

# MACHINE TRANSLATION AS HELP IN THE COMMUNICATION PROCESS

## STROJNO PREVOĐENJE KAO POMOĆ U PROCESU KOMUNIKACIJE

Zdravko DOVEDAN, Sanja SELJAN, Kristina VUČKOVIĆ

[zdovedan@hotmail.com](mailto:zdovedan@hotmail.com), [sseljan@ffzg.hr](mailto:sseljan@ffzg.hr), [kvuckovi@ffzg.hr](mailto:kvuckovi@ffzg.hr)

Odsjek za informacijske znanosti

Filozofski fakultet Zagreb

### Sažetak

Danas, više no ikada, prilike zahtijevaju kompjutersku podršku za brzo pronalaženje informacija i prevođenje različitih tekstova. Otvaranje novih tržišta, ekonomski razvoj, kulturna, znanstvena i politička suradnja, kao i medicinsko liječenje zahtijevaju nadilaženje lingvističkih i nacionalnih barijera. Potražnja za strojnim prevođenjem sve je veća. Da bi se preveli samo dokumenti na 12 glavnih jezika EEZ-a, potrebno je prevođenje u 132 smjera. U Hrvatskoj, također, zbog približavanja i suradnje evropskim institucijama i kompanijama, postoji potreba za prevođenjem ogromnih količina dokumentacije uglavnom sa/na engleski jezik. Problem strojnog prevođenja nadilazi granice zemalja. Postavljaju se pitanja: u kojoj mjeri je ono moguće? da li je isplativo? kakva je kvaliteta prijevoda? u kojim slučajevima može zamijeniti čovjeka? da li je moguće potpuno automatsko prevođenje?

Ključne riječi: strojno prevođenje, prirodni jezik, problemi formalizacije, višeznačnost, ograničenost, podjezik, automatizacija, kvaliteta teksta

### Abstract

*Today, more than ever, circumstances require computer systems for retrieving the information and for translating various texts. New markets, economy development, cultural, scientific and political cooperation, medical treatments require surpassing linguistic and national boundaries. Demand for Machine Translation (MT) is high and growing. In order to translate documents in 12 major languages of EEC, it is necessary to translate in 132 directions. Because of approaching and cooperation with European institutions and companies, in the Croatia there is also important demand for translating huge quantities of documentation from/to the English language*

*The problem of machine translation (MT) surpasses national boundaries. The questions are obvious: What are possibilities of MT? Does it pay off? What is the quality of machine translated texts? Can it replace the human being? Is fully automated translation possible?*

*Key words: Machine Translation, Natural Language, formalization problems, ambiguity, restriction, sublanguage, automatization, text quality*

## 1 Uvod

Strojno prevođenje (eng. Machine Translation) podrazumijeva postupak pretvaranja informacija s jednog prirodnog jezika na drugi pomoću računala, ali brojna elektronička pomagala (rječnike, tezauruse, baze) kojima se ubrzava ili olakšava postupak prevođenja.

Što je potaknulo njegov razvitak? Danas, više nego ikada, postoji potreba za brzim dostupom do informacija. Za bavljenje znanstvenim radom neophodno je konzultiranje strane literature, kao i za praćenje novosti u medicini, za zajednički rad na projektima između zemalja, za konzultiranje pravnog sustava u drugim zemljama, za organizaciju određenih segmenata kompanija itd. Međunarodna ekonomska i politička stabilnost također ovise o zajedničkim informacijama. Nova zajednička tržišta i rast trgovinske razmjene izazvali su veliku potražnju za brzim prevladavanjem jezičnih barijera. Samo prevođenje dokumenata na dvanaest glavnih jezika Evropske ekonomske zajednice, zahtijevalo bi prijevode u 132 različita smjera ( $n(n-1)$ ).

### 1.1 Kvaliteta prijevoda

*Kvaliteta prijevoda* jedan je od najvažnijih čimbenika. Dobri prevodioci pažljivim prevođenjem mogu prevesti i nijanse izvornoga teksta, no vrlo često takva kvaliteta nije potrebna. Vrijednost informacija se s vremenom mijenja ili se prevode tekstovi namijenjeni užem krugu ljudi, a kvaliteta prijevoda ne mora biti savršena. Ponekad je dovoljan grubi prijevod teksta kojim se prenose najvažnije informacije.

Najveći problem u strojnom prevođenju predstavlja višeznačnost riječi ili rečenica. S. Nirenburg također navodi dva načina za *prevladavanje dvosmislenosti*: ograničiti gramatiku i vokabular izvornog teksta i tako stvoriti podjezik ili odbaciti zahtjev za potpunom automatizacijom i dozvoliti posredovanje čovjeka

Kvaliteta prijevoda je uglavnom tim bolja što sustav sadrži više pravila i veće rječnike, ali uz uvjet da je tekst napisan iz uže definiranog područja koristeći pri tome jednoznačne pojmove, ne predugačke rečenice, bez

mного zamjenica itd. Jedan od važnih elemenata jest i strukturna bliskost dvaju jezika, vrsta teksta i tematika.

Kvaliteta prijevoda također je usko vezana i uz mogućnost interaktivnog sudjelovanja, jer je gotovo uvijek potrebno pred- ili post-uređenje teksta.

Stoga se govori o različitim *stupnjevima automatizacije*: moguće je dobiti potpuno automatski prijevod slabije kvalitete koji ponekad više služi od nikakvog prijevoda, zatim prijevod nekog teksta iz uže definiranog područja ili poluautomatski prijevod uz posredovanje *pred-urednika*, *post-urednika* ili interaktivnim sudjelovanjem.

Prema već navedenom, MT najbolje radi ako je tema usko ograničena (npr. upute za rukovanje, tekstovi iz područja fizike, kemije, medicine i sl.) i uz uvjet korištenja jednoznačno definiranih pojmova. Prema Muriel Vasconcellos, koja je predsjednica "Association for Machine Translation in the Americas" i savjetnica za strojno prevođenje u Washingtonu, Peter Wheeler iz Antler Translation Services koristi MT za prevođenje automobilskih priručnika s engleskog na francuski za potrebe General Motorsa. "Automobilski priručnici su idealni tekstovi za MT – suhoparni, objektivni, činjenični, često se ponavljaju i vrlo su dosadni. S takvom vrstom tekstova ljudi teško izlaze na kraj. Pomoću MT-a postigli smo trostruko povećanje produktivnosti."

## 2 Lingvistički i statistički pristup u kombinaciji

Strojno prevođenje spada u skupinu koju zajednički nazivamo obradom prirodnog jezika (eng. Natural Language Processing). No, zbog mnogih karakterističnih osobina ono se smatra i zasebnim područjem. Takav odnos proizlazi iz činjenice da je strojno prevođenje najranija grana NLP-a. Uz ovo područje, usko je vezan i pojam kompjuterske lingvistike (eng. Computational Linguistics) koji je širi od pojma strojnog prevođenja jer podrazumijeva sve moguće uporabe računala u jeziku, statistici i obradama koje proizlaze iz značenja jezika. Još jedan blizak pojam je strojno razumijevanje prirodnih jezika (eng. Machine Understanding of Natural Language). To je naoko uži pojam jer se ograničava na razumijevanje, ali to i jest bit problema.

U posljednje vrijeme sve se više govori o *jezičnom inženjeringu* (eng. *Lanugage Engineering – LE*), čime se udaljava od navedenih pojmova NLP i kompjuterske lingvistike. O navedenome govore i promjene u nazivu iz *Department of Language and Linguistics* and *Centre for Computational Linguistics* u *Department of Language Engineering*, kao i ime programa evropske komisije iz *Linguistic Research and Engineering* u *Language Engineering*. Jedan od primjera je i časopis *Natural Language Engineering*.

Prema nedavnim istraživanjima, lingvistička teoretska pravila kombinirana sa statističkim metodama daju najbolje rezultate. Novi naziv *jezični inženjering* održava obnovljeni interes za kvantitativnu procjenu

koja se temelji na lingvističkim podacima (korpusi tekstova, rječnici) dostupnima na računalu.

Ime *jezični inženjering* otkriva novi odnos prema temeljima istraživačkog postupka: povećan je interes za statističku analizu, za metode kvantitativne procjene, a osobito za jezične podatke dostupne na računalu (korpusi, rječnici). U posljednjih deset godina, područje strojnog prevođenja ispituje se više empirijskim metodama, nego na a priori danim lingvističkim teorijama i pravilima. Jedan od najzačudnijih primjera jest isključivo statistički pristup IBM kompanije – Watson Research Laboratories koji su koristili veliki kanadsko englesko – francuski korpus (Brown et al. 1980). Iako u navedenom istraživanju nisu korišteni gotovo nikakvi klasični lingvistički izvori (leksikoni, sintaksa, semantika), sistem je preveo pravilno između 50-65% rečenica. Unatoč dobrim postignutim rezultatima, Wilks (1994) ističe da uspješnost ostaje statična ukoliko se koriste isključivo statističke metode bez lingvističkog pristupa. Iz navedenog je vidljivo da su za razvoj strojnog prevođenja potrebna daljnja teoretska lingvistička istraživanja, koja u kombinaciji sa statističkim metodama mogu dati najbolje rezultate (primjer Systran), za razliku od teoretski zasnovanog sistema EUROTRA (Johnson et al. 1985.). IBM-ova statistička istraživanja pokazala su da nije potrebno znati apsolutno ništa o jeziku na koji se prevodi, što većina istraživača intuitivno smatra neprihvatljivim .

## 3 Razvoj strojnog prevođenja

Prvi uređaji za strojno prevođenje stvoreni su početkom tridesetih godina kroz patente koje su neovisno iznijeli George Artsrouni (francusko-armenskog prijekla) i Petr Smirnov-Troyanskii (Rusija). George Artsrouni je 1937. patentirao shemu pohrane na papirnatoj vrpici koja bi se mogla koristiti za pronalaženje ekvivalentnih riječi u drugome jeziku. P. Troyanskii je imao viziju dvojezičnog ili višejezičnog mehaničkog prevođenja koje bi se provodilo u tri koraka: logička i sintaktička analiza izvornog jezika, transformacija temeljnog oblika i funkcija u ekvivalentan slijed ciljnog jezika i prijevod u govorni ciljni jezik.

Ozbiljniji pokušaji započeli su nakon pojave ENIAC-a (eng. Electronic Numerical Integrator and Calculator) 1946. godine. Prema autorima Barr i Feigenbaum, pojam strojnog prevođenja javio se već prije uporabe prvih računala. Kako je to novo područje, nastaju mnoge teorije vezane za samo područje strojnog prevođenja, ali i za informacijske znanosti općenito.

U to vrijeme *Claude Shannon* postavlja tezu o informacijskoj teoriji i komunikacijskom kanalu, *Norbert Wiener* određuje pojam kibernetike, a *Terry Winograd* govori o razvoju kompjuterske lingvistike već nakon pojave prvih računala. Prve obrade tekstova sastojale su se u popisima pojavljivanja riječi ili konkordancija (riječi u okruženju).

Jedan od začetnika teorije komunikacije, *Warren Weaver* iznio je teoriju 1949. da bi se tehnike vojnog

šifriranja mogle primijeniti i na strojno prevođenje. Dakle, računalo je trebalo simulirati rad prevodioca: za svaku riječ pogledati u dvojezični rječnik, usporediti riječi iz ulaznog teksta s onima pohranjenima u rječniku (binarno ili slučajno pretraživanje), odabrati odgovarajući prijevod i nakon obrade cijele rečenice, prevedene riječi složiti u određeni poredak prema pravilima ciljnoga jezika. Za razliku od programa koji samo "gledaju" riječi, strojno prevođenje treba raščlaniti tekst izvornog jezika i slagati rečenice u ciljnome jeziku. Ulaz u računalo mora biti tekst u obliku pogodnom za kompjutersku obradu.

Već 1951. godine započela su istraživanja u SAD-u na Massachusetts Institute of Technology na temu strojnog prevođenja i godinu dana kasnije sazvana je prva konferencija. Bilo je očigledno da automatizirano prevođenje ne može biti razvijeno bez dugoročnih istraživanja. Tako su se naizgled jednostavnom rješenju ispriječile su se brojne teškoće u odabiru odgovarajućeg prijevoda i u poretku riječi u rečenici ciljnoga jezika. Naime, uvidjelo se da kvaliteta prijevoda ne ovisi samo o veličini rječnika, te je prevođenje riječ za riječ nestalo, a zamijenila su ga istraživanja s ciljem "razumijevanja" teksta.

### 3.1 Razdoblje sveopćeg optimizma (1946-1965)

60-ih godina nastali su programi koji su obilježili prve rezultate na području obrade prirodnoga jezika. To je vrijeme općeg zanosa i ideje o računalima kao o "svemogućim" strojevima koji misle.

U ovom razdoblju počeli su se razvijati ili su već javno prikazani najpoznatiji programa za konverzaciju: program PARRY (1966) autora Kenneth Colby koji simulira razgovor s poremećenom osobom; program ELIZA (1966) autora Josepha Weizenbaum koji simulira razgovor psihijatra i pacijenta; program SHRDLU (1972) autora Terry Winograda koji je započeo u ovom razdoblju, a simulira rad robotske ruke koja razmješta geometrijska tijela na stolu; program BASEBALL (1963) autora Berta Greena koji se bavi pronalaženjem informacija; program STUDENT (1968) autora Daniela Bobrow koji rješava jednadžbe visokoškolskog stupnja itd.

U doba prvog spominjanja strojnog prevođenja u uporabi je bio samo strojni jezik. Također nitko nije govorio o beskontekstnim i kontekstnim gramatikama, o transformacijskoj gramatici i gramatici proširenih prijelaznih mreža. Prvo mjesto u kompjuterskoj lingvistici zauzimali su statistički eksperimenti na jeziku, kao npr. frekvencije slova ili frekvencije slijeda od dva slova. Takve analize bile su korisne u stvaranju pseudo-koda.

Pravi razvoj strojnog prevođenja započinje polovicom 40-ih godina kada su W.Weaver i A.D.Booth (1946), obojica stručnjaci za razbijanje kodova pomoću računala, koristeći pri tome frekvencijske analize znakova i riječi, vodili javne rasprave. Prema njihovom mišljenju metode dešifriranja mogle bi se upotrijebiti i u

prevođenju, a osnovni problem predstavlja ugradnja cjelokupnih rječnika dvaju jezika, iako su i sami priznali da postojanje rječnika neće riješiti sve probleme.

Ubrzo su uočene sljedeće poteškoće:

- mnoge riječi imaju više prijevoda, što ovisi o kontekstu
- redoslijed riječi u rečenici je u različitim jezicima različito definiran
- lokucije i idiomatski izrazi prevode se kao značenjska cjelina

U to vrijeme činilo se vrlo vjerojatnim da je osnovni problem u prevođenju između dva jezika predstavljao rječnik. 1947. Booth i D.H.V.Britten napravili su program za pretraživanje rječnika. 1949. godine Weaver je razaslao okružnicu na temu "Prijevod" na 200 mjesta i time izazvao veliki interes. U dopisu navodi ideju da svi jezici imaju neke zajedničke crte, kao i ideju da se prevođenje između dva jezika obavlja pomoću međukoda (*metajezika*) koji je zajednički svim ljudima. Time bi se broj prijevoda znatno smanjio, pa bi npr. za prevođenje dokumenata na 12 jezika trebalo  $2n$  ili 24 prijevoda, što je znatno manje od direktnog prevođenja i 132 prijevoda. Istovremeno se postavlja i problem prikaza znanja, a time i kvalitete prijevoda, koja se preko međukoda svakako smanjuje, no ovisno o vrsti teksta i strukturi rečenica, ona za određeni krug ljudi može biti dovoljna. Ovakav pristup pogodan je ukoliko se prevodi između 4 i više jezika.

Nakon Weaverovog dopisa započeo je rad u nekoliko centara SAD-a. Također se uvode pojmovi *pred-urednika* i *post-urednika*. Pred-urednik priprema tekst za prevođenje kako bi u što većoj mjeri smanjio jezične i strukturne nejasnoće: uklanja višeznačnosti, dugačke rečenice rastavlja na kraće, smanjuje broj zamjenica i jasnije postavlja veze među riječima. Post-urednik nadograđuje prevedeni tekst. Priprema tekst se isplati kada se prevodi na više jezika jer se smanjuje potreba za naknadnim doradama ili kada su pravila izvornog jezika prilično zamršena. Interaktivnim uređivanjem korisnik smanjuje broj naknadnih ispravaka

1954. godine na Georgetown University, Leon Dostet je udružio snage sa IBM-om na projektu koji je rezultirao prvom javnom demonstracijom sustava strojnog prevođenja. Pažljivo odabrani uzorak od 49 rečenica na ruskom jeziku preveden je na engleski, uz uporabu vrlo ograničenog rječnika od 250 riječi i 6 gramatičkih pravila. Iako je ta demonstracija imala mali znanstveni značaj, bila je ipak dovoljno zadovoljavajuća da potakne visoke donacije za razvoj sustava za strojno prevođenje u SAD-u i nadahne projekte u drugim zemljama.

God. 1954. osnovan je časopis *Machine Translation* (MT) i zadobio veliku pažnju.

Doba intenzivne aktivnosti, kada su dominirali sustavi tzv. 'prve generacije' za izranim pristupom (eng. direct approach), vođeni rječnikom (eng. dictionary-driven) i niskom razinom sintaktičke analize (eng. low-level of syntactic analysis) završeno je pojavom publikacije ALPAC-ovog izvješća 1966. godine, čime je strojno prevođenje ušlo u tiho razdoblje.

### 3.2 *Tih period (1966-1975)*

Nakon ovog velikog zanosa, nastupilo je razdoblje u kojem je MT gotovo potpuno nestalo sa scene. Za vrijeme "tihog" perioda (1966-1975) kompjuter su postali snažniji, brži i većih mogućnosti, te se ponovno javlja interes za strojno prevođenje, ali u okviru određenog područja.

Hillel je 1960. kritizirao stav da bi strojno prevođenje trebalo biti kreacija potpuno automatiziranog prevođenja. Tvrdio je da se na takav način dolazi do rezultata koji su nejasni, dvosmisleni i loši, što je argumentirao na primjeru riječi koja može imati najmanje dva značenja, npr. eng. pen što može značiti kontejner za životinje i olovku.

1964. godine osnovana je u SAD-u ALPAC komisija (Automatic Language Processing Advisory Committee) sa zadatkom da ispita perspektive strojnog prevođenja. U svom poznatom izvješću iz 1966. godine ALPAC je zaključio da je strojno prevođenje sporije, lošije i dvostruko skuplje od živih prevoditelja, bez perspektive u bližoj budućnosti, istaknuvši i kritiku post-uredništva.. Nije bila prepoznata nikakva daljnja potreba za nastavkom ulaganja u strojno prevođenje, nego se više ulaže u podršku živim prevoditeljima kao što su automatski rječnici.

ALPAC-ovo izvješće je pretrpjelo žestoke kritike kao usko i kratkovidno i uvjetovalo je zatišje u narednom desetljeću, kao i potpuni prekid istraživanja u SAD-u.

U ovom razdoblju dovršeni su neki programi na kojima je počeo rad u prethodnom razdoblju (program SHRDLU (1972), program LUNAR (1973) autora Williama Wood koji služi kao pomoć geolozima u analizi stijena donesenih s mjeseca).

### 3.3 *Racionalna primjena (1975 - 1988)*

Rani pokušaji *sintaktičke analize* doveli su do stvaranja prijevoda slabe kvalitete i do velikih razočarenja. Postalo je jasno da je za dobar prijevod potrebno osigurati *razumijevanje* ulaznog teksta. Po prvi puta postalo je vidljivo koliko je potrebno znanja o svijetu za prijevod najjednostavnijih rečenica. Zatim slijedi ideja da sistemi za prevođenje trebaju biti opskrbljeni ne samo rječnikom, nego i univerzalnom enciklopedijom.

80-ih godina u strojnom prevođenju dominirao je sintaksn pristup i metoda 'transfera' što obilježava drugu generaciju sistema za prevođenje, kao npr. ARIANE razvijen na Grenoble University, METAL u Texasu, SUSY na Saarbrücken, MU na Kyoto University i jedan od najpoznatijih višejezičnih projekata EUROTRA u Evropskoj zajednici.

Godine 1973. Yorick Wilks razvio je program za strojno prevođenje (Machine Translation System) koji prevodi manje paragrafe s engleskog na francuski jezik. U programu je prisutna ideja Weaverova međukoda i prikaza pomoću *osnovnih semantičkih jedinica*.

Godine 1975. Roger Schank izradio je program MARGIE (Meaning Analysis, Response Generation and Inference on English) koji koristeći teoriju pojmovne

zavisnosti predstavlja pokušaj intuitivnog modela u postupku razumijevanja prirodnog jezika.

Programi za strojno prevođenje izlaze iz laboratorija na tržište i u urede. Prvu takvu upotrebu započela je 1976. kanadska vlada programom METEO i Evropska zajednica programom SYSTRAN. Ubrzo nakon toga Logos i suradničke kompanije (ALPS i Weidner) počinju proizvodnju tržišno orijentiranih sistema za strojno prevođenje. Američka zdravstvena organizacija Pan American Health Organization svakodnevno koristi SPANAM i ENGSPAN sisteme za prevođenje između španjolskog i engleskog. Razvijaju se i brojni tržišni sistemi za prevođenje sa engleskog na japanski i obratno.

Najznačajnija karakteristika ovoga razdoblja jest da su nakon ALPAC izvještaja očekivanja znatno realističnija. Programi za strojno prevođenje promatraju se kao pomoć u prevođenju, a njihovi nedostaci su očekivani, obznanjeni i lakše prihvaćeni.

Jedan od poznatijih programa s primjerom podjezika koji se najviše približio potpuno automatskom strojnom prevođenju jest TAUM-METEO (Traduction Automatique de l'Université de Montréal, 1976) program koji prevodi meteorološke izvještaje s engleskog na francuski, koristeći pri tome vokabular od 1.500 riječi. Bez naknadnih intervencija postiže se prijevod s oko 80% točnosti. U program je ugrađeno pet vrsta stabala koja prevode:

- samo jednu riječ
- stereotipne naslove (ime, datum, dan)
- rečenicu tipa: pridjev – priložna oznaka – vrijeme (Uglavnom sunčano – s umjerenim vjetrovima – danas)
- rečenicu tipa: imenica – vrijeme za min i max temperaturu - vrijeme (Temperatura – 15 do 18 stupnjeva – sutra)
- rečenicu tipa: imenska grupa – pridjev – priložna oznaka (Prognoza za petak – sunčano – s umjerenim vjetrovima)

Jedna od vodećih kompanije u svijetu koja se bavi strojnim prevođenjem jest SYSTRAN (<http://www.systranet.com>). Projekt se koristi od 1970. u Wright-Patterson zračnim snagama za potrebe prevođenja s engleskog na ruski i obratno, kao i u EEZ-u od 1976. godine. Iako je postignuta vrlo visoka kvaliteta prijevoda, u programu je naznačeno da se ne garantira 100%tna točnost prijevoda, te da ovakvo strojno prevođenje ni u kojem slučaju ne može zamijeniti profesionalnog prevodioca. Proizvod SYSTRAN PRO namijenjen je stručnjacima u prevođenju, a prevodi između engleskog, njemačkog, francuskog, talijanskog, portugalskog i španjolskog jezika. Za svaki jezik korišteno je oko 200.000 osnovnih pojmova, a u sekundi obradi preko 10.000 pravila. Osim općeg rječnika postoji 21 specijalizirani rječnik iz područja tehnologije, medicine, avijacije, kompjuterskih znanosti, ekonomije, matematike itd., kao i mogućnost stvaranja vlastitog specijaliziranog rječnika.

Kvalitetu strojno prevedenih tekstova projekt Systran zasniva se na nekoliko principa: 1) bogati i dobro razrađeni morfološki rječnici 2) posredovanje post-urednika prije raspodjele dokumenata 3) dobro razrađeni programi za obradu riječi i tekstova koji služe kao pomoć post-uredniku.

Alta Vista sustav za prevođenje (<http://www.altavista.digital.com>) prevodi sa pet svjetskih jezika čime informacije sa Web stranica postaju dostupnije široj populaciji (automatski prijevodi novinskih članaka, pristup turističkim informacijama, prijevodi voznog reda vlakova, prijevodi istraživačkih projekata i multinacionalnih kompanije koje imaju predstavništva u različitim zemljama, itd.). Iako su prijevodi nesavršeni i zahtijevaju doradu, omogućuju razumijevanje općeg sadržaja.

Na adresi <http://www.travlang.com/languages> moguće je pronaći program *Foreign Language for Travelers* (1995) autora Michael C. Martin koji prevodi između više od 45 jezika, među kojima i s hrvatskog. Omogućuje učenje isključivo osnovnih riječi i fraza koje stranac koristi u inozemstvu, no pogodan isključivo za početnike.

Globalink Translation Services osnovan je 1989. godine, otkada izdaje software za strojno prevođenje (<http://www.globalink.com>), a omogućuje prevođenje Web stranica sa španjolskog, francuskog, njemačkog u engleski i obratno, prevođenje cijelih tekstova koristeći 31 rječnik za nekoliko svjetskih jezika. Cilj ovih programa je približiti nove proizvode, povećati prodaju i upoznati novu publiku s novim artiklima. Tekst dobiven strojnim prevođenjem nadoraduje prevoditelj.

Napredak u razvoju MT mjeri se njegovom sposobnošću da postepeno prevodi sve teže vrste tekstova uz što manju pomoć čovjeka. Idući ključni cilj, prema M. Vasconecellos je mogućnost prevođenja između evropskih jezika i jezika koji nema abecedu i strukturu rimskog tipa (npr. japanskog, kineskog, korejskog, arapskog).

### 3.4 Nova era (1989- )

1989. godina uzima se kao početak novog razdoblja na području strojnog prevođenja. Jedna od najznačajnijih pojava jest primjena 'korpusne lingvistike' (corpus-based) i poznatijih metoda temeljenih na pravilima (eng. rule-based methods). U korpusnu lingvistiku uključene su statističke metode, prevođenje temeljeno na primjerima (eng. example-based approach) i primjena korpusa ka izvora najrazličitijih lingvističkih podataka i baza znanja. Za ovo razdoblje karakterističan je pomak sa sintakse na više 'leksičke' prikaze.

#### 3.4.1 Metode temeljene na ograničenjima

Osnovna pretpostavka u primjeni strojnog prevođenja temeljenog na pravilima (eng. rule-based) jest da strojno prevođenje zahtijeva analizu i prikaz značenja teksta izvornog jezika i generiranje ekvivalentnog teksta ciljnoga jezika. Prikazi bi pri tome trebali biti leksički i strukturalno nedvosmisleni. Postoje dva osnovna pristupa: metoda 'transfera koja se realizira u tri faze (analiza, transfer, generiranje) ili 'interlingua' model kod kojeg postoje dvije faze (analiza i generiranje). O prednostima i nedostacima svake od metoda neće se sada raspravljati jer je potrebna obuhvatnija analiza.

Iako u ovome razdoblju završavaju dva najznačajnija projekta druge generacije (Ariane i Eurotra), koja zahtijevaju naknadno uređenje teksta, a temeljena su na sintaktičkom pristupu, bez interaktivne komponente. Iako se analiza i generiranje provode na različitim razinama (morfolškoj, sintaktičkoj, semantičkoj), nema primjene na pragmatičkoj ili diskurznoj razini. Navedeni projekti ipak žive u novootvorenom projektu EUROLANG francuske kompanije SITE i njemačke kompanije Siemens-Nixdorf. Cilj projekta je bio proizvesti sistem za strojno prevođenje sa 'user-friendly' sučeljem, pogodan za prevodioce, uz direktan pristup rječnicima, sa engleskog na francuski, njemački, talijanski, španjolski i obratno i francusko-njemački i obratno.

IBM kompanija također je započela sredinom 80-ih godina niz istraživanja unutar LMT projekta (eng. Logic programming MT) gdje se prevođenje odvija u četiri faze opisane u Prolog-u: 1) leksička analiza i opisi ulaznih riječi i njihovih transfera 2) sintaktička analiza ulaznog teksta, te proizvodnje površinske i dubinske (logičke) strukture 3) transfer i restrukturiranje informacija 4) morfolško generiranje ciljnog teksta. Temeljna ideja je kombiniranje leksičkog pristupa i logičkog programiranja.

Najznačajnije istraživanje temeljeno na interlingua pristupu temeljenom na pravilima (eng. rule-based interlingua approach) provedeno je na Carengie Mellon University. od 1985-1989 tim stručnjaka radio je na MT sistemu temeljenom na znanju (eng. knowledge-based MT system), čija je osnovna pretpostavka da prevođenje mora nadići lingvističke barijere i uključiti 'razumijevanje', te primjenjuje model leksičko-funkcionalne gramatike (LFG).

Općeniti okvir za brojne pogramе koje se temelje na pravilima uključuje formalizme koji se kategoriziraju kao *varijante 'unifikacijskih' formalizama* ili *formalizama temeljenih na ograničenjima* (eng. constraint-based formalisms), prema Sadler i Arnolds, 1992. U navedene formalizme pripadaju Leksičko-funkcionalna gramatika LFG (Kaplan i Bresnan 1982), Generalized Phrase Structure Grammar (Gazdar 1985) i Head-driven Phrase Structure Grammar (Pollar i Sag, 1985). Drugu struju čini logičko programiranje razvijeno

iz formalizam TAUM projekta, pisanog u Prologu-u, iz čega se razvio formalizam Definite Clause Grammar.

Formalne gramatike temeljene na ograničenjima primijenjene su u nastavka Eurotra projekta, kroz manje potprojekte koje provode isti istraživači, kao što su MiMo (van Noord, 1991), CAT2, transfer-based formalism of Eurotra (Sharp & Streiter, 1992), ParGram Project. Osnovna karakteristika unifikacijskih formalizama i formalizama temeljnih na ograničenjima je inherentna reverzibilnost, što je osobito poželjno kod dvosmjernih sistema za prevođenje.

Jedan od MT sistem za prevođenje temelji se na *teoriji principa i parametara*, poznatijima kao *Government Binding Theory*, prema kojoj postoje općeniti ili univerzalni principi koji vrijede za sve jezika. Razlike među jezicima opisuju se sintaktičkim i leksičko-semantičkim parametrima. Najpoznatiji sistem koji koristi navedenu teoriju naziva se UNITRAN (Dorr, 1992). Temelji se na interlingua modelu i sadrži dvije osnovne komponente: 1) sintaktičku obradu za prihvati i proizvodnju gramatički točne rečenice i 2) leksičko-semantičku obradu za izvođenje pojmovnih prikaza i za usklađivanje sa ciljnim jezikom. Na sintaktičkoj razini koristi se teorijom principa i parametara, a na leksičko-semantičkoj razini nadopunjuje se leksičko-semantičkim primitivima, sličnima kao u Schankovim prikazima pojmovne zavisnosti. Informacije specifične za jezik kombiniraju se iz leksičko-semantičkih primitiva i sintaktičkih ograničenja.

Kao ekstremni slučaj ideje pomicanje od sintaktičkog prema leksičkom pristupu je model za strojno prevođenje u kojem se pravila transfera ukidaju i uvode jednostavni dvojezični ekvivalenti. Primjer je ACQUILEX projekt (Sanfilippo 1992) namijenjen izgradnji višejezičnog leksikona. U ovom projektu napuštaju se strukturalni pristup u korist skupa semantičkih, sintaktičkih i leksičkih ograničenja.

### 3.4.2 Korpusna lingvistika

Dok se prevođenje temeljeno na pravilima može promatrati kao logički nastavak u obradi jezika sa sintakse prema leksičkoj obradi, primjena korpusa (eng. corpus-based) u metodama strojnog prevođenja označava ulaza u novu eru. Prema J. Jutchins, u primjeni korpusa moguće je razlikovati dva osnovna pristupa:

- a) izravno korištenje informacija iz korpusa za analizu, transfer i generiranje prijevoda, što uključuje prevođenje temeljeno na statistici i na primjerima.
- b) indirektna primjena korpusa kao izvora informacija za stvaranje leksičkih ili gramatičkih baza ili baza znanja.

Korpus može poslužiti kao izvor za primjenu različitih *statističkih postupaka* (eng. statistics-based MT). Najznačajniji takav projekt je IBM-ov CANDIDE (Brown et al., 1990). Istraživanje se temelji na korpusu koji sadrži parlamentarne debate na engleskom i francuskom. Program se temelji na ideji koji se

proračunava vjerojatnost pojavljivanja neki riječi i da ta riječi odgovara dvama, jednome ili nitijednome prijevodu u drugom jeziku. Model se temelji na bigramima (dvije konsektivne riječi) jednoga jezika usporedivih sa bigramima drugoga jezika. Na taj su način obrađene 73 rečenice, od koji je 48% točnih, alternativni ili donekle točnih. Istraživači unutar IBM-a predlažu da se program poboljša ne novim statističkim metodama (npr. trigramima) nego uključivanjem lingvističkih informacija (morfologije i sintaktičkih transformacija za diskontinuirane strukture).

*Metoda temeljena na primjeru* (eng. example-based method) pojavila se prvi puta sredinom 80-ih godina, ali je primijenjena tek krajem istoga desetljeća. Osnovna ideja ovoga pristupa jest da se u prevođenju često javlja analogni primjeri već ranije prevedeni koji mogu poslužiti u postupku novog prevođenja. Zbog toga se pristupilo izradi većih dvojezičnih baza sa primjerima rečenica i sa bržim pristupom. Najpoznatiji takav projekt je ATR projekt ili 'Prevođenje upravljano transferom' za prevođenje govornog jezika (Sumita et al. 1990), i to iz uže definiranog područja kao registracija na konferencijama ili rezerviranje hotelskog smještaja.

Jedno od novijih istraživanja tiče se paralelne računalne obrade, odnosno neuralnih mreža (i.e. connectionist approach).

O važnosti korpusa također se govori prilikom izgradnje podjezika specifičnog za određeno područje zbog izgradnje vokabulara, gramatičkih obilježja, uočavanje sintakse i sl.

Izgradnje leksičke baze i baze znanja osobito je važna kod 'interlingua' pristupa koji koristi pojmovne prikaze (Unitran system; Knowledge-based system na Carnegie-Mellon University, Pangloss).

Kombinacije pristupa temeljenog na pravilima i statističke obrade osnovna je karakteristika ARCH-TRAN sistema za prevođenje sa engleskog na kineski. Prekomjernim povećavanjem pravila nastaju nove teškoće u formalizaciji, zbog čega se jezična baza nadograđuje statističkim informacijama (eng. corpus-based statistics-oriented). Na taj način lingvističkim se pravilima definiraju moguća pojavljivanja leksičkih jedinica, statističkim metodama procjenjuje čestota supojavlivanja određenih jedinica.

## 4 Novosti u tehnologiji i buduća primjena

Prema knjizi Douglasa Adamsa "Autosoperski vodič kroz galaktiku" cilj je napraviti ekvivalent "brbljavca" u obliku prijenosnog uređaja koji simultano prevodi s bilo kojeg svjetskog jezika na bilo koji drugi svjetski jezik. Iako ovo zvuči teško prihvatljivo, tehnologija strojnog prevođenja je već prilično visoko dosegnula. [8]

Danas postoje dva osnovna faktora koja omogućuju lakše korištenje MT-a. Unos teksta u računalo više nije problem – postoje elektronske datoteke, CD-ovi, prijenos teksta putem elektronske pošte (eng. e-mail) itd.

Isto tako, osobna računala i radne stanice danas su dovoljno snažni za potrebe MT-a. Kako tehnologija vodi sve manjim računalima, tako MT postaje dostupan sve širem krugu ljudi. Prvi takav sustav – MicroCAT - pojavio se 1983. a danas se više ne proizvodi. Unix radne stanice, 386, 486, Pentium itd. imaju dovoljno snage za korištenje velikih MT programa. Tako je Pan American Health organizacija u Washingtonu koristila 386 računalo za sustav koji prevodi s engleskog na španjolski. Sustav je 1992. prenesen na računalo sa 33MHz, radi pod DOS-om, 2 MB RAM-a, HARD 80 MB.

Što se tiče osobne upotrebe stolno prevodenja, moguće je koristiti OCR programe (Optical Character Recognition), CD-ROM, modeme i faksove za unos teksta i grafike itd. Mnoge baze podataka sadrže podatke koji nisu na engleskom jeziku, te ih je potrebno prevesti. MT se može koristiti i preko telefonske službe na način da se preko modema pošalje tekst s vlastitog računala na neki veći sustav. U SAD-u se može nazvati Systran i izabrati prevodenje na neki od ponuđenih jezika.

Prema nekim predviđanjima za 10-ak godina nekoliko će novih tehnologija će sudjelovati u svakodnevnici: OCR, automatsko raspoznavanje govora i strojno prevodenje – od telefona koji prevode, do višejezične elektronske pošte, uređaja koji skeniraju i prevode pisma. Na taj način moći će se kupovati proizvod na stranom jeziku (npr. korejskom) preko telefona koji prevodi razgovor.. Postoje sustavi koji raspoznaju govor s točnošću većom od 90%. Jedna od najsloženijih tehnologija jest za raspoznavanje govora, koji mogu prepoznati ograničenu količinu govora a da ih prethodno nije potrebno prilagoditi glasu određene osobe, a temelje se na statistici koji uspoređuju izgovorene foneme sa pohranjenima i daju nizove teksta. Na projektima strojnog prevodenja ulažu se veliki napor i sredstva. Tako se u Japanu u ART Laboratories radi na projektu telefona koji bi prevodio konferencije s engleskog na japanski i uz rječnik od 1.500 riječ mogao prevesti bilo čiji govor. U Njemačkoj se radi na projektu Vermobila - prijenosnog uređaja za prevodenje poslovnih razgovora.

Prema Eduardu Hovy, voditelju projekta u Information Sciences Institute, University of Southern California, na sveučilištu Carnegie Mellon napravljena su tri izuma: Illumina - uređaj za strojno prevodenje koji se koristi u uredu ili knjižnici pored pisača ili fotokopirnog stroja. Primjenom OCR tehnologije, Illumina prevodi tekstove koji se stave na ploču za skeniranje. Drugi je Passport – ručni uređaj za prevodenje riječi koji koristi fotografiju ulaznog teksta i preko dvojezičnog rječnika daje prijevod. Ovaj izum izuzetno je koristan na putu po stranoj zemlji. Kada se uoči neki strani natpis, snimi se pomoću uređaja Passeport, na ekranu se odabere riječ koja nas zanima i traži se prijevod. Treći je projekt koji pomaže osobama oštećenog sluha. Singspeak System prevodi američki jezik znakova i govorni engleski. Sastoji se od rukavica sa senzorima i uređaja koji prate položaj ruku. Senzori šalju signale u računalo koje iz pretvara u engleske znakove.

## 4.1 Podjezik

Iako su sistemi za strojno prevodenje 70-ih i 80-ih godina bili su namijenjeni prevodenju sa bilo kojeg područja, odnosno imali su opću namjenu (kao Systran, Log, Ariane, Eurotra, Metal itd.), što također vrijedi za većinu japanskih istraživačkih, komercijalnih i PC-sistema, u praksi su se izgrađivali rječnici vezni uz određeno područje (npr. računarstvo, zdravstvo, medicina i sl.).

Danas se izgrađuju sistemi vezani za određeno područje, iako ideja nije nova (Meteo sistem za prevodenje vremenske prognoze; ART projekt za prevodenje rezervacija hotelskog smještaja i registracije za konferencije; Critter sistem za prevodenje tjednih izvještaja vezanih u prodaju mesa u Kanadi; Pangloss za prevodenje novinskih izvještaja o poslovanju; sistemi za prevodenje itinerera za vozače; Volmac Lngware Services izradio je program za prevodenje tekstova za tekstilnu kompaniju, za osiguravajuća društva, za prevodenje priručnika za avione itd.; Tradex prevodi vojne telex poruke, ARGO sistem se koristi u financijama i ekonomiji zbog nadgledanje japanskog; Hess sistem za prevodi komande u zrakoplovstvu itd.) Mogućnost nadogradnje rječnika i izgradnja vlastitih rječnika osnovna je karakteristika sistema za prevodenje i to je upravo ono područje u kojem korisnik može najviše intervenirati. Tehnike izgradnje rječnika postaju dostupnije širem krugu korisnika

Karakteristika je brojnih sistem da su izgrađenih za određenu vrstu korisnika uz mogućnost proširenja rječnika.

## 5 Strojno prevodenje i Internet

Posljednjih godina utjecaj Interneta je ogroman i nezaobilazan. Internet se koristi u poslovanju, u istraživanjima, u edukaciji, za razonodu itd. Potreba za brzim dostupom informacija više je nego očigledna. Istraživanje iz 1997. godine o distribuciji jezika na Internetu pokazala su da engleski jezik dominira sa 84% na web-stranicama i da 43% korisnika nisu govornici engleskog jezika.

Globalizacija potiče potrebu za komunikacijom, odnosno za što bržim prevodenjem. Jedan od razloga jesti i taj što veći dio populacije ne govori engleski koji je dominantan jezik u međunarodnoj trgovini i poslovanju.

Iako se govori o globalizaciji, Evropska komisija za očuvanje kulture i jezika (Comité Européen pour le Respect des Cultures et des Langues en Europe) ima web-stranicu prevedenu na 11 jezika, zalažući se za identitet kulture i jezika, za očuvanje tradicije različitih zemlja, ali i potičući jedinstvo u suradnji.

Na Internetu je očita potreba za prevodenjem najrazličitijih poruka, e-mailova, web-stranica itd. Internet je također odredio i značenje brže dostave prijevoda pojedincima i manjim tvrtkama. Mnoge kompanije nude svoje proizvode i usluge, povezani na neki od programa za strojno prevodenje, ali u suradnji sa post-urednicima. Sigurno je da će se s vremenom pojačati potreba za što kvalitetnijim sustavima, a time i



porasti ponuda sistema za strojno prevođenje na Internetu. Pretraživači također donose odgovore na pitanja postavljena na različitim jezicima.

Internet danas možda ima najjači utjecaj na 'globalni svijet' zahtijevajući proširenje vidika na području strojnog prevođenja. Nedavne pretpostavke da će jednoga dana osobna računala bit zamijenjena mrežnim računalima, već su se obistinila. S Interneta se skidaju najrazličitije vrste priručnika, tekstova, obavijesti, sustava za strojno prevođenje i dr., čime će i strojno prevođenje dobiti odgovarajuće mjesto. Ovakav način upotrebe i trgovanja svakako ne znači da će to postati jedini način doticaja, ali će svakako povećati zanimanje i zaintrigirati korisnike.

## 6 Konferencije, udruženja

Navedene aktivnosti institucija, potreba za suradnjom te za brzim i istovremenim dostupom do informacija na različitim jezicima, doveli su do povećanog interesa za strojno prevođenje od strane publike, vlada zemalja, obrazovnih ustanova, tržišno orijentiranih kompanija i najrazličitijih većih institucija. Upravo iz tog razloga održavaju se različite konferencije vezane uz strojno prevođenje kao

- MT – Machine Translation Summit Conference,
- AAAI Spring Symposium
- EAMT Conference
- TMI – Theoretical and Methodological Issues in Machine Translation of Natural Languages
- COLING – International Conference on Computational Linguistics
- International Symposium on Natural Language Understanding (NLU) and Artificial Intelligence (AI)
- ECAI – European Conference on artificial Intelligence

Uz brojne časopise navedene tematike, postoje i udruženja za korisnike, istraživače i stvaraoce aplikacija s nastojanjima da se predstavi nova tehnologija - IAMT (International Association for Machine Translation) i unutar njega AMTA (Association for Machine Translation Americas) koji izdaju časopis MT News International koji izlazi svaka četiri mjeseca, udruženje EAMT European Association of Machine Translation.

Isto tako Evropska unija je sponzorirala niz seminara posvećenih 'jezičnoj industriji' u kojima strojno prevođenje i obrada jezika imaju važnu ulogu. Još jedan od znakova oživljavanja je izvještaj o japanskim istraživanjima na području strojnog prevođenja.

Jedno od udruženja jest EAGLES (eng. Expert Advisory Group on Language Engineering Standards) unutar Evropske zajednice koje se bavi istraživanjima i utvrđivanjem standarda za korpuse, računarske leksikone, gramatičke formalizme, govorni jezik i za procjenu sistema.

Mnoštvo sistema, teorija i projekata pokazuje potrebu za izradom standarda (Japanska elektronička industrija izdala je skup kriterija za ekonomsku i tehničku procjenu sistema za prevođenje od strane korisnika, JEIDA, 1992.

## 7 Zaključak

Čini se da strojno prevođenje više ne nosi u sebi onu poznatu dilemu: može li zamijeniti čovjeka? U svijetu se rade brojna istraživanja, kojima se upravo to ispituje. Strojno prevođenje treba promatrati upravo kao i samo računalo: ono služi kao pomoć u komunikacijskom procesu, zbog olakšavanja ljudskog rada, no za uspješan razvoj ovog područja potrebna su ogromna ulaganja (suradnja stručnjaka iz različitih područja, suvremena tehnologija, financijska podrška, vremenska raspoloživost rada isključivo na projektima strojnog prevođenja). U postupku strojnog prevođenja čovjek može više ili manje posredovati, ovisno o željenoj kvaliteti prevedenog teksta, jer točnost strojno prevedenih tekstova može dosegnuti i 95% točnosti.

Potpuno automatsko strojno prevođenje jest cilj svih ambicioznijih sistema za strojno prevođenje, no gotovo svaki od njih zahtijeva ljudsku intervenciju, tako da se može govoriti o stupnjevima automatizacije u procesu strojnog prevođenja. Za smanjenje problema višeznačnosti, tekstovi se usko ograničavaju na određeno područje, sužavajući na taj način gramatiku i vokabular na određeni podjezik.

Strojno prevođenje predstavlja još jedno pomoćno sredstvo u cilju bolje komunikacije između kompanija, zemalja ili interesnih grupa, a o važnosti ovog interdisciplinarnog područja govore brojne konferencije i projekti kojima se nastoji unaprijediti komunikacija.

Navedene primjene samo su mali dio revolucije strojnog prevođenja i strojne obrade teksta, koji bi polako trebali unesti pomoć u svakodnevni život, ali i koje traže novo opismenjivanje.

Strojno prevođenje je nakon svojih razvojnih faza vjerojatno pronašlo put racionalizma. Lišeni ideala o 'svemogućim' strojevima koji pored ostalog mogu i prevoditi kao živi prevoditelji, postali smo svjesni njihovih ograničenja. Treba pronaći načine kako ih koristiti za olakšavanje ili ubrzanje rada. Svjestan mogućnosti i ograničenja, korisnik je spremniji prihvatiti nesavršen tekst koji će ipak ispuniti onaj osnovni zadatak da komunicira informaciju.

### Literatura:

1. Allen, J.: Natural Language Understanding, The Benjamin/Cumming Publishing Company, Inc.
2. Brigham, N. Global Communications and the Global Language Divide: Implications of New Translation Technologies. Midterm paper, 2000.
3. Brown, P.F., Cocke, J., Della Pietra, S.A., Jelinek, F., Lafferty, J.D., Mercer, R.L., Roossin, P.S.: A statistical approach to machine translation, Computational Linguistics 16 (1990), p. 79-85.
4. Brown, R.D.: Example-Based Machine Translation in the Pangloss System. In Proceeding of the 16<sup>th</sup> International Conference on Computational Linguistics (COLING-96), p. 169-174.
5. Carbonell, J.G., Tomita, M.: Knowledge-based machine translation, the CMU approach. In



- Nirenberg,S. (ed.): Machine Translation: Theoretical and Methodological Issues, Cambridge University Press, 1985, p. 136-144.
6. Cunningham,H., Wilks,Y., Gaizauskas,R.J.: New Methods, Current Trends and Software Infrastructure for NLP
  7. Deville,G., Herbignaux,E.: Natural Language Modeling in a Machine Translation Prototype for Healthcare Applications: a Sublanguage Approach. In Proceedings of the 6<sup>th</sup> International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-95).
  8. Elliston,J.S.: Computer-aided translation – a business viewpoint. Snell,B.M. (ed): Translating and the Computer, North-Holland, Amsterdam, 1979, p. 149-158.
  9. Hovy, E.: MT u vašoj službi, Byte 1993., p.72-73.
  10. Hutchins,W.J., Somers,H.L.: An Introduction to Machine Translation, Academic Press, London 1992.
  11. Hutchins, J.: Latest Developments in Machine Translations Technology: Beginning a New Era in MT research. MT Summit, Kobe, Japan, 1993.
  12. Johnson,R.L., Whitelock,P.: Machine translation as an expert task. Nirenberg,S. (ed.): Machine Translation: Theoretical and Methodological Issues, Cambridge University Press, 1985, p. 136-144.
  13. Kittredge,R.I.: The significance of sublanguage for automatic translation. In Nirenburg,S. (ed): Machine translation: theoretical and methodological issues, Cambridge University Press, 1987.p.59-76
  14. Lehrberger,J.: Automatic translation and the concept of sublanguage. Kittredge,R.I., Lehrberger,J. (eds.) Sublanguage: Studies of Language in Restricted Semantic Domains, De Gruyter, Berlin, 1982, p. 81-106.
  15. Melby,A.: On human-machine interaction in translation.In Nirenburg,S. (ed): Machine translation: theoretical and methodological issues, Cambridge University Press, 1987. p.22-42
  16. Nirenburg,S.: Knowledge and choices in machine translation.In Nirenburg,S. (ed): Machine translation: theoretical and methodological issues, Cambridge University Press, 1987. p.1-21.
  17. Oettinger,A.G.: A Survey of Soviet work on automatic translation. Mechanical Translation 5 (3), p. 101-110.
  18. Quantz,J.J., Küssner,U., Gehrke,M.: Common Modeling for Machine Translation, In Proceedings of the 6<sup>th</sup> International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-95).
  19. Rosetta,M.T.: Compositional Translation, Kluwer, 1994.
  20. Tucker,A.B.: Current strategies in machine translation research and development. In Nirenburg, S. (ed.): Machine translation: theoretical and methodological issues, Cambridge University Press, 1987. p.22-42.
  21. Van Eynde,F. (ed.): Linguistic Issues in Machine Translation, Pinter, London, 1993.
  22. Vascellous, M.: Univerzalni brbljavac. Byte, 1993., p.69-71,73-74.
  23. Weaver,W.: Translation. In Locke,W.N., Booth,A.D. (eds.) Machine Translation of Languages: Fourteen Essays, The Technology Press of MIT, 1955, p. 15-23
  24. Wilks,Y.: Some Notes on State of the Art: Where are We Now in MT: What Works and What Doesn't and the Role of MT as an International Collaborative Activity.