

ŽELJKO BUJAS (Filozofski fakultet, Zagreb)

STVARANJE DVOJEZIČNE LEKSIKOGRAFSKE DATOTEKE RUČNIM PREOKRETANJEM I NJEZINO VREDNOVANJE

Utvrdivši postojanje nepoklapanja između građe protusmjernih dvojezičnih rječnika (englesko-hrvatskih i hrvatsko-engleskih), autor opisuje leksikografski potencijal tako skrivene građe i ističe potrebu da se ta korisna građa učini dostupnom preokretanjem spomenutih rječnika. Slijedi opis vlastite metode i iskustava ručnog preokretanja DEH 70 (Englesko-hrvatskog rječnika M. Drvodelića iz 1970 /rev. Z. Bujas/). Dobivena građa — konverzijska datoteka od oko 250.000 kartica hrvatsko-engleskih prijevodnih parova — dalje se formalno analizira i razbija na 6 kategorija nepoklapanja. Tako se određuje visok udio leksikografski korisne građe (74.3%). Prikupljena korisna građa ilustrira se na primjeru dviju natuknica iz DEH 70. Na kraju se prikazuje upotrebljivost konverzijske datoteke za leksikološka istraživanja kontrastivnom analizom hrvatskog glagolskog tipa *na-* + Vb osnova + *se* i njegovih engleskih ekvivalenata.

1.

Po svojoj prirodi i funkciji, takozvana lijeva strana dvojezičnog rječnika ima dublje korijene u civilizacijskim aspektima leksika od njegove desne strane. Mi tako očekujemo da natuknice tipične za anglosasku civilizaciju — kao na primjer *sporrán*, *coleslaw*, *Christmas cracker*, *gerrymander*, *dude ranch* — budu uključene već u englesko-hrvatskom rječniku srednjeg opsega. Mi međutim ne očekujemo, niti ćemo normalno naići na te engleske riječi kao glose nekih hrvatskih natuknica ili složenih izraza u hrvatsko-engleskom rječniku slične veličine. Isto, naravno, vrijedi i pri obratnom smjeru. Hrvatske riječi specifične za domaću civilizaciju — *klijet*, *zuba(ta)c*, *prvoborac*, *dohodovni*, *zurovski* — neće se tako pojaviti kao glose ni uz jednu englesku natuknicu i u najopsežnijem englesko-hrvatskom rječniku, jer se jednostavno ne uklapaju u pojmovni inventar anglosaksonske civilizacije.

Kad se međutim ne radi o pojedinim riječima ili izrazima specifičnim za neku civilizaciju, nego o masi redovite građe dvojezičnih rječnika, teže je — a i nepotrebno — prihvaćati disparatnosti sadržaja uočive usporedbom »protusmjernih« rječnika. Pogotovu kad je riječ o paru dvojezičnih rječnika od istog autora, i o zaista uspjelim leksikografskim rješenjima, prisutnim u dvojezičnom rječniku jednog smjera, a odsutnim u protivnom smjeru. Većina pisaca dvojezičnih rječnika neizbježno je svjesna tog potencijala građe koji se oslobađa promjenom smjera, to jest preokretanjem (konvertiranjem) takvih rječnika. Unatoč tome leksikografi se često ne upuštaju u preokretanje dvojezičnih rječnika; vjerojatno zbog golema posla što ga takvo konvertiranje iziskuje, ali možda i nedovoljno uvjereni u korist od takva rada.

Koliki je, dakle, zapravo taj potencijal — a koliki (i kakav) posao? Kako je i laiku jasno da se tu uvijek radi o velikom naporu, pouzdan odgovor na oba pitanja bitan je svakome tko planira konverziju dvojezičnog rječnika bilo sam ili kao timski projekt.

2.

Pokušat ću odgovoriti na oba pitanja opisom postupaka i ocjenom dosadašnjih rezultata vlastitog pothvata preokretanja opsežne dvojezične rječničke građe (započetog 1970. i još nedovršenog).

Odlučio sam da, kao prvo, na malom oglednom uzorku dvojezičnog rječničkog teksta ocijenim i konverzijski potencijal građe i kategorije jedinica kojima će trebati baratati u toku postupka. Za takvu je procjenu najprikladniji par protusmjernih dvojezičnih rječnika srednjeg opsega, od istog autora i istovremeno objavljenih. Izabrao sam stoga englesko-hrvatski i hrvatsko-engleski rječnik Milana Drvodelića,¹ oba iz 1970 (u daljem tekstu DEH 70 i DHE 70).

Metodološki bi postupak bio doduše nešto dosljedniji da sam uzeo isti rječnički par iz 1953, odnosno 1954, kad njegov tekst još nije bio podvrgnut tuđoj reviziji (osobito opsežnoj /mojoj/ u DEH 70). Tako bi se naime dobila vjernija slika o konverzijskom potencijalu u dvojezičnim rječnicima od istog autora. Nad tim teoretskim obzirima prevladali

¹ Pregled dosadašnjih izdanja ovih rječnika (izdavač uvijek Školska knjiga, mjesto izdanja uvijek Zagreb):

Englesko-hrvatski rječnik, 1954, 976 str.

Englesko-hrvatskosrpski rječnik, 1962 (1104 str.) i 1970 (1198 str.), oba preradio Ž. Bujas

Hrvatskosrpsko-engleski rječnik, 1961, (913 str.) i 1970 (913 str.), oba preradio za tisak L. Spalatin

U obratnom smjeru:

Hrvatsko-engleski rječnik, 1953, 940 str.

Hrvatskosrpsko-engleski rječnik, 1961, (913 str.) i 1970 (913 str.), oba preradio za tisak L. Spalatin.

Hrvatsko ili srpsko-engleski rječnik, 1978, 847 str., preradio Ž. Bujas

su međutim praktični leksikografski razlozi, to jest mogućnost da se dobije pouzdanija, suvremenija i opsežnija rječnička građa. Konačno, zahvati revidiranja dvojezičnih rječnika — podložni i sami općim metodološkim načelima dvojezične leksikografije — ne mogu bitnije izmijeniti prirodu konverzijskog potencijala.

Birajući, kao najmanji mogući ogledni uzorak, »tipičnu« natuknicu u DEH 70, zaustavio sam se na *lot* zbog njezine »srednje« dužine i prosječne (da se tako izrazim) strukture:

lot [lot] *n* ždrijeb, ždrebanje; udio, sudjelovanje; sudbina; parcela, gradilište; sila, mnoštvo; količina istovrsne robe, partija; *sl* filmski grad, studio

Ako ovu natuknicu zamislimo kao eksplicitni niz englesko-hrvatskih prijevodnih parova (*lot* : *ždrijeb*, *lot* : *ždrebanje*, *lot* : *udio* itd.), lako ćemo te parove pretvoriti u ovakvu hrvatsko-englesku listu parova:

ždrijeb : *lot*
ždrebanje : *lot*
udio : *lot*
sudjelovanje : *lot*
sudbina : *lot*
parcela : *lot*
gradilište : *lot*
sila : *lot*
mnoštvo : *lot*
količina (istovrsne robe) : *lot*
(količina) *istovrsne* (robe) : *lot*
(količina istovrsne) *robe* : *lot*
partija : *lot*
filmski (grad) : *lot*
(filmski) *grad* : *lot*
studio : *lot*

Ako sada ovih 16 parova usporedimo s građom u paralelnom hrvatsko-engleskom rječniku (DHE 70), vidjet ćemo da već ova jedna natuknica — i to bez svog frazeološkog dijela — dovoljno ilustrira i opseg i raznovrsnost konverzijskog potencijala u protusmjernim parovima dvojezičnih rječnika. Ona naime sadržava primjere za sva tri osnovna tipa nepoklapanja između građe u takvim rječnicima: (a) odsutnost natuknice, (b) odsutnost sklopa i (c) odsutnost glose.

a) *Potpuna odsutnost hrvatske natuknice u DHE 70*

parcela : *lot*

b) Odsutnost istovjetnog tekstovnog sklopa u DHE 70

količina (∼ istovrsne robe) : *lot*
istovrstan (količina ∼ne robe) : *lot*
roba (količina istovrsne ∼e) : *lot*
filmski (∼ grad) : *lot*
grad (filmski ∼) : *lot*

c) Odsutnost engleske glase (*lot*) u DHE 70

ždrebanje : *lot*
udio : *lot*
gradilište : *lot*
partija : *lot*
studio : *lot*

(Iz trećeg tipa ispuštena su tri para */sila : lot, mnoštvo : lot* i *sudjelovanje : lot/* koji su nejasni bez konteksta */sudjelovanje u borbi : throwing in one's lot with/* ili nepotpuni */sva sila : lot* i *mnoštvo : a lot of people/*).

Frazeološki dio natuknice *lot* prestrukturiran u hrvatsko-engleske prijevodne parove izgledao bi ovako:

mnogo je : *it's a lot*
mnogo ljudi : *a lot of people, lots of people*
vrlo mnogo : *lots and lots*
vući ždrijeb, ždrebati : *to cast (draw) lots*
ždrijeb je pao na mene : *the lot fell upon me*
dijeliti sudbinu s nekim : *to cast (throw) in one's lot with*
zapalo me : *it fell to my lot*
na meni je da : *it is my lot to*
moja životna sudbina : *my lot in life*
težak život (sudbina) : *hard lot*
on je ništarija : *he is a bad lot*
građevna parcela : *building lot*
ove se slike prodaju zajedno : *these pictures form a single lot*
uzmi sve : *take the whole lot*
briga tebe! : *a lot you care*

Kako vidimo, ovdje nije moguće izdvojiti hrvatsko-engleske prijevodne parove tipa *ždrijeb : lot, ždrebanje : lot* itd. kao dosad (osim u slučaju *ždrebati : cast /draw/ lots*). Naravno, može se tvrditi da je u paru *vući ždrijeb : cast (draw) lots* nosilac smisla riječ *ždrijeb*, pa se tu zapravo radi o leksikografskom paru *ždrijeb (vući ∼) : cast (draw) lots*, pri čemu je *vući* smislovno podređen. Intuitivno ipak znamo da *vući* može i samo za sebe imati puno značenje *ždrebanja* (*»da vučemo!«, »šta*

si izvukao?» i sl.). Još je teže s *na meni je da : it is my lot to*. Koja je tu »glavna« riječ — *meni* ili *na*? Ovdje se radi o sklonjivoj frazi, jer je moguće reći *na meni/tebi/njemu/njoj/nama/vama/njima je*, pa bi možda čvršće — i prostorno ekonomičnije — uporište ovoj frazi u hrvatsko-engleskom rječniku bila natuknica *na*. Konačno, je li *je iz na meni je da* sasvim prazno? Da li su baš svi slučajevi *je* koje bismo dobili preokretanjem — a bilo bi ih vjerojatno na stotine — sasvim beskorisni? Slučajevi *je = was* (npr. *rekao je da je siguran : he said he was sure*), *je = am* (*vruće mi je : I am hot*) ili *je = Ø* (*odjeća što je nose pokajnici*) nesumnjivo zavređuju da se leksikografski zabilježe.

Načelno je dakle bitno da pri preokretanju dvojezičnog rječnika ne promakne nijedna riječ jezika na koji se konvertira (jezika cilja), s time da se o njezinom konačnom korištenju — uvrštavanju ili neuvrštavanju u novi (ili revidirani postojeći) dvojezični rječnik — donese odluka kasnije, nakon što je pouzdano, bez propusta prikupljena totalna građa. Prikupljanje građe dakako znači da svaku riječ teksta jezika cilja prati i prijevod (prijašnji izvornik, prijašnja lijeva strana). Taj prijevod dosta često ne pruža jasne prijevodne parove, to jest ne omogućuje da se nedvojbeno odredi neposredni prijevodni ekvivalent svake riječi jezika cilja. Koja je engleska riječ, na primjer, neposredan prijevod hrvatskom *vrlo* u frazeološkom paru *vrlo mnogo : lots and lots* iz analizirane natuknice? S druge strane, i tamo gdje je moguće točno odrediti prijevodni par — na primjer (iz iste natuknice) *sudbinu : one's lot* iz para *dijeliti sudbinu s nekim : to cast (throw) in one's lot with* — važno je imati uvida u kontekst obaju članova prijevodnog para, jer bez konteksta *vrlo* često ne možemo suditi o prihvatljivosti, primjenljivosti, kolokabilnosti, strukturnom potencijalu prijevoda i sličnom.

Preokretanje dvojezičnog rječnika ukratko podrazumijeva takav postupak koji osigurava:

- a) da svaka riječ jezika cilja bude obuhvaćena,
- b) da svaku tu riječ prati što opsežniji i neizmijenjen dvojezični kontekst izvorne rječničke građe.

Kako to postići? Neka na to odgovore moja iskustva stečena ručnim preokretanjem spomenutog rječničkog para.

3.

Razmišljajući o najdjelotvornijem načinu preokretanja DEH 70, a svjestan mehaničke i repetitivne prirode postupka, odmah mi se nametnula pomisao na kompjutor kao idealno pomagalo. I metodološki bilo je očito da je najbolji put izrada specifične kompjutorske konkordancije cijelog teksta DEH 70 (pogotovu pošto sam već imao vlastitog iskustva s izradom kompjutorskih konkordancija hrvatskih, doduše književnih, tekstova). U to vrijeme, međutim, kompjutori su u nas još

uvijek bili rijetki i (kao i danas) financijski nedostupni istraživačima pojedincima. To, kao i želja da leksikografski postupak preokretanja dvojezične građe posluži i pedagoški — u radu mojih jezičnih seminara 3. godine studija anglistike na Filozofskom fakultetu u Zagrebu — doveli su do odluke da se upustim u ručno preokretanje DEH 70.

Razmatrajući kako da se to ostvari sa što manje napora i što ekonomičnije — ne žrtvujući spomenuta dva osnovna zahtjeva (obuhvaćanje svake hrvatske riječi i čuvanje punog obojezičnog konteksta) — počeo sam tako da sam odbacio kao nepraktično i neekonomično svako prepisivanje, ekscerpiranje na kartice ili fotokopiranje rječničkog teksta. Odlučio sam da se služim samom tiskanom stranicom i pretvaram je neposredno u kartice. Jednostavnije rečeno, DEH 70 trebalo je razrezati u kartice — i to tako da svaka riječ hrvatskog teksta ima svoju karticu, s time da se na istoj kartici nađe i engleski prijevod. Abecediranjem tih kartica po hrvatskoj riječi, dobit će se konačna hrvatsko-engleska leksikografska datoteka višestruke namjene: za revidiranje DHE 70, za rad na zamišljenom velikom hrvatsko-engleskom rječniku, ali i za druga leksička istraživanja (o čemu više kasnije). Sve je to međutim bila tek okvirna zamisao i nije rješavala niz praktičnih i tehničkih problema koji su se javljali.

Kako, na primjer, postići da na svakoj kartici bude uvijek označena samo jedna, i to uvijek druga, hrvatska riječ teksta kao nosilac kartice — tako da nijedna hrvatska riječ ne ostane bez svoje kartice?

Kako, s druge strane, razrezati rječnik u kartice? Najjednostavnije bi to bilo obaviti strojnim nožem za papir, koji bi za jednu minutu pretvorio rječnik u 11.630 kartica formata 4 x 7 cm (po okomitoj razdjelnici dvostupačnog teksta, a zatim vodoravno svaki stupac na pet kartica jednake visine). Unatoč svojoj privlačnoj jednostavnosti, nedostaci tog postupka su u tome što blok tiskanog teksta nije idealno centriran na svakoj stranici, pa bi se okomitim rezanjem (po razdjelnici) mogla odrezati sva početna slova desnog, odnosno završna slova lijevog stupca na stanovitom broju stranica. Dalje, kod dužih rječničkih članaka vodoravno rezanje odvojilo bi neke engleske natuknice od hrvatske glose (sada natuknice).

Zbog svega toga, kao i zbog spomenute odluke da se u ovaj proces uključuje studenti, konačni postupak bio je kako slijedi:

(1) Nabavljeno je (novčanom pomoću tadašnje Komisije za znanstveno-istraživački rad Filozofskog fakulteta) 30 primjeraka DEH 70, tako da je svaka stranica teksta postojala u 15 kopija. (Drugih 15 kopija padalo je pri razrezivanju kao poledina stranice.)

(2) Sada je kosom crtom, uočljivo, križana prva riječ hrvatskog teksta na prvoj od 15 kopija svake stranice. Iduća hrvatska riječ jednako je obilježena na drugoj stranici, ona za njom na trećoj, riječ za ovom na četvrtoj stranici i tako dalje, sve dok na petnaestoj stranici nije bila označena 15. riječ hrvatskog teksta od početka stranice. Sada se vra-

ćamo na prvu stranicu i tamo obilježavamo 16. riječ hrvatskog teksta, na drugoj stranici 17. riječ, na trećoj 18. i tako dalje.

(3) Takvim obilježavanjem postizemo dvoje: (a) ne preskače se nijedna riječ hrvatskog teksta (osim kratica tipa *itd.*, *i sl.*, *i dr.*, *npr.* koje namjerno izostavljamo); (b) broj obilježenih riječi po stupcu iznosi prosječno 7 (dakle 14 po stranici), a razmak među njima — premda promjenljiv ovisno o gustoći hrvatskog teksta — relativno je ujednačen i iznosi 3–4 cm. Razrezavši svaku stranicu najprije po okomici u dva stupca, sada režemo između dvije markirane hrvatske riječi, pazeći samo da se ne prekine potreban kontekst. Isto tako nastojimo da gdje god je to moguće na kartici bude i izvorna engleska natuknica. Iz iskustva sada znamo da se u prosjeku događa da je ta izvorna natuknica odsutna s najmanje dvije kartice po stranici, pa se mora upisati rukom u marginu (najbolje prije razrezivanja).

(4) Sada se dobivene kartice abecediraju po hrvatskoj riječi (u kanonskom obliku).

4.

Time smo dobili leksikografsku građu koja nam je bila svrhom i prikaz preokretanja DEH 70 mogao bi time završiti, da nismo sebi postavili za cilj i procjenu potencijala tipične građe dvojezičnog rječnika. Studenti analizatori imali su stoga kao dalji zadatak da usporede dobivenu građu u obliku abecediranih hrvatsko-engleskih kartica sa sadržajem DHE 70. Svaki student je, preokrenuvši tri izvorne stranice DEH 70, stvorio vlastitu leksikografsku datoteku od prosječno 600 hrvatsko-engleskih kartica (3 stranice x 15 kopija = 45 stranica x 14 kartica po razrezanoj stranici). Uspoređivanje tih kartica s građom DHE 70 moralo se obaviti metodologijom primjerenom sposobnostima analizatora, dakle postupnom, eksplicitnom i jednoznačnom.

Studenti analizatori trebali su prema tome odgovoriti na tri pitanja, i to ovim redom:

(1) Je li obilježena hrvatska riječ na kartici sama ili dio tekstovnog sklopa (skupine, rečenice)?

(2) Postoji li ta riječ (sama ili sa sklopom) u DHE 70?

(3) Ako postoji, ima li isti ili različit engleski prijevod?

Udruženi, ti parametri daju šest mogućih kategorija:

Obilježena hrvatska riječ	U DHE 70		
	ne postoji	postoji	
		engl. isti	engl. različit
sama	1	2	3
sa svojim sklopom	4	5	6

Ilustrirat ću svaku kategoriju jednim primjerom:

Ktg. 1 *argumentiran* : *reasoned*

Ktg. 2. *mlinar* : *miller*

Ktg. 3 *kalan* : *oozy* (u DHE 70 samo *kalan* : *muddy, dirty, thick, turbid, troubled, dreggy*)

Ktg. 4 *glavna prometna arterija* : *trunk-road* (u DHE 70 samo *arterija* : *artery*)

Ktg. 5 *koliko je god* moguće : *as much as possible*

Ktg. 6 *kad god* : *as often as* (u DHE 70 samo *kad god* : *whenever, at any time whatsoever*)

Počevši u šk. g. 1970/71. pa do kraja šk. g. 1979/80 (uz jednogodišnji prekid), studenti analizatori su u 265 referata konvertirali 795 stranica od ukupno 1162 u DEH 70. Tako je preokrenuto 68.4% ukupnog opsega rječnika (preko dvije trećine), što je više nego dovoljno da se na osnovi tako dobivene datoteke (od oko 170.000 kartica) ukaže na zakonitosti u ustroju te građe. To sam i učinio na uzorku od 144 stranice, s dovoljnom disperzijom da bi uzorak bio reprezentativan. Pokazala se ovakva distribucija u 6 zadanih kategorija:

Ktg.	1	2	3	4	5	6	Ukupno
hrv.-engl. prijevodnih parova	2.180	3.345	4.702	19.743	622	722	31.364

Izraženo u proporcijama:

Ktg.	1	2	3	4	5	6	Ukupno
udio kategorije	7%	10.7%	15%	63%	2%	2.5%	100.0%

Ako dobivene proporcije projiciramo na ukupnu masu DEH 70 od 1162 stranice, dobit ćemo ovu distribuciju očekivanog konačnog broja prijevodnih parova:

Ktg.	1	2	3	4	5	6	Ukupno
hrv.-engl. prijevodnih parova	12.700	27.100	38.000	159.200	5.000	6.300	235.500

Isključimo li sada kategorije 2 i 5, koje obuhvaćaju slučajeve potpunog poklapanja DEH i DHE 70, preostaje čak 87.3% građe iskoristive za dopunu nekog postojećeg hrvatsko-engleskog rječnika (konkretno DHE 70), ili za pisanje novog takvog rječnika. Već to više nego dovoljno opravdava svaki napor uložen u preokretanje dvojezičnih rječnika. U našem slučaju, već dosad — nakon konverzije 68.4% ukupne rječničke građe — dobivena je datoteka od preko 170.000 kartica, od čega oko 150.000 leksikografski korisnih kartica! Po završetku konverzije (predvidivo koncem 1982), može se očekivati konačan opseg korisne datoteke od 220.000 kartica hrvatsko-engleskih prijevodnih parova.

5.

Posebnu pozornost privlači kategorija 4 kao kudikamo najopsežnija od šest kategorija moje analize. Njezin predviđeni konačni udio iznosi 63% čitave datoteke (a čak 72.3% od korisne datoteke), odnosno 159.000 kartica od čega je dosad prikupljeno blizu 110.000. Ta je kategorija dakle sama gotovo trostruko veća od ukupne mase svih ostalih korisnih kategorija ($1 + 3 + 6 = 24.5\%$ ukupne datoteke, odnosno 57.000 kartica). Razlog nesrazmjerno većem opsegu kategorije 4 treba tražiti u definiciji sklopa kojom smo se služili na seminarima (»svaki segment teksta od dvije ili više riječi između konvencionalnih grafičkih ili interpunkcijskih granica«) i u definiranju odsutnosti sklopa (»odsutnim iz DHE 70 smatra se svaki sklop koji nije tamo doslovno zastupljen, od riječi do riječi i od slova do slova«). Stanovita mehanička eksplicitnost ovih definicija pokazala se vrlo korisnom i djelotvornom za studente analizatore, dakle metodološki opravdanom. Ona je međutim dovela do jedne hipertrofirane kategorije kojom su obuhvaćene vrlo različite sintaktičke i druge potkategorije segmenata hrvatskog teksta nejednake leksikografske iskoristljivosti — od doista neupotrebljivih, preko posredno iskoristljivih, do nedvojbeno i neposredno korisnih. Određivanje međusobnih omjera tih kategorija zahtijeva opsežne analize, što prelazi okvire ovog rada. Već sada može se ipak reći na osnovi pažljive procjene dosad prikupljene građe da potpuno neupotrebljivih parova u materijalu kategorije 4 ima najviše jedna petina, a sve drugo je posredno i neposredno korisna leksikografska građa. Neupotrebljivi parovi čine veliku većinu slučajeva u slijedeće četiri osnovne skupine:

- (a) Obilježena riječ izrazito je »gramatička« (funkcionalna, neleksička).

npr. *staviti u rasol* : *pickle*
dio koji zahvaća preko drugoga : *lap*
usoliti, sušiti ili dimiti ribu : *kipper*
sudariti se : *knock*

- (b) Obilježena riječ dio je dužeg opisa bez značenjski središnje leksičke jedinice.
npr. *slikovit planinski predio s brojnim jezerima u sjeverozap. Engleskoj : the Lake District*
- (c) Obilježena riječ dio je dodatnog objašnjenja (obično u zagradi).
npr. *iskvarena engleština (prije u ustima Kineza, sada urođenika) : pidgin English*
- (d) Obilježena riječ (premda leksička) potrebna je samo da se popuni sintaktička struktura, a semantičko težište je na drugoj riječi.
npr. *jedini mu je spas (izlaz) bio u bijegu : flight was his only resource*

Smanjimo li sada za jednu petinu ukupni očekivani udio kategorije 4 od 63%, dobit ćemo 50.4%, što možemo (budući da se i onako radi o procjeni) zaokružiti na 50%. Nakon toga ispravka, konačni očekivani omjer između upotreblijive i neupotreblijive građe dobivene preokretanjem DEH 70 iznositi će:

ktg. 2 + 5 + neupotreblijivi parovi iz 4 = 10.7% + 2% + 13%

ktg. 1 + 3 + (smanjena) 4 + 6 = 7% + 15% + 50% + 2.3%

Dakle, 25.7% neupotreblijive naprama 74.3% upotreblijive građe. Još uvijek eto — i nakon pažljivog ispravka udjela kategorije 4 — vidimo da su praktički tri četvrtine konvertirane građe leksikografski koristan materijal. U našem slučaju to znači da od ukupne konačne mase od 253.500 prijevodnih parova treba očekivati leksikografsku datoteku od 188.350 upotreblijivih parova (kartica).

6.

Kolika je to građa, shvatit ćemo najbolje ako je zamislimo raspoređenu na tekst DEH 70 s njegove 1162 stranice čistog rječničkog teksta. Na svaku stranicu otpalo bi u prosjeku oko 160 hrvatsko-engleskih prijevodnih parova, ili oko 80 parova na svaki stupac, odnosno 3 para na svaka 2 poluretku rječničkog teksta. Kad bismo, pri konačnoj odluci o uvrštavanju ili neuvrštavanju, odbacili i jednu trećinu svih leksikografskih parova, i u tom slučaju bismo konačno uvrstili blizu 120.000 parova ili *jedan par po poluretku rječničkog teksta!*

Šta to znači za obogaćivanje i poboljšavanje postojećeg rječnika (DHE 70), bit će sada jasno i laiku, a kamoli leksikografu. Ipak, ilustracije ra-

di, navest ću dvije nasumce izabrane kraće natuknice iz DHE 70 (*bič* i *bijeg*) i popratiti ih popisom upotrebljivih prijevodnih parova dosad dobivenih preokretanjem dvije trećine građe. To će — unatoč nepotpunosti konvertirane građe — vjerujem bolje od ičega drugoga predočiti korisnost preokretanja dvojezičnih rječnika kao leksikografskog postupka.

(Neizmijenjena natuknica iz DHE 70)

bič ~*m* whip; (*jahaći*) horse-whip, hunting-crop, riding-whip; (*mornarski /kazna/*) cat-o'-nine-tails; (*konop*) whip-lash, whip-cord; (*remen*) thong; (*udarac*) lash, stroke; *fig* scourge, calamity, plague | *pucketati* ~*em* crack (smack) a whip

bijeg ~*m* (*trk*) run, running; (*uteći*) flight, escape; *mil* retreat, rout; (*vojnika*) desertion; (*otmica*) elopement | *nagnuti u* ~ flee, take to flight, take to one's heels, run away; *natjerati u* ~ put to flight, rout, disperse; *umaći* ~*om* slip, give one the slip

Komentar:

Ostavit ću po strani neke nezgrapnosti i netočnosti u izvornom tekstu DHE 70. (Uostalom, one su uglavnom i uklonjene u revidiranom

(Korisni parovi iz DEH 70 /uz naznaku kategorije/)

bič : lash (3)
jahaći ~ : riding-crop (3)
držak ~*a* : crop (4)
švigir ~*a* : thong (4)
švigir ~*a* : whip-cord (4)
oštar i *brz udarac* ~*em* : lash (4)
kratak (oštar) udarac (trzaj) ~*em* : flick (4)
udarac ~*em* : swish (3)
zvižduk ~*a* : swish (4)
ošnuti ~*em* : whip (4)
istući ~*em* : horsewhip (4)
pucnuti ~*em* : flick (4)
zvrk koji se udara ~*em* : whipping-top (4)

bijeg : get-away (3)
bijeg : bolt (3)
 ~ *iz vojske* : desertion (4)
 ~ *iz tamnice (zatvora)* : prison-breaking (4)
nagli ~ : dive (4)
 ~ *glavom bez obzira* : precipitate flight (4)
to pruža priliku za ~ : this provides an escape (4)
čete naginju u ~ : troops break (4)
dati se u ~ : betake oneself to flight (4)
dati se u ~ : take to one's heels
natjerati u ~ : put to rout
jedini mu je spas (izlaz) bio u ~*u* : flight was his only escape (4)

izdanju iz 1978.) O gornja dva popisa prijevodnih parova treba međutim reći nekoliko načelnih stvari. Kao prvo, to je korisna građa koja se djelomično može odmah uvrstiti (npr. *jahači bič* : *riding-crop*; *bijeg* : *get-away*; *dati se u bijeg* : *take to one's heels, betake oneself to flight*), a dijelom doraditi (npr. *pucnuti bičem* : *flick* > *flick one's whip*), pa tek onda uvrstiti. Zatim, uvrštavanje može uz dopunu engleskog značiti i ispravak hrvatskog teksta, npr. *švigir biča* : *thong* spojiti će se s izvornim *konop (biča)* : *whip lash, whip-cord* u novo i bolje *švigir (biča)* : (općenito) *whip-lash*, (od uzice) *whip-cord*, (od kože) *whip-thong*. Neki parovi neće jednostavno biti uvršteni, jer su nepotpuni bez konteksta (npr. *bijeg* : *bolt*, što je jasno samo u *make a bolt for*) ili su nategnuti prijevodi (*nagli bijeg* : *dive*). Konačno, neki drugi parovi neće ući u gornje natuknice ni u najpsežnijem hrvatsko-engleskom rječniku, budući da leksički dominantna riječ nije ni *bič* ni *bijeg* (npr. *jedini mu je spas (izlaz) bio u bijegu* : *flight was his only escape*, što očito mora biti svrstano pod *spas* ili *izlaz*).

7.

Dobivena datoteka od četvrt milijuna jedinica neće se primjenjivati samo za usavršavanje postojećih ili rad na novim hrvatsko-engleskim rječnicima. Zahvaljujući svojem ustrojstvu i specifičnim osobinama, ova »banka leksičkih podataka« poslužit će i za teoretska, leksikološka, istraživanja.

Kako je svaka jedinica datoteke fizički nezavisna kartica, koja se lako razvrstava i premješta u uvijek nove skupine, datoteka je vrlo fleksibilan medij. Već s pomoću postojećeg abecednog reda u njoj, može se na primjer izraditi abecedna lista hrvatskih elemenata s frekvencijama, zatim čestotna lista tih elemenata. Time bismo, kao prvo, dobili kvantitetni opis frekvencijske strukture lijeve strane dvojezičnog rječnika, i mogućnost da te podatke usporedimo s čestotnim leksičkim odnosima u prirodnom tekstu istog jezika. Naša datoteka omogućila nam je dakle da ispitamo jedan važan aspekt rječničke strukture i tako vršimo bazična leksička istraživanja.

Zatim: izdvajanjem po abecedi »neleksičkih« parova u skupine po njihovoj sintaktičkoj funkciji, možemo dobiti specifične datoteke prijedloga, veznika, zamjenica, čestica, glagolskih oblika poput *je, ću* itd. Budući da su to visokofrekventne jedinice — a popraćene ekvivalentom na drugom jeziku — dobit ćemo bez posebnih napora na raspolaganje bogatu građu za razna kontrastivna istraživanja. Rezultati takvih specifičnih leksikoloških istraživanja bit će međutim i neposredno korisni za leksikografiju.

Prikazat ću to samo jednim primjerom. Među stotinama parova s hrvatskom povratnom zamjenicom *se* moći ćemo lako izdvojiti glagole koji počinju prefiksom *na-*. Većina tih glagola ima značenje »izvršiti radnju u znatnoj (pretjeranoj, zadovoljavajućoj) mjeri, ili do kraja«.

Postavlja se pitanje: postoji li na engleskoj strani takvih prijevodnih parova neka slična strukturna zakonitost? Ako postoji, našli smo područje s kontrastivnim potencijalom, u našem slučaju jedan uski (ali ne nevažni) kontrastivno proizvodni sektor leksika. Pregled građe ubrzo će nam pokazati da su većina engleskih prijevoda hrvatske glagolske formule *na- + Vb baza + se* i sami formule: *to have done -ing, to finish -ing, to have had enough of -ing, to Vb to one's heart's content, to Vb one's fill, to Vb a long time* i slično. Pun raspon ovih formula, uz značenjsko diferenciranje, prikazat ću na primjeru glagola *naprati se*:

a) u značenju »u znatnoj mjeri«

wash a long time
wash many times
wash often (enough)
wash repeatedly
wash again and again
wash a lot of
wash an awful lot of (+ b)
do plenty of washing
have done plenty of washing

b) u značenju »u pretjeranoj mjeri«

be weary of washing
get tired of washing
have (had) enough of washing (+ c)
have done one's share of washing (+ c)
have one's fill of washing (+ c)
wash one's fill (+ c)

c) u značenju »u zadovoljavajućoj mjeri«

wash to one's heart's content

d) u značenju »do kraja«

have done washing
finish / stop / cease washing

Bez obilne datoteke dobivene preokretanjem dvojezičnih rječnika prikupljanje ovako proizvodno iscrpne građe ne bi bilo moguće.

Doduše, gornje formule prikazuju samo primjenu engleskih kontrastivnih formula nepodložnih restrikcijama na hrvatsku leksičku strukturu *na- + Vb baza + se*. Postoje naime i engleske formule ograničene primjenljivosti. Na primjer, *naplakati se* : *have a good/long nice cry* ili *naspavati se* : *have a good / long sleep*; odnosno *nastrahovati se* : *suffer* (ili *live through*) *much fear*. Oba tipa kontrastivnih parova mo-

guća su, očito, samo kod engleskih glagola popraćenih homomorfnom imenicom (*to cry, to sleep, to fear* > *a cry, a sleep, a fear*).

Potpunosti radi, treba napomenuti da postoje i pojedinačni frazeološko-idiomatski, engleski ekvivalenti hrvatske strukture *na-* + Vb baza + *se*, koji nisu svodljivi na kontrastivno upotrebljive formule (*naspati se* : *have a good night's rest, get some sleep*; *naplakati se* : *cry one's eyes / heart out*; *naraditi se* : *to have (had) a long day*; *nakrasti se* : *line one's nest*; i slično).

Nije teško zamisliti mogućnost da se dobiju i jednako vrijedne kontrastivno-leksikografske proizvodne formule s hrvatskim *sufiksom* kao ishodištem. To bi međutim zahtijevalo da se naša četvrtmilijunska datoteka preabecedira po završetku obilježene hrvatske riječi na kartici. To bi značilo razgraditi dosadašnju datoteku ustrojenu po normalnoj abecedi. Fleksibilniji medij gdje se stvaranjem nove datoteke ne razgrađuje dotadašnja jest zapis na magnetskoj vrpci. Drugim riječima — kompjutorska obrada.

A to prelazi okvire ovoga rada.

Summary

BUILDING A BILINGUAL LEXICOGRAPHIC FILE BY MANUAL CONVERSION AND ITS EVALUATION

Establishing the existence of disparity between the contents of bilingual dictionaries of opposite direction (English-Croatian and Croatian-English), the author describes the lexicographic potential of the useful material thus hidden, stressing the need to make it available through dictionary conversion. This is followed by a description of the author's own method and experience in the manual conversion of DEH 70 (English-Croatian dictionary by M. Drvodelić of 1970 /rev. by Ž. Bujas/). The resulting material — a Conversion File of some 250,000 slips of Croatian-English translation pairs — is formally analyzed further and broken down into 6 formal categories of disparity. The share of lexicographically useful slips is determined to be 74.3%. This useful material is illustrated by two entries from DEH 70. In conclusion, the applicability of Conversion File to lexicological research is illustrated by contrastive analysis of the Croatian verb type *na-* + Vb + *se* and its English equivalents.