

Jasmina D. Moskovljević Popović*

Filološki fakultet, Beograd**

Milica D. Dinić Marinković

Filološki fakultet, Beograd

811.163.41'26-053.5

doi 10.18485/analiff.2015.27.2.16

O RAZVOJNOM KORPUSU PISANOG JEZIKA I MOGUĆNOSTIMA NJEGOVE PRIMENE U PRAĆENJU JEZIČKOG RAZVOJA

Osnovni cilj ovog rada je da ukaže na značaj postojanja reprezentativnog korpusa koji na sistematičan način prati razvoj sposobosti za pisanu komunikaciju, kao i da predstavi jedan takav korpus u nastajanju – Razvojni korpus pisanog srpskog jezika (RAKOPS).

U uvodnom delu rada ukazano je na važnost formiranja razvojnih korpusa pisanog jezika i izloženi su osnovni podaci o malobrojnim trenutno postojećim korpusima ove vrste. U centralnom delu rada predstavljen je RAKOPS i izneti su osnovni podaci o trenutnoj i budućoj strukturi ovog korpusa, načinu njegovog formiranja i digitalizaciji. U završnom delu rada razmotrene su mogućnosti primene RAKOPS-a u praćenju različitih aspekata jezičkog razvoja.

Ključne reči: razvojni korpus, pisani jezik, osnovnoškolski uzrast, srpski jezik, RAKOPS

1. Uvod.

Dugi niz decenija razvojna jezička istraživanja bila su usmerena mahom na ispitivanje usvajanja govora i različitih jezičkih kompetencija neophodnih za uspešnu i svrsishodnu usmenu komunikaciju. U najvećem broju slučajeva ispitanici u ovim istraživanjima bila su deca čiji se uzrast kretao od prvih meseci života do predškolskog uzrasta (0-5 god.). Ovak-

* Katedra za opštu lingvistiku, Filološki fakultet, Studentski trg 3, 11000 Beograd, jasmina.moskovljevic@fil.bg.ac.rs.

** Rad prvog autora ovog članka podržan je u okviru projekta „Standardni srpski jezik: sintaksička, semantička i pragmatička istraživanja“ (178004), koji finansira Ministarstvo prosvete i nauke Republike Srbije.

va situacija bila je, između ostalog, odraz teorijskih shvatanja u najvećoj meri inspirisanih teorijom Noama Čomskog, dominantnim lingvističkim modelom tokom čitave druge polovine 20. veka, u okviru koga je proces usvajanja jezika sagledavan kao: „brz, lagan, i u velikoj meri završen do uzrasta od tri do pet godina...” (Berman 2004: xi). Tek poslednjih decenija dvadesetog veka otpočela su istraživanja usmerena na kasniji jezički razvoj (Nippold 1988), koja su omogućila praćenje različitih aspekata usvajanja jezika tokom osnovnoškolskog i starijih uzrasta, uključujući, između ostalog, praćenje početnih faza opismenjavanja, kao i sveukupni razvoj sposobnosti za pisanu komunikaciju.

Značaj i uticaj pisane komunikacije u savremenom dobu teško je preценiti. Nagli i ubrzan razvoj elektronskih medija još više u prvi plan ističe upravo pisanu komunikaciju i čini je izuzetno važnim oblikom opštenja u 21. veku. Otuda je u poslednjih 20-ak godina značajno poraslo interesovanje za istraživanja specifičnih karakteristika pisanog jezika na različitim strukturnim nivoima (posebno na sintaksičkom, leksiko-sintaksičkom i tekstualnom), kao i za istraživanja koja se bave unutrašnjom strukturom i organizacijom različitih tipova i žanrova tekstova. Dok su studije posvećene odlikama pisanog jezika odraslih govornika relativno brojne, razvojnih studija koje se bave praćenjem procesa usvajanja daleko je manje.¹ Još je manji broj ovakvih studija koje su zasnovane na sistematskom praćenju podataka iz razvojnog korpusa iako bi upravo ti podaci mogli da nadomeste praznine u eksperimentalnim nalazima, koji su, po samoj svojoj prirodi, uvek parcijalni, i omoguće kreiranje jedne celovite i koherentne slike, kako o opštem toku razvoja pisanog jezika, tako i o specifičnostima razvoja njegovih pojedinih komponenti.

Nema sumnje da je osnovni razlog što broj ovakvih istraživanja nije veći to što ni razvojnih korpusa pisanog jezika koji bi pratili usvajanje kompetencija za pisano izražavanje izvornih korisnika jezika nema mnogo, a i oni koji postoje sadrže isključivo tekstove pisane na engleskom jeziku. Lee (2010: 116) navodi ukupno četiri razvojna korpusa pisanog jezika i u svim navedenim slučajevima autori tekstova koji su uključeni u korpus su izvorni govornici engleskog jezika. Prvi od njih, LUCY², sakupljen je na

1 Za izuzetke v. Indrisiano, R. & Squire, J. (eds.) (2000), MacArthur, C.A, Graham, S. & Fitzgerald, J. (eds.) (2006).

2 Sampson, G. (2005).

Univerzitetu u Saseksu i pored tekstova odraslih autora, koji čine najveći deo građe, sadrži i dva segmenta čiji su autori deca uzrasta 9–12 godina (veličine 30.000 reči), odnosno adolescenti (veličine 33.000 reči). Drugi, LOCNESS, nastao je kao rezultat istraživačkih napora u Centru za korpusnu lingvistiku Katoličkog univerziteta u Luvenu, veličine je 324.304 reči, i sastoji se od tekstova čiji su autori britanski maturanti (60.209 reči), britanski studenti (95.695 reči) i američki studenti (168.400 reči), tako da se samo uslovno može smatrati razvojnim korpusom.³ Slična je situacija i sa preostala dva korpusa – i BAWE (*British Academic Written English Corpus*)⁴, veličine 6.5 miliona reči, i njegov američki pandan, MICUSP (*Michigan Corpus of Upper-level Student Papers*), veličine 2 miliona reči, nastali su kompilacijom tekstova čiji autori reprezentuju samo jedan deo populacije, studentsku populaciju, i samo određene uzraste. Tako su u sastav korpusa BAWE ušli radovi studenata od prve godine dodiplomskih do prve godine master-studija, dok korpus MICUSP sačinjavaju radovi studenata završne godine dodiplomskih i prve godine postdiplomskih studija, i to isključivo oni radovi koji su ocenjeni najvišom ocenom. Otuda se i ova dva korpusa, a posebno MICUSP, samo uslovno mogu posmatrati kao razvojni. Uz to, odluka autora MICUSP-a da u njegov sastav uključe samo studentske radove ocenjene najvećom ocenom značajno smanjuje mogućnosti za praćenje napredovanja u razvoju pisanog jezičkog izraza, kao što sužava i mogućnosti za pun opseg međusobnog poređenje tekstova unutar samog korpusa.

Nasuprot tome, osnovni motiv pri kreiranju *Razvojnog korpusa pisanog srpskog jezika (RAKOPS-a)* bio je upravo da se omogući dosledno i sistematsko praćenje procesa usvajanja i razvoja pisanog jezika od njegovih ranih stadijuma pa sve do okončanja srednje škole.

2. Struktura RAKOPS-a, njegovo prikupljanje i digitalizacija.

Ideja o formiranju RAKOPS-a nastala je u okviru zajedničkog projekta Katedre za opštu lingvistiku i Centra za primenjenu lingvistiku

3 The Louvain Corpus of Native English Essays (LOCNESS)– <http://www.learner-corpusassociation.org/resources/tools/locness-corpus/>

4 BAWE (British Academic Written English) and BAWE Plus Collections – <http://www2.warwick.ac.uk/fac/soc/al/research/collections/bawe/>. V. i Nesi et al. (2004, 2010).

Filološkog fakulteta Univerziteta u Beogradu. Osnovni cilj ovog projekta je da se prikupljanjem reprezentativnog uzorka tekstova koji nastaju u prirodnim, uobičajenim uslovima na koje su deca navikla, znači u školi, tokom časa, obezbedi dovoljno velika i pouzdana građa koja će omogućiti sistematsko i dosledno praćenje razvoja pisanog srpskog jezika na uzrastu od 9 do 18 godina, odnosno na uzrastu od 3. razreda osnovne škole do 4. razreda srednje škole. Prva dva razreda osnovne škole isključena su jer se deca tokom njih tek obučavaju čitanju i pisanju, pa bi pribavljanje odgovarajućeg uzorka teksta bilo teško obezbediti.

Prikupljanje građe za korpus odvija se uvek po istoj proceduri i podrazumeva uvek isti sled koraka. Ispitivač dolazi u pripremnu posetu školi tokom koje se sa nadležnim nastavnicima i upravom škole dogovara o preciznom vremenu i toku ispitivanja, kao i o tome u kojim odeljenjima će se ono vršiti. Na taj način se sprečava da se prilikom samog ispitivanja gubi vreme na organizacione i administrativne začkoljice. Svako dete u istraživanju učestvuje u dva navrata, u razmaku od 10 do 15 dana. Istraživanje u oba navrata traje jedan školski čas. U svakom navratu dete dobija unapred pripremljeno pisano uputstvo da napiše jedan od četiri moguća tipa teksta – narativni, deskriptivni, ekspozitorni, ili argumentativni. Pošto svako dete piše dva različita rada, koja se svrstavaju u dva različita tipa teksta, prikupljena građa obezbeđuje mogućnost poređenja razlika koje se u pisanom jezičkom izrazu javljaju kao posledica variranja žanra, odnosno tipa teksta, kao što omogućava i praćenje razvoja sposobnosti za pisanje tekstova različitih žanrova.

Uključenost i ravnopravna zastupljenost različitih tipova tekstova predstavlja posebno značajno obeležje RAKOPS-a. U literaturi, naime, postoje brojni podaci o tome da kreiranje različitih tipova teksta od autora zahteva aktivaciju različitih kako jezičkih, tako i kognitivnih sposobnosti. Svaki oblik pisanja, kao i svaki žanr teksta odlikuju specifičnosti koje ga karakterišu i razlikuju od ostalih tipova teksta i otuda se proces pisanja ne sme posmatrati i analizirati kao uniforman (Deane *et al.* 2008:15–16). Uz to, do ovladavanja različitim tipovima teksta ne dolazi se na istom uzrastu, niti ono ima isti razvojni tok – dok se sposobnost za pisanje narativnih tekstova razvija relativno rano (uz značajno oslanjanje na naraciju u govoru), razvoj sposobnosti za pisanje ekspozitornih, a posebno argumentativnih tekstova teče daleko sporije i na njega utiču brojni dodatni činioci kao što

su dužina školovanja, stepen izloženosti različitim vrstama i žanrovima tekstova, količina vremena koja se posvećuje pisanju, kao i vrsta zadataka, uputstava i podsticaja koje deca dobijaju. Zbog svega navedenog, odluka da se u RAKOPS uključuju tekstovi različitog tipa u međusobno balansiranom odnosu, višestruko je opravdana.

Kada je o upotrebi pisma reč, učenici radove pišu pismom koje sami odaberu, što omogućava praćenje njihovih preferencija za upotrebu ćirilice ili latinice, kao i da li se i na kom uzrastu ove preferencije menjaju. Oni sami daju i naslov svome radu pošto su uputstva za pisanje koja dobijaju relativno uopštena i upućuju samo na širu temu, kao i očekivani tip teksta, ali ne specificuju precizan naslov rada. Kako je naslov teksta istovremeno i njegov (ekstremni) sažetak, ovakav istraživački postupak pruža priliku za praćenje razvojnog toka sposobnosti za sažimanje (sumarizaciju) teksta.

Tokom istraživanja prikupljaju se i podaci o uzrastu i polu ispitanika, razredu i školi koju pohađa, opštem uspehu u prethodnom razredu i oceni iz srpskog jezika. Ovakvom procedurom omogućava se dosledno praćenje većine relevantnih sociodemografskih varijabli koje karakterišu uzorak ispitanika, kao i kontrola njegove reprezentativnosti.

U narednoj fazi kompilacije korpusa, prikupljeni tekstovi prebacuju se u digitalni format. Ova faza sastoji se od nekoliko podfaza. U prvoj od njih radovi se pripremaju za digitalizaciju – svakom radu dodeljuje se odgovarajuća šifra koja kodira informaciju o autoru rada, razredu, školi i tipu teksta, i uparuju se dva rada koje je pisao isti učenik. Radovi koje nije moguće upariti (zato što je učenik tokom jednog od dva časa kada je uzorak prikupljan bio odsutan) odvajaju se i obeležavaju na poseban način. Formiraju se i na odgovarajući način kodiraju imena datoteka u koje se radovi smeštaju. U posebnu datoteku unosi se skup metapodataka koji obezbeđuje uspostavljanje veze između originalnih tekstova i dodeljenih kodova. Potom se vrši prekucavanje radova u procesoru *Microsoft Word*. Tokom prekucavanja verno se prenose sva obeležja originalnog teksta – pismo kojim je napisan (uključujući i mešanje alfabeta, ukoliko ga u radu ima), pravopisne i interpunkcijske oznake, grafički oblik teksta (raspored i pozicija naslova, pozicija paragrafa, i sl.) ostaje očuvan. Verno se prenose i sve greške koje se u radu javljaju. Korektnost prekucanog teksta i njegova saobraznost sa originalom dva puta se proveravaju od strane dva nezavisna pregledača. Tek po uklanjanju svih nesaglasnosti između originalnog i

prekucanog teksta, tekst se skladišti u odgovarajuću datoteku. U narednoj fazi digitalizacije tekstovi se iz .doc(x)–formata prebacuju u jedinstvenu XML bazu podataka, što korpus čini mašinski čitljivim, nezavisnim od platforme, odnosno od softvera i hardvera koji se koristi, i pogodnim za korišćenje programa za dalju obradu i anotaciju korpusa.

U svom finalnom obliku RAKOPS bi trebalo da sadrži 10.000 tekstova – po hiljadu tekstova sa svakog uzrasta od 3. razreda osnovne do 4. razreda srednje škole, odnosno po 2.500 tekstova narativnog, deskriptivnog, ekspozitornog i argumentativnog tipa. Smatra se da je ova veličina korpusa više nego dovoljna za sprovođenje validnih istraživanja i dobijanje pouzdanih podataka o razvoju pisanog jezičkog izraza (Chipere *et al.*, 2001).

U trenutku pisanja ovog rada korpus sadrži 3270 radova, odnosno 454.787 reči („tokens“). Pošto se svake godine uvećava za oko 800-1000 radova, svoj finalni oblik trebalo bi da dobije u narednih pet do sedam godina. Međutim, i u svom sadašnjem vidu korpus je moguće koristiti za različite vrste istraživanja i analiza odlika pisanog jezika, posebno kada je reč o trećem i osmom razredu, odnosno o uzrastima od 9 i 14 godina, za koje je broj radova uključenih u korpus već sada više nego dovoljan (v. Tabelu 1).

Razred:	Broj radova	% radova	Broj reči	% reči
3.	963	29%	109248	22%
4.	193	6%	31433	6%
5.	283	9%	44539	9%
6.	353	11%	54144	11%
7.	326	10%	57358	12%
8.	1152	35%	193072	39%

Tabela 1 – Trenutna distribucija tekstova po razredima

Dostupne su i početne analize kvantitativnih žanrovskih obeležja tekstova, kao što je moguće pristupiti i ispitivanju njihovih kvalitativnih obeležja. Kada je o kvantitativnim karakteristikama reč, iz Tabele 2 jasno se uočava da se tekstovi različitih tipova znatno razlikuju po prosečnoj dužini teksta izraženoj brojem reči – narativni tekstovi su najduži, slede

ih ekspozitorni i argumentativni, dok su deskriptivni tekstovi upadljivo najkraći (v. Tabelu 2). Iako su dopuna i ujednačavanje uzorka neophodni, kao i primena odgovarajućih statističkih procedura za utvrđivanje značajnosti uočenih razlika, različita prosečna dužina tekstova različitog tipa/žanra svakako predstavlja interesantan rezultat koji će tek daljim i detaljnijim istraživanjem moći da dobije svoje puno tumačenje. Naime, ne postoji ništa inherentno deskriptivnom tipu teksta što bi, bar na prvi pogled, moglo da prouzrokuje da njegova dužina bude za toliko manja od dužine ostalih tipova teksta. Naprotiv, pisanje deskriptivnih tekstova različitih vrsta jedan je od prvih zadataka sa kojim se deca sreću već na samom početku školovanja (1. – 2. razred), kao što predstavlja i jedan od najčešćih i najstandardnijih zadataka tokom čitavog perioda školovanja, ali su deskriptivni tekstovi koje učenici produkuju ipak upadljivo kraći od svih ostalih tipova teksta.

Tip teksta	Broj radova	% radova	Broj reči	% reči	Prosečna dužina teksta (broj reči)
Narat.	1413	43%	255080	52%	180
Deskr.	276	8%	29939	6%	108
Ekspo.	1293	40%	170365	35%	132
Argum.	288	9%	34410	7%	120

Tabela 2 – Trenutna distribucija različitih tipova teksta u korpusu i njihova prosečna dužina

I kada je reč o praćenju razvoja sposobnosti za pisano jezičko izražavanje u okviru različitih žanrova/tipova teksta, rezultati preliminarne analize podataka iz korpusa su interesantni, u nekim slučajevima i iznenađujući. Tako, na primer, poređenje prosečne dužine narativnih tekstova (izražene brojem reči) između trećeg i osmog razreda pokazuje da su narativi četrnaestogodišnjaka gotovo dvostruko duži od narativa devetogodišnjaka, što svakako nije neočekivano. Međutim, prilikom poređenja dužine ekspozitornih tekstova koje su pisali ti isti devetogodišnjaci i četrnaestogodišnjaci, razlike u dužini teksta gotovo da i nema (v. Tabelu 3). Ovo svakako nije ono što bi se moglo očekivati i utvrđivanje mogućih uzroka za ovakvo neobičnu distribuciju rezultata kada je reč o ekspozitornim tekstovima tek predstoji.

Razred:	III	VIII
Tip teksta:		
Narativni	106	200
Ekspozitorni	136	133

Tabela 3 – Uzasne razlike u prosečnoj dužini narativnog i ekspozitornog teksta izražene brojem reči

3. Mogućnosti primene RAKOPS-a.

Mogućnosti upotrebe RAKOPS-a mnogobrojne su i raznovrsne. Kao i kada je o drugim korpusima pisanog jezika reč, otvaraju se mogućnosti za iscrpne automatske i poluautomatske kvantitativne analize koje obuhvataju, ali se ne ograničavaju na kreiranje frekvencijskih lista leksema za pojedine uzrasne i/ili žanrovske segmente korpusa, ispitivanje obima vokabulara, praćenje razlika u dužini rečenice, pasusa i teksta, kao i poređenje tekstova različitog žanra po svim navedenim parametrima. Posebno je značajno to što će izvršene analize omogućiti i definisanje raspona u okviru koga se kreću vrednosti individualnih razlika, kao i praćenje uticaja sociodemografskih varijabli na te razlike.

Uz to, način na koji je RAKOPS osmišljen, kompiliran i digitalizovan omogućava kreiranje i izvođenje gotovo neograničenog broja različitih razvojnih istraživanja. Čak i njegov trenutni obim od 3270 tekstova, od kojih otprilike po hiljadu potiče iz trećeg i osmog razreda, dopušta analize (od kojih su neke već u toku) brojnih jezičkih odlika koje karakterišu pisani jezički izraz dece na ova dva uzrasta.

Tako je, na primer, kada je o trećem razredu reč (uzrast od 9 godina) interesantno pratiti greške koje nastaju kao posledica mešanja dva alfabeta (ćirilice i latinice) i ispitati okolnosti (reči, pozicije u reči, glasovni sklop) u kojima se ove greške najčešće javljaju. Leksičke greške, njihova tipologija i frekvencija njihove pojave u različitim tipovima teksta takođe predstavljaju plodno tlo za isitivanje. S obzirom da deca ovog uzrasta još uvek nisu u potpunosti usvojila (bar kad je o pisanom jeziku reč) kompletni inventar sintagmatskih i rečeničnih konstrukcija, posebno onih složenijih, zanimljivo je pratiti ne samo one koje su u njihovim tekstovima javljaju, već i one kojih nema, ili se u njihovoj upotrebi dosta greši. Posebno su interesantne greške koje se javljaju u segmentaciji jezičkog izraza jer omogućavaju da se prati kolike i koje jedinice informacija deca smeštaju u istu rečenicu.

Istraživanja u kojima se prati rano usvajanje različitih žanrovskih obeležja tekstova predstavljaju poseban izazov. Mada na ovom uzrastu većina dece pravi uočljivu razliku između konstrukcionih zahteva koje pred njih postavljaju tekstovi različitih žanrova, nije zanemarljiv ni broj one dece koja, bez obzira kakvu su temu i uputstvo dobila, uvek produkuju tekstove iste strukture i istog tipa (najčešće narativne).

Kada je reč o tekstovima koje su pisali učenici osmog razreda (uzrast 14–15 godina), praćenje vrednosti svih dosad pobrojanih parametara i njihovo poređenje sa korespondentnim vrednostima dobijenim za uzrast od 9 godina omogućava omeđavanje prostora unutar koga se odvija razvoj pisanog jezičkog izraza tokom osnovne škole. Uz to, analiza najboljih, najlošijih i prosečnih postignuća učenika osmog razreda pruža pouzdanu i empirijski verifikovanu osnovu za uspostavljanje obrazovnih standarda u oblasti *Pisanje* za kraj obaveznog obrazovanja.

Naravno, kada prikupljanje i digitalizacija RAKOPS-a budu u potpunosti okončani, odnosno kada korpus bude dostupan u svom punom obimu (10.000 tekstova) mogućnosti njegove primene dodatno će se proširiti. RAKOPS će biti relativno velika, dobro izbalansirana i reprezentativna baza tekstova, u prigodnom digitalnom formatu, koja će omogućiti detaljno i sistematsko praćenje različitih pokazatelja razvoja sposobnosti za pisano jezičko izražavanje uključujući (ali ne ograničavajući se na) obim i adekvatnost upotrebe vokabulara, upotrebu pravopisa i interpunkcije, pojavu i dinamiku razvoja različitih (posebno složenih) sintaksičkih konstrukcija, parametre semantičkog i pragmatičkog razvoja, kao i različite pokazatelje strukturne i funkcionalne organizacije teksta.

Literatura:

- Berman, R. A. (ed.) (2004): *Language Development across Childhood and Adolescence*. Amsterdam: John Benjamins.
- Chipere, N., Malvern, D., Richards, B., Duran, P. (2001): "Using a corpus of school children's writing to investigate the development of vocabulary diversity." *UCREL Technical Paper number 13. Special issue*. Iz: Rayson, P., Wilson, A., McEnery, T., Hardie, A., Khoja, S. (eds.), *Proceedings of the Corpus Linguistics 2001 Conference*. Lancaster

- University: UCREL, 126 – 133.
- Deane, P., Odendahl, N., Quinlan, T., Fowles, M., Welsh, C., & Bivens-Tatum, J. (2008): *Cognitive models of writing: Writing proficiency as a complex integrated skill*. ETS Research report RR-08-55.
- Indrisiano, R. & Squire, J. (eds.) (2000): *Perspectives on Writing: Research, Theory, and Practice*. Newark, DE: International Reading Association.
- Lee, D. Y. W. (2010): "What corpora are available?". Iz: O'Keeffe, A. & McCarthy, M., *The Routledge Handbook of Corpus Linguistics*. London: Taylor & Francis, 107–121.
- MacArthur, C.A., Graham, S. & Fitzgerald, J. (eds.) (2006): *The handbook of writing research*. New York: Guilford Press.
- Michigan Corpus of Upper-level Student Papers*. (2009). Ann Arbor, MI: The Regents of the University of Michigan.
- Nesi, H., Sharpling, G. & Ganobcsik-Williams, L. (2004) "Student papers across the curriculum: Designing and developing a corpus of British student writing". *Computers and Composition* Vol. 21, 4, 401-503.
- Nesi, H., Gardner, S., Thompson, P., Wickens, P., (2015): *British Academic Written English Corpus (BAWE)*. <http://www.coventry.ac.uk/research/research-directories/current-projects/2015/british-academic-written-english-corpus-bawe/?theme=main>
- Nippold, M. A. (1988): *Later Language Development – The School-Age and Adolescent Years*. Austin, TX: Pro-Ed.
- Rizzo, Camina Reo (2010): "Getting on with corpus compilation: from theory to practice." *ESP World*, Vol . 9, 1 (27). <http://www.esp-world.info>.
- Sampson, G. (2005): "The LUCY Corpus – Structure in Written English in the UK" <http://www.grsampson.net/rlucy.html>.

Jasmina Moskovljević Popović
Milica Dinić Marinković

Summary

ON DEVELOPMENTAL CORPUS OF WRITTEN LANGUAGE AND ITS POSSIBLE USES IN STUDYING CHILDREN'S LANGUAGE DEVELOPMENT

The major objective of this paper is to point out the significance of the existence of a representative corpus which provides reliable and systematic data about the written language development, as well as to introduce one such corpus in the making – the developmental corpus of Serbian written language, or RAKOPS.

The first part of the paper provides general information about the existing (and relatively few) corpora of this kind. In the next section, the structure and the design of the RAKOPS has been introduced and the process of its compilation and digitization has been outlined. In the final part of the paper the anticipated possible uses of the RAKOPS in the monitoring of various aspects of written language development are further elaborated and discussed.