

KOMPARATIVNOSLAVISTIČKE LINGVOKULTURALNE TEME



Uredile

Neda Pintarić

Ivana Čagalj

Ivana Vidović Bolt



srednja europa

Zagreb 2019.

409

MREŽNO GENERISANJE TEKSTOVA NA SLOVENSKIM JEZICIMA

Referat se sastoji iz sljedećih dijelova: definisanje predmeta analize, tumačenje pojma generator teksta, utvrđivanje istraživačkog korpusa, izdvajanje vrsta generatora, razlikovanje offline i online generatora. U prvoj se dijelu razmatraju lingvistički aspekti vještačkog generisanja tekstova, između ostalog u odnosu na prirodnu generaciju tekstova, u drugoj se vrši analiza osnovnih generatora, strukture i tipova tekstova koji se generišu na slovenskim jezicima. Istraživački korpus čini građa svih slovenskih jezika. Pošto se izvori za pojedine jezike nigdje ne mogu naći na istom mjestu u internetu, potrebno je dosta vremena da se stvori relevantan višejezički korpus (analizu otežava nepostojanje kompleksnih i sistemskih istraživanja date teme). Modeli generacije tekstova identični su tipološki za sve slovenske jezike (razlikuju se po načinu i u čestoti upotrebe), a čine ga 1) standardni jezici, 2) kategorije generisanja, 3) odnos polaznog (ulaznog) i završnog (izlaznog) teksta, 4) formula i shema automatske proizvodnje tekstova. Zbog ograničenog prostora u radu su izostavljeni svi citati i primjeri.

Ključne riječi: *generator, slovenski jezici, tekst, input, output, online, offline*

Online generator je program koji automatski proizvodi na bazi jednog teksta (T_1), ili nultog teksta (T_0), drugi tekst (T_{2gen}), primjenjujući pri tome procedure automatizacije, analize i sinteze, rješavanje formula i dr. za različite ciljeve: utilitarne (poboljšavanje ili ubrzavanje komunikacije, razmjena informacija, porast rejtinga...), ludističke, zabavne, stilističke (auto)prezentacione, demonstracione i sl. Prirodna sredina generatora je internet. Oni funkcionišu, po pravilu, online, rjeđe offline (instaliraju se iz Mreže i aktiviraju sa kompjutera). Generisani tekstovi odlikuju se heterogenošću (kreatori i korisnici mogu biti i profesionalci i amateri). Nekim programima postoji slobodan pristup, za korišćenje drugih treba platiti.

Generacija (G_c) pretpostavlja postojanje fakultativnog generirajućeg (ulaznog) teksta (T_1) i obligatornog generisanog (izlaznog) teksta (T_{2gen}). Ona obuhvata tri prostora: prostor lokacije T_1 (pozitorijum), prostor generacije T_2 (depozitorijum) i prostor distribucije T_2 (repozitorijum). Polazni T_1 ("izlazni") nalazi se u pozitorijumu (šablonu, formularu, tabeli, bazi podataka). Pomoću različitih procedura u depozitorijumu T_1 modifikuje se, transformiše T_{gen2} (G_{proz}), a onda prebacuje, odlaže u repozitorijum (na server, u elektronsku poštu, socijalnu mrežu i dr.). Generisanje se sastoji iz tri karike: $T_1/T_0 \rightarrow G_{proz} \rightarrow T_{gen}$. U procesu generacije stupaju

u interakciju dva teksta: motivirajući (T_1) i motivisani (T_{2gen}). Prvi je generirajući, drugi je generisani. Ulazni je tekst (T_1) jedan tekst (između ostalog i nulti – T_0), izlazni (T_{2gen}) – jedan tekst ili nekoliko tekstova (od dva do čak nekoliko miliona, pa i više). Stoga navedenu formulu treba precizirati: $T_1/T_0 \rightarrow G_{proz} \rightarrow T_{2gen[a,b,c,...x]}$. U Gt zapažaju se sljedeće zakonitosti: 1) što je manje riječi u T_{gen1} time je lakše generisati $T_{2gen[a,b,c,...x]}$, 2) što je viši prag predvidljivosti konstituenata T_1 time je lakše proizvesti $T_{2gen[a,b,c,...x]}$, 3) što je manje ekspresije, emocije, semantičkih i stilističkih pomjeranja u T_{gen1} time je lakše dobiti osmišljene $T_{2gen[a,b,c,...x]}$. Od T_1 stvara se mnoštvo $T_{2gen[a,b,c,...x]}$, ali obrnuto je rjeđe (da iz mnoštva T_1 nastaje jedan T_{2gen}). U funkciji T_1 može dolaziti precedentni tekst.

ArtifGen nema analoga u istorijskom razvoju jezika svijeta. Aktuelno stanje sugerise nekoliko mogućih scenarija razvoja: 1) broj generatora TxIntGn sve će se više povećavati, 2) objekti ArtifGen širiće se na sve diferencijalne vrste jezika, 3) TxIntGn sve će više zamjenjivati i (u nekim slučajevima, djelimično ili potpuno) potiskivati/istiskivati tekstove NatGen, 4) kvalitet ArtifGen postepeno će se približavati kvalitetu NatGen, dok se u bližoj i daljoj budućnosti ne izjednače, 5) što će više vremena prolaziti sve će više ArtifGen biti u stanju da obavlja danas (gotovo) nerealni zadatak: generisanje potpunih metafora (u širokom smislu), stvaranje razumne ekspresije i sasvim smislenih tekstova, između ostalog i književnomjetničkih.

Postoje dva tipa ArtifGen: monolingvalni (u okvirima jednog jezika) i multilingvalni (u okvirima nekoliko jezika). Prvi se zatvara u granice jednog jezika, dok drugi predstavlja transformaciju teksta T_1 jezika L_1 u tekst $T_2, T_3... T_x$ jezika $L_2, L_3... L_x$ i odlikuje se time što uvijek pretpostavlja postojanje najmanje dvaju tekstova: originalnog i prevodnog (u monolingvalnom ArtifGen fakultativno je postojanje motivirajućeg teksta na istom jeziku). Multilingvalno ArtifGen može biti sinhrono (trenutno pretvaranje teksta T_1 jezika L_1 u tekst T_2 jezika L_2) i konsektivno (prvo se prevodi tekst T_1 jezika L_1 na tekst T_2 jezika L_2 , zatim tekst T_1 jezika L_1 na jezik L_3 teksta T_3 itd.). Multilingvalno ArtifGen razmatra se u okvirima teorije prevođenja, mašinskog, odnosno automatskog prevođenja.

Na nivou grafike razrađuju se mrežni generatori koji su u stanju da transformišu tekst na jednom pismu u tekst na drugom. Ovdje postoje različite kombinacije: a) tekst na slovnom pismu **A** (npr. ćirilici) ↔ tekst na slovnom pismu **B** (npr. latinici), b) štampani tekst → rukopisni tekst, c) štampani tekst → multimedijalni (animacioni) tekst, d) tekst na slovnom pismu (npr. ćirilici) ↔ tekst na piktografiji (pismu čiji znakovi predstavljaju shematske crteže koji odlikavaju predmete stvarnosti i prirodne pojave), e) tekst na slovnom pismu (npr. ćirilici) ↔ tekst na ideografiji (njome se ne slika predmet, kao kod piktograma, već sadržaj, značenje riječi: egipatsko, kinesko pismo), f) tekst na slovnom pismu (npr. ćirilici) ↔ tekst

na slogovnom pismu (recimo, indijskom), g) tekst na fonografiji (slogovnom ili slovnom pismu) ↔ tekst na ideografiji (egipatskom, kineskom pismu).

Na fonetsko-fonološkom nivou centralni je problem pitanje mogu li mrežni generatori proizvoditi cjelovite usmene tekstove. Za to postoje povoljni uslovi pošto su u okviru strukturne i primijenjene lingvistike razrađeni i intenzivno se razrađuju odgovarajući instrumentariji. Ovdje centralnu opoziciju čini par pismeni govor ↔ usmeni govor. Proces ArtifGen može teći u oba pravca: 1) pismeni tekst → usmeni tekst, 2) usmeni tekst → pismeni tekst.

U drugom slučaju javljaju se dvije mogućnosti: a) transformisanje usmenog teksta u pismeni, b) generisanje suprasegmentarnih elemenata (akcenta, intonacije, sloga, prije svega) i njihovo fiksiranje u pismenom obliku.

Na tvorbenome se planu ne generiše cjelovit tekst, već proizvode pojedine derivacione jedinice. Za ArtifGen posebno su interesantne razrade u okviru aplikativne generativne analize tvorbe riječi 70-ih i 80-ih godina XX v.

Za ArtifGen semantika je jedan od najvažnijih nivoa jezika. Ona je veoma dobro proučena pa je razrađen niz modela značenja u različitim lingvističkim pravcima – strukturalizmu, funkcionalnoj gramatici, generativnoj gramatici, kognitivnoj lingvistici. Za ArtifGen posebno su interesantni: interpretaciona (generativna) gramatika N. Čomskog, generativna semantika Dž. Lakoffa, Dž. Mak-Kolija, M. P. Postala, model od smisla ka tekstu (jezičke sinteze) I. A. Meljčuka, J. D. Apresjana i A. K. Žolkovskog (model Smisao ↔ Tekst), model semantičkih mreža, model semantičkog kodiranja i dr.

Na leksičkom nivou posebno interesovanje izaziva sinonimizacija. Ona je važna za ArtifGen pa su razrađeni sinonimizatori (sinonimajzeri, distributeri kontenta, randomizatori, generatori unikatnih tekstova) – programi za automatsko generisanje rerajt tekstova pomoću sinonima iz baze podataka. ArtifGen operiše pojmom slučajne (randomske) riječi (neologizma) za čije se dobijanje razrađuju i posebni generatori – randomizatori. Važne su i ključne riječi (dolaze kao oporni elementi u generisanju i percipiranju teksta), stoga se prave generatori ključnih riječi, da se provjeri da li napisani tekst odgovara ključnim riječima, a takođe u cilju dobijanja višeg rejtinga u pretraživačima.

U procesu ArtifGen uzima se u obzir činjenica da su morfološke kategorije fiksirane za riječi nezavisno od njihove sintaksičke pozicije (pozicije u rečenici) i odvojene od konkretnog iskaza. Pošto morfološki oblici ne mogu da samostalno obrazuju tekst, ArtifGen je ograničeno na: a) vrste riječi, b) strukturne elemente riječi (morfeme su prvenstveno nastavci), c) morfološka variranja, d) kategorijalna obilježja (rod, broj, padež, glagolski način, vrijeme, vid i dr.). Dobru teoretsku osnovu za ArtifGen čine modeli opšte morfologije prirodnih jezika i modeli konkretnih jezika.

Rečenica je prva jezička jedinica koja ima predikativnost (dovođenje u vezu saopštenja sa vremenskim planom stvarnosti), smisaonu i intonacionu završenost, stoga je automatsko generisanje ove sintaksičke jedinice ključno za TxIntGn. Nije slučajno da zbog njenog velikog značaja postoje raznovrsni modeli formalizacije i interpretacije koji na ovaj i onaj način mogu biti iskorišćeni u ArtifGen.

Na stilističkom nivou ArtifGen dotiče pitanja: a) koliko su mrežni generatori u stanju da proizvode ekspresivnost, b) mogu li oni da izađu na kraj sa slikovitošću, c) mogu li se stvarati tekstovi na bazi osmišljenog narušavanja jezičkih normi, d) kako se i u kojoj mjeri generišu tekstovi različitih stilova. Što se tiče posljednjeg pitanja, već postoje generatori koji izbacuju tekstove svih funkcionalnih stilova. Najlakše je sa tekstovima zvaničnog karaktera, stilistički suvim, neekspresivnim (prvenstveno tipskim dokumentima), a najsloženije sa književnoumjetničkim.

Specifičnost slovenskih jezika – flektivni karakter, postojanje velikog broja gramatičkih oblika, razvijenost padežnog i vidskog sistema stvara znatne probleme za ArtifGen. Ovi jezici spadaju u sintetičke u kojima gramatički markeri nastavci, prefiksi, sufixi, unutrašnja fleksija, promjena akcenta, supletivni oblici zahtijevaju ozbiljnu obradu u procesu ArtifGen i suštinski ga usložnjavaju. Za analitičke jezike (romanske, germanske i dr.) ove teškoće nisu tako velike pošto se kod njih gramatičko značenje izvodi izvan okvira riječi i izražavaju pomoću veznika, prijedloga, članova i drugih pomoćnih riječi, a takođe reda riječi.

U analizi smo obuhvatili više vrsta tekstova (ne samo književnih, odnosno poetskih i njihovih sedam modela) na različitim jezicima te izdvojili 22 tipa. Ti se modeli sastoje od pet elemenata: naziv (slikoviti) modela, tip generacije, formula, primjer(i) iz baze podataka, shema. Postoje polazni (ulazni) i završni (izlazni) tekstovi. Polazni, ulazni (za generisanje) tekstovi T_1 mogu biti integralni ($T_{1Integr[a,b,c,...x]}$), segmentarni ($T_{1Segm[a,b,c,...x]}$) i elementarni ($T_{1Elem[a,b,c,...x]}$). U integralne ($T_{1Integr[a,b,c,...x]}$) spadaju potpuni tekstovi, pjesme, dokumenti, pisma i sl. Segmentarni tekstovi ($T_{1Segm[a,b,c,...x]}$) predstavljaju dijelove $T_{1Integr[a,b,c,...x]}$ tipa rime. Elementarni tekstovi ($T_{1Elem[a,b,c,...x]}$) dolaze, recimo, kao linkovi u koje se stavljaju generisani ankor (između $\langle a \rangle$ i $\langle /a \rangle$). Izlazni $T_{2gen[a,b,c,...x]}$ mogu biti, kao i T_1 , integralni ($T_{2genIntegr[a,b,c,...x]}$), segmentarni ($T_{2genSegm[a,b,c,...x]}$) i elementarni ($T_{2genElem[a,b,c,...x]}$). Odnos ulaznih i izlaznih tekstova najčešće (u 7 modela od 22) sa integralnim T_2 . Zbog ograničenog prostora nismo u mogućnosti da ih ukratko predstavimo.

Izdvajaju se četiri tipa generatora teksta: čovjek (H-generator), životinja (Z-generator), biljka, automat (A-generator); ovaj posljednji podrazumijeva i simbiozu čovjek – automat. Ako je tačna pretpostavka o tome da postoje vanzemaljci (čak da se među nama već nalaze), onda bi se izdvojila još jedna kategorija – X-generatori.

A-generatori dolaze u tri verzije: online, offline i on/offline. Offline generatori odlikuju se kinetičnošću (pokretljivošću) i manipulativnošću (mogućnošću upravljanja), vizualnošću (imaju vidljivu formu), dok se online generatori nalaze

skriveni od korisnika (s one strane monitora). Automati (A-generatori) mogu biti u obliku čovjeka (humanoidi), a mogu uopšte da ne liče na njega (nehumanoidi, npr. tzv. automatske sekretarice). Obje se kategorije realizuju u obliku lingvoida (koji su u stanju da proizvedu tekst) i nelingvoida (koji nemaju nikakvu jezičku funkciju, to su npr. nijemi roboti, koji recimo igraju fudbal, prepoznaju ukus piva, udaraju sa više ruka po dobošu).

Glavni predstavnici offline generatora su roboti (ovaj termin neki koriste i za oznaku pojedinih online generatora). Među njima za nas su najzanimljiviji A-lingvoidi, koji samostalno proizvode tekst, odnosno autonomno razvijaju sredstvo komuniciranja, mimo čovjekove volje.

Za online generatore bitno pitanje jeste kakva je budućnost interneta. Sada on ima oko tri milijarde korisnika (što je impozantno u odnosu na ukupan broj stanovnika Zemlje – sedam milijardi), do 2020. uvešće se tehnologija 5G, a do 2040. tehnologija G6. Na mjesto Weba 2.0 dolazi Web 3.0 kojim će se neutralisati razlika između offline i online. U dogledno vrijeme pojaviće se kvantni internet (zahtijeva temperaturu od -272 celzijusa i jako magnetno polje), bežična veza će sve više potiskivati žičanu, bez gadžeta će se ulaziti u internet (njih će zamijeniti čipovi u glavi, rukama i nogama), nestaće personalni računari (njihovu funkciju preuzeće terminali), možda će se ostvariti i mašta vezana za kretanje u vremenu (u prošlost i budućnost), koristiće se svjetlost za prenošenje podataka, pojaviće se masa novih izmišljenih jezika (sada im ima oko hiljadu), možda će nastali i univerzalni jezik – i sve to daće novu suštinsku dimenziju internetu, ljudskom jeziku i komunikaciji.

Izvršena analiza ukazuje na to da generatori proizvode različite tekstove: smisaone, polusmisaone, nalik na smisaone, pseudosmisaone i besmislene. TxIntGn odlikuje se različitom tematikom, različitom intencionalnošću, nejednakom praktičnom vrijednošću i kvalitetom, a takođe raznorodnim ciljevima: utilitarnim, ludističkim, stilističkim, razonodnim, (auto)prezentacionim, demonstracionim. Prirodna generacija tekstova (NatGen) razlikuje se od vještačke (ArtifGen), između ostalog, time što a) NatGen teče po prirodnim algoritmima, b) ArtifGen zahtijeva razradu pravila i komandi automatizacije i generacije, c) za ArtifGen koriste se modeli koji su za NatGen nedovoljni. TxIntGn često liči na prirodne tekstove, stoga od tri tipa odnosa prirodnih i vještačkih jezika: identičnosti ($A = B$), sličnosti ($A \approx B$) i različitosti ($A \neq B$) ovdje prevladuje sličnost. Npr. u mrežnoj poeziji izdvajaju se dvije kategorije: pravi stihovi i stihovi koji liče na prave. TxIntGn sve više prodire u svakodnevni život i ljusku djelatnost. Dobra osnova za ArtifGen su raznovrsni metodi i modeli formalizacije NatGen. Oni su više orijentisani na analizu nego sintezu, što je jedan od razloga zašto se rjeđe koriste za ArtifGen. Za stvaranje vještačkog intelekta u oblasti jezika potrebno je vrijeme (takve razrade ne zadiru duboko u daleku prošlost) i ulaganje velikih napora (ljudskih i tehničkih). ArtifGen za veoma kratak period kompjuterizacije i internetizacije (pola stoljeća)

nije mogla dostići visoki kvalitet i u potpunosti riješiti složena pitanja vezana za smisao, metaforizaciju i ekspresivnost. Mrežni generatori jedno su od sredstava provjere validnosti lingvističkih hipoteza i formalizovanih instrumentarija. To je svojevrsan lakmus papir za utvrđivanja prikladnosti lingvističkih modela za njihovo praktično korišćenje. ArtifGen otvara nove puteve u razvoju jezika i njegovom istraživanju.

Polazeći od rezultata izvršene analize može se dati niz prognoza i hipoteza. U bližoj budućnosti (malo vjerovatno) ili u dalekoj budućnosti (više vjerovatno) kvalitet i funkcionalnost ArtifGen približice se NatGen toliko da će biti veoma teško razlikovati prirodne tekstove od vještačkih (to se i sada, ali u vrlo ograničenim slučajevima, dešava). TxIntGn postepeno će se usavršavati i povećavati kvalitet i efikasnost, što će dovesti do toga da će u mnogim oblastima života i djelatnosti potpuno ili djelimično zamijeniti/potisnuti tekstove nastale klasičnim putem. Ekspanzija TxIntGn imaće nepovratan karakter. Predmet ArtifGen biće i ono što se danas uopšte ne dâ generisati. Tehnološki napredak u dalekoj/bliskoj budućnosti objediniće u jednu cjelinu dva automata – prirodni (ljudski mozak) i vještački (generator) tako da će razlika između njih uveliko biti neutralizovana. Tome će doprinostiti hibridizacija čovjeka, čiji će organizam biti načičkan čipovima, elektronskim mrežama, bazama podataka, kompleksima znanja i najrazličitijih generatorima, između ostalog, i onim koji će proizvoditi osmislene tekstove (teško je odgovoriti na pitanje da li će time čovjek prestati da bude čovjek, a automat mašina). Generatori će autonomno klonirati sami sebe, poboljšavajući i ubrzavajući komunikaciju. Hibridizovani (čipizovani i automatizovani) čovjek biće u stanju da proizvodi tekstove znatno različite od postojećih. Generacija tekstova od strane takvog simbioznog mehanizma (čovjeka-automata) dovešće do pretvaranja prirodno-vještačke dvojezičnosti u jednojezičnost, čije je i obrise i obim danas teško zamisliti. Generatori tekstova biće toliko razvijeni da će istisnuti mnoge profesije i vrste ljudske djelatnosti (već se danas postavlja pitanje o tome da li generatori „autopjesnici” ubijaju žive pjesnike). Umjetničko stvaralaštvo biće suštinski promijenjeno i prošireno tematski, metodološki, strukturno i estetski.

Izvori i literatura:

- Automat za pisanje poezije-www: *Automat za pisanje poezije*. <http://www.dw.com/hr/automat-za-pisanje-poezije/a-2282176> (15. 02. 2018.)
- Blindtextgenerator-www: *Blindtextgenerator*. <http://www.blindtextgenerator.de/> (01. 06. 2017.)
- Božićne pjesme-www: *Umjetna inteligencija piše božićne pjesme (loše)*. <https://www.bug.hr/umjetna-inteligencij/umjetna-inteligencija-pise-bozicne-pjesme-lose-2034> (10. 01. 2018.)

- Dubay, W. H. 2004. *The Principles of Readability*. Costa Mesa, CA: Impact Information.
- Generator AFS-pl-dokumentów DamDom-www: *Generator dokumentów DamDom.pl*. <https://www.damdom.pl/generator-dokumentow.html> (15. 02. 2018.)
- Generator AFS-uk-www: *Форма генерації документів*. <https://cesaris.itsway.kiev.ua/gendoc> (15. 02. 2018.)
- Generator KFS-cs-ePicka-www: *Generátor poezie ePicka*. <http://benesnet.cz/epicka/> (15. 02. 2018.)
- Generator KFS-cs-Próza-www: *Náhodný Generátor Próza APK*. <https://apkname.com/cs/cc.stever1975.androidmarkovgenerator> (15. 02. 2018.)
- Generator KFS-pl-Wiersz_tekstu-www: *Generator z dostosować wiersz tekstu i wybór tle*. <https://pl.pixiz.com/template/3353> (15. 02. 2018.)
- Generator KFS-sl-www: *Generator poezije*. <http://www.pesnik.net/content/view/8814/44/> (15. 02. 2018.)
- Generator KFS-uk-www: *Онлайн генератор рифм: подбор рифмы к словам*. <https://rymy.com.ua/опублікувати-вірш/> (15. 02. 2018.)
- Generator NFS-pl-artykuły naukowe-www: *Wygenerowane losowo „artykuły naukowe” zdarzają się częściej, niż mogłoby się wydawać*. <http://giznet.pl/wygenerowane-losowo-artykuly-naukowe-zdarzaja-sie-czesciej-niz-mogloby-sie-wydawac/> (15. 02. 2018.)
- Generator PFS-pl-wiadomości-www: *Generator Wiadomości Więziennych*. <http://areszto.wany.pl/> (15. 02. 2018.)
- Generator PFS-sk-www: *Generátor vedeckých článkov*. <https://dispersus.blog.sme.sk/c/260580/Generator-vedeckych-clankov.html> (15. 02. 2018.)
- Generator poruka-www: *Generator poruka*. <https://www.nbshop.rs/dokumentacija/poruke/generator-poruka> (10. 01. 2018.)
- Generator RFS-cs-syntetizator-www: *Česky hlasový syntetizátor*. <http://www.gb-soft.cz/podpora-tts.htm> (15. 02. 2018.)
- Generator RFS-pl-www: *Old spice generator tekstu poczty głosowej*. <https://www.wykop.pl/link/407243/old-spice-generator-tekstu-poczty-glosowej/> (15. 02. 2018.)
- Gordić-Petković, V. 2004. *Virtuelna književnost*. Beograd: Zavod za izdavanje udžbenika i nastavna sredstva.
- Jezik robota_Facebook-www: *Umjetna inteligencija razvila vlastiti jezik: Facebook hitno gasi kontroverzne botove: „Počeli su razgovarati!”*. <http://net.hr/tehnoklik/ovo-je-zabrinjavajuće-facebook-hitno-gasi-kontroverzne-botove-je-je-umjetna-inteligencija-razvila-vlastiti-jezik/> (10. 01. 2018.)
- Kafanabot-www: *Kafanabot*. <http://www.kafanabot.me/> (10. 01. 2018.)
- Komputerska poezija-www: *Kompiuterska poezija povezuje žanrove*. <http://www.upnt.org/kompiuterska-poezija-povezuje-zanrove> (10. 01. 2018.)

- Novinar robot-www: *Umjesto novinara, vijesti u Južnoj Koreji piše robot*. [https://www.express.hr/tehnou/umjesto-novinara-vijesti-u-juznoj-koreji-pise-robot-11921](https://www.express.hr/tehnou/umjesto-novinara-vijesti-u-juznoj-koreji-pise-robot) (10. 01. 2018.)
- Peović Vuković, K. 2004. *Književnost i tehnologija novih medija*. Neobjavljeni magistarski rad. Zagreb: Filozofski fakultet Sveučilišta u Zagrebu.
- Robot_žalbe-www: *Robot besplatno piše žalbe sudu*. <https://www.penzin.rs/britanija-robot-besplatno-pise-zalbe-sudu/> (10. 01. 2018.)
- Roman_nagrada-www: *Roman koji je napisao kompjuter zamalo dobio književnu nagradu*. <http://www.glif.rs/blog/roman-koji-je-napisao-kompjuter-zamalo-dobio-knjizevnu-nagradu/> (10. 01. 2018.)
- Saračević, N. 2017. *Internet-književnost: Književni tekstovi u hipermedijalnom diskursu*. Sarajevo: University Press.
- SCIgen-www: *SCIgen*. <https://ru.wikipedia.org/wiki/SCIgen> (01. 06. 2017.)
- TextRandomizer-www: *TextRandomizer*. <http://freelancers-tools.com/?p=2434> (08. 01. 2017.)
- Ušćumlić-www: Ušćumlić, Iva. *Ugrožava li hipertekstualna književnost na Internetu klasične romane u knjigama?* <http://www.monitor.hr/clanci/ugrozava-li-hipertekstualna-knjizevnost-na-internetu-klasicne-romane-u-knjigama/6249/> (10. 01. 2018.)
- Автопохвалитель-www: *Автопохвалитель*. <http://nepom.ru/usefulness/auto-praise.php> (10. 02. 2015.)
- Автопоэт-www: *Автопоэт*. <https://yandex.ru/autopoet/monorim/20/> (01. 06. 2017.)
- Балаболка-www: *Балаболка*. <http://www.cross-plus-a.ru/balabolka.html> (08. 01. 2017.)
- Генератор – цепи Маркова-www: *Генератор текста на основе цепей Маркова*. http://www.manhunter.ru/webmaster/358_generator_teksta_na_osnove_cerey_markova.html (01. 06. 2017.)
- Генератор биографий-www: *Генератор биографий*. <http://ainte.ru/talkbox/person/> (10. 02. 2015.)
- Генератор любовных писем-www: *Генератор любовных писем* [1]. http://www.libo.ru/test_libe.html. Состояние (10. 02. 2015.)
- Генератор поздравлений-www: *Составить поздравление в стихах*. <http://yubiley.org/sostavit-pozdravlenie-v-stihah.html> (03. 02. 2015.)
- Генератор смешных стихов-www: *Генератор смешных стихов*. <http://risovach.ru/kartinka/2906356> (03. 02. 2015.)
- Генератор уникальных текстов-www: *Генератор уникальных текстов для политических дорвеев*. <http://zeddy.ru/besplatnyj-generator-unikalnyh-tekstov/> (14. 03. 2017.)

- Помощник поэта-www: *Помощник поэта*. <https://www.stihi.ru/assist/> (17. 03. 2017.)
- Рандомизатор текста-www: *Рандомизатор текста*. <http://miniwebtool.ru/> (08. 01. 2017.)
- Рефераты Yandex: Яндекс-www: *Яндекс*. <http://referats.yandex.ru/referats/> (15. 02. 2015.)
- Тошович 2015 = Тошович, Бранко. 2015. *Интернет-стилистика: монография*. Москва: Флинта, Наука.
- Тошович 2017 = Тошович Бранко. 2017. *Структура интернет-стилистики: монография*. Москва: Флинта, Наука.
- Убийца поэзии-www: *Генератор текстов – убийца поэзии*. <https://yvision.kz/post/439760> (14. 03. 2017.)

WEB-BASED GENERATION OF TEXTS IN SLAVIC LANGUAGES

This paper consists of the following parts: defining the subject of analysis, interpreting the term text generator, determining the research corpus, isolating generator types, distinguishing offline and online generators. The research corpus encompasses all Slavic languages (with the most number of examples being in Russian and the least in Macedonian). Since sources for individual languages differ, it takes a long time to create a relevant multi-language corpus (the analysis is made more difficult by lack of complex and systematic research on the subject). What we have collected for a short period of time, however, gives a global idea of the contemporary state of automatic generation of texts in those languages. The text generation models have identical typology for all Slavic languages (they differ in the way and in the frequency of use), and consist of 1) standard languages, 2) generation categories, 3) the relation between the input and the output, 4) formula and scheme for automatic text production.

Keywords: *generator, Slavic languages, text, input, output, online, offline*