
Sintaksičko-semantička anotacija elektronskog korpusa srpskog jezika

Naučni rad

DOI: 10.18485/judig.2025.1.ch20

Nataša Kiš¹,  0000-0001-8627-0316

Apstrakt

U radu će biti predstavljene neke od mogućnosti korišćenja elektronskog korpusa kao primarnog izvora građe za različita istraživanja iz oblasti sintakse i semantike savremenog srpskog jezika. Dva osnovna zadatka rada jesu da se prikažu dosadašnja istraživanja vezana za proces komplementacije prideva i pridevskih imenica, odnosno njihovog rekcijuskog potencijala na sintagmatskom planu, a koja počivaju na relevantnoj građi preuzetoj prvenstveno iz elektronskog *Korpusa savremenog srpskog jezika*, SrpKor2013 (<https://noske.jerteh.rs/#dashboard?corpname=SrpKor2013>). Drugi zadatak je da se ukaže na potrebu proširenja mogućnosti pretrage elektronskih korpusa vezanu za sintaksičko-semantičku analizu na nivou proste rečenice. Pretrage ovog tipa zasnivaju se prvenstveno na sintaksičkoj i semantičkoj anotaciji korpusa i uključivale bi, između ostalog, podatke o valencijskim osobinama predikata rečenice i svim obaveznim i fakultativnim aktantima u konkretnoj situaciji, a važan segment jeste izdvajanje semantičkih klasa reči i njihovih sintaksičko-semantičkih osobenosti na osnovu kojih bi se mogli pretraživati različiti rečenični modeli. Jedan od ciljeva rada jeste da se ukaže na odgovarajuće lingvističke analize koje bi omogućile izdvajanje i uključivanje podataka potrebnih za dalji razvoj u oblasti sintaksičke i semantičke anotacije korpusa.

Ključne reči: sintaksa, semantika, anotacija, elektronski korpus, rekcija, srpski jezik

¹ Odsjek za srpski jezik i lingvistiku, Filozofski fakultet, Univerzitet u Novom Sadu, natasab@ff.uns.ac.rs

1. Uvod

1. U opisu svakog pojedinačnog prirodnog jezika, bilo da se radi o deskriptivnoj odnosno sinhronijskoj lingvistici, istorijskoj odnosno dijahronijskoj lingvistici, komparativnoj ili pak kontrastivnoj lingvistici, od izuzetnog je značaja uvid u što obimniju i stilski što raznovrsniju jezičku građu. Razvoj korpusne lingvistike umnogome doprinosi relevantnosti istraživanja i utemeljenosti zaključaka analiza u svim oblastima lingvističkih istraživanja. Ovom prilikom ukazaćemo na mogućnosti korišćenja elektronskog korpusa kao primarnog izvora građe za određena istraživanja iz oblasti sintakse i semantike savremenog srpskog jezika.

Dva osnovna zadatka rada jesu da se prikažu dosadašnja istraživanja vezana za proces komplementacije prideva i pridevskih imenica, odnosno njihovog rekcijaskog² potencijala na sintagmatskom planu, a koja počivaju na relevantnoj građi preuzetoj prvenstveno iz elektronskog *Korpusa savremenog srpskog jezika*, SrpKor2013 (<https://noske.jerteh.rs/#dashboard?corpname=SrpKor2013>). Drugi zadatak je da se ukaže na potrebu proširenja mogućnosti pretrage korpusa zasnovanu na sintaksičkoj i semantičkoj anotaciji. Pretrage ovog tipa omogućavale bi pristup podacima o valencijskim karakteristikama predikata rečenice i svim obaveznim i fakultativnim aktantima u konkretnoj situaciji, rekcijским особеностима глагола, prideva, imenica i priloga, a važan segment takve analize može biti izdvajanje semantičkih klasa reči i njihovih sintaksičko-semantičkih особености на основу којих би се могли пре tražивати и различити реcheničни модели.

2. Primena korpusa SrpKor2013 u lingvističkim istraživanjima

2. U studiji *Dopune prideva u savremenom srpskom jeziku* (Kiš 2016) posmatran je odnos komplementacije u pridevskoj sintagmi sa primarnim ciljem da se ukaže na osnovne sintaksičko-semantičke karakteristike odnosa komplementacije prideva u savremenom srpskom jeziku. Istraživanje je podrazumevalo da se najpre izdvoje oni pridevi koji dobijaju dopune prema inherentnoj leksičkoj semantici, u svim ili samo u nekim od svojih značenja, kako bi se mogla izvršiti semantička kategorizacija ovih prideva. Stoga je bilo je potrebno odabrati recentni rečnik srpskog jezika, mi smo se opredelili za šestotomni rečnik Matice

² U domaćoj i stranoj literaturi ima različitih određenja pojma *rekcija*. Mi ćemo ovaj odnos posmatrati kao „посебан, граматикализован, облик зависног односа који се првенствено сагледава на синтагматском нивоу, између две пунозначне (аутосемантичне) речи. Једна граматички доминантна реч – глагол, именица, придев – без обзира на то у ком је облику употребљена, управља датом везом, а друга је регирана” (Ружић у: Пипер и др., 2005: 537).

srpske (RMS), i na osnovu značenja prideva izdvojili one za koje se moglo očekivati da zahtevaju dopunu. Sledeći korak bio je pronalaženje svih izdvojenih prideva upotrebljenih u različitim jezičkim kontekstima. Pokazalo se da je svakako najadekvatniji izvor građe upravo elektronski *Korpus savremenog srpskog jezika*, SrpKor2013, dostupan na Matematičkom fakultetu Univerziteta u Beogradu,³ budući da je omogućavao pretragu u tekstovima koji pripadaju svim funkcionalnim stilovima savremenog jezika, a što je omogućilo praćenje aktuelnih zbivanja i tendencija u savremenom srpskom jeziku, posebno kada se uzmu u obzir i oni primeri koji nisu u skladu sa jezičkom normom. Pri tom, prilikom formiranja upita postoji mogućnost da se pretražuju samo izvorni tekstovi, kao i da budu uključeni i prevodi. Analizirano je blizu 600 prideva, dok korpus samog rada sačinjava 383 prideva,⁴ a iz elektronskog korpusa je ekscerpirano gotovo 9000 rečenica.

2.1. Prilikom pretrage korpusa bilo je potrebno izdvajati čitavu paradigmu jednog prideva, kao i sve rečenice u kojima se data lema u korpusu pojavljivala budući da je u analizi odnosa komplementacije relevantan prvenstveno jezički kontekst.⁵ Regularni izraz u pretrazi predstavlja gramatičku osnovu⁶ prideva, međutim kako bi bili izdvojeni i oblici prideva u nominativu muškog roda sa nepostojanim A, primenjivan je zasebni upit: **sigurn[a-z]*** i **siguran**. Korpus omogućava više načina pretraživanja ukoliko se koriste operatori grupisanja, npr. ako se žele pretražiti svi oblici flektivne paradigme imenice deca, ponuđena su dva načina: **deca|dece|deci|decu|deco|decom** ili **dec(a|e|i|u|o|om)**. Međutim, kako je paradigma prideva veoma razuđena zbog kongruencije sa imenicama u tri gramatička roda i dva broja, opredelili smo se za navedeni oblik upita iako je njegov nedostatak što su na taj način izdvajane i druge leme sa istim korenom, npr. *sigurnost*, *sigurnosni* i sl. Nakon uvida u moguće realizacije tražene pridevske lekseme i sintagmatske spojeve čiji je upravni ili glavni član pridev, pretraga je bila sužena tako što su desno od prideva dodavani ili predlozi ili pak prototipični veznik dopunske klauze

³ U svojoj doktorskoj disertaciji Miloš Utvić (2014: 247–303) detaljno predstavlja nastanak i razvoj elektronskog *Korpusa savremenog srpskog jezika*. Autor izdvaja parametre za verziju SrpKor2013 i pokazuje faze u kreiranju ovog korpusa: prikupljanje tekstova i njihovu digitalizaciju, obradu elektronskih tekstova, izvršene anotacije – bibliografska, strukturna i morfološka, indeksiranje tekstova i pripremu paralelizovanih korpusa.

⁴ Za realizaciju dopuna uz neke od prideva, uprkos njihovoj inherentnoj leksičkoj semantici, koja za to ima potencijala, i na osnovu koje su pridevi inicijalno izdvajani, nisu pronađene potvrde u korpusu, pa tako nisu bili ni obuhvaćeni analizom.

⁵ Uticaj nejezičkog konteksta svakako je važan, ali sam elektronski korpus ne pruža uvid u njega.

⁶ Gramatička osnova prideva dobija se kada se pridevu u genitivu jednine, bez obzira na rod, ukloni nastavak za genitiv (sigurn-og, sigurn-e, sigurn-og).

da. Sami predlozi bili su deo očekivane predložko-padežne konstrukcije u funkciji dopune prideva: **sigurn[a-z]* u; sigurn[a-z]* da**.

Iste večeri pevala sam sa ranjenicima , [sigurnija u](#) sebe , kao da sam raskinula sa nečim što je ležalo duboko u meni .

Prvo, kao što će biti objašnjeno u narednom poglavlju , mi još nismo sasvim [sigurni u](#) to koja tačno teorija uspešno objedinjuje opštu relativnost i kvantnu mehaniku,

Da li ste [sigurni da](#) želite da SAJLAS izvrši ovaj zadatak?

Na ovaj način iz pretrage bi bili isključeni komplementi u slobodnoj padežnoj formi – bez predloga, ili blokirani obaveznim determinatorom, a koji bi mogli biti izdvajani samo prilikom šire pretrage koja bi podrazumevala prvobitni regularni izraz upita.⁷ Sintaksička i semantička anotacija⁸ korpusa mogle bi unaprediti i pojednostaviti pretrage potrebne za različite vrste analiza na ovom jezičkom nivou, o čemu će biti više reči u narednom odeljku.

Nakon formiranja korpusa za istraživanje zadatak je bio opisivanje sintaksičkih i semantičkih osobina samih prideva radi utvrđivanja semantičkih i sintaksičkih tipova dopuna koji se pojavljuju u funkciji komplemenata pojedinih semantičkih klasa prideva.

2.1.2. Već ova faza istraživanja pokazala je potrebu izdvajanja *tvorbenih gnezda/porodica* kako bi se mogle pratiti sintaksičko-semantičke karakteristike korenske/motivne morfeme i motivisanih reči. Mogućnost pretrage u elektronskim korpusima preko formiranih tvorbenih gnezda omogućila bi različite vrste analiza i na drugim jezičkim nivoima. U tom smislu, kada se radi o morfološkoj anotaciji⁹ korpusa, od značaja bi bilo grupisanje lema koje predstavljaju prvo-stepene, drugo-stepene ... derivate izvedene iz jedne proste, korenske morfeme. Kao problem u pretragama tog tipa mogu se pojaviti *alomorfi* iste morfeme,¹⁰ na primer:

⁷ Sam pridev *siguran* ne regira dopune u vidu slobodnih padežnih formi, kao što je to slučaj sa pridevima tipa *pun, ravan, siromašan, svestan* i sl.

⁸ Pod pojmom anotacije korpusa podrazumeva se „postupak kojim se dodatne informacije pridružuju delovima korpusa (tekstovi, logičke celine u okviru teksta, tokeni) (Utvić 2014: 27). Prema vrstama podataka razlikuje se više nivoa anotacije, a ovom prilikom su u fokusu *sintaksička anotacija* (pridruživanje jednočlanom ili višečlanom nizu korpusnih reči zajedničke informacije o funkciji u rečenici (subjekat, predikat, objekat, glagolska odredba) ili obeležavanje takvog niza kao sintagme (imenička, pridevska, itd.)) i *semantička anotacija* (pridruživanje oznaka odgovarajućeg značenja svakom tokenu) (Utvić 2014: 28).

⁹ Miloš Utvić predstavio je u svom magistarskom radu osnovne modele morfološke obrade elektronskog teksta (2008: 22–25).

¹⁰ Na ovaj segment morfološke realizacije leksema skreće pažnju i Miloš Utvić (2008: 20–21).

ZVAti – ZOVnuti – ZIVkati; PLOT – PLETivo – prePLITati;
REĆi – REKoh – REĆi – REČ; VUĆi – VUKoh – VUČem – VUCi;
NOGa – NOŽi – NOŽni

2.1.3. Još jedan važan parametar prilikom analize procesa komplementacije i rekcijaskog potencijala prideva jeste rečenična funkcija u kojoj se realizuju. Naime, pridev može da dobije dopunu ukoliko se nalazi u poziciji leksičkog jezgra predikata. Stoga su analizom obuhvaćeni pridevi koji se pre svega nalaze u leksičkom jezgru kopulativnog i semikopulativnog predikata, ali i pridevi koji u rečenici imaju funkciju apozitiva ili predikativnog atributa, a koji su inače kondenzatori rečeničnog sadržaja. Navešćemo prideve upotrebljene u ovim pozicijama u rečenici, a primerima ćemo pokazati njihovu sposobnost vezivanja dopuna:

- Pridev u kopulativno-imenskom predikatu (Popularnu kulturu stvaraju podređeni koji *nisu pomirljivi* sa svojim položajem u društvu)
- Pridev u semikopulativnom predikatu (Kao *merodavne* za dokazivanje vlasništva, Udruženje smatra gruntovnice iz 1939. godine)
- Pridev kao apozitiv (Aprilska penzija, skromna i *nedovoljna* za najosnovnije potrebe, stigla je i već je, kažu, potrošena)
- Pridev kao predikativni atribut (Vojnici je s ljubavlju primaju, *zahvalni* na njenoj pažnji)

2.2. Govoreći o parsiranju, Miloš Utvić navodi da se ovaj proces u užem smislu „poistovećuje sa (automatskom) sintaksičkom analizom pomoću računara ‘čiji je zadatak prepoznavanje rečenice i pridruživanje (odgovarajuće) sintaksičke strukture’” (navedeno prema Utvić 2014: 106 [Jurafsky & Martin, 2008: 427]).

Parsiranje je bitan zadatak u automatskoj obradi teksta na prirodnom jeziku i često je neophodan preduslov za naknadne obrade i primene poput semantičke analize, analize diskursa, ekstrakcije informacija, mašinskog prevođenja, itd. Takođe, parsiranje je složen zadatak i postoji više različitih pristupa rešavanju tog problema, okupljenih u klase sintaksičkih teorija, među kojima su najznačajnije *gramatike frazne strukture* ili *konstituentne gramatike* (eng. phrase structure grammars, constituency grammars) i *gramatike zavisnosti* (eng. Dependency grammars)) (Utvić 2014: 106).

Kada se radi o analizi odnosa komplementacije na sintagmatskom nivou, naša dosadašnja istraživanja upućuju na zaključak da je parsiranje zasnovano na gramatici zavisnosti adekvatnija teorijska osnova.

Rezultat sintaksičke analize gramatikom zavisnosti se takođe predstavlja drvetom zavisnosti, pri čemu:

- čvorovi drveta, izuzev korena, su morfosintaksičke reči u rečenici;
- (usmerene) grane drveta predstavljaju zavisnosti i usmerene su ka zavisnoj reči,
- svaka grana je obeležena odgovarajućom gramatičkom funkcijom ili semantičkom relacijom;
- koren drveta je specijalan čvor koji je granom povezan sa glagolom kao centralnom („glavnom”) rečju u rečenici (Utvić 2014: 108).

U slučaju gramatika zavisnosti ([Mel’čuk, 1988]) nad skupom morfosintaksičkih reči u okviru rečenice uspostavlja se binarna relacija takva da je morfosintaksička reč x u relaciji sa morfosintaksičkom rečju y ako i samo ako y „zavisí” od x , pri čemu svaka od zavisnosti predstavlja neku gramatičku funkciju (na primer subjekat, objekat, itd.) ili opštu semantičku relaciju (agent, cilj, itd.) (Utvić 2014: 108).

Prethodna precizna lingvistička analiza procesa komplementacije, kao i drugih sintaksičko-semantičnih procesa i relacija u prostoј rečenici neophodan je uslov za preciznu sintaksičku anotaciju elektronskih korpusa.¹¹ Time bi se, na primer, prilikom određivanja svake od zavisnosti, odnosno svake gramatičke funkcije, mogla praviti razlika između njihovih podtipova: subjekat (gramatički i logički), objekat (direktni i indirektni – tipični i netipični) i dopune (obavezne, fakultativne i obavezni determinatori).

Istraživanja sintaksičkih i semantičkih karakteristika deadjektivnih odnosno pridevskih imenica trenutno su u toku. Jedan od važnih segmenata ovog istraživanja, koji se velikim delom odnosi na rekcijski potencijal deadjektivnih imenica, ima za cilj takođe izradu rekcijskog rečnika. Analizom je obuhvaćena 361 imenica (broj imenica je manji od broja prideva koji regiraju dopune jer se, primera radi, od nekih prideva ne mogu derivirati deadjektivne imenice (*drugačiji, znan, izmirljiv, isti, kadar, konten, ravan, saodnosan*), a sama pretraga elektronskog korpusa otežana je činjenicom da se rekcijski potencijal deadjektivnih imenica razlikuje u manjoj ili većoj meri od rekcijskog potencijala motivnih prideva. Za to postoje brojni, prvenstveno jezički faktori, ali se ovom prilikom na tome nećemo zadržavati.

¹¹ Osnova za sintaksičko parsiranje može se tražiti, između ostalog, u rekcijskim rečnicima koji su urađeni za veliki broj slovenskih jezika. Radi se mahom o rekcijskim rečnicima glagola, ali i imenica i prideva. Na konferenciji Leksikografija i leksikologija u svetlu aktuelnih problema održanoj 2020. godine u Beogradu, u organizaciji Instituta za srpski jezik SANU, autorka je predstavila model budućeg rekcijskog rečnika prideva (Киш 2021). Pregled ove vrste specijalnih rečnika u slovenskim lingvističkim tradicijama videti u: Киш 2021; Штрбац 2010.

Osnovni upit predstavlja regularni izraz koji čini paradigma jedne leme – to su u ovom slučaju svi padežni oblici deadjektivne imenice u jednini i množini. Upit čini gramatička osnova deadjektivne imenice koja se dobija odbijanjem padežnog nastavka u genitivu jednine: **bezobzirnost[a-z]***. Na ovaj način dobija se uvid u moguće sintaksičke, odnosno gramatičke funkcije ovih imenica u rečenici, inventar potencijalnih predložko-padežnih formi dopuna, uvid u dopune realizovane bez predloga, a moguće je pratiti u kojem se jezičkom kontekstu date imenice realizuju sa dopunama ali i bez njih. Precizna sintaksička anotacija korpusa mogla bi olakšati ovakve vrste pretraga ukoliko bi, na primer, upit omogućavao da se izdvajaju samo imenice prema rečeničnim funkcijama ili pak prema sintagmama čiji su upravni, glavni član.¹² Kako je moguće pretragu suziti samo na slučajeve kada deadjektivne imenice vezuju uz sebe dopune sa strukturnim likom B (imenica blokirana predlogom),¹³ regularni izraz formiran je na isti način kao prilikom pretrage prideva: npr. **bezobzirnost[a-z]* prema**.

Nasilnička vožnja među najtežim prekršajima Nasilnička vožnja, kao jedan od najtežih prekršaja prema novom zakonu, predstavlja nepoštovanje pravila i [bezobzirnost prema](#) ostalim učesnicima u saobraćaju.

Pri formiranju upita mora se uzeti u obzir i činjenica da, za razliku od prideva, dopune deadjektivnih imenica ne moraju biti u kontaktnom položaju sa imenicama (često se imenice realizuju sa različitim determinatorima, npr. determinator kojim se iskazuje nosilac osobine, stanja ili raspoloženja i sl.), tako da se podrazumeva drugačiji regularni izraz: **velikodusnost[a-z]* [{0,6} prema**.

To bi vrlo lepo istaklo nesebičnost i [velikodusnost pređašnjih očeva i uopšte pređašnjih ljudi prema](#) novoj, lakomislenoj socijalnoj omladini.

Ono što je za pretragu elektronskog korpusa relevantno jeste i mogućnost izvođenja više deadjektivnih imenica, najčešće dve, od istog

¹² Prilikom sintaksičke anotacije korpusa bilo bi korisno, na primer, uzeti u obzir dekomponovane i perifrastične predikate, koji bi se mogli posmatrati kao spojevi u kojima se ostvaruje različit stepen kolokacijske veze između semikopulativnih glagola i deverbativnih i deadjektivnih imenica u njihovom sastavu (npr. *držati govor, izneti predlog, nemati vrednost, uliti hrabrost* i sl.) (Радовановић / Бугарски, 2005: 206–207). O kolokacijama u korpusnoj lingvistici videti u: Utvić 2014: 195–198).

¹³ O strukturi padežnog sistema srpskog jezika videti detaljnije u: Ivić 1995: 195–203.

prideva. Ove imenice razlikuju se u manjoj ili većoj meri prema svojim semantičkim i sintaksičkim karakteristikama, a neretko jedan od oblika nije zabeležen u RMS, što je svakako predmet naše analize. Ovakvi slučajevi su, čini se, još jedna potvrda potrebe da se u elektronskom korpusu pretrage rade prema tvorbenim gnezdima/porodicama. Samo jedan od brojnih primera jesu imenice *adaptivnost* (ovaj oblik nije zabeležen u RMS) i *adaptacija*:

Bez obzira na diktirane uslove , neophodno je imati u vidu koliko duh naše tradicije ima realne [adaptivnosti](#) za recepciju onog što se zvanično prihvata.

Sredstva ove aproprijacije namenjena su za rekonstrukciju i [adaptaciju](#) verskih objekata na teritoriji Republike Srbije , obuhvatajući AP Kosovo i Metohija

... donošenje sektorskih planova i programa mera za [adaptaciju](#) na klimatske promene i usaglašavanje sektorskih strategija sa strategijama , belom knjigom EU za adaptaciju i relevantnim direktivama EU ;

2.3.1. Naredni primer dovodi nas do još jednog problema koji bi mogao dobiti svoje razrešenje u sintaksički i semantički anotiranom elektronskom korpusu. Od prideva *vredan* izvedene su dve imenice – *vrednost* i *vrednoća*.

vredan, -dna, -dno

1. marljiv, radan.
2. koji vredi, koji ima vrednost, dragocen.
3. a. koji zasluđuje (nešto), dostojan (čega) .
b. kadar, sposoban.

vrednost

1. svojstvo onoga što je vredno, što je na ceni.
2. ono što vredi.
3. važnost, značaj.
4. valjanost, punovažnost.
5. veličina, znatnost; dar, sposobnost.
6. mat. količina.

vrednoća

- osobina onoga koji je vredan, marljivost, radinost, valjanost.

Imenica *vrednoća* može se dovesti u semantičku vezu sa pridevima koji imaju značenje *marljiv, radan*, i poput prideva, dobija dopunu kojom se data osobina vezuje samo za određene situacije, domene života neke osobe u kojima se ispoljava. Značenja imenice *vrednost* motivisana su drugim značenjskim likom istog prideva, a sve se to ogleda na sintaksičko-semantičkom planu i u rekcijским konstrukcijama koje grade ove lekseme.

Ovo je rezultat vredan *uvažavanja*.

Te dve knjige su vredne da se pročitaju.

Kupili su akcije vredne 14 miliona dolara.

Ovo zapažanje je *od velike* vrednosti za istoričare

Svoj visoki položaj duguje svojoj vrednoći u učenju.

Kao relevantan parametar za razlikovanje ovih imenica pojavljuje se semantička kategorija *animatnosti* (živo+/-) pojma nosioca osobine. Naime, nosilac osobine označene imenicom *vrednoća* markiran je obeležjem živo+, dok s druge strane, za pojam kojem se pripisuje neka *vrednost* ovo obeležje nije relevantno.

3. Mogućnosti pretrage u sintaksički i semantički anotiranim korpusima

3. Ovi primeri vode nas do drugog zadatka ovog rada. Namera nam je da se iz ugla analize na nivou proste rečenice ukaže na potrebu proširenja mogućnosti pretrage anotiranih korpusa.

Pretraga korpusa bila bi svakako preciznija ukoliko bi se pri sintaksičkoj anotaciji uzeli u obzir podaci o valencijskim osobinama¹⁴ predikata rečenice i svim obavezanim i fakultativnim aktantima u konkretnoj situaciji, pri čemu bi trebalo uključiti i različite, relevantne sintaksičko-semantičke parametre kao što su: *animatnost, usmerenost/direktivnost, recipročnost, ablativnost, obuhvatnost, partitivnost* i sl. Na taj način, mogle bi biti izdvojene grupe prideva koji su markirani, primera radi, obeležjem *recipročnosti* i koji, upotrebljeni kao leksičko jezgro različitih vrsta analitičkog predikata, utiču na sintaksičku strukturu rečenice i semantičke relacije u njoj: A) pridevi **jednakosti** – *ekvivalentan, identičan, istovetan, jednak, kongruentan, homologan, korespondentan, podudaran, proporcionalan, ravan, ravnopravan, recipročan, saglasan, sazvučan, saobrazan, srodan, konkurentan, isti, izjednačen, konkordantan, konforman, sinoniman, ujednačen, uporediv, usklađen, adekvatan*; B) pridevi **nejednakosti, različitosti**

¹⁴ Valentnost u užem, jezičkom smislu, može se odrediti kao spajanje, povezivanje ili kombinovanje glagolskih i imenskih reči i izraza na sintaksičko-semantičkom planu. Ovakvo udruživanje reči zasnovano je na leksičosemantičkoj vezi među njima, ali i na gramatičkim pravilima koja važe za svaki jezik posebno. Broj učesnika u situaciji odnosno broj aktanata zavisi prvenstveno od valencije predikata (Ружић у: Пипер и др., 2005: 517).

– *kontradiktoran, nesličan, protivan, protivrečan, obrnut, suprotan, nesaglasan, nesrodan, oprečan, kontraran, različit, drugačiji, nepodudaran*; C) pridevi **prostornih odnosa** – *naporedan, paralelan, simetričan, pripijen, priljubljen*; D) pridevi sa značenjem **socijativnosti** – *dobar, intiman, nerazdvojan, nesložan, blizak, zaručen, inkompatibilan, komplementaran, korelativan, uzdržan, nepovezan, nespojiv, nevezan, neuporediv, suplementan, kompatibilan, saodnosan*.

Kada je pridev *jednak* deo imenskog predikata, u rečenici će biti dva obavezna aktanta između kojih se uspostavlja relacija iskazana pridevom – nosilac osobine i drugi nosilac osobine. Oni mogu biti različito raspoređeni u rečenici, odnosno mogu biti u tačno utvrđenim funkcijama u rečenici, pri tom sintaksičko-semantičko obeležje *animatnost* (živo+/-) utiče na mogućnost uspostavljanja relacije između pojmova sa obeležjem živo+ ili živo-:

- oba učesnika u poziciji gramatičkog subjekta (*Ana i Ivana su jednake, Trougao A i trougao B su jednaki*) ili je prvi učesnik u poziciji gramatičkog subjekta, a drugi je u poziciji dopune sa značenjem koučesnika u situaciji (*Ana je jednaka Ivani / sa Ivanom; Trougao A jednak je trouglu B / sa trouglom B*).

Relacioni pridevi *paralelan* i *zaručen* pojavljuju se u rečenicama strukturiranim prema istom modelu, ali pridev *paralelan* podrazumeva da učesnici u relaciji imaju obeležje živo-, dok pridev *zaručen* uspostavlja odnos među učesnicima sa oznakom živo+:

- *Duž A i duž B su paralelne; Duž A paralelna je duži B / sa duži B*
- *Marko i Ivana su zaručeni; Marko je zaručen za Ivanu / sa Ivanom*.

3.1. Ako pogledamo pridev *veran* i jedan od primarnih derivata, imenicu *vernost*, videćemo razlike u njihovom rekcijском potencijalu. Relacioni pridev *veran* vezuje uz sebe objekatsku dopunu u dativu bez predloga sa oznakom živo+/- (on je *veran prijatelju / toj ideologiji*), imenica *vernost* dobija isti tip objekatske dopune realizovane različitim formama (pokazao je *vernost prijatelju / prema prijatelju*), ali se može u određenom kontekstu upotrebiti i bez dopune (oni ne mogu da govore o *vernosti*).

3.2. Posebnu pažnju treba skrenuti na višeznačnost leksema koja se odražava i na rekcijски potencijal reči budući da jedna leksema u svojim različitim semantičkim realizacijama može imati i različita sintaksičko-semantička obeležja što može biti problem kako u semantičkoj tako i u sintaksičkoj anotaciji korpusa.¹⁵ Pridev *veran* i imenica *vernost*, ukoliko

¹⁵ Višeznačnost i obrada „nep(rep)oznatih” reči jesu glavni problemi u automatskoj morfosintaksičkoj anotaciji korpusa, pri čemu se u ovom kontekstu može reći da „višeznačnost (eng. ambiguity) nastupa kada se jednom tokenu može pridružiti više morfosintaksičkih opisa koji se međusobno isključuju” (Utvić 2014: 143).

nosilac osobine ima oznaku živo-, mogu imati značenje „biti jednak nečemu“, što implicira da će dopuna koju regiraju pripadati drugačijem semantičkom tipu, odnosno imaće značenje drugog nosioca osobine (tekst je *veran originalu*; izdavač se držao *vernosti prema tekstu*).

3.3. Iz ugla govornika srpskog kao maternjeg ili kao nematernjeg/stranog jezika bilo bi korisno da se u pretragama elektronskog korpusa može poći od semantičkih klasa reči čiju bi formalizaciju, odnosno strukturne modele u kojima se realizuju pokazali primeri ekscerpirani iz elektronskog korpusa (npr. u klasi reči sa značenjem optativne modalnosti nalaze se pridev *željan*, imenica *želja* i glagol *želeti*, koji se mogu realizovati u sledećim rečeničnim modelima: Ja sam *željna odmor* – Moja *želja za odmorom* je velika / Ja imam *želju za odmorom* – Ja *želim da se odmorim / odmor*).

3.4. Posebno važno i svakako kompleksno pitanje jeste način formalizacije različitih semantičkih uloga u rečenici, o čemu govori Miloš Utvić predstavljajući proces parsiranja zasnovanog na gramatici zavisnosti (2014: 108–111). Tako je, na primer, u strukturiranju rečenice, relevantno i pitanje kojim se jezičkim sredstvima iskazuju podaci o agensu/pseudoagensu. U savremenom srpskom jeziku izdvajaju se različiti strukturni tipovi s obzirom na ove parametre, tako da, ako u rečenici postoji doživljavač/eksperiment, rečenica može biti strukturirana na više načina: *ja bih jela jabuke* / *ja želim da jedem jabuke* / *jedu mi se jabuke* i sl.).

4. Zaključak

Iz svega rečenog, može se videti koliko su sintaksička i semantička anotacija elektronskog korpusa izuzetno kompleksni procesi koji zahtevaju postojanje teorijskih osnova analize vezanih za svaki pojedinačni prirodni jezik. Gramatike pojedinačnih jezika, čak i ako se radi o istim jezičkim porodicama, na sintaksičko-semantičkom planu često se značajno razlikuju. To svakako ima uticaja na način obrade i parsiranja tekstova. Stoga je, s jedne strane, saradnja između lingvista i onih koji se bave programerskim poslovima neophodna, kako bi se ukazalo na potrebne jezičke analize koje bi omogućile i dale osnov za efikasniju i precizniju anotaciju korpusa. S druge strane, mogućnosti primene sintaksički i semantički anotiranih korpusa u različitim jezičkim istraživanjima, ali i u nastavi jezika, ne samo kao maternjeg, već i kao nematernjeg, odnosno stranog, u značajnoj bi meri unapredile i učinile efikasnijim istraživačke i nastavne procese.

Izvori

- [1] *Речник српскохрватског књижевног језика* (1967–1976). Нови Сад / Загреб: Матица српска / Матица хрватска (Матица српска – Матица хрватска 1-3, Матица српска 4-6).
- [2] Vitas, D. i Utvić, M. *Korpus savremenog srpskog jezika* (verzija SrpKor2013) Copyright Grupa za jezičke tehnologije Univerziteta u Beogradu. <https://noske.jerteh.rs/#dashboard?corpname=SrpKor2013>

Literatura

- [1] Ivić, M. (1995). *Lingvistički ogledi*. Biblioteka XX vek, knj. 61. Slovoğraf.
- [2] Jurafsky, D. & Martin, J. H. [2008]. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*, second ed., Prentice Hall.
- [3] Киш, Н. (2016). *Допуне придева у савременом српском језику*. Едиција Е-дисертација, књ. 11. Филозофски факултет. <http://digitalna.ff.uns.ac.rs/sadrzaj/2016/978-86-6065-397-2>
- [4] Киш, Н. (2021) Рекцијски речник придева. У С. Ристић, И. Лазе Коњик, Н. Ивановић (ур.) *Лексикографија и лексикологија у светлу актуелних проблема* (стр. 969–983). Институт за српски језик САНУ.
- [5] Mel'čuk, I. (1988). *Dependency Syntax: Theory and Practice*. Suny Series in Linguistics. State University Press of New York.
- [6] Пипер, П. и др. (2005). *Синтакса савременог српског језика. Проста реченица*. Институт за српски језик САНУ – Београдска књига – Матица српска.
- [7] Радовановић, М. и Бугарски, Н. (2007). Номинализације: два лика. У *Зборник Матице српске за славистику* 71–72 (стр. 199–209).
- [8] Штрбац, Г. (2010). Специјални граматички речници у лексикографији. У В. Ружић, Д. Маринковић (ур.) *Годишњак Филозофског факултета у Новом Саду XXXV/2* (стр. 369–379). Филозофски факултет.
- [9] Utvić, M. (2008). *Konačni automati u regularnoj imenskoj derivaciji* (nepublikovana magistarska teza). Univerzitet u Beogradu chromeextension: //efaidnbmnnnibpccajpcglclefindmkaj/<https://informatika.fil.bg.ac.rs/misko/bibliografija/2008/msMilosUtvic.pdf>
- [10] Utvić, M. (2014). *Izgradnja referentnog korpusa savremenog srpskog jezika* [Doktorska disertacija, Univerzitet u Beogradu]. [NaRDuS] <https://nardus.mpin.gov.rs/handle/123456789/4091>

Syntactic-semantic Annotation of Serbian Language Electronic Corpora

Nataša Kiš

Summary

This paper explores the potential of using electronic corpora as a primary source of data for various studies in the fields of syntax and semantics of contemporary Serbian. It has two main objectives. The first is to present previous research on the complementation of adjectives and adjectival nouns – specifically, their government potential at the syntagmatic level – based on relevant material drawn primarily from the *Corpus of Contemporary Serbian Language*, SrpKor2013 (www.korpus.matf.bg.ac.rs).

The second objective is to emphasize the need to expand the search capabilities of electronic corpora to support syntactic-semantic analysis at the level of the simple sentence. Such searches rely primarily on syntactic and semantic corpus annotation and would include, among other features, data on the valency properties of predicates and on both obligatory and optional arguments within specific contexts. A key component of this process is the classification of semantic word classes and their syntactic-semantic properties, which would enable targeted searches for different sentence models.

One of the paper's goals is to outline the types of linguistic analyses required to extract and integrate relevant data for the further development of syntactic and semantic corpus annotation. The study aims to demonstrate the complexity of these annotation processes, which necessitate solid theoretical foundations specific to each natural language. Even within the same language families, grammars may vary significantly at the syntactic and semantic levels – differences that directly impact text processing and parsing strategies.

Therefore, collaboration between linguists and computational experts is essential to ensure that linguistic insights inform more accurate and efficient annotation. At the same time, the application of syntactically and semantically annotated corpora in linguistic research and language education – whether in teaching Serbian as a first, second, or foreign language – could substantially enhance the effectiveness and scope of both research and pedagogical practices.

Keywords: syntax, semantics, annotation, electronic corpora, governing ability, Serbian language