

UDK 811.163.42'373.611

811.163.42'322.2

Izvorni znanstveni članak

Prihvaćeno za tisak 22. ožujka 2013.

Krešimir Šojat
Sveučilište u Zagrebu
ksojat@ffzg.hr
Matea Srebačić
Sveučilište u Zagrebu
msrebaci@unizg.hr
Vanja Štefanec
Sveučilište u Zagrebu
vstefane@ffzg.hr

CroDeriV i morfološka raščlamba hrvatskoga glagola

U radu se prikazuje izrada leksikona hrvatskih glagola CroDeriV i teorijske postavke na kojima ona počiva. CroDeriV je računalni leksikon koji sadržava podatke o morfološkoj strukturi gotovo 14.000 hrvatskih glagola. U prvom dijelu članka prikazuju se postupci morfološke analize glagola i međusobnoga povezivanja glagola s istim korijenom. Glagoli su u prvoj fazi izrade CroDeriV-a automatski segmentirani s pomoću pravila. U drugoj fazi rezultati segmentacije i svodenja na isti korijen ručno su provjereni. U drugome dijelu članka obrazlaže se uopćeni prikaz morfološke strukture hrvatskoga glagola utemeljen na podatcima dobivenima iz CroDeriV-a, pri čemu se posebna pozornost pridaje vrstama, funkciji i značenju sufikasa. Naposljetku se iznose empirijski podatci o mogućim kombinacijama afikasa koji sudjeluju u tvorbi hrvatskih glagola, kao i o njihovoj frekvenciji utvrđenoj analizom glagola iz CroDeriV-a.

1. Uvod¹

Razvoj računalne lingvistike i računalnih aplikacija koje se koriste prirodnim jezikom tijekom posljednjih tridesetak godina potaknuo je izradu jezičnih resursa za čitav niz jezika, pa tako i za hrvatski. Računalni resursi za hrvatski jezik obuhvaćaju jednojezične i paralelne korpuse većega ili manjega opsega, banku stabala i leksikone poput morfološkoga leksikona, leksikona glagolskih

1 Zahvaljujemo anonimnim recenzentima na iscrpnim i inspirativnim recenzijama i mnogobrojnim dobronamjernim uputama, koje su, nadamo se, ovaj rad učinile jasnijim i kvalitetnijim. Dakako, sva odgovornost za eventualne pogreške samo je naša.

valencija i semantičke mreže.² Takvi se resursi mogu koristiti u razvoju alata za obradu prirodnoga jezika i raznim računalno potpomognutim jezikoslovnim istraživanjima. Izrada jezičnoga resursa, bez obzira na njegovu svrhu, pretpostavlja prikupljanje i obilježavanje jezičnih podataka. Vrsta i način obrade jezičnih podataka u pravilu se razlikuju prema jezičnoj razini na koju se odnose (npr. morfološki, sintaktički ili semantički leksikoni). Na pojedinoj razini jezični resursi često su specijalizirani, odnosno obuhvaćaju samo određene podatke (npr. flektivni leksikon na morfološkoj razini ili valencijski leksikon na sintaktičkoj razini). Na tako uže određenim područjima jezični se resursi mogu razlikovati prema teorijskome okviru unutar kojega se podatci analiziraju i obilježavaju. U ovome članku opisuje se izrada računalnoga leksikona CroDeriV koji sadržava morfološke podatke o hrvatskim glagolima. U prvome dijelu članka obrazlaže se motivacija za sastavljanje takve vrste jezičnoga resursa i opisuju se teorijske pretpostavke na kojima počiva analiza i obilježavanje jezičnih podataka. Na temelju analiziranih glagola izvodi se i argumentira uopćena morfemska struktura hrvatskoga glagola. Morfemska struktura hrvatskoga glagola sastoji se od konačnoga broja utora s pripadajućim značenjima. Utori se dijele na prefiksalsne, sufiksalsne, korijenske i interfiksalsne. Posebno se detaljno opisuju tri sufiksalsna utora predviđena unutar morfemske strukture hrvatskih glagola. U drugome dijelu članka prikazuju se moguće kombinacije prefikasa i sufikasa dobivene analizom glagola u CroDeriV-u. Te su kombinacije prvi put u hrvatskoj lingvistici sustavno prikazane i potkrijepljene egzaktnim podacima o njihovoj frekvenciji dobivenima analizom opsežna uzorka glagola. Na kraju, raspravlja se o daljnjim mogućnostima korištenja prikazanoga morfološkog leksikona.

2. Motivacija

Inicijalni poticaj za izradu morfološkoga leksikona CroDeriV pojavio se tijekom rada na Hrvatskome WordNetu, računalnome leksikonu čija je temeljna jedinica skup sinonima – *sinskup*. Sinskupovi su međusobno povezani odnosima poput hiponimije, meronimije i antonimije. U izradi takve leksičko–semantičke mreže, a posebice tijekom izrade glagolskih sinskupova, uočeno je da se značenjske veze između hrvatskih glagola koji služe kao tvorbena osnova za tvorbu drugih glagola derivacijskim afiksima u velikome broju slučajeva ne mogu obuhvatiti odnosima koji se koriste između glagolskih sinskupova kao cjelina. Odnosi između glagolskih sinskupova, koje dalje nazivamo semantičkim odnosima, obuhvaćaju hiponimiju/hiperonimiju (A je B na određeni način, npr. *žderati* – *jesti*), antonimiju (A je suprotno od B, npr. *postiti* – *jesti*), uzrok (A

2 Svi su resursi dostupni u bazi META-SHARE (<http://meta-share.ffzg.hr/repository/search/>). Više o izradi Hrvatskoga morfološkoga leksikona (<http://hml.ffzg.hr/hml/>) v. u Tadić i Fulgosi (2003), o Hrvatskoj ovisnosnoj banci stabala (<http://hmk.ffzg.hr/hobs/>) u Tadić (2007), o Leksikonu valentnosti hrvatskih glagola (<http://theta.ffzg.hr/crovallex/>) u Mikelić Preradović i dr. (2009), a o Hrvatskome WordNetu u Raffaelli i dr. (2008).

uzrokuje B, npr. *hraniti – jesti*) i poddogadaj (A je dio B, npr. *žvakati – jesti*). Ti semantički odnosi pokazali su se nedovoljnima, odnosno preuskima za obuhvaćanje cijeloga spektra odnosa koji proizlaze iz tvorbene, a time i značenjske povezanosti hrvatskih glagola. Temeljni odnos na kojemu počiva struktura leksičkih hijerarhija u Hrvatskome WordNetu jest hiperonimija/hiponimija. Prilikom izrade leksičkih hijerarhija za hrvatske glagole može se postaviti pitanje je li glagol *uplivati* hiponim glagola *plivati* ili *ući*, s obzirom na to da je semantički povezan s obama. Nadalje, može se postaviti pitanje kako označiti odnose između glagola poput *plivati* i *isplivati*, *preplivati*, *uplivati*, kao i *preplivavati*, *uplivavati* itd. Zbog specifična i unaprijed zacrtana oblika leksikona svaki sinskup može imati samo jedan hiperonim pa je sinskup s glagolom *uplivati* označen kao hiponim sinskupa s glagolom *ući*.³ Takva su se rješenja pokazala nedostatnima za opis glagolskoga dijela hrvatskoga leksika i potaknula razradu dodatnoga skupa odnosa koji se nazivaju morfosemantičkim odnosima.⁴ Njihova karakteristika jest da se ne protežu između cjelokupnih sinskupova, već između njihovih pojedinih članova.⁵ Morfološka se komponenta tih odnosa temelji na zajedničkome korijenu i tvorbenoj povezanosti članova različitih sinskupova. Njihova se semantička komponenta temelji na vrsti semantičke veze između tvorbeno povezanih članova različitih sinskupova. Morfološki uvjetovana semantička interpretacija i označavanje morfosemantičkih odnosa relativno su jednostavni kod predvidljive i jednoznačne modifikacije osnovnoga oblika, kao npr. kod tvorbe imeničnih umanjenica, mocijskih parova, imenica koje označavaju vršitelja radnje itd. U izradi semantičke mreže prema ocrtanim načelima znatno je veći problem određivanje odnosa između tvorbeno povezanih glagola. To vrijedi ponajprije za glagole koji se derivacijskim afiksima tvore od drugih glagola jer se od osnovnih oblika u pravilu razlikuju i u glagolskome vidu. Kako bi se uopće moglo pristupiti razradi morfosemantičkih odnosa, npr. između neprefigiranih i prefigiranih oblika glagola, a i proširenju i obogaćivanju Hrvatskoga WordNeta, potrebno je znati koji derivirani oblici određenoga glagola uopće postoje, odnosno koji afiksi mogu sudjelovati u derivaciji određenih glagola. Budući da konkretni podatci o tvorbenim procesima veće skupine glagola dosad nisu objavljeni ili nisu javno dostupni, nametnula se potreba za izradom jezičnoga resursa koji bi pružio uvid u što veći broj ovjerenih kombinacija derivacijskih afikasa i polaznih oblika. Takav bi resurs s jedne strane zadovoljio gore navedene potrebe, dok bi s druge strane barem djelomično popunio nedostatak računalnoga resursa posvećenoga izuzetno bogatoj tvorbi riječi u hrvatskome jeziku.

3 O strukturi Hrvatskoga Wordneta i metodama njegove izrade vidi Raffaelli i dr. (2008).

4 O morfosemantičkim poljima u hrvatskome jeziku vidi Raffaelli i Kerovec (2008) i Katunar i Šojat (2011).

5 Treba napomenuti da se sinskup može sastojati i od samo jednoga člana.

2.1. Računalna obrada morfologije hrvatskoga jezika

Računalna obrada morfologije hrvatskoga jezika dosad je primarno bila usmjerena na fleksiju. Na tome području prednjači Hrvatski morfološki leksikon⁶ (dalje u tekstu HML), koji sadržava 120.000 lema i sve njihove flektivne oblike. Podsastavnica HML-a jest Hrvatski lematizacijski poslužitelj,⁷ koji omogućava generiranje i prepoznavanje svih oblika riječi iz HML-a. Tadić i Fulgosi (2003) razmatraju mogućnost proširenja HML-a uvođenjem derivacijske sastavnice leksikona, odnosno njegovom podjelom na flektivni i derivacijski dio. Buduća derivacijska sastavnica HML-a temeljila bi se na automatskome generiranju kombinacija leksičkih i derivacijskih morfema propisanim pravilima. Preduvjeti za proširenje HML-a derivacijskom komponentom jesu popis morfema koji se kombiniraju i popis pravila za generiranje ovjerenih kombinacija, pri čemu bi popis morfema koji se kombiniraju trebao biti razdvojen na derivacijske i leksičke. Popis derivacijskih morfema može se relativno jednostavno utvrditi pregledom postojećih gramatika i Babićeve *Tvorbe* (2002), ali je znatno teže skupiti opsežan popis leksičkih morfema. Čavar i dr. (2008) navode da u Institutu za hrvatski jezik i jezikoslovlje postoji popis od 250.000 leksičkih, derivacijskih i flektivnih morfema. Ta leksička baza, koja između ostaloga sadržava gotovo 77.000 imenica i 25.000 glagola segmentiranih na morfeme, služi kao podloga alatu za morfološku analizu i lematizaciju *CroMo* (Čavar i dr. 2008; Čavar i dr. 2009). Nažalost, resursi na temelju kojih je izgrađen taj preobličivač s konačnim brojem stanja nisu javno dostupni, kao ni način na koji su ti resursi prikupljeni.

Navedena leksička baza i HML ručno su izgrađeni resursi. Šnajder (2008) prikazuje postupke strojne obrade morfologije hrvatskoga leksika i automatske izrade morfološkoga leksikona. Flektivna se sastavnica modela koristi za izradu flektivnoga leksikona iz neoznačena korpusa, a derivacijska za grupiranje tvorbeno povezanih unosaka flektivnoga leksikona. Derivacijska sastavnica obuhvaća imenice, glagole i pridjeve povezane sufiksalsnom tvorbom. Obje se sastavnice radi automatskoga crpljenja podataka koriste za morfološku normalizaciju, odnosno lematizaciju i korjenovanje riječi. Korjenovanje označava postupak uklanjanja afikasa iz oblika tvorbeno povezanih riječi kako bi se dobio zajednički korijen. Pritom su afiksi koji se uklanjaju u pravilu sufiksi, a dobiveni oblici nisu leksički morfemi riječi nego oblične ili tvorbene osnove. Korjenovatelji za hrvatski jezik koji funkcioniraju na sličnim principima prikazani su u Ljubešić i dr. (2007) i Pandžić (2012). Svi navedeni korjenovatelji usmjereni su na prepoznavanje derivacijskih i flektivnih sufikasa, dok povezanost osnova i prefiksalsno deriviranih oblika ni kod jednoga nije obuhvaćena. Iako su ti alati u većoj ili manjoj mjeri temeljeni na lingvističkim pravilima, njihov cilj nije morfološka raščlamba i stoga se ne mogu koristiti za opsežnija istraživanja tvorbe riječi. Činjenica da za jezik s izrazito razvijenim i razgra-

6 Pretraživ na <http://hml.ffzg.hr/hml/>. Više o leksikonu v. u Tadić i Fulgosi (2003).

7 Pretraživ na <http://hml.ffzg.hr/hml/info.php?show=hlp> (v. i Tadić 2006).

natim derivacijskim procesima, kao što je hrvatski, ne postoji računalni resurs koji bi dao uvid u tvorbene uzorke i tvorbene porodice,⁸ poticaj je za njegov razvoj.

2.2. Teorijske postavke

Na samome početku, prije analize dosadašnjih postavki na području tvorbe riječi hrvatskih jezikoslovaca i vlastita opisa morfološke strukture hrvatskoga glagola, važno je razgraničiti osnovne pojmove kojima ćemo se koristiti. Ponajprije valja razlikovati morfem i morf. Pod pojmom *morf* podrazumijevaju se različiti fizički ostvaraji nekoga morfema u površinskoj postavi, a za prikaz, predodčenje toga morfema u dubinskoj postavi uzima se morf koji se smatra osnovnim ili temeljnim (Marković 2012: 38) i iz kojega se glasovnim promjenama uz primjenu pravila mogu izvesti svi njegovi alomorfi u hrvatskome jeziku (v. npr. Marković 2012: 37, 361, Babić, Z. 1991: 10). *Morfem* čini skup svih morfova istoga sadržaja u komplementarnoj distribuciji. U CroDeriV-u prikazane su i površinska i dubinska postava svake leme, pri čemu se na površinskoj postavi nalaze morfovi koji su spojeni na zajednički morfem u dubinskome prikazu. Drugim riječima, kada govorimo o površinskome prikazu, govorimo o morfovima, a kada govorimo o dubinskome prikazu, govorimo o morfemima koji obuhvaćaju sve svoje potvrđene ostvaraje – *alomorfe* – u hrvatskome jeziku. Na kraju, valja razlikovati između tvorbene osnove i korijena (leksičkoga morfa). *Tvorbena* je *osnova* ona koja sudjeluje u jednome derivacijskome procesu i na koju se izravno dodaje derivacijski afiks, a *korijen* je dio riječi koji nastaje kad se odstrane svi ostali morfovi i koji je nositelj temeljnoga značenja riječi (v. i Marković 2012: 50). *Tvorbena osnova* i *korijen* mogu se i ne moraju podudarati (npr. *bac-* u *baciti* istovremeno je i korijen i tvorbena osnova, dok je u *razumjeti* tvorbena osnova *razum*, a korijen *um*). Termini *morf* i *morfem*, *osnova* i *korijen* nerijetko u literaturi nisu jasno definirani, pa se čini da bi jasno razgraničenje moglo pridonijeti boljem razumijevanju teorijskih postavki, a onda i iz njih izvedene morfološke strukture hrvatskoga glagola koji će se prikazati u ovome i sljedećem poglavlju.

Osim Babićeva kapitalnoga djela *Tvorba riječi u hrvatskome književnome jeziku* (1986, ovdje se koristi treće izdanje iz 2002) i prilično opsežna poglavlja posvećena tvorbi riječi u Maretićevoj *Gramatici i stilistici hrvatskoga ili srpskoga jezika* iz davne 1899. opisu je tvorbenih fenomena hrvatskoga jezika posvećeno relativno malo prostora. Ni samo mjesto tvorbe riječi u gramatičkome opisu nije jednoznačno određeno. Neki tvorbu riječi promatraju kao dio morfologije, koju onda dijele na flektivnu i derivacijsku, a neki kao

8 Tvorbene porodice ovdje obuhvaćaju sve lekseme sa zajedničkim korijenom, neovisno o značenjskim pomacima koji se događaju pri tvorbenim procesima. Dakle, u tvorbenu porodicu sa zajedničkim korijenom *bac-* uključeni su i glagoli *baciti*, *ubaciti*, *prebaciti* (odnosno glagoli koji i dalje označavaju bacanje), ali i *nabaciti* u značenju 'natuknuti' i *pobaciti* kod kojega dolazi do potpunoga značenjskog odmaka od polaznoga oblika u tvorbi.

izdvojenu od morfologije ili kao dio leksikologije.⁹ U Barić i dr. (2003) tvorba se riječi obrađuje kao zasebno poglavlje, odnosno odvojeno od morfologije. U Akademijinoj gramatici tvorba riječi i fleksija obrađuju se u zasebnim svescima.¹⁰ U nastavi hrvatskoga jezika tvorba riječi tradicionalno se poučava kao dio leksikologije.¹¹ Silić i Pranjković (2005) tvorbu riječi promatraju kao dio morfologije, koju dijele na tvorbu oblika riječi i tvorbu riječi, dok Raguž (1997) u svojoj *Praktičnoj gramatici* uopće ne spominje tvorbu. Osim što je tvorbenim procesima hrvatskoga jezika u cjelini posvećeno samo jedno cjelovito djelo, opis tih procesa i jezičnih jedinica koje sudjeluju u njima može se razlikovati kod različitih autora. Autori se znatno razlikuju u opisu morfološke strukture hrvatskih glagola, odnosno u određivanju jezičnih jedinica i njihovih funkcija u tvorbenim procesima. Te razlike odnose se na 1) nerastavljanje tvorbenih sufikasa na morfove ili 2) različito rastavljanje tvorbenih sufikasa na morfove, što za izravnu posljedicu ima 3) različit broj morfova, a onda i morfema infinitivne osnove.

S jedne strane, nerastavljanje sufikasa karakteristično je u prikazu tvorbenih procesa kod Babića (2002) i Barić i dr. (2003). U opisu sufiksalske tvorbe u pravilu se koriste kombinacije sufikasa, pri čemu se kod opisa glagolske sufiksalske tvorbe također nerijetko ne odvajaju ni infinitivni sufiks, pa se kao sufiksi navode npr. *-uckati*, *-uškati*, *-injati*. S druge strane, Silić i Pranjković (2005) i Marković (2009, 2012) tvorbene sufikse rastavljaju na morfove, ali se njihove kombinacije prikazuju kao *uc-k-a-ti* i *uck-a-ti*, a glagoli se prikazuju segmentirani na morfove kao *kup-ov-a-ti*, ali i *kup-o-va-ti*. Prilikom promišljanja strukture derivacijskoga leksikona CroDeriV zaključeno je da se tvorbeni postupci koji postoje kod hrvatskih glagola najbolje mogu uočiti i opisati na temelju raščlambe veće skupine glagola na morfove i supostavljanjem njihove strukture, iz čega će biti moguće izvesti uopćenu dubinsku morfemsku strukturu hrvatskoga glagola. Tako pripremljena građa znatno olakšava prepoznavanje i opis tvorbenih sredstava koja se koriste u određenome derivacijskom procesu, kao i slijed pojedinih tvorbenih postupaka. Slijed tvorbenih postupaka kod glagola može se odnositi na određivanje koja se tvorbeni osnovi koristi za prefiksaciju ili za sufiksaciju, odnosno koja osnovi uopće može derivacijom dobiti prefiks ili sufiks. Osim toga, tako pripremljena građa omogućuje i uvid u razne kombinacije prefikasa i sufikasa u tvorbi hrvatskih glagola. Napokon, tako pripremljena građa vrijedna je podloga za istraživanje tvorbenih procesa u hrvatskome jeziku koji nadilaze samo prefiksalsnu tvorbu glagola od glagola, što je bio početni poticaj za izradu CroDeriV-a. Specifični postupci obrade jezične građe i postavke na kojima se temelje navode se u sljedećemu poglavlju.

9 Iscrpan pregled pristupa tvorbi riječi u hrvatskome jezikoslovlju daje Kuna (2006).

10 U svesku *Povijesni pregled, glasovi i oblici hrvatskoga književnoga jezika* (poglavlje o morfologiji napisali su Stjepan Babić i Stjepko Težak) obrađena je fleksija, a u već spomenutoj Babićevoj *Tvorbi riječi u hrvatskome književnome jeziku* derivacija.

11 Usp. Samardžija (1995), Tafra, Košutar (2009), Kuna (2006), Marković (2009).

3. CroDeriV – morfološki leksikon hrvatskih glagola¹²

Na samom je početku izrade CroDeriV-a bilo potrebno donijeti nekoliko odluka povezanih:

- a) s polaznom jedinicom u leksikonu,
- b) s dubinom tvorbene, odnosno morfološke raščlambe,
- c) s dubinskom i površinskom postavom morfološke strukture jedinica.

Polazni oblik u izradi leksikona jest glagol u infinitivu. U početnoj je fazi prikupljeno oko 14.000 glagolskih lema iz postojećih dostupnih rječnika za hrvatski jezik.¹³ Taj će se popis u sljedećoj fazi izrade nadopuniti lemama iz hrWaC-a, lematiziranoga i morfosintaktički označenoga hrvatskoga web-korpusa.¹⁴ Radi što veće buduće iskoristivosti CroDeriV-a, rastavljanje lema na morfove i analiza morfološke strukture zahtijevali su nešto više promišljanja. Kako bi se u budućnosti omogućio uvid u cjelovite tvorbene porodice, odnosno porodice koje obuhvaćaju različite vrste riječi, odlučeno je da se morfološka raščlamba infinitiva preko tvorbene osnove protegne i do korijena. Budući da izravna raščlamba do korijena često rezultira tvorbenim uzorcima koji uključuju sufikse koji se ne koriste u tvorbi glagola, taj je postupak razdijeljen na dva koraka. Primjerice, glagol *godištiti* tvoren je od imenice *godište* sufiksom *-i-* pa njegova tvorbena raščlamba izgleda ovako: *godišt-i-ti*. S obzirom na to da je imenica *godište* tvorena sufiksalsnom tvorbom od imenice *god* i sufiksa *-išt-e*, potpuna bi morfska raščlamba glagola *godištiti* izgledala ovako: *god-išt-i-ti*. Zbog takvih je primjera odlučeno da se u prvome koraku odvajaju afiksi koji se koriste u tvorbi glagola (dakle: *godišt-i-ti*), a da se u drugome koraku lema spaja na korijen zajednički cijeloj tvorbenoj porodici, u ovome slučaju na *god-*. Na sličan je način riješen problem alomorfije, i prefiksalsne i sufiksalsne, a i korijenske, odnosno alomorfije osnove. Za svaku je lemu predviđen dvorazinski prikaz. Na prvoj razini površinska je postava lema rastavljena na morfove od kojih je svaki na drugoj razini, u dubinskoj postavi, spojen na zajednički, temeljni morfem. Za dubinski prikaz morfema odabran je onaj njegov morf iz kojega se glasovnim promjenama u hrvatskome jeziku uz primjenu pravila mogu izvesti ostali morfovi. Tako se npr. površinska raščlamba glagola *raščišćavati* može prikazati kao *raš-čišć-ava-ti*, dok je u dubinskoj strukturi, primjerice, morf *raš-* povezan s *raz-*, a *-čišć-* s *-čist-*. Korijen *čist* služi kao poveznica s tvorbeno povezanim lemama poput *čistiti* ili *pročišćavati*. Takva dvorazinska notacija omogućava: a) detekciju tvorbenih uzoraka jer se prefiksalsni i sufiksalsni alomorfi svode na zajednički oblik; b) detekciju tvorbeno povezanih riječi jer se prepoznaju riječi s istim korijenom.

Ako se lema sastoji od kombinacija nekoliko prefikasa i sufikasa, prefiksi i sufiksi međusobno su razdvajani. Popisi prefikasa i sufikasa koji sudjeluju u tvorbi glagola preuzeti su iz Babićeve *Tvorbe riječi* (2002: 503–557) i dodatno su modificirani. Znatnija modifikacija odnosila se na analizu sufikasa, što je

¹² Pretraživanje CroDeriV-a u skoroj će budućnosti biti javno dostupno na internetu.

¹³ Ponajprije iz *Velikoga rječnika hrvatskog jezika* Vladimira Anića.

¹⁴ V. Ljubešić i Erjavec (2011). hrWaC se sastoji od 1,2 milijarde pojava prikupljenih s .hr domene.

detaljnije objašnjeno u poglavlju 4.3. Osnovna modifikacija Babićeva pristupa jest u trojumu:

1. infinitivni je sufiks odvojen (pri čemu se *-ći* promatra kao alomorf od *-ti*),
2. sufiksi koji se dobivaju nakon odvajanja infinitivnoga sufiksa promatraju se kao dvomorfski¹⁵ (npr. *-uck-a-*, *-uš-i-*, *-Ø-a-*),
3. sufiksi koji se razlikuju samo u početnome *j* (npr. *-java-/ -ava-*, *-jiva-/ -iva-*) promatraju se kao alomorfi.¹⁶

Na temelju utvrđenih popisa derivacijskih afikasa izrađena su pravila za automatsku segmentaciju glagola na morfove. Jedan skup pravila odnosio se na odvajanje prefikasa, a drugi na odvajanje sufikasa od tvorbene osnove.¹⁷ Nažalost, rezultat primjene tih postupaka velik je broj pogrešno segmentiranih glagola zbog glasovne podudarnosti afikasa i korijena. Automatskim prepoznavanjem i odvajanjem prefikasa, između ostaloga, glagol *privilegirati* rastavljen je na *pri-vileg-ir-a-ti* umjesto na *privileg-ir-a-ti*, a glagol *sniježiti* na *s-nijež-i-ti* umjesto na *snijež-i-ti*. Problemi kod automatskoga odvajanja sufikasa i osnove mogu se ilustrirati primjerima poput glagola *pobjeći* koji je rastavljen na *po-b-Ø-je-ći* umjesto na *po-b-je-Ø-Ø-ći* ili glagola *urastati* koji je rastavljen na *u-ras-t-a-ti* umjesto na *u-rast-Ø-a-ti*. Zbog velikoga broja sličnih slučajeva rezultati automatske raščlambe ručno su provjereni.

U velikom broju slučajeva također nije bilo moguće automatski prepoznati homografnе korijenske morfove i s njima točno povezati sve glagole koji ih sadržavaju, odnosno raščlambom je često dobiveno nekoliko značenjski različitih korijena istoga fonemskog sastava. Na primjer, rastavljanjem glagola *bojati se* 'osjećati strah' i glagola *bojiti/bojati* 'nanositi boju' dobiveni su homografni oblici. Prilikom ručne provjere dobivenih podataka ti su oblici označeni različitim oznakama kako bi se naznačile značenjske razlike, a time i pripadnost različitim tvorbenim porodicama. Dakle, korijen glagola *bojati se* označen je kao *boj1*, a korijen glagola *bojiti* i *bojati* označen je kao *boj2*. Dodatni problem u povezivanju lema na zajednički korijen bili su homografi koji uz isti korijen imaju istu morfološku strukturu, ali se razlikuju naglasno i vidski. Takvi su slučajevi riješeni unosom dviju lema koje se razlikuju u obilježju vida, a naglasna razlika nije vidljiva u ovako strukturiranome leksikonu (npr. *pogledati* (IPF) i *pògledati* (PF)). Osim površinske i dubinske postave, svakoj je lemi pridano i obilježje glagolskoga vida. Nesvršenost je označena s IPF, svršenost s PF, a dvovidnost s BI, čime je mogućnost višestruke interpretacije kao u navedenom primjeru uklonjena.

15 Iako bi se mogli promatrati i kao tromorfski, pa čak i četveromorfski, o čemu će više biti riječi u sljedećemu poglavlju.

16 Za utvrđivanje jasnih pravila prema kojima bi se iz dubinske nejotirane postave mogla izvesti površinska jotirana postava trebalo bi provesti detaljno zasebno istraživanje, no evo nekoliko argumenata za promatranje tih varijanata kao alomorfa: Petar Skok (1971: 77) kao zasebne natuknice navodi samo sufikse *-av-a-ti*, *-ev-a-ti*, *-iv-a-ti*, *-ov-a-ti*, dok jotirane varijante ne navodi; Silić (2002: 19) kaže da bismo osnovski morfem (!) *-ja-* mogli nazvati alomorfom morfema *-a-*, a ostale morfeme navodi kao *-(j)ava-* i *-(j)iva-*, iz čega se može zaključiti da ih smatra alomorfima.

17 Detaljan prikaz izrade i primjene tih pravila nalazi se u Šojat, Srebačić i Tadić (2012).

4. Struktura CroDeriV-a

Pri izradi CroDeriV-a vodilo se računa o njegovoj iskoristivosti u daljnjim istraživanjima hrvatskoga jezika, a samim time i o različitim mogućnostima pretraživanja podataka. Rastavljanjem lema na sve morfove, a ne samo na tvorbene osnove i afikse, omogućen je uvid u dosad slabo istraženo područje kombiniranja prefikasa i sufikasa s istim ili različitim korijenom. Preduvjet za utvrđivanje položaja i funkcije nekoga derivacijskog afiksa i usporedbe s njegovim položajem i funkcijom kod drugih glagola jest uniformno prikazana morfološka struktura svih glagola u leksikonu. To znači da je za konzistentan opis hrvatskih glagola prema navedenim načelima na dubinskoj razini bilo potrebno utvrditi jedinstvenu morfemsku strukturu u koju se mogu uklopiti svi glagoli. Prikaz takve uopćene morfemske strukture glagola u CroDeriV-u, utemeljene na prethodno napravljenom morfološkom analizom glagola, donosi se u sljedećim poglavljima.

4.1. Tvorbene osnove i korijeni

Razdioba glagola na morfove odvijala se u dvama koracima. U prvome koraku odvojeni su svi prefiksi i svi sufiksi koje navodimo u poglavljima 4.2. i 4.3. Tako su dobivene tvorbene osnove glagola koje mogu biti a) jednomorfske – oblikom jednake korijenu i b) dvomorfske (u nekim slučajevima i višemorfske) – sastavljene od korijena i derivacijskoga sufiksa za tvorbu oblične osnove neke vrste riječi koja nije glagol i koja služi kao osnova za tvorbu glagola. Prvi slučaj nije problematičan jer se korijen lako može povezivati s drugim riječima različitih vrsta riječi kojima je zajednički. U drugome je slučaju detektiranje tvorbenih porodica otežano jer se ne radi o korijenu zajedničkome cijeloj tvorbenoj porodici. Na primjer, tvorbena osnova od koje se izvodi glagol može biti imenica, poput *ribar* – *ribariti* pa je sufiks *-ar* u tome slučaju sufiks za tvorbu imenice, a ne glagola. Kako bi se olakšalo buduće stvaranje tvorbenih porodica među riječima različitih vrsta riječi, u drugome su koraku morfološke raščlambe glagola izdvojeni korijeni i svi su glagoli koji dijele isti korijen povezani na njega. U slučaju homografnih korijena, kao što je naznačeno u poglavlju 3, uvedene su brojčane oznake za njihovo razlikovanje (npr. *skupiti* i *kupovati* spojeni su na korijene *kup1* i *kup2*).

4.2. Prefiksacija

Pri odvajanju prefikasa od osnova koristio se skup od 24 prefiksa iz Babićeve *Tvorbe riječi* (2002: 536–557) i svi njihovi alomorfi koji su povezani na temeljni oblik. Ti su prefiksi: *do-*, *bez-*, *iz-*, *mimo-*, *na-*, *nad-*, *naj-*, *o-/ob-*,¹⁸ *od-*, *pa-*, *po-*, *pod-*, *pre-*, *pred-*, *pri-*, *pro-*, *protu-*, *raz-*, *s-*, *su-*, *suprot-*, *u-*, *uz-*, *za-*. Njima je dodan i prefiks *ne-*, koji Babić ne navodi, a može se pojavljivati na različitim prefiksanim pozicijama unutar prefiksanih kombinacija: *nestati*, *ne-*

18 Pitanje jesu li prefiksi *o-* i *ob-* alomorfi ili zasebni morfemi nije jednoznačno riješeno, posebice jer postoje valjani argumenti za oba motrišta (usp. npr. Belaj 2008: 31 ili Babić 2002: 543). Ovdje, u skladu sa Srebačić (2011), *o-* i *ob-* promatramo kao alomorfe.

mati, onemogućiti, onespokojiti. Iako Babić (ibid.) navodi prefiks *obez-* kao cjelinu, u CroDeriV-u je taj prefiks rastavljen na *o-* + *-bez-*. Prefiksi u glagolima stranoga podrijetla (*de-*, *dis-*, *re-* itd.) nisu odvajani jer u hrvatski sustav ulaze kao cjelina s glagolima i u pravilu ne sudjeluju u tvorbi glagola slavenskoga podrijetla. Kako bi se omogućio uvid u ovjerene kombinacije prefikasa pri tvorbi hrvatskih glagola, svaki je prefiks odvajani kod osnova na koje je dodano više prefikasa. U CroDeriV-u je zabilježeno gotovo 10.000 glagola u čijoj tvorbi sudjeluje samo jedan prefiks, dok je glagola s dvama prefiksima petstotinjak. Svega trinaest glagola tvoreno je s pomoću pet potvrđenih tročlanih kombinacija prefikasa (*iz-po-raz-*, *po-iz-pre-*, *ob-raz-po-*, *pre-raz-po-*, *ob-ne-s-*). Glagoli s četirima prefiksima potvrđeni su u dvama slučajevima (kombinacije: *ob-ne-raz-po-*, *pre-po-iz-od-*¹⁹). Glagoli s pet prefikasa u CroDeriV-u nisu potvrđeni. Detaljniji podatci o kombinacijama prefikasa navode se u poglavlju 5.1. Prikaz prefiksalske strukture na temelju analize veće skupine glagola omogućuje uvid u potvrđene kombinacije prefikasa,²⁰ ali i empirijske podatke o plodnosti pojedinih prefikasa te mogućnosti njihova spajanja s pojedinim korijenima. Drugim riječima, CroDeriV omogućuje detekciju svih glagola sa zajedničkim korijenom i svih prefikasa koje određene glagolske osnove mogu primiti. Vezane osnove,²¹ odnosno osnove koje se dobivaju nakon odbacivanja prefikasa i služe u tvorbi najmanje dvaju glagola, a ne mogu stajati kao samostalni glagoli, posebno su izdvojene i označene kao nesamostalne kombinacije.

4.3. Sufiksacija

Nakon analize i odvajanja prefikasa, sličan je postupak primijenjen na glagolske sufikse. Iako su se prvotne postavke za sufiksalsnu analizu temeljile na morfološkim razdiobama koje su za hrvatski najbolje razradili Silić (2002, 2005), Zrinka Babić (1991) i Marković (2012), nakon segmentacije sufikasa i strukturiranja podataka u CroDeriV-u postalo je jasno da te postavke treba modificirati. Postojeće razdiobe kreću od pretpostavke o minimalnoj tromorfskoj strukturi hrvatskih glagola, odnosno o minimalnoj dvomorfskoj strukturi glagolskih osnova (Silić 2002: 17, Marković 2012: 371). No, pritom se ne uzi-

19 *Oneraspoložiti* i *prepoizodnositi*. Drugi je glagol, kao što je poznato, iz Slamnigove pjesme, no govornik bi se vrlo lako mogao prisjetiti i drugih glagola koji dodatno svjedoče o mogućnosti dodavanja do četiriju prefikasa na hrvatske glagolske osnove, npr. *pre-is-po-na-pijati* ili *na-is-pri-povijedati se*, koji navodi i Marković (2012: 132).

20 Moguće je da će se s unosom novih lema u CroDeriV broj potvrđenih kombinacija povećavati.

21 Babić (2002: 37). Iako je termin nespretn jer su gotovo sve osnove promjenjivih riječi u hrvatskome jeziku vezane, ovdje se misli na glagolske osnove koje ne mogu stajati samostalno, a da im se ne doda prefiks. Npr. **vijestiti* kao osnova glagola *navijestiti*, *izvijestiti* ili **četi* kao osnova glagola *početi*, *začeti*, *načeti* ne postoje kao samostalni glagoli. Marković (2012: 45) za korijene tipa **če* kaže da ih »kadšto smatramo jedincatima«. »Jedincati ili unikatni morf, ili unifiks (engl. *unique morph*), onaj je koji se pojavljuje samo u jednoj riječi ili u jednoj kolokaciji.« No, značenja korijena *-če-*, kako kaže Marković (2012: 45), »domišljamo tek iz izvedenica«, pa ga je zbog toga opravdano smatrati jedincatim morfom. Značenje, pak, korijena *-vijest-* jasno je i iz samoga korijena jer se radi o imeničkoj tvorbenoj osnovi za tvorbu glagola. Stoga, uz navedenu ogradu, smatramo termin *vezana osnova* ipak prikladnijim od *jedincatoga morfa*.

maju u obzir sufiksi sa specifičnim, ponajprije deminutivnim značenjem, koji tu strukturu povećavaju za jedno sufiksno mjesto (npr. *pjev-uš-i-ti*). Oblična bi se osnova u tom slučaju trebala prikazivati kao sastavljena od korijenskoga morfa i dvaju sufikasa, no opći konsenzus o tromorfskoj strukturi osnova ne postoji. Naime, osim što se može naići na raščlambe poput *rad-i-ti* ili *kop-a-ti* (Silić i Pranjković 2005: 45), odnosno na raščlambe koje pretpostavljaju dvomorfsku strukturu osnove, postoje i raščlambe poput *bič-ev-a-ti* ili *smanj-iv-a-ti* (Silić i Pranjković 2005: 45), odnosno *klju-v-a-ti* (Barić i dr. 2003: 235²²), koje pretpostavljaju tromorfsku strukturu osnove, a kad bi se tomu dodao i deminutivni sufiks, došlo bi se do četveromorfske strukture (npr. *jad-ik-ov-a-ti*, *za-muc-k-iv-a-ti*). Iako Marković (2012: 366) ističe da bi opis morfološke strukture glagola trebao biti dovoljno apstraktan, ali istovremeno i sveobuhvatan kako bi omogućio »promatranje svih hrvatskih glagolskih vrsta na jednak način (korijen + tematski sufiks + fleksijski sufiks)«, položaj onoga što se naziva tematskim i fleksijskim sufiksom nije precizno definiran u sufiksnoj strukturi glagola. Tako se na istome mjestu, odmah uz korijen, navode sufiksi koji se koriste za promjenu glagolskoga vida, ali i derivacijski sufiksi kojima primarna funkcija nije vidska ili uopće ne utječu na njegovu promjenu, npr. *kup-ov-a-ti* vs. *pjev-uš-i-ti* (v. Silić i Pranjković 2005: 45, 151²³). Analizom sufiksne strukture infinitivnih osnova²⁴ odlučeno je da se za sve glagole u CroDeriV-u ona modelira na temelju glagola s najkompliciranijom strukturom, odnosno glagola s najviše zabilježenih sufikasa. Na temelju analize zaključeno je da je potrebno predvidjeti tri mjesta za sufiksne morfove, odnosno morfeme²⁵ kako bi se sustavno zabilježila i prikazala struktura glagola u takvu računalnom leksikonu. Počevši od korijena, prvo mjesto do njega predviđeno je za derivacijske sufikse koji uglavnom izražavaju deminutivnost i rjede pejorativnost. Sljedeća su dva mjesta predviđena za sufikse koje Marković (2012: 366–367) naziva tematskima i fleksijskima. U CroDeriV-u je drugo mjesto od korijena predviđeno za sufiks kojim se izražava komponenta trajanja radnje (svršenost ili nesvršenost). Treće mjesto zauzima sufiks koji određuje glagolsku vrstu i koji se kadšto naziva i tematskim samoglasnikom ili sufiksom.²⁶ Morfemska struktura glagola s desne strane korijena u CroDeriV-u izgleda ovako:

- 22 Jedan od rijetkih primjera morfološke raščlambe u toj gramatici, dok se u Babićevoj *Tvorbi riječi* (2002) glagoli ne rastavljaju na morfove. Zbog toga smo se u početnoj razradi teorijskih postavki najčešće oslanjali na Silića i Pranjkovića (2005) i Markovića (2012).
- 23 Glagoli se u poglavlju o deminutivnim sufiksima ne rastavljaju na morfove, nego se *ušiti* navodi kao *-uš(i)ti*, iz čega je jasno da su dva sufiksna mjesta popunjena na način na koji je ovdje prikazano. Barić i dr. (2003: 375), primjerice, u deminutivnih sufikasa ne razdvajaju ni tematski sufiks (*-uš(i)ti*). Marković (2009: 221) ima *greb-uc-k-a-ti*. Time se još jednom potvrđuje neusustavljenost morfološkoga opisa u hrvatskih gramatika.
- 24 U CroDeriV-u se zasad barata samo infinitivnim osnovama, koje se dobivaju odbacivanjem infinitivnoga sufiksa *-ti* ili njegova alomorfa *-ći* (usp. Babić, Z. 1991: 12). Prezentskim se ili drugim oblicima dosad nije bavilo.
- 25 Morfem 'deminutivnost/pejorativnost', morfem 'glagolski vid', morfem 'glagolska vrsta'.
- 26 »Izraz osnove plodnih glagola jasno je označen završetkom – jednim od četiri samoglasnika, tzv. tematskim samoglasnikom, koji ga upravo obilježava kao glagol.« (Babić 1991: 19). Marković (2012: 367) taj samoglasnik naziva tematskim sufiksom.

tvorbena osnova + sufiks₁ + sufiks₂ + sufiks₃ + infinitivni sufiks

Primjeri sufiksalne strukture glagola na površinskoj razini u CroDeriV-u prikazani su u tablici 1.

	PREFIKS	TVORBENA OSNOVA	SUFIKS 1	SUFIKS 2	SUFIKS 3	INFINITIVNI SUFIKS
GLEDATI		gled	∅	∅	a	ti
	po	gled	∅	av	a	ti
		gled	uck	∅	a	ti
PRATI		pra	∅	∅	∅	ti
	pre	pra	∅	∅	∅	ti
		per	uck	∅	a	ti
VIDJETI		vid	∅	∅	je	ti
	pred	vid	∅	∅	a	ti
DREMUCKATI		drem	uck	∅	a	ti
		drem	uc	n	u	ti
MICATI		mic	∅	∅	a	ti
		mak	∅	n	u	ti
KRIVUDATI		kriv	ud	∅	a	ti

Tablica 1. Primjeri morfske raščlambe glagola u CroDeriV-u

Razlozi za takvu podjelu obrazlažu se u poglavljima 4.3.1., 4.3.2. i 4.3.3.

4.3.1. Sufiks₁

Kao što je navedeno, utor sufiks₁ zauzimaju sufiksi koji se obično opisuju kao sufiksi za tvorbu deminutivnih, odnosno pejorativnih glagola (npr. *-k-*, *-ak-*, *-uc-*, *-uck-*, *-uš-*, *-ušk-* itd.).²⁷ Ti sufiksi dolaze i u glagola nesvršenoga i u glagola svršenoga vida (npr. *u-šut-k-∅-a-ti* – *u-šut-k-av-a-ti*). U literaturi se katkad nailazi na tumačenja o dvomorfskome sastavu takvih sufikasa. Primjerice, Marković (2009: 221) glagol *grebuckati* rastavlja kao *greb-uc-k-a-ti*. Takva interpretacija mogla bi biti potaknuta oprekom glagolskovidskih parnjaka poput *greb-uck-∅-a-ti* – *greb-uc-n-u-ti*.²⁸ U tim se primjerima odsječak *uc*, s obzirom na to da je zajednički nepromjenjiv dio, nameće kao samostalna jedinica u analizi. To bi pak značilo da se *-uck-* i *-ušk-* trebaju rastavljati na dva morfa: *-uc-* + *-k-*, odnosno *-uš-* + *-k-* (dakle *greb-uc-k-∅-a-ti*, a prema tome onda za naše potrebe i *greb-uc-∅-n-u-ti*). U konačnici bi analiza prema takvim načelima zahtijevala uvođenje peteromorfske strukture infinitivne osnove u CroDeriV. Budući da je značenjskom analizom glagola koji sadržavaju sufikse poput *-uc-* ili *-uck-* utvrđeno da nema razlike u značenju između glagola

27 Smatramo da je kod glagola primjerenije govoriti o smanjenu intenzitetu radnje osnovnoga glagola, a ne o umanjenoj ili deminutivnosti te radnje (usp. Šojat, Srebačić, Tadić 2012).

28 Razlozi za razdvajanje morfema *-nu-* navode se u poglavlju 4.3.2.

koji sadržavaju jedan ili drugi²⁹ te da potencijalni sufiks *-k-* ne nosi nikakvu dodatnu značenjsku razliku, odlučeno je da se navedeni primjeri promatraju kao različiti, odnosno zasebni sufiksi. Navedeno se može ilustrirati i sufiksima *-uš-* vs. *-uš-* u primjerima poput *ljulj-ušk-Ø-a-ti* vs. *pjen-uš-Ø-a-ti*. I u tim primjerima ti sufiksi nose samo dodatno značenje smanjena intenziteta glagolske radnje.³⁰ Uvođenje dodatnoga utora previše bi zakompliciralo opis, posebice kada se ima na umu izrada računalnoga resursa i računalna obradba u njemu zapisanih podataka. Osim sufikasa s deminutivnim i pejorativnim značenjima, u utoru sufiks₁ nalaze se i sufiksi koji služe u tvorbi glagola od ostalih vrsta riječi, npr. *-ud-* (*krivudati*), *-č-* (*kmečati*), *-inj-* (*glavinjati*) i sufiksi koji imaju specifična značenja poput *-ik-* (*voz-ik-Ø-a-ti*). Broj glagola kod kojih u tome utoru nije *-Ø-*, kao i distribucija pojedinih sufikasa navode se na dijagramu 5. u poglavlju 5.2.

4.3.2. Sufiks₂

Utor sufiks₂ zauzimaju sufiksi koji nose podatak o glagolskome vidu.³¹ Ti su sufiksi: *-Ø-*, *-n-*, *-av-*, *-iv-*, *-ov-*, *-ev-*, *-ijev-*, *-ir-*, *-ficir-*, *-ificir-*, *-izir-*. Osim *-Ø-*, navedeni sufiksi odgovaraju sufiksima *-nu-*, *-ava-*, *-iva-*, *-ova-*, *-eva-*, *-ijeva-*, *-ira-*, *-ficira-*, *-ificira-*, *-izira-* koji se u takvim oblicima navode u literaturi (npr. Babić 2002: 518ff, Barić i dr. 2003: 372ff). U skladu sa Silićem i Pranjkovićem (2005: 45) i Markovićem (2012: 367), smatramo da je riječ o dvomorfskim³² strukturama, odnosno da sufikse poput *-ava-* ili *-ificira-* treba promatrati kao sastavljene od dvaju morfova: *-av-+-a-*, odnosno *-ificir-+-a-*. Silić i Pranjković (2005: 45) sufikse *-ava-*, *-iva-*, *-ova-*, *-eva-* prikazuju kao dvomorfske i navode da se »jedino tako mogu uspostaviti odnosi *-ov- // -ev- / -iv- : -u-*«. Takvim iskazivanjem međusobnoga odnosa navedenih sufikasa u infinitivnoj i prezentskoj osnovi uspostavljena je jasna ekvivalencija prema njihovu značenju i funkciji. Međutim, takav odnos nije uspostavljen za glagole druge vrste jer se infinitivni sufiks *-nu-* u pravilu promatra kao jednomorfski. Iako Silić i Pranjković (2005: 44) ne odvajaju *-nu-* u infinitivu na morfove *-n-* i *-u-*, odnosno tumače ga kao jednomorfsku strukturu, daljnji prikaz i diskusija o prezentskoj osnovi ostavljaju mogućnost drugačijih interpretacija. Silić i Pranjković (ibid.) navode:

29 Smatramo da glagoli *prduckati* i *prducati* ili *kašljuckati* i *kašljucati* znače isto i da govornik ima slobodan izbor između sufikasa *-uck-* i *-uc-*.

30 Ispadanje suglasnika *k* u supostavljenim primjerima poput *greb-uck-Ø-a-ti* i *greb-uc-n-u-ti* ili *drem-uck-Ø-a-ti* i *drem-uc-n-u-ti* može se također tumačiti fonološki uvjetovanim ispadanjem unutar suglasničkoga skupa *ckn*. Naime, medijalni slijed *ckn* nije potvrđen u hrvatskome jeziku (Turk 1992). Osim toga, ispadanje suglasnika *k* ispred sufiksa *-nu-* može biti ostatak staroga pravila prema kojemu »su ispadali suglasnici *p, t, k, b, d, g* ispred *n* (u sufiksu *-nu-*)« (Turk 1992: 152).

31 Pod pojmom glagolskoga vida ovdje se misli samo na svršenost i nesvršenost glagolske radnje. Prema kategoriji glagolskoga vida kako je ovdje shvaćena, hrvatski glagoli mogu biti svršeni, nesvršeni i dvovidni.

32 I dvomorfskim jer odgovaraju morfemima 'glagolski vid' i 'glagolska vrsta' u dubinskoj strukturi.

»Infinitivna osnova druge vrste završava sufikslnim morfemom *nu*, a prezentska sufikslnim morfemom *ē*. Između prezentskoga sufikslnog morfema *ē* i korijenskoga morfema intervenira umetak *n*. Druga vrsta ima jedan razred: *-nu-ti* / *n-ē-*: *mřz-nu-ti* / *mřz-n-ē-m.*«

Iz toga proizlazi da razlika između infinitivne i prezentske osnove nije samo u sastavu morfova, nego i u njihovu broju, budući da ostaje nejasno je li navedeni »umetak *n*« zaseban morf (ali i zaseban morfem, što će se pokazati iz daljnje rasprave). Dalje u tekstu, u tablici sa svim glagolskim vrstama, pa tako i s drugom glagolskom vrstom, Silić i Pranjković (2005: 47) prezentski sufikslni morf *ē* i spomenuti »umetak *n*« prikazuju kao jedan morf u prezentu glagola koji supostavljaju infinitivnomu *-nu-*. Na taj su pak način dobiveni jednomorfski sufiksi i u infinitivu i u prezentu pa se može postaviti pitanje djeluje li umetak *n* i u infinitivu, ili pak, gledano s druge strane, postoji li u infinitivnoj osnovi morf *nu*, a u prezentskoj morf *ē* ispred kojega se umeće *n*. S obzirom na to da je ono što Silić i Pranjković (2005: 44) nazivaju umetkom nepromjenjiv dio i u infinitivnoj i u prezentskoj osnovi, smatramo da je prihvatljivije promatrati infinitivno *nu* kao sastavljeno od dvaju morfova, a istovremeno i od dvaju morfema. Prvi je od njih *-n-* koji daje podatak o glagolskome vidu i načinu vršenja glagolske radnje, a drugi je *-u-* koji daje podatak o konjugacijskome tipu, odnosno glagolskoj vrsti, analogno sufiksima *-a-* u *čitati* ili *-i-* u *misliti* (dakle: *met-Ø-n-u-ti* ili *drem-uc-n-u-ti* prema *čit-Ø-Ø-a-ti* i *misl-Ø-Ø-i-ti*).

Takvom segmentacijom utor sufiks₂ u CroDeriV-u određen je za sufikse koji nose značenje glagolskoga vida. Sufiksi *-av-*, *-iv-*, *-ov-*, *-ev-*, *-ijev-* služe za tvorbu nesvršenih glagola (npr. *kup-Ø-Ø-i-ti* – *kup-Ø-ov-a-ti*), a sufiksi *-ir-*, *-ficir-*, *-ificir-*, *-izir-* u pravilu za tvorbu dvovidnih glagola stranoga podrijetla.

Sufiksom *-n-* u utoru sufiks₂ tvoreno je ukupno 775 glagola koji se trenutno nalaze u CroDeriV-u. Od toga je tek 15 nesvršenih glagola. To su glagoli: *brinuti*, *čeznuti*, *čvrsnuti*, *gasnuti*, *ginuti*, *greznuti*, *mrknuti*, *mrznuti*, *plahnuti*, *sahnuti*, *tonuti*, *trnuti*, *trnuti*, *tvrdnuti*, *venuti*, što čini tek 1,9% od ukupnoga broja glagola sa sufiksom *-n-*. Zbog toga se čini opravdanim tvrditi da sufiks *-n-* ima primarno značenje svršenosti glagolskoga vida, uz iznimku tih 15 zabilježenih glagola. Osim svršenosti, *-n-* nerijetko ima i značenjsku komponentu jednokratnosti i umanjenosti, koje mogu postati i jedina njegova značenja. Također je zabilježeno 20 glagola koji završavaju na *-njivati*, a kojima *-n-* nije dijelom osnove, nego sufiksa. Svi su takvi glagoli nesvršeni, a njihov se vid kodira sufiksom *-jiv-*.³³ Sufiks *-n-* u tim glagolima ne sadržava podatak o glagolskome vidu, nego samo podatak o višestrukome ponavljanju kratke radnje. Tim svojim specifičnim značenjem odgovara općemu sadržaju sufiksa₁, i u tim slučajevima smješten je u taj utor. Takvi su nesvršeni glagoli koji označavaju višestruko ponavljanje radnje, npr. *zabljesnjivati*, *nadahnjivati*, *otvrdnjivati*.³⁴

³³ Alomorf sufiksa *-iv-*.

³⁴ Činjenica da se u tih glagola sufiks *-n-* ne nalazi u okolini sufiksa *-u-* dodatno opravdava njihovo razdvajanje na dva morfa. Glagola s *-njiva-* vrlo je malo (tek 20) i nerijetko su nerazumljivi izvornomu govorniku hrvatskoga jezika (npr. *pričveknjivati*, *oseknjivati*).

Kao što je rečeno, sufiks₂ daje podatak o glagolskome vidu. Pritom je najslabiji slučaj sufiks –Ø– jer može imati i značenje svršenosti, nesvršenosti i dvovidnosti, iako je većina glagola tvorena njime nesvršenoga vida. Iza sufiksa –Ø–, u utoru sufiks₃ mogu doći sufiksi –a–, –i– i –Ø–.

Glagola kod kojih iza sufiksa –Ø– u utoru sufiks₂ slijedi –Ø– u utoru sufiks₃, odnosno sa sufiksalsnom strukturom –Ø–Ø–Ø–ti, vrlo je malo i čine zatvorenu skupinu od 81 glagola, od kojih se mnogi upotrebljavaju vrlo rijetko. Ta skupina obuhvaća 63 nesvršena glagola, 15 svršenih i 3 dvovidna glagola. Svih 15 svršenih glagola sa sufiksom –Ø– u utorima sufiks₂ i sufiks₃ ima prezentski oblik na –ne– (npr. *dići – dignem, sresti – sretnem*, osim *dati*, koji ima *dam* uz rjeđe *dadnem*). Mnogi od njih infinitivni su oblik prilagodili prema skupini glagola sa sufiksom –n– na mjestu sufiksa₂, o čemu svjedoče dvostruki infinitivi, npr. *dići – dignuti* ili *taći – taknuti*. Ti primjeri svjedoče o tendenciji uklapanja glagola iz te zatvorene vrste u paradigme otvorenih i brojnijih vrsta, kao i o tome da se vrsta očisti od svršenih glagola koji već supostoje u dvama oblicima.

Sufiks –Ø– u utoru sufiks₂ gotovo uvijek znači nesvršenost, osim u vrlo malome broju iznimaka. Glagoli kojima iza sufiksa –Ø– slijedi –a– svi su nesvršenoga vida, uz desetak dvovidnih iznimaka (npr. *ručati, večerati, čestitati, probati*). Glagoli kojima iza sufiksa –Ø– slijedi –i– mogu biti i svršeni i nesvršeni, no i tu je svršenih glagola tek oko 6%. Velik broj takvih svršenih glagola ima svoj nesvršeni parnjak tvoren kombinacijom sufikasa –Ø–a– (*bacati – baciti; vraćati – vratiti; udarati – udariti; lupati – lupiti*) i svi su praslavenškoga podrijetla.³⁵

Broj glagola koji opreku u vidu iskazuju alternacijom sufikasa –a– i –i– u utoru sufiks₃ iza –Ø– u utoru sufiks₂ vrlo je malen (u CroDeriV-u ukupno 46 neprefiguriranih glagola, odnosno 23 vidska parnjaka). Naime, način izražavanja vidske opreke sufiksima –a– i –i– u utoru sufiks₃ danas više nije produktivan i može se pretpostaviti da je takav način izražavanja vidske opreke ostatak starijega stanja koji se zadržao u toj zatvorenoj skupini.³⁶ Na temelju tih primjera također se može pretpostaviti da je u starijim stanjima jezika osnovna glagolska struktura bila dvomorfska, odnosno da se podatak o vidu nije iskazivao posebnim sufiksalsnim morfološkim oblikom kao što su –av–, –iv– ili –ov–, nego je

35 Može se postaviti pitanje o opravdanosti uvođenja sufiksa –Ø– na utoru sufiks₂, s obzirom na to da nije jednoznačan, odnosno glagoli sa sufiksom –Ø– u tome utoru mogu biti svih vidova. Njegova je opravdanost posebno upitna u sustavu u kojemu postoje sufiksi koji jednoznačno označavaju samo jedan glagolski vid, poput –iv–, –av– i sl. No, kod takvih glagola (npr. *iš-čit-Ø-av-a-ti*) sufiks –a– u utoru sufiks₃ označava samo glagolsku vrstu, pa smatramo da ima isto značenje i u ostalih glagola. U tim glagolima glagolski vid označava sufiks –Ø–, koji u neprefiguriranih glagola ima osnovno značenje nesvršenosti, a u prefiguriranih svršenosti i stoji na istome mjestu u strukturi kao i spomenuti sufiksi –iv–, –av– i sl. Isto tako, iako bi se prema strukturi amalgamiranih značenja sufikasa moglo tvrditi da sufiksi –a– i –i– koji dolaze u utoru sufiks₃ nose podatak o glagolskome vidu, oni to također ne čine jednoznačno (npr. *misliti* vs. *kupiti*).

36 Ovo su zabilježeni glagoli koji opreku u vidu tvore na opisani način: *baciti – bacati, đipiti – đipati, hititi – hitati, javiti – javljati, lupiti – lupati, obratiti – obraćati, opojiti – opajati, oporaviti se – oporavljati se, platiti – plaćati, primiti – primati, pružiti – pružati, pustiti – puštati, razoriti – razarati, roditi – rađati, sjetiti se – sjećati se, snimiti – snimati, skočiti – skakati, staviti – stavljati, stupiti – stupati, turiti – turati, udariti – udarati, vratiti – vraćati, zabaviti – zabavljati.*

bio amalgamiran u morfu koji u CroDeriV-u zauzima utor sufixs₃, odnosno u onome sufixu koji daje podatak o vrsti glagolske konjugacije. Zbog činjenice da se radi o vrlo ograničenoj skupini glagola u kojoj je podatak o vidu i glagolskoj vrsti amalgamiran u jednome morfu, radi dosljednosti unutar sustava odlučeno je da će se i njihova sufixalna struktura prikazivati kao tromorfška ($-\emptyset-\emptyset-a/i-ti$).

4.3.3. Sufiks₃

U utoru sufixs₃ mogu se naći sufixi $-u-$, $-a-$, $-i-$, $-je-$ i $-\emptyset-$, koje Zrinka Babić (1991:19ff) naziva tematskim samoglasnicima, odnosno oni sufixi koji glagol obilježavaju »kao glagol«. No, ako oni obilježavaju glagol kao glagol, drugim riječima, ako je njihov sadržaj 'glagol' i govore o glagolskoj vrsti, preciznije bi bilo govoriti o sufixsima, a ne o tematskim samoglasnicima.³⁷ Alternacije morfova u utoru sufixs₃ između infinitivne i prezentske osnove prikazane su u tablici 2.³⁸ Glagoli sa sufixsom $-\emptyset-$ u utoru sufixs₃ nemaju pravilnu podjelu pa prezentski morf ovisi o pojedinačnome glagolu.

Infinitivna osnova		Prezentska osnova	
$-u-$	→	$-e-$	<i>puk-\emptyset-n-u-ti</i> → <i>puk-\emptyset-n-e-m</i>
$-a-$	→	$-a-$ $-e-$ $-i-$	<i>čit-\emptyset-\emptyset-a-ti</i> → <i>čit-\emptyset-\emptyset-a-m</i> <i>pis-\emptyset-\emptyset-a-ti</i> → <i>piš-\emptyset-\emptyset-e-m</i> <i>muč-\emptyset-\emptyset-a-ti</i> → <i>muč-\emptyset-\emptyset-i-m</i>
$-i-$	→	$-i-$	<i>misl-\emptyset-\emptyset-i-ti</i> → <i>misl-\emptyset-\emptyset-i-m</i>
$-je-$	→	$-i-$	<i>vol-\emptyset-\emptyset-je-ti</i> → <i>vol-\emptyset-\emptyset-i-m</i>

Tablica 2. Alternacija morfova u infinitivu i prezentu

Na temelju iznesenoga sufixalnu strukturu infinitivne osnove glagola u CroDeriV-u prikazujemo kao tromorfšku i tromorfemsku: utor sufixs₁ zauzimaju sufixi koji imaju specifična značenja poput deminutivnosti ili pejorativnosti, utor sufixs₂ sufixi kojima se kodira glagolski vid, a utor sufixs₃ sufixi koji govore o glagolskoj vrsti. Iako ne moraju sva tri predviđena sufixalna mjesta kod svih glagola biti popunjena nekim drugim sufixsom osim $-\emptyset-$, ona su u sustavu predviđena radi lakšega prikazivanja morfološke strukture svih zabilježenih glagola. Zaključno, cjelokupna predviđena struktura glagola u CroDeriV-u izgleda ovako:

**prefiks₄ + prefiks₃ + prefiks₂ + prefiks₁ +
(tvorbena osnova₂ + interfiks) + tvorbena osnova₁ +
sufiks₁ + sufixs₂ + sufixs₃ + infinitivni sufixs**

Tvorbena osnova₂ uvedena je zbog glagolskih složenica poput *zlopamtiti* (*zl-o-pamt- \emptyset - \emptyset -i-ti*), no njihov je broj neznatan³⁹ pa se ovdje nismo detaljno bavili njihovom strukturom.

³⁷ V. i bilješku 26.

³⁸ U CroDeriV-u se nalaze samo leme u infinitivu, prezentski oblici ili oblici drugih glagolskih vremena i načina, barem zasad, nisu uneseni ni obrađeni.

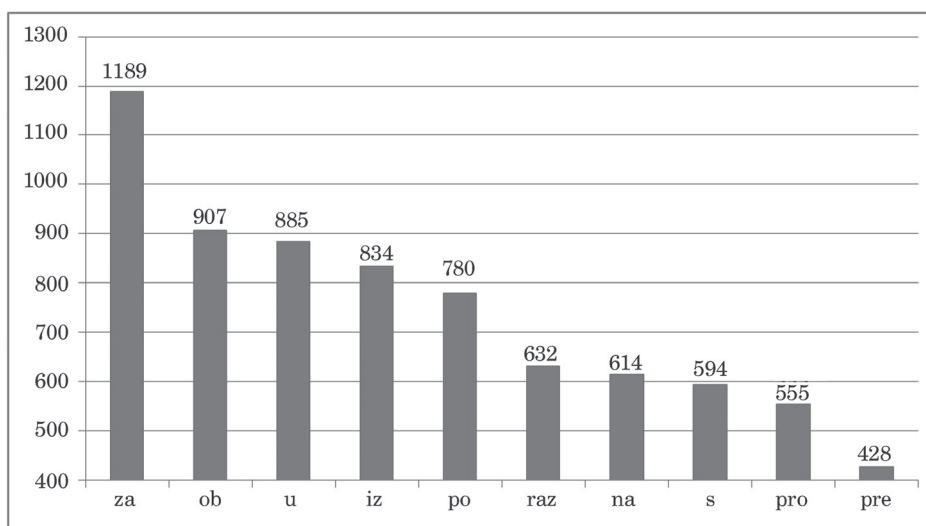
³⁹ 128 od 13.780 glagola, što čini tek 0,92% ukupnoga broja lema.

5. Dobiveni podatci

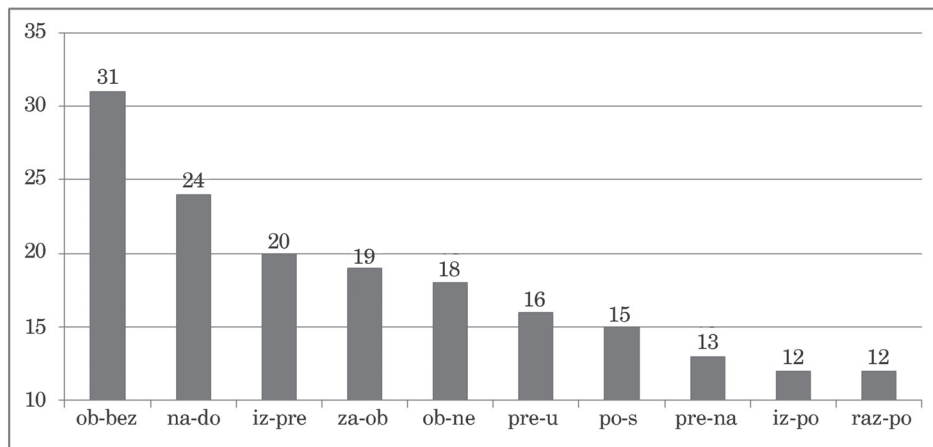
U sadašnjemu se obliku CroDeriV sastoji od 13.780 lema. Od toga je broja 11.746 prefiguriranih glagola i 2034 neprefigurirana glagola. Vezanih tvorbenih osnova, to jest osnova koje nisu samostalne riječi, ali služe u tvorbi najmanje dvaju glagola, ima 1175. Segmentacijom na morfove ustanovljena su 3292 korijena. Kao što smo naveli, prednost koju ovakav leksikon omogućava jest uvid u podatke o frekvenciji i mogućnostima kombiniranja pojedinih afikasa. U ovome poglavlju iznose se podatci o kombinacijama pojedinih prefikasa i sufikasa u tvorbi hrvatskih glagola. Koliko je nama poznato, podatci takve vrste dosad nisu sustavno obrađeni i prikazani u literaturi o tvorbi hrvatskih glagola.

5.1. Kombinacije prefikasa u CroDeriV-u

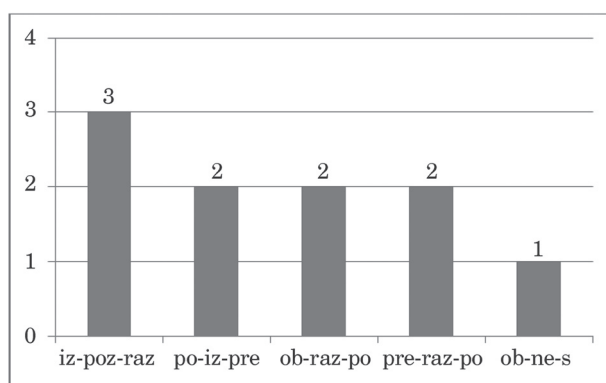
Najčestotniji su prefiksi u kombinacijama tvorbenih osnova i jednoga prefiksa: *za-*, *ob-*, *u-*, *iz-*, *po-*, *raz-*, *na-*, *s-*, *pro-*, *pre-*. Čestotnost tih prefikasa prikazana je na dijagramu 1. Deset najčešćih kombinacija dvaju prefikasa i broj njihovih pojavljivanja u CroDeriV-u prikazani su na dijagramu 2. Sve zabilježene kombinacije triju i četiriju prefikasa i njihova frekvencija prikazane su na dijagramima 3. i 4.



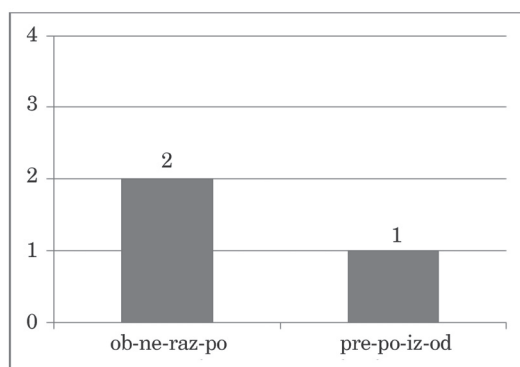
Dijagram 1. Deset najčešćih prefikasa u CroDeriV-u



Dijagram 2. Deset najčešćih kombinacija dvaju prefikasa u CroDeriV-u



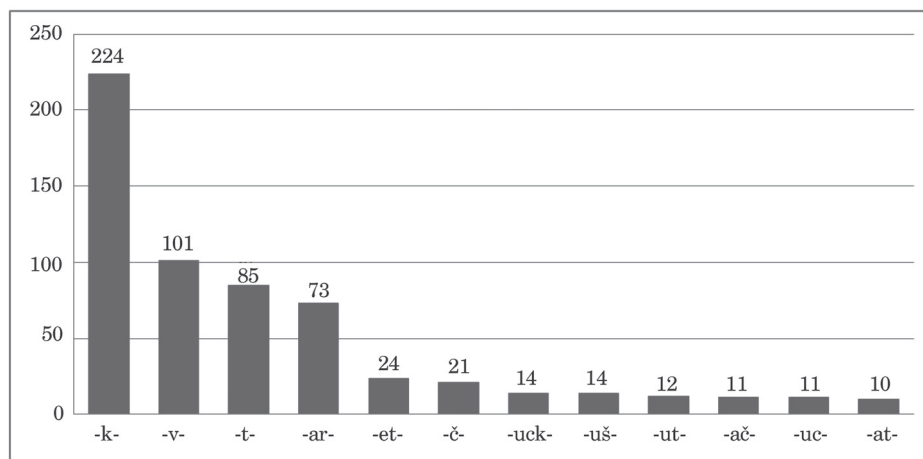
Dijagram 3. Kombinacije triju prefikasa u CroDeriV-u



Dijagram 4. Kombinacije četiriju prefikasa u CroDeriV-u

5.2. Kombinacije sufikasa u CroDeriV-u

Popunjen utor sufiks₁ nekim drugim sufiksom osim $-\emptyset-$ zabilježen je kod 707 glagola. Sufiksi u utoru sufiks₁, a da taj sufiks nije $-\emptyset-$, koje sadržava najmanje deset glagola iz CroDeriV-a i čestotnost njihova pojavljivanja u takvim slučajevima prikazani su na dijagramu 5.



Dijagram 5. Deset najčešćih sufikasa sa specifičnim značenjem u CroDeriV-u

Utor sufiks₂ i sufiks₃ u CroDeriV-u zauzimaju sufiksi koji označavaju dvije kategorije inherentne svim glagolima u hrvatskome – glagolski vid i glagolsku vrstu. U utoru sufiks₂ najbrojniji je sufiks $-\emptyset-$ (9867), a u utoru sufiks₃ sufiks $-a-$ (9376 glagola). Među kombinacijama sufiksa₂ i sufiksa₃ najčešće su kombinacije $-\emptyset-a-$ (5207 glagola), $-\emptyset-i-$ (3614 glagola) i $-av-a-$ (1058 glagola).

6. Zaključak

U radu su prikazani postupci primijenjeni pri izradi leksikona hrvatskih glagola CroDeriV i morfološka struktura hrvatskih glagola dobivena na temelju analiziranih podataka. CroDeriV u sadašnjemu obliku sadržava otprilike 14.000 glagolskih lema rastavljenih na morfove. Nakon uvodnoga dijela u kojemu se argumentira potreba za takvim resursom, iznesene su teorijske postavke na kojima počivaju obrada i prikaz podataka. Za sve je glagole utvrđen zajednički korijen i svi glagoli koji ga dijele preko njega su međusobno povezani. S obzirom na to da su se teorijske postavke preuzete iz postojeće morfološke literature pokazale nesustavnima, posebice u opisu sufiksne strukture hrvatskih glagola, uvodi se morfološka raščlamba kojom je moguće obuhvatiti i konzistentno prikazati sve glagole u CroDeriV-u. Takvom je raščlambom omogućeno jasno definiranje izraza i sadržaja svakoga afiksa u morfološkoj strukturi glagola. Smatramo da predloženi prikaz sufiksne strukture glagola u CroDeriV-u može barem donekle pridonijeti rasvjetljavanju njezina katkad maglovita prikaza u hrvatskome jezikoslovlju.

Plodnost derivacijskih afikasa u literaturi posvećenoj tvorbi hrvatskih glagola u pravilu se opisuje izrazima kao »vrlo je plodan« ili »slabo je plodan« (usp. npr. Babić 2002 ili Barić i dr. 2003), dok empirijski podatci o kombinacijama dvaju ili više afikasa gotovo da i ne postoje. Prikazanom je raščlambom omogućen uvid u kombinacije derivacijskih afikasa u sustavu hrvatskih glagola te jasno iskazivanje njihove zastupljenosti na temelju analize i opisa većega broja primjera. Također, prikazanom će raščlambom nakon unošenja drugih vrsta riječi u CroDeriV biti omogućeno prepoznavanje tvorbenih porodica s pomoću popisa korijena. Ta sljedeća faza izrade CroDeriV-a obuhvatit će imenice, pridjeve i priloge. Naposljetku, jasno utemeljeni kriteriji za morfološku raščlambu čine CroDeriV pogodnim za unapređenje računalnih alata na području obrade prirodnoga jezika i iskoristivim u svim istraživanjima tvorbenih procesa u hrvatskome jeziku.

Literatura

- Anić, Vladimir (2006). *Veliki rječnik hrvatskog jezika*. Zagreb: Novi Liber.
- Babić, Stjepan, Dalibor Brozović, Milan Moguš, Slavko Pavešić, Ivo Škarić, Stjepko Težak (1991). *Povijesni pregled, glasovi i oblici hrvatskoga književnoga jezika*. Zagreb: HAZU : Nakladni zavod Globus.
- Babić, Stjepan (2002). *Tvorba riječi u hrvatskome književnome jeziku*. Zagreb: HAZU : Nakladni zavod Globus.
- Babić, Zrinka (1991). *Generativni opis konjugacijskih oblika*. Zagreb: Hrvatsko filološko društvo.
- Barić, Eugenija, Mijo Lončarić, Dragica Malić, Slavko Pavešić, Mirko Peti, Vesna Zečević, Marija Znika (2003). *Hrvatska gramatika*. Zagreb: Školska knjiga.
- Belaj, Branimir (2008). *Jezik, prostor i konceptualizacija. Shematična značenja hrvatskih glagolskih prefiksa*. Osijek: Filozofski fakultet u Osijeku.
- Čavar, Damir, Ivo-Pavao Jazbec, Siniša Runjaić (2008). Interoperability and Rapid Bootstrapping of Morphological Parsing and Annotation Automata. U: Erjavec, Tomaž, Jerneja Žganec Gros (ur.) *Proceedings of the Sixth Language Technologies Conference, October 16th–17th, 2008: Proceedings of the 11th International Multiconference Information Society – IS 2008, volume C*. Ljubljana: Institut »Jožef Stefan«, 2008. 80–85.
- Čavar, Damir, Ivo-Pavao Jazbec, Tomislav Stojanov (2009). CroMo – Morphological Analysis for Standard Croatian and its Synchronic and Diachronic Dialects and Variants. U: Piskorski, Jakub, Bruce Watson, Anssi Yli-Jyrä (ur.) *Finite-State Methods and Natural Language Processing – Post-proceedings of the 7th International Workshop FSMNLP 2008*. Italija: IOS Press. 183–190.
- Katunar, Daniela, Krešimir Šojat (2011). Morphosemantic fields in the building of the Croatian WordNet: The verbs of movement. U: Brdar, Mario, Marija Omazić, Višnja Pavičić Takač, Tanja Gradečak-Erdeljić, Gabrijela Buljan (ur.) *Space in Time and Language*. Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien: Peter Lang GmbH. 79–89.
- Kuna, Branko (2006). Proučavanje tvorbe riječi u hrvatskom jeziku tijekom 20. stoljeća. U: *Hrvatski jezik u XX. stoljeću*. Zagreb: Matica hrvatska. 339–365.
- Ljubešić, Nikola, Damir Boras, Ozren Kubelka (2007). Retrieving information in Croatian: Building a simple and efficient rule-based stemmer. U: Seljan, Sanja, Hrvoje Stančić (ur.) *INFUTURE2007: Digital Information and Heritage*. Zagreb: Odsjek za informacijske znanosti Filozofskoga fakulteta. 313–320.
- Ljubešić, Nikola, Tomaž Erjavec (2011). hrWaC and slWaC: Compiling Web Corpora for Croatian and Slovene. U: Habernal, Ivan, Vaclav Matousek (ur.) *Text, Speech and Dialogue 2011. Lecture Notes in Computer Science*. Berlin/Heidelberg: Springer. 395–402.

- Maretić, Tomo (1899). *Gramatika i stilistika hrvatskog ili srpskog književnog jezika*. Zagreb: Kugli.
- Marković, Ivan (2009). Tri nehrvatske tvorbe: infiksacija, reduplikacija, fuzija. *Rasprave Instituta za hrvatski jezik i jezikoslovlje*, 35, 217–241.
- Marković, Ivan (2012). *Uvod u jezičnu morfologiju*. Zagreb: Disput.
- Mikelić Preradović, Nives, Damir Boras, Sanja Kišiček (2009). CROVALLEX: Croatian Verb Valence Lexicon. U: *Proceedings of the 31st International Conference on Information Technology Interfaces*, 533–538.
- Pandžić, Ivan (2012). *Oblikovanje korjenovatelja za hrvatski jezik u svrhu pretraživanja informacija*. Diplomski rad. Sveučilište u Zagrebu, Filozofski fakultet, Odsjek za lingvistiku.
- Raffaelli, Ida, Barbara Kerovec (2008). Morphosemantic fields in the analysis of Croatian vocabulary. *Jezikoslovlje* 9.1–2, 141–169.
- Raffaelli, Ida, Marko Tadić, Božo Bekavac, Željko Agić (2008). Building Croatian WordNet. U: *Proceedings of the Fourth Global WordNet Conference*, Szeged, 349–359.
- Raguž, Dragutin (1997). *Praktična hrvatska gramatika*. Zagreb: Medicinska naklada.
- Samardžija, Marko (1995). *Leksikologija s poviješću hrvatskoga jezika: udžbenik za 4. razred gimnazije*. Zagreb: Školska knjiga.
- Silić, Josip (1998). Morfonologija hrvatskoga glagola: Tipovi osnova. *Riječki filološki dani* 2, 241–274.
- Silić, Josip (2002). Ustrojstvo glagolske osnove uvjetovane glagolskovidskim promjenama. U: *Zbornik Zagrebačke slavističke škole*. Zagreb: FF Press. 16–21.
- Silić, Josip, Ivo Pranjeković (2005). *Gramatika hrvatskoga jezika za gimnazije i visoka učilišta*. Zagreb: Školska knjiga.
- Skok, Petar (1971). *Etimologijski rječnik hrvatskoga ili srpskoga jezika*. Deanović, Mirko, Ljudevit Jonke (ur.). Knjiga prva: A–J. Zagreb: JAZU.
- Srebačić, Matea (2011). *Morfosemantički opis glagola promjene u Hrvatskome WordNetu*. Diplomski rad, Sveučilište u Zagrebu, Filozofski fakultet, Odsjek za lingvistiku i Odsjek za kroatistiku.
- Šnajder, Jan (2008). *Morfološka normalizacija tekstova na hrvatskome jeziku za dubinsku analizu i pretraživanje informacija*. Doktorska disertacija. Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva.
- Šojat, Krešimir, Nives Mikelić-Preradović, Marko Tadić (2012). Generation of Verbal Stems in Derivationally Rich Language. U: *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC '12)*, 928–933.
- Šojat, Krešimir, Matea Srebačić, Marko Tadić (2012). Derivational and Semantic Relations of Croatian Verbs. *Journal of Language Modelling*, 0(1), 111–142.
- Tadić, Marko, Sanja Fulgosi (2003). Building the Croatian Morphological Lexicon. U: *Proceedings of the EACL2003 Workshop on Morphological Processing of Slavic Languages* (Budimpešta 2003), ACL, 41–46.
- Tadić, Marko (2006). Croatian Lemmatization Server. U: Koeva, Svetla, Mila Dimitrova–Vulchanova (ur.) *Proceedings of the 5th Formal approaches to South Slavic and Balkan languages Conference (FASSBL2006)*. Sofija: Bugarska akademija znanosti, 140–146.
- Tadić, Marko (2007). Building the Croatian Dependency Treebank: the initial stages. *Suvremena lingvistika*, 63, 85–92.
- Tafra, Branka, Petra Košutar (2009). Rječotvorni modeli u hrvatskome jeziku. *Suvremena lingvistika*, 67, 87–107.
- Turk, Marija (1992). *Fonologija hrvatskoga jezika*. Rijeka – Varaždin: Izdavački centar Rijeka – Tiskara Varaždin.

CroDeriV and the morphological analysis of Croatian verb

The paper discusses the procedures in building of CroDeriV, the computational lexicon containing data on morphological structure of Croatian verbs. CroDeriV is the first morphological resource dealing with derivational phenomena of Croatian. In the first part of the paper, after the motivation for building this kind of lexicon and the brief overview of the existing morphological resources for Croatian, procedures for morphological segmentation of verbs in CroDeriV are presented. Each verb in CroDeriV is segmented into lexical and derivational morphemes. Verbs of the same root are mutually linked. This procedure enables the recognition of derivationally related families of verbs and, at the same time, the detection of full derivational spans of particular base forms. The second part of the paper focuses on the morphological structure of Croatian verbs based on the analysis of almost 14000 verbal lemmas currently included in CroDeriV. The analysis enabled the recognition of generalized morphological structure applicable to all Croatian verbs. It consists of four slots for derivational prefixes and three slots for derivational suffixes on each side of a lexical morpheme, and these slots are provided for every verbal lemma in CroDeriV. This structure is compared with other approaches dealing with morphology of Croatian verbs. The three suffixal slots and their semantics are explained in more detail, since this kind of segmentation has so far not been introduced in Croatian morphological literature. First suffixal slot comprises suffixes with specialized meanings (e.g. diminutive, pejorative), second slot suffixes with aspectual meaning, and third slot suffixes denoting conjugational class. The final part of the paper describes attested combinations of derivational affixes in CroDeriV and indicates the frequency of their occurrence.

Ključne riječi: CroDeriV, morfologija glagola, računalni leksikon, hrvatski jezik

Key words: CroDeriV, morphology of verbs, computational lexicon, Croatian language