besedila, ki bi jim težko sledili z ročnim označevanjem, in tako analizirati veliko besedil.

ABSTRACT: The author presents the use of a morphological tagger developed at the Institute of the Slovene language at SRC SASA, in textual research. The application enables fast and reliable detection and display of grammatical patterns which occur over the course of a text. Such patterns would be rather difficult to identify by tagging texts manually and thus it would hardly be possible to analyse large quantities of texts.

Uvod

Pričujoča razprava obsega dva dela. Najprej so v prvem razdelku predstavljena teoretična izhodišča za besediloslovno analizo. V glavnih potezah bomo poskušali načrtati nove metodološke usmeritve pri analizi besedil. Iz tega kratkega pregleda bo razvidno, da je za opis besedilnega pomena treba opisati oblikoskladenjske vzorce. Ročno beleženje oblikoskladenjskih značilnosti v besedilu je zelo zamuden postopek že za večji niz stavkov, zato bomo v drugem razdelku opisali, kako lahko s pomočjo oblikoslovno označenega korpusa izpišemo oblikoslovne vzorce in rezultate predstavimo grafično.

1. Teoretična izhodišča za besediloslovno analizo

Jezikoslovci si pri odgovoru na vprašanje, kaj je pomen, niso edini. Tisti, ki zagovarjajo ožje pojmovanje pomena, pravijo, da mora teorija pomena »omogočati sistematičen opis besednih in stavčnih pomenov ter razmerij med njimi« (Golden

1997, 207). Gre torej za opis pomena jezikovnih znakov, ki pripadajo jezikovnemu sistemu, kar postavlja preučevanje besedilnega pomena zunaj pomenoslovja.

Tisti, ki pomen pojmujejo širše, vidijo notranjo povezavo med pomenom in sporazumevanjem, zato jih zanima tisti pomen jezikovnega znaka, ki ga opredeljuje jezikovna raba. Ne zanima jih torej stavek kot enota sistema, ampak stavek kot enota, ki se uresniči v besedilu. Čeprav je besedilu tako odmerjena večja teoretična pozornost, pa vprašanje, kako sistematično preučevati besedilne pomene, še zdaleč ni rešeno.

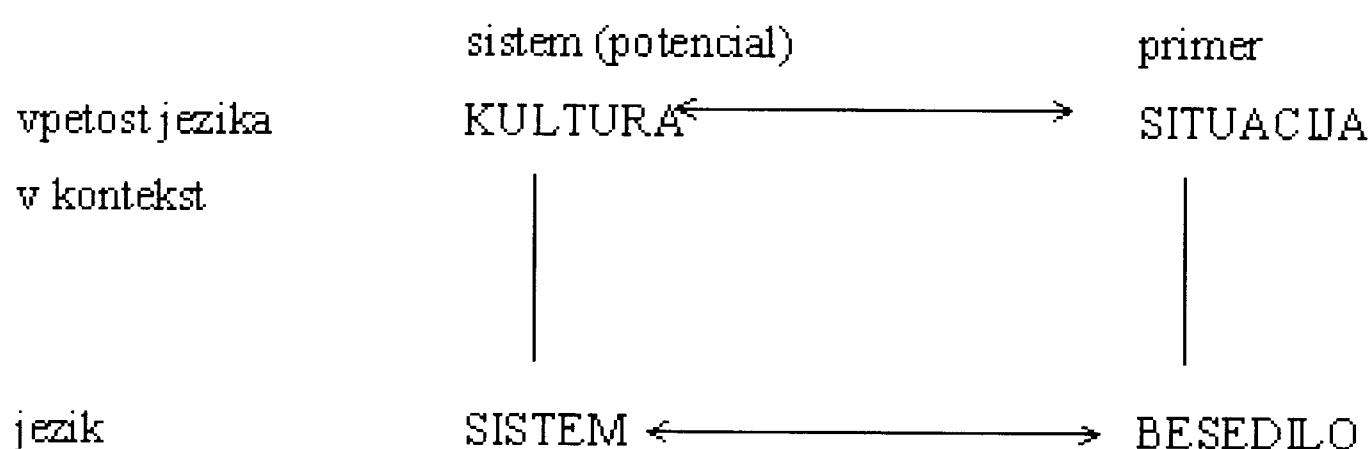
Preučevanje jezikovnega pomena v konkretnih okoliščinah rabe se začne šele v 60. letih s pragmatično teorijo govornih dejanj. In čeprav ta teorija zaobjame tudi nekatere zunajjezikovne dejavnike, ki vplivajo na pomen izrekov, zaradi omejenosti na stavčno ravnino ne daje možnosti za raziskovanje različnih oblik sporočanja. Govorci namreč v različnih sporazumevalnih okoliščinah ne tvorijo povedi, ampak besedila. Zato je za vpeljavo pojma besedilo v jezikoslovje odločilna predvsem analiza diskurza, ki je z raziskovalnim zanimanjem za jezikovne pojave na nadpovedni ravni prispevala k oblikovanju koncepta besedila. V razvoju pomenoslovja po letu 1970 je poleg analize diskurza nedvomno odigrala pomembno vlogo tudi sistemsko-funkcijska teorija jezika.

Ker Hallidayeva funkcijska teorija jezika odmerja besedilu večjo teoretično pozornost, je izbrana za teoretično izhodišče besediloslovne analize. V nadaljevanju si bomo pogledali Hallidayeve teoretične trditve o besedilu kot pomenski enoti.

1.1 Besedilo in Hallidayeva sistemsko-funkcijska teorija jezika

Pri obravnavi vprašanja pomena v Hallidayevi teoriji jezika je treba postaviti v izhodišče dejstvo, da gre za sistemsko-funkcijsko teorijo jezika, ki poudarja, da jezik ni samo abstrakten sistem znakov, ampak da je njegova osnovna funkcija sporazumevalna. Po Hallidayu je vsaka jezikovna dejavnost vpeta v družbeno-kulturni kontekst. Razmerje med kontekstom na eni strani in jezikom na drugi strani je prikazano na sliki 1.

Slika 1: Razmerje med jezikom in kontekstom (po Cloran 1994, 64)



Opomba: levo – desno = upo samičenje
 zgoraj – spodaj = uresničitev

Iz Hallidayeve opredelitve dinamičnega razmerja med kontekstom in jezikom izhaja spoznanje, da je jezik skupek vseh pomenskih izbir, ki so govorcju na voljo v določenem kulturnem kontekstu. Od tudi Hallidayeva oznaka, da je jezik »pomenski potencial« (angl. 'meaning potential') (Halliday 1978, 39). Pri sporazumevanju, to je v konkretnem situacijskem kontekstu, pa govorec iz mreže jezikovnih pomenov izbere samo nekatere pomene. Proces in rezultat pomenskega izbora v določenem situacijskem kontekstu je besedilo.

Tako je v Hallidayevem modelu jezikovne dejavnosti največja enota besedilo in ne poved. Besedilo kot jezikoslovna enota je v štiriravninskem sistemu, ki ga Halliday predlaga za opis jezika, uvrščeno na semantično ravnino, kakor je prikazano na sliki 2.

Slika 2: Jezikoslovne ravnine in jezikoslovne enote (po Cloran 1994 priredila A. B.)

RAVNINE JEZIKOSLOVNEGA OPISA	ENOTA
1. KONTEKST Področje, ton in način diskurza	?
2. SEMANTIKA Predstavna, medosebna in besedilna metafunkcija	besedilo
3. SLOVNICA	stavek
4. Glasoslovje	ko diranje

Vsak posamezen stavek v besedilu je v specifičnem pomenskem razmerju z nekaj stavki ali vsemi stavki v besedilu. Ta specifična pomenska razmerja v besedilu dajejo besedilu pomensko enotnost, in zato Halliday opredeli besedilo kot semantično enoto (Halliday 1977). Opisati besedilo kot semantično enoto torej pomeni opisati pomensko zgradbo besedila.

Pri analizi besedil ne smemo namreč pozabiti, da smo na ravnini besedila, ki v sistemski teoriji jezika ni slovnična, ampak pomenska enota. Tako kakor druge jezikovne enote ima tudi besedilo strukturo, in to pomensko strukturo. Hasanova predlaga, da je najmanjša strukturna enota pomena, v kateri se še izraža žanrska struktura besedila, sporočilo in da so osnovne pomenske prvine, ki sestavljajo sporočilo: entiteta, dogodek in okoliščine (1991, 81, po Cloran 1994, 250).

Pri pomenskem opisu besedila se bomo naslonili na razprave R. Hasan o semantiki sporočilnih enot (1991) in monografijo C. Cloran o retoričnih enotah (1994) ter

oblikovali semantično mrežo, v kateri je poudarjeno razmerje med entiteto in dogodkom na eni strani ter osebkom in povedkom na drugi strani. Za tak opis pa moramo poleg semantične ravnine upoštevati slovnično ravnino, saj so besedilni pomeni dosegljivi le kot ubesedeni pomeni, torej kot oblikoskladenjski vzorci ali s Hallidayevimi besedami: »Slovnična izbira je uresničitev semantične« (Halliday 1973, 93). Zato je prvi korak besediloslovne analize opis oblikoskladenjskih značilnosti besedila.

V razdelku 2 bomo opisali, kako smo si pri nadrobni jezikoslovni analizi pomagali z uporabo računalniškega orodja.

2. Opis oblikoskladenjskih značilnosti osebkovih entitet in povedkovih dogodkov

Da bi ponazorili, kako zapleteno je beleženje besedilnih pomenov, smo izbrali odlomek iz televizijske pridige iz Murske Sobote, ki je bila predvajana 1. marca 1998. Ročno beleženje oblikoskladenjskih značilnosti osebkovih entitet in povedkovih dogodkov je zelo zamuden postopek že za večji niz stavkov, kar ponazarja spodnji primer iz pridige, v katerem so oblikoskladenjske značilnosti osebkov in povedkov zabeležene v oglatih oklepajih.

Primer 1: Odlomek besedila, stavki 1– 6

1. Drage sestre in bratje, pred dnevi sem slišal na radiu misel napovedovalca:

entiteta = [glagolsko morfemsko določilo, 1. oseba ednine, moški spol]

dogodek = [glagolska oblika preteklik]

2. »Utrujeni smo od pustnih noči.«

entiteta = [glagolsko morf. določilo, 1. oseba množine, moški spol]

dogodek = [glagolska oblika sedanjik]

3. In tisti dan sem že bil na čelu zaznamovan z znamenjem smrti.

entiteta = [glagolsko morf. določilo, 1. oseba ednine, moški spol]

dogodek = [glagolska oblika preteklik]

4. Čas je,

entiteta = [glagolsko morf. določilo, 3. oseba ednine, srednji spol]

dogodek = [glagolska oblika sedanjik]

5. da se zaustavimo

entiteta = [glagolsko morf. določilo, 1. oseba množine, moški spol]

dogodek = [glagolska oblika sedanjik]

6. in premislimo.

entiteta = [glagolsko morf. določilo, 1. oseba množine, moški spol]

dogodek = [glagolska oblika sedanjik]

Ker postopek ročnega vnašanja oblikoskladenjskih značilnosti ni primeren za analizo večjega števila besedil, smo se odločili, da bomo besedilo oblikoslovno označili.

2.1 Oblikoslovno označevanje

Tako kot imamo v jezikoslovju različne ravnine, na katerih preučujemo jezikovne enote, je tudi v korpusnem jezikoslovju mogoče označevanje na različnih ravninah: glasoslovno označevanje (angl. 'phonological annotation'), oblikoslovno označevanje (angl. 'grammatical annotation' ali 'part-of-speech tagging'), skladenjsko označevanje (angl. 'syntactical annotation'), semantično označevanje (angl. 'semantic annotation') in pragmatično označevanje (angl. 'pragmatic annotation') (O'Donnell 1999, 74).

Najbolj razširjena oblika korpusnega označevanja (angl. 'corpus annotation') je oblikoslovno označevanje. Zanj potrebujemo, tako kot za vsako korpusno označevanje, posebno programsko opremo in nabor oznak. Kakor na večini drugih področij kvantitativnega jezikoslovja so bili prvi oblikoslovni označevalniki narejeni za angleški jezik. Označevalnik za slovenščino je začel nastajati 1996 na Inštitutu za slovenski jezik ZRC SAZU in je bil prvič predstavljen v članku *O strojno podprtem oblikoslovnem označevanju slovenskega besedila* (Jakopin in Bizjak 1997).

Oblikoslovno označeni korpusi se ponavadi uporabljajo za leksikografske, oblikoslovne in skladenjske raziskave posameznih jezikovnih značilnosti. Da je oblikoslovno označeni korpus mogoče uporabiti tudi za besediloslovne raziskave, je pokazal Biber (1998). Prešteval je pogostnost posameznih glagolskih oblik za čas in način v znanstvenih člankih ter ugotovil, da se deli besedila, to so uvod, metode, rezultati in obravnava, precej razlikujejo glede na dobljene rezultate. Kako se glagolske oblike nizajo v besedilu, je prikazal z diagramom, iz katerega je jasno vidno, da meje med deli besedila niso ostro začrtane, ampak zajemajo konec ene enote in začetek nove enote (Biber 1998, 122–130). Biber je s to raziskavo potrdil, da oblikoslovno označeni korpusi omogočajo samodejno beleženje jezikoslovnih značilnosti, nanizanih v linearnem poteku besedila in ustvarjajočih besedilne vzorce, ki bi jim težko sledili z ročnim označevanjem.

V nadaljevanju tega razdelka bomo na kratkem odlomku pridige pokazali, kako smo z uporabo urejevalnika EVA iz oblikoslovno označenega besedila dobili izpisek niza glagolskih oblik, pri čemer z nizom mislimo na zaporedje glagolskih oblik, kakor si sledijo v linearnem poteku besedila.

Izhodišče za analizo je torej oblikoslovno označeno besedilo. Za ponazoritev je izbran odlomek iz pridižnega besedila, ki je prikazan na sliki 3.

V oblikoslovno označenem besedilu ima vsaka beseda pripisano pravilno, od sobesedila odvisno oblikoskladenjsko oznako. Nabor oznak je podrobneje opisan v članku *O strojno podprtem oblikoslovnem označevanju slovenskega besedila* (Jakopin in Bizjak 1997). Tu pa si bomo pogledali samo oznake za glagolske oblike, ki so temelj za raziskovanje referencialnosti in časovnosti, na katerih gradimo opis

Slika 3: Odlomek oblikoslovno označenega besedila

Drege sestre in bratje, pred dnevi sem slišal na radiu
 P2p1 S2p1 Vpr Sep1 E6 Sep6 GPLae GLae E5 Sae5
 misel napovedovalca: »Utrujeni smo od pustnih noči.« in
 S2e4 Sae2 GNmp1 GPap EZ P2p2 S2p2 Vpr
 tisti dan smo že bili na čelu zaznamovan z znamenjem serti.
 ZKae4 Sae4 GPBae Č GLBae E5 Sae5 GNme1 E4 Sae4 S2e2
 »Čes je, da se zaustavimo in premislimo.« »Soj vendar v
 Sae1 GPce Vpo Gap Gap Vpr Gap Č Vpr E5
 življenju ni vse tako razbato.« »Ljudi ni vse tako črno, tako
 Sae5 OZPce ZTae1 RZ Pae1 Č OZPce ZTae1 RZ Pae1 RZ
 žalostno, kot se mi to pogosto dozdeva v mojem življenju.«
 Pae1 Vpo Gap ZKae3 ZKae1 R Oce E5 ZVaeae5 Sae5
 »Minuli teden sta zaznamovala pust in pepelnica.« »Življenje
 PLae4i Sae4 GPLcd GLcd Sae1 Vpr S2e1 Sae1
 pa ni samo pust in ni samo pepelnica.« »Življenje je splet
 Vpr OZPce Č Sae1 Vpr OZPce Č S2e1 Sae1 GPce Sae4
 napetosti med obojima.« in v postu se kristjani bolj
 S2e2 E6 ZTae6 Vpr E5 Sae5 Gap Sep1 R
 odločeno za pepelnico, in sicer povsem hoto, povsem
 Gap E4 S2e4 Vpr Vpr R R R
 zavezano, kajti to so dnevi očiščenja, dnevi prenove.«
 R Vpr ZKae1 GPcp Sep1 Sae2 Sep1 S2e2
 ■

»Kakor smo slišali v današnjem prebrancu evangelija, je
 Vpo GPLap GLap E5 Pae5 PHae5 Sae5 GPLae
 Kristus, ki je bil pri krstu v Jordanu napolnjen s
 IVae1 ZVR GPBae GLBae E5 Sae5 E5 IZae5 GNae1 E6
 svetlo duhom, odšel v puščavo.« »Puščava.« »Kdo na pozna tega
 IVae6 GLae E4 S2e4 S2e1 ZVae1 CZ Oce ZKae2

entitet in dogodkov v drugem koraku analize. Oznake so v preglednici 1 razvrščene tako, kakor si sledijo v besedilu, le da je os branja navpična in ne vodoravna kakor v besedilu.

Preglednica 1: Razlaga glagolskih oznak na sliki 3

GPLae (sem)	glagol, pomožni, ob opisnem deležniku na -l, prva oseba, ednina
GLme (slišal)	glagol, opisni deležnik na -l, moški spol, ednina
GNmp1 (utrujeni)	glagol, trpni deležnik na -n, moški spol, množina, prvi sklon
Gpap (smo)	glagol, pomožni, prva oseba, množina
GPBae (sem)	glagol, pomožni, ob opisnem deležniku na -l (biti), prva oseba, ednina
GLBme (bil)	glagol, pomožni, ob opisnem deležniku na -l (biti), moški spol, ednina
GNme1(zaznamovan)	glagol, trpni deležnik na -n, moški spol, ednina, prvi sklon
Gpce (je)	glagol, pomožni, tretja oseba, ednina
Gap (zaustavimo)	glagol, povedni sedanjik, prva oseba, množina
Gap (premislimo)	glagol, povedni sedanjik, prva oseba, množina
GZPce (ni)	glagol, nikalna oblika, pomožni, tretja oseba, ednina
GZPce (ni)	glagol, nikalna oblika, pomožni, tretja oseba, ednina
Gce (dozdeva)	glagol, povedni sedanjik, tretja oseba, ednina
GPLcd (sta)	glagol, pomožni, ob opisnem deležniku na -l, tretja oseba, dvojina

GZPce (ni)	glagol, nikalna oblika, pomožni, tretja oseba, ednina
GZPce (ni)	glagol, nikalna oblika, pomožni, tretja oseba, ednina
Gpce (je)	glagol, pomožni, tretja oseba, ednina
Gap (odločamo)	glagol, povedni sedanjik, prva oseba, množina
GPcp (so)	glagol, pomožni, tretja oseba, množina
GPLap (smo)	glagol, pomožni, ob opisnem deležniku na -l, prva oseba, množina
GLmp (slišali)	glagol, opisni deležnik na -l, moški spol, množina
GPLce (je)	glagol, pomožni, ob opisnem deležniku na -l, tretja oseba, ednina
GPBce (je)	glagol, pomožni, ob opisnem deležniku na -l (biti), tretja oseba, ednina
GLBme (bil)	glagol, pomožni, ob opisnem deležniku na -l (biti), moški spol, ednina
GNmel(napolnjen)	glagol, trpni deležnik na -n, moški spol, ednina, prvi sklon
GLme (odšel)	glagol, opisni deležnik na -l, moški spol, ednina

Oblikoslovno oznako sestavljajo: kratica za besedno vrsto, ki je pisana z velikimi tiskanimi črkami in vedno na prvem mestu, sledijo kratice, pisane z velikimi tiskanimi črkami, ki označujejo besednovrstne skupine, in kratice za oblikoslovne značilnosti, ki so pisane z malimi tiskanimi črkami ali števki. Glagolske oznake so bile za nalogo dopolnjene tako, da je pri sestavljenih glagolih pomožnemu glagolu (oznaka GP) dodana kratica L (torej GPL), ko glagolu sledi pretekli deležnik, in kratica B (GPB), ko glagolu sledi pretekli deležnik biti. Te spremembe so bile uvedene zato, da smo poenostavili računalniško obdelavo in delo z makroprogrami v urejevalniku EVA.

Z urejevalnikom Eva smo najprej razdelili besedilo na povedi, ker je členjenje besedila na povedi v Evi samodejen postopek. Odlomek oblikoslovno označenega besedila, razdeljen na povedi, je prikazan na sliki 4.

Razdelitev na povedi je bila v pomoč pri nadaljnji računalniški obdelavi, saj smo želeli ohraniti linearno zaporedje glagolskih oblik v besedilu, kakor kaže slika 5, na kateri je besedilo zapisano v navpični smeri, v vodoravni pa imamo razporejene oznake, lekseme in številke povedi.

Z makroprogrami smo nato iz navpičnega linearnega seznama celotnega besedila izbrali glagolske in samostalniške oblike in ohranili zaporedje povedi v besedilni razporeditvi ter tako dobili tudi stavčno zaporedje. Urejevalnik Eva namreč še nima izdelanega postopka za samodejno razčlenjevanje besedila na stavke.

Seznami, kakršen je na sliki 6, so bili oblikovani z urejevalnikom Eva za celotno besedilo. V vodoravni smeri imamo razvrščene številke, ki označujejo mesto stavka v besedilu, sledijo glagolske oblike in njihove oblikoskladenjske oznake ter v primeru, ko je v stavku ubesedeni osebek, tudi samostalniške oblike in njihove oblikoskladenjske oznake. S tem je narejen prvi korak analize, to je opis oblikoskladenjskih značilnosti entitet in dogodkov v besedilih. Opis ima obliko seznama,

Slika 4: Odlomek oblikoslovno označenega besedila, razdeljen na povedi

.1. Drage sestre in bratje, pred dnevi sem... slišal na radiu misel napovedovalca: »Utrujeni smo od pustnih noči.«
 \$1. Pzpl Szpl Vpr Sapl E6 Sep6 GPLae GLae ES Sae5 Sze4 Sae2 GHap1 GPap E2 Pzp2 Szp2 *
 .2. In tisti dan sem... že bil... na čelu zaznamovan z... znamenjem serti.
 \$2. Vpr ZKae4 Sae4 GPBae Č GLBae ES Sae5 GHae1 E6 Sae6 Sze2 *
 .3. Čas je... da se zaustavimo in premislimo.
 \$3. Sae1 GPce Vpo Gap Gap Vpr Gap *
 .4. Saj vendar v življenju ni... vse... tako rožnato.
 \$4. Č Vpr E5 Sae5 GPce ZTse1 AZ Pse1 *
 .5. Tudi ni... vse... tako črno, tako žalostno, kot se ni... to... pogosto dozdeva v... mojem... življenju.
 \$5. Č GPce ZTse1 AZ Pse1 AZ Pse1 Vpo Gap ZDae3 ZKse1 R Gce ES ZSvae5 Sae5 *
 .6. Minuli teden sta... zaznamovala pust in pepelnica.
 \$6. PLae41 Sae4 GPLcd GLad Sae1 Vpr Sze1 *
 .7. Življenje pa... ni... samo pust in... ni... samo pepelnica.
 \$7. Sse1 Vpr GPce Č Sae1 Vpr GPce Č Sze1 *
 .8. Življenje je... splet napetosti med obojim.
 \$8. Sse1 GPce Sae4 Sze2 E6 ZTse6 *
 .9. In v postu se kristjani bolj odločajo za pepelnico, in sicer povsem hote, povsem zavestno, kajti to... so... dnevi očiščevanja, dnevi prenove.
 \$9. Vpr E5 Sae5 Gap Sapl A Gap E4 Sze4 Vpr Vpr A R A R Vpr ZKae1 GPap Sapl Sae2 Sapl Sze2 *
 .10. Kakor smo... slišeli v današnjem praznem evangeliju, je... Kristus, ki je... bil... pri krstu v Jordanu napolnjen s... Svetim Duhom, odšel v puščavo.
 \$10. Vpo GPLap GLap ES Pae5 PHe5 Sae5 GPLce IVae1 ZVR GPBae GLBae ES Sae5 ES IZae5 GHae1 E6 IVae6 GLae E4 Sze4 *
 .11. Puščava.
 \$11. Sze1 *

v katerem so glagolske oblike navedene v navpičnem linearnem zaporedju zaradi računalniške obdelave podatkov.

Tak prikaz opisa bralcu ne da predstave o kontinuiteti, premikih in prepletih entitet in dogodkov, ki prispevajo k pomenskemu valovanju besedila, zato bomo opise entitet in dogodkov pretvorili v številčna zaporedja, ki so osnova za grafični prikaz pomenov, premočrtno snujočih se v besedilu.

2.2 Grafični prikaz osebkovih entitet in povedkovih dogodkov

Najprej smo z makroprogrami urejevalnika Eva glagolske oblikoskladenjske oznake glede na kategoriji osebe in števila pretvorili v številčne oznake po naslednjem ključu:

1. oseba ednine dobi številčno oznako 1;
1. oseba dvojine dobi številčno oznako 2;
1. oseba množine dobi številčno oznako 3;
2. oseba ednine dobi številčno oznako 4;
2. oseba dvojine dobi številčno oznako 5;
2. oseba množine dobi številčno oznako 6;
3. oseba ednine dobi številčno oznako 7;
3. oseba dvojine dobi številčno oznako 8;
3. oseba množine dobi številčno oznako 9.

Pretvorbo glagolskega zaporedja oblikoskladenjskih oznak glede na kategoriji osebe in števila v številčno zaporedje prikazuje primer 2.

Primer 2: Pretvorba kategorije osebe in števila v številčno oznako v odlomku besedila

- 1.sem·GPLae → 1
- 2.smo·GPap → 3
- 3.sem·GPBae → 1
- 4.je·GPce·čas·Sze1 → 7
- 5.zaustavimo·Gap → 3

Pžpl·drage·1	A·pogosto·5	Sme5·evangeliju·10
Sžpl·sestre·1	Gce·dozdeva·5	GPLce·je·10
Vprin·1	E5·v·5	IVme1·Kristus·10
Smp1·bratje·1	ZS·Vaese5·mojem·5	ZVR·ki·10
E6·pred·1	Sse5·življenju·5	GPBce·je·10
Smp6·dnevi·1	PLme4i·minuti·6	GLBme·bil·10
GPLae·sem·1	Sme4·teden·6	E5·pri·10
GLme·slišal·1	GPLcd·sta·6	Sme5·kristu·10
E5na·1	GLmd·zaznamovala·6	E5·v·10
Sme5·radni·1	Smel·pust·6	IZme5·Jordani·10
Sže4·misel·1	Vprin·6	GNme1·napolnjen·10
Sme2·napovedovaka·1	Sže1·pepelnica·6	E6·s·10
GNmp1·utrujeni·1	Ssel·življenje·7	IVme6·Svetim·Duhom·10
GPap·smo·1	Vprpa·7	GLme·odšel·10
E2od·1	GZPce·ni·7	E4·v·10
Pžp2·pustnih·1	Č·samo·7	Sže4·puščavo·10
Sžp2·noči·1	Smel·pust·7	
Vprin·2	Vprin·7	
ZKme4·tisti·2	GZPce·ni·7	
Sme4·dan·2	Č·samo·7	
GPBae·sem·2	Sže1·pepelnica·7	
Č·že·2	Ssel·življenje·8	
GLBme·bil·2	GPce·je·8	
E5na·2	Sme4·splet·8	
Sse5·čelu·2	Sže2·napetosti·8	
GNme1·zaznamovan·2	E6·med·8	
E6z·2	ZTse6·obojim·8	
Sse6·znanemjem·2	Vprin·9	
Sže2·snrti·2	E5·v·9	
Smel·čas·3	Sme5·postu·9	
GPce·je·3	Gmp·se·9	
Vpo·da·3	Smp1·kristjani·9	
Gmp·se·3	A·bolj·9	
Gap·zaustavimo·3	Gapodločamo·9	
Vprin·3	E4·za·9	
Gap·premislimo·3	Sže4·pepelnico·9	
Č·saj·4	Vprin·9	
Vpr·vendar·4	Vpr·sicer·9	
E5·v·4	A·povsem·9	
Sse5·življenju·4	A·hote·9	
GZPce·ni·4	A·povsem·9	
ZTsel·vse·4	A·zavestno·9	
AZ·tako·4	Vpr·kajti·9	
Psel·rožnato·4	ZKsel·to·9	
Č·tudi·5	GPep·so·9	
GZPce·ni·5	Smp1·dnevi·9	
ZTsel·vse·5	Sse2·očiščevalja·9	
AZ·tako·5	Smp1·dnevi·9	
Psel·črno·5	Sže2·penove·9	
AZ·tako·5	Vpo·kakor·10	
Psel·žalostno·5	GPLap·smo·10	
Vpo·kot·5	GLmp·slišali·10	
Gmp·se·5	E5·v·10	
ZOae3·ni·5	Pme5·današnjem·10	
ZKsel·to·5	PNme5·prebranjem·10	

Slika 5: Odlomek oblikoslovno označenega besedila, razdeljen na povedi in prikazan v navpični linearni razporeditvi

- 6·premislimo·Gap → 3
- 7·ni·GZPce → 7
- 8·ni·GZPce → 7
- 9·dozdeva·Gce → 7
- 10·sta·GPLcd·pust·Smel·pepelnica·Sžel → 8
- 11·ni·GZPce·življenje·Ssel → 7
- 12·ni·GZPce → 7
- 13·je·GPce·življenje·Ssel → 7
- 14·odločamo·Gap·kristjani·Smp1 → 3
- 15·so·GPcp → 9
- 16·smo·GPLap → 3
- 17·je·GPLce·Kristus·IVmel → 7
- 18·je·GPBce → 7

Ker se osebkove entitete in povedkovi dogodki zapisujejo v stavčne enote sočasno, smo se odločili, da bomo osebkove entitete zapisali v graf točkovno, časovno usmeritev povedkovih dogodkov pa prikazali z barvnim poljem po naslednjem ključu:

- a) preteklik dobi številčno oznako 10, ki predstavlja modro barvo;

Slika 6: Glagolski niz odlomka besedila

- 1·sem·GPLae
- 2·smo·GPap
- 3·sem·GPBae
- 4·je·GPce·čas·Sžel
- 5·zaustavimo·Gap
- 6·premislimo·Gap
- 7·ni·GZPce
- 8·ni·GZPce
- 9·dozdeva·Gce
- 10·sta·GPLcd·pust·Smel·pepelnica·Sžel
- 11·ni·GZPce·življenje·Ssel
- 12·ni·GZPce
- 13·je·GPce·življenje·Ssel
- 14·odločamo·Gap·kristjani·Smp1
- 15·so·GPcp
- 16·smo·GPLap
- 17·je·GPLce·Kristus·IVmel
- 18·je·GPBce

- b) sedanjik dobi številčno oznako 10·10, ki predstavlja zeleno barvo;
- c) prihodnjik dobi številčno oznako 10·10·10, ki predstavlja rumeno barvo.

Pretvorbo glagolskega zaporedja oblikoskladenjskih oznak glede na kategorijo časa v številčno zaporedje prikazuje primer 3.

Primer 3: Pretvorba kategorije časa v številčno oznako v odlomku besedila

- 1·sem·GPLae → 10
- 2·smo·GPap → 10·10
- 3·sem·GPBae → 10
- 4·je·GPce → 10·10
- 5·zaustavimo·Gap → 10·10
- 6·premislimo·Gap → 10·10
- 7·ni·GZPce → 10·10
- 8·ni·GZPce → 10·10
- 9·dozdeva·Gce → 10·10
- 10·sta·GPLcd· → 10
- 11·ni·GZPce → 10·10
- 12·ni·GZPce → 10·10
- 13·je·GPce → 10·10
- 14·odločamo·Gap → 10·10
- 15·so·GPcp → 10·10
- 16·smo·GPLap → 10
- 17·je·GPLce → 10
- 18·je·GPBce → 10

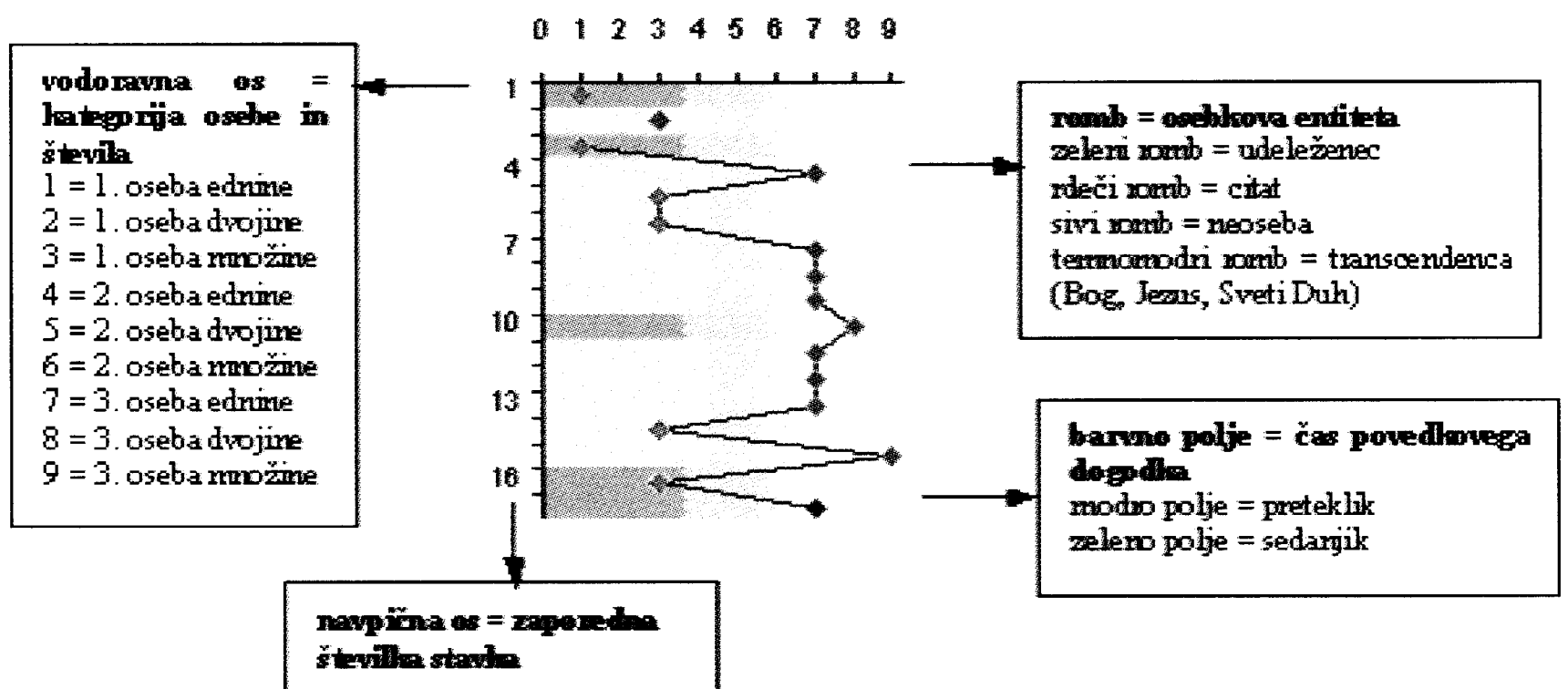
Številčna zaporedja za kategorijo osebe, števila in časa smo vnesli v program MsExcel in oblikovali graf, v katerem točke od 1 do 9 na vodoravni osi predstavljajo kategorijo osebe in števila, točke na navpični osi zaporedne številke stavkov v besedilu in barvni stolpci kategorijo časa.

Graf, ki je bil izdelan na osnovi podatkov o osebkovih entitetah in povedkovih dogodkih v odlomku besedila (stavki 1–18), je prikazan na sliki 7 in je opremljen s pojasnili za branje.

1. Drage sestre in bratje, pred dnevi sem slišal na radiu misel napovedovalca:
2. »Utrujeni smo od pustnih noči.«
3. In tisti dan sem že bil na čelu zaznamovan z znamenjem smrti.
4. Čas je,
5. da se zaustavimo
6. in premislimo.
7. Saj vendar v življenju ni vse tako rožnato.
8. Tudi ni vse tako črno, tako žalostno,
9. kot se mi to pogosto dozdeva v mojem življenju.
10. Minuli teden sta zaznamovala pust in pepelnica.
11. Življenje pa ni samo pust

12. in ni samo pepelnica.
 13. Življenje je splet napetosti med obojim.
 14. In v postu se kristjani bolj odločamo za pepelnico, in sicer povsem hote, povsem zavestno,
 15. kajti to so dnevi očiščevanja, dnevi prenove.
 16. Kakor smo slišali v današnjem prebranem evangeliju,
 17. je Kristus,
 18. ki je bil pri krstu v Jordanu napolnjen s Svetim Duhom, odšel v puščavo.

Slika 7: Razlaga grafa za odlomek besedila, stavki 1–18



Za lažje razumevanje grafa je pred njim izpisano besedilo. Enote od 1 do 9 na vodoravni osi označujejo kategorijo osebe in števila. Enote na navpični osi predstavljajo zaporedno številko stavka v besedilu. Kadar v besedilu nimamo na začetku vrstice oznake za številko stavka, to pomeni, da se stavek nadaljuje (glej 17. stavek na sliki 7) ali da gre za neglagolske stavke in polstavčne tvorbe. Osebkove entitete so v grafu predstavljene kot barvne točke v obliki romba in povedkovi dogodki kot barvna polja.

V prvem stavku, ki je na navpični osi označen z zaporedno številko 1, je entiteta v smeri vodoravne osi oddaljena za eno enoto, kar pomeni, da je nanosnik govorec (1. oseba ednine), barvno polje pa je modro, kar nam sporoča, da je dogodek prvega stavka usmerjen v preteklost. Tudi nanosnik tretjega stavka je govorec in dogodek usmerjen v preteklost. V petem, šestem in štirinajstem stavku je entiteta v smeri vodoravne osi oddaljena za tri enote, kar pomeni, da je nanosnik govorec in poslušalci (1. oseba množine), barvno polje pa je zeleno, torej so dogodki usmerjeni v sedanost. Prvoosebni in drugoosebni glagoli ednine, dvojine in množine označujejo v besedilu udeležence govornega dogodka in imajo nanosnik v situacijskem kontekstu konkretnega govornega dogodka, zato so na grafu vedno označeni z zelenim romбом.

Razpon od 1. do 6. enote na vodoravni osi je razpon netretje glagolske osebe in na tem območju grafa se gibljejo zeleni rombi, ki predstavljajo pomensko skupino, imenovano udeleženci govornega dogodka. Razpon tretje glagolske osebe je od 7. do 9. enote na vodoravni osi in na tem območju so razvrščene entitete četrtega, sedmega, osmega, devetega, desetega, enajstega, dvanajstega, trinajstega in petnajstega stavka, ki imajo nanosnika –človeško in so dogodkovno vse usmerjene v sedanost (polje zelene barve), razen osmega stavka, ki je usmerjen v preteklost (polje modre barve). V sedemnajstem stavku je v besedilo uvedena nova entiteta, ki se nanaša na nadnaravno bitje in je označena z modrim romбом.

Posebej je treba omeniti še drugi stavek, ki je označen z rdečim romбом. Rdeči romb se od drugih barvnih rombov razlikuje, ker sporoča, da gre v besedilu za dobesedno navajanje. Citiranje uvaja besedilo drugega govornega dogodka in tako vpeljuje entiteto in dogodek drugega časa in prostora. Specifična kombinacija entitete in dogodka tvori enoto, ki je na ravni pomena višja od entitete in dogodka ter se imenuje retorična enota. Rdeči romb tako označuje enoto kompozicijske ravnine.

Da bi pojasnili, kako so v grafu upodobljene osebkove entitete z barvnimi točkami in povedkovi dogodki z barvnimi polji in kako je zabeleženo njihovo gibanje, ali drugače povedano, kako je treba brati graf, smo v tem razdelku podrobno opisali entitete in dogodke za posamezne stavke odlomka pridižnega besedila. Vendar namen grafa ni branje stavka za stavkom. Prav grafična upodobitev s svojo statičnostjo naj bi nam pomagala, da tisto, kar ostaja v ubeseditvenem procesu neopazno in prikrito, pusti sled, pride na površje ter nam da predstavo o kontinuiteti in premikih v pomenskem valovanju besedila.

Sklepne ugotovitve

Graf je torej nekakšen slikovni model, ki beleži gibanje entitet in dogodkov in je tako izhodišče za opis strukturnih vzorcev, nastajajočih s kontinuiteto, premiki in prepleti entitet in dogodkov v besedilu. Z računalniškimi postopki, opisanimi v tem razdelku, lahko oblikujemo grafe za cel korpus besedil in jih uporabljamo kot slikovne modele za opazovanje strukturnih vzorcev, ki jih ustvarja zaporedje entitet in dogodkov ter tako določamo pomensko-strukturno zgradbo besedil.

Prav grafična upodobitev s svojo statičnostjo naj bi nam pomagala, da tisto, kar ostaja v ubeseditvenem procesu neopazno in prikrito, pusti sled, pride na površje ter nam da predstavo o kontinuiteti in premikih v pomenskem valovanju besedila.

Navedena literatura

Biber, Douglas, Conrad, Susan, Reppen, Randi, 1998, *Corpus linguistics: Investigating language structure and use*, Cambridge University Press, Cambridge.

- Cloran, Carmel 1994, *Rhetorical Units and Decontextualization: An Enquiry into some Relations of Context, Meaning and Grammar*, Department of English Studies, University of Nottingham (Monographs in Systemic Linguistics, 6), Nottingham.
- Golden, Marija, 1997, *O jeziku in jezikoslovju*, Filozofska fakulteta, Oddelek za splošno in primerjalno jezikoslovje, Ljubljana.
- Halliday, Michael A. K., 1973, *Explorations in the Functions of Language*. Edward Arnold, London.
- Halliday, Michael A. K., 1977, Text as semantic choice in social contexts, v: T. Van Dijk in J. Petofi (ur.), *Grammars and Descriptions*. Berlin, de Gruyter, 176–225.
- Halliday, Michael A. K., 1978, *Language as Social Semiotic, The Social Interpretation of Language and Meaning*, Edward Arnold, London.
- Hasan, Ruqaiya, 1991, Questions as a mode of learning in everyday talk. *Language Education: Interaction and Development*, Ur. T. McCausland, Launceston, University of Tasmania, 70–119.
- Jakopin, Primož, Bizjak, Aleksandra, 1997, O strojno podprtem oblikoslovnem označevanju slovenskega besedila, *Slavistična revija* 45/3–4, 513–532.
- O'Donnell, Matthew, 1999, The Use of Annotated Corpora for New Testament Discourse Analysis: A Survey of Current Practice and Future Prospects, *Discourse Analysis and the New Testament, Approaches and Results*, Journal for the Study of the New Testament, Supplement Series 170, Studies in New Testament Greek 4, Ur. S. E. Porter in J. T. Reed. Sheffield Academic Press, Sheffield, 71–117.

Morphological Tagging and Textual Analysis Summary

In order to describe textual meanings the author employs the systemic-functional linguistic theory. A semantic network emphasizing the relation between the entity and the event on the one hand and the subject and the predicate on the other is presented. A sermon from Murska Sobota which was broadcast on March 1st 1998 on TV Slovenia serves as an exemplification for this process. The basic structure for the presentation of entities and events is the notation of morpho-syntactic patterns which occur in a linear text string. In the notation process the linguistic features of a text, which would be rather difficult to identify by tagging texts manually, the morphological tagger for Slovene, developed at the Institute of the Slovene language at SRC SASA, is of great help.

Every string of morpho-syntactic tags is converted into a graphical presentation. The graph functions as a sort of a map, describing the movements of entities and events, and as such it becomes the starting point for the description of the semantic construction of a given text. It is exactly the graphical presentation that enables the identification of semantic shifts, which are hidden and perceived only intuitively by

the listener/reader, and provides the access to this information for further linguistic research.

Aleksandra Bizjak, Inštitut za slovenski jezik Frana Ramovša ZRC SAZU, Novi trg
2, 1000 Ljubljana
E-pošta: aleksandra.bizjak@zrc-sazu.si