

ЗБОРНИК У ЧАСТ

Љиљани Суботић

**ТЕМЕ ЈЕЗИКОСЛОВНЕ У СРБИСТИЦИ
КРОЗ ДИЈАХРОНИЈУ И СИНХРОНИЈУ**

Уредници:

Јасмина Дражић
Исидора Бјелаковић
Дејан Средојевић



Нови Сад
2016

MORFOLOŠKA PRAVILA GENERISANJA KAO OSNOVA NOVOG TIPa GRAMATIKE

U radu se razmatraju morfološka pravila za generisanje oblika promjenljivih vrsta riječi i ukazuje na važnost tih pravila za izradu novog tipa gramatike koji autor naziva totalnim (jer obuhvata sve riječi, bez izuzetka, iz velike i reprezentativne leksičke građe) i korpusnim (jer se bazira na leksičkim spiskovima i tekstovima). U prvom dijelu se govori o postupku i sredstvima automatskog generisanja paradigmi, u drugom o tipovima morfoloških pravila, u trećem o glasovnim promjenama koje zahtijevaju morfološka pravila, a u četvrtom se prezentira gramatički pristup koji autor razvija.

Ključne riječi: gramatika, morfologija, pravilo, generisanje, anotiranje, paradigma, alternacija, korpus.

1. Za istraživanje gramatičke strukture riječi (morfoloških, semantičkih, sintaksičkih i stilističkih vrijednosti oblika), osmišljavanje širokih i raznorodnih načina upotrebe, a takođe radi učenja srpskog jezika kao maternjeg i stranog (izrade obrazaca za lakšu edukaciju) važno je razraditi kompleksan sistem generisanja maksimalnog i po mogućnosti potpunog broja morfoloških oblika pomoću minimalnog broja pravila. Morfološki tipovi koji se susreću u gramatikama nisu za to pogodni pa je potrebno stvoriti širi i kompleksniji sistem.

2. Postoje dvije vrste morfološkog pristupa: (1) klasični, tradicionalni, zasnovan na izdvajanju osnovnih paradigmatskih tipova, (2) morfogeneratorski, zasnovan na utvrđivanju pravila za automatsko generisanje paradigmi sa svim međusobnim odnosima, alternacijama i varijantama (dubletima, tripletima...).

Prednost klasičnog pristupa (deduktivnog: opšte → pojedinačno) jeste u tome što se sve lekseme u okviru jedne vrste riječi morfološki raščlanjuju na minimalni broj vrsta (recimo, imenice na pet vrsta, glagoli na osam vrsta i sl.), ali mu je nedostatak što se na osnovu takvog ograničenog broja tipova ne mogu automatski proizvesti paradigme svih leksema date vrste riječi. Za potrebe morfološkog obilježavanja elektronskog korpusa i edukaciju neophodno je imati minima-

lan broj morfoloških pravila. Prednost generatorskog pristupa je sveobuhvatnost i preciznost, jer se na osnovu njega mogu dobiti paradigme za sve lekseme, bez izuzetka. Morfogeneratorski pristup je zasnovan na indukciji (pojedinačno → opšte: svaka riječ se individualno obrađuje, a zatim uvrštava u određeno pravilo). Njegova je vrijednost u tome što daje sve paradigme, ali mu je nedostatak što je broj pravila znatno veći (i na stotine) u odnosu na broj tipova u klasičnoj tipologiji pa su stoga razrađuju kumulativna pravila (vrši se kategorijalno sužavanje na što manji broj pravila), koja tu slabost neutrališu. Za razliku od klasičnog (tradicionalnog) opisa ovaj se može iskoristiti za automatsko morfosintaksičko obilježavanje i pretraživanje.

3. Pravila generisanja koja smo razradili nastala su u periodu od 2008. do 2016. te provjerena tokom izrade „Gralis-MorfoGeneratora“ – analitičko-sintetičkog sistema za (1) automatsko morfosintaksičko anotiranje riječi u Gralis-Korpusu, (2) morfološku analizu srpskog, hrvatskog i bosanskog/bošnjačkog jezika, (3) automatsko generisanje paradigmi promjenljivih vrsta riječi. Morfo-Generator predstavlja online analitičko-sintetički sistem za obilježavanje, izradu paradigmi i pretraživanja po kodnim pozicijama u Gralis-Korpusu (MorphoGenerator-www). Morfološko anotiranje, koje se vrši pomoću njega, započeto je aprila 2008. u okviru projekta *Razlike između srpskog, hrvatskog i bosanskog/bošnjačkog jezika* (FWF-Projekt, P19158-G03, 2006–2010) austrijskog Fonda za podršku naučnih istraživanja, a nastavljeno na projektu *Gralis-Lexikarium*, koji je finansirala Pokrajinska vlada Štajerske (2008–2014).¹ Lingvistički sistem obilježavanja i proces generisanja razradio je autor ovih redova, a programski paket na bazi lingvističkog modela razvila Olga Lehner. Budući da naš kopus obuhvata niz jezika, bilo je neophodno izabrati kodiranje koje bi najviše odgovaralo njihovom kontrastiranju i proučavanju. Konceptiji Korpusa najbliži je bio sistem Multext-East (Multilingual Texts and Corpora for Eastern and Central European Languages – multilingual dataset for language engineering research and development: MultiText East-www), razrađen 2004. od strane grupe autora na čelu sa

¹ Pomoću MorpfoGeneratora može se dobiti svaka paradigma, što je posebno važno za istraživanja i proces obrazovanja. Međujezičke razlike u upotrebnoj frekvenciji ne određuju se paušalno, u opštim formulacijama, već objektivno pomoću korpusnih podataka. U MorpfoGeneratoru može se izabrati za pretragu pojavnica ili lema, kraj ili početak riječi, njen segment, određena vrstu riječi. MorpfoGenerator nudi takođe obrnuto sortiranje i paralelne podatke o učestalosti u tri jezika (srpskom, hrvatskom, bošnjačkom). Imenice, pridjevi i glagoli, koji se odlikuju širokim sistemom promjene, daju se u odvojenim maskama. MorpfoGenerator daje statističku informacija o svakoj vrsti riječi u Gralis-Korpusu. U MorpfoGeneratoru može se izabrati bilo koja riječ s ciljem da se dobije potpuna deklinacija, konjugacija i komparacija.

Tomažem Erjavcem budući da je bilo prilagođeno prirodi slovenskih jezika. Ono se odnosi na klasične vrsta riječi, skraćnice i tzv. rezidual (residual).²

4. Za pretraživanje u Gralis-Korpusu na osnovu gramatičke anotacije koristi se CQP-sintaksa, koja nudi široke i raznorodne mogućnosti.

5. Gramatičko kodiranje za Gralis-Korpus sastoji se od sljedećih pozicija: 1. vrsta riječi, 2. podtip vrste riječi, 3. glagolski način, 4. vrijeme, 5. lice, 6. broj, 7. rod, 8. dijateza, 9. negacija, 10. određenost, 11. refleksivnost, 12. padež, 13. kategorija živog, 14. klitika, 15. vid, 16. etikecija, 18. akcenat, 19. način vršenja glagolske radnje, 20. leksičko-semantička grupa, 21. spojivost (reakcija, kongruencija), 22. ekspresija, 23. funkcionalni stil, 24. analitički oblik, 25. destrukcija, 26. derivacija, 27. broj pravila, 28. gramatički tip generisanja oblika.

Na osnovu opšteg modela stvorena je konkretna shema sa 20 pozicija koja obuhvata sve vrste riječi i njihove kategorije. Za oznaku taksona koriste se mala slova (samo za ukazivanje na lice i pravilo daju se brojke). Pošto ima više pozicija nego slova, neke grafeme se ponavljaju, ali ne dolazi do zbrke budući da prvu poziciju uvijek zauzima vrsta riječi, što onemogućuje pojavu dvosmislenosti. To izgleda ovako:

1. VRSTA RIJEČI **n** – imenice, **v** – glagoli, **a** – pridjevi, **p** – zamjenice, **r** – prilozi, **s** – prijedlozi, **c** – veznici, **m** – brojevi, **i** – uzvici, **q** – riječce, **y** – skraćnice;

2. PODTIP VRSTE RIJEČI – I m e n i c e: **c** – zajedničke, **p** – vlastite, **m** – gradivne, **l** – zbirne. G l a g o l i: **m** – sa punim leksičkim značenjem, **a** – pomoćni, **o** – modalni, **c** – kopulativni, **b** – bazni. P r i d j e v i: **f** – opisni, **r** – odnosni, **m** – gradivni, **s** – prisvojni, **o** – redni, **m** – količinski. Z a m j e n i c e: **p** – lične, **d** – pokazne, **i** – neodređene, **s** – prisvojne, **q** – upitne, **r** – odnosne, **x** – povratne, **z** – odrične, **g** – opšte, **y** – upitno-odnosne, **j** – određene, **t** – pokazno-odnosne. P r i l o z i: **g** – opšti, **z** – odrični, **a** – pridjevski, **v** – glagolski, **q** – upitni. P r i j e d l o z i: **p** – prepozitivni, **t** – postpozitivni. V e z n i c i: **c** – naporedni, **s** – zavisni. B r o j e v i: **c** – glavni, **o** – redni, **m** – zbirni, **l** – multiplikativni, **s** – posebni. R i j e č c e: **z** – odrične, **q** – upitne, **o** – modalne, **r** – potvrđne. S k r a ć e n i c e: **n** – imenske, **r** – priloške;

3. TIP OBLIKA – G l a g o l i: **i** – indikativ, **m** – imperativ, **c** – kondicional/potencijal 1, **h** – kondicional/potencijal 2, **n** – infinitiv, **p** – particip, **g** – glagolski prilog 1 (sadašnjeg vremena), **w** – glagolski prilog 2 (prošlog vremena), **u** – supin, **t** – prelazni, **q** – citirani, **s** – hipotetički. P r i d j e v i: komparacija – **p**: pozitiv, **c** – komparativ, **s** – superlativ. P r i l o z i: komparacija – **p**: pozitiv, **c** – komparativ, **s** – superlativ, **e** – elativ;

4. VRIJEME: **p** – prezent, **i** – imperfekt, **f** – futur I Sr (srpski), **w** – futur I Hr (hrvatski), **z** – futur II Sr/Hr (srpski i hrvatski), **q** – futur II, **s** – perfekt, **l** – pluskvamperfekt 1, **t** – pluskvamperfekt 2, **a** – aorist;

² Pod time se podrazumijeva ono što je ostalo nakon nekog procesa, događaja ili neke radnje.

5. LICE: **1** – prvo, **2** – drugo, **3** – treće;
6. BROJ: **s** – jednina, **p** – množina, **d** – dual, **l** – zbirnost;
7. ROD: **m** – muški, **f** – ženski, **n** – srednji, **l** – opšti;
8. DIJATEZA: **a** – aktiv, **p** – pasiv;
9. NEGACIJA: **n** – da, **y** – ne;
10. ODREĐENOST: **n** – da, **y** – ne;
11. REFLEKSIVNOST: **n** – da, **y** – ne;
12. PADEŽ: **n** – nominativ, **g** – genitiv, **d** – dativ, **a** – akuzativ, **v** – vokativ, **i** – instrumental, **l** – lokativ;
13. KATEGORIJA ŽIVOG: – da, **y** – ne;
14. KLITIKA: **n** – da, **y** – ne;
15. VID: **p** – nesvršeni, **e** – svršeni, **b** – dvostruki;
16. ETIKECIJA: **n** – da, **y** – ne;
17. PRELAZNOST: **n** – da, **y** – ne;
18. DESTRUKCIJA: **n** – da, **y** – ne;
19. DERIVACIJA: **s** – nemotivirani, **c** – složeni;
20. BROJ PRAVILA: 01, 02, 03...

Za sada nisu popunjene sljedeće pozicije: 11. refleksivnost, 16. etikecija, 18. destrukcija, 19. derivacija. Od 28 planiranih mjesta biće još obrađeni: 18. akcenat (koje se daje u Akcentarijumu³), 19. način vršenja glagolske radnje, 20. leksičko-semantička grupa, 21. spojivost (rekcija, kongruencija), 22. ekspresija, 23. funkcionalni stil, 24. analitički oblik.

6. Morfološka anotacija za srpski, hrvatski i bošnjački jezik izvršena je za 108.297 riječi. Njihove paradigme generisane su (za sada) pomoću 822 pravila.

7. Postupak anotiranja izgleda ovako.

- a) Priprema se spisak svih leksičkih jedinica u okviru jednog standardnog jezika.
- b) Riječi se dijele na vrste riječi.
- c) U okviru njih se lekseme objedinjavaju u poseban spisak po gramatičkom obilježju važnom za generisanje paradigme (za imenice kategorija roda, za pridjeve komparacija, za glagole vid i sl.).
- d) Dobijeni spisak se raščlanjuje na podspiskove, uzimajući u obzir dopunsko gramatičko obilježje (za imenice kategorija živog, za glagole prelaznost itd.). Time se završava priprema leksičke građe za gramatičku obradu i dolazi nova faza sa sedam procedura.
- e) Vrš se analiza spiska pod (d) i utvrđuju paradigmatski markeri (padeža, lica i sl.) relevantni za generisanje paradigmi (npr. za imenice muškog roda na suglasnik nominativ, genitiv, vokativ i instrumental jednine te nominativ i genitiv množine).

³ To je program za akcentovanje riječi (<http://www-gewi.uni-graz.at/gralis-alt/php/en/Akcentarium/suche.php>). Postoji direktna veza između MorfoGeneratora i Akcentarijuma.

f) Sastavlja se spisak riječi sa opštim gramatičkim markerima (recimo za imenice muškog roda na suglasnik koje imaju u nominativu jednine nepostojano **a**).

g) Bira se tipična riječ u toj grupi i za nju izrađuje tabela sa 20 pozicija i svim oblicima. Po toj riječi i broju tipa naziva se svako pravilo (npr. 174^{mudrac}).

Time se završava lingvistički rad i počinje programski.

h) Uzima se materijal dobijen u procedurama (a) – (g) i pretvara u relacionu bazu podataka MySQL. Pri tome se izrađuju dvije tabele. Jedna se sastoji od četiri stupca u kojima se nalaze leme, njihovi nepromjenljivi dijelovi, važno gramatičko obilježje (npr. za imenice kategorija živog, za glagole prelaznost), broj lingvističkog pravila. U osnovi druge tabele nalazi se tabela sa kodovima, kodnim pozicijama i paradigmama za određeno pravilo, ali se ona razlikuje od prve novim stupcem sa nepromjenljivim segmentima riječi.

i) Ove dvije tabele pretvaraju se u format .csv, pri čemu se sadržaj stubaca razdvaja se tačkom i zapetom: šetati;še;y;45; plesati;ple;y;48; platiti;pla;y;90e. Rezultati se unoše u MySQL u kome se generišu oblici i paradigme. Iz dviju generisanih tabela (recimo, za glagole sadrži više od 3 miliona redova) dobija se rječnik/leksikon oblika. Napokon, nastaje zbirna tabela za sve vrste riječi.

j) Pripremaju se dvije maske – jedna za pretragu informacije po gramatičkoj anotaciji i druga za odražavanje rezultata pretrage.

k) MorfoGenerator se povezuje sa Gralis-Korpusom i otvaraju informacioni kanali u dva pravca – od MorfoGeneratora prema Korpusu i od Korpusa prema MorfoGeneratoru s ciljem da se bilo koji oblik iz MorfoGeneratora može naći u Korpusu kako bi se dobila informacija o njegovoj frekvenciji u Korpusu i kompletna paradigma.

8. Centralni pojam ovog gramatičkog opis je pravilo. Pod tim se podrazumijeva (a) procedura utvrđivanja morfoloških promjena i variranja, određivanja načina alterniranja i izrade njihovog opisa u obliku kratkih elementarnih uputa, (b) matrica saodnosa između kodnih pozicija i paradigmatskih punjenja, (c) postupak razotkrivanja sistema promjenljivosti riječi, njihovog variranja i prezentiranja u cilju generisanja svih oblika i cjelovitih paradigmi. Sistem pravila je dosta širok i kompleksan jer je sastavljen od više potkategorija.

9. Prije svega, postoje sintagmatska i paradigmatska pravila. Prva se tiču linijske organizacije dviju ili više riječi. Druga se odnose na mogućnost izbora jedinica za njihovo linearno ulančavanje. Za dobijanje potpune paradigme riječi u istom tipu promjene (deklinaciji, konjugaciji i/li komparaciji) važna su paradigmatska pravila. Ona mogu biti pojedinačna i opšta.

10. Pojedinačna pravila se odnose na odvojene gramatičke kategorije i potkategorije i ukazuju na specifičnost promjene u okviru određene gramatičke kategorije, recimo roda. Tako, prilikom generisanja oblika imenica muškog roda

na nulti nastavak treba uzimati u obzir četiri padeža jednine (nominativ, genitiv, vokativ i instrumental) i dva padeža množine (nominativ i genitiv).

11. Opšta pravila tiču se triju tipova promjene: deklinacije, konjugacije i komparacije. Npr. u deklinaciji pravila se baziraju na padežima bitnim za dobijanje oblika i paradigme nastaju na bazi pet od sedam padeža jednine (nominativa, genitiva, akuzativa, vokativa i instrumentala) i triju od sedam padeža množine (nominativa, genitiva i dativa).

Opšta pravila se dijele na strukturna, kategorijalna i interkategorijalna. Strukturalna pravila ukazuju na to kakva se formalna sredstva koriste za generisanje paradigmi i do kakvih alternacija dolazi na spoju nastavka i prethodne morfeme. Ovdje npr. spada pravilo vezano za broj nastavaka. Tako, u sistemu imenica svi se oblici tvore pomoću 11 nastavaka te uzimanjem u obzir supletivnih oblika (*čovjek – ljudi*), proširene osnove (*vuk – vukovi*) i postakcenatske dužine (up.: gen. jedn. i mn. *Ovo je žena – Nema ženā*). Kategorijalna pravila se odnose na parametre bitne za generisanje morfoloških oblika. Intrakategorijalna pravila obuhvataju dvije ili više kategorija.

12. Kategorijalna pravila predstavljaju procedure generisanja oblika u okviru određene potkategorije (gramatičkog lica, vida, glagolskog roda i sl.). Tako se paradigme imenica generišu na osnovu roda, broja i kategorije život/neživotg.

13. Interkategorijalna pravila određuju paradigmatsku relevantnost saodnosa morfoloških i gramatičkih kategorija. Jedno od takvih pravila ukazuje na to da se generisanje oblika akuzativa za imenice m. r. na nulti nastavak vrši na bazi obilježja živo/neživo pošto imenice za neživo dobijaju oblik nom. jedn. (vidim *predmet*), a imenice za živo oblik gen. mn. (vidim *brata*).

14. Posebnu grupu čine (a) ekskluzivna pravila (koja pokrivaju riječi na određenu završnu fonemu) i inkluzivna pravila (koja pokrivaju imenice sa različitim finalnim fonemama), (b) fleksijska pravila (čiji su markeri nastavci), (c) vertikalna pravila, koja mogu biti padežna (osnovu čine padeži), komparacijska (osnovu čine stepeni poređenja), konjugacijska (osnovu čine lica), (d) kumulativna pravila (više pravila se konusiraju i objedinjavaju) i (e) izdvojena pravila (sa osobenom strukturom).

15. Odnos ekskluzivnih i inkluzivnih pravila varira u svakoj vrsti riječi i njihovom tipu. Npr. kod imenica muškog roda za neživo na nulti nastavak izrazito prevladavaju prva – 50 : 14. Ekskluzivna pravila tiču se sljedećih završnih

suglasnika: 22 za imenice na **-c** [ts]) – 45^{stric}, 52^{zec}, 174^{mudrac}, 350^{štruc}, 46^{kukac}, 54<sup>vra-
bac</sup>, 43^{otac}, 49^{svetac}, 50^{sudac}, 60^{bratac}, 73^{očevdac}, 159^{škrtac}, 197^{gačac}, 51^{čitalac}, 324^{lažac}, 322^{pisac},
323^{knjigovezac}, 345^{Leskovac}, 346^{Štrbac}, 347^{Beogradac}, 348^{Dvizac}, 349^{Branežac}, 11 za imenice na
-k – 180^{bak}, 184^{vuk}, 178^{seljak}, 05^{čovjek}, 351^{Dodik}, 179^{predak}, 181^{patuljak}, 182^{mačak}, 183^{cvrčak},
296^{patak}, 291^{sinak}, 5 za imenice na **-r** – 39^{ribar}, 299^{krčmar}, 328^{car}, 186^{tigar}, 297^{dabar}, 4 za
imenice na **-n** – 185^{Srbin}, 246^{Turčin}, 325^{gospodin}, 191^{ovan}, 2 za imenice na **-g** – 176^{strateg},
329^{bog}, 2 za imenice na **-d** – 175^{gospod}, 2 za imenice na **-t** – 188^{brat}, 326^{gost}, 1 za
imenice na **-š** [ʃ] – 177^{siromah}. 1 za imenice na **-z** – 250^{vitez}. Inkluzivnih pra-
vila ima 14: 65^{turist}, 67^{mladić}, 158^{zer}, 69^{galeb}, 18^{kralj}, 63^{duh}, 320^{morž}, 327^{derviš}, 280^{Vasiljev},
344^{Malik}, 187^{pas}, 189^{asistent}, 352^{Bosanac}, 353^{Konjarek}. Date imenice ukazuju na kvanti-
tativnu disharmoniju: inkluzivna pravila obuhvataju mnogo više imenice nego
inkluzivna – najviše 67^{mladić} (5.032), zbog velikog broja prezimena na **-ć**, i 65^{turist},
(1.714), dok se među ekskluzivnim pravilima izdvaja 345^{Leskovac} (553), jer pokriva
značajan dio antroponima, i 178^{seljak} (420).

16. Što se tiče fleksijskih pravila, ona su specifična za svaku vrstu riječi. Na primer, u sistemu imenica mogu se, prije svega, izdvojiti tri takva pravila. Prvo fleksijsko pravilo ima opšti karakter: svaki padež, svaka paradigmatska jedinica ima neki nastavak. Drugo fleksijsko pravilo glasi: svaka grupa imenica (muškog roda za neživo, muškog roda za živo, srednjeg roda, ženskog roda, opšteg roda, pluralia tantum) posjeduje svoj sistem nastavaka. Tako se fleksijski sistem imenica muškog roda za živo na nulti nastavak sastoji od osam nastavaka: **-ø**, **-a**, **-e**, **-u**, **-om**, **-em**, **-i**, **-ima**. Treće fleksijsko pravilo se odnosi na jednu grupu – na imenica za živo sa nultim nastavkom i glasi: nijedan se nastavak ne veže isključivo za jedan padež, tj. postoji fleksijski sinkretizam (nastavci ulaze u sastav najmanje dvaju padeža, s tim što **-u** dolazi u tri). Nastavci ovako koreliraju sa padežima: nulti nastavak (**-ø**) pojavljuje se u nom. i vok. jedn. (nom. *turist*, nom./vok. *Konjalek*), **-a** dolazi u gen. jedn., akuz. jedn. i gen. mn., nom. mn. (*turista*, *gospoda*), **-e** veže se za vok. jedn. i akuz. mn. (*turiste*), **-u** se nalazi u dat. jedn. (*turistu*), vok. i lok. jedn. (*mladiću*), **-om** i **-em** služe za obrazovanje inst. jedn. (*turistom*, *mladićem*), **-i** se susreće u nom. i vok. mn. (*turisti*), **-ima** je fiksiran za dat., instr. i lok. mn. (*turistima*).

Vertikalna pravila se odnose na padeže, lica i stepene poređenja pa se razlikuju padežna, konjugacijska i komparacijska pravila. Recimo, padežna pravila ukazuju na to kakva je struktura nastavaka u pojedinim padežima. Ilustracije radi, postoje četiri pravila za imenice muškog roda za živo. Prvo pravilo isključuje izuzetke: svaki padež ima najmanje jedan nastavak. Drugo pravilo je kvantitativno determinisano: postoje padeži sa jednim nastavkom, sa dva nastavka i sa tri nastavka. U grupi sa jednim nastavkom nalazi se nom. jedn. (nulti nastavak **-ø**), gen. i akuz. jedn. (**-a**), dat. i lok. jedn. (**-u**), akuz. mn.

(-e), dat., instr. i lok. mn. (-ima). Nekoliko padeža imaju dva nastavka: vok. jedn. -e, -u (turiste, mladiću), instr. jedn. -om, -em (turistom, mladićem), nom. i vok- jedn. mn. -i, -a (turisti, gospoda). Gen. mn. dolazi sa tri nastavka: -a, -i, -iju (turista, gosti, gostiju). Treće pravilo je sinkretičko: u jednini se podudara genitiv sa akuzativom, a u množini dativ, instrumental i lokativ. Četvrto pravilo se odnosi na jediničnost: postoji bar jedno precizirajuće pravilo u jednini (gen. akuz., vok. i instr.) i množini (nom. i gen.).

17. U generisanje morfoloških oblika zapažaju se krajnji slučajevi: jedna pravila se koriste za obrazovanje minimalnog broja likova (jedne, dvije ili tri riječi, npr. pravilo 80^{veoce} pokriva 12 imenica sr. r. na -e), druga za generisanje maksimalnog broja (nekoliko hiljada, kao što je pravilo 03^{strana}, pomoću koga se obrazuje paradigma za 10.532 imenice ž. r. na -a). Polarne pozicije zauzimaju pravila za izdvajanje paradigmi bez bilo kakvih alternacija samoglasnika i/li suglasnika kao i pravila za lekseme sa maksimalnom morfonološkim variranjem. Ponekad jedna jedina riječ traži posebno pravilo, npr. *veo* (pravilo 310^{eeo}).

18. Potreba za izradu velikog broja pravila, posebno za imenice i glagole, rezultat je raznovrsnih alternacija. One se mogu svesti na dvije osnovne – fonetsko-fonološke (u prvom redu nepostojano *a*, palatalizacija, sibilizacija i jotovanje) i prozodijske (prvenstveno postakcenatske dužine u gen. mn. nekih imenica).

19. Globalno postoje dvije vrste glasovnih promjena: 1) lingvalne, koje dolaze (a) u okviru lekata (varijanata) jednog jezika (npr. ekavskoj i ijekavskoj varijanti srpskog jezika tipa *htio* – *hteo*) ili (b) u korelaciji dvaju ili više jezika (npr. srpskog i drugih slovenskih i neslovenskih jezika), 2) lingvostrukturne, koje se javljaju u okviru paradigmi određenog jezika, npr. u glagolskom sistemu srpskog jezika.

20. Lingvalne alternacije (skraćeno α -alternacije) nastaju na nivou varijanata jednog jezika ili u korelacijskoj mreži dvaju i više jezika te mogu biti a) α -intrakorelacione – u okviru lekata istoga jezika: idiolekata, sociolekata (žargona), spaciolekata (dijalekata), funkciolekata (funkcionalnih stilova), npr. srpskog jezika, b) α -interkorelacione – u okviru genetski i tipoloških veoma bliskih jezika sa istom dijalekatskom bazom, npr. futur I u srpskom i hrvatskom (*čitaću* – *čitat ću*), c) α -suprakorelacione – u okviru teritorijalno bliskih jezika iste zajednice različite dijalektske osnovice, npr. južnoslovenskih (npr. srpskog i makedonskog), d) α -superkorelacione – u okviru teritorijalno razdvojenih jezika iste zajednice, npr. južnoslovenskih i istočnoslovenskih (recimo srpskog i ruskog), e) α -ekstrakorelacione – u okviru genetski različitih jezika (npr. srpskog i njemačkog).

21. Lingvostrukturne alternacije (skraćeno β -alternacije) imaju za ishodište paradigmu određenog jezika i mogu biti trostruke prirode.

A. Po tome da li dolaze u istom ili različitom obliku razlikujemo: 1) morfološki homogene alternacije – glasovne promjene u istom morfološkom liku (npr. prezentu), 2) morfološki heterogene alternacije – glasovne promjene u različitim morfološkim likovima (npr. infinitivno-prezentske: *moći* – *mogu*).

B. Paradigmatski alternacija može biti kompletna i parcijalna. Kompletna zahvata čitavu paradigmu (npr. prezent *naći* – *nađem*, *nađeš*, *nađe*; *nađemo*, *nađete*, *nađu*), a djelimična samo jedan njen segment (npr. jedninu ili množinu) ili elemenat (npr. pojedina lica ili padeže), recimo 2. i 3. licu jednine i 1. i 2. licu množine: *mogu*, *možeš*, *može*; *možemo*, *možete*, *mogu*.

22. Postoji monoalternacija, bialternacija i polialternacija. Monoalternacija dolazi u slučajevima pojave samo jednog oblika sa glasovnim variranjem (recimo od *biti* je *bijem* u 1. licu jednine prezenta, *biješ* u 2. licu jednine, *bije* u 3. licu jednine itd.). To je, dakle, glasovna promjena bez druge morfološke alternative. Bialternacija je dubletna, dvostruka alternacija i javlja se kada postoje dva različita gramatička lika u istoj paradigmatskoj jedinici (padežu, licu, stepenu komparacije), od kojih oba mogu biti neutralna ili je jedan od njih stilski obilježen (kao razgovorni, regionalni, dijalekatski, žargonski...). Tako od glagola *moći* imamo dublete u 1. licu jednine sadašnjeg vremena: *mogu*, *možeš*. Polialternacija podrazumijeva pojavu trostruke ili višestruke alternacije u istoj paradigmatskoj jedinici. Ona može biti, ter-, kvatro-, kvinto-... alternacija. Teralternacija (trostruko, tripletno variranje) nastaje u slučajevima postojanja triju varijanta istog oblika u istoj paradigmatskoj jedinici. Npr. u 1. licu jednine sadašnjeg vremena od glagola *iskati* imamo *ištem*, *iščem*, *iskam*. Za kvatroalternaciju (četvorostruko variranje) nemamo primjere, ali postoje slučajevi kada u istoj paradigmatskoj jedinici postoje četiri varijantska oblika sa dvije alternacije, recimo u 1. licu prezenta glagola *ht(j)eti* sa odričnim i potvrdnim oblicima: *hoću*, *ću*, *neću*, *ht(j)ednem*. Kvintoalternaciju (petorostruku alternaciju) isto tako ne nudi naš korpus. U istom paradigmatskoj jedinici može da se pojavi čak pet varijanata istog glagolskog oblika. Takav je slučaj sa pomoćnim glagolom *biti/jesam* u 1. licu jedn. sad. vr.: *jesam*, *sam*, *nisam*, *nijesam*, *budem*. Ovdje bi spadao i glagol *ht(j)eti* ukoliko bismo u jednu paradigmu objedinili njegove (a) odrične i neodrične, (b) ekavske i ijekavske likove: *hoću*, *ću*, *neću*, *htednem*, *htjednem*. Nisu nam poznati slučajevi formalne raznovrsnosti veće (od pet) u okviru iste paradigmatske jedinice.

23. U sistemu promjenljivih vrsta riječi, postoje tri vrste paradigmi: deklinacija, konjugacija i komparacija. U deklinacijske vrste riječi (sa sistemom pade-

ža) spadaju imenice, zamjenice, pridjevi i dio glagola (participi). Konjugacijsku vrstu (sa oblicima za lice) čine samo glagoli. U komparacijske vrste riječi (sa stepenima poređenja) dolaze pridjevi i prilozi. Dakle šest vrsta riječi pripadaju paradigmatskim vrstama riječi (imenice, zamjenice, pridjevi, brojevfi, glagoli, prilozi), a četiri neparadigmatskim: prijedlozi, veznici, partikule i uzvici.

24. Što se tiče kvantitativnih vrijednosti vrsta riječi, imenice čine jednu trećinu (34,43%), slijede pridjevi (32,34%) i glagoli (29,89%). Ove tri vrste riječi obuhvataju 92,48% ukupnog leksičkog fonda. Najviše pravila traže glagoli (378), zatim imenice (353), a mnogo manje pridjevi (71) i zamjenice (50), dok minimalan broj zahtijevaju brojevi (12). Međutim, ako se ta pravila dovedu u vezu sa brojem riječi na koje se odnose, dobija se sasvim druga slika: za 135 zamjenica potrebno je imati 50 pravila, za 198 brojeva 12, dok za 37.606 imenica 353, za 30.030 glagola 378, za 32.492 pridjeva 71. Ako se dovedu u vezu 353 pravila za 37.606 lema⁴, dobiće se koeficijent 0,0094, čime imenice zauzimaju četvrtu poziciju – iza zamjenica (0,370), brojeva (0,061), glagola (0,013), a jedino iza njih se nalaze pridjevi (0,0022). To znači da imenice spadaju u grupu riječi koja u odnosu na ukupan broj leksema zahtijeva najmanji broj pravila. Vrste riječi koje obrazuju zatvoreni skup (koji se ne popunjava novim riječima, u datom slučaju zamjenice i brojevi)⁵ zahtijevaju veći broj pravila nego one koje čine otvoreni skup (imenice, pridjevi i glagoli).⁶

25. Ovaj gramatički opis karakterišu sljedeće crte. Prvo, u njemu nema izuzetaka niti se tim pojmom bilo gdje operiše, budući da se u datom korelativnom pristupu oni ne priznaju: svaka jedinica se tretira potpuno ravnopravno (unikatne jedinice se tumače kao obični elementi morfološkog sistema). Drugo, dati pristup predstavlja oblik totalne gramatike (jer se prvo stvara konkretan zatvoreni leksički i tekstualni korpus, potpuno relevantan za utvrđivanje zakonitosti na nivou čitavog jezika, a onda se svaka jedinica iz njega gramatički opisuje i statistički obrađuje; pri tome se daje morfološka struktura riječi bez obzira na njihovu upotrebnu vrijednost, stilsku markaciju i leksičko-semantičku kategoriju), korpusne gramatike (u osnovi leži korpus od oko stotinu hiljada riječi). Treće, vrši se jasno

⁴ Lema je osnovni, polazni oblik u paradigmi, oblik koji se navodi u rječnicima kao natuknica. Sve ostale paradigmatske jedinice spadaju u pojavnice (token).

⁵ Pomenuti koeficijent kod zamjenica iznosi čak 37,04, a kod brojeva 6,06.

⁶ Druge vrste riječi (7.495) ovako su predstavljene: prilozi 6.901, skraćenice 243, prijedlozi 197, riječice 104, veznici 50. Odnos dviju grupa vrsta riječi (ukupno 108.297) ima sljedeće vrijednosti: imenice, pridjevi, zamjenice, brojevi, glagoli – 100.802, prilozi, skraćenice, prijedlozi, riječice, veznici – 7.495.

terminološko razgraničenje pa se recimo, strogo razlikuje nastavak (ono što se mijenja u paradigmi na kraju riječi) od završetka (koji može biti sastavljen od jedne foneme ili više fonema te obuhvatati nastavak, sufiks i postfiks). Četvrto, dati pristup daje podroban tvorbeni opis i pregled krajnjih pozicija svih promjenljivih vrsta riječi (sa nastavcima i završecima). Peto, bitan dio pristupa je valorizacija pravila na bazi tekstualnih korpusa: ogroman broj riječi i veliki broj pravila zahtijeva da se ona provjere u konkretnim tekstovima. Valorizaciju posebno traže morfološke varijante (dubleti i tripleti), kako bi se utvrdilo šta preovladava, kakva je stilska vrijednost i kakve postoje tendencije.

IZVORI

- Anić, V. (2000). *Rječnik hrvatskoga jezika*. Zagreb: Novi Liber.
- Gralis-Korpus. <http://www-gewi.uni-graz.at/gralis/korpusarium/gralis_korpus.html> 15. 10. 2015.
- Gralis-Lexikarium. <<http://www-gewi.uni-graz.at/gralis-alt/php/en/Lexikarium/index.html>> 15. 10. 2015.
- Gralis-Morphogenerator. <<http://www-gewi.uni-graz.at/gralis-alt/0.Projektarium/MorphoGenerator/morpho.php>> 15. 10. 2015.
- Hrvatski enciklopedijski rječnik (2002). Zagreb: Novi Liber.

CITIRANA LITERATURA

- Babić, S., D. Brozović, M. Moguš; S. Pavešić, I. Škarić, S. Težak, (1991). *Povijesni pregled, glasovi i oblici hrvatskoga književnog jezika*. Nacrt za gramatiku. Zagreb: Hrvatska akademija znanosti i umjetnosti.
- Silić, J., I. Pranjković (2005). *Gramatika hrvatskoga jezika za gimnazije i visoka učilišta*. Zagreb: Školska knjiga.
- Tošović, B. (2010) . *Korrelative Grammatik des Bosni(aki)schen, Kroatischen und Serbischen. Teil I: Phonetik – Phonologie – Prosodie*. Münster: LIT.
- Zečević, V., M. Peti, S. Pavešić, D. Malić, M. Lončanić, E. Barić; M. Znika, (1995). *Hrvatska gramatika*. Zagreb: Školska knjiga.

*

- Николић, М. (2000). *Обратни речник српског језика*. Нови Сад, Београд: Матица српска – Институт за српски језик САНУ.
- Николић, М. (ред. и ур.) (2007). *Речник српског језика*. Нови Сад: Матица српска.
- Пипер, П., Х. Клајн (2014). *Нормативна граматика српског језика*. Нови Сад: Матица српска.

- Речник МС/МХ: *Речник српскохрватскога књижевног језика*. Књ. I–VI (1969). Нови Сад – Загреб: Матица српска – Матица хрватска.
- РСАНУ: *Речник српскохрватског књижевног и народног језика*: I–XIX (1959–2014). Београд: Институт за српски језик.
- Станојчић, Ж., Љ. Поповић (2002⁸). *Граматика српскога језика*: Уџбеник за I, II, III и IV разред средње школе. Београд: Завод за уџбенике и наставна средства, Завод за издавање уџбеник.
- Стевановић, М. (1970²). *Савремени српскохрватски језик (Граматички системи и књижевнојезичка норма)*: I – Увод, Фонетика, Морфологија. Београд: Научна књига.
- Тошовић Б. (2011). *Коррелационна граматика сербског, хрватског и бошњачког језиков: Часть I: Фонетика – Фонология – Просодия*. Москва: Языки славянской культуры.

Branko Đ. Tošović

MORPHOLOGICAL RULES OF GENERATING AS THE BASIS FOR A NEW TYPE OF GRAMMAR

S u m m a r y

This paper has discussed morphological rules for generating varying parts of speech and has pointed out the importance of these rules for creating a new type of grammar which the author calls total (because it includes all words, without exception, from large and representative lexical material) and corpus (because it is based on corpus material composed of words and texts). The first part has dealt with the procedure and means of automatically generating paradigm, the second part with types of morphological rules, the third with sound changes that require morphological rules, the fourth with the presented grammatical description that the author has developed.

Keywords: Grammatik, Morphology, Rules, Generation, Annotation, Paradigm, Alternation, Corpus.