

Konjugacijska vrsta glagola uzimati u hrvatskim računalnim korpusima

Kirasić, Žana

Undergraduate thesis / Završni rad

2016

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Rijeka, Faculty of Humanities and Social Sciences / Sveučilište u Rijeci, Filozofski fakultet u Rijeci**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:186:816043>

Rights / Prava: [In copyright / Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-04-26**



Repository / Repozitorij:

[Repository of the University of Rijeka, Faculty of Humanities and Social Sciences - FHSSRI Repository](#)



SVEUČILIŠTE U RIJECI
FILOZOFSKI FAKULTET

Žana Kirasić

Konjugacijska vrsta glagola uzimati u hrvatskim računalnim korpusima

(ZAVRŠNI RAD)

Rijeka, 2016.

SVEUČILIŠTE U RIJECI
FILOZOFSKI FAKULTET
Odsjek za kroatistiku

Žana Kirasić

Matični broj:

Konjugacijska vrsta glagola uzimati u hrvatskim računalnim korpusima

ZAVRŠNI RAD

Prediplomski studij: Hrvatski jezik i književnost

Mentor: doc. dr. sc. Kristian Novak

Rijeka, 8. rujna 2016.

SADRŽAJ:

1. UVOD	3
2. KONJUGACIJSKE VRSTE U HRVATSKOME JEZIKU	4
2.1. Hrvatske glagolske vrste u Hrvatskoj gramatici (Barić i suradnici).....	5
2.2. Hrvatske glagolske vrste u Gramatici hrvatskoga jezika (Silić – Pranjković)	7
3. GLAGOL UZIMATI U KONTEKSTU HRVATSKIH KONJUGACIJSKIH VRSTA.....	9
4. HRVATSKI RAČUNALNI KORPUSI I HNK v3.0	10
4.1. Računalna obrada hrvatskih korpusa	10
4.2. Počeci računalne obradbe hrvatskoga i nacionalni korpus	13
4.3. Raspon, opseg i sastav korpusa hrvatskoga suvremenog jezika.....	15
4.4. Zašto nam je potreban višemilijunski referentni korpus?	17
4.5. Hrvatski nacionalni korpus (HNK).....	19
4.6. Hrvatska jezična riznica.....	20
5. METODOLOGIJA ISTRAŽIVANJA	21
5.1. Struktura istraživanja.....	21
6. ANALIZA	23
6.1. Grafikon: usporedba uporabe oblika iz konjugacijske vrste V.1 i V.2.....	23
6.2. Analiza grafikona	24
7. ZAKLJUČAK	25
8. SAŽETAK.....	26
9. KLJUČNE RIJEČI	27
10. POPIS LITERATURE	28

1. UVOD

U ovom će radu biti prikazana problematika konjugacijske vrste glagola *uzimati*, i to na temelju zastupljenosti pojedinih oblika u hrvatskom računalnom korpusu HNK v3.0.

Najprije ću predstaviti dvije podjele glagolskih vrsta prema dvjema gramatikama, Gramatici hrvatskoga jezika čiji su autori Josip Silić i Ivo Pranjković, te Hrvatskoj gramatici autora Eugenija Barića i drugih suradnika.

U nastavku ću glagol *uzimati* smjestiti u kontekst hrvatskih konjugacijskih vrsta te dati osnovne podatke o povijesti, stanju i perspektivama hrvatskih računalnih korpusa gdje će biti prikazan pregled projekata u Republici Hrvatskoj smještenih u Zavodu za lingvistiku Filozofskoga fakulteta sveučilišta u Zagrebu. Zatim ću na temelju literature prikazati važnost računalne tehnologije u istraživanju prirodnoga jezika te reći nešto o statusima koja ta istraživanja moraju imati.

Nakon kratkog povijesnoga pregleda doći ćemo do korpusa hrvatskoga suvremenog jezika pa će biti objašnjen njegov raspon, opseg i sastav. Bit će također prikazano zašto nam je potreban višemilijunski referentni korpus i naposljetku ću reći nešto o Hrvatskom nacionalnom korpusu i Hrvatskoj jezičnoj riznici.

U poglavlju *Metodologija istraživanja* usporedit ću frekventnost pojavljivanja prezentskih oblika glagola *uzimati* prema konjugacijskim vrstama V.1 i V.2 te navesti primjere dobivene pretraživanjem Hrvatskoga nacionalnog korpusa (HNK) v3.0.

U nastavku je grafički prikazan i komentiran kvantitativni odnos zastupljenosti pretraživanih oblika u računalnom korpusu.

2. KONJUGACIJSKE VRSTE U HRVATSKOME JEZIKU

Želimo li govoriti o osvještavanju hrvatskih glagolskih vrsta moramo krenuti od stranih jezika, kao što su latinski, francuski, engleski, njemački itd. (Marković 2012: 215)

Govornici hrvatskoga jezika nisu niti svjesni kako konjugacijski tipovi postoje u hrvatskom jeziku, ali im je jasno da pri učenju stranih jezika treba svladati konjugacijske tipove. (Isto: 215)

U svim slavenskim jezicima glagoli se raspoređuju u nekoliko fleksijskih vrsta, nekoliko konjugacija ili sprega. (Isto: 215)

Tradicionalno, ustalile su se dvije vrste prema dvjema osnovama ili bazama, a to su infinitivno–aoristna od koje se u hrvatskom jeziku tvore infinitiv, aorist, glagolski prilog prošli, te glagolski pridjevi radni i trpni, te prezentska osnova od koje se u hrvatskome jeziku tvore prezent, imperativ i glagolski prilog sadašnji. (Isto: 215)

U hrvatskoj gramatičkoj tradiciji preteži razvrstavanje prema infinitivnoj osnovi na šest, odnosno sedam vrsta. Sedma vrsta skup je nepravilnih glagola (biti, ići, htjeti). Takvo se razvrstavanje zasniva na podjeli iz staroslavenske gramatike. (Isto: 215)

Glagolske vrste prema infinitivnoj osnovi imamo u gramatikama V. Babukića, P. Budmanija, A. V. Tkalčevića, T. Maretića i dr., a glagolske vrste prema prezentskoj osnovi razvrstane su u Silić – Pranjković. (Isto: 216-217)

2.1. Hrvatske glagolske vrste u Hrvatskoj gramatici (Barić i suradnici)

Kao što je već navedeno glagoli se prema infinitivno–aoristnoj osnovi dijele na šest vrsta, odnosno sedam ako se nepravilni glagoli uzimaju kao posebna vrsta.

U prvoj vrsti infinitivno-aoristna osnova nema tvorbenog morfa, a dijeli se u sedam razreda. U prvom razredu jednake su prezentska i infinitivna osnova, a završavaju na *t* ili *d*. Infinitiv završava na *–sti*. Glagoli koje ubrajamo u prvi razred su: *bosti – bodem, cvasti – cvatem, jesti – jedem, krasti – kradem, dovesti – dovedem* itd. (Barić 1997: 248-249)

U drugom razredu jednake su prezentska i infinitivna osnova i završavaju na *s* i *z*. Infinitiv završava na *–sti*. Takvi su glagoli: *gristi – grizem, pasti – pasem, musti – muzem, vesti – vezem* itd. (Isto: 250-251)

U trećem razredu i prezentska i infinitivna osnova završavaju na *b* i *p*, a infinitiv završava na *–sti*. Primjeri takvih glagola su: *crpsti – crpem, dupsti – dubem, grepsti – grebem, zepsti – zebem* itd. (Isto: 251-252)

Infinitivna i prezentska osnova u četvrtom razredu završavaju na *k*, *g* i *h*. Infinitiv završava na *–ći*. To su glagoli: *peći – pečem, sjeći – siječem, strići – strižem* itd. (Isto: 252-253)

U petom razredu infinitivna osnova završava na *e* i *u*, a prezentska na *n* i *m*. Primjeri su glagoli: *kleti – kunem, oteti – otmem, uzeti – uzmem*. (Isto: 253-254)

Glagoli šestog razreda imaju u prezentskoj osnovi korijenske morfe na *r* i *lj*, a u infinitivnoj osnovi na *rije/re*. Takvi su glagoli: *klati – koljem, mljeti – meljem, zreti – zrem (zrim)* itd. (Isto: 254-255)

Naposljetku, u sedmom razredu i infinitivna i prezentska osnova završavaju na *i*, *u*, *je/i*. To su glagoli: *biti – bijem, šiti – šijem, obuti – obujem, smjeti – smijem, umjeti – umijem* itd. (Isto: 255-256)

Drugu vrstu čine glagoli čija infinitivna osnova završava na *nu*, a prezentska na *n*. To su glagoli: *tonuti – tonem* (u imperfektu se provodi alternacija *n>nj*: *tonjah*), *dignuti – dignem, venuti – venem*. (Isto: 256-257)

U trećoj vrsti prezentski nastavci imaju morf *i* (-*im*), a infinitiv ima nastavak –*eti* ili –*ati*. Treća se vrsta dijeli na dva razreda, u prvi razred idu glagoli na –*eti* s morfom *je* i glagoli s osnovom na *lje*, *nje*, *re*. U prezentskoj se osnovi događa alternacija *lj>l*, *nj>n*. (Isto: 257-258)

U drugom su razredu glagoli s morfom *a* čija osnova završava palatalima *č*, *ž*, *j*, te skupinama *št* i *žd*, a u prezentu imaju nastavke s alomorfom -*i*- (*čućiati* – *čućim*, *držati* – *držim*). (Isto: 258-259)

Četvrta vrsta ima infinitivnu osnovu na morf *i*, a infinitiv završava na –*iti* (*voziti* – *vozim*, *ljubiti* – *ljubim*). (Isto: 259-260)

Petu vrstu čine glagoli čija infinitivna osnova ima morf *a*, a infinitiv završava na –*ati*. Glagoli pete vrste razvrstavaju se u četiri razreda. U prvom su razredu glagoli čiji su prezentski nastavci s morfom *a* (*polijevati* – *polijevam*, *potkivati* – *potkivam*, *darivati* – *darivam*). (Isto: 260-265)

U drugom su razredu glagoli s morfom *e* u prezentskim nastavcima i s jotiranom osnovom u prezentu, imperativu i glagolskom prilogu sadašnjem (*lagati* – *lažem*, *micati* – *mičem*, *brisati* – *brišem*). (Isto: 265-266)

U trećem se razredu u prezentskim nastavcima nalazi morf -*e*- s time da se glagoli dijele u dvije skupine, u prvoj skupini korijen ima fonemski isti morf u objema osnovama (*derati* – *derem*, *orati* – *orem*), u drugoj su skupini glagoli kojima korijenski alomorf u infinitivnoj osnovi ima dva suglasnika, a u prezentskoj se osnovi oni proširuju samoglasnikom. Primjeri takvih glagola su: *brati* – *berem*, *prati* – *perem*). (Isto: 266-267)

U četvrtom su razredu glagoli s infiksima -*j(a)*- i -*v(a)*- u infinitivnoj osnovi, a infinitiv završava na –*jati*, –*vati*. Prezentska osnova nema sufiksa, a nastavci su osnovu počinju s –*j* (*sijati* – *sijem*, *kljuvati* – *klujem*). (Isto: 267-268)

U šestu vrstu ubrajamo glagole čija se infinitivna osnova tvori morfovima *ova*, *eva*, *iva*, a infinitiv završava na -*o/e/i-vati*, a prezentska osnova morfom *u*: *darovati* – *darujem*, *kazivati* – *kazujem*. (Isto: 269-271)

Nepravilni glagoli mogu tvoriti sedmu vrstu. To su glagoli koji tvore svoje oblike od supletivnih osnova: *biti* – *jesam* / *budem*, *htjeti* – *hoću*, *ići* – *idem*. (Isto: 271)

2.2. Hrvatske glagolske vrste u Gramatici hrvatskoga jezika (Silić – Pranjković)

U Silić – Pranjkovićevoj gramatici glagoli su podijeljeni u šest vrsta od kojih se svaka dijeli na nekoliko razreda, prva na osamnaest, druga na jedan, treća na tri, četvrta također na tri, dok peta i šesta imaju po dva razreda.

U prvoj vrsti infinitivna osnova završava nultim (-0-) sufiksanim morfemom, a prezentska osnova sufiksanim morfemom -e-. Između korijenskog i sufiksalnog morfema nalaze se umeci *n*, *d*, i *dn*. Glagoli se prve vrste dijele na osamnaest razreda, a neki od tih glagola su: *grebem – grebu*, *počnem – počnu*, *kunem – kunu*, *klati – koljem*, *tarem – taru*, *pospem – pospu*, *stadem – stadu*, *zrem – zru...* (Silić – Pranjković 2005: 42-43)

Infinitivna osnova u drugoj vrsti završava sufiksanim morfemom *nu*, a prezentska osnova sufiksanim morfemom *e*. Između prezentskog morfema *e* i korijenskog morfema intervenira umetak *n*. Druga vrsta ima samo jedan razred (-*nu-ti* / -*n-e-m*: *mrznuti – mrznem*). (Isto: 44)

Glagoli treće vrste imaju infinitivnu osnovu koja završava sufiksanim morfemom *a* i njegovim inačicama -*va*, -*ja* i prezentsku osnovu koja završava sufiksanim morfemom -*je*-. Završne suglasnike korijenskog morfema *d*, *t*, *g*, *k*, *c*, *h*, *z*, *s*, *b*, *p*, *m*, *v*, *sk* iz sufiksalnog morfema -*je*- jotira u *đ*, *ć*, *ž*, *č*, *š*, *ž*, *š*, *blj*, *plj*, *mlj*, *vlj*, *šč/št*. Infinitivno *ov* ispred *a* alternira s prezentskim *u*, infinitivno *ev* ispred *a* s prezentskim *u* i infinitivno *iv* ispred *a* s prezentskim *u*. U skladu s time treća se vrsta dijeli na tri razreda (*glodati – glođem – glođu*, *kljuvati – kljujem – kljuju*, *grijati – grijem – griju*). (Isto: 44)

Glagoli četvrte vrste imaju infinitivnu osnovu koja završava sufiksanim morfemom -*i*-, -*je*- i -*a*-, te prezentsku osnovu koja završava sufiksanim morfemom -*i*-. Korijenski morfem ispred sufiksalnog morfema -*a*- završava na *č*, *ž*, *št* i *žd*. Stoga se četvrta vrsta dijeli na tri razreda (*raditi – radim – rade*, *vidjeti – vidim – vide*, *bježati – bježim*). (Isto: 44-45)

Glagolima pete vrste infinitivna osnova završava sufiksanim morfemom -*a*-, a prezentska osnova sufiksanim morfemom -*ā*-. U nekih se glagola ispred sufiksanih morfema -*a*- i -*ā*- pojavljuje *av* pa se stoga peta vrsta dijeli na dva razreda (*kopati – kopam – kopaju*, *proučavati – proučavam – proučavaju*). (Isto: 45)

Glagoli šeste vrste imaju infinitivnu osnovu koja se završava sufiksanim morfemima –ova- / -eva- i –iva-, te prezentsku osnovu koja se završava sufiksanim morfemom –je-. U skladu s time glagoli se šeste vrste dijele na dva razreda (*kupovati* – *kupujem* – *kupuju*, *smanjivati* – *smanjujem* – *smanjuju*). (Isto: 45)

Na samom kraju Silić i Pranjković navode glagole tipa *dati* i *spati* za koje govore kako se nalaze izvan svih vrsta i razreda i ne ubrajaju ih u neku sedmu vrstu kao što je to u *Hrvatskoj gramatici* Eugenije Barić i suradnika.

3. GLAGOL UZIMATI U KONTEKSTU HRVATSKIH KONJUGACIJSKIH VRSTA

Danas je moguće vidjeti 1. lice prezenta glagola *uzimati* u obliku *uzimam*, ali i u obliku *uzimljem*. Dakle, mogući su i oblici po vrsti V. 1. i po vrsti V. 2. Ivan Marković u svom *Uvodu u jezičnu morfologiju* piše da glagoli „bježe“ iz jedne vrste u drugu, što bi značilo da su oblici po vrsti V. 1. stariji, a po vrsti V. 2. mlađi.

Glagol *uzimati* je, znači, u kontekstu konjugacijske vrste V.2 izgledao ovako: *uzim-a-ti* / *uzim-lje-m*. Sada se ustalilo korištenje glagola *uzimam* pa glagol *uzimati* počinjemo promatrati unutar konjugacijske vrste V.1 gdje on izgleda ovako: *uzim-a-ti* / *uzim-a-m*. Zbog toga Ivan Marković u *Uvodu u jezičnu morfologiju* govori o „bježanju“ glagola iz jedne vrste u drugu, u našem slučaju iznimno plodnu konjugacijsku vrstu V.1 što je slučaj u gramatici Eugenije Barić i suradnika.

Isti se pomak događa i u Silić – Pranjkovićevoj gramatici. Jedina je razlika što se glagol *uzimati* nalazio u konjugacijskoj vrsti III.1 (*uzim-a-ti* / *uzim-lje-m*), a sada ga smještamo u konjugacijsku vrstu V.1 (*uzim-a-ti* / *uzim-a-m*).

U petu se konjugacijsku vrstu uključuju i glagoli s posuđenim osnovama koji su nedavno ušli u hrvatski jezik. Neki od tih glagola su sljedeći: *surf-a-ti* / *surf-a-m* (*surf*), *gugl-a-ti* / *gugl-a-m* (*Google*), *fejk-a-ti* / *fejk-a-m* (*fake*), *fejs-a-ti* / *fejs-a-m* (*Facebook*), *čil-a-ti* / *čil-a-m* (*chill*) i brojni drugi. (Marković 2012: 223-224)

4. HRVATSKI RAČUNALNI KORPUSI I HNK v3.0

4.1. Računalna obrada hrvatskih korpusa

Marko Tadić u članku *Računalna obrada hrvatskih korpusa: povijest, stanje i perspektive*¹ želi dati pregled projekata u Republici Hrvatskoj kojima je metodologija ona iz domene korpusne lingvistike. Većina je tih projekata smještena u *Zavodu za lingvistiku Filozofskoga fakulteta Sveučilišta u Zagrebu*.

„Prvi je hrvatski korpus, veličine 100 000 pojava, sastavio je frekvencijski obradio psiholog Ivan Furlan u svojoj disertaciji *Raznolikost rječnika. Struktura govora*.“²

Prvi je računalno obrađeni korpus bio barokni ep *Osman* Ivana Gundulića koji je frekvencijski obradio i konkordancijama popratio Željko Bujas 1967. godine. Ubrzo je nakon toga Željko Bujas pripremio i *Povratak Filipa Latinovicza* Miroslava Krleže te *Suzanu* Marka Marulića.³

1968. godine u *Zavodu za lingvistiku Filozofskoga fakulteta Sveučilišta u Zagrebu* započinje veliki kontrastivni projekt čiji je voditelj bio Rudolf Filipović pod nazivom *Yugoslav Serbo-Croatian – English Contrastive Project* u kojem se prvi puta kod nas počinje računalna obradba korpusa. „Projekt je trajao do 1971. i osim izravnih svjetskih dostignuća u kontrastivistici, rezultirao je još za svoga trajanja i snažnim metodološkim utjecajem na druge projekte u Zavodu koji su počeli usvajati računalnu obradu korpusa kao standardnu istraživačku proceduru. Sama je obrada uvijek, ispravno, bila smatrana prije svega sredstvom obrade jezične građe, a ne ciljem.“⁴

1968. godine u Zavodu je pokrenut projekt nazvan *Jezik Marka Marulića* čiji je voditelj bio Milan Moguš. Dvije je godine poslije projekt proširen i preimenovan u

¹ Tadić, Marko, *Računalna obradba hrvatskih korpusa: povijest, stanje i perspektive*, 1997., <http://www.hnk.ffzg.hr/txts/mt4hmk3.pdf> (posjećeno: 22.7.2016.)

² Tadić, Marko, *Računalna obradba hrvatskih korpusa: povijest, stanje i perspektive*, 1997., <http://www.hnk.ffzg.hr/txts/mt4hmk3.pdf> (posjećeno: 22.7.2016.)

³ Tadić, Marko, *Računalna obradba hrvatskih korpusa: povijest, stanje i perspektive*, 1997., <http://www.hnk.ffzg.hr/txts/mt4hmk3.pdf> (posjećeno: 22.7.2016.)

⁴ Tadić, Marko, *Računalna obradba hrvatskih korpusa: povijest, stanje i perspektive*, 1997., <http://www.hnk.ffzg.hr/txts/mt4hmk3.pdf> (posjećeno: 22.7.2016.)

Kompjutorska analiza stare hrvatske književnosti i s njime počinje pravi zamah korpusne obrade hrvatskih tekstova.⁵

Do 1981. konkordirana su sva hrvatska djela Marka Marulića, djela Barne Karnarutića, *Planine* Petra Zoranića, *Jejupka* Mikše Pelegrinovića, djela Hanibala Lucića i Petra Hektorovića, *Hvarkinja* Martina Benetovića, *Ranjinin zbornik*, Držićeve komedije, djela Ivana Bunića Vučića, Vitezovićeve djela, *Sveta Rožalija* Antuna Kanižlića, komedije Tituša Brezovačkoga i *Razvod istarski*, ali i *Balade Petrice Kerempuha* Miroslava Krleže.⁶

Dobiveni su rezultati korišteni za nova kritička čitanja, potvrđivanje autorstva gdje je ono bilo dvojbeno, ali i za istraživanja na svim jezičnim razinama.

U razdoblju od 1972. do 1975. godine trajao je projekt *Englesko-hrvatski leksikografski korpus* čiji je voditelj bio Željko Bujas. Polazna je ideja projekta bila tretirati tekst dvojezičnog a rječnika kao korpus.⁷

Ubrzo se pojavila potreba za korpusom koji bi bio reprezentativan za suvremeni hrvatski jezik pa je 1976. godine pod vodstvom Milana Moguša pokrenut projekt nazvan *Korpus suvremenog hrvatskog književnog jezika* čiji je cilj bio sastaviti jednomilijunski korpus koji je nazvan Mogušev korpus. Stjecajem okolnosti korpus je završen 1996. godine, a to je bio „prvi pokušaj u hrvatskoj lingvistici da se na temeljima reprezentativnoga korpusa započne usustavljivanje i istraživanje jezične građe“. Najznačajnijim rezultatom obrade Moguševa milijunskog korpusa smatra se *Hrvatski čestotni rječnik*.⁸

1989. godine u Dubrovniku je organizirana konferencija *Language industries – Needs and Perspectives* na kojoj su se prvi put skupili stručnjaci iz zapadnoeuropskih,

⁵ Tadić, Marko, Računalna obradba hrvatskih korpusa: povijest, stanje i perspektive, 1997., <http://www.hnk.ffzg.hr/txts/mt4hmk3.pdf> (posjećeno: 22.7.2016.)

⁶ Tadić, Marko, Računalna obradba hrvatskih korpusa: povijest, stanje i perspektive, 1997., <http://www.hnk.ffzg.hr/txts/mt4hmk3.pdf> (posjećeno: 22.7.2016.)

⁷ Tadić, Marko, Računalna obradba hrvatskih korpusa: povijest, stanje i perspektive, 1997., <http://www.hnk.ffzg.hr/txts/mt4hmk3.pdf> (posjećeno: 22.7.2016.)

⁸ Tadić, Marko, Računalna obradba hrvatskih korpusa: povijest, stanje i perspektive, 1997., <http://www.hnk.ffzg.hr/txts/mt4hmk3.pdf> (posjećeno: 22.7.2016.)

srednjoeuropskih i istočnoeuropskih zemalja. Tada je prvi put uspostavljena značajnija suradnja Zavoda sa zapadnoeuropskim projektima i centrima u razmjeni stručnjaka i ideja.⁹

„U travnju 1996. pod vodstvom Vesne Muhvić-Dimanovski pri Ministarstvu znanosti i tehnologije prijavljen je projekt *Računalna obradba hrvatskoga jezika* kojem je jedan od primarnih zadataka sastavljanje više desetaka milijuna velikoga korpusa hrvatskoga jezika.“ Predloženo je da se korpus, po uzoru na neke strane korpuse, nazove *Hrvatski nacionalni korpus* (HNK).¹⁰

Hrvatski nacionalni korpus sastoji se od dvije sastavnice. Prva je *30-milijunski korpus suvremenoga hrvatskoga jezika* (30M) kao reprezentativni, uravnoteženi korpus s tekstovima nastalim 1990. ili kasnije, a druga *Hrvatski tekstovni elektronski arhiv* (HETA) koji čini neuravnotežena zbirka korpusa s tekstovima starijim od 1990. ili tekstovima koji ne odgovaraju reprezentativnosti 30M, a sami su dovoljno veliki.¹¹

⁹ Tadić, Marko, *Računalna obradba hrvatskih korpusa: povijest, stanje i perspektive*, 1997., <http://www.hnk.ffzg.hr/txts/mt4hmk3.pdf> (posjećeno: 22.7.2016.)

¹⁰ Tadić, Marko, *Računalna obradba hrvatskih korpusa: povijest, stanje i perspektive*, 1997., <http://www.hnk.ffzg.hr/txts/mt4hmk3.pdf> (posjećeno: 22.7.2016.)

¹¹ Tadić, Marko, *Računalna obradba hrvatskih korpusa: povijest, stanje i perspektive*, 1997., <http://www.hnk.ffzg.hr/txts/mt4hmk3.pdf> (posjećeno: 22.7.2016.)

4.2. Počeci računalne obradbe hrvatskoga i nacionalni korpus

Istraživanje prirodnoga jezika u mnogome nam olakšava računalna tehnologija, imajući pritom na umu kako je nama računalno pretraživanje korpusa sasvim uobičajena aktivnost, za razliku od 1996. godine kada je Marko Tadić tek počeo naglašavati važnost računalne obradbe hrvatskoga jezika.

Istraživanja s područja obradbe hrvatskoga jezika moraju imati status fundamentalnoga istraživanja u humanističko-društvenim znanostima te status strateškoga istraživanja za Republiku Hrvatsku.¹² Za razliku od nekih drugih istraživanja koja imaju status fundamentalnoga istraživanja, na primjer istraživanje subatomske čestice, začudno je kako takvo istraživanje i pristup izostaju na području humanističko-društvenih znanosti. Tu ulogu tako preuzima računalna obradba prirodnoga jezika.¹³

Temelj svakoga istraživanja teksta je korpus pa je korpusna lingvistika određena metodološkom osnovicom na kojoj se temelji istraživanje. Korpusna se metodologija može primjenjivati u različitim lingvističkim disciplinama.¹⁴ S obzirom na to da se danas uvid u korpus ne može niti zamisliti bez pomoći računala valja razlikovati *predračunalnu korpusnu lingvistiku* čija se teorijska osnovica nalazi u američkih deskriptivista i *računalnu korpusnu lingvistiku*.¹⁵

Računala se mogu koristiti i za provjeru istraživačkih hipoteza pa zbog toga njihova uporaba u lingvistici mora steći status kakav ima u prirodnim znanostima. Računalo stoga služi kao alat za prikupljanje i obradu istraživačkih podataka te provjeru istraživačkih

¹² Tadić, Marko, Računalna obradba hrvatskoga i nacionalni korpus, 1996., <http://www.hnk.ffzg.hr/txts/mt4hmk.pdf> (posjećeno: 22.7.2016.)

¹³ Tadić, Marko, Računalna obradba hrvatskoga i nacionalni korpus, 1996., <http://www.hnk.ffzg.hr/txts/mt4hmk.pdf> (posjećeno: 22.7.2016.)

¹⁴ Tadić, Marko, Računalna obradba hrvatskoga i nacionalni korpus, 1996., <http://www.hnk.ffzg.hr/txts/mt4hmk.pdf> (posjećeno: 22.7.2016.)

¹⁵ Tadić, Marko, Računalna obradba hrvatskoga i nacionalni korpus, 1996., <http://www.hnk.ffzg.hr/txts/mt4hmk.pdf> (posjećeno: 22.7.2016.)

hipoteza. Primijenimo li napisano na lingvistiku, istraživačke podatke čini jezična građa, a računalni modeli odgovaraju modelima dijelova jezičnih podsustava.¹⁶

„Ako je računalo potrebno pri skupljanju i obradbi jezične građe, a to je valjda danas već neupitno, onda se računalna obradba hrvatskoga u potpunosti uklapa u nacionalni istraživački prioritet istraživanja hrvatskoga jezika.“¹⁷

Zbog česte uporabe računala u svim područjima ljudskoga života javila se i potreba za jezičnim alatima koji su pomagala jezikoslovcima u istraživanju jezika informatičarima u dizajniranju i programiranju sustava za obrade teksta, ali i svakodnevnim korisnicima računala pri manipulaciji tekstem. Takvi se jezični alati izrađuju za svaki jezik posebno što znači da svaka jezična zajednica mora pokrenuti projekte čiji će rezultat biti upotrebljivi jezični alati.¹⁸ Situacija hrvatskoga jezika u tom smislu nije jednostavna jer dok se drugi, mnogoljudniji jezici financiraju zahvaljujući zainteresiranosti velikoga tržišta, hrvatskom je jeziku jedino rješenje pomoć jezične zajednice, odnosno države koja bi trebala pokriti većinu troškova tih projekata.¹⁹

Zbog svega toga, navodi Marko Tadić, kako računalna obradba hrvatskoga mora imati status strateškoga istraživanja za Republiku Hrvatsku. Sastavljanje korpusa moralo je biti primaran zadatak istraživanja hrvatskoga jezika s jedne strane zbog nedostatka leksikonskih podataka, a kako bi se nedostatak uklonio korpus je morao biti sastavljen tako da se na njegovoj osnovi mogu izraditi veliki leksikoni i za ljude kao korisnike i za strojeve, s druge strane razlog za postavljanje korpusa u ishodište računalne obrade hrvatskoga jezika je zahtjevnost njegova sastavljanja pa se javlja potreba za njegovom višestrukom iskoristivošću i višesvrhovitošću.²⁰

¹⁶ Tadić, Marko, Računalna obradba hrvatskoga i nacionalni korpus, 1996., <http://www.hnk.ffzg.hr/txts/mt4hmk.pdf> (posjećeno: 22.7.2016.)

¹⁷ Tadić, Marko, Računalna obradba hrvatskoga i nacionalni korpus, 1996., <http://www.hnk.ffzg.hr/txts/mt4hmk.pdf> (posjećeno: 22.7.2016.)

¹⁸ Tadić, Marko, Računalna obradba hrvatskoga i nacionalni korpus, 1996., <http://www.hnk.ffzg.hr/txts/mt4hmk.pdf> (posjećeno: 22.7.2016.)

¹⁹ Tadić, Marko, Računalna obradba hrvatskoga i nacionalni korpus, 1996., <http://www.hnk.ffzg.hr/txts/mt4hmk.pdf> (posjećeno: 22.7.2016.)

²⁰ Tadić, Marko, Računalna obradba hrvatskoga i nacionalni korpus, 1996., <http://www.hnk.ffzg.hr/txts/mt4hmk.pdf> (posjećeno: 22.7.2016.)

4.3. Raspon, opseg i sastav korpusa hrvatskoga suvremenog jezika

Vremenski raspon korpusa određuje se na temelju razlike između najstarijega i najmlađeg teksta koji se nalaze u korpusu, iako korpus ne mora nužno imati obje granice, on može sezati od ili do neke točke.²¹

Ako se želi sastaviti korpus suvremenoga hrvatskog jezika valja krenuti od godine koja je važna osim za hrvatski jezik i za Hrvatsku kao državu i za Hrvate kao narod, a to je 1990. Naglašava pri tome Marko Tadić kako je hrvatski jezik postojao i puno prije te godine, ali bitno je da se korpus suvremenoga hrvatskog jezika ograniči i započne od neke određene točke u vremenu.

Sve što je u hrvatskome jeziku nastalo prije 1990. godine uključuje se u dio korpusa koji se zove Zbirka hrvatskih testova koja je dio Hrvatskoga nacionalnoga korpusa (HNK) isto kao i Korpus suvremenoga hrvatskoga jezika.²²

Opseg korpusa ovisi o njegovoj namjeni. Namjera projekta *Računalna obradba Hrvatskoga* u Zavodu za lingvistiku Filozofskoga fakulteta Sveučilišta u Zagrebu bila je sastaviti korpus koji je mogao ući u mrežu europskih referentnih korpusa, a kako je sastavljanje složen i skup pothvat trebalo ga je zamisliti i organizirati kao višesvrhovito pomagalo koje bi služilo većem broju istraživača.²³

Iako su generacije korpusa stranih jezika rasle od milijun pojava, preko desetak milijuna pa sve do stotinjak milijuna pojava, za prvi hrvatski korpus to su bili preveliki brojevi pa se je bilo potrebno ograničiti na korpus ne veći od 30 milijuna pojava.²⁴

Ako se želi sastaviti korpus koji će biti reprezentativan za neki jezik moraju se pri sastavljanju imati na umu različita područja uporabe jezika, tipovi i dužina tekstova, žanrovi,

²¹ Tadić, Marko, Raspon, opseg i sastav korpusa hrvatskoga suvremenog jezika, 1998., <http://www.hnk.ffzg.hr/txts/mt4hmk2.pdf> (posjećeno: 22.7.2016.)

²² Tadić, Marko, Raspon, opseg i sastav korpusa hrvatskoga suvremenog jezika, 1998., <http://www.hnk.ffzg.hr/txts/mt4hmk2.pdf> (posjećeno: 22.7.2016.)

²³ Tadić, Marko, Raspon, opseg i sastav korpusa hrvatskoga suvremenog jezika, 1998., <http://www.hnk.ffzg.hr/txts/mt4hmk2.pdf> (posjećeno: 22.7.2016.)

²⁴ Tadić, Marko, Raspon, opseg i sastav korpusa hrvatskoga suvremenog jezika, 1998., <http://www.hnk.ffzg.hr/txts/mt4hmk2.pdf> (posjećeno: 22.7.2016.)

medij ostvarivanja jezične poruke, autorove osobine te vrijeme nastanka teksta. Prema idealnoj strukturi svaki bi se korpus trebao dijeliti na određeni broj potkorpusa koji se onda dijele na sastavnice koje se pak dijele na uzorke, s time da su sastavnice i uzorci odsječci tekstova koji više nisu korpusi.²⁵

Za korpus standardnoga hrvatskog jezika treba u obzir uzeti i neka ograničenja kao što su pisani jezik, tekstovi izvornih hrvatskih govornika, a ne prijevodi, razdoblje od 1990., tekst stvoren u stvarnoj komunikaciji, a ne drama, proza a ne poezija, jezik osoba starijih od šesnaest godina te standardni hrvatski, a ne narječja.²⁶

Popunjenost uzoraka treba planirati na temelju rezultata dvaju jednostavnih istraživanja, a to su ispitivanje zastupljenosti muških, ženskih i skupnih autora u novinama, časopisima i katalozima pa se prema tim rezultatima raspoređuje autorska zastupljenost u korpusu te ispitivanje čitanosti pojedinih knjiga i na taj način vrednovanje njihove kandidature za ulazak u korpus.²⁷

„Za postavljanje strukture korpusa najlakši bi kriterij na prvoj razini mogao biti medij ostvarivanja, potom tip teksta, a ostale bi razine slijedili ostali kriteriji.“²⁸ Tako bi korpus standardnoga hrvatskog jezika obuhvaćao pisani tekst koji se sastoji od knjiga, novina, časopisa, brošura i slično te govoreni tekst koji može biti formalni i neformalni.

²⁵ Tadić, Marko, Raspon, opseg i sastav korpusa hrvatskoga suvremenog jezika, 1998., <http://www.hnk.ffzg.hr/txts/mt4hmk2.pdf> (posjećeno: 22.7.2016.)

²⁶ Tadić, Marko, Raspon, opseg i sastav korpusa hrvatskoga suvremenog jezika, 1998., <http://www.hnk.ffzg.hr/txts/mt4hmk2.pdf> (posjećeno: 22.7.2016.)

²⁷ Tadić, Marko, Raspon, opseg i sastav korpusa hrvatskoga suvremenog jezika, 1998., <http://www.hnk.ffzg.hr/txts/mt4hmk2.pdf> (posjećeno: 22.7.2016.)

²⁸ Tadić, Marko, Raspon, opseg i sastav korpusa hrvatskoga suvremenog jezika, 1998., <http://www.hnk.ffzg.hr/txts/mt4hmk2.pdf> (posjećeno: 22.7.2016.)

4.4. Zašto nam je potreban višemilijunski referentni korpus?

U radu nazvanom *Zašto nam je potreban višemilijunski referentni korpus* Marko Tadić u samom uvodu postavlja bitno terminološko razgraničenje. Tadić želi objasniti kako se termin *korpus* neće koristiti u tradicionalnom, općelingvističkom smislu, već se misli na korpus onakav kakvim ga definira korpusna lingvistika, a to je računalom podržan korpus.

Opis jezika u cijelosti nije moguć zbog metodoloških ograničenja pa se jeziku prilazi preko njegovih ostvaraja. Također, ne mogu se istraživati sve realizacije nekog jezika već je potrebno sastaviti korpus koji reprezentira realnost određenih jezičnih varijeteta ili jezik u cjelini. Sastavljanje korpusa koji će biti referentan za neki jezik nije jednostavan projekt pa se korpus mora graditi tako da se ima u vidu što veći broj korištenja tog istog korpusa u budućnosti.²⁹

Većina, Tadićevim riječima, „ozbiljnih“ europskih jezika posjeduje ili gradi takve korpuse sa višesvrhovitom primjenom. Razlikuje Tadić dvije skupine primjena, prvi je korpus koji služi kao referentni za neki jezik, a drugi je korpus kao leksikografsko pomagalo. Referentni korpus tako služi kao „materijal za istraživanja raznolikosti na svim jezičnim razinama“³⁰. Rezultati koje korpus nudi su dvojaki. S jedne strane oni su inventarskog tipa što podrazumijeva kako neki jezični entitet ili postoji ili ne postoji u tom korpusu, a s druge strane rezultati su čestotnoga tipa što znači da neki jezični entitet postoji s određenom zastupljenošću u tom korpusu, a s obzirom na to da je korpus reprezentativan za jezik u cjelini onda taj jezični entitet postoji i s određenom zastupljenošću i tom jeziku. Mi takav korpus nemamo, ali Tadić piše kako nam je jedan takav korpus potreban jer bismo njime dobili osnovne, polazne vrijednosti.

Promatramo li korpus kao leksikografsko pomagalo uočljivo je da se i tu radi o inventariziranju i prebrojavanju jezičnih jedinica „na tradicionalno omeđenom području s relativno strogo definiranom metodologijom“³¹. I prije su se rječnici radili pomoću korpusa,

²⁹ Tadić, Marko, *Zašto nam je potreban višemilijunski referentni korpus?*, 1990., <http://www.hnk.ffzg.hr/txts/mt4hdpl1990.pdf> (posjećeno: 22.7.2016.)

³⁰ Tadić, Marko, *Zašto nam je potreban višemilijunski referentni korpus?*, 1990., <http://www.hnk.ffzg.hr/txts/mt4hdpl1990.pdf> (posjećeno: 22.7.2016.)

³¹ Tadić, Marko, *Zašto nam je potreban višemilijunski referentni korpus?*, 1990., <http://www.hnk.ffzg.hr/txts/mt4hdpl1990.pdf> (posjećeno: 22.7.2016.)

ali taj su korpus činili drugi rječnici koje su autori konzultirali i natuknice iz njih uvrštavali su u novi rječnik više ili manje prerađene. Unatoč tome smatralo se potrebnim napraviti rječnik na osnovi korpusa u koji će ulaziti samo one riječi i značenja koja su potvrđena u tom korpusu pa su od presudne važnosti veličina korpusa, kriterij biranja žanrovskih i tematskih područja i principi uzorkovanja.³²

³² Tadić, Marko, Zašto nam je potreban višemilijunski referentni korpus?, 1990., <http://www.hnk.ffzg.hr/txts/mt4hdpl1990.pdf> (posjećeno: 22.7.2016.)

4.5. Hrvatski nacionalni korpus (HNK)

„Hrvatski nacionalni korpus (HNK) usustavljena je zbirka odabranih tekstova pretežito suvremenoga hrvatskoga jezika koji pokrivaju razne medije, žanrove, stilove, područja i tematiku. Sâm je korpus, popraćen dodatnim lingvističkim i nelingvističkim podacima, smješten u datobazu kojoj se pristupa s pomoću programa za pretraživanje Bonito.“³³ Hrvatski nacionalni korpus je javno i slobodno dostupan za istraživanje, obrazovanje i ostale nekomercijalne uporabe, dok se komercijalni korisnici moraju predbilježiti kako bi dobili korisnički račun.

Trenutno HNK v3.0 sadrži 216,8 milijuna pojava.³⁴

Zavod za lingvistiku Filozofskoga fakulteta Sveučilišta u Zagrebu kao referentna institucija za korpusna istraživanja hrvatskoga jezika 1996. godine počeo je sa sastavljanjem *Hrvatskoga nacionalnoga korpusa (HNK)* po uzoru na neke strane nacionalne korpusne kao što je Britanski nacionalni korpus, Češki nacionalni korpus i brojni drugi.³⁵

Hrvatski nacionalni korpus sastavlja se od odabranih tekstova na hrvatskome jeziku iz svih područja, struka, žanrova i stilova: od književnih i znanstvenih djela do udžbenika, od novinskih tiskovina do mrežnih novina.³⁶

Hrvatski nacionalni korpus složen je od dvije sastavnice. Prva je *30-milijunski korpus suvremenoga hrvatskoga jezika* (30M) kao reprezentativni, uravnoteženi korpus s tekstovima nastalim 1990. ili kasnije, a druga *Hrvatski tekstovni elektronski arhiv* (HETA) koji čini neuravnotežena zbirka korpusa s tekstovima starijim od 1990. ili tekstovima koji ne odgovaraju reprezentativnosti 30M, a sami su dovoljno veliki.³⁷

³³ Hrvatski nacionalni korpus, <http://www.hnk.ffzg.hr/> (posjećeno: 29.7.2016.)

³⁴ Hrvatski nacionalni korpus, <http://www.hnk.ffzg.hr/> (posjećeno: 29.7.2016.)

³⁵ Tadić, Marko, Hrvatski računalni korpusi, <http://www.matica.hr/vijenac/158/Hrvatski%20ra%C4%8Dunalni%20korpusi/> (Posjećeno: 19.7.2016)

³⁶ Tadić, Marko, Hrvatski računalni korpusi, <http://www.matica.hr/vijenac/158/Hrvatski%20ra%C4%8Dunalni%20korpusi/> (Posjećeno: 19.7.2016)

³⁷ Tadić, Marko, Hrvatski računalni korpusi, <http://www.matica.hr/vijenac/158/Hrvatski%20ra%C4%8Dunalni%20korpusi/> (Posjećeno: 19.7.2016)

4.6. Hrvatska jezična riznica

Hrvatska jezična riznica projekt je Instituta za hrvatski jezik i jezikoslovlje koji je Ministarstvo znanosti, obrazovanja i športa odobrilo u svibnju 2005.³⁸

Temeljni je zadatak toga projekta stvaranje javno dostupnih internetskih resursa za hrvatski jezik, pružanje temeljnih obavijesti o hrvatskoj jezičnoj normi te ostalih obavijesti o hrvatskome jeziku.³⁹

U sklopu projekta izrađuje se nekoliko korpusa različitih razvojnih faza hrvatskoga jezika i digitaliziraju se objavljeni i rukopisni rječnici hrvatskoga jezika, prikuplja se i digitalizira građa za reprezentativni korpus hrvatskoga standardnog jezika koji je podloga za izradu Velikoga rječnika hrvatskoga jezika i druga jezikoslovna istraživanja.⁴⁰

Korpus hrvatskoga jezika prikuplja se od odabranih tekstova hrvatskoga jezika svih struka i funkcionalnih usmjerenja. Obuhvaća književne i druge pisane izvore iz razdoblja od konačnoga oblikovanja hrvatske standardnojezične norme u drugoj polovici 19. stoljeća do danas.⁴¹

Korpus hrvatskoga standardnog jezika sadrži temeljna djela hrvatske književnosti, publicistička djela, znanstvena djela raznih struka i sveučilišne udžbenike, osnovnoškolske i srednjoškolske udžbenike, prijevodnu literaturu, dnevni, tjedni i mjesečni tisak te knjige iz predstandardnoga razdoblja hrvatskoga jezika, a koje su prilagođene današnjemu hrvatskom standardnom jeziku.⁴²

³⁸ Hrvatska jezična riznica, <http://riznica.ihj.hr/index.hr.html> (posjećeno: 19.7.2016.)

³⁹ Hrvatska jezična riznica, <http://riznica.ihj.hr/index.hr.html> (posjećeno: 19.7.2016.)

⁴⁰ Hrvatska jezična riznica, <http://riznica.ihj.hr/index.hr.html> (posjećeno: 19.7.2016.)

⁴¹ Hrvatska jezična riznica, <http://riznica.ihj.hr/index.hr.html> (posjećeno: 19.7.2016.)

⁴² Hrvatska jezična riznica, <http://riznica.ihj.hr/index.hr.html> (posjećeno: 19.7.2016.)

5. METODOLOGIJA ISTRAŽIVANJA

U ovom će dijelu završnoga rada biti uspoređene frekventnosti pojavljivanja prezentskih oblika za svako lice glagola *uzimati* prema konjugacijskoj vrsti V.1 i konjugacijskoj vrsti V.2. Oblici konjugacijske vrste V.2 su: *uzimljem*, *uzimlješ*, *uzimlje*, *uzimljemo*, *uzimljete*, *uzimlju*, a oblici konjugacijske vrste V.1 su: *uzimam*, *uzimaš*, *uzima*, *uzimamo*, *uzimate*, *uzimaju*.

5.1. Struktura istraživanja

Na samom početku usporedit ćemo kolokacije za 1. lice jednine riječi *uzimljem* i *uzimam*. Pretraživanjem korpusa⁴³ HNK v3.0 vidljivo je kako kao rezultat pretraživanja riječi *uzimljem* nismo dobili niti jednu kolokaciju te riječi, za razliku od riječi *uzimam* za koju smo dobili 133 kolokacije te riječi. Neki od primjera uporabe riječi *uzimam* dobiveni pretraživanjem jesu: „(...) to samo **uzimam** kao primjer (...)“, „**Uzimam** život u svoje ruke (...)“, „(...) ali ja sebi ne **uzimam** za pravo komentirati ga (...)“, „(...) to nikomu ne **uzimam** za zlo (...)“ itd.

Nastavljamo usporedbu ostalih parova prezentskih oblika glagola *uzimati*. Slijedi par za 2. lice jednine, a to je par *uzimlješ* iz konjugacijske vrste V.2 i *uzimaš* iz konjugacijske vrste V.1. Na temelju pretraživanja korpusa HNK v3.0 utvrdili smo kako se oblik *uzimlješ* ne pojavljuje niti jednom, dok se oblik *uzimaš* pojavljuje 7 puta i to u ovim primjerima: „(...) to znači da **uzimaš** novac od nekog (...)“, „Bože, zašto k sebi uvijek **uzimaš** krive?“, „(...) **uzimaš** 'zdravo za gotovo' (...)“, „(...) **uzimaš** olovku i papir sa stola (...)“ i slično.

Par kojim označavamo 3. lice jednine je *uzimlje* iz konjugacijske vrste V.2 i *uzima* iz konjugacijske vrste V.1. Nakon pretraživanja korpusa HNK v3.0 rezultati upućuju na to kako se oblik *uzimlje* pojavljuje 3 puta i to u ovim primjerima: „(...) Nacional **uzimlje** kao razlog za ukidanje Vjesnika?“, „(...) sve drugo što zbrke i različitosti **uzimlje** ozbiljno (...)“. Oblik *uzima* pojavljuje se 6 665 puta u primjerima kao što su: „Obično se zbog nepouzdanosti ne **uzima** u obzir (...)“, „(...) takav mogući razvoj događaja ne **uzima** u obzir ni povijest ni

⁴³ http://filip.ffzg.hr/cgi-bin/run.cgi/first_form (posjećeno: 19.7.2016.)

psihologiju“, „Europska unija još **uzima** kao opravdanje (...)“, „Taj datum se **uzima** kao početak nove (...)“.

Nakon što smo usporedili oblike glagola *uzimati* za jedninu, usporedit ćemo i oblike za množinu prezenta.

Za 1. lice množine uspoređujemo par *uzimljemo* iz konjugacijske vrste V.2 i *uzimamo* iz konjugacijske vrste V.1. Rezultat pretraživanja dokazuje kako se oblik *uzimljemo* ne pojavljuje niti jednom, dok se oblik *uzimamo* pojavljuje 115 puta. Potvrđujemo to sljedećim primjerima: „Pritom kao prvu pretpostavku **uzimamo** (...)“, „Vrlo često za primjer **uzimamo** zapadne medije (...)“, „(...) te **uzimamo** uzorke za DNA analizu“, „Antibiotike **uzimamo** kao bombone (...)“.

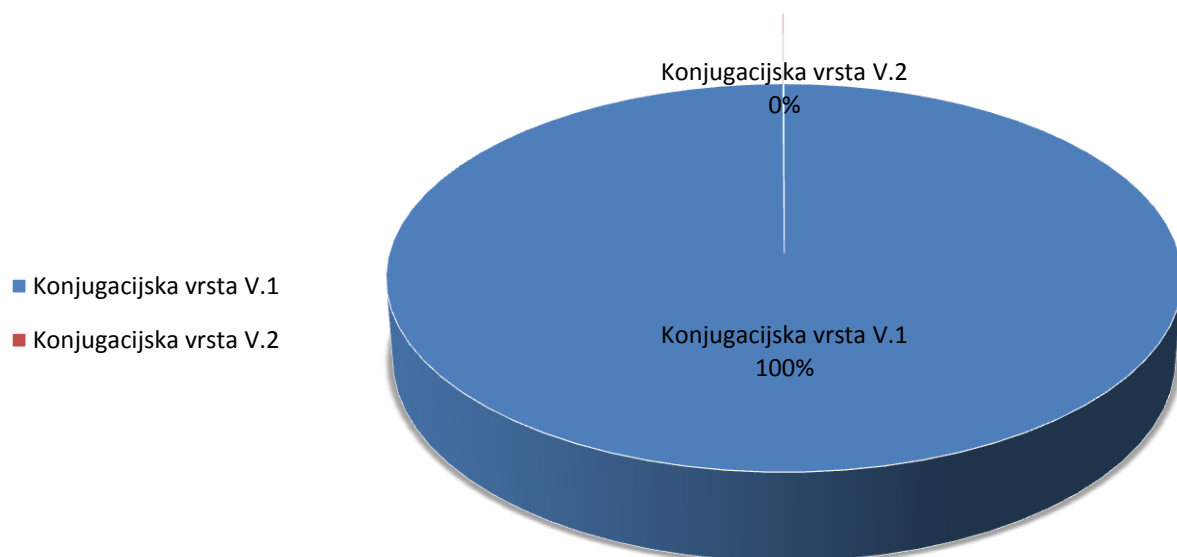
Par kojim označavamo 2. lice množine je oblik *uzimljete* iz konjugacijske vrste V.2 i oblik *uzimate* iz konjugacijske vrste V.1. Na temelju rezultata dobivenih pretraživanjem vidimo kako se oblik *uzimljete* nije pojavio niti jednom, a oblik *uzimate* pojavio se 41 put što vidimo u navedenim primjerima: „Ako **uzimate** kredit s rokom otplate (...)“, „Jednima **uzimate** zemlju (...)“, „(...) za godinu osnivanja **uzimate** 1903.“, „(...) onda ih **uzimate** prije natjecateljske sezone (...)“.

Naposlijetku promatramo par za 3. lice množine. Oblik *uzimlju* iz konjugacijske vrste V.2 i oblik *uzimaju* iz konjugacijske vrste V.1. Rezultatima pretraživanja dokazano je kako se oblik *uzimlju* pojavljuje 3 puta i to u sljedećim primjerima: „(...) filozofske etike danas **uzimlju** za svoje teško (...)“, „(...) koje stalno 'kopni' jer ga **uzimlju** ljudi (...)“, „(...) dogovori o knjigama koje jedni od drugih **uzimlju** za prevođenje (...)“, dok se oblik *uzimaju* pojavljuje 3 173 puta. Neki od primjera koji to potvrđuju su: „(...) koji u obzir **uzimaju** potrošnju turista (...)“, „Pritom si **uzimaju** pravo da i one (...)“, „(...) gotovo uopće ne **uzimaju** u obzir (...)“, „(...) svjedoci smo pripremljenih radova koji sve više **uzimaju** maha (...)“ itd.

6. ANALIZA

6.1. Grafikon: usporedba uporabe oblika iz konjugacijske vrste V.1 i V.2

Usporedba uporabe oblika iz konjugacijske vrste V.1 i V.2



6.2. Analiza grafikona

Kao što je vidljivo u prethodnome poglavlju metodologija istraživanja, odnosno struktura istraživanja oblici prezenta glagola *uzimati* koje ubrajamo u konjugacijsku vrstu V.2 jesu *uzimljem, uzimlješ, uzimlje, uzimljemo, uzimljete, uzimlju*, dok su oblici prezenta za konjugacijsku vrstu V.1 *uzimam uzimaš, uzima, uzimamo, uzimate, uzimaju*.

Gore predloženim rezultatima pretraživanja korpusa HNK v3.0 dobili smo konkretne brojeve prikazivanja određenog oblika glagola *uzimati*.

Rezultati su pokazali kako su se oblici iz konjugacijske vrste V.2 pojavili ukupno 6 puta, s time da se pokazalo kako se koriste samo oblici za 3. lice jednine *uzimlje* i 3. lice množine *uzimlju*.

Potpuno su suprotni rezultati pretraživanja za konjugacijsku vrstu V.1. Oblici za prezent glagola *uzimati* iz ove konjugacijske vrste ukupno su se pojavili 10 134 puta s time da najveći dio toga rezultata zauzima oblik za 3. lice jednine *uzima* čak 6 665 pojava te oblik za 3. lice množine *uzimaju* s 3 173 pojava.

Prethodno prikazani grafikon, usporedba uporabe oblika iz konjugacijske vrste V.1 i V.2, pokazuje kako je uporaba oblika prezenta glagola *uzimati* iz konjugacijske vrste V.2 toliko zanemarivo malena da su rezultati pojavnosti tih oblika doslovno svedeni na 0%, dok su rezultati pojavnosti oblika iz konjugacijske vrste V.1 100% u uporabi.

Mogli bismo na temelju toga zaključiti kako se je glagol *uzimati* gotovo u potpunosti pomakao iz konjugacijske vrste V.2 u V.1.

7. ZAKLJUČAK

U završnom sam radu prikazala podjelu glagola u konjugacijske vrste u dvjema hrvatskim gramatikama i kriterije na temelju kojih se oni svrstavaju u određenu vrstu odnosno razred.

Naglasak je u radu bio na glagolu *uzimati* i cilj je bio dokazati da se je taj glagol preselio iz konjugacijske vrste V.2 u konjugacijsku vrstu V.1 prema gramatici Eugenije Barić i suradnika odnosno iz konjugacijske vrste III.1 u konjugacijsku vrstu V.1 prema Silić – Pranjkovićevoj gramatici.

Kako bih dokazala da se glagol *uzimati* uistinu preselio iz jedne konjugacijske vrste u drugu svoje sam istraživanje provodila u Hrvatskom nacionalnom korpusu (HNK) v3.0. pretraživala sam korpus i uspoređivala sam oblike za prezent glagola *uzimati* iz konjugacijskih vrsta V.1 i V.2.

Rezultat pretraživanja korpusa prikazala sam pomoću grafikona iz kojega je vidljivo kako su se oblici za prezent iz konjugacijske vrste V.2 pojavili svega 6 puta dok su se oblici prezenta iz konjugacijske vrste V.1 pojavili 10 134 puta. Analogno tim kvantitativnim podacima nastao je rezultat u postocima prikazan u grafikonu gdje je prikazano kako je zastupljenost oblika glagola *uzimati* iz konjugacijske vrste V.1 stopostotna, dok su rezultati pojavnosti oblika iz konjugacijske vrste V.2 doslovno svedeni na 0%.

Na temelju kvantitativnih podataka kasnije pretvorenih u postotke zaključujem kako se je glagol *uzimati* gotovo u potpunosti preselio iz konjugacijske vrste V.2, odnosno III.1 u konjugacijsku vrstu V.1. Gotovo u potpunosti samo zato što ipak kvantitativno postoji u korpusu 6 pojava oblika iz konjugacijske vrste V.2, odnosno III.1 iako je to grafički svedeno na nulu.

8. SAŽETAK

U hrvatskom se jeziku glagoli dijele u nekoliko konjugacijskih vrsta prema infinitivno-aoristnoj, odnosno prezentskoj osnovi. Razvrstavanjem glagola dobivamo šest, odnosno sedam vrsta s time da sedmu vrstu čine nepravilni glagoli.

Posebno je u ovom radu izdvojen glagol *uzimati* koji je nekada bio dio konjugacijske vrste V.2 u gramatici Eugenije Barić i suradnika, odnosno dio konjugacijske vrste III.1 u Silić – Pranjkovićevoj gramatici, a sada se njega ubraja u konjugacijsku vrstu V.1 u obje gramatike.

Bitnu ulogu u spoznaji o pomaku glagola *uzimati* iz jedne konjugacijske vrste u drugu imaju računalni korpusi, odnosno u ovom slučaju, za moje istraživanje korišten je Hrvatski nacionalni korpus (HNK) v3.0. Pretraživajući taj korpus dobila sam rezultate, odnosno frekventnost pojavljivanja prezentskih oblika za svako lice glagola *uzimati*. Uspoređivala sam tako frekventnost pojavljivanja oblika iz konjugacijske vrste V.1 (*uzimam, uzimaš, uzima, uzimamo, uzimate, uzimaju*) i konjugacijske vrste V.2 (*uzimljem, uzimlješ, uzimlje, uzimljemo, uzimljete, uzimlju*).

Usporedbom kolokacija za sva lica jednine i množine prezenta vidljivo je kako su brojke uvijek na strani konjugacijske vrste V.1 jer ukupno gledano pojavnih oblika iz konjugacijske vrste V.2, odnosno III.1 ima 6 i to samo u 3. licu jednine i 3. licu množine, dok oblika iz konjugacijske vrste V.1 ima 10 134, najviše u 3. licu jednine i 3. licu množine.

U skladu s time na kraju prikazani grafikon daje stopostotnu prednost oblicima iz konjugacijske vrste V.1 što znači da se glagol *uzimati* gotovo u potpunosti preselio u konjugacijsku vrstu V.1.

9. KLJUČNE RIJEČI

Popis ključnih riječi u završnome radu:

- glagol
- konjugacijska vrsta
- računalni korpus
- Hrvatski nacionalni korpus (HNK)
- Hrvatska jezična riznica

10. POPIS LITERATURE

1. Barić, Eugenija i suradnici, *Hrvatska gramatika*, Školska knjiga, Zagreb, 1997.
2. Marković, Ivan, *Uvod u jezičnu morfologiju*, Disput, Zagreb, 2012.
3. Silić, Josip, Pranjković, Ivo, *Gramatika hrvatskoga jezika*, Školska knjiga, Zagreb, 2005.
4. Tadić, Marko, *Računalna obradba hrvatskih korpusa: povijest, stanje i perspektive*, 1997., <http://www.hnk.ffzg.hr/txts/mt4hmk3.pdf> (posjećeno: 22.7.2016.)
5. Tadić, Marko, *Računalna obradba hrvatskoga i nacionalni korpus*, 1996., <http://www.hnk.ffzg.hr/txts/mt4hmk.pdf> (posjećeno: 22.7.2016.)
6. Tadić, Marko, *Raspon, opseg i sastav korpusa hrvatskoga suvremenog jezika*, 1998., <http://www.hnk.ffzg.hr/txts/mt4hmk2.pdf> (posjećeno: 22.7.2016.)
7. Tadić, Marko, *Zašto nam je potreban višemilijunski referentni korpus?*, 1990., <http://www.hnk.ffzg.hr/txts/mt4hmk11990.pdf> (posjećeno: 22.7.2016.)
8. Tadić, Marko, *Hrvatski računalni korpusi*, <http://www.matica.hr/vijenac/158/Hrvatski%20ra%C4%8Dunalni%20korpusi/> (Posjećeno: 19.7.2016)
9. Hrvatski nacionalni korpus, <http://www.hnk.ffzg.hr/> (posjećeno: 29.7.2016.)
10. Hrvatska jezična riznica, <http://riznica.ihjj.hr/index.hr.html> (posjećeno: 19.7.2016.)
11. http://filip.ffzg.hr/cgi-bin/run.cgi/first_form (posjećeno: 19.7.2016.)