

Nataša Hirci

Nov pripomoček pri prevajalskem delu: računalniški korpusi

Povzetek

V prispevku avtorica opisuje nekaj možnosti za uporabo računalniških besedilnih korpusnih virov pri prevajalskem delu. Na kratko so predstavljene nekatere prednosti in težave pri rabi takih pripomočkov.

Abstract

The article gives an outline of some applications of computer corpora for translation, and a brief survey of the advantages and problems involved in using such translation tools.

1 Uvod

V informacijski dobi pri čedalje hitrejšem tehnološkem razvoju postajajo jezikovne tehnologije pomemben pripomoček pri prevajalskem delu. Pri reševanju prevajalskih težav dvojezični slovarji ponavadi ne zadostujejo, saj so kot vir podatkov iz resničnega sveta, ki jih potrebujemo pri prevajanju, včasih popolnoma neustrezni. Hartmann (1997: 292) pravi, da semantična abstrakcija, ki je sestavni del leksikalnega inventarja v slovarju, besede pravzaprav prikrajša za njihovo naravno sobesedilo; tako mora prevajalec nadomestiti pomanjkanje sobesedilnih informacij s svojo dvojezično diskurzno kompetenco, še posebej glede besedišča, ki je specifično za posamezno kulturo.

Vendar kompetenca posameznega prevajalca ni nujno dovolj, da bi zapolnili vse vrzeli, ki so nastale zaradi nezadostnega sobesedila in rabe jezika v dvojezičnih slovarjih. V idealnih razmerah bi morali imeti prevajalci dostop do velikih, sistematično urejenih zbirk zanesljivih podatkov, ki bi potrjevali uporabo semantično enakovrednih leksikalnih enot med jeziki glede na slog, zvrst in področje. To bi bila dodatna informacija k temu, kar je dostopno v slovarjih. Uporaba velikih besedilnih korpusov bi lahko precej olajšala tudi prevodoslovno delo, saj bi pokazala pragmatično rabo besed v enem ali več jezikih, s tem pa empirično podkrepila domneve o neki besedi oziroma besedni zvezi in prevodih te besede oziroma besedne zveze.

Obravnavanje resničnih besedil prispeva zelo dragocene podatke, zato je analiza besedil s pomočjo korpusa lahko v veliko pomoč; prevajalcu se ni treba več zanašati le na intuicijo o tem, katera beseda ali besedna zveza v maternem ali tujem jeziku je pravilna. Med mnogimi možnostmi, ki jih korpusni viri ponujajo prevajalcem, v članku omenjam le nekatere.

2 Pozitivne strani besedilnih korpusov in uporabnost teh korpusov

Biber in sodelavci (1998: 5, 8–9) pravijo, da imata tako imenovani korpusni pristop in analiza z besedilnimi korpusi več prednosti, saj gre za velike, načrtno zbrane baze avtentičnega gradiva v naravnem diskurzu, to pa omogoča empirično analizo pristnih in uporabljenih vzorcev jezikovne rabe.

McEnery in Wilson (1996: 90) sicer pravita, da sta prevladujoči področji uporabe besedilnih korpusov leksikologija in predvsem leksikografija; odkar so se pojavili različni tipi korpusov, ki so dostopni širšemu krogu ljudi, pa se njihova uporabnost širi na vsa jezikoslovna področja, tudi v prevodoslovje. Obsežne zbirke različnih tipov besedil, kjer najdemo primere resnične rabe besed in njihovega sobesedila, pa lahko pri delu koristno uporabljajo tudi prevajalci.

2.1 Tipologija besedilnih korpusov

Meyer in Mackintosh (1996: 260) korpus opredelita kot enovito, notranje strukturirano in standardno označeno zbirko besedil v elektronski¹ obliki, zbranih po izhodiščnih kriterijih glede na namen samega korpusa. Besedilni korpusi so zbrani tako, da kažejo stanje ali raznovrstnost katerega jezika in omogočajo temeljit vpogled v jezik na več ravneh in področjih, zelo uporabni pa so tudi v jezikoslovju.

Glede na funkcijo in uporabnost lahko korpusne vire delimo različno (prim. Baker, 1995; Erjavec, 1996/97; Hirci, 1999; Sinclair, 1991). S širjenjem računalniških korpusov se je pojavila potreba po razvrstitvi in ovrednotenju takih besedilnih zbirk. EAGLES, Expert Advisory Group for Language Engineering Standards, se ukvarja z opisom lastnosti, s katerimi lahko besedilno zbirko ovrednotimo, in z definiranjem zvrsti korpusov, ki jih je smiselno med seboj razlikovati. Glavne korpusne deli na:

- **referenčne** – ti pomenijo osnovno vrsto korpusa; uporabljajo se za jezikovne standarde (Erjavec, 1996/97) in naj bi predstavljali celovito podobo jezika in bili s tem izhodišče za temeljne jezikoslovne raziskave;
- **specializirane** – vsebujejo specifična besedila s posameznega področja;
- **korpusne govornih oziroma govorjenih besedil** – ti vsebujejo govorjena besedila oziroma transkripcijo govora;
- **spremljevalne** – to je dokaj nov tip korpusa. So dinamični, odsevajo dejansko stanje jezika, ker se tako kot jezik sam nenehno spreminjajo. Zastarela besedila sproti spravljajo v arhivih, nova besedila pa dodajajo;
- **vzorčne** – taki korpusi niso sestavljeni iz celotnih besedil, ampak iz fragmentov, ki so približno enako dolgi (fragmenti v Brownovem korpusu ameriške angleščine vsebujejo okrog 2000 besed);
- **prevajalske dvo- ali večjezične vzporedne** (sestojijo iz izvirkov in prevodov v enega ali več tujih jezikov) in **primerljive korpusne** (vsebujejo besedila, ki so primerljiva glede na žanrsko opredelitev, register, temo ipd.).

2.2 Možnosti rabe različnih tipov korpusov pri prevajalskem delu

Jezika se nenehno učimo, zato lahko prevajalci že s poznavanjem resnične rabe besed in tipa sobesedila, v katerem se pojavljajo, izboljšajo poznavanje, razumevanje in rabo tujega jezika, to pa lahko ustrezno uporabijo pri prevajanju. Pri prevajalskem delu so koristni **referenčni** in **specializirani** korpusi ter **dvo- ali večjezični vzporedni** in **primerljivi** korpusi.

Referenčni korpusi so enojezične zbirke besedil, ki pomenijo obsežen vir informacij o jeziku in njegovih lastnostih. Biti morajo dovolj veliki, da so reprezentativna slika relevantnih jezikovnih različic in tipičnega besedišča. Največji referenčni korpusi so korpusi angleškega jezika, na

¹ Poimenovanje v drugih vedah se razlikuje: korpus ni nujno zbirka besedil v elektronski obliki.

primer The Bank of English in BNC, British National Corpus, za slovenščino pa imamo za zdaj korpusa *Beseda* in *FIDA* ter nekatere manjše korpusse. *Beseda* je korpus leposlovnih besedil in nekaterih prevodov leposlovja v slovenski jezik, od septembra 2000 pa vsebuje tudi neleposlovna besedila (časopis Delo), *FIDA* je referenčni korpus sodobnih, predvsem neleposlovnih besedil (različni časopisi, revije, strokovna, naravoslovno-tehnična, humanistično-družbena besedila ipd.), čeprav vsebuje tudi manjši podkorpus leposlovnih besedil – približno 10 odstotkov (prim. Stabej, 1998).

Kot referenčni jezikovni viri so referenčni korpusi najkoristnejši – so velike zbirke besedil (korpus *FIDA* vsebuje 100 milijonov besed, *Beseda* sestoji iz 48 milijonov besed, The Bank of English pa iz 415 milijonov besed), ki kažejo sliko naravnega jezika in njegovih specifičnih značilnosti (npr. *kibbitzers* (Johns, 1998)). S poznavanjem ne le resnične rabe besed, ampak tudi tipa sobesedila, v katerem se pojavljajo, se prevajalci laže odločajo o pravilni rabi besede ali fraze v tujem jeziku.

Specializirani korpusi so taki kot referenčne enojezične zbirke besedil, vendar sestojijo iz besedil s specifičnega področja, s tem pa je tudi besedišče omejeno na to področje. Take korpusse lahko prav tako koristno uporabljamo pri prevajalskem delu, saj se pri prevajanju ponavadi ukvarjamo z besedili s posameznega področja, kjer najdemo specifično besedišče.

Pri prevajanju so lahko v pomoč *vzporedni korpusi* – primer slovensko-angleškega vzporednega korpusa je IJS-ELAN² – saj pokažejo specifične strukture v nekem jeziku ali nekatere prevodne zakonitosti, ali pa *primerljivi korpusi*, večjezični omogočajo primerjavo dveh ali več jezikov oziroma besedil in njihovih jezikovnih struktur, *enojezični* pa primerjavo prevodov s prostim tvorjenjem besedil v istem jeziku.

Vzporedni in primerljivi korpusi so za zdaj najprimernejši za prevajalsko delo, vendar pri prevajanju splošnih besedil koristijo tudi referenčni korpusi, pri delu s strokovnim jezikom pa predvsem specializirani korpusi. Večjezikovni korpusi omogočajo raziskave jezikovnih struktur v njihovem domačem okolju in osvetljujejo jezikovne vzorce v naravnem jeziku. Eden od glavnih prispevkov vzporednih besedilnih korpusov je, da omogočajo vpogled v modele, ki kažejo, kako se obnašajo poklicni prevajalci, ko rešujejo prevodna vprašanja, in h kakšnim rešitvam se zatekajo, hkrati pa tudi vpogled v modele zelene kakovosti.

V Sloveniji je za zdaj najbolj problematično, da nimamo reprezentativnih računalniških korpusov, ki bi jih lahko uporabljali za različne vrste raziskav. Za slovenski jezik za zdaj obstajajo: **ORW-SI**, slovenski prevod Orwelove knjige 1984, del korpusa **MULTEXT-East**; **ELAN** (še neoznačen korpus slovenskih in angleških besedil, ki obsega nekaj več kot milijon besed³); zbirka leposlovnih besedil Mirana Hladnika; **VAYNA** (Verbal Attacks on the Yugoslav National Army); **GORE** (korpus zapiskov s konference o gorah iz leta 1995); nekateri drugi manjši korpusi, ki niso prosto dostopni; **FIDA**, ki v času pisanja tega članka (še) ni javno dostopna, in **Beseda**. Pri nas tudi še ni dobro razvitih orodij za delo s korpusi, saj je dostop do tistih, ki obstajajo, omejen (obstaja prosto dostopen konkordančnik⁴ na internetu, vendar ga moramo zaradi nereprezentativnosti različnih besedilnih tipov uporabljati zelo pazljivo, sicer lahko dobimo zavajajoče podatke).

² IJS-ELAN je dober začetek, vendar vsebuje le omejeno število različnih besedilnih tipov. Več o tem korpusu v 3.1.

³ Zbrali Erjavec, Skubic in Vintar.

⁴ O konkordančnih programih glej 4.

3 Vloga prevajalskih korpusov

Že prej je bilo omenjeno, da je glavna vloga primerljivih korpusov v prevajalstvu identifikacija vzorcev, ki so specifični za prevedena besedila ne glede na izhodiščni oziroma ciljni jezik, glavna funkcija vzporednih korpusov pa je ugotavljanje prevodnih norm v specifičnih kontekstih in odkrivanje metod in rešitev, h katerim se zatekajo poklicni prevajalci.

Vendar prevajalci niso nujno tudi računalniški strokovnjaki, zato za iskanje natančnih informacij o prevodnih ustreznicih J1 in J2 potrebujejo do uporabnika prijazen program, ki omogoča enostaven dostop in uporabo, na primer konkordančne programe. Bolj problematičen je dostop do večjih računalniško berljivih zbirk, v katere je vključena tudi slovenščina.

3.1 Primer slovensko-angleškega vzporednega korpusa ELAN

Z dostopom do konkordanc in odstavkov v korpusu ELAN, ki je dvojezični korpus besedil v slovenskem in angleškem jeziku, se je stanje v prevajalstvu nekoliko spremenilo. Pri delu s tem korpusom opazujemo prevodne ustreznice na ravni povedi in konkordance posameznih besed ali besednih zvez v obeh delih korpusa, čeprav korpus še ni označen.

Slovensko-angleški vzporedni korpus ELAN je del evropskega korpusnega projekta ELAN. Sestoji iz približno milijona besed, ki so del petnajstih besedil. V glavnem so to terminološko bogata besedila (govori predsednika Kučana, uradni dokumenti službe slovenske vlade za evropske zadeve in navodila za Linux, katalog izdelkov Lekove domače lekarne): gradnja korpusa je eden prvih poskusov zbiranja in označevanja besedil, ki so vzporedna, zato korpus (zaradi omejenosti različnih tipov besedil) ni reprezentativen. Kljub temu ima vse bistvene lastnosti pravega vzporednega korpusa in je izredno pomemben korak v gradnji korpusov na Slovenskem (več o sestavi korpusa, označevanju in drugih značilnostih korpusa v Erjavec, 1999, in Vintar, 1999).

Pri vzporednih korpusih lahko vse naštetu ugotavljamo za dva ali celo več jezikov hkrati in iz dobljenih podatkov sklepamo na primer o pogostosti posamezne prevodne ustreznice v primerjavi z drugo prevodno ustreznico. Prikaz jezikovnih elementov je v dveh poravnanih oknih KWIC (KeyWord In Context). Vendar pa so vzporedni korpusi primerni ne le za večjezične, ampak tudi za enojezične raziskave. Tako na primer iskanje vseh pojavitev neke besede, katere prevod se ne pojavi v prevodu stavka, od koder je beseda, hitro pokaže na idiomatsko rabo te besede. Spodaj je v dvojezičnih oknih prikazan zadetek za 'acquis' v slovensko-angleškem delu korpusa ELAN (tako neko besedo lahko opazujemo tudi v primerjavi z angleškimi prevodi). Zadetki so prikazani s sobesedilom.

IJS-ELAN English part

Search for "acquis" as PARA

134 hits

structure, to incorporate the **acquis** into its national legislation and to

PzP še navaja, da se bo morala Slovenija lotiti vseh vprašanj, ki so opredeljena v Mnenju, zahteva uskladitev upravno administrativne strukture, vključitev pravnega reda EU v zakonodajo ter zagotovitev njegovega dejanskega izvajanja.

on the adoption of the Community **acquis** and on the priorities identified in

Pomoč iz sredstev PHARE je v prvi vrsti namenjena uveljavljanju pravnega reda EU in izvajanju prednostnih nalog iz PzP in DPSA.

preparations for the adoption of the **acquis** in the area of Economic Monetary

Predstavlja tudi priprave za prevzem pravnega reda Ekonomsko monetarne unije.

policy. Slovenia accepts the **acquis** for the seven chapters and requires a

Za vseh sedem področij Slovenija sprejema pravni red EU, prehodno obdobje smo zahtevali le na področju kultura in avdiovizualna politika, na treh izmed sedmih področij pa bo potrebno slovensko zakonodajo dodatno uskladiti s pravnim redom EU in okrepiti inštitucionalno ureditev (str. 4).

has to be brought into line with the **acquis** and the institutional framework needs

Za vseh sedem področij Slovenija sprejema pravni red EU, prehodno obdobje smo zahtevali le na področju kultura in avdiovizualna politika, na treh izmed sedmih področij pa bo potrebno slovensko zakonodajo dodatno uskladiti s pravnim redom EU in okrepiti inštitucionalno ureditev (str. 4).

legislation's compliance with the **acquis** was completed for the chapters on

Pregled usklajenosti slovenske zakonodaje s pravnim redom EU je bil v septembru opravljen na področjih zunanji odnosi (26. delovna skupina) in carinska unija (15. delovna skupina).

be harmonised with the community's **acquis**. INTEGRATION INTO THE EU The first

Prvi od šestih pregledov usklajenosti zakonodaje pa je potekal tudi na področju kmetijstvo (7. delovna skupina), ki je glede prilagoditve pravnemu redu EU med vsemi najobširnejše in verjetno najzahtevnejše.

of the Slovenian legislation with the **acquis** was carried out this spring. After

Na teh področjih je bil pregled usklajenosti slovenske zakonodaje s pravnim redom EU opravljen že spomladi.

(i) does Slovenia accept the **acquis** in the chapters concerned; (ii)

i) ali Slovenija sprejema pravni red EU za ta področja,

concerned in conformity with the **acquis**, or will there be a need to change

iii) ali je zakonodaja RS za ta področja usklajena s pravnim redom EU ali pa bodo potrebne spremembe slovenske zakonodaje ter morebitna ustanovitve novih inštitucij in upravnih organov za izvajanje pravnega reda EU.

Primer 1: Prikaz desetih od 134 zadetkov pri iskanju vzporednih konkordanc besede 'acquis' (obravnavan tudi na Forumu prevajalcev novembra 2000) v angleško-slovenskem delu korpusa IJS-ELAN⁵

Čeprav korpus ELAN ni bil narejen za potrebe prevajanja in prevodoslovja, lahko na podlagi so-besedila in konkordanc opazujemo nekatere jezikovne značilnosti in prevodno ustreznost vsaj na povedni ravni. Zaradi različnih razlogov (korpus ELAN vsebuje besedila, ki so omejena na po-samezna področja; besedila v njem niso oblikoslovno in skladenjsko označena; za zdaj ni javno dostopen)⁶ je težko govoriti o možnosti aplikacije tega korpusnega vira na prevodoslovje. Poleg tega je slovenski del korpusa del večjega evropskega projekta ELAN, ki je še v gradnji. Šele ko bo projekt končan, bomo z gotovostjo vedeli, kje in kako ga lahko uporabljamo, čeprav je že v takšni obliki, kot ga poznamo, uporaben pri prevajalskem delu.

⁵ Primerjaj <http://nl2.ijs.si/corpus/index.html>.

⁶ Kljub temu je mogoče dostop do konkordančnika prek interneta.

4 Delo s korpusi

Z uporabo računalniških tehnologij lahko hitro obdelamo veliko količino podatkov, vendar hkrati lahko opazujemo le tiste jezikovne pojave, ki jih je mogoče nekako formalizirati. Uporabnost korpusa je torej močno odvisna od števila in tipa dodatnih jezikovnih podatkov, ki jih korpus vsebuje. Med najpogostejše uporabljani metode pri korpusnem delu sta izdelava frekvenčnih seznamov (na njih lahko opazujemo pojavljanje posamezne besede ali besedne oblike in na podlagi statističnih podatkov za celoten korpus izdelamo frekvenčni seznam) ter iskanje konkordanc in kolokacij (prim. Sinclair, 1991). Za pregledovanje označenih besedilnih korpusov lahko uporabljamo več programov; s tako imenovanimi pregledovalniki lahko poiščemo želene dele korpusa. Najbolj znana oblika predstavitve informacij so konkordance, in sicer v obliki ključne besede s sobesedilom oziroma KWIC (angl. *Key Word In Context*). Konkordance pokažejo pojavite izbrane besede oziroma besedne zveze v korpusu, izpisane so na sredini, skupaj s sobesedilom. Naslednji primer kaže nekaj pojavitev slovenske besedne zveze 'pravni red' (za primerjavo z *acquis*).

Search for "pravni" "red" as KWIC

25 hits (limited to 500)

kuca: sožitja in sodelovanja. Načet je **pravni red** in okrnjena moralna in razvojna moč
kuca: družba, v kateri bo okrepljen **pravni red** in uveljavljen preglednejši pravni
ecmr: sedem področij Slovenija sprejema **pravni red** EU, prehodno obdobje smo zahtevali le
ecmr: i) ali Slovenija sprejema **pravni red** EU za to področje, ii) ali Slovenija
ecmr: utemeljitev. Slovenija sprejema **pravni red** EU za vseh sedem področij, ki so
ecmr: politike, pa jih formalno **pravni red** EU v tem področju ne ureja, bodo
ecmr: na tem področju v celoti prevzeli **pravni red** EU. Z namenom uspešnejšega
ecmr: spadajo v področje kulture, pa jih **pravni red** EU za to področje formalno ne ureja,
ecmr: razvojna pomoč. Slovenija sprejema **pravni red** EU za vseh pet področij drugega sklopa
ecmr: področja Slovenija v celoti sprejema **pravni red** EU, vendar zahteva prehodno obdobje

Primer 2: KWIC-izpis desetih od 25 zadetkov pri iskanju konkordanc besedne zveze 'pravni red' v slovenskem delu korpusa ELAN

Za primerjavo s slovenskimi zadetki lahko potem poiščemo konkordance v angleškem korpusu, na primer v CollinsCobuild Online⁷ (beseda se pojavi s krajšim sobesedilom, izpiše se 40 pojavitev besede, ki so naključno izbrane v 50-milijonskem korpusu, prosto dostopnem na internetu).

Koristno je tudi iskanje sobesedij ali kolokacij, saj te lahko pokažejo vezave besed z drugimi besedami, s skladenjskega in pomenskega stališča. Način pregledovanja je podoben kot pri KWIC, obstajajo pa še programi, ki samodejno izberejo kolokacije. Poznamo več programov, ki naredijo seznam vseh oblik besed v celotnem korpusu ali v delu korpusa, in sicer v **frekvenčnem** ali **abecednem zaporedju** (podrobnejši pregled v Sinclair, 1991).

Za prevajalce je za zdaj najdostopnejše iskanje konkordanc. WordSmith Tools je konkordančni program, ki ga je Mike Scott izdelal za OUP⁸. Ta program išče konkordance, seznime besed

⁷ Podrobnejša referenca v virih.

⁸ Oxford University Press.

(te lahko med seboj primerja), poišče ključne besede v besedilih in nato sestavi seznam najpogostejših ključnih besed, vsebuje pa tudi program za spreminjanje besedil ipd. ParaConc⁹ je prav tako večjezični konkordančnik, ki ga je programiral Barlow (1996), uporabljal pa ga je kot pripomoček pri poučevanju tujih jezikov. Ker je prosto dostopen na internetu, zelo razumljiv in preprost za uporabo, je zelo koristno orodje tudi za tiste, ki niso prav spretni pri delu z računalnikom. Potrebujemo le še reprezentativen besedilni korpus (danes je nekakšna osnovna meja za reprezentativnost približno milijon besed), besedila pa morajo biti poravnana (na primer s programom Win Align, ki ga ponuja TRADOS¹⁰, ali z Atrilovim DejaVujem ipd.). Ko imamo korpusni vir, lahko s ParaConcem iščemo zelene dele korpusa, posamezna besedila in prevode teh besedil, konkordance, kolokacije ipd.

4.1 Jezikovne posebnosti: težave pri strokovnih izrazih

Delo s korpusom omogoča izpeljevanje sklepov na podlagi resničnih primerov. Pri prevajanju v tuj jezik pogosto naletimo na problem ustreznosti posamezne besede ali besedne zveze v kakem komunikacijskem položaju. Za primer vzemimo *electric* in *electrical* – v angleščini je raba teh dveh besed problematična, saj se v različnih okoljih vežeta z različnimi besedami, vendar slovensko-angleški slovar kot prevedek za 'električen' navaja oboje, *electric* in *electrical*. Poglejmo, kako nerojenemu govorniku v taki dilemi pri prevajanju lahko pomaga korpus.

Search for "electric" as PARA

19 hits

ecmr: achieved despite the fact that some **electric** generators are being renovated. The

Visoko proizvodnjo so v hidroelektrarnah dosegli kljub dejstvu, da nekatere agregate dravskih elektrarn trenutno obnavljajo.

ecmr: still cut, resulting in a surplus of **electric** energy despite lower production and

Oddaja (izvoz) polovične proizvodnje Jedrske elektrarne Krško na Hrvaško je še vedno prekinjena, zato smo kljub znižanju proizvodnje in povečanju porabe imeli presežek električne energije, ki smo ga izvozili.

ecmr: to January 1999. Slovenia had an **electric** energy surplus of 774 GWh due to its

Zaradi manjšega izvoza elektrike na Hrvaško smo imeli v zadnjem dvanajstmesečnem obdobju do januarja 1999 774 GWh presežkov električne energije.

ekol: etc). 47 Simultaneous production of **electric** energy and heat. 48 Hydro-electric

47 Sočasne proizvodnje električne energije in toplote.

ekol: 34% of the total consumption of **electric** power. 50 According to the

49 NE Krško zagotavlja ca. 34 % sedanje porabe elektrike v Sloveniji.

ekol: refrigerating or freezing equipment, **electric** or other; heat pumps other than air

Hladilniki, zmrzovalniki in druge naprave za hlajenje ali zmrzovanje, električni in drugi;

ekol: in their enclosures; audio-frequency **electric** amplifiers; electric sound amplifier

avdio-frekvenčni električni ojačevalniki;

anx2: audio-frequency electric amplifiers; **electric** sound amplifier sets – where the value

⁹ Več o tem na spletnih straneh: <http://www.ruf.rice.edu/~barlow/norway.html>.

¹⁰ TRADOS je nemško podjetje, ki se intenzivno ukvarja s programi, ki nam pomagajo pri računalniško podprtem prevajanju (med drugim tudi Translator's Workbench in MultiTerm).

kompletne električne enote za ojačevanje zvoka

anx2: of heading No 8535 or 8536. for **electric** control or the distribution of

Table, plošče, pulti, mize, omare in druge osnove, opremljene z dvema ali več aparati iz tar. št. 8535 ali 8536, za električno krmiljenje ali razdeljevanje električnega toka, vštevši tiste z vdelanimi instrumenti ali aparati iz 90. poglavja, razen komutacijskih aparatov iz tar. št. 8517

anx2: coaxial cable) and other insulated **electric** conductors, whether or not fitted

Izolirana žica (vštevši lakirano in anodizirano žico), kabli (vštevši koaksialne kable) in drugi izolirani električni vodniki, s konektorjem ali brez;

Primer 3: Prikaz prvih desetih primerov za *electric* in prevodi v slovenščino v korpusu IJS-ELAN

Search for "electrical" as PARA

23 hits

ecmr: by 2.7 %, and the manufacture of **electrical** and optical equipment (real growth of

Izmed največjih izvoznikov je izvoz najbolj pospešil proizvodnjo strojev in naprav, ki je po fizičnih pokazateljih porasla za 2,7 %, ter proizvodnjo električne in optične opreme (realni porast za 7,2 %).

ecmr: machinery and equipment, paper, **electrical** and optical equipment. For some

Največji delež v izvozu industrijskih proizvodov predstavljajo kemični izdelki s skoraj 40 %, sledijo pa proizvodi kovinske industrije, stroji in naprave, papir, električna in optična oprema.

ecmr: products (up by 11.7 %), **electrical** machines and devices (17.4 %), road

Polovico realne rasti izvoza so prispevali medicinski in farmacevtski proizvodi (11,7-odstotna rast), električni stroji in naprave (17,4 %), cestna vozila (13,1 %) ter pohištvo in deli (14,1 %).

ecmr: have been recorded in chemicals, **electrical** equipment, metal products and wood

Podobno je tudi v industrijskih panogah, pri čemer po najvišjih koeficientih izstopajo predvsem kemična, elektro, kovinska in lesna industrija, torej panoge z visoko tehnologijo oziroma večjo kapitalno intenzivnostjo.

ecmr: but it was only in the **electrical** and optical equipment, and rubber

V petih panogah so rezultati leta 1997 tesno povezani s spremembami proizvodnega programa, vendar le proizvodnja električne in optične opreme ter gumarska industrija v nadpovprečni rasti dodane vrednosti na zaposlenega.

ecmr: to CEFTA markets lies in the car, **electrical**, metal, leather, wood and machinery

Prav tako je očitno, da imajo največji potencial za povečanje izvoza na trg držav CEFTA predvsem avtomobilska, elektro, kovinska, usnjarska, lesna in strojna industrija.

ecmr: their upper structural limits. The **electrical**, machinery, leather and textiles

Največji potencial za povečanje uvoza iz držav CEFTA je v elektro, strojni, usnjarski in tekstilni industriji.

ecmr: were recorded in the manufacture of **electrical** and optical equipment, manufacture of

Še precej višje ravni obsega zalog na enoto proizvodnje so ustvarili proizvajalci električne in optične opreme, izdelkov iz gume in plastičnih mas ter usnja in usnjenih izdelkov.

ekol: emissions are mostly the result of **electrical** power activity (81%); NOx emissions

K emisiji SO₂ največ prispeva elektroenergetika (81 %), k emisiji NO_x pa promet (66 %).
 ekol: shares of CO₂ emissions are held by **electrical** power activity 17 (35%) and traffic
Največji delež pri emisijah CO₂ imata elektroenergetika 17 (35 %) in promet (32 %).

Primer 4: Prikaz prvih desetih primerov za *electrical* in prevodi v slovenščino v korpusu IJS-ELAN

Vendar je zadetke, ki smo jih dobili iz dvojezičnega dela korpusa ELAN, smiselno primerjati tudi z rezultati iskanja po enojezičnem angleškem referenčnem korpusu, na primer v The Bank of English na internetu¹¹, saj tako hitro ugotovimo, s katerimi besedami se ta dva izraza najpogosteje vežeta. ELAN namreč zaradi svoje nereprezentativnosti ni najkoristnejši za strokovne prevajalce, saj bi ti pravzaprav potrebovali dvojezični korpus s posameznega (strokovnega) področja. Prav zaradi pomanjkanja praktično uporabnih dvojezičnih korpusov, kjer je vključena slovenščina, si moramo pomagati z enojezičnimi.

[p] Battery-operated or **electric** breast pumps are the fastest and
 of it and what will they do with it **electric** cars speed silently past solar-
 inmate is scheduled to die in the **electric** chair at 11 p.m. Eastern time
 of the newly-privatised water and **electric** companies have bought shares worth
 p] Now the spotlight is on General **Electric** Company (GEC), where last week
 these days it is usually done by an **electric** compressor. With the handpump, the
 problems... [p] Easy hand-held **electric** control unit enables smooth four-
 ELECTRIC HEATING [h] [p] Direct **electric** heating is efficient, flexible,
 of shiny stones glimmering like **electric** lights. He touched her cheek. `Can
 will be puzzled not so much at his **electric** mix of what he terms R & B, rock
 to buildings; and [p] 5.2 provide **electric** power needed by us in installing,
 a whole two-and-one-half times. The **electric** power industry, which had not even
 that electromagnetic fields from **electric** power lines could increase the
 of about 660,000 kilowatts. (An **electric** range and oven together use about
 every time. [p] Belling 170M [p] **Electric** Single Oven [p] A stunning white

Primer 5: Konkordance za *electric*¹²

the bath and/or shower? [p] [p] No **electrical** appliances in the bathroom! [p]
 and hand blenders and everyday **electrical** appliances including irons,
 cramped, with unsafe windows and **electrical** appliances, and not enough
 THOMPSON: Sounds of domestic **electrical** appliances which can be seen and
 replacement of old and inefficient **electrical** appliances. [p] It says 69 of
 because it changes light into **electrical** energy. But the dictionary leaves
 lean tissue is more conductive of **electrical** energy than fat. [p] Another way
 of Essex, a 2 year MSc course in **Electrical** Engineering is offered which
 was appointed to a lectureship in **Electrical** Engineering at the same time that

¹¹ CollinsCoubuild Online.

¹² Zaradi omejenosti s prostorom so v prispevku predstavljeni le delni rezultati iskanja (omejitev na največ 15 primerov).

published by the Institution of **Electrical** Engineers stated that by the turn years ago, joining from Volex, the **electrical** group, where he had also been 000 a year chairman of high street **electrical** group Dixons, Roger Boissier, Spark ignition - does not need an **electrical** supply [p] Variable heat settings [h] The Final Stage [h] [p] The **electrical** work has just been completed.

Ideal Home always recommends that **electrical** work is carried out by a

Primer 6: Konkordance za *electrical*

Že na prvi pogled ugotovimo, da so najpogostejše rabljeni izrazi z *electric* accumulator, amplifier, cars, conductor, power in nekateri drugi, z *electrical* pa appliances, equipment, engineering, machinery, systems itd. Podobno velja za leksiko-gramatikalno vezljivost nekaterih drugih problematičnih izrazov, na primer: kdaj uporabiti *atomic*, kdaj *nuclear*, kdaj *each* in kdaj *every* ali pa kdaj *at the beginning* in kdaj *in the beginning* (prim. Pisanski in Hirci, 1999).

Včasih pri zadetkih, dobljenih v korpusnih virih, takoj vidimo, kdaj in v kakšnih položajih se posamezne besede pojavljajo oziroma s kakšnimi členi ali predlogi se leksemi vežejo, to pa je smiselno upoštevati pri prevajalskem delu.

Poglejmo še primer za slovenski jezik: rezultate iskanja pogostosti izrazov *plošča*, *CD-plošča*, *plošča CD*, *plošča DVD* in *zgoščenka*. Vprašanje o rabi teh izrazov se je pred kratkim pojavilo na Forumu prevajalcev. Rezultati iskanja v korpusu Beseda so na prvi pogled presenetljivi: plošča se pojavi 385-krat, vendar v vseh mogočih kombinacijah (npr. glasbena, spominska, prva, nova ipd.), zgoščenka 87-krat, CD-plošča se v korpusu pojavi le dvakrat, plošča DVD enkrat, zadetkov za plošča CD pa sploh ni. Upoštevati moramo, da korpus Beseda vsebuje literarna besedila in besedila iz časopisa Delo. Vendar ugotovitve vseeno niso zanemarljive, saj je prav jezik v časopisu tisti, ki naj bi odseval dejansko stanje jezika. Navedena sta še dva primera rabe iz časopisa Dela.

DELO, 13. januarja 1998, sentence 1669 in context:

A, oprostite, zakaj ne bi potemtakem rekli kar plošča, tako kot smo prej? Recimo CD-plošča ali CD-ROM-plošča? Hm, tale zgoščenka je vsak dan bolj videti kot nepotrebna izmišljaja. Ljudstvo v pogovoru tako uporablja kar tlačenka, a je ta plemenita beseda žal pomensko že zasedena z manj plemenitim izdelkom.

DELO, 24. februarja 1999, sentence 1072 in context:

Odveč so tudi računalniška plošča in skrajnopruristične skovanke v slogu zgoščenka in plošček. Kaj pa CD plošča (compact disk), kompaktna, digitalna avdio plošča itd.? Strokovno korrektno ji sicer rečemo laserska plošča (kar pomeni digitalna avdio plošča, s katere nam informacijo o glasbenem zapisu ne odčitava diamantna gramofonska igla, pač pa laserski žarek), toda za vsakdanjo časopisno, pisarno, oglasno in drugo podobno rabo bo enako dobro, če bomo govorili in pisali samo – plošča.

Primer 7: Dva primera iz korpusa Beseda, ki vsebujeta iskane besedne zveze

Rezultate takega iskanja moramo uporabljati zelo previdno, saj bi iskanje po zbirki besedil iz računalniških revij gotovo dalo drugačne podatke. Spet smo pred problemom, s katerim se spopri-

jemam že pri slovarjih: v Sloveniji je malo dobrih in reprezentativnih dvojezičnih slovarjev, pa tudi enojezični Slovar slovenskega knjižnega jezika ne kaže več dejanskega stanja v jeziku. Pri upoštevanju rezultatov iskanja po korpusnih virih se je tega treba nenehno zavedati.

4.2 Pripomoček za dopolnjevanje jezikovne kompetence v tujem jeziku

Jezikovna kompetenca v tujem jeziku je omejena, prav tako je omejen vsak pripomoček. Korpus je dragocen vir podatkov o jezikovni rabi, na katerega se lahko opremo. Pri tujem jeziku pogosto naletimo na primere, kjer vezljivosti izraza ne moremo preveriti v nobenem od drugih pripomočkov, ker je ti enostavno ne navajajo. Podatki, ki jih dobimo iz korpusa, tedaj dopolnjujejo našo jezikovno kompetenco. Če skušamo na primer uporabiti različne besedne zveze, ki jih slovarji navajajo kot sopomenke, v praksi naletimo na težave, saj jih v nekaterih komunikacijskih položajih ne moremo uporabljati popolnoma izmenljivo. Pri delu s študenti, ki niso rojeni govornici angleškega jezika, se je Johns (1998) med drugim ukvarjal tudi z glagoli povzročanja. V angleščini obstajajo 'cause', 'result in', 'bring about', 'lead to' in njihove različice v pasivu.

room for **environmental destruction** *caused by* war. Please help us stop it.
cent of **deaths from lung cancer** are *caused by* smoking. The younger people
Immune Deficiency Syndrome. It is *caused by* a virus (called HIV) which
to adult teeth than **decay**. It's *caused by* bacteria that form plaque and
and the **stress and bitterness** *caused by* divorce. A single mother by
Paulo says **mounting labour unrest** *caused by* President Collor de Mello's
enterprises, or subsidise the **damage** *caused by* the **incompetence** of their
only source of **East European loss** *caused by* the crisis. There is the loss
to resolve a **political crisis** *caused by* the action of truck and taxi
third place. The by-election was *caused by* the **death** of the sitting MP, Mr
given on any **damage or casualties** *caused by* the **reported attack**. This is
of experiments in the paranormal *caused by* different states of belief. The
[f] This is a condition which is *caused by* a congenital weakness in a blood
to The 61. The **earlier leak** had been *caused by* **unintended carelessness**; the new
find that **much of their stress** was *caused by* keeping their problem a secret.

Primer 8: Nekaj primerov konkordanc za zvezo *caused by*

Že na prvi pogled ugotovimo, da se izraz *caused by* praviloma uporablja v zvezi z nečim negativnim, in to na levi in desni strani same zveze (na levi: environmental destruction, decay, the stress and bitterness, a political crisis, damage or casualties, leak, stress, na desni pa: war, smoking, virus, incompetence, death, unintended carelessness). Tega se mora prevajalec pri delu zavedati, saj lahko nenamerno doda negativno konotacijo prevodu, ki bi sicer moral biti nevtralne narave. Za primerjavo si oglejmo še nekaj drugih primerov rabe glagolov povzročanja, na primer *lead to*.

[p] But drinking too much can also *lead to* crime. Drink-driving is one
on the lookout for these tricks can *lead to* self-delusion. here at La Crete
s ceremony at Joseph's tomb might *lead to* tension in the West Bank's most

to such events. Feelings of panic may *lead to a person feeling anxious* most of mothers or grandmothers. This can *lead to inter-generational stress* with an by itself for "peace of mind" may *lead to harmful self-absorption*. There plans would not work and would *lead to more murders*. They would fetter Swiss population and a merger would *lead to mass layoffs*. [p] CS Holding the month they spent apart would *lead to a reconciliation*. [p] But Fergie that humanitarian aid could *lead to an improvement* in relations and Pope said he hoped the meeting would *lead to a just and peaceful solution* of exchange rate mechanism could *lead to significantly higher* of the beating heart, of which could *lead to better understanding* of avoid job dissatisfaction they do not *lead to job satisfaction*. For this the dreaded Sonderaktionen do not *lead to reprisals* against the innocent and followed the moonlit path that *led to* the clearing beside the river. actively interventionist, and this *led to* expansionism and finally to reduced. This had inevitably *led to* some tension in the marriage and both now and in the future, has *led to* considerable work on defining and traditions of Solomonic authorship *led to* canonical status: association butane straight down the throat has *led to* deaths through freezing of the as they crossed to the steps which *led to* a tiled patio at the front of the the economic needs of the Cold War *led to* his release in 1950 from gradual recovery in the economy has *led to further improvements* in both an understatement. [p] [h] Muslims *led to* slaughter in torture pits

Primer 9: Nekaj primerov konkordanc za zvezi *lead to/led to*

Rezultati iskanja so presenetljivi, saj ima nekaj več kot pol primerov negativno konotacijo (crime, self-delusion, tension, a person feeling anxious, more murders), preostali pa imajo nevtralno ali pozitivno konotacijo (a reconciliation, an improvement, a just and peaceful solution, better understanding). Tudi pri iskanju konkordanc za zvezo *led to* se zgodba ponovi – večina primerov je z negativno konotacijo (izjema je kombinacija 'led to further improvements'). Zanimiva je Johnsova ugotovitev (1998), da *lead to* ponavadi najdemo v aktivu in ne v pasivu. Prav taki primeri kažejo, kako lahko korpusne vire koristno uporabimo tudi pri prevajalskem delu.

5 Težave pri uporabi korpusov

Pri uporabi primerov iz korpusov obstaja možnost, da naletimo na napake ali primere nereprezentativne rabe (Sinclair, 1991). Ne moremo pričakovati, da bo vse, kar najdemo v korpusu, samodejno pravilno. Korpus pogosto vsebuje nepregledana besedila, v teh pa najdemo tudi čudne primere, tipkarske napake, včasih kje manjka kakšna beseda itd. Toda če se pri delu tega zavedamo, ni pričakovati večjih težav.

Posebna težava pri uporabi korpusov je potrebna oprema, poleg tega je težaven dostop do njih, saj so le nekateri dostopni brezplačno: med angleškimi je mogoč delen dostop do The Bank of English ali do predstavitvene različice BNC prek interneta, za slovenski jezik pa imamo na voljo konkordančnik IJS. Za slovenščino je položaj trenutno še zmeraj precej omejen: korpusi so premajhni, da bi bili reprezentativna slika slovenskega jezika, ali pa do njih nimamo prostega dostopa.

6 Sklep

V prispevku so bile predstavljene nekatere možnosti za rabo računalniških besedilnih korpusov pri prevajanju, ki jim je sledil pregled nekaterih načinov uporabe glede na tip besedilnega korpusa. Nato so bile na kratko predstavljene nekatere prednosti in težave pri rabi jezikovnih korpusov pri prevajalskem delu.

Ker slovenskih korpusnih virov nimamo veliko, je morda smiselno razmišljati še v drugo smer: prevajalci se pogosto specializirajo za prevajanje strokovnih besedil s posameznega področja. Če imajo na razpolago veliko takih besedil (izvirnikov in prevodov), bi bilo smiselno, da si zgradijo svoj korpus besedil, za iskanje po njem pa lahko uporabijo enega od že znanih iskalnikov ali konkordančnikov (npr. WordSmith Tools).

Ob številnih nastajajočih korpusih in teh, ki so že na voljo, se pojavlja vprašanje, kdo jih bo uporabljal. Izdelava korpusov, še posebej široko dostopnih, je smiselna samo, če jih kasneje kdo uporablja. Na Oddelku za prevajanje in tolmačenje Filozofske fakultete v Ljubljani bodoče prevajalce že pripravljamo na delo s korpusi, verjamem pa, da tak način dela lahko koristi tudi izkušenim prevajalcem. Prevajalce lahko štejemo za močno skupino potencialnih uporabnikov, vendar bo priljubljenost dela s korpusnimi viri zrasla šele, ko bodo ti večji in besedilno bolj raznoliki, orodja za delo z njimi pa lažje dostopna.

7 Viri

- Baker, M. 1995. Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *Target*, 7/2, 223–243.
- Biber, D., Conrad, S., in Reppen, R. 1998. *Corpus Linguistics*. Cambridge: Cambridge University Press.
- BNC, British National Corpus: <http://info.ox.ac.uk/bnc>
- Erjavec, T. 1996/97. Računalniške zbirke besedil. *Jezik in slovstvo*, 42/2–3, 81–96.
- Erjavec, T. 1999. Making the ELAN Slovene/English Corpus. *Language Technologies – Multilingual Aspects* (ur. Vintar, Š.). Zbornik konference 32nd Annual Meeting of the Societas Linguistica Europaea, 8.–11. julij, 1999, Ljubljana, 23–30.
- EAGLES: Expert Advisory Group on Language Engineering Standards: <http://www.ilc.pi.cnr.it/EAGLES96/>.
- Gorjanc, V. 1999. Korpusi v jezikoslovju in korpus slovenskega jezika FIDA. *Zbornik predavanj 35. seminarja slovenskega jezika, literature in kulture*. Ljubljana, 28. 6.–17. 7. 1999, 47–59.
- Hartmann, R. R. K. 1997. From Contrastive Textology to Parallel corpora: Theory and Applications. *Trends in Linguistics. Studies and Monographs* 101. *Language, History and Linguistic Modelling. A Festschrift for Jacek Fisiak on his 60th Birthday Vol II*, (ur. Hickey, R. in Puppel, S.), Berlin: Mouton de Gruyter, 1973–1987.
- Hirci, N. 1999. Pogled v prihodnost: vloga prevodoslovnih besedilnih korpusov v Sloveniji. *Uporabno jezikoslovje* (ur. Kovačič, I.), 7–8, 135–155.
- ICAME: International Computer Archive of English: <http://www.hd.uib.no>
- Johns, T. 1998. Kibbitzing One-to-One Consultations in English for Academic Purposes. Zbornik prispevkov na konferenci *New Challenges in Teaching ESP in Slovenia* (ur. Čebren, N.), Portorož, junij 1998, 67–75. Ali: <http://sun1.bham.ac.uk/johnstf/timeap3.html>
- Konkordančnik IJS-ELAN: <http://nl2.ijs.si/corpus/index.html>
- Korpus Beseda: <http://bos.zrc-sazu.si/beseda.html>

Korpus FIDA: <http://www.fida.net>

Meyer, I., Mackintosh, K. 1996. The Corpus from a Terminographer's Viewpoint. *International Journal of Corpus Linguistics* 1/2, 257–285.

McEnery, T. in Wilson, A. 1996. *Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Pisanski, A. in Hirci, N. 2000. Vloga besedilnih korpusov pri izboljševanju jezikovne kompetence. Zbornik konference *Informacijska družba – Jezikovne tehnologije za slovenski jezik* (ur. Gros, J. in Erjavec, T.), Ljubljana, 17.–19. oktober 2000, 64–69.

Sinclair, J. H. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Stabej, M. 1998. Besedilnovrstna sestava korpusa FIDA. *Uporabno jezikoslovje: Jezikovne tehnologije*. Tematska številka referatov na II. kongresu »Jezik za danes in jutri« (ur. Kačič, Z.), Ljubljana, 8.–10. oktober 1998, 96–106.

The Bank of English/ CobuildDirect OnLine: <http://www.titania.cobuild.collins.co.uk/>

Vintar, Š. 1999. A Lexical Analysis of the IJS-ELAN Slovene-English Parallel Corpus. *Language Technologies – Multilingual Aspects* (ur. Vintar, Š.). Zbornik konference 32nd Annual Meeting of the Societas Linguistica Europaea, 8.–11. julij, 1999, Ljubljana, 63–70.

Novi člani DZTPS, sprejeti od aprila 2000 do februarja 2001

Mojca Arh (angleščina, hrvaščina)	Aleš Mustar (romunščina)
Helena Biffio Zorko (ruščina, francoščina)	Taša Okrajšek Kurnik (nemščina, angleščina)
Emil Blagus (nemščina, hrvaščina)	Davor Okrožnik (nemščina, ruščina)
Indira Činč (angleščina)	Snežana Pak Dvorak (angleščina)
Primož Debenjak (nemščina)	Sabina Pinter (nemščina, angleščina)
Simona Dečko (nemščina, angleščina)	Sigrid Pogačar Pirnat (italijanščina, hrvaščina)
Cilka Demšar (angleščina)	Andrej Primc (nemščina, angleščina)
Jasna Djordjević (angleščina, srbsščina)	Bojan Rebolj (angleščina, nemščina)
Rok Dobnikar (nemščina, angleščina)	Aleksandra Rekar (srbsščina, angleščina)
Peter Gabrovšek (nemščina)	Vesna Skvarča (nemščina)
Andreja Galinec (angleščina)	Mitja Slane (nemščina)
Nada Holc (nemščina)	Joanna Slawinska (poljščina)
Špela Hrastnik (francoščina, španščina)	Marina Spanring Poredoš (ruščina)
Klemen Jelinčič (hebrejščina)	Špela Urbas (nemščina, angleščina)
Vincenc Juteršek (angleščina)	Andreja Vide Hladnik (angleščina, nemščina)
Vital Klabus (nemščina, francoščina, angleščina)	Maja Visenjak Limon (angleščina)
Aleksandra Kraner (nemščina, francoščina)	Maja Vitežnik (angleščina, italijanščina)
Aleksandra Kuhar (srbsščina)	Nena Zeilhofer Bruketa (francoščina)
David Limon (angleščina)	Bojana Zgonec (nemščina, hrvaščina, francoščina)
Jerneja Lipičnik Matijašević (angleščina, hrvaščina)	Irena Žakelj (nemščina, angleščina)
Amalija Maček Mergole (nemščina, španščina)	Marina Barbara Žlender (francoščina, španščina)
Anželika Medved - Oniško (ruščina, nemščina)	Katja Žumer (angleščina, nemščina, italijanščina)
Evgenij Meleščenko (ruščina, ukrajinščina)	
Tomaš Mladen (hrvaščina)	