

Pripomočki na spletu za prevajalce zakonodaje EU

Miran Željko

Center Vlade RS za informatiko
Langusova 4, Ljubljana
miran.zeljko@gov.si

Povzetek

Prevajanje zakonodaje Evropske unije je izjemno obsežno delo, pri katerem sodeluje veliko prevajalcev. Del prevajalcev je zaposlen v službi Vlade RS za evropske zadeve, večina prevajalcev pa dela pogodbeno. Prevajalci, ki jih zaposluje država, imajo na voljo sodobne prevajalske pripomočke podjetja Trados, zunanji prevajalci pa so glede tega bistveno slabše opremljeni. Da bi zagotovili usklajeno terminologijo tudi pri zunanjih prevajalcih in izboljšali kakovost teh prevodov, smo podatkke, ki nastajajo ob prevajanju, prenesli na splet (<http://www.gov.si/evroterm>). Tam so prevajalcem na voljo v dveh oblikah: kot slovar in korpus, pri čemer je slovar povezan s korpusom.

Abstract

Due to the country's endeavours to become an EU member, Slovenian translators have to translate an enormous volume of documents. This work is organised by the Translation Unit of the Government Office for European Affairs (GOEA). Although there are 25 in-house translators employed at GOEA, the majority of translations are done by translation agencies and freelance translators. The translators at the GOEA Translation Unit have created a terminology database which is continuously maintained and updated. In order to make this database available to translators outside the GOEA we have developed our own software for web access (<http://www.gov.si/evroterm/e/>). This solution allows us to customise output, to develop a web interface in several languages, to obtain the desired data about a certain term (origin, subject, definition...) and so on. We recently incorporated a concordance function, which shows the translator how a particular word is used in context.

1. Uvod

Pred vključitvijo Slovenije v Evropsko unijo bodo morali slovenski prevajalci prevesti ogromno količino pravnih dokumentov (Krstič, 1999), večinoma iz angleščine v slovenščino in obratno. Pri tem nastaja več težav: pojavljajo se novi izrazi, obstoječi splošni slovarji (npr. Veliki angleško-slovenski in Veliki slovensko-angleški slovar) so pomanjkljivi in zastareli, v novejših pravnih slovarjih (npr. Pravni slovensko-angleški prevajalski slovar, Pravni terminološki slovar, Slovenski pravni leksikon) terminologija še ni povsem usklajena (Velkaverh, 2001), terminologija posameznih strok pa je razdrobljena v množici malih terminoloških slovarjev (na spletu npr. (Željko, 2002)). Pri tako obsežnem projektu je torej usklajeno terminologijo možno zagotoviti le tako, da vzpostavimo lastno terminološko bazo.

Prevod zakonodaje Evropske unije v Sloveniji poteka v prevajalskem oddelku Službe Vlade RS za evropske zadeve (SVEZ, 2002). V prevajalskem oddelku dela 25 prevajalcev, dve terminologinji in tri lektorice, poleg njih pri prevajanju pogodbeno sodeluje približno 150 prevajalcev (samostojnih in tistih, ki delajo prek prevajalskih agencij). Zakonodajo z istega področja torej prevajajo različni prevajalci – pri tem pa je zelo pomembno, da se vsi držijo enotne terminologije.

V prevajalskem oddelku SVEZ-a je to rešeno z uporabo skupne terminološke baze (uporabljajo program MultiTerm nemškega podjetja Trados), poleg tega prevajalci pri delu uporabljajo tudi program s pomnilnikom prevodov (Tradosov program Translator's WorkBench (TWB)). Programi podjetja Trados so bili izbrani, ker jih uporabljajo tudi druge članice EU in kandidatke za članice, nakup za prevajalce, zaposlene v SVEZ-u, pa je financirala EU. Prevajalci, zaposleni v SVEZ-u, so vsi na isti lokaciji in si ob reševanju težav lahko medsebojno pomagajo.

Večje težave so z zunanjimi prevajalci. Ker gre za novo področje, za katero niti nimamo popolnih slovarjev, je izkušenih prevajalcev silno malo, prevajalci so bistveno slabše medsebojno povezani in so praviloma tudi slabše opremljeni z računalniškimi pripomočki. Tradosova orodja so za naše razmere precej draga (okoli 200.000 SIT za osnovno verzijo programa), zato lahko njihov nakup ekonomsko upraviči le malo prevajalcev. V zadnjem letu omogoča delo s Tradosovimi bazami tudi brezplačen program Wordfast (<http://www.champollion.net/>), a je ta program še v razvoju. Na srečo se je v zadnjih letih povečala izobrazenost prevajalcev pri uporabi interneta, zato smo že leta 2000 začeli s prenosom terminologije na internet (Krstič in Željko, 2000), leta 2002 pa smo terminologijo dopolnili še s korpusom prevodov (Željko in Krstič, 2002).

2. Namen članka

V uvodu je bila omenjena osnovna težava, ki jo je treba rešiti pri prevajanju zakonodaje EU: uporaba enotne terminologije pri vseh sodelujočih prevajalcih. Rešitev za prevajalce, zaposlene na SVEZ-u, je opisana v (Krstič, 1999) in (Krstič in Belc, 1999), v tem prispevku pa bom opisal lastne rešitve, ki smo jih razvili za pomoč zunanjim prevajalcem.

Prevajalci v SVEZ-u so priključeni na lokalno računalniško omrežje, zaradi varnosti pa zunanji prevajalci prek interneta nimajo dostopa do tega omrežja; glede na običajno hitrost dostopa tak način dela tudi ne bi bil posebej učinkovit. Podatke je zato treba preoblikovati in shraniti na poseben spletni strežnik.

Leta 2000, ko smo začeli s prenosom podatkov na internet, je podjetje Trados nudilo prenos terminologije na splet, a je ta program deloval le na strežnikih, ki so tekli na operacijskem sistemu Windows NT (ta omejitev je veljala tudi leta 2002, glejte (Trados, 2002)). Na Centru

Vlade RS za informatiko (CVI, 2002) smo prvi spletni strežnik postavili leta 1994 (Mat'Kurja, 1994) in smo takrat izbrali operacijski sistem unix. Tekom let se je ta izbira izkazala kot zanesljiva, stabilna in varna in je nismo želeli spreminjati samo zaradi terminološke zbirke, ki v celotni ponudbi podatkov državne uprave predstavlja razmeroma majhen delež – to pa je pomenilo, da bomo morali rešitev razviti sami.

Zadeve smo se lotili v treh korakih:

- prenos terminologije na internet,
- pretvorba pomnilnikov prevodov v dvojezični korpus in prenos korpusa na internet,
- povezava terminologije in korpusa.

Programne za izvedbo te rešitve smo napisali v programskem jeziku perl, uporabljamo pa tudi nekatere programe operacijskega sistema unix.

3. Prenos terminologije na internet

SVEZ-ovi prevajalci uporabljajo program MultiTerm za urejanje in iskanje po terminologiji. Program (MultiTerm, 2002) ima podatke urejene v lastni bazi, če pa želimo te podatke obdelovati po svoje, jih lahko izvozimo v tekstni obliki, ki je že prirejena za nadaljnjo obdelavo: v vsaki vrstici je navedeno ime parametra med znakoma "<" in ">" ter vrednost tega parametra. Različni vpisi imajo lahko različne parametre, obvezna sta le dva: zaporedna številka vpisa (ki jo dodeli program MultiTerm ob vnosu novega podatka) in beseda ali pojem v enem jeziku (z oznako jezika), zadeva pa ima smisel šele, če je naveden prevod te besede vsaj v enem jeziku (z oznako jezika). Pri vsakem vpisu je možno imeti še celo množico drugih podatkov, zasnova baze pa je zelo fleksibilna, saj imajo različni vpisi lahko različne podatke (npr. prevodi v enega ali več jezikov, različni komentarji za različne jezike ...), pri enem vpisu se nekateri podatki lahko ponavljajo (npr. če za besedo v enem jeziku obstaja več različnih prevedkov v drugi jezik) in nekatera polja so medsebojno povezana (npr. komentarjev je lahko več, vsak pa se nanaša na polje pred njim). Pomen posameznih polj je naslednji:

<i>Creation Date</i>	datum vpisa v originalno bazo na SVEZ-u
<i>Created by</i>	ministrstvo ali vladna služba, kjer je bil prevod opravljen (če ni znano, na katerem ministrstvu je prevod nastal ali je bilo virov več, je oznaka <i>super</i>)
<i>Entry number</i>	zaporedna številka vpisa v originalno bazo
<i>Subject</i>	področje (v angleščini s kodo)
<i>Subj</i>	področje (v slovenščini)
<i>Project</i>	projekt, kjer so prevodi nastajali ali so se preverjali izrazi (neobvezno)
<i>SourceDoc&Lang</i>	naslov dokumenta, ki se je prevajal oz. vir izrazja (neobvezen podatek); iz tega podatka je mogoče sklepati, ali se je dokument prevajal iz tujega jezika v slovenščino (npr. akti ES kot direktive, uredbe, odločbe, včasih označeni s številko Celex) ali iz

	slovenščine v tuj jezik (slovenska zakonodaja, državni program za prevzem pravnega reda ES)
<i>Definition</i>	opredelitev pojma (priporočen podatek)
<i>TermRef</i>	vir prevedka (dokument, ministrstvo ali oseba; priporočen podatek)
<i>Note</i>	opombe
<i>Reliability</i>	zanesljivost prevoda (od 1 do 5); te podatke imajo vpisi, ki so v bazi od leta 2001 dalje

Da bi zagotovili primerno hitrost iskanja ob uporabi preproste baze, osnovno bazo očistimo podatkov, ki niso nujno potrebni (odstranimo npr. polja <graphic>, <Entry Class>, <NoteRef> in <TermTyp>). Indeksne datoteke naredimo za štiri najuporabnejše jezike: slovenščino, angleščino, francoščino in nemščino. Program za iskanje po terminološki zbirki na spletu smo imenovali Evroterm (Željko in Krstič, 2002) in je dostopen na naslovu <http://www.gov.si/evroterm>.

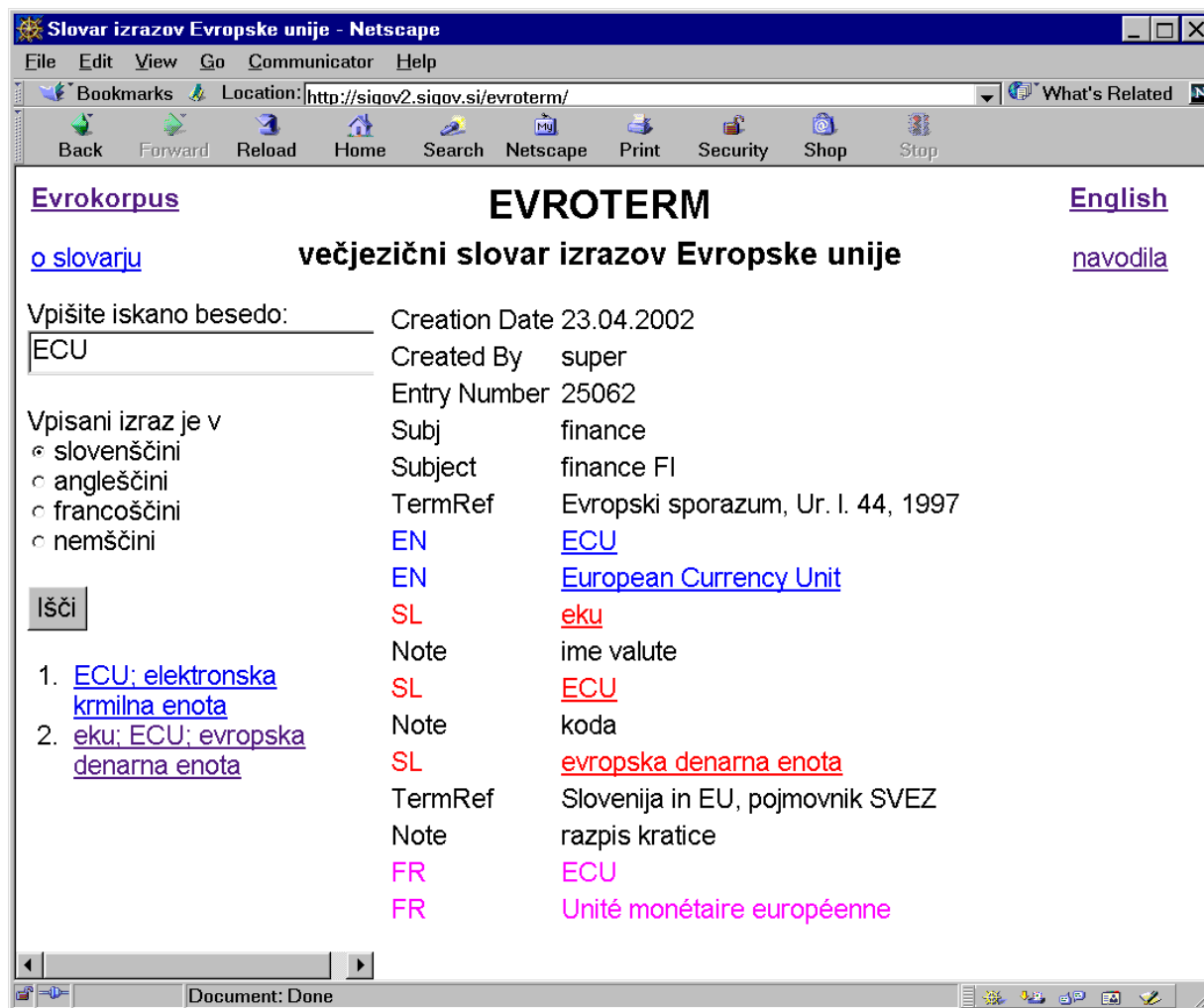
Spletni uporabniški vmesnik je sestavljen iz treh delov (slika 1): v zgornjem delu je vedno dostopen opis slovarja, navodila za iskanje, preklop med slovenskim in angleškim uporabniškim vmesnikom ter preklop med terminologijo in korpusom; levo vpisujemo iskano besedo in označimo jezik, kot rezultat pa dobimo seznam zadetkov, sortiran po abecedi in ko kliknemo enega od zadetkov, se vsi podatki o določenem vpisu izpišejo v desnem okviru. Za večjo preglednost so prevodi v štirih najpomembnejših jezikih napisani v štirih različnih barvah in jih tako prevajalec hitreje najde med množico drugih (praviloma manj pomembnih) podatkov. Pri iskanju program najde vse besede ali nize besed, ki se začnejo z vpisanim nizom znakov, s pomočjo unixovih regularnih izrazov pa lahko izvedemo bolj kompleksna iskanja, kar je včasih koristno, ko npr. iščemo besede, ki imajo drugačen zapis v britanski in ameriški angleščini (npr. z iskalnim nizom "organi[sz]ation" bi našli besede, ki vsebujejo *organisation* in *organization*; z nizom "colou?r" pa bi našli *color* in *colour*) – podrobnejša navodila za iskanje so dostopna na spletni strani Evroterma.

4. Pretvorba pomnilnikov prevodov v dvojezični korpus in prenos na internet

SVEZ-ovi prevajalci med prevajanjem uporabljajo program Translator's WorkBench (TWB, 2002). To je program s pomnilnikom prevodov (Vintar, 1999; Željko, 2000), ki mu kot uporabniški vmesnik služi urejevalnik besedil MS Word: ko prevajalec začne prevajati, dobi v okvirju prevodno enoto (običajno je to stavek) izvornega besedila, v naslednji okvir pa napiše prevod. Ko izoblikuje prevod in želi prevesti naslednji stavek, se izvornik in prevod prejšnjega shranita v pomnilnik. Pred prevajanjem naslednjega stavka program pregleda svojo bazo, če je morda v njej enak ali vsaj zelo podoben izvornik in ponudi prevajalcu izvornik in prevod. Prevajalec se potem odloči, ali bo obdržal prevod iz baze, ga morda popravil ali začel prevajati čisto od začetka. Pomnilnik prevodov je uporaben z dveh vidikov: če se besedila ponavljajo, gre prevajanje hitreje (kar je

uporabno npr. pri spremembah obstoječe zakonodaje), poleg tega pa je v pomnilniku prevodov skrito veliko

uporabne terminologije, ki je v terminološki zbirki morda ni.



Slika 1: Spletni uporabniški vmesnik Evroterma

Tudi bazo programa Translator's WorkBench lahko pretvorimo v besedilno obliko. Primer zapisa:

```
<TrU>
<CrD>
<CrU>BRANKA
<Att L=Status>Prvi prevod
<Att L=Področje>KMETIJSTVO
<Att L=Podpodročje>varstvo rastlin
<Att L=Vrsta dokumenta>Directive
<Att L=Naročnik>SVEZ-MKGP
<Txt L=Celex številka>32000L0029
<Txt L=Besedilna zbirka – TWB>en_sl_KMETIJSTVO
<Seg L=UKE>Having regard to the opinion of the
Economic and Social Committee,
<Seg L=EUR>ob upoštevanju mnenja Ekonomsko-
socialnega odbora,
</TrU>
```

Za naše potrebe sta najpomembnejši vrstici z

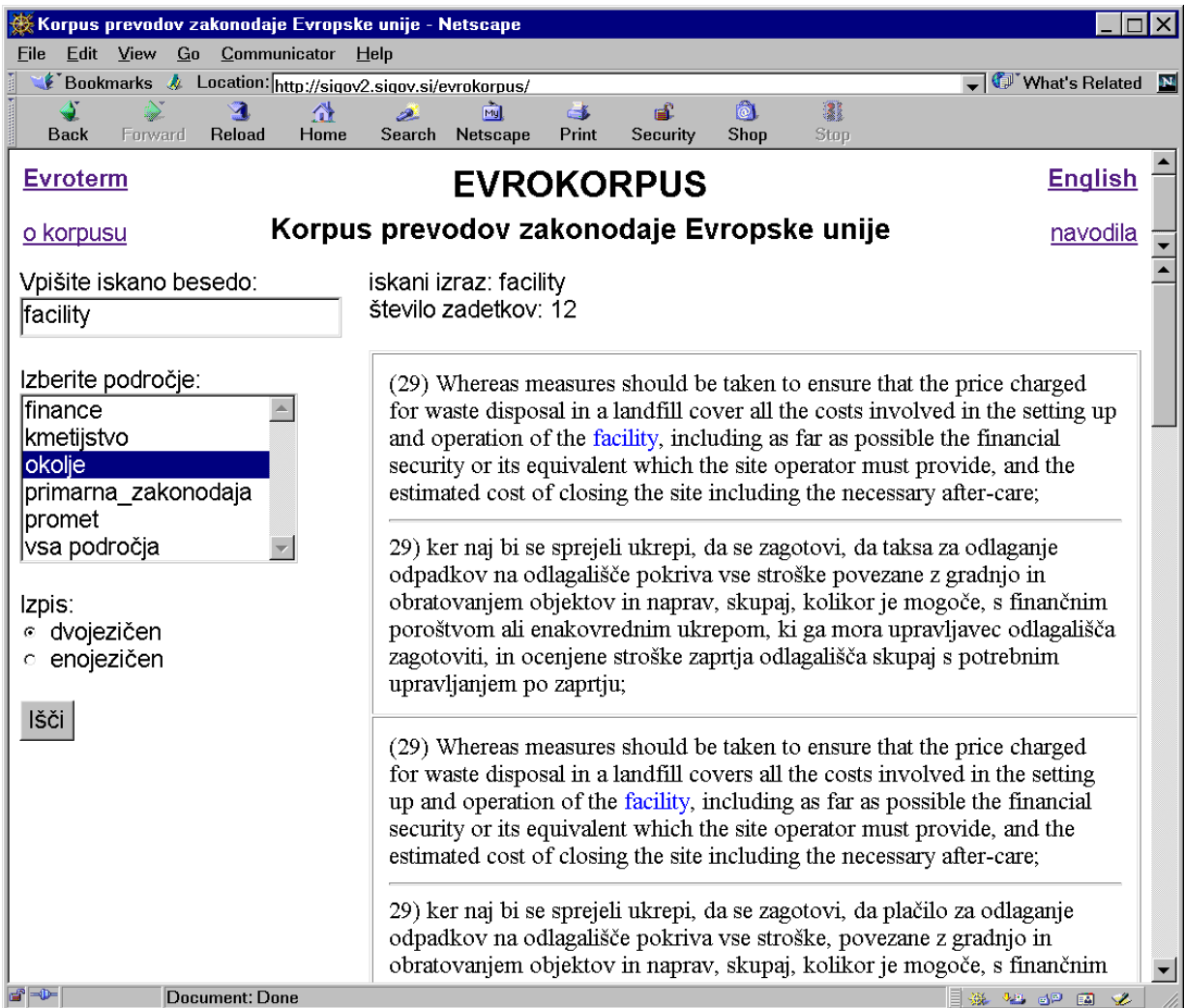
oznakama <Seg L=UKE> in <Seg L=EUR>: v prvi je prevodna enota v angleščini, v drugi pa je njen prevod v slovenščino. S tem smo dobili najpreprostejšo obliko dvojezičnega korpusa. Posamezni pomnilniki prevodov so ločeni po področjih in vsako področje je shranjeno v svoji datoteki. Avgusta 2002 so bila na voljo naslednja področja: finance, kmetijstvo, okolje, primarna zakonodaja EU in transport.

Uporabniški vmesnik je podoben kot prej (slika 2): zgoraj je stalni del z enako navigacijo kot pri Evrotermu, na levi uporabnik določi iskalne pogoje in izbere način izpisa, desno pa se izpišejo rezultati. Iskanje je možno omejiti na posamezno področje, lahko pa iščemo po vseh bazah. Izpis je lahko dvojezičen (v tem primeru dobimo celotno prevodno enoto v izvorniku in prevodu) ali enojezičen (v tem primeru dobimo v izpisu samo okoli 50 znakov levo in desno od iskanega pojma). Dvojezični izpis je koristen takrat, kadar prevajalec ne pozna prevoda določenega pojma v danem kontekstu (npr. te besede ni v slovarju) in ga zanima, kako so ta pojem v preteklosti prevajali drugi

prevajalci. Enojezični izpis (slika 3) je v bistvu konkordančni in je koristen takrat, ko prevajalec sicer pozna pomen določenega pojma, a ga ob tem zanimajo druge posebnosti uporabe, ki iz slovarja ali jezikovnih priročnikov niso razvidne (ta podatek je posebej koristen

pri prevajanju v tuj jezik).

Program za iskanje po korpusu na spletu smo imenovali Evrokopus.



Slika 2: Dvojezični izpis v korpusu prevodov

5. Povezava terminologije in korpusa

V SVEZ-ovi terminološki zbirki se občasno pod enim vpisom pojavlja več besed v enem ali drugem jeziku – taka sta npr. izraza *facility* ali *sustainable* (v obeh primerih imamo za en izraz v angleščini kar štiri izraze v slovenščini). Pri tem dostikrat ne gre za sinonime, temveč je vsak izraz primeren le v določenem sobesedilu. Prevajalec, ki mu je neko področje novo, pa nima občutka za izbiro pravega izraza. V slovarjih v knjižni obliki je ob večpomenskih izrazih navedenih nekaj praktičnih primerov, v našem slovarju pa si lahko pomagamo kar s korpusom: pri podrobnem izpisu podatkov o določenem izrazu v programu Evroterm je besede v angleščini in slovenščini možno klikniti; ob tem dobimo na ekran rezultat iskanja dotične besede po celotnem korpusu, z

navedbo področja, v katerem se je pojavila najdena beseda.

6. Rezultati

Najnovejše podatke s SVEZ-ovega lokalnega omrežja prenesemo na splet približno enkrat mesečno. V terminološki bazi je bilo septembra 2002 več kot 26.000 vpisov, v korpusu pa okoli 4 milijone besed.

V primerjavi s Tradosovo rešitvijo ima naša naslednje bistvene prednosti:

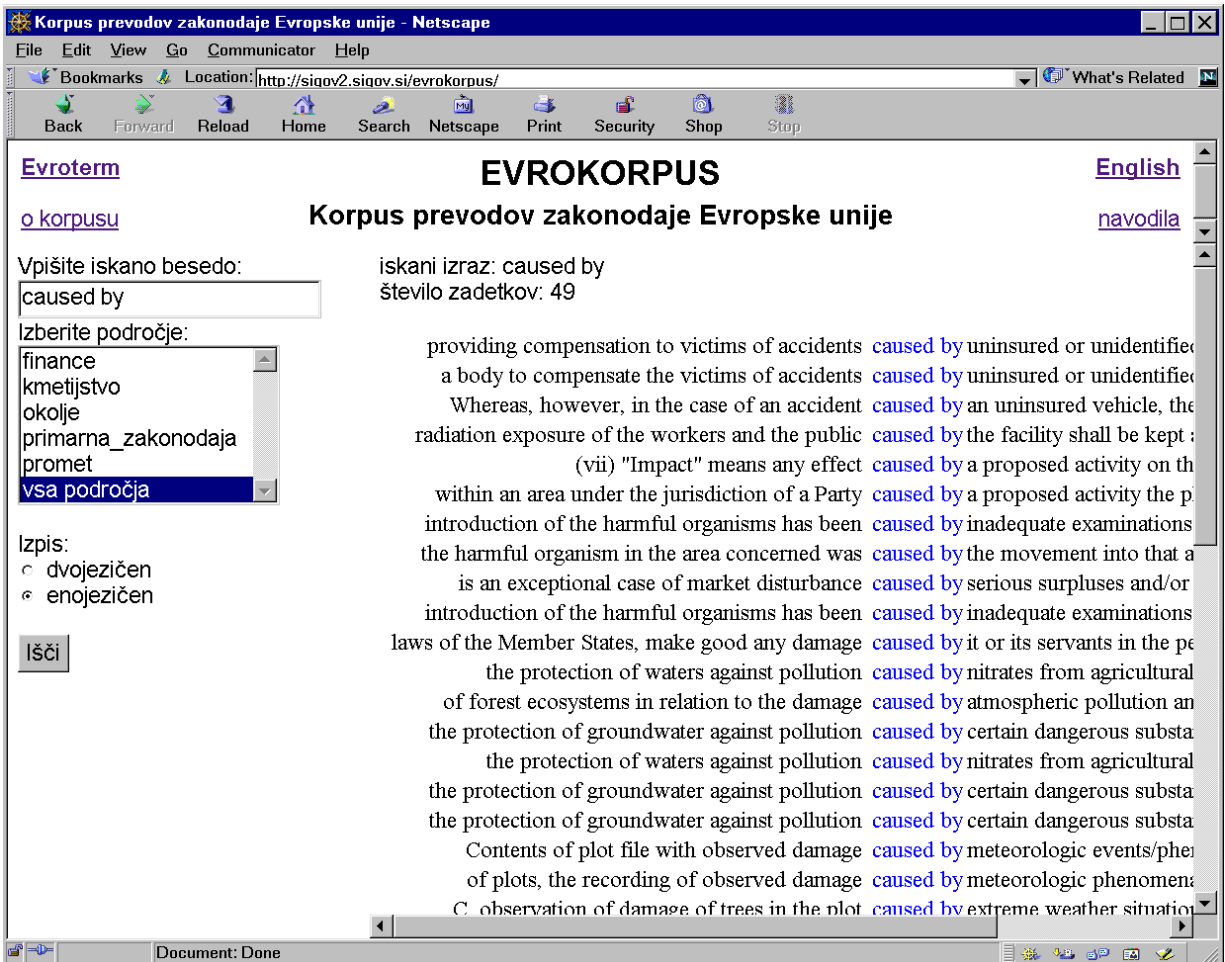
- na splet smo prenesli terminologijo in korpus iz pomnilnikov prevodov ter povezali terminologijo s korpusom;

- lastna rešitev nam omogoča, da jo postopno nadgrajujemo, pri čemer skušamo obdržati preprost in konsistenten uporabniški vmesnik; ta je shranjen ločeno

od programa, zato lahko preprosto naredimo novo jezikovno različico, če bi se pokazale take potrebe;

• podatki so v obliki datotek z oznakami besedila ("tagged text files"), kar pomeni, da predstavljena rešitev

ni omejena na prenos Tradosovih baz na splet, temveč bi lahko uporabili katerikoli drug program s podobno funkcijo, ki zna izvoziti podatke v podobni obliki.



Slika 3: Enojezični izpis v Evrokorpusu

Z vidika SVEZ-a ima prenos omenjenih podatkov na splet naslednje koristi:

• zunanji prevajalci imajo dostop do istih podatkov kot prevajalci, zaposleni na SVEZ-u;

• terminologija v prevodih je bolj usklajena in prevodi zunanjih prevajalcev so boljši;

• prevode na internetu vidi več ljudi, zato hitreje najdejo napako in se ta lahko hitreje odpravi;

• zaradi boljših prevodov je delo strokovnih in pravnih redaktorjev ter lektorjev hitrejšo in lažje;

• ker so vsi podatki javno dostopni, je to najboljše zagotovilo, da se bo v Sloveniji uporabljala enotna terminologija v zvezi z zakonodajo EU – pomembno je, da je terminologija na voljo tudi novinarjem, študentom prevajanja in prevajalcem, ki podobna besedila prevajajo za druge naročnike.

Ob takih in podobnih rešitvah se je seveda treba zavedati, da so te vredne toliko, kolikor jih ljudje uporabljajo v praksi. Statistika dostopov kaže, da so spletne strani Evroterma (med vsemi stranmi državnih

ustanov), med tistimi, pri katerih število obiskov najhitreje raste: v zadnjem letu se je mesečno število dostopov povečalo približno za trikrat, skupno pa je bilo v zadnjem letu več kot 230.000 poizvedb v slovarju (podrobnosti so na naslovu: <http://sigov1.sigov.si/evroterm/stat.html>); na lestvici dostopov do vseh strani na strežniku www.gov.si (podrobnosti so na: <http://sigov1.sigov.si/stat.html>) pa je bila septembra 2002 najbolj obiskana stran Evroterma na 5. mestu (pred njo so bile datoteke Ministrstva za zunanje zadeve in Državnega zbora; vseh datotek na strežniku je okoli 50.000). K temu rezultatu je nekaj verjetno prispevalo tudi dejstvo, da je DZS leta 2002 umaknil spletno različico Velikega angleško-slovenskega slovarja in je zato Evroterm trenutno najboljše spletni slovar, ki vključuje slovenščino. Ti rezultati ponovno potrjujejo mnogokrat napisano in izrečeno trditev, da ljudje na spletu iščejo predvsem konkretno uporabno vsebino, ki je na spletnih straneh državnih ustanov, kljub množici podatkov in slik, žal, obupno primanjkuje.

7. Zaključek

V članku je opisana izvedba prenosa Tradosove terminološke baze in pomnilnika prevodov na spletni strežnik, ki teče pod operacijskim sistemom unix. Ker razvijalec in prodajalec programov Trados ni nudil ustrezne rešitve, smo jo razvili sami. S prenosom podatkov na splet smo omogočili dostop do terminološke zbirke in korpusa prevodov vseh zainteresirani javnosti. Služba Vlade RS za evropske zadeve največje koristi pričakuje pri poenotenju terminologije, če bodo te podatke dosledno uporabljali njihovi zunanji sodelavci – prevajalci zakonodaje EU.

8. Literatura

- CVI (Center Vlade RS za informatiko). 2002.
<http://www.gov.si/cvi>
- Krstič Adriana. 1999. Prevajanje zakonodaje Evropske unije, *Mostovi 1999*, str. 55-59
- Krstič Adriana, Belc Jasna. 1999. Prevajanje pravnih aktov Evropske unije. *Uporabno jezikoslovje 1999* (7-8), str. 191-205
- Krstič Adriana, Željko Miran. 2000. Evroterm – terminologija EU na internetu, *Zbornik referatov posvetovanja INDO 2000*, str. 112-116.
<http://www.gov.si/cvi/slo/indo/indo2000/referati/referat14.html>
- Mat'Kurja. 1994. Kaj je novega za december 1994.
<http://www.matkurja.com/slo/new/1994/12/>
- MultiTerm. 2002.
http://www.trados.com/mt_en.asp?lang=en&cat=8&action=mt&show=1&
- SVEZ (Služba Vlade RS za evropske zadeve). 2002.
<http://www.gov.si/svez>
- Trados. 2002. <http://www.trados.com>
- TWB (Translator's WorkBench). 2002.
http://www.trados.com/cts_en.asp?lang=en&cat=8&action=cts&site=21&show=trans&
- Velkaverh Gregor. 2001. Pravni slovensko-angleški prevajalski slovar. *Mostovi 2001*, str. 109-112
- Vintar Špela. 1999. Računalniška orodja za prevajanje, *Mostovi 1999*, str. 47-54
- Željko Miran, Krstič Adriana. 2002. Web-based Trados Databases – An Alternative Approach, *Translation: New Ideas for a New Century (Proceedings of the XVI FIT World Congress)*, str. 303-308.
- Željko Miran. 2000. Pomnilniki prevodov v praksi, *Mostovi 2000*, str. 75-90
- Željko Miran. 2002. Spletni slovarji.
<http://www.gov.si/slovar.html>
- Željko Miran, Krstič Adriana. 2002. Evroterm – terminologija EU na internetu, *Zbornik posvetovanja Dnevi slovenske informatike 2002*, str. 408-412.