

# RAČUNALNIŠKE TEHNOLOGIJE ZA PREVAJANJE

Špela Vintar  
 Oddelek za prevajanje in tolmačenje  
 Filozofska fakulteta  
 e-mail: spela.vintar@guest.arnes.si

## Povzetek

Sodobnemu prevajalcu pomeni računalnik mnogo več kot pisalni stroj. Z različnimi orodji za računalniško podprto prevajanje, kamor poleg črkovalnikov, elektronskih slovarjev in tezavrov sodijo predvsem programi s pomnilnikom prevodov in terminološke banke, je pri nekaterih vrstah besedil mogoče prevajalski proces pospešiti, izboljšati in poceniti, poleg teh pa v prevajalsko prakso prinaša korenite spremembe tudi internet. Prispevek podaja pregled jezikovnih tehnologij in virov za prevajalce od zametkov strojnega prevajanja do programskih orodij, ki so se razvila v zadnjem desetletju, spregovori pa tudi o uporabi korpusov pri prevajanju in vplivu sodobnih informacijsko-komunikacijskih medijev na to področje. V zaključku orišemo položaj v Sloveniji in nakažemo smernice za učinkovitejše uvajanje teh tehnologij v naš prevajalski prostor, predvsem z oblikovanjem ustreznih študijskih programov.

## Abstract

*To a technologically aware translator a computer is more than simply a typewriter. Tools for Computer-Assisted Translation - an umbrella term embracing spellcheckers, on-line dictionaries and thesauri as well as Translation Memory and Terminology Management systems - have the potential to raise the quality, speed and cost effectiveness of the translation process. It is also undergoing profound changes due to the explosive growth of the Internet. The paper gives an overview of translation technologies and resources from their beginnings in machine translation to the software products developed during the past few years. Described are applications of textual corpora in translation and the impact of new information and communication media in this field. The concluding section outlines the present situation in Slovenia and suggests ways for further implementation of translation technologies, especially by (re-)designing translator-training curricula.*



## UVOD

Naravni jezik je eden najkompleksnejših sistemov, kar se jih je razvilo skozi človekovo evolucijo, prevajanje pa kot prenašanje sporočil med temi sistemi eden najbolj zapletenih miselnih procesov, pri katerem ostajajo mnoga vprašanja s psihološkega in jezikoslovnega vidika nepojasnjena. Razvijanje računalniških tehnologij, ki bi znale ta proces simulirati ali vsaj optimizirati nekatere njegove dele, je zato že dolgo izziv tako jezikoslovcem kot računalničarjem. Prva desetletja živahnega raziskovanja so skupaj z mnogimi razočaranji in težavami utemeljila nastanek novih znanstvenih področij računalniškega jezikoslovja in prevodoslovja in prispevala pomembna spoznanja, ki so med drugim povzročila preobrat v »klasični«  
lingvistiki in povratek k empiričnim metodam v jezikoslovju.

Danes so strojni prevajalniki človeka povsem ali delno nadomestili že v številnih industrijskih vejah, kjer se rutinsko prevajajo določeni enostavni tipi

besedil, in pri mnogih prevajalskih nalogah, kjer uporabniku zadošča grobi prevod. Na drugi strani imajo jezikovni profesionalci, med katere sodijo tudi prevajalci, danes na voljo številna računalniška orodja, ki omogočajo hitrejšo in bolj kakovostno prevajanje nekaterih vrst besedil, sodobne informacijske in komunikacijske tehnologije pa prevajalcem nudijo hiter dostop do jezikovnih virov in podatkov z vsega sveta. V času globalizacije in evropskih integracij, ko produkcija novih besedil in zahteve po prevodih nezadržno naraščajo, se od prevajalcev na eni strani pričakuje, da bodo svoje delo opravljali hitro in kakovostno, po drugi strani pa zaradi velikih količin besedil sedanjih visokih cen prevodov ne bo več mogoče ohranjati. Poznavanje prevajalskih orodij in drugih jezikovnih tehnologij bo v prihodnje tako nujen del prevajalčeve izobrazbe in pogoj za njegovo konkurenčnost.

## STROJNO PREVAJANJE NEKOČ IN DANES

Izraz strojno prevajanje (*MT – Machine Translation*) navadno označuje računalniške sisteme za prevajanje naravnih jezikov, pri katerih je prevajalski proces do največje možne mere avtomatiziran. Ti sistemi vključujejo lastne eno- in večjezične leksikone, programe za morfološko in sintaktično analizo in sintezo, razreševanje večpomenskosti, prepoznavanje večbesednih semantičnih enot in druge kompleksne mehanizme, ki naj bi omogočali avtomatski prevod s čim manj napakami. Vloga uporabnika pri takih sistemih ni vselej povsem odpravljena, pri interaktivnih sistemih je za razreševanje večpomenskosti predvidena človekova pomoč, skoraj pri vseh sistemih pa je potrebna predpriprava (*pre-editing*) izvirnega besedila in poprava (*post-editing*) rezultata (Hutchins, 1992).

Prve resne raziskave na tem področju so se pričele kmalu po vojni v ZDA na MIT-u, leta 1954 pa sta IBM in Georgetown University že predstavila prvi sistem za strojno prevajanje iz ruščine v angleščino, katerega besednjak je bil omejen na 250 besed in je obsegal le šest slovničnih pravil. Čeprav so bili rezultati skromni, je ta poskus prepričal ameriško vlado in širšo javnost, da je zamisel uresničljiva, zato so v naslednjih letih v ZDA predvsem v strateške namene finančno podprli vrsto tovrstnih projektov, živahno raziskovanje pa se je pričelo tudi v Evropi in Sovjetski zvezi.

Že v začetku šestdesetih let pa se je začetno navdušenje pričelo krhati, saj se je že na leksikalni ravni pojavila vrsta težav, povezanih predvsem z razreševanjem večpomenskosti, tj. homografije in polisemije. Leta 1966 je posebni odbor ameriške vlade ALPAC, ki je imel nalogo ovrednotiti dotdanje rezultate in oceniti možnosti za nadaljnji razvoj, podal znamenito uničujoče poročilo, katerega posledica je bila, da je bila finančna pomoč ukinjena skoraj vsem projektom na tem področju, težišče raziskav pa se je vsaj za naslednje desetletje preselilo v Evropo.

Tu se je razvoj nadaljeval, vendar pogojen s povsem drugačnimi potrebami. V Evropi kot izrazito mnogojezičnem prostoru je razvoj narekovala predvsem Evropska gospodarska skupnost, ki je najprej podprla sistem za angleško-francoske prevode Systran (ta je bil mnogo prej zasnovan v ZDA za potrebe US Air Force), nato pa velikopotezni projekt EUROTRA za prevajanje med vsemi jeziki tedanje EGS.

Hkrati s novimi projekti so se razvijali tudi različni pristopi k izgradnji sistemov za strojno prevajanje. Prvi poskusi so temeljili na t.i. direktnem pristopu, ki razen zelo osnovne morfološke analize ni vseboval drugih modulov za jezikovno analizo in je bilo tako prevajanje bolj ali manj omejeno na zamenjavo posameznih besed s pomočjo dvojezičnega leksikona. Na podlagi pomanjkljivosti tega pristopa so pri projektu Systran razvili t.i. transferni pristop, ki prevajanje na ravnih

povedi razdeli v fazo analize, dvojezičnega transfera in sinteze. Posamezni moduli pri transfernem sistemu so tako zasnovani za določeni par jezikov, pri čemer se v prvi fazi analizira izhodiščna stavčna struktura (*parsing*), v drugi se ta prevede v ustrežno strukturo ciljnega jezika, v tretji fazi pa se oblikuje ciljna poved. Vzporedno s tem poteka proces oblikoslovne analize in sinteze. Projekt EUROTRA je zaradi svoje izrazito mnogojezične usmerjenosti skušal razviti lasten pristop, t.i. *interlingua*, katerega osnovna zamisel je prevajanje vseh jezikov v nekakšen medjezik oziroma univerzalno, jezikovno neodvisno strukturo in semantično ponazoritev stavčnih členov ter odnosov med njimi v odvisnosti od povedka. V svoji idealni različici bi taka vmesna oblika omogočala prevajanje v vse jezike sveta, saj bi bilo iz nje mogoče neposredno oblikovati poved v ciljnem jeziku. Žal se je kmalu pokazalo, da stvar ni tako preprosta, zato se je razvoj preusmeril v iskanje medjezikovnih elementov v skupini izbranih jezikov, s čimer je pri izgradnji sistemov za strojno prevajanje mogoče izkoristiti podobnosti med jeziki in se pri odpravljanju napak osredotočiti na razlike (Hutchins, 1992).

Danes večina znanih sistemov temelji na transfernem pristopu, pri tem pa se pri boljših sistemih vse bolj uveljavljajo empirični pristopi. Tako se za razjasnjevanje skladenjske ali leksikalne večpomenskosti uporabljajo vzporedni korpusi, ki računalniku pri vsaki obstoječi negotovost nudijo zbirko primerov iz resničnih besedil, na podlagi katerih se je sistem s pomočjo primerjave frekventnosti sposoben odločiti za prevodno ustreznico.

Predvsem za večje jezike je na trgu vse več komercialnih prevajalskih programov, ki se med seboj v kakovosti in ceni zelo razlikujejo. Boljši med njimi imajo vgrajene tudi specialne leksikone za posamezna področja, ki olajšujejo razreševanje večpomenskosti, ponujajo pa tudi možnost hranjenja že prevedenih in popravljenih besedil, ki programu nato služijo za vzor pri novih prevodih (Schubert, 1998). Strojni prevajalniki so prisotni tudi na internetu, ponekod vsebovani v iskalnih orodjih (Systran<sup>1</sup> v povezavi z Altavisto<sup>2</sup>, Langenscheidt T1 Professional<sup>3</sup>, Translation Experts<sup>4</sup> itd.).

V podjetjih, kjer prevajanje deloma ali v celoti poteka strojno, se v produkcijo besedil uvaja t.i. nadzorovani jezik (*controlled language*), ki z vnaprej definiranim besediščem in stavčnimi strukturami zagotavlja kakovost in terminološko ustreznost strojnih prevodov. V ta namen so nekateri uporabniki

1 URL: <http://babelfish.altavista.digital.com/>

2 URL: <http://www.systranmt.com/>

3 URL: <http://www.langenscheidt.de/ubersetzt1.htm>

4 URL: <http://www.tranexp.com/>

razvili programska orodja za prevajanje besedil v nadzorovani jezik oziroma za preverjanje le-tega pred strojnim prevajanjem.

Za slovenščino uporabnih strojnih prevajalnikov še nimamo, čeprav je to področje izziv za mnoge, ki se ukvarjajo z jezikovnimi tehnologijami. Poleg tega, da je slovenščina že zaradi svoje morfološke bogatosti s tega vidika zapleten jezik, pa je mnogo hujša ovira dejstvo, da za slovenščino šele gradimo osnovne jezikovne vire in orodja, ki so za izdelavo strojnega prevajalnika neizogibno potrebni, npr. program za oblikoskladenjsko analizo, dvojezične leksikone, korpus za izdelavo eno- ali dvojezičnega leksikona itd. Nekaj poskusov izgradnje sistema za strojno prevajanje se je zgodilo na Institutu Jožef Stefan, od komercialnih ustanov pa tu brez dvoma prednjači podjetje Amebis s svojim angleško-slovenskim prevajalnikom Presis, ki je zaenkrat še v eksperimentalni fazi (Romih, 1998). V zvezi s strojnim prevajanjem se postavlja vprašanje, ali je ta orodja sploh smiselno razvijati pri nas, saj imajo tuji proizvajalci potrebno znanje in izkušnje že zbrane, po drugi strani pa je pridobivanje lastnih izkušenj na tem področju pomembno zaradi morebitnega sodelovanja z drugimi v bodoče, kar je bil pri podjetju Amebis tudi odločilni argument za razvijanje lastnega sistema.

## PROGRAMSKA ORODJA ZA RAČUNALNIŠKO PODPRTO PREVAJANJE

Programi za strojno prevajanje prevajalce v nekaterih specifičnih segmentih sicer že nadomeščajo, za veliko večino prevajanih besedil pa še vedno velja, da obstoječa orodja ne morejo izdelati zadovoljivih prevodov oziroma bi bila predpriprava besedil in poprava rezultatov strojnega prevajanja mnogo preveč zamudna. Za jezikovne profesionalce so se tako v zadnjih desetih letih razvila orodja, ki skušajo prevajalski proces olajšati, optimizirati in poceniti, ne pa simulirati. Med orodja za računalniško podprto prevajanje (*CAT – Computer-Aided Translation*) v širšem smislu sodijo vse jezikovne tehnologije, ki prevajalcu služijo kot pripomoček na poti do prevoda, se pravi tudi črkovalniki, tezavri, elektronski slovarji in drugi elektronski podatkovni viri, v ožjem pomenu pa so pomemben korak na tem področju predvsem programi za izdelavo in vzdrževanje terminoloških bank ter programi s pomnilnikom prevodov (*TM – Translation Memory*).

### Pomnilniki prevodov

Pomnilnik prevodov je podatkovna zbirka vzporednih dvo- ali večjezičnih prevodnih enot (*translation unit*). Prevodna enota je navadno poved, lahko pa tudi večji (npr. odstavek) ali manjši del besedila (npr. naslov,

polje v tabeli, alineja itd.), ki je v izvorniku in prevodih zapisana v pomnilnik. Temeljna predpostavka pri uporabi tega orodja je torej, da se pri nekaterih vrstah besedil določene strukture, formulacije ali cele povedi ponavljajo, zato pomnilnik prevodov med prevajanjem v ozadju išče podobne ali enake enote in jih ponudi prevajalcu. Kadar sta si nova prevodna enota in enota iz pomnilnika le podobni, gre za t.i. megleni zadetek (*fuzzy match*), ki ga program poišče na podlagi ujemanja posameznih besed ali besednih nizov. Stopnja ujemanja oziroma podobnosti je navadno izražena v odstotkih, prag ujemanja pa lahko uporabnik nastavi sam. Kadar je nova prevodna enota povsem enaka enoti, ki je že zapisana v pomnilnik, gre za polni zadetek (*exact match*).

### Primer:<sup>5</sup>

Prevodna enota v pomnilniku prevodov:

*Repairs must be made by trained fitters only.*

*Vsa popravila mora izvajati ustrezni strokovnjak.*

Nova prevodna enota:

*Repairs on electrical appliances must be made by trained fitters only. (73%)*

Pomnilnik prevodov nastaja sproti med prevajanjem, če pa imamo na razpolago že prevedena besedila v izvorniku in prevodu v elektronski obliki, lahko pomnilnik ustvarimo tudi s pomočjo orodja za vzporejanje (*alignment*), ki obe besedili razdeli na segmente, te poravnava v pare izvornik-prevod in jih shrani v pomnilnik.

Ker pomnilnik prevodov nima vgrajenih modulov za morfosintaktično analizo prevodnih enot, ampak je sposoben prepoznavati podobnost le na ravni besed ali besednih nizov, je teoretično uporaben za vse jezikovne pare, praktično pa za vse jezike, za katere je zagotovljena znakovna podpora.

Dejanska uporabnost pomnilnikov prevodov je v veliki meri odvisna od značilnosti besedil, ki jih želimo prevajati, saj temelji na predpostavki, da se v določenih vrstah besedil stavki, besedni nizi ali posamezni izrazi v bolj ali manj nespremenjeni obliki ponavljajo. To velja predvsem za nekatere besedilne vrste iz družine tehničnih in strokovnih besedil, na primer navodila za uporabo, tehnične opise, nekatera pravna besedila itd. Poleg t.i. faktorja ponavljanja na učinkovitost dela s pomnilnikom prevodov vplivajo še drugi dejavniki v okviru besedila in izven njega, npr. povprečna dolžina prevodnih enot, obseg prevajalskega projekta (dolžina besedila), velikost, število in kakovost terminoloških bank in pomnilnikov prevodov, ki jih imamo na razpolago pred prevajanjem, ipd.

<sup>5</sup> Za primer je bil uporabljen program TRADOS Translator's Workbench.

Dejanska uporabnost tega orodja je torej močno odvisna od narave prevajalskega dela oziroma od značilnosti besedil, ki jih prevajamo. To je razvidno tudi iz prakse, saj se programi s pomnilnikom prevodov najhitreje uveljavljajo prav v velikih industrijskih podjetjih, kjer se prevajajo velike količine besedil z istega področja in kjer se že prevedena besedila pogosto posodablja in izdajajo na novo (tipičen primer so navodila za uporabo oz. uporabniški priročniki). Druga velika skupina uporabnikov zajema mednarodne in državne upravne organe (npr. Evropski parlament).

Program s pomnilnikom prevodov je lahko del urejevalnika besedil (Word, WordPerfect), lahko pa ima lastno delovno namizje, v katerega uvozimo dokument, ki ga želimo prevesti. Med najpomembnejše tovrstne proizvode sodijo TRADOS Translator's Workbench<sup>6</sup> (zraven sodita še terminološki program MultiTerm in orodje za vzporejanje WinAlign), IBM Translation Manager<sup>7</sup>, STAR Transit in TermStar<sup>8</sup>, ATRIL DéjaVu<sup>9</sup> in drugi.

### Programi za gradnjo, vzdrževanje in izmenjavo terminoloških bank

Vsi strokovni in tehnični prevajalci se vsakodnevno srečujejo s termini, ki jih v splošnih slovarjih ni najti, pogosto pa tudi v specialnih in področnih slovarjih ne. Večina strokovnih in znanstvenih področij se danes razvija tako hitro, da novih izrazov slovaropisje ne more spremljati. Hkrati je terminološka doslednost in natančnost prav pri strokovnih in tehničnih besedilih ključnega pomena, in tako tudi ključni pogoj za funkcionalen prevod. Prevajalec ali prevajalski tim si zato pogosto ustvarja lastne terminološke glosarje ali banke, ki naj bi pomagale pri zagotavljanju terminološke enotnosti vsaj za določeni prevajalski projekt, če že ne za celo stroko.

V ta namen so na voljo številna programska orodja, s katerimi je mogoče terminološke vnose strukturirati, posodablja, povezovati med seboj in po njih iskati. Standardni terminološki vnos je sestavljen iz številnih polj (Arntz/Picht, 1995: 239), ki določeni termin ustrezno opredeljujejo, npr. izvorni termin, definicija, primer iz besedila in vir, klasifikacijski ključ, prevodi, sinonimi in antonimi, grafični prikaz itd. Pri računalniških programih za gradnjo terminoloških bank se polja pri vnosu v grobem delijo na glavo vnosa, kjer so shranjeni upravni podatki (datum vnosa, ime vnašalca, datum zadnje spremembe, ime spreminjevalca, zaporedna številka, podatki o prevajalskem projektu

itd.), indeksna polja (termin v različnih jezikih, sinonimi itd.), besedilna polja (definicija, opomba, primer itd.) in atributna polja (slovnični kvalifikatorji itd.).

Starejši sistemi (npr. Eurodicautom<sup>10</sup>, terminološka banka Evropske komisije) so sicer omogočali vnašanje številnih jezikovnih, strokovnih in upravnih podatkov, niso pa zagotavljali navzkrižnih povezav med vnosi in posameznimi izrazi ter izmenljivosti podatkov. Sodobni programi za gradnjo terminoloških bank omogočajo oblikovanje vnosne sheme v skladu s specifičnimi potrebami prevajalskega projekta, navzkrižne povezave, ki terminološke vnose strukturirajo v semantične mreže, grafične prikaze, v zadnjem času pa so bili oblikovani tudi standardi za zapis tovrstnih podatkovnih zbirk, ki zagotavljajo njihovo izmenljivost (MARTIF; ISO 12200) (Reinke/Schmitz, 1998).

Programi za gradnjo terminoloških bank so navadno vgrajeni v programe s pomnilnikom prevodov ali so del istega programskega paketa. Med najpogosteje uporabljanimi je MultiTerm podjetja TRADOS, sledijo pa mu STAR-ov Termstar, ATRIL-ov TermWatch in drugi.

Tudi na področju terminologije se odpirajo nove poti s spoznanji korpusnega jezikoslovja. Računalniška zbirka besedil z določenega strokovnega področja je izredno dragocen vir pri gradnji terminoloških bank, če pa imamo na razpolago ustrezna računalniška orodja, je mogoče le-to celo delno ali povsem avtomatizirati. Z uporabo vzporednih korpusov je mogoče iz izhodiščnega besedila avtomatsko izdelati seznam potencialnih terminoloških izrazov in nato na podlagi vzporednih konkordanc iz ciljnega besedila izluščiti predloge za prevodne ustreznice. To bi bil nedvomno hiter in učinkovit način za gradnjo terminoloških bank za določeno področje, saj nam hkrati omogoča tudi vpogled v dejansko rabo termina v besedilu. Žal je predpogoj za to ustrezna zbirka vzporednih besedil v elektronski obliki, ki je pogosto ni enostavno dobiti.

### Druga programska orodja

Za specifična področja prevajanja se razvijajo tudi posebna orodja, posebej tam, kjer je povpraševanje veliko in je od njih mogoče pričakovati tržne prednosti. Nekaterim velikim podjetjem se celo splača razvijati lastne programe s pomnilnikom prevodov ali sisteme za upravljanje s terminologijo, ki so prirejeni posebnim zahtevam uporabnikov in vgrajeni v celotni pretok besedil in delovne procese.

V zadnjih letih je ena od najhitreje rastočih vej na področju računalništva lokalizacijska industrija, ki s seboj prinaša tudi ogromne količine prevajalskega dela. Za prevajanje programskih paketov in njihove dokumentacije se na široko uporabljajo prej omenjeni programi s pomnilniki prevodov v povezavi s terminološkimi bankami, poleg tega pa je na tržišču na voljo

6 URL: <http://www.trados.com/>

7 URL: <http://www.software.ibm.com/ad/translat/eqfn0b02.html>

8 URL: <http://www.translit.se/star/transit.htm>

9 URL: <http://www.atril.com/>

10 URL: <http://www2.echo.lu/edic/>

nekaj posebnih lokalizacijskih programskih orodij, ki so še posebej koristna pri prevajanju programskih datotek in vsebujejo poleg gornjih komponent tudi validacijske funkcije. Med najbolj uporabljanimi orodji so Corel Catalyst, Applocalize, Accent Global Development Kit, Microsoft RLToolset itd. (Esselink, 1998).

## KORPUSI PRI PREVAJANJU

Računalniške zbirke besedil ali korpusi so za izdelavo vsakršnih jezikovnih virov - najsi gre za eno- ali dvojezične slovarje, slovnice, črkovalnike, tezavre, terminološke banke ali pomnilnike prevodov - neprecenljivega pomena, saj naj bi odražali stanje določenega jezika ali podjezika v njegovi besedilni manifestaciji v danem času in prostoru. Z razvojem zmogljivih računalnikov se je uresničila tudi možnost gradnje velikih računalniških korpusov z več sto milijonov besedami, vzporedno s tem pa se razvijajo tudi vse boljša računalniška orodja za njihovo izgradnjo, analizo, upravljanje in iskanje po njih. Za prevajalce so zanimivi predvsem vzporedni korpusi, ki vsebujejo poravnane stavke izvirnika in njegovega prevoda v enega ali več jezikov, pri prevajanju v tuji jezik pa tudi enojezični korpusi bodisi splošnega jezika bodisi kakšne jezikovne podvrste (npr. korpusi literarnih, strokovnih ali publicističnih besedil). Ker za slovenski jezik večjih enojezičnih ali vzporednih korpusov še nimamo, si slovenski prevajalci trenutno lahko še najbolj pomagajo s tujejezičnimi korpusi, ki so - skupaj z iskalnimi orodji - dosegljivi na internetu. Za angleški jezik sta takšna npr. British National Corpus<sup>11</sup> in The Bank of English<sup>12</sup>, za nemški daje korpus na razpolago projekt Cosmas<sup>13</sup>.

Med najpogosteje uporabljanimi metodami pri delu s korpusi so izdelava frekvenčnih seznamov in iskanje konkordanc in kolokacij (Sinclair, 1991). Čeprav frekvenčni seznam ni nič drugega kot štetje, kolikokrat se določena beseda ali besedna zveza v korpusu pojavi, nam lahko že ta preprosta statistična obdelava nudi uporabne - in pogosto presenetljive - rezultate. V angleško govorečem svetu - za angleški jezik obstaja tudi največ tovrstnih virov in raziskav - so na primer rezultati raziskav o frekventnosti posameznih časovnih glagolskih oblik pošteno pretresli dotlej veljavne temelje angleške slovnice in posledično tudi metodike pouka angleškega jezika.

Konkordanca je prikaz pojavov določene besede ali niza skupaj s sobesedilom (KWIC - keyword in context). Orodje za iskanje konkordanc v korpusu nam tako za vsako poizvedbo prikaže zadetke v korpusu

skupaj z neposrednim kontekstom v besedilu, pri čemer je ta lahko omejen na število znakov ali besed levo in desno od iskanega niza, na poved, verz ali odstavek, pri nekaterih konkordančnih orodjih je tudi uporabniško nastavljen.

Pregled konkordanc za določeni niz, besedo ali besedno zvezo je pri prevajanju lahko zelo koristen. Posebej pri prevajanju v tuji jezik nas pogosto zanima, v kakšnem kontekstu je določeno besedo ali frazo mogoče uporabiti, s katerim predlogom se veže, ali jo je mogoče uporabiti figurativno in podobno. Če imamo na voljo dovolj velik korpus ciljnega jezika, si lahko ogledamo primere rabe iskane besede v dejanskih besedilih, kar je še posebej koristno pri neologizmih, izposojenkah in drugih besedah, ki jih v običajnih slovarjih (še) ni mogoče najti. Kadar imamo na voljo vzporedni korpus, lahko s pomočjo vzporednih konkordanc iščemo možne prevodne ustreznice. Če imamo za iskano besedo v izhodiščnem jeziku na voljo npr. petdeset primerov njene pojavitve v stavku skupaj s prevodom, lahko primerjamo frekventnost posameznih prevodnih ustreznic in njihove kontekste, kar močno olajša izbiro primerne prevoda v našem besedilu.

Ker večina orodij za iskanje po korpusih podpira tudi divje iskanje ali regularne izraze, je mogoče iskati tudi dele besed, morfeme, besede, ki se z določenim nizom končajo ali začnejo, sopojav dveh ali več besed in podobno. Tudi ta možnost je pri prevajanju pogosto dragocena, na primer kadar nas zanima, kako se v sodobnem jeziku obnaša določeni besedotvorni morfem ali kadar za določeno besedo iščemo rimo.

Med najbolj znanimi orodji za iskanje po korpusih, ki so za sprejemljivo ceno naprodaj tudi prek interneta, so Wordsmith<sup>14</sup>, MonoConc<sup>15</sup> za iskanje po enojezičnih in ParaConc<sup>16</sup> za iskanje po vzporednih korpusih.

Uporaba korpusov pri prevajanju se uveljavlja šele zadnja leta, ko je vse več korpusov - vsaj za večje jezike - javno dostopnih preko interneta. Za lastne potrebe si lahko prevajalec ustvari tudi svoj manjši korpus, ki ga nato uporablja bodisi s pomočjo orodij, ki jih nudijo programi s pomnilnikom prevodov (tudi ti pogosto omogočajo iskanje konkordanc), bodisi s pomočjo enega od zgoraj naštetih orodij.

## INTERNET - ZAKLADNICA INFORMACIJ ZA PREVAJALCE

Internet je v zadnjem desetletju revolucionarno vplival na številna področja človekovega delovanja in življenja, v zadnjih letih pa postaja tudi vse bolj nepogrešljivo orodje za prevajalce. Za učinkovito izkoriščanje

11 URL: <http://info.ox.ac.uk/bnc/>

12 URL: <http://titania.cobuild.collins.co.uk/>

13 URL: <http://www.ids-mannheim.de/kt/corpora-me.html>

14 URL: <http://www.ndirect.co.uk/lexical/>

15 URL: <http://www.athel.com/new.html>

16 glej prejšnji

možnosti, ki jih ta medij ponuja, je bistveno predvsem, kako dobro se prevajalec znajde v internetnem morju informacij in storitev.

Pri prevajanju se prevajalec srečuje z najrazličnejšimi besedili s številnih področij, in prevajalski problemi, ki se ob tem pojavljajo, pogosto niso rešljivi zgolj s pomočjo slovarjev in leksikonov na domači polici ali trdem disku. Včasih se delo zatakne zaradi nerazumevanja izhodiščnega besedila, zaradi manjkajočega izraza v tujem ali maternem jeziku, pogosto pa si je za kakovosten prevod potrebno razširiti strokovno in jezikovno znanje z določenega področja. To znanje prevajalci tradicionalno iščejo v knjižnicah, pri ustreznih ustanovah, ki se s področjem ukvarjajo, pri strokovnih svetovalcih in prevajalskih kolegih. Danes lahko dobršen del tega - časovno izredno zamudnega - iskanja opravimo preko interneta, ki hkrati služi kot ogromna knjižnica in komunikacijski medij za vzpostavljanje stikov z viri informacij. Izmed številnih možnosti, ki jih ta medij prevajalcu ponuja, jih naštejmo le nekaj:

### Iskanje izrazja

Na svetovnem spletu je na voljo množica eno- in večjezičnih splošnih in področnih slovarjev, nekaj tudi slovenskih<sup>17</sup>, za tuje jezike pa tudi številne terminološke banke. Obstaja kar nekaj spletnih strani, ki takšne povezave zbirajo, pri nas je ena takšnih domača stran Društva znanstvenih in tehniških prevajalcev Slovenije<sup>18</sup>. Poleg tega številne strokovne ustanove, ministrstva, podjetja ipd. na svojih straneh objavljajo manjše terminološke slovarje za svoje področje ali povezave do njih.

### Iskanje primerljivih besedil

Predvsem kadar prevajamo besedilne vrste, ki nam niso povsem domače, je v veliko pomoč, če imamo na razpolago sorodno besedilo v ciljnem jeziku, iz katerega lahko sklepamo o besedilnih in stilnih konvencijah, ki za določeno besedilno vrsto veljajo, se seznanimo s področjem in izrazjem itd. Internet je neizčrpen vir primerljivih besedil, znati jih moramo le najti, hkrati pa se zavedati, da je njihova kakovost dostikrat vprašljiva.

### Vzpostavljanje stikov z ustanovami, kontaktnimi osebami

Včasih lahko prevajalski problem razreši le strokovnjak z določenega področja. Prek interneta lahko po vsem svetu iščemo ustanove in kontaktne osebe ter jih po elektronski pošti prosimo za nasvet.

17 URL: <http://www.sigov.si/slovar.html>

18 URL: <http://www.drustvo-ztps.si/>

### Poštni sezname in novičarske skupine

Jezikoslovcem in prevajalcem je na voljo več mednarodnih poštnih seznamov (mailing lists) in novičarskih skupin (newsgroups), v okviru katerih je mogoče postavljanje vprašanj in obveščanje o novostih stroke. Prek poštnih seznamov imamo hkrati na voljo rojene govorce vseh mogočih jezikov, ki lahko hitro razrešijo dvome pri prevajanju v tuji jezik. Nekateri poštni sezname imajo tako veliko članov, da lahko na zastavljeno vprašanje pričakujemo več deset odgovorov že v roku ene ure.

### Spremljanje stroke in tehnološkega razvoja

Prek interneta lahko prevajalci prebirajo strokovne publikacije in revije, izvejo za dogodke in novosti na področju stroke in spremljajo razvoj na področju jezikovnih tehnologij. Pogosto je nove programe mogoče preskusiti ali si naložiti preskusne različice.

### Trženje lastnih storitev na internetu

Oglaševanje preko interneta je v tujini že nekaj let ena najbolj cvetočih dejavnosti, pri nas pa to postaja šele v zadnjem času. Podobno kot je danes za vsakega ponudnika storitev samoumevno, da je njegova telefonska številka v imeniku, bo čez nekaj let verjetno veljalo za internetno prisotnost. Predstavljanje svojih znanj in dejavnosti na internetu ima tudi to prednost, da doseže morebitne naročnike izven meja Slovenije, prevajanje pa sodi med dejavnosti, ki jih je v celoti mogoče opravljati na daljavo.

## ZAKLJUČEK

V Sloveniji si prevajalske tehnologije le počasi utirajo pot v prakso, tako da je položaj težko presojati na podlagi konkretnih izkušenj. Kar se tiče programskih orodij, se ta postopoma uvajajo v proces prevajanja slovenske zakonodaje in drugih meddržavnih dokumentov na prevajalskem oddelku Službe Vlade RS za evropske zadeve (Krstič, 1998), vse več pa jih uporabljajo tudi prevajalske agencije. Nekoliko počasneje se zanje odločajo velika izvozno usmerjena podjetja, pri katerih tudi narava prevajalskega dela ni vselej primerna za uvajanje teh orodij. Po drugi strani je Slovenija očitno premajhno tržišče, da bi proizvajalci prevajalskih programov te dejavno tržili pri nas, tako da velika večina prevajalcev v podjetjih, pa tudi samostojnih, ne pozna prednosti, ki jih ti programi lahko nudijo.

Računalniško podprto prevajanje v širšem smislu od prevajalca zahteva splošno računalniško in informacijsko osveščenost, ki zajema različne spretnosti in znanja od urejevalnika besedil, elektronskih slovarjev in interneta do učinkovite izrabe drugih

jezikovnih virov, kot so korpusi. Pri razvijanju te osveščenosti imajo pomembno vlogo na eni strani študijski programi za prevajalce, na drugi pa pospešena gradnja in zagotavljanje osnovnih jezikovnih virov in tehnologij v slovenskem prostoru.

Če se na kratko posvetimo prvim, študij prevajanja pri nas nudi pred dvema letoma ustanovljeni Oddelek za prevajanje in tolmačenje na Filozofski fakulteti v Ljubljani. Znanja s področja prevajalskih tehnologij si študenti pridobijo v okviru predmeta Prevajalska orodja,<sup>19</sup> ki mu je v drugem letniku študija namenjenih 30 ur, v tem času pa skušamo študentom posredovati vsaj površen vpogled v področje. Večina prevajalskih institutov po Evropi področju jezikovnih tehnologij namenja večjo pozornost, takšna pa so tudi priporočila Evropske komisije, ki so bila izoblikovana v okviru projekta LETRAC<sup>20</sup> (Language Engineering for Translator Curricula). V skladu s temi naj bi v okviru izobraževanja prevajalcev jezikovnim tehnologijam namenili vsaj 10 odstotkov vseh ur po predmetniku, poleg tega pa naj bi predmetniki vsebovali še možnost izbirnih predmetov za poglobljeno spoznavanje področij, kot so strojno prevajanje, računalniško jezikoslovje, terminologija, korpusno jezikoslovje, osnove programiranja za jezikoslovce, programska orodja za lokalizacijo itd. Na Oddelku za prevajanje in tolmačenje si prizadevamo, da bi bilo vsaj nekatere od omenjenih vsebin v naslednjih dveh letih mogoče ponuditi tudi tukajšnjim bodočim prevajalcem.

Tudi izgradnji jezikovnih virov in tehnologij se v zadnjih letih pri nas posveča precej pozornosti. V teku je projekt FIDA<sup>21</sup> (Erjavec et al, 1998), v okviru katerega nastaja prvi referenčni korpus za slovenski jezik, Institut Jožef Stefan pa se vključuje v različne evropske projekte, v sklopu katerih se gradijo večjezični jezikovni viri - vzporedni korpusi in leksikoni (MULTEXT-East,<sup>22</sup> ELAN<sup>23</sup>, CONCEDE<sup>24</sup>). Nad področjem bo odslej bdelo pred kratkim ustanovljeno Slovensko društvo za jezikovne tehnologije<sup>25</sup>.

Slovenija se vključuje v evropske politične in gospodarske tokove, z njimi pa naraščajo tudi potrebe po prevajanju in hkrati njegovi optimizaciji. V Evropi je postalo uvajanje, razvoj, standardizacija in evalvacija prevajalskih tehnologij ena od prioritet v procesu evropske integracije, to pa bo v naslednjih letih - če želimo nadoknaditi zamujeno - še toliko bolj veljalo za slovenski prevajalski prostor.

## VIRI

1. Arntz, R. in Picht, H. (1995) *Einführung in die Terminologearbeit*. Hildesheim, Zürich, New York: Georg Olms.
2. Brungs, B. (1996) *Translation Memories als Komponente integrierter Übersetzungssysteme*. Saarbrücker Studien zu Sprachdatenverarbeitung und Übersetzen. Saarbrücken: Universität des Saarlandes.
3. Erjavec, T. (1997) *Računalniške zbirke besedil*. *Jezik in slovstvo*, 42/2-3, str. 81-96. <http://nl.ijs.si/et/Bib/SIKorpus/siKorpus-la2/>
4. Erjavec, T. et al (1998) *Korpus Fida*. V: Erjavec, T., Gros, J. (ur.) *Zbornik konference Jezikovne tehnologije za slovenski jezik, v sklopu simpozija Informacijska družba '98*. Ljubljana: Institut Jožef Stefan.
5. Esselink, B. (1998) *A Practical Guide to Software Localization*. Amsterdam, Philadelphia: John Benjamins.
6. Falcone, S. (1998) *Translation Aid Software*. *Translation Journal* 2 (1). <http://www accurapid.com/journal/03TM2.htm>
7. Heyn, M. (1997) *Present and future needs in the CAT-world*. <http://www.trados.ch/english/pressrel.htm>
8. Heyn, M. (1998) *Integrating machine translation into translation memory systems*. <http://www.trados.ch/english/pressrel.htm>
9. Hirci, N. (1998) *Korpusi v prevodoslovju*. V: Erjavec, T., Gros, J. (ur.) *Zbornik konference Jezikovne tehnologije za slovenski jezik, v sklopu simpozija Informacijska družba '98*. Ljubljana: Institut Jožef Stefan.
10. Holloway, T. (1996) *Translation Memory Software*. *ITI Bulletin*, August 1996, str. 16-27.
11. Hutchins, W. J.; Somers, H. L. (1992) *An Introduction to Machine Translation*. London: Academic Press.
12. Krstič, A. (1998) *Problematika prevajanja zakonodaje Evropske unije*. V: Erjavec, T., Gros, J. (ur.) *Zbornik konference Jezikovne tehnologije za slovenski jezik, v sklopu simpozija Informacijska družba '98*. Ljubljana: Institut Jožef Stefan.
13. Pearson, Jennifer (1998) *Terms in Context*. Amsterdam, Philadelphia: John Benjamins.

19 URL: <http://www2.ames.si/čsvinta/index.htm>

20 URL: <http://www.iai.uni-sb.de/LETRAC/home.html>

21 URL: <http://www.fida.net>

22 URL: <http://nl.ijs.si/ME/>

23 URL: <http://solaris3.ids-mannheim.de/elan/>

24 URL: <http://nl.ijs.si/tomaz/Concede.html>

25 URL: <http://nl.ijs.si/sdjt/>

14. Reinke, U. in Schmitz, K.-D. (1998)  
*Testing the Machine Readable Terminology Interchange Format (MARTIF).*  
Saarbrücker Studien zu Sprachdatenverarbeitung und Übersetzen. Saarbrücken: Universität des Saarlandes.
15. Romih, M. (1998)  
*Amebis in jezikovne tehnologije.*  
V: Erjavec, T., Gros, J. (ur.)  
Zbornik konference Jezikovne tehnologije za slovenski jezik, v sklopu simpozija Informacijska družba '98.  
Ljubljana: Institut Jožef Stefan.
16. Schubert, K. (1998)  
*Zur Automatisierbarkeit des Übersetzens.*  
Zbornik kongresa Modelle der Übersetzung - Grundlagen für Methodik, Bewertung, Computermodellierung, Saarbrücken, november 1998.  
Saarbrücken: Universität des Saarlandes.
17. Seybold, M. (1995)  
*Terminologieverwaltung unter Windows - Eine vergleichende Untersuchung.* Saarbrücker Studien zu Sprachdatenverarbeitung und Übersetzen.  
Saarbrücken: Universität des Saarlandes.
18. Sinclair, J. (1991)  
*Corpus, Concordance, Collocation.* Oxford: Oxford University Press.
19. Skubic, A. (1997)  
*Računalniški programi za prevajanje.*  
Mostovi XXXI, str. 28-37.
20. Spies, C. (1995)  
*Vergleichende Untersuchung von integrierten Übersetzungssystemen mit Translation-Memory-Komponente.* Saarbrücker Studien zu Sprachdatenverarbeitung und Übersetzen.  
Saarbrücken: Universität des Saarlandes.
21. Vintar, Š.; Erjavec, T. (1998)  
*Prednosti in omejitve programov s pomnilnikom prevodov.* Zbornik II. kongresa „Jezik za danes in jutri“, Društvo za uporabno jezikoslovje Slovenije.

◆  
 Špela Vintar je diplomirala iz nemščine in angleščine na Filozofski fakulteti v Ljubljani z diplomskim delom o uporabnosti programov s pomnilniki prevodov. Od oktobra 1998 je zaposlena na Oddelku za prevajanje in tolmačenje kot asistent stažist in poučuje predmet Prevajalska orodja.  
 ◆