



NARAVOSLOVNOTEHNIŠKA FAKULTETA

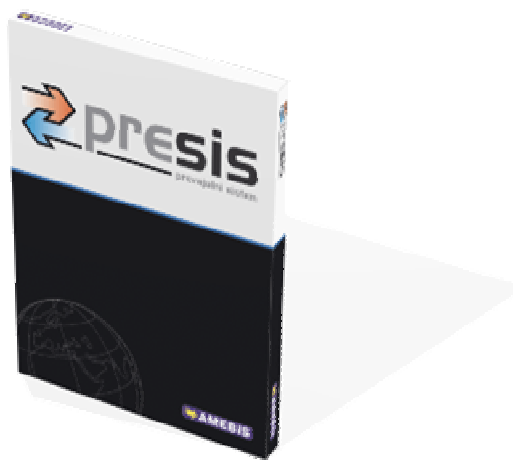
ODDELEK ZA TEKSTILSTVO

GRAFIČNA TEHNOLOGIJA

SEMINARSKA NALOGA PRI PREDMETU

JEZIKOVNE TEHNOLOGIJE

PREVAJALNI SISTEM AMEBIS PRESIS



Mentor:

mag. Andrej ISKRA

Avtorici:

Nataša AHAČIČ

Silva GRILJ

Ljubljana, maj 2006

KAZALO

1 UVOD	2
2 STROJNO PREVAJANJE	3
2.1 Slovenščina	4
3 PREDSTAVITEV PODJETJA AMEBIS	5
4 PREVAJALNI SISTEM AMEBIS PRESIS	6
4.1 O Presisu	6
4.1.1 Prevajalni sistem je prilagojen uporabniku	7
4.2 Paketi	8
4.2.1 Osebni paket	8
4.2.2 Poslovni paket	9
4.2.3 Mrežni paket	10
4.2.4 Primerjava paketov glede na funkcionalnost	11
4.3 Uporabniški vmesniki	12
4.3.1 Amebis Presis Mini	12
4.3.2 Amebis Presis za Internet Explorer	13
4.3.3 Amebis Presis Pro	14
4.4 Delovanje	16
4.4.1 Korpusi	16
4.4.2 Osnovni koraki prevajanja	17
4.4.3 Uporaba glagolskih predlog	18
4.4.4 Internetna in časovno omejena predstavljena verzija	20
4.4.5 Napotki pri prevajanju	20
4.5 Ustreznost prevajanja	22
4.5.1 Primeri prevajanja	22
4.5.2 Analiza prevodov	24
4.5.3 Ugotovitve po analizi	26
5 ZAKLJUČEK	27
6 LITERATURA	28

1 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, zato že dolgo predstavlja 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.

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.

V seminarski nalogi sva se osredotočili na prevajalni sistem. In sicer sva želeli pogledati, kako je na področju strojnega prevajanja na slovenski strani. Tako sva naleteli na Amebis, ki je, kar se strojnega prevajanja tiče, v Sloveniji vodilni. V seminarski nalogi sva pod drobnogled vzeli njihov prevajalni sistem Presis. Skušali sva dobiti sliko, kako prevajalni sistem deluje in tudi, kako je sestavljen. Omenjava tudi korpuse, ki so najpomembnejši del vsakega prevajalnega sistema. Na koncu pa sva želeli teorijo združiti še s prakso. Tako sva opisali še najine osebne izkušnje s prevajalnim sistemom s praktičnega vidika.

2 STROJNO PREVAJANJE

Stojno prevajanje postaja eno izmed pomembnejših področij jezikovnih tehnologij. Velika večina prevajalnih sistemov je namenjena predvsem podpori prevajanju, kar pomeni, da so le v pomoč prevajalcu pri njegovem delu, še vedno pa je prevajalec tisti, ki besedilo dokončno prevede. Ob izredno hitrem napredku strojne in programske opreme pa nekateri prevajalni sistemi postajajo tako dobri, da lahko govorimo o izključno strojnem prevajanju, saj poseg prevajalca pogosto ni več potreben.

Pod izrazom **strojno podprto prevajanje** razumemo prevajanje, ki ga v prvi vrsti izvaja prevajalec ob podpori različnih računalniških orodij, med katerimi igrajo pomembno vlogo pomnilniki prevodov. Ti so jezikovno neodvisni, delujejo pa tako, da prevajalcu poskušajo pomagati na podlagi prevodov celih stavkov, ki jih je treba prej vnesti v prevajalno bazo. Za splošno prevajanje je taka vrsta programa precej neuporabna, se pa toliko bolj uporabljajo na omejenih področjih, kjer se določeni stavki precej ponavljajo (npr. navodila za uporabo, zakonodaja). Pomnilniki prevodov v primeru, da konkretnega stavka ni v prevajalni bazi, predlagajo prevod, ki je po obliki oz. po uporabljenih besedah najbližji. Prevajalec mora prevod ustrezno spremeniti, kar pa je še vedno hitreje, kot če bi moral stavek prevesti v celoti.

Strojno prevajanje poteka brez pomoči prevajalca. Taki sistemi so jezikovno odvisni, njihov razvoj pa zahteva izredno veliko znanja in časa. Brez poprejšnjega učenja so sposobni prevesti poljubno besedilo. V splošnem njihov prevod ni vedno najboljši, zato so pregledi in popravki s strani uporabnika še vedno zaželeni.

Prevajalne sisteme s področja strojnega prevajanja lahko razvrstimo po načinu dela na:

- prevajalne sisteme, ki le menjajo besedno obliko za besedno obliko,
- prevajalne sisteme, ki znajo vhodno besedo lematizirati (poiščejo osnovno obliko besede),
- prevajalne sisteme, ki poleg morfologije upoštevajo tudi sintakso obeh jezikov in
- prevajalne sisteme, ki prevajajo na podlagi morfološke, sintaktične in pomenske analize.

Zadnji prevajalni sistemi dajejo najboljše rezultate. Da bi pravilno razumeli stavek, del informacij prenašajo iz prejšnjih stavkov, saj se pogosto zgodi, da se kak stavek nanaša na prejšnje. Osnovna enota prevajanja so tako odstavki.

2.1 Slovenščina

Prevajanje iz enega jezika v drugega ni bil problem zgolj za jezikoslovce, temveč tudi za matematike, statistike, informatike in druge. Osnova vsega so obsežne zbirke besedil, t.i. **korpusi**, ki vsebujejo poved v izvirnem jeziku in istočasno njen prevod. Računalniki ugotavljajo pojavljanje posameznih besed v kombinaciji z drugimi in izračunavajo verjetnost tovrstnega ponavljanja.

Slovenščina je jezik s posebnimi značilnostmi. Je visoko **pregiben** jezik, torej z velikim številom končnic besed z besednim repom, ki ni natančno določen. Ob tem imamo še slovensko značilnost, **dvojino**, ki je posebnost med večino slovanskih jezikov. Večina samostalnikov se lahko tvori **edninsko**, **dvojinsko** ter **množinsko**, in to v **šestih sklonih**. Večina pridevnikov lahko tvori tri spole, vsa tri števila, šest sklonov in tri osnovne ravni stopnjevanja, ob tem pa ima slovenski jezik še mnoga izpuščanja.

Sodobni in zmogljivi računalniki bi zagotovo bili kos tem zahtevam, vendar se osnovnim jezikovnim pravilom sintakse, gramatike, sinonimom besed oz. fraz priključuje še zapletena **semantika** oz. pomen prevajanih sporočil. Pot do uspešnega samodejnega oziroma strojnega prevajanja bo zagotovo še dolgotrajna.

3 PREDSTAVITEV PODJETJA AMEBIS



Podjetje Amebis, ki se nahaja v Kamniku, je bilo ustanovljeno v začetku devetdesetih let. Ukvarja se predvsem z razvijanjem programske opreme na področju jezikovnih tehnologij in elektronskega založništva, kjer tudi velja za vodilno. Znano pa je tudi po elektronskih slovarjih, leksikonih, črkovalnikih, kakor tudi po raznih podatkovnih zbirkah in drugih sorodnih izdelkih.

Med njihove največje novosti na trgu trenutno spadajo:

- ***Klepec*** – Program, ki je nastal v okviru projekta Kolos in služi kot univerzalni uporabniški vmesnik za komunikacijo z računalnikom v naravnem jeziku.
- ***Srbski- ali Hrvaško-slovenski in Slovensko-hrvaški ali -srbski slovar***. Avtor obeh slovarjev je Janko Jurančič, založnik založba DZS, d.d., programski del pa je delo Amebisa.
- ***Veliki splošni leksikon*** - Gre za najobsežnejšo slovensko zbirko na elektronskem mediju (okoli 26 milijonov znakov ali 5 milijonov besed).
- ***Programski paket BesAna*** – Slovnični pregledovalnik za slovenski jezik (samostojen program za operacijski sistem Windows in vgradni modul za Microsoft Office).

Prelomen trenutek za podjetje, še posebej pa za Slovenijo je bil, ko so razvili slovnični pregledovalnik za slovenski jezik – *BesAna* in pa vedno bolj uveljavljeni prevajalni sistem *Presis*.

4 PREVAJALNI SISTEM AMEBIS PRESIS



4.1 O Presisu

Prevajalni sistem Presis je namenjen strojnemu prevajanju besedil, za zdaj sta izdelana prevajalnika med slovenščino in angleščino in obratno. Zaposleni pri Amebisu pa namigujejo, da se v ozadju že pripravljata tudi nemška modula. Vendar zaradi obsežnosti projekta zagotovo še ne bosta na voljo v letu 2006.

Presis je zgrajen iz prevajalnega jedra, kamor se vključujejo prevajalni moduli za posamezne smeri prevajanja in odjemalcev oziroma uporabniških vmesnikov.

Odjemalci so na voljo le kot namestitvene aplikacije, čeprav Amebis na svoji spletni strani ponuja prevajalnik tudi kot spletno aplikacijo (<http://presis.amebis.si/>), vendar le v predstavitvene namene.

V Amebisu so Presis razvili s programskim orodjem Microsoft Visual Studio ter programskim jezikom C++, za povezovanje odjemalcev in strežnika pa so uporabili predmete COM. Presis je za nameščanje nezahteven, saj potrebuje le osebni računalnik, ki poganja operacijske sisteme Microsoft Windows 98, ME 2000 ali XP. Najmanjše strojne zahteve obsegajo procesor Intel Pentium, 64 MB RAM-a, pogon CD-ROM, disk z vsaj 150 MB prostora ter miško ali drugo kazalno napravo. Omrežna različica programa ob tem zahteva še podporo za omrežja NetBIOS oziroma omrežne strežnike Microsoft Networking, Novell Netware 5 ali Samba. Program ponuja tudi vtičnik za brskalnik Microsoft Internet Explorer 5.5 ali katero od novejših različic. V Amebisu lahko po dogovoru z naročnikom izdelajo vtičnike tudi za druge programe.

Pri gradnji slovarjev in prevajalne baze si v Amebisu pomagajo s korpusom Fida (ki je pravzaprav njihov). Kot pravijo, trenutno pripravljajo celo FIDO plus, ki naj bi bila še boljša. Oba prej omenjena modula (slo-ang in ang-slo) pri Amebisu redno nadgrajujejo. Tako program, kot trenutno ocenjujejo, pozna že približno 200.000 besed v različnih oblikah (tako na slovenski kot angleški strani). Vendar je prevodov pri slednji manj. Prevodov je približno

50.000. Kar navsezadnje ni tako malo, sploh ob dejstvu, da slovar nastaja po pogostosti in sproti. Torej je jezik precej »živ«, pravijo pri Amebisu.

Amebis Presis je pri prevajanju hiter in učinkovit, saj namesto uporabnika opravi celo vrsto rutinskih opravil. Nudi tudi dodatne nasvete, ki pomagajo pri oblikovanju končnega prevoda. Za razliko od ostalih klasičnih prevajalskih orodij, pa tudi upošteva slovnična pravila slovenskega in drugih jezikov. Zaradi vgrajenega pomnilnika prevodov je primeren za različne skupine uporabnikov.

Vendar pa je kljub vsemu nemogoče pričakovati, da bo Amebis Presis vsako besedilo vedno in v celoti pravilno prevedel, kajti jezik in pomenski odtenki se stalno spreminjajo, ravno tako pa tudi želje uporabnikov.

4.1.1 Prevajalni sistem je prilagojen uporabniku

Prevajalni sistem Presis ponuja tri uporabniške odjemalce oziroma vmesnike: Mini, Pro in IE, ki ponujajo različne funkcionalnosti in način dela. Na voljo je v treh paketih, Osebni, Poslovni in Mrežni. Osebni paket je namenjen fizičnim osebam ter vključuje vmesnika Presis Mini in Presis IE, medtem ko sta poslovni ter omrežni paket namenjena pravnim osebam in ponujata vse tri uporabniške vmesnike.

4.2 Paketi

Pri Amebisu obstajajo tri vrste paketov prevajalnega sistema Presis:

- Osebni
- Poslovni
- Mrežni paket

4.2.1 Osebni paket

Namenjen je le fizičnim osebam, zato ga pravne osebe ne morejo kupiti, niti ne smejo uporabljati v poslovne namene. Za uporabo je enostaven in nudi osnovno pomoč pri prevajanju besedil. Z njim se lahko prevaja poljubno dolga besedila, ki se jih običajno preko odložišča prenese iz katerega koli programa, pomaga pa tudi pri prevajanju spletnih strani. Vsebovana programska vmesnika sta: *Amebis Presis Mini* in *Amebis Presis za Internet Explorer*.



Sistemske zahteve:

- računalnik, združljiv z IBM modeli PC (procesor Pentium ali novejši)
- operacijski sistem Windows 98, ME, 2000 ali XP
- vsaj 64 MB pomnilnika
- enota CD-ROM
- trdi disk z vsaj 100 MB praznega prostora
- miška ali druga podobna vhodna enota
- Internet Explorer 5.5 ali novejši (opcija)

4.2.2 Poslovni paket

Namenjen je tako fizičnim kot pravnim osebam. Omogoča nekatere dodatne funkcije, ki jih osebni paket nima. Med najpomembnejšimi je predvsem možnost dodajanja novih prevodov besed in stavkov v osebni slovar, kar lahko zelo vpliva na kakovost prevajanja.

Vsebovani programski vmesniki so: *Amebis Presis Mini*, *Amebis Presis za Internet Explorer* in *Amebis Presis Pro*.



Sistemske zahteve:

- računalnik, združljiv z IBM modeli PC (procesor Pentium ali novejši)
- operacijski sistem Windows 98, ME, 2000 ali XP
- vsaj 64 MB pomnilnika
- enota CD-ROM
- trdi disk z vsaj 100 MB praznega prostora
- miška ali druga podobna vhodna enota
- Internet Explorer 5.5 ali novejši (opcija)

4.2.3 Mrežni paket

Namenjen je predvsem tistim pravnim osebam, ki želijo z minimalnim vzdrževanjem ponuditi možnost uporabe prevajalnika vsem uporabnikom v mreži. V osnovi omogoča uporabo le enemu (poljubnemu) uporabniku v mreži, z dokupom dodatnih licenc pa ga lahko uporablja tudi več uporabnikov hkrati.

Vsebovani programski vmesniki: *Amebis Presis Mini*, *Amebis Presis za Internet Explorer* in *Amebis Presis Pro*.



Sistemske zahteve:

- računalnik, združljiv z IBM modeli PC (procesor Pentium ali novejši)
- operacijski sistem Windows 98, ME, 2000 ali XP
- vsaj 64 MB pomnilnika
- enota CD-ROM
- trdi disk z vsaj 100 MB praznega prostora
- miška ali druga podobna vhodna enota
- Internet Explorer 5.5 ali novejši (opcija)
- omrežje NetBIOS (Microsoft Networking, Novell Netware 5 ali Samba)

4.2.4 Primerjava paketov glede na funkcionalnost

Funkcije:	Osebni paket	Poslovni paket	Mrežni paket
prevajanje besedil	da	da	da
nasveti o možnih prevodih	da	da	da
možnost neposrednega urejanja prevoda	da	da	da
vgrajen črkovalnik	-	da	da
barvno označevanje tipa prevoda	-	da	da
vgrajena povezava s slovarji ASP32	-	da	da
dodajanje prevodov besed	-	da	da
dodajanje prevodov stavkov	-	da	da
možnost dodatnega rednega osveževanja	-	da	da
večuporabniško delovanje na mrežnem strežniku	-	-	da

4.3 Uporabniški vmesniki

Prevajalni sistem Amebis Presis omogoča delo s tremi uporabniškimi vmesniki. Vsi trije uporabljajo isto prevajalno jedro, zato se prevodi med uporabniškimi vmesniki ne razlikujejo. Različna je le njihova uporabnost in načini dela.

4.3.1 Amebis Presis Mini

Osnovni programski vmesnik za pomoč pri prevajanju. Odlikujejo ga predvsem enostavnost in hitrost dela, prilagodljivost uporabniškega vmesnika ter odlično sodelovanje z drugimi programi preko odložišča.

Ko poženemo Amebis Presis Mini, se nam prikaže v obliki enovrstične orodne vrstice nad sistemsko orodno vrstico, ki vsebuje naslednje osnovne gumbe in polja:

1. Menu
2. Gumb za brisanje
3. Vnosno polje
4. Gumb za prevajanje
5. Gumb za shranjevanje vsebine prevodnega polja v odložišče
6. Polje s prevodom



Vmesnik omogoča, da imamo slovar stalno aktiven na namizju, pri čemer ne moti običajnega dela. Če nam ena vrstica pri prevajanju ne zadostuje, si lahko velikost okna poljubno prilagodimo.

Amebis Presis Mini se v osnovi uporablja tako, da v vnosno polje ročno vpišemo ali preko odložišča prenesemo iskano besedo, frazo, stavek ali besedilo in sprožimo prevajanje. Če v meniju vključimo spremljanje odložišča, Amebis Presis Mini vsako spremembo v odložišču avtomatsko prevede.

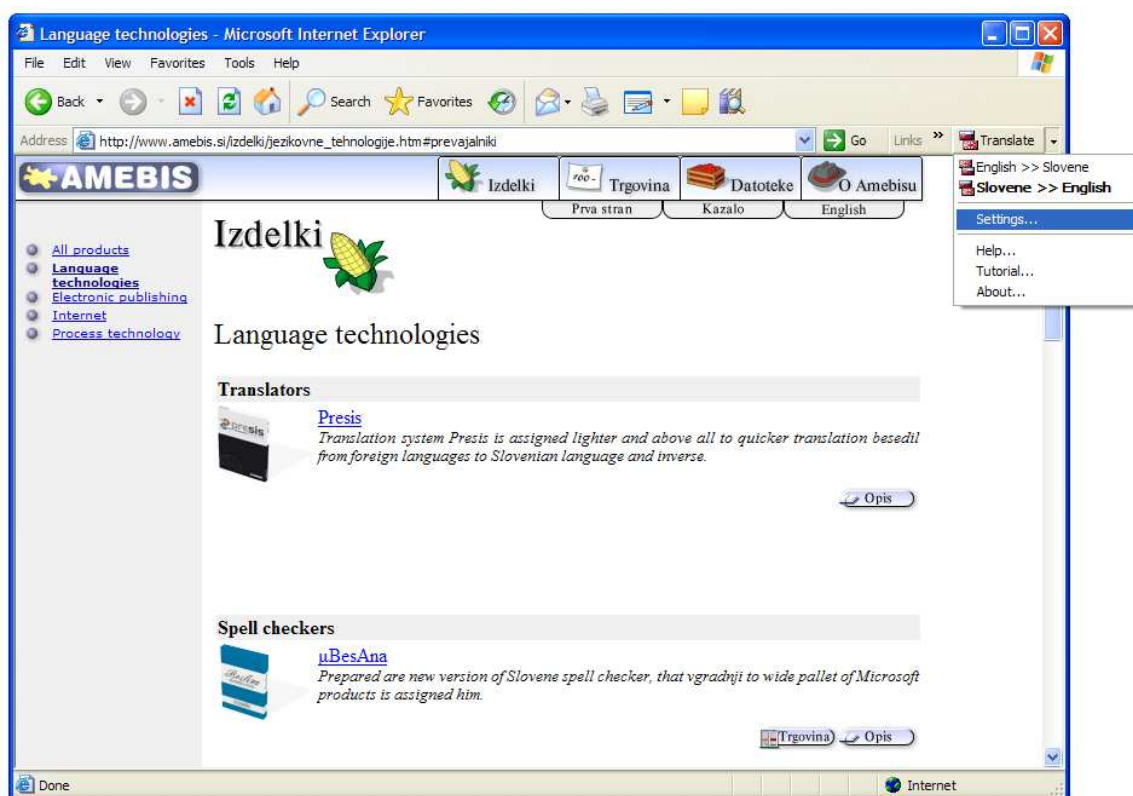
Amebis Presis Mini lahko uporabljamo tudi kot slovar, pri čemer nam besed ni potrebno vpisovati v osnovni obliki, saj prevajalnik upošteva tudi slovnico. Ker se število prevodov postopoma povečuje in ker prevajalnik lahko prevede tudi posamezno besedo, je uporabnost prevajalnika kot slovarja na dlani.

4.3.2 Amebis Presis za Internet Explorer

Programski vmesnik je namenjen prevajanju spletnih strani v pregledovalniku Internet Explorer. Za prevod trenutno prikazane spletne strani zadostuje pritisk na gumb in prevedla se bodo vsa besedila.

Po namestitvi se v meniju View/Toolbars doda vrstica Presis. Če želimo prevajalnik uporabljati, ga moramo najprej vključiti.

S puščico na desni strani gumba prikličemo meni, v katerem lahko nastavimo smer prevajanja, spremenimo nekatere nastavitve, prikličemo pomoč ali informacije o programu in uporabniku.



S pritiskom na gumb sprožimo prevajanje spletne strani, ki je trenutno prikazana. Čas prevajanja je odvisen od hitrosti računalnika ter dolžine in zahtevnosti spletne strani. Po nekaj trenutkih se prikaže prevedena stran. Pri tem pa ostanejo aktivne vse povezave, kar pomeni, da nas Amebis Presis ne ovira pri nadaljevanju brskanja.

4.3.3 Amebis Presis Pro

Najmočnejše orodje prevajalnega sistema Amebis Presis. S pomočjo nekaterih dodatnih funkcij nudi še večjo pomoč pri prevajanju, zato je delo z njim še udobnejše in hitrejše. Edini omogoča dodajanje besed in stavkov. Ta uporabniški vmesnik je na voljo samo v poslovnem paketu.

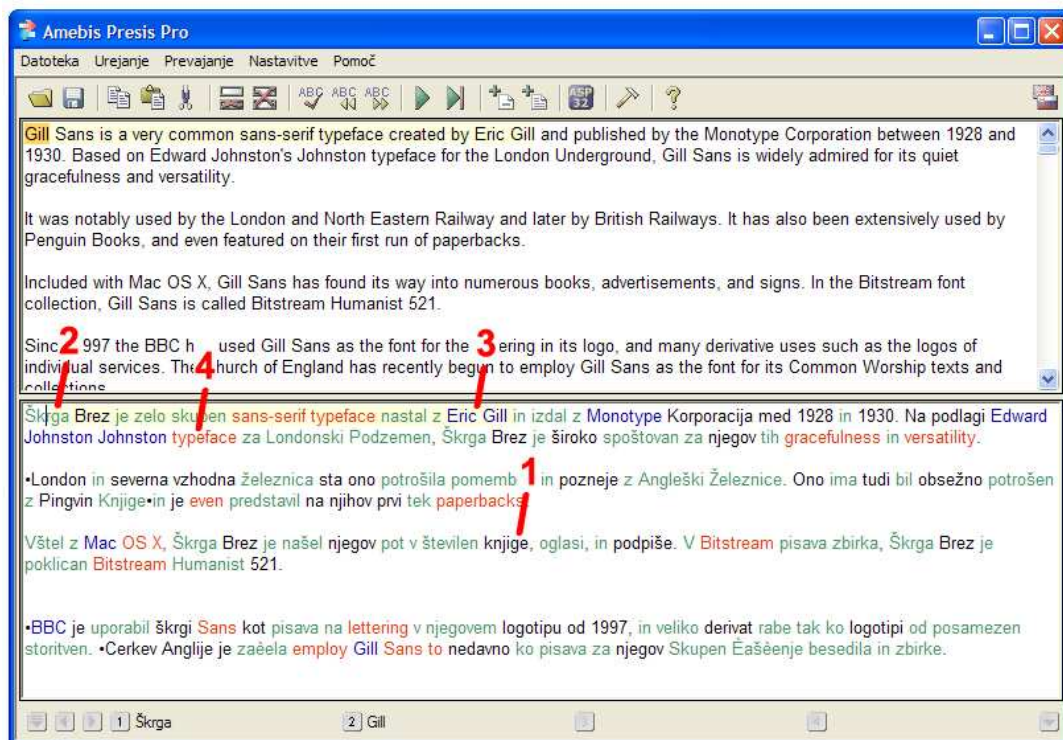
Besedilo, ki ga želimo prevesti, lahko vtipkamo v vnosno polje, lahko ga prenesemo s pomočjo odložišča, ali pa odpremo datoteko v formatu .txt ali .rtf. Pred prevajanjem lahko z vgrajenim črkovalnikom preverimo pravilnost zapisa in po potrebi popravimo napake, saj te lahko poslabšajo prevod.

Prevajamo lahko po stavkih, pri čemer naenkrat vidimo le en stavek. Ko ga s pomočjo programa in njegovih nasvetov dokončno prevedemo, skočimo na prevajanje naslednjega stavka. Z ustreznim ukazom lahko sprožimo tudi avtomatsko prevajanje celotnega besedila. V tem primeru se po določenem času, ki je odvisen predvsem od hitrosti računalnika ter dolžine besedila, v prevodno okno izpiše prevod celotnega besedila. Tipična hitrost prevajanja je nekaj 10 strani besedila formata A4 v minuti.*

* Hitrost prevajanja sva preverili. Amebis Presis Pro je prevajal točno 10 strani A4 formata malo več kot minuto.

Amebis Presis Pro z barvami označi štiri tipe prevedenih besed:

1. Besede z enim znanim prevodom
2. Besede z več možnimi prevodi
3. Nprevredena lastna imena
4. Neznane besede



S pomočjo barvnega označevanja lahko hitreje opazimo napake in jih zato lažje in hitreje ustrezno popravimo.

Prevod lahko poljubno spreminjamo, pri čemer nam program pomaga z nasveti. Pri besedah z več možnimi prevodi se ti nasveti pojavijo avtomatsko v trenutku, ko pridemo s kazalcem na določeno besedo ali pa jih priključimo s pritiskom na desni miškin gumb, ki odpre ustrezno okence.

Če je na računalniku nameščen en ali več elektronskih slovarjev iz sistema ASP32, nam delo olajša tudi vgrajena povezava z njimi. Če smo s kazalcem na neki besedi v enem izmed obeh oken in pritisnemo ustrezen gumb, se nam odpre izbrani slovar in avtomatsko prikaže prevod besede v njem.

Če program ne prevede določene besede oz. zanjo nima ustreznega prevoda, lahko prevod (v osebni slovar) dodamo sami. Prav tako lahko v pomnilnik prevodov shranimo cele stavke, ki se pogosto pojavljajo v naših besedilih. Tako si lahko z nekaj vnosi, ki pokrivajo naše pogoste izraze, zelo izboljšamo kakovost prevoda.

4.4 Delovanje

4.4.1 Korpusi

Korpus: Zbirka kosov jezika, ki so izbrani in urejeni po eksplicitnih jezikoslovnih kriterijih, z namenom, da služijo kot vzorec jezika.

Računalniški korpus: korpus, ki je zapisan na standarden in enovit način za poljubne iskalne naloge.

Korpusi so zbirke besedil, urejene po določenih pravilih in pretvorjene v enotno obliko zapisa, ter ustreznih programov, s katerimi preiskujemo besedila in nad njimi izvajamo določene statistične operacije. Pri jezikoslovnem raziskovalnem delu se je v zadnjih nekaj letih izkazalo, da so besedilni korpusi (še zlasti, če so obsežni in pravilno uravnoteženi), edini pravi pokazatelj dejanskega stanja jezika.

Sodobni besedilni korpusi so zelo veliki in navadno obsegajo kar nekaj sto milijonov besed. Gradnja tako obsežnih korpusov je zahtevna in dolgotrajna. Potrebna je organizacija pridobivanja ustreznih besedil. Dobljena besedila je treba iz različnih vhodnih formatov pretvoriti v en sam končni format zapisa, ter ob tem v glavo besedila zapisati osnovne podatke o izvoru. Zaradi različnega namena uporabe obstaja več različnih vrst korpusov (zgodovinski, govorni, referenčni, totalni, ...). Temu primerno morajo prilagoditi tudi izbiro materiala in način gradnje.

Gradnja korpusa:

Zbrane materiale je potrebno na podlagi določenih pravil uravnotežiti. Za uravnoteženje se uporabljajo določena razmerja glede lektorske obdelave (lektorirano, nelektorirano), izvora (dnevniki, tedniki, ...) in zvrsti besedil (drame, leposlovje, proza, ...), tako da ti predstavljajo pravo sliko jezika.

Pri gradnji korpusa je običajno še oblikoslovno označevanje besed, nazadnje pa je treba korpus spraviti v obliko, ki je primerna za hitro iskanje in različne statistične operacije nad njim.

Korpusi so v glavnem nameščeni na strežnikih in so uporabnikom dostopni preko spleta. Trenutno je v Sloveniji najbolj uravnotežen in največji besedilni korpus slovenskega jezika korpus FIDA, ki vsebuje dobrih 100 milijonov besed.

4.4.2 Osnovni koraki prevajanja

Prevajanje se izvaja v treh osnovnih korakih:

1. Najprej se z **analizo** poskuša čim bolj natančno ugotoviti, kakšen je pomen posameznih besed v izvornem stavku ter kako so te med seboj pomensko povezane. Analizator mora najti in upoštevati vse mogoče pomenske in sintaktične možnosti.

Primer: *To je Miro Cerar*. Pomen stavka si lahko razlagamo, kot da je to oseba moškega spola z imenom in priimkom Miro Cerar. Vendar pa si isti stavek lahko razlagamo še v dveh drugih pomenih, ki nista običajna, vendar sintaktično in pomensko povsem možna. Stavek lahko razumemo tudi tako, kot da to (nekaj) je žensko z imenom in priimkom Mira Cerar, ali pa, kot da oseba z imenom Miro Cerar je to (nekaj).

Če bi imeli le sintaktični analizator, potem so vse tri pomenske možnosti sintaktično pravilne in enakovredne. Šele z dodatno pomensko analizo, ki si pomaga s statističnimi informacijami, pridobljenimi s pomočjo analize posameznega jezika (korpusa), ter informacijami pridobljenimi na podlagi prejšnjih stavkov, lahko te možnosti utežimo po verjetnosti ter s tem določimo najverjetnejši prevod.

2. Po izluščenju določenih pomenov posameznih besed ali fraz v izvornem stavku sledi **pretvorba** osnovnih oblik teh besed v besede in fraze ciljnega jezika. Pri tem si prevajalnik običajno pomaga z vgrajenim splošnim slovarjem. Ker ima lahko določena beseda ali fraza več možnih prevodov, je treba vse prenesti najprej do tretjega modula, ki se potem odloči za pravi prevod. Poleg osnovnega splošnega slovarja ima prevajalnik lahko še dodatne terminološke slovarje, osebne slovarje uporabnika ali vzorce prevodov celih stavkov, katerih naloga je čim bolje in s čim več možnimi besedami prevesti v ciljni jezik glede na področje, s katerega je prevajano besedilo. Vsako področje ima namreč svoje zakonitosti tvorjenja stavkov in specifične prevode določenih besed.

3. Ko je prevajalnik opremljen s podatki o pomenih posameznih besed in njihovih možnih prevodih, mora opraviti še **sintezo**, ki oblikuje stavek po slovničnih pravilih ciljnega jezika. Poleg tega se mora odločiti za najboljši prevod ustreznega pojma. Pri tem lahko preprosto vrne tisti prevod, ki je v splošnem (statistične analize) najpogostejši, lahko pa na podlagi drugih besed v stavku (statistična analiza kolokacij) določi prevod, ki je v primeru konkretnega stavka verjetno boljši. Vsekakor ima statistična analiza obeh jezikov, tako izvirnega kot ciljnega, bistveno vlogo pri kakovostnem prevajanju. To je znanje, ki se pri človeku (prevajalcu) odraža v obliki izkušenj, in na podlagi katerih se odloča tudi sam.

4.4.3 Uporaba glagolskih predlog

Po začetni fazi, ko je bilo prevajanje večinoma na nivoju besed z nekaj osnovne stavčne analize, je Presis zdaj prešel na uporabo glagolskih predlog pri stavčni analizi in generaciji prevodov. Trenutno je vnesenih dobrih 10.000 glagolskih predlog.

Predloga opiše, kako se določen glagol uporablja. Pove, s katerimi predmeti se veže, katere omejitve so pri tem, katera prislovna določila so lahko ob glagolu, katere predložne zveze so tipično povezane z glagolom in podobno. Predloge so potem povezane v pomene, pri čemer so pripisane tudi potrebne stilne oznake, ki jih prevajalnik uporablja za boljše izbiranje prevodov. Da bi se izognili prevelikemu ponavljanju pri predlogah, je del informacij, ki so značilne za veliko število glagolov v posameznem jeziku, rešen programsko na nivoju prevajalnika (analizatorja oz. generatorja) in so v teh primerih v predlogah poudarjene le izjeme. Tak primer je npr. vezava s prislovnimi določili, kjer so tipično dovoljena dodatno k predlogu poljubna prislovna določila, razen prislovnega določila kraja, po katerem se vprašamo kam. Analizator in generator tudi skrbita za pravilen vrstni red prislovnih določil v posameznem jeziku.

S pomočjo predlog lahko razrešimo pri prevajanju dvoumnosti, ki se slovnično razlikujejo, recimo glede prehodnosti/neprehodnosti, povratnosti pri slovenskih glagolih, frazne glagole v angleščini... Prav tako je možno stalne fraze v celoti vpisati v predloge, s čimer lahko dosežemo pravilne prevode fraz. Ta zapis fraz je veliko bolj prilagodljiv, kot bi bil zapis z enostavnim zaporedjem besed, saj so elementi predloge lahko v poljubnem vrstnem redu.

Tako lahko Presis zdaj brez težav prevede tudi stavek *Bobu je včeraj Janez rekel bob.* v *Janez called a spade a spade yesterday.* Pri tem najde tudi druge možne prevode, vendar da na prvo mesto rešitev, kjer predloga pokrije čim večji del stavka.

Predloge določijo tudi stavčne člene, tako da je zaradi tega mogoče povezovati pomene glagolov s pomeni samostalnikov glede na vlogo, v kateri se pojavljajo. Tako je mogoče določiti, da sta glagol *skuhati* in samostalnik *kosilo* povezana tako, da je kosilo tipično predmet v tožilniku pri tem glagolu, ne pa npr. osebek ali predmet v dajalniku.

Pri predlogah je možno napisati tudi tipične predložne zveze, s katerimi se povezujejo. S tem lahko po eni strani izboljšamo prepoznavanje predlog, po drugi strani pa tudi dosežemo, da se predloge ob določenem glagolu prevaja drugače, kot bi se sicer običajno. V našega Janeza se je zaljubila lepa Micka. tako Presis prevede v *Beautiful Micka fell in love with our Janez.*

Presis zna predloge tudi prilagoditi različnim časom in trpniku. Morebitne izjeme so označene z uporabo zastavic pri predlogah. Če stavka, ki je bil v originalu v trpniku, v ciljnim jeziku ni mogoče zapisati na ta način, ga generator pretvori v tvorni način. Primer za to je prevod iz *John was made happy by George.* v *George je osrečil Johna.*

Posebnost slovenščine so tudi glagoli, ki dobijo v neosebni rabi drug pomen. Tak primer je *Gre za ta problem.*, kar Presis s pomočjo predloge prevede v *It is this problem.* oz. kot druga možnost *It is about this problem.* Najde pa seveda tudi kopico drugih možnosti, recimo *It goes behind this problem.*

Predloge so lahko tudi take, da kot del zahtevajo nov glagol. Tipičen primer so predloge za modalne glagole. V takem primeru zna Presis sestaviti predlogi. Primer za tak prevod je *Why could John fall in love with Mary?*, kar se prevede v *Zakaj bi se John lahko zaljubil v Mary?*

4.4.4 Internetna in časovno omejena predstavitevna verzija

Na spletni strani Amebis ponuja predstavitevno verzijo, kjer lahko preizkusimo kakovost prevajanja trenutne verzije prevajalnega sistema Amebis Presis (verzija 1.51). Pri tem smo omejeni na dolžino največ 200 znakov, ki jih lahko vpišemo v polje, nato pa izberemo smer prevajanja. V predstavitveni verziji ni mogoče dodajanje lastnih prevodov. Vsa besedila, ki jih prevajamo na teh straneh, si pri Amebisu shranjujejo kot pomoč pri nadaljnjem razvoju prevajalnih tehnologij.

Pri Amebisu pa nam ponujajo tudi časovno omejeno predstavitevno verzijo prevajalnika*. Za tovrstno brezplačno verzijo prevajalnika se je potrebno obrniti na info@amebis.si.

* Nama so ponudili predstavitevno verzijo poslovnega paketa (vsi trije uporabniški vmesniki) z rokom uporabe 94 dni.

4.4.5 Napotki pri prevajanju

Prevajalni sistem Amebis Presis ponavadi ne prevaja povsem pravilno. Prevod je namenjen približnemu razumevanju originalnega besedila ali osnovi za nadaljnje prevajanje. Vendar je kakovost prevoda odvisna tudi od dejavnikov, na katere lahko vplivamo. Za boljše prevajanje se je potrebno držati spodnjih navodil:

- Stavke in lastna imena je potrebno pisati z veliko začetnico (imena jezikov, dni v tednu in mesecev v slovenščini niso lastna imena).
- Ne smemo uporabljati SAMIH VELIKIH ČRK.
- Besedilo za prevajanje naj bo slovnično pravilno, saj to pomaga pri boljši analizi.
- Pisati je potrebno »kaša« ali »ka^sa« in ne »kasa«, »kasha« ali »kas"a«.
- Pred ločili ne smemo pisati presledkov, potrebno jih je pisati za ločili. Pred oklepajem in za zaklepajem naj bo presledek, razen če zaklepaju sledi ločilo.

Zgradba stavka, velika začetnica, ločila, po možnosti brez neznanih besed... vse to vpliva na prevajanje. Čim bolj lahko program določi posamezne možnosti v analizi, tem bolj je lahko uspešen.

Nekaj primerov, ki kažejo zakaj je pravilna raba začetnic in ločil pomembna:

- *lahko noč* Presis prevede v *can night*,
Lahko noč. prevede pravilno *Good night*.
- *Rada posluša Ravelov Bolero?* -> *Does she like to listen to Ravel's bolero?* ali
Rada posluša Ravelov Bolero. -> *She likes listening to Ravel's bolero.*
- *Jaz sem matic.* -> *I am nuts.* ali
Jaz sem Matic. -> *I am Matic.*
- *francoz je šel domov* -> *spanner went home* ali
Francoz je šel domov. -> *Frenchman went home.*
- *kdaj boste poslali denar* -> *you will send money sometimes* ali
Kdaj boste poslali denar? -> *When will you send money?*
- *Ali veste kje je?* -> *Or know somewhere her?* ali
Ali veste, kje je? -> *Do you know, where is he?*

Poleg tega velja poudariti, da je prevajanje iz slovenščine v angleščino občutno boljše kot iz angleščine v slovenščino. Razlogov je več, najpomembnejše pa je dejstvo, da se s slovensko slovnico pri Amebisu ukvarjajo dlje in je zato tudi analiza bolj uspešna.

4.5 Ustreznost prevajanja

4.5.1 Primeri prevajanja

Da bi videli, kako prevajalni sistem Presis dejansko deluje, na lastni koži, sva preizkusili poskusno verzijo Amebis Presisa.

S poskusno verzijo prevajalnega sistema Presis sva prevedli nekaj stavkov oziroma besednih zvez in želeli preveriti, kakšna bo kvaliteta prevoda. Prevod, ki sva ga dobili, sva skušali umestiti kot »ustreznega« (oz. »sprejemljivega«) ali kot »napačnega«/»neustreznega«. Vsak prevod sva nato analizirali. Pregledali sva katere dele stavka je prevedel (oz. ni prevedel) in katere besede so imele poleg danega prevoda še druge možne prevode.

Prevode sva naredili za nekaj primerov v obeh smereh. Najprej iz slovenščine v angleščino in nato še obratno.

Treba je poudariti, da so besede prevedenega stavka različno obarvajo in vsaka barva pomeni sledeče:

črna = Besede z enim znanim prevodom.

zelena = Besede z več možnimi prevodi.

modra = Nprevedenjena lastna imena.

rdeča = Neznane besede.

- Prevod iz slovenščine v angleščino:

USTREZNI/SPREJEMLJIVI PREVODI	
<i>Slovenščina:</i>	<i>Angleščina:</i>
Danes je lepo vreme.	There is fine weather today.
Mojca je zaspana.	Mojca is sleepy.
Pes je klobaso.	Dog is eating sausage.
Povedala mi je vse o svojem otroštvu.	She said all concerning her childhood to me.

NEUSTREZNI/NAPAČNI PREVODI	
<i>Slovenščina:</i>	<i>Angleščina:</i>
Živijo, kako si?	Do they live , how are you?
Pojdi se učiti!	Go to learn!
Lije kot iz škafa.	Pours as from škafa .
Pokvaril se mi je računalnik. Zato sem ga nesel v popravilo.	Broken down me is computer. That is why I carried him in repair.

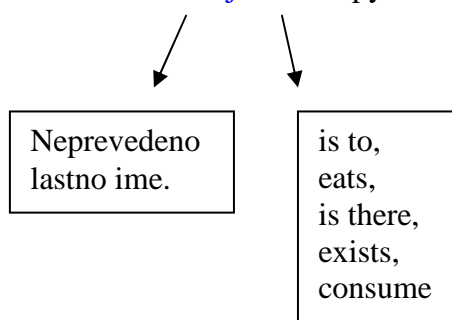
- Prevod iz angleščine v slovenščino:

USTREZNI/SPREJEMLJIVI PREVODI	
<i>Angleščina:</i>	<i>Slovenščina:</i>
A baby is calling a mother.	Dojenček kliče mater.
Sun is shining.	Sonce sije .
I am freezing.	Zmrzujem .
Hello, how are you?	Živio , kako obstajaš .

NEUSTREZNI/NAPAČNI PREVODI	
<i>Angleščina:</i>	<i>Slovenščina:</i>
Would you like some more coffee? There's still a few left.	Ali bi ti kakšen več kave bilo všeč? There's še vedno malo levi.
Take an umbrella in case it rains while you are out.	Vzeti dežnik v primer ono dežji medtem ko ti so ven.
I am Miro Romih.	Sem Miro Romih .
Good luck!	Dobra sreča!

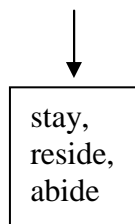
4.5.2 Analiza prevodov

1. Mojca je zaspana. → Mojca is sleepy.



Prvi stavek je kratek in enostaven. Sestavljen je iz osebk, povedka in pridevnika. Prevajalni sistem je stavek pravilno prevedel. Posebnost pri tem prevodu: ime *Mojca* ne prevaja, pravilno jo označi kot neprevedeno lastno ime. Pri povedku *is* se ponudi več različnih prevodov; med njimi vsi nekako ne sodijo v kontekst, oziroma pomensko ne sodijo v ta stavek. Zato je prvi prevod, ki ga je našel prevajalnik še najbolj primeren za naš prevod.

2. Živijo, kako si? → Do they live, how are you?

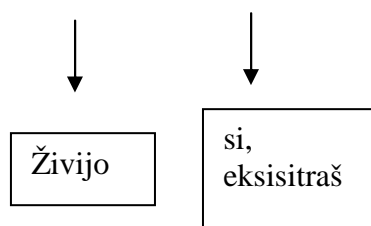


Postavili sva vprašanje, ki se začne s pozdravom. Stavek je ločen z vejico. Zanimivo je, da prvi del stavka prevede napačno, drugega pa pravilno. Naredili sva preizkus in osamili oba dela stavka. Tako sva v prevajalnik vpisali le besedo *Živijo*, da bi videli, če bo isto besedo prevedel enako. V tem primeru je prevedel besedo kot *hello* ali *hi*. Se pravi, da je besedo, kot ločeno, zaznal pravilno. Poleg tega, da je prevajalnik prvi del stavka napačno prevedel, je zanimivo tudi to, da je namesto trdilnega stavka uporabil kar vprašalnega. Se pravi, da se v angleškem prevodu nahajata kar dve vprašanji.

Še ena zanimivost, ki sva jo opazili pri tem prevodu (oziroma bolje rečeno pri načinu prevajanja) je, da prevajalni sistem vedno najprej poišče prevode, ki so vsaj z delom besede nekako povezani ali pa podobni besedi, ki jo želimo prevesti. Nato poišče prevode, ki so pomensko sorodni.

Ta stavek sva uvrstili v neustreznega.

3. Hello, how are you? → Živio, kako obstajaš?



Ravno zaradi prejšnjega primera sva naredili prevod še v obratni smeri. Popolnoma isti stavek (vprašanje) sva prevedli iz angleščine v slovenščino. Rezultat je obrodil precej boljše sadove. Prevod je veliko bolj smiseln. Poleg tega je prvi del stavka celo pravilno preveden. Pri drugem stavku pa je najprej uporabil soroden prevod. Druga varianta, ki jo ponuja prevajalnik, je bolj ustrezna. Načeloma bi lahko rekli, da je prevod sprejemljiv, saj ponuja pravilno rešitev (*Živijo, kako si?*).

4. Would you like some more coffee?

There's still a few left.

→

Ali bi ti kakšen več kave bilo všeč?

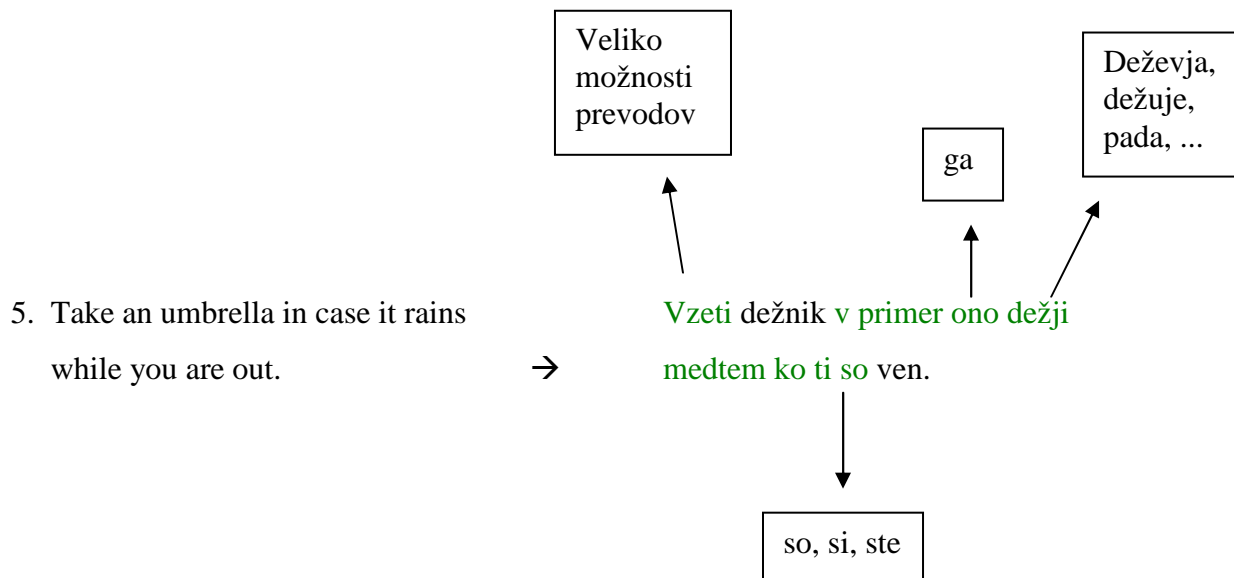
There's še vedno malo levi.

Ni prevoda.

puščen,
ostal,
odpeljal

Posebnosti pri teh dveh stavkih so: če dobro pogledamo vsak del stavka in potem vsak njegov prevod, opazimo sledeče. Prvi del *Would you like* prevede pravilno, potem *some more coffee* tudi pravilno in *like* tudi pravilno. Če pa pogledamo celoten prevod/stavek, pa vse skupaj izgleda, kot da prevajalnemu sistemu manjka le to, da teh prevedenih delov še ne zna nekako ustrezno zložiti skupaj (na pravo mesto).

Pri drugem stavku okrajšane besede *there's* ne pozna, če pa vtipkamo v okno ločeno besedo *there is*, pa bo besedo znal prevesti.



Namenoma sva želeli prevesti tudi kakšen malo daljši stavek. Posamezne dele stavka prevaja preveč dobesedno. Posamezne besede so ustrezno prevedene, vendar stavek še vedno ni slovnično usklajen.

4.5.3 Ugotovitve po analizi

Med prevajanjem sva ugotovili, da so besede, ki jih ne prevede (označene so z rdečo ali modro barvo), tiste besede, ki jih prevajalnik ne pozna/jih nima shranjenih. V našem primeru so te besede bile večinoma lastna imena, priimki, pa tudi razne okrajšave (npr.: *there's*). Tudi besede, ki jih prevajalnik ne prepozna.

Pod nekaterimi prevodi se pojavi še nekaj drugih možnih rešitev, ki uporabniku omogočajo, da prilagodi prevod tako, kot se njemu zdi najbolj primerno. Seveda ni nujno, da bodo tudi te dodatne možne besede kaj bolj v pomoč. Kajti prevajalnemu sistemu še vedno primanjkuje prevodov. Zato so mnogi prevodi togi in nerodni.

Dejstvo je, da Presis bolj natančno prevaja iz angleščine v slovenščino, saj je v primeru, ko sva prevajali stavke iz angleščine v slovenščino, sistem večkrat pravilno prevedel, kot pri prevodu iz slovenščine v angleščino.

5 ZAKLJUČEK

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 hitrejša 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 vzdrževati. Poznavanje prevajalskih orodij in drugih jezikovnih tehnologij bo v prihodnje tako nujen del prevajalčeve izobrazbe in pogoj za njegovo konkurenčnost.

6 LITERATURA

1. *Prevajalni sistem Amebis Presis* [online]. Amebis d.o.o., 2006 [citirano aprila 2006]. Dostopno na svetovnem spletu: <<http://presis.amebis.si/>>.
2. ROMIH, M. *Jezikovne tehnologije. Življenje in tehnika*, 2003, januar, str. 20-29.
3. BARTOLINI, B. *Babilonski računalnik*. Moj Mikro, 2004, št. 1, str. 64-65.
4. ŽORŽ, J. *Prevajanje brez gore slovarjev*. Moj Mikro, 2004, št. 12, str. 44-45.
5. *Uporaba glagolskih predlog pri strojnem prevajanju* [online]. HOLOZAN, P. [citirano maja 2006]. Dostopno na svetovnem spletu: <<http://nl.ijs.si/isjt04/zbornik/sdjt04-04holozan.pdf/>>.
6. *Uvod v korpusno jezikoslovje* [online]. ERJAVEC, T. Odsek za tehnologije znanja, Institut Jožef Stefan, 2004 [citirano maja 2006]. Dostopno na svetovnem spletu: <<http://nl.ijs.si/et/talks/korpus/>>.
7. *Končno prevajanje tudi na slovenskem spletu* [online]. Mestna občina Celje, 2006 [citirano maja 2006]. Dostopno na svetovnem spletu: <<http://www.celje.si/novica.asp?id=565>>.
8. VINTAR, Š. *Računalniške tehnologije za prevajanje*. Uporabna informatika. 1999, let. 7, št.1, str. 17-24.
9. Informacije, ki sta nama jih posredovala Iztok Grilc in Peter Holozan iz podjetja Amebis d.o.o.