

Označavanje natuknice u della Bellinu rječniku i povezivanje autora izvora citata s vanjskim resursima

Petra Bago

Filozofski fakultet Sveučilišta u Zagrebu

SAŽETAK: U posljednjih nekoliko desetljeća sve se više povijesnih tekstova digitalizira. Međutim kako se razvoj odvijao izolirano na projektima, sveučilištima i institutima, ili među pojedincima, nedostajala je komunikacija među članovima zajednice, što je rezultiralo resursima, alatima i sustavima koji međusobno ne mogu razmjenjivati informacije. Stoga se u novije vrijeme pokazala potreba za standardizacijom tehnologija, ali i cijelih procesa razvoja. Kao ključni pojam u fazi digitalizacijskih projekata javlja se interoperabilnost, kojom se želi omogućiti međusobna komunikacija digitaliziranih resursa. U radu je prikazan način kodiranja rječničkih natuknica iz della Bellina rječnika iz XVIII. st. korištenjem TEI (*Text Encoding Initiative*) standarda te njihovo povezivanje na vanjske resurse (mrežno izdanje *Hrvatske enciklopedije* Leksikografskog zavoda Miroslav Krleža i hrvatsku inačicu *Wikipedije*).

Ključne riječi: *povijesni rječnici; interoperabilnost; kodiranje; Text Encoding Initiative; društveno-humanistička informatika; Ardelio della Bella*

Uvod

Razvoj informacijsko-komunikacijske tehnologije (*Information and communications technology* – ICT) omogućuje jednostavniji pristup i brži protok informacija, pohranu velikih količina podataka te sve jednostavnije rukovanje istima. Godine 2011. Hilbert i López proveli su istraživanje u kojem su, prateći 60 analognih i digitalnih tehnologija u razdoblju od 1986. do 2007., procijenili sposobnost tehnologije da pohrani, komunicira i obradi informacije. Prema istraživanju, sposobnost računala opće namjene da obradi podatke rasla je godišnje po stopi od 58%. Također je na svjetskoj razini sposobnost dvosmjernih telekomunikacija godišnje rasla po stopi od 28%, a pohrana informacija po stopi od 23%.¹ Osim što je iz toga vidljivo da se spo-

¹ Prema navedenom istraživanju (Hilbert, López 2011), 2007. godine ljudska vrsta mogla je pohraniti $2,9 \times 10^{20}$ optimalno sažetih bajtova i komunicirati gotovo 2×10^{21} bajtova te obaviti $6,4 \times 10^{18}$ naredbi u sekundi na računalima opće namjene.

sobnosti informacijsko-komunikacijske tehnologije povećavaju, njezina sve lakša primjena omogućava prodiranje u sve segmente ljudskoga djelovanja, pa tako i u kulturnu djelatnost. Stoga ne čudi činjenica da se sve više projekata koji uključuju kulturnu baštinu okreće upravo informacijsko-komunikacijskoj tehnologiji kako bi omogućili pristup, protok, pohranu i rukovanje baštinom. Da bi se omogućila računalna obrada baštine,² prvi korak koji je potrebno učiniti jest digitalizacija baštine, odnosno pretvorba analognih podataka ili signala (npr. tiskani dokument) u digitalni, odnosno brojčani oblik.³

Digitalizacija kulturne baštine

U posljednjih nekoliko desetljeća sve se više povijesnih tekstova digitalizira. U Hrvatskoj postoje četiri osnovne baštinske djelatnosti: knjižnična, arhivska, muzejska i konzervatorsko-restauratorska, a glavni im je cilj prikupljanje, obrađivanje, istraživanje, pohranjivanje, zaštita i komuniciranje baštine (Seiter-Šverko, Križaj 2005). Digitalizacija građe u arhivima, knjižnicama i muzejima, ali i u drugim ustanovama, primjerice na sveučilištima, fakultetima i institutima, provodi se iz više razloga. Osnovni razlog pristupanju digitalizacije građe jest zaštita izvornika, koja može imati dva oblika. Kod aktivnoga oblika digitalne se preslike koriste umjesto samih izvornika, kako bi se izvornici očuvali od mogućih oštećenja prilikom rukovanja, dok se kod pasivnoga oblika digitalne preslike koriste samo kao sigurnosne kopije koje mogu (barem djelomično) nadoknaditi gubitak u slučaju oštećenja ili uništenja izvornika. Sljedeći, temeljni razlog digitalizacije građe je poboljšanje dostupnosti i mogućnosti korištenja građe, pogotovo ako se radi o rijetkim ili jedinstvenim objektima. Omogućavanjem mrežnoga pristupa digitaliziranoj građi osigurava se njezina veća dostupnost, kako stručnjacima tako i široj publici. Budući da pretvorba građe u digitalni oblik omogućava njezinu bržu i jednostavniju računalnu obradu, digitalizacijom se mogu stvoriti nove ponude i usluge korisnicima koje su moguće jedino zahvaljujući novom mediju/okruženju u kojem se nalazi. Konačno, razlog pristupanju digitalizaciji građe može biti upotpunjavanje postojećega fonda institucije, ako ta institucija iz različitih razloga procijeni da su joj određene digitalne preslike potrebne (*Prijedlog Nacionalnog programa digitalizacije arhivske, knjižnične i muzejske građe*. Radna grupa za digitalizaciju arhivske, knjižnične i muzejske građe Ministarstva kulture Republike Hrvatske, 2006).

² Prema natuknici *digitize*. Oxford Dictionaries. Oxford University Press. <http://oxforddictionaries.com/definition/english/digitize?q=digitise> (pristupljeno 10. V. 2015).

³ Prema natuknici *digitalizirati*. Hrvatski jezični portal. Novi Liber. URL: http://hjp.novi-liber.hr/index.php?show=search_by_id&id=fldmXxQ%3D&keyword=digitalizirati (pristupljeno 10. V. 2015).

Za digitalizaciju su prikladne različite vrste građe pa je tako moguće digitalizirati tekstove, slike, zvuk, video, trodimenzionalne objekte, simulacije događaja itd. Digitalne se preslike za takvu građu izrađuju uglavnom radi zaštite originala.

Tekstovi se pak mogu pretraživati, ali i automatski obrađivati. Upravo zato današnji projekti digitalizacije uglavnom teže pretvorbi tiskanih dokumenata u digitalne tekstove. Zahvaljujući sve lakšoj i širokoj upotrebi informacijsko-komunikacijske tehnologije u znanosti, ali i svakodnevnom životu, projekti digitalizacije obuhvaćaju razna interdisciplinarna područja kao što su informacijske znanosti, društveno-humanistička informatika (engl. *digital humanities*, stariji naziv *humanities computing*), računalna lingvistika itd. Sva navedena područja uglavnom se oslanjaju na istraživanja digitalnih tekstova, neovisno o tome radi li se o digitaliziranim tekstovima ili o tekstovima koji su izvorno nastali u digitalnom obliku.

Posljedica velikoga interesa za digitalizaciju razvoj je metoda, resursa i alata za računalnu obradu povijesnih tekstova. Međutim, kako se razvoj odvijao izolirano na projektima, sveučilištima i institutima, ili među pojedincima, nedostajala je komunikacija među članovima digitalizacijske zajednice, što je rezultiralo resursima, alatima i sustavima koji ne mogu međusobno razmjenjivati informacije. Stoga se u novije vrijeme pokazala potreba za standardizacijom tehnologija, ali i cijelih procesa razvoja. Kao ključni pojam u fazi digitalizacijskih projekata javlja se interoperabilnost, kojom se želi omogućiti međusobna komunikacija digitaliziranih resursa.

Osim omogućavanja interoperabilnosti resursa, javlja se i potreba za automatizacijom procesa digitalizacije te naknadnoga obogaćivanja tih resursa. Posljedica sve veće dostupnosti povijesnih tekstova sve je veće zanimanje za primjenu metoda, a time i alata, za obradu prirodnoga jezika pri obradi povijesnih tekstova. Ipak, povijesni tekstovi imaju nekoliko jezičnih značajki koje predstavljaju prepreke za primjenu alata za obradu prirodnoga jezika u onom obliku u kakvom trenutačno postoje. Računalna obrada povijesnih tekstova naići će vjerojatno na jednake probleme na koje nailazi kada se obrađuju suvremeni tekstovi, kao što su homonimija i polisemija (Piotrowski 2012:1). Međutim dodatne značajke otežavaju primjenu alata za obradu prirodnoga jezika na povijesnim tekstovima: nedostatak standardne inačice jezika, nedostatak standardnoga pravopisa, nedostupnost tekstova u digitalnom obliku te nepostojeći resursi i alati za obradu prirodnoga jezika za povijesne tekstove. Stoga je potrebno razviti navedene metode, a time i alate, za takvu vrstu tekstova.

Interoperabilnost

Kako bi se riješio problem nedostatka komunikacije između jezičnih podataka i postigla interoperabilnost jezičnih resursa, udružila su se dva velika međunarodna projekta: američki projekt *Sustainable Interoperability for Language Technology*

(SILT)⁴, kojemu je cilj postojeću, ali fragmentiranu tehnologiju te resurse razvijene kao podrška jezičnoj tehnologiji pretvoriti u pristupačne, stabilne i interoperabilne resurse koji se u različitim područjima mogu lako ponovno koristiti,⁵ i europski projekt *Fostering Language Resources Network* (FLaReNet)⁶, kojemu je cilj razviti zajedničku viziju u području jezičnih resursa i jezičnih tehnologija te promicati europsku strategiju za konsolidaciju sektora, povećavajući time konkurentnost na razini Europske unije i svijeta.⁷ Cilj toga međunarodnoga udruživanja jest suradnja članova zajednice iz područja jezičnih tehnologija, ali i ostalih koji djeluju u srodnim područjima. Suradnja bi trebala rezultirati konsenzusom vezanim uz dijeljenje podataka i tehnologija za jezične resurse i aplikacije, zajedničkim radom na interoperabilnosti postojećih podataka te, gdje je to moguće, promicanjem standarda za označavanje i stvaranje resursa (Ide i sur. 2009). Jedno od glavnih postignuća navedene međunarodne suradnje jest ponuđena definicija pojma »interoperabilnost« unutar područja jezičnih tehnologija (Pustejovsky, Ide 2010).

Općenito se interoperabilnost može definirati kao stupanj do kojega različiti sustavi, organizacije i/ili pojedinci mogu raditi zajedno kako bi postignuli zajednički cilj. U računalnim sustavima interoperabilnost se uglavnom dijeli na dvije vrste: sintaktičku i semantičku. Svrha sintaktičke interoperabilnosti jest omogućiti komunikaciju i razmjenu podataka putem određenih formata podataka, komunikacijskih protokola i slično. Važno je da dođe do razmjene informacija, ali nema jamstva da će tumačenje razmijenjenih informacija biti jednako. S druge strane, kod semantičke interoperabilnosti sustavi imaju mogućnost automatski tumačiti razmijenjene informacije suvislo i točno kako bi dobili korisne rezultate, a tumačenje je moguće putem poštivanja zajedničkoga referentnoga modela za razmjenu informacija. Kod ove vrste interoperabilnosti bitno je da tumačenje razmijenjenih informacija bude jednako s obje strane komunikacije (Pustejovsky, Ide 2010).

Kako bi se omogućila interoperabilnost jezičnih resursa glavni smjer razvoja područja ide prema specificiranju apstraktnoga modela podataka za strukturiranje jezičnih podataka, na koji se zatim mogu povezati sintaktička ostvarenja zajedno s povezivanjem na skup kategorija jezičnih podataka koji komunicira sadržaj informacija, odnosno jezični sadržaj. Stoga u kontekstu jezičnih resursa možemo sintaktičku interoperabilnost definirati kao mogućnost različitih sustava da obrade (odnosno čitaju) razmijenjene podatke, bilo direktno bilo preko jednostavne konverzije. S dru-

⁴ Održiva interoperabilnost jezičnih tehnologija (prijevod autorice).

⁵ Sustainable Interoperability for Language Technology. URL: <http://www.anc.org/LAPPS/SILT/index.html> (pristupljeno 12. V. 2015).

⁶ Mreža za promicanje jezičnih resursa (prijevod autorice).

⁷ Fostering Language Resources Network (FLaReNet). URL: <https://www.flarenet.eu/> (pristupljeno 12. V. 2015).

ge strane semantičku interoperabilnost možemo definirati kao mogućnost sustava da razmijenjene jezične informacije tumači suvislo te ih dosljedno upućuje na zajednički skup referentnih kategorija. U tom smislu jezični se resursi moraju više usredotočiti na semantičku nego na sintaktičku interoperabilnost (Pustejovsky, Ide 2010).

Osim navedenih projekata SILT i FLaReNet, na europskoj su razini pokrenute dvije istraživačke mreže CLARIN i NeDiMAH. Cilj je mreže CLARIN (*Common Language Resources and Technology Infrastructure*⁸) pružiti istraživačima društvenih i humanističkih znanosti jednostavan i održiv pristup digitalnim jezičnim podacima (u pisanom, govorenom, video ili kombiniranom obliku) te napredne alate za pronalaženje, istraživanje, iskorištavanje, označavanje ili analizu takve vrste podataka neovisno o tome gdje su pohranjeni⁹. Ključna pretpostavka uspješnosti ove infrastrukture je interoperabilnost jezičnih podataka i alata različitih istraživačkih centara. Mreža NeDiMAH (*Network for Digital Methods in the Arts and Humanities*¹⁰) imala je za cilj organizaciju niz aktivnosti usmjerenih prema poticanju i razvoju digitalnih istraživanja u području humanističkih znanosti i umjetničkog područja u Europi¹¹. Posljedica mreže nije bila interoperabilnost digitalnih projekata, već umrežavanje istraživača kako bi se potaknula komunikacija među članovima zajednice.

U istraživanju su korištene TEI oznake za označavanje strukture rječničkih natuknica. TEI (*Text Encoding Initiative*¹²) konzorcij razvija i održava smjernice (*Guidelines*) koje su namijenjene svima koji se bave pripremom i/ili obradom tekstualnih resursa u digitalnom obliku. Smjernice su ustvari preporuke konzorcija kako na odgovarajući način prikazati svojstva onih tekstualnih resursa koji se radi lakše računalne obrade moraju izričito označiti. Preporuke dolaze u obliku detaljnoga opisa oznaka (*tag*) koje se unose u digitalni prikaz teksta, čime se obilježava struktura teksta i moguća druga svojstva koja se radi kasnije obrade žele eksplicitno označiti. Na taj se način implicitna tekstualna svojstva, u procesu koji se naziva kodiranje (*encoding*), čine eksplicitnima. Kodiranje prema unaprijed definiranoj shemi omogućava razmjenu podataka bez gubitaka informacija i bez mogućnosti da će neka informacija biti krivo interpretirana. Budući da se TEI smjernice temelje na XML-u, kao što se objašnjava u nastavku, također su neovisne o računalnim programima i operativnim sustavima¹³.

⁸ Infrastruktura zajedničkih jezičnih resursa i tehnologija (prijevod autorice).

⁹ <http://clarin.eu/>

¹⁰ Mreža digitalnih metoda u umjetnostima i humanističkim znanostima.

¹¹ <http://www.nedimah.eu/>

¹² Inicijativa za kodiranje teksta (prijevod autorice).

¹³ TEI Consortium (ur.): »iv. About These Guidelines«. TEI P5: Guidelines for Electronic Text Encoding and Interchange. (Verzija 2.6.0). URL: <http://www.tei-c.org/Guidelines/P5/> (pristupljeno 17. V. 2015).

Smjernice se mogu primijeniti na tekstove svih prirodnih jezika, iz bilo kojega razdoblja i bilo koje književne vrste te bez ikakva ograničenja na oblik i sadržaj. Upravo su zbog toga smjernice odabrane kao temelj za označavanje strukture starih hrvatskih rječnika. Osim na digitalizirane tekstove, smjernice se mogu primijeniti i na tekstove nastale u digitalnom formatu¹⁴.

Temelj smjernica je *eXtensible Markup Language* (XML), jednostavan i fleksibilan jezik za označavanje koji se i inače koristi za razmjenu podataka na webu i drugdje.¹⁵ XML ima tri temeljne prednosti:¹⁶

1. stavlja naglasak na opisne oznake, a ne proceduralne
2. razlikuje koncepte sintaktičke točnosti i valjanosti s obzirom na upute za konkretan dokument
3. neovisan je o hardveru i softveru.

Uz to, XML omogućava prikaz i upis znakova prema standardu Unicode, što je nužno za starije hrvatske rječnike jer sadrže razne znakove i ligature koji danas ne pripadaju hrvatskoj abecedi, ali pripadaju tradiciji europske leksikografije. Navedenim standardom moguće je prikazati gotovo sve znakove koji se javljaju u starim hrvatskim rječnicima.

Kodiranje rječnika

Sadržaj smjernica vrlo je širok pa obuhvaća općenita poglavlja namijenjena svima koji se bave kodiranjem teksta, ali i specifična poglavlja za određene vrste tekstova te razne druge sadržaje potrebne korisnicima. Općenita poglavlja bave se temama kao što su uvod u XML za potrebe korisnika smjernica, jezici i znakovi, infrastruktura TEI-a, zaglavlja u TEI-u, elementi dostupni u svim dokumentima u TEI-u, te nestandardni znakovi i glifovi. U posebnim poglavljima smjernica zasebno se opisuju stihovi, izvedbeni tekstovi, transkripcija govora, rječnici, djela u rukopisu, prikazi primarnih izvora, kritička izdanja, imena, datumi, ljudi i mjesta, tablice, formule, slikovni prikazi i bilježenja glazbe, te jezični korpusi.¹⁷

¹⁴ Ibid.

¹⁵ eXtensible Markup Language (XML). URL: <http://www.w3.org/XML/> (pristupljeno 10. VII. 2014).

¹⁶ TEI Consortium (ur.): »v. A Gentle Introduction to XML«. TEI P5: Guidelines for Electronic Text Encoding and Interchange (Verzija 2.6.0). URL: <http://www.tei-c.org/Guidelines/P5/> (pristupljeno 15. V. 2015).

¹⁷ TEI Consortium (ur.): TEI P5: Guidelines for Electronic Text Encoding and Interchange (Verzija 2.6.0). URL: <http://www.tei-c.org/Guidelines/P5/> (17. V. 2015).

Za ovaj rad najvažnije je deveto poglavlje, koje se bavi rječnicima, a u kojem se definiraju oznake koje se mogu koristiti za označavanje rječničkih natuknica digitaliziranih rječnika, ali i drugih leksičkih resursa, kao što su primjerice glosariji.¹⁸ Upute iz navedenoga poglavlja korištene su za označavanje rječničkih natuknica starih hrvatskih tiskanih rječnika.

Pri obradi i označavanju rječnika treba se suočiti s dva glavna problema. Prvi je problem vezan uz strukturu natukničkog članka, a drugi uz informaciju dostupnu u natukničkom članku. Struktura natukničkog članka razlikuje se od rječnika do rječnika, a najčešće i unutar istoga rječnika. Kako bi shema za kodiranje mogla biti prikladna za različite strukture, jedini način je dozvoliti da se bilo koji element može pojaviti bilo gdje u članku. Međutim postoje i rječnici koji imaju dosljednu strukturu te stoga shema mora podržavati i takve slučajeve. Zbog toga je u okviru TEI standarda bilo potrebno definirati dva različita elementa za natuknički članak: *entry* i *entryFree*. Elementom *entry* prikazuju se strukturirani natuknički članci, dok element *entryFree* sadrži iste sastavnice kao i element *entry*, ali se mogu slobodnije kombinirati. Preporuka smjernica jest da se, gdje god je to moguće, prednost daje elementu *entry* naspram elementa *entryFree*.¹⁹ Za kodiranje natukničkih članaka starih hrvatskih rječnika upotrebljavat će se element *entry*.

Drugi problem koji pri označavanju rječničkoga teksta treba razriješiti jest način predstavljanja informacija u natukničkom članku. Većina je informacija u rječnicima implicitna ili vrlo sažeta. Stoga je potrebno odabrati treba li prikazati informacije kao u izvorniku ili implicitne informacije učiniti eksplicitnima. Budući da su istraživačima oba načina jednako potrebna i zanimljiva, shema mora podržavati i jedan i drugi način prikaza te također njihov međuodnos. Za neka je istraživanja važno zadržati izvorni prikaz natukničkog članka, dok će za druga biti važne samo leksičke informacije, što za posljedicu može imati promijenjen redoslijed pojavljivanja informacija, pri čemu je čak moguće i proširivanje, ali i izbacivanje određenih dijelova teksta. Moguća su istovremeno i oba pristupa pa je potrebno paralelno označiti tekst na oba načina i postaviti ih u međuodnos. Svi ti pristupi mogu se ostvariti upotrebom načela sadržanih u TEI smjernicama.²⁰

Za kodiranje tekstova natukničkih članaka starih hrvatskih rječnika najprikladniji je način zadržati izvorni redoslijed pojavljivanja informacija, dok će se dodatne informacije (koje se ne nalaze u izvornom tekstu) označiti u atributima elemenata. Na taj način razdvojiti će se izvorni sadržaj, koji se pojavljuje kao sadržaj elementa, od

¹⁸ TEI Consortium (ur.): »9. Dictionaries«. TEI P5: Guidelines for Electronic Text Encoding and Interchange (Verzija 2.6.0). URL: <http://www.tei-c.org/Guidelines/P5/> (pristupljeno 17. V. 2015).

¹⁹ Ibid.

²⁰ Ibid.

novoga sadržaja, odnosno značenja, koji će se dodati u atributu elementa. Kada se oznake nalaze u samom izvornom dokumentu i okružuju tekst koji želimo kodirati, takav način označavanja naziva se unutarnje označavanje (engl. *internal markup*).

Natukničke članke je moguće označiti i takozvanim udaljenim ili vanjskim označavanjem (engl. *stand-off markup*), metodom koja omogućava da oznake upućuju na izvorni dokument, ali se nalaze izvan samog izvornog teksta. Oznake se mogu nalaziti u istom izvornom dokumentu, ali ne okruživati tekst koji želimo kodirati, već se nalaziti u drugom dijelu XML dokumenta. Osim toga, oznake se mogu nalaziti i u vanjskom dokumentu te upućivati na izvorni dokument, koji u tom slučaju može biti jednostavna tekstna datoteka ili bilo koja druga vrsta zapisa (npr. zvuk, video, slika i sl.). Udaljeno označavanje ne sadrži izravno izvorni tekst, već samo upućuje na dijelove izvornog dokumenta. Na taj je način moguće jedan dokument označiti različitim hijerarhijski nekompatibilnim shemama, bez mijenjanja samog izvornog dokumenta.²¹

Međutim, s obzirom na to da je kod unutarnjeg označavanja izvorni tekst izravno povezan sa samim oznakama, prednost je ove metode označavanja jednostavnost korištenja i održavanja. Povrh toga, kod udaljenog označavanja u Smjernicama nije jasno izloženo daje li se prednost oznakama unutar istog dokumenta gdje se nalazi i izvorni dokument ili trebaju oznake biti u drugom dokumentu. U slučaju da se oznake nalaze unutar istog dokumenta, nije jasno ozloženo gdje se točno u XML dokumentu treba nalaziti izvorni tekst, a gdje oznake. Konačno, udaljeno označavanje predstavlja problem provjere valjanosti XML dokumenta²². Iz navedenih razloga odlučeno je da će u ovoj fazi istraživanja stari hrvatski rječnici biti označeni metodom unutarnjeg označavanja. Međutim, u slučaju da se riješe navedene nedoumice, moguće je izvršiti eksternalizaciju već označenih natukničkih članaka, odnosno podijeliti jedan XML dokument na dva dokumenta, jedan izvorni dokument i jedan XML dokument s udaljenim oznakama.

²¹ TEI Consortium (ur.): »16. Linking, Segmentation, and Alignment«. TEI P5: Guidelines for Electronic Text Encoding and Interchange (Verzija 2.6.0). URL: <http://www.tei-c.org/Guidelines/P5/> (pristupljeno 15. X. 2015).

²² Dobro oblikovan (engl. well-formed) XML dokument se pridržava sintakse XML jezika, dok se valjan (engl. valid) dokument pridržava dodatnih pravila konkretnog jezika temeljenog na XML-u, u našem slučaju TEI smjernica.

O della Bellinu rječniku

Ovaj se rad bavi kodiranjem natukničkih članaka drugoga izdanja della Bellina²³ rječnika *Dizionario italiano-latino-ilirico*, tiskanoga u Dubrovniku 1785. godine (della Bella 1785).

Ardelio della Bella (Foggia, 1655 – Split, 1737) bio je član Družbe Isusove, propovjednik i misionar, znanstvenik i pisac, leksikograf i gramatikograf. Tijekom dvadesetak godina boravka u Hrvatskoj prikupljao je građu za svoj rječnik *Dizionario italiano-latino-ilirico*, koji je objavio u Veneciji 1728. godine. Petar Bašić iz Dubrovnika poboljšao je i priredio za tisak drugo izdanje della Bellina rječnika koje je objavljeno u dva sveska u Dubrovniku 1785. godine (Bago 2008).

Rječnik je namijenjen talijanskim isusovačkim misionarima, kao priručnik koji će im pomoći u širenju vjere na narodnom, odnosno hrvatskom jeziku, ali i ostalim slavenskim jezicima. Rječnik spada u skupinu višejezičnih rječnika. Sastavljen je od sveukupno 899 stranica. Sastoji se od dva dijela: uvoda, koji je pisan je na talijanskom jeziku i nalazi se na prvih 56 stranica, paginiranih rimskim brojevima, i popisa riječi na 843 stranice, paginirane arapskim brojevima, koji sadrži približno 19 000 rječničkih natuknica. Prvi svezak rječnika ima 451 stranicu te obuhvaća uvodni dio i popis riječi od A do H. Drugi svezak ima 448 stranica i obuhvaća popis riječi od I do Z. Rječnik po prvi put u hrvatskoj leksikografiji donosi primjere upotrebe rječničkih natuknica iz različitih književnih djela i usmene književnosti (Bago, Boras 2012). Budući da su neke riječi i naglašene, rječnik može poslužiti i kao izvor za istraživanje povijesti akcentuacije u hrvatskom jeziku.

Označavanje natuknice

Na primjeru natukničkog članka iz della Bellina rječnika prikazat ćemo shemu korištenu za označavanje rječničkih natuknica. U Prilogu se nalazi cijeli izvorni tekst della Bellina natukničkog članka čije označavanje je prikazano u ovom radu.

Kao što je prethodno navedeno, za označavanje natukničkih članaka odabran je element <entry>. Element <form> označava sve oblike glavne natuknice, pisane i govorene. Atributom *type* može se odrediti o kojem je obliku riječ. Prilikom označavanja rječničke natuknice korišten je atribut *type*=”lemma”, koji označava da se radi o glavnoj rječničkoj natuknici u osnovnom obliku, što je u većini slučajeva i točno. Dodatnim se atributom definira jezik glavne natuknice, važan podatak za dvojezične i višejezične rječnike. Budući da odabrani rječnik ima samo pisanu varijantu glavne natuknice, a nedostaje mu izgovorna varijanta, korišten je element <orth>.

²³ Nives Sironić-Bonefačić napominje da postoje dvije različite grafije Ardelijeve prezimena: della Bella i Della Bella te smatra da je prihvatljiviji prvi oblik (Sironić-Bonefačić 1986: 12), koji je i prihvaćen u ovom radu.

Element koji se može naći u svim dijelovima TEI dokumenta jest <pc>, koji označava interpunkcijske znakove. Kod označavanja natuknica izričito je označen.

Prema gornjem opisu, natuknica *Anello* izgleda ovako:

```
<form type="lemma" xml:lang="it">  
  <orth>Anello</orth>  
  <pc>,</pc>  
</form>
```

Unutar glavne natuknice može se nalaziti i podnatuknica koja nije u osnovnom, već u nekom od flektivnih oblika. Uglavnom se u rječniku ne nalazi potpuni oblik riječi u fleksiji (type="inflected"), već samo nastavci. Razlog tomu je ušteda prostora. Tako je genitiv jednine imenice *Parften* prikazan samo kao »na«²⁴, a označili bismo ga na sljedeći način:

```
<form type="inflected">  
  <gramGrp>  
    <case value="genitive"/>  
    <number value="singular"/>  
    na<pc>.</pc>  
  </gramGrp>  
</form>
```

Ovdje se pojavljuju tri nova elementa. Element <gramGrp> grupira morfosintaktičke informacije jednoga pojma. Element <case> navodi u kojem je padežu imenica te njegovu vrijednost putem atributa (value="genitive"), dok element <number> navodi broj imenice te njegovu vrijednost putem atributa (value="singular"). Budući da posljednja dva podatka nisu izričito izrečena, već samo implicitno putem nastavka »na«, dodani su kao prazni elementi bez sadržaja. Kao što je već prethodno spomenuto, sadržaji elemenata bit će samo oni dijelovi koji se nalaze u izvorniku.

Nadalje, elementom <form> može se označiti i kolokacija, odnosno podnatuknica koja sadrži glavnu natuknicu kao noseću riječ (type="compound"). Element <oRef/> služi za referiranje na glavnu natuknicu u osnovnom obliku. Naprimjer u glavnoj natuknici *Anello* nalazi se kolokacija *Anello fenza pietra*:

```
<form type="compound" xml:lang="it">  
  <orth><oRef/> fenza pietra <pc>.</pc></orth>  
</form>
```

²⁴ Iako je nastavak za genitiv jednine imenice muškog roda u hrvatskome –a, della Bella je u pravilu navodio posljednji ili čak i posljednja dva sloga imenice kao genitivni nastavak. Budući da je rječnik bio namijenjen talijanskim isusovačkim misionarima, pretpostavljamo da je della Bella smatrao da će taj način navođenja genitivnog nastavka olakšati misionarima sklonidbu imenica.

U rječniku se nakon nastavka genitiva jednine kod imenica nalazi oznaka za rod (<gen>), a vrijednost roda može se zabilježiti atributom (value="m"). Osim oznake za rod, dodana je i oznaka za vrstu riječi (<pos>), a vrijednost vrste riječi može se zabilježiti atributom (norm="noun"). Oznaka za vrstu riječi je prazan element jer nije izričito izrečen u rječniku. Naprimjer, nakon već spomenute imenice *Parten* nalazi se nastavak za genitiv jednine te oznaka za muški rod:

```
<gramGrp>
  <pos norm="noun"/>
  <gen value="m">m<pc>.</pc></gen>
</gramGrp>
```

Glagoli se označavaju elementom <pos> te atributom @norm vrijednosti "verb". Za nastavak prvog lica jednine sadašnjega i prošloga vremena dodani su odgovarajući prazni elementi.

```
<cit type="translate" xml:lang="hr">
  <quote>Ski-<lb/>nuti <pc>,</pc></quote>
  <pos norm="verb"/>
  <form type="inflected">
    <gramGrp>
      <per value="1"/>
      <number value="singular"/>
      <tns value="present"/>
      kinivam <pc>,</pc>
    </gramGrp>
    <gramGrp>
      <per value="1"/>
      <number value="singular"/>
      <tns value="perfect"/>
      nuòfam <pc>.</pc>
    </gramGrp>
  </form>
</cit>
```

Element <sense> grupira različita značenja natukničkog članka, kao što su definicije, primjeri i prijevodi. Prema TEI Smjernicama ovim se elementom također grupiraju različite leksičke sveze glavne natuknice s njihovim prijevodima, što je možebitno zbunjujuće, jer se značenje glavne natuknice unutar leksičkih sveza ne mora nužno mijenjati. Kod označavanja natukničkih članaka primijenili smo ovu prihvaćenu praksu, iako bi u budućim istraživanjima bilo zanimljivo istražiti kako razlučiti stvarna različita značenja glavne natuknice. Atributom *n* može se označiti broj različitih značenja. Unutar toga elementa, elementom <cit> označuju se prijevodi

(type="translation") glavne natuknice ili podnatuknice te primjeri (type="example"). Obavezan element unutar elementa <cit> je <quote>. Tako je prijevod glavne natuknice *Anello* na latinskom označen ovako:

```
<cit type="translation" xml:lang="la">
  <quote>Anulus <pc>,</pc></quote>
</cit>
```

Primjeri hrvatskih citata u della Bellinu rječniku označeni su na sljedeći način:

```
<cit type="example" xml:lang="hr">
```

Unutar citata nalazi se i oznaka izvora primjera označena elementom <bibl>. Naprimjer:

```
<cit type="example" xml:lang="hr">
  <quote>I prid bojnóm sabgljom gljutóm <lb/>
    Parsten pirni cjùse uljesti <pc>.</pc></quote>
  <bibl><ref target="gundulic">Ofm</ref><pc>.</pc></
bibl>
</cit>
```

Ponekad u rječnicima postoji uputa na koje se značenje riječi misli. Takve upute označene su elementom <usg> i atributom type="hint":

```
<usg type="hint">
  che fi porta in dito <pc>.</pc>
</usg>
```

U rječnicima su česte uputnice na druge natuknice, a označene su elementom <xr> i atributom type="see". Unutar toga elementa nalazi se element <ref>, kojim je označen natuknički članak na koji se referira. Budući da taj element upućuje na natukničke članke unutar rječnika, ali i na vanjske resurse, potrebno je definirati lokaciju na koju se upućuje. To je moguće atributom *target*, gdje se u vrijednost atributa stavi lokacija reference, kao što je to u ovom primjeru:

```
<xr type="see">V.
  <ref target="#barbare">Barba-re</ref>
  <pc>.</pc>
  <ref target="#radicare">Radicare</ref>
  <pc>.</pc>
</xr>
```

Konačno, označeni su i prelasci u novi red (<lb/>), u novi stupac (<cb/>) i na novu stranicu (<pb/>).

Struktura natukničkog članka iz della Bellina rječnika označena je u skladu s TEI smjernicama, što je provjereno vanjskim DTD-om (*Document Type Declaration*), zasebnim dokumentom u kojem se definiraju pravila korištenja elemenata i atributa prema unaprijed dogovorenoj strukturi. U Prilogu se nalazi primjer cijelog natukničkog članka *Anello*, označenog odgovarajućim TEI oznakama.

Povezivanje autora izvora citata s vanjskim resursima

Elementom `<cit type="example" xml:lang="hr">` označeni su primjeri hrvatskih citata u della Bellinu rječniku, dok su elementom `<bibl>` označeni izvori tih citata. Elementom `<ref>` i vrijednošću atributa `@target` definiran je autor.

```
<cit type="example" xml:lang="hr">
  <quote>I prid bojnóm sabgljom gljutóm <lb/>
    Parsten pirni cjùse uljesti <pc>.</pc></quote>
  <bibl><ref target="gundulic">Ofm</ref><pc>.</pc></
bibl>
</cit>
```

U zaglavlju digitalnoga dokumenta nalazi se popis autora izvora citata koji se mogu naći u della Bellinu rječniku. U sljedećem primjeru vidljiv je način na koji je označen autor Ivan Gundulić:

```
<teiHeader>
...
<fileDesc>
...
<sourceDesc>
...
<listPerson>
  <person xml:id="gundulic">
    <persName>Ivan Gundulić
    <ptr target="http://www.enciklopedija.hr/natuknica.
aspx?ID=23793"/>
    <ptr target="https://hr.wikipedia.org/wiki/Ivan_
Gunduli%C4%87"/>
    </persName>
  </person>
</listPerson>
...
</sourceDesc>
...
```

</fileDesc>

...

</teiHeader>

XML dokument može se povezati elementom <ptr> s vanjskim resursima. U primjeru su dodane poveznice na mrežno izdanje natuknice »Gundulić, Ivan« *Hrvatske enciklopedije* Leksikografskog zavoda Miroslav Krleža te na natuknicu »Ivan Gundulić« hrvatske inačice *Wikipedije*. Budući da je tablica popisa autora i djela korištenih kao korpus za izradu rječnika napisana na talijanskom jeziku, potrebno je zabilježiti hrvatske inačice imena, kao što je to i zabilježeno u gornjem primjeru. Dodatni problem automatizaciji ovoga procesa stvaraju inačice imena koja danas više nisu uvriježena za određene autore. Naprimjer Ivan Gundulić je zabilježen kao »Ivan Franov Gundulić«, što se ne nalazi ni u jednom od dvaju navedenih mrežnih izvora kao oblik natukničkog članka. Iako se popis imena ne može lako automatizirati, proces ručnoga sastavljanja popisa nije trajao dugo jer je popis autora i njihovih djela ograničen.

Zanimljiv je podatak da je Gundulićev *Osman* tiskan gotovo stotinu godina nakon rječnika, pa su citati iz njega ispisani iz rukopisnih predložaka. Njihovo navođenje u della Bellinu rječniku prvo je objavljivanje njihova teksta (Sironić-Bonefačić 1993: 64).

Zaključak

U novije se vrijeme pokazala potreba za standardizacijom tehnologija, ali i cijelih procesa razvoja. Interoperabilnost se javlja kao ključni pojam u fazi digitalizacijskih projekata, čime se želi omogućiti međusobna komunikacija digitaliziranih resursa.

Kako bi se omogućila interoperabilnost povijesnih rječnika, rječnici moraju biti označeni određenim standardom. U radu je prikazano uspješno kodiranje natukničkog članka della Bellina rječnika iz XVIII. st. korištenjem sheme za kodiranje rječnika iz TEI (*Text Encoding Initiative*) smjernica koja predstavlja *de facto* standard u društveno-humanističkoj informatici. U slučaju da se riješe u radu spominjane nedoumice vezane za udaljeno označavanje, moguće je izvršiti višerazinsko označavanje natuknica prilagođavanjem i proširivanjem već postojećih XML dokumenata. Osim označavanja informacija unutar natuknica, navedena shema omogućava povezivanje s vanjskim resursima, kao što je u radu prikazano povezivanje s mrežnim izdanjem *Hrvatske enciklopedije* Leksikografskoga zavoda Miroslav Krleža i hrvatskom inačicom *Wikipedije*. U budućem istraživanju moguće je ovom metodom pove-

zivanje s dodatnim vanjskim resursima, kao što su *online* rječnici (npr. Hrvatski jezični portal²⁵), Linked Data²⁶ ili DBpedijom²⁷.

Formalizacijom prikaza rječničkih natuknica koristeći TEI oznake, omogućeno je upućivanje elemenata na LMF (*Lexical Markup Framework*) (Romary, Wegstein 2012). LMF²⁸ je ISO standard za označavanje leksikona u području obrade prirodnog jezika i strojno čitljivih rječnika. Označavanje resursa još jednim standardom, omogućit će se povezivanje s dodatnim leksikografskim resursima.

Dodatno je moguće buduća istraživanja usmjeriti prema što većoj automatizaciji svih navedenih koraka, bilo da se istraži mogućnost automatskog povezivanja s vanjskim resursima, bilo da se automatizira proces označavanja formaliziranih natuknica dodatnim standardima.

Prilog

Prikaz cijeloga izvornoga teksta della Bellina natukničkoga članka *Anello*, označavanje kojega je prikazano u ovom radu:

Anello , che fi porta in dito . *Anulus* , *li* . *m.* Parften , na . m. Anello fenza pietra . *Anu-* *lus purus* . Parften beş ofcza . Anello fpo- *falizio* . *Anulus pronubus* . Rukovni par- *ften* . Parften blagofovjeni . Şarucni par- *ften* . Şcenidbeni parften . Parften pirni . *I prid bojnom sabgljom gljutom* *Parsten pirni cjuse uljesti* . Ofm. Anello *d'oro* . *Anulus aureus* . Şlatni parften- *Luogo* , o parte dell'anello , in cui s'in- *cafra* la gemma . *Pala* , *æ . f.* Bok od- *parftena* . Işnofito od *parftena* . Anello in- *gemmato* , che ha la gemma . *Anulus gemmatus* . Parften s'ofczem , s'draghjem ka- *menom* . Anello da figillare . *Anulus fi-* *gnatorius* . Pecjatni parften . Parften pe- *cjatnik* . Anello da cucire . V. Ditale . *Anello* , o anelletto da cortina . *Anuli-* *cortinales* . Biocjuh , cjuha . m. Anello da- *catena* . *Catene anulus* . Biocjuh , ha . m. *Porfi* , o metterfi l'anello . *Anulum indue-* *re* . Staviti parften na ruku . Parften na- *taknuti* , ticem , nuòfam , illi natakòfam . *Levarfi* l'anello . *Deponere anulum* . Ski- *nuti* , kinivam , nuòfam . Staknuti , kni- *vam* , knuòfam . Smaknuti , knivam , knuò- *fam* parften s'ruke . Portar l'anello in- *dito* . *Digito anulum gestare* . Nofiti , nofim , *nofiòfam* parften na ruzi . Fare , o lavo- *rare* anelli . *Fabricare anulos* . Kovati , ku- *jem* , kovòfam parftene .

²⁵ <http://hjp.novi-liber.hr/>

²⁶ <http://linkeddata.org/>

²⁷ <http://wiki.dbpedia.org/>

²⁸ <http://www.lexicalmarkupframework.org/>

Prikaz natukničkoga članka *Anello*, koji je označen odgovarajućim TEI oznakama:

```
<entry>
  <form type="lemma" xml:lang="it">
    <orth>Anello</orth>
    <pc>,</pc>
  </form>
  <usg type="hint">
    che
    fi
    porta
    in
    dito
  </usg>
  <sense n="1">
    <cit type="translation" xml:lang="la">
      <quote>Anulus <pc>,</pc></quote>
      <form type="inflected">
        <gramGrp>
          <case value="genitive"/>
          <number value="singular"/>
          li<pc>.</pc>
        </gramGrp>
      </form>
      <gramGrp>
        <pos norm="noun"/>
        <gen value="m">m<pc>.</pc></gen>
      </gramGrp>
    </cit>
    <lb/>
    <cit type="translation" xml:lang="hr">
      <quote>Parfthen <pc>,</pc></quote>
      <form type="inflected">
        <gramGrp>
          <case value="genitive"/>
          <number value="singular"/>
          na<pc>.</pc>
        </gramGrp>
      </form>
```

```
<gramGrp>
  <pos norm="noun"/>
  <gen value="m">m<pc>.</pc></gen>
</gramGrp>
</cit>
</sense>
<sense n="2">
  <form type="compound" xml:lang="it">
    <orth><oRef/> fenza pietra <pc>.</pc></orth>
  </form>
  <cit type="translation" xml:lang="la">
    <quote>Anu<lb/>lus purus <pc>.</pc></quote>
  </cit>
  <cit type="translation" xml:lang="hr">
    <quote>Parften beš ofcza <pc>.</pc></quote>
  </cit>
</sense>
<sense n="3">
  <form type="compound" xml:lang="it">
    <orth><oRef/> fpo-<lb/>falizio <pc>.</pc></orth>
  </form>
  <cit type="translation" xml:lang="la">
    <quote>Anulus pronubus <pc>.</pc></quote>
  </cit>
  <cit type="translation" xml:lang="hr">
    <quote>Rukovni par-<lb/>ften <pc>.</pc></quote>
  </cit>
  <cit type="translation" xml:lang="hr">
    <quote>Parften blagofovjeni <pc>.</pc></quote>
  </cit>
  <cit type="translation" xml:lang="hr">
    <quote>Šarucni par-<lb/>ften <pc>.</pc></quote>
  </cit>
  <cit type="translation" xml:lang="hr">
    <quote>Šcenidbeni parften <pc>.</pc></quote>
  </cit>
  <cit type="translation" xml:lang="hr">
    <quote>Parften pirni <pc>.</pc></quote>
  </cit>
<lb/>
<cit type="example" xml:lang="hr">
```

```
<quote>I prid bojnom sabgljom gljutom <lb/>
  Parsten pirni cjùse uljesti <pc>.</pc></quote>
<bibl><ref target="gundulic">Ofm</ref><pc>.</pc></bibl>
</cit>
</sense>
<sense n="4">
  <form type="compound" xml:lang="it">
    <orth><oRef/> <lb/>d' oro <pc>.</pc></orth>
  </form>
  <cit type="translate" xml:lang="la">
    <quote>Anulus aureus <pc>.</pc></quote>
  </cit>
  <cit type="translate" xml:lang="hr">
    <quote>Šlatni parften <pc>.</pc></quote>
  </cit>
  <lb/>
</sense>
<sense n="5">
  <form type="compound" xml:lang="it">
    <orth>Luogo <pc>,</pc> o parte dell' <oRef/> <pc>,</pc> in cui
s' in-<lb/>caftra la gemma <pc>.</pc></orth>
  </form>
  <cit type="translate" xml:lang="la">
    <quote>Pala <pc>,</pc></quote>
    <form type="inflected">
      <gramGrp>
        <case value="genitive"/>
        <number value="singular"/>
        æ<pc>.</pc>
      </gramGrp>
    </form>
    <gramGrp>
      <pos norm="noun"/>
      <gen value="m">f<pc>.</pc></gen>
    </gramGrp>
  </cit>
  <cit type="translation" xml:lang="hr">
    <quote>Bok od <lb/> parftena <pc>.</pc></quote>
  </cit>
  <cit type="translation" xml:lang="hr">
    <quote>Işnofito od parftena <pc>.</pc></quote>
```

```
</cit>
</sense>
<sense n="6">
  <form type="compound" xml:lang="it">
    <orth><oRef/> in-<lb/>gemma <pc>,</pc> che ha la gemma
<pc>.</pc></orth>
  </form>
  <cit type="translate" xml:lang="la">
    <quote>Anulus gem-<lb/>matus <pc>.</pc></quote>
  </cit>
  <cit type="translate" xml:lang="hr">
    <quote>Parftens' ofczem <pc>,</pc> s' draghjem ka-<lb/>menom
<pc>.</pc></quote>
  </cit>
</sense>
<sense n="7">
  <form type="compound" xml:lang="it">
    <orth><oRef/> da figillare <pc>.</pc></orth>
  </form>
  <cit type="translation" xml:lang="la">
    <quote>Anulus fi-<lb/>gnatorius <pc>.</pc></quote>
  </cit>
  <cit type="translation" xml:lang="hr">
    <quote>Pecjatni parften <pc>.</pc></quote>
  </cit>
  <cit type="translation" xml:lang="hr">
    <quote>Parften pe-<lb/>cjatnik <pc>.</pc></quote>
  </cit>
</sense>
<sense n="8">
  <form type="compound" xml:lang="it">
    <orth><oRef/> da cucire <pc>.</pc></orth>
  </form>
  <xr type="see">V.
  <ref target="#ditale">Ditale</ref>
  <pc>.</pc>
  </xr>
  <lb/>
</sense>
<sense n="9">
  <form type="compound" xml:lang="it">
```

```
<orth><oRef/> <pc>,</pc> o anelletto da cortina <pc>.</pc></orth>
</form>
<cit type="translate" xml:lang="la">
  <quote>Anuli <lb/> cortinales <pc>.</pc></quote>
</cit>
<cit type="translate" xml:lang="hr">
  <quote>Biocjuh <pc>,</pc></quote>
  <form type="inflected">
    <gramGrp>
      <case value="genitive"/>
      <number value="singular"/>
      cjuha<pc>.</pc>
    </gramGrp>
  </form>
  <gramGrp>
    <pos norm="noun"/>
    <gen value="n">m<pc>.</pc></gen>
  </gramGrp>
</cit>
</sense>
<sense n="9">
  <form type="compound" xml:lang="it">
    <orth><oRef/> da <lb/> catena <pc>.</pc></orth>
  </form>
  <cit type="translation" xml:lang="la">
    <quote>Catenæ anulus <pc>.</pc></quote>
  </cit>
  <cit type="translate" xml:lang="hr">
    <quote>Biocjuh <pc>,</pc></quote>
    <form type="inflected">
      <gramGrp>
        <case value="genitive"/>
        <number value="singular"/>
        ha<pc>.</pc>
      </gramGrp>
    </form>
    <gramGrp>
      <pos norm="noun"/>
      <gen value="n">m<pc>.</pc></gen>
    </gramGrp>
  </lb/>
```

```
</cit>
</sense>
<sense n="10">
  <form type="compound" xml:lang="it">
    <orth>Porfi <pc>,</pc> o metterfi l' <oRef/> <pc>.</pc></orth>
  </form>
  <cit type="translation" xml:lang="la">
    <quote>Anulum indue-<lb/>re <pc>.</pc></quote>
  </cit>
  <cit type="translation" xml:lang="hr">
    <quote>Staviti parften na ruku <pc>.</pc></quote>
  </cit>
  <cit type="translation" xml:lang="hr">
    <quote>Parften na-<pb/>taknuti <pc>,</pc></quote>
    <form type="inflected">
      <pos norm="verb"/>
      <gramGrp>
        <per value="1"/>
        <number value="singular"/>
        <tns value="present"/>
        ticem <pc>,</pc>
      </gramGrp>
      <gramGrp>
        <per value="1"/>
        <number value="singular"/>
        <tns value="perfect"/>
        nuòfam <pc>,</pc>
      </gramGrp>
      ili
      <gramGrp>
        <per value="1"/>
        <number value="singular"/>
        <tns value="perfect"/>
        natakòfam <pc>.</pc>
      </gramGrp>
    </form>
  </cit>
</sense>
<sense n="11">
  <form type="compound" xml:lang="it">
```

```
<orth>Levarfi l' <oRef/> <pc>.</pc></orth>
</form>
<cit type="translate" xml:lang="la">
  <quote>Deponere anulum <pc>.</pc></quote>
</cit>
<cit type="translate" xml:lang="hr">
  <quote>Ski-<lb/>nuti <pc>,</pc></quote>
  <pos norm="verb"/>
  <form type="inflected">
    <gramGrp>
      <per value="1"/>
      <number value="singular"/>
      <tns value="present"/>
      kinivam <pc>,</pc>
    </gramGrp>
    <gramGrp>
      <per value="1"/>
      <number value="singular"/>
      <tns value="perfect"/>
      nuòfam <pc>.</pc>
    </gramGrp>
  </form>
</cit>
<cit type="translate" xml:lang="hr">
  <quote>Staknuti <pc>,</pc></quote>
  <pos norm="verb"/>
  <form type="inflected">
    <gramGrp>
      <per value="1"/>
      <number value="singular"/>
      <tns value="present"/>
      kni-<lb/>vam <pc>,</pc>
    </gramGrp>
    <gramGrp>
      <per value="1"/>
      <number value="singular"/>
      <tns value="perfect"/>
      knuòfam <pc>.</pc>
    </gramGrp>
  </form>
</cit>
```

```
<cit type="translate" xml:lang="hr">
  <quote>Smaknuti <pc>,</pc></quote>
  <pos norm="verb"/>
  <form type="inflected">
    <gramGrp>
      <per value="1"/>
      <number value="singular"/>
      <tns value="present"/>
      knivam <pc>,</pc>
    </gramGrp>
    <gramGrp>
      <per value="1"/>
      <number value="singular"/>
      <tns value="perfect"/>
      knuò-<lb/>fam <pc>.</pc>
    </gramGrp>
    parften s' ruke <pc>.</pc>
  </form>
</cit>
</sense>
<sense n="12">
  <form type="compound" xml:lang="it">
    <orth>Portar l' <oRef/> in <lb/> dito <pc>.</pc></orth>
  </form>
  <cit type="translation" xml:lang="la">
    <quote>Digito anulum gestare <pc>.</pc></quote>
  </cit>
  <cit type="translation" xml:lang="hr">
    <quote>Nofiti <pc>,</pc></quote>
    <pos norm="verb"/>
    <form type="inflected">
      <gramGrp>
        <per value="1"/>
        <number value="singular"/>
        <tns value="present"/>
        nofim <pc>,</pc>
      </gramGrp>
      <lb/>
      <gramGrp>
        <per value="1"/>
        <number value="singular"/>
```

```
<tns value="perfect"/>
nofiòfam <pc>.</pc>
</gramGrp>
parften na ruzi <pc>.</pc>
</form>
</cit>
</sense>
<sense n="13">
  <form type="compound" xml:lang="it">
    <orth>Fare <pc>,</pc> o lavo-<lb/>rare anelli <pc>.</pc></orth>
  </form>
  <cit type="translation" xml:lang="la">
    <quote>Fabricare anulos <pc>.</pc></quote>
  </cit>
  <cit type="translation" xml:lang="hr">
    <quote>Kovati <pc>,</pc></quote>
    <pos norm="verb"/>
    <form type="inflected">
      <gramGrp>
        <per value="1"/>
        <number value="singular"/>
        <tns value="present"/>
        ku-<lb/>jem <pc>,</pc>
      </gramGrp>
      <lb/>
      <gramGrp>
        <per value="1"/>
        <number value="singular"/>
        <tns value="perfect"/>
        kovòfam <pc>.</pc>
      </gramGrp>
      parftene <pc>.</pc>
    </form>
  </cit>
</sense>
</entry>
```

LITERATURA

- Bago**, Petra: *Označavanje strukture rječnika teksta prvog sveska drugog izdanja della Bellinog talijansko-latinsko-hrvatskog rječnika iz 1785.* (diplomski rad), Filozofski fakultet Sveučilišta u Zagrebu, 2008.
- Bago**, Petra, **Boras**, Damir: *A framework for consolidating most important digitized Croatian dictionaries from 1595 to 1945*, 17. International Information Technology and Journalism conference, Dubrovnik 2012.
- Common Language Resources and Technology Infrastructure (CLARIN)*. URL: <http://clarin.eu/> (pristupljeno 10. X. 2015).
- DBpedia*. URL: <http://dbpedia.org/About> (pristupljeno 8. X. 2015).
- della Bella**, Ardelio: *Dizionario italiano-latino-ilirico*. Nella Stamparia Privilegiata, Prima edizione, Ragusa 1785.
- digitalizirati*. Hrvatski Jezični Portal. Novi Liber. URL: http://hjp.novi-liber.hr/index.php?show=search_by_id&id=fldmXxQ%3D&keyword=digitalizirati (pristupljeno 10. V. 2015).
- digitize*. Oxford Dictionaries. Oxford University Press. URL: <http://oxforddictionaries.com/definition/english/digitize?q=digitise> (pristupljeno 10. V. 2015).
- Hilbert**, Martin, **López**, Priscila: The World's Technological Capacity to Store, Communicate, and Compute Information. *Science* 332(2011) 6025, str. 60–65.
- Hrvatski jezični portal*. URL: <http://hjp.novi-liber.hr/> (pristupljeno 8. X. 2015).
- eXtensible Markup Language (XML)*. URL: <http://www.w3.org/XML/> (pristupljeno 10. VII. 2014).
- Fostering Language Resources Network (FLaReNet)*. URL: <https://www.flarenet.eu/> (pristupljeno 12. V. 2015).
- Ide**, Nancy, **Pustejovsky**, James, **Calzolari**, Nicoletta, **Soria**, Claudia: The SILT and FLaReNet International Collaboration for Interoperability. *Proceedings of the Third Linguistic Annotation Workshop*, ACL-IJCNLP 2009, str. 178–181.
- Lexical Markup Framework*. URL: <http://www.lexicalmarkupframework.org/> (pristupljeno 10. X. 2015).
- Linked Data – Connect Distributed Data Across the Web*. URL: <http://linkeddata.org/> (pristupljeno 8. X. 2015).
- Network for Digital Methods in the Arts and Humanities (NeDiMAH)*. URL: <http://www.nedimah.eu/> (pristupljeno 10. X. 2015).
- Piotrowski**, Michael: *Natural Language Processing for Historical Texts*. Edited by Graeme Hirst. Synthesis Lectures on Human Language Technologies 17. Morgan & Claypool Publishers, 2012.
- Prijedlog Nacionalnog programa digitalizacije arhivske, knjižnične i muzejske građe*. Radna grupa za digitalizaciju arhivske, knjižnične i muzejske građe Ministarstva kulture Republike Hrvatske, 2006. URL: <http://daz.hr/bastina/NacionalniProgramDigitalizacije.pdf> (pristupljeno 10. V. 2015).
- Pustejovsky**, James, **Ide**, Nancy: What Does Interoperability Mean, Anyway? Toward an Operational Definition of Interoperability. *Proceedings of the Second International Conference on Global Interoperability for Language Resources*, Hong Kong 2010. URL: <http://www.cs.vassar.edu/~ide/papers/ICGL10.pdf> (pristupljeno 15. V. 2015).
- Romary**, Laurent, **Wegstein**, Werner: Consistent Modeling of Heterogeneous Lexical Structures. *Journal of the Text Encoding Initiative* [Online] 3(2012). URL: <http://tei.revues.org/540> (pristupljeno 15. X. 2015).
- Seiter-Šverko**, Dunja, **Križaj**, Lana: Digitalizacija kulturne baštine u Republici Hrvatskoj: od trenutne situacije prema nacionalnoj strategiji. *The 2. NAPLE konferencija: Uloga narodnih knjižnica u nacionalnim politikama izgradnje društva znanja (Digitalizacija kulturne baštine i provedba multikulturalizma u knjižnicama)*, Supetar 2005. URL: <http://www.min-kulture.hr/naple/hrv/index.htm>

- Sironić-Bonefačić**, Nives. *Analiza talijanske i hrvatske leksičke građe u rječniku Ardelija della Belle: Dizionario italiano, latino, illirico. Venezia, 1728*. (doktorska disertacija). Zagreb 1986.
- Sironić-Bonefačić**, Nives. DELLA BELLA, Ardelio. *Hrvatski biografski leksikon*. Leksikografski zavod Miroslav Krleža, Zagreb 1993. URL: <http://hbl.lzmk.hr/clanak.aspx?id=4507> (pristupljeno 31. VIII. 2014).
- Sustainable Interoperability for Language Technology*. URL: <http://www.anc.org/LAPPS/SILT/index.html> (pristupljeno 12. V. 2015).
- TEI Consortium (ur.): »9. Dictionaries«. *TEI P5: Guidelines for Electronic Text Encoding and Interchange* (Verzija 2.6.0). URL: <http://www.tei-c.org/Guidelines/P5/> (pristupljeno 17. V. 2015).
- TEI Consortium (ur.): »16. Linking, Segmentation, and Alignment«. *TEI P5: Guidelines for Electronic Text Encoding and Interchange* (Verzija 2.6.0). URL: <http://www.tei-c.org/Guidelines/P5/> (pristupljeno 15. X. 2015).
- TEI Consortium (ur.): »iv. About These Guidelines«. *TEI P5: Guidelines for Electronic Text Encoding and Interchange* (Verzija 2.6.0). URL: <http://www.tei-c.org/Guidelines/P5/> (pristupljeno 15. V. 2015).
- TEI Consortium (ur.): *TEI P5: Guidelines for Electronic Text Encoding and Interchange* (Verzija 2.6.0). URL: <http://www.tei-c.org/Guidelines/P5/> (pristupljeno 17. V. 2015).
- TEI Consortium (ur.): »v. A Gentle Introduction to XML«. *TEI P5: Guidelines for Electronic Text Encoding and Interchange* (Verzija 2.6.0). URL: <http://www.tei-c.org/Guidelines/P5/> (pristupljeno 15. V. 2015).

MARKING DICTIONARY ENTRIES IN DELLA BELLA'S DICTIONARY AND LINKING AUTHORS OF THE QUOTATION SOURCES TO EXTERNAL RESOURCES

Petra Bago

University of Zagreb, Faculty of Humanities and Social Sciences

ABSTRACT: The digitization process of historical texts is increasing in the last several decades. However, the digitization projects are usually isolated to the project teams, universities and institutes, but also to individuals developing the project. There was a lack of communication between the community members, which resulted in resources, tools and systems that are not able to exchange information. Consequently, recently it was demonstrated that there is demand for standardization of technologies, but also all processes of development. Interoperability emerges as the key concept at this phase of digital humanities, which aims to facilitate communication of data. In this paper we present encoding scheme of dictionary entries of della Bella's 18th century dictionary using the TEI (*Text Encoding Initiative*) standard and a method of linking to external resources (web edition of *Croatian encyclopedia* of the Miroslav Krleža Institute of Lexicography and the Croatian version of *Wikipedia*).

Keywords: *historical dictionaries; interoperability; encoding; Text Encoding Initiative; digital humanities; Ardelio della Bella*