

# Frazemi i druge višerječne lekseme u dvojezičnoj leksikografiji

(Danko Šipka, Chandler)

## 1. Uvodne napomene

Ovdje se, na primjeru dva završena leksikografska projekta (Šipka, 2000 i Šipka, 2002) te jednog projekta u toku (Šipka, 2008) pokazuju specifičnosti obrade frazema u dvojezičnoj leksikografiji. Iskustva pokazuju da mnoga razgraničenja relevantna u leksikologiji, pa čak i u jednojezičnoj leksikografiji, u sastavljanju dvojezičnih rječnika nisu relevantna.

Dva dovršena projekta (Šipka, 2000 i Šipka, 2002) namijenjena su izvornim govornicima engleskog jezika, kao pomoć u razumijevanju leksemske mase koja tada nije bila zabilježena u standardnim, tada najčešće korištenim srpskohrvatsko-engleskim rječnicima (takvim kao Benson: 1992 i Drvodelić: 1989). Prvi rječnik (Šipka: 2000) uključivao je nestandardnu leksiku (iz slenga, šire rasprostranjene dijalekatske osobine, itd.) i obuhvatio je preko 28 hiljada odrednica, a drugi (Šipka, 2002) leksiku koja je, u talasu nacionalizma i nastajanja novih država, unesena u jezik u devedesetim godinama dvadesetog vijeka (sam rječnik na kraju je sveden na blizu 5 000 odrednica, ali je u bazu podataka na osnovu koje je sastavljan ubačeno blizu 45 000 odrednica – naknadno je izbacivano sve ono što bi potencijalnim korisnicima govornicima engleskog jezika bilo jasno). Projekt koji je u toku (Šipka, 2008) ograničen je na srpsku varijantu, namijenjen i srpskim i engleskim govornicima kao pomoć u razumijevanju i produkciji oba jezika. Radi se dakle o opštem rječniku obima 25 hiljada odrednica.

## 2. Višerječne lekseme kao leksikografski problem

Pitanje razgraničavanja frazema od drugih višerječnih leksema u literaturi je često raspravljano, npr. u Allerton (2004), Burger et al. (1982), a i rješavano u

frazeološkim monografijama (npr. Mršević-Radović, 1987) ili jednojezičkim rječnicima (npr. Matešić, 1982). Mogu se primijeniti različite definicije frazema, već u zavisnosti od pristupa, a ovdje ćemo poći od neutralne opšte definicije, koju daje Crystal (2001: 155):

**idiom** A sequence of words which is semantically and often syntactically restricted, so that it functions as a single unit. The meanings of the individual words cannot be combined to produce the meaning of the idiomatic expression as a whole. For example, the meanings of go, fly, and kite cannot account for the use of the sentence Go fly a kite! in its sense of “Go away” or “Don’t be silly”

Kako se vidi, frazemima se ne smatraju druge višerječne kombinacije, tipa: *divlja svinja*, *pasta za zube*, itd. Pogledaćemo sada kakve su potrebe korisnika dvojezičnih rječnika i time zadaci dvojezičnog leksikografa, što će dati osnovu za to da u trećem dijelu, na materijalu tri pomenuta projekta, pretresemo pitanje koliko je u toj oblasti relevantno razgraničenje.

Korisnik bilo kojeg rječnika ima sljedeće dvije potrebe:

- a. Naći što više informacija
- b. Naći tražene informacije što brže

U dvojezičnom rječniku primarne potrebe jesu nalaženje izvorišnih riječi i njihovih ekvivalenata, odnosno objašnjenja, svih onih koji su korisniku potrebni. U Šipka (2000) i Šipka (2002) primarna potreba korisnika bila je da govornik engleskog provjeri značenje srpskohrvatskih riječi, pomoću engleskih ekvivalenata i/ili objašnjenja. U Šipka (2008), primarne potrebe su kompleksnije i izgledaju ovako:

- a. Govornik srpskog
  - i. Provjeriti značenje engleskih leksema
  - ii. Naći engleski ekvivalent za srpsku leksemu
- b. Govornik engleskog
  - i. Provjeriti značenje srpskih leksema
  - ii. Naći srpski ekvivalent za englesku leksemu.

Ove potrebe iste su za sve lekseme, jednorječne i višerječne. Ono po čemu se višerječne lekseme razlikuju od jednorječnih stvar je leksikografske prakse.

U identifikaciji i selekciji teže je doći do frekvencijskih podataka za višerječne, nego za jednorječne lekseme. U rječničkoj makrostrukturi javlja se pitanje njihovog smještanja u alfabetskom poretku. U rječničkoj mikrostrukturi postavlja se pitanje kako prikazati njihovo variranje, teže je, a često i nemoguće, naći ekvivalente, itd. Kratko govoreći, mada se u smislu potreba korisnika višerječne lekseme ni po čemu ne razlikuju od jednorječnih (njena/njegova očekivanja su ista u oba slučaja), leksikografima ih je znatno teže

obraditi. Stoga je i pitanje koliko ima smisla razgraničavati i posebno obrađivati frazeme u odnosu na druge višerječne lekseme u praktičnom leksikografskom radu izuzetno važno – upravo tom pitanju posvetićemo se u sljedećem dijelu ovog rada.

### 3. Frazemi u odnosu na druge višerječne lekseme u leksikografskom radu

Odnos frazema i drugih višerječnih leksema razmotrićemo iz sljedeća dva ugla:

- a. leksikografskog procesa – upravo, kako pomenuti odnos utiče ili ne utiče na pojedine faze leksikografskog rada,
- b. leksikografskog proizvoda preciznije, da li je pomenuti odnos jednako bitan ili nebitan u svim tipovima dvojezičnih rječnika (standardnih, nestandardnih, za razumjevanje, za jezičku produkciju, itd.)

U razmatranju leksikografskog procesa poći ćemo od pitanja prikupljanja leksema. U sva tri navedena projekta proces je bio eklektičan – korišteni su korpusi, postojeći rječnici, frekvencijske liste i drugi izvori. U prikupljanju višerječnih kombinacija ključnu ulogu odigrale su konkordance i frekvencijske liste za kolokacije. Prvi korak u procesu prikupljanja bilo je odvajanje višerječnih leksema (tipa *popiti batine* ili *kisela voda*) od kolokacija bez tog statusa (tipa *popiti pivo* ili *hladna voda*). U ovoj fazi nema nikakve razlike između različitih vrsta višerječnih kombinacija – bitna je leksikalizacija, to da višerječna kombinacija ima status lekseme ili bar neki status koji bi bio na putu za leksemski.

U sljedećem koraku, selekciji prikupljenih leksema, bitna je čestoća upotrebe višerječne kombinacije i nivo potrebe za upotrebom i ovdje postoje određene razlike između frazema i drugih višerječnih leksema u smislu da je potreba za upotrebom frazema u principu niža nego potreba za upotrebom drugih višerječnih leksema. Po skali vladanja jezikom ILR (više informacija o ovoj i drugim skalama može se naći u Šipka, 2006), upotreba frazema jedna je od karakteristika nivoa 4 i 5, dakle, dobro obrazovanih izvornih govornika (i ispitivači su upravo obučeni da „love” frazeme pri utvrđivanju vladanja jezikom na tim nivoima). Korisnik dvojezičnog rječnika u principu je neko ko je na nivou 3 ili nižem (oni na nivou 4 i 5 koristiće uglavnom jednojezične rječnike, a po dvojezični će sezati samo sporadično, npr. oni koji su profesionalni prevodioci). Mada bi bilo teško utvrditi direktnu korelaciju, ipak se može reći da je potreba uvrštavanja frazema u dvojezični rječnik unekoliko niža nego potreba unošenja odgovarajućih drugih višerječnih leksema. Ovo je posebno

izraženo u određenim tipovima rječnika, o čemu se govori u narednom dijelu rada.

Sljedeća faza leksikografskog rada sa višerječnim leksemama jeste njihova makrostrukturalna obrada – prije svega, utvrđivanje njihovog vezivanja za leksičku jedinicu od jedne riječi. I ovdje je problem u osnovi isti za sve višerječne lekseme, a specifičnost frazema nije to što su frazemi, nego je kompleksnost njihove obrade nešto viša jer se među frazemima javlja viši stepen varijabilnosti nego kod ostalih višerječnih leksema. Mada i kod drugih višerječnih leksema mogu da se jave alternativni oblici, npr. *blorovodonična kiselina* ili *sona kiselina*, oni su, u poređenju sa frazemima, dosta rijetki. Kod frazema se slučajevi kakvi su, npr. *lavli/Lavovski deo*, *ispravljati/ispraviti krivu Drinu* (gdje imamo formalno variranje), *u dobru i (u) zlu* (gdje se javlja opcioni element), te *u dubini duše/srca* (gdje se javlja alternativna metafora) izuzetno česti. U sva tri pomenuta leksikografska projekta, a na osnovu principa navedenih u drugom dijelu ovog rada, višerječne lekseme smještane su u alfabetskom poredku pod svim svojim jednorječnim djelovima gdje bi ih korisnici mogli tražiti, npr. *u dobru i u zlu* navođeno je i pod *dobro* i pod *zlo* i pod *u*. U praktičnom radu u bazi podataka višerječna leksema navodila se samo na jednom mjestu (drugi elemenat, ukoliko se radi o punoznačnoj leksemi), a onda su višerječne lekseme skriptovima kopirane i na druga mjesta. To što se nešto skriptovima kopira na više mjesta trivijalan je problem. Međutim, varijabilnost frazema stvara jedan drugi, ključni, metaleksikografski problem. Kako je jedna od dvije osnovne potrebe korisnika rječnika naći tražene informacije to prije, rječnik mora imati ograničen obim (problem je, razumije se, izraženiji u rječnicima manjeg obima, o čemu će biti riječi u sljedećem dijelu teksta). Tako onda frazemi, znatno češće nego druge višerječne lekseme, ostavljaju leksikografa pred dilemom da li uvrstiti sve varijetete ili neke izostaviti da bi se smanjio obim rječnika.

Konačno, u fazi mikrostrukturne obrade osnovni je zadatak pronalaženje ekvivalenata i ovdje se zapravo javljaju najveće razlike između frazema i drugih višerječnih leksema. Osnovna razlika je u tome da su frazemi znatno više vezani za kulturu, pa znatno češće nemaju direktne ekvivalente od drugih višerječnih leksema. Tu je onda leksikograf znatno češće primoran da pribjegne objašnjavanju nego kod nekih drugih višerječnih leksema, kakvi su na primjer, složeni termini, gdje su pojmovi univerzalni, pa, po pravilu, postoji direktni ekvivalent. Često se dešava da izostaje cio referencijalni okvir frazema. Ilustrativan je tu primjer frazema *ko kamen za kupus* ‘veliki, obično o glavi’, obrađenog u Šipka (2000). Američkim korisnicima (kojima je rječ-

nik i namijenjen) izvorišni domen metafore potpuno je nepoznat jer se kupus ne kiseli u glavicama nego se isključivo koristi ribanac a i on se kupuje u kesama u samoposluzi. Tako je onda uz objašnjenje 'big in size' odrednica morala i da objasni činjenicu da se kupus kiseli u glavicama na koje se stavljaju daske pritisnute ogromnim kamenom, da glavice ne bi isplivale iz vode. Nešto su jednostavniji za objašnjavanje bili frazemi tipa *Radojka i Tine* 'zglobni autobus', gdje je dovoljno pomenuti da se radi o harmonikaškom duetu da bi se razumio izvorišni domen metafore. Ni ova razlika između frazema i drugih višerječnih jedinica ne proizlazi iz njihove leksikološke odjelitosti od drugih višerječnih leksema nego je dio šire pojave metafora vezanih za kulturu, koje se javljaju i kod jednorječnih leksema tipa *vartburđini* 'automobil marke Wartburg' *škodilak* 'automobil marke škoda', gdje američkim korisnicima nisu poznate dane marke automobila ni njihov društveni status.

Kad je u pitanju leksikografski proizvod, upravo tip rječnika, status frazema varira u po sljedeća tri parametra:

- a. rječnik standardne prema rječniku nestandardne leksike,
- b. recepcijski u odnosu na produkcijski (odnosno produkcijsko-recepcijski) rječnik,
- c. rječnik užeg obima prema rječniku šireg obima.

Kako pokazuju podaci iz Šipka (2000), preko 42 % nestandardne leksičke mase rezultat je semantičkih proširenja (najčešće metaforskih). Veliki broj semantičkih proširenja uz činjenicu da su nestandardni jezički oblici ukotvljeni u uzak socijalni i geografski prostor, uslovljavaju pojavljivanje znatno većeg broja metafora vezanih za kulturu nego što je slučaj u rječnicima standardne leksike.

Kao ilustraciju ove razlike dajemo višerječne lekseme sa osnovnim elementom baciti/bacati iz sirovih baza podataka (prije selekcije) u rječniku nestandardne leksike, Šipka (2000) i rječniku standardne leksike, Šipka (2002). Kako se vidi, u prvom slučaju, većina tih leksema su frazemi, u drugom, u većini slučajeva imamo nefrazemske višerječne lekseme

bacanje plamena, bacanje u brazdu, bacati divljake, bacati drvlje i kamenje, bacati farove, bacati folove, bacati govna, bacati kipce, bacati klamfe, bacati lire, bacati litre, bacati loptice, bacati malog u nesvest, bacati malog u nesvjest, bacati u kljun, bacati okoslave, bacati pijesak u oči, bacati pikule u oči, bacati prašinu u oči, bacati se pod iglu, bacati u zahod, bacati uši, baca plamena, baciti ga u nesvest, baciti kadar, baciti kašiku u govna, baciti kašiku u pekmez, baciti koplje u trnje, baciti liru, baciti luk, baciti oko, baciti peglu, baciti pod nameštaj, baciti pušlek, baciti se na kop, baciti

sidro, baciti u neznanje, baciti u pjenu, baciti u sevdah, baciti u trošak, baciti udicu, baciti uši, baciti valter, baciti vunu, baciti šminku, baciti štanglu u usta (Šipka, 2000)

bacač granata, bacanje blata, bacanje diska, bacanje kladiva, bacanje koplja, bacanje kugle, bacanje letaka, bacanje novca u vjetar, bacanje rukavice, bacati novčić, bacati sjenu, baciti kocku, baciti oko, baciti pod noge, baciti rukavicu, baciti rukavicu u lice, baciti se na, baciti u zasjenak, baciti u zatvor (Šipka, 2002)

Kad je u pitanju parametar recepcije i produkcije, osnovna razlika je u tome da receptivni rječnik ne mora da predstavi savršene funkcionalne ekvivalente – tamo gdje ih nema, dovoljna su objašnjenja. Dobar primer za ovo su kombinacije sa glagolom *jebati* u Šipka (2000). Nastranu to što se može diskutovati o tome koliko je koja kombinacija leksikalizovana i treba li ih uopšte uvrštavati (veoma široko uvrštavanje ovdje je bilo direktan proizvod jedne specifične leksikografske politike uslovljene potrebama primarnih korisnika rječnika), sama činjenica da njihov niz obuhvata strane 184–200 ovog rječnika, dok će se u recepcijsko-produkcijskom rječniku (Šipka, 2008) to ograničiti na jedan kratak rječnički članak, jasno govori o različitim potrebama recepcije i produkcije. Razlike su dalje i u mikrostrukturnoj obradi. Pododrednica *jebem mu mater* pokazuje specifičnosti obrade za recepcijski rječnik, koja sadrži objašnjenja i informacije potrebne da se shvati „težina“ višerječne kombinacije u kontekstu.

**Jēbēm mu mātēr!** {S} (lit. I fuck his mother!) **a. col** *Off* used as a strong curse, Fuck him! **b. Obs** used to refer to the futility of all efforts, it's no use: *Ne znam šta da radim. Jebem mu mater! I do not know what to do. It's fucking hopeless!* (Šipka, 2000:187)

Kako se vidi, obje upotrebe, od kojih je druga zasigurno leksikalizovana, popraćene su objašnjenjima i etiketama (*col* – kolokvijalno, *Off* – uvredljivo, *Obs* – opsceno), čija je osnovna namjena da korisnik pri recepciji govorenog ili napisanog teksta shvati implikacije upotrebe ovih elemenata. U produkcijski rječnik ne bi imalo smisla ni uvrštavati ovu odrednicu jer se ista ideja može prenijeti i drugim jezičkim sredstvima koje bi imale direktan i potpun funkcionalni prevodni evivalent, da se ne govori o tome da, čak i na nivou ILR 5 ne postoji uslov korištenja opscenih leksema, dovoljno je njihovo poznavanje, dakle, recepcija, ne produkcija, pa i najobimniji produkcijski rječnik ne mora da sadrži takve riječi. Recepcija je, opet, neophodna, uzevši u obzir činjenicu da je upotreba ovih elemenata, pogotovo u kolokvijalnom izrazu, izuzetno česta. I ovdje se ne radi o nekoj specifičnosti frazema u odnosu na druge višerječne kombinacije – radi se o osobini određenog leksemskog sloja, kako jednorječnog tako i višerječnog, a taj leksemski sloj javlja se češće kod frazema nego kod drugih višerječnih kombinacija.

Konačno, obim rječnika u velikoj mjeri utiče na disproporciju frazema u odnosu na druge višerječne kombinacije u konačnom leksikografskom proizvodu. Ovdje se može reći da manji broj odrednica u rječniku automatski znači i proporcionalno manji broj frazema u odnosu na druge višerječne kombinacije. Ovo je posljedica činjenice da je dobar dio frazema zapravo obilježavač vladanja jezikom na nivou ILR 4 i 5, dok se druge višerječne kombinacije u velikoj mjeri smještaju na nivoe ILR 1–3. Tako onda Šipka (2008), dakle rječnik obima 25 hiljada odrednica, ne uključuje niz frazema koji se mogu naći u rječnicima dvostruko većeg obima, tipa Benson (1992) ili Dvodelić (1989). Kako Benson (1992) obilježava frazeme zvjezdicom, može se odrediti njihov broj i on je 3 705 u rječniku od 36 964 glavne odrednice (dakle, one koje su obrađene, zajedno sa onima kod kojih se samo unakrsno upužuje, ima ih 49 347). Dakle, broj frazema je nešto preko 1 % od ukupnog broja glavnih odrednica. Preliminarna ispitivanja na osnovu građe Šipka (2008) pokazuju da će ova razmjera biti znatno manja i da neće premašiti 0.2 %. I ovdje se ne radi o činjenici da je nešto frazem, nego o frekvencijskim činjenicama, upravo distribuciji leksemskih elemenata (u ovom slučaju višerječnih) po nivoima vladanja jezikom.

#### 4. Zaključci

Iskustva iz tri pomenuta leksikografska projekta upućuju na sljedeće zaključke o statusu frazema u odnosu na druge višerječne lekseme u leksikografiji.

- a. Relevantna leksikološka razgraničenja ne preslikavaju se direktno ni u leksikografsku praksu ni u leksikografske proizvode.
- b. Za korisnika rječnika nema neke veće razlike između frazema i drugih višerječnih leksema u smislu da se korisničke potrebe što potpunijeg i što bržeg pronalaženja ne mijenjaju u zavisnosti od tipa višerječne lekseme. Za sve višerječne lekseme važe oni parametri koje je obradio Zgusta (1971), koji već preko trideset godina ostaje neprikosnoven u ovom i drugim metaleksikografskim promišljanjima.
- c. Frazemi u leksikografskom radu predstavljaju dodatni izazov (u odnosu na druge višerječne lekseme), ne zbog leksikoloških razgraničenja, nego zato što se za njih češće vezuju metaforska proširenja i zato što je njihov veći broj vezan za više nivoe vladanja jezikom.

Iz ovoga proizlazi da bi frazemima trebalo posvetiti pažnju u obuci leksikografa, prije svega tako što bi se ukazivalo na pojave tipa metafora vezanih za kulturu, vezanost frazema za više nivoe vladanja jezikom i nudila konkretna

rješenja obrade koja bi na optimalan način odgovorila na ove pojave već u zavisnosti od tipa leksikografskog proizvoda. Ovaj segment leksikografske obuke ne bi trebalo da bude vezan za frazeme nego bi trebao da bude dio šireg sagledavanja ovih pojava kako kod višerječnih tako i kod jednorječnih leksema. Posebno obilježavanje frazema u rječniku, kako to rade neki rječnici, nije produktivno jer korisnika opterećuje informacijom koja iz nje-gove/njene perspektive nema nekog posebnog značaja.

## Literatura

- Allerton, D.J. et al. (2004): *Phraseological Units: Basic concepts and their application*. Bern.
- Benson, M. (1992): *Srpskohrvatsko-engleski rečnik*. Beograd.
- Burger, H. et al. (1982): *Handbuch der Phraseologie*. Berlin, New York.
- Crystal, D. (2001): *A Dictionary of Language*. Chicago.
- Drvodelić, M. (1989): *Hrvatsko ili srpsko engleski rječnik*. Zagreb.
- Matešić, J. et al. (1982): *Frazeološki rječnik hrvatskog ili srpskog jezika*. Zagreb.
- Mršević-Radović, D. (1987): *Glagolsko-imeničke frazeološke sintagme*. Beograd.
- Šipka, D (2000): *SerboCroatian-English Colloquial Dictionary*. Springfield.
- Šipka, D (2002): *Dictionary of Serbian, Croatian, and Bosnian New Words*. Springfield.
- Šipka, D (2005): *Uvod u leksikologiju i srodne discipline*. Novi Sad.
- Šipka, D (2006): Ocenjivanje vladanja jezikom skale i procedure. In: *Jezik danas* 21–22. 12–15.
- Šipka, D (2008): *Srpsko-engleski i englesko-srpski rečnik*. (25 000 reči, u izradi).
- Zgusta, L (1971): *Manual of Lexicography*. The Hague