

Ivana Filipović Petrović

Zavod za lingvistička istraživanja

Hrvatska akademija znanosti i umjetnosti

Ulica književnika Ante Kovačića 5, HR-10000 Zagreb

ifilipovic@hazu.hr

## KAKO IGGYJU REĆI POP, A DYLANU BOB: IZ KORPUSNIH ISTRAŽIVANJA FRAZEMA

O računalnim se korpusima od njihove pojave govori kao o korisnim alatima kojima se služe lingvisti, a do danas su gotovo nezaobilazni u suvremenim istraživanjima većine lingvističkih poddisciplina. Jedna je od njih i frazeologija. Među temeljnim obilježjima frazema smatraju se konvencionaliziranost, (ne)promjenjivost i desemantiziranost, a da bi se ta obilježja mogla sustavnije proučavati i stupnjevati, potreban je ne samo veliki broj primjera upotrebe frazema u jeziku, nego i alat s opcijama koje daju različite statističke podatke, poput onih o frekvenciji pojavljivanja i karakterističnosti pojedine višerječnice te najčešćoj okolini, ali i brzo i pregledno dostupan širi kontekst u kojem se frazemi javljaju. U ovome radu istražuju se obilježja i ponašanje frazema u najvećem hrvatskom računalnom korpusu, mrežnome korpusu hrWaC. Riječ je o korpusnim istraživanjima koja se provode u sklopu izrade *Frazeološkoga rječnika hrvatskoga jezika*. Rezultatima pokazanima u ovome radu cilj je utvrditi osobitosti frazema koje koriste govornici hrvatskoga jezika vezane uz njihovu promjenjivost, značenja i upotrebu, čime će se pridonijeti potpunijanju i detaljnijem poznavanju hrvatske frazeologije.

### 1. Uvod

Korpusna je lingvistika nesumnjivo zaslužna za sve brojnija frazeološka istraživanja u mnogim jezicima iz dvaju razloga: računalni su korpusi pružili mogućnost detaljnog istraživanja frazema dajući veliki broj primjera upotrebe, kao i podatke o frekvenciji, a dobiveni podaci iz korpusa pokazali su da frazeologija prožima jezik u upotrebi, što je potaknulo daljnja istraživanja (usp. *Corpas Pastor–Mitkov 2019: v*). Spoznaje o fra-

zemima nastale korpusnim istraživanjima ponovo su otvorile pitanja o temeljnim obilježjima frazema, konkretno ukazale su na potrebu njihova revidiranja. Obilježja poput čvrste strukture, okamenjenosti oblika i desemantiziranosti koja su definirana u začecima frazeologije kao lingvističke discipline (v. npr. Fleischer 1997; Menac 2007) doživjela su preispitivanje temeljeno na jezičnim dokazima iz korpusa o stvarnoj upotrebi. Primjerice, prema modificiranom poimanju desemantizacije, riječ je samo o određenom stupnju odmaka od prototipnog značenja leksema (Stanojević–Parizoska 2005:706), odnosno o svojevrsnom proširenju toga značenja. Također, pokazalo se da su varijante frazema ustaljene, sistematične i zapravo vrlo česte, a mogu se objasniti pomoću kognitivnih procesa, baš kao i modifikacije, namjerne kreativne preoblake frazema u svrhu postizanja određenog stilističkog efekta. Kao rezultat svega toga, o frazemima se može govoriti kao o konvencionaliziranim višerječnim svezama koje se reproduciraju kao cjelina, relativno su promjenjive i moguće ih je semantički interpretirati (usp. Omazić 2003:29), a frazeologija je počela pokrivati veći broj višerječnica od kojih su mnoge varijantne i semantički prozirne (v. Granger–Paquot 2008:27). Osim na frazeologiju, korpusna su istraživanja utjecala na još jednu disciplinu, s njom usko povezanu – leksikografiju. Korpusna je revolucija uopće znatno promijenila leksikografiju, konkretno načine sastavljanja rječnika i shvaćanje uloge rječnika u društvu, a te su promjene dotaknule i leksikografsku obradu frazema.

Korpusna istraživanja provedena u ovome radu pokazuju vrstu podataka koje o frazemima daje računalni korpus uz pomoć specifičnih načina pretrage i opcija alata u kojem se korpus nalazi. Pogledajmo sljedeći primjer:

(1) *Nikada dosad nisam osobito obraćao pozornost na potrebe, jer su oni u Japanu kao dobar dan, ali sada je bilo posve drukčije.*

Riječ je o jednome od mnogih primjera upotrebe frazema *biti kao dobar dan* u značenju ‘nešto posve normalno, uobičajeno’ koji se pokazao vrlo frekventnim u najvećem računalnom korpusu hrvatskog jezika, mrežnom korpusu hrWaC. Njegovo je značenje transparentno na temelju doslovnih značenja sastavnica: zahvaljujući našem konvencionalnom znanju o svijetu o tome da je »dobar dan« uobičajen, čest i jednostavan pozdrav koji svakodnevno upotrebljavamo u komunikaciji, možemo razumjeti značenje frazema. Nadalje, rezultati su pokazali da se taj frazem javlja i u obliku s rekcijom:

(2) *Nisam od onih kojima je psovka kao dobar dan*  
te u obliku *doći kao dobar dan* komu:

(3) *Marš gradom praćen je urlikom "pjesama" voljenog kluba uz pokoji standardni rasističko-šovinistički usklik, koji "navijačima" dođe kao dobar dan.*

Drugim riječima, riječ je o varijantnome frazemu. Međutim, postojeći hrvatski frazeološki, ali i opći rječnici, kao ni *Baza frazema hrvatskoga jezika*,<sup>1</sup> ne bilježe taj frazem. Korpus je u ovom slučaju otkrio postojanje i frekventnost upotrebe toga frazema, primjeri upotrebe olakšali su razumijevanje njegova značenja, a sistematično javljanje varijantnih oblika ukazalo je na njegovu promjenjivost. Osim za frazeologiju, ti su podaci važni za leksikografiju i usmjerit će rječničku obradu toga frazema kako bi se u konačnici korisniku dala potpuna informacija o frazemu.

Cilj je ovoga rada na primjerima rezultata istraživanja frazema u hrWaCu donijeti zaključke o temeljnim obilježjima i ponašanju frazema u hrvatskome jeziku: o njihovoj konvencionaliziranosti, varijantnosti, semantičkoj (ne)prozirnosti, ali i drugima, poput specifične upotrebe i višeznačnosti te na koncu otkriti nove ustaljene frazeme.

U nastavku rada najprije se govori o ulozi računalnih korpusa u frazeološkim istraživanjima te o metodama upotrebe korpusa, a zatim se opisuje teorijski okvir i metodologija istraživanja provedenih za ovaj rad. Nakon toga slijede rezultati podijeljeni u pet skupina prema vrsti informacije koju o frazemu saznajemo i na kraju se nalazi zaključak.

## 2. Korpusi u frazeološkim istraživanjima

Korpusni pristup empirijski je pristup jezičnim podacima, odnosno ispitivanje i zaključivanje o jeziku na temelju njegove stvarne upotrebe (Biber–Conrad–Reppen 1998; Tognini-Bonelli 2001; Meyer 2002; McEnery–Gabrielatos 2006). Osim uvida u stvarnu upotrebu, veliki računalni korpusi donijeli su brzo dostupnu veliku količinu podataka i podatke o frekvenciji upotrebe koja je važna za zaključke o konvencionaliziranosti izraza, posebno višerječnih sveza s figurativnim značenjem kao što su frazemi.

Počeci korpusnog pristupa u frazeologiji vežu se uz Johna Sinclaira i njegov projekt Cobuild u okviru kojeg se izrađivao rječnik utemeljen na podacima iz korpusa *Bank of English*, prvi put objavljen 1987. Sinclairov je pristup polazio od korpusa u kojemu se na velikom broju primjera utvrđivala najčešća okolina riječi i figurativna značenja u kontekstu te se tako dolazilo do frazema. Taj se pristup naziva korpusom vođen (engl. *corpus-driven*), dok je drukčiji pristup nazvan korpusno utemeljen (engl. *corpus-based* i *corpus-illustrated*), a podrazumijeva konzultiranje korpusa radi pri-

<sup>1</sup> <http://frazemi.ihj.hr/>

mjera ili utvrđivanja da se riječ koristi, ali prema kriterijima određene lingvističke teorije i jezične intuicije pojedinca (v. više Tognini-Bognelli 2001).

U počecima korpusne lingvistike, kritike su upućivane prema nereprezentativnosti korpusa s obzirom na njihovu veličinu. Smatralo se, primjerice, da prvi korpusi engleskog jezika, poput korpusa Brown koji je sadržavao milijun riječi, ne mogu dati statistički značajne podatke o ponašanju riječi i njihovoj okolini jer su premali (Hanks 2009:214). I u frazeološkim istraživanjima, da bi se pristupilo sistematično potrebno je imati korpus što je moguće veći (v. Sinclair 2004). Što je veći korpus, mehanizmi njegova pretraživanja imaju veće šanse pokazati pravilnosti u upotrebi, što je posebno važno za frazeološke varijante. Ako se opis sličnog događaja ponavlja uz određene varijacije, tada što se češće ponavlja, lakše je uvidjeti pravilnosti i ustanoviti konvencionalizirane varijante. Danas kada se riječi u suvremenim korpusima broje u milijardama, takva veličina korpusa više ne dovodi u pitanje njegovu reprezentativnost.

Kada je riječ o korpusnim istraživanjima, raspravljalo se o još nekim značajnim pitanjima. Naprimjer, ako su korpusi zbirke tekstova, može se postaviti pitanje u čemu je zapravo razlika između korištenja tekstova za pronalaženje frazema, kao što su se primjerice u leksikografiji prije korpusa (ali i poslije korpusa) za rječnike frazema koristili razni skupljeni tekstovi, poput književnih djela, novina, časopisa i raznih pisanih tekstova na mreži, i korištenja računalnog korpusa. O tome je raspravljao začetnik korpusnoga pristupa u frazeologiji John Sinclair. Tekst je, navodi Sinclair, čak i onaj prilično dugačak, nešto što se može detaljno proučiti, analizirati i u potpunosti raščlaniti. Onaj tko analizira tekst ima potpunu kontrolu nad njime: vidi njegov početak, sredinu i kraj, jedinstvo cjeline. Na taj se način proučava tekst, a kada se računalno sastavi kolekcija od više (mnogo više) tekstova, tada to prerasta u korpus i počinje zaokret u metodološkom pristupu. Početak, sredina, kraj korpusa, kao i poredak tekstova u korpusu u pravilu su proizvoljni. Štoviše, cijeli smisao proučavanja jezika u korpusu, a ne u pojedinačnim tekstovima, leži u ostvarivanju mogućnosti da se primijete i promatraju jezične pojave koje su inače suviše udaljene da bi se uočile bez konkordancije, ili su izrazito frekventne ili nisu uopće frekventne, ili napokon, nisu ni vidljive bez određenih statističkih pokazatelja (Sinclair 2004:188–189). U korpusu koristimo alate za promatranje jezika, poput konkordancije, kolokacija, filtera i drugo koji pružaju razinu tumačenja drukčiju od one koja se zbiva kada gledamo u cjeloviti tekst i čitamo ga određenim redom. Stoga ključna razlika dvaju pristupa nije u količini jezika koji se proučava niti u vrsti tekstova i njihovu sadržaju – nego u metodologiji, dakle načinu na koji pristupamo jeziku. Dakako, uz napo-

menu da i pritom postoji više pristupa, kao što su već spomenuti korpusom vođen i korpusno utemeljen pristup, koji utječu na konačne rezultate.

Napokon, za jezične pojave poput frazema za koje su iznimno važna dva faktora: kontekst i okolina u kojoj se javljaju te veliki broj primjera upotrebe kako bi se mogle uočavati pravilnosti i ponavljanja, računalni su korpusi pouzdani, plodonosni i – pokazuju to suvremene frazeološke publikacije<sup>2</sup> – neizostavni u njihovim istraživanjima.

### 3. Frazemi u korpusu hrWaC

Frazeološka istraživanja u hrvatskom mrežnom korpusu hrWaC provode se kontinuirano u sklopu izrade *Frazeološkoga rječnika hrvatskoga jezika*.<sup>3</sup> Jezičnim se podacima pristupa prema načelima korpusom vođenog pristupa, što znači da se ne pretpostavljaju unaprijed oblici frazema, nego se pomoću opcija alata za pretragu korpusa Sketch Engine kao što su skica riječi, filter i kolokacije pronalaze frazemi i mnogi podaci povezani s njima poput frekventnosti oblika, specifične upotrebe, varijanata i drugo. Polazište za dobivanje frazema tisuću je najfrekventnijih imenica i glagola u korpusu koji se pretražuju pomoću spomenutih funkcija u Sketch Engineu, a namjera je i proširivati polazišne vrste riječi. Rječnik u izradi trenutačno broji pet stotina frazema, a korpusna istraživanja frazema nastavljaju se. U ovome se radu pokazuju pretraživanja i rezultati koji umnogome utječu na sliku suvremenog hrvatskog frazeološkog fonda, kao i na teorijska promišljanja o svojstvima frazema. Štoviše, pokazuju se spoznaje o frazemima koje postojeći frazeološki rječnici ne bilježe, a potrebno ih je uključiti u buduće frazeološke rječnike kako bi korisnicima pružili potpunije i detaljnije informacije o frazemima u upotrebi.

U nastavku izdvajamo rezultate podijeljene u kategorije prema vrstama informacija koje o frazemima saznajemo: varijante, modifikacije, novi konvencionalizirani frazemi, frazemi s više značenja i specifična upotreba. Dakako, iz obzira prema prostoru i opsegu rada, odabrane su kategorije i primjeri koji pokazuju novosti i specifičnosti u ponašanju frazema koje nije moguće utvrditi bez istraživanja u računalnom korpusu.

#### 3.1. Varijante

Najprije ćemo se osvrnuti na varijantnost frazema ponajviše zato što je taj segment frazeologije možda najjače *uzdrman* spoznajama iz korpusnih istraživanja. Frazeološka istraživanja u više jezika provedena na veli-

<sup>2</sup> Poput npr. Mitkov (2017); Corpas Pastor–Mitkov (2019).

<sup>3</sup> frazeoloski-rjecnik.eu

kim računalnim korpusima (npr. Moon 1998; Cignoni–Coffey–Moon 1999; Langlotz 2006; Omazić 2015, Filipović Petrović–Parizoska 2017; Parizoska 2019) pokazala su da je broj frazema koji se javljaju u istom obliku razmjerno mali te da jezik u upotrebi zapravo prožimaju konvencionalizirane i nekonvencionalizirane varijante, odnosno modifikacije. Nadalje, drukčiji pristup varijantnosti od onog koji je ponudila tradicionalna frazeologija (v. Menac 2007 i usporedbu dvaju pristupa u Parizoska–Filipović Petrović 2020) naći će se u kognitivnoj lingvistici. Prema tom pristupu, frazeološke varijante opisuju specifičnu situaciju i sudionike koji su dio nje, dakle imaju istu konceptualnu osnovu koja može biti realizirana u obliku različitih predodžbi. To se u jeziku odražava promjenama leksičko-gramatičkog ustrojstva frazema. Zbog toga među varijantnim oblicima mogu postojati određene razlike u leksičkom sastavu, strukturi, značenju i upotrebi. Na takvom se shvaćanju varijantnosti temelje i primjeri izdvojeni u ovome radu.

Podaci iz mrežnog korpusa hrvatskoga jezika pokazali su da se frazemi javljaju u nekoliko vrsta varijanata: leksičkim, leksičko-sintaktičkim, gramatičkim i tvorbenim. Primjerice, pojedini frazem može imati oblik glagolskog i imenskog spoja riječi (*boriti se s vjetrenjačama* i *borba s vjetrenjačama*), zatim kod poredbenih frazema<sup>4</sup> kao *tertium comparationis* javljaju se pridjev i glagol (*miran kao bubica* i *mirovati kao bubica*), a neki se glagolski frazemi javljaju u više različitih leksičko-sintaktičkih oblika (*upasti, ući u kolotečinu; vratiti se u kolotečinu; izaći, ispasti iz kolotečine; izbaciti iz kolotečine koga*). Istraživanje frazeoloških varijanata u korpusu hrWaC detaljno je opisano u Filipović Petrović–Parizoska (2020), a ovdje izdvajamo nekoliko istaknutih primjera.

Među brojnijim varijantama pokazale su se leksičke varijante. Sljedeći primjeri pokazuju promjenjivost frazema *bježati, trčati koliko noge nose koga*:

- (4) *U nekom trenutku primjećujem da povećavam razliku i na kraju ih gubim iz vida. Trčao sam koliko me noge nose.*
- (5) *Šakal je brzo odgurnuo kamen, a mala antilopa mu je bacila prašinu u oči, iskočila iz šupljine i počela je bježati koliko su je noge nosile.*
- (6) *Stoga, braćo, noktom u ledinu, pa iz ove grebene države što te noge nose.*
- (7) *Začujem njezin glas, naglo otvorim vrata, bacim pismo i pobjegnem što su me noge nosile.*
- (8) *Dali su se u bijeg koliko su ih noge nosile.*
- (9) *Prvo sam odradila šta sam imala na poslu i otperjala kud me noge nose.*

<sup>4</sup> Vidi detaljnije u Parizoska–Filipović Petrović (2017).

Iz primjera je vidljivo da se frazem, osim s najfrekventnijim glagolima *bježati* i *trčati*, javlja i s drugim glagolima, ali i bez glagola te s drugim pri-lozima osim *koliko*, poput *kud* i *što*. Takve je promjene moguće pouzdano utvrditi samo na velikom broju primjera kakav pruža računalni korpus.

Nadalje, posebno su zanimljivi primjeri frazema koji se javljaju u više leksičko-sintaktičkih oblika, poput *biti u zapečku* (10), *ostati u zapečku* (11), *sjediti u zapečku* (12), *gurnuti u zapećak što* (13), *baciti u zapećak što* (14) i *pa-sti u zapećak* (15):

(10) *No, kroz povijest filma, posebno u zlatno doba Hollywooda, melodrama je bila zakon, a sad je u zapečku.*

(11) *U međuvremenu, druge su se struke razvijale, vrednovale, cijenile, a primalj-stvo je ostalo negdje u zapečku.*

(12) *A u svojoj 70. godini života nije sjedila u zapečku i kačkala prisjećajući se dana stare slave, već je napravila novu modnu revoluciju pomevši konkurenciju mlade generacije.*

(13) *Premda mu je životno opredjeljenje s vremenom postala medicina, fotoaparati nipošto nije gurnuo u zapećak.*

(14) *Onda je došao rat i tričave teme bacio u zapećak.*

(15) *Odlaskom Engleza, kriket je pao u zaborav, a i Vis je nekako pao u zapećak, po-gotovo nakon Drugog svjetskog rata, kada je čitav otok pretvoren u vojnu bazu za-branjenu za strance.*

U pokazanim se primjerima nevažno, zaboravljeno mjesto konceptu-alizira kao zapećak, pri čemu se zapećak može doživjeti kao spremnik, a varijante frazema odražavaju boravak u spremniku, stavljanje u spre-mnik, ostanak u spremniku i dolazak u spremnik, profilirajući pritom ra-zličite aspekte istog događaja.<sup>5</sup> Takvi frazemi koji se u kontekstu realizi-rajaju u nizu konstrukcija nazivaju se shematičnima i imaju relativno fleksi-bilnu strukturu.

Također varijantnima, no manje shematičnima pokazali su se frazemi koji odražavaju prijenos (npr. *dobiti batine* i *dati batine*). Budući da je tu ri-ječ o specifičnome događaju, dosljedno se takvi frazemi javljaju u dvjema konstrukcijama: *dati crveni karton komu* i *dobiti crveni karton*, *dati žuti kar-ton komu* i *dobiti žuti karton*, *dati nogom u guzicu komu* i *dobiti nogom u guzi-cu*, *dati po nosu komu* i *dobiti po nosu*, *dati na pladnju komu što* i *dobiti na plad-nju što*.

Takve spoznaje o varijantnosti frazema, osim u frazeologiji, važne pos-ljedice imaju i u leksikografiji. Naime, rječnici utemeljeni na korpusima i

<sup>5</sup> Događajne sheme (shema prijenosa, shema premještanja, shema samostalnog kretanja i druge) opisuju se u Radden–Dirven (2007:284–299).

kognitivnom pristupu varijantnosti, u natukničkom će članku varijantne frazeme obraditi tako da korisniku pokaže njihovu povezanost i zajedničku konceptualnu motivaciju (usp. Parizoska–Filipović Petrović 2020).

### 3.2. Modifikacije

Premda su se u počecima frazeologije kao jezične discipline promjene u frazemima, posebno one kompleksnije, jednostavno smatrale izlaskom iz sfere frazeologije, na što je ponajviše utjecao stav o okamenjenosti kao ključnom obilježju frazema, modifikacije su ipak pronašle svoje mjesto u frazeologiji. Kognitivna je lingvistika modifikacije objasnila kao kreativnu upotrebu frazema koja je u uskoj vezi s konvencionalnom upotrebom i originalnim oblikom frazema, dok je korpusna lingvistika dala zamah njihovim istraživanjima u stvarnoj jezičnoj upotrebi.

Frazeološke se modifikacije opisuju kao namjerne i svjesne preoblike frazema radi određene komunikacijske svrhe. Slušatelj značenju modifikacije pristupa s obzirom na značenje originalnog frazema pa je pri takvim nekonvencionalnim promjenama frazema važno zadržati barem djelomičnu prepoznatljivost temeljnoga oblika (lekseme, strukturu, pozadinske kognitivne koncepte) (Langlotz 2006:216; Omazić 2008:72). Promjene koje govornici čine najčešće su zamjena sastavnica drugim riječima ili dodavanje novih, zatim zamjena mjesta frazemskih sastavnica te spajanje dvaju frazema u jedan. Također, u nekim primjerima ne dolazi ni do kakvih značajnijih leksičkih ni strukturnih promjena izvornog oblika frazema, već se ponovno aktualiziraju leksička značenja riječi u njegovu sastavu. Takvo istovremeno aktiviranje leksičkog značenja pojedinih riječi u sastavu frazema i frazeološkog značenja naziva se miješanjem razina (engl. *level switching*) (Omazić 2003:112). Primjer (16) takvog modificiranja uočili smo istražujući primjere upotrebe frazema *servirati na pladnju (tanjuru) komu što* u korpusu hrWaC:

(16) *Tebi sve što mediji lijepo serviraju na tanjur, ti to fino popapaš.*

Uz figurativno značenje tog frazema ‘osigurati sve potrebno za koga ili pružiti komu što važno na način da tko sve to dobije bez truda’ istovremeno je aktivirano doslovno značenje sastavnice *tanjur* uz koje se veže konzumacija hrane tako što je dodan glagol *popapati* koji znači *jesti*. Time je humorom ublažen prigovor sugovorniku i efekt modifikacije je postignut.

Dvije velike frazeološke i frazeografske tradicije – engleska i ruska – uključuju modifikacije u istraživanja i rječnike, pa u ruskome postoji rječnik modifikacija frazema (Melerović–Mokienko 2005), a engleski frazeološki rječnici *Longman Idioms Dictionary* (1998) i *Collins Cobuild Idioms Dicti-*



*onary* (2002) bilježe i modifikacije. Štoviše, ti rječnici svjedoče da je upotreba modifikacija u engleskome jeziku vrlo raširena, posebno u novinskome stilu i općenito u medijima. U hrvatskome jeziku istraživanja modifikacija nisu brojna (Omazić 2003; Kovačević–Mihaljević 2005; Parizoska 2006), a rječnici ih ne bilježe, odnosno bilježi ih tek jedan rječnik – Benešićev *Rječnik hrvatskoga književnoga jezika od preporoda do I. G. Kovačića*. Premda ih ne naziva modifikacijama jer rječnik nastaje u doba prije same frazeologije kao discipline, sredinom 20. stoljeća, Benešić u primjerima upotrebe (citirama iz književnih djela), uz konvencionalne oblike frazema bilježi i modifikacije, poput primjera (17) u kojem se frazem *vratiti šila za ognjilo* ‘osvetiti se’ javlja u modificiranome obliku *vratiti tri šila za jedno ognjilo* u značenju ‘višestruko se osvetiti’ (Julije Benešić: *Rječnik...* 2017:369):

(17) ... opljačkaše Herbersteinovi vojnici banova imanja oko Ozlja, zapališe kuće, odvedoše marvu [...] Knez Fran nije mislio na drugo, nego kako da vrati tri šila za jedno ognjilo (Kumičić, *Urota zrinsko-frankopanska*, 1916, 303).

Istraživanja frazema u korpusu hrWaC pokazuju da su modifikacije prisutne kod hrvatskih govornika. Često je riječ o javnom diskursu, izjava-ma govornika u medijima, naslovima u novinama ili na portalima te sarkastičnim i ironičnim raspravama o nekoj temi na forumu ili u komentari-ma ispod članaka. Potvrđuje se time Langlotzov (2006) stav da je varijantnost frazema fenomen jezika u upotrebi, pa se stoga mora promatrati u korpusu. Konkretno, primjeri modifikacija u hrvatskom mrežnom korpusu u najvećoj su mjeri semantičke prirode: istovremeno se aktivira frazeološko značenje i doslovno značenje sastavnica frazema, kao u sljedećim primjerima:

(18) Ništa, krenuo ja trbuhom za kruhom u špajz, uzeo pola kile grahama i pola kile bijelog brašna.

(19) Ovom IP-u sam kazao što mislim. Možda mi je to loša osobina: ne znam zamati u celofan već isporučujem u originalnom pakiranju.

(20) Našim građanima očito nije dovoljno da im voda dođe do grla nego moraju biti već dobrano pod vodom da bi se pobunili.

(21) Tebi stari ne fali daska u glavi, fali ti cijela baraka.

U modifikaciji (18) frazema *krenuti, otići trbuhom za kruhom* ‘napustiti zavičaj u potrazi za boljim životnim uvjetima’ aktivirano je doslovno značenje sastavnica *trbuh* i *kruh* tako što su dodani pojmovi povezani s prehranom: integralno i bijelo brašno. Namjerni humoristični efekt modifikacije postignut je i dodavanjem *špajze* kao mjesta na koje se odlazi, a za koje se može pretpostaviti da je govorniku vrlo blizu, u njegovu životnom prostoru, nasuprot najčešće vrlo dalekom i stranom mjestu na koje se odno-

si značenje originalnog frazema. U primjeru (19) originalni frazem *zamotati u celofan što* koji znači 'reći ili prikazati što neugodno ili nepovoljno na lijep način, kako bi se prikrile loše strane; izbjegavati reći istinu' modificiran je tako što je aktivirano doslovno značenje sastavnice *celofan* 'ukrainski papir kojim se zamataju pokloni, cvijeće i slično' da bi ga se suprotstavilo *originalnom pakiranju* koje u značenju modifikacije predstavlja surove ili neugodne činjenice koja se otvoreno iskazuju. Nadalje, često se modifikacije koriste kako bi se hiperboliziralo značenje originalnog frazema, kao u primjerima (20) i (21). Značenje frazema *došla je voda do grla komu* 'u jako teškom položaju je tko' pojačava se aktiviranjem doslovne slike frazema (*voda do grla*) kojoj se dodaje slika u kojoj je tko *dobrano pod vodom* te se time modificira u značenje 'u neizdrživom položaju je tko'. Također, značenje frazema *fali daska u glavi komu* 'nije posve normalan tko' na sličan je način naglašeno do značenja 'u potpunosti je lud tko' tako što je u modifikaciji dodana riječ *baraka* koja je aktivirala doslovno značenje sastavnice originala *daska* 'plosnat i dug komad drveta', jer se baraka kao improvizirana građevina najčešće sastoji od mnogo dasaka.

Korpusna su istraživanja pokazala još jednu tendenciju govornika pri upotrebi frazeoloških modifikacija. Naime, frazemi u kojima se govori o veličini, jačini, kvaliteti ili važnosti čega (poput *za koplje iznad (ispred)*; *raditi od muhe slona*) često se modificiraju u situacijama u kojima se želi stupnjevati to obilježje. Pritom se ti frazemi namjerno preinačuju tako što se umjesto jedne sastavnice umeću druge koje ukazuju na povećanje obilježja, ponekad do preuveličavanja kako bi se postigao željeni efekt modifikacije. Takvi su primjeri frazem *raditi od muhe slona* koji se javlja u modifikacijama *napraviti od muhe vola*, *raditi od muhe avion*, *napraviti od muhe krdo pobješnjelih mamuta*, *praviti od muhe Kilimandžaro* (22) te frazem *[za] koplje ispred (iznad)* u modifikacijama *dva koplja iznad*, *tri koplja ispred*, *deset kopalja iznad* (23), *za sto kopalja iznad*:

(22) *Umjesto da svoje propuste smjestimo u realne okvire, naučimo nešto iz njih i zaboravimo neugodne epizode, mi od muhe pravimo Kilimandžaro.*

(23) *Charlize Theron je apsolutno deset kopalja iznad svih ostalih, i po obleki i po ljepoti i po držanju.*

Rjeđe se javljaju modifikacije u svrhu igre riječima te promjene u redoslijedu sastavnica i dodavanje novih dijelova kao u primjerima modifikacija frazema *reći popu pop, a bobu bob* (24) i (25):

(24) *Kako Iggyju reći pop, a Dylanu bob: Ogleđi iz rock i pop-kulture.*

(25) *Kvragu, ja jesam osoba koja voli reći bobu bob, a popu pop, ali spremna sam i reći bobu pop, a popu bob ako za to postoji logički opravdan razlog.*

### 3.3. Novi konvencionalizirani frazemi

Osim novih spoznaja o karakteristikama frazema, njihovoj upotrebi i varijantnosti, korpusni su podaci u velikoj mjeri utjecali na frazeografiju, odnosno sastavljanje rječnika frazema. U predgovoru engleskog frazeološkog rječnika *Longman Idioms Dictionary* navodi se da se rječnik temelji na korpusu iz triju važnih razloga: korpusi omogućuju pronalazak frazema, korpusi daju informacije o varijantnim oblicima frazema i njihovoj tipičnoj upotrebi i konačno, s pomoću korpusâ može se utvrditi koliko se često frazemi upotrebljavaju (1998:xi).

Istraživanja frazema provedena *corpus-driven* pristupom u korpusu hrWaC rezultirala su frazemima i značenjima koji nisu zabilježeni u postojećim hrvatskim frazeološkim rječnicima. Tako su se konvencionaliziranima pokazali izrazi *čist kao kristal*, *puknuti kao kokica* i *izgledati kao avion*, pa ih treba kao frazeme uvrstiti u rječnike. Njihovo figurativno značenje iščitava se iz konteksta u primjerima upotrebe (26), (27) i (28):

(26) *Njen nastup je siguran, smion, njena dikcija čista kao kristal.*

(27) *Nije da smo ikada bili sasvim normalni, ali sada smo totalno prolupali, ono, pukli ko kokice.*

(28) *Aktivan sportaš mora izgledati kao avion, a svaka žena hoće mišić na mišiću.*

Osim novih frazema, podaci iz korpusa pokazali su i dosad u leksikografiji nezabilježena značenja, što je primijećeno u frazemima sa sastavnicom *nebo*. Skica riječi u alatu za pretragu korpusa Sketch Engine za imenicu *nebo* ukazala je na vrlo frekventnu sintagmu *do neba*, a daljnja je pretraga pokazala da su među čestim glagolima koji se nalaze ispred te sintagme *uzdizati* i *dizati*. Uvidom u primjere odnosno njihovim pomnim iščitavanjem ustanovljeno je da se frazemi s tim sastavnicama javljaju u više značenja:

(29) *Posebno ju je nerviralo kada su oca pred njom uzdizali do neba i neprestano hvalili, jer to, naime, nije baš bilo dobro za njegovu taštinu.*

(30) *Dvostruki strijelac sa sinoćnje utakmice, kojeg u Turskoj dižu do nebesa, uvjeren je kako Hrvatska nema nimalo šanse i bahato najavljuje pobjedu.*

(31) *Trebali bismo dići glas do neba, uzbuniti čitavu zemaljsku kuglu zbog nepravde koja nam se nanosi.*

(32) *Hrvati ne razlikuju bitne stvari od nebitnih. Dižu galamu do neba kad je riječ o nekoj beznačajnosti.*

(33) *Naravno, farmaceuti su odmah digli buku do nebesa, jer bi im to izbilo ogromne količine profita iz ruku.*

(34) *Baluba je postala ime tek kad su digli cijene do neba pa su se snobovi polakomili za skupom robom i svima trubili koliko su para dali za namještaj.*

(35) *Ne treba dizati tenzije do neba jer ne ide se u rat nego na nogometni stadion.*

(36) *U Italiji se – osobito u novinarskim i političarskim krugovima – digla panika do neba.*

(37) *U Prvoj HNL nastavlja se po uhodanom putu: čim momčad izgubi dvije ili tri utakmice zaredom, trener dobiva otkaz, a ambicije s novim stručnjakom dižu se do nebesa.*

(38) *Znam, ima onih koji kad završe fakultet dignu nos do neba i sve im je teško raditi.*

(39) *Kad smo mu pokušale objasniti da žene koje nama dolaze na kavu dolaze pričati o šemama, kolačima, ogovarati susjede, komentirati gluposti, uvrijedio se i digao nos do neba.*

Iz navedenih se primjera mogu iščitati četiri frazema: *dignuti/dizati (uzdizati) do neba (nebesa) koga, što* (primjeri 29 i 30) u značenju ‘pretjerano se oduševljavati kime ili čime, jako pohvalno govoriti o kome, čemu’, zatim *dignuti/dizati glas (galamu, buku) do neba (nebesa)* (primjeri 31, 32 i 33) u značenju ‘intenzivno se zalagati i boriti za što ili protiv čega’, *dignuti/dizati (dignulo se / diže se) što do neba* (primjeri 34, 35, 36 i 37) u značenju ‘potencirati, uvećavati što’ te *dignuti nos do neba* s dvama značenjima: ‘uobraziti se, biti nadmen’ (primjer 38) i ‘naljutiti se’ (primjer 39). Može se zaključiti da je veliki broj primjera upotrebe koji pruža računalni korpus nužan uvjet za pouzdano i jasno određivanje oblika frazema te razgraničavanje njihovih značenja, posebno kad su u pitanju iste sastavnice. Ključnim se pokazao i *corpus-driven* pristup zahvaljujući kojem se nisu ciljano tražili poznati oblici i značenja frazema, jer bi tada oni drugi neizbježno ostali izvan dosega istraživanja, nego je početak bio jedna imenica, a zatim se dalje utvrđivala njezina česta okolina i moguća figurativna značenja.

### 3.4. Frazemi s više značenja

Pretražujući u korpusu rezultate koji pokazuju najčešću okolinu riječi i uočavajući pritom konstrukcije s prenesenim značenjem, često je u daljnjim koracima potrebno detaljno čitati primjere upotrebe kako bi se sa sigurnošću utvrdilo značenje frazema. Pritom se ponekad pokaže da pojedini frazemi imaju dva značenja. I u tim se primjerima uočava sustavnost kao posljedica konceptualne motivacije u pozadini značenja frazema. Naime, frazemi sa somatskom sastavnicom, poput noge, glave, nosa, prsta javljaju se u dvama značenjima, što se može uočiti na primjeru frazema *ostati na nogama* koji se javlja u značenjima ‘ostati uspravan, uspjeti stajati’ (40) i ‘opstati, održati se’ (41):

(40) *Stajao je u sredini, optužen za krađu. Nije dugo ostao na nogama. Kad su ga udarcima srušili na tlo, počeli su ga šutati.*

(41) *Kada su agresori polomili sve druge odašiljače, Radio Posušje je ostao čvrsto na nogama i prenosio sve najvažnije događaje u tim teškim danima.*

Općenito su frazemi sa somatskom sastavnicom u jeziku brojni, zbog važne uloge tijela u čovjekovu shvaćanju svijeta. Naše konvencionalizirano znanje o dijelovima tijela i za što su pojedini dijelovi važni ili kojeg su oblika, omogućuje nam njihovo preslikavanje u sferu apstraktnog. U primjeru frazema *ostati na nogama* prvo je značenje motivirano znanjem o važnosti nogu za stajanje, konkretno stajanje se ostvaruje isključivo nogama, dok se u drugom značenju dogodio pomak s fizičkog stajanja na simbolično stajanje koje predstavlja opstanak, održavanje čega stabilnim. Isti se pomak dogodio i u primjeru frazema *biti na nogama*, pri čemu prvo značenje glasi 'ustati se' (42), a drugo 'uzbuniti se, aktivirati se' (43):

(42) *Ja ne mogu spavati dugo, jutros nisam imala nikakve obaveze, a već sam u 7 bila na nogama.*

(43) *U subotu je helsinška policija bila na nogama nakon što je dojavljeno da se u jednomu kamionu koji brodom stiže u luku nalazi – bomba.*

Također, frazem *doći glave* koga javlja se u dvama značenjima:

(44) *Samouvjerenost i podcjenjivanje planine mnoge je došlo glave, a tu opasnu kombinaciju na svojoj koži osjetio sam i sam.*

(45) *Dakle, još jedno neozbiljno shvaćanje protivnika moglo bi Dinamo doći glave, ali možda je ipak manja sramota ispasti sada nego ponoviti blamažu iz prošle sezone.*

Ti se frazemi razumiju zahvaljujući središnjem mjestu koje glava ima u čovjekovu poimanju sebe i njezinoj važnosti za život, pa u primjeru (44) metonimijski odnos GLAVA ZA ČOVJEKA čini motivacijsku pozadinu za razumijevanje značenja 'umrijeti, izgubiti život', dok se u primjeru (45) znaenje o glavi kao najvažnijem dijelu tijela, presudnom za život preslikava na druge pojavnosti osim čovjeka, koje tada bez svoje metaforičke glave ne mogu opstati, propadaju ili zapadaju u velike probleme.

U dvama se značenjima javljaju i frazemi *dobiti po nosu* sa značenjima 'dobiti batine, biti istučen' (46) i 'biti oštro iskritiziran, ukoren' (47):

(46) *Dobro upućeni u stanje stvari u šali ističu kako bi se taj problem mogao izgleda riješiti samo još jednom kavanskom tučnjavom, ali takvom tučnjavom u kojoj bi neki od hrvatskih čelnika dobio po nosu.*

(47) *Svaki put kad sam ušao u rat s medijima dobio sam po nosu. Što mi sada treba da se opet nekom zamjerim.*

i *ne vidjeti prst pred nosom* sa značenjima 'ne vidjeti ispred sebe zbog mraka, magle ili kakve druge zapreke' (48) i 'ne shvaćati širinu problema, ne vidjeti stvari kakvima jesu' (49), vidljivima iz primjera:

(48) *Gospođa Šantek vozi po Zelenom valu, kiša lije, brisači rade punom parom, a ona svejedno ne vidi prst pred nosom jer je to, podsjetimo se, izrazito tmuran dan, dan koji ne obećava.*

(49) *Tko od nas nije bio u situaciji da se zaljubio, da od silne zaljubljenosti nije vidio prst pred nosom, stvari koje su drugima bile očite?*

Konačno, treba spomenuti da podaci iz korpusa ukazuju na to da su frazemi u značenjima koja prelaze fizička obilježja somatizama i proširuju se u sferu apstraktnog – frekventniji, što potvrđuje sada već znamenitu postavku Lakoffa i Johnsona (1980) da čovjek u velikoj mjeri svakodnevno koristi metafore u razmišljanju i govoru.

### 3.5. Specifična upotreba

Na velikom broju primjera upotrebe kakav donosi računalni korpus moguće je primijetiti još neke specifičnosti vezane uz frazeme. Najčešće je u pitanju određen kontekst u kojem se frazem vrlo često javlja, kao u primjeru frazema *biti na izmaku* u značenju 'biti pri kraju, pri završetku' (*Čitav tjedan koji je na izmaku potrošen je na liječničke preglede*) za koji se pokazalo da ga govornici često koriste kada govore o završetku kalendarske godine (50) i (51):

(50) *Kraj godine je vrijeme kad se povlače bilance i daje inventura godine koja je na izmaku i najjavljaju događanja u novoj.*

(51) *Prigodnim poklonom u sva domaćinstva nastojimo unijeti dio vlastitog kraja: kulturu, tradiciju, običaje, krajolik i događaje koji su obilježili godinu na izmaku.*

Također, frazem *pokrenuti s mrtve točke* što u značenju 'pokrenuti što nakon duže pasivnosti i mirovanja' (*Akcija banaka, kojom su ponudile nekretnine na kupnju po nešto povoljnijim uvjetima, pokrenula je tržište s mrtve točke*) često se upotrebljava s imenicom 'stvar', odnosno 'stvari' (52) i (53):

(52) *Nakon skoro godinu dana djelovanja pokrenuli smo neke stvari s mrtve točke.*

(53) *Na tom području novi predsjednik Republike neće moći učiniti mnogo bez obzira na njegove želje da stvar pokrene s mrtve točke.*

Takvi podaci o specifičnoj upotrebi posebno su se odrazili na leksikografiju. Rječnici koji se temelje na korpusima počeli su u natuknicu uključivati i informacije o upotrebi (engl. *usage notes*) koje uglavnom upućuju na određena ograničenja u upotrebi riječi i višerječnica ili na specifične kontekste u kojima se pojavljuju. Kada je o frazemima riječ, informacije o upotrebi posebno su korisne onima koji uče strani jezik, jer im bez korpusnih podataka takva obilježja frazema ostaju nepoznata. Navodimo još jedan primjer iz korpusa koji pokazuje da se frazem *dobiti žuti karton* ponekad javlja u kombinaciji s frazomom *dobiti crveni karton*:

(54) *U slučaju da tvrtka dobije žuti karton, počinje teći rok u kojem mora ispraviti propust. Imate osam dana da popravite stvar. Nakon toga dobijete zeleni ili crveni i kaznu.*

(55) *Povećanje plaća ne dolazi u obzir. Uprava HŽ Holdinga ovim je slučajem zadržala prvi žuti karton. A dva će biti crveni što znači otkaz upravi.*

#### 4. Zaključak

Brojna suvremena frazeološka istraživanja temeljena su na podacima iz velikih računalnih korpusa. Razlog za to su jedinstvene mogućnosti pretraživanja korpusa koje daju specijalizirani alati u kojima se korpusi nalaze, kao što je npr. Sketch Engine, a koji pružaju statistički značajne podatke o karakteristikama i ponašanju frazema u jeziku, kao što su frekvencija upotrebe, sistematično pojavljivanje varijanata i moguća specifična okolina u kojoj se javljaju. Također, jedno od istaknutijih obilježja frazema – varijante i modifikacije – fenomen su jezika u upotrebi pa ih je moguće istraživati jedino na velikom broju primjera stvarne upotrebe, kakav sadržavaju računalni korpusi. Dosad provedena istraživanja u mnogim jezicima dala su rezultate važne za spoznaje o opsegu frazema, njihovoj promjenjivosti i stupnju prozirnosti značenja. Budući da računalni korpusi daju nebrojene mogućnosti istraživanja tog jezičnog fenomena, frazeologija i korpusna lingvistika tek trebaju dosegnuti svoj puni potencijal istraživanja.

U ovome radu bavili smo se frazemima koje pronalazimo u korpusu i podacima o frazemima koje saznajemo iz korpusa. Primjeri su pokazali da računalni korpusi: 1) donose podatke o postojanju frazema, odnosno čestoti njegove upotrebe, 2) putem primjera stvarne upotrebe olakšavaju razumijevanje značenja frazema i daju uvid u mogući specifični kontekst u kojem se frazem češće ili jedino javlja, 3) pokazuju sistematičnost u upotrebi varijanata određenog frazema, što je važno za uvištavanje varijanata u rječnik i 4) pokazuju kreativnu upotrebu frazema u komunikacijske svrhe u obliku namjernih modifikacija. U pitanju su znanja o frazemima koja ne možemo saznati čitajući odabrane zasebne tekstove bilo koje vrste ili konzultirajući rječnike, pa se korpusi opravdano mogu smatrati neizostavnim jezičnim alatima u frazeološkim istraživanjima i frazeografskoj praksi.

## Literatura

- Biber, Douglas; Susan Conrad; Randi Reppen. 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Cignoni, Laura; Stephen Coffey; Rosamund Moon. 1999. Idiom variation in Italian and English: Two corpus-based studies. *Languages in Contrast*, 2/2, 279–300.
- Collins COBUILD Idioms Dictionary*. 2002. Glasgow: HarperCollins Publishers.
- Computational and Corpus-Based Phraseology. Second International Conference, Europhras 2017, London, UK, November 13-14, 2017. Proceedings*. 2017. Ur. Mitkov, Ruslan. Cham: Springer.
- Computational and Corpus-Based Phraseology. Third International Conference, Europhras 2019, Malaga, Spain, September 25–27, Proceedings*. 2019. Ur. Corpas Pastor, Gloria; Ruslan Mitkov. Cham: Springer.
- Filipović Petrović, Ivana; Jelena Parizoska. 2017. Leksikografska obrada frazema s promjenjivom glagolskom sastavnicom u hrvatskome. *Jezikoslovlje*, 18, 2, 245–278.
- Fleischer, Wolfgang. 1997. *Phraseologie der deutschen Gegenwartssprache*. Tübingen: Max Niemeyer Verlag.
- Granger, Sylviane; Magali Paquot. 2008. Disentangling the phraseological web. Prir. Granger, Sylviane; Fanny Meunier. *Phraseology: an interdisciplinary perspective*. Amsterdam: John Benjamins Publishing, 27–49.
- Hanks, Patrick. 2009. The impact of corpora on dictionaries. Prir. Baker, Paul. *Contemporary Corpus Linguistics*. London: Continuum, 214–236.
- Julije Benešić: *Rječnik hrvatskoga književnoga jezika od preporoda do I. G. Kovačića*. 2017. Ur. Moguš, Milan; Anja Nikolić-Hoyt. Prir. Filipović Petrović, Ivana; Milan Moguš; Anja Nikolić-Hoyt; Bojana Schubert. Sv. 14. *spužvica–švrljuga*. Zagreb: Hrvatska akademija znanosti i umjetnosti.
- Kovačević, Barbara; Milica Mihaljević. 2005. Frazemi u publicističkome funkcionalnom stilu (modifikacija, kontaminacija i transformacija frazeološke strukture). Prir. Granić, Jagoda. *Semantika prirodnog jezika i metajezik semantike: zbornik Hrvatskoga društva za primijenjenu lingvistiku*. Zagreb – Split: HDPL, 393–404.
- Lakoff, George; Mark Johnson. 1980. *Metaphors We Live By*. Chicago – London: The University of Chicago Press.
- Langlotz, Andreas. 2006. *Idiomatic Creativity*. Amsterdam – Philadelphia: John Benjamins.
- Longman Idioms Dictionary*. 1998. Harlow: Longman.



- McEnery, Tony; Costas Gabrielatos. 2006. *English Corpus Linguistics*. Prir. Aarts, Bas; April McMahon. *The Handbook of English Linguistics*. Blackwell Publishing Ltd, 33–71.
- Melerovič, Mokienko 2005 = Мелерович, А. М.; В. М. Мокиенко. 2005. *Фразеологизмы в русской речи. Словарь*. Москва: Астрель.
- Menac, Antica. 2007. *Hrvatska frazeologija*. Zagreb: Knjigra.
- Meyer, Charles F. 2002. *English corpus linguistics: an introduction*. Cambridge: Cambridge University Press.
- Moon, Rosamund. 1998. *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Oxford: Clarendon Press.
- Omazić, Marija. 2003. *Modifications of phraseological units in English*. Zagreb: Sveučilište u Zagrebu (disertacija).
- Omazić, Marija. 2008. Processing of idioms and idiom modifications: A view from cognitive linguistics. Prir. Granger, Sylviane; Fanny Meunier. *Phraseology: An interdisciplinary perspective*. Amsterdam: John Benjamins Publishing, 67–80.
- Omazić, Marija. 2015. *Phraseology through the looking glass*. Osijek: Josip Juraj Strossmayer University of Osijek, Faculty of Humanities and Social Sciences.
- Parizoska, Jelena. 2006. Semantičke modifikacije frazema u novinskim naslovima. Ur. Granić, Jagoda. *Jezik i mediji. Jedan jezik: više svjetova. Zbornik HDPL-a*. Zagreb – Split: Hrvatsko društvo za primijenjenu lingvistiku, 533–540.
- Parizoska, Jelena. 2019. *Promjenjivost glagolskih frazema u engleskom i hrvatskom jeziku – kognitivnolingvistička analiza*. Doktorski rad. Zagreb: Filozofski fakultet Sveučilišta u Zagrebu.
- Parizoska, Jelena; Ivana Filipović Petrović. 2017. Variation of Adjectival Slots in kao ('as') Similes in Croatian: A Cognitive Linguistic Account. Prir. Mitkov, Ruslan. *Computational and Corpus-Based Phraseology. Second International Conference, Europhras 2017, London, UK, November 13-14, 2017. Proceedings*. Cham: Springer, 348–362.
- Parizoska, Jelena; Ivana Filipović Petrović. 2020. Varijantni frazemi u e-rječniku. *Rasprave Instituta za hrvatski jezik i jezikoslovlje*, 46, 2, 941–955.
- Radden, Günther; René Dirven. 2007. *Cognitive English Grammar*. Amsterdam – Philadelphia: John Benjamins.
- Sinclair, John. 2004. *Trust the text. Language, Corpus and Discourse*. Routledge.
- Stanojević, Mateusz-Milan; Jelena Parizoska. 2005. Konvencionalne konceptualne metafore i idiomatičnost. Ur. Granić, Jagoda. *Semantika prirodnog jezika i metajezik semantike*. Zagreb – Split: Hrvatsko društvo za primijenjenu lingvistiku, 701–712.

Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Amsterdam – Philadelphia: John Benjamins.

## A Corpus-Driven Study of Idioms in Croatian

### Abstract

Since their advent, electronic corpora have been considered a useful linguistic tool, but today they are virtually a matter of course in research conducted in most linguistic subdisciplines, including phraseology. Idioms, the subject matter of phraseology, are characterized by conventionality, (in)variability and desemantization. Their systematic study requires not only numerous examples of their use, but should also rely on statistical trends, including their frequency, characteristic combinations, the environments where they appear and, ideally, instantaneous access to their wider context.

This paper presents a study into the characteristics of idioms by relying on data from the largest Croatian electronic web corpus, hrWaC. The study has been conducted within the project to compile the *Dictionary of Croatian Idioms*. The aim of the paper is to determine the characteristics of idioms used by speakers of Croatian, particularly their variability, meaning and use, in an attempt to contribute to gaining a deeper and a more complete insight into Croatian phraseology.

Ključne riječi: frazemi, računalni korpus, frekvencija, konvencionaliziranost, varijantnost, primjeri stvarne upotrebe

Keywords: idioms, corpus, frequency, conventionality, variability, examples of actual use