

Nas jezik

SLJK

ISSN 0027-8084
UDK 808.61/.62

ИНСТИТУТ ЗА СРПСКИ ЈЕЗИК САНУ

НАШ ЈЕЗИК

XXXV/1—4

**БЕОГРАД
2004**

НАШ ЈЕЗИК

Књига XXXV

Св. 1–4 (2004)

САДРЖАЈ

Ч л а н ц и

Књижевни језик

- Појам *VERUM FOCUS* и његово сигналзовање у савременом стандардном српском (МИЛКА ИВИЋ) 1
- Морфонолошке алтернације и морфолошка анализа (на примерима из српског књижевног језика) (МИРОСЛАВ НИКОЛИЋ) 6
- Статистички и кориснички оријентисана граматика: пример именица мушког рода (ДАНКО ШИПКА) 13
- Анализа приступа придевима у сербокroatистичким граматикама (ВЕСНА ЛОМПАР) 23
- Придеви с наставком *-ић* (МИЛАН ШИПКА) 30
- Придеви и прилози с префиксом *без-* (ЂОРЂЕ ОТАШЕВИЋ) 46
- Називи за животиње у српском језику с обзиром на њихов род (САЂА ЂУРОВИЋ) 54

Историја књижевног језика

- Реконструкција самогласника у критичком издању Горског вијенца (РАДМИЛО МАРОЈЕВИЋ) 67

Стилистика

- Паремиолошки стил(ови) (ЈЕЛЕНА ЈОВАНОВИЋ) 117

П р и л о з и

- Перципирање финалних вокала у деκлинацији страних мушких имена и презимена на вокал (МАРИНА СПАСОЈЕВИЋ) 137

П р и к а з и

- Српски језички приручник (РАДА СТИЈОВИЋ) 155
- Стана Ристић, Експресивна лексика у српском језику (БОЈАНА МИЛОСАВЉЕВИЋ) 159
- Нови језички савјетник (СРЕТО ТАНАСИЋ) 163
- Маја Ђукановић — Жељко Марковић, Основи граматике словеначког језика (ТАЂА САМАРЦИЋ) 167
- Семантичко-деривациони речник, Свеска 1: Човек — делови тела, (МИЛИЦА РАДОВИЋ-ТЕШИЋ) 169

Д о д а т а к

- Сугестије за израду WEB сајта Института за српски језик САНУ (СОФИЈА МИЛОРАДОВИЋ — НЕНАД МИЛОРАДОВИЋ) 175

ДАНКО ШИПКА
(Феникс, САД)

СТАТИСТИЧКИ И КОРИСНИЧКИ ОРИЈЕНТИСАНА ГРАМАТИКА: ПРИМЕР ИМЕНИЦА МУШКОГ РОДА

У овом раду аутор, на примеру именица мушког рода, представља алтернативни начин граматичке обраде засноване на корпусној лингвистици и уобличене тако да корисницима олакша приступ информацијама.

Увод

Започети још у осамдесетим годинама прошлог века, два битна друштвена процеса развила су се у потпуности током претходне деценије. С једне стране, нове информацијске технологије омогућују извлачење најразноврснијих информација из различитих дигитализованих текстуалних и нетекстуалних извора. С друге пак стране, у свим областима живота све се више брине о кориснику — Мајкрософтови програми, који толико угађају кориснику да се од њега очекује једино да перципира различите сличице и реагује на њих без готово икакве когнитивне обраде информација, добар су пример овог другог тренда.

Оба ова процеса постепено продиру у различите области примењене лингвистике, првенствено лексикографију. Све више речника сачињава се на основу електронских корпуса, све је више електронских речника и лексичких база података. Истовремено, посебно у англосаксонској лексикографији, израста грана речника намењених потребама оних који уче језик (два водећа британска издавача у овој области, Oxford University Press и Longman, имају серије таквих речника). У граматици овај приступ, мада мање изражен, ипак је присутан — добар пример је новија Ученичка граматика грузијског језика за говорнике енглеског (George Hewitt: *Georgian A Learner's Grammar*, Routledge, 1996).

Наша граматиологија, мислим ту на приручнике српскохрватског језика у целини или једног од његових етничких стандарда (видети: Вагић, Е. et al. (1995), *Хрватска грамајлика*, Школска књига, Загреб; Мразовић П. и Вукадиновић З. (1988), Мразовић, П., Вукадиновић, З., *Грамајлика српскохрватског језика за странце*, Универзитет у Новом Саду, Нови Сад; Станојчић Ж. и Поповић Љ. (1992), *Грамајлика српског језика*, Завод за уџбенике и наставна средства, Београд; Стевановић, М. (1981), *Савремени српскохрватски језик*, Научна књига, Београд), остаје

изван оба ова процеса — и по томе што тврдње нису засноване на статистичким подацима и по томе што се не води рачуна о кориснику.

Овде ћу, на примеру именица мушког рода, представити алтернативни начин граматичке обраде засноване на корпусној статистици и обличене тако да за кориснике олакша приступ информацијама.

Начелни проблеми

Пре него што се позабавимо конкретним проблемом и његовом граматичком обрадом, ваља се запитати зашто уопште граматичку обраду треба узети за два поменута тренда. Какве су од тога користи?

Потенцијалне користи које доноси информатизација граматичког описа и његово прилагођавање кориснику могу се свести на пораст вредности ове три особине: исцрпност, прецизност, примењивост.

Укратко, ако граматички опис заснивамо на статистичким корпусним подацима, у ситуацији смо да њиме обухватимо све случајеве који се у корпусу јављају. Надаље, информација коју дајемо прецизна је у смислу да знамо колико конкретних лексичких јединица потпада под одређени флексијски образац. Коначно, информација ће бити примењива тако што ће корисник у кратком времену моћи пронаћи одговарајуће информације.

Изнесени аргуменат можемо представити и с друге стране. Наиме, непримењивање статистички и кориснички оријентисаног приступа води у ситуацију где се не обухватају сви случајеви (честе су у постојећим граматицама квазифреквенцијска одређења типа: многе, неке, нпр. именице, и сл.), не зна се колико је сваки флексијски образац релевантан у смислу лексичких јединица које га следе, а корисници граматица тешко могу да нађу релевантне информације.

Изнесене предности статистички и кориснички оријентисаног граматичког описа могу се остварити уколико опис следи ове опште принципе: а) исцрпност, б) прецизност, в) недвосмисленост, г) штедљивост, д) прегледност описа. Другим речима, описом треба обухватити све постојеће случајеве, све њихове облике и само те облике, недвосмислено одредити који случајеви потпадају под које правило, употребити најкраћа могућа средства описа представљена у прегледном облику.

Овде ћемо на конкретном примеру именица мушког рода, једног од најкомплекснијих сегмената наше морфологије, показати како би требало да изгледа статистички и кориснички оријентисан граматички опис.

Испитивање конкретног случаја — деклинација именица мушког рода

Ако се погледа опис именица мушког рода у било којој од постојећих штампаних граматица (видети примере у претходном делу текста), приметимо да се не зна који удео конкретне именичке масе следи који флексијски тип. Даље, флексијски типови организовани су непрегледно и без општег принципа који би одредио које алтернативе чине засе-

бан флексијски тип, а које су варијација једног те истог флексијског типа. Присутне су, надаље, бројне псеудофреквенцијске квалификације типа „већном“, „многе“, „често“ итд. Све ово чини да је опис за корисника крајње збуњујући и тако мало примењив.

Овде се предлаже опис који уклања ове недостатке. Опис је заснован на сопственој бази података од преко 120 хиљада одредница са означеним флексијским типом за сваку од њих. База података обухвата 79270 именица, од чега 35285 мушког рода. Опис узима у обзир сваку од ових именица.

Теоријска позадина описа подразумева следеће:

Принцип 1: Опис је теоријски неутралан, другим речима како постојеће теорије морфосинтаксичког описа, на пример: ХПСГ, минимализам, теорија оптималности, когнитивна лингвистика, тако и теорије морфолошког описа, нпр. бешавна морфологија, дворазинска морфологија и сл. нису узете у обзир.

Образложење 1.1: Граматика намењена општем кориснику не може од њега очекивати познавање одређеног лингвистичког приступа.

Образложење 1.2: Теоријски приступи, заинтересовани консицентношћу властитог описа, најчешће представљају непотребно преоптерећивање корисника количином информација (шира елаборација може се наћи у Šipka et al. *NLP is Different: Meaningful is Meaningless, Meaningless is Meaningful*, in: Gerhild Zybatow et al. (ed.) 3. Europäische Konferenz „Formale Beschreibung slavischer Sprachen“, Leipzig 1999, (= *Linguistische Arbeitsberichte*; 75), Institut für Linguistik der Universität Leipzig, pp. 213–220, 2000).

Принцип 2: Опис је заснован на глобалном и локалном успостављању подразумеване вредности. На пример, ако се не каже друкчије, именица која се завршава на консонант припадаће основном деклинацијском типу.

Образложење 2.1: На овај начин олакшава се усвајање информација — за подразумевану вредност не треба памтити никакве додатне информације.

Принцип 3: Опис подразумева утврђен редослед представљања елемената. На пример, регуларна флексија представља се пре изузетачне, одступање од целе флексијске парадигме пре одступања у поједином падежу итд.

Образложење 3.1: На овај начин корисник има јасно утврђену мапу тражења и усвајања информација.

Принцип 4: Успостављање подразумеване вредности и редослед представљања засновани су на статистичком уделу елемената. Подразумевана вредност је увек најчешћи елеменат, чешћи елеменат претходи ређем.

Образложење 4.1: На овај начин корисник се поново растеређује у погледу усвајања информација јер се додатне информације не траже уз најчешће сретане флексијске типове.

Принцип 5: Опис подразумева хијерархију елемената флексије. Тако је одступање у једном флексијском наставку хијерархијски ниже од таквог одступања коме подлеже цела парадигма.

Образложење 5.1: На тај начин корисник добија јасну слику о релативној тежини информација које усваја.

Принцип 6: Опис тежи минималном напору корисника при усвајању информација. Опис је успостављен тако да се подразумевана вредност изводи на најједноставнији могући начин (канонска основа + наставци уз евентуалне морфолошке алтернације).

Образложење 6.1: Корисник на овај начин лакше усваја информације.

Имајући све ово у виду, грађу именица мушког рода требало би представити у оваквом редоследу.

подразумевани флексијски образац	три најчешћа изузетачна обрасца	други изузетачни обрасци	изузеци у појединачним наставцима и алтернацијама	регионалне алтернације
----------------------------------	---------------------------------	--------------------------	---	------------------------

Ако се погледају прва три елемента представљања, статистике јасно говоре у прилог оваквом редоследу. Подразумевани флексијски образац (именице типа *ћрозор*, *ћрофесор*, *ћрад* и сл.) налазимо у 80% свих наших именица. Три следећа најчешћа обрасца обухватају 12,5% именица (образак непостојаног а, нпр. *ћисац*, *ћисца*), 3,5% (именице на -а у номинативу, нпр. *судија*) те 2% (номинатив на -о, -е, нпр. *радио*, *геније*), што значи да свега четири обрасца обухватају 98% свих именица. Сви други обрасци, узети заједно, покривају 2% именица. Ово значи да ће, рецимо, неко ко учи наш језик, научивши свега један флексијски образац, правити грешке у само 20% именица. Ови подаци сасвим јасно говоре какав треба да је редослед представљања информација и шта треба нагласити.

Изузеци у појединачним наставцима и алтернацијама, сасвим разумљиво, треба да буду представљени након флексијских парадигми јер обухватају знатно уже подручје. Уз то, врло често је тако да је могућа и регуларна вредност и одступање, па непримењивање одступања не води у грешку него у непостојање стилске или семантичке нијансе (нпр. *вуци* поред *вукови*).

Регионални изузеци (типа ијекавског ије : је, нпр. цвијет : цвјетови) обухватају још мањи број случајева, па их треба дати на крају.

Треба рећи и ово. Подразумевани флексијски образац следи начело: основа = канонски (речнички) облик
облик = основа (уз могуће алтернације) + наставак

Сви случајеви који следе ово начело иду у подразумевани флексијски образац. Остали случајеви (они код којих се канонски облик разликује од основе и они где се облици творе од више основа иду у засебне флексијске образце. Тако онда, оно што се у традиционалним граматицама зове дуга множина иде у више образаца (и основни и образац непостојаног *a*) јер се ради о интерфиксу који додајемо у множини на канонску основу и јер се та алтернација протеже у више флексијских типова. Оно што се традиционално зове кратка множина засебан је флексијски тип јер поред основе једнаке канонском облику треба да имамо и другу основу за множину.

Подразумевани флексијски образац представимо овде у целини, остале дајући само сажете информације и примере.

Подразумевани флексијски образац мушког рода (80% именица)
 основа = канонски (речнички) облик
 облици = основа (уз могуће алтернације) + наставци

Пад.	Јед.	Примери	Мн. ^{5 7}	Примери
<i>Ном.</i>	- \emptyset^8	аутомобил, сат, зец, војник, краљ, део	- <i>и</i> ³	аутомобили, сатови, зечеви, војници, краљеви, делови
<i>Ген.</i>	- <i>a</i>	аутомобила, сата, зеца, војника, краља, дела	- <i>a</i>	аутомобила, сатова, зачева, војника, краљева, делова
<i>Дат.</i>	- <i>у</i>	аутомобилу, сату, зецу, војнику, краљу, делу	- <i>има</i> ³	аутомобилима, сатовима, зечевима, војницима, краљевима, деловима
<i>Ак.</i>	- \emptyset^2 ⁸ - <i>a</i>	аутомобил, сат, зеца, војника, краља, део	- <i>e</i>	аутомобиле, сатове, зечеви, војнике, краљеве делове,
<i>Вок.</i>	- <i>e</i> ^{4 6} - <i>у</i>	аутомобиле, сате, зече, војнике, краљу, деле,	- <i>и</i> ³	аутомобили, сатови, зечеви, војници, краљеви, делови
<i>Инс.</i>	- <i>ом</i> ¹ - <i>ем</i>	аутомобилом, сатом, зецом, војником, краљем, делом,	- <i>има</i> ³	аутомобилима, сатовима, деловима, зечевима, војни- цима, краљевима
<i>Лок.</i>	- <i>у</i>	аутомобилу, сату, зецу, војнику, краљу, делу,	- <i>има</i> ³	аутомобилима, сатовима, зечевима, војницима, краљевима, деловима

Алтернације:

1. тврда основа: -о... мека: -е...
војником, столом, краљем
2. живо: -а неживо -∅
Видим војника. Видим аутомобил.
3. Алтернација к, г, х : ц, з, с
к-ц: војник, војници, војницима;
г-з: подвиг, подвизи, подвизима;
х-с: монарх, монарси, монарсима
4. тврда основа -е и алтернација к, г, х : ч, ж, ш
к-ч: војник, војниче;
г-ж: враг, враже;
х-ш: дух, душе

мека основа -у

5. један слог, тврда основа: интерфикс -ов- воз : возови
један слог, мека основа: интерфикс -ев- краљ : краљеви
два и више слогова: интерфикс -∅- капут : капутџи
6. алтернација ц, з : ч, ж
ц-ч: зец : зече
з-ж: књаз : књаже
7. ако је интерфикс -ев-, алтернација је ц : ч
зец : зечеви
8. алтернација о : л
о-л: део : дела

Најчешћи нзузетачни обрасци

Ови и сви други обрасци у конкретной граматици дају се с истим степеном елаборације као основни тип. Овде су, ради уштеде у простору, представљени у скраћеном облику.

Образац неностојаног а (12,5% именица)

основа 1 = *[консонант]а[консонант], нпр: сан/* = било шта/

основа 2 = *[консонант][консонант], нпр. *сн*

облици = основа{1, 2}(уз могуће алтернације) + наставци:

Примери: *иринач, мачак, палац, орао, руководилац, отац*

Образац номинативног -а (3,5% именица)

Основа = канонски облик — (завршно -а)

Облици = основа (уз могуће алтернације) + наставци

Примери: *војвода, владика, архиепископ*

Образац номинативног -о, -е или -и (2% именица)

Основа = канонски облик — (завршно -о, -е, или -и)

Облици = основа (уз могуће алтернације) + наставци

 Примери: *аушо, радио, несџашко, геније, јули*
Остати изузетачни обрасци (2% узети заједно)

Назив обрасца	Начин мењања	Примери
Номинативно -о/-е Генитивно -е	основа = канонски облик — (завршно -о или -е); облици = основа (уз могуће алтернације) + наставци	<i>Владо, Владе;</i> <i>Сџише, Сџише;</i> <i>фишо, фише</i>
Номинатив јединице -о, Номинатив мно- жине -а	основа = канонски облик — (завршно -о); облици = основа (уз могуће алтернације) + наставци	<i>ехо, мн. еха;</i> <i>жиголо, мн.</i> <i>жигола</i>
Скраћена множина	основа 1 = канонски облик; основа 2 = канонски облик — (завршно -ин); облици = основа {1,2}(уз могуће алтернације) + наставци	<i>Србин</i> <i>мн. Срби;</i> <i>Турчин</i> <i>мн. Турци</i>
Придевска промена	основа = канонски облик — (завршно -и); облици = основа (уз могуће алтернације) + придевски наставци	<i>руски,</i> <i>руског (језик);</i> <i>ближњи,</i> <i>ближњег</i>
Образац јединског проширења основе	основа = канонски облик — (завршно -о или -е); облици = основа (уз могуће алтернације) + наставци	<i>Раде, Радеша;</i> <i>креле, крелеша</i>
Основа на -о, -е, -и, или -у	основа = канонски (речнички) об- лик; облици = основа (уз могуће алтернације) + наставци:	<i>биро, бироа;</i> <i>бифе, бифеа;</i> <i>џакси, џаксија;</i> <i>гуру, гуруа;</i> <i>круџије, круџијеа</i>
Основа на страни графем	Исто као подразумевани тип, али уз факултативно додавање цртице	<i>Drew, Drewa</i> <i>или Drew-a</i>
Верзалне скраћенице	Исто као подразумевани тип, али уз обавезно додавање цртице	<i>РАК, РАК-а</i> <i>(Рачунарски</i> <i>амајтерски</i> <i>клуб)</i>
Обрасци који обухватају само по једну именицу	Различити начини	<i>човек мн. људи;</i> <i>господин мн.</i> <i>господа;</i> <i>браћ мн. браћа</i>

Након представљања свих ових образаца могу се дати случајеви кад се одступа од појединих падежних наставака, кад постоји више могућности, кад се не примењује алтернација и сл. Тако се онда нано-во успоставља подразумевана вредност (у свим флексијским обрасцима подразумева се одговарајући наставак, алтернација и сл.), а изузеци се дају у засебном делу. Тамо где не постоји правило које може да обухвати изузетке, наводе се све именице које су изузетачне. На пример, правило је да једносложне именице у множини имају интерфикс *-ов/-ев-* (узима се множинска основа, где нема непостојаног *a* и где је ијекавска вредност јата *je*). Наравно, има једносложних именица које немају или не морају имати тај интерфикс, и двосложних који га имају. Онда се у овом делу мора навести:

Једносложне са кратком множином (* значи могућност дуге и кратке множине)

*брк, *цол, црв, чак, дан, *дух, *двор, ђак, град ('сѿејен'), гром, јад, кал, *кам, *крал, мѿ(а)к, мош, мрав, њок, *сан, *скуш, сѿис, *шаш, шѿић, *шрак, шрой, шрѿ(а)к, *шруд, *вук, зуб, *звук, *жал, *жбир, жмул, жрец.

Неизведене двосложне са дугом множином (* значи могућност дуге и кратке множине)

*бусен, *чокањ, *голуб, *камен, *лабуд, *лежај, *округ, *ѿаун, *ѿаук, *сѿечај, *шечај, *шешр(и)еб.

Све изведене именице са једносложном у другом делу, нпр. *ѿа-роброд*, *сѿѿермеч*, итд.

Следило би представљање регионалних алтернација, типа: *цвијетѿ* : *цвѿетѿови*.

Уколико би граматика обухватала и образце мењања акцента, тај део следио би након материјала о коме је била реч.

Треба напоменути и ово. Овде представљени подаци дати су на грађи српскохрватских стандарда узетих у целини. Ако бисмо обрачунали статистику за српски језик у целини, она би одступала од хрватске око 1% у расподели именица на *-а* у номинативу. Наиме, та група била би у српском стандарду већа због разлика типа *архѿшекѿѿ* : *архѿшекѿѿа*.

Закључци

Овде је показано како се применом статистике и кориснички оријентисаног приступа може представити исцрпан и прегледан опис једног веома компликованог флексијског обрасца (именице мушког рода далеко се теже подају опису од оних женског и средњег рода). Успостављање статистички условљених подразумеваних вредности и редоследа представљања може се применити и на бројне друге области, рецимо на представљање глаголских флексија. Ако се глаголи класификују само по инфинитивној основи (што је за корисника знатно јед-

ноставније јер му се каже да у речнику погледа који је задњи вокал инфинитивне основе), онда добијамо овакве вредности.

Образац	Инфинитив	Подразумевани презент		Подразумевани придеви	% свих глагола	Пример подразумеваног обрасца	% подразумевани образац
		1. л. јед.	3. л. мн.				
<i>a</i> -конјугација	а	а	ају	а	59%	причати, причам, причају, испричан	67%
<i>e</i> -конјугација	е	и	е	'е	5%	желети, желим, желе, жељен	74.5%
<i>и</i> -конјугација	и	и	е	'е	28%	радити, радим, раде, израђен	96.5%
<i>у</i> -конјугација	у	е	у	у	6%	куцнути, куцнем, куцну, куцнут	96%
<i>∅</i> -конјугација	∅	е	у	е	2%	помести, пометеи, помету, пометен	100% (са варијацијом у придеву трпном)

На пример, 59% свих глагола припада *a*-конјугацији. Од тих 59%, њих 67% следи подразумеване наставке. Из овога се види да *и*-, *у*-, и *∅*- конјугација готово да и немају изузетака, да је *e*-конјугација релативно малобројна, па већину обраде у настави и сл. треба посветити *a*-конјугацији.

Примена принципа заснованости на статистици и узимања у обзир корисника може се применити и у граматичком опису речничких одредница. Ту се онда за подразумеване вредности не би давале додатне информације. Насупрот томе, такве информације би се изразито истицале код примера који следе изузетачне парадигме.

S u m m a r y*Danko Šipka***STATISTICALLY AND USER ORIENTED GRAMMAR:
THE EXAMPLE OF THE MASCULINE NOMINAL INFLECTION**

The author offers a model of grammatical description which draws upon statistical data and presents inflectional data in a user friendly format. It is argued that this model of grammatical description brings about a considerable enhancement in the acquisition of complex inflectional and other grammatical information. The example of the masculine nominal inflection is used to test the model of the analysis on a highly intricate set of inflectional patterns.