

Odgovor na članek Davorja Petrovića o težavah pri uporabi Evroterma in Evrokorpusa

A Reply to the Article by Davor Petrović on Evroterm and Evrokopus Databases

Evroterm in Evrokopus sta največji zbirki na svojem področju. Glede na to, kako sta nastajali (nastanek je opisan na spletnih straneh www.gov.si/evroterm), je jasno, da zaradi tega verjetno vsebujeta tudi največ napak. Gre za načelno vprašanje: ali naj pilimo stvari do perfekcije in naj slovarji izhajajo vsakih nekaj desetletij (npr. slovenski pravopis, veliki ang.-slo. slovar, tehniški slovar ...) ali pa uporabimo možnost, ki se je pokazala nekako pred 15 leti, in slovar ali korpus gradimo postopno, ga popravljamo, dopolnjujemo, odstranjujemo napake, dodajamo funkcionalnosti tudi pri programski opreми ipd., izdelek pa je ves čas dostopen na internetu. Če je naš cilj čim večja kakovost, se zdi, da je prva pot boljša – a ko pogledamo rezultate, v vsakem slovarju najdemo napake, tudi če je narejen še tako skrbno in tudi če so pri nastajanju sodelovali še taki strokovnjaki. Pri prevajanju pravnega reda Evropskih skupnosti ta možnost sploh ni prišla v poštev, ker je bilo za prevajanje časa le približno šest let. Edina možnost je bila, da je terminološka zbirka nastajala sproti, med prevajanjem. Velik del Evroterma je zbirka izrazja, ki smo ga dobili od različnih strokovnjakov; vir težav (nedoslednosti, več izrazov za isti pojem, uporaba tujk) je torej v strokah, tega pa Evroterm ne more odpraviti (morda bi zdaj šlo, ko so neskladja jasneje vidna – a očitno je, da stroke dopuščajo več izrazov za isti pojem). Nestandardizacija izrazja v Evrokopusu pa je preprosto posledica hitrega vzporednega dela res številnih prevajalcev (bolj ko je šel projekt h koncu, hitreje je teklo prevajanje), pri čemer je razpoložljiva terminologija stalno zaostajala za potrebami (po domače bi se reklo, da smo se z avtomobilom začeli voziti, čim je delal motor in so se vrtela kolesa, avtomobil pa smo dograjevali kar med vsakodnevno uporabo), uredniki pa na neskladje niso bili pozorni, ker dokumentov niso videli na tak način, kot jih je zdaj mogoče videti v Evrokopusu (kar nekaj težav je bilo, preden so ljudje sprejeli idejo spletnega korpusa za potrebe prevajanja – moj glavni razlog za uvedbo korpusa pa je bil prav poenotenje terminov – torej to, kar se zdaj kaže kot problem).

Toliko na splošno, zdaj pa še nekaj komentarjev k trditvam, zapisanim v članku. Gospod Petrović piše: »... se v analizi nisem oziral na stopnjo obdelave, ampak sem enakovredno obravnaval vse ustreznice.« Tale stavek je verjetno ključen pri vsej zgodbi in pojasnjuje, zakaj so rezultati analize taki, kot so. Pri običajnih slovarjih (kjer domnevamo, da so vsi izrazi popolnoma preverjeni) bi bil tak postopek v redu. Zakaj je v Evrotermu pri številnih izrazih podatek o zanesljivosti (številka od 1 do 5)? Zakaj je v Evrokorpusu pri vsakem zadetku označeno, iz katere faze prevajalskega postopka je dotični stavek? Prav zaradi tega, ker smo se ustvarjalci teh dveh baz ves čas zavedali, da ne moremo zagotavljati popolnosti – lahko pa prevajalcem ponudimo že uporabljene rešitve in jim z oceno zanesljivosti ali kakovosti vnaprej damo indikacijo glede vrednosti zadetka. Če je torej g. Petrović presodil, da prevodi terminov, ki imajo oceno manj kot 5, niso popolni, in je enako ugotovil tudi za zadetke iz korpusa, ki so na nižji stopnji od *finalizirane*, je ugotovil samo to, kar so napisali že snovalci obeh baz: tako označeni zbrani podatki niso popolnoma preverjeni. Je take stvari treba preverjati v analizi? Bi bilo bolje, da iz Evroterma odstranimo vse izraze, ki imajo oceno zanesljivosti manj kot 5, in iz Evrokorpusa vse prevode, ki imajo stopnjo redakcije nižjo od *finalizirane*? V prvi različici Evrokorpusa je bilo le okoli milijon besed, vse so bile iz delovnih prevodov (najnižja stopnja) – pa vendar sem tudi v tej zbirki pri svojem prevajalskem delu našel nekaj rešitev, ki jih ni bilo mogoče najti nikjer drugje, zato sem menil, da mora biti Evrokorpus čim prej javno dostopen.

Naslednji problematičen postopek pri analizi je ta, da g. Petrović enakovredno uporablja terminološko zbirko in korpus ter sešteva in primerja različne slovenske izraze za en angleški izraz. Jasno je, da je slovar (glosar, terminološka zbirka ipd.) narejen veliko skrbneje kot katerikoli korpus. V slovarju je zato praviloma manj napak, pri korpusu pa so napake, nedoslednosti, različni izrazi za isto stvar in podobne težave tako rekoč neizogibne. Slovar nam zato kaže idealen svet, korpus pa nam kaže dejansko stanje stvari – in tega ne moremo direktno primerjati. Ko presojamo vrednost zadetkov iz korpusa, se je zadeve torej treba lotiti statistično: če dobimo 100 zadetkov in se v 80 primerih uporablja en izraz, v 20 primerih pa drugi (za isto besedo v prvem jeziku), je verjetnost, da smo na pravi poti, če izberemo prvi izraz, približno 80-odstotna. Nič več. Gotovosti ni.

V tabeli avtor navaja npr. tale primer:

Community legislation	4	zakonodaja Skupnosti ^o	predpisi Skupnosti; pravila Skupnosti; pravo Skupnosti
-----------------------	---	-----------------------------------	--

V prvem stolpcu je angleški izraz, v drugem je število ustreznice v obeh bazah, v tretjem stolpcu naj bi bile primerne ustreznice, v četrtem pa neprimerne.

Če iskani izraz poiščemo v Evrotermu, dobimo le en prevod: zakonodaja Skupnosti (nazadnje popravljen 25. 2. 2003 – torej je bil avtorju na voljo tudi v času pisanja njegove diplomske naloge). To je edini prevod, ki ga predlaga Evroterm. Če potem iščemo v Evrokorpusu, dobimo 619 zadetkov (tale članek je bil napisan maja 2006, zato bodo bralci ob preverjanju po branju morda dobili drugačne rezultate), program pa je narejen tako, da izpiše samo prvih 200 zadetkov, pri čemer jih sortira po fazi redakcije in najprej izpiše najboljše. Od teh 200 jih je 58 v fazi »finalizirano«, preostali so iz faze »pravna redakcija«. Med temi 200 sem štirikrat našel prevod »predpisi Skupnosti«, nobenega prevoda »pravilo Skupnosti« in nobenega prevoda »pravo Skupnosti« – vsi drugi zadetki so bili »zakonodaja Skupnosti!«. Če iskanje omejim na finalizirane prevode (najvišja stopnja redakcije – tik pred objavo), dobim 58 zadetkov, od teh je 57 »zakonodaja Skupnosti« in eden »predpisi Skupnosti«. Statistično vzeto je rezultat odličen – in ga pri iskanju po korpusih redko dosežemo. Ta podatek kaže še na dvojje: v tabeli g. Petrovića manjka podatek o številu posameznih zadetkov – iz priložene tabele se zdi, da se ustrezne pojavi kar enakomerno pomešano, dejansko pa iz tega primera vidimo, da je število pojavitev priporočenega izraza iz Evroterma pravzaprav presenetljivo visoko. Prikazani rezultat kaže še na nekaj: korpus se osvežuje, prevodi v njem se izboljšujejo, čeprav morda ne na tak način, kot je predlagal avtor; po tistem, ko je delal analizo za diplomsko naloge, se je število prevodnih enot v korpusu povečalo približno za 2,5-krat, pri čemer smo dodajali predvsem prevode z višjih stopenj redakcije.

G. Petrović neke v svojem članku zapiše trditev: *če malce poenostavimo, bi lahko rekli, da moramo približno vsak tretji prevod v obeh bazah jemati z rezervo, saj obstaja verjetnost, da se njegov pomen le delno ujema z izvirnim terminom ali se sploh ne ujema*. Vem, da so v obeh zbirkah napake, vendar na oko (ob vsakdanji uporabi Evroterma in Evrokorpusa) nimam občutka, da je kar tretjina prevodov napačnih. Glede na prej navedena dejstva pa vidimo, zakaj je avtor prišel do tega sklepa: ker je upošteval število različnih prevodov nekega izraza, ne pa število ponovitev posameznega prevoda. Če se v 200 prevedenih stavkih za en angleški izraz uporabljata dva različna slovenska izraza, se na prvi pogled zdi (ob zahtevi, da za en koncept uporabljamo le en izraz), da je pol prevodov problematičnih. Slika pa je povsem drugačna, če upoštevamo še pogostost pojavljanja: potem (v prej omenjenem primeru) razmerje ni več 50 : 50 %, temveč 4 : 196 oz. 2 : 98 % (v finalizirani verziji celo 1 : 57, kar je še nekaj desetink odstotka boljši rezultat) – to je pa že nekaj čisto drugega; sam menim, da je napak in pomanjkljivosti v korpusu v resnici več kot 2 %. Celotne tabele seveda nisem preverjal (in verjetno v vseh primerih ne bi dobil tako dobrega rezultata), lahko pa preverimo še en izraz z dna tabele: za »unlawful« Evroterm predlaga prevod »nezakonit« ali »protipraven« (zadnji popravek novembra 2005 – torej

ne vemo, kakšno je bilo stanje v času pisanja diplomske naloge). V Evrokorpusu je med 30 finaliziranimi zadetki 29 prevodov (96,7 %) »nezakonit« in eden »nedovoljen«, če iskanje razširimo na vse prevode, se med 119 zadetki 83-krat pojavi prevod »nezakonit« (69,7 %), drugih nisem analiziral. Iz tega drugega primera vidimo, kakšna je razlika, če analiziramo vse prevode ali le finalizirane.

Glede več prevodov za en izraz ima avtor tudi nekaj pomislekov. Vzemimo primer iz analize:

Community law	6	pravo Skupnosti ^o	zakonodaja Skupnosti; pravni red Skupnosti; pravni akti Skupnosti; pravni predpisi Skupnosti; zakon Skupnosti
---------------	---	------------------------------	---

Za »Community law« sta v Evrotermu dve prevodni ustreznici: pravo Skupnosti in zakonodaja Skupnosti. Angleščina je precej ohlapen jezik (zato npr. vodja slovenske skupine prevajalcev za prevajanje sodb Evropskega sodišča predlaga, da bi sodbe prevajali iz jezika postopka, če to ni mogoče, pa raje iz francoščine ali nemščine kot iz angleščine) in lahko en izraz pogosto prevedemo na več načinov. Zanimivo pri tem je, da sta za omenjeni izraz dva prevoda tudi v francoščini in nemščini – torej stanje v slovenščini sploh ni tako grozno!

Če se mi ustreznice, ki so navedene v obeh bazah, niso zdele primerne, sem predlagal nove ustreznice, piše g. Petrović. Obdelal je 82 terminov, od teh se mu jih zdi povsem neustreznih sedem in zanje predlaga nov izraz, drugačen od tistega v bazi. Ker je teh primerov malo, si lahko ogledamo vse:

1. Izraz *illegal trade* je v Evrotermu preveden kot nedovoljena trgovina in nezakonito trgovanje (zadnja sprememba leta 2001). G. Petrović predlaga izraz nezakonito trgovanje. Torej smo usklajeni (in mi ni jasno, kako se je ta izraz sploh znašel na seznamu problematičnih).
2. Za *legal regulation* g. Petrović predlaga izraz pravna ureditev, v Evrotermu najdemo zakonska ureditev (z opombo, da gre za prevod iz slovenščine v angleščino), v Evrokorpusu pa je v stanju »finalizirano« le en zadevek: ... set out under a special legal regulation ... je prevedeno kot ... iz posebnih predpisov ...
3. Legislation je v Evrotermu prevedeno kot zakonodaja (z opombo: Služba za zakonodajo – Office for Legislation). G. Petrović tu predlaga uzakonitev (???). V Evrokorpusu je 960 finaliziranih zadetkov, v katerih se pojavlja izraz legislation, že pregled prvih desetih pa kaže, da uzakonitev pri njih ne pride v poštev: harmonization of legislation se v slovenščini čisto v redu sliši kot uskladitev (ali usklajevanje) zakonodaje, national legislation je nacionalna zakonodaja ipd.

Izraz uzakonitev ne pride v poštev v nobenem od teh primerov (verjamem pa, da bi bil morda kdaj tudi to primeren prevod). Tomaž Longyka je v svojem slo.-ang. pravnem slovarju napisal, da izrazu zakonodaja ustreza legislation, za uzakonitev (česa) pa ima izraz regulation (of sth.) by statute. Evroterm pravi, da je uzakonitev enactment.

4. *Izraza legally committed ni v Evrotermu, v Evrokorpusu sta dva zadetka (oba na stopnji »finalizirano«), prevedena kot zavezovati. Strinjam se s predlogom g. Petrovića, da je v tem primeru natančnejši izraz pravno zavezovati. Izraz se v 700.000 prevodnih enotah pojavlja le dvakrat, torej ga verjetno ne srečamo prav pogosto – in ga najbrž zaradi tega tudi ni v Evrotermu.*
5. *Tudi izraz legally defined position je zelo redek: v Evrokorpusu se pojavi le enkrat; v stavku, ki je šel skozi pravno redakcijo, je preveden kot zakonski položaj. V Evrotermu tega izraza ni, g. Petrović predlaga pravni položaj ali z zakonom določen položaj.*
6. *Legally-held weapons je že tretji silno redek izraz: obstaja le en zadetek v Evrokorpusu (pravna redakcija) in tam je national inventories of legally-held weapons owned by the country's authorities preveden kot nacionalni register zakonito hranjenega orožja v lasti državnih organov. G. Petrović predlaga dovoljeno orožje, a se mi zdi, da v tem kontekstu to ne gre.*
7. *Zadnji očitek leti na legislative approximation, ki naj bi bil po Petrovićevem predlogu preveden kot približevanje zakonodaje. V Evrotermu so pod enim vpisom združeni trije izrazi: approximation of laws, approximation of legislation in legislative approximation, predlagana prevoda pa sta dva: približevanje zakonodaje in prilagajanje zakonov (tudi v nemščini sta dva prevoda). Zadnji popravek tega vpisa je res novembra 2005, vendar je citat iz Evropskega sporazuma o pridružitvi med RS in Evropskimi skupnostmi in njihovimi članicami, torej je bil podatek vpisan že prej (ne vemo pa, ali je bil g. Petroviću na voljo med pisanjem diplomske naloge).*

In kakšen je rezultat? Od sedmih »neustreznih« prevodov se dva v Evrotermu skladata s Petrovićevimi predlogi. Predlog, da bi legislation prevedli kot uzakonitev, je mimo (morda gre za napako v tabeli). Ostanjejo trije silno redki izrazi (ki se vsi skupaj v Evrokorpusu pojavijo le štirikrat(!)). Poleg teh je še legal regulation, pri katerem ne morem biti razsodnik. Ampak vseeno: kaj je že pisalo o analizi: če malce poenostavimo, bi lahko rekli, da moramo približno vsak tretji prevod v obeh bazah jemati z rezervo. Ali pa moramo malo z rezervo jemati tako mnenje?

Pravo ni moja specialnost, si pa lahko predstavljam, da se tudi tu, podobno kakor npr. v tehniških strokah, za en pojem lahko uporablja več kot en izraz. Konkretno za izraz »illegal immigration« v Evrotermu obstajata dva prevoda: »ilegalne migracije« in »nezakonito priseljevanje«. Izraz »ilegalne migracije« ima oceno zanesljivosti 5, kot vira sta navedena Ministrstvo za zunanje zadeve in Služba vlade za zakonodajo, povsem slovenska ustreznica nima ocene. Domnevam, da so v tem primeru oba izraza potrdili pravniki – in ti imajo pač različne poglede na stvari. Problem pri tej zadevi je, da v času prevajanja zakonodaje Evropskih skupnosti ni bilo strokovnega glosarja ali slovarja, ki bi ga naredila ustanova s primerno težo (npr. Pravna fakulteta). Verjetno bi bili prevodi pravnih besedil najboljši, če bi jih prevajali pravniki prevajalci – a v Sloveniji je takih ljudi premalo in po besedah Darje Erbič, ki je bila prvi vodja prevajalske ekipe za prevajanje pravnega reda EU, je bil velik problem v projekt pritegniti pravnike. Kot zanimivost: Pravna fakulteta v Ljubljani se je problematike pravne terminologije v Evrotermu lotila šele po vstopu Slovenije v EU, analiza pa še vedno ni končana.

Še en citat iz članka g. Petrovića: ... *se na predlagane ustreznice v Evrotermu in v Evrokorpusu prevajalci ne morejo povsem zanesti, ampak jih morajo preverjati še pri drugih virih – to pa je zelo zamudno opravilo, ki si ga večina strokovnih prevajalcev zaradi časovne stiske ne more privoščiti, zato trpi kakovost prevodov in jasnost prevedenih besedil.*

Tristo hudičev! Ali sem pravilno razumel navedeni stavek? Zaradi napak v Evrotermu trpi kakovost prevodov?! Če je tako, bi bilo pa treba Evroterm prepovedati! Ali pa vsaj odstraniti s spleta! Seveda bi bilo idealno, če bi obstajal en sam slovar, v katerem bi prevajalec našel vse odgovore na svoja vprašanja. Ampak česa takega za zdaj še ni. Za prevajalce bi bil zelo uporaben *OneLook Dictionary* (www.onelook.com) za slovenščino. Upal sem, da se bom lahko sam lotil tega, a za zdaj zaradi pomanjkanja časa slabo kaže. Tudi tak »enogledni« slovar ne bi dal vseh odgovorov, bi pa skrajšal čas iskanja. Viri niso nikoli stoddotno zanesljivi. Edina rešitev za prevajalca je specializacija za eno ali dve področji, katerih terminologijo pozna (in razume!) v obeh jezikih. Če jo razume, je slovar zanj dodatni vir informacij in lahko utemelji svoj prevod, tudi če je kakšna beseda zapisana drugače kot v slovarju (potem tudi ne potrebuje 30 splošnih in posebnih slovarjev ter priročnikov). Če se zanaša le na slovar in nima svojega znanja, je pa hudo, se strinjam. No, test je preprost: po mojih izkušnjah s prevajanjem računalniških in strojniških besedil je v današnjih besedilih, ko vsi pisci predvsem hitijo, na vsakih pet strani besedila vsaj ena vsebinska napaka. Če jo prevajalec najde in ve, kaj bi v resnici moralo pisati tam, potem obvlada zadevo. Če ne, se pa bojim, da bo poleg Evroterma potreboval še veliko dodatnih virov.

Končam lahko tam, kjer sem začel: Evroterm bo toliko dober (uporaben, zanesljiv, standardiziran ...), kolikor ga bodo sooblikovali njegovi uporabniki. Pri slovarju, ki je natisnjen v knjigi, so napake nepopravljive do natisa nove izdaje, pri Evrotermu pa so uporabniki povabljeni k izboljševanju trenutnega stanja; v zadnji tretjini zapisa o Evrotermu na spletu piše: *Evroterm je živa zbirka, v kateri se terminologija stalno izboljšuje in dopolnjuje, pri čemer so dobrodošli vaši strokovni predlogi. Komentarje glede vsebine terminološke zbirke ipd. pošljite terminologiji Adriani Krstič - Sedej ali Nataši Zaviršek - Žorž, pripombe glede delovanja programa pa avtorju programa Miranu Željku.* Klik na ime vam (ob pravilni nastavitvi programa) odpre program za pošiljanje pošte z vpisanim elektronskim naslovom ciljne osebe. Več od tega ne moremo narediti. Ko pošljete sporočilo, pa je od stanja vpisa odvisno, kaj se bo zgodilo - če podatek še ni vpisan, bo verjetno dodan brez težav; če najdete tipkarsko napako, bo odpravljena prej kot v enem tednu; če pa že obstaja prevod, ki ima zanesljivost ocenjeno s 5, bo treba predlog za spremembo malo bolj utemeljiti.