

Digitalizacija korpusa starohrvatskih tekstova i kritika teksta

Amir Kapetanović

Institut za hrvatski jezik i jezikoslovlje
Ulica Republike Austrije 16, Zagreb, Hrvatska
akapetan@ihjj.hr

Summary

Digitalization of Old Croatian texts was initiated within the framework of the project “The Old Croatian Dictionary”. The purpose of the initiative was to make a dictionary that would lexicographically describe the lexis of the oldest periods of Croatian literacy in the Croatian language (from first records to the end of the 15th century). Digitalization of Old Croatian texts is more complex than digitalization of contemporary texts, not only owing to predominantly non-authored corpus in manuscript form, but also owing to the fact that the collected Old-Čakavian / Old-Štokavian corpus, originally written in three scripts, first needs to be critically analyzed. The Croatian medieval texts have been published since the 19th century in different forms (photograph, transcript, transliteration, transcription), and the publications have varying degree of quality. In addition, some texts that may become integral parts of the corpus have not been published yet. The paper will present the corpus structure, and major issues and principles in corpus analysis and design.

Key words: digitalization, corpus, the Old-Croatian language, Middle Ages, lexicography, textual criticism

Uvod

Danas, u vremenu sveopće digitalizacije, oko nas nastaju različiti tekstovni arhivi (digitalne biblioteke) i korpusi različite kvalitete. Pri dohvaćanju neke aproksimativno određene milijunske brojke u utrci za kolikoćom pojavnica određenoga korpusa kadšto se zaboravi na kakvoću primarnih korpusnih podataka. Jamačno kvaliteta korpusa, koji se sastavlja s ciljem njegova iskorištavanja u lingvističke svrhe, ovisi ponajprije o valjanoj digitalizaciji odabrane građe. U toj prvoj fazi rada u izradi korpusa moraju se primarni podatci (tekstovi) kvalitetno obraditi i pohraniti, a tek potom može se korpus nadograđivati metapodacima.

Put do označenoga (anotiranoga) korpusa ipak je kraći i lakši ako stvaramo *korpus suvremenoga jezika* nego *jezičnopovijesni korpus*. Prilikom izrade korpusa suvremenoga jezika mora se oprezno pristupiti izboru reprezentativnih inačica

tekstova (ako je poznato nekoliko autoriziranih izdanja teksta) i vjerno iz izvornika prenijeti primarne podatke u elektronički oblik (ako je riječ o tekstovima iz neelektroničkih medija). Ako kanimo sastaviti jezičnopovijesni korpus, priprema je za digitalizaciju veoma zahtjevna. Primarni podatci u takvu korpusu, za razliku od korpusa suvremenoga jezika, nisu *izvorni*, nego *posredni* (prijepis ili transliteracija ili transkripcija izvornih tekstova). Primjerice, digitalne slike izvornih srednjovjekovnih rukopisnih vrela hrvatske pisane baštine mogu poslužiti samo za dokumentaciju. OCR-programi do danas nisu tako usavršeni da bi mogli izvršiti prepoznavanje rukom pisanih znakova, osobito u hrvatskim srednjovjekovnim tekstovima izvorno pisanim glagoljicom, ćirilicom i starom latinicom. Nijedan stroj danas ne može osim *prepoznavanja* znakova izvršiti i filološku *interpretaciju* znakova, a bez toga se ne može izgraditi korpus baštinskih tekstova pisanih trima pismima, nenormiranim pravopisom i slovopisom. Posredni primarni podatci mogu se naći u dosadašnjim izdanjima baštinskih tekstova (preslovljeni izvori), ali mnoga izdanja u manjoj ili većoj mjeri nisu pouzdana. Zato je potrebno vratiti se opet originalima i provjeravati valjanost preslovljenih izvora. Ako se jezičnopovijesni korpus izgrađuje sa zadatkom da omogućiti izradu kakva povijesnoga rječnika, treba predvidjeti koje informacije za leksikografsku obradu u konačnici mora pružati korpus.

Povijest projekta

U Institutu za hrvatski jezik i jezikoslovlje (IHJJ) u Zagrebu izgrađuje se Korpus starohrvatskih tekstova (KST) koji će u prvom redu biti upotrijebljen za izradu *Starohrvatskoga rječnika*. Hrvatska filologija nema ni pokusni računalno čitljiv korpus starohrvatskih tekstova ni na suvremenim leksikografskim načelima izrađen rječnik koji bi nudio informacije o hrvatskom leksiku starijih razdoblja. Mažuranićevi *Prinosi za hrvatski pravno-povjestni rječnik* obuhvatili su samo pravna vrela, a za *Akademijin rječnik* (AR) iskorišteni su kao izvori samo neki starohrvatski tekstovi iz kojih nije ekscerpiran sav leksik. Danas jedino AR u hrvatskoj filologiji ima ulogu opsežnoga papirnoga korpusa (obilno su navođene potvrde), premda je to vrlo varljiva baza podataka, osobito ako se u njemu traže podatci o prvim ili jedinim potvrdama određenoga leksema. Tako je, primjerice, prema *Akademijinu rječniku* topomim *Evropa/Europa* potvrđen u hrvatskim tekstovima tek od 17. stoljeća¹, a prividni hapakslegomenon *taržan* rezultat je nedovoljno proučene grafije Zoranićevih *Planina*².

Izradu rječnika najstarijih tekstova otpočeli su neki slavenski narodi u 20. stoljeću (npr. nedovršeni dvotomni Gebauerov *Slovník staročeský*, nedavno dovršeni višetomni *Staropolski słownik*), a zamisao o starohrvatskom rječniku po-

¹ Uvid u starohrvatske tekstove, kao što su različiti hrvatski prijevodi *Lucidara*, pokazuju da se potvrđenost toga toponima može pomaknuti barem u 15. stoljeće; o tome v. Kapetanović 2005: 3.

² O tome v. Kapetanović 2006: 23.

tječe iz ranih 90-ih godina 20. stoljeća, kada se u tadašnjem Zavodu za hrvatski jezik razmišljalo o izradi *Rječnika hrvatskoga srednjovjekovlja* (Malić 1998: 61). Godine 1998. dr. Dragica Malić ponudila je u jednom članku skicu za izradu rječnika, a 2001. i nacrt s popisom izvora i pravilima leksikografske obrade u knjižici *Nacrt za Hrvatski rječnik do Marulića i njegovih suvremenika*. Od te godine do 2005. obavljena su u IHJJ-u preliminarna istraživanja pod vodstvom dr. D. Malić, koja su ponajviše bila usmjerena na prikupljanje i provjeravanje predviđenih izvora za rječnik kako bi se moglo otpočeti s izgradnjom strojno čitljiva korpusa. U tim su se istraživanjima mnogi predviđeni izvori (transkripcije i transliteracije starih tekstova) pokazali nepouzdanima i moralo se ulagati mnogo truda, znanja i vremena u provjeravanje nekih izvora prema originalima. Taj opsežni dio posla nije dovršen. Unatoč tim spoznajama i rezultatima intenzivnoga trogodišnjega istraživanja malobrojnih suradnika, projekt je kritiziran na temelju polaznoga popisa izvora: “*Hrvatski rječnik do Marulića i njegovih suvremenika*, izrađen na korpusu samo objavljene građe, ne bi mogao biti temeljno djelo hrvatske filologije. Jer, kao što je već istaknuto, objavljen je tek manji dio građe. K tomu objavljena je građa – ponavljam – dobrim dijelom nepouzdana i nužno je usporediti je s originalima. Pa kad je tako, uputno bi bilo građu ekscerpirati iz originala ili njihovih preslika, kako su činili priređivači Akademijina povijesnoga *Rječnika*. (Nazor 2005: 453)”

Drugim riječima, time je predloženo da se na početku 21. stoljeća leksik iz originala ili preslika ekscerpira kao što se nekoć kadšto činilo pri izradi AR-a i uvijek pri izradi *Rječnika crkvenoslavenskoga jezika hrvatske redakcije*³. Na taj način stvarao bi se *papirni korpus* (ili u modernijoj verziji: *računalna zbirka potvrda*), a ne strojno čitljiv korpus starohrvatskih tekstova. Takav korpus ne bi mogao biti podlogom iscrpne leksikografske obrade, otežano bi bilo provjeravanje i povezivanje podataka prilikom leksikografske obrade i teško bi se taj korpus mogao u budućnosti iskorištavati za druga deficitarna jezičnopovijesna istraživanja hrvatskoga jezika u IHJJ-u i izvan njega. I u skicama i nacrtima za rječnik njegova se izrada povezivala s izradom računalno čitljiva korpusa: “Svi će se izvori konkordirati, što znači da će sve potvrđene riječi ući u obradu. Time će se izbjeći samovoljan odabir riječi, kojim se npr. uvelike odlikuje *Akademijin rječnik*.” (Malić 1998: 68); “Treba nabaviti/napraviti računalni program za konkordiranje izvora i za optimalno iskorištavanje građe, odnosno za stvaranje korpusa izvora kao osnove za leksikografsku obradu.” (Malić 2002: 152)

³ O ekscerpiranju za *Rječnik crkvenoslavenskoga jezika hrvatske redakcije* osnovnom i komparativnom metodom v. u uvodu toga rječnika na str. III.

Planovi i opis korpusa

Rad na izradi korpusa nastavljen je 2007. u IHJJ-u u sklopu projekta *Starohrvatski rječnik* (voditelj A. Kapetanović), koji je dio institutskoga programa *Hrvatska jezična riznica* (voditelj Dunja Brozović Rončević), uz potporu Ministarstva znanosti, obrazovanja i športa Republike Hrvatske. Polazni popis vrela koji je izradila dr. Malić reduciran je s obzirom na gornju granicu starohrvatskoga jezika (u korpus ulaze izvori do 1500, uz neke iznimke mladih prijepisa). Sustavno se pristupilo izradi kritičkih izdanja hrvatskih srednjovjekovnih tekstova i njihovoj digitalizaciji. Odlučeno je da se pri izgradnji KST-a primijene isti ali korpusu prilagođeni računalni standardi i sustavi kao i u izradi ostalih korpusa u okviru današnje *Hrvatske jezične riznice* IHJJ-a, kako bi u budućnosti KST mogao postati potkorpusom krovnoga hrvatskoga jezičnoga korpusa u IHJJ-u. Do tada će KST funkcionirati kao samostalan korpus, razdijeljen na sastavnice prema vrstama tekstova, u čemu se ogleda i struktura građe budućega rječnika:

SASTAVNICE	IME	VRSTE TEKSTOVA
SASTAVNICA 1	<i>Pjesništvo i drama</i>	pjesme, plačevi, prikazanja
SASTAVNICA 2	<i>Proza I</i>	djelomični prijevodi Biblije, molitvenici, lektionari
SASTAVNICA 3	<i>Proza II</i>	apokrifi, vizije, čudesa Marijina, svetačke legende, romani, priče, poučna proza...
SASTAVNICA 4	<i>Pravni tekstovi</i>	isprave, statuti, zakoni, urbari, regule
SASTAVNICA 5	<i>Povijesni tekstovi</i>	ljetopisi, kronike
SASTAVNICA 6	<i>Kratki napisi</i>	natpisi, grafiti, različiti zapisi i bilješke
SASTAVNICA 7	<i>Pabirci</i>	glose, izbor tipičnoga hrvatskoga općega leksika iz crkvenoslavenskih vrela i starohrvatskih imena i titula iz hrvatskih latinskih isprava

Raznolikost tekstova razvrstanih u sedam sastavnica osigurava jednu drugu ne manje važnu kvalitetu korpusa (uzorkovanje različitih tipova i vrsta pisanih vrela). Korpus će u konačnici biti tako označen da se može pretraživati čitav korpus ili samo dijelovi korpusa (po određenim sastavnicama, po stoljećima: 11-15, po starohrvatskim književnim idiomima: čakavski/štokavski). U pretposljednjoj navedenoj sastavnici bit će dosta zapisa koji nisu veći od sintagme i jedne rečenice. Budući da su sintagma i jedna rečenica premaleni korpusni uzorci, morat će se odlučiti kako će takva građa, nastala u rasponu od pet stoljeća, biti organizirana. Posljednja navedena sastavnica razlikovat će se od svih ostalih jer neće biti sastavljena od cjelovitih tekstova. To će biti pomoćna sastavnica koja će se iskorištavati u leksikografskoj obradi ako ne bude (dovoljno) potvrda iz hrvatskih vrela. Istina je da u hrvatskom srednjovjekovlju nije postojao "čvrst zid" između hrvatskih narodnih idioma i crkvenoslavenskoga jezika, a to se odnosi ponajviše na proznu beletristiku. Tekstovi koji su pisani pretežno crkvenoslavenskim jezikom, neće se uzimati kao vrela, a crkvenoslavizmi

u tekstovima pisanim pretežno hrvatskim obrađivat će se u *Starohrvatskom rječniku* ravnopravno kao i drugi *izmi* (talijanizmi, germanizmi, bohemizmi...). Tako će u korpusu i u rječniku biti zastupljen neizbrisiv trag utjecaja crkvenoslavenskoga jezika na hrvatske književne idiome srednjovjekovlja. Zapisi starohrvatskih imena i titula u hrvatskim latinskim ispravama navodit će se u rječniku kao rekonstrukcije s izvornim zapisom, npr. **Budac* (Budeč, Budicius, Budičo, Budiz, Budizo).

Korpus starohrvatskih tekstova mogao bi se ovako definirati: računalno čitljiv, označen, jednojezični, nespecijalni statični povijesni korpus pisanoga jezika sastavljen od cjelovitih tekstova (osim pomoćne sastavnice br. 7). Što se raspona korpusa tiče (od prvih zapisa do 1500), i tu postoji problem s gornjom vremenskom odrednicom. Bit će iznimaka u vezi s nekim tekstovima, pa će tako ući u korpus neki srednjovjekovni tekstovi sačuvani samo u prijepisima iz 16. stoljeća. Nezamislivo je, primjerice, da *Vinodolski zakon* sastavljen 1288. (sačuvan u najstarijem prijepisu iz 16. stoljeća) ne bude dio korpusa jer on predstavlja važno vrelo starohrvatske pravne terminologije. Opseg korpusa (broj pojavnica) teško je za sada predvidjeti, ali se prema grubim procjenama može pretpostaviti da bi *Starohrvatski rječnik* mogao sadržati oko 50 000 natuknica.

Prijenos informacija i izbor načina transponiranja primarnih podataka

Andrea Hofmeister analizirala je u jednom teorijski usmjerenom tekstološkom prilogu (2005) odnos “nalaza” (*Befund*) i “tumačenja” (*Deutung*) s obzirom na polja informacija i stupnjeve (vrste) izdanja. Na primjer, ona posve točno zaključuje kada kaže da bazična transliteracija kumulativno sadrži informacije o pisanom tekstu svih polja (od paleografskoga do semantičkoga), ali upravo zbog toga prijeti nepreglednost i manja pristupačnost korisniku (*benutzerfreundlich*).⁴ Ipak se ne možemo složiti s njezinom tvrdnjom da originalni sadržaj informacija može ponuditi jedino rukopis sam ili visokokvalitetni faksimil rukopisa jer faksimil ne može za razliku od originala pružiti paleografima podatke o vodenim žigovima, što nije nevažna informacija istraživaču koji želi datirati rukopis. A. Hofmeister u svojem radu nudi shemu odnosa polja informacija i stupnjeva (vrsta) izdanja, a mi ćemo ovdje ponuditi sličnu, prilagođenu našim potrebama i terminologiji (umjesto polja informacija navest ćemo izbor nekih konkretnih informacija koje se posreduju različitim načinima transponiranja teksta):

4 V. A. Hofmeister 2005: 5-6.

POSREDOVANJE INFORMACIJA U RAZLIČITO TRANSPONIRANOM TEKSTU					
	<i>original</i>	<i>snimka</i>	<i>prijepis</i>	<i>transliteracija</i>	<i>transkripcija</i>
materijal za pisanje	+	–	–	–	–
oblik slova i duktus	+	+	–	–	–
izvantekstni znaci	+	+	–	–	–
kontinuirano pisanje	+	+	+/-	–	–
slovopis	+	+	+	+	–
jezik	+	+	+	+	+

Takva editorička pitanja nisu nevažna za korpusnu lingvistiku. Osiguravanje zalihosnih informacija ne samo da je neekonomično u ciljanim istraživanjima (nepotreban utrošak vremena i novca) nego može pomutiti i transparentnost podataka te tako smanjiti razinu primanja važnih informacija. Stoga, nije nevažno hoćemo li primarne podatke u korpus prenositi transliteracijom, fonološkom transkripcijom, prijepisom⁵.

	<i>snimka</i>	<i>prijepis</i>	<i>lat. transliteracija</i>	<i>lat. transkripcija</i>
<i>glagoljica</i>	ТВОЈА ТВОЈА ТВОЈА ТВОЈА	ТВОЈА ТВОЈА ТВОЈА ТВОЈА	tvoč t'voa tvoia tvoja	tvoja ⁶
<i>ćirilica</i>	ТВОЈА ТВОЈ ТВОЈА ТВОЈА	ТВОЈА ТВОЈ ТВОЈА ТВОЈА	tvoja tvoč tvoa tvoja	
<i>latinica</i>	TVOIA tuoa tuoija tuoya	<i>TVOIA</i> <i>tuoa</i> <i>tuoija</i> <i>tuoya</i>		

⁵ *Transliteracija* (lat. *translitteratio*) označuje postupak prenošenja znakova jednoga pisma adekvatnim znakovima drugoga pisma, npr. *краль* > *kral'*. *Fonološka transkripcija* (lat. *transcriptio*) označuje postupak kojim se grafemi za određene foneme iz jednoga slovopisa jednoga pisma prenose u slovopis istoga ili drugoga pisma, npr. iz 13. stoljeća *краль* > *kralj*. *Prijepis* označuje vjerno i dosljedno prenošenje znakova originalnoga slovopisa jednoga pisma bez konverzija i interpretacije znakova, npr. *краль* > *краль*.

⁶ Komentar tablice: Vodoravno su ispisani načini prenošenja izvornoga teksta, okomito su poredana hrvatska pisma. Fingira se zapis riječi *tvoja* različitim pismima (uglavnom "mladi" načini zapisa glagoljicom i ćirilicom), kao što "snimka" "originala" pokazuje. Potom se zapisi prepisuju

Odluka o tome mora se donijeti prije otpočinjanja digitalizacije jer su posljedice te odluke dalekosežne. Mora nam biti jasno koje informacije želimo prenijeti i kojem krugu korisnika (užem/širem). Budući da se KST izrađuje s ciljem istraživanja *jezika* (najprije leksika, potom gramatike), a ne različitih slovopisa hrvatskih triju pisama, sve potrebne informacije može ponuditi fonološka transkripcija. Osmišljena su unutar projekta pravila transkripcije za izradu KST-a jer ih za srednjovjekovne tekstove nijedan hrvatski tekstološki priručnik do danas nije ponudio.⁷ Odbacivanjem zalihosnih informacija olakšava se računalna obrada podataka i ne ugrožava se transparentnost jezičnih podataka u korpusu i u rječniku. Na primjer, zbog višenamjenskoga *s* u staroj hrvatskoj latinici zapis *kosa* mogao bi se ovisno o kontekstu transkribirati kao *koža / koza / kosa* (posljednja mogućnost može odražavati tri homonima). Ako se primjer ne bi transkribirao, korisnik korpusa i rječnika morao bi se samostalno upustiti u filološku interpretaciju potvrda *kosa* bez uvida u grafijski sustav čitava teksta iz kojega je zapis. Promjena originalnoga slovopisa i pisma nije krivotvorenje jezika. Obično se u tradicionalnoj filologiji transliteracija povezuje sa znanstvenim, a transkripcija s popularnim izdanjima tekstova⁸, ali se pritom zaboravlja da su za znanstvena istraživanja neiskoristive one transliteracije koje sadrže puno pogrešaka i krivih interpretacija znakova, a one transkripcije koje maksimalno čuvaju izvorni jezik teksta i koje prati kritički aparat, mogu biti korisne za znanstvena istraživanja, kao npr. neka izdanja u *Starim piscima hrvatskim*. Loša su ona izdanja starih tekstova u kojima se miješaju dva postupka: transliteracije s primjenom nekih transkripcijskih načela (npr. transliterira se *kral* > *kral*, ali *Maria* > *Marija*) i transkripcije s nekim transliteracijskim načelima (npr. transkribiranje: *moji očči* = posebno markiranje grafema *s* obzirom na to kako je fonem zapisan u izvorniku i čuvanje ortografskih geminata).

(“prijepis”) i pritom tip pisma ne mora biti isti kao na “snimci”. Nadalje se izvodi latinička transliteracija svih zapisa osim latiničkih, jer “transliteracija iz latinice u latinicu nije moguća, tj. ona bi bila čisti preslik originala u grafemskom smislu” (Damjanović 1998: 61; Vončina 1999: 25). Teorijski bi bilo moguće ćirilicom i latinicom transliterirati glagoljički tekst i ćirilički tekst glagoljicom i latinicom, ali danas to više nije uobičajeno (o tome v. Bratulić 1981). Različito zapisana riječ *tvoja* (ovdje na 12 načina, i to nisu svi mogući) svedena je latiničkom transkripcijom na samo jedan zapis, i to tako da jezik ni u jednoj pojedinosti nije promijenjen.

⁷ U jedinom hrvatskom tekstološkom priručniku (Vončina 1999) ne daju se pravila transkripcije srednjovjekovnih tekstova, premda primjeri iz nekoliko rukopisa pokazuju što bi mogla biti uzorna transkripcija. D. Kapetanić ne bavi se u svojem članku (1970) editoričkim problemima srednjovjekovnih rukopisa. Pomoć se ne može naći ni u drugim južnoslavenskim priručnicima, na primjer, knjiga srpskoga tekstologa D. Ivanića (2001) nema pravila transkripcije i ne bavi se starosrpskim tekstovima.

⁸ Na primjer: “Latiničku transkripciju ili transliteraciju prihvaćali su kasnije i drugi izdavači [poslije I. Črnčića, op. A. K.], a danas je već postalo redovitom praksom da se hrvatskoglagoljski tekstovi izdaju u latiničkoj transliteraciji ili transkripciji, ovisno o tome da li se radi o znanstvenom ili popularnom izdanju.” (Grabar 1978: 40)

Leme u korpusu i u rječniku također će biti transkribirane (normalizira se grafija, a ne jezik) jer je zbog različitosti idioma, slovopisa i pisama nemoguće stvoriti grafijski i jezično idealnu⁹ natuknicu (npr. dvojna natuknica *meja/međa* predstavljat će zapise *meiam*, *meyu*, *medji*, *medyom*...). Osim interpretacije rad na korpusu predviđa i rad na dokumentaciji, i to će biti dodatne informacije u računalno pretraživu korpusu. Planira se izrada opisa svih vrela i povezivanje digitalnih slika izvornika s transkripcijama. Na taj način nadoknadila bi se redukcija informacija za filologe koji žele provjeriti valjanost transkribiranoga zapisa. Osim toga, korpus će u konačnici imati i statističke podatke o čestotnosti.

Srednjovjekovni tekst i nova kritika teksta

Izrada KST-a i *Starohrvatskoga rječnika* treba se suočiti ne samo s primjenom novih računalnih tehnologija¹⁰ nego i s novim odnosom prema srednjovjekovnom tekstu i novom kritikom teksta koju je razvila tekstologija/editorika 20. stoljeća.

KST temelji se uglavnom na rukopisnoj srednjovjekovnoj baštini (iznimka su inkunabule). Srednjovjekovni tekst nema autora (s biografijom) i sve su njegove inačice ravnopravne.¹¹ Autorstvo u modernom smislu nije starije od konca 18. stoljeća, stoga tekstovi korpusa nisu autorizirani i nema autografa. U srednjovjekovnim anonimnim tekstovima autorska bi se svijest doduše mogla tražiti u određenim slojevima tekstova¹², ali, dakako, ne autorstvo kako ga shvaćamo od početka 19. stoljeća. Ne može se ipak reći da su srednjovjekovni ljudi zazirali od isticanja vlastitih imena; pokazuju to brojni hrvatski srednjovjekovni glagoljički natpisi i grafiti od Istre do Dubrovnika (v. Fučić 1982).

Srednjovjekovni je tekst, pisan narodnim jezicima, nestabilan i otvoren, javlja se u brojnim varijantama i verzijama jer se prevodi, kompilira, prepisuje, prerađuje i širi bez kontrole prvotnoga sastavljača. Egzistira u svojim mijenama sve dok se širi prepisivanjem ili usmenim prenošenjem. Jedina iznimka je biblijski tekst. U srednjem vijeku važnije je bilo što više umnožiti tekst prepisivanjem, manje se u to doba pazilo na kritiku teksta, kao u antici ili u renesansi. Otkriće

⁹ U *Rječniku crkvenoslavenskoga jezika hrvatske redakcije* natuknice se navode "u normaliziranom ('idealnom') liku crkvenoslavenskoga jezika hrvatske redakcije". V. prvi svezak rječnika (2000), str. VIII. i dalje.

¹⁰ Novija hrvatska stručna (teorijska) literatura vezana uz probleme izrade korpusa usmjerena je na izradu korpusa suvremenoga hrvatskoga jezika, npr. Tadić 2003. Mora se reći da je nedovoljno hrvatskih priručnika s postupnim uvođenjem u probleme korpusne lingvistike, kao što su npr. dva novija njemačka: Scherer 2006; Lemnitzer–Zinsmeister 2006. U posljednjem navedenom priručniku postoji popis i osnovni opis različitih njemačkih korpusa i arhiva, v. Lemnitzer–Zinsmeister 2006: 113-126.

¹¹ Diskusiju o tome v. npr. u Stackmann 1994: 412 i dalje.

¹² Kao što to pokazuje R. Schnell (1998) kritikom neofiloloških stajališta o *autoru* i o *djelu* u njemačkom srednjovjekovlju.

tiska uvelike je promijenilo odnos prema tekstu. Riječima B. Cerquiglinia: “Izum knjigotiska je puno više mentalna nego tehnička revolucija.” (Cerquiglini 2005: 125; prev. A. K.)

Varijantni srednjovjekovni tekstovi nekoć su se promatrali kao i novovjekovni autorski tekstovi. Danas je realno prikazivanje teksta (izvora) zamijenilo neka- dašnje *rekonstruiranje arhetipa* teksta (na temelju svih rukopisa sastavljao se jedan koji bi mogao biti najbliži originalu) ili konstruiranje *idealnoga* teksta. U novijoj tekstologiji smatra se da su svi sačuvani oblici nekoga teksta (od njegova nacrtu do tiskanoga izdanja) ravnopravne inačice jer svaka pokazuje određeni stupanj razvoja teksta. Zato će različite varijante i verzije srednjovjekovnih tekstova ravnopravno ući kao izvori u KST.

Mnoga dosadašnja izdanja srednjovjekovnih tekstova nisu loši izvori za KST samo zato što su dosadašnji njihovi “čitači” činili omaške i prezentirali ih u svojim izdanjima bez ujednačenih tekstoloških načela nego i zato što su i neka pouzdana izdanja izrađena “starom” kritikom teksta. Riječ je o izdanjima u kojima se, primjerice, rukopis iz 15. stoljeća ispravljao prema prijepisu iz 17. stoljeća ili su se lakune srednjovjekovnih rukopisa popunjavale prema vrlo mladim prijepisima.

U KST-u će dakle ravnopravno biti zastupljene sve poznate inačice jednoga hrvatskoga srednjovjekovnoga teksta nastale do konca 15. (ili početka 16. stoljeća). Na primjer, srednjovjekovna pjesma *Bog se rodi v Vitliomi* bit će u korpusu predstavljena ravnopravno dvjema inačicama (jedna iz *Pariške pjesmarice*, druga iz *Zbornika duhovnoga štiva IV a 92*)¹³: neće se od dvije inačice izradivati jedan kontaminiran tekst s namjerom da se rekonstruira arhetip te pjesme i neće se pristupiti “popravljanju” tih dviju najstarijih inačica na temelju njihovih mlađih prijepisa iz 17. i 18. stoljeća. Ako se budu činile nužne emendacije i konjektуре, svaka od njih morat će biti dobro argumentirana i dokumentirana.

Sve nam to govori da će kvaliteta KST-a i *Starohrvatskoga rječnika* ovisiti o koordiniranoj suradnji računalne lingvistike, tekstologije/editorike i leksikografije.

Zaključak

U članku je usmjerena pozornost na kritiku teksta prilikom izrade korpusa, i to u prvoj fazi rada: prikupljanje primarnih podataka za korpus. U izradi KST-a to je otežano zbog niza faktora (dosadašnja nedovoljno dobra izdanja tekstova, primjena stare kritike teksta, neujednačenost tekstoloških postupaka). Stoga je potrebno vratiti se izvornicima i kritički ih transkripcijom prenijeti u elektronički oblik. Struktura građe omogućuje drugu kvalitetu korpusa (raznovrsnost). KST treba biti računalno čitljiv, označen, jednojezični, nespecijalni statični povijesni korpus pisanoga starohrvatskoga jezika.

¹³ O izvorima i čitanju te pjesme v. Ivšić 1939.

Literatura

- AR = Rječnik hrvatskoga ili srpskoga jezika. I-XXIII. Zagreb: JAZU, 1880-1976.
- Bratulić, Josip. Edicijska praksa hrvatskih istraživača i izdavača srednjovjekovnih tekstova u XIX i XX stoljeću (Historijski prikaz). // *Međunarodni naučni skup Tekstologija srednjovjekovnih južnoslovenskih književnosti*. / Bogdanović, Dimitrije (ur.). Beograd: SANU, 1981, 137-147.
- Carquiglini, Bernard. Textuäre Modernität. // *Texte zur Theorie des Textes*. / Kammer, Stephan, Lüdeke, Roger (priredili i komentirali). Stuttgart: Philipp Reclam jun., 2005, 116-131. [Fragmentarni njemački prijevod francuskoga izvornika *Éloge de la variante*, Paris, 1989.]
- Damjanović, Stjepan. Što (ne) smijemo mijenjati u piščevu tekstu? // *Radovi Zavoda za slavensku filologiju*. 32 (1998); 57-65.
- Fučić, Branko. *Glagoljski natpisi*. // *Djela JAZU*. 57 Zagreb: JAZU, 1982.
- Grabar, Biserka. O nekim problemima kritičkog izdavanja hrvatskoglagoljskih tekstova. // *VIII međunarodni slavistički kongres: Prilozi*. // Kuzmanović, Mladen; Stamać, Ante; Šojat, Antun (ur.). Zagreb: Hrvatsko filološko društvo, 39-44.
- Hofmeister, Andrea. Textkritik als Erkenntnisprozeß: sehen – verstehen – deuten. // *Editio*. 19 (2005); 1-9.
- Ivanić, Dušan. Osnovi tekstologije. Beograd: Narodna knjiga–Alfa, 2001.
- Ivšić, Stjepan. O tobožnoj "najstarijoj sačuvanoj hrvatskoj pjesmi prije god. 1920". // *Grada za povijest književnosti hrvatske*. XIV (1939); 1-10.
- Kapetanić, Davor. Kako pripremati izdanja djela novijih hrvatskih pisaca. // *Croatica*. 1 (1970), 1; 237-259.
- Kapetanović, Amir. Toponimi u hrvatskim "Lucidarima". // *Folia onomastica Croatica*. 14 (2005); 1-22.
- Kapetanović, Amir. Leksik Marulićeve *Suzane*. // *Colloquia Maruliana*. 15 (2006); 15-25.
- Lemnitzer, Lothar; Zinsmeister, Heike. *Korpuslinguistik: Eine Einführung*. Tübingen: Gunter Narr Verlag, 2006.
- Malić, Dragica. Skica za *Hrvatski rječnik do Marulića i njegovih suvremenika*. // *Filologija*. 30-31 (1998); 61-70.
- Malić, Dragica. *Nacrt za Hrvatski rječnik do Marulića i njegovih suvremenika*. Zagreb: Institut za hrvatski jezik i jezikoslovlje, 2002.
- Mažuranić, Vladimir. *Prinosi za hrvatski pravno-povjestni rječnik. I-II*. Zagreb: JAZU, 1902-1922.
- Nazor, Anica. Dragica Malić i glagoljski tekstovi. // *Rasprave Instituta za hrvatski jezik i jezikoslovlje*. 31 (2005); 449-454.
- Rječnik crkvenoslavenskoga jezika hrvatske redakcije. I. svezak. Zagreb: Staroslavenski institut, 2000.
- Scherer, Carmen. *Korpuslinguistik*. Heidelberg: Universitätsverlag Winter, 2006.
- Schnell, Rüdiger. 'Autor' und 'Werk' im deutschen Mittelalter: Forschungskritik und Forschungsperspektiven. // *Neue Wege Mittelalter-Philologie*. / Heinzle, Joachim; Johnson, L. Peter; Vollmann-Profe, Gisela (ur.). Berlin: 1998, 12-73.
- Stackmann, Karl. Neue Philologie? // *Modernes Mittelalter. Neue Bilder einer populären Epoche*. / Heinzle, Joachim (ur.). Frankfurt a. M., Leipzig: Insel-Verl., 1994, 398-427.
- Tadić, Marko. *Jezične tehnologije i hrvatski jezik*. Zagreb: Ex Libris, 2003.
- Vončina, Josip. *Tekstološka načela za pisanu baštinu hrvatskoga jezičnog izraza: Posebni prilog Stoljećima hrvatske književnosti*. Zagreb: Matica hrvatska, 1999. (2006)