

## **Nova verzija CROVALLEX-a : pristupi nadogradnji valencijskoga leksikona glagola hrvatskoga jezika**

U ovom poglavlju bit će opisane dvije metode kojima se početna verzija leksikona CROVALLEX tijekom posljednjih nekoliko godina dopunjavala i proširivala. CROVALLEX trenutno sadrži 6129 glagolskih lema grupiranih u sinskupove na čijem se sustavnom opisu i dalje radi, a leksikon u ovoj fazi obuhvaća valencijski opis 3460 glagola u sinskupovima kojima je pridružena oznaka sintaktičko-semantičke klase. U novoj verziji leksikona CROVALLEX preko valencijskih okvira eksplicitno su opisane sintaktičke i semantičke značajke svakoga glagolskog sinskupa koji se temelji na sposobnosti prefigiranoga glagola da se pojavi u valencijskim okvirima koji su značenjski očuvani. Skup okvira koji je pridružen svakom sinskupu odražava značenjske komponente koje ograničavaju broj i vrstu obveznih dopuna i tipičnih izbornih dodataka. Posljedično, svaki glagolski sinskup je opisan tematskim, tj. semantičkim ulogama (dubinskim padežima) i selekcijskim ograničenjima svojih prefigiranih glagola. Svaki sinskup je također definiran valencijskim okvirima svojih glagola, budući da okviri sadrže skup sintaktičkih opisa glagolskih dopuna.

Na temelju češkog VerbaLex leksikona, dopune koje su dio valencijskih okvira u leksikonu CROVALLEX obogaćuju se detaljnijim (primarno semantičkim) opisom, koristeći hrvatske i češke sintaktičko-semantičke klase, koje postoje u oba leksikona i koje su u oba leksikona inspirirane Levininom klasifikacijom.

## 1. Klasteri prefigiranih glagola

Prvotni plan proširenja leksikona obuhvaćao je grupiranje glagola u klaster vidskih parnjaka, tj. stvaranje klastera glagola nastalih procesom perfektivizacije (npr. klaster *pisati* bi obuhvaćao glagole *pisati* i *napisati*). Klastere su trebali sačinjavati glagoli čije značenje je gotovo identično značenju glagola od kojega su nastali i identično njegovomu valencijskom ponašanju, a procesom prefiksacije se mijenja samo vid glagola.

No, analizom 1739 glagola u leksikonu, izdvojena su samo 453 neprefigurirana glagola, kojima se dodao istoznačni perfektivni parnjak. Svi ostali prefigurirani glagoli u leksikonu nastali su procesom perfektivizacije koji je rezultirao promjenom značenja prefiguriranog glagola ili čak sekundarnom imperpektivizacijom, tj. dodavanjem infiksa kojima se vid iz svršenoga promijenio u nesvršeni (npr. *pisati-zapisati-zapisivati*; *pisati-upisati-upisivati*; *pisati-opisati-opisivati*; *pisati-popisati-popisivati*; *pisati-potpisati-potpisivati*).

Daljnijim istraživanjem došlo se do zaključka da su određeni prefiksi povezani s vrlo specifičnim značenjima koja je moguće definirati za svaki prefiks te da upravo ta značenja mijenjaju značenje glagola kojemu se dodaju.

Analizom se dodatno utvrdilo da se neki prefiksi javljaju samo uz određene klase glagola, tj. da neprefigurirani glagoli posjeduju značenjske sličnosti.

Konačno, pristupilo se proširenju leksikona na način da se svima nesvršenim neprefiguriranim glagolima dodaju svi produktivni prefiksi u hrvatskome standardnom jeziku te se u leksikonu zadrže nastali prefigurirani glagoli koji uistinu postoje u standardnom jeziku.

Kao polazišna točka za proširenja leksikona postavljena su 453 neprefigurirana glagola i njihovi vidski parnjaci, a za ostale prefigurirane glagole u leksikonu ručno su definirani neprefigurirani glagoli od kojih su nastali te vidski parnjaci tih polaznih glagola (budući da prefigurirani glagoli u leksikonu većinom nisu bili vidski parnjaci neprefiguriranih glagola i značenjem su se razlikovali od neprefiguriranog glagola od kojih su nastali). Dodavanje vidskih parnjaka postojećima neprefiguriranim glagolima te dodavanje novih neprefiguriranih glagola i njihovih vidskih parnjaka (koji nisu od ranije bili u CROVALLEX-u), rezultiralo je leksikomom od 1453 polazna glagola (neprefiguriranih glagola i čistih perfektivnih parnjaka s identičnim značenjem i valencijskim ponašanjem). Nadalje, 1453 polazna glagola kombinirana su sa svakim prefiksom iz skupa od

20 prefiksa koji su produktivni u hrvatskome jeziku (do-, iz-, na-, nad-, o-/ob-, obez-, od-, po-, pod-, pre-, pred-, pri-, pro-, raz-, s-, su-, u-, uz-, za-).

Budući da određeni prefiksi završavaju na suglasnik i posjeduju različite alomorfe, ovisno o početnom slovu polaznoga glagola, kao i alternirajuće oblike sa samoglasnikom -a (raza-, iza-, uza, nada-, oba-, oda -, poda-), definiran je skup od 50 pravila za kombiniranje predmetka i polaznog glagola, uzimajući u obzir karakteristične morfonološke prilagodbe (npr. raz+plakati = rasplakati, iz+čupati = iščupati, itd.).

Konačan popis sastojao se od 28 947 glagola, čija je ovjerenost provjeravana u nekoliko hrvatskih jednojezičnih i dvojezičnih rječnika (Parčić, 1901; Šonje, 2000; Anić, 1998; Šarić & Wittschen, 2003 te Filipović, 1996).

Parčić (1901) je bio jedini rječnik dostupan u elektroničkome obliku, dok su pretraživanje ostala četiri rječnika obavljalo ručno.

Što se tiče potvde glagolskih leksema: 4600 glagola pronađeno je u Parčićevom rječniku (1901), 3591 glagol je pronađen u (Šonje, 2000), 4262 glagola u (Anić, 1998), 3327 glagola u (Šarić & Wittschen, 2003), a 2541 glagol u (Filipović, 1996).

Kada su se preklopili, 5 rječnika dalo je 6129 potvrđenih glagola ili 21,2% od svih glagola. Nadalje, 73 glagola pronađena su samo u (Šonje, 2000), 360 glagola samo u (Anić, 1998), 56 glagola samo u (Šarić & Wittschen, 2003), a 38 glagola samo u (Filipović, 1996).

U sljedećem koraku detaljno su analizirana sva značenja 20 produktivnih prefikasa, uspoređena svojstva prefiguriranih i polaznih glagola i izgrađeni modeli za određene skupine glagola koji dijele isto značenje prefiksa i isto valencijsko ponašanje (Tablica 1).

Prefiks s najvećim skupom značenja je **za-**, s 20 različita značenja, dok je prosječan broj značenja po prefiksu 11.

Dodavanjem prefiksa polaznim glagolima uspostavlja se veza među različitim glagolima koji počinju dijeliti ista značenjska obilježja upravo zbog pridodanoga prefiksa. Ova veza se posebno očituje u valencijskim okvirima tih glagola jer prefigurirani glagoli počinju dijeliti iste valencijske okvire, bez obzira na valencijsko ponašanje glagola od kojih su nastali.

Na primjer, neki glagoli zbog prefiksa nasljeđuju značenje "premjestiti nešto ili nekog iznutra van ili odozdo prema gore", neki dobivaju značenje "pridruživanja, sastavljanja", itd.

Nadalje, prefigurani glagoli mogu se smatrati hiponimima polaznih glagola, ako se njihovo značenje ne razlikuje značajno od značenja polaznih glagola. Međutim, značenje prefiguiranih glagola nije jednostavna kombinacija značenja polaznoga glagola i značenja prefiksa, jer su i polazni glagoli i prefigurani glagoli obično višeznačni. Prema tome, dodavanje prefiksa može se promatrati samo kao veza između polaznoga i prefiguiranoga glagola u jednom od njihovih značenja, a ne između glagolskih leksema.

Višeznačnost polaznoga glagola može uzrokovati višeznačnost prefiguiranoga glagola, ali prefigurani glagol može postati višeznačan i zbog višeznačnosti samoga prefiksa koji mu je dodijeljen.

Prefiks iz-		
Značenje	Valencijski okvir	Primjer glagola
1. Ekstralokativnost I: pomicanje vršitelja iz unutrašnjosti čega prema van ili odozdo prema gore	AGT <sub>-1</sub> :obl [DIR1 <sub>-iz+2</sub> ] [DIR3 <sub>-na+4</sub> ]	izići, istrčati, iskočiti, izletjeti, izjuriti
2. Ekstralokativnost II: pomicanje trpitelja iz unutrašnjosti čega prema van ili odozdo prema gore	AGT <sub>-1</sub> :obl PAT <sub>-4</sub> :obl [DIR1 <sub>-iz+2</sub> ] [DIR3 <sub>-na+4</sub> ]	izvući, izvaditi, izbaciti, izgurati, izbiti, iščakati
3. Partitivnost: odvajanje trpitelja od čega ili dijela od cjeline	AGT <sub>-1</sub> :obl PAT <sub>-4</sub> :obl [ORIG <sub>-iz+2</sub> ]	isključiti, izuzeti, ispisati (iz čega), ispustiti
4. Rezultativnost koja ukazuje na to da se vršenjem radnje trpitelj potrošio, nema ga više	AGT <sub>-1</sub> :obl PAT <sub>-4</sub> :obl [RESL <sub>-do+2</sub> ]	ispiti, izjesti, izgaziti, iznositi, izlizati
5. Rezultativnost i sativnost: povratni glagoli označuju da je vršitelj ono što označuje izvršio do kraja, da više nema mogućnosti vršiti to dalje	AGT <sub>se</sub> :obl [RESL <sub>-do+2</sub> ]	istrčati se, isplesati se, isplivati se, izgrmjeti se, iskakati se, izdivljati se, ispavati se
6. Rezultativnost kao konačna točka tijeka, dovršenje radnje nad trpiteljem bez mogućnosti za daljnju akciju	AGT <sub>-1</sub> :obl PAT <sub>-4</sub> :obl	iscijeliti, izbrojiti, izigrati, izliječiti, iznenaditi
7. Glagoli označuju da se radnjom djelotvori stvaranje	AGT <sub>-1</sub> :obl PAT <sub>-4</sub> :obl [ORIG <sub>-iz/od+2</sub> ]	isklesati, istesati, istkati, izmisliti, izliti (lijevanjem odliti)

Prefiks iz-		
Značenje	Valencijski okvir	Primjer glagola
	[RESL <sub>_na+4</sub> ]	
8. Distributivnost I: radnja se izvršila na većem broju trpitelja redom	AGT <sub>_1</sub> :obl PAT <sub>_4</sub> :obl [RESL <sub>_do+2</sub> ]	izujedati, izubijati, izlijepiti, izljubiti, ispremiješati, ispresijecati
9. Distributivnost II: radnju je izvršio veći broj vršitelja redom ili obuhvaća cijelo mnoštvo	AGT <sub>_1</sub> :obl	izumirati, ispogibati, izginuti
10. Glagoli označuju da se radnjom nešto steklo, zadobilo	AGT <sub>_1</sub> :obl PAT <sub>_4</sub> :obl [ORIG <sub>_iz/od+2</sub> ]	iznuditi, izmamiti, izmoliti, isprostiti, izmoljakati, izvabiti
11. Glagoli kretanja označuju da se radnjom osvojilo ono na čemu se kretanje zbivalo	AGT <sub>_1</sub> :obl PAT <sub>_4</sub> :obl	ispenjati (planinski vrh) istrčati (stazu) isplivati (bazen)
12. Glagoli označuju promjenu stanja, najčešće nepovoljnu po vršitelja	AGT <sub>_1</sub> :obl [RESL <sub>_do+2</sub> ]	iščeznuti, izginuti, izgorjeti
13. Glagoli označuju promjenu stanja, najčešće nepovoljnu po trpitelja	AGT <sub>_1</sub> :obl PAT <sub>_4</sub> :obl	ispržiti, istopiti, iscijediti
14. Intenzitet I: vršitelj je radnju izvršio obilno, do krajnje mjere, intenzivno, najčešće s negativnim prizvukom	AGT <sub>_se</sub> :obl [RESL <sub>_do+2</sub> ]	izopijati se, izležavati se, iščuđavati se
15. Intenzitet II: vršitelj je nad trpiteljem radnju izvršio do krajnje mjere, intenzivno, najčešće s negativnim prizvukom	AGT <sub>_1</sub> :obl PAT <sub>_4</sub> :obl	izmoriti, izmučiti, izudarati

**Tablica 1. Prefiks iz- sa značenjima i primjerima prefigiranih glagola**

Značenje prefigiranog glagola može se u potpunosti razlikovati od značenja polaznog glagola zbog pridodanog prefiksa (primjeri 1-6).

Primjeri:

(1) dati – iz+dati / otkriti ili iz+dati / predati na raspolaganje

- (2) baciti - po+baciti / prekinuti trudnoću; pred+baciti / prigovoriti; pod+baciti / iznevjeriti
- (3) sjesti- na+sjesti / zasjesti ili na+sjesti/ biti žrtva prijevare; pri+sjesti / zapeti u grlu ili pri+sjesti / osvojiti mjesto uz koga
- (4) paliti - is+paliti / izbaciti zrno ili sačmu iz vatrenog oružja; s+paliti / sagorjeti; za+paliti / izazvati plamen
- (5) gubiti - po+gubiti / izgubiti sve redom ili po+gubiti / lišiti života; izgubiti / ostati bez čega
- (6) pasti - na+pasti / pasti u većoj količini ili na+pasti / izvršiti napad

Kao dio istraživanja, ručno smo analizirali sintaktičke argumente prefigiranih glagola i njihov odnos s polaznim glagolima, koristeći popis od 20 produktivnih prefikasa te popis od 6129 potvrđenih glagola, kao i odgovarajuće polazne glagole u CROVALLEX-u.

Naši rezultati pokazuju da učinak prefikasa na strukturu argumenata glagola u hrvatskome jeziku pokazuje pravilnost koja se može iskoristiti za značajno smanjivanje ručnoga leksikografskog rada potrebnoga za stvaranje opsežnoga valencijskog leksikona.

Glagoli koji dijele isti prefiks (točnije, isto značenje određenoga prefiksa) također dijele iste valencijske okvire te broj obveznih dopuna i tipičnih izbornih dodataka, kao što su DIR1 (smjer od) ili DIR3 (smjer prema) sa svojim površinskih oblicima (objašnjeno u primjerima 7-13 na prefiksu **od-** u značenju "smjer kretanja od").

Primjeri 12 i 13 pokazuju prefiks **do-** u značenju "nastavljanje radnje preko njenog logičnog kraja", pridružen različitim glagolima s istim rezultatom (usporedimo do+dati / reći još nešto; do+pisati / dodati onome što je napisano; do+liti / dodati tekućine ondje gdje je već uliveno).

Primjeri:

- (7) Jedrilica se otisnula iz uvale na pučinu AGT<sub>1</sub>:obl [DIR1<sub>iz+2</sub>] [DIR3<sub>na+4</sub>]
- (8) Voda je otekla s krova u oluk AGT<sub>1</sub>:obl [DIR1<sub>s+2</sub>] [DIR3<sub>u+4</sub>]
- (9) Zmija je otpuzala iz sobe na livadu AGT<sub>1</sub>:obl [DIR1<sub>iz+2</sub>] [DIR3<sub>na+4</sub>]
- (10) Mačka je odskočila s kreveta na pod AGT<sub>1</sub>:obl [DIR1<sub>s+2</sub>] [DIR3<sub>na+4</sub>]
- (11) Damir je odšetao iz kazališta u mračnu ulicu AGT<sub>1</sub>:obl [DIR1<sub>iz+2</sub>] [DIR3<sub>u+4</sub>]

(12) Maja je dočrtala još cvjetova na svoj pergament. AGT<sub>1</sub>:obl PAT<sub>4</sub>:obl [DIR3<sub>na+4</sub>]

(13) Ivana je dolila mlijeko u šalicu AGT<sub>1</sub>:obl PAT<sub>4</sub>:obl [DIR3<sub>u+4</sub>]

Posljedično, moguće je sustavnije dodati tipične izborne dodatke valencijskomu okviru pojedinoga glagola.

Na primjerima možemo primijetiti i semantički sklad između prefikasa, prijedloga i tipičnih izbornih dodataka. Značenje prefiksa se pojačava značenjem prijedloga. Na primjer, prefiks *pre-* i prijedlog *preko* dijele značenje "kretanja preko nečega ili nekoga". Nadalje, prijedlog tipa *preko* uvijek je dio izborne dopune (tipa DIR2<sub>-preko+2</sub>) koje glagoli s prefiksom *pre-* obično vežu uz sebe.

Taj sklad je najočitiiji u glagolima kretanja, gdje neki prefiksi zadržavaju različita značenja smjera ili mjesta. Stalna uporaba istoga prijedloga (npr. *u-*, *na-*, *uz-*) s istim prefiksom naglašava specifično značenje toga prefiksa (primjeri 14-19).

Primjeri:

(14) **u + akuzativ**, u valencijskom okviru glagola s prefiksom **u-** (*uroniti, uskočiti, uputiti*)

(15) **na + akuzativ** u valencijskom okviru glagola s prefiksom **na-** (*naskočiti, navaliti, naslagati*)

(16) **za + akuzativ** u valencijskom okviru glagola s prefiksom **za-** (*zamaknuti, zbaciti, zaći*)

(17) **preko + genitiv** u valencijskom okviru glagola s prefiksom **pre-** (*prebaciti, prenijeti, premazati*)

(18) **kroz + akuzativ** u valencijskom okviru glagola s prefiksom **pro-** (*provući, progurati, prokopati*)

(19) **uz + akuzativ** u valencijskom okviru glagola s prefiksom **uz-** (*uzdići, uspraviti, uzviti*).

Možemo zaključiti da su valencijska svojstva glagola usko povezana s njegovom semantikom. Dodavanjem prefiksa mijenja se aspekt glagola, ali i njegovo značenje te njegova valencijska svojstva.

U početnome dijelu ovoga rada opisano je **180** sintaktičko-semantičkih klasa u CROVALLEX-u, tj. izrađena je klasifikacija utemeljena na hipotezi po kojoj bi obrada cijelih skupina glagola sa sličnim semantičkim obilježjima mogla

poslužiti kao pokazatelj njihove sličnosti na sintaktičkoj razini, što bi doprinijelo usustavljenosti leksikona.

Drugim riječima, motivacija za uvođenje ovakve klasifikacije bila je interpretirati odnos između sintakse i semantike hrvatskih glagola i omogućiti generalizaciju nad nekim jezičnim svojstvima s ciljem smanjenja redundantnosti u leksikonu.

Drugi motiv je bio dokaz da je na glagole kretanja u hrvatskome jeziku moguće sustavno primijeniti metodologiju za analizu glagola kretanja u engleskome jeziku (Žic-Fuchs, 1989).

Levinina je klasifikacija poslužila kao dobro polazište za sintaktičko-semantičku klasifikaciju hrvatskih glagola, jer je otkrila da različiti polazni glagoli u CROVALLEX-u koji pripadaju istoj semantičkoj klasi, također dijele identične valencijske okvire (kao i broj i vrstu obveznih i izbornih argumenata te površinske morfosintaktičke oblike). Levinina klasifikacija također nam je pomogla otkriti da promjena u valencijskome okviru glagola ukazuje na moguće promjene u semantici glagola. Problem s ovom klasifikacijom je što ne rezultira dosljednim klasama, što je vidljivo i iz problema na semantičkoj razini u prvoj verziji CROVALLEX-a. Veliki broj glagola u hrvatskome jeziku je višeznačan, i iako većina glagola u postojećim korpusima hrvatskoga jezika posjeduje jedno dominirajuće značenje, postoji značajan broj visokofrekventnih glagola koji se ne mogu adekvatno opisati samo jednim značenjem.

Tablica 2 pokazuje da prefigurirani glagoli, u odnosu na jednovalentne polazne glagole od kojih su nastali, pripadaju različitim sintaktičko-semantičkim klasama, uzimaju specifične tipične izborne dodatke (smjer) ili imaju različit broj obveznih dopuna u svome valencijskom okviru čime postaju dvovalentni glagoli (primjerice, glagoli s prefiksom ob- ili na-).

Promjena prefiksa (primjeri 20-22) ukazuje na promjenu i obvezne glagolske dopune i glagolskoga značenja (usporedimo: *misлити* / *s+mišljati* / *za+mišljati*).

Primjeri:

(20) Marija je **mislila** AGT<sub>1</sub>:obl

(21) Marija je **smišljala** priču AGT<sub>1</sub>:obl PAT<sub>4</sub>:obl

(22) Marija je **zamišljala** more AGT<sub>1</sub>:obl PAT<sub>4</sub>:obl



Dodatno, primjeri u tablici 2 pokazuju da većina prefigiranih glagola koji dijele isti prefiks pripadaju jednoj (istoj) klasi. Stoga nam je cilj bio interpretirati semantiku glagolskih prefikasa u hrvatskome jeziku i prikazati ove specifične značajke prefigiranih glagola u novoj verziji CROVALLEX-a.

Lema	Valencijski okvir	Klasa
Polazni glagoli: juriti, trčati, letjeti, ploviti	AGT_1:obl	glagoli načina kretanja
Prefigirani glagoli: pojuriti, potrčati, poletjeti	AGT_1:obl [DIR3_prema+3]	glagoli žurbe
Prefigirani glagoli: odjuriti, otrčati, odletjeti	AGT_1:obl [DIR1_od+2] [DIR3_prema+3]	apkurzivni glagoli
Prefigirani glagoli: izjuriti, istrčati, isploviti	AGT_1:obl [DIR1_iz+2] [DIR3_prema+3]	ekskurzivni glagoli
Prefigirani glagoli: utrčati, uploviti, uletjeti	AGT_1:obl [DIR3_u+4]	inkurzivni glagoli
Prefigirani glagoli: oprčati, obletjeti, oploviti	AGT_1:obl PAT_4:obl	glagoli kružnoga kretanja
Prefigirani glagoli: najuriti	AGT_1:obl PAT_4:obl	glagoli tjeranja

**Tablica 2.** Promjene u valencijski i značenju uzrokovane prefiksima dodanima glagolima kretanja

Nova verzija leksikona CROVALLEX temelji se na pretpostavci da glagoli koji dijele isti prefiks (točnije, isto značenje određenoga prefiksa) također imaju iste ili slične dopune s istim ili sličnim morfosintaktičkim oblicima u svojim valencijskim okvirima. Nadalje, neke dopune sa specifičnom sintaktičkom realizacijom također su poslužile kao kriterij za uspostavljanje dosljednih sintaktičko-semantičkih klasa, gdje nekoliko prefikasa (kao što su *iz-*, *s-* i *od-*) dijeli isto značenje.

Primjerice, moguće je dokazati pojavljivanje tipičnoga dodatka DIR1 (smjer od) s određenim morfosintaktičkim oblicima (**iz** + genitiv, **s** + genitiv, **od** + genitiv) u valencijskim okvirima glagola koji pripadaju sljedećima sintaktičko-semantičkim klasama: *glagoli nestajanja*, *glagoli uklanjanja*, *glagoli tjeranja*, *glagoli primjene sile*, *glagoli stavljanja tekućih entiteta u spremišta odnosno na površine*, *glagoli bacanja*, *glagoli stjecanja i dobivanja*, *glagoli prijenosa poruke*, *glagoli oslobođenja*, *apkurzivni glagoli*.

Također, prefiksi *iz-*, *s-* i *od-* dijele sljedeća značenja: odvajanje, udaljavanje, dovršavanje radnje u kojoj je trpitelj potrošen, dovršenje radnje bez mogućnosti za daljnju akciju, stvaranje, stjecanje ili dobivanje te promjenu stanja.

Ove kombinacije prefikasa i polaznih glagola pružaju uvid u blisku vezu sintakse i semantike glagola koji dijele isti prefiks, kao i mogućnost generalizacije različitih lingvističkih značajki, što bi za posljedicu moglo imati smanjivanje redundantnosti u leksikonu CROVALLEX.

U novoj verziji leksikona CROVALLEX preko valencijskih okvira eksplicitno su opisane sintaktičke i semantičke značajke svakoga glagolskog sinskupa. Sinskup se temelji na sposobnosti prefigiranoga glagola da se pojavi u parovima valencijskih okvira koji su na neki način značenjski očuvani. Skup sintaktičkih okvira koji je pridružen svakom sinskupu odražava značenjske komponente koje ograničavaju broj i vrstu obveznih i tipičnih izbornih dodataka. Posljedično, svaki glagolski sinskup je opisan tematskim ulogama (dubinskim padežima) i selekcijskim ograničenjima svojih prefigiranih glagola. Svaki sinskup je također definiran valencijskim okvirima svojih glagola, budući da okviri sadrže skup sintaktičkih opisa glagolskih dopuna.

Budući da takav glagolski sinskup sadrži sve relevantne karakteristike pojedinoga glagola, omogućava generalizaciju sintaktičkih i/ili semantičkih obilježja glagola u sinskupu. Ako ih grupiramo po prefiksu, glagoli stvaraju semantički i sintaktički koherentne sinskupove, čiji članovi dijele raspon obilježja, počevši od implementacije i tumačenja određenih dopuna do postojanja morfološki povezanih oblika.

Uz svaki sinskup nalazi se popis pojedinačnih valencijskih okvira, a za svaki valencijski okvir specificirani su elementi sinskupa na koje se taj okvir odnosi.

Na ovaj način opisana je cijela grupa glagola koji su međusobno povezani sintaktičko-semantičkim obilježjima, što dodatno pridonosi usustavljenosti i preglednosti leksikona.

Kao rezultat, moguće je smanjiti napor potreban za proširenje leksikona i umanjiti vjerojatnost uvođenja pogreške prilikom dodavanja novih glagola i njihovih valencijskih opisa u postojeći leksikon.

Analizirajući sve produktivne prefikse glagola u hrvatskome jeziku, dokazali smo da prefiks podrazumijeva ne samo promjenu u glagolskome aspektu, nego (puno češće) i promjenu u leksičkome značenju glagola. Ukoliko prefiks podrazumijeva isključivo promjenu u glagolskome vidu (slučaj vidskih parnjaka), novonastali prefigirani glagol nasljeđuje i značenje i valencijski opis s identičnim valencijskim okvirima polaznoga glagola. Ukoliko kod novonastalog prefigiranog glagola dolazi do promjene u leksičkome značenju, ta promjena sa sobom donosi i promjenu u valencijskome opisu (i broju argumenata i morfosintaktičkoj realizaciji), kao i promjenu sintaktičko-semantičke klase.

Pri određivanju tipologije značenjskih obilježja svakoga produktivnog prefiksa u hrvatskome jeziku otkriveno je da obilježja koja proces prefigiranja implicira obično znače temporalnost (inkoativnost, punktualnost, iterativnost), lokativnost, smjer (adlativnost, ablativnost, perlativnost), dimenziju (intenzitet), promjenu stanja (rezultativnost, mutativnost), transformativnost, distributivnost, komitativnost, itd.

Nažalost, ovakvom klasifikacijom ne možemo u potpunosti opisati glagolsku semantiku. Stoga je plan za budući rad na leksikonu uvesti značenjske uloge koje će biti distinktivnije od postojećih dubinskih padeža (funktora) u leksikonu CROVALLEX.

Trenutni cilj je poboljšati sintaktičko-semantičku klasifikaciju glagola u leksikonu i smanjiti redundantnost u CROVALLEX-u. Konačni cilj cijeloga istraživanja nad sintaktičko-semantičkim klasama i gore opisanim sinskupovima je omogućiti automatsko predviđanje valencijskih okvira, kao i sintaktičko-semantičkih svojstava prefigiranih glagola koji još uvijek nisu opisani i pokriveni leksikonom CROVALLEX.

## 2. Dvojezični češko-hrvatski valencijski leksikon glagola

Iako postoje studije koje opisuju proširenje postojećih valencijskih leksikona ekstrahiranjem informacija iz običnih dvojezičnih rječnika (Fujita & Bond, 2007), vrlo malo dosadašnjih istraživanja proučava transjezični potencijal Levinine klasifikacije glagola za izgradnju valencijskoga leksikona. Nadzirani transjezični pristup klasifikaciji iz engleskoga u talijanski jezik sa 59 glagola i 3 klase (Merlo i sur., 2002) ostvario je točnost od 86,3%. Drugi znanstvenici (Sun i sur., 2008; Sun i sur., 2010) otkrili su sličnosti klasifikacije u engleskome i francuskome jeziku, što podržava ranije rezultate (Merlo i sur., 2002). Također, preuzimanjem glagolskih okvira iz češkoga WordNet-a u rumunjski WordNet (Tufis i sur., 2006) i bugarski WordNet (Pala & Horák, 2008) ostvaren je visok postotak podudarnosti - 25% bugarskih valencijskih okvira je bilo identično okvirima u češkome jeziku, dok je 500 od 600 rumunjskih valencijskih okvira preuzeto bez ili uz male modifikacije.

Svi gore navedeni rezultati upućuju na činjenicu da Levinina klasifikacija ima transjezičnu osnovu te da su opća metodologija i određena obilježja prenosivi iz jezika u jezik.

U svrhu daljnjeg proširivanja i usavršavanja leksikona CROVALLEX, a i radi istraživanja mogućnosti prekojezičnoga razvoja valencijskih leksikona glagola za srodne jezike, pristupilo se izgradnji češko-hrvatskoga dvojezičnoga valencijskoga leksikona.

Na temelju češkoga VerbaLex leksikona, dopune koje su dio valencijskih okvira u leksikonu CROVALLEX obogaćuju se detaljnijim (primarno semantičkim) opisom, koristeći hrvatske i češke sintaktičko-semantičke klase, koje postoje u oba leksikona i koje su, u oba leksikona, inspirirane Levininom klasifikacijom.

Češki i hrvatski jezik, budući da pripadaju istoj obitelji slavenskih jezika, posjeduju morfološke, sintaktičke i semantičke sličnosti. Glavna razlika u valencijskim opisima glagola u tim jezicima leži u površinskoj (morfosintaktičkoj) realizaciji obveznih glagolskih dopuna (njihov broj i funkcija su vrlo slični ili čak isti).

Što se tiče sintaktičko-semantičke klasifikacije glagola u CROVALLEX-u, već smo naveli da sadrži **180** sintaktičko-semantičkih klasa (točnije, 72 klase s još dvije razine potpodjele) te 6129 glagolskih lema grupiranih u sinkupove na čijem se sustavnom opisu i dalje radi. Sintaktičko-semantičke klase u leksikonu

izvorno su preuzete iz VerbNet leksikona, koji se temelji na Levininoj klasifikaciji glagola, a klase su prilagođene hrvatskomu jeziku na način da je prednost dana semantičkom kriteriju nad sintaktičkim alternacijama.

Glede valencijskoga leksikona čeških glagola, VerbaLex leksikon razvijen je na Sveučilištu u Brnu (Hlaváčková & Horák, 2005) i obuhvaća oko 10 500 čeških glagola kojima su dodijeljene sintaktičko-semantičke klase inspirirane Levininom klasifikacijom te tu leži sličnost s CROVALLEX-om.

VerbaLex valencijski okviri su obogaćeni zamjenicama (npr. kdo - tko, co - što), uz brojeve morfoloških padeža koji postoje i u leksikonu CROVALLEX. Ovakva vrsta bilježenja dopuna pokazala se iznimno važnom za razlikovanje živoga od neživoga entiteta i bilo bi korisno u budućnosti uvesti taj dodatak u leksikon CROVALLEX.

Primjer valencijskih okvira glagola u CROVALLEX i VerbaLex leksikonu prikazan je u tablici 3.

<b>Češki okvir</b>	<b>frame:AGENT</b> <person:1> <sup>obl</sup> <sub>kdo1</sub> <b>PATIENT</b> <person:1> <sup>obl</sup> <sub>koho4</sub>
<b>Češki primjer</b>	<i>adoptovali chlapečka</i>
<b>Hrvatski okvir</b>	<b>frame:AGENT</b> <sup>obl</sup> <sub>1</sub> <b>PATIENT</b> <sup>obl</sup> <sub>4</sub>
<b>Hrvatski primjer</b>	<i>posvojiše dječaka</i>

**Tablica 3.** Usporedba valencijskoga okvira glagola u CROVALLEX i VerbaLex leksikonu

CROVALLEX i VerbaLex posjeduju gotovo jednak broj dubinskih padeža prve razine, čak im je i inventar funkтора vrlo sličan.

No, VerbaLex dodatno razlikuje oko 1000 dubinskih padeža druge razine, od kojih je svaki povezan s Princeton WordNet-om i češkim WordNet-om. Dubinski padeži u VerbaLex-u se temelje na entitetima prvoga i drugoga reda EuroWordNet Top ontologije (Vossen i sur, 1998) i omogućavaju visok stupanj razlikovanja značenja. Primjerice, za najčešći dubinski padež - AGENT, VerbaLex zbog dodatne razine semantike razlikuje radi li se o osobi, životinji, skupini ljudi, instituciji ili stroju.

Dakle, bilježenje dubinskih padeža prve razine u dvama leksikonima te odsutnost dubinskih padeža druge razine u leksikonu CROVALLEX predstavlja glavne razlike između CROVALLEX i VerbaLex valencijskih okvira. Osnovna

usporedba dubinskih padeža prve razine dovodi do zaključka da najfrekventnije značenje glagola dijeli iste ili vrlo slične dubinske padeže (semantičke uloge) u oba leksikona.

Usporedbom češko-hrvatskih parova valencijskih okvira glagola za sve Levinine klase, analiziran je semantički dio čeških valencijskih okvira i prenesena sintaktička obilježja na hrvatske okvire uz manje morfosintaktičke izmjene.

Polazna hipoteza za ovakav tip proširivanja leksikona CROVALLEX je da ovakav pristup može riješiti probleme koji proizlaze iz višeznačna prirode glagola, a za koje u prvotnoj verziji CROVALLEX-a nije bilo rješenja.

Promjena značenja glagola u hrvatskome jeziku ne utječe uvijek na njegovu valenciju, ali CROVALLEX 2.008 ne posjeduje mogućnost bilježenja tih razlika (primjeri 1-4). Sve rečenične konstrukcije (1-4) su sintaktički i morfološki moguće, ali rečenice (2) i (3) su semantički moguće samo u poetskom smislu, budući da vršitelj radnje nije živi entitet (tj. more kao prostor slane vode ne može osjećati). Ovaj tip problema može se riješiti uvođenjem druge razine semantike koja dodatno razlikuje podvrste semantičkih uloga (primjeri 5 i 6).

- (1) Dolores nedostaje more [AGT<sub>3</sub>:obl PAT<sub>1</sub>:obl]
- (2) Dolores nedostaje moru [PAT<sub>1</sub>:obl AGT<sub>3</sub>:obl]
- (3) Moru nedostaje Dolores [AGT<sub>3</sub>:obl PAT<sub>1</sub>:obl]
- (4) More nedostaje Dolores [PAT<sub>1</sub>:obl AGT<sub>3</sub>:obl]
- (5) Dolores nedostaje more [AGT<sub>3</sub>:obl(**person:1**) ENT<sub>1</sub>:obl(**body of water:1**)]
- (6) More nedostaje Dolores [ENT<sub>1</sub>:obl(**body of water:1**) AGT<sub>3</sub>:obl(**person:1**)]

Stoga je odlučeno u CROVALLEX uvesti podtipove semantičkih uloga na temelju VerbaLex leksikona kako bi se dobila detaljnija i točnija semantička klasifikacija, ali i izgradio dvojezični valencijski leksikon.

Početno proširivanje leksikona provedeno je kroz sedam koraka:

- (1) 19 Levininih klasa je uzeto kao zlatni standard za usporedbu glagolske semantike u češkome i hrvatskome jeziku (tablica 4)
- (2) 7 klasa (od ukupno 19) koje sadrže glagola oba jezika izabrano je za transjezičnu usporedbu
- (3) ručno su dizajnirana morfosintaktička pravila transjezičnog prijenosa

- (4) za češke glagole u svakoj od tih 7 klasa su pronađeni njihovi hrvatski ekvivalenti ili su prevedeni na hrvatski jezik (ako nisu bili pronađeni na postojećem popisu)
- (5) valencijski okviri čeških glagolskih skupova su primijenjeni na hrvatske glagole, uzimajući u obzir morfosintaktičke razlike te primjenom morfosintaktičkih pravila
- (6) cijeli niz obilježja valencijskih okvira 312 hrvatskih glagola koji su rezultat koraka (5) ručno je analiziran
- (7) budući da su 2 klase (*18.1 hit-glagoli gađanja* i *37.3 manner of expression – glagoli načina izražavanja*) sadržavale više hrvatskih glagola nego čeških, valencijski okviri prevedenih čeških glagolskih skupova su primijenjeni na sve hrvatske glagole u klasi kako bi se ispitala mogućnost automatskoga stvaranja obogaćene verzije hrvatskoga valencijskog leksikona.

Što se tiče implementacije VerbNet klasifikacije u drugim jezicima, upravo je 17 klasa iz Levinine klasifikacije korišteno kao engleski zlatni standard (Swift, 2005) u automatskoj izradi klastera engleskih glagola. Iste klase su također korištene za istraživanje prekojezične primjenjivosti klasa za francuski jezik (Sun i sur., 2008). Stoga je i kod proširenja hrvatskoga leksikona korišten isti zlatni standard za usporedbu semantike glagola češkoga i hrvatskoga jezika, budući da je Levinina klasifikacija već implementirana u oba valencijska leksikona (tablica 4).

Od 17 klasa engleskoga i francuskoga zlatnoga standarda, samo 5 klasa sadržavalo je glagole i češkoga i hrvatskoga jezika, budući čak za 12 Levininih klasa u VerbaLex-u ne postoje podatci (tj. pridruženi glagoli). Odlučeno je, ipak, da se zadrži ovaj zlatni standard za buduće usporedbe i analize valencijskih okvira u hrvatskome, francuskome i engleskome jeziku. Još dvije klase dodane su zlatnomu standardu (*33. judgement - glagoli osude* i *17.1 throw – glagoli bacanja*), kako bi se dobio dovoljan broj čeških i hrvatskih glagola za usporedbu. Pokrivenost glagola u hrvatskim klasama bolja je nego u češkome, ali hrvatski leksikon u ovoj fazi obuhvaća 3460 klasificiranih glagola koji su opisani valencijskim okvirima, a VerbaLex 10 500 glagola, koji nisu svi klasificirani, ali svi posjeduju valencijski opis. Budući da značajnom broju čeških glagola još nisu dodijeljene sintaktičko-semantičke klase, bilo bi zanimljivo ispitati mogu li se morfosintaktičkim pravilima izrađenima u svhu proširenja hrvatskoga leksikona koristiti reverzibilno za sistematsku klasifikaciju čeških glagola koji još

nisu pridruženi odgovarajućim klasama. Iako jedini postojeći češko-hrvatski dvojezični rječnik (Merhaut, 1998) nudi samo mogućnost prevođenja pojedinačnih glagola, a ne cijelih sinskupova, prevođenje glagola i preslikavanje valencijskih okvira iz izvornoga (češkoga) u ciljni (hrvatski) jezik nije bio nemoguć zadatak. Svaki češki glagol u rječniku u prosjeku je preveden sa 3 hrvatska glagola. Ako značenje bilo kojega glagola u češkome sinskupu (izraženo kroz definiciju sinskupa) nije odgovaralo značenju hrvatskoga glagolskog sinskupa, glagol je označen kao nejasan i izostavljen iz klase. S druge strane, višeznačni glagoli koji pripadaju većem broju čeških i hrvatskih sinskupova, također pripadaju i većem broju (različitih) semantičkih klasa, što je bitno olakšalo cijeli proces prevođenja i preslikavanja valencijskih okvira.

Levinine klase	Broj čeških glagola	Broj hrvatskih glagola
9.1 Put	75	46
10.1 Remove	327	97
11.1 Send	86	62
13.5.1 Get	-	68
18.1 Hit	6	56
22.2 Amalgamate	-	50
29.2 Characterize	-	26
30.3 Peer	-	96
31.1 Amuse	-	120
36.1 Correspond	-	34
37.3 Manner of expression	9	42
37.7 Say	-	13
40.2 Nonverbal express	-	33
43.1 Light emission	-	36
45.4 Change of state	-	249
47.3 Modes with motion	-	32
51.3.2 Run	-	65
<b>33. Judgement</b>	<b>233</b>	<b>70</b>
<b>17.1 Throw</b>	<b>23</b>	<b>26</b>

**Tablica 4.** Testne klase za češki i hrvatski jezik na temelju Levininih izvornih klasa



Budući da su se rezultati morali analizirati ručno, mali skup klasa i glagola koji pripadaju svakoj klasi omogućili su temeljitu evaluaciju dobivenih rezultata. I klase i pripadajući glagoli u češkome i hrvatskome jeziku su ručno uspoređeni kako bi se otkrio niz usporedivih značajki (morfoloških padeža, sintaktičkih pozicija i značenja).

Rezultat ručnoga definiranja skupa pravila bila su 103 morfosintaktička tranzicijska pravila za prijedložne sintagme, 7 za priložne, 3 za pridjevske i 18 za zavisne rečenice. Neka od pravila koja daju najbolji uvid u sintaktičke razlike između dva jezika prikazane su u tablicama 5 i 6.

Značenjska uloga	Površinska struktura u češkome	Površinska struktura u hrvatskome	Primjer
LOC (location:1)	čemu3	čim7	královna kraluje Anglii / kraljevna vlada Engleskom
PAT (person:1)	komu3	kim7	král kraluje svému lidu / kralj kraljuje svim narodima
PAT (person:1)	komu3	koga4	rozuměl mu / razumio ga je
FEEL (feeling:1)	čim7	od+čega2	srđce mu plesalo radostí / srce mu je plesalo od radosti
EVEN (event:1)	čeho2	čemu3	nadál se úspěchu / nadao se uspjehu
PAT (person:1)	koho2	komu3	nadál se bratra / nadao se bratu
PAT (person:1)	kym7	koga4	mrštil svým sokem na zem / oborio je protivnika na tlo
OBJ (object:1)	čim7	koga4	pohyboval nábytkem / pomicao je namještaj
CAUS (cause)	čim7	od+čega2	trpěl epilepsií / bolovao je od epilepsije

**Tablica 5.** Primjer morfosintaktičkih tranzicijskih pravila za morfološke padeže

Značenjska uloga	Češka sintaksa	Hrvatska sintaksa	Primjer
PAT (person:1), SOC (person:1)	proti+komu3	protiv+koga2	agitovao proti ředitelce / lobirao je protiv direktora
ATTR (attribute:2); SUBS (substance:1); KNOW (knowledge:1); EXT (number:2)	a+co4, s+ćim7	i+što4, s+ćim7	porovnal jejich vřhody a nevřhody / usporedio je njihove prednosti i nedostatke
SUBS (substance:1); ACT (act:2); STATE (state:4), TIME (start:2);	z+ćeho2	od+ćega2	těžii z jejich práce / profitirala je od njihovog rada
OBJ (object:1)	z+ćeho2	s+ćega2	běhal z kopce / trći s brda
ART (publication:1), LOC (location:1); OBJ (ground:7)	z+ćeho2	iz+ćega2	vybřiraly odpadky z kontejnerů / skupljali su smeće iz kontejnera
OBJ (object:1); LOC (space:2); OBJ (natural object:1); ART (covering:2)	zpod+ćeho2	ispod+ćega2	vytáhla zpod řatů časopis / ispod haljine je izvukla časopis
PAT (person:1), SOC (person:1)	z+koho4	od+koga2	syn tahal z rodićů nemalé finanćní částky / sin je od roditelja izvlaćio znaćajne kolićine novca
SOC (person:1),	po+kom6	od+koga2	zđedil dům po babiće / naslijedio je kuću od bake
PAT (person:1), ENT (person:1)	po+kom6	za+kim7	bernardýn řtěkl po kocourovi / bernardinac je lajao na maćka
TIME (occasion:2); ACT (act:2)	k+ćemu3	za+što4	k narozeninám si přeje veliký dort / za rođendan želi veliku tortu
REAS (reason:1)	kvuli+ćemu3	zbog+ćega2	kradl kvůli dluhům / krao je zbog dugova

**Tablica 6a.** Primjer morfosintaktićkih tranzicijskih pravila za prijedložne sintagme

Značenjska uloga	Češka sintaksa	Hrvatska sintaksa	Primjer
REAS (reason:1)	z+čeho2	od+čega2	zbláznil se z lásky / poludio je od ljubavi
DIFF (difference)	o+co4	za+što4	prohrály/vyhrál o jednu branku / izgubili su / pobijedili su za 1 gol
PAT (person:1), SOC (person:1); ENT (person:1)	o+koho4	za+koga4	bojovala o své děti / borila se za svoju djecu
OBJ (object:1); OBJ (natural object:1);	o+co4	za+što4	zajíma se o historii / zanima se za povijest
PAT (person:1)	na+koho4	koga2	pamatoval se na ni / sjećao se nje
OBJ (object:1); OBJ (natural object:1);	na+co4	čega2	pamatoval se na její ruce / sjećao se njene ruke
ACT (act:2);	proti+ čemu3	od+ čega2	bránil ho proti nájezdům Turků / branio ga je od najeзде Turaka
SUBS (substance:1); OBJ (object:1); STATE (state:4),	po+čem6	za+čim7	toužil po úspěchu / čeznuo je za uspjehom
PAT (person:1)	do+koho2	za+kim7	manžel se zbláznil do moderátorky / muž je poludio za moderatoricom

**Tablica 6b.** Primjer morfosintaktičkih tranzicijskih pravila za prijedložne sintagme

Usporedbom je ustanovljeno da se glagoli u najvećem broju mogu prevesti direktno, tj. i češki i hrvatski glagol u istoj semantičkoj klasi dijele isti valencijski okvir (prvu razinu semantike), i najčešće posjeduju dvije obvezne dopune - vršitelja radnje (AGT) u nominativu te trpitelja radnje (PAT) u akuzativu (analizovat / analizirati; absorbovat / apsorbirati; bombardovat / bombardirati; napodobovat / oponašati; ranit / raniti; rodit / roditi; kouřit / pušiti; uklízet /

čistiti; umět / znati; bít (tlouci) / biti (udarati); hodit / baciti; ozdobit / okititi itd.).

Usporedbom je također ustanovljeno da glagoli koji u češkome jeziku posjeduju više od dvije obvezne dopune, najčešće imaju dopunu vršitelja radnje, primatelja te trpitelja radnje. U češkome jeziku su sve te tri dopune često obvezne, dok je u valencijskome opisu hrvatskih ekvivalenata primatelj radnje obično tipični izborni dodatak (slíbit / obećati, vysvětlit / objasniti, darovat / darovati, umožnit / omogućiti, upozornit / upozoriti, vrátit / vratiti, zakázat / zabraniti, popřát / zaželjati, obětovat / žrtvovati).

Dosljednost postoji i kod glagola kretanja kojima se izražava promjena mjesta ili položaja vršitelja radnje. Kod sljedećih glagola su funktori smjera (DIR1 ili DIR3) u češkome jeziku obvezna dopuna, dok su u hrvatskome jeziku tipični izborni dodatak (posadit se / sjesti, odnést / odnijeti, odvést / odvesti, vracet se / vraćati se).

Kao što je vidljivo iz navedenih primjera, češki glagoli gotovo uvijek imaju više obveznih dopuna nego hrvatski.

Kao rezultat prvoga pokušaja proširivanja hrvatskoga leksikona preslikavanjem čeških sinskupova odabranih klasa na hrvatske sinskupove, dobili smo valencijske okvire 312 hrvatskih glagola (idiomi su izostavljeni). Izraz "identični valencijski okvir" odnosi se na valencijski okvir glagola koji ima isti broj dopuna s identičnom prvom i drugom razinom semantike u češkome i hrvatskome jeziku. Što se tiče morfosintaktičke realizacije semantičkih uloga, glagoli mogu dijeliti iste površinske oblike ili se primjenjuju morfosintaktička pravila. Rezultati evaluacije prikazani su u tablici 7.

Točnost glagolske valencija ( $ACC_{vv}$ ) je izračunana za svaki glagol u klasi, kao i točnost valencije klase ( $ACC_{cv}$ ) za svih 7 klasa.  $ACC_{vv}$  predstavlja udio točnih valencijskih okvira pridruženih glagolu, u odnosu na sve okvire dodijeljene glagolu. To je suma ovjerenih valencijskih okvira podijeljena s brojem svih valencijskih okvira prenesenih sa češkoga na hrvatski glagol.  $ACC_{cv}$  predstavlja zbroj  $ACC_{vv}$ -a svih glagola u svakoj klasi podijeljen s brojem glagola u klasi. Iako je dokazana visoka točnost dodjele čeških valencijskih okvira hrvatskim glagolima (293 od ukupno 312 hrvatskih glagola dijeli identične okvire s češkim ekvivalentima, a jedina razlika je morfosintaktičko ponašanje semantičkih uloga), prijenos je realiziran na istom skupu glagola za koje su ručno razvijana

pravila, pa se ova točnost može tretirati samo kao gornja granica koja se može dostići ovakvim pravilima.

Može se također zaključiti da Levinina klasifikacija pomaže u razrješavanju nejasnoća glede primjene morfosintaktičkih pravila. Na primjer, češki izraz "z + čeho" može biti preveden na hrvatski kao "od + čega" i "odakle". Ali, ako glagol u čijem se okviru izraz nalazi pripada klasi *glagola gradnje i stvaranja*, prevodi se kao "od + čega". Ako pak glagol pripada klasi *glagola uklanjanja ili čišćenja*, prevodi se kao "odakle".

Levinine klase	Broj uspoređenih glagola	Broj glagola s identičnim valencijskim okvirima	ACC <sub>cv</sub>
9.1 Put	46	41	0.96
10.1 Remove	97	89	0.97
11.1 Send	62	60	0.99
18.1 Hit	6	6	1
37.3 Manner of speaking	9	9	1
33. Judgement	70	68	0.98
17.1 Throw	22	20	0.97

**Tablica 7.** Rezultati evaluacije za eksperimente na svim usporedivim klasama

U zaključku eksperimenta s transjezičnim tranzicijskim pravilima i poluautomatskim dodjelivanjem valencijskih okvira možemo istaknuti:

1. Najčešće semantičke uloge (tipa vršitelj i trpitelj) i njihovi valencijski okviri mogu se automatski prenijeti iz jednoga jezika u drugi, ako glagolski sinskupovi pripadaju istoj semantičko-sintaktičkoj klasi (jer time dijele isto značenje i valencijski okvir). Primjerice, češki glagoli *hodit* i *hazet* (*baciti* i *bacati*) pripadaju klasi *17.1 throw – glagoli bacanja* i imaju sljedeće značenje "zamahom mijenjati mjesto čemu ili komu u određenom smjeru". Njihov okvir u češkome jeziku ima obveznu semantičku ulogu AGENT (person:1) i OBJECT (object:1) sa sintaktičkim oblicima AG (kdo1), OBJ (co4, čim7). U hrvatskome jeziku je broj i redosljed uloga ista, s jednom razlikom: OBJ se može realizirati samo kao *što4* (hrvatski ekvivalent za *co4*), ne kao *čim7*, koji je specifičan za češki.

2. Glagoli u oba jezika mogu dijeliti značenje i broj semantičkih uloga koje se sintaktički realiziraju različitim morfološkim padežima. Stoga su potrebna morfosintaktička pravila transfera čeških valencijskih okvira na hrvatske glagole i obrnuto. Primjerice, češki glagoli *blahopřát*, *gratulovat*, *poblahopřát*, *pogratulovat*, *popřát*, *přát* pripadaju Levininoj klasi 33. *judgement-glagoli osude* i prevedeni su na hrvatski jezik kao *čestitati*, *poželjeti*, *željeti*, *zaželjeti* sa sljedećim značenjem "prenijeti zadovoljstvo, odobrenje ili pohvalu osobi". Njihovi se okviri u češkome i hrvatskome sastoje od identičnih obveznih semantičkih uloga AGENT (person: 1) i PAT (person: 1) s istim sintaktičkim oblicima, ali izborni dodatak TIME se u češkome jeziku realizira kao *k + čemu3* ili *na + co4*, dok se u hrvatskome jeziku može realizirati samo kao *što4* ili *na + čemu6*.

Posljednji korak eksperimenta obuhvatio je dvije Levinine klase (*18.1 hit-glagoli gađanja* i *37.3 manner of expression – glagoli načina izražavanja*) koje su sadržavale više hrvatskih glagola nego čeških. Valencijski okviri prevedenih čeških glagolskih sinkupova su primijenjeni na sve hrvatske glagole u klasi kako bi se ispitala mogućnost automatskoga stvaranja obogaćene verzije hrvatskoga valencijskog leksikona.

Rezultat u tablici 8 pokazuje da se od 6 valencijskih okvira (koji vrijede za svih 6 čeških glagola u klasi HIT), prvi okvir može automatski primijeniti na većinu hrvatskih glagola u klasi HIT (53 glagoli ukupno), a drugi okvir na ostala 3 glagola u toj klasi. Ostala četiri okvira (3) – (6) u tablici 8 ne vrijede za sve hrvatske glagole zbog druge razine semantike koju unose u opis, ali se mogu dodijeliti hrvatskim glagolima u klasi poluautomatski.

Istraživanje prekojezičnih svojstava Levininih klasa i poluautomatske dodjele valencijskih okvira glagolima u CROVALLEX i VerbaLex leksikonu otkrilo je velike sličnosti između češke i hrvatske klasifikacije glagola, ali također i novu mogućnost obogaćivanja hrvatskoga leksikona glagola valencijskim opisima na poluautomatski način sa semantičkim ulogama druge razine, uspostavljajući vezu s Princeton WordNet i CzechWordNet hijerarhijama i uvođenjem iznimno relevantnog razlikovanja živih i neživih sastavnica. Naš eksperiment pokazuje da je, čak i za jezike koji nemaju dovoljan broj računalno podržanih jezičnih resursa, moguće stvoriti osnovni valencijski leksikon primjenom maloga broja pravila za prekojezični morfosintaktički prijenos uz korištenje Levinine

klasifikacije glagola. U sljedećem koraku pokušat će se poboljšati kvaliteta transfera uvođenjem uvjetnih pravila koja će se temeljiti na supojavljanju dopuna glagolu u valencijskome okviru.

<b>Levinina klasa: 18.1 HIT</b>	
Češki glagoli u klasi HIT i hrvatski ekvivalenti:	atakovat-atakirati, dorážet-navaljivati, napadat-napadati, napadnout-napasti, utočit-nasrtati, zautočit-nasrnuti
Hrvatski glagoli u klasi HIT:	<i>atakirati</i> , batinati, bičevati, biti, bombardirati, dohvatiti, gađati, grebati, grepsti, gristi, ispljuskati, isprebijati, istući, izlemati, izlupati, izudarati, izujedati, kamenovati, kvrcnuti, lemati, lupati, lupiti, mitraljirati, mlatiti, munuti, nagristi, <i>napadati</i> , <i>napasti</i> , <i>nasrnuti</i> , <i>nasrtati</i> , <i>navaljivati</i> , ogrebat, ogrepsti, okrznuti, ošamariti, ošinuti, pendrečiti, pljunuti, pljusnuti, pogoditi, potući, prebiti, raketirati, razbijati, razbiti, strijeljati, šamarati, šibati, tući, udarati, udariti, ujesti, trknuti, upucati, zveknuti, zviznuti
Valencijski okviri u VerbaLex-u za klasu HIT	(1) AG <sub>kdo1</sub> :obl(person:1) VERB PAT <sub>koho4</sub> :obl(person:1) (2) AG <sub>kdo1</sub> :obl(person:1) VERB PAT <sub>na+koho4</sub> :obl(person:1) (3) ENT <sub>co1</sub> :obl(animal:1) VERB ENT <sub>co4</sub> :obl(animal:1) (4) ENT <sub>co1</sub> :obl(animal:1) VERB PAT <sub>koho4</sub> :obl(person:1) (5) ENT <sub>co1</sub> :obl(animal:1) VERB PAT <sub>na+koho4</sub> :obl(person:1) (6) GROUP <sub>co1</sub> :obl(group:1) VERB PAT <sub>koho4</sub> :obl(person:1)
Hrvatski valencijski okviri za klasu HIT prilagođeni prema češkim okvirima	(1) AG <sub>tko1</sub> :obl(person:1) VERB PAT <sub>koga4</sub> :obl(person:1) (2) AG <sub>tko1</sub> :obl(person:1) VERB PAT <sub>na+koga4</sub> :obl(person:1) (3) ENT <sub>tko1</sub> :obl(animal:1) VERB ENT <sub>koga4</sub> :obl(animal:1) (4) ENT <sub>tko1</sub> :obl(animal:1) VERB PAT <sub>koga4</sub> :obl(person:1) (5) ENT <sub>tko1</sub> :obl(animal:1) VERB PAT <sub>na+koga4</sub> :obl(person:1) (6) GROUP <sub>tko1</sub> :obl(group:1) VERB PAT <sub>koga4</sub> :obl(person:1)
Češko-hrvatski glagolski parovi u klasi HIT s brojem okvira koji se na njih odnosi	atakovat-atakirati (1, 3, 4, 6), dorážet-navaljivati (2, 5), napadat-napadati (1, 3, 4, 6), napadnout-napasti (1, 3, 4, 6), utočit-nasrtati (2, 5), zautočit-nasrnuti (2, 5)

**Tablica 8.** Primjer eksperimenta s hrvatskim glagolima