

Tomislava Bošnjak Botica¹, Dario Lečić², Gordana Hržica³, Jurica Budja⁴

^{1,4} Institut za hrvatski jezik i jezikoslovlje, ² Hrvatska zaklada za znanost, ³ Edukacijsko-rehabilitacijski fakultet Sveučilišta u Zagrebu

¹ tbošnjak@ihj.hr, ² dario.lecic@gmail.com, ³ gordana.hrzica@erf.hr, ⁴ jbudja@ihj.hr

Baza hrvatskih morfoloških dubleta (*DvojBa*)¹

Morfološka je dubletnost pojava dvaju (ili više) oblika jedne riječi istoga gramatičkoga značenja (npr. *šetam/šećem, putovi/putevi*). Potaknuti nedostatkom jezičnoga resursa u kojem bi se lako mogli pronaći primjeri pojedinih dubletnih oblika u hrvatskome, autori ovoga teksta pokrenuli su projekt izrade podatkovne baze upravo s tim ciljem – načiniti popis gramatičkih kategorija u kojima se pojavljuje morfološka dubletnost i popis leksema koji se pojavljuju s dubletnim oblicima zajedno s pripadajućim pojavnicama pojedinih oblika pronađenih u hrvatskim jezičnim vrelima (računalnim korpusima, leksikografskim izvorima, književnosti i drugoj relevantnoj građi). Primarni je cilj stvaranje lako dostupna i pretraživa resursa koji bi ubrzao istraživanja morfološke dvostrukosti u hrvatskome i koji bi bio otvoren hrvatskoj znanstvenoj zajednici. Svakoj pojavnici bit će pridruženi podatci o djelu iz kojega potječe, godini nastanka toga djela, autoru djela, njegovu zemljopisnom arealu, vrsti teksta te odgovarajuće morfološke oznake. Na temelju tih podataka moći će se pratiti pojava od trenutka njezina ulaska u jezik kroz sve promjene do suvremenoga statusa, utvrditi distribucija pojedinih oblika u suvremenom jeziku te konačno odrediti uloga morfoloških dubleta unutar jezičnoga sustava. Baza može omogućiti preciznije buduće opise jezika.

1. Uvod

Kad govorimo o strukturi riječi, u lingvistici dvadesetoga stoljeća (u svim vodećim pristupima) dugo je bilo rašireno mišljenje da jedan oblik odgovara samo jednom značenju, odnosno gramatičkoj funkciji (i obratno). Aronoff (1976) tu pojavu naziva načelom morfološkoga blokiranja (engl. *morphological blocking principle*), odnosno tvrdi da postojanje jednoga oblika automatski sprječava (tj. blokira) pojavu drugoga oblika istoga (gramatičkog) značenja². Na primjer, u hrvatskom jeziku

1 Projekt izrade baze provodi se u Institutu za hrvatski jezik i jezikoslovlje. Ovaj je rad sufinancirala i Hrvatska zaklada za znanost u okviru projekta »Višerazinski pristup govornom diskursu u jezičnom razvoju« (UIP-2017-05-6603).

2 Za slična načela vidjeti i Clark 1987 (načelo kontrasta, engl. *principle of contrast*), Giegerich 2001 (učinak blokiranja sinonimije, engl. *synonymy blocking effect*) itd.

supletivni oblik *ljudi* kao nominativ množine leksema čovjek blokira pojavu očekivanoga oblika *čovjeci (Marković 2012) itd. Drugim riječima, sinonimije u gramatici ne bi trebalo biti, za razliku od leksika, gdje je ona prisutna u velikoj mjeri. Ipak, jezično nam iskustvo pokazuje da to ne vrijedi uvijek, odnosno da za isto značenje katkad postoje dva ili više oblika (npr. genitiv množine imenice *naranča*) te se u takvim primjerima opravdano može govoriti o morfološkoj sinonimiji³, za razliku od pojave u kojoj jedan oblik ima više gramatičkih značenja (npr. sinkretizam triju množinskih padeža).

Ta je jezična pojava donedavno bila relativno zanemarena u lingvističkim istraživanjima. Nekoliko je razloga za to. Prvi je razlog to što se nije uklapala u prethodno opisane tradicionalne teorijske okvire prema kojima je odnos između oblika i značenja 1:1, a varijabilnost oblika nepoželjna pojava. Primjerice, Babić (1962: 62) tvrdi da, »ako dva oblika imaju isto značenje i ako se mogu podjednako upotrijebiti u svakom kontekstu, jedan je od njih balast u jeziku«, koji predstavlja (nepotrebno) dvostruko opterećenje za naš mozak. Silić (1998: 175), pak, koristeći se pojmovima iz teorije informacija, tvrdi da »dublete, za razliku od sinonima, otežavaju komunikaciju jer povećavaju entropiju: što ih ima više, predvidljivost je informacije manja«. Svaki primjer koji se ne bi uklapao u takav teorijski okvir pokušao se eliminirati uvođenjem značenjske, funkcionalne, stilske, dijalektološke ili koje druge uvjetovanosti dvaju oblika.⁴ Međutim, budući da ta uvjetovanost nije nužno utemeljena u stvarnom jeziku, nego može biti proizvoljna ocjena pojedinih autora, to je dovelo do situacije da se u priručnicima mogu naći različite preporuke o tome koji je oblik »pravilan«.

Drugi je razlog relativno rijetke prisutnosti ove teme u literaturi taj što u jezicima koji su predmet većine lingvističkih istraživanja i literature morfološka dubletnost zaista jest rijetka pojava te su morfološke dublete prisutne u ograničenom broju paradigama i izoliranim kategorijama. Primjerice, u engleskom se jeziku dubletni oblici pojavljuju samo u množini nekoliko imenica latinskoga podrijetla (*cactus* > *cactuses/cacti*, *formula* > *formulas/formulae*) ili prošlom vremenu pojedinih glagola (*leap* > *leapt/leaped*, *dive* > *dived/dove*). U njemačkom su jeziku dublete moguće u malom broju imenica u genitivu jednine (*Brot* > *Brottes/Brots*) ili u množini (npr. *Mund* > *Münder/Münde*). Primjeri iz romanskih jezika (npr. imperativ u talijanskome: *andare* > *vai/va'*, particip prošli u talijanskome: *seppellire* > *sepolto/seppellito*, 1. lice prezenta u francuskome: (*je*) *peux/puis*) također su sporadični.

3 Naziv navodi Marković (2012). Ostali nazivi koji se u literaturi upotrebljavaju za tu jezičnu pojavu jesu: preobilje (hrvatski naziv uveden u Bošnjak Botica i Hržica 2016 prema engl. *overabundance*, Thornton 2011), konkurentni oblici (engl. *competing forms*, Bermel i Knittl 2012), suparnički oblici (engl. *rival forms*, Baayen i sur. 2013), varijabilne izlazne vrijednosti (engl. *variable outputs*, Bauer 2014). U ovom članku upotrebljavamo nazive *morfološka dubletnost/dvostrukost* ili *morfološki dubletizam* za cjelokupnu pojavu, odnosno *morfološke dublete* za pojedine oblike. Uzmemo li u obzir da se u nekim kategorijama pojavljuju i trostruki oblici, ti se nazivi mogu činiti neprikladnima. Međutim, dublete (i dubletnost) zbog raširenosti shvaćamo kao krovni naziv za pojavu o kojoj govorimo.

4 Naravno, postoje i autori koji drukčije razmišljaju. Primjerice, Jonke (1965: 199) tvrdi da »uvođenje dvostrukosti, pa i trostrukosti, znači i teoretski i praktički znatan korak naprijed«, dok bi »njihovo nasilno uklanjanje prouzrokovalo teško podnošljiv poremećaj norme književnog jezika«.

U slavenskim jezicima, koji imaju izrazito bogat fleksijski sustav, takvi se primjeri mogu pronaći u gotovo svim kategorijama promjenjivih riječi, pri čemu morfološka dubletnost obuhvaća cijele kategorije leksema, a ne samo izolirane primjere. Primjerice, u češkom jeziku sve imenice muškoga roda koje završavaju na tvrdi suglasnik (njih više od tisuću) mogu imati dva oblika u lokativu jednine: oblik koji završava na *-u* i oblik koji završava na *-ě* (usp. Bermel i Knittl 2012). Hrvatski jezik u tom smislu ne zaostaje za svojim slavenskim srodnicima. Štoviše, broj kategorija u kojima se dvostrukosti pojavljuju te raznovrsnost oblika toliki su da zahtijevaju ozbiljnu analizu. Dublete se u hrvatskom jeziku pojavljuju u sklonidbi imenica i zamjenica, sprezanju glagola, stupnjevanju pridjeva, tvorbi pojedinih kategorija imenica itd. U odjeljku 3.1. donosimo popis kategorija u kojima se pojavljuju morfološke dublete u hrvatskome, što će potvrditi našu tezu da je hrvatski jezik izrazito bogat morfološkim dubletama, odnosno jedan od slavenskih jezika u kojem je ta pojava najraširenija.

U svojem poimanju dubletnosti razlikujemo dvije vrste: (a) *dublete* i (b) *inačice* (*varijante*). U prvu kategoriju ulaze dubletni oblici riječi koji su nastali dodavanjem dvaju (različitih) gramatičkih morfema (npr. genitiv množine: *mrav-i/mrav-ā*), dok u drugu kategoriju ulaze oblici koji su nastali dodavanjem alomorfa jednoga (istoga) morfema (npr. instrumental jednine: *ribar-em/ribar-om*). Time se priklanjamo dosadašnjim gramatičkim opisima (npr. Silić i Pranjković 2005). S jedne strane držimo da je u primjeru genitiva množine riječ o dvama morfemima ili dvama nastavcima koji doduše imaju isto značenje (sadržaj) na konkretnom mjestu, ali im je izraz različit (zato i jesu u sinonimnom odnosu). S druge strane, u primjeru instrumental polazimo od fonološke uvjetovanosti koja je dovela do alternacije (prije glasa) jednoga morfema stvorivši dva alomorfa. Brojni primjeri pokazuju da se ta uvjetovanost koleba, odnosno da se alternacija proširuje i na nepalatalne osnove. Postoje i tumačenja (Marković 2012: 120) da se u obama slučajevima radi o alomorfima, tj. realizacijama istoga morfema (kao jedinice) jer im je značenje isto (padež, broj, sklonidbeni tip) te u tom smislu ne bi bilo razlike između nastavaka *-a* i *-i* u odnosu na nastavke *-om* i *-em*. Ako se tako gleda, ni navedena distinkcija *dubleta vs. inačica* ne bi bila opravdana. Zasad zadržavamo iznesenu distinkciju, a sama kategorija daje se dodatno stupnjevati pa možemo govoriti o *jakoj* (dublete) i *slaboj* (inačice) *dubletnosti* (usp. jaku i slabu supletivnost, npr. Bošnjak Botica i Hržica 2016), pri čemu bi se samo u prvom primjeru radilo o gramatičkoj, tj. morfološkoj sinonimiji.

2. Prethodna istraživanja

Za suvremeni pogled na organizaciju morfoloških paradigama nezaobilazan je pristup razvijen u okviru kanonske tipologije (engl. *canonical typology*) Grevillea Corbetta (2009, 2010). Corbett (2009: 1) definira kanonski fleksijski sustav (engl. *canonical inflectional system*) kao sustav u kojem postoji »jedinstveno preslikavanje s oblika na funkciju i s funkcije na oblik« (engl. *a unique mapping from form to function and from function to form*). Drugim riječima, u kanonskom fleksijskom sustavu

svaka kategorija u fleksijskoj paradigmi ostvaruje se drugim oblikom. Prema toj definiciji kanonski deklinacijski sustav imenica hrvatskoga jezika sastojao bi se od 42 različita oblika (7 padeža * 2 broja * 3 deklinacijska obrasca). Međutim, sam Corbett (2010: 142) priznaje da su primjeri koji se uklapaju u kanon u svjetskim jezicima »vrlo rijetki, možda i nepostojeći«. Argument kojim se u prošlosti opravdavalo nepostojanje kanonskih sustava jest upravo to što bi takva količina oblika bila neekonomična za jezik i korisnika. No, kao što je pokazao npr. Kapović (2012: 29–30), sam pojam ekonomije u jeziku vrlo je problematičan. U svakom slučaju, pojava morfološkoga dubletizma odnosno preobilja predstavljala bi otklon (engl. *deviation*) od kanoničnosti.

Kao što je rečeno, u svjetskom je jezikoslovlju morfološkoj dubletnosti posvećeno malo pozornosti (od radova nastalih u 20. stoljeću ističemo tek Haber 1976), dok je više zanimanja bilo za sintaktičke inačice. Tek su noviji radovi (Fehringer 2004, Thornton 2011, Cappellaro 2013, Santilli 2014, Naghzhguy–Kohan i Kuteva 2016, Guzman Naranjo 2017 itd.), odmaknuvši se od engleskoga jezika kao glavnoga predmeta interesa, pokazali da ta pojava nije tako rijetka i da je vrijedi istraživati.

U području slavenske filologije morfološka dubletnost najdetaljnije je analizirana u češkom jeziku (Cummins 1995, Bermel i Knittl 2012) te nešto manje u poljskome i ruskome (Kottum 1981, Chumakina, Hippisley i Corbett 2004, Baayen i sur. 2013). U hrvatskom su jeziku dosad analizirani uglavnom pojedini leksemi (npr. Mokuter 1969 o rodu imenice *bol*, Okuka 1982 o oblicima komparativa *strog*, Ham 2002 o oblicima imenice *put*, Vukojević i Hudeček 2003 o rodu imenice *jetra*) ili pojedina skupina leksema (Moguš 1977 o dubletnim komparativnim oblicima, Težak 1982 o vokativu jednine imenica e–deklinacije, Grčević 2006, 2007 o instrumentalima jednine imenica i–deklinacije itd.). U novije vrijeme tim se područjem nešto opsežnije uglavnom bave autori okupljeni na projektu izrade baze hrvatskih morfoloških dubleta (Bošnjak 2005, Lečić 2015, Bošnjak Botica i Hržica 2016, Lečić 2016, Polančec 2017, Bošnjak Botica, Budja i Hržica 2018).

Razlozi slabije istraženosti ove teme, osim činjenice da je ipak posrijedi rubna jezična pojava, mogli bi biti i brojni teorijsko–metodološki problemi koji se pojavljuju jednom kad istraživanje počne. Kao što je rečeno, sama je pojava rubna u jeziku te ju je zbog te rubnosti teško uočiti, ali i definirati koje sve jedinice obuhvaća (npr. je li dubletnost imenica u sklonidbi kao rezultat dvorodnosti leksičke ili morfološke naravi, jesu li dubletni oblici varijante jednoga morfema ili su to dva morfema, može li se po različitim navescima govoriti o dubletama u dativu i lokativu jednine itd.). Osim toga, u nekim je primjerima teško povući jasnu granicu između standardnih i nestandardnih oblika riječi jer se različiti normativni priručnici razilaze u mišljenjima (a i kad su ona dana, nisu uvijek uvjerljivo potkrijepljena). Istraživanja u suvremenoj lingvistici sve više ovise o modernim tehnologijama, što posebno vrijedi za morfologiju, u kojoj se morfološki opis jezika često dobiva s pomoću tehnoloških alata; međutim, upravo zbog spomenute rubnosti same pojave jezični

alati često pogrešno ili nepotpuno označuju takve primjere, stoga dobiveni rezultati mogu biti nepouzdana.

3. Projekt Baza hrvatskih morfoloških dubleta

Najtežim i najdugotrajnijim dijelom prethodnih istraživanja morfoloških dubleta u hrvatskome pokazalo se prikupljanje relevantnih podataka (pojavnica dubletnih oblika) i određivanje njihovih vrela. Stoga se najprije počelo s objedinjavanjem na jednom mjestu različitih vrela iz kojih su crpljeni podatci o dubletnim oblicima. Primarni je cilj stvoriti pouzdan i javno dostupan jezični resurs namijenjen ponajprije znanstvenomu istraživanju promatrane pojave, ali koji može služiti i kao referentna točka u raspravi o kojoj god konkretnoj jedinici ili kategoriji dubletnosti. Projekt se od prosinca 2017. provodi u okviru redovite djelatnosti Instituta za hrvatski jezik i jezikoslovlje.

Pri izradi podatkovne baze autori su definirali četiri koraka:

1. popis i opis kategorija dubletnosti u hrvatskome
2. popis i opis svih vezanih kategorija relevantnih za istraživanje
3. izrada strukture baze
4. međusobno povezivanje dijelova baze i definiranje načina pretraživanja.

3.1. Kategorija dubletnosti u hrvatskome

Prvi je korak prije izrade same baze odrediti sve gramatičke kategorije u kojima se pojavljuju dublete u hrvatskom jeziku. Posao je olakšan zahvaljujući prethodnim radovima autorâ. Pregledom hrvatskih gramatika, rječnika, jezikoslovnih članaka, jezičnih savjetnika i ostale literature izdvojili smo gramatičke funkcije u kojima se dvostrukosti pojavljuju te kategorije leksema koje ih ispunjaju (popis obuhvaća dublete u pravom smislu riječi i varijante/inačice). Pri sastavljanju takva popisa moguće je poći od različitih kriterija: od jezičnih kategorija (pa tako opisati sve dubletnosti koje se nalaze, primjerice, u množini imenskih riječi), od čestote određene kategorije dubletnosti ili od opsežnosti kategorije. Radi lakšega snalaženja odlučili smo se na pristup koji polazi od vrsta riječi te unutar svake vrste riječi daje što iscrpniji popis kategorija. Dakle, nadređena je podjela lako dostupna i jasna (i korisnicima izvan uže struke), dok je daljnja podjela nešto određenija. Podjelom nisu nužno iscrpljene sve vrste dubletnosti u hrvatskome, ali se teži visokomu stupnju opisa svih mogućnosti. Isto tako, zbog prostornoga ograničenja i koncepcije rada ne donosimo detaljniju klasifikaciju građe (primjerice, prema dodatnim fonološkim, morfološkim ili morfosemantičkim kriterijima).

1) imenički sustav:

- I1: imenice a–sklonidbe s postojanim *a* u GD(A)VLI jd. i množini: *Katar, Juraj* itd.
- I2: imenice a–sklonidbe s umetanjem *a* u NA jd.: *subjek(a)t, objek(a)t, koncer(a)t* itd.

- I3: imenice a–sklonidbe s umetanjem *a* u G mn.: *prefiks, sufiks* itd.
- I4: imenice a–sklonidbe s postojanim *e* u GD(A)VLI jd. i množini: *Jupiter, pater* itd.
- I5: imenice a–sklonidbe s prijegasom u I jd.: *ribar, mesar, vijenac, car, maslac, muzej, radio, stric* itd.
- I6: imenice a–sklonidbe s prijegasom u I jd. i množini: *kut, put, šut, bus, bris, pir, mraz, plašt, prišt, sprej* itd.
- I7: imenice a–sklonidbe s prijegasom u množini: *kos, nos, sir, zec* itd.
- I8: imenice a–sklonidbe s vokalizacijom u NA jd.: *predio/predjel, seoce/selce* itd.
- I9: imenice a–sklonidbe s vokalizacijom u GDVLI jd. i NDAVLI mn.: *stolac, kolac, zaselak* itd.
- I10: imenice a–sklonidbe u V jd.: *leptir, med, svijet, predak, Englez, grašak, Boris, Domagoj* itd.
- I11: imenice a–sklonidbe s dugom i kratkom množinom: *golub, prsten, slučaj, vuk, znak, vrag, velegrad, prototip* itd.
- I12: imenice a–sklonidbe u G mn.: *mrav, crv, prst, gost, nokat, zub* itd.
- I13: imenice e–sklonidbe sa sibilizacijom u DL jd.: *točka, tvrtka, rešetka, zaprška, maska, juha* itd.
- I14: imenice e–sklonidbe u V jd.: *izdajica, varalica, kukavica, Ana, Sanja, Marija, Lana* itd.
- I15: imenice e–sklonidbe u G mn.: *naranča, kruška, breskva, crkva, izložba, sorta* itd.
- I16: imenice e–sklonidbe u G mn.: *torba, noga, ruka, majka, torta, cesta* itd.
- I17: imenice i–sklonidbe u I jd.: *misao, noć, ljubav, mladost, radost, zrelost, mast* itd.
- I18: imenice srednjega roda proširene osnove u GDLI jd.: *jaje, zvonce, seoce* itd.
- I19: imenica *prsa* u G mn.: *prsā/prsiju*
- I20: imenice m. i ž. r. na suglasnik u cijeloj paradigmi: *bol, glad, čar, splav, varoš, trulež, gnjilež, pelud* itd.
- I21: imenice e–sklonidbe i i–sklonidbe u cijeloj paradigmi: *vrst/a, žeđ/a* itd.

2) zamjenički sustav:

- Z1: posvojne zamjenice za 1. i 2. lice jd. sa stezanjem u GDAL: *moj, tvoj, svoj*
- Z2: posvojne zamjenice za 3. lice jd. ž. r. u cijeloj paradigmi: *njezin/njen*
- Z3: posvojne zamjenice u GDL jd. m. i s. r.: *njegov, njen, njezin*
- Z4: neodređena zamjenica *sav* u G mn.: *svih/svijju*

3) pridjevski sustav⁵:

- P1: jednosložni pridjevi u komparativu: *čist, mrk, strog, lijen, nijem, plav, bijel, suh, dug* itd.

⁵ Prve dvije kategorije preslikavaju se i na priloge, koje zasad ne analiziramo.

- P2: dvosložni pridjevi u komparativu: *gorak, mrzak, krotak, gladak, blizak, bijesan, kasan* itd.
- P3: pridjevi s prijeglasom u cijeloj paradigmi: *ribarev/ribarov, carev/carov, trešnjev/trešnjov, višnjev/višnjov* itd.
- P4: pridjevi s vokalizacijom u cijeloj paradigmi: *anđelski/andeoski* itd.

4) glagolski sustav⁶:

- G1: glagoli 1. i 5. vrste u prezentskoj osnovi: *skitati, šetati, uzimati, vagati, sipati* itd.
- G2: glagoli 1. i 6. vrste u prezentskoj osnovi: *tapšati, škakljati* itd.
- G3: glagoli 2. i 0. vrste u prezentskoj osnovi: *vapiti, slomiti* itd.
- G4: glagoli 3. i 0. vrste u prezentskoj osnovi: *bdjeti, vreti, zreti* itd.
- G5: glagoli 2. i 0. vrste u prezentskoj osnovi: *mniti* itd.
- G6: glagoli 4. i 5. vrste u prezentskoj osnovi: *drhtati* itd.
- G7: glagoli 0. vrste u prezentskoj osnovi: *dati, znati, reći, moći* itd.
- G8: glagoli 9. i 0. vrste u infinitivnoj osnovi: *maći/maknuti, taći/taknuti, navići/naviknuti, nageti/nagnuti* itd.
- G9: glagoli 2. i 0. vrste u cijeloj paradigmi: *crpsti/crpiti, podignuti/podići, šiti/šivati* itd.
- G10: glagoli 0. vrste u cijeloj paradigmi: *izaći/izići, unići/ući* itd.
- G11: glagoli 0. vrste u gl. pridjevu radnom: *potegnuti, stegnuti* itd.
- G11: glagoli 0. vrste u gl. pridjevu trpnom: *naduti, šiti* itd.
- G12: glagoli 1. i 7. vrste u cijeloj paradigmi: *označavati/označivati, naglašavati/naglašivati* itd.
- G13: glagoli 1. i 4. vrste u cijeloj paradigmi: *sadržati/sadržavati* itd.
- G14: glagoli 2. i 4. vrste u infinitivnoj osnovi: *bojiti/bojati, zujiti/zujati, kliziti/klizati* itd.
- G15: glagoli 2. i 3. vrste u infinitivnoj osnovi: *veljeti/veliti, mrzjeti/mrziti* itd.

Broj je kategorija mogao biti i veći (kad bi se npr. dodatno razložili glagoli 0. vrste ili imenice e–sklonidbe s dubletom u V jd.) ili manji (kad bi se npr. u jednu kategoriju okupili svi leksemi s dvostrukim oblicima u instrumentalu jednine). Gornji popis zacijelo nije konačan.

3.2. Ostale kategorije relevantne za istraživanje

Za svaki unos u bazu (tj. potvrdu pojedinoga dubletnog oblika u jezičnim vreli-ma) određujemo sljedeće vrijednosti:

- relevantne morfološke kategorije (vrsta riječi, rod, broj, padež, lice itd.) ko-dirane prema normi MULTEXT–East (Erjavec i sur. 2003)
- djelo u kojem se pojavljuje
- autor djela i areal autora

⁶ Glagolske vrste (konjugacije) slijede podjelu Z. Jelaske (npr. Jelaska 2003).

- vrijeme nastanka
- vrelo u kojem je pronađen.

Hrvatski jezik zbog stupnja svoje mnogorazinske složenosti zahtijeva dijakronijsko proučavanje koje će početi od partikularnoga da bi, eventualno, došlo do totalnoga. Zato smo kao jedinice partikularnoga istraživanja uveli zemljopisne areale.

Arealna razdioba idealna je za period prije stvaranja općehrvatskoga, odnosno prije procesa jezičnoga objedinjavanja u sklopu ilirskoga pokreta. Nakon toga njezina upotrebljivost biva sve manja. Razlikujemo areal rođenja i materinskoga govora autora (Areal 1) te areal u kojem je autor djelovao (Areal 2), ako je različit od Areala 1. To zato što su neki pisci promijenivši prebivalište, morali promijeniti i jezik. Međutim, pretpostavlja se da pisac poprimanjem novoga idioma ne zapušta posve svoj materinski idiom.⁷ Razdioba na areale uzima u obzir više kriterija:

1. Dijalektna trodioba. Hrvatski se od doseljenja Slavena na jugozapad konstituirao kao trodijalektan jezik, a razlike među dijalektima, kao i među poddijalektima jednoga dijalekta, znaju biti velike. Čakavština se, na primjer, prostire od krajnjega sjeverozapada hrvatskoga jezičnog prostora daleko na jugoistok, a na sjeveru prodire dosta i na kopneni sjeveroistok. Sjeverna će čakavština, nadati se je, imati više doticajnih točaka sa slovenskim jezikom ili hrvatskom kajkavštinom, a južna s kontaktnim štokavskim govorima.
2. Politički entitet. Kako je hrvatski jezični prostor bio mnogostruko politički razdijeljen, iskonom različiti govori u kojoj god političkoj jedinici nužno su morali više ili manje konvergirati.
3. Seobe. Za vrijeme i poslije turskih osvajanja dijalektološka se slika umnogom promijenila. Jedni su se dijalekti iskorijenili i raselili, raspršili unutar i izvan hrvatskoga jezičnog prostora, dok su drugi na njihovo mjesto pridošli.
4. Jezično susjedstvo. Određeni okoliši hrvatskoga jezika stoje u neposrednom kontaktu s genetički veoma srodnim, a onda i veoma sličnim jezicima, nehrvatskim istočnoštokavskim i slovenskim kajkavskim. Time je put ulaženju jezičnih fenomena iz tih jezika u hrvatski otvoren.

Na temelju postavljenih kriterija utvrdili smo sljedeće areale:

1. Bačka i Budim
2. Slavonija
3. istočna Bosna
4. zapadna Bosna, zapadna Hercegovina, dalmatinsko zaleđe (tj. štokavska Dalmacija)
5. Makarsko primorje i Neretva
6. Dubrovnik

⁷ Na primjer, pisac V. Došen rođen je u zapadnom novoštokavskom ličkobunjevačkom kraju, ali je svoja djela napisao u Slavoniji, u okružju slavonske štokavštine.

7. Boka kotorska
8. čakavska Dalmacija
9. čakavska uža Hrvatska i Istra
10. štokavska uža Hrvatska
11. kajkavska uža Hrvatska.

Kad je riječ o vrelima u kojima pronalazimo pojavnice, u prvoj fazi izrade baze relevantna su nam vrela dostupni korpusi hrvatskoga jezika: Hrvatska jezična riznica (Čavar i Brozović Rončević 2012), Hrvatski nacionalni korpus (Tadić 1998) i Hrvatski mrežni korpus hrWaC (Ljubešić i Klubička 2014). Izvlačenje podataka iz navedenih korpusa nije nimalo lak zadatak i svakoj kategoriji dubletnosti iz poglavlja 3.1. trebalo je pristupiti na poseban način. U većini je kategorija dubletnosti jedan od dubletnih nastavaka jednoznačan, tj. upotrebljava se samo na tom mjestu u paradigmi, dok drugi ima nekoliko gramatičkih značenja, tj. upotrebljava se na više mjesta u paradigmi. Ove prve oblike moguće je pretraživati nekim jednostavnijim i općenitijim CQL (engl. *Corpus Query Language*) upitom, dok je za pronalazak potonjih oblika potrebno dodatno ograničiti upite ili pribjeći »ručnoj« analizi. Primjerice, jednostavnim upitom [word=«.*arem«] možemo dohvatiti sve pojavnice oblika –em u instrumentalu jednine imenica muškoga roda na –ar (uz nekoliko izoliranih leksema, kao što su *harem* i *barem*), dok će se među rezultatima upita [word=«.*arom«] naći jako mnogo pojavnica instrumentala jednine imenica ženskoga roda na –ara, koje nisu predmet našega interesa. Iako su sva tri korpusa morfosintaktički označena, svjesni smo nedostataka automatskoga označavanja, stoga ograničavanju upita prema vrsti riječi, rodu i drugim morfosintaktičkim kategorijama također pristupamo oprezno i radije pribjegavamo »ručnim« metodama pretraživanja (leksem po leksem, oblik po oblik i sl.).

U kasnijim fazama izrade baze potvrdama iz navedenih korpusa pridodat ćemo potvrde pronađene u drugim mrežnim vrelima, starijoj hrvatskoj leksikografiji i književnosti, jezikoslovnim člancima, Hrvatskom korpusu govornoga jezika (Kuvač Kraljević i Hržica 2014) itd.

Autori su svjesni da će neke vrijednosti u kategorijama navedenima u poglavlju 3.2. ostati prazne jer je za neke pojavnice u vrelima nemoguće utvrditi, primjerice, autora ili godinu nastanka. Ovo će osobito biti slučaj za pojavnice iz korpusa hrWaC, koji obiluje nepotpisanim i neredigiranim internetskim upisima.

3.3. Izrada strukture baze⁸

Za potrebe projekta izrađena je jedinstvena relacijska baza podataka u okruženju MySQL te pripadajuće sučelje za unos i obradu građe. Shema baze sastoji se od triju temeljnih tablica, gdje je tablica *obličnica* noseća te se ostale tablice vežu na nju. Osnovni unos sastoji se od jedanaest polja od kojih su četiri vezana na vanjske tablice te se popunjavaju odabirom vrijednosti polja iz padajućega izbornika.

8 Zahvaljujemo kolegi dr. sc. Bruni Nahodu na pomoći pri oblikovanju i izradi prve inačice baze.

Ovako razrađena shema osigurava veliku razinu sustavnosti obrade te će u budućnosti omogućiti kvalitetne temelje za izradu mrežne tražilice, ali i za analizu građe u daljnjim istraživanjima.

3.4. Međusobno povezivanje dijelova baze i definiranje načina pretraživanja

Sučelje baze koje će biti vidljivo korisnicima podijeljeno je u dva dijela, koja predstavljaju dva različita oblika upotrebe baze: pretraživanje i pregledavanje. Prvi dio sučelja bit će organiziran po načelu *pretraživanja* prema ključnim riječima. Korisnik će u polje za pretraživanje moći upisati bilo leksem bilo oblik riječi nakon čega će dobiti sve dubletne pojavnice toga leksema/oblika koje se nalaze u bazi. Pretraživanje će se moći dodatno ograničiti prema autoru, arealu, godini, vrsti teksta, vrelu, morfološkim kategorijama i ostalim kategorijama navedenima u odjeljku 3.2.

Drugi dio sučelja omogućuje *pregledavanje* baze prema kategorijama. U tom dijelu sučelja bit će navedene vrste riječi koje su unesene u bazu (imenice, pridjevi, zamjenice, glagoli).⁹ Označivanjem jedne od tih četiriju kategorija korisniku će se prikazati abecedni popis svih leksema koji joj pripadaju i koji su uneseni u bazu. Odaabirom jednoga leksema korisniku će se prikazati isti rezultati kao da im je pristupio putem pretraživanja. Osim po vrstama riječi korisnik će moći pregledavati bazu i po kategorijama navedenima u odjeljku 3.1. Primjerice, odabirom kategorije I7 korisniku će se prikazati abecedni popis svih unesenih imenica a–sklonidbe koje se pojavljuju s dubletnim oblicima u dugoj množini. Postupak pristupanja pojedinim potvrdama u bazi isti je kao i u prethodno opisanim načinima korištenja bazom.

4. Koristi od baze i moguća primjena u istraživanjima

Početa zamisao s kojom se pošlo u stvaranje ovakve baze bila je stvaranje lako dostupna i pretraživa resursa koji bi ubrzao istraživanja morfološke dvostrukosti u hrvatskome i koji bi bio otvoren hrvatskoj znanstvenoj zajednici. Koliko znamo, baza morfoloških dubleta nema u drugim jezicima, stoga je ovakav projekt novina u jezikoslovlju.

Posebno je važna dijakronijska dimenzija baze jer ona omogućuje praćenje pojave od trenutka njezina ulaska u jezik kroz sve promjene do suvremenoga statusa (potkrijepljeno podacima iz najnovijih općejezičnih korpusa). Držimo da će se tek okupljanjem veće građe te sučeljavanjem dijakronijskih i sinkronijskih podataka moći odgovoriti na pitanje o tome je li posrijedi privremeno ili trajno jezično obilježje.

Osim dijakronijske uzima se u obzir i dijatopijska perspektiva da se odrede, ondje gdje je to moguće, mjesto ulaska, smjer širenja i današnja rasprostranjenost pojedinih dubletnih oblika. Usporedbom jednih i drugih podataka mogli bismo doći do odgovora na pitanje koliko je promatrana pojava ishod vlastita jezičnoga razvo-

9 Dubletnost u stupnjevanju priloga razumijemo kao zatečeno stanje u odnosu na polazni pridjevni oblik te je zasad ne analiziramo.

ja, a koliko je potaknuta utjecajem srodnih jezičnih sustava. Na temelju baze moći će se odrediti i tipična mjesta ulaska, smjer širenja i današnja rasprostranjenost pojedinih dubletnih oblika.

Na osnovi unesenih podataka korisnici baze moći će odgovoriti, na primjer, na ova pitanja: Pojavljuje li se imenica *bol* češće u ženskom ili u muškom rodu? Ovisi li distribucija dvaju oblika o razdoblju? Je li ta prevlast različita u različitim padežima? Ovisi li o sintaktičkom okruženju? Pojavljuju li se različite tendencije s obzirom na značenje (npr. fizička/duševna *bol*)?¹⁰ Ima li razlike u čestoti dviju osnova glagola šetati s obzirom na lice i broj? Kako izgleda distribucija triju nastavaka za genitiv množine imenica ženskoga roda?

Uz odgovore na pitanja o distribuciji dubletnih oblika konkretnih imenica ili kategorija držimo da baza može doprinijeti davanju odgovora i na općenitija jezična pitanja. Primjerice, ako dijakronijski možemo pratiti razvoj pojedinoga dubletnoga oblika na temelju njegova mjesta i vremena ulaska u jezik, mogli bismo pretpostaviti smjerove budućega razvoja morfološke sinonimije u hrvatskom jeziku. Isto tako na temelju podataka o arealima autorâ možemo pratiti ishodište dvostrukosti kao unutarjezične pojave ili kao posljedice jezičnih kontakata. Korisnici baze imat će dovoljno podataka s pomoću kojih će moći sudjelovati u stručnim raspravama o prirodi morfološke dvojnosti te njezine uloge u samom shvaćanju jezika. Autori smatraju da će u konačnici *DvojBa* doprinijeti raspravi o jednom od temeljnih pitanja: je li jezik savršen sustav u kojem su dublete samo prolazne pojave prelaska iz jednoga u drugi oblik ili je to znatno labavije uređen sustav čiji su dublete sastavni dio.

5. Zaključak

Izrada jezičnoga resursa morfoloških dubleta nekoga jezika inovativan je pristup proučavanju ovoga područja. Istraživanjima se obično pristupa prikupljanjem podataka o jednoj kategoriji dubletnosti te njezinim opisivanjem, često samo na deskriptivnoj razini, bez podataka o udjelu pojedinih oblika. Time se gubi mogućnost odgovora na važna pitanja povezana s našim shvaćanjem jezika, to jest određivanja uloge morfološke dubletnosti u jezičnom sustavu. Radom na jedinstvenoj bazi koja omogućuje pretragu prema kategorijama ili pojedinim riječima, a uključuje ključne informacije (npr. autorstvo, godina, vrsta djela, areal) bitno se olakšava dohvat relevantnih podataka i provođenje metodološki utemeljenih istraživanja. Baza može

10 Smještanje imenica poput *bol* u kategorije dubletnosti sinkronijski gledano ne uklapa se u predloženi model jer po opisu u suvremenim rječnicima hrvatskoga jezika proizlazi da postoje dva leksema *bol*, koji u jednom rodu (i sklonidbi) imaju jedno značenje, a u drugom drugo. Međutim, prethodna istraživanja pokazala su da je razjednačenje značenja na temelju oblika (ili oblika na temelju značenja, što se normativno danas provodi) nastalo prilično kasno. Muški se rod u hrvatskom jeziku proširio pod utjecajem pisaca s istočnoga ruba hrvatske štokavštine, a njegovo zauzimanje značenja fizičke patnje pod utjecajem medicinskoga žargona tek u 20. st. (usp. Mokuter 1969, Bošnjak Botica, Budja i Hržica 2018).

služiti i kao dodatan izvor drugim jezičnim resursima, primjerice, može se povezati s rječničkim¹¹ ili gramatičkim resursima te tako omogućiti preciznije opise jezika.

Način organizacije podataka, trenutačno teže dostupnih i dohvatljivih, u jedinstvenu bazu omogućuje njihovo taksativno navođenje i pretragu. Podrobnija klasifikacija građe (primjerice, dodatnim fonološkim, morfološkim ili morfosemantičkim kriterijima) preduvjet je temeljitije znanstvene obrade ove jezične pojave. Stoga ova baza može znanstvenicima omogućiti da lakše dohvate podatke koje će dodatno klasificirati i obrađivati, eliminirajući dugotrajan i naporan put prikupljanja podataka. Sama struktura baze takva je da dopušta proširenja novim kategorijama klasifikacija pa je tako moguće i da se tijekom razvoja uvedu nove kategorije.

6. Literatura

- Aronoff, Mark (1976). *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.
- Baayen, R. Harald, Anna Endersen, Laura A. Janda, Anastasia Makarova i Tore Nessel (2013). Making choices in Russian: Pros and cons of statistical methods for rival forms. *Russian Linguistics* 37(3): 253–291.
- Babić, Stjepan (1962). Duljina ili dužina? *Jezik* 11: 62–63.
- Bauer, Laurie (2014). Grammaticality, acceptability, possible words and large corpora. *Morphology* 24: 83–103.
- Bermel, Neil i Luděk Knittl (2012). Corpus frequency and acceptability judgments: A study of morphosyntactic variants in Czech. *Corpus Linguistics and Linguistic Theory* 8(2): 241–275.
- Bošnjak, Tomislava (2005). Gramatička i semantička kolebljivost dvosložnih imenica sa sufiksom –ež. *Rasprave Instituta za hrvatski jezik i jezikoslovlje* 31: 1–18.
- Bošnjak Botica, Tomislava i Gordana Hržica (2016). Overabundance in Croatian dual-class verbs. *Fluminensia: Journal for philological research* 28(1): 44–59.
- Bošnjak Botica, Tomislava, Jurica Budja i Gordana Hržica (2018). Rodovna i sklonidbena dvojnost imenica u hrvatskom jeziku. *Rasprave Instituta za hrvatski jezik i jezikoslovlje* 44(1): 57–88.
- Cappellaro, Chiara (2013). Overabundance in diachrony: a case study. S. Cruschina, M. Maiden, and J. C. Smith, ur. *The Boundaries of Pure Morphology: Diachronic and Synchronic Perspectives*. Oxford: Oxford University Press, 209–220.
- Chumakina, Marina, Andrew Hippisley and Greville G. Corbett (2004). Historical changes in the Russian lexicon: a case of alternating suppletion. *Russian Linguistics* 28: 281–315.
- Clark, Eve V. (1987). The Principle of Contrast: A constraint on language acquisition. Brian MacWhinney, ur. *Mechanisms of Language Acquisition*. Hillsdale, NJ: Erlbaum, 157–193.
- Corbett, Greville G. (2009). Canonical Inflectional Classes. *Proceedings of the Decembrettes* 6: 1–11.

11 Na primjer, planirano je povezivanje s Hrvatskim mrežnim rječnikom (Mrežnikom) koji se izrađuje u Institutu za hrvatski jezik i jezikoslovlje.

- Corbett, Greville G. (2010). Canonical derivational morphology. *Word Structure* 3(2): 141–155.
- Cummins, George M. (1995). Locative in Czech: –u or –e: Choosing locative singular endings in Czech nouns. *Slavic and East European Journal* 39: 241–260.
- Ćavar, Damir i Dunja Brozović Rončević (2012). Riznica: The Croatian Language Corpus. *Prace filologiczne* LXIII: 51–65.
- Erjavec, Tomaž, Cvetana Krstev, Vladimir Petkevič, Kril Simov, Marko Tadić i Duško Vitas (2003). The MULTEXT–East Morphosyntactic Specifications for Slavic Languages. *Proceedings of the EACL2003 Workshop on Morphological Processing of Slavic Languages*. Budimpešta: ACL 2003, 25–32.
- Fehringer, Carol (2004). How Stable are Morphological Doublets? A Case Study of /ə/ ~ ø variants in Dutch and German. *Journal of Germanic Linguistics* 16(4): 285–329.
- Giegerich, Heinz (2001). Synonymy blocking and the Elsewhere Condition: Lexical morphology and the speaker. *Transactions of the Philological Society* 99: 65–98.
- Grčević, Mario (2006). Instrumental jednine imenica i–sklonidbe u hrvatskome književnom jeziku od 16. do 19. stoljeća. *Jezik* 53: 50–59.
- Grčević, Mario (2007). Instrumental imenica i–vrste na –st u suvremenome hrvatskome književnom jeziku. *Jezik* 54: 15–22.
- Guzmán Naranjo, Matias (2017). The se–ra alternation in Spanish subjunctive. *Corpus Linguistics and Linguistic Theory* 13(1): 97–134.
- Haber, Lyn R. (1976). Leaped and leapt: A theoretical account of linguistic variation. *Foundations of Language* 14: 211–238.
- Ham, Sanda (2002). Putem ili putom, putovima ili putevima. *Jezik* 49: 134–142, 183–190.
- Jelaska, Zrinka (2003). Hrvatski glagoli. Zagreb: Sveučilišna škola hrvatskoga jezika i kulture.
- Jonke, Ljudevit (1965). *Književni jezik u teoriji i praksi*. Zagreb: Znanje.
- Kapović, Mate (2012). Čiji je jezik? Zagreb: Algoritam.
- Kottum, Steinar E. (1981). The genitive singular form of masculine nouns in Polish. *Scando-Slavica* 27(1): 179–186.
- Kroch, Anthony (1994). Morphosyntactic variation. K. Beals i sur., ur. *Papers from the 30th regional meeting of the Chicago Linguistic Society*. Chicago: University of Chicago Press.
- Kuvač Kraljević, Jelena i Gordana Hržica (2014). Croatian Adult Spoken Language Corpus (HrAL). *Fluminensia* 28(2): 87–102.
- Lečić, Dario (2015). Morphological doublets in Croatian: the case of the Instrumental singular. *Russian Linguistics* 39. 375–393.
- Lečić, Dario (2016). *Morphological Doublets in Croatian: A multi-methodological analysis* (PhD dissertation). University of Sheffield.
- Ljubešić, Nikola i Filip Klubička (2014). {bs,hr,sr}WaC – Web corpora of Bosnian, Croatian and Serbian. F. Bildhauer, & R. Schäfer ur. *Proceedings of the 9th Web as Corpus Workshop (WaC-9)*, Gothenburg, 29–35.
- Marković, Ivan (2012). *Uvod u jezičnu morfologiju*. Zagreb: Disput.
- Moguš, Milan (1977). Još o neujednačenosti komparativnih oblika. *Jezik* 25: 146–148.

- Mokuter, Ivan. 1969. Rod imenice *bol* u novijoj hrvatskoj književnosti. *Zbornik za filologiju i lingvistiku* 12: 183–194.
- Naghzguay–Kohan, Mehrdad i Tanya Kuteva (2016). On competition and blocking in inflectional morphology: Evidence from the domain of number in New Persian. *Folia Linguistica* 50(1): 65–96.
- Okuka, Miloš (1982). I stroži i strožiji. *Književni jezik* 11: 93–97.
- Polančec, Jurica (2017). Naknadni prijedlog iza glasova s t z u hrvatskom jeziku. *Rasprave Instituta za hrvatski jezik i jezikoslovlje* 43(1): 197–225.
- Santilli, Enzo (2014). *Italian Comparatives: a Case of Overabundance?* Magistarski rad. Sveučilište u L'Aquili.
- Silić, Josip (1998). Leksik i norma. *Radovi Zavoda za slavensku filologiju* 32: 169–175.
- Tadić, Marko (1998). Raspon, opseg i sastav korpusa suvremenoga hrvatskoga jezika. *Filologija* 30–31: 337–347.
- Težak, Stjepko (1982). Vokativ jednine imenica e vrste. *Radovi Zavoda za slavensku filologiju* 17: 31–40.
- Thornton, Anna M. (2011). Overabundance (multiple forms realizing the same cell): a non-canonical phenomenon in Italian verb morphology. M. Maiden i sur., ur. *Morphological Autonomy: Perspectives from Romance Inflectional Morphology*. Oxford: Oxford University Press, 358–381.
- Thornton, Anna M. (2012). Reduction and maintenance of overabundance: A case study on Italian verb paradigms. *Word Structure* 5(2): 183–207.
- Vukojević, Luka i Lana Hudeček (2003). Jetra – imenica ženskog ili srednjeg roda? *Rasprave Instituta za hrvatski jezik i jezikoslovlje* 29: 355–362.

Database of Croatian Morphological Doublets (DvojBa)

Morphological doubletism refers to a phenomenon whereby two (or more) forms of the same lexeme have the same grammatical meaning (e.g. *leapt/leaped*, *cactuses/cacti*). Inspired by a lack of a linguistic resource where examples of specific doublet forms in the Croatian language could be found, the authors of this paper launched the project of creating a database whose ultimate goal is to create a list of grammatical categories in which morphological doubletism occurs and a list of lexemes which appear with doublet forms, as confirmed by tokens of the respective forms retrieved from Croatian language sources (computer corpora, lexicographic sources, literature and other relevant materials). The primary aim is to create an easily accessible and searchable resource that would enhance the research of morphological doubletism in Croatian and that the Croatian scientific community would be able to access for free. Each token would be accompanied by information regarding the work it occurs in, the year the work was created, the author of the work, his/her geographic area, type of text and the relevant morphological labels. These data can be used to trace the phenomenon from the point of its entry into the language through all changes up to its present-day status, determine the distribution of individual forms in the modern language and ultimately determine the role of morphological doublets in the language system. The database can contribute to providing more precise descriptions of the language, which mostly disregard the phenomenon in question.

Ključne riječi: morfološke dublete, hrvatski jezik, podatkovne jezične baze

Keywords: morphological doublets, Croatian, linguistic databases