

УНИВЕРЗИТЕТ У БЕОГРАДУ
ФИЛОЛОШКИ ФАКУЛТЕТ

**У ПРОСТОРУ
ЛИНГВИСТИЧКЕ
СЛАВИСТИКЕ**

Зборник научних радова

Поводом 65 година живота академика Предрага Пипера

Уредници:

Људмила Поповић

Дојчил Војводић

Мотоки Нوماћи

Београд, 2015

BRANKO TOŠOVIĆ
Institut za slavistiku
Univerzitet „Karl Franc“ u Gracu

GENERISANJE MORFOLOŠKIH OBLIKA I PARADIGMI IMENICA MUŠKOG RODA ZA ŽIVO U SRPSKOM JEZIKU

Rad se sastoji iz tri dijela. U prvom se govori o razlozima i potrebi automatskog generisanja morfoloških oblika i paradigmi u istraživačke i edukacijske svrhe, ističu pravila pomoću kojih se to vrši, ukazuje na specifičnost promjenljivih vrsta riječi. U drugom se kratko predstavlja model generisanja oblika i paradigmi u imenskom sistemu. U trećem se prezentira struktura pravila za imenice muškog roda koje označavaju živa bića.

Ključne riječi: generisanje, vrste riječi, imenice, imenice za živo, muški rod, pravilo.

Za istraživanje morfološke strukture riječi, osmišljavanje širokih i raznorodnih načina upotrebe, a takođe radi učenja jezika kao maternjeg i stranog 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 u cilju istraživanja morfoloških, semantičkih, sintaksičkih i stilističkih vrijednosti gramatičkih oblika, izrade obrazaca za lakše tumačenje i učenje jezika, pretraživanja po gramatičkim parametrima u Gralis-korpusu, višjezičkom paralelnom korpusu za proučavanje slovenskih jezika utemeljenom 2007. na Univerzitetu „Karl Franc“ u Gracu (Gralis-Korpus-www).

Pravila generisanja nastala su u periodu od 2008. do 2015. i provjerena tokom izrade „Gralis-MorfoGeneratora“ – programa za (1) automatsko morfosintaksičko anotiranje riječi, (2) morfološku analizu srpskog, hrvatskog i bošnjačkog jezika, (3) automatsko generisanje paradigmi promjenljivih vrsta riječi. MorfoGenerator predstavlja online analitičko-sintetički sistem za obilježavanje oblika, generisanje paradigmi i pretraživanja po kodnim pozicijama. Gramatička, leksikografska, ortografska i ortoepska informacija se objedinjuje u online rječniku Gralis-Leksikarium (Gralis-Lexikarium).

Kao što je poznato, promjenljive vrste riječi imaju tri vrste paradigmi: deklinaciju, konjugaciju i komparaciju. U deklinacijske vrste riječi (sa sistemom padeža) 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, brojevi, glagoli, prilozi), a četiri neparadigmatskim (prijedlozi, veznici, partikule i uzvici). Najviše pravila za generisanje oblika traže glagoli (378), zatim imenice (343), a mnogo manje pridjevi (71) i zamjenice (50), dok je minimalan broj neophodan za brojeve (12). Međutim, ako se ta pravila dovedu u vezu sa brojem riječi na koje se odnose, dobija se druga slika: 135 zamjenica traže 50 pravila, 198 brojeva 12, 37.606 imenica 343, 30.030 glagola 378, 32.492 pridjeva 71. Vrste riječi koje obrazuju zatvoreni skup (koji se ne popunjava novim jedinicama, u datom slučaju zamjenice i brojevi) imaju manji broj pravila nego one koje čine otvoreni skup (imenice, pridjevi i glagoli). Kvantitativni odnos promjenljivih i nepromjenljivih vrsta riječi u našem korpusu (ukupno 108.297 leksema) izgleda ovako: imenice, pridjevi, zamjenice, brojevi, glagoli (100.802), prilozi, skraćenice, prijedlozi, riječce, veznici (7.495). Imenice čine jednu trećinu (34,43%), slijede pridjevi (32,34%) i glagoli (29,89%).¹ Među 108.802 leksema za koje je dobijena paradigma pomoću MorfoGeneratora nalazi se 37.606 imenica. Ova vrsta riječi nudi bogat sistem oblika, odlikuje se kategorijom roda (muški, ženski, srednji, zajednički), broja (jednina i množina), padeža (sedam), ima u akuzativu jednine m. r. kategoriju živog/neživog, izražava gramatičko značenje nizom nastavaka pretežno sinkretičnog tipa i posjeduje raznorodne morfonološke alternacije – samoglasničke (refleks jata, prijevoj, preglas, vokalizaciju, ispadanje samoglasnika, fakultativno **a**, **e**, **u**) i suglasničke (jednačenje po zvučnosti, mjestu i načinu obrazovanja, palatalizaciju, sibilizaciju, jotovanje, eliziju suglasnika), a takođe sadrži prozodijske specifičnosti (promjenu u mjestu i vrsti akcenta, postakcenat-ske dužine i sl.).

U obrazovanju imenskih oblika učestvuje 11 nastavaka. To su: **-a**, **-o**, **-e**, **-i**, **-u**, **-ø**, **-oj**, **-em**, **-om**, **-ama**, **-ima**. Međutim, oni se ne vežu isključivo za jedan padež, već se pojavljuju u više dijelova paradigme. Najveći sinkretizam ispoljava **-i** (11 padeža) i **-a** (8), a najmanji **-ø** i **-oj** (dva padeža). Za generisanje 37.606 imenica (njenih oblika i paradigmi) potrebno je imati 308 pravila.

¹ Ove tri vrste riječi obuhvataju 92,48% ukupnog leksičkog fonda.

Ta su pravila prilično složena jer se mora uzimati u obzir raznovrsnost nastavaka, široki sistem morfoloških oblika i interakcija gramatičkih kategorija. U srpskom jeziku gotovo da nema imenica sa nultom deklinacijom, odnosno bez promjene po padežima pa stoga one ne zahtijevaju nikakva pravila. Takve su: *deci* 'decilitar', *buržoa*, žensko ime *Miki*, okamenjeni glagolski oblik *vjerujem* u funkciji imenice. Složenost generisanja imenskih oblika i paradigmi sastoji se u tome što se u nizu slučajeva pojavljuju dubleti kako u jednini (zbog čega je, recimo, potrebno pravilo 225^{zaselak}: gen. jedn. *zaselak* – *zaselka/zaseoka*), tako i u množini (npr. 01^{put}: nom. mn. *puti/putevi/putovi*).² Ponekad postoje trostruki oblici pa su neophodna pravila tipa 230^{lisac} – *lisac*: nom. mn. *lisci/liskovi/lisaci*. Posebnu teškoću izazivaju različita rješenja u srpskoj, hrvatskoj i bošnjačkoj standardizaciji. Ona ponekad dovode do dubletnosti pa nastaje „dvostruka dvostrukost“ (varijativnost u okviru jednog jezika i u okviru dvaju/triju jezika). Pravila za generisanje imenskih oblika razvrstavaju se prema gramatičkom rodu te posebna pravila imaju imenice (1) muškog roda, (2) ženskog roda, (3) srednjeg roda, (4) zajedničkog roda i (5) pluralila tantum.

Među 9.034 imenica ženskog roda na **-a** najviše ih dolazi sa završetkom **-ca**: 1995 (22,08%), **-ja**: 1.664 (18,42%), **-ka**: 1.617 (17,90%) i **-na**: 1.218 (13,48%). Imenice na **-a** čine 71,88% (6.494) svih imenica. Manje pravila (6) traže imenice ženskog roda na **-e**, **-i/-er**, **-o/-ol**, **-ao**. Imenice srednjeg roda zahtijevaju 35 pravila. Najviše ih je potrebno za imenice na **-e** (22), slijede imenice na **-o** (11). Samo jedno pravilo se tiče imenica na **-a** i na **-m**. Za generisanje imenica sa dva roda (muški i ženski, muški i srednji, ženski i srednji) i tri roda (muški, ženski, srednji) neophodno je imati 16 pravila, među kojima preovladavaju ona koja se odnose na imenice muškog i ženskog roda (12). Imenice pluralia tantum pokriva 14 pravila. One mogu biti muškog, ženskog, srednjeg roda i dvostrukog roda. Nepromjenljive imenice imaju samo jedno pravilo.

Imenice muškog roda (7.696) zahtijevaju više pravila nego imenice ženskog i srednjeg roda: od 308 pravila na muški rod otpadaju dvije trećine – 193 (62,66%), pri čemu dominiraju pravila za imenice muškog roda na nulti nastavak (141).

1. Kad je u pitanju kategorija živog i neživog, najviše pravila se odnosi na imenice za živo sa pokretnim **a** (49,25%)³, a dva puta manje na imenice za

² Leksema *put* je muškog roda ('staza') i ženskog roda ('boja kože').

³ On se obično pojavljuje u nom. jedn. i gen. mn.

neživo sa istim samoglasnikom (23,88%). Minimalan broj pravila nalazimo u sistemu imenica bez pokretnog samoglasnika **a** – 14,93%: za neživo 5,97%, za živo i neživo 5,97%. Po broju pravila izdvajaju se ona koja pokrivaju imenice za neživo (92 pravila): mnogo manje ima imenica za živo (43), a još manje za živo/neživo (6).

2. Postoje dvije osnovne grupe imenica muškog roda na suglasnik – sa nul-tim nastavkom (-**ø**) i sa nastavkom **-a** u nominativu jednine. Za imenice muškog roda na nulti nastavak (5.410 leksema) treba imati pravila za imenice koje označavaju živo (1.678), neživo (3.732), živo i neživo (25), dok takva diferencijacija nije potrebna za imenice na samoglasnik (2.261). Pri tome je neophodno znati najmanje četiri padeža jednine (nom., gen., vok. i instr.) i dva padeža množine (nom. i gen.), jer u njima dolazi do samoglasničkih i/li suglasničkih alternacija i/ili se pojavljuju dubleti.

3. Postoje tri grupe imenica muškog roda za neživo: a) imenice bez pokretnog samoglasnika **a**, b) imenice sa pokretnim samoglasnikom **a** počev od gen. jedn., b) imenice sa pokretnim samoglasnikom **a** samo u gen. mn., c) imenice bez pokretnog samoglasnika **a** i sa njim (mješoviti tip). Ove imenice (ukupno 3.732 imenica) traže 102 pravila, od toga se 56 odnosi na imenice bez pokretnog vokala **a**, 40 sa njim, a 6 pokriva imenice mješovitog tipa. Pravila za imenice bez pokretnog samoglasnika **a** obuhvataju mnogo više riječi nego pravila sa pokretnim **a**: odnos je 3.152 : 1.678. Samo deset riječi se nalazi u domenu pravila za imenice sa i bez **a**. Imenice bez pokretnog samoglasnika obrazuju dvije grupe – homogeni (u kojoj se imenice završavaju na isti suglasnik) i heterogeni (u kojoj se imenice završavaju na različit suglasnik).

4. Za obrazovanje paradigmi imenica muškog roda za živo (1.678)⁴ potrebna su 52 pravila. Ona uzimaju u obzir činjenicu da se u jednini podudara akuzativ sa genitivom. Za generisanje oblika važni su nom., gen., vok. i instr. jedn. te nom. i gen. mn. Zbog složenosti morfološke strukture (varijantnosti i širine alternacija) izdvajaju se imenice na **-c** [ts], **-r** i **-k**.

5. U pojedinim padežima postoji niz nastavaka koja traže posebna pravila. Najviše fleksija (ispred kojih dolazi alternacija **-ci**, **-zi**, **-si** ili je nema) nalazimo u nom. i gen. mn. (4).

⁴ U ovu brojku nisu uključena sva prezimena na **-ić**, **ić**, **-ov**, **-ev**, **-in**.

a) Nom. mn.: **-i** (*turist – turisti, mladić – mladići, ribar – ribari, mudrac – mudraci, Rac – ‘pogrđan mađarski naziv za Srbina’, luc – ‘riba’, seljak – seljaci, strateg – stratezi, siromah – siromasi, krmčar – krmčari, akušer – akušeri*), **-ovi** (*zer – zerovi*), **-i/-ovi** (*galeb – galebi/galebovi, bak – baci/bakovi, vuk – vuci/vukovi, duh – dusi/duhovi, stric – strici – stričevi, tigar – tigri/tigrovi, ovan – ovni/ovnovi, patak – paci/patkovi, dabar – dabri/dabrovi, mačak – maći – mačkovi*), **-i/-evi** (*kralj – kralji/kraljevi, stric – strici/stričevi, otac – oci/očevi, svetac – sveći/svečevi*), **-i/-ovi/-evi** (*zec – zeci/zecovi/zečevi, vitez – vitezi/vitezovi/viteževi*). Postoji ortografski dubleti (*sudac – suci/sudci*)⁵. Kod nekih imenica u množini se skraćuje završetak: *Srbin – Srbi*. Tu je i pluralia tantum (*čovjek – ljudi*).

b) Gen. mn. **-a** (*turist – turista, mladić – mladića, ribar – ribara, zer – zero-va, mudrac – mudraca, seljak – seljaka, strateg – stratega, siromah – siromaha, krčmar – krčmara, kukac – kukaca, vrabac – vrabaca, sudac – sudaca, bratac – brataca, očevdac – očevdaca, škrtac – škrtaca, gačac – gačaca, čitalac – čitaoca, patuljak – patuljaka, cvrčak – cvrčaka, pas – pasa, asistent – asistent/asistenata*), **-a/-ova** (*galeb – galeba/galebova, bak – baka/bakova, vuk – vuka/vukova, mačak – mačaka/mačkova, tigar – tigra/tigrova, ovan – ovana/ovnova, patak – pataka/patkova, dabar – dabara/dabrova*), **-a/-eva** (*stric – strica/stričeva, otac – oceva/očeva, svetac – svetaca/svečeva*), **-a/-ova/-eva** (*zec – zeca/zecova/zečeva, vitez – viteza/vitezova/viteževa*).

c) Manje ima vok. jedn. (3): **-e** sa prvom palatalizacijom ili bez nje (28 pravila: *turist – turiste, ribar – ribare, zer – zere, galeb – galebe, stric – striče, zec – zeče, mudrac – mudrače, čovjek – čovječe, bak – bače, vuk – vuče, seljak – seljače, strateg – strateže, duh – duše, siromah – siromaše, gospod – gospode, Srbin – Srbine, brat – brate, vitez – viteže/viteze, kukac – kukče, vrabac – vrapče, otac – oče, svetac – sveče, sudac – suče, očevdac – očeviče, škrtac – škvrče, čitalac – čitaoče, patuljak – patuljče, tigar – tigre, ovan – ovne, dabar – dobre*), **-u** (10 pravila: *mladić – mladiću, kralj – kralju, krčmar – krčmaru, akušer – akušeru, asistent – asistentu, patak – patku, gačac – gačcu, predak – pretku, mačak – mačku, cvrčak – cvrčku*), **-e/-u** (3 pravila: *pas – pselpsu, ribar – ribare/ribaru, Miloš – Miloše/Milošu*), d) a takođe instr. jedn. (3): **-om** (27 pravila: *turist – turistom, zer – zerom, galeb – galebom, zec – zecom, čovjek – čovjekom, bak – bakom, vuk – vukom, seljak – seljakom, strateg – strategom, duh – duhom*,

⁵ Analizom su obuhvaćene i riječi tipične za hrvatski jezik koje se u srpskom manje koriste ili uopšte ne upotrebljavaju.

siromah – *siromahom*, *gospod* – *gospodom*, *Srbin* – *Srbinom*, *brat* – *bratom*, *krčmar* – *krčmarom*, *akušer* – *akušerom*, *vitez* – *vitezom*, *predak* – *pretkom*, *patuljak* – *patuljkom*, *mačak* – *mačkom*, *cvrčak* – *cvrčkom*, *tigar* – *tigrom*, *pas* – *psom*, *ovan* – *ovnom*, *patak* – *patkom*, *dabar* – *dabrom*, *asistent* – *asistentom*), **-em** (14 pravila *mladić* – *mladićem*, *kralj* – *kraljem*, *stric* – *stricem*, *mudrac* – *mudracem*, *sudac* – *sucem/sudcem*, *kukac* – *kukcem*, *vrabac* – *vrapcem*, *otac* – *ocem*, *svetac* – *svecem*, *bratac* – *bracem*, *očevdac* – *očevicem*, *škrtac* – *škrcem*, *gačac* – *gačcem*, *čitalac* – *čitaocem*), **-om/-em** (2 pravila *ribar* – *ribarom/ribarem*, *Miloš* – *Milošom/Milošem*). U genitivu jednine imenica na **-lac** suglasnik **l** prelazi u **o**: *čitalac* – *čitaoca*. Imenica *brat* nema množine pa se kao množinski oblik koristi zbirna imenica *braća*.

6. Pravila za imenice bez pokretnog samoglasnika **a** obuhvataju mnogo više riječi nego pravila sa pokretnim **a**: odnos je 1.400 : 151. Imenice bez pokretnog samoglasnika obrazuju dvije grupe – homogenu (u kojoj se imenice završavaju na isti suglasnik) i heterogenu (u kojoj se imenice završavaju na različit suglasnik). Većinu završetaka pokrivaju pravila 65^{turist} (15), 69^{galeb} (13) i 57^{mladić} (7). Najveći broj leksema obuhvata pravilo 65^{turist} (391), 39^{ribar} (333) i 178^{seljak} (300).

7. Među imenicama muškog roda za živo sa nultim nastavkom (**-ø**) kvantitativno se izdvajaju alternacije kod imenica na **-c[tʃ]** pa stoga one traže i maksimalan broj pravila (16).

8. Imenice na **-c [tʃ]** sa pokretnim samoglasnikom i bez njega. Postoje tri tipa ovih imenica, od kojih su dva homogena (jedan čine imenice sa pokretnim **a**, drugi imenice bez njega) i jedan heterogeni (imenice i sa pokretnim **a**, i bez njega). Zbog postojanja varijanata u nom. jedn. (*stric*: *strici/stričevi*; *zec*: *zečevi/zecovi/zeci*) i u gen. mn. (*stric*: *strica/stričeva*; *zec*: *zecova/zečeva*) potrebna su pravila 45^{stric} i 52^{zec}. Pravilo 174^{mudrac} se odnosi na imenice bez dubleta i sa instr. jedn. na **-em** (*mudrac/mudrac* – *mudracem/mudracem*).

Imenice sa pokretnim samoglasnikom **a** zahtijevaju 13 pravila. Deset od njih se tiču imenica bez dvostrukih oblika, pri čemu pokretni samoglasnik dolazi samo u nom. jedn. (*vrabac*, *gačac*, *očevdac*, *kukac*, *čitalac*, *bratac*, *škrtac*, *lažac*, *pisac*, *knjigovezac*) i u gen. mn. (*vrabaca*, *gačaca*, *očevdaca*, *kukaca*, *čitalaca*, *brataca*, *škrtača*, *lažaca*, *pisaca*, *knjigovezaca*). Tri pravila su izazvana postojanjem dubleta. Ovdje se izdvaja 50^{sudac} sa četiri varijante ortografskog karaktera: *sudac* – gen. jedn. *suca* – *sudca*, vok. jedn. *suče* – *sudče*, instr. jedn. *sucem* –

sudčem, nom. mn. *suci* – *sudci* (gen. mn. samo *sudaca*).⁶ Dvije padežne varijante obuhvaćene su pravilima 43^{otac} i 49^{svetac}: *otac* – nom. mn. *oci* – *očevi*, gen. mn. *otaca* – *očeva*, *svetac* – nom. mn. *sveci*/*svečevi*, gen. mn. *svetaca*/*svečeva*.

Što se tiče suglasnika ispred pokretnog samoglasnika **a**, preovladava **t**: 43^{otac}, 49^{svetac}, 60^{bratac}, 159^{škrtac}. U svim oblicima bez pokretnog samoglasnika **a** on stupa u dodir sa **c** [tʃ] i rastvara se, npr.: *otac* – *otca* > *oca*, *svetac* – *svetca* > *sveca*, *bratac*/*bratac* – *bratca* > *braca*, *škrtac* – *škrtca* > *škrcu*. U dva pravila uzima se u obzir to da se ispred pokretnog samoglasnika nalazi **d**, koje prelazi u oblicima bez pokretnog samoglasnika **a** u **t**, a zatim u **c** [tʃ]: *sudac* – *sudca* > *sutca* > *suca*, *očevdac* – *očevdca* > *očevitca* > *očevica*; takođe se zvučno **b** transformiše u bezvučno **p**: *vrabac* – *vrabca* > *vrapca*. Po jedno pravilo pokrivaju alternaciju **ž** [ʒ] > **š** [ʃ] (*lažac* > *lašca*), **s** > **š** [ʃ] (*pisac* > *pišće*), **z** > **s** (*knjigovezac* > *knjigovesca*).

To što se suglasnici **č** [tʃ] i **k** ne mijenjaju čini osnovu pravila 197^{gaćac} (*gaćac*, *gaćca*), 46^{kukac} (*kukac*, *kukca*).

Zamjena **l** > **o** u kosim padežima (osim gen. mn.) stvara potrebu za pravilom 51^{čitalac} (*čitalac* – *čitaoca*).

Najviše leksema pokriva pravilo 46^{kukac} (172 riječi) i 51^{čitalac} (37). Ostala pravila (njih 25) obuhvataju svega 40 riječi.

Pravilo 46^{kukac} generiše paradigme za 172 lekseme po sljedećoj formuli: **-_wca**, **-_wče**, **-_wcem** | **-_wci**, **-_waca** (gen. jedn., vok. jedn., instr. jedn. | nom. mn., gen. mn.).⁷ Ove imenice imaju više završetaka. Najviše je imenica na **-vac** (65), čija se pravila obrazuju po obrascu **-vca**, **-vče**, **-vcem** | **-vci**, **-vaca**: *poslodavac*, *stanodavac*, *zakonodavac*. Imenice na **-nac** čine drugu veću grupu (39). Njihova paradigmatska formula glasi: **-nca**, **-nče**, **-ncem** | **-nci**, **-naca**. Ovdje spadaju imenice tipa: *pijanac*, *padobranac*, *stranac*.⁸ Manje ima imenica na **-rac** (17) sa markerima **-rca**, **-rče**, **-rcem** | **-rci**, **-raca**: *magarac*, *jarac*, *muškarac*.

Još je manje imenica na **-lac** (12) sa markerima **-lca**, **-lče**, **-lcem** | **-lci**, **-laca** (*provincijalac*, *genijalac*, *gimnazijalac*), **-mac** 11: **-mca**, **-mče**, **-mcem** | **-mci**, **-maca** (*samac*, *Nijemac*, *ljubimac*), **-kac**, **-jac**, **-ljac** (7): **-_wca**, **-_wče**, **-_wcem**

⁶ Pisanje **dc** susreće se samo u nekim hrvatskim pravopisnim priručnicima.

⁷ Donji indeks _w označava da se na toj poziciji nalazi suglasnik.

⁸ Zbog prostornog ograničenja u daljem tekstu navodimo za veće grupe imenica samo po tri ilustrativna primjera.

| -_wci, -_waca (*kukac, -kca, -kče, -kcem* | *-kci, -kaca*), (*policajac, belogardejac/ bjelogardejac, vanpartijac*), *-jca, -jče, -jcem* | *-jci, -jaca* (*dobrovoljac, vanzemaljac, moljac*) *-ljca, -ljče, -ljcem* | *-ljci, -ljaca, -pac* (5): *-pca, -pče, -pcem* | *-pci, -paca* (*slijepac/slepac, kupac, nakupac*).

Po broju leksema koje se generišu slijedi pravilo 51^{čitalac} sa 37 jedinica zasnovano na formuli *-oca, -oče, -cem* | *-oci, -laca*: *čitalac, činilac, davalac*.

Sve ostale grupe su manje zatupljene. To su 422^{pisac} (6) *-sca, -šče, -scem* | *-sci, -saca*: *vatrogasac, lisac, člankopisac*, 54^{vrabac} (5) *-pca, -pče, -pcem* | *-pci, -baca*: *vrabac, knjigoljubac, rodoljubac*, 324^{lažac} (4) *-šca, -šče, -ščem* | *-šci, -šaca* *lažac, debelokožac, crnokožac*, 174^{mudrac} (3) *-ca, -če, -cem* | *-ci, -ca* *mudrac, takmac, mrtvac*, 323^{knjigovezac} (2) *-sca, -šče, -scem* | *-sci, -zaca* *knjigovezac, ženomrzac*, 73^{očevdac} (3) *-ca, -če, -cem* | *-ci, -daca* *očevdac, mladac, konjovodac*, 43^{otac} (2) *-ca, -če, -cem* | *-ci, -čevi, -čeva, -taca* *otac, praotac*, 45^{stric} (2) *-ca, -če, -cem* | *-ci, -čevi, -ca, -čeva* *stric, princ*, 52^{zec} (2) *zec -ca, -če, -com* | *-ci, -covi, -čevi, -ca, -cova, -čeva* *knez -za, -že, -zom* | *-zi, -zovi, -ževi, -za, -zova, -ževa* 49^{svetac} (1) *svetac, -ca, -če, -cem* | *-ci, -čevi, -čeva, -taca*, 50^{sudac} (1) *sudac, -ca, -če, -cem* | *-ci, -daca*, 60^{bratac} (1) *bratac, -ca, -če, -cem* | *-ci, -taca*, 159^{škrtac} (1) *škrtac, -rca, -rče, -rcem* | *-rci, -taca*, 197^{gačac} (1) *gačac* 'vrana' *-čca, -čcu, -čcem* | *-čci, -čaca*.

9. Imenice na **-r** sa pokretnim samoglasnikom i bez njega. Ove imenice čine drugu grupu po broju pravila (13). Sedam pravila se bazira na manjem broju dubleta (tri slučaja): 39^{ribar} – *ribar*: *ribare/ribaru* (vok. jedn.), *ribarom/ribarem* (instr. jedn.), 297^{dabar} – *dabar*: *dabri/dabrovi* (nom. mn.), 186^{tigar}: *tigar/tigar*: *tigri/tigrovi* (nom. mn.), *tigra/tigrova* (gen. mn.).

Nepostojanje dvostrukih oblika čini osnovu za pravila 158^{zer}, 99^{krčmar}, 302^{akušer}.

Pravilo 158^{zer} je izazvano proširenjem osnove u množini (*zer/zer* 'velika i snažna ptica' – *zerovi, los – losovi*).

Nepostojanje alternacije fiksiraju pravila 299^{krčmar} – *krčmar*: *krčmara* i 302^{akušer} – *akušer*: *akušera*.

Najveći broj imenica na **-r** generiše oblike po pravilu 39^{ribar} (267). Na drugom mjestu je 39^{cmužičar} (60).

Pravilo 39^{ribar} obuhvata imenice sa različitim završetkom, čiji su generatorski orijentiri **-a, -u, -om** | **-i, -a** (gen., vok. i instr. jedn. | nom. i gen. mn.). Najviše ima imenica na **-ar**: (115), npr. *akcionar, alhemičar, bečar*. Slijede imenice na

-er (68): *amater, bokser, dizajner*. Na trećem mjestu su one koje se završavaju na **-tor** (52): *administrator, ambasador, diktator*. Manje je imenica na **-ir** (24): *brigadir, kasir, kurir*, **-ur** (5): *brontosaur* 'izumrla džinovska životinja iz perioda krede', *Grgur, kengur, Pompadur, tiranosaur*, **-or** (3): *Isidor, Nikanor*. Paradigmatički markeri su **-a, -u, -om/-em | -i, -a** (gen., vok. i instr. jedn. | nom. i gen. jedn.): *alkoholičar, analitičar, bankar*.

Po drugim pravilima se tvore oblici za znatno manji broj imenica – 302^{akušer} (1) **-a, -u, -om | -i, -a** *akušer*, 299^{krčmar} (3) **-ra, -ru, -rom | -ri, -ra** *krčmar, baloner, hipnotizer*, 186^{dabar} (1) **-bra, -bre, -brom | -bri, -brovi, -bara, -brova** *dabar*, 297^{tigar} (1) **-gra, -gre, -grom | -gri, -grovi, -gra, -grova, -gara** *tigar*, 299a^{car} (1) **-a, -e, -em | -i, -evi, -a, -eva** *car*, 299b^{gospodar} (1) **-a, -e, -u, -om, -em | -i, -a** *gospodar*.

10. Imenice na **-k** sa pokretnim samoglasnikom i bez njega. Po broju pravila (10) one čine treću grupu. Ovdje su potrebna pravila za imenice sa pokretnim **a**, bez alternacije i sa alternacijom.

Imenice sa pokretnim samoglasnikom **a** traže pet pravila. Imenice bez dubleta (kod kojih se ak. podudara sa gen.) zahtijevaju tri pravila u zavisnosti od toga na koji se suglasnik završavaju: **-bak**: 180^{bak} (*bak*: vok. jedn. *bače*, nom. mn. *bakovi*, gen. mn. *baka/bakova/bakova*), **-ljak** [-ĺak]: 178^{seljak} (*seljak*: vok. jedn. *seljače*, nom. mn. *seljaci*) ili na **-rak**: 294^{rak} (*rak*: vok. jedn. *rače*, nom. mn. *rakovi*).

Neke imenice na **-čak** [-tʃak] i **-ljak** [-ĺak] nemaju dublete i za generisanje paradigme potrebna su dva pravila: 183^{cvrčak} (*cvrčak*: nom. mn. *cvrčci*, gen. mn. – *cvrčaka*): 181^{patuljak} (*patuljak*: vok. jedn. *patuljče*, nom. mn. *patuljci*, gen. mn. – *patuljaka*).

Na imenice sa varijativnošću odnose se dvije grupe pravila: **a**) imenice na **-čak** sa dva dubleta traže pravilo 182^{mačak} (*mačak*: nom. mn. *mačci/mačkovi*, gen. mn. *mačaka/mačkova*), a takođe imenice na **-tak**: 296^{patak} (*patak*: nom. mn. *paci/patkovi*, gen. mn. *pataka/patkova*), **b**) imenice na **-dak** sa jednim dubletom obuhvata pravilo 179^{predak} – *predak*: nom. mn. *preci/pretc*; ovdje se radi o ortografskoj varijanti u nom. mn. – jednoj u srpskom jeziku (*predak* – *preci*) i dviju u hrvatskom (*predak*: *preci* – *pretci*)⁹.

⁹ Gen. mn. glasi *predaka*.

Najviše varijanti (5) pokriva pravilo 230^{lisak}: 1. *lisak* 'konj sa celom' gen. jedn. *lisaka*, vok. jedn. *lisače*, instr. jedn. *lisakom*, nom. mn. *lisci*, gen. mn. *liska-ka*, 2. *lisak* (zoonim: vrsta močvarnih ptica / fitonim: 'glavni dio lista, pojedini list') – gen. jedn. *liska*, vok. jedn. *lišće*, instr. jedn. *liskom*, nom. mn. *liskovi*, gen. mn. *liskova*.

U generisanju oblika za imenice na **-k** postoje pravila pomoću kojih se generišu oblici za više stotina imenskih jedinica i pravila koja služe za obrazovanje formi samo jedne lekseme.

Najviše riječi (90%) od ukupnog broja imenica na **-k** pokriva pravilo 178^{se-ljak}. Ono se bazira na sljedećim markerima: **-ka**, **-če**, **-kom** | **-ci**, **-ka** (gen. jedn., vok. jedn. | nom. mn., gen. mn.). Ove imenice imaju završetke: **-nik**, **-ak** i **-ik**. Najviše je zastupljen **-nik** (237): *podanik*, *konjanik*, *izaslanik*. Slijedi **-ak**: (64): *seljak*, *veseljak*, *zemljak*.

Nema mnogo imenica na **-ik** (2): *laik*, *šeik*.

Ostala pravila pokrivaju manje imenica. Pravilo 181^{patuljak} je zasnovano na markerima **-_wka**, **-_wče**, **-_wkom** | **-_wci**, **-_waka** i obuhvata 13 leksema: *patuljak*, *čovječuljak*, *čovečuljak*, *momak*, *potomak*, *sinak*, *pastorak*, *čvorak*, *curak*, *djevojčurak*, *posinak*, *ljevak*, *levak*.

Po pravilu 182^{cvrčak} generiše se paradigma za *cvrčak* i *siromašak* po formuli **-_wka**, **-_wku**, **-_wkom** | **-_wci**, **-_waka**.

Druga pravila se odnose na jedinične lekseme – 179^{predak}: *predak*, **-tka**, **-tku**, **-tkom** | **-ci**, **-tci**, **-daka**, 49^{curetak}: *curetak* **-tka**, **-če/-tče**, **-ktu**, **-tkom** | **-ci**, **-tci**, **-taka**, 49a^{tetak}: *tetak*, **-tka**, **-tku**, **-tkom** | **-tci**, **-taka**, 182^{mačak}: *mačak*, **-čka**, **-čku**, **-čkom** | **-čci**, **-čovi**, **-čaka**, **-čkova**, 296^{patak}: *patak*, **-tka**, **-tku**, **-tkom** | **-ci**, **-tkovi**, **-taka**, **-tkova**.

11. Imenice bez alternacija i dubleta tipa *turist*. Po broju primjera izdvaja se pravilo 65^{turist}. Ono obuhvata velik broj imenica (391) koje nemaju ni alternaciju, ni dubletne oblike. Njihovi su paradigmatski markeri: **-a**, **-e**, **-om** | **-i**, **-a** (gen., vok. i instr. jedn. | nom. i gen. mn.). Pretežni broj takvih imenica (149) završava se na **-n**: *ataman*, *baron*, *baštovan*. Drugu veću grupu čine imenice na **-t** (117): *ađutant*, *advokat*, *agent*. Slijede imenice na **-v** (26): *crv*, *detektiv*, *lažov*, *lopov*. Manje su česte imenice na **-l** (19) *admiral*, *apostol*, *civil*, **-s** (18) *gorostas*, *kolos*, *kustos*, **-m** (17) *agronom*, *astronom*, *fantom*, **-d** (14) *izrod*, *leopard*, *medved*/*medvjed*, **-z** (10) *Englez*, *uljez*, *oroz*, **-f** (6) *fotograf*, *filozof*, **-p** (6) *Arap*, *arhiepiskop*,

mangup, **-b** (2) *domoljub*, *rodoljub*, **-r** (2) *batler*, *Kerber*, **-ž** [3] (2) *lupež*, *Serž*, **-c** [ʦ] (2) *luc*, **-đ** [ǰ] (1) *vođ*.

12. Ostale imenice sa pokretnim vokalom **a** i bez njega. Pravila za tu grupu pokriva malen broj leksema. U njoj najviše pravila (5) zahtijevaju imenice koje se završavaju na **-n**. Posebna pravila su potrebna kada postoje varijantni oblici (više od dva). Jedno od njih je 191^{ovan} – *ovan/ovan*: *ovni/ovnovi* (nom. mn.), *ovana/ovnova/ovnova* (gen. mn.). U osnovi 185^{Srb} leži nulta varijativnost – *Srb*: *Srbina* (gen. jedn.), *Srbi* (nom. mn.), *Srba* (gen. mn.). U grupi imenica na **-n** zastupljena su dva tipa alternacija: pokretni samoglasnik **a** (191^{ovan}), refleks jata (62^{korijen}) i proširenje osnove u mn. (191^{ovan} – *ovan/ovan*: gen. jedn. *ovna*). Najviše se imenica generiše po pravilu 185^{Srb} (23) sa paradigmatskim markerima **-ina** **-ine**, **-inom** | **-i**, **-a** (gen., ak. i instr. jedn. | nom. i jen. mn.) *Srb*, *državljanin*, *građanin*. Ostala pravila odnose se na pojedinačne imenice: 325^{gospodin} *gospodin*, **-dina** **-dine**, **-dinom** | nom. mn. **-da**, **-dini**, gen. mn. **-de**, **-dina**, vok. mn **-do**, 191^{ovan} *ovan*, **-vna** **-vnu**, **-vnom** | **-vni**, **-vnovi**, **-vana** **-vnova**.

Završetak **-g** traži dva pravila, od kojih 176^{strateg} generiše oblike za najveći broj leksema (27) po formuli **-ga**, **-že**, **-gom** | **-zi**, **-ga**: *strateg*, *ideolog*, *suprug*. Pravilo 329^{bog} pokriva imenice *bog*, *polubog*, *drug*, *vrag*, *beg*, *beglerbeg* ‘vrhovni zapovjednik’ sa markerima **-ga**, **-že**, **-gom** | **-zi**, **-govi**, **-ga**, **-gova**. Neka su pravila vezana za imenice bez dubleta: 176^{strateg}. U oblicima vok. jedn. **g** prelazi u **ž** [3]: *strateg/strateg* – *strateže/strateže*.

Neka pravila za imenice sa završetkom **-t** uzimaju u obzir postojanje dvostrukih oblika u dva padeža, recimo 329^{bog} nom. mn. **-zi**, **-govi**, gen. mn. **-ga**, **-gova**. Varijativnost u gen. mn. leži u osnovi pravila tipa 189^{asistent} (*asistent*: *asistent*/*asistenata*). Postoje pravila koja se odnose na imenice bez dubleta: 188^{brat}. Pomoću ovih pravila generišu se oblici za malen broj imenica: 189^{asistent}, formula **-nta**, **-nte**, **-ntom** | **-nti**, **-nata**: *asistent*, *demonstrant*, *sektant*, *apstinent*, 188^{brat}, formula **-ta**, **-te**, **-tom** | **-ća**, **-će** vok. **braćo**: *brat*, *polubrat*, 326^{gost}, formula **-a**, **-e**, **-om** | **-i**, **-i**, **-iju** *gost*.

Sve ostale imenice imaju manje od tri pravila. To su one koje se završavaju na **-š** [ʃ]: 190^{Miloš}, formula **-še**, **-še**, **-šu**, **-šem**, **-šom** | **-ši**, **-ša** (4) *Miloš*, *Radoš*, *Tomaš*, *Uroš*. Pravila za ove imenice su povezana sa varijativnošću u dva padeža – vok. i instr. jedn.: 190^{Miloš} – *Miloš*: *Miloše/Milošu*, *Milošem/Milošom*. Za pravilo 177^{siromah} formula glasi: **-ha**, **-še**, **-hom** | **-si**, **-ha** (8) *siromah*, *Alah*, *Kazah*, *padišah* ‘titula bivših vladara (sultana)’, *šejh*, *patrijarh*, *monarh*, *očuh*.

Tri se pravila odnose na imenice bez padežnih varijanti, ali sa mogućim prozodijskim specifičnostima: 177^{siromah} – *siromah*: gen. jedn. *siromaha*. U svim imenicama obuhvaćenim ovim pravilom u nom. mn. dolazi do sibilizacije u oblicima sa neproširenom osnovom: *siromah* – *siromasi*.

Pravilo 63^{duh}, po formuli **-ha, -še, -hom | -si, -hovi, -ha, -hova** (1), uzima u obzir alternaciju (prvu palatalizaciju) u vok. jedn.: *duh* – *duše*, *siromah* – *siromaše*.

U osnovi pravila 18^{kralj} leže dvostruki oblici u gen. mn. po formuli **-a, -u, -em | -i, -evi, -a, -eva** (6): *kralj* – gen. jedn. *kralja*, vok. jedn. *kralju*, nom. *kraljevi/kralji*, gen. mn. *kralja, kraljeva* (*kralj, ekskralj, gej, zmaj, miš, vrač*).

Imenice koje se završavaju na **-č [tʃ]** pokriva pravilo 57^{mladić} po formuli **-a, -u, -em | -i, -a** (79) *birač, čistač, crtač*. Ono se odnosi i na imenice na **-č [tʃ]** koje imaju dvostruke oblike i nastavak **-em** u instr. jedn. (*brijač: brijačem*).

Riječi na **-ć [tɕ]** nemaju dublete, a u instr. jedn. javlja se nastavak **-em** (*mladić: mladićem*). One su takođe obuhvaćene pravilom 57a^{mladić} po formuli **-a, -u, -em | -i, -a**: *mladić, konjić, kunić*; ovdje se nalaze i mnogobrojna prezima na **-ić** (*Popović, Branković* i sl.).

Pravilo 69^{galeb} dolazi po formuli **-a, -e, -om | -i, -ovi, -a, -ova** (34): *galeb, golub, jastreb*. U ovoj grupi se pojavljuju varijante u nom. i gen. mn.: 69^{galeb} (*galeb: galebi/galebovi, galeba/galebova*). Imenice koje se završavaju na **-j [j]** pokriva pravilo 57^{mladić} (13): *lakej, genij, heroj*. Pravilo 175^{gospod} (formula **-da, -de, -dom**) odnosi se samo na jednu imenicu čiju paradigmatiku čini oblik množine: *gospod: gospoda*.

Imenice koje obuhvata pravilo 187^{pas} (3) traže posebno pravilo zbog postojanja dvostrukih oblika i pokretnog **a**: 187^{pas} *pas – psa* (*pas, cipal, psihijatar*).

Za imenice na **-z** dovoljno je pravilo 250^{vitez} po formuli **-za, -že, -zom | -zi, -zovi, -ževi, -za, -zova, -ževa** (250^{vitez}). Ono je iznuđeno dubletima u nom. mn. (*vitezi/vitezovi/viteževi* i gen. mn. *viteza/vitezoval/viteževo*), a takođe sibilizacijom u vok. jedn. – **z > ž [ʒ]**: (*vitez – viteže*).

Imenice na **-đ [ʒ]**, **-f, -p, -v, -z** ne zahtijevaju posebna pravila.

LITERATURA

- Gralis-Korpus-www: Gralis-Korpus http://www-gewi.uni-graz.at/gralis/korpusarium/gralis_korpus.html. Stanje 15. 8. 2015. (Utemeljio i razvio Branko Tošović)
- Klajn, Ivan. 2005. *Gramatika srpskog jezika*. Beograd: Zavod za udžbenike i nastavna sredstva.
- Maretić, Tomo. 1901. *Gramatika hrvatskoga jezika za niže razrede srednjih škola*. Zagreb: Štampa i naklada Knjižare L. Hartmana (Kugli i Deutsch).
- Mrazović, Pavica, Vukadinović, Zora. 1990. *Gramatika srpskohrvatskog jezika za strance*. Sremski Karlovci – Novi Sad: Izdavačka knjižarnica Zorana Stojanovića.
- Silić, Josip, Pranjković, Ivo. 2005. *Gramatika hrvatskoga jezika za gimnazije i visoka učilišta*. Zagreb: Školska knjiga.
- Stanojčić, Živojin. 2010. *Gramatika srpskog književnog jezika*. Beograd: Kreativni centar.
- Stevanović, Mihailo. 1970. *Savremeni srpskohrvatski jezik (Gramatički sistemi i književnojezička norma)*: I – Uvod, Fonetika, Morfologija. 2. izd. Beograd: Naučna knjiga.
- Tošović, Branko. 2009. *Korrelative Grammatik des Bosni(ak)ischen, Kroatischen und Serbischen. Teil I: Phonetik – Phonologie – Prosodie*. Münster: LIT. (Slawische Sprachkorrelationen, Bd. 4)
- Tošović, Branko. 2010^a. *Gralis: Das linguistische Slawistik-Portal der Karl-Franzens-Universität Graz (2000–2010)*. Graz: Institut für Slawistik der Karl-Franzens-Universität.
- Tošović, Branko. 2010^b. *Das Gralis-Korpus // Branko Tošović, Arno Wonisch (Hg./Ur.) Srpski pogledi na odnose između srpskog, hrvatskog i bošnjačkog jezika*. Graz – Beograd: Institut für Slawistik der Karl-Franzens-Universität Graz – Beogradska knjiga. Knjiga I/2: 491–519.
- Tošović, Branko. 2013^a. Морфологическое порождение существительных сербского, хорватского, бошняцкого и черногорского языков // *Wiener slawischer Almanach*. München: Verlag Otto Sagner, c/o Kubon & Sagner: 113–134.
- Tošović, Branko. 2014^b. Automatsko kodiranje pomoću MorfoGeneratora // *Slavistika*. Beograd: Slavističko društvo Srbije. Knj. XVIII: 207–214.
- Zečević, Vesna, Peti, Mirko, Pavešić, Slavko, Malić, Dragica, Lončarić, Mijo, Barić, Eugenija, Znika, Marija. 1995. *Hrvatska gramatika*. Zagreb: Školska knjiga.
- Тошович, Бранко. 2014. Грамматическая аннотация для Гралис-Корпуса // *Казанская школа-семинар по компьютерной и когнитивной лингвистике: TEL'2014 „Языковая семантика: модели и технологии“*. Казань: Научно-исследовательский институт „Прикладная семиотика“ Академии наук Республики Татарстан: 142–147.
- Тошович, Бранко. 2011. *Корреляционная грамматика сербского, хорватского и бошняцкого языков: Часть I: Фонетика – Фонология – Просодия*. Москва: Языки славянской культуры.

Бранко Тошович

ПОРОЖДЕНИЕ МОРФОЛОГИЧЕСКИХ ФОРМ И ПАРАДИГМ
ДЛЯ ОДУШЕВЛЕННЫХ СУЩЕСТВИТЕЛЬНЫХ
МУЖСКОГО РОДА В СЕРБСКОМ ЯЗЫКЕ

Резюме

Анализ состоит из трех частей. В первой (а) говорится о причинах и необходимости автоматического порождения морфологических форм и парадигм для изменяемых частей речи в целях исследования и обучения, (б) выделяются правила, при помощи которых проводятся генераторские процедуры, (в) указывается на своеобразие изменяемых частей речи. Во второй части коротко представлена модель порождения форм и парадигм в системе имен существительных. Третья часть посвящена правилам порождения для одушевленных существительных мужского рода.