

## КОЛОСТРУКЦИОНА АНАЛИЗА И МОГУЋНОСТИ ЊЕНЕ ПРИМЕНЕ У КОРПУСНИМ ПРОУЧАВАЊИМА СРПСКОГ ЈЕЗИКА<sup>2</sup>

У овом раду прегледног типа даје се осврт на колострукциону анализу, као групу квантитативних (статистичких) метода којима се мери степен привлачења односно одбијања између речи и синтаксичких конструкција у којима се оне јављају. Она се типично користи у оквиру конструкционе граматике, примењује се на материјалу из великих електронских корпуса, а увиди који се добију оваквим начином анализе не само што могу да пруже емпиријску проверу постојећих знања, већ и да та знања неретко значајно допуне увидима који без такве анализе не би били могући. Рад даје преглед три основна типа колострукционе анализе: 1) анализе простих колексема, 2) анализе дистинктивних колексема и 3) анализе коварирајућих колексема, на примерима из енглеског језика, са циљем да поменути облике анализе приближи домаћој стручној јавности и укаже на то како се они потенцијално могу применити и на грађу из српског језика.

*Кључне речи:* колострукциона анализа, конструкциона граматика, српски језик.

### 1. Увод

Рад има следеће циљеве: 1) да представи теоријске и методолошке основе колострукционе анализе, 2) да представи типове колострукционе анализе, 3) да изложи изабране облике примене и домете колострукционе анализе у литератури и 4) да укаже на могуће начине примене колострукционе анализе на грађу из српског језика, посебно у светлу чињенице да је дати облик анализе већ дао значајне теоријске увиде, а да аутору није познато да је примењен на грађу из српског језика.

Колострукциона анализа представља породицу квантитативних (статистичких) корпуснолингвистичких метода за проучавање односа између лексе-

<sup>1</sup> vladan.pavlovic@filfak.ni.ac.rs

<sup>2</sup> Рад је изложен на Међународној научној конференцији *Лексикографија и лексикологија у светлу актуелних проблема* одржаној у организацији Института за српски језик САНУ (Београд), од 28. до 30. октобра 2020. године.

ма и граматичких структура у којима се оне јављају. Њоме се мери степен привлачења и одбијања између појединачних лексема, са једне стране, и синтаксичких конструкција у којима се оне јављају, са друге стране. У великој се мери ослања на строгу квантификацију и инференцијалну статистику, а сам назив *колострукција* представља сливеницу насталу од речи *колокација* и *конструкција*.

Овај облик анализе развили су пре непуне две деценије Штефан Т. Грис и А. Стефанович (STEFANOWITSCH, GRIES 2003; в. и STEFANOWITSCH 2013, *inter alia*). Колострукционом анализом дати аутори су проширили постојеће квантитативне методе у чијем су фокусу колокације, односно групе од две или више речи које се јављају једне с другима у синтагматски типично непрекинутом низу. За разлику од таквог облика анализе, колострукциона анализа прилагођена је истраживању лексичко-граматичког интерфејса.

Методолошки оквир колострукционе анализе представља квантитативна корпусна лингвистика. Другим речима, дати облик анализе се ослања на језичке податке који су заступљени у типично великим балансираним и репрезентативним електронским корпусима. Ти подаци се подвргавају строгој квантификацији и статистичкој обради, са циљем уочавања образаца на масама података. Њен теоријски оквир чине граматичке теорије које: 1) посматрају граматичке структуре различитих нивоа сложености и схематичности као смислене јединице, односно као језичке знакове (тј. јединице у којима долази до упаривања форме и значења), које се као такве битно не разликују од других јединица лексикона, и 2) признају значај фреквенцијске дистрибуције лингвистичких јединица.

Другим речима, колострукциона анализа се примењује у контексту различитих конструкционистичких приступа језику и то оних њихових варијанти које заступају став о постојању дијалектичког односа између менталних језичких структура и конкретне језичке употребе. Један од таквих приступа је голдбергијанска конструкциона граматика (GOLDBERG 1995). Примењује се не само у области истраживања питања која се непосредно тичу системског описа језика и синхронијских лекталних варијација у језику, него и истраживања оних питања која се тичу усвајања првог и другог језика, дијахронијских промена у језику и когнитивне обраде језика. Уз то, њени резултати могу бити од користи и у лексикологији и лексикографији јер омогућују бољи увид у међузависност лексема и синтаксичких конструкција у којима се оне јављају (CROFT, SUTTON 2017).

## 2. Циљ, типови и ток колострукционе анализе

Основни циљ колострукционе анализе јесте установити које лексичке јединице су „типичне“ за дату граматичку конструкцију, тј. (не)појављивање којих лексичких јединица у датој конструкцији је за ту конструкцију статистички значајно.

Колострукциона анализа има три типа, од којих је сваки прилагођен је истраживању специфичних аспеката односа између лексичких јединица и граматичких (пре свега синтаксичких) конструкција у којима се оне јављају. То су: анализа простих колексема (*simple collexeme analysis*), анализа дистинктивних колексема (*distinctive collexeme analysis*) и анализа коварирајућих колексема (*covarying collexeme analysis*).

Анализом *простих* колексема истражује које се лексичке јединице јављају у статистички значајној мери на једном месту у једној граматичкој конструкцији. На пример, може се истраживати који се глаголи јављају у императивној конструкцији, или у конструкцији *keep on V-ing* на месту јединице *V*, у статистички значајној мери.

Анализом *дистинктивних* колексема утврђују се разлике између две перифрастичне конструкције или више њих када су у питању њихове колострукционе преференције. То се чини идентификовањем тога које лексема показују преференцију ка употреби у једној спрам друге перифрастичне конструкције, што даље омогућује идентификацију суптилних разлика између две наизглед семантички идентичне конструкције или више њих. Из тог угла су, поред осталог, у литератури проучаване активна спрам пасивне конструкције, као и конструкција *will V* у односу на конструкцију *be going to V* (упор. *I will finish it soon* и *That's not going to happen*). Алтернативно, овај тип анализе може се користити и у установљавању разлика у употреби формално једне исте синтаксичке или морфосинтаксичке конструкције, на пример конструкције *too ADJ to V* (нпр. *too early to talk about it*) у различитим регионалним или социјалним варијететима језика.

Анализом *коварирајућих* колексема истражује се међузависност између лексичких јединица које се налазе на два различита места у истој синтаксичкој конструкцији. Другим речима, она служи да се њоме идентификују парови речи које се у датој конструкцији заједно јављају у већој мери од очекиване, како би се потом приступило квалитативној анализи тако добијених података. У том смислу је, поред осталог, проучавана конструкција *VN2 into V-ing* (нпр. *He tricked me into drinking decaff*), односно међузависност јављања јединица *V* и *V-ing* у њој.

Сам ток колострукционе анализе састоји се из више корака.

Прво, потребно је одабрати одговарајући велики електронски корпус, посебно с обзиром на чињеницу да је колострукциона анализа, као и други облици статистичке анализе података, управо и намењена анализи маса података. Примери таквих корпуса у енглеском језику јесу корпуси доступни овде: <https://www.english-corpora.org/>, док је такав могућ корпус српског језика доступан на линку <https://www.sketchengine.eu/srwac-serbian-corpus/>.

У другом кораку врши се електронско (аутоматизовано) издвајање примера употребе дате конструкције коришћењем одговарајуће претраживачке синтаксе. У трећем кораку обавља се ручна провера тако издвојених приме-

ра, односно поткорпуса за анализу. У четвртом кораку врши се аутоматско утврђивање фреквенција неопходних за колострукциону анализу, при чему тачан тип податка зависи од типа колострукционе анализе који се обавља. У петом кораку тако прикупљени подаци подвргавају се одговарајућем типу колострукционе анализе, типично у оквиру отвореног програмског језика и окружења за статистичка израчунавања под називом R. Тако добијени резултати укључују три групе података: 1) очекиване фреквенције одговарајуће јединице (типично глагола) у датој синтаксичкој конструкцији у односу на јављање дате јединице у другим контекстима, а на основу корпусних података (применом теста *chi na квадрат*); 2) вредности *p*, које су најчешћи индикатор статистичке вероватноће, који се добијају применом Фишеровог теста, као једног од статистичких тестова. Као што је познато, те вредности могу се кретати између 0 и 1. Што је вредност *p* ближа нули, може се сматрати да је повезаност одређене лексичке јединице (попут глагола), са једне стране, и дате (морфо-)синтаксичке конструкције, са друге стране, већа, односно да је одговарајућа нулта хипотеза нетачна; 3) негативни логаритми са основом 10 за вредности *p* (*индекс јачине колострукционе везе*, у даљем тексту *IJKB*). Што је та вредност виша, то је јачина такве колострукционе везе већа. Напокон, у шестом кораку, врши се квалитативна анализа тако добијених података на основу релевантних лингвистичких теорија и модела.

### 3. Примери употребе колострукционе анализе

У овом одељку најпре ћемо навести примере употребе сва три типа колострукционе анализе на материјалу из енглеског језика. После сваке групе примера даћемо осврт на то како се сличне анализе могу применити и на грађу из српског језика.

#### 3.1 Примери употребе анализе простих колексема

3.1.1 Као први пример анализе простих колексема наводимо пример императивне конструкција (STEFANOWITSCH, GRIES 2003). У питању је конструкција која се практично састоји из једног елемента – глагола, и за коју се типично сматра да има директивну функцију, односно да се користи како би говорник од слушаоца затражио извршење неке радње која је говорнику од значаја.

Међутим, резултати који се добију применом анализе простих колексема на дату конструкцију, указују на то да поменута претпоставка (која је неретко заступљена и у литератури) често није потпуно тачна.

Ранг	Колексема	Учесталост јављања колексема у императивној конструкцији	Вредност $p$	ИЈКВ
1	Let	86	1,99E-97	96,70114692
2	See	171	7,47E-80	79,1266794
3	Look	47	1,18E-24	23,92811799
4	Listen	26	4,05E-23	22,39254498
5	Worry	31	5,18E-22	21,28567024
6	Fold	16	9,25E-22	21,03385827
7	Remember	35	1,83E-18	17,73754891
8	Check	21	2,09E-17	16,67985371
9	Process	15	2,16E-17	16,66554625
10	Try	47	5,13E-17	16,28988263
11	Hang on	17	7,90E-17	16,10237291
12	Tell	46	1,30E-15	14,88605665
13	Note	16	2,96E-15	14,52870829
14	Add	21	2,65E-12	11,57675413
15	Keep	28	1,13E-11	10,94692156

Табела I. 15 колексема са највишим ИЈКВ у императивној конструкцији у енглеском језику

Као што се може видети из Табеле I, 4 од 15 глагола са најјачим ИЈКВ уопште не кодирају никакву радњу већ ментални процес (*see, worry, remember, note*). Највећи број глагола којима се изражава нека радња јесу такви глаголи код којих извршење те радње не доводи до било каквих опипљивих и јасних резултата (*look, listen, hang on, check (it) out, try, keep*). Наравно, има глагола који могу довести до извршења радње са конкретним, физичким резултатима (нпр. *add*, као у примеру *Add this to the mixture*). Међутим, они нису доминантни у мери у којој би се то могло очекивати на основу литературе која се бави овим глаголским начином. Додатно се може приметити да радње означене глаголима јесу пожељне пре свега из слушаочевог, пре него ли говорничког угла, што показује следећи пример: *Just try it and see what happens*. Овакви примери имају и функцију усмеравања пажње слушаоца. Уз то, оваквим примерима изражава се упућивање савета некоем него ли заповест или молба. Преостали глаголи – *fold, check, process, tell* – генерално се могу сматрати каноничким императивима будући да означавају такве радње које доводе до неких конкретних резултата, и то таквих за које није очигледно да на било који директан начин служе примаоцу поруке (*Check the condition of the drive belt periodically*). Глагол *tell* несумњиво има директивну употребу (*Tell him we are waiting for the order*), али исто тако и употребу којом се врши организација дискурса (*Tell us about Barcelona then*).

На основу наведеног произилази да се императив заправо типично *избегава* уз глаголе којима се означава нека конкретна радња (нпр. *do, take* и сл.), вероватно из разлога што се такви изричаји могу сматрати превише директним и наметљивим. Такође произилази да је једна од његових основних функција она која се односи на организацију дискурса и усмеравање нечије пажње на начин који је релативно ненаметљив.

Иако представљени резултати свакако не дају потпуну слику о употреби императивне конструкције у енглеском језику, и иако захтевају додатну проверу, може се рећи да анализа простих колексема, као један од типова колострукционе анализе, заиста може да пружи увид у језичке податке на које се можда не би обратила пажња без оваквог облика анализе.

Добар пример граматике српског језика где се већини оваквих аспеката употребе императива поклања пажња јесте ПИПЕР, АНТОНИЋ и др. (2005: 453–457). Било би корисно сличну анализу спровести и за императив у српском језику и на основу такве анализе по потреби допунити постојеће описе ове конструкције у датој и сличним граматикама.

3.1.2 Најзначајнији резултати који се добијају применом анализе простих колексема на конструкцију *V N2 N3* (нпр. *told her a story*) представљени су у Табели II.

Ранг	Колексема	Учесталост јављања колексема у корпусу	Учесталост јављања колексема у конструкцији <i>V N2 N3</i>	Очекивана учесталост јављања колексема у конструкцији <i>V N2 N3</i>	Вредност <i>p</i>	ИЈКВ
1	give	1091	560	14,57	0	∞
2	tell	792	493	10,58	0	∞
3	send	295	78	3,94	4,13E-076	75,384
4	ask	504	92	5,73	9,65E-074	73,015
5	show	628	84	8,39	5,15E-056	55,288
6	offer	196	54	2,62	3,73E-054	53,428
7	convince	32	23	0,43	1,70E-036	35,770
8	cost	65	23	0,87	9,04E-027	26,044
9	inform	55	20	0,73	9,57E-024	23,019
10	teach	92	23	1,23	7,94E-023	22,100

Табела II. 10 колексема са највишим ИЈКВ у конструкцији *V N2 N3* у енглеском језику

Подаци из Табеле II показују да су глаголске колексема које су најтешње везане за дату конструкцију оне којима се означава трансфер, било физички или метафорички. Иако је поменути чињеница свакако позната и одраније, представљени подаци су значајни јер показују да се колострукционом анализом могу открити високо конвенционализовани односи привлачења између глагола и

конструкција чак и под претпоставком да такви односи између дате две јединице не постоје. Уз то, представљени подаци указују и на значај разликовања између конкретних, сирових, фреквенција, са једне стране, и таквих фреквенција подврнутих статистичком тесту (у конкретном случају – Фишеровом тесту). На пример, може се видети да се глагол *send* у датој конструкцији јавља укупно 295 пута, односно ређе него глагол *ask*, који је јавља 504 пута, као и глагол *show*, који се јавља 628 пута. И поред тога, а на основу датог статистичког поступка, произилази да је управо глагол *send*, пре него ли два преостала глагола, тешње везан за дату конструкцију. Поменуто је у складу са чињеницом да статистички тестови и постоје како би помогли откривању образаца у подацима које иначе не бисмо могли да откријемо ако бисмо се ослањали само на сирове фреквенције. У том смислу, било би корисно сличну анализу извршити како за еквивалентну конструкцију у српском језику (као у примеру *Дао сам му књигу.*), тако и у другим конструкцијама, и том приликом утврдити колико дати тип анализе може да послужи и као емпиријски доказ постојећих теоријских становишта, а колико може да такве ставове и унапреди, понекад чак и оспори.

### 3.2 Примери употребе анализе дистинктивних колексема

3.2.1 Најзначајнији резултати добијени применом анализе дистинктивних колексема перифрастичних конструкција *V N2 N3* (*I gave her a book*), која је горе већ поменута, и *V N3 to N2* (*I gave a book to her*), представљени су у Табели III, и то према STEFANOWITSCH 2013.

Конструкција <i>V N2 N3</i>				Конструкција <i>V N3 to N2</i>			
Ранг	Колексема	Вредност <i>p</i>	ИЈКВ	Ранг	Колексема	Вредност <i>p</i>	ИЈКВ
1	give (461:146)	1,84e-120	119.74	1	bring (7:82)	1,47E-009	8.83
2	tell (128:2)	8,77e-058	57.06	2	play (1:37)	1,46E-006	5.84
3	show (49:15)	8,32e-012	11.08	3	take (12:63)	2,00E-004	3.70
4	offer (43:15)	9,95e-010	9.00	4	pass (2:29)	2,00E-004	3.70
5	cost (20:1)	9,71e-009	8.01	5	make (3:23)	6,80E-003	2.17
6	teach (15:1)	1,49e-006	5.83	6	sell (1:14)	1,39E-002	1.86
7	wish (9:1)	5,99e-004	3.22	7	do (10:40)	1,51E-002	1.82
8	ask (12:4)	1,30e-003	2.89	8	supply (1:12)	2,91E-002	1.54
9	promise (7:1)	3,60e-003	2.44	9	read (1:10)	5,99E-002	1.22
10	deny (8:3)	1,22e-002	1.91	10	hand (5:21)	6,36E-002	1.20

Табела III. 10 најдистинктивнијих колексема у конструкцији *V N2 N3* спрам конструкције *V N3 to N2* у енглеском језику

Као што се може видети, дате перифрастичне конструкције заправо имају сасвим другачије колострукционе профиле. Дистинктивне глаголске колексема за конструкцију *V N2 N3* јесу оне које се односе на физички или метафорички трансфер у ситуацијама када су и агенс (референт имплицитне јединице *N1*) и реципијент (референт јединице *N2*) физички близу једно

другом. У питању су глаголи попут *give, tell, show, offer, teach, wish, ask, deny, award*, као у примеру: *She gave me a book*. Код конструкције *V N3 to N2*, оне се односе на такав трансфер где агенс мора да превали неко растојање или уложи изванредан труд, физички или метафорички, да би се трансфер објекта означеног јединицом *N3* уопште остварио. У питању су глаголи *bring, play, take, pass, sell, do, supply*, као у примеру: *We will definitely pass the queries to the authorities*. На овакво разликовање типично се не скреће пажња у референтним књигама, укључујући ту и педагошке граматике.

3.2.2 У Табели IV приказани су резултати анализе дистинктивних колексема за конструкције *get N2 V-en* (*I must get it done as soon as possible*) и *have N2 V-en* (*I had my car repaired*), и то према GILQUIN 2006.

Конструкција <i>get N2 V-en</i>			Конструкција <i>have N2 V-en</i>		
Ранг	Колексема	ИСКЈ	Ранг	Колексема	ИСКЈ
1	sort out (74)	44,99	1	cut (53)	24,95
2	do (278)	40,17	2	perm (22)	16,12
3	sort (16)	10,43	3	do (183)	13,04
4	finish (17)	8,33	4	build (16)	8,99
5	organise ize (15)	7,85	5	cut off (10)	7,33
6	dress (9)	5,87	6	put down (10)	7,33
7	wash (13)	5,50	7	put (17)	7,09
8	fix (14)	5,14	8	service (8)	4,99
9	book (7)	4,56	9	clean (11)	4,45
10	start (11)	3,08	10	remove (7)	3,73

Табела IV. 10 најдистинктивнијих колексема у конструкцији *get N2 V-en* спрам конструкције *have N2 V-en* у енглеском језику

Упркос чињеници да се дате конструкције уобичајено описују као синонимне, горњи подаци указују на то да оне имају другачије коластрукционе профиле будући да лексичке јединице *V-en* које су дистинктивне за једну конструкцију нису дистинктивне и за другу.

Наиме, јединице *V-en* у конструкцији *get N2 V-en* генерално се односе на организацију (нпр. *sort / sort out, finish, organise|ize*) и свакодневне активности (нпр. *dress, wash, cook*). Такође, горњи подаци указују на то да се она типично користи када имплицитно постоји нека врста напора који треба уложити, или нека тешкоћа коју треба савладати, или када постоји неки рок до када одређену радњу треба обавити (нпр. *We must get it done by Sunday afternoon at 2:00*). Поред тога, датом конструкцијом типично се имплицира да одређену радњу неко ради сâм (*I must get it finished by midnight*).

Са друге стране, када је у питању конструкција *have N2 V-en*, горњи подаци показују да се она типично користи онда када се пружа нека услуга



субјекатском референту (нпр. *have one's hair cut*). Дакле њено основно значење јесте *унајмити некога да пружи неку услугу субјекатском референту*.

Поменуто не искључује могућност да се дате две конструкције могу користити напоре, као синонимне (*He got / had his car fixed yesterday*). Ипак, анализа дистинктивних колексема показује да то јављање глагола *get* у датој ситуацији, иако је могуће, није статистички вероватно. Уз то, када се колексема *get* појави у датој конструкцији са неком од колексема које су дистинктивне за глагол *have*, као што је глагол *build* (односно његов прошли партицип *built*), целом конструкцијом поново се изражава то да се нешто ради уз напор и потешкоће, као и да неретко постоји и неки рок до када дату радњу треба обавити. У референтној литератури, и то посебно оној која се редовно користи и на студијама англистике, на дате разлике практично се не уопште обраћа пажња или се то ради половично.

3.2.3 Наведени примери су значајни у светлу теме овог рада јер могу пружити увид у могуће правце додатног емпиријског проучавања алтернативних конструкција у српском језику. Неке од таквих конструкција могле би бити следеће. Прво, то би могле бити *декомпоноване конструкције* у српском језику и *конструкције без декомпоновања*. Наведени тип анализе, дакле анализа дистинктивних колексема, тако би могла допринети провери и могућем проширењу увида доступних у изворима попут РАДОВАНОВИЋ 1977; ТАНАСИЋ 1995; ЈАКИЋ 2011; ЈАКИЋ, ШЕВА 2011, или СТОЈАНОВИЋ 2006. Друго, то би могле бити *конструкције са перифрастичким суперлативом и синтетичким суперлативом*, где би овај облик анализе могао помоћи у провери и проширењу налаза доступних у изворима попут ЈОСИЈЕВИЋ 2015. Треће, то могу бити случајеви факултативне дистрибуције *да-конструкције* и *конструкције са инфинитивом* (*Можете да уђете / Можете ући*), уз пратеће истраживање употребе сваког од дата два типа перифрастичних конструкција у регионалним варијететима српског језика, што је такође један од могућих видова употребе датог типа анализе, односно датог методолошког поступка. Четврто, на овај начин би се додатно могле проучити *конструкције где се у клаузама финалних реченица користи контранегација и где се не користи*, а обе имају слично (ненегирано) значење (в. КОВАЧЕВИЋ 2015а: 218–220), као у примерима *Пружићемо максимум не бисмо ли победили / да бисмо победили Литванце*. У свим таквим и сличним случајевим алтернативних конструкција у српском језику од значаја би било придржавање *принципа несинонимичности*, који гласи да ако су две конструкције синтаксички различите, онда оне морају бити семантички или прагматички различите (GOLDBERG 1995: 67).

## 3.3 Примери употребе анализе коварирајућих колексема

3.3.1 Најзначајнији резултати примене анализе коварирајућих колексема на конструкцију *It BE ADJ to V* (*It is hard to be a corpus linguist*) дати су у Табели V, и то према HILPERT 2014.

Ранг	Парови колексема (ADJ-V)	Учесталост јављања парова колексема у конструкцији <i>It BE ADJ to V</i>	Очекивана учесталост јављања парова колексема у конструкцији <i>It BE ADJ to V</i>	ИЈКВ
1	Interesting – note	244	12,60	278,86
2	Fair – say	103	3,72	133,29
3	Imporant – remember	152	21,42	105,44
4	True – say	49	1,33	75,72
5	reasonable – assume	55	1,89	66,40
6	Hard – believe	93	12,74	55,89
7	Hard – imagine	102	16,09	54,33
8	Important – note	164	51,31	43,98
9	Unrealistic – expect	23	0,19	43,92
10	Reasonable – suppose	30	0,76	40,76

Табела V. 10 парова колексема ADJ-V са највишим ИЈКВ у конструкцији *It BE ADJ to V* у енглеском језику

Дати резултати показују да се придев *important* јавља у паровима са комуникативним и когнитивним глаголима (*note, remember, realize, stress, compare, recognize*). Придеви који се односе на тежину неке радње или стања (*difficult, easy, hard*), као и придеви *reasonable* и *unreasonable*, јављају са глаголима који се односе на когнитивне процесе и когнитивна стања (*see, forget, imagine, believe, assume, suppose, expect*). Представљени резултати указују и на то да се дата конструкција првенствено користи да уведе нову информацију у дискурс. Наиме, структуре попут *It is interesting to note / fair to say / important to remember / true to say / reasonable to assume*, итд, свакако не носе информацију која представља основни пропозициони садржај и типично су мање наглашене од материјала који након таквих структура следи. Све поменуто заправо је већ описано у литератури (в. KALTENBÖCK 2005) и може да послужи као додатна емпиријска потврда таквог закључка.

Обратимо, међутим, пажњу и на примере *hard to believe, hard to imagine, hard to see, unrealistic to expect, unreasonable to expect*. За њих би се могло рећи да означавају негативан епистемички став према садржају језичког материјала (односно садржају пропозиције) који следи. Овакав увид нигде се не помиње у теоријској литератури везаној за дату конструкцију. Такође, комбинације са глаголима *expect* и *suppose* (*It is unrealistic to expect...*, *It is reasonable to suppose...*), показују да се дата конструкција може користити

и да се њоме реферира на валидност неког става или закључка, о чему се такође не говори у литератури.

У том смислу, резултати анализе коварирајућих колексема свакако такође могу допринети адекватнијој анализи значења и употребе конструкције *It be ADJ to V*, односно допунити постојећа теоријска становишта у вези са њом, што и јесте један од темељних циљева обављања и овог и преостала два облика колострукционе анализе.

У светлу теме овог рада значајно је нагласити да слична конструкција постоји и у српском језику. У питању је *субјекатска клауза уз копулативне прилошке предикате* (ПИПЕР, АЛАНОВИЋ и др. 2018: 149–154). Другим речима у питању је конструкција *ПРИЛОГ (евидентно, упадљиво, разумљиво, нетачно, извесно, логично, значајно, итд.) + БИТИ ДА (СУБЈЕКАТ) + ГЛАГОЛ У ЛИЧНОМ ГЛАГОЛСКОМ ОБЛИКУ + ОСТАЛИ КЛАУЗАЛНИ ЕЛЕМЕНТИ*, као у примеру: *Упадљиво је да он не жели да дође.*

Иако је дата конструкција врло квалитетно описана у поменутом и сличним изворима, било би корисно проверити да ли дата анализа може да пружи додатан увид у њу, нпр. у смислу тога који се парови колексема *прилог – глагол у личном глаголском облику* у њој јављају у статистички значајној мери и шта нам то може рећи о семантици дате конструкције. Можда би таква анализа у примени на дати тип конструкције у српском језику не само потврдила већ присутне теоријске ставове, већ и довела до увида на које се у литератури није обратила довољна пажња, као што и јесте случај са поменутом еквивалентном конструкцијом из енглеског језика.

3.3.2 Други пример употребе анализе коварирајућих колексема који овде наводимо јесте онај који се односи на конструкцију *too ADJ to V* (JENSEN 2014a, 2014b, 2014c), као у примеру *too shy to ask her out*.

У питању је скаларна конструкција, у којој се модификатором *too* изражава висок степен присуства атрибута који се означен елементом *ADJ* (*booster-type degree modifier*, v. Paradis, 2000: 149). За тај висок степен присуства атрибута означен елементом *ADJ* може се сматрати да има такав утицај на ситуацију (радњу или стање) изражену елементом *V* да ту ситуацију онемогућава (Jensen, 2014a, 2014b).

Тај однос онемогућавања / блокаде може бити: (1) „природан“, када нешто физички и објективно може бити рецимо *too big to fit into place* (*превелико да се негде смести*) или *too small to be visible* (*немало да би се видело*), и (2) културно условљен, када се избор пара *ADJ-V* може сматрати таквим да представља рефлекс имплицитних образаца карактеристичних за дату културу.

Примена анализе коварирајућих колексема на дату конструкцију дала је, поред осталог следеће резултате. Придев *young* привлачи когнитивне и евалуативне глаголе (*know, remember, understand, grasp*). Поменуто може да укаже на сило-динамички однос који се може описати као то да млађи уз-

раст, тј. мањи број година, блокира, или бар умањује, интелектуални капацитет. Другим речима, поменуто се може сматрати рефлексивом несвесних културом условљених образаца понашања у вези са годиштем / узрастом у америчкој култури (с обзиром на то да су резултати изведени на основу корпуса који се односи на америчку варијанту енглеског). Придев *polite* привлачи глаголе интеракције и комуникације (нпр. *ask, add, disagree, inquire, comment, question*), што показује да се у појединим контекстима комуникацијски чинови могу схватити као угрожавајући по лице (у прагматичком смислу) говорника у оквиру културног модела учтивости у америчкој култури. Придев *busy* привлачи глаголе интеракције, комуникације и друштвености (*talk, attend, meet, visit*), што указује на то да у америчкој култури постоји тенденција да се друштвени контакти потискују у други план онда када је неко заузет. Употреба придева *macho* у датој конструкцији показује да је висок ниво мачизма у колизији са емоционалношћу, осетљивошћу, ситуацијама у којима долази до компромитовања лица (такође у прагматичком смислу), као и са традиционално женским пословима. Ово, даље, може бити повезано са нерационалним и неконструктивним понашањем онога који се карактерише мачизмом (у поменутом контексту америчке културе), чему сведочи следећи пример: *too macho to see a doctor* (JENSEN 2015: 147).

У светлу циљева овог рада важно је истаћи да функционално сличне конструкције постоје и у српском језику: *Пре-ПРИДЕВ БИТИ да ГЛАГОЛ* (*Премлад је да бисмо му дали тај положај*) и *Превише БИТИ ПРИДЕВ да ГЛАГОЛ* (*Превише је лепа да би ...*). У том смислу сматрамо да би и такве конструкције, односно парове *придев-глагол* у њима, било корисно подврћи наведеном типу анализе како би се проверило да ли би се слична запажања могла применити и на одговарајући материјал из српског језика. Поменуто би се чинило са свешћу да се, шире посматрано, у језичкој употреби могу приметити рефлексиве подлежаћих културом условљених концептуализација (JENSEN 2014с, 2015). Том приликом, корпусни подаци и релеватни статистички методи (као што је то колострукциона анализа уопште, као и анализа коварирајућих колексема посебно) могу да послуже као основа за теоријско-методолошки ригорозно и систематично емпиријско истраживање односа између културе и језика (JENSEN 2015).

#### 4. Закључак

На основу представљеног можемо понудити више закључака. Прво, колострукционом анализом могу се открити високо конвенционализовани односи привлачења између лексичких јединица и синтаксичких конструкција чак и под претпоставком да такви односи између дате две јединице не постоје. Друго, обична дескриптивна статистика неретко није довољно по-

узdana да би се установило да ли су одговарајуће лексичке јединице заиста и статистички значајне за дату конструкцију. У том смислу се колострукционом анализом може унапредити дескриптивна адекватност анализе. Треће, анализа међусобне зависности између конструкција и лексичких елемената који се у њима јављају може осветлити семантику граматичких конструкција. Четврто, ослањање на статистичке методе и строге опсервационе и експерименталне протоколе може помоћи у емпиријској провери постојећих теоријских ставова, што такве ставове не само да може да потврди, већ и да их унапреди, а понекад и доведе у питање. Пето, колострукциона анализа јесте један од облика анализе који захтева примену типично великих репрезентативних електронских корпуса, пре него ли ослањање на интроспекцију, која неретко није довољна (в. КОВАЧЕВИЋ 2015б: 237–239). И шесто, најзначајније у оваквој врсти анализе свакако нису сама израчунавања, него осмишљавање таквог истраживачког питања на које колострукциона анализа може да одговори пружањем релевантних квантитативних података и потом квалитативна (теоријска) обрада таквих података. Све поменуто може да буде од несумњивог значаја и у изучавању српског језика. Овај рад управо је имао за циљ да пружи подстрек таквим изучавањима српског језика у којима ће и колострукциона анализа, као метод статистичке обраде језичких података, наћи своје место, посебно и у светлу изразитог развоја корпусне лингвистике, која такву обраду материјала и природно захтева.

### Цитирана литература

- ЈАКИЋ, Милена и Нада Шева (2011). „Квантитативно испитивање декомпонованог / перифрастичног предиката у српском језику”. *Зборник Матице српске за славистику*, 80, 2011: 117–136.
- ЈАКИЋ, Милена. „О типологији декомпонованих конструкција у српском језику”. *Зборник Матице српске за филологију и лингвистику* LIV(1), 2011: 151–165.
- ЈОСИЈЕВИЋ, Јелена (2015). „Модел перифрастичког суперлатива у српском књижевном језику”. *Philologia Mediana* 7(7), 2015: 319–331.
- КОВАЧЕВИЋ, Милош. *О реченици и њеним члановима*. Јасен: Београд, 2015а.
- КОВАЧЕВИЋ, Милош. *Кроз синтагме и реченице*. Јасен: Београд, 2015б.
- ПИПЕР, Предраг, Ивана Антонић, Владислава Ружић, Срето Танасић, Људмила Поповић, Бранко Тошовић. (2005). *Синтакса савременог српског језика: проста реченица* (у редакцији Милке Ивић). Београд: Институт за српски језик САНУ, Београдска књига; Нови Сад: Матица српска, 2005.
- ПИПЕР, Предраг, Миливој Алановић, Слободан Павловић, Ивана Антонић, Дојчил Војводић, Људмила Поповић, Срето Танасић, Биљана Марић, Марина Николић. *Синтакса сложене реченице у савременом српском језику*. Нови Сад: Матица српска; Београд: Институт за српски језик САНУ, 2018.

- РАДОВАНОВИЋ, Милорад. „Декомпоновање предиката (на примерима из српскохрватског језика)”. *Јужнословенски филолог* XXXIII, 1977: 53–78.
- СТОЈАНОВИЋ, Андреј. (2006). „Декомпоновани предикат у научном стилу српског језика”. *Српски језик* XI (1–2), 2006: 223–239.
- ТАНАСИЋ, Срето. „Декомпоновање глагола и структура просте реченице”. *Јужнословенски филолог* LI, 1995: 157–166.
- CROFT, William and Logan Sutton. “Construction grammar and lexicography”. In: Hanks, Peter and Gilles-Maurice de Schryver (eds). *International Handbook of Modern Lexis and Lexicography*. Springer: Berlin, Heidelberg. 2017, 1–12.
- GILQUIN, Gaëtanelle (2006). “The verb slot in causative constructions. Finding the best fit”. *Constructions* 1(3), 2006: p. 1–46.
- GOLDBERG, Adele E. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press, 1995.
- HILPERT, Martin. “Collostructional analysis: Measuring associations between constructions and lexical elements.” In: Glynn, Dylan and Justyna A. Robinson (Eds.), *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy*, Amsterdam and Philadelphia: John Benjamins Publishing Company, 2014, p. 391–404.
- JENSEN, Kim Ebensgaard. (2014a). “This construction is too hot to handle: A corpus study of an adjectival construction”. In: *Proceedings from 14th Annual Meeting of the Japanese Cognitive Linguistics Association*. Kyoto: The Japanese Cognitive Linguistics Association. 2014a: 743–748.
- JENSEN, Kim Ebensgaard. “Too female to be ruthless and too pregnant to argue: Semantic conflict and resolution in the [too ADJ to V]-construction”. *Suvremena lingvistika* 40(77), 2014b: 1–26.
- JENSEN, Kim Ebensgaard (2014c). “Force-dynamic cultural models in a scalar adjectival construction”. [Paper presentation]. 5th UK Cognitive Linguistics Conference, Lancaster, England, 2014. <[https://vbn.aau.dk/ws/files/206865129/Force\\_dynamic\\_manuscript.pdf](https://vbn.aau.dk/ws/files/206865129/Force_dynamic_manuscript.pdf)>. 25<sup>th</sup> May 2020.
- JENSEN, Kim Ebensgaard. “Inferring cultural models from corpus data: Force-dynamic cultural models reflected in the discursive behavior of a scalar adjectival construction.” *Globe: A Journal of Language, Culture and Communication* 1, 2015: 126–151.
- KALTENBÖCK, Gunther. “It-extraposition in English: A functional view”. *International Journal of Corpus Linguistics* 10(2), 2005: 119–59.
- PARADIS, Carita. “It’s well weird: Degree modifiers of adjectives revisited: The nineties”. In: John M. Kirk (Ed.), *Corpora Galore: Analyses and Techniques in Describing English*. Amsterdam: Rodopi. 2000: 147–160.
- STEFANOWITSCH, Anatol, and Stefan Thomas Gries. “Collostructions: Investigating the interaction between words and constructions.” *International Journal of Corpus Linguistics* 8(2), 2003: 209–243.
- STEFANOWITSCH, Anatol. 2013. “Collostructional Analysis.” In: Hoffman, Thomas, and Graeme Trousdale (eds). *The Oxford Handbook of Construction Grammar*, Oxford: OUP, 290–306.

Vladan O. Pavlović

## COLLOSTRUCTIONAL ANALYSIS AND THE POSSIBILITIES OF ITS USE IN THE STUDY OF SERBIAN

The paper gives an overview of collostructional analysis – a group of quantitative (statistical) methods used for measuring the degree of attraction and repulsion between words and the syntactic constructions they appear in. It is typically used within construction grammar, it gets applied to linguistic material from large electronic corpora, and yields results which need not only help empirically verify the already existing theoretical knowledge, but which can also contribute to such knowledge in a way which would not be possible otherwise. The paper presents an overview of the three types of collostructional analysis (*simple collexeme analysis*, *distinctive collexeme analysis* and *covarying collexeme analysis*) applied to various constructions from English, with the aim of acquainting the professionals in the Serbian language with it, so that similar research can also be undertaken related to the given language as well.

*Key words and phrases:* collostructional analysis, construction grammar, the Serbian language.