

BRANKA BARČOT

FILOZOFSKI FAKULTET SVEUČILIŠTA U ZAGREBU

LIDIJA MILKOVIĆ

FILOZOFSKI FAKULTET SVEUČILIŠTA U ZAGREBU

FRAZEOGRAFIJA U MREŽNOME RUHU – OD KARTOTEKE DO *SKETCH ENGINEA*

IZVORNI ZNANSTVENI RAD

UDK 81'373.7'374

811.163.42'373.7:811.112.2

811.112.2'373.7:811.163.42

[HTTPS://WWW.DOI.ORG/10.17234/9789533790121.07](https://www.doi.org/10.17234/9789533790121.07)

U radu se najprije donosi pregled razvoja frazeografije u Hrvatskoj: od tiskanih rječnika nastalih „u radionici“ *zagrebačke frazeološke škole* do e-frazeografije digitalnog doba, a potom se predstavljaju rezultati rada na pilot-projektu koji se temelji na ručno prikupljenoj kartoteci njemačkih frazema. Opisuju se sve projektne faze rada na frazeološkoj građi: digitalizacija kartoteke te obrada digitaliziranih dokumenata; frazeološka obrada građe; pronalazak hrvatskih frazeoloških ekvivalenata; anotiranje frazema; izrada usporedivoga publicističkog potkorpusa njemačkih i hrvatskih frazema pomoću alata *Sketch Engine*. Cilj je ovoga rada predstaviti razvojni put frazeografije u nas posljednjih 50-ak godina te prikazati na koji se način može kartoteka, prikupljena na tradicionalan način, iskoristiti i „odjenuti u suvremeno mrežno ruho“. Dakle, krajnja je svrha rada predstaviti utabanu novu metodološku stazu za nastanak dvojezičnog korpusa koji bi bio od najveće koristi ponajprije studentima njemačkog jezika u nas i prevoditeljima s njemačkog jezika i na njemački, ali jednako tako i izvornim govornicima njemačkog jezika te slavistima na njemačkom govornom području. Spomenuta metodološka staza potencijalno će jednako tako biti zanimljiva i našim istraživačima koji se bave područjem frazeologije i frazeografije budući da, koliko nam je poznato, ovakav korpus u nas ne postoji.

Ključne riječi: tradicionalna frazeografija; e-frazeografija; publicistički potkorpus njemačkih frazema; publicistički potkorpus hrvatskih frazema; usporedivi frazeološki korpus; Sketch Engine

1. UVOD

Dobro je poznato da je potreba za bilježenjem nerazumljivih, novih riječi stara onoliko koliko i kontakti među jezicima. Općenito se može reći da je leksikografski rad u jezično-kulturnim zajednicama započeo onda kada su se pojavili prvi popisi riječi s kratkim objašnjenjima. O formalnim leksikografskim počecima u svijetu ali i u nas detaljnije piše Ivana Filipović

Petrović (2018: 5–40) u svojoj monografiji *Kada se sretnu leksikografija i frazeologija. O statusu frazema u rječniku*. Mi bismo u ovom radu pozornost željeli usmjeriti na frazeografiju u nas. Poput velikog Henrija Bédjonta i mi ćemo uzeti slobodu te parafrazirati Stevea Pinkera te se usuditi ustvrditi kako nikada nismo susreli filologa koji se nije zainteresirao za frazeme. Da bilježenje frazema u rječnicima nije nimalo lagan zadatak, najbolje ilustrira citat Sue Atkins koja navodi da su oni „močvara za leksikografe“ (Filipović Petrović 2018: 83).

Kako su se u toj „močvari“ od samih početaka izdvajanja frazeologije kao samostalne lingvističke discipline snalazili naši frazeolozi i frazeografi opisuje se u prvom dijelu rada naslovljenom *Frazeografija s pečatom zagrebačke frazeološke škole*. Potom slijedi opis naše recentnije frazeografske situacije u dijelu *Frazeografija i izazovi digitalnog doba*. Budući da je cilj ovoga rada osim predstavljanja razvojnog puta frazeografije u Hrvatskoj tijekom posljednjih 50-ak godina, jednako tako i prikazivanje načina na koji se kartoteka, prikupljena na tradicionalan način, može iskoristiti i „odjenući u suvremeno mrežno ruho“, slijede poglavlja u kojima se opisuju faze rada na pilot-projektu pod nazivom *Publicistički potkorpus mrežnoga njemačko-hrvatskog frazeološkog rječnika*, zatim se opisuje alat *Sketch Engine* kao i specifičnosti rada s tim alatom te znanja i vještine potrebne za pronalazak željenog frazema u usporedivom publicističkom potkorpusu njemačkih i hrvatskih frazema. O tom se usporedivom potkorpusu također detaljnije piše u zasebnom dijelu ovoga rada, a jedna od tema kojima se ovdje bavimo su i razlike između anotacije frazema u potkorpusu s jedne strane, te njihova grafičkog izdvajanja u tiskanim frazeološkim rječnicima, s druge strane.

2. FRAZEOGRAFIJA S PEČATOM ZAGREBAČKE FRAZEOLOŠKE ŠKOLE

Zahvaljujući akademkinji Antici Menac i njezinom pionirskom radu *O strukturi frazeologizma* s početka 70-ih godina XX. st. utabao se put frazeologiji kao lingvističkoj disciplini na našim prostorima. Navedeni je rad označio početak frazeoloških istraživanja u Hrvatskoj temeljenih na teoriji ruske frazeologije, kolijevci frazeološke znanosti. Njezin je frazeografski rad također pionirski na ovim prostorima¹. Akademkinja Menac odigra-

1 Sa suradnicima (Tatjana Korać, Milenko Popović, Miho Skljarov, Radimir Venturin, Renata Volos) sastavlja dvosveščani *Rusko-hrvatski ili srpski frazeološki rječnik* (1979, 1980). U razdoblju od 1985. do 1998. g. tiskani su dvojezični i trojezični frazeološki rječnici u seriji *Mali frazeološki rječnici*. Idejni je začetnik tog projekta Antica Menac, a uz nju je još cijeli niz suradnika bio uključen u realizaciju te ideje: Menac, A., Trostinska, R. I. *Hrvatskosrpsko-rusko-ukrajinski frazeološki rječnik* (1985); Menac, A., Sesar, D., Kuchar, R. *Hrvatskosrpsko-češko-slovački frazeološki rječnik* (1986); Menac, A., Pintarić, N. *Hrvatskosrpsko-poljski frazeološki rječnik* (1986); Menac, A., Vučetić, Z. *Hrvatskosrpsko-talijanski frazeološki rječnik* (1988); Menac, A., Blaževac, K. *Hrvatskosrpsko-francuski frazeološki rječnik* (1988); Menac, A., Rojs, J. *Hrvatsko-slovenski frazeološki rječnik* (1992); Cvjetković-Kurelec, V. *Hrvatsko-novogrčki frazeološki rječnik* (1994); Bricko, M., Salopek, D. *Hrvatsko-grčki frazeološki rječnik* (1994); Novaković, D., Perić, O., Tajčević, L., Vratović, V. *Hrvatsko-latinški frazeološki rječnik* (1994). Drugo izdanje su imali: *Hrvatsko-rusko-ukrajinski frazeološki rječnik* (1993); *Hrvatsko-talijanski frazeološki rječnik* (1995) i *Hrvatsko-češko-slovački frazeološki rječnik s indeksom čeških i slovačkih frazema* (prošireno i dopunjeno izdanje) (1998).

la je ključnu ulogu vezano za frazeologiju na hrvatskoj jezikoslovnoj sceni te je svojim znanstvenoistraživačkim radom u toj disciplini stekla velike zasluge. Osim Antice Menac i frazeologa okupljenih oko nje u *zagrebačku frazeološku školu* ne smije se zaboraviti znanstveni i frazeografski rad i doprinos Josipa Matešića².

Frazeografska je djelatnost u Hrvatskoj i danas najvećim dijelom vezana uz *zagrebačku frazeološku školu* koju karakterizira kontinuitet u radu te širenje leksikografskog spektra ne samo na jednojezičnom i dvojezičnom planu već i na višejezičnom planu s tim da su uključeni kako slavenski tako i neslavenski europski jezici, a svim je tim rječnicima zajedničko to da je hrvatski polazišni jezik. Upravo je glavni pokretač frazeografske djelatnosti u okviru *zagrebačke frazeološke škole* slavljena u čiju je čast ovaj zbornik i pripremljen – Željka Fink Arsovski. *Hrvatski frazeološki rječnik* koji Željka Fink Arsovski objavljuje 2003. godine u suautorstvu s Anticom Menac i Radomirom Venturinom izuzetan je leksikografski događaj jer predstavlja prvi takav rječnik suvremenoga hrvatskog jezika. Željka Fink Arsovski je frazeološki *perpetuum mobile* i idejni začetnik koncepcije kontrastivne frazeografske obrade poredbenih frazema koja se donosi u dvama pionirskim leksikografskim radovima s nizom suautora: *Hrvatsko-slavenski rječnik poredbenih frazema* (2006) i *Hrvatsko-romansko-germanski rječnik poredbenih frazema* (2016). Ukupno je osam frazeoloških rječnika koje Željka Fink Arsovski u suautorstvu donosi na hrvatsku leksikografsku scenu³.

Što je zajedničko ovim frazeološkim rječnicima? Koja su se frazeografska načela primjenjivala „u radionici“ *zagrebačke frazeološke škole* prilikom njihova sastavljanja? Ako dvjema stazama – tradicionalnoj i kognitivnoj – kojima se kretala i kreće frazeologija kao lingvistička disciplina (*usp.* Filipović Petrović 2018: 41–81), dodamo i treću stazu – lingvokulturološku, onda valja zaključiti da se metodologija frazeografskog rada *zagrebačke frazeološke škole* kreće po tradicionalnoj stazi. Frazeograf na početku treba donijeti odluku o tome za koga se sastavlja rječnik, koji će frazemi ući u rječnik, koja će riječ imati status nadnatuknice, kako će izgledati natuknički lik te kakva će biti tzv. arhitektura rječničkog članka. Takve se odluke opisuju u uvodnom dijelu rječnika, a *zagrebačka frazeološka škola* u frazeološkim rječnicima u uvodnom dijelu uvijek dosljedno navodi osnovne podatke o rječniku te se prilikom opisa rječničkog članka navode sljedeće informacije i to prema algoritmu: nadnatuknica, natuknica, definicija frazema i primjer njegove upotrebe. Korisniku se uvijek donosi informacija o ukupnom broju frazema u rječniku te informacija o njihovoj generalnoj strukturi, no princip se njihova odabira za uključivanje u rječnik podrobnije ne objašnjava (izuzetak je u tom smislu tematski *Rječnik hrvatskih animalističkih frazema* iz 2017.). Budući da je riječ o tiskanim rječnicima, nadnatuknica igra važnu ulogu prilikom

2 On je autor prvoga *Frazeološkog rječnika hrvatskoga ili srpskog jezika* (1982), a u njegovoj redakciji 1988. g. izlazi *Hrvatsko-njemački frazeološki rječnik*. U radu na tom rječniku surađivali su: Renate Hansen, Jürgen Petermann, Stefan Rittgasser, Martina Steiger i Irena Zimanji-Hofer.

3 Uz navedena tri valja spomenuti i ostalih pet rječnika: *Hrvatsko-engleski frazeološki rječnik* (2008) (u suautorstvu s D. Vrgočem); *Hrvatsko-ruski frazeološki rječnik* (2011) (u suautorstvu s A. Menac, I. Mironovom Blažina, R. Venturinom); *Hrvatski frazeološki rječnik* (2014) (u suautorstvu s A. Menac i R. Venturinom); *Rječnik hrvatskih animalističkih frazema* (2017) (u suautorstvu s I. Vidović Bolt, B. Barčot, B. Kovačević, N. Pintarić, A. Vasung); *Rusko-hrvatski frazeološki rječnik* (2019) (u suautorstvu s V. Mokienko, A. Hrnjak, B. Barčot).

organizacije i prezentacije frazeološke građe. *Zagrebačka frazeološka škola* slijedi tzv. morfološki princip kada je riječ o navođenju nadnatuknice. U praksi to izgleda na sljedeći način: ako frazem sadrži imenicu, obrađen je pod tom imenicom, te se ona smatra glavnom riječju, odnosno nadnatuknicom. Ako frazem sadrži dvije imenice, za nadnatuknicu se uzima ona prva. Kao imenice se tretiraju i druge vrste riječi kada su poimeničene. Po nadnatukničkoj hijerarhiji nakon *imenice* i *poimeničene riječi* slijede *pridjev*, *prilog*, *glagol*, *broj* pa *zamjenica*. O uporabi vrsta slova (primjerice, debelih verzalnih slova za nadnatuknice, debelih slova za natuknice, običnih kosih za oznaku rekcije), o dosljednom i vrlo minuciozno razrađenom principu uporabe izlomljenih zagrada (za fakultativne dijelove frazema), okruglih zagrada (za zamjenjive sastavnice frazema) te uglatih zagrada kojima se označuje moguća kolokacija frazema, potom o gramatičkim pitanjima vezano za lik natuknice kao i pravopisna pitanja – jednom riječju, o mikrostrukturi frazeoloških rječnika *zagrebačke frazeološke škole* – studiozno i znalački se osvrnula Branka Tafra u radu naslovljenom *Natuknica i nadnatuknica u frazeološkom rječniku iz 2021. godine*.

Natuknički lik frazema odluka je autora rječnika s pečatom *zagrebačke frazeološke škole* i nije nužno rezultat najveće frekvencije upotrebe. Osim toga, iz rječničkog članka nužno se ne iščitavaju ni podaci o varijantama i modifikacijama frazema, već je to prepušteno korisniku za kojeg se pretpostavlja da jezikom vlada u dovoljnoj mjeri kako bi mogao razumjeti te samostalno proizvesti varijaciju⁴ i/ili modifikaciju⁵ određenog frazema, pa čak i kao rezultat ludičke čovjekove aktivnosti kojoj podliježu frazemi jednako kao i leksemi.

Imajući na umu tri načina na koja se mogu koristiti korpusni podaci u leksikografske svrhe, valja reći da *zagrebačka frazeološka škola* koristi korpusno oprimljeni pristup (*corpus-illustrated approach*), odnosno ona pristupa korpusu kako bi se ciljano pronašli primjeri u službi ilustracije određenog frazema (za razliku od pristupa vođenog korpusom (*corpus-driven approach*) i korpusno utemeljenog pristupa (*corpus-based approach*)) (Tognini-Bonelli 2001). U uvodnom dijelu rječnika uvijek se navodi kako su izvori za primjere u rječniku u prvom redu prozna djela, potom publicistika, a onda i primjeri do kojih se došlo pomoću internetskih pretraživača i naposljetku korpusni primjeri.

3. FRAZEOGRAFIJA I IZAZOVI DIGITALNOG DOBA

Elektronička leksikografija ili e-leksikografija razvija se od sredine prošloga stoljeća, no valja istaknuti kako su računala tada bila korištena samo za obradu podataka nakon što su podaci prethodno prikupljeni konvencionalno, odnosno ručno (Štrkalj Despot, Ostroški Anić 2020: 6). Prijelomni trenutak za razvoj e-leksikografije bio je prijelaz prema leksičkoj bazi podataka. Kristina Štrkalj Despot i Ana Ostroški Anić u svom članku iz 2020. godi-

4 Frazeološke su varijante izrazi koji su se ustalili dugom upotrebom, što znači da su određene izmjene leksičkog sastava i strukture frazema prihvaćene jezičnom konvencijom (Parizoska 2019: 2).

5 Frazeološke se modifikacije objašnjavaju kao namjerne i svjesne preinake frazema, pri čemu govornici namjerno mijenjaju leksički sastav, strukturu i značenje frazema u danom kontekstu radi određene komunikacijske svrhe (Parizoska 2019: 2). Treba znati da kognitivnolingvistička istraživanja pokazuju kako modifikacijama frazema upravljaju kognitivni mehanizmi te da ona nisu samo odraz jezične kreativnosti individualnih govornika (*ibid.*).

ne *Pregled razvoja hrvatske e-leksikografije* ističu kako je prijelaz revolucionirao i leksikografsku obradu i prezentaciju građe korisniku, ali i način na koji se korisnik podacima služi (*ibid.*). Osamdesete godine XX. st. donose rječnike u CD-ROM obliku, a u devedesetim godinama pojavljuju se prvi pravi izvorno digitalni mreži rječnici. S obzirom na činjenicu da se u to doba tiskani rječnici najprije konvertiraju u digitalni oblik, još uvijek se ne može govoriti o velikoj razlici između tiskanih i e-rječnika. Korpusna revolucija, koja se smješta u vremenske okvire posljednja dva desetljeća prošloga stoljeća, „pripremila je teren“ za revoluciju u leksikografiji. Za razliku od svjetske leksikografske scene koja se brzo prilagodila korpusnoj revoluciji, hrvatska leksikografija dobila je prvi izvorno digitalni rječnik hrvatskoga jezika tek s tridesetak godina zakašnjenja, iako je hrvatska leksikografija bila pravovremeno uključena u prve korpusne projekte i začetke razvoja discipline e-leksikografije, no svoj golemi leksikografski potencijal, kako zaključuju Štrkalj Despot i Ostroški Anić (*ibid.*: 18), ona nažalost nije iskoristila.

„Ipak, doba suvremene leksikografije doba je korpusne lingvistike i računalne tehnologije, a to su područja u koja se hrvatska frazeografija tek treba uključiti“ (Filipović Petrović 2018: 138). S obzirom na digitalno doba i digitalno okruženja u kojem živimo i kojeg smo dio, i frazeografija iznalazi način da odgovori izazovima digitalnog doba koji se pred nju stavljaju. O potencijalu i uključenosti hrvatske elektroničke frazeografije ili e-frazeografije u svjetske trendove može se ustvrditi isto, odnosno mogu se donijeti vrlo slični zaključci kao i općenito za hrvatsku e-leksikografiju.

Autorski dvojac Parizoska i Filipović Petrović stava su da je svrha frazeološkog e-rječnika prenijeti korisniku informaciju o stvarnoj upotrebi frazema, a to je po mišljenju dviju autorica ona informacija o upotrebi koja proizlazi iz korpusnih podataka. Njihov *Mrežni frazeološki rječnik hrvatskoga jezika* na kojem autorice rade upravo je vođen prezentacijom podataka prikupljenih u hrvatskome mrežnom korpusu hrWaC (Ljubešić, Klubička 2014). Dakle, način bilježenja frazema ovisi o frekvenciji njegova pojavljivanja u korpusu. Sastavljačice rječnika na temelju te frekvencije donose odluku o tome koje oblike nekog frazema zabilježiti kao ustaljene varijante, te ako ih ima više, kojim ih redoslijedom navesti. Prednosti e-rječnika, kako ističu Parizoska i Filipović Petrović, jest to da oni nisu prostorno ograničeni, da natuknice uključuju poveznice, da se frazemi mogu pretraživati po bilo kojoj riječi, a moguće su i dodatne informacije koje ranije nisu nalazile mjesto u frazeološkim rječnicima tradicionalnog tipa. One u svom radu i koncepciji mrežnoga rječnika hrvatskog jezika veliku pozornost posvećuju varijantnim oblicima frazema, njihovim modifikacijama i konceptualnoj organizaciji. Njihov način obrade i bilježenje leksičkih, tvorbenih i leksičko-sintaktičkih varijanata u skladu je s teorijskim postavkama kognitivne gramatike. „Varijante su one konstrukcije koje imaju istu konceptualnu osnovu i sličan leksički sastav, a na različite načine opisuju situaciju koju jezično oblikuje dani frazem. Zbog toga se sve vrste varijanta navode u istoj natuknici“ (Parizoska, Filipović Petrović 2020: 948–949).

4. OD KARTOTEKE DO SKETCH ENGINEA

Željka Fink Arsovski zajedno sa Stefanom Rittgasserom, uvažanim njemačkim slavistom, od 2002. godine radila je na sastavljanju pionirskoga *Hrvatskog frazeološkog kor-*

*pusa*⁶. Koliko nam je poznato, rad na tom korpusu zaustavljen je, a na mrežnim je stranica dostupna probna verzija koja obuhvaća slova A, B, C, Č, Ć, D, DŽ, Đ, E, F, R, Z. Na dostupnoj se stranici navodi da se korisnicima prezentira frazeološko jezično blago hrvatskoga jezika iz suvremenih tekstova (u rjeđim se slučajevima navode primjeri iz XIX. st. te iz prve polovice XX. st.), odnosno iz književnih djela te iz publicističkih tekstova. Vezano za način navođenja frazema, autori slijede sličan princip kao i kod tiskanih frazeoloških rječnika o kojima je bilo riječi ranije u tekstu, odnosno slijede tzv. morfološki princip. Dakle, riječ je o abecedno organiziranoj digitalnoj zbirci frazema koji se crvenim slovima ističu unutar tekstova. Taj popis hrvatskih frazema s ilustracijama iz književnih i publicističkih tekstova nije anotiran, pa stoga nije pogodan za pretraživanje prema, primjerice, nekoj od sastavnica frazema.

U međuvremenu umirovljeni profesor Rittgasser osim što je za svoga radnoga vijeka vrijedno ručno prikupljao hrvatsku frazeološku građu, isto je to posljednjih pedesetak godina činio i s njemačkim frazeološkim blagom i to ponajprije iz raznih novina i časopisa, a potom je svoju građu, koju sam nije leksikografski obrađivao, velikodušno ustupio autorima ovih redaka.

Svjescni korpusne revolucije i digitalnih izazova s kojima se frazeografija susrela, petoročlana je istraživačka skupina prihvatila izazov i odlučila je iznaći način te utabati metodološki put s ciljem da tu kartoteku učini dostupnom širokom krugu korisnika: kako studentima njemačkog jezika u nas i prevoditeljima, tako i izvornim govornicima njemačkog jezika (budući da prikupljena građa predstavlja svojevrsni raritet i za njemačku frazeografiju) te slavistima na njemačkom govornom području, a jednako tako i istraživačima u nas koji se bave područjem frazeologije i frazeografije. Infrastruktura institucijskog projekta pod nazivom *Publicistički potkorpus mrežnoga njemačko-hrvatskog frazeološkog rječnika* u trajanju od svega nekoliko mjeseci omogućila je istraživačkoj skupini da ostvari navedeni cilj za jedan manji dio spomenute kartoteke.

Metodologija rada nad ručno prikupljenim karticama s frazemima iz njemačkih tiskovina uključivala je sljedeće etape: 1) digitaliziranje i uređivanje prikupljene publicističke građe odgovarajućim alatom za optičko prepoznavanje znakova (riječ je o programu *ABBYY FineReader*). Točnost rezultata dobivenih automatskim prepoznavanjem dodatno je provjerena pomnim iščitavanjem teksta i usporedbom s izvornom građom, nakon čega su greške u rezultatima optičkog prepoznavanja ručno ispravljene. Iz novinske građe izbačeni su neiskoristivi ulomci: rukom pisani primjeri, nepotpune rečenice, isječki iz francusko-njemačkoga rječnika i sveze riječi koje prema našem shvaćanju nisu nužno frazemi, o čemu će detaljnije biti riječi u dijelu naslovljenom *Usporedivi publicistički potkorpus njemačkih i hrvatskih frazema*; 2) obradu publicističke građe na način da su se iz novinskih isječaka najprije ekstrahirale frazeološke jedinice, odnosno da su se razdvojile rječničke natuknica od ilustrativnog primjera upotrebe. Izdvajanje njemačkih frazema iz primjera provedeno je uz oslanjanje na njemački frazeološki rječnik *DUDEN Band 11: Redewendungen, Wörterbuch der deutschen Idiomatik.*; 3) pronalazak ekvivalentnih hrvatskih frazema te ilustrativ-

6 *usp.* www.ffzg.hr/hfk/old/ (Datum pristupa: 17. 12. 2021.)

nih primjera upotrebe frazema u publicističkom potkorpusu hrWaC-a (Ljubešić, Klubička 2014). Hrvatski ekvivalentni frazemi zabilježeni su uz oslanjanje na *Hrvatski frazeološki rječnik* (2014) autorskog trojca A. Menac - Ž. Fink Arsovski - R. Venturin; 4) anotiranje frazeološke građe; 5) izradu usporedivoga publicističkog potkorpusa njemačkih i hrvatskih frazema pomoću alata *Sketch Engine*. O posljednjim dvjema etapama rada, odnosno o anotiranju frazeološke građe kao i o izradi usporedivoga publicističkog potkorpusa njemačkih i hrvatskih frazema podrobnije će biti riječi u daljnjem tekstu.

U ovoj je fazi rada u okviru višemjesečnog institucijskog projekta obrađeno ukupno oko 800 kartica iz kojih je ekscerpirano 500-tinjak njemačkih frazema koji su zatim obrađeni i posloženi abecednim redom prema glavnoj sastavnici. Glavni sastavnica je ukupno 200. U usporedivom hrvatskom potkorpusu pridodano je zasad svega 10 % potpunih frazeoloških prijevodnih ekvivalenata. Ipak, svaki hrvatski prijevodni ekvivalent pridružen je samo jednom njemačkom paragrafu. To znači da svoj frazeološki prijevodni ekvivalent ima samo 6,25 % njemačkih segmenata u korpusu.

5. O SKETCH ENGINEU

Sketch Engine korpusni je upitni sustav koji u sebi ima integrirane proširene mogućnosti pretraživanja i ekscerpiranja jezičnih podataka, a sve operacije unutar sustava temelje se na korpusnom upitnom jeziku (*Corpus Query Language - CQL*) (Kilgarriff i sur. 2004: 107). Već je 2014. godine, deset godina od osnutka, postao vodeći alat za stvaranje, obradu i pretraživanje korpusa (*ibid.*: 7). Osnovne mogućnost koje *Sketch Engine* korisniku, zainteresiranome za različite aspekte funkcioniranja jezika, pruža su sljedeće: skica riječi, to jest sažeti prikaz gramatičkih i kolokacijskih odlika određene riječi, razlike u skicama riječi, konkordancija, n-gram, tezaurus te stvaranje jednojezičnih i višejezičnih korpusa.⁷ Sve opcije integrirane u sustav mogu se primijeniti na korpuse sadržane u *Sketch Engineu*, ali i na korpuse koje sami korisnici stvaraju (*Sketch Engine* 2021). Korisnik alata može stvoriti korpus koristeći vlastite dokumente ili prikupljanjem javno dostupne tekstove s Interneta. Korpusi koje korisnici samostalno učitavaju u *Sketch Engine* prolaze kroz proces lematizacije, označivanja vrsta riječi (*part-of-speech tagging - POS-tagging*) i morfosintaktičkih oznaka (*morpho-syntactic tagging*). Kako bi se mogao zadati upit pomoću leme ili oznake, svakoj pojavnici unutar korpusa mora biti pridodana informacija o njenom cjelovitom flektivnom obliku i pripadnosti vrsti riječi. Kako je riječ o jezično specifičnim informacijama, alati se za svaki jezik razvijaju neovisno o *Sketch Engineu* te se mogu naknadno uključiti u taj sustav. Kada korisnik stvara novi korpus, njemu se predlaže koji sustav lematiziranja, označivanja vrsta riječi i morfosintaktičkih oznaka treba primijeniti za pojedini jezik. Ukoliko korisnik ne želi koristiti sustave anotiranja dostupne na *Sketch Engineu*, on može samostalno označiti tekst i, formatiravši ga prema zadanim parametrima, učitati ga u program. Za hrvatski jezik predlaže se skup oznaka *MULTEXT-East Morphosyntactic Croatian Spe-*

7 Prijevodi engleskih termina vezanih za korištenje *Sketch Engineom* temelje se na pojmovniku sastavljenom u okviru projekta Hrvatski mrežni rječnik (MREŽNIK) koji financira Hrvatska zaklada za znanost. Pojmovnik je dostupan na mrežnim stranicama projekta: <http://ihjj.hr/mreznik/page/pojmovnik/6/> (Datum pristupa: 16. 12. 2021.)

cification, puni popis gramatičkih oznaka sastavili su Nikola Ljubešić i Tomaž Erjavec i dostupan je na mrežnoj stranici Grupe za obradu prirodnoga jezika pri Filozofskom fakultetu Sveučilišta u Zagrebu.⁸ Za njemački jezik pri izradi korpusa u *Sketch Engineu* predložen je skup oznaka *German RFTTagger part-of-speech tagset*,⁹ kojega su razvili Helmut Schmid i Florian Laws (Sketch Engine 2021). Detaljan popis tumačenja morfosintaktičkih oznaka dostupan je u vodiču za korištenje *Sketch Enginea*.¹⁰ Poznavanje popisa morfosintaktičkih oznaka potrebno je korisniku za napredno pretraživanje korpusa pomoću upita zadanoga CQL-om.

5.1. Anotacija prikupljene publicističke građe

Anotacija ili označavanje korpusa proces je dodavanja dodatnih informacija o tekstu. Informacije se mogu dodavati strukturama (dokumentima, paragrafima, rečenicama itd.) ili pojavnica (Sketch Engine 2021). Pojavnica se obično pridodaju informacije o pripadnosti vrsti riječi ili njihovoj sintaktičkoj funkciji, a strukturama se može dodati sijaset informacija. Unutar *Sketch Enginea* postoji ugrađeni anotacijski alat. Pomoću njega mogu se anotirati pojavnice i dodati novi tipovi metapodataka o učitanim dokumentima (godina objavljivanje, izbor, stilski registar itd.), ali ne i strukture u korpusu (*ibid.*). K tome, anotirati se mogu isključivo korpusi koje korisnici sami stvaraju. Budući da anotacija frazeološke građe podrazumijeva označavanje skupa pojavnica kao određene strukture, ugrađeni anotacijski alat unutar *Sketch Enginea* nije moguće koristiti za tu svrhu, već je neophodno anotirati dokumente izvan njega. Dokumenti, rečenice i paragrafi obično se automatski anotiraju pri učitavanju korpusa, ali korisnik može po potrebi ručno označiti njihove granice. Anotacijska sintaksa nije proizvoljna, već se zasniva na proširivom jeziku za označavanje, XML-u (*eXtensible Markup Language - XML*). Anotacijske oznake elementi su XML dokumenta. Početak elementa označuje se početnom oznakom, to jest, odabranim nazivom zapisanim unutar šiljastih zagrada (<frazem>). Završna oznaka razlikuje se od početne samo po kosoj crti prije naziva elementa (</frazem>), a unutar njih nalazi se sadržaj elementa (Kirašić 2005: 80). Korisnik XML-a samostalno odlučuje kako želi nazvati svoje anotacijske oznake, te što će u njih uključiti. U anotaciju je moguće je uključiti i informaciju o elementima. Ona podrazumijeva dodavanje atributa i vrijednosti elementima. Primjerice, pri anotiranju frazema mogle bi se posebno označivati glavne i sporedne sastavnice, što bi korisniku omogućilo zadavanje upita o položaju jedne sastavnice frazema u odnosu na bilo koju drugu jedinicu korpusa. Primjer tako označenoga tekstnog odsječka izgleda ovako:

```
Ganz <frazem type="sporedne_sastavnice"> von der </frazem> <frazem
type="glavna_sastavnica"> Hand </frazem> <frazem type="sporedne_sastav-
nice"> zu weisen ist </frazem> diese Möglichkeit <frazem type="sporedne_sa-
stavnice"> nicht </frazem>.
```

8 <http://nlp.ffzg.hr/data/tagging/msd-hr.html> (Datum pristupa: 26. 1. 2022.),

9 Više o razvoju i primjeni alata može se pročitati u članku: Schmid; Helmut; Laws, Florian, „Estimation of Conditional Probabilities with Decision Trees and an Application to Fine-Grained POS Tagging“, COLING 2008, Manchester, Great Britain.

10 <https://www.sketchengine.eu/german-rftagger-part-of-speech-tagset/> (Datum pristupa: 16. 12. 2021.)

U danome prikazu **frazem** je oznaka, **type** je atribut, a **glavna i sporedne sastavnice** vrijednosti su atributa **type**. Pri anotaciji naše publicističke građe nije provedena dodatna klasifikacija, već je u građi dodijeljena samo oznaka **frazem**. U trenutnoj fazi rada na pilot-projektu oznaci je dan naziv **frazem** u skladu s ustaljenom frazeološkom terminologijom u hrvatskom jeziku, ali, ukoliko bi se rad nad projektom nastavio, nazivi anotacijskih oznaka uskladili bi se sa smjernicama udruženja *Text Encoding Initiative*¹¹, kako bi i resursi stvoreni u okviru projekta mogli komunicirati s drugim sličnim resursima. Ako će se u kasnijim fazama rada nad građom pojaviti potreba za detaljnijom klasifikacijom, ona će se također uvesti u skladu sa smjernicama spomenutog udruženja. U do sada obrađenoj građi nije se pojavio slučaj da se u istom odlomku pojavljuju dva frazema. Ako bi se pojavio takav slučaj radu nad građom, pokušao bi se pronaći način nadilaženja tog problema u anotacijskoj praksi primijenjenoj u projektima koji nalikuju našem.

Svaka kategorizacija korpusnih podataka proces je interpretacije, a za uspješnu interpretaciju potrebno je jasno izdvojiti i opisati parametre koji se pri njoj uzimaju u obzir. Taj metodološki postupak olakšava precizno definirana anotacijska shema. Anotacijska se shema može predočiti kao iscrpna operativna definicija određene varijable s detaljnom uputom kako treba vrijednosti varijable pripisati jezičnim, tj. korpusnim podacima. Ona obično uključuje shemu kodiranja koja specificira oznake kojima će se prikazati jezične kategorije u korpusu (Stefanowitsch 2020: 120–127). U publicističkoj građi koju je sabrao profesor Rittgasser pridružene su anotacije za tri tipa struktura: frazeme, paragrafe i dokument. Čitava građa anotirana je kao jedan dokument, rečenice nisu ručno anotirane jer se oznake za početak i kraj rečenice automatski pridodaju pri učitavanju korpusa u *Sketch Engine*. Anotacijska oznaka za paragraf dodana je svakom novinskom isječku, a o sastavu anotiranih frazema više će biti riječi u sljedećem odlomku. Kako bi se izbjegle greške pri ručnom unošenju anotacijskih oznaka u korpus, koristio se uređivač jasnopisa (*plan text editor*) Notepad++. U Notepad++ instaliran je WebEdit dodatak te je Notepad++ pomoću tog dodatka konfiguriran tako da se može lako označiti niz pojavnica zadanim oznakama. Taj dodatak također omogućuje postavljanje dodatnih tipki pomoću kojih se bez tipkovnih prečaca pridjeljuju anotacijske oznake. Paragrafi se omeđuju oznakama `<p></p>`, dokumenti `<d></d>` i frazemi `<frazem></frazem>`. Prilikom pretrage pojedine sastavnice u anotiranom korpusu potrebno je zadati upit koji uključuje anotacijsku oznaku za frazem.

5.2. Sastav anotiranih frazema

U ruskoj sintaktičkoj tradiciji za svrstavanje skupa tagmema u rečenične dijelove ključno je razlikovanje slobodnih i neslobodnih sintagmi. Za slobodne sintagme tipično je da leksičke jedinice koje ulaze u njen sastav čuvaju svoje samostalno leksičko značenje i sastavnice takve sintagme mogu se zamijeniti riječima koje pripadaju istoj kategoriji. Neslobodne sintagme sastoje se od leksički nesamostalnih riječi. U rečenici se sastavnice neslobodnih sintagmi ne raščlanjuju, nego funkcioniraju kao jedan cjelovit rečenični dio (Валгина 2003: 35–36). Frazemi pripadaju neslobodnim sintagmama i upravo je to njihovo svojstvo jedna

11 <https://tei-c.org/> (Datum pristupa: 15. 12. 2021.)

od ključnih osobina pri definiraju frazema.¹² Primjerice, u frazemu *dati / davati u dobre ruke* koga, što (HFR 2014: 508), sastavnica *u dobre ruke* nije priložna oznaka, već dio nerazdvojive cjeline. Bez obzira na to što se iz dijakronijskoga gledišta ponekad mogu odrediti tipovi gramatičkih veza između pojedinih tagmema u sastavu frazema, te veze u suvremenom jeziku gube svoju produktivnost. Frazemi se u rečenici ponašaju kao nerazdvojive cjeline i imaju vlastita valencijska svojstva (Шведова и др. 1980: 25). Stoga, riječi označene kurzivom u frazeološkoj rječničkoj natuknici koja se navodi niže ukazuju na produktivna valencijska svojstva frazema i klasificiraju se kao samostalni rečenični dijelovi pri sintaktičkoj analizi. U primjerima upotrebe frazema u *Hrvatskom frazeološkom rječniku* izdvajaju se samo sastavnice frazema, a nisu uključene riječi kojima frazem kao značenjski nedjeljiva cjelina otvara mjesto u rečenici, tj. sadržaj ispunjenih valencijskih mjesta frazema. To je zamjetno u sljedećim primjerima:

dati / davati (staviti / stavljati) ruku u vatru <za koga, za što; da...>
 jamčiti (garantirati) <za koga; da>, imati puno povjerenje <u koga, da...>
 ...gore nema nitko ništa protiv tebe, čak ni Učo, jer on nije zloпамtilo, iako, naravno, nitko u ovoj situaciji NE BI za tebe STAVLJAO RUKU U VATRU.
 (Bauer)
 Za Banovine nije više značio ništa, ustaše bi ga vjerojatno ubili da nije već bio na samrti. To je, naravno, samo pretpostavka za koju ne bih mogao STAVITI RUKU U VATRU. (HFR 2014: 508)

U prvom primjeru u sastav frazema uključena je negacija i puni flektivni oblik glagola, ali ne i dopune kojima mjesto u rečenici otvara frazem. U drugom primjeru modalni glagol u kondicionalu I. nije uključen u sastav frazema, iako on u rečeničnom ustrojstvu zajedno s infinitivom suznačnoga glagola čini složeni glagolski predikat (Silić, Pranjković 2005: 289).

Takav princip označivanja frazema, međutim, nije prihvaćen pri anotiranju sastava frazema tijekom rada na spomenutom pilot-projektu. Unutar anotacijskih oznaka uključene su sve dopune kojima frazem otvara mjesto u rečenici kao što se vidi iz sljedećeg primjera:

<p> *Paladino se doima pošten čovjek, a da je stručnjak, to svatko zna. No, <frazem> ne bih stavio ruku u vatru ni za koga </frazem>, a stoji i ministрова da Rebro ima goleme gubitke i da se mora raditi po zakonu, ali sam Hebrangov nastup, koji djeluje osvetnički, apsolutno ide na živce.* </p>

U njemačkom frazeološkom rječniku *DUDEN Band 11: Redewendungen, Wörterbuch der deutschen Idiomatik* sastavnice frazema u primjerima upotrebe nisu ni na koji način označene.

die / seine Hand für jmdn., etw. ins Feuer legen: sich vorbehaltlos und uneingeschränkt für jmdn., etw. verbürgen: Wir kennen sie, seit sie in die Stadt gekommen ist, und legen jederzeit die Hand für sie ins Feuer (Brecht, Mensch 25). (DUDEN: 319)

12 Fink-Arsovski (2002: 6–7) vršenje sintaktičke funkcije kao dio rečeničnoga ustrojstva uvrštava u kriterije koje je neophodno zadovoljiti kako bi se neki skup riječi mogao smatrati frazedom.

Pri anotiranju njemačkih frazema primjenjivao se isti princip kao i pri anotiranju hrvatskih frazema; označene su sve dopune i skupovi dopuna koje ispunjavaju valencijska mjesta frazema.

<p> Aber hast du auch vermutet, daß er dich betrügt? RUTH: Was einen Seitensprung angeht, na ja, <frazem> **die Hand hätte** </frazem> ich <frazem> **für ihn** </frazem> wohl <frazem> **nicht ins Feuer gelegt** </frazem>. Aber ich habe ihn nicht für einen notorischen Fremdgeher gehalten. </p>

Označivanje dopuna frazema pruža korisniku mogućnost olakšane klasifikacije gramatičkih i ontoloških svojstava frazeoloških dopuna i pomaže odrediti kojim tematskim ulogama pripadaju dopune te otvara mogućnost odgovaranja na mnoga istraživačka pitanja, naprimjer, jesu li češće dopune koje označuju živoga ili neživoga referenta te mijenja li se značenje frazema ovisno o njegovim dopunama.

5.3. Pretraživanje korpusa

Za pretraživanje anotiranih frazema u *Publicističkom potkorpusu mrežnoga njemačko-hrvatskog frazeološkog rječnika* najvažnija je opcija unutar *Sketch Enginea* konkordancija. Ona služi za pretraživanje korpusa zadavanjem upita o lemapa, pojavnica, skupovima riječi i oznakama te o dokumentima i korpusnim strukturama unutar kojih se tražene jezične jedinice nalaze. Za pretraživanje lema i pojava dovoljno je upisati traženu riječ na prazno polje unutar sučelja pod opcijom *concordance*, a za pristup svim ostalim podacima u korpusu potrebno je zadati upit pomoću CQL-a. Rezultati bilo koje pretrage pokazuju se u kontekstu upotrebe i korisnik ih dodatno može obrađivati razvrstavanjem, anotiranjem, izradom frekvencijske distribucije i filtriranjem. Integracija metapodatka o sastavu frazema u korpus provodila se uzimajući u obzir mogućnosti pretrage koje CQL nudi korisniku unutar *Sketch Enginea*. CQL u *Sketch Engineu* temelji se na inačici koju su razvili Oliver Christ i Bruno M. Schulze¹³. Za pretraživanje anotiranih struktura u korpusu najvažniji operatori su *within* i *containing* (Jakubiček i sur. 2010: 742). Za pretraživanje struktura i/ili nizova anotiranih pojava, upit treba sadržavati atribute i njima pridružene vrijednosti te podatak o njihovom položaju unutar određene strukture. Za pronalazak leme *ruka* potrebno je zadati sljedeći upit.

[attribute="value"] → [lemma="ruka"]

Prikazani upit pronaći će sve primjere pojavljivanja leme *ruka* u korpusu. Ukoliko korisnik želi pronaći sve primjere te leme u sastavu frazema, u upitu je neophodno označiti da se lema nalazi unutar anotacijske oznake **frazem**. To se postiže uvrštavanjem operatora *within* ili *containing* u upit. Operator *within* označuje da se pronađeni niz znakova nalazi unutar granica zadane strukture (*ibid.*). Pretraga pomoću operatora *within* zadaje se tako

13 Više o korištenju te inačice CQL vidi u: Oliver Christ, Oliver; Schulze, Bruno M.; Hofmann, Anja and Esther Koenig, *The IMS Corpus Workbench: Corpus Query Processor (CQP): User's Manual. Technical Report*, University of Stuttgart: Institute for Natural Language Processing, Stuttgart, Germany, 1999.

da se prije njega upisuje pojava, lema ili oznaka, a nakon njega naziv tražene strukture, npr.

`[tag="N.*"] within <frazem/>`

Operator *containing* ima gotovo jednaku sintaksu kao i operator *within*. Upit se razlikuje po redoslijedu upisivanja podataka; naziv strukture potrebno je upisati prije operatora, a traženu jezičnu jedinicu nakon operatora, npr.

`<frazem/> containing [word="Hand"] [tag="V.*"]`

Tako zadani upit kao rezultat vraća zadani obrazac omeđen strukturom `<frazem/>`, odnosno sastavnicu *Hand* nakon koje u linearnom poretku slijedi bilo koji flektivni oblik glagola u sastavu frazema. Za prikaz granica frazema u rezultatu pretrage potrebno je odabrati strukturu *frazem* pod *View options* u sučelju *Sketch Enginea*. Isti upit može se zadati za jednojezični, ali i za usporedivi korpus. Za pretraživanje usporedivoga korpusa treba odabrati opciju *Parallel Concordance*, zatim kao tip upita označiti CQL, a ostatak pretrage jednak je kao i u jednojezičnom korpusu.

S obzirom na obilježja *Publicističkoga potkorpusa mrežnoga njemačko-hrvatskog frazeološkog rječnika*, on više odgovara definiciji usporedivoga, nego usporednoga korpusa. Prema Bratanić (1998: 174) usporedivi korpusi jesu korpusi različitih jezika sastavljeni prema identičnom obrascu, a usporedne (paralelne) korpuse čine tekstovi na jednom jeziku i njihovi prijevodi na drugi jezik, a namijenjeni su ponajprije uspostavljanju prijevodne ekvivalencije. Građa je po jednakom kriteriju odabrana za uvrštavanje u *Publicistički potkorpus mrežnoga njemačko-hrvatskog frazeološkog rječnika* (primjeri upotrebe frazema u publicistici), a i frazemi su po istoj anotacijskoj shemi obilježeni u rečenicama na hrvatskome i njemačkome jeziku. Ipak, primjeri upotrebe nisu u potpunosti prevedeni, već samo sadrže frazeološke prijevodne ekvivalente. Iz tih razloga *Publicistički potkorpus mrežnoga njemačko-hrvatskog frazeološkog rječnika* zapravo bi valjalo nazvati usporedivim korpusom anotiranih njemačkih i hrvatskih frazema.

6. USPOREDIVI PUBLICISTIČKI POTKORPUS NJEMAČKIH I HRVATSKIH FRAZEMA

Tijekom rada na pripremanju njemačkoga korpusa u svrhu sastavljanja usporedivoga publicističkog potkorpusa njemačkih i hrvatskih frazema u *Sketch Engineu* otvorio se niz frazeoloških pitanja među kojima je i ono o donjoj i gornjoj granici frazema. Unatoč činjenici da se višerječnost izdvaja kao jedna od osnovnih formalnih značajki frazema, u ručno prikupljenoj Rittgasserovoj frazeološkoj građi zamijetili smo i jednorječne (leksičke) frazeme koje smo objeručke prihvatili kao donju, minimalnu granicu frazema jer dijelimo Čermákov stav o jednorječnim (leksičkim) frazemima koje on naziva *lexical idioms* ili *cranberry idioms* (usp. Čermák 2007), te stav Elisabeth Piirainen koja naglašava kako bi uključivanje metaforičkih jednorječnih jedinica (*figurative one-word units*) u polje istraživanja moglo obogatiti naše znanje o metaforičnosti jezika i njegovoj vezi s kulturološki relevantnim tekstovima (usp.

Piirainen 2012: 37). Tako su u korpus uključeni njemački leksički frazеmi poput *Blaustrumpf*, *butterweich*, *glasklar* i *Schlitzohr*. Kada je riječ o gornjoj granici frazеma, odnosno, što se tiče omeđivanja maksimalne širine frazeološkog polја, Rittgasserovo i naše se shvaćanje razilaze. Usprkos tomu mi smo ipak odlučili ostaviti poslovice u korpusu, dok smo lingvokultureme¹⁴ ostavili postrani, tj. nismo ih uključili u korpus. Tako je u konačnici, u korpus, primjerice, uključena paremiološka jedinica *Man muss das Eisen schmieden, solange es heiß ist* koja ima hrvatski ekvivalent *Željezo se kuje dok je vruće*, a iz korpusa su ispali sljedeći lingvokulturemi: *die Gretchenfrage*; *wie Hans Glück-in-die-Luft*;, *ach du, mein Sohn Brutus*. Potonji primjer ima hrvatski ekvivalent *Zar i ti, sine Brute?* To je jedan od najpoznatijih citata Julija Cezara, odnosno riječi koje je Cezar na samrti uputio jednom od urotnika, Marku Juniju Brutu, a radilo se o dramatičnim okolnostima atentata i izdaji. I danas se taj lingvokulturem koristi u situaciji kada se komentira nečija neočekivana izdaja ili ponašanje. Što se tiče preostala dva navedena lingvokulturema, za razliku od Cezarovog citata oni nisu dio internacionalnoga jezičnog sloја, već su usko vezani za kulturu njemačkoga govornog područја te ih poznaju nositelji njemačke kulture. *Die Gretchenfrage* ima značenje ‘odlučujuće, ključno pitanje’, a etimološki je vezano uz *Fausta* Johanna Wolfganga von Goethea. Mlada djevojka Margareta, odnosno Greta, kojoj religija igra veliku ulogu u životu, Faustu postavlja sljedeće pitanje: *Nun sag, wie hast du's mit der Religion? Du bist ein herzlich guter Mann, allein ich glaub, du hältst nicht viel davon*. „Gretino pitanje“ tako je postalo sinonim za svako odlučujuće, tj. ključno pitanje, i nije, dakako, vezano samo za pitanje o stavu prema religiji. S obzirom da ga mi poimamo kao lingvokulturem, nismo ga uključili u korpus, a nije zanemariv ni podatak da ta jezična jedinica nije zabilježena u njemačkim frazeološkim rječnicima kao ni iduća poredbena jedinica. Poredbena konstrukcija *wie Hans Glück-in-die-Luft* ima značenje ‘poput nepažljivog dječaka koji ne gleda kuda hoda pa posrće; poput sanjara’, a etimološki se vezuje uz dječju priču *Die Geschichte vom Hanns Guck-in-die-Luft* iz XIX. st. Heinricha Hoffmanna. Riječ je o dogodovštinama dječaka Hansa koji gleda uvis (pa otuda i dio *Guck-in-die-Luft*), hoda uokolo te posrće. No u navedenoj se poredbenoj konstrukciji *wie Hans Glück-in-die-Luft* dogodila kontaminacija s naslovom bajke braće Grimm *Hans im Glück*.

U korpusu su ostali uključeni tzv. uzvični frazеmi kojima se, kako je u frazeologiji poznato, izražava odnos govornika prema onome što je izrečeno (i to u obliku replike), ili prema nekom zbivanju ili pojavi. Vrijedi napomenuti kako se za tu vrstu frazеma u rječnicima ne donosi značenje, već se uz njih pronalazi pragmatička informacija o situacijama u kojima se obično upotrebljavaju. Većina je njih u ovom dijelu Rittgasserovo ručno prikupljene građe sa sastavnicom *Gott: Gott hab ihn selig!; Gott im Himmel!; Gott sei Dank!; um Gottes Willen!*.

S obzirom na to da je njemačka publicistička građa prikupljana posljednjih 50-ak godina, vrijedi napomenuti da se frazеmi bilježe onako kako smo ih zatekli u primjerima upotrebe, odnosno ne slijede pravila koje je unijela Reforma njemačkoga pravopisa iz 1996. godine. To su primjeri poput: *aus einem Guß* (nakon pravopisne reforme: *aus einem Guss*) ili pak primjer *an Hand von* (nakon pravopisne reforme: *anhand*) koji danas nema status frazеma.

14 Za više informacija o lingvokulturemima v. Barčot 2017: 45–68.

Prema ranije opisanoj metodologiji te imajući na umu netom opisane kriterije za uključivanje ili neuključivanje određenih jedinica u naš korpus, u ovoj smo fazi rada nad Rittgasserovom kartotetokom uspjeli obraditi točno 445 frazema. Za nešto više od 10 % tih frazema pronađeni su potpuni ekvivalenti u hrvatskom jeziku koji su potom ilustrirani publicističkim primjerima iz hrWaC-a (Ljubešić, Klubička 2014), te su u konačnici spojeni u usporedivi publicistički potkorpus njemačkih i hrvatskih frazema u *Sketch Engineu*. U svrhu opimjerivanja svega ranije opisanog u nastavku donosimo tablicu s nekoliko primjera iz usporedivog potkorpusa njemačkih i hrvatskih frazema:

<p><p> Hier <frazem> zeigt </frazem> die Politik der maoistischen Gruppen <frazem> ihr wahres Gesicht</frazem>. </p></p>	<p><p>Koalicija <frazem> je pokazala svoje pravolice </frazem>u manje od 6 mjeseci. Socijalno su neosjetljivi, na izazove i probleme u gospodarstvu odgovaraju najprimitivnijim metodama, teroriziraju građane, obrtnike i gospodarstvenike sa ekstremno visokim i novim nametima, prijete se svima i prodaju loše priče u koje više nitko živ ne vjeruje.</p></p>
<p><p> Aus der Sicht der Linken kommt der Zwist im Regierungslager wie <frazem> ein Geschenk des Himmels </frazem>. Die Streiter Mitterands planen für den Fall eines Sieges bei den Parlamentswahlen eine Konfrontation mit dem Präsidenten, um ihn zum Rücktritt zu zwingen. </p></p>	<p><p>A njegov govor je ono za čim čezne svaki novinar jer zna da je to tema za naslovnicu medija, da je to <frazem> dar s neba </frazem> i dovoljan materijal da sam ispuni naslovnicu i osigura čitanost u nekoliko sljedećih brojeva.</p></p>
<p><p> WIDDER 21. März - 20. April: Alle geldlichen Angelegenheiten nur nach reiflicher Überlegung abwickeln. Glückschancen im Beruf. In der Liebe dürfen Sie sich von Ihrer großzügigsten Seite zeigen! Vorsicht vor falschen Freunden und großen Anschaffungen! Das <frazem> bringt Sie </frazem> nämlich schnell <frazem> aus dem Gleichgewicht </frazem>. </p></p>	<p><p>Nedostatak energije i poslovi koji vas očekuju mogu <frazem> vas izbaciti iz ravnoteže </frazem>. Nemojte gubiti samokontrolu i nastojte se organizirati.</p></p>
<p><p> Mein Leben ist so einsam - in Hasselwert fehlt die Herrin – ich <frazem> würde Sie auf Händen tragen </frazem> - Sie sind mir sehr, sehr teuer - jeden Wunsch würde ich Ihnen erfüllen. </p></p>	<p><p>U Splitu mi je bilo fenomenalno, tamo <frazem> su me </frazem> maltene <frazem> nosili na rukama </frazem>, osjećao sam ljubav na svakom koraku dok sam prolazio rivom “, kaže pjevač, koji je tamo nedavno snimio spot s Jelenom Rozgom za njihov duet “Ima nade”.</p></p>

7. UMJESTO ZAKLJUČKA

Važno je istaknuti da je ovaj kratkoročni projekt ispunio sve planirane ciljeve te utabao metodološki put za obradu ukupne zaprimljene kartoteke s frazeološkom građom. Stoga ga preferiramo nazivati pilot-projektom.¹⁵ S obzirom na ambicioznost iz idejne faze projekta, a uzimajući u obzir mogućnosti i realno ostvarene ciljeve, projekt bi valjalo preimenovati u *Usporedivi publicistički potkorpus njemačkih i hrvatskih frazema*. Za potpunu obradu cjelokupne Rittgasserove kartoteke te pronalazak hrvatskih frazeoloških ekvivalenata s ilustrativnim primjerima upotrebe, a potom anotiranje građe i popunjavanje postojećega usporedivog publicističkog potkorpusa u *Sketch Engineu*, potrebno je puno duže vremensko razdoblje i angažman veće radne skupine.

Ipak, inovativnost i uspješnost sastavljanja zasebnog tipa paralelnoga potkorpusa, odnosno usporedivoga publicističkog potkorpusa s anotiranom frazeološkom građom odlika je ovog pilot-projekta te ga zato smatramo zalogom za budućnost, odnosno zalogom da se aspiracije za jasnim okvirom i infrastrukturnom podrškom ovoj zamisli u vidu višegodišnjega znanstvenog projekta u budućnosti i ostvare.

Upravo ovakav jedan usporedivi frazeološki potkorpus može prevoditeljima dobro poslužiti kao dopuna postojećim frazeološkim rječnicima – ponajprije onim tiskanima – jer uključuje frazeološke varijante ali i modifikacije frazema u odnosu na kanonske oblike koji se bilježe u rječnicima. U usporedivom frazeološkom potkorpusu ovoga tipa pokazuju se frazemi u njihovom „živom okruženju“ jednog stilskog registra u dvama jezicima (u ovom konkretnom slučaju u publicističkom). Stoga bi u budućnosti bilo poželjno nastaviti s radom na ovom projektu te obogatiti njemačku i hrvatsku frazeološku i frazeografsku scenu jednim dodatnim izvorom vrijednih publicističkih oprimjerenja i to u mrežnome ruhu.

15 Više o istraživačkoj skupini, te o rezultatima pilot-projekta i njihovoj diseminaciji može se saznati na sljedećoj poveznici: http://www.ffzg.unizg.hr/slaven/istocna/?page_id=4849 (Datum pristupa: 15. 12. 2021.)

LITERATURA

- Barčot, Branka, *Lingvokulturologija i zoonimska frazeologija*, Hrvatska sveučilišna naklada, Zagreb, 2017.
- Bratanić, Maja, „Korpusna lingvistika na kraju 20. stoljeća i implikacije za suvremenu hrvatsku leksikografiju“, *Filologija*, 30-31, 1998., str. 171–177.
- Čermák, František, „Idioms and morphology“, u: *Phraseologie: ein internationales Handbuch der zeitgenössischen Forschung*, Burger, H., Dobrovolskij, D., Kühn, P., Norrick, N.R. (Hg.), 1. Halbband, Walter de Gruyter, Berlin – New York, 2007., str. 20–27.
- Duden = Duden. *Redewendungen. Wörterbuch der deutschen Idiomatik*. Band 11, hrsg. von der Dudenredaktion, Mannheim, Leipzig, Wien, Zürich: Dudenverlag, 2002.
- Filipović Petrović, Ivana, *Kada se sretnu leksikografija i frazeologija: o statusu frazema u rječniku*, Srednja Europa, Zagreb, 2018.
- Fink-Arsovski, Željka, *Poredbena frazeologija: pogled izvana i iznutra*, FF press, Zagreb, 2002.
- HFR = Menac, Antica; Fink-Arsovski, Željka; Venturin, Radomir, *Hrvatski frazeološki rječnik*, Naklada Ljevak, Zagreb, 2014.
- Jakubiček, Miloš; Kilgarriff, Adam; McCarthy, Diana; Rychlý, Pavel, „Fast syntactic searching in very large corpora for many languages“, u: *Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation*, Otoguro, Ryo; Ishikawa, Kiyoshi; Umemoto, Hiroshi; Yoshimoto, Kei; Harada, Yasunari (ur.), Waseda University, Sendai, 2010., str. 741–747.
- Kilgarriff, Adam; Rychlý, Pavel; Smrz, Pavel; Tugwell, David, „The Sketch Engine“, u: *Proceedings of the 11th EURALEX International Congress*, Williams, Geoffrey; Vessier, Sandra (ur.), Université de Bretagne-Sud, Lorient, 2004., str. 105–116.
- Kilgarriff, Adam; Baisa, Vít; Bušta, Jan; Jakubiček, Miloš; Kovář, Vojtěch; Michelfeit, Jan; Rychlý, Pavel; Suchome, Vít, „The Sketch Engine: ten years on“, *Lexicography Journal of ASIALEX* 1, 2014., str. 7–36.
- Kirasić, Damir, „XML tehnologija i primjena u sustavima procesne informatike“, u: *Proceedings of the 28th International Convention MIPRO 2005*, Kljaić, Josip; Baranović, Neven (ur.), Croatian Society for Information and Communication Technology, Electronics and Microelectronics - MIPRO, Rijeka, 2005., str. 79–88.
- Ljubešić, Nikola; Klubička, Filip, „{bs,hr,sr}WaC - Web Corpora of Bosnian, Croatian and Serbian“, u: *Proceedings of the 9th Web as Corpus Workshop (WaC-9)*, Bildhauer, Felix; Schäfer, Roland (ur.), 2014., str. 29–35.
- Parizoska, Jelena, *Promjenjivost glagolskih frazema u engleskom i hrvatskom jeziku – kognitivnolingvistička analiza*, Sveučilište u Zagrebu, 2019. (doktorski rad)
- Parizoska, Jelena; Filipović Petrović, Ivana, „Varijantni frazemi u e-rječniku“, *Rasprave Instituta za hrvatski jezik i jezikoslovlje*, vol. 46, br. 2, 2020., str. 941–955.
- Piirainen, Elisabeth, *Widespread Idioms in Europe and Beyond: Towards a Lexicon of Common Figurative Units*. Peter Lang Publishing Inc, New York, 2012.
- Pranjković, Ivo; Silić, Josip, *Gramatika hrvatskoga jezika: za gimnazije i visoka učilišta*, Školska knjiga, Zagreb, 2005.
- Stefanowitsch, Anatol, *Corpus linguistics: A guide to the methodology*, Language Science Press, Berlin, 2020.
- Štrkalj Despot, Kristina; Ostroški Anić, Ana, „Pregled razvoja hrvatske e-leksikografije“, u: *Svijet od riječi. Terminološki i leksikografski ogledi*, Brač, Ivana; Ostroški Anić, Ana (ur.), Institut za hrvatski jezik i jezikoslovlje, Zagreb, 2020, str. 5–24.

Tafra, Branka, „Natuknica i nadnatuknica u frazeološkom rječniku“, *Južnoslovenski filolog*, 77(2), 2021, str. 35–59.

Tognini-Bonelli, Elena, *Corpus linguistics at work*, John Benjamins Publishing Company, Amsterdam i Philadelphia, 2001.



Валгина, Нина Сергеевна, *Современный русский язык: Синтаксис, 4-е издание*, Высшая школа, Москва, 2003.

Шведова, Наталия Юльевна; Арутюнова, Нина Давидовна; Бондарко, Александр Владимирович; Иванов, Валерий Васильевич; Лопатин, Владимир Владимирович; Улуханов, Игорь Степанович; Филин, Федот Петрович, *Русская грамматика. Том II. Синтаксис*, Наука, Москва, 1980.



Sketch Engine. Dostupno na: <https://www.sketchengine.eu/> (Datum pristupa: 15. 12. 2021.)

ФРАЗЕОГРАФИЯ В ЦИФРОВОМ НАРЯДЕ – ОТ КАРТОТЕКИ ДО SKETCH ENGINE

В настоящей статье представляется развитие фразеографии в Хорватии: от бумажных словарей, созданных «в мастерской» *Загребской фразеологической школы*, до электронной фразеографии цифровой эры, а потом рассматриваются результаты работы над пилотным проектом, основывающемся на вручную собранной картотеке немецких фразеологизмов. Описываются все фазы пилотного проекта, т.е. все фазы работы над фразеологическим материалом: оцифровка картотеки и обработка оцифрованных документов; фразеологическая обработка материала; поиск хорватских фразеологических эквивалентов; аннотирование фразеологизмов; создание сравнимого публицистического подкорпуса немецких и хорватских фразеологизмов с помощью Sketch Engine. Цель данной работы – во-первых, представить путь развития фразеографии в Хорватии за последние 50 лет и, во-вторых, показать каким способом можно картотеку, собранную традиционно, использовать и «одеть в современный цифровой наряд». Значит, основная выгода текста состоит в том, что мы хотим представить новый методологический путь для создания двуязычного корпуса, который прежде всего пригодится студентам немецкого языка в Хорватии и переводчикам с немецкого языка и на немецкий, а также и носителям языка, и славистам в немецкоговорящих странах. Упомянутый методологический путь возможно заинтригурует и наших исследователей, интересующихся фразеологией и фразеографией, так как, поскольку нам известно, в Хорватии такого корпуса не существует.

Ключевые слова: традиционная фразеография; электронная фразеография; публицистический подкорпус немецких фразеологизмов; публицистический подкорпус хорватских фразеологизмов; сравнимый фразеологический корпус; Sketch Engine