

Vida Vukoja  
Staroslavenski institut  
Demetrova 11, HR-10000 Zagreb  
vida.vukoja@stin.hr

## O KORPUSU RJEČNIKA CRKVENOSLAVENSKOGA JEZIKA HRVATSKE REDAKCIJE I NJEGOVU ODNOSU PREMA KORPUSIMA HRVATSKOGA JEZIKA

U prilogu se najprije iznose osnovne značajke korpusa za Rječnik crkvenoslavenskoga jezika hrvatske redakcije, koji se čuva u Staroslavenske institutu u Zagrebu. Potom se one uspoređuju s odgovarajućim značajkama korpusa koji se u Institutu za hrvatski jezik i jezikoslovlje sastavlja za izradu Starohrvatskoga rječnika. Svrha je te usporedbe, kao i upoznavanja stanja procesa digitalizacije hrvatskih crkvenoslavenskih tekstova, da se razmotri hoće li ti različiti korpusi moći doprinijeti stvaranju solidne osnove za različita istraživanja jezične prakse u srednjovjekovnome hrvatskome društvu.

### I. Uvod

Termin »korpus« ima više određenja. Bo Svensén (2009:43) razlikuje starije i suvremeno značenje toga termina. Dok se prije naglašavala iscrpnost na razini cjeline korpusa i na razini tekstova koji tvore korpus te na postojanju neke vrste srodnosti ili sličnosti tih tekstova (čime korpus stječe svojstvo homogenosti),<sup>1</sup> suvremeno razumijevanje toga termina naglašava da je korpus zbirka jezičnoga materijala koji je ishod namjernoga izbora, u skladu s određenim kriterijima, i uporabljen kao osnova za jezikoslovno istraživanje. Drugo je vrijedno određenje termina »korpus« dao John Sinclair. Prema njemu, korpus je skupina prikupljenih tekstova koji

<sup>1</sup> Georges Kleiber uz iscrpnost i homogenost kao kriterij za procjenu kakvoće korpusa navodi reprezentativnost (Kleiber 1978:65–71, usp. Raffaelli 2001:91–95).

svjedoče o izborima i kombinacijama izbora korisnika određenoga jezika. Osim toga, stvaranje je korpusa jedna od središnjih radnji vezanih za izradu rječnika (Sinclair 2003:167).

Do vremena široke primjene računalne tehnologije veliki su se ugledni rječnici stvarali na osnovi zbirke navoda<sup>2</sup> ekscerpiranih iz odabranih tekstova, i to tako da se nije ekscerpivala svaka pojavnica određenoga lematiziranoga leksema, nego ih se ekscerpivalo onoliko koliko je bilo određeno za vrijeme planiranja izrade korpusa. U vrijeme kada su se korpusi izrađivali ručno i na papiru takvi su korpusi bili prihvatljivije rješenje od obrade cjelovitih tekstova. Kada je računalna tehnologija omogućila da se relativno lako i brzo u računalo mogu unijeti brojni i cjeloviti tekstovi, pa i kada su obimni, rječnike se uobičajilo pisati na osnovi korpusa. Prvi je veći digitalni korpus, od oko milijun riječi, bio tzv. Brown Corpus sastavljen 1963–1964. na američkome sveučilištu Brown.

U Staroslavenskome se institutu u Zagrebu od kraja 1950-ih do početka 1990-tih sastavljao korpus hrvatskoga crkvenoslavenskoga jezika na osnovi kojega se od 1991. objavljuje Rječnik crkvenoslavenskoga jezika hrvatske redakcije (u nastavku teksta: RCJHR). Korpus sadrži tekstove napisane na hrvatskome crkvenoslavenskome idiomu nastale u razdoblju od XI/XII. do sredine XVI. stoljeća. Osim za izradu RCJHR<sup>3</sup> korpus se rabi i u jezikoslovnim istraživanjima hrvatskoga crkvenoslavenskoga te u istraživanjima glagoljske književnosti. Građa je toga korpusa vrlo važan izvor podataka za istraživanja crkvenoslavenskih idioma, a potencijalno je vrlo vrijedan i dosada nedovoljno korišten izvor usporednih podataka za dijakronijska i povijesna istraživanja hrvatskoga jezika.

Među istraživačima hrvatskoga jezika česta su dva nevjerodostojna stava o odnosu hrvatskoga crkvenoslavenskoga prema srednjovjekovnome hrvatskome. Prvi stav previše naglašava sastavnicu »hrvatski«, pa se pretpostavlja da je hrvatski crkvenoslavenski gotovo nešto kao »rana faza u razvoju hrvatskoga«. Drugomu je stavu prevelik naglasak na crkvenoslavenskoj sastavnici čime se propušta uvidjeti da nije riječ o bilo kakvom

<sup>2</sup> O zbirkama navoda v. npr. Svensén 2009:40–42.

<sup>3</sup> Od 1991. do 2012. u papirnome je obliku objavljeno 19 sveščića toga rječnika (svaki sadrži 64 stranice). Planom je projekta izrade RCJHR predviđeno da se svake godine pripremi po jedan sveščić i po mogućnosti i objavi. Prvih je deset sveščića 2000. ukoričeno u prvi svezak (leme: *a – vrêdb*). Više je puta taj prvi svezak RCJHR-e ocijenjen kao izvrsno leksikografsko djelo, npr. na predstavljanju prvoga sveska RCJHR tijekom XIII. međunarodnog slavističkog kongresa u Ljubljani 2003, također Грковић-Мејдор 2007:187. Time se neizravno pohvaljuje i korpus za izradu toga rječnika. Dosada objavljeni sveščići zasada su elektronski dostupni kao slike u PDF formatu. Od 14. se sveščića tekst priprema tako da se može digitalno pretraživati.

crkvenoslavenskom idiomu, nego onomu na koji je utjecao hrvatski.

Pouzdana istraživanja srednjovjekovne jezične prakse u hrvatskim zemljama, čak niti kada se ona žele baviti samo hrvatskim jezikom, ne mogu biti provođena bez uzimanja u obzir hrvatske crkvenoslavenske jezične baštine.<sup>4</sup> Mjesto bi hrvatskoga crkvenoslavenskoga idioma i korpusa toga idioma u okviru istraživanja hrvatske pismenosti i jezika trebalo postajati sve jasnije kako napreduje izrada korpusa za Starohrvatski rječnik (u nastavku teksta: SHR) u okviru istoimenoga projekta Instituta za hrvatski jezik i jezikoslovlje jer bi korpus za SHR trebao pružiti do sada naj sustavnije obrađenu građu iz srednjovjekovnih pisanih spomenika na hrvatskome jeziku.

Među postojećim korpusima hrvatskoga jezika (Hrvatska jezična mrežna riznica, Hrvatski nacionalni korpus, korpus koji je pripreman za dopune Akademijina rječnika, korpus za izradu Rječnika hrvatskoga kajkavskog književnog jezika<sup>5</sup> i drugi)<sup>6</sup> korpus za RCJHR najviše dodirnih točaka ima s korpusom za izradu SHR-a. Riječ je o korpusu hrvatskih tekstova kojim su obuhvaćeni zapisi od najstarijih zabilježenih stadija hrvatskoga jezika do tekstova s kraja XV. i početka XVI. stoljeća.<sup>7</sup> Taj se vremenski raspon gotovo posve podudara s onim obuhvaćenim korpusom za RCJHR. Prema tomu, korpus bi starohrvatskih tekstova mogao služiti i kao

<sup>4</sup> Ponekad se u istraživanju hrvatskoga jezika naiđe na oblike koji se teško mogu primjereno opisati bez uvida u hrvatski crkvenoslavenski jezični materijal. Npr. s poteškoćama toga tipa se u izradi svoje doktorske radnje susrela Ivana Vrtič. Opisujući sintaksu Kašićeva jezika naišla je na oblike koji su neobični za hrvatski jezik, ali su njihovi pojavi razumljiviji kada se uzme u obzir hrvatski crkvenoslavenski (usp. Vrtič 2009:117–118, 218–219, 291–293).

<sup>5</sup> Kao primjer uzajamne koristi koje bi takvo supostojanje moglo donijeti mogu poslužiti leksemi hrv. osnove *privrg-* u smislu političke ili uopće društvene naklonjenosti nekome ili nečemu. Posredstvom istraživanja hrvatske crkvenoslavenske osnove *privrg-* otkriveno je da je takvo značenje potvrđeno u hrvatskome kajkavskome idiomu barem od sredine XVII. stoljeća i barem do sredine XVIII. stoljeća. S obzirom da je korpus za izradu Rječnika hrvatskoga kajkavskog književnog jezika statičan (zatvoren), a u njemu nisu postojale potvrde o takvome značenju, ono neće biti uvršteno iako su autori rječnika upozoreni na pismene potvrde toga značenja. Da je slično otkriveno u hrvatskome crkvenoslavenskome novootkriveno bi značenje bilo uvršteno, jer korpus za RCJHR ima odlike dinamičnoga korpusa – stalno je otvoren za primanje novih zanimljivih potvrda. V. Lučić 2004.

<sup>6</sup> Popis se većine važnijih korpusa hrvatskoga jezika može naći na mrežnoj stranici: <http://jthj.ffzg.hr/korpusi.htm>.

<sup>7</sup> Donja je vremenska granica određena vremenom nastanka najstarijih starohrvatskih tekstova. Gornja je vremenska granica građe pomaknuta do u XVI. stoljeće, jer su neki raniji tekstovi sačuvani samo u tako kasnim prijepisima. Takav je slučaj s Vinodolskim zakonikom čiji je tekst izvorno iz 1288, ali najstariji njegov prijepis potječe iz XVI. st.

neprocjenjiva pomoć u što jasnijem razgraničavanju hrvatskoga i crkvenoslavenskoga u hrvatskome crkvenoslavenskome idiomu, iako se takvo razgraničenje teško može posve jednoznačno odrediti, jer srednjovjekovni govorni hrvatski idiom s hrvatskim crkvenoslavenskim čini jedan dijasistem (Katičić 1992).<sup>8</sup> S druge strane, istraživačima bi hrvatskoga jezika korpus za RCJHR trebao biti pomoćan izvor podataka u dijakronijskim i uopće povijesnim istraživanjima hrvatskoga jezika.

## II. Opis korpusa za RCJHR<sup>9</sup>

Tijekom Četvrtoga međunarodnoga slavističkoga kongresa održanoga u Moskvi 1958. iznesen je i usvojen prijedlog da se izradi tezaurus crkvenoslavenskoga jezika koji bi objedinjavao sve redakcije toga jezika.<sup>10</sup> Upravo je u vrijeme održavanja toga kongresa u Pragu počeo izlaziti *Slovník jazyka staroslověnského* (u nastavku teksta: *Slovník*; v. Štefanić 1962, Nazor 1991:I) izrađen na osnovi korpusa kanonskih starocrkvenoslavenskih tekstova. Na izradi je toga rječnika dugo radio i František Václav Mareš, koji je na osnovi iskustva leksikografa *Slovníka* napisao podroban nacrt (Mareš 2007 (1962)) za izradu RCJHR te korpusa na osnovi kojih bi se taj rječnik stvarao. Između ostaloga, određeni su kriteriji za izbor izvora (prednost je data potpunijim i starijim cjelovitim spomenicima), posebno je određen način ekscerpcije, struktura kartica, način ustrojavanja kartoteka. Kao u korpusu za izradu *Slovníka*, i u korpusu za RCJHR odlučeno je da se ustroje kartoteke ekscerpata po izvorima građe (u žargonu: mjesni katalog) i po azbučnome redoslijedu hrvatskih crkvenoslavenskih lema (u žargonu: azbučni katalog) te kartoteka

<sup>8</sup> Ovdje se neće ulaziti u raspravu povezanu s opravdanosti i značenjem nazivka »dijasistem«, nego će se on rabiti onako kako je određen u Katičićevu radu (1992), gdje se kao i u istupima Dalibora Brozovića slijedi poimanje Uriela Weinreicha, koji je taj nazivak uveo: »*Diasystem is the placement of discrete varieties in a kind of continuum determined by their partial similarities.*« (Weinreich 1954:393).

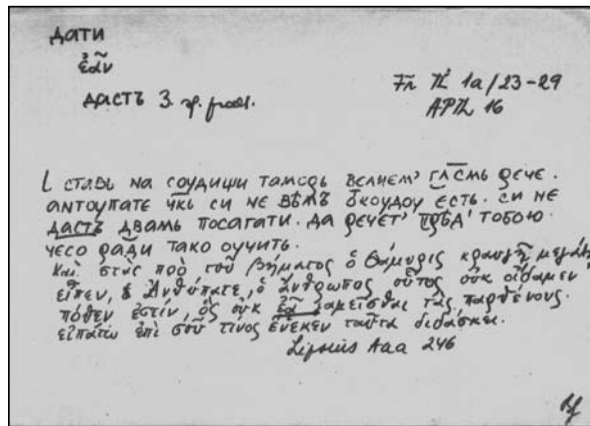
U posljednje se vrijeme Milan Mihaljević (2011) bavio dvojnom jesu li u hrvatskoj srednjovjekovnoj jezičnoj praksi hrvatski i hrvatski crkvenoslavenski dva dosljedno razdvojena jezika ili je riječ o dva povezana idioma. Drugim riječima, je li riječ o pravom dvojezičnjaštvu (kao u slučaju hrvatskoga i latinskoga jezika) ili dvojezičnjaštvu tipa diglosije (kada u društvu supostojе dva jezika, od kojih jedan uživa veći ugled od drugoga). Ustanovio je kako je riječ o diglosiji prema prihvaćenim kriterijima Charlesa Fergusona (1959).

<sup>9</sup> Osnovni se podaci o RCJHR uključeni u ovo poglavlje nalaze i u opširnome uvodu RCJHR, objavljenome najprije kao prvi sveščić RCJHR 1991., ponajprije u tekstovima Nazor 1991. i Grabar, Mareš, Mulc 1991.

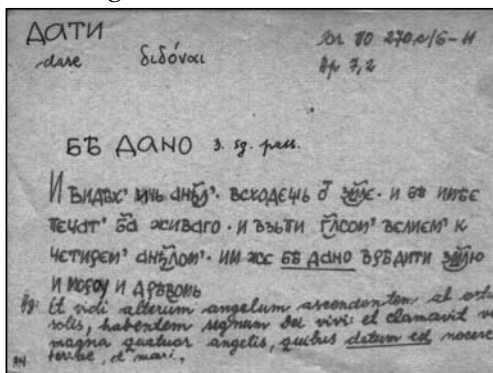
<sup>10</sup> O okolnostima te tzv. moskovske inicijative te o njezinu razvoju u različitim zemljama, uključujući i Hrvatsku, v. Nazor 2008.

hrvatskih crkvenoslavenskih lema (u žargonu: geslar), kartoteka grčkih usporednica hrvatskih crkvenoslavenskih lema (u žargonu: grčki geslar) i kartoteka latinskih usporednica hrvatskih crkvenoslavenskih lema (u žargonu: latinski geslar). Korpus za izradu RCJHR počeo se sastavljati 1959, a prvu su ekipu ekscerptora predvodili Josip Hamm i Vjekoslav Štefanić.<sup>11</sup> Prilozi 1. i 2. prikazuju osnovne kartice mjesnoga i azbučnoga kataloga, a prilog 3. prikazuje usporednu karticu.

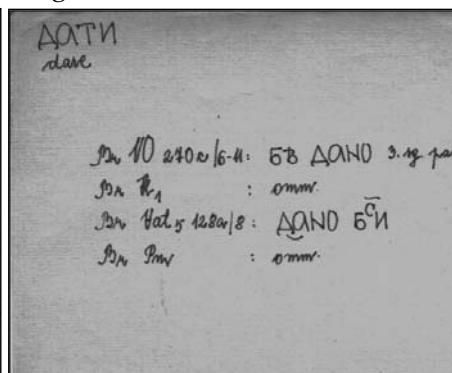
Prilog 1.



Prilog 2.



Prilog 3.



Korpus je sastavljan s težnjom da na osnovi njega bude moguće sastaviti tezaurus — rječnik koji bi sadržavao ukupno jezično blago sačuvanih hrvatskih crkvenoslavenskih tekstova. Svakako, riječ je o referentnome, općemu korpusu jer ispunjava uvjete za različita istraživanja: sa-

<sup>11</sup> Prethodno je 1950. godine započeto ekscerpiranje za rječnik hrvatskoglagoljskih starina. No nakon nastanka tzv. moskovske inicijative ono je zaustavljeno kako bi počelo ekscerpiranje za novoplanirani veliki rječnik crkvenoslavenskoga jezika. O tome v. Nasop 1966.

drži normiran rječnik hrvatskih crkvenoslavenskih jezika i dovoljno je velik da obuhvaća dostatan broj različitih okolina u kojima se određeni leksem rabi. Korpus sadrži tekstove različitih žanrova, a posebna je pozornost usmjerena tomu da bude i reprezentativan — da omjerom različitih skupina tekstova koje sadrži zrcali omjer različitih skupina tekstova u ukupnome zbiru poznatih hrvatskih crkvenoslavenskih tekstova. Strukturiran je ponajprije na osnovi sadržaja kodeksa (misali, časoslovi, psaltiri, obrednici, odlomci, zbornički tekstovi), ali nalaze li se inačice jednoga teksta u različitim skupinama izvora, one su uspoređene, a razlike zabilježene na usporednim karticama. Zbog toga se ipak može reći da korpus nema strogo odvojene sastavnice.

Hrvatski crkvenoslavenski nije organski jezik, nikada nije bio govorni idiom i nikomu nije bio materinski jezik, već je riječ o knjiškome jeziku, živu samo u rukopisno tradiranim tekstovima i tiskanim izdanjima. Zbog toga činjenica da se neka jezična činjenica ne nalazi u ekscerpiranom tekstovima ima veću vrijednost nego što bi bio slučaj da je riječ o nekom organskom idiomu čiji je rječnik sastavljen na zapisima toga idioma.<sup>12</sup> U slučaju valjano ekscerpiranoga korpusa knjiškoga jezika činjenica da neke jezične značajke nema u korpusu vrlo vjerojatno znači da ju određeni jezik niti nema, a u slučaju organskoga idioma činjenica da određenoj značajci nema traga u korpusu uvijek može biti plod puke slučajnosti (usp. Zgusta 1991:224). U korpusu nisu sve pojavnice gramatički obilježene. Onim lemmama koje su potvrđene vrlo brojnim primjerima, desetinama ili čak stotinama, gramatički je obilježen velik broj pojava, ali ipak ne svaka. Time se izbjegava izlišna zalihost u korpusu (npr. ne navode se sve potvrde oblika *estb*, trećega lica jednine prezenta glagola *biti*, ili *bože*, vokativa imenice *bogb*).

Izvori su za korpus RCJHR rukopisi i tiskana izdanja među kojima samo jedno izdanje nije inkunabula (Senjski obrednik). Za korpus RCJHR građa je ekscerpirana izravno s izvornika ili s preslika izvornika. Popis 62 izvora nastalih u razdoblju od XI/XII. do sredine XVI. stoljeća i uključenih u korpus sadrži (potpun se popis izvora nalazi u Nazor 1991:XXXI–XXXVI): 26 odlomka (od toga četiri odlomka Apostola<sup>13</sup>, pet odlomaka misala, sedam odlomaka časoslova, deset odlomaka legendarnih i apokrifnih tekstova te sermona među kojima neki sadrže odlomke iz nekoliko različitih tekstova), četiri misala, jedanaest časoslova, tri psaltira, tri obrednika te tekstove pisane hrvatskim crkvenoslavenskim iz petnaest zbornika. Uz njih se kao pomoćni izvori rabe još dva misala i dva časoslova, ali i

<sup>12</sup> O odnosu govorenoga i pisanoga jezika v. Žagar 2005:224–237.

<sup>13</sup> »Apostol« je bogoslužna knjiga koja sadrži Djela apostolska i poslanice.

drugi izvori kada se u njima nađe potvrda nekoga leksikografski zanimljivoga podatka. Izdana su i kritička izdanja ili opisi gotovo svih hrvatskih crkvenoslavenskih tekstova koji služe kao korpusna građa za RCJHR (podatke o izdanjima i opisima v. Nazor 1991:XXXI–XXXIX). Najveći su pothvati u području objavljivanja izvora vjerojatno Hrvojev misal u faksimilu i s kritičkim izdanjem teksta u latiničnoj transliteraciji te Drugi novljanski brevijar u fototipskome izdanju s uvodom, iscrpnom bibliografijom teksta i popisom biblijskih tekstova te tekstova homilija i sermona.<sup>14</sup> Navedeni izvori sadrže velik broj različitih tekstova od kojih se mnogi nalaze u više inačica. U takvim je slučajevima jedna inačica, obično najduža i s najizrazitijim svojstvima hrvatskoga crkvenoslavenskoga jezika, izabrana da bude osnovni tekst i bilježi se na osnovnim karticama, a razlike koje se uoče u drugim inačicama upisuju se na usporedne kartice.

Za brojne se hrvatske crkvenoslavenske tekstove prevedene s grčkoga ili latinskoga, a takvi su u većini, uspio utvrditi prijevodni predložak ili usporedni grčki ili latinski tekst i on je tada pribilježen kao što je vidljivo na prilogima 1–2. Iako se više sustavno ne traga za grčkim i latinskim prijevodnim predlošcima tekstova koji su uključeni u korpus, stalno je otvorena mogućnost da se novopronađeni takvi prijevodni predlošci upišu u postojeće kartice. Time je korpus za RCJHR zapravo postao usporedan, a ne jednojezičan korpus.<sup>15</sup> Važno je znati da hrvatski crkvenoslavenski tekst nije u svakoj pojedinosti uskladi<sup>16</sup> s grčkim ili latinskim usporednim tekstom koji je upisan kao prijevodni predložak, što je vezano za uglavnom rukopisnu tradiciju hrvatskoglagoljskih tekstova — prepisivač je imao slobodu

<sup>14</sup> U izradi RCJHR Hrvojev misal se rabi kao pomoćni izvor, a Drugi novljanski brevijar je u popisu osnovnih izvora za RCJHR.

<sup>15</sup> O usporednim korpusima v. npr. Svensén 2009:55–57, također Borin 2002.

<sup>16</sup> Usklađivanje (eng. *alignment*) proces je svojstven radu s usporednim korpusima kojim se određuje koja rečenica, sintagma ili leksem u tekstu pisanome jednim jezikom odgovara kojoj rečenici, sintagmi ili leksemu u tekstu pisanome drugim jezikom. Teubert i Čermáková (2004:124) navode da zbog poteškoća u usklađivanju na razini rečenica i leksema postoji svega nekoliko većih takvih korpusa, većih od otprilike 5 000 000 pojavnica po jeziku. S obzirom da korpus za RCJHR ima oko 420 000 kartica mjesnoga kataloga, a u to su ubrojane i usporedne kartice, ni on se ne bi mogao ubrojiti u takve veće usporedne korpusne. Grčkih i latinskih pojavnica svakako ima daleko manje od 5 000 000, jer nemaju svi hrvatski crkvenoslavenski odsječci teksta na kartici prijevod, a kad ga i imaju on je ili na latinskome ili na grčkome. Svakako, latinskih pojavnica ima daleko više nego grčkih što je očito i iz procjene broja kartica u kartoteci grčkih usporednica crkvenoslavenskih lema (60 000) i broja kartica u kartoteci latinskih usporednica hrvatskih crkvenoslavenskih lema (200 000).

I dva su usporedna korpusa suvremenoga hrvatskoga jezika, hrvatsko-slovenski i hrvatsko-engleski, daleko od velikih. Hrvatsko-slovenski npr. ima oko 500 000 pojavnica po jeziku.

unijeti određene novine u svoj tekst, npr. skraćivanjem teksta, unošenjem novih jezičnih elemenata, mijenjanjem jednih leksema drugima. S obzirom na to da su srednjovjekovni latinski i grčki predlošci također nastajali prepisivanjem, katkada upisani grčki ili latinski tekst i nije stvarni prijevodni predložak hrvatskoga crkvenoslavenskoga teksta nego inačica toga teksta — najbliža koju su sastavljači korpusa pronašli. Ipak, u većini primjera hrvatski crkvenoslavenski tekst uglavnom vjerno prati tekst upisanoga grčkoga ili latinskoga teksta. Štoviše, u mnogim primjerima upravo on olakšava razumijevanje hrvatskoga crkvenoslavenskoga teksta.

S obzirom na vrijeme početka izrade korpusa za RCJHR, on je očekivano u obliku papirne kartoteke ekscerpata, kao što je inače slučaj sa zbirkama navoda. No, ovdje je riječ o korpusu, jer se u njemu ne nalaze samo odabrani navodi lematiziranih leksema iz različitih tekstova (kao u zbirkama navoda), nego se na karticama nalaze ekscerpirani čitavi odabrani tekstovi pisani hrvatskim crkvenoslavenskim jezikom, pa i čitavi rukopisni kodeksi. Svi su leksemi zabilježeni u korpusu lematizirani. Jedna je kartica ciklostilom preslikana u dvostruko više primjeraka od broja pojava. Jedan niz ispisanih kartica poredanih redosljedom određenim lematizacijom umetan je u kutije kartoteke ustrojene po izvorima građe (tvori ga oko 420 000 kartica<sup>17</sup>), a kartice su drugoga niza po azbučnome redosljedu lema umetane na svoje mjesto tvoreći kartoteku hrvatskih crkvenoslavenskih lema (tvori ga više od 400 000 kartica).<sup>18</sup> Azbučni katalog ima manje kartica nego mjesni jer nije svaka pojava određene leme odabrana da bude i nositeljicom primjera uporabe te leme, a kartice s takvim pojavnicama se onda niti ne pojavljuju u azbučnome katalogu. Ustrojena je kartoteka hrvatskih crkvenoslavenskih lema (po azbučnome redosljedu). Iako tu kartoteku tvori nešto više od 18 100 kartica<sup>19</sup> (od kojih je oko 8 500 imenica, 5 400 glagola, 2 500 pridjeva; v. primjer kartice u prilogu), rječničkih bi članaka u RCJHR trebalo biti manje (oko 17 100), jer će neki leksemi koji se u toj kartoteci pojavljuju kao leme u RCJHR biti uputom spojeni s nekim drugim leksemom koji se pojavljuje kao lema

<sup>17</sup> Procjene broja kartica u kartotekama ekscerpata, te kartotekama grčkih i latinskih usporednica hrvatskih crkvenoslavenskih lema navedene su prema Čunčić i Magdić 2008. Procjene broja kartica po vrstama riječi te kartoteke hrvatskih crkvenoslavenskih lema vlastite su.

<sup>18</sup> Nevenka Linarić taj je zahtjevan posao, kao i usporedbu svih ekscerpata s izvornim tekstovima, predano obavljala od uspostave kartoteka do kraja svoga života 1976. godine. Poslije je brigu o kartotekama korpusa marno vodila Ivana Mulc, a danas to jednako tako čini Marija Klenovar.

<sup>19</sup> Prebrojila sam 18 144 kartica. Možda će taj broj po završetku izrade RCJHR biti nešto drugačiji zbog novopronađenih riječi i odluka vezanih za normalizaciju.



rječničkoga članka.<sup>20</sup> Osim toga, za nekolicinu leksema u toj kartoteci već sada stoji da su lažni *hapax legomena* (riječi koje se pojavljuju samo jednom) ili da je za određenu skripciju iz izvora određena drugačija normalizacija<sup>21</sup>, pa onda takva skripcija neće biti lemom u RCJHR. Kao što je već navedeno, pored kartoteke hrvatskih crkvenoslavenskih lema tu su i kartoteke grčkih usporednica hrvatskih crkvenoslavenskih lema (poredanih po grčkome alfabetu; tvori ga oko 60 000 kartica; v. kartice toga geslara u prilogu 5.) i latinskih usporednica hrvatskih crkvenoslavenskih lema (poredanih po latinskoj abecedi; tvori ga oko 200 000 kartica; v. kartice toga geslara u prilogu 6.).

Prilog 4.

ΔΑΤΙ, ΔΑΤΙ CE			
date	δίδομαι	conferre	συνομιλεῖν
donare	χαρίζεσθαι	mittere	ἀφίεμαι
largiri	παραδίδωμι	offerre	τολμαῖν
tribuere	ἀποδίδωμι	deponere	ἐκφέρειν
praestare	παραδίδωμι	procurare	μαρτυρεῖν
reddere	εἶναι	efficere	προσφέρειν
concedere	συγγιγρῆναι	restituere	ἔχειν
tradere	ἐπιδίδωμι	commutare	ἀφαιρῆσθαι
praebere	μεταδίδωμι	Sl. = ΔΑΤΙ, ΔΑΤΙ CA	

locare	μιλεῖν	potius dare	τοῖς εἶναι
reparare	restituere	donum dare	
promittere	committere	incensum autem	ἀφίεμαι
distribuire	mittere	dare	προσδίδωμι
restituere	donationum	meten	συγγιγρῆσθαι
indulgere	damnae	dari	ἔχειν
ferre	mittere	accipere	
condonare	pollari	subdari	
trahere	dedere	cedere	
praeparare	preparari		

Prilog 5.

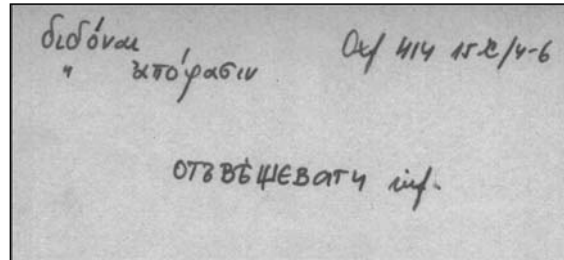
δίδομαι	Въ во 166 г. / 24-2
	Гр 28, 4
ПОБЛАЖИТИ	
ПОБЛАЖИЛ' ЕСТ' 3. сг. пф.	

δίδομαι	Въ во 3. 132 / 2-0
	129 2. 16
ΔΑΤΙ	
ΔΑΤΙ 2. γ. гр.	

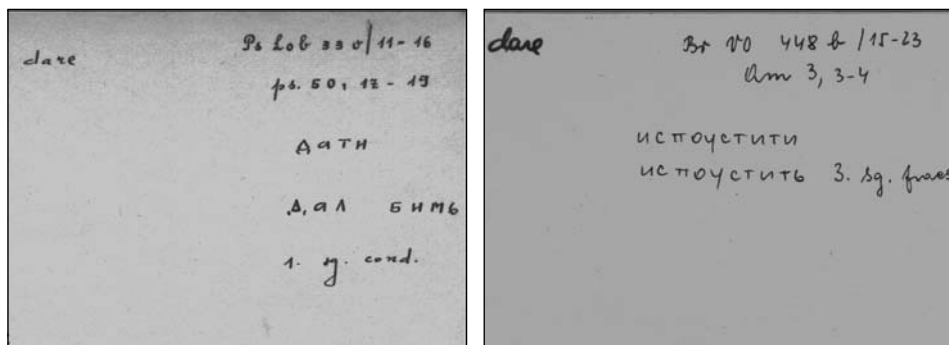
<sup>20</sup> Leme je računalno unijela Ana Kovačević. Glagolske je leme računalno unijela Vida Vukoja, a pridjevske Sandra Sudec.

<sup>21</sup> O normalizaciji koja je primijenjena u glavi rječničkih članaka v. Grabar, Mareš, Mulc 1991:VIII–XIII.

Prilog 5.



Prilog 6.



Ovdje se neće niti pokušati točnije procijeniti broj osnovnih i usporednih kartica, jer je to zapravo nemoguće. Tek se za orijentaciju može reći da je osnovnih kartica vjerojatno oko jedne trećine, a usporednih oko dvije trećine od ukupnoga broja kartica mjesnoga kataloga. Isto je tako trenutno nemoguće procijeniti koliko hrvatskih crkvenoslavenskih pojava ima na jednoj osnovnoj kartici, ali za orijentaciju se može reći da ih je najčešće između deset i petnaest.<sup>22</sup> Izračunom na osnovi takvih *ad hoc* procjena dolazi se do brojke od otprilike između 1 400 000 i 2 100 000 pojava.<sup>23</sup>

Korpus sadrži podjednako svojstva statičnih i dinamičnih korpusa. Poput statičnih korpusa (još zvanih i korpusom primjera, eng. *sample corpus*) on je u svom azbučnom katalogu strukturirana zbirka ograničenoga

<sup>22</sup> Neke kartice imaju svega dva ili tri leksema, a neke ih imaju i po nekoliko desetina. Kratkoća je ekscerpta obično povezana sa strukturom ekscerpirana teksta, npr. kada je riječ o kratkoj rubrici, naslovu, početnim riječima psalma koji u časoslovu glagoljašu naznačuje koji će psalam moliti, dok tekst psalma nije zapisan jer ga glagoljaš zna napamet.

<sup>23</sup> Naravno, u slučajevima starih (stadija) jezika, referentni korpus ne može brojiti milijune pojava kao što je slučaj s korpusima suvremenih (stadija) jezika. Npr. British National Corpus sadrži 100 000 000 riječi, a Hrvatski nacionalni korpus trenutno sadrži 101 300 000 pojava (<http://www.hnk.ffzg.hr/>). Nacionalni jezični referentni korpusi obično sadrže između 50 i 500 000 000 pojava.

broja primjera lematiziranih leksema koji se jednom uvršteni više ne izuzimaju iz korpusa. Poput dinamičnih korpusa (još zvanih i nadgledani korpus, eng. *monitor corpus*), njegov se mjesni katalog sastoji od čitavih tekstova i otvoren je za uvrštavanje novih tekstova ili potvrda ako se to ocijeni poželjnim. Korpus za RCJHR posebno je vrijedan za paleoslaviste, jer sadrži tekstove koji su u izvornome obliku teško dostupni, a nerijetko i teško čitljivi. Najveći se broj hrvatskih crkvenoslavenskih tekstova nalazi u zbirci foto-preslika knjižnice Staroslavenskoga instituta u Zagrebu, tek ih se manji broj čuva na mikrofilmu u istoj knjižnici. Tekstovi hrvatskoga crkvenoslavenskoga jezika pisani su različitim tipovima glagoljice, a ponegdje se među inicijalima nađe latinice i ćirilice.<sup>24</sup> Na marginama ili na drugim mjestima pergamena odnosno listova ispisanih glagoljičnim tekstom nađu se i zapisi pisani latinicom ili ćirilicom, ali tu je redovito riječ o hrvatskome jeziku pa takvi zapisi nisu uključeni u korpus za RCJHR.<sup>25</sup> Kako je moskovskom inicijativom odlučeno da bi veliki planirani rječnik crkvenoslavenskoga jezika bio pisan starom ćirilicom, odlučeno je da će i hrvatski crkvenoslavenski ekscerpti u korpusu, kao i leme rječničkih članaka u RCJHR, biti iz glagoljice transliterirani<sup>26</sup> u staru ćirilicu<sup>27</sup> i to prema Jagićevu transliteracijskome sustavu.<sup>28</sup> Za transliteraciju se moglo odlučiti jer je prva svrha korpusa upravo izrada rječnika, a ne npr. paleografska, filološka i druga ne-leksikografska istraživanja u kojima korpus također nalazi svoje mjesto.<sup>29</sup> Valja napomenuti da su hrvatski

<sup>24</sup> Latinični se inicijali pojavljuju u različitim hrvatskim crkvenoslavenskim tekstovima i pri ekscerpiranju su na kartice upisani latinicom (v. sliku 1. na kojoj je vidljivo da je latinični inicijal upisan latinicom, za razliku od ostaloga teksta ekscerpta koji je transliteriran s glagoljice i upisan starom ćirilicom). Neki hrvatskoglagoljski spomenici imaju mnogo latiničnih inicijala, npr. u Drugom vrbničkom misalu većina je inicijala latiničnih (Štefanić 1960:350).

<sup>25</sup> U glagoljičnim se tekstovima najčešće nalazi tzv. hrvatska ćirilica, no nađu se i zapisi ustavnom ćirilicom. O ćirilicnim zapisima u Fatevićevu zborniku i Brevijaru Vida Omišljanina v. Nazor 2000:295. Tzv. hrvatska ćirilica, pismo koje se rabilo u Hrvatskoj i Bosni, naziva se i *bosančica*. Za te i druge nazive, posebno u starijoj paleografskoj literaturi, v. Nazor 2000:291, Šimić 2010:238–239.

<sup>26</sup> Transliteracijom se naziva postupak prenošenja grafema jednoga pisma u grafeme drugoga pisma (»znak za znak«), pri čemu se ne vodi računa o izgovoru.

<sup>27</sup> Staro ćirilsko pismo usustavljeno je prema naputcima moskovske inicijative na osnovi kojih su trebali biti usklađeni planirani rječnici različitih crkvenoslavenskih idioma.

<sup>28</sup> V. Jagić 1879:XXXVII, gdje se nalazi tablica s glagoljskim i odgovarajućim staroćirilskim grafemima; o skripciji Jagićevih izdanja v. Bratulić 1981:145–146.

<sup>29</sup> »[N]aš je cilj leksikografska obrada, a ne mehanička reprodukcija teksta.« (RCJ-

crkvenoslavenski primjeri u rječničkim člancima, a oni čine većinu teksta u RCJHR, transliterirani u latinicu.

## II.1. Trenutno stanje procesa digitalizacije korpusa za RCJHR i poteškoće na koje se u tom procesu nailazi<sup>30</sup>

Korpus za RCJHR obiman je kartičan korpus u kojemu su bilježene grafijske, fonološke, morfološke, leksičke i tekstološke razlike inačica istoga teksta, a uz to je i riječ o usporednome korpusu. Zato bi njegova pretvorba u računalno pretraživ korpus bila složena i dugotrajna operacija koja bi zahtijevala veliku novčanu potporu i više osoba osposobljenih za taj posao, a tomu se Staroslavenski institut u doglednoj budućnosti nažalost ne može nadati. Zbog toga se pristupilo skromnijem, ali dohvatljivijem cilju – računalno u obliku JPEG slika snimiti osnovne i usporedne kartice mjesnoga kataloga. Do konca rujna 2009. to je i učinjeno, a tako snimljena građa računalno se pretražuje slično ručnome pretraživanju – klikom računalnoga miša lista se kartica za karticom. Iako sa stajališta korpusnoga jezikoslovlja time nije učinjeno mnogo, omogućena je sigurnija pohrana izvornoga kartotečnoga mjesnoga kataloga koji više nije nužno i radni materijal, jer se sadržaj kartica može pregledavati na računalu (trenutno samo s računala koja se nalaze u Institutu). S obzirom na to da je korpus u takvu obliku nedostatno prilagođen složenome poslu izrade rječničkih članaka RCJHR-e, Lucija Turkalj učinila je posebnu inačicu kako bi sastavljačima rječničkih članaka olakšala pretraživanje elektronične inačice mjesnoga kataloga.

Poteškoće se na putu potpune digitalizacije koji bi svojstvima odgovarala suvremenim računalnim korpusima mogu grupirati s obzirom na tekstološko-filološke kriterije, na one koje proistječu iz pretežno rukopisne tradicije hrvatskih crkvenoslavenskih tekstova i one koje proistječu iz pretežno prijevodne naravi tih tekstova. Vezano za rukopisnu tradiciju hrvatskih crkvenoslavenskih tekstova u prvome redu treba istaknuti da se mnogi tekstovi nalaze razasuti u ekscerpiranim izvorima u više inačica, a sadašnji kartični korpus u takvim slučajevima bilježi ne samo tekstovne ili leksičke razlike nego i morfofonološke, pa i grafijske. Možda je najveći uzrok poteškoća vezanih za digitalizaciju bilježenje grafijskih razlika zbog čega se u mnogome otežava ostvarenje vrlo važne i korisne mogućnosti računalnoga pretraživanja korpusa. Ipak, trenutno je praktički nemoguće

HR 2000:XVI–XVII).

<sup>30</sup> Zahvaljujem Marici Čunčić koja mi je omogućila saznati pojedinosti o provedbi procesa digitalizacije hrvatskih crkvenoslavenskih tekstova i mjesnoga kataloga korpusa za RCJHR.

zanemariti bilježenje grafijskih razlika, jer se one bilježe i u izradi rječničkih članaka za RCJHR. Osim toga, grafijske su razlike predmet određenih paleografskih, književno-filološko-tekstoloških, gramatičkih, leksikoloških istraživanja (a opravdano zanemarena npr. u istraživanjima značenja). Vezano za prijevodnu narav hrvatskih crkvenoslavenskih tekstova, vjerojatno je najveća poteškoća što on ne prati uvijek vjerno upisani grčki ili latinski tekst usporedni tekst. Usklađivanje na razini rečenice ili nekoliko sadržajno povezanih rečenica uglavnom se relativno lako može izvršiti, ali nije uvijek tako i na razini sintagme ili pojedinih leksema.

S obzirom da korpus trenutno nije moguće digitalizirati (a to vjerojatno nikada niti neće biti učinjeno), u Staroslavenskome se institutu odvojeno od projekta izrade RCJHR prionulo digitalizaciji hrvatskih crkvenoslavenskih tekstova pomoću programa koji hrvatske crkvenoslavenske tekstove čini dostupnima u latiničnoj transliteraciji te omogućuje usklađivanje hrvatskoga crkvenoslavenskoga teksta s utvrđenim usporednim latinskim ili grčkim tekstom (ako takvoga ima).<sup>31</sup> U tako se uređen program upisuju različiti tekstovi, a zbog praktičnih razloga prednost za sada imaju stariji i kraći tekstovi. Nastojat će se upisati svi poznati hrvatski crkvenoslavenski tekstovi, što znači da bi bili upisani i svi tekstovi koji trenutno čine mjesni katalog korpusa za RCJHR. Kako će taj posao napredovati uvelike ovisi o količini raspoloživih novčanih sredstava i o tomu koliko se osoba sposobnih za zahtjevan posao upisivanja i gramatičkoga obilježavanja hrvatskih crkvenoslavenskih, grčkih i latinskih tekstova te usklađivanja hrvatskih crkvenoslavenskih tekstova s odgovarajućim usporednim grčkim i latinskim tekstovima tome poslu može posvetiti. Trenutno se digitalno snima azbučni katalog, dok se hrvatski crkvenoslavenski, grčki i latinski geslari ne planiraju digitalno snimiti, barem ne zasada.

Kako bi korpus koji nastaje prepisivanjem tekstova bio što pouzdaniji, metodologija izrade takvih korpusa nalaže da dvije osobe neovisno jedna o drugoj unose isti tekst (usp. Svensén 2009: 42). Nakon toga u izvorniku bi trebalo provjeriti mjesta u kojima se unosi razlikuju,<sup>32</sup> zbog čega se osnovano može posumnjati da je u jednome ili u oba prijepisa došlo do pogreške. Nažalost, to će u slučaju digitalizacije hrvatskih crkvenoslavenskih tekstova biti vrlo teško izvedivo, pa će se vjerojatno morati ostati pri samo jednome unosu koji se zbog toga treba pažljivo izvršiti.

<sup>31</sup> Autor je toga programa Antonio Magdić, a urednica Marica Čunčić.

<sup>32</sup> Takva mjesta u kojima se unosi razlikuju utvrđuju se računalnim preklapanjem tekstova.

### III. Svojstva korpusa za SHR<sup>33</sup> i njihova usporedba sa svojstvima korpusa za RCJHR

Amir Kapetanović, voditelj projekta izrade SHR-a<sup>34</sup> za koji se pretpostavlja da će imati oko 50 000 lema, korpus za taj rječnik ovako određuje: »računalno čitljiv, označen, jednojezični, nespecijalni statični povijesni korpus pisanoga jezika sastavljen od cjelovitih tekstova« (2007:177). Navodeći da je korpus za SHR nespecijalan, autor kani reći da je riječ o tzv. (općem) referentnome korpusu kako se obično takvi korpusi nazivaju u leksikološkoj literaturi (npr. Teubert i Čermáková 2004:118 i dalje; Pearson 1998:44; Baker 2006<sup>8</sup>:43 i dalje; u više članaka u Wilson, Archer, Rayson 2006). Potrebno je pojasniti Kapetanovićevo opisivanje korpusa za SHR kao statičnog. Autor time zapravo želi reći da ne mogu nastati novi starohrvatski tekstovi. S obzirom da se novootkriveni ili previđeni starohrvatski tekstovi mogu unijeti u korpus, on ipak sadrži i svojstva dinamičkoga korpusa, baš kao što je slučaj i s korpusom za RCJHR. Također, s obzirom da se korpus za SHR opisuje kao označen, dobro je reći da se sastavljači korpusa nadaju korpus sintaktički raščlaniti<sup>35</sup> i označiti, ali trenutno do sada uneseni tekstovi u korpus nisi označeni.

Struktura se korpusa za SHR zasniva na sadržaju tekstova, pa se korpus sastoji od književnih tekstova pisanih srednjovjekovnim hrvatskim jezikom raspoređenih u sedam sastavnica (Kapetanović 2007:177): *pjesništvo i drama; proza I*, koju čine djelomični prijevodi Biblije, molitvenici i lekcionari; *proza II*, koju čine apokrifi, vizije, Marijina čudesa, svetačke legende, romani, priče, poučna proza i drugo; *pravni tekstovi; povijesni tekstovi; kratki natpisi*, koju čine natpisi, grafiti i drugovrsni zapisi; *pabirci* koji sadrže glose, hrvatske jezične jedinice sakupljene iz hrvatskih crkvenoslavenskih izvora te hrvatska imena i titule iz latinskih izvora. Samo sadržaj sastavnice *pabirci* ne udovoljava posljednjemu dijelu Kapetanovićeve određenja korpusa za SHR navedenoga na početku ovoga poglavlja, jer nije riječ o cjelovitim tekstovima. Ta je sastavnica ipak samo pomoćna i rabi se samo kada nema dovoljno potvrda u izvorima drugih sastavnica. Za usporedna istraživanja, ona koje će uzimati u obzir i materijal iz korpusa za RCJHR, sastavnica *pabirci* bit će također važna, jer bi ona trebala sadržava-

<sup>33</sup> Zahvaljujem Amiru Kapetanoviću na pomnom iščitavanju ovoga poglavlja i korisnim prijedlozima vezanima za tekst čitavoga ovoga priloga.

<sup>34</sup> Prethodnicom se ovoga projekta može držati nastojanje da se izradi rječnik na osnovi spomenika do Marulićeva vremena, o čemu v. Malić 2002.

<sup>35</sup> Taj se postupak u žargonu naziva parsiranjem (eng. *parsing*; od lat. *pars orationis*). Riječ je o određivanju granica jedinica teksta (riječi, leksema) te njihovih gramatičkih obilježja.

ti i »izbor tipičnoga hrvatskoga općega leksika iz crkvenoslavenskih vre-  
la« (Kapetanović 2007:176).

Korpus za SHR sastoji se od tekstova iz rukopisa i inkunabula nastalih između XI. i XV. odnosno XVI. stoljeća. Riječ je o tekstovima pisanima hrvatskom književnom čakavštinom, a u dubrovačkim su tekstovima zabilježeni i štokavski elementi. Iako se tekstovi u korpus za SHR nužno ne unose iz izvornika ili preslika (kao što je slučaj u ekscerpriji za RCJHR), točnost se unesenih tekstova provjerava upravo usporedbom s izvornicima. Tekstovi se ne ekscerpiraju, a sve se inačice jednoga teksta u korpusu tretiraju jednako — nijedna se inačica ne izdvaja kao osnovni tekst. U digitaliziranome obliku tekstovi postaju dijelom korpusa za SHR, a nakon kritičke obrade kane se objavljivati i kritička izdanja pojedinih sastavnica. Nedavno su tiskani tekstovi prve sastavnice korpusa (Kapetanović, Malić, Štrkalj Despot 2010), a po odluci priređivača objavljeni su kao zbirka svih pripadajućih tekstova.<sup>36</sup> Trenutno se radi na zbirci biblijskih tekstova te tekstova iz molitvenikâ i lekcionarâ.

Korpus za SHR stvara se na standardima koja mu omogućuju da jednom s ostalim računalnim korpusima koji se izrađuju u Institutu za hrvatski jezik i jezikoslovlje može funkcionirati kao sastavnica velikoga korpusa hrvatskoga jezika. Tomu pogoduje i odluka da tekstovi toga korpusa budu grafijski normalizirani na osnovi fonološke transkripcije<sup>37</sup>. Grafijske su se razlike mogle zanemariti, odnosno mogla su se odrediti stroga normalizacijska pravila, jer korpusu nije nakana da služi i za potrebe paleografskih istraživanja (u tome bi slučaju prednost trebalo dati prijepisu<sup>38</sup>) kao što je slučaj s korpusom za RCJHR. Obradene i uspoređene značajke dvaju korpusa o kojima je ovdje riječ mogu se sažeti kako je prikazano u tablici 1.

<sup>36</sup> S obzirom da nije riječ o odabranim tekstovima — nije riječ o hrestomatiji, a kako nema razlikovanja filološke vrijednosti pojedinih inačica istoga teksta niti je učinjen kritički aparat — nije riječ niti o kritičkome izdanju, pa se čini da se (barem zasad) odustalo od pripreme kritičkih izdanja. Usp. »Sustavno se pristupilo izradi kritičkih izdanja hrvatskih srednjovjekovnih tekstova i njihovoj digitalizaciji.« (Kapetanović 2007:176).

<sup>37</sup> Fonološka je transkripcija postupak prijenosa grafema kojim se bilježi određeni fonem iz jednoga slovopisa određenoga pisma u drugi slovopis istoga ili drugoga pisma (npr. »czaašt« (Vrančić 1971), »čast« (Kašić 1990), »cháfzt« (Habdelić 1670) za suvremeni hrv. »čast«; podaci prema: Portal hrvatske rječničke baštine).

<sup>38</sup> Postupkom se prijepisa označuje dosljedno prenošenje grafema iz izvornika (npr. kada se Vrančićeva skripcija »czaašt« zapisuje vjerno, kao »czaašt«). O određenjima fonološke transkripcije, transliteracije i prijepisa te njihovoj primjeni u korpusu za SHR v. Kapetanović 2007:177 – 180).

Tablica 1. Značajke korpusa za RCJHR i korpusa za SHR

podudarna	svojstva korpusa za RCJHR i korpusa za SHR			korpus za SHR
	djelomično razlikovna korpus za RCJHR	korpus za SHR	razlikovna korpus za RCJHR	
referentni	struktura korpusa zasnovana na sadržaju kodeksa i sadržaju tekstova (načelno jedinstvena struktura)	struktura korpusa zasnovana na sadržaju tekstova (sedam relativno nezavisnih sastavnica)	korpus za RCJHR načelno dovršen	korpus za SHR u izradi
povijesni	neorganski idiom kojemu je osnova ne-hrvatski idiom — starocrkvenoslavenski, a uz njega i hrvatski organski idiom (ponajprije čakavski)	neorganski idiom u kojemu osnovu čine hrvatski organski idiomi (čakavski, tragovi štokaštine)	oko 18 100 lema	pretpostavlja se da će biti oko 50 000 lema
sadrži svojstva i statičnih i dinamičkih korpusa	objavljuju se kritička izdanja tekstova uključenih u korpus	objavljuju se zbirke tekstova koje čine pojedine sastavnice korpusa (barem zasad nije riječ o kritičkim izdanjima)	računalno nečitljiv	računalno čitljiv
pisani jezik (rukopisno tradiranih tekstova i inkunabula <sup>39</sup> )			usporedan	jednojezičan
tekstovi počinju iz (XI/) XII – XVI. st.			transliteracija	fonoška transkripcija
cjeloviti tekstovi			tekstovi su ekscerpirani	tekstovi nisu (niti neće biti) ekscerpirani
			razlikuje osnovni tekst i inačice toga teksta	sve inačice jednoga teksta tretira jednako
			osnovni, hrvatski crkvenoslavenski tekst, pisan starom ćirilicom; usporedni latinski tekstovi; pisani latinicom; a usporedni grčki tekstovi; pisani grčkim alfabetom	pisan latinicom
			uglavnom označen	neoznačen (trenutno)

<sup>39</sup> Među tiskanim izvorima, inkunabula nije samo Senjski obrednik (izvor za RCJHR), tiskan 1507. ili 1508. godine.



## V. Zaključak

Korpus za RCJHR trenutno se rabi za izradu RCJHR, ali i za druga jezikoslovna i filološka istraživanja hrvatskoga crkvenoslavenskoga idioma. S obzirom da je hrvatski crkvenoslavenski — crkvenoslavenski idiom na koji je utjecao hrvatski jezik, on je vrijedan izvor podataka za istraživanja jezične prakse u srednjovjekovnom hrvatskom društvu. Trenutno se u takvim istraživanjima korpus za RCJHR rabi tek u iznimnim slučajevima. Tako je vjerojatno i zbog nekoliko praktičnih razloga: pisan je starom ćirilicom, organiziran je kao papirna kartoteka ekscerpata, ustrojen je kao usporedan korpus (uz hrvatske crkvenoslavenske tekstove uglavnom se nalaze njihovi grčki i latinski prijevodni predlošci ili usporedni tekstovi).

Među različitim korpusima hrvatskoga jezika korpus za RCJHR najviše dodirnih točaka ima s korpusom za SHR koji je u izradi. Ta dva korpusa sadrže jezični materijal bliskih idioma iz praktički podudarnih vremenskih razdoblja (XI/XII. do XVI. st.). Osim toga oba su korpusa referentna, sadrže svojstva i statičnih i dinamičnih korpusa, sastavljena su od cjelovitih pisanih tekstova od kojih se mnogi nalaze u više inačica, a svi su izlučeni iz rukopisa i inkunabula. Djelomično se podudaraju u osnovama strukture. Također, i hrvatski crkvenoslavenski i starohrvatski knjiški su (neorganski) idiomi, ali imaju određene veze s organskim idiomima: hrvatski crkvenoslavenski je pod stalnim utjecajem hrvatskoga govora (ponajprije čakavskoga),<sup>40</sup> dok se starohrvatski obuhvaćen korpusom za SHR zasniva na hrvatskim organskim idiomima — na čakavskome s tragovima dubrovačke štokavštine bez utjecaja (barem ne većega) drugih slavenskih idioma.

Ne podudaraju se u tomu što korpus za RCJHR nije računalno čitljiv, uglavnom je označen, ustrojen je kao usporedan korpus u kojemu se razlikuje osnovni tekst od njegovih inačica, sastoji se od ekscerpiranih čitavih tekstova, poštivana su načela transliteracije na osnovi koje je jedino u lematizaciji provedena normalizacija oblika. Korpus za SHR računalno je čitljiv, trenutno neoznačen (iako se označavanje planira), jednojezičan korpus koji sadrži cjelovite tekstove svih inačica jednoga teksta i to zapisanih u skladu s načelima fonološke transkripcije. Također, hrvatski crkvenoslavenski tekstovi uključeni u korpus za RCJHR objavljuju se u kritičkim izdanjima, dok se tekstovi uključeni u korpus SHR (barem zasad) izdaju kao zbirke svih tekstova određene sastavnice korpusa. Konačno, RCJHR trebao bi imati više nego dva puta manji broj rječničkih članaka od SHR: u

<sup>40</sup> Pri tome treba uzeti u obzir da je i starocrkvenoslavenski neorganski idiom, i to nastao na osnovi organskoga idioma koji nije hrvatski, nego govor drugih južnih Slavena — u prvome redu onih koji su živjeli u okolici Soluna, s utjecajem slavenskoga govora Moravljana.

RCJHR ima ih manje od 18 100, a za SHR se pretpostavlja da će ih biti oko 50 000.

Sretna je okolnost da se korpus za SHR oblikuje tako da se u prve tri njegove sastavnice (od sedam) nalazi većina vrsta tekstova od kojih se sastoji korpus za RCJHR. Zato se možda neće morati čekati do kraja izrade svih sedam sastavnica korpusa za SHR za prva ogledna usporedna istraživanja. Osim navedene tri sastavnice, istraživače dodira hrvatskoga crkvenoslavenskoga i srednjovjekovnoga hrvatskoga posebno će zanimati posljednja sastavnica (*pabirci*) koja bi trebala sadržavati starohrvatske elemente pronađene u crkvenoslavenskim tekstovima. S obzirom da je korpus za SHR jednojezičan, a u prve tri sastavnice (možda i u ostalima) on sadrži nezanemarljiv broj tekstova prevedenih s latinskoga, za istraživače je srednjovjekovne jezične prakse govornika hrvatskoga jezika koristan korpus za RCJHR i utoliko što sadrži brojne latinske tekstove — prijevodne predloške ili usporedne tekstove — od kojih potječu tekstovi sakupljeni u korpusu za SHR. K tomu, tu se nalaze i hrvatske crkvenoslavenske inačice prijevoda istih tih latinskih tekstova, čime se stvara relativno zaokružena cjelina u pogledu dodira hrvatskoga crkvenoslavenskoga idioma i srednjovjekovne književne čakavštine.

Navedene razlike utječu na mogućnosti istraživanja koja bi obuhvaćala dva navedena idioma na osnovi dvaju korpusa o kojima je ovdje riječ. Nastave li se digitalizirati hrvatski crkvenoslavenski i starohrvatski tekstovi može se očekivati da će u budućnosti istraživanja koja bi obuhvaćala podatke iz hrvatskoga crkvenoslavenskoga i starohrvatskoga biti lakše izvediva, brojnija i bolje kakvoće. Računalna bi pretraživost tih dvaju korpusa omogućila mnoga usporedna jezična, leksikografska istraživanja i istraživanja iz područja znanosti o književnosti koja su sada teško izvediva ili zapravo neizvediva. Sadašnji, papirni korpus za RCJHR još je, a vjerojatno će još dugo i biti, nezamjenjiv izvor hrvatskoga crkvenoslavenskoga jezičnoga materijala. Prije svega jer sadrži vrlo veliku količinu pouzdano transliteriranih hrvatskih crkvenoslavenskih tekstova i pitanje je kada će se toliki jezični materijal stići unijeti u računalnu zbirku hrvatskih crkvenoslavenskih tekstova. Kada se uz postojeći korpus za RCJHR dovrši digitalizacija tekstova u okviru korpusa za SHR te digitalizacija hrvatskih crkvenoslavenskih tekstova, nastat će izdašna osnova za različita istraživanja dijasistema (onako kako se taj pojam shvaća u Katičić 1992) u okviru kojega se odvijala jezična praksa u srednjovjekovnim hrvatskih zemljama.<sup>41</sup>

<sup>41</sup> Prikazani rezultati proizašli su iz znanstvenoga projekta »Rječnik crkvenoslavenskoga jezika hrvatske redakcije«, provođenog uz potporu Ministarstva znanosti, obrazovanja i športa Republike Hrvatske.

## Literatura

- Baker, Paul. 2006<sup>8</sup>. *Using corpora in discourse analysis*. London – New York: Continuum. 198str.
- Borin, Lars (ed.): *Parallel corpora, parallel worlds. Selected papers from a symposium on parallel and comparable corpora at Uppsala University, Sweden, 22–23 April, 1999*. Amsterdam : Rodopi. i,220str.
- Bratulić, Josip. 1981. Ediciona praksa hrvatskih istraživača i izdavača srednjovjekovnih tekstova u XIX i XX stoljeću (Historijski prikaz). U zb. *Международни научни скуп: Текстологија средњовековних јужнословенских књижевности, 14–16. новембра 1977. Димитрије Богдановић* (ур.). Београд : САНУ. 137–147.
- Čunčić, Marica, Antonio Magdić. 2008. The scientific program of the Old Church Slavonic Institute Zagreb, Croatia. Heinz Miklas, Anissava Miltenova (edited). U zb. *Slovo: Towards a Digital Library of South Slavic Manuscripts. Proceedings of the international conference 21–26 February 2008 Sofia, Bulgaria*. Sofia : »Bojan Penev« Publishing Center. 265–280.
- Drugi novljanski brevijar = II. *novljanski brevijar. Hrvatskoglagoljski rukopis iz 1495. Župni arhiv Novi Vinodolski. Fototipsko izdanje. Uvod i bibliografija* (na hrvatskom i engleskom). Priredile Marija Pantelić i Anica Nazor. Zagreb : Staroslavenski institut »Svetozar Ritig«, Turistkomerc. 46, XXV str., 500 listova s faks., [4] lista s tablama.
- Ferguson, Charles A. 1959. Diglossia. *Word* 15. 325–340.
- Grabar, Biserka, Franjo Većeslav Mareš, Ivana Mulc. 1991. Oblikovanje i sastav natuknice. *Rječnik crkvenoslavenskoga jezika hrvatske redakcije. (sveščić 1., Uvod)*. Zagreb : Staroslavenski zavod Hrvatskoga filološkog instituta. VIII–XVIII. (prijevod na engleski XIX–XXX.)
- Грковић-Мејдор, Јасмина. 2007. Rječnik crkvenoslavenskoga jezika hrvatske redakcije, I. svezak (а–врѣдъ), Staroslavenski institut, Zagreb, 2000, 576 str. (prikaz), *Прилози за књижевност, језик, историју и фолклор LXXII/1-4*, 185–187.
- Habdelić, Juraj. 1670. *Dictionar ili rechi szlovenszke zvezega ukup zebrane, u red posztaulyene, i Diachkemi zlahkotene trudom Jurja Habdelicha, masnika Tovarustva Jesusevoga, na pomoch napredka u diachkom navuku skolneh mladenczeu horvatszkoga i szlovenszkiego naroda*. Nemski Gradecz, pri odvetku Widmanstadiusa. 458+27str.
- Hrvatska jezična mrežna riznica. Dostupna na mrežnim stranicama: <http://riznica.ihj.hr/dokumentacija/index.html.hr>
- Hrvatski nacionalni korpus. Dostupan na mrežnim stranicama: <http://www.hnk.ffzg.hr/>.
- Hrvojev misal = *Missale Hervoiae Ducis Spalatensis Croatico-glagoliticum*.

- Transcriptio et Commentarium*. Zagreb – Ljubljana – Graz: Staroslavenski institut »Svetozar Ritig« – Mladinska knjiga – Akademische Druck- und Verlagsanstalt, 1973. Sv. 1.: potpuno faksimilirano izd. originala iz Topkapi Sarayi muzeja u Istanbulu. 494 str. (faks., inicijali i minijature u bojama). Sv. 2.: *Transcriptio et commentarium / editionem curaverunt Biserka Grabar, Anica Nazor, Marija Pantelić; sub redactione Vjekoslav Štefanić*. xiii, 544 str.
- Jagić, Vatroslav. 1879. *Quattuor evangeliorum Codex glagoliticus olim Zographensis nunc Petropolitanus. Characteribus cyrillicis transcriptum notis criticis prologomenis appendicibus auctum adiuvante summi ministerii Borussici liberalitate edidit V. Jagić*. Berolini: Apud Weidmannos. 174 str. (dostupno i na mrežnoj stranici: <http://logol.info/files/zograf.pdf>).
- Kapetanović, Amir. 2007. Digitalizacija korpusa starohrvatskih tekstova i kritika teksta. U zb. *The Future of Information Sciences: INFUTURE2007 - Digital Information and Heritage*. Seljan, Sanja i Stančić, Hrvoje (ur.). Zagreb: Odsjek za informacijske znanosti, Filozofski fakultet. 173–182.
- Kapetanović, Amir, Dragica Malić, Kristina Štrkalj Despot. 2010. *Hrvatsko srednjovjekovno pjesništvo, Pjesme, plačevi i prikazanja na starohrvatskom jeziku*. Zagreb: Institut za hrvatski jezik i jezikoslovlje. lix, 884 str.
- Kašić, Bartol. 1990[1599]. *Hrvatsko-talijanski rječnik s konverzacijskim priručnikom*. Zagreb: Kršćanska sadašnjost i Zavod za jezik IFF (izvornik tiskan u Rimu 1599). 198 str.
- Katičić, Radoslav. 1992. 'Slovênski' i 'hrvatski' kao zamjenjivi nazivi jezika hrvatske književnosti. U knj. *Novi jezikoslovni ogledi*. Zagreb: Školska knjiga. 312–328.
- Kleiber, Georges. 1978. *Le mot »ire« en ancien français (XIe – XIIe siècles). Essai d'analyse sémantique*. Paris: Klincksieck. 488 str.
- Korpus za izradu RCJHR. Zagreb: Staroslavenski institut.
- Korpus za izradu *Starohrvatskoga rječnika*. Zagreb: Institut za hrvatski jezik i jezikoslovlje.
- Lučić, Vida. 2004. Primjer leksikografski zanemarene metaforizacije. *Filologija* 43, 55–75.
- Malić, Dragica. 2002. *Nacrt za Hrvatski rječnik do Marulića i njegovih suvremenika*. Zagreb: Institut za hrvatski jezik i jezikoslovlje. 154 str.
- Mareš, František Václav. 2007(1962). Návrh přípravných prací pro slovník jazyka církevněslovanského. *Církevněslovanská lexikografie 2006*. Sestavil Václav Čermák; Editoři: Emilie Bláhová, Eva Šlaufová, Václav Čermák. Praha: Slovanský ústav AV ČR, v. v. i. i Euroslavica. 64–84. (ruski prijevod: Мареш, Франтишек Вацлав. 1966. Проект подготовки словаря церковнославянского языка. *Вопросы языкознания XV*. Москва: Академия наук СССР, Институт языкознания. 86–99.)

- Mihaljević, Milan. 2011. Položaj crkvenoslavenskoga jezika u hrvatskoj srednjovjekovnoj kulturi. U: *Zbornik na трудови од Меѓународниот научен собир »Свети Наум Охридски и словенската духовна, културна и писмена традиција«* (организиран по повод на 1100-годишнината от смртта на св. Наум Охридски). Охрид, 4–7 ноември 2010. Велев, Илија, Ацо Гиревски, Лилијана Макаријоска, Илија Пиперкоски, Костадина Мокрова (ur.). Скопје. 229–238.
- Nazor, Anica. 1991. Uvod; Popis izvora; Navedena literatura. *Rječnik crkvenoslavenskoga jezika hrvatske redakcije. (sveščić 1., Uvod)*. Zagreb : Staroslavenski zavod Hrvatskoga filološkog instituta. I–III (prijevod na engleski IV–VII.); XXXI–XXXVI; XXXVII–XXXIX.
- Назор, Аница. 1996. О словаре хорватско-глаголической редакции общеславянского литературного (церковнославянского) языка. *Вопросы языкознания XV*. Москва : Академия наук СССР, Институт языкознания. 99–105.
- Nazor, Anica. 2000. Prožimanje glagoljice i ćirilice na hrvatskom prostoru. U knj. *Hrvatska i Europa. Kultura, znanost i umjetnost*. Svezak II. (Srednji vijek i renesansa (XIII–XVI. stoljeće). Uredio Eduard Hercigonja. Zagreb : Hrvatska akademija znanosti i umjetnosti – Školska knjiga. 288–297.
- Nazor, Anica. Rječnik crkvenoslavenskoga jezika iniciran na IV. međunarodnom slavističkom kongresu u Moskvi 1958. godine (u povodu 50. obljetnice inicijative). U zb. Vidjeti Ohrid. *Referati hrvatskih sudionika i sudionika za XIV. međunarodni slavistički kongres (Ohrid, 10.–16. rujna 2008.)*. Uredio Marko Samardžija. Zagreb : Hrvatsko filološko društvo – Hrvatska sveučilišna naklada, 2008. 65–82.
- Portal hrvatske rječničke baštine. Dostupan na mrežnoj stranici: <http://crodip.ffzg.hr/>.
- Raffaelli, Ida. 2001. *Polisemija i sinonimija u leksiku klasičnoga starofrancuskog*. Neobjavljena doktorska radnja. Zagreb : Sveučilište u Zagrebu, Filozofski fakultet. 350str.
- RCJHR 2000 = *Rječnik crkvenoslavenskoga jezika hrvatske redakcije. I. sv. A–vrêdb*. Zagreb : Staroslavenski institut. 2000.
- Sinclair, John. 2003. Corpora for lexicography. U knj. van Sterkenburg, Piet. (ed.) *A Practical Guide to Lexicography*. Amsterdam/Philadelphia : Benjamins Publishing Company. 167–178.
- Svensén, Bo. 2009. *A Handbook of Lexicography*. Cambridge : Cambridge University Press. 535 str.
- Šimić, Marinka. 2010. Hrvatska ćirilica. U knj.: *Knjige od likarije. Faksimil i obrada dviju ljekaruša pisanih hrvatskom ćirilicom*. Zagreb : HAZU. 233–248.

- Štefanić, Vjekoslav. 1960. *Glagoljski rukopisi otoka Krka. (Djela JAZU, Knjiga 51., ur. Marko Kostrenčić)* Zagreb : JAZU. 456 str.
- Štefanić, Vjekoslav. 1962. Problem rječnika južnoslavenskih redakcija staroslavenskog jezika. *Slovo* 11–12. 181–187.
- Teubert, Wolfgang, Anna Čermáková 2004. Directions in corpus linguistics. U knj. Halliday, Michael Alexander Kirkwood, Wolfgang Teubert, Colin Yallop, Anna Čermáková. *Lexicology and Corpus Linguistics. An Introduction*. London – New York : Continuum. 113–165.
- Vrančić, Faust. 1971. *Rječnik pet najuglednijih evropskih jezika, latinskog, talijanskog, njemačkog, hrvatskog i mađarskog*, prvotisak Mleci 1595, reprint izdanje. Zagreb : Liber. (izvornik: Faustus Verantius. *Dictionarium quinque nobilissimarum Europae linguarum Latinae, Italicae, Germanicae, Dalmaticae et Ungaricae cum vocabulis Dalmaticis quae Ungari sibi usurparunt*. Venecija, 1595.). 6+135 str.
- Vrtič, Ivana. 2009. *Sintaksa Kašićeva prijevoda Svetoga pisma*. Neobjavljena doktorska disertacija. Zagreb : Filozofski fakultet Sveučilišta u Zagrebu. 458 str.
- Weinreich, Uriel. 1954. Is a structural dialectology possible. *Word* 10, 388–400.
- Wilson, Andrew, Dawn Archer, Paul Rayson (edited). 2006. *Corpus linguistics around the world* (Language and computers: Studies in practical linguistics. No. 56.) Amsterdam – New York: Rodopi. 241 str.
- Zgusta, Ladislav. 1991. *Priručnik leksikografije*. Sarajevo : Svjetlost. 335 str.
- Žagar, Mateo. 2005. Grafolingvistika i stari tekstovi: osebujnosti pisanog jezika. U zb. *Od fonetike do etike. Zbornik o sedamdesetogodišnjici prof. Dr. Josipa Silića*. Uredio Ivo Pranjković. Zagreb : Disput. 223–258.

On the Corpus of the *Dictionary of Church Slavonic Language of Croatian Redaction* and its Relation  
Towards the Croatian Language Corpora

Abstract

The corpus for the Dictionary of the Croatian Redaction of Church Slavonic (acronym: DCRCS) is a corpus of Croatian Church Slavonic excerpted texts created in the Old Church Slavonic Institute in Zagreb not only for the dictionary compilation but also for paleographic, grammatical, philological and semantic research. Among corpora of the Croatian language, the greatest number of characteristics it shares with the corpus for the compilation of the Dictionary of Old Croatian (acronym: DOC) which has been created in

the Institute of Croatian Language and Linguistics in Zagreb. They consist of the language material of close idioms (Croatian Church Slavonic and medieval Croatian, respectively), from practically identical period (XI/XII–XVI c.). Also, they are both referential corpora, having characteristics of static and dynamic corpora; both consist of integral written texts (of which many in several variants) extracted from the manuscripts and incunabulas. They partially correspond in terms of their basic internal structure: the structure of the corpus of the DCRCS is based on the codex contents and the contents of particular texts, while the structure of the corpus for the DOC is primarily based on the contents of the texts.

Corpora for the DCRCS and for the DOC differ in a number of characteristics. In contrast to the corpus for the DOC, the corpus for the DCRCS is organized as a computer-unreadable paper card-file (only partially it is available in JPEG format). In addition, it is mostly annotated (the corpus for the DOC is not currently annotated, but hoped to be in the future), parallel (the corpus for the DOC being monolingual). The texts for the corpus for the DCRCS are transliterated in Old Cyrillic script and normalized forms are present only in lemmatization (the texts of the corpus of the DOC are inscribed in Latin script and standardized according to the principles of phonological transcription).

The existing corpus for the DCRCS is, and most probably will remain for long time or for evermore, an irreplaceable source of the Croatian Church Slavonic language data even though a valuable corpus of digitalized Croatian Church Slavonic texts has been created in the Old Church Slavonic Institute. Hopefully, the corpus for the DCRCS, combined with the corpus for the DOC and the mentioned corpus of the Croatian Church Slavonic texts (both in progress) will once create a very solid ground for the extensive research on the linguistic praxis in the medieval Croatian lands.

**Ključne riječi:** korpus, Rječnik crkvenoslavenskoga jezika hrvatske redakcije, Starohrvatski rječnik

**Key words:** corpus, Dictionary of the Croatian Redaction of Church Slavonic, Dictionary of Old Croatian

