

KORPUSI V PREVODOSLOVJU

Nataša Hirci

Oddelek za prevajanje in tolmačenje

Filozofska fakulteta

Aškerčeva 2

Ljubljana

tel: 061/121 32 40

e-pošta: natasa.hirci@guest.arnes.si

POVZETEK

Tehnologija postaja neizogiben del teoretičnega in uporabnega jezikoslovja. Članek predstavlja tipologijo računalniško berljivih besedilnih korpusov in njihovo uporabnost v jezikoslovju, še posebej v prevodoslovju. Predstavljena so tudi orodja za delo z njimi. Prevodoslovje se je znašlo pred vprašanjem, ali je besedilne korpuse mogoče koristno uporabljati tudi pri prevajanju in kako. Vse več prevodoslovcev se zavzema, da bi jih vpeljali v prevodoslovne študije, saj nam lahko pomagajo ugotavljati prevodno ustreznost, uporabni pa so tudi kot pripomoček pri poučevanju prevajanja.

ABSTRACT

Technology is becoming an inevitable part of theoretical and applied linguistics. This article presents a typology of computer-readable textual corpora and their applicability to linguistics, and in particular in translation. Tools used to search corpora are also presented. Translation studies are faced with the question of whether it is possible to successfully apply corpora to translation and how to do so. More and more translation experts and theoreticians favour the introduction of textual corpora in translation studies as they can be of great use in the search for translation equivalence. They can also be useful as an aid in translator training.

1 OPIS BESEDILNIH KORPUSOV

Korpusna lingvistika se vse bolj uveljavlja, saj omogoča kvantitativne študije (pogostnost besed, kolokacije, itd), obenem pa daje možnost za empirične študije sintaktičnih in leksikalnih pojavov.

Meyer, Mackintosh [1] pravi, da je korpus enovita, notranje strukturirana in standardno označena zbirka besedil v elektronski obliki, zbranih po izhodiščnih kriterijih glede na namen samega korpusa (povzeto po Gorjanc [2]). Take zbirke besedil postajajo danes vse bolj nepogrešljiv vir podatkov v jezikoslovju.

Besedilni korpusi so zbrani tako, da karakterizirajo stanje ali raznovrstnost nekega jezika. Uporabljamo jih kot "osnovo, na kateri gradimo opise jezika, ali pa kot sredstvo ugotavljanja različnih hipotez o jeziku" (Erjavec [3]). V zadnjem času so taki korpusi že standardno

shranjeni na računalnikih: ti nam izredno hitro in kompleksno pomagajo obvladovati ogromne količine besedil in jih tudi učinkoviteje izkoriščati.

Snovalci Fide [4], prvega referenčnega korpusa v slovenskem jeziku, pravijo, da je osnovni namen korpusov omogočanje temeljitega vpogleda v jezik na najrazličnejših ravneh in področjih; tako jih lahko s pridom uporabljamo v jezikoslovju, v humanističnih in družboslovnih vedah in celo v informatiki in matematiki.

2 NASTANEK RAČUNALNIŠKIH KORPUSOV

Obdelave računalniških korpusov so bile aktualne in priljubljene že v petdesetih in šestdesetih letih, nato pa so zaradi pretvorbene-tvorbene slovnice Noama Chomskega in njegovih vplivnih teorij doživele skoraj popolni zaton. Vendar pa se je v osemdesetih zaradi hitrega razvoja najrazličnejših tehnologij vnovič pojavilo zanimanje za računalniško obdelavo in empirične raziskave velikih količin besedil. Tehnološki napredek v računalništvu je vplival na bliskovit dvig količine in kakovosti računalniških korpusov. Največ raznovrstnih korpusov je za angleški jezik. Samo za primerjavo rasti lahko navedem korpus Lancaster-Oslo/Bergen, izdelan proti koncu sedemdesetih, ki je vseboval milijon besed in British National Corpus iz leta 1994, ki ima okrog sto milijonov besed. V zadnjih letih se je povečalo zanimanje za takšne jezikovne vire v Evropski uniji, ki tudi finančno podpira izdelavo besedilnih korpusov (različni projekti, npr. Pendant, Intersect, Interlingua [5], itd.).

V nadaljevanju bom na kratko predstavila različne tipe korpusov, natančneje pa obdelala primerljivi in vzporedni tip besedilnih zbirk, ker so prav te najbolj neposredno povezane s prevodoslovjem.

2.1 Tipologija korpusov

S širjenjem računalniških korpusov se je pojavila tudi potreba po razvrstitvi in ovrednotenju besedilnih zbirk. EAGLES [6], Expert Advisory Group for Language Engineering, se ukvarja z opisom karakteristik, s katerimi lahko neko besedilno zbirko ovrednotimo, in z definiranjem zvrsti korpusov, ki jih je smiselno med seboj razlikovati.

Po tipologiji originalnega modela EAGLES so glavne značilnosti nekega korpusa sledeče:

1. **velikost:** korpus naj vsebuje takšno količino podatkov, kot mu to dopuščata najnovejša tehnologija in finančna podpora različnih projektov;
2. **aventičnost:** pomembni so kriteriji, po katerih so besedila izbrana in potem uvrščena v korpus;
3. **kakovost** izdelave in klasifikacija v različne tipe besedil, poravnava besedil, ustrezna anotacija, itd.;
4. **enostavnost** zapisa besedilnega korpusa (SGML in TEI/CES);
5. **reprezentativnost;**
6. in njegova **dokumentiranost.** (Več o zapisu in označevanju korpusov v Erjavec [3] in EAGLES [6].)

Za doseganje reprezentativnosti so nujno potrebni že vnaprej določeni parametri, po katerih je besedilni korpus sestavljen. Tako so besedila vključena v korpus glede na jezikovno zvrst, besedilno vrsto, žanrsko pripadnost, medij, v katerem se je neko besedilo pojavilo, ipd. (Gorjanc [2]). Korpus lahko zajema umetnostna (poezijo, prozo, dramatik) in neumetnostna besedila (velikokrat so to članki dnevnega in periodnega tiska). Korpusi so zgrajeni po različnih kriterijih, tako pri projektu Pendant [5] med pomembne zunajjezikovne dejavnike štejejo medij, stil, žanr, temo, datum publikacije, jezik in jezikovne povezave.

V naslednjih dveh poglavjih bom na kratko predstavila različne tipe besedilnih korpusov in njihovo uporabnost.

2.2 Vrste besedilnih korpusov

EAGLES deli glavne korpuse na:

- **referenčne:** ti predstavljajo osnovno vrsto korpusa in pomemben vir informacij o jeziku in njegovih značilnostih. Tak korpus mora biti dovolj velik, če želimo z njim pokazati razlike v rabi jezika in tipične značilnosti besedišča. Uporablja se predvsem kot osnova za pisanje slovnice, slovarjev, tezavrov in drugih jezikovnih referenčnih materialov. Reprezentativna primerka referenčnega korpusa sta The Bank of English [7], ki je do julija 1998 vseboval že okrog 326 milijonov besed, in CNC [8], Češki nacionalni korpus. Tudi v Sloveniji je delo s korpusi pritegnilo skupino ljudi, ki sestavljajo prvi slovenski referenčni besedilni korpus - Fida [4]. Gre za projekt, v katerem sodelujejo Filozofska fakulteta Univerze v Ljubljani, Institut Jožef Stefan, založba DZS in podjetje Amebis. Kot cilj projekta so opredelili "...oblikovanje obsežnega referenčnega besedilnega korpusa slovenščine, ki bo primerljiv z že obstoječimi korpusi drugih evropskih jezikov" [4];
- **govorjene oz. govorne:** ti za razliko od pisnih besedil vsebujejo govor oziroma transkripcijo govora. V večjem obsegu se pojavljajo šele v zadnjem času, ker so zanimivi za avtomatsko procesiranje govora, tj. elektronsko analizo in sintezo govora;
- **spremljevalne:** ti so dokaj nov tip korpusa. Za razliko od drugih so dinamični, saj se spreminjajo. Odražajo

dejansko stanje jezika, ker se tako kot jezik sam nenehno spreminjajo. Zastarela besedila sproti spravljajo v arhivih, nova besedila, dostopna v računalniško berljivih oblikah, pa dodajajo;

- **vzorčne:** taki korpusi niso sestavljeni iz celotnih besedil, ampak iz fragmentov, ki so približno enako dolgi (t.j. fragmenti v Brownovem korpusu ameriške angleščine vsebujejo okrog 2000 besed);
- **primerljive in**
- **vzporedne.**

Več o zadnjih dveh tipih, ki sta najpomembnejša za prevodoslovje, v tretjem poglavju. Najprej pa oris uporabnosti korpusov.

2.3 Uporabnost besedilnih korpusov

McEnery in Wilson [9] 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.

Lestvica torej ni sestavljena glede na pomembnost, v glavnem pa jih dandanes lahko koristno uporabljamo:

1. v *slovaropisju* (Colins Cobuild English Language Dictionary iz leta 1987 je, na primer, prvi slovar, ki je bil zgrajen izključno na osnovi računalniškega besedilnega korpusa);
2. *pri sestavljanju terminoloških glosarjev;*
3. *pri jezikovnih študijah (v besediloslovju, sociolingvistiki, psiholongvistiki, uporabnem jezikoslovju in prevodoslovju)*
4. *in pri razvoju jezikovnih tehnologij,* predvsem kot:
 - a. pripomočke za učenje tujih jezikov (celo za pisanje empirično usmerjenih učbenikov),
 - b. pripomočke pri prevajanju,
 - c. pripomočke za poučevanje prevajanja (npr. konkordančni programi),
 - č. kot programe za analizo in sintezo govora.

McEnery in Wilson [9] pravita tudi, da ima analiza s pomočjo besedilnih korpusov več prednosti:

- besedilni korpusi so velikanske baze podatkov v naravnem diskurzu, kar nam omogoča empirično analizo pristnih in uporabljenih jezikovnih vzorcev;
- ko jih uporabljamo s pomočjo najnovejše tehnologije, torej (pol)avtomatskih računalniških orodij, nam besedilni korpusi omogočajo analize besedil v takem obsegu, ki bi bile v kakšnem drugačnem formatu popolnoma neizvedljive (Biber et al [10]);
- korpusi so obogateni z dodatnimi jezikovnimi informacijami (npr. anotacija oblikoslovnih in skladenjskih značilnosti, prozodična transkripcija, itd.).

Korpuse lahko apliciramo tudi na področje prevajalstva. Prevodoslovcem se zdita še posebej zanimiva primerljivi in vzporedni tip besedilnih korpusov. Preden se osredotočim na vlogo korpusov v prevodoslovju naj

omenim, da se prevodoslovcem še ni posrečilo izoblikovati enotne terminologije, ki bi uporabljali v korpusnih raziskavah. Več o tem v poglavjih, ki sledijo.

3 VLOGA KORPUSOV V PREVODOSLOVJU

Pri prevajanju se srečujemo z mnogimi težavami. Uporaba večjezičnih besedilnih korpusov bi nam lahko v veliki meri olajšala delo, saj bi nam pokazala pragmatično rabo besed v enem ali več jezikih, s tem pa empirično podkrepila naše domneve o neki besedi oziroma besedni zvezi in njenih prevodih.

Za izdelavo besedilnih korpusov so nujno potrebni selekcijski kriteriji. Bakerjevi [11] se zdijo najpomembnejši naslednji:

- *splošni jezik/ specifično področje*
- *pisni/govorjeni jezik*
- *sinhrono/diahrono*
- *tipičnost* glede na *vir* (pisec/govorec) in *zvrst* (npr. časopisni uvodnik, radijski intervju, članki iz revij, leposlovje, strokovna literatura, zapisi iz sodišča, itd.)
- *geografske omejitve* (npr. britanska in ameriška angleščina)
- *eno-/dvo- ali večjezičnost*

Vendar pa takoj dodaja, da so te kriterije v glavnem postavili jezikoslovci, in to na osnovi enojezičnega korpusa. Takšna klasifikacija korpusov je sicer pravilna, a nepopolna za specifične potrebe v prevodoslovju, predvsem glede na kriterije za eno- oziroma večjezične besedilne zbirke. Pravi, da bi morali prevodoslovci zdaj natančneje opredeliti kriterije za korpus, ki so najbolj neposredno vezani na raziskave v prevodoslovju. Predlaga, da kriteriju za tipičnost dodajo tudi, kdo je prevajal posamezna besedila (profesionalni prevajalci ali amaterji, samostojni/člani prevajalske agencije, je prevod iz/v materni jezik, itd.). Tudi glede zvrsti bi bilo potrebno dodati, da je prevod lahko drugačnega žanra kot besedilo v izvorniku, kar bi seveda vplivalo na razvrstitev v dvo- ali večjezičnih korpusih.

Ponavadi korpus, ki so namenjeni za potrebe prevajanja in prevajalskih študij, imenujemo vzporedni korpusi. Bakerjevi tako poimenovanje ni všeč, saj izraz vzporedni korpus asociira na besedila in njihove prevode v tuji jezik (na protistavno analizo besedil in njihovih prevodov) s tem pa zavaja v miselnost, da so vsi prevajalski korpusi tudi večjezični. Seveda ni nujno tako (glej primerljivi korpusi).

3.1 Tipi korpusov v prevajalstvu

Tako lahko torej besedilne korpus, spet delimo na (povzeto po Baker [11]):

1. **vzporedne**
2. **večjezične**
3. **primerljive**

a) Vzporedni korpusi - gre za zbirke besedil, sestavljene iz besedil v izvornem jeziku in njihovih prevodov v tujem

jeziku. Največ zanimanja za take besedilne zbirke se pojavlja v prevodoslovju. Tak tip korpusa postaja vse bolj nepogrešljiv ne le pri prevajanju samem, ampak tudi kot orodje pri programih za poučevanje prevajanja in pri učenju jezikov s pomočjo jezikovnih tehnologij.

Prav tako jih s pridom uporabljajo pri izdelavi terminoloških baz podatkov za velike mednarodne korporacije (npr. AT&T). Verjetno bodo koristili tudi pri izboljšavah sistemov za strojno prevajanje, saj naj bi se stroji 'učili' iz takšnih korpusov. Največji porabniki vzporednih korpusov so Združeni narodi, Nato, Evropska unija in nekatere države, ki imajo dva uradna jezika, npr. Kanada (Canadian Hansard Corpora [12]).

Kot primer dvojezičnega besedilnega korpusa naj navedem projekt Intersect [5] (izdelan na univerzi v Brightonu in sestavljen iz besedil v francoščini, vzeti iz časopisa *Le Monde*, prevodov člankov v angleščini, ki so se pojavili v *Guardian Weekly*, navodil za uporabo, tehničnih besedil, ipd.). Projekt Lingua [5] je konkordančni projekt, ki ga je finančno podprla Evropska unija, vključuje pa Dansko, Grčijo, Italijo, Veliko Britanijo in Francijo. V okviru tega projekta zbirajo in obdelujejo večjezični vzporedni korpus, ki je namenjen razvoju prevajalskih pripomočkov in orodij za učenje in poučevanje tujih jezikov.

b) Večjezični korpusi - tak tip korpusov vsebuje nize dveh ali več enojezičnih besedilnih korpusov v različnih jezikih, ki so bili izdelani na podlagi podobnih kriterijev. Taki korpusi omogočajo vpogled v jezikovne strukture v njihovem domačem okolju in ne v prevedenem besedilu. Nudijo dostop do naravnih vzorcev v jeziku; pomembno vlogo igrajo predvsem pri materialih za pisanje, poučevanju prevajalcev in pri izboljšavi programov za strojno prevajanje. V večjezičnih korpusih so lahko identificirani tudi termini oziroma njihovi prevodi, bodisi besede ali fraze. Vendar pa se s tem že zelo približujejo terminološkim bankam in pomnilnikom prevodov (Translation Memories).

Kot primer takega tipa korpusa naj navedem The Council of Europe Multilingual Lexicography Project (Baker [11]), ki je vseboval besedilne zbirke v angleščini, nemščini, švedščini, italijanščini, španščini, madžarščini in srbohrvaščini.

c) Primerljivi korpusi - ti so pravzaprav križanci med vzporednimi in večjezičnimi korpusi. Bakerjeva [11] pravi, da so sestavljeni iz dveh posameznih ločenih zbirk besedil v istem jeziku: eno zbirko sestavljajo besedila v izvornem jeziku, medtem ko je druga zbirka prevodov v ta jezik iz enega ali več drugih jezikov. Tako je korpus izvornih besedil pravzaprav enojezični korpus. Seveda pa morata imeti oba korpusa skupne značilnosti, ne le pri zapisu ampak tudi pri izbiri sami, sicer sta le besedilna vira v različnih jezikih. Primerljiva morata biti glede na velikost, področje, obdobje, žanr in, če je le mogoče, tudi temo. Korpus prevodov mora biti reprezentativen ne le

glede na avtorje besedil ampak tudi na prevajalce, ki so besedila prevajali.

Zgornja delitev korpusov zaenkrat v prevodoslovju ni popolnoma uveljavljena, čeprav se vse več prevodoslovcev zaveda, da je treba terminologijo čimprej poenotiti. Ulrych [13] na primer predlaga delitev na **enojezične primerljive korpus**e (pri Bakerjevi so to primerljivi korpusi), **dvo-** ali **večjezične primerljive korpus**e (teksti napisani v različnih jezikih - brez prevodov - in zbrani na podlagi podobnih kriterijev: glede žanrske delitve, komunikacijske funkcije, dolžine, itd.), **dvo-** ali **večjezične vzporedne korpus**e (ki so sestavljeni iz besedil v izvornem jeziku in njihovih prevodov v enega ali več tujih jezikov), in dodaja novo kategorijo **enojezičnih vzporednih korpusov**, ki naj bi bili sestavljeni iz dveh ali več različnih prevodov določenega besedila, ki so jih prevedli različni prevajalci, lahko pa so tudi diahrone verzije prevodov enega samega prevajalca. Tako bi lahko primerjali različne prevode nekega besedila v enem samem jeziku.

3.2 Vloga primerljivih in vzporednih korpusov v prevajalstvu

Glavna vloga primerljivih korpusov v prevajalstvu je identifikacija vzorcev, ki so specifični za prevedena besedila ne glede na izvorni oziroma ciljni jezik. Raziskave na tem področju bi lahko pomagale pri novih hipotezah o procesu prevajanja ali pa jih ovrgle. Glavna funkcija vzporednih korpusov pa je ugotavljanje prevodnih norm v specifičnih kontekstih, in odkrivanje metod in rešitev, h katerim se zatekajo profesionalni prevajalci.

Kot primerka primerljivih in vzporednih korpusov, kjer je bil vključen tudi slovenski jezik, naj navedem TELRI in MULTEXT-East [14], ki sta na voljo tudi na zgoščenki. MULTEXT-East zajema šest srednje- in vzhodnoevropskih jezikov: bolgarski, češki, estonski, madžarski, romunski in slovenski, in je nadaljevanje projekta MULTEXT, ki je zajemal šest jezikov Evropske unije: angleščine, francoščine, nemščine, italijanščine, španščine in nizozemščine, pozneje pa je bil dodan še švedski korpus.

nejša pobočja , kjer je speljana <mulatjera> . Po njih brez težav rahlo p a so skorajda brez snega , le po <mulatjeri> moraš večino časa gaziti do d Kanceljni proti lovski koči ob <mulatjeri> na Kriške pode . Po treh ura guje proti Vratom. Na Vrata nas <mulatjera> pripelje v dveh urah. V vro na južno stran in jo uberemo po <mulatjeri> strmo vsekani navzgor v vršn . To je bila očitno oskrbovalna <mulatjera> in je zato speljana ves čas elo blizu vršne piramide Krna . <Mulatjera> se tu neopazno zgubi . Na se e Zaprikraj . Pot je speljana po <mulatjerah> iz prve svetovne vojne mimo lovska stezica okoli Pihavca do <mulatjere> , ki pelje iz Trente na Kriš Robon) in povratkom po daljši <mulatjeri> čez pode na Nevejsko sedlo . zrazite doline pod Hudim vrhom . <Mulatjera> na Lipnik je namreč speljana

PRIMER 1: Primer KWIC izpisa konkordanc besede "mulatjera"

3.3 Orodja za delo s takimi korpusi

Za pregledovanje označenih besedilnih korpusov lahko uporabljamo vrsto programov. S tako imenovanimi pregledovalniki lahko poiščemo zelene dele korpusa. Najbolj znana oblika predstavitve informacij iz korpusa so konkordance in sicer v obliki KWIC (Key Word In Context). Pokažejo pojavitve izbrane besede oziroma besedne zveze v korpusu, izpisane pa so na sredini, skupaj s svojim sobesedilom. Primer 1 nam kaže pojavitev besede mulatjera (več o tem v Erjavec [3]).

Zanimivo je tudi iskanje sobesedij ali kolokacij, saj le-te lahko pokažejo vezave določenih besed z drugimi besedami tako s skladijskega kot s pomenskega stališča. Način pregledovanja je podoben kot pri KWIC, obstajajo pa tudi programi, ki avtomatsko izberejo kolokacije. Nekateri konkordančni programi pokažejo cele stavke ali pa celo odlomke. Drugi, npr. Microconcord, pa frekvenčne sezname kolokacije v določenem obsegu (recimo tri besede levo ali desno od konkordance).

Obstaja tudi vrsta drugih programov, ki naredijo seznam vseh oblik besed v celotnem korpusu ali v določenem delu korpusa, in sicer v frekvenčnem ali pa abecednem zaporedju (podrobnejši pregled v Sinclair [15]).

Pri vzporednih korpusih je prikaz v dveh poravnanih KWIC oknih. 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, v katerem se beseda nahaja, hitro pokaže na idiomatske uporabe te besede.

Kot primer naj navedem nekaj različnih konkordančnih programov. Multiconcord je večjezični vzporedni konkordančni program, ki so ga uporabljali v projektu Lingua. Ta konkordančnik omogoča avtomatsko primerjavo besedil in njihovih prevodov v deset evropskih jezikov. Glavni cilj projekta je bil razvoj konkordančnega programa, ki je zgrajen na osnovi operacijskega sistema Windows, in predstavitve možnosti za eksperimentiranje z njim. Konkondarčnik naj bi namreč v prihodnje uporabljali pri učenju tujih jezikov in pri prevajanju. (Več o tem Ulrych [13], in King [16]).

Tudi ParaConc je večjezični konkordančni program, ki ga je Barlow [17] uporabljal kot pripomoček pri poučevanju jezika.

Precej konkordančnih programov dostopnih tudi na trgu. V naslednjih poglavjih bom orisala pomen in koristnost korpusov v prevodoslovju.

3.4 Pomen vzporednih in primerljivih besedilnih korpusov v prevodoslovju

Vse več je prevodoslovcev, ki so prepričani, da je uporaba takšnih besedilnih zbirk za prevodoslovje velikega pomena. Bakerjeva [18] trdi, da bo resnični preobrat v prevajalskih študijah pravzaprav nastopil kot neposredna posledica dostopa do velikih korpusov izvornih in prevedenih besedil ter razvoja ne le specifičnih metod ampak tudi orodij za uporabo korpusov na načine, ki so najprimernejši za potrebe prevajanja. S korpusi si bomo lahko pomagali tudi pri opredeljevanju načel, po katerih se prevajalci ravnavajo, ko prevajajo.

Čeprav je koristnost besedilnih korpusov za potrebe prevodoslovja nesporna in je bilo o tem napisanih veliko teoretičnih razprav, je do njihove praktične uporabe za iskanje boljših prevodnih ustreznic in odkrivanje specifičnih struktur nekega jezika še dolga pot. Vsekakor pa je večjezične in vzporedne korpuse možno uporabljati pri pisanju materialov, pri poučevanju prevajalcev, pri izboljšavi sistemov za strojno prevajanje in pri terminoloških študijah, kjer so terminološki izrazi zbrani na podlagi korpusov ((Shuttleworth in Cowie 1997: 120) v Ulrych [13]).

3.5 Uporaba vzporednih in primerljivih besedilnih korpusov pri poučevanju prevajanja

Uporaba vzporednih in primerljivih besedilnih korpusov v prevodoslovju postaja ena najnovejših metod, ki korpuse uveljavlja kot koristno orodje tudi pri poučevanju bodočih prevajalcev. Eden prvih, ki so predlagali uporabo besedilnih korpusov pri poučevanju prevajanja in pri pripravi učnih materialov za tako poučevanje, je bil Hans Lindquist. Tudi J. Sinclair [15] v svojih delih poudarja pomen korpusnih virov v prihodnjih letih in njihov velik vpliv na razvoj prevajanja. Marsikje konkordančni programi predstavljajo obvezni del poučevanja bodočih prevajalcev (Maia, Barlow, King [16]) in upamo lahko, da bo v bližnji prihodnosti tako tudi pri nas.

Največji prispevek primerljivih korpusov za prevajalske študije je identificiranje struktur, ki so značilne za prevedena besedila ne glede na njihov izvorni in ciljni jezik. Omogoča nam torej primerjavo ne le dveh jezikov (IJ in CJ) oziroma besedil in njunih jezikovnih struktur, ampak tudi prevode s prostim tvorjenjem besedil v istem jeziku. Primarna funkcija vzporednih korpusov pa je ugotavljanje prevodnih norm v specifičnih kontekstih in odkrivanje metod in rešitev, h katerim se zatekajo

profesionalni prevajalci, ko skušajo premagati težave, s katerimi se srečujejo pri prevajanju nekega besedila.

4 ZAKLJUČEK

Predstavila sem različne tipe besedilnih korpusov in nekatere vidike uporabe takih besedilnih zbirk tudi s pomočjo konkordančnih programov. Še posebej zanimiva se mi zdi možnost njihove rabe pri poučevanju prevajanja. V članku sem se osredotočila na kratek pregled besedilnih zbirk, natančneje pa primerljivih in vzporednih besedilnih korpusov in njihove uporabnosti. MULTEX-East in TELRI sta dober začetek. V kratkem se nam obeta tudi enojezični referenčni besedilni korpus Fida. Ti projekti so nedvomno spodbuda za nadaljnje delo na tem področju, predvsem kar zadeva dvo- ali večjezične besedilne korpuse, ki bi se jih dalo koristno uporabljati v prevodoslovju. Slovenija pred vstopom v Evropsko unijo vsekakor rabi takšne pridobitve. Večjezične korpuse vzporednih in primerljivih besedil bi namreč lahko s pridom uporabljali na več področjih, predvsem pa pri ugotavljanju prevodne ustreznosti.

5 ZAHVALA

Za korisne pripombe, nasvete in potrpežljivost se zahvaljujem Neži, Mateji in Tatjani, še posebej pa Tomažu Erjavcu iz IJS. Za vse morebitne napake v članku je odgovorna avtorica.

6 VIRI

- [1] [Meyer, I., Mackintosh, K., (1996). "The Corpus from a Terminographer's Viewpoint", v: *International Journal of Corpus Linguistics*, 1/2, 257-285.
- [2] Gorjanc, V. (1997). "Češki narodni korpus". *Slavistična revija*, 45/3-4, julij-december, 578-580.
- [3] Erjavec, T. (1996/97). "Računalniške zbirke besedil". *Jezik in slovstvo*, 42/2-3, 81-96 ali na: <http://nl.ijs.si/tomaz/Bib/S1Korpus/>
- [4] Slovenski korpus FIDA: <http://www.fida.net/>
- [5] PEDANT: <http://logos.svenska.gu.se/PEDANT/>
- [6] EAGLES: <http://www.ilc.pi.cnr.it/EAGLES96/>
- [7] Bank of English: <http://www.titania.cobuild.collins.co.uk/>
- [8] CNC, Češki nacionalni korpus: <http://ucnk.ff.cuni.cz/cnc>
- [9] McEnery, T., Wilson, A. (1996). *Corpus Linguistics*. Edinburgh: Edinburgh University Press, 90-111.
- [10] Biber, D., Conrad S. and Reppen R. (1994). *Corpus-based Approaches to Issues in Applied Linguistics*, v: *Applied Linguistics*, OUP, 15/2: 169-175.
- [11] Baker, M. (1995). "Corpora in Translation Studies: An Overview and Some Suggestions for Future Research", v *Target* 7:2, Amsterdam: John Benjamins Publishing Co., 223-243.
- [12] Canadian Hansard Corpora: <http://www.ruf.rice.edu/~barlow/para.html>

- [13]Ulrych, M. (1997). "The Impact of Multilingual Parallel Concordancing on Translation", v: *Practical Applications in Language Corpora* (ur. Lewandowska-Tomaszczyk, B. in Melia, P.J.), Lodz University Press, Lodz, 421-435.
- [14]MULTEXT-East in TELRI: <http://www.ids-mannheim.de/telri/cdrom.html>, ali v: Erjavec T., Ide N., Petkevič V. in Veronis J. (1996). Multext-east: Multilingual text tools and corpora for central and eastern European Languages, v: *Proceedings of the First TELRI European Seminar: Language Resources for Language Technology*, Sept. 1995, Tihany, Madžarska, 87-98.
- [15]Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press, Oxford, 1991.
- [16]King, P. (1997). "Parallel Corpora for Translator Training", v: *Practical Applications in Language Corpora* (ur. Lewandowska-Tomaszczyk, B. in Melia, P.J.), Lodz University Press, Lodz, 393-402.
- [17]Barlow, M. (1996). "Parallel Texts in Language Teaching", v: *Proceedings of Teaching and Language Corpora* (ur. Botley, S., Glass, J., McEnery T. in Wilson A.), 45-56.
- [18]Baker, M. (1993). "Corpus Linguistics and Translation Studies. Implications and Applications", v: *Text and Technology. In Honour of John Sinclair*, ur. M Baker et al. Amsterdam, Philadelphia: John Benjamins Publishing Co., 233-250.