

Agrovoc descriptors: brassica oleracea capitata; cabbages; terminology; nomenclature; identification; taxonomy; classification; research; interdisciplinary research; newspapers

Agris category code: C30, F01, U30

COBISS Code 1.01

Grafični prikaz besedilnih pomenov v znanstvenem članku o varstvu zelja (*Brassica oleracea* L. var. *capitata*)

Aleksandra BIZJAK KONČAR¹, Dragan ŽNIDARČIČ²

Delo je prispelo 7. maja 2007; sprejeto 20. junija 2007.

Received May 7, 2007; accepted June 20, 2007.

IZVLEČEK

Glavni namen pričujočega prispevka je osvetliti vprašanje besedilnih vzorcev v znanstvenem članku s področja varstva zelja (*Brassica oleracea* L. var. *capitata*). Teoretično izhodišče za analizo je sporočilna semantika, ki se naslanja na sistemsko-funkcijski model jezika. V prispevku so uporabljene metode besedilne analize v povezavi z orodji korpusnega jezikoslovja. Pri beleženju jezikoslovnih značilnosti, nanizanih v linearnem poteku besedila, ki bi jim težko sledili z ročnim označevanjem, nam je v pomoč oblikoslovni označevalnik za slovenščino, izdelan na Inštitutu za slovenski jezik ZRC SAZU, pri izpisovanju jezikoslovnih značilnosti besedil pa je uporabljena grafična predstavitev. Graf je nekakšen slikovni model in je tako izhodišče za opis besedilne zgradbe znanstvenega članka.

Ključne besede: zelje, *Brassica oleracea* L. var. *capitata*, varstvo rastlin, znanstveni članek, besedilni vzorci, oblikoslovni označevalnik za slovenščino

ABSTRACT

MAP OF TEXTUAL PATTERNS IN RESEARCH ARTICLE OF PROTECTION OF CABBAGE (*Brassica oleracea* L. var. *capitata*)

This paper aims to explore textual patterns of research article of protection of cabbage (*Brassica oleracea* L. var. *capitata*). The theoretical construct for the analysis is provided by concept of message semantics as defined in Systemic Functional linguistics. In the study, the specificity of discourse analysis is combined with tools of corpus analysis. The linguistic features of texts which would be difficult to keep track of by manual tagging are recorded by the automated part-of-speech tagger for Slovenian developed at the Institute of Slovenian language at ZRC SAZU and the graphic display is used for the presentation of data. The graph functions as a sort of map and as such becomes the starting point for the description of the discourse structure of research article.

Key words: cabbage, *Brassica oleracea* L. var. *capitata*, crop protection, research article, textual patterns, part-of speech tagger for Slovenian language

¹ dr. znan., Inštitut za slovenski jezik Frana Ramovša ZRC SAZU, SI-1000 Ljubljana, Novi trg 4, aleksandra.bizjak@zrc-sazu.si

² mag. znan., Univerza v Ljubljani, Biotehniška fakulteta, SI-1000 Ljubljana, Jamnikarjeva 101, dragan.znidarcic@bf.uni-lj.si

UVOD

Da besedilo ni samo linearna združba stavkov, ampak hierarhično urejena struktura, je teoretično izhodišče vseh novejših smeri jezikoslovja, sociologije in psihologije, ki se ukvarjajo s preučevanjem besedila. Zgodovinski pregledi nam kažejo, da ima preučevanje zgradbe besedil v literarni vedi dolgo tradicijo in je spodbudilo tudi raziskovanje strukturne zgradbe besedil v jezikoslovju. S problematiko strukturne zgradbe besedil se na evropskih tleh ukvarjajo vse od antike. Tako je že Aristotel v *Poetiki* (ok. 330 pr. Kr.) obravnaval zgradbo tragedije in epa (Aristotel, 1982). Strukturni zgradbi se posvečata predvsem dve jezikoslovni šoli: sistemsko-funkcijska in pragmatična.

Teoretično izhodišče prispevka je model za besedilno analizo, kakršnega je izoblikovala Hasanova (Hasan, 1984). Zaradi tesne navezanosti njene teorije na Hallidayevo teorijo jezika smo najprej zarisali širši tloris Hallidayeve misli o jeziku. 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, zato je tudi razumevanje pomena zelo široko. Po Hallidayu je vsaka jezikovna dejavnost vpeta v družbeno-kulturni kontekst. Pri sporazumevanju govorec iz mreže jezikovnih pomenov izbere samo nekatere pomeni. Proces in rezultat pomenskega izbora v določenem kontekstu je besedilo. In čeprav se nam govornica in pisna besedila na prvi pogled kažejo, kakor da so iz besed in stavkov, so v resnici zgrajena iz ubesedenih pomenov (Halliday, 1989). V sistemsko-funkcijski teoriji jezika je besedilo pomenska enota.

Opis pomenskega ustroja besedila je zahtevna naloga, saj je besedilna semantika novo področje jezikoslovja, ki je v primerjavi z besedno in stavčno semantiko dokaj neraziskano. Glavni razlog za težko razumevanje semantičnega ustroja besedila je neizdelan metajezik za opis te ravnine (Birch in O'Toole, 1988).

METODE DELA

Metajezik za opis besedilnega pomena

Podlaga za analizo semantičnega ustroja znanstvenega članka je bila sporočilna semantika in njeni semantični pojmi: sporočilo, entiteta in dogodek (Hasan, 1991). V sistemskem jezikoslovju je temelj za oblikovanje pomenskih mrež opis oblikoskladenjskih vzorcev, saj je med pomensko in slovnično ravnino dialektično razmerje. Pomeni so dosegljivi samo kot ubesedeni pomeni na slovnični ravnini (Halliday, 1973). Človekove predstave postanejo v besedilu jezikovni izrazi ali modeli: entiteta, dogodek in okoliščine, ki so v stavku izraženi kot samostalniška zveza, glagolska zveza in prislovne ali predložne zveze. Zato je prvi korak k analizi besedil opis oblikoskladenjskih značilnosti osebkovih entitet in povedkovih dogodkov. Ker postopek ročnega vnašanja oblikoskladenjskih značilnosti ni primeren za analizo daljšega besedila, smo se odločili za oblikoskladenjsko označevanje, ki je predstavljeno v naslednjem razdelku.

Oblikoslovno označevanje

Za bolj detajlno jezikoslovno analizo besedil moramo računalniško besedilo označiti dodatno. Tako kot imamo v jezikoslovju različne ravnine, na katerih preučujemo jezikovne enote, je tudi v korpusnem jezikoslovju mogoče označevanje na različnih ravneh: 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).

Najbolj razširjena oblika jezikoslovnega označevanja je oblikoslovno označevanje. 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). 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. Za oblikoslovno označevanje potrebujemo posebno programsko opremo in nabor oznak. Označevalnik za slovenščino je začel nastajati leta 1996 na Inštitutu za slovenski jezik ZRC SAZU in je bil prvič predstavljen v Slavistični reviji (Jakopin in Bizjak, 1997).

Izhodišče za našo analizo znanstvenega članka s področja varstva zelja je torej oblikoslovno označeno besedilo. Na sliki je 1 ponazorjen odlomek oblikoslovno označenega besedila na računalniškem zaslonu.

```

§V_ letu 2002 smo_ preučevali vpliv števila škropljenj z_ deltametrinom na zmanjšanje
§E5 Sse5 Š GPap GLmp Sme4 Sse2 Ssp2 E6 Sme6 E5 Sse4
škodljivosti tobakovega resarja na zgodnjem zelju.
Sže2 Sme1 E5 Pse5 Sse5
§V_ bločnem poskusu smo_ na zunanjih listih glave občutljivega cv. Parel ugotavljali odstotek
§E5 Pme5 Sme5 GPap E5 Pmp5 Smp5 Sže2 Pme2 K IRme1 GLmp Sme4
poškodovanosti listne površine.
Sže2 Pže2 Sže2
§Ob upoštevanju nižjega gospodarskega praga škodljivosti (do 1% poškodovane listne površine) med
§E5 Sse5 Pme2j Pme2 Sme2 Sže2 E2 PNže2 Pže2 Sže2 E6
obravnavanja, v_ katerih smo_ rastline enkrat in_ dvakrat poškropili z_ insekticidom, ne
Ssd6 E5 ZRsd5 GPap Sžp4 A Vpr A GLmp E6 Sme6 ČZ
ugotavljamo razlik.
Gap Sžp2
§Omenjeni prag tam ni_ bil_ presežen šele pri osmem zunanjem listu v_ glavi, kar je_
§PNme1i Sme1 ZK GZPce GLBme GNme1 Č E5 ŠVme5 Pme5 Sme5 E5 Sže5 ZR GPce
veljalo tudi za neškropljene rastline, pri katerih smo_ na zunanjih listih ugotovili največji
GLse Č E4 Pžp4 Sžp4 E5 ZRžp5 GPap E5 Pmp5 Smp5 GLmp Pme4JJ
povprečni indeks poškodb.
Pme4i Sme4 Sžp2

```

Slika 1: Odlomek oblikoslovno označenega besedila

Figure 1: Section of grammatically tagged text

V oblikoslovno označenem besedilu ima vsaka beseda pripisano pravilno, od sobesedila odvisno oblikoskladenjsko oznako. 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. Nabor oznak sta podrobneje opisala Jakopin in Bizjak (1997).

Glavni namen oblikoslovnega označevanja je bil pojasniti teoretične pojme, ki so vključeni v pojmovni sestav sporočilne semantike in oblikoslovnega označevanja.

Analiza znanstvenega članka

Zapleteno vprašanje besedilnega pomena, to je prepoznavanje semantične kontinuitete in premikov ter s tem povezano razumevanje strukturnih enot, smo ponazorili z besedilno analizo znanstvenega članka: Vpliv števila škropljenj z insekticidom na zmanjšanje škodljivosti tobakovega resarja (*Thrips tabaci* Lindeman, Thysanoptera, Thripidae) na zgodnjem zelju (Trdan in Žnidarčič, 2004).

Določanje osebkovih entitet in povedkovih dogodkov je bilo v prvem koraku opisovanje oblikoskladenjskih uresničitev osebkov in povedkov v stavčnih enotah. Postopek ročnega vnašanja oblikoskladenjskih značilnosti je zelo dolgotrajen že za eno besedilo. Oblikoslovno označeno besedilo nam je omogočilo samodejno beleženje jezikoslovnih značilnosti,

nanizanih v linearnem poteku besedila. Tako v nadaljevanju predstavljamo računalniški postopek, s katerim smo iz oblikoslovno označenega korpusa dobili izpisek zaporedja glagolskih in samostalniških oblik, kakor si sledijo v linearnem poteku besedila, in predstavili postopek pretvorbe niza oblikoskladenjskih oznak v grafični zapis.

Grafični prikaz osebkovih entitet in povedkovih dogodkov

Z oblikoslovnim označevalnikom za slovenščino, vgrajenim v urejevalnik EVA, smo označili izbrani znanstveni članek, kar omogoča samodejno zapisovanje jezikoslovnih značilnosti, nanizanih v linearnem poteku besedila. Glagolske oblikoskladenjske oznake smo glede na kategorijo osebe, števila in časa pretvorili v številčne oznake in v programu MsExcel oblikovali graf.

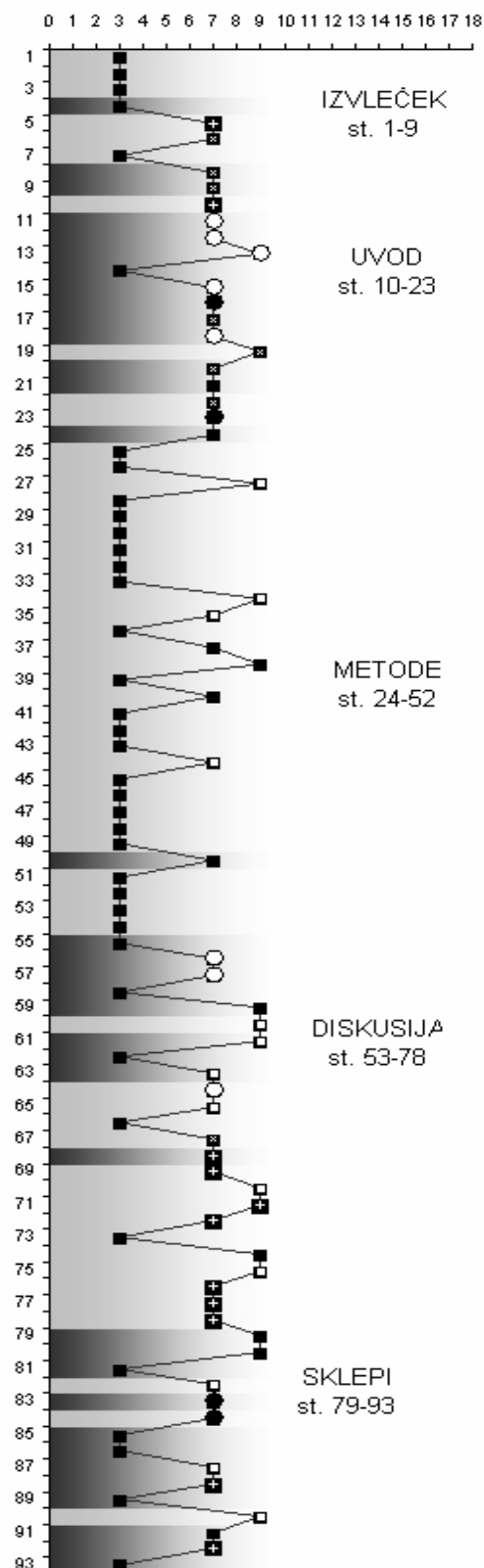
REZULTATI

Graf, ki je bil izdelan na osnovi podatkov o osebkovih entitetah in povedkovih dogodkih v članku, je prikazan na sliki 2 in je opremljen s pojasnili za branje.

Enote od 1 do 9 na vodoravni osi označujejo kategorijo osebe in števila. Enote na navpični osi pomenijo zaporedno številko stavka v besedilu. Osebkove entitete so predstavljene kot točke in povedkovi dogodki kot polja.

V prvem stavku, ki je na navpični osi označen z zaporedno številko 3, je entiteta v smeri vodoravne osi oddaljena za tri enote, kar pomeni, da je nanosnik govorec in raziskovalci (1. oseba množine), barvno polje pa je svetlo, kar nam sporoča, da je dogodek prvega stavka usmerjen v sedanjost. Tudi nanosniki drugega, tretjega in četrtega stavka so raziskovalci, v četrtem stavku pa je dogodek usmerjen v preteklost, saj je barvno polje temno. Razpon od 1. do 6. enote na vodoravni osi je razpon netretje glagolske osebe in na tem območju grafa se gibljejo črni kvadrati, ki sestavljajo pomensko skupino, imenovano raziskovalci. S črnim kvadratom so označene tudi tiste entitete, ki se gibljejo v razponu tretje glagolske osebe, a se nanašajo na raziskovanje (na primer 21. stavek). Razpon tretje glagolske osebe je od 7. do 9. enote na vodoravni osi in na tem območju so razvrščene še naslednje pojmovne entitete: zelje (beli kvadrat), tobakov resar (beli krog), deltametrin (črni krog) in pojem škodljivost (plus).

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 nekatere stavke besedila. Vendar namen grafa ni branje stavka za stavkom. V nadaljevanju smo graf uporabljali kot slikovni model, ki smo ga vzeli pod drobnogled in opazovali besedilne vzorce, ki jih ustvarja zaporedje entitet in dogodkov v tem besedilu.



LEGENDA:

vodoravna os = kategorija osebe in števila

- 1 = 1. oseba ednine
- 2 = 1. oseba dvojine
- 3 = 1. oseba množine
- 4 = 2. oseba ednine
- 5 = 2. oseba dvojine
- 6 = 2. oseba množine
- 7 = 3. oseba ednine
- 8 = 3. oseba dvojine
- 9 = 3. oseba množine

navpična os = zaporedna številka stavka

osebkovala entiteta

- črni kvadrat = raziskovalec, poskus
- beli kvadrat = zelje
- beli krog = tobakov resar
- črni krog = deltametrin
- plus = škodljivost
- križec = drugo

čas povedkovega dogodka

- temno polje = preteklik
- svetlo polje = sedanjik

Slika 2: Graf za znanstveni članek o varstvu zelja

Figure 2: Map of research article of protection of cabbage

Opis besedilnih vzorcev

S pomočjo slikovnega modela na sliki 2 smo osvetlili besedilne vzorce, ki jih v besedilu zarisuje gibanje entitet in dogodkov. Že bežen pogled na slikovni model znanstvenega članka je dovolj, da opazimo razlikovanje med entitetami na začetku in koncu besedila ter entitetami v osrednjem delu. Kakšne so torej razlike med začetnim in osrednjim delom? Kaj pa med začetnim in končnim delom?

Poglejmo si enoto, imenovano izvleček, in opazujmo vzorce, ki jih ustvarjajo konfiguracije entitet in dogodkov. Od prvega do devetega stavka prevladuje entiteta, ki označuje raziskovalce v prvi osebi množine (črni kvadrat), dogodki pa so usmerjeni in sedanost in preteklost. Pomeni se torej gibljejo v pojmovnem polju raziskovanja.

Prehod k novi enoti, uvodu, uvaja deseti stavek. Iz grafa lahko razberemo, da v uvodu ustvarjata kontinuiteto na osi entitet raziskovanje (črni kvadrat) in uničevanje (entiteta tobakov resar, ki je predstavljena z belim krogom, in entiteta deltametrin, ki je predstavljena s črnim krogom). Na časovni osi pa prevladuje usmerjenost v preteklost. Naslednja enota se razteza od 24. do 52. stavka in se imenuje metode. Tudi metode so tako kakor izvleček raziskovalno osredinjena enota (črni kvadrat), ki uvede entiteto zelje (beli kvadrat). Lastnosti, ki gradijo pomenski premik v metodah, so vezane tudi na os dogodka, na kateri je izbrana možnost za sedanost, to je istost časa govorjenja in pomenske podstave.

Od 53. do 55. stavka imamo prehod k enoti diskusija z rezultati, za katero je značilno, da so entitete bolj razpršene. V linijo kratkih sledij, najprej tobakov resar in zelje (beli krog in beli kvadrat), nato pojmovna entiteta škodljivost (plus), se nenehno vpleta entiteta raziskovanje (črni kvadrat). Povedkovi dogodki so v prvi polovici enote usmerjeni v preteklost (temno polje), v drugi polovici v sedanost (svetlo polje). Taka razpršena linija sledij se nadaljuje tudi v sklepni enoti (stavki 79–93), vendar z osredinjenostjo na preteklost.

Z analitičnim opazovanjem besedilnih vzorcev, ki jih ustvarjajo entitete in dogodki, smo za znanstveni članek o varstvu zelja opisali pet jasno zamejenih delov besedila. Ugotovili smo, da se deli besedila, to so izvleček, uvod, metode, diskusija z rezultati in sklepi, precej razlikujejo glede na konfiguracije entitet in dogodkov. Osredotočenost na raziskovanje, sledja pojmov s področja varstva zelja ter raba preteklega in sedanjega časa so lastnosti, ki oblikujejo znanstveni članek.

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. 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.

SKLEPI

Z besediloslovno analizo smo ugotovili, da pozorno opazovanje nizanja osebkovih entitet in povedkovih dogodkov pripelje do strukturnega tlora znanstvenega članka s področja varstva zelja. Poseben pomen pri opisu pomena sta imela semantična pojma entiteta in dogodek, ki se navezujeta na sporočilno semantiko (Hasan, 1991) in s katerima je mogoče opisati strukturne enote besedila in narediti pomenske premike, ki so v jeziku zakriti in jih poslušalec/bralec zaznava le intuitivno, razvidne in dostopne jezikoslovnemu preučevanju.

Čeprav je namen prispevka grafično prikazati besedilne vzorce, se ob analizi postavljajo nova vprašanja o besedilnih vzorcih v znanstvenih člankih s področja varstva zelja. Da bi dobili gradivo za analizo, ki s svojim obsegom zagotavlja tudi ustrezno reprezentativnost vzorca za dovolj zanesljivo in veljavno posplošitev rezultatov na populacijo znanstvenih člankov s tega področja, bi morali oblikovati korpus, ki bi obsegal najmanj 50 besedil. Tako je prispevek tudi spodbuda za nadaljnje zbiranje gradiva in ugotavljanje besedilnih značilnosti znanstvenih člankov s področja vrtnarstva in varstva rastlin.

VIRI

- Aristoteles, 1982. *Poetika*. Ljubljana: Cankarjeva založba.
- Biber, D., Conrad, S., Reppen, R. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Butt, D. 1972. Randomness, order and the latent patterning of text. *Functions of Style* (ur.: Birch, D., O'Toole, M.). London: Pinter Publishers.
- Halliday, M.A.K. 1973. *Explorations in the Functions of Language*. London: Edward Arnold.
- Halliday, M.A.K. 1989. Context of situation. V: Halliday, M. A. K., Hasan, R.: *Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective*. 3. izd. Oxford: Oxford University Press: 3–12.
- Hasan, R. 1984. The Nursery Tale as a Genre. *Nottingham Linguistic Circular*, 13: 1–51.
- Hasan, R. 1991. Questions as a mode of learning in everyday talk. *Language Education: Interaction and Development* (ur.: McCausland, T.). Launceston: University of Tasmania: 70–119.
- Jakopin, P., Bizjak, A. 1997. O strojno podprtem oblikoslovnem označevanju slovenskega besedila. *Slavistična revija*, 45/3–4: 513–532.
- O'Donnell, M. B. 1999. The Use of Annotated Corpora for New Testament Discourse Analysis: A Survey of Current Practice and Future Prospects. V: *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.: Porter, S.E., Reed, J.T.). Sheffield: Sheffield Academic Press: 71–117.
- Trdan, S., Žnidarčič, D. 2004. Vpliv števila škropljenj na zmanjšanje škodljivosti tobakovega resarja (*Thrips tabaci* Lindeman, Thysanoptera, Thripidae) na zgodnjem zelju. *Novi izzivi v poljedelstvu*: 513–532.