

UDK 811.163.42(091)
811.163.42'322
Prethodno priopćenje
Rukopis primljen 8. I. 2020.
Prihvaćen za tisak 18. III. 2020.
doi.org/10.31724/rihjj.46.2.10

Marijana Horvat

Institut za hrvatski jezik i jezikoslovlje
Ulica Republike Austrije 16, HR-10000 Zagreb
mhorvat@ihjj.hr

ISTRAŽIVANJE POVIJESTI HRVATSKOGA JEZIKA U DIGITALNO DOBA¹

Hrvatska ima dugu tradiciju priređivanja djela u području povijesnoga jezikoslovlja i raspolaže mnogim vrijednim izdanjima te je veoma važno da se ta tradicija nastavi i prilagodi današnjemu digitalnom dobu. U ovome radu ponudit će se model retrodigitalizacije starih hrvatskih gramatika jer postojeći retrodigitalizirani resursi ne uključuju gramatike.

1. Uvod

Digitalizacija starih tekstova u stalnome je porastu, posebice od početka ovoga stoljeća. Razlozi digitalizacije arhivske, knjižnične, muzejske građe te one u znanstvenim ustanovama ili pohranjene u samostanima višestruki su. Jedan je od osnovnih razloga digitalizacije starih tekstova zaštita izvornika i očuvanje baštine jer se pri korištenju rabe digitalne preslike, a ne sami izvornici, pa se na taj način čuvaju od mogućih oštećenja. Digitalne su preslike ujedno i način izrade sigurnosne kopije samih izvornika (usp. Bago 2014: 78). Važan je razlog digitalizacije građe njezina veća dostupnost i stručnjacima i široj javnosti, što, s jedne strane, omogućava jednostavniji i brži pristup traženoj informaciji, pa time i istraživačkim aktivnostima, znanstvenim prinosima i spoznajama jer

¹ Ovaj je rad financirala Hrvatska zaklada za znanost projektom *Retrodigitalizacija i interpretacija hrvatskih gramatika do ilirizma* IP-2018-01-3585.

su stručnjacima mogućnosti koje otvaraju nove tehnologije nemjerljive s onima kojima raspolažu tiskana djela. S druge strane, izvan kruga stručnjaka, širemu krugu korisnika, posebice mladima koji su otvoreni u prihvaćenju elektroničkih medija, daje se veća mogućnost upoznavanja s kulturnom baštinom.

Unatoč spomenutom porastu digitalizacije kulturnopovijesno važnih tekstova u Hrvatskoj se tek stvaraju retrodigitalizirani resursi, npr. pojedinačni projekti Instituta za hrvatski jezik i jezikoslovlje (<http://ihjj.hr>); projekt Filozofskoga fakulteta Sveučilišta Josipa Jurja Strossmayera u Osijeku *EDICIJA – digitalna knjižnica hrvatske tiskane baštine* (www.ffos.unios.hr/EDICIJA/oediciji.php; projekt je dovršen, a zbirka sadržava pet tekstova); projekti Filozofskoga fakulteta Sveučilišta u Zagrebu *Croatiae auctores Latini* (CroALa) (www.ffzg.unizg.hr/klafil/croala) i *Hrvatska rječnička baština i prikaz rječničkoga znanja*, unutar kojega je izrađen Portal hrvatske rječničke baštine (<http://croqip.ffzg.hr>). Informacije o digitalizaciji hrvatske rječničke baštine nalazimo primjerice u članku Petre Bago *Označavanje natuknice u della Bellinu rječniku i povezivanje autora izvora citata s vanjskim resursima* (Bago 2014), u kojemu se opisuje način kodiranja rječničkih natuknica iz Della Bellina rječnika iz XVIII. st. uporabom TEI (Text Encoding Initiative) standarda te njihovo povezivanje na vanjske resurse (mrežno izdanje *Hrvatske enciklopedije* Leksikografskoga zavoda Miroslav Krleža i hrvatsku inačicu *Wikipedije*).

Međutim, ne postoji repozitorij koji bi sadržavao digitalizirane i pretražive stare hrvatske gramatike ni model koji bi se na njih mogao primijeniti, kao ni sustavno provedena znanstvena istraživanja u tome području. Spomenimo da digitalizirana građa Zbirke rukopisa i starih knjiga Nacionalne i sveučilišne knjižnice u Zagrebu sadržava faksimile nekoliko gramatika s osnovnim metapodatcima (<http://digitalna.nsk.hr/?object=list&filter=11688&mr%5B544759%5D=a>).

U Sloveniji, primjerice, postoji portal slovenskih gramatika i pravopisa od 1584. do današnjih dana (www.fran.si/slovnice-in-pravopisi). Portal sadržava slike izvornika gramatika i pravopisa (faksimile) te metapodatke uz svaki faksimil (npr. stoljeće, ime autora i inačicu zapisa imena autora, podatke o godini rođenja i smrti autora, kratak životopis autora, izvorni naslov i prijevod naslova (ako je primjenjivo), podatke o jeziku i metajeziku, vrsti, važnosti i namjeni djela i dr.), ali tekstovi nisu pretraživi. Slična ideja digitalne baze povijesnojezičnih tekstova potvrđena je na mrežnoj stranici *Classiques Garnier Numérique* (www.classiquesgarnier.com).

ssiques-garnier.com/numerique). Sadržava velik korpus gramatika francuskoga jezika, komentare i jezične rasprave od 14. do 17. stoljeća, s pretraživim bazama, ali bez slobodnoga pristupa. Na mrežnim stranicama Vokabulár webový / Mluvnické oddělení vývoje jazyka (Odjel za razvoj jezika) Ústavu pro jazyk český AV ČR nalazi se korpus staročeških gramatika od 16. do 19. stoljeća (<https://vokabular.ujc.cas.cz>). Gramatike su predstavljene faksimilima i popisom naziva (tj. ključnih riječi). Preko ključnih riječi moguće je pretraživanje sadržaja gramatika tako da se njima upućuje na faksimil svih gramatika koje na određenoj fotografiji sadržavaju traženi sadržaj.

Budući da je u Institutu za hrvatski jezik i jezikoslovlje 2001. godine pokrenut niz *Pretisci*, u kojemu je dosad objavljeno petnaestak gramatika, njihova se retrodigitalizacija i mrežna dostupnost nametnula kao sljedeći korak. Stoga je na natječaj Hrvatske zaklade za znanost prijavljen projekt *Retrodigitalizacija i interpretacija hrvatskih gramatika do ilirizma*, koji je HRZZ odobrio za financiranje te je projekt započeo s radom 25. studenoga 2019. godine.² U ovome radu opisat će se metodologija i ciljevi toga projekta, koji će donijeti inovacije u području istraživanja povijesnojezičnih tekstova, konkretno dopreporodnih gramatika.

2. Metodologija i ciljevi

Retrodigitalizacija hrvatskih gramatika razumijeva prijenos tiskanoga medija u računalno čitljiv i pretraživ tekst. Ona obuhvaća i višerazinsko označivanje transkribiranoga ili prevedenoga teksta hrvatskih dopreporodnih gramatika te njihovo povezivanje sa slikama izvornoga teksta (faksimilima). Korisnik će moći pregledavati faksimile priređenih gramatika, čitati transkribirani ili prevedeni tekst gramatika te ga pretraživati po unaprijed zadanim kriterijima pretrage (kojima će se omogućiti pretraživanje sklonidbenih paradigma i paradigma sprezanja).

Projektom je obuhvaćeno dvanaest gramatika, štokavskih i kajkavskih, objavljenih do hrvatskoga narodnog preporoda, odnosno do izbora zajedničkoga stan-

² Projekt će trajati četiri godine. Istraživačku skupinu čine: dr. sc. Marijana Horvat (voditeljica projekta), doc. dr. sc. Petra Bago, dr. sc. Martina Kramarić, dr. sc. Ivana Lovrić Jović, dr. sc. Ana Mihaljević, dr. sc. Sanja Perić Gavrančić, prof. emeritus Ivo Pranjković, prof. dr. sc. Diana Stolac, prof. dr. sc. Ljiljana Šarić, dr. sc. Barbara Štebih Golub, izv. prof. dr. sc. Tamara Tvrtković (suradnici), dr. sc. Toma Tasovac (savjetnik), dr. sc. Vuk-Tadija Barbarić (izvanprojektni savjetnik).

dardnog jezika i slovopisne norme kao poveznice književne baštine na trima stilizacijama hrvatskoga književnog jezika i tronarječnome hibridu. Priredili bi se faksimili svih dvanaest gramatika. Tekstnim digitalnim izdanjem bilo bi obuhvaćeno osam gramatika, odnosno transkripcijama ili prijevodima na kojima bi se provodilo kodiranje segmenata gramatičkoga teksta i morfoloških paradigma TEI oznakama. To su: Bartol Kašić, *Institutionum linguae Illyricae libri duo* (1604.), Jakov Mikalja, *Gramatika talijanska ukratko ili kratak nauk za naučiti latinski jezik* (1649.), Ardelio Della Bella, *Istruzioni grammaticali della lingua illirica* (1728.), Blaž Tadijanović, *Svašta po malo iliti kratko složenje imena, riči u ilirski i njemački jezik* (1761.), Ignacije Szentmártony, *Einleitung zur kroatischen Sprachlehre für Deutsche* (1783.), Josip Voltić, *Grammatica illirica* (1803.). Riječ je o knjižnim izdanjima koja je Institut za hrvatski jezik i jezikoslovlje objavio u nizu *Pretisci* te su za njih već izrađene transkripcije ili prijevodi.³ Njima su dodane gramatike Francesca M. Appendinija *Grammatica della lingua Illirica* (1808.) i Marijana Lanosovića *Uvod u latinsko riči slaganje s nikima nimačkog jezika biliškama za korist slovinskih mladića složen* (1776.), za koje su prijevod s talijanskoga jezika na hrvatski, odnosno transkripcija izrađeni u okviru internoga institutskoga projekta Dopreporodne hrvatske gramatike (<http://ihjj.hr/projekt/dopreporodne-hrvatske-gramatike/32/>), ali su još u rukopisu.⁴ Iz gramatika pisanih hrvatskim metajezikom ispisivat će se gramatičko nazivlje. Hrvatsko gramatičko nazivlje ispisivat će se i iz četiriju priručnika koji neće biti uključeni u označivanje: Josip Matijević, *Pomum granatum* (1771.), Tomo Mikloušić, *Syllabus vocabulorum Grammaticae Emmanuelis Alvari, in Croaticam*

³ Podatci o pretisnutim izdanjima s transkripcijama ili prijevodima i pratećim studijama: 1. Kašić, Bartol. 2002. *Institutionum linguae Illyricae libri duo / Osnove ilirskoga jezika u dvije knjige*. Pretisak izdanja iz 1604. Prijevod: Sanja Perić Gavrančić. Pogovor: Darija Gabrić-Bagarić. Institut za hrvatski jezik i jezikoslovlje. Zagreb.; 2. Mikalja, Jakov. 2008. *Gramatika talijanska ukratko ili kratak nauk za naučiti latinski jezik*. Pretisak izdanja iz 1649. Transkripcija: Marijana Horvat. Studija: Darija Gabrić-Bagarić i Marijana Horvat. Institut za hrvatski jezik i jezikoslovlje. Zagreb.; 3. Della Bella, Ardelio. 2006. *Istruzioni grammaticali della lingua illirica / Gramatičke pouke o ilirskome jeziku*. Pretisak izdanja iz 1728. Prijevod: Nives Sironić-Bonefačić. Pogovor: Darija Gabrić-Bagarić. Institut za hrvatski jezik i jezikoslovlje. Zagreb.; 4. Horvat, Marijana; Ramadanović, Ermina. 2012. *Jezikoslovni priručnik Blaža Tadijanovića Svašta po malo iliti kratko složenje imena i riči u ilirski i njemački jezik (1761.)*. Pretisak izdanja iz 1761., transkripcija i studija. Transkripcija: Marijana Horvat. Institut za hrvatski jezik i jezikoslovlje. Zagreb.; 5. Štebih Golub, Barbara. 2014. *Ignacije Szentmártony i Uvod u nauk o horvatskome jeziku*. Pretisak izdanja iz 1783., prijevod i studija. Institut za hrvatski jezik i jezikoslovlje. Zagreb.; 6. Voltić, Josip. 2016. *Grammatica Illirica / Ilirska gramatika (1803.)*. Pretisak izdanja iz 1803. Prijevod: Sanja Perić Gavrančić. Transkripcija i studija: Marijana Horvat i Sanja Perić Gavrančić. Institut za hrvatski jezik i jezikoslovlje. Zagreb.

⁴ Appendinijevu gramatiku prevela je Ivana Lovrić Jović, a Lanosovićevu je transkribirala Sanja Perić Gavrančić.

linguam conversorum (1796.), Antun Rožić, *Kratko napatjenje vu kruto hasnoviteh, i zevsema potrebnih temeljih dijačkoga jezika* (1820.), Antun Rožić, *Prvi temelji dijačkoga jezika za početnike vu domorodnom jeziku* (1821.).

Nisu sve gramatike pisane hrvatskim jezikom niti su sve gramatike hrvatskoga jezika. Hrvatske gramatike sa stranim metajezikom sadržavaju obilnu hrvatsku građu u sklonidbenim paradigmama, paradigmama sprezanja i hrvatskim primjerima te u različitim uputama, poukama, tumačenjima, sintaktičkim konstrukcijama, što je i očekivano s obzirom na to da se opisuje hrvatski jezik. Gramatike stranih jezika (talijanskoga, latinskoga, njemačkoga) zbroju hrvatskih gramatika pridružuju odlomci i primjeri na hrvatskome te hrvatski kao metajezik. Hrvatski se jezik u tim gramatikama može rekonstruirati na temelju kontrastivno uvedениh hrvatskih primjera, sklonidbenih obrazaca i obrazaca sprezanja te iz hrvatski pisanih gramatičkih uputa i pouka, s tim što opis stranoga jezika diktira i uvjetuje opseg hrvatske gramatičke građe. Za označivanje je stoga iznimno važna analiza strukture svake gramatike, na temelju koje će se izraditi popis oznaka za retrodigitalizaciju. Oznake će se raditi prema smjernicama TEI-ja za humanističke znanosti za strojno čitljiv format XML (eXtensible Markup Language). U početnim mjesecima rada na projektu nužno je detaljno proučiti svaku pojedinu gramatiku i provesti dubinsku raščlambu teksta gramatike kako bi se njezina struktura točno prenijela u format XML. Nakon toga slijedi odabir TEI oznaka i izrada prilagođene TEI sheme, koja bi bila unificirana za sve gramatike kako bi se osigurala dosljednost obilježavanja. Potom slijedi najduža faza, koja uključuje prenošenje teksta gramatika u format XML i njihovo tematsko označivanje (primjerice prema kriterijima kao što su padežne paradigme ili glagolski oblici i vremena). Označivanje će se, kao što je spomenuto, provoditi na transkribiranim tekstovima i prijevodima. Označivanjem odabranih gramatičkih kategorija korisniku bi bilo omogućeno tematsko pretraživanje po unaprijed zadanim elementima pretrage prema standardnojezičnome hiperonimu (npr. *a*-sklonidba, pluskvamperfekt) svih gramatika u jednome prikazu, a time i uspoređivanje sadržaja među gramatikama s obzirom na odabrane elemente pretrage.

Također se planira sastaviti kazalo hrvatskoga dopreporodnog gramatičkog nazivlja koje bi sadržavalo popis potvrđenih naziva s kraticom gramatike u kojoj se nazivi nalaze (npr. kratica Mik. uz naziv odnosit će se na potvrdu u *Gramatici talijanskoj ukratko ili kratkom nauku za naučiti latinski jezik* (1649.) Jakova Mi-

kalje; kratica Lan. = Marijan Lanosović: *Uvod u latinsko riči slaganje s nikima nimačkog jezika biličkama za korist slovinskih mladića složen* (1776.)). Nazivi bi bili okupljeni pod standardnojezičnim hiperonimom, npr. **pridjev**: *ime ganutivo* (Mik.), *ime primitljivo* (Lan.).

Metodologija obuhvaća ove korake:

1. izrada faksimila odabranih gramatika
2. priprema transkripcija ili prijevoda odabranih gramatika
3. analiza strukture svake gramatike i njihova dubinska raščlamba
4. uspostava smjernica za kodiranje teksta: odabir TEI oznaka ujednačenih za sve gramatike; odabir strukture metapodataka u okviru TEI zaglavlja i izrada prateće dokumentacije
5. izrada prilagođene TEI sheme (u formatu ODD i RelaxNG) sa skupom predviđenih oznaka za kodiranje gramatika
6. definiranje standardnojezičnih hiperonima za pretraživanje
7. prenošenje teksta gramatika u format XML i označivanje segmenata gramatičkoga teksta i morfoloških paradigmi TEI oznakama
8. izrada kazala hrvatskoga dopreporodnog gramatičkog nazivlja
9. testiranja postojećih rješenja (npr. TEI Boilerplate, TEI Publisher) za izradu korisničkoga okruženja za čitanje i pretragu
10. izrada tehničkoga programskog rješenja za povezivanje dokumenata u TEI-ju sa sučeljem i za povezivanje označenih dijelova teksta s faksimilima gramatika
11. prilagođivanje konkretnoga programerskog rješenja za izradu portala hrvatskih gramatika do ilirizma
12. objavljivanje obrađenih gramatika s metapodacima te kazalo nazivlja na portalu hrvatskih gramatika do ilirizma.

Cilj je projekta uspostava portala hrvatskih gramatika do ilirizma sa slikama izvornika, metapodacima i tekstovima transkripcija/prijevoda odabranih gramatika, povezivanje faksimila gramatika s transkripcijama/prijevodima, omogućivanje pretraživosti slovnica na morfološkoj razini po unaprijed zadanim elementima pretrage te izrada kazala hrvatskoga dopreporodnog gramatičkog nazivlja i metapodataka o svakoj gramatici. Projekt će omogućiti sustavno istra-

živanje dopreporodnih hrvatskih gramatika i pridonijeti stvaranju cjelovitoga uvida u hrvatsku dopreporodnu gramatičku tradiciju. Planirano je i povezivanje s drugim relevantnim projektima i ustanovama.⁵

Kako ne postoji repozitorij koji bi sadržavao digitalizirane i pretražive stare hrvatske gramatike ni model koji bi se na njih mogao primijeniti, ujedno će se razviti proces i ponuditi model za buduća slična istraživanja.

Portal će biti otvorenoga pristupa, dostupan znanstvenicima i široj publici, a zamišljen je za trajni razvoj i rast nakon dovršetka projekta.

3. Zaključak

Rad na projektu zahtijeva interdisciplinarni pristup i suradnju proučavatelja iz različitih područja, ne samo iz jezikoslovne kroatistike, kao i aktivnost računalnih stručnjaka, što pokazuje i sastav istraživačke skupine.

Projektom se nastoje intenzivirati istraživanja dopreporodne gramatičke građe, upotpuniti spoznaje o morfološkom razvoju hrvatskoga jezika u dopreporodnom razdoblju i spoznaje o razvoju normativnoga opisa te ponuditi znanja primjenom modernih lingvističkih disciplina i tehnologija. Projekt će omogućiti nova istraživanja, posebice komparativna i traduktološka i ne samo kroatistička, te utjecati na očuvanje navedene građe i baštine, koju će ujedno učiniti dostupnom širem krugu korisnika. Također će ponuditi model za buduća slična istraživanja u području digitalne humanistike. U tome smislu vjerujemo da će projekt donijeti inovacije ne samo u području hrvatske, nego i u području slavenske filologije.

Literatura

BAGO, PETRA. 2014. Označavanje natuknice u della Bellinu rječniku i povezivanje autora izvora citata s vanjskim resursima. *Studia lexicographica* 8/2(15). 77–102.

BAGO, PETRA; LJUBEŠIĆ, NIKOLA. 2015. Using machine learning for language and structure annotation in an 18th-century dictionary. *Electronic lexicography in the 21st century*:

⁵ S tim u vezi već je uspostavljena suradnja s projektom *Hrvatsko jezikoslovno nazivlje – JENA* (<http://ihj.hr/jena>), a planirano je povezivanje s Manuscriptoriumom (www.manuscriptorium.com), drugom najvećom digitalnom knjižnicom u Europi za stare rukopise i dokumente prije 18. stoljeća, iz Nacionalne knjižnice u Pragu te s Nacionalnom i sveučilišnom knjižnicom u Zagrebu.

linking lexical data in the digital age. Ur. Kosem, Iztok; Jakubiček, Miloš; Kallas, Jelena; Krek, Simon. Trojina, Institute for Applied Slovene Studies – Lexical Computing Ltd. Ljubljana – Brighton. 427–442.

BOSANČIĆ, BORIS. 2011. *Označavanje teksta starih knjiga na hrvatskom jeziku pomoću TEI standarda*. Doktorski rad. Sveučilište u Zadru. Zadar. 415 str.

JOVANOVIĆ, NEVEN. 2008. Digitizing Croatian Latin Writers. *Trends in Computational and Formal Philology 2008: An Italian Overview*. Venecija – Padova. 22–24.

JOVANOVIĆ, NEVEN. 2012. CroALa. *Journal of the Text Encoding Initiative 2*. 1–9. doi.org/10.4000/jtei.425.

Mrežni izvori

Classiques Garnier. www.classiques-garnier.com/numerique (pristupljeno 20. prosinca 2019.).

CroALa: Croatiae auctores Latini. www.ffzg.unizg.hr/klafil/croala (pristupljeno 20. prosinca 2019.).

Manuscriptorium – Digital Library of Written Cultural Heritage. www.manuscriptorium.com (pristupljeno 20. prosinca 2019.).

EDICIJA – digitalna knjižnica hrvatske tiskane baštine. www.ffos.unios.hr/EDICIJA/oediciji.php (pristupljeno 20. prosinca 2019.).

Hrvatska rječnička baština i prikaz rječničkoga znanja. <http://croqip.ffzg.hr> (pristupljeno 20. prosinca 2019.).

Hrvatsko jezikoslovno nazivlje – JENA. <http://ihjj.hr/jena> (pristupljeno 20. prosinca 2019.).

PIERAZZO, ELENA. 2014. Digital Scholarly Editing: Theories, Models and Methods. *HAL Open Archives System*. <http://hal.univ-grenoble-alpes.fr/hal-01182162/document> (pristupljeno 20. prosinca 2019.).

Slovenske slovnice in pravopisi. Portal slovenskih slovnice in pravopisov od 1584 do danes. www.fran.si/slovnice-in-pravopisi (pristupljeno 20. prosinca 2019.).

TEI – Text Encoding Initiative. <http://www.tei-c.org/index.xml> (pristupljeno 20. prosinca 2019.).

Vokabulář webový. <https://vokabular.ujc.cas.cz> (pristupljeno 20. prosinca 2019.).

Zbirka rukopisa i starih knjiga. <https://digitalna.nsk.hr/pb/?object=list&filter=11688> (pristupljeno 20. prosinca 2019.).

Croatian Language History Research in the Digital Era

Abstract

Croatian philology has a long tradition of publishing reference books in the field of historical linguistics; it is very important to continue and foster this tradition in the future, as well as to adapt it to the modern digital era.

Recent large-scale digitization initiatives have focused heavily on historical texts. Many reasons have led to the increasing digitization of archival documents, library and museum material, and collections of old books and manuscripts stored in monasteries. One of the main reasons for the digitization of old texts is to protect the original document from possible damage resulting from inappropriate handling. The use of digital copies instead of original documents is necessary to protect and preserve old written texts, as they are the most reliable way to back up original documents. Another very important reason for digitizing old texts is to make them more accessible to scholars, experts, and the public. In this way, young people who prefer using digital media will become more easily acquainted with cultural heritage that would otherwise be inaccessible to them. The digitization model of old Croatian grammar books will be discussed, as the existing resources do not include grammar books from the pre-standard period of the Croatian language.

Ključne riječi: digitalna humanistika, retrodigitalizacija, dopreporodne hrvatske gramatike, označivanje teksta, TEI oznake

Keywords: digital humanities, retro-digitization, pre-standard Croatian grammar books, text encoding, TEI tags

