



UDK 811.163.42'373.74

Pregledni rad

Primljen 20. 08. 2015.

Prihvaćen za tisak 19. 01. 2016.

Ivana Matas Ivanković

Goranka Blagus Bartolec

Institut za hrvatski jezik i jezikoslovlje, Zagreb

Što nam korpus govori o imenskim skupinama (na primjeru imenice *mreža*)

Imenske skupine razlikuju se po stupnju povezanosti među sastavnicama. Neke se sastavnice slobodno povezuju na opisnoj razini (*velik stol, sun čan dan, slatka kava, hladne ruke, mnogo ljudi*), a druge su leksički povezane u čvrstu svezu kojom se označuje izvanjezični sadržaj koji se ne može izraziti samo jednom riječju (*crna kava, krevet na kat, kreditna kartica, radno mjesto*). Sastavnice imenske skupine primarno se povezuju prema sročnosti s imenicom koju određuju (*velika soba, radno mjesto*), dok se neke imenske skupine temelje na nesročnim sastavnicama koje se prema funkciji s osnovnom imenicom povezuju pridruživanjem ili upravljanjem (*korak naprijed, mnogo ljudi, malo prijatelja, četkica za zube, roba s greškom*). U radu će se, na primjeru imenice *mreža* koja ima bogat sintagmatski i semantički potencijal, pokazati opis imenske skupine na temelju korpusa te će se obrada u rječnicima usporediti s podacima iz korpusa (*hrWaC, Hrvatski nacionalni korpus v3.0*). Analizirat će se računalno dobiven popis kolokata imenice koja je u korpusnoj pretrazi zadana kao čvor te čiji se kolokati razlikuju prema frekvenciji i prema čvrstoći sveze. Korpusnom analizom dobit će se popis imenskih skupina te će se proučiti koliko se on podudara s postojećim leksikografskom opisom i s postojećim teorijskim načelima obrade sveza riječi u hrvatskom jeziku. Cilj je rada utvrditi na koji način korpusna analiza pridonosi leksikografskoj obradi sveza riječi.

Ključne riječi: korpus; hrvatski jezik; kolokacijske sveze; slobodne sveze.

1. Uvod

Leksikografski priručnici hrvatskoga jezika u definicijama, tj. opisu značenja, ali i u navođenju primjera (slobodnih sveza, čvrstih sveza, idioma itd.) nedovoljno ili uopće ne polaze od korpusnih potvrda. Na taj se način u rječnicima ne navode sve značenjske i upotrebne mogućnosti riječi, tj. rječničkih natuknica. Rad se temelji na frekvencijskoj analizi konstrukcija prema podacima dobivenim iz korpusa.¹ Namjera je uputiti na različite značenjske i upotrebne mogućnosti koji se iščitavaju iz frekvencijskih vrijednosti konstrukcija s imenicom kao čvorom. Za analizu je odabrana imenica *mreža* koja ima velik sintagmatski i semantički potencijal na više ra-zina: u doslovnom i prenesenom značenju. U radu će se postojeći rječnički opisi imenice *mreža* usporediti s rezultatima pretrage na korpusu. U drugom dijelu pri-kazat će se dobiveni kolokacijski kandidati te će se na temelju toga opisati sintak-tički potencijal korpusne pretrage te njezin semantički potencijal. U trećem dijelu na temelju dobivenih rezultata prikazat će se uloga korpusa u jezikoslovnim istra-živanjima.

U trima suvremenim rječnicima hrvatskoga jezika, dvama tiskanima (*Školskom rječniku hrvatskoga jezika* (Birtić i dr. 2012), u nastavku ŠKRJ, i *Rječniku hrvatskoga jezika* (2000), u nastavku RHJ) te jednom mrežnom (*Hrvatskome jezi čnom portalu*, <http://hjp.novi-liber.hr>, u nastavku HJP), *mreža* je na prvome mjestu definirana u primarnom značenju iako to znač enje u današnjoj upotrebi više nije dominantno. S obzirom na svoj izgled manjih, uglavnom jednakih, međusobno poveza-nih dijelova *mreža* ima i bogat metaforički potencijal koji proizlazi i iz višestruke namjene konkretnoga predmeta (hvatanje, čuvanje...). U RHJ-u *mreža* je definirana kao „1) pletivo načinjeno od različitoga tvoriva kojemu niti tvore jednaka oka: ~ za ptice, ribarska ~, paukova ~ 2) **grad** a) konstrukcija dobivena preplitanjem linij-skih elemenata vlakana, pruća, armature: željezna ~ b) sustav povezanih linija prometnica ili vodova: električna ~, telefonska ~, cestovna ~, željeznička ~, vodo-vodna ~, kanalizacijska ~ 3) **inf** sustav povezanih računala 4) sustav povezanih je-dinica: trgovinska ~, školska ~, policijska ~”. Izdvojeni su frazemi „paukova ~ zamka, klopka; **stvoriti ~u** organizirati ljude za što; **uloviti u svoju ~u** namamiti; **uplesti se u čiju ~u** postati žrtvom čije spletko, čijih nečasnih nakana”. HJP na pr-

¹ Trenutačno postoje tri pretraživa korpusa hrvatskoga jezika: *hrWaC*, <http://nlp.ffzg.hr/resources/corpora/hrwac> (vidi više Ljubešić – Klubička 2014), *Hrvatski nacionalni korpus*, http://filip.ffzg.hr/bonito2/run.cgi/first_form (vidi više Tadić 2009) i *Hrvatska jezična riznica*, <http://riznica.ihjj.hr/index.hr.html> (vidi više Brozović Rončević – Čavar 2010). S obzirom na to da *Hrvatska jezična riz-nica* nije lematizirana, a *hrWac* je najveći (s ukupno 2 028 788 248 pojavnica u odnosu na 216,8 milijuna pojavnica *Hrvatskoga nacionalnog korpusa*), daljnja pretraga rađena je na *hrWacu*.



vome mjestu donosi definiciju „naprava od različitih materijala ispletena na jednaka oka [*ribarska mreža; teniska (odbojkaška) mreža*]”, zatim „2. *pren.* raširen povezan sistem cesta, kanala itd., ukupnost povezanih ustanova i poduzeća [*vodovodna mreža; električna mreža*] 3. organizacija više izvora podataka ili radnih i informativnih cjelina; sistem, sustav 4. *int. a. razg.* naziv za Internet *b. razg.* naziv za web, sustav za prijenos podataka na Internetu *c.* niz međusobno povezanih računalnih sustava”. Kao frazemi izdvojeni su „**uloviti u svoju mrežu** namamiti koga da radi za čije ciljeve; **uplesti se u čiju mrežu** nepromišljeno se upustiti u čije kombinacije”. U ŠKRJ-u definirana je kao „1. pletivo velikih oka [*odbojkaška ~, ribarska ~*]; 2. *pren.* povezan sustav čega [*električna ~, prometna ~, telefonska ~*]”. Kao ustaljene sveze izdvojene su „~ **tijela mat.** lik sastavljen od stranica tijela položenih u ravninu; **paukova ~ zool.** mreža koju pauk sam izrađuje od niti ljepljive tekućine iz posebne žlijezde, a služi mu za hvatanje kukaca, sin. paučina; **računalna ~ inform.** komunikacijski sustav između dvaju ili više računala ili računalnih sustava”, a kao frazemi „**uloviti (uhvatiti) koga u <svoju> mrežu** staviti koga pod svoj utjecaj; **upasti (uhvatiti se) u čiju mrežu** potpasti pod čiji utjecaj”.

2. Statistička podloga u korpusnoj pretrazi

Statističke metode uvelike se primjenjuju u korpusnoj lingvistici. U suvremenoj je korpusnoj lingvistici ustaljeno nekoliko načina izračunavanja jačine veze između kolokacijskih kandidata i čvora² (više o statističkim metodama vidi u Church i dr. 1991: 4–7). Računalni korpus hrvatskoga jezika *hrWac* i *Hrvatski nacionalni korpus* temelje se na takvim izračunima i donose podatke o frekvenciji, tj. običnoj čestoti pojavljivanja, *T-scoreu*, MI-u (*mutual information*) koji su primjenjivi u široj, izvankorpusnoj analizi dobivenih jezičnih podataka te drugim preciznijim mjerama za posebne izračune kao što je *logDice*. Pritchard (1998: 292) *T-score* opisuje kao mjeru kojom se „mjeri stupanj vjerojatnosti da će se dvije riječi supojaviti u odnosu na njihovo slučajno pojavljivanje”, a MI određuje kao mjeru koja „pokazuje količinu informacije/obavijesti koju jedna riječ pruža o vjerojatnosti supojavlivanja druge riječi i tu vrijednost uspoređuje s utvrđenom čestotom pojavljivanja svake od tih riječi”. Statističke mjere upućuju na različite čvrstoće veza među riječima, od slobodnih sveza preko kolokacija i naziva do frazema. Takvi statistički podatci korisni su u prevođenju i poučavanju jezika, u leksikografskom radu te posebno u proučavanju višerječnih skupina hrvatskoga jezika. U ovom je radu opis rezultata dobive-

² Nazivi *kolokacijski kandidat* i *čvor* prijevodi su naziva *collocation candidate* i *node* koji se primjenjuju u korpusima *hrWac* i *Hrvatski nacionalni korpus*, izrađenima u sustavu NoSketch Engine.



nih korpusnom pretragom temeljen na frekvencijskoj analizi, a naglasak je stavljen na analizu dvorječnih imenskih skupina s imenicom *mreža* kao čvorom.

Budući da imenica *mreža* u rečenici može doći u raznim oblicima, u pretraživanju kolokata koristila se lema, koja pokriva cijelu paradigmu, a ne samo određene oblike. U slučaju pretrage po riječima odvojeno se broje padežni oblici, npr. *društvene* se pojavljuje 5468 puta, *društvenim* 4926 puta, a *društvenih* 4687 puta i oni su po učestalosti kolokata čvora *mreža* redom na 3., 4. i 5. mjestu. Lema nam daje podatak da je *društven* najučestaliji kolokat s lijeve strane imenice *mreža*, neovisno o tome u kojemu se obliku pojavljuje, a pojavljuje se ukupno 21071 put. I *HrWac* i HNK interpunkcijske znakove također prikazuju kao kolokacijske kandidate pa su u daljnjem prikazu interpunkcijski znakovi isključeni iz analize i tabličnoga prikaza, a kao potencijalni korpusni kandidati donose se samo samoznačne i suznačne riječi te kratice i pokrate (G, tv).

2.1. Korpusna pretraga po kolokacijskim kandidatima

Rad je usmjeren na imenske skupine kojima je u funkciji glavne sastavnice (čvora) imenica, a zavisna je sastavnica (kolokacijski kandidat) sročna imenica ili pridjev-na riječ koja obično dolazi ispred glavne sastavnice ili nesročna riječ koja je obično iza glavne sastavnice, pa je poč etno pretraživanje korpusa ograničeno na prvi kolokacijski kandidat s lijeve strane (-1) glavne sastavnice (čvora) i na prvi kolokacijski kandidat s desne strane (+1). Pretragom je obuhvaćeno prvih pedeset kolokacijskih kandidata.³ Preliminarna pretraga pokazala je da se s desne strane često pojavljuju prijedlozi te su u obzir uzete i veze s najčešćim prijedlozima, koje su zatim pretraživane kao čvor (npr. *mreža za* i daljnja pretraga usmjerena je na +1).

Prilog 1. Sročni kolokacijski kandidati imenice *mreža* kao čvora

Query *mreža*, 0 208,948 > GDEX 208,948 (103.0 permillion), pretraga je ograničena na jedan kolokat nalijevo od čvora (s opcijom P.DUPLICATE 0).

³ Korpus je postavljen tako da izlistava po pedeset kolokacijskih kandidata, međutim, broj je navedenih kandidata manji jer su s popisa isključeni pravopisni znakovi.



1. društven	21,071	17. taj	1,444	33. zatresati	856
2. u	14,977	18. plinski	1,372	34. sebe	855
3. na	6,622	19. električan	1,335	35. telekomunikacijski	855
4. vodovodan	3,414	20. tv	1,327	36. televizijski	844
5. biti	2,932	21. komunikacijski	1,254	37. drugi	825
6. i	2,702	22. prodajan	1,219	38. pogoditi	801
7. mobilan	2,656	23. ženski	1,191	39. međunarodan	778
8. svoj	2,469	24. europski	1,085	40. globalan	771
9. bežičan	2,253	25. iz	1,082	41. teroristički	769
10. kanalizacijski	2,242	26. vlastit	1,045	42. hrvatski	741
11. računalan	1,991	27. distribucijski	991	43. nov	723
12. socijalan	1,852	28. preko	961	44. telefonski	715
13. ekološki	1,716	29. čuvar	959	45. vodoopskrban	714
14. lokalan	1,696	30. G	952	46. maloprodajan	680
15. s	1,464	31. dio	951	47. ovaj	678
16. sav	1,449	32. prazan	858	48. fiksna	662

Pregled prvih pedesetak kolokacijskih kandidata pokazuje da slijeva uz imenicu prevladavaju sročni pridjevni kolokati s kojima imenica *mreža* tvori dvije skupine sveza: čvrste sveze (kolokacijske sveze i dvorječni nazivi) i slobodne sveze. Podjela na čvrste i slobodne sveze ustaljena je u hrvatskoj (Pritchard 1998, Blagus Bartolec 2014) i angloameričkoj leksikološkoj literaturi (Meljčuk 2001, Benson i dr. 1997).⁴ Kao kolokacijske sveze među prvih pedesetak kandidata izdvajaju se: *društvena mreža, vodovodna mreža, mobilna mreža, bežična mreža, kanalizacijska mreža, računalna mreža, plinska mreža, električna mreža, tv mreža, komunikacijska mreža, prodajna mreža, ženska mreža*. Kao slobodne sveze izdvajaju se: *svoja mreža, sve mreže, lokalna mreža, ta mreža, europska mreža, vlastita mreža*. Sve navedene sveze upućuju na to da se imenica *mreža* ne pojavljuje u svojemu primarnom značenju, nego u jednom od svojih prenesenih značenja.⁵ U pregledanim

⁴ Pritchard (1998: 287–288) razlikuje slobodne ili otvorene kombinacije riječi koje imaju širok i najslobodniji izbor drugih leksičkih jedinica s beskonačnim mogućnostima njihova stvaranja te kolokacije kao sveze s „više ili manje predvidljivim sintagmatskim supojavlivanjem riječi” s malom mogućnošću zamjene kolokacijskih sastavnica.

⁵ Preneseno značenje sveza *svoja mreža, sve mreže, ta mreža* i *vlastita mreža* iščitava se iz širega konteksta (*Interneta nema. Kod onih ovisnih o mreži svih mreža nastupit će panika...; Vipnet u svojoj mreži ima 50 odašiljača...; Kroz tu mrežu ostvaruje se mogućnost razmjene e-mailova...; Ramos*

rječnicima sekundarna (izvedena) značenja imenice *mreža* uglavnom se određuju kao prenesena (s odrednicom *pren.*), stoga se i u radu ta značenja opisuju kao prenesena.

Među prvih pedeset kandidata koji s imenicom *mreža* tvore svezu izdvaja se nekoliko nesročnih, imenskih i glagolskih, kolokata (*čuvar mreže, dio mreže, zatresati mrežu, pogoditi mrežu*).

Prilog 2. Nesročni kolokacijski kandidati imenice *mreža*

Query *mreža*, 0 208,948 > GDEX 208,948 (103.0 permillion), pretraga je ograničena na jedan kolokat nadesno od čvora (s opcijom P.DUPLICATE 0).

1. i	11,437	18. imati	618	35. javan	362
2. u	8,044	19. ne	611	36. natura	351
3. za	5,834	20. udruga	598	37. zdrav	350
4. koji	5,262	21. kako	592	38. jer	350
5. biti	5,258	22. poput	532	39. taj	346
6. na	3,811	23. gost	519	40. protiv	341
7. s	2,066	24. da	516	41. a	341
8. te	1,978	25. do	474	42. preko	324
9. od	1,504	26. iz	468	43. morati	323
10. ili	1,493	27. po	455	44. trebati	316
11. moći	1,237	28. htjeti	427	45. europski	307
12. facebook	1,222	29. twitter	412	46. carnet	300
13. sebe	953	30. škola	411	47. nov	291
14. kao	885	31. što	403	48. hrvatski	289
15. mlad	866	32. svoj	401	49. oko	289
16. hrvatska	857	33. nakon	393		
17. sav	704	34. pa	367		

Kolokacijski potencijal po frekvenciji ima zamjениčni pridjev *sav* kada dolazi kao dio sintagme *mreža svih mreža* (*Na to još treba dodati i ekonomsku krizu koja tresе gotovo cijeli svijet, mrežu svih mreža, pametne i one malo manje pametne telefone...*), zatim *mreža gostiju* u značenju 'suparnički gol' (...*iz jedne gužve Grubi-*

je zatresao vlastitu mrežu. itd. I u primjerima kao *zatresti mrežu* riječ je o prenesenom značenju imenice *mreža*, metonimiji za gol (ovdje u značenju 'zabiti gol').



šić *trese mrežu gostiju i otvara nadu za osvajanje barem jednog boda.*). Među prvih deset kandidata zdesna nalaze se prijedlozi *u, za, na, s, od*, što upućuje na njihov visok kolokacijski potencijal. Daljnja je pretraga usmjerena na te prijedloge.

Kolokacijski čvor *mreža u* pojavljuje se 8044 puta. Prijedlog *u* uza se vezuje genitiv, akuzativ i lokativ. Pregled korpusno dobivenih kolokacijskih kandidata pokazuje da je većinom riječ o spojevima koji nemaju kolokacijski potencijal kao čvrste sveze, već uglavnom imaju lokativno značenje u slobodnim svezama poblize određujući sadržaj na koji se odnose (*mreža u Hrvatskoj, mreža u naselju, mreža u svijetu, mreža u ulici, mreža u Europi, mreža u SAD-u*).

Na drugom je mjestu po učestalosti pojavljivanja kolokacijski čvor *mreža za* koji se pojavljuje 5834 puta. Velik broj imenica s frekvencijom većom od 20 u vezi s *mreža za* ima značenje namjene i velik broj čine čvrste kolokacijske sveze: *mreža za praćenje (kakvoće zraka / TV programa), mreža za prijenos (informacija i usluga / znanja i iskustava), mreža za razmjenu (informacija), mreža za zaštitu (staništa/okoliša), mreža za zaštitu (od komaraca / od tuče), mreža za distribuciju (električne energije / plina),⁶ mreža za spavanje, mreža za lov, mreža za ležanje*.

Kolokacijski čvor *mreža na* pojavljuje se 3811 puta, ali prema primjerima učestalijih kolokacijskih kandidata vidljivo je da je i čvor *mreža na* nositelj slobodnih sveza s lokativnim značenjem (*mreža na području, mreža na svijetu, mreža na internetu*) koje samo poblize određuju neki sadržaj.

Od 2066 kolokacijskih kandidata čvora *mreža s* među prvih dvadeset po učestalosti nijedan nema kolokacijski potencijal, već je riječ o slučajnim supojavljanjima s opisnim značenjem (*Danas Internet ima izgled paukove mreže s mnogo krajnjih izdanaka*).

Iza mreža od (1504 pojavnica) kao najčešć i kolokacijski kandidat dolazi prijedlog *preko*, zatim *oko* (*Tvrtka Sonik djeluje u tri županije s mrežom od preko 130 prodavaonica., Organizirao je mrežu od oko 5000 suradnika...*).

2.2. Korpusna pretraga po sintaktičkim obilježjima

Pretraga hrvatskih korpusa, osim jednostavnih upita prema zadanoj lemi, riječi ili višerječ noj svezi, omogućuje složenije načine pretraživanja s pomoću regularnih izraza i tagova na temelju kojih se uspostavlja ciljano traženje sintaktičkih struktu-

⁶ S obzirom na to da su primarno u analizu uključeni samo prvi kolokati zdesna ili slijeva od čvora, genitivne dopune pojedinih kolokata stavljene su u zagrade.



ra. Takav pregled pokazuje da u korpusu *hrWac* prevladavaju imenske sveze s temeljnom strukturom pridjev + imenica *mreža*⁷ (zabilježeno je 108 748 pojavnica). Dolaze kao:

čvrste kolokacijske sveze: *društvena mreža, vodovodna mreža, mobilna mreža, bežična mreža, kanalizacijska mreža, računalna mreža, socijalna mreža, ekološka mreža, plinska mreža, električna mreža, komunikacijska mreža, prodajna mreža, maloprodajna mreža, ženska mreža, telekomunikacijska mreža, televizijska mreža, teroristička mreža, vodoopskrbna mreža, telefonska mreža, ribarska mreža, zaštitna mreža, cestovna mreža, niskonaponska mreža, prometna mreža, elektroenergetska mreža...*

slobodne ustaljene sveze: *lokalna mreža, europska mreža, vlastita mreža, distribucijska mreža, prazna mreža, druga mreža, međunarodna mreža, globalna mreža, hrvatska mreža, nova mreža, poslovna mreža, nacionalna mreža, istraživačka mreža, velika mreža, svjetska mreža...*

Pretraga ograničena na lemu *mreža* i imenicu *slijeva* od leme⁸ izdvojila je 31038 pojavnica. Kao najčešći kolokacijski kandidati izdvajaju se nesročne imenice koje tvore slobodne ustaljene sveze: *dio mreže, pristup mreži, čuvar mreže, upravljanje mrežom, širenje mreže, izgradnja mreže, stvaranje mreže, rad mreže, korištenje mreža, održavanje mreža, član mreže, broj mreža, racionalizacija mreže, uspostava mreže ...* Većina je imenskih sveza temeljena na odnosu upravljanja imenice *slijeva*. Oblik imenice *mreža*, uvjetovan rekcijским obilježjima imenice *slijeva*, može biti u genitivu (*širenje mreže*), dativu (*pristup mreži*), instrumentalu (*upravljanje mrežom*). Neke imenice motivirane su glagolom pa je padežni oblik uvjetovan rekcijским obilježjima toga glagola (*pristup mreži*), dok druge primarno otvaraju mjesto imenici u genitivu (*dio mreže*). Lijevo od čvora *mreža* među prvih pedeset kandidata ne pojavljuje se sveza sa sročnom (komplementarnom) imenicom.⁹

Mogućnost izvlačenja ciljanih sintaktičkih konstrukcija pokazuje i njihovu upotrebnu vrijednost. Primjerice, u hrvatskome standardnom jeziku preporučuje se upotreba sročnoga atributa, tj. sveze pridjev + imenica, iako je, na primjerima s imenicom *mreža*, vidljivo da su češće sinonimne skupine s nesročnim atributom, tj. sveze imenica + imenica u genitivu: *mreža korisnika* (93 primjera) / *korisnička mreža* (2 potvrde), *mreža protivnika* (155 potvrda) / *protivnička mreža* (5 potvrda), *mreža potrošača* (9 potvrda) / *potrošačka mreža* (0 potvrda), *mreža računala* (221 potvr-

⁷ Pretraga regularnim izrazom [tag="A.*"] [lemma="mreža"].

⁸ Pretraga regularnim izrazom [tag="N.*"] [lemma="mreža"].

⁹ O svezama s komplementarnom sastavnicom usp. Blagus Bartolec (2014).



da) / *računalna mreža* (155 potvrda) . Iznimka je kolokacijska sveza *mreža žena* ko-ja se pojavljuje 159 puta, a imenica sa sročnim pridjevom *ženska mreža* pojavljuje se 493 puta (u korpusu dominira upotreba imena udruge *Ženska mreža Hrvatske*, a između *mreža žena* i *ženska mreža* postoji i značenjska razlika).

Posve su rijetke imenske sveze s glagolom – među prvih pedeset kandidata slije-va samo su dva glagola u sportskom (nogometnom) žargonu (*zatresti mrežu*, *pogo-diti mrežu*).

Sveze s prijedlogom kao kolokatom (*mreža u*, *mreža na*), unatoč velikoj frekvenciji, nemaju velik kolokacijski potencijal, dolaze kao kontekstna dopuna čvrstoj svezi slijeva (*Ovih je dana započela rekonstrukcija vodovodne mreže u ulici Petra Svačića...*, *Facebook je danas najveća društvena mreža na svijetu...*). Izdvaja se konstrukcija *mreža za* + imenica u akuzativu koja ima velik kolokacijski potencijal u značenju namjene (*mreža za komarce*, *mreža za ležanje*, *mreža za prtljagu* itd.).

Pretraga ograničena na lemu *mreža* i imenicu *zdesna*¹⁰ izdvojila je 33 863 pojavnice, a kao najčešći kolokacijski kandidati pojavljuju se: *mreža stanica*, *mreža odašiljača*, *mreža gostiju*, *mreža ljudi*, *mreža poslovnica*, *mreža domaćina*, *mreža udruga*, *mreža autocesta*, *mreža škola*, *mreža organizacija*, *mreža gradova*, *mreža institucija*, *mreža znanja*, *mreža nasilja*, *mreža kriminala*, *mreža zločina*, *mreža podrške*, *mreža odnosa*... Genitivni oblik imenice *zdesna* obvezna je značenjska dopuna imenici *mreža*, a najčešće se ne može zamijeniti sročnim pridjevom jer bi došlo do promjene značenja (razlikuju se *mreža gradova* i *gradska mreža*) ili je riječ o imenici od koje se ne može tvoriti pridjev (*mreža podrške*). Među prvih pedesetak kandidata desno od čvora ne pojavljuju se sveze sa sročnom (komplementarnom) imenicom. Od svih kandidata desno od imenice *mreža* samo su tri takve potvrde u primarnom značenju imenice *mreža*, s niskom frekvencijom: *mreža stajačica/stajačica* (36 potvrda), *mreža potegača* (36 potvrda), *mreža plivarica* (28 pojavnica).

2.3. Semantički potencijal korpusne pretrage

Podatci dobiveni korpusnom pretragom pomažu nam istaknuti različita značenja imenskih sveza. Neka su značenja doslovna i odnose se na konkretan objekt: *ribarska mreža*, *mreža za prtljagu*, *mreža za ležanje*, *mreža za zaštitu od komaraca*. Ta se značenja u rječnicima donose pod primarnim značenjem natuknice, dok analiza pokazuje da su u današnjoj upotrebi češća prenesena značenja temeljena na pre-

¹⁰ Pretraga regularnim izrazom [lemma="mreža"][tag="N.*"].

dodžbenoj metafori (*image metaphors*), što je uobičajen na čin širenja značenja s prototipnoga na izvedena značenja (usp. Mihaljević i Šarić 1996: 443–444). Predodžba mreže kao pletiva povezanih niti i velikih oka vidljiva je u brojnim kolokacijskim svezama dobivenim korpusnom pretragom u kojima prevladava upravo izvedeno, tj. preneseno značenje imenice *mreža* i koje označuju povezane sustave u kojima su te niti katkad jasnije predočljive (*mreža autocesta, vodovodna mreža, kanalizacijska mreža, računalna mreža*), za razliku od kolokacijskih sveza *društvena mreža, mreža za borbu protiv bolesti, ženska mreža* u kojima je ta predodžba apstraktnija.

Pretraživanje korpusa i dobiveni kolokacijski kandidati omogućuju grupiranje imenskih sveza u semantička polja u kojima imenica *mreža* značenjski potencijal najčešće ostvaruje s pomoću predodžbene metafore:

informacijska i komunikacijska tehnologija: *bežična mreža, računalna mreža, mobilna mreža, TV mreža, komunikacijska mreža, telekomunikacijska mreža, televizijska mreža, telefonska mreža, fiksna mreža, mreža za prijenos znanja/podataka, mreža za razmjenu informacija, mreža za povezivanje uređaja, mreža za pristup internetu, mreža za pristup računalnim datotekama*

opskrba i distribucija: *vodovodna mreža, kanalizacijska mreža, plinska mreža, električna mreža, prodajna mreža, distribucijska mreža, vodoopskrbna mreža, maloprodajna mreža, mreža za distribuciju električne energije, mreža za opskrbu potrošača, mreža za opskrbu električnom energijom, mreža za punjenje električnih automobila, mreža za prodaju cigareta*

predmetno značenje (ne temelji se na predodžbenoj metafori): *mreža za komarce, mreža za spavanje, mreža za lov, mreža za ležanje, mreža za prtljagu*
sportsko značenje: *prazna mreža, zatresti mrežu, pogoditi mrežu, mreža za gol*

prostorno određenje: *lokalna mreža, europska mreža, međunarodna mreža, globalna mreža, hrvatska mreža*

očuvanje prirode: *ekološka mreža, mreža za praćenje kakvoće zraka, mreža za zaštitu staništa / okoliša*

društveno djelovanje: *društvena mreža, socijalna mreža, ženska mreža, mreža za suradnju, mreža za istraživanje, mreža za borbu protiv bolesti.*

Među navedenim skupinama sveza grupiranih prema semantičkim poljima najzastupljenija je prva skupina, što je odraz današnje svakodnevice u kojoj dominira informacijska i komunikacijska tehnologija.



3. Uloga korpusa u jezikoslovnim istraživanjima

S obzirom na neiscrpane mogućnosti korpusne pretrage koju nam omogućuju razvijeni korpusni alati hrvatski su korpusi izvor velikoga broja podataka koji mogu pomoći u leksikografskom radu, leksikološkim istraživanjima te u gramatičkom opisu hrvatskoga jezika, njegova sintaktičkog ustrojstva i upotrebe:

1. Korpusna je pretraga pokazala da se upotreba imenice i njezin opis u rječnicima mogu razlikovati. Na primjeru imenice *mreža* pokazalo se da je ona u primarnom značenju ‘pletivo s jednakim okama’ mnogo rjeđe u upotrebi u odnosu na sekundarna, najčešće prenesena značenja (‘sustav čega’). Prenesena značenja temeljena na metafori danas su posve dominantna i ustaljena. Npr. *društvena mreža* u korpusu sa svim svojim oblicima ima 21 071 pojavnicu, *vodovodna mreža* ima 3414 pojavnica, *mobilna mreža* 2656, a *ribarska mreža* ima samo 614 pojavnica i ne nalazi se među prvih pedeset najfrekventnijih kandidata. U trima pregledanim rječnicima hrvatskoga jezika, primjerice, ne opisuje se značenjski potencijal imenice *mreža* koji se ostvaruje u semantičkom polju društvenoga djelovanja i očuvanja prirode. Iako korpusni rezultati upućuju na slabu zastupljenost primarnoga značenja imenice *mreža*, to bi značenje i dalje trebalo ostati na prvome mjestu u rječničkom članku (zbog toga što se nova značenja temelje upravo na tome primarnom značenju), ali bi ostali korpusni pokazatelji svakako trebali ući u leksikografski opis. Takvu bi opisu pridonijeli mrežni rječnici hrvatskoga jezika u kojima bi se podatci mogli nadograđivati i redovito posuvremenjivati jer je to u tiskanim izdanjima nemoguće.
2. Korpusi mogu poslužiti za relativno jednostavno i brzo izvlačenje kolokacijskih sveza. Među najfrekventnijim kolokacijskim kandidatima imenice *mreža* kao čvora prevladavaju oni koji s njom tvore čvrstu kolokacijsku svezu. To se posebno odnosi na pretragu usmjerenu na prvo mjesto s lijeve strane čvora *mreža*, koja nam daje sročne atribute i velik broj kolokacija, kao što su *društvena mreža*, *vodovodna mreža*, *mobilna mreža*, *bežična mreža*¹¹, *kanalizacijska mreža*, *računalna mreža*, *socijalna mreža*, *ekološka mreža*, *plinska mreža*, *električna mreža*, *komunikacijska mreža*, *prodajna mreža*, *ženska mreža*, *telekomunikacijska mreža*, *televizijska mreža*, *teroristička mreža*, *vodo-opskrbljena mreža*, *telefonska mreža*, *ribarska mreža*, *zaštitna mreža*, *cestovna mreža*, *niskonaponska mreža*, *prometna mreža*, *elektroenergetska mreža*.

¹¹ Sveza *bežična mreža* granični je primjer jer se, s obzirom na veću mogućnost kombiniranja, npr. *bežična računalna mreža*, *bežična telefonska mreža*, može smatrati i slobodnom svezom.



Također su prema korpusnim rezultatima veoma plodne slobodne ustaljene sveze: *lokalna mreža, sve mreže, europska mreža, vlastita mreža, distribucijska mreža, prazna mreža, druga mreža, međunarodna mreža, globalna mreža, hrvatska mreža, nova mreža, maloprodajna mreža, poslovna mreža, nacionalna mreža, istraživačka mreža, velika mreža, svjetska mreža.*

3. Korpus omogućuje i pretragu sintaktičkih struktura koja je u primjeru s imenicom *mreža* pokazala da dominiraju sveze sa sročnim atributom (sveze s temeljnom strukturom pridjev + imenica čine 108 748 pojavnica). Pretraga čvora *mreža* s drugom imenicom neposredno slijeva ili zdesna od nje pokazala je da se pojavljuju u podjednakom broju. Imenice desno od čvora *mreža* dolaze u genitivu i najčešće se ne mogu zamijeniti pridjevom: *mreža oda-šiljača, mreža gostiju, mreža ljudi, mreža poslovnica, mreža domaćina, mreža udruga, mreža autocesta, mreža škola, mreža organizacija, mreža grado-va, mreža institucija, mreža znanja, mreža nasilja, mreža kriminala, mreža zločina, mreža podrške, mreža odnosa.* Neke se, pak, mogu zamijeniti sročnim atributom, ali su u upotrebi češće s nesročnim, imeničnim atributom, npr. *mreža korisnika* (93 potvrde) / *korisnička mreža* (2 potvrde). Imenica *mreža* može doći s desne strane i tada njezin oblik ovisi o imenici koju određuje (*upravljanje mrežom*).

Kolokacijske sveze sa sročnom (komplementarnom) sastavnicom s desne strane pojavljuju se samo u primarnom značenju imenice *mreža* (*mreža stajačica/stajačica* 36 pojavnica, *mreža potegača* 36 pojavnica, *mreža plivarica* 28).

U pretraživanju desne strane nametnula se potreba za pregledavanjem sveza s prijedložno-padežnim izrazom, no pokazalo se da većina primjera ima lokativno ili opisno značenje, jedino *mreža za* + imenica u akuzativu u značenju namjene ima izražen kolokacijski potencijal. Razlog je vjerojatno u tome što je prostorno značenje previše općenito da bi moglo biti distinktivno obilježje u stvaranju vezanih skupina. Prema korpusnim rezultatima mogu se izdvojiti:

- prijedložne kolokacijske sveze bez dopune: *mreža za komarce, mreža za spavanje, mreža za ležanje, mreža za prtljagu, mreža za gol* itd.
- prijedložne kolokacijske sveze s dopunom kojom se označuje hiponimni odnos prema općem hiperonimu: *mreža za povezivanje uređaja / s kućnim pametnim uređajima, mreža za borbu protiv prijevара/bolesti/siromaštva, mreža za prodaju cigareta / naftnih derivata,*



mreža za razmjenu informacija/datoteka/video/videouradaka, mreža za prijenos znanja/podataka/plina.

Iako glagolske skupine nisu bile cilj ovoga rada, pretragom po korpusu izdvojene su i sveze glagola s imenicom *mreža*: *zatresti mrežu, pogoditi mrežu.*

4. Zaključak

U radu je naglasak stavljen na primjenu korpusa u istraživanju imenskih skupina. Kao ogledna uzeta je imenica *mreža* s velikim značenjskim i kolokacijskim potencijalom. U prvom je planu bila analiza kolokacijskih kandidata po frekvenciji. Na imenici *mreža* pokazalo se da je korpus dobar izvor za opis semantičkih i sintaktičkih obilježja imenske riječi i. Dobiveni kolokacijski kandidati mogu se proučavati sa semantič koga aspekta koji pokazuje da je u upotrebi raspored značenja drugačiji od onoga u rječnicima. Rječnici tradicionalno polaze od osnovnoga značenja, dok u upotrebi dominiraju različita prenesena značenja koja se mogu podijeliti u različita semantička polja prema domenama ljudskoga djelovanja. Najviše ih pripada in-formacijskim i komunikacijskim tehnologijama, pa je korpus pokazatelj utjecaja izvanjezične stvarnosti na jezična kretanja, odnosno preslikavanja suvremenih društvenih trendova na jezik. Korpus također nudi i pretraživanje različitih sintak-tičkih obilježja neke riječi ili skupine. Upotrebom regularnih izraza te se mogućnosti povećavaju, omogućujući vrlo brzu ciljanu pretragu određenoga sintaktičkog uzorka hrvatskoga jezika.

Literatura

- Benson, Morton; Benson, Evelyn; Ilson, Robert. 1997. *The BBI dictionary of English word combinations*. John Benjamins Publishing Co. Amsterdam –Philadelphia.
- Birtić, Matea; Blagus Bartolec, Goranka; Hudeček, Lana; Jojić, Ljiljana; Kovačević, Barbara; Lewis, Kristian; Matas Ivanković, Ivana; Mihaljević, Milica; Miloš, Irena; Ramadanović, Ermina; Vidović, Domagoj. 2012. *Školski rječnik hrvatskoga jezika*. Zagreb: Institut za hrvatski jezik i jezikoslovlje; Školska knjiga.
- Blagus Bartolec, Goranka. 2014. *Riječi i njihovi susjedi: Kolokacijske sveze u hrvatskom jeziku*. Zagreb: Institut za hrvatski jezik i jezikoslovlje.
- Brozović Rončević, Dunja; Čavar, Damir. 2010. Riznica: the Croatian language corpus. *Prace filologiczne* 63. 51–65. <http://riznica.ihj.hr/CLC-Slavicorp.pdf>.
- Church, Kenneth; Gale, William; Hanks, Patrick; Hindle, Donald. 1991. Using statistics in lexical analysis. In Zeri, Urnik (ur.), *Lexical acquisition: Using on-line resources to build a lexicon*, 115–164. Lawrence Erlbaum.



Hrvatski jezični portal. <http://hjp.novi-liber.hr/>.

Ljubešić, Nikola; Klubička, Filip. 2014. {bs,hr,sr} WaC – Web corpora of Bosnian, Croatian and Serbian. U Bildhauer, Felix; Schäfer, Roland (ur.), *Proceedings of the 9th Web as Corpus Workshop (WaC-9)*, 29–35. Gothenburg: Association for Computational Linguistics.

Meljčuk (=Mel'čuk), Igor. 2001. Collocations and lexical functions. U Cowie, A.P. (ur.), *Phraseology theory, analysis, and applications*, 23–53. Oxford University Press. Oxford – New York.

Mihaljević, Milica, Šarić, Ljiljana. 1996. Metaforizacija kao terminološki postupak u engleskom i hrvatskom. *Suvremena lingvistika* 41–42. 437–449.

Pritchard, Boris. 1997. O kolokacijskom potencijalu rječničkog korpusa. *Filologija* 30–31. 285–304.

Šonje, Jure (ur.). 2000. *Rječnik hrvatskoga jezika*. Zagreb: Leksikografski zavod Miroslav Krleža; Školska knjiga.

Tadić, Marko. 2009. New version of the Croatian National Corpus. U Hlaváčková, Dana; Horák, Aleš; Osolsobě, Klara; Rychlý, Pavel (ur.), *After half a century of Slavonic natural language processing*, 199–205. Brno: Masaryk University.

Adrese autora:

Ivana Matas Ivanković
Institut za hrvatski jezik i jezikoslovlje
Republike Austrije 16
10000 Zagreb
E-mail: imatas@ihjj.hr

Goranka Blagus Bartolec
Institut za hrvatski jezik i jezikoslovlje
Republike Austrije 16
10000 Zagreb
E-mail: gblagus@ihjj.hr

WHAT CAN CORPUS TELL US ABOUT NOUN PHRASES (ON THE EXAMPLE OF THE NOUN *MREŽA*)

Noun phrases in Croatian can differ in the degree of correlation between its constituents. Some constituents form a descriptive free word combinations (*velik stol* 'large table', *sunčan dan* 'sunny day', *slatka kava* 'sweet coffee', *hladne ruke* 'cold hands'), while others form multiword units which concretize extra-linguistic content that can not be expressed in one word (*crna kava* 'black coffee', *krebet na kat* 'bunk bed', *kreditna kartica*



'credit card', *radno mjesto* 'workplace'). Dependent constituents can be adjectives, which are congruent with a noun (*velika soba* 'big room', *radno mjesto* 'working place'), or they can be adverb phrase or prepositional phrase (*korak naprijed* 'step ahead', *mnogo ljudi* 'many people', *malo prijatelja* 'a few friends', *četkica za zube* 'toothbrush', *roba s greškom* 'faulty good'). This paper will analyze noun *mreža* (with reach syntagmatic and semantic potential) and its co-occurrences – they can either form a collocation or a free combination of words. The lexicographic description will be compared with the corpus-data. The analyses will take into consideration a list of computationally obtained collocates (*collocation candidates*) of a node noun. The frequency and the strength between the words occurring within a particular span can differ. The list of collocates obtained from the corpus will be taken into account and we will examine how it coincides with the existing lexicographic description and with theoretical principles of word combination interpretations in Croatian. The aim of the study is to determine how the corpus analysis can improve the treatment of word-combination entries in lexicographic work.

Key words: corpus; Croatian; collocation; free combination.