



Matea Birtić

Siniša Runjaić

Institut za hrvatski jezik i jezikoslovlje, Zagreb

USTROJ BAZE HRVATSKIH GLAGOLSKIH VALENCIJA

Prethodno priopćenje

U ovom je radu predstavljena struktura Baze hrvatskih glagolskih valencija koja se izrađuje u Institutu za hrvatski jezik i jezikoslovlje. Baza će biti podloga javno dostupnom sustavu za pretraživanje i učenje glagolskih valencija, a poslužit će i za izradu tiskanoga rječnika glagolskih dopuna. Osmišljen je trofazinski način unosa obrađenih podataka, a podatci se unose u pripremljeni obrazac u programskom paketu Tschwanelex. Podatci uneseni u aplikaciju zapisuju se u obliku strukturirane baze podataka u okruženju SQL. Nakon proučavanja različitih modela odlučeno je da se naš model osloni na već dostupne valencijske opise hrvatskoga jezika i na njemačku valencijsku tradiciju, pa je kao predložak našem modelu izabran njemački valencijski rječnik VALBU.

Ključne riječi: hrvatski jezik, sintaksa, valentnost, baza glagolskih valencija, valencijski rječnik

1. Kako je nastala Baza hrvatskih glagolskih valencija?

Baza hrvatskih glagolskih valencija izrađuje se u Institutu za hrvatski jezik i jezikoslovlje kao interni projekt u sklopu temeljnih djelatnosti Odjela za opće jezikoslovlje. Projekt se nastavlja na projekt *Valencijski rječnik hrvatskih glagola* započet u Odjelu za hrvatski standardni jezik. Na sintaktičkoj i leksikografskoj obradbi sudjeluje šestero suradnika, uz dvojicu kolega kao informatičku potporu.

U ljeto 2013. godine započelo se s osmišljavanjem strukture baze i cjelokupnoga projekta. Trebalo je uzeti u obzir dotadašnja istraživanja i obradbe glagola, ali i misliti o budućem što jednostavnijem radu, odnosno izabrati model koji će se i lingvistički i informatički najlakše pretvoriti u bazu podataka. Istodobno smo se bavili iskustvima i rezultatima do kojih se došlo pri uspostavi sličnih baza u suvremenom okruženju diljem Europe, posebice rezultatima rada na valencijskim bazama u slavenskim i bliskim jezicima. Na umu smo imali sljedeće zahtjeve:

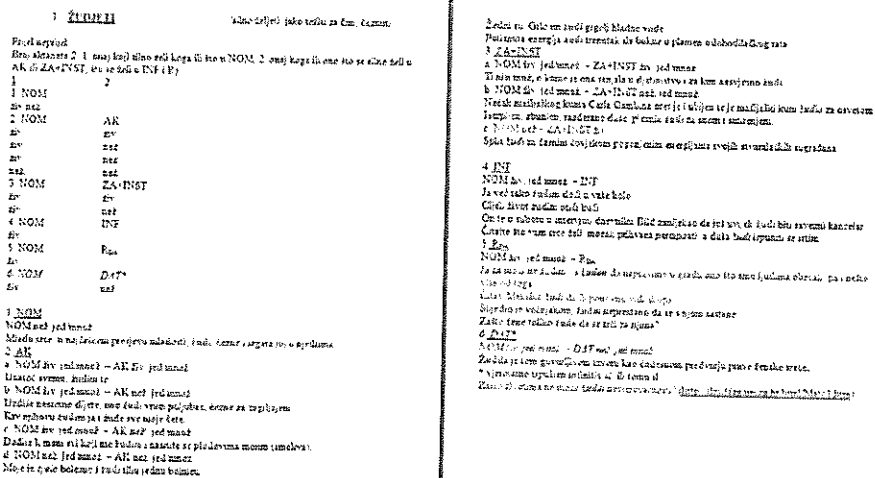
1. djelomice slijediti strukturu podataka u opisnim datotekama s glagolima obrađenima na projektu *Valencijski rječnik hrvatskih glagola*
2. izbjeći pogreške maksimalističkoga pristupa obradbi velikoga broja lema u odnosu na broj ljudi i planiranih radnih sati
3. odabrati lingvistički model koji će slijediti hrvatsku tradiciju opisa valentnosti, a također biti jednostavan u teoriji i primjeni
4. u dogovoru s informatičarima odlučiti je li produktivnije i ekonomičnije izraditi sučelje za unos i tablice u zamišljenoj bazi podataka ili odabrati leksikografski programski paket u kojem bi sve relacije i vrste podataka bile izvedive.

U nastavku ćemo prikazati kako smo ispunili zahtjeve i dobili strukturu podataka u obrascu za unos kakvim se danas služimo te ustroj pretražive baze podataka.

1.1. Prijašnja struktura podataka

Do kraja 2012. godine na projektu *Valencijski rječnik hrvatskih glagola* bilo je obrađeno oko 2000 glagola podijeljenih u semantičke skupine i zapisanih u obliku linearnih tekstnih podataka (v. slika 1). Takvi zapisi nisu mogli biti podloga tiskanom rječniku, a nisu se mogli ni izravno pretvoriti u pretraživu bazu podataka. Stoga smo od ljeta 2013. godine počeli razrađivati strategiju pretvorbe obrađenih glagola u bazu glagolskih valencija i na kraju odlučili pojedine skupine jezičnih podataka uzeti u obzir kao dio strukture za novi način unosa.

S jezikoslovne smo strane utvrdili da dotadašnji način obradbe nije bio ekonomičan jer se izravno povezivao morfološki sa semantičkim opisom te se za svaku valencijsku dopunu tražilo živo i neživo te jedninsko i množinsko ostvarenje. Traženje žive i nežive inačice za svaku dopunu iziskuje previše korpusnoga istraživanja, a ne daje odgovarajuće rezultate. Uvidjeli smo da se množina i jednina ili živost i neživost moraju tretirati kao ograničenja, tj. utvrdilo se da se takav opis može uopćiti uvođenjem pojmovnih (semantičkih) kategorija koje se pripisuju valencijskim dopunama. Na slici 1. prikazan je prijašnji način obradbe, za koji danas smatramo da nije prikladan.



Slika 1. Ilustracija obradbe glagola *žudjeti* u *Microsoft Wordu*

Premda smo htjeli izbjeći uvođenje klasičnog sustava dopuna¹, postalo je jasno da izravno povezivanje morfologije sa semantikom opterećuje sustav velikim brojem primjera. Tako se primjerice u obradbi glagola *stanovati* jedno sintaktičko mjesto opisivalo trima različitim obrascima:

Ona stanuje <u>u gradu</u> .	NOM--u+lok
Ona stanuje <u>tamo</u> .	NOM--pril
Ona stanuje <u>negdje oko kolodvora</u> .	NOM--pril. skupina

Takvi se obrasci mogu sažeti jednom (u ovome slučaju priložnom) obvezatnom valencijskom dopunom te su se naša promišljanja okrenula prema opisu sa sustavom sintaktičkih dopuna. Detaljnije raslojavanje značenja te određivanje pripadnosti kojega glagola semantičkoj skupini različitoj od temeljne² bili su nužni koraci u budućoj obradbi glagola.

1.2. Preinake abecedarija i ograničavanje broja glagola

Istodobno s kreiranjem nacrtu baze odlučili smo smanjiti broj glagola za obradbu (abecedarij). Smatrali smo da abecedarij s jedne strane mora zadovoljiti uvjet što bolje funkcionalnosti, tj. potrebe određenih korisnika, a s druge mora zadovoljiti zahtjeve vremenskoga ograničenja.³ Budući da smo stekli negativno iskustvo s maksimalističkim pristupom broju glagola, ograničili smo se na otprilike 700 glagola. Odluka je motivirana dvjema činjenicama: prvo, rječnik *VALBU* kao naš lingvistički uzor obrađuje 638 glagola.⁴ Drugo, prema istraživanjima članova našega projekta, teorijski je za razumnije 90 % tekstova određenoga jezika potrebno poznavati oko 3500 riječi. Među 3500 najčešćih riječi glagoli obuhvaćaju 20 % vokabulara te ih je među 3500 riječi oko 700.⁵ Stoga smo tijekom istraživanja krenuli sučeljavati popis najčešćih glagola iz *Hrvatskoga čestotnog rječnika* (Moguš, Bratanić i Tadić 1999) i kontrolnoga popisa glagola koji su suradnici izradili na temelju udžbenika za učenje hrvatskoga kao drugoga i stranoga jezika. Presjekom tih izvora dobili smo popis glagola za koji smatramo da predstavlja razmjerno realnu sliku najfrekventnijih glagola u hrvatskom jeziku te sada radimo s operativnim popisom od 900 glagola. Svi su glagoli s toga popisa raspoređeni u revidirane semantičke skupine,

¹ Pod pojmom klasičnoga sustava dopuna podrazumijevamo sustav sintaktičkih dopuna koji se rabi u opisima valentnosti nastalim u okvirima njemačke valencijske tradicije (npr. Schumacher et al. 1986; Engel 1988; Schumacher et al. 2004) i njegovim izdancima za hrvatski jezik kao što je npr. Samardžija 1986a.

² Tako primjerice glagol *dobaciti* u svojem temeljnom značenju pripada skupini glagola bacanja, ali jednim svojim značenjem pripada glagolima govorenja.

³ U prvome smo redu na umu imali naprednije korisnike kojima je hrvatski materinski jezik, ali u širem kontekstu i potencijalne studente ili kolege istraživače koji uče ili proučavaju hrvatski jezik kao strani jezik. Potencijalno valencijska baza u suradnji sa zainteresiranim stručnjacima može predstavljati i doprinos razvoju jezičnih tehnologija, odnosno poboljšanju točnosti alata za sintaktičko obilježavanje korpusa hrvatskoga jezika.

⁴ U međuvremenu je dovršen i objavljen rječnik (Đorđević i Ulrich 2013) sa sličnom teorijskom podlogom i s oko 800 glagola srpskoga jezika.

⁵ Sažet i prilagođen dio iz Bošnjak Botica 2011.

izradene prema modelu Beth Levin (Levin 1993)⁶ te dodatno prilagođene hrvatskom jezičnom materijalu.

1.3. Novija povijest istraživanja valencija u hrvatskome jeziku

Budući da se naš projekt djelomice nastavlja i na hrvatsku tradiciju te nije prvi valencijski rječnik, a ni prva računalna baza glagolskih valencija u Hrvatskoj, prije konačnoga stvaranja baze usporedili smo naš valencijski opis s opisima valentnosti glagola u hrvatskom jezikoslovlju. Novija povijest istraživanja valentnosti započinje osamdesetih godina 20. stoljeća s doktoratom Marka Samardžije *Valentnost glagola u suvremenom hrvatskom književnom jeziku* iz 1986. godine. Slijedi *Rječnik valentnosti hrvatskih glagola* (RVHG) iz 1993. godine⁷ nastao kao rezultat rada na projektu *Kontrastivno proučavanje hrvatskoga jezika prema stranim jezicima*. U novije doba značajan je doprinos doktorat Krešimira Šojata (Šojat 2008), koji uz informativan teorijski uvod sadržava primjenu sintaktičkih opisa u glagolskom leksikonu *Hrvatskoga WordNeta (CroWN)*.⁸ Iste godine nastaje i *Crovallex*⁹, javno dostupan leksikon hrvatskih glagolskih valencija koji je na temelju svoje disertacije izradila Nives Mikelić Preradović. Leksikon je revidiran 2010. godine, a modelom slijedi češku valencijsku tradiciju iz druge polovice 20. st., koja se temelji na *funkcijskom generativnom opisu* (engl. *functional generative description*) i rječniku *Valency Lexicon of Czech Verbs VALLEX*¹⁰.

1.4. Informatičko rješenje

Usporedno s izborom jezikoslovnoga modela i slaganjem sheme pretvorbe, odlučivali smo o pristupu: izraditi vlastito sučelje za unos i potom podatke povezivati u relacijske tablice u okruženju SQL ili izabrati leksikografski programski paket koji bi ispunjavao sve naše zahtjeve. Budući da je u pozadini ipak izrada rječnika, a i zbog višegodišnjega iskustva s radom u softveru *Tschwanelex*, odustali smo od ideje stvaranja vlastitoga sučelja za unos jer smo istraživanjem utvrdili da se podatci uneseni u aplikaciju *Tschwanelex* zapisuju u željeni oblik relacijske baze (SQL) iz koje možemo razviti budući sustav

⁶ Taj smo model izabrali s obzirom na to da smo tijekom prijašnje obradbe raspoređivali glagole u skupine prema vezi semantičkih i sintaktičkih obilježja istraživanih glagola. Pri reviziji popisa glagola i podjeli na semantičke skupine, što je bio i temeljni kriterij za plan obradbe u *Bazi hrvatskih glagolskih valencija*, prilagodili smo imena skupina temeljne klasifikacije zbog otprije uočene povezanosti semantičkih i sintaktičkih obilježja. Pritom smo imali na umu dvije činjenice: a) da nam višerazinski opis omogućuje i pridruživanje određenoga značenja pojedine glagolske leme, odnosno podleme drugoj semantičkoj skupini u odnosu na temeljno značenje te b) da ćemo u budućnosti tako povezane podatke o semantičko-sintaktičkim vezama moći prilagoditi i prikazati kao pretražive modele učenja za neizvorne govornike hrvatskoga jezika. (Za detaljan opis toga cjelokupnog procesa v. Brač i Bošnjak Botica 2015).

⁷ Rječnik sa 70 glagola, kojem su autori Ivan Bauer, Nives Opačić, Ivo Pranjković i Marko Samardžija objavljen je 1993. kao dio 4. sveska knjige *Kontrastivna analiza engleskog i hrvatskog jezika (Contrastive analysis of English and Croatian)* (od 75. do 109. str.) u izdanju Zavoda za lingvistiku Filozofskoga fakulteta Sveučilišta u Zagrebu. Ta četverodijelna knjiga (od 1975. do 1993.) predstavlja rezultate projekta *Kontrastivno proučavanje hrvatskoga jezika prema stranim jezicima*. Posljednji 4. svezak pod naslovom *Teorija valentnosti i Rječnik valentnosti hrvatskih glagola (The valence theory and the valency dictionary of Croatian verbs)* uredio je Rudolf Filipović te uz RVHG sadržava i pet studija o valentnosti hrvatskih jezikoslovaca.

⁸ Dostupno na <http://meta-share.ffzg.hr/repository/browse/croatian-wordnet/32d93d48703d11e28a985cf2c4e6c59e166cc06132a740ebb36c515b093096b2/#1>.

⁹ Dostupno na <http://theta.ffzg.hr/crovallex/> u inačici 2.0008.

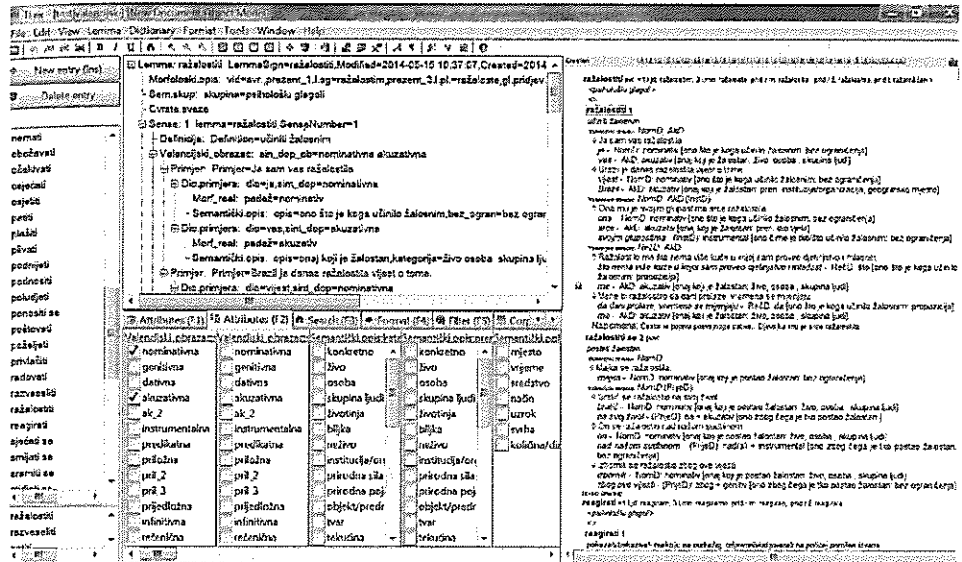
¹⁰ Dostupno na <http://ufal.mff.cuni.cz/legacy/vallex/2.6.3/doc/home.html> u inačici 2.6.3.

3)⁶ te dodatno prilagođene hrvatskom jeziku

na projektu *Kontrastivno proučavanje hrvatskoga jezika* u Hrvatskoj, pri čijem je izradi sudjelovao i profesor Ivan Preradović. Leksikon je revidiran 2010. godine i izdano pod nazivom *1. funkcional generative description*

slaganjem sheme pretvorbe, odlučivši se za programski paket koji bi ispunjavao zahtjeve za izradu rječnika, a i zbog višegodišnjeg iskustva u stvaranju vlastitoga sučelja, odlučili su se za aplikaciju *Tschwaneler* iz koje možemo razviti budući sustav obradbe raspoređivali glagole u skupine prema vrsti i podjeli na semantičke skupine *hrvatskih valencija*, prilagodili smo imena skupina i sintaktičkih obilježja. Pritom smo imali na umu i potrebu za pružanje određenoga značenja pojedine glagolske skupine, a to je da ćemo u budućnosti ta značenja prikazati kao pretražive modele učenja u okviru projekta *Kontrastivno proučavanje hrvatskoga jezika* i *Rječnik valentnosti hrvatskih glagola* (uradio je Rudolf Filipović te uz RVHG sadržava

pretraživanja. Ubrzo se odluka pokazala ispravnom jer je proces od ideje do realizacije pretvorbe bio kratak i prva je inačica operativnoga sučelja dovršena do kraja 2013. godine. Slijedilo je razdoblje početkom 2014. u kojemu su se polja i relacije u sučelju za unos dodatno doradivali i unapređivali. Današnji je oblik stabilna radna verzija (v. slika 2) u koju ćemo uvoditi dodatna poboljšanja, ali zacijelo ne više na razini lingvističkoga ustroja već pojedinačnih polja.



Slika 2. Ilustracija obradbe glagola *ražalostiti* u *Bazi hrvatskih glagolskih valencija*

2. Konačno oblikovanje lingvističkoga modela

Nakon proučavanja literature i konačnih rezultata naš se lingvistički model općenito temelji na modelu njemačkoga valencijskog rječnika *VALBU* (Schumacher et al. 2004), čija se pretraživa mrežna inačica nalazi na stranicama Instituta za njemački jezik u Mannheimu.¹¹ Time nastavljam i domaću valencijsku tradiciju jer se u *Rječnik valentnosti hrvatskih glagola* teorijski oslanjao na njemačke uzore, a time i posredno na Tesnièreovu sintaktičku ovisnosnu teoriju. No postoje važne razlike u kojima se naš opis na svim razinama udaljuje od ostalih.¹² Rješenja smo uspoređivali i s modelima koje su odabrali suvremeni autori (Mikielić Preradović 2008; Šojat 2009; 2012), a temelje se na različitim teorijskim pogledima od ovisnosne teorije i sukladno tomu različitim rješenjima (*teorija okvira* u hrvatskome *Wordnetu* i *funkcijski generativni opis* u *Crowallexu*).

Pri tome naglašavam da izborom valencijskoga modela nismo automatski razriješili opća jezikoslovna pitanja koja su prisutna u svim valencijskim opisima, iako se djelomice razlikuju s obzirom na jezik koji se opisuje valencijskom bazom ili rječnikom (ili polazišni jezik u višejezičnoj bazi ili rječniku):

¹¹ Dostupno na <http://hypermedia.ids-mannheim.de/evalbu/index.html>.

¹² Pojedine se razlike opisuju unutar prikaza razina opisa u bazi podataka u sljedećem poglavlju.

ian-wordnet/32d93d48703d11e28a985cf2c4c6c59

ne.html u inačici 2.6.3.

- a) Hoće li povratne inačice glagola (ili povratni glagoli) koji imaju i prijelazni oblik biti zasebne leme (leksički unos) ili ne?
- b) Hoće li se vidski parnjaci tretirati kao različiti glagoli, tj. hoće li vidski parnjaci istoga glagola biti uvedeni jednom lemom (leksičkim unosom) ili će se obrađivati kao posebne leme (leksički unos)?
- c) Ako se uvodi sustav sintaktičkih dopuna, kako osmisliti jasne kriterije određivanja obvezatnosti i neobvezatnosti pojedine dopune?

Naša su rješenja tih pitanja uvjetovana hrvatskim jezičnim materijalom te slijede predložena rješenja proučenih valencijskih opisa. Stoga smo odlučili kao posebne leme uvoditi samo nepravne povratne glagole, ali i svaki vidski parnjak. *Crovallex* primjerice slijedi češku tradiciju unošenja svakoga povratnoga glagola kao posebne leme, ali kao i naš pristup ne slijedi češku tradiciju u svođenju svih glagolskih vidskih izvedenica pod jedan leksički unos.¹³ Budući da se vidski parnjaci mogu razlikovati u značenjima, pa time i pripadajućim valencijskim obrascima, smatrali smo opravdanim uvođenje svakoga vidskog parnjaka kao posebne leme. Sustav sintaktičkih dopuna, koji s jedne strane olakšava valencijske opise dopuštajući veću apstraktnost od izravnoga pridruživanja morfologije sintaksi, s druge strane usložnjuje valencijske opise razlikovanjem obvezatnih i neobvezatnih dopuna,¹⁴ za koje još uvijek ne postoje jednoznačni testovi koji mogu obuhvatiti sve slučajeve.

3. Glavne značajke trirazinske Baze hrvatskih glagolskih valencija

3.1. Prva razina

Prva razina sastoji se od glagolske leme uz koju se vežu tri osnovna dijela: morfološki blok, pripadnost semantičkoj skupini i rubrika čvrste sveze. Glagolske su leme infinitivi glagola, osim u slučaju tzv. nepravilnih povratnih glagola kada se uz infinitiv piše i čestica *se* (*diviti se*).

Morfološki blok uključuje oznaku vida i glagolske oblike, a iza morfološkoga bloka slijedi pripadnost glagola temeljnoj semantičkoj skupini. Pripadnost glagola semantičkoj skupini bilježi i *Crovallex*, ali ne i *VALBU*, što je prva točka u kojoj se razlikujemo od njemačkoga predloška. U rubriku *Čvrste sveze* upisujemo kolokacijske i frazemske veze glagola i drugih riječi. Uz svaku čvrstu vezu donosimo objašnjenje i primjer (*mrziti iz dna duše* 'jako mrziti koga ili što' ◊ Ako idete na posao, mrzit ćete budilicu iz dna duše.).

¹³ *Leksički unos* prijevod je termina *lexical entry* koji u češkome modelu označuje ulaznu lemu koja se može ostvariti s više glagolskih vidskih izvedenica.

¹⁴ Obvezatne i neobvezatne dopune obično se u hrvatskoj literaturi nazivaju obvezatnim/obaveznim i fakultativnim dopunama (Samarđžija 1986; 1993; Šojat 2008). Šojat (2008) kaže kako se pojam obvezatnih i neobvezatnih dopuna eksplicitno ne razrađuje kod Tesničera već ga detaljnije obrađuju 1973. Helbig i Schenkel u uvodu svojemu rječniku. „[Nužni rečenični elementi se]... dijele na one koji se obavezno moraju pojaviti u okolini određenog glagola u nekoj rečenici, te na one koji su također u dosegu glagolske valentnosti, ali njihovo pojavljivanje nije obavezno (tzv. fakultativni)...” (Šojat 2008, 22). U uvodu rječnika *VALBU* (Schumacher et al. 2004) tvrde da su obvezatne i neobvezatne (fakultativne) dopune posljedica strukturnih svojstava glagola, a ne porabe glagola u određenome kontekstu (isto, 27). U našem pristupu razlikovanje obvezatnih i neobvezatnih dopuna nastoji se što je više moguće smanjiti izdvajanjem sintaktičkih dopuna na razini značenja glagola.

3.2. Druga razina

Druga razina obuhvaća izdvajanje značenja glagola, a pojedina se značenja uvode upisom podleme s obročanjem (*sramiti se 1, sramiti se 2*). Uz podlemu se veže oznaka povratnosti, stilski oznaka (ako je različita od neutralne), mogućnost promjene semantičke skupine glagola i definicija glagola. Samo tzv. nepravni povratni glagoli imaju upisanu česticu *se već* u lemi, a svi se ostali glagoli upisuju kao nepovratne leme. Povratna se inačica upisuje kao jedno od značenja nepovratne leme, primjerice glagolska lema *dosadivati* ima sljedeće podleme: *dosadivati 1, dosadivati 2, dosadivati se 3*. U njemačkom se modelu svaki povratni glagol, bez obzira na to ima li ili nema nepovratnu inačicu, upisuje kao posebna lema i stoga je obradba povratnih glagola druga značajka prema kojoj se razlikujemo od njemačkoga predloška. Pojedini podlemama pridružuje se opisna definicija. Značenja se definiraju parafrazno, a potom se upisuje eventualni sinonim.

3.3. Treća razina

Poslije opisa značenja ulazi se na treću razinu koja sadržava valencijski opis koji se ostvaruje valencijskim obrascem ili obrascima (tj. valencijskim uzorkom ili uzorcima). Svakom se značenju može pridružiti više valencijskih obrazaca, a unutar jednoga valencijskog obrasca može se opisati više primjera. Vidljivo je da pretpostavljamo da se valencijski uzorci moraju izdvajati na razini značenja, a ne na razini glagolske leme. Uz valencijski obrazac veže se oznaka normativnosti i rubrika ograničenja (u koju se opisuju ograničenja koja može imati određena dopuna, primjerice množina ili zbirnost akuzativne dopune glagola *sabrati*).

Valencijski uzorci sastoje se od sintaktičkoga, morfološkoga i semantičkoga opisa. Sintaktički opis slijedi sustav njemačkih dopuna s razumljivim prilagođivanjima koja su posljedica hrvatske sintakse. Kao i u većini njemačkih tradicionalnih valencijskih opisa,¹⁵ pretpostavlja se postojanje obvezatnih i neobvezatnih dopuna i dodataka. Samo su obvezatne i neobvezatne dopune dijelovi valencijskog uzorka. Uobičajeni i česti dodatci uz neki glagol mogu biti zabilježeni u napomenama ili eventualno u čvrstim svezama. Svi se hrvatski valencijski uzorci mogu opisati s pomoću deset skupina dopuna: *nominativna, genitivna, dativna, akuzativna, instrumentalna, predikatna, priložna, prijedložna, infinitivna i rečenična dopuna*. U rječniku *VALBU* sve se rečenice njemačkoga jezika mogu valencijski opisati s pomoću osam skupina dopuna: *nominativna, akuzativna, genitivna, dativna, prijedložna, priložna, predikatna i glagolska dopuna*. Posljedično svojem padežnom sustavu njemački model nema instrumentalnu dopunu. Iz njihova modela preuzeli smo predikatnu dopunu, ali nismo preuzeli glagolsku dopunu koja obuhvaća rečenične i infinitivne dopune. Rečenične dopune koje pripadaju glagolskoj dopuni, prema autorima rječnika *VALBU*, samo su one rečenice koje se ne mogu shvatiti kao ostvarenje koje padežne dopune.¹⁶ Usporedimo li naš sa Samardžijinim sustavom iz 1986., uočava se podudarnost u broju skupina dopuna (10) te u sadržaju osam skupina dopuna (*nominativna, akuzativna, genitivna, dativna, instrumentalna, prijedložna,*

¹⁵ Usp. Samardžija 1986a; Šojat 2008; Schumacher et al. 2004.

¹⁶ Sve ostale rečenične dopune pripadaju padežnim dopunama.

priložna i *infinitivna dopuna*). U skladu s njemačkom tradicijom Samardžija ne izdvaja rečeničnu dopunu, ali pretpostavlja imeničku i pridjevsku dopunu, koje smo obuhvatili natkategorijom predikatne dopune jer smatramo kako time povećavamo stupanj sintaktičke apstrakcije.

Sadržaji prvih četiriju skupina dopuna razmjerno su prozirni. Budući da je u hrvatskome jeziku subjekt često neizrečen u 1. i 2. licu, postoje valencijski opisi slavenskih jezika koji prvi sudionik radnje, tj. nominativnu dopunu stavljaju u zgrade. Smatramo da neizricanje subjekta nije povezano sa semantičkoleksičkim svojstvima glagola, nego je ono šira sintaktička pojava koja nema veze s pojedinim glagolom, već s cjelokupnim sustavom hrvatskoga jezika. Zato smo odlučili bilježiti nominativnu dopunu i kada subjekt prvoga ili drugoga lica nije izrečen. Tehničko je rješenje da rečeničnom primjeru dodajemo subjektu zamjenicu u zgradama i bilježimo kao nominativnu dopunu (*Bojim se uči.* → *Bojim se (ja) uči.*). Partitivne i negacijske (slavenske) genitive smatramo ostvarenjem akuzativne dopune.¹⁷ Instrumentalna dopuna uglavnom obuhvaća dalje objekte u instrumentalu, a predikatna dopuna odnosi se na imenice ili pridjeve koji se u hrvatskom smatraju dijelovima imenskoga predikata ili proširka predikata.¹⁸ Za razliku od njemačkoga modela, u našem su modelu sve rečenične dopune izdvojene u posebnu skupinu. U tome pogledu model nije homogen jer su rečenične dopune često¹⁹ ostvarenja neke od padežnih dopuna.²⁰

Prema našem nacrtu valencijskoga obrasca svakoj sintaktičkoj dopuni odgovara morfološko ostvarenje. Morfološki opis djelomice je prisutan i u RVHG, ali ga nema u njemačkom rječniku *VALBU*. U naš je opis uvršten radi što bolje razumljivosti rječnika za inozemne učenike hrvatskoga jezika. Morfološki opis obuhvaća sve morfološke kategorije koje se javljaju kao ostvarenja svih sintaktičkih dopuna. Pretpostavili smo četiri osnovne kategorije: 1. *prijedlozi i prijedložne skupine*, 2. *padeži*, 3. *rečenične* i 4. *ostale*. Prijedlozi i prijedložne skupine čine zatvoren popis svih prijedloga i prijedložnih skupina. Padeži obuhvaćaju sve padeže osim vokativa. Budući da vokativ nikad nije ovisan rečenični dio, tj. argument glagola, nije ga potrebno navoditi među morfološkim padežima. Načelno u takvu popisu ne bi trebalo biti mjesta ni za lokativ jer se nikad ne nalazi izvan prijedložnih izraza, no uvrstili smo ga u morfološki popis padeža radi funkcionalne strukture naše baze. Naime, pojedini prijedložni izrazi unose se izborom prijedloga i izborom odgovarajućega padeža koji slijedi iza njega. Morfološka ostvarenja rečenične dopune obuhvaćaju popis

¹⁷ Jedan je od recenzenata upozorio na značenjsku neistovjetnost genitiva i akuzativa na akuzativnome mjestu, koje smo svjesni. Upravo zbog takvih razlika uveden je i morfološki sloj u opisu. Dakle u rečenici *'Dijete je popilo mlijeka.'* imenica *mlijeka* bit će određena u sintaktičkome opisu kao akuzativna dopuna, ali će u morfološkome opisu njezin opis biti genitiv, a ne akuzativ.

¹⁸ Predikatnom dopunom smatramo imenice i pridjeve u nominativu i instrumentalu te *kao*-skupinu ili rjede prijedložne izraze s prijedlogom *za* u sljedećim rečenicama: *'Postao je vojnik.'*, *'Svi su ga smatrali svecem.'*, *'Svi su ga doživljavali kao sveca.'*, *'Našu budućnost doživljavam svijetlom.'*

¹⁹ Prema nekim shvaćanjima svaka je rečenična dopuna ostvarenje padežne dopune. Premda to vrijedi za najveći broj rečeničnih dopuna, nismo sigurni mogu li se padežnim dopunama obuhvatiti sve rečenične dopune. Primjerice, glagol *misliti* najčešće nema akuzativnoga objekta (osim veoma rijetko), ali kao dopunu može imati zamjenicu *to* koja zamjenjuje cijelu rečenicu i rečenicu. Stoga se na rečeničnu dopunu glagola *misliti* može gledati dvojako (kao ostvarenje akuzativne dopune ili kao na posebnu kategoriju). Na temelju sličnih (rijetkih) glagola i zbog što jednostavnijega i lakšega opisa u našem se modelu sve rečenice koje su dopune glagolu određuju kao pripadnici kategorije rečenične dopune.

²⁰ Upozorili su nas da bi to trebalo eksplicitno navesti jer time naš model nije ustrojen na istovrsnim načelima i ne podudara se s njemačkim.

m tradicijom Samardžija ne izdvajamo evsku dopunu, koje smo obuhvatili ime povećavamo stupanj sintaktičke prozirni. Budući da je u hrvatskom valencijski opisi slavenskih jezika avljaju u zagrade. Smatramo da kim svojstvima glagola, nego je m glagolom, već s cjelokupnim iti nominativnu dopunu i kada rješeno da rečeničnom primjeru ao nominativnu dopunu (*Bojim* (slavenske) genitive smatramo na uglavnom obuhvaća dalje i imenice ili pridjeve koji se u oširka predikata.¹⁸ Za razliku dopune izdvojene u posebnu ne dopune često¹⁹ ostvarenja

ntaktičkoj dopuni odgovara n i u RVHG, ali ga nema u lje razumljivosti rječnika za a sve morfološke kategorije ostavili smo četiri osnovne nične i 4. ostale. Prijedlozi jedložnih skupina. Padeži nije ovisan rečenični dio, im padežima. Načelno u nalazi izvan prijedložnih alne strukture naše baze. zborom odgovarajućega pune obuhvaćaju popis na akuzativnome mjestu, koje u rečenici *'Dijete je popilo* ali će u morfološkom opisu lu te kao-skupinu ili rjeđe a smatrali *svećem.*', *'Svi su* emda to vrijedi za najveći nične dopune. Primjerice, u može imati zamjenicu *liti* može gledati dvojako etkih) glagola i zbog što određuju kao pripadnici stovrsnim načelima i ne

vezničkih sredstava (*da, što, kako, li, neka, upitna riječ* i *kao*), oznaku za navođenje (*navod*) i nulu kao oznaku uvođenja zavisnih rečenica bez veznika (*0*). Ostale morfološke kategorije sadržavaju *prilog* i *priložne skupine*, *infinitiv*, *kao-skupinu*, *kvantifikacijsku skupinu* i *pridjev*. Popis morfoloških kategorija napravljen je prema dosadašnjem iskustvu rada s glagolima i općenitim jezikoslovnim spoznajama, no moguće je da će daljnji rad zahtijevati dodavanje novih morfoloških kategorija.

Semantički opis također slijedi njemački uzor koji svaku dopunu opisuje dvorazinski: definicijom pojedinačne semantičke funkcije i pojmovnom kategorijom koja se izabire iz popisa semantičkih kategorija. Opisna definicija pojedinačne semantičke funkcije temelji se na parafraznoj definiciji glagola i mora sadržavati iste značenjske dijelove kao i glagolska definicija. Ako je definicija glagola *brinuti se* 'biti u brizi ili zabrinut', onda je opis semantičke funkcije nominativne dopune 'onaj koji je u brizi'. Popis je pojmovnih kategorija preuzet iz njemačkoga rječnika bez većih promjena jer pojmovne kategorije ne pripadaju jezičnom, već izvanjezičnom znanju. Pri opisu rečeničnoga dijela s pomoću pojmovnih kategorija obrađivač ima izbor između dviju nadređenih kategorija: *bez ograničenja* i *pojmovna kategorija*. Dakle, pretpostavka je da dopune nekih glagola nisu ograničene na određeni broj izvanjezičnih pojmova, odnosno da nemaju semantičkih ograničenja. Tako će nominativna dopuna 'ta haljina' u rečenici „*Ne sviđa mi se ta haljina.*” biti dopuna bez semantičkih ograničenja, dok će dativna dopuna 'mi' biti ograničena na [živo, osoba]. Ne zapisuju se dopune koje realiziraju neočekivanu pojmovnu kategoriju proizišlu iz poetske ili neuobičajene upotrebe glagola. Zapisuju se dopune uz glagole kojima se izriču uobičajena prenesena značenja i ispred pojmovne kategorije stavlja se oznaka [pren.]. Stoga je nominativna dopuna 'cijela dvorana' u rečenici „*Njegovoj šali nasmijala se cijela dvorana.*” označena kao [pren.: mjesto]. Pojmovne kategorije na koje se mogu odnositi priložne dopune odvojili smo od ostalih pojmovnih kategorija i trenutačno ih je sedam (*mjesto, vrijeme, sredstvo, način, uzrok, svrha* te *količina/dimenzija/stupanj/veličina*).²¹

4. Trenutačno stanje i kratkoročni plan

Cilj je omogućiti jednostavno pretraživanje baze prema glagolu kao nositelju leksičkoga unosa, a u kasnijim ćemo etapama domisliti načine složenoga pretraživanja (primjerice prema vrstama dopuna, valencijskim obrascima, semantičkoj skupini ili morfološkim realizacijama). Do sada smo obradili semantičku skupinu psiholoških glagola. Sva su 53 glagola s tom oznakom temeljnoga značenja u punom obliku u bazi, a razvrstano je njihovih 168 značenja na drugoj razini opisa sa svim pripadajućim valencijskim obrascima. Trenutačno istražujemo 48 izdvojenih rečeničnih uzoraka te njihovu distribuciju unutar semantičke skupine. Nakon zaključne redakture omogućit ćemo mrežno pretraživanje i publikaciju koja bi sadržavala valencijski rječnik psiholoških glagola i njihovih rečeničnih uzoraka.

²¹ Predvidjeli smo i moguće nove realizacije priložnih dopuna.

Literatura

- Bošnjak Botica, Tomislava. 2011. „Kategorija glagolske vrste u hrvatskom jeziku”. Doktorska disertacija, Sveučilište u Zagrebu.
- Brač, Ivana i Tomislava Bošnjak Botica. 2015. „Semantička razdioba glagola u *Bazi hrvatskih glagolskih valencija*”. *Fluminensia* 27 (1): 105–120.
- Đorđević, Miloje i Ulrich Engel. 2013. *Srpsko-nemački rečnik valentnosti glagola = Wörterbuch zur Verbvalenz Serbisch-Deutsch*. München: Verlag Otto Sagner.
- Engel, Ulrich. 1988. *Deutsche Grammatik*. Heidelberg: Groos.
- Filipović, Rudolf, ur. 1993. *Teorija valentnosti i rječnik valentnosti hrvatskih glagola*. Zagreb: Zavod za lingvistiku Filozofskog fakulteta Sveučilišta u Zagrebu.
- Helbig, Gerhard i Wolfgang Schenkel. 1973. *Wörterbuch zur Valenz und Distribution deutscher Verben*. Leipzig: VEB Bibliographisches Institut.
- Levin, Beth. 1993. *English verb classes and alternations: a preliminary investigation*. Chicago: The University of Chicago Press.
- Mikelić Preradović, Nives. 2008. „Pristupi izradi strojnog tezaurusa za hrvatski jezik”. Doktorska disertacija, Sveučilište u Zagrebu.
- Moguš, Milan, Maja Bratanić i Marko Tadić. 1999. *Hrvatski čestotni rječnik*. Zagreb: Zavod za lingvistiku Filozofskog fakulteta / Školska knjiga.
- Samardžija, Marko. 1986a. „Valentnost glagola u suvremenom hrvatskom književnom jeziku”. Doktorska disertacija, Sveučilište u Zagrebu.
- Samardžija, Marko. 1986b. „Dopune u suvremenom hrvatskom književnom jeziku”. *Radovi Zavoda za slavensku filologiju* 21: 1–32.
- Schumacher, Helmut et al. 1986. *Verben in Feldern: Valenzwörterbuch zur Syntax und Semantik deutscher Verben*. Berlin: Walter de Gruyter.
- Schumacher, Helmut et al. 2004. *VALBU – Valenzwörterbuch deutscher Verben*. Tübingen: Gunter Narr Verlag.
- Šojat, Krešimir. 2008. „Sintaktički i semantički opis glagolskih valencija u hrvatskom”. Doktorska disertacija, Sveučilište u Zagrebu.
- Šojat, Krešimir. 2009. „Morfosintaktički razredi dopuna u Hrvatskom WordNetu2”. *Suvremena lingvistika* 35 (2): 305–338.
- Šojat, Krešimir. 2012. „Struktura glagolskog dijela Hrvatskog WordNeta”. *Filologija* 59: 153–172.

THE STRUCTURE OF THE VALENCY BASE OF CROATIAN VERBS

SUMMARY

This paper presents the Valency base of Croatian verbs that has been created at the Institute of Croatian Language and Linguistics. The database and the valency data entry form have been designed with the aim of transferring verbs' valency descriptions that had previously been written as unstructured data into a structured machine-readable form. The aim of this work is to create retrieval and learning system of Croatian verbs' valencies as the basis for a comprehensive valency dictionary of Croatian verbs. A three-level data structure of the valency data entry form has been developed using the Tschwanelex software application, and the data entered has been stored in the form of a structured database in SQL.

environment. The first level of an entry contains a lemma form in the infinitive, morphological information, the basic semantic class of the verb described and the verb's collocations. The second level contains a verb definition, a stylistic label if it is present, the reflexive form of the verb (if possible), and other verb semantic classes, if needed. The third level begins with a sentence example linked to a valency pattern, which unites a syntactic, morphological and semantic description. The syntactic description includes ten complement classes. The morphological description refers to a particular example and it is primarily useful in learning Croatian as a foreign language. The semantic description consists of a verb-specific description of participants and semantic categories to which the complement refers, followed by a comment. Such descriptions are repeated until each verb meaning is listed and described. After studying various approaches to verb valency, it was decided that our approach will follow valency descriptions for Croatian verbs and the German valency tradition. Taking this into account, VALBU – Valenzwörterbuch deutscher Verben has been selected as a reference model.

Keywords: *Croatian language, syntax, verb valence, verb valence database, valence dictionary*

eziku". Doktorska
 i u Bazi hrvatskih
 igola = Wörterbuch zur
 glagola. Zagreb: Zavod
 istribution deutscher Verben.
 'gation. Chicago: The
 i jezik". Doktorska
 agreb: Zavod za
 evnom jeziku".
 ku". Radovi Zavoda za
 x und Semantik
 Tübingen: Gunter Narr
 kom". Doktorska
 ". *Suvremena*
 'ija 59: 153–172.

OF

*een created
 ise and the
 'ring verbs'
 data into a
 trieval and
 prehensive
 he valency
 plication,
 ise in SQL*