

Alica Arnaut

MAŠINSKO PREVOĐENJE

Sažetak

U savremenom svijetu rješenje velikog broja problema, ili barem smjernice za njihovo rješenje, mogu se pronaći putem računarske tehnologije. Postalo je gotovo nezamislivo baviti se bilo kakvom djelatnošću bez učešća računara. Računari i sve naprednije računarske tehnologije postaju osnovna potreba i imperativ savremenog doba.

Tako je i lingvistika pokušala pronaći efikasan način za rješenje nekih svojih problema putem računarskih tehnologija. Komunikacija ljudi iz različitih jezičkih područja podrazumijeva i znanje različitih jezika. Stoga se ukazala i potreba za pojednostavljenjem prevođenja tekstova različitih sadržaja iz jednog u drugi jezik. Rješenje se pokušalo naći u računarskim programima koji prevode tekstove iz jednog jezika u drugi. Međutim, da li su time riješeni svi problemi i dokle se došlo u pokušajima da mašina zamijeni čovjeka u prevodilačkim poslovima?

Ključne riječi: mašinsko prevođenje, teorija informacije, računarske tehnologije

O mašinskom prevođenju uopće

Mašinsko prevođenje predstavlja granu računarske lingvistike koja prevodi zadati tekst iz jednog prirodnog jezika u drugi, koristeći se pritom računarskim softverom. Da bi mašinsko prevođenje moglo biti djelotvorno, potrebna je potpuna saradnja lingvista i programera. Kada je u drugoj polovini 20. vijeka mašinsko prevođenje stavljeno u središte pažnje lingvističkog svijeta, očekivali su se značajni rezultati i neki vid revolucije u prevodilačkim poslovima. Nastojalo se doprinijeti univerzalizaciji kulture, olakšavajući i ubrzavajući proces prevođenja tekstova iz različitih oblasti – nauke, tehnike, politike. Međutim, o mašinskom prevođenju ne bi se uopće moglo govoriti da prethodno nisu precizno određeni metodi jezičkog opisa, te razvojni stepen tehničkih nauka. Zahvaljujući metodološkim principima strukturalizma, jezički opis je dostigao preciznost koja je bila dovoljna za njegovo prilagođavanje tehnikama mašinskog

prevođenja. S druge strane, brzi razvoj tehničkih nauka omogućio je jednostavno korištenje računara i njihovih mogućnosti za potrebe ovakvog prevođenja.

Postavlja se pitanje da li je moguće prevesti sve tipove tekstova (od onih unificiranih kakvi su administrativni, preko manje standardiziranih, neformalnih tekstova, pa do svakodnevne konverzacije). Razvijaju se specijalizirani softveri koji omogućavaju mašinsko prevođenje tekstova uz ograničenja mogućnosti zamjena u okvirima pojedinih tipova teksta. Ova tehnika posebno je efikasna kada su u pitanju tekstovi pisani formalnim jezikom, pa su takvi prijevodi vjerodostojniji i bliži originalnom tekstu. I pored silnih nastojanja i napredaka koji se svakim danom postižu, mašine, ipak, neće biti u mogućnosti da vjerodostojno prevedu književnoumjetničke tekstove. Da bi to bilo moguće, mašina bi morala u sebi sadržavati sve moguće uneobičajene riječi i izraze, koji su katkad i samom čovjeku teško razumljivi i često neprevodljivi.

Da bi se moglo prevoditi sa jednog jezika na drugi, potrebno je napraviti rječnike sa upotrebnom leksikom. Radi se na pravljenju tzv. mikroglosara. Oni predstavljaju rječnike leksema i izraza jedne struke (npr. medicine, geografije i sl). I pored pažljivog odabira riječi koje će ući u mašinske glosare, može doći do problema u prevođenju zbog eventualnog prisustva sinonima. Tada se najčešće pribjegava teoriji vjerovatnoće, po kojoj se prednost daje onoj riječi koja je upotrebnom najfrekventnija. Tako se riječi klasificiraju na obične, manje obične i rijetke. Postoji i mogućnost navođenja svih mogućih prijevoda sporne riječi u prevedenom tekstu, pa se čitaocu prepušta da izabere onu najpogodniju za dati kontekst.

Pravilan, precizan i detaljan opis jezičke strukture osnovni je imperativ postojanja i funkcioniranja složenog procesa mašinskog prevođenja. Tu do izražaja dolaze sposobnosti i umijeće lingvista da jezičke pojave svoga jezika prodube i definiraju maksimalno jasno i koncizno u isto vrijeme. Najprije je potrebno uporediti jezičke strukture jezika između kojih će se vršiti prevodilačka aktivnost kako bi se odredila mjera kongruentnosti, odnosno stepen tipološke srodnosti ta dva jezika. U tu svrhu lingvisti vrše popisivanje inventara jezičkih jedinica i navode principe njihove distribucije. Posebno pitanje predstavlja način „organizacije *pamćenja*: da li će mašina posebno *pamtiti* samo cijele reči (= *leksička* organizacija mašine) ili će pamtiti posebno osnove, a posebno nastavke reči (= *gramatička* organizacija mašine)“ (Ivić,

1978:222). Lingvisti nisu saglasni oko toga koji način organizacije mašine je bolji, jednostavniji i istovremeno daje vjerodostojnije rezultate. Naročit problem predstavlja gramatička organizacija zbog velikog broja gramatičkih osobina jezika, te njihove nedovoljne razrađenosti u samim jezicima. Uzmimo za primjer bosanski jezik i njegovu gramatičku strukturu u odnosu na gramatičku strukturu engleskog jezika. Broj padeža, padežna homonimija i sinonimija, kao i rod imenica samo su neki od eventualnih problema koji bi se mogli javiti kao nekongruirajući faktori između ova dva jezika.

Uz dobar lingvistički opis površinske strukture jezika može se omogućiti dosta efikasan rad i dobro funkcioniranje mašina u procesu prevođenja. Ipak, u dubinsku strukturu teže je proniknuti na način njenog svođenja na neke univerzalne obrasce (barem za taj jezik). Teško je svesti sve moguće semantičke varijante pojedinih jezičkih elemenata na matematičke obrasce i odrediti njihove ekvivalente u drugom jeziku kako bi se olakšalo prevođenje. Čini se da semantika predstavlja najveći problem mašinskog prevođenja. Programirati računar koji će tekst razumjeti na način na koji ga razumije čovjek i koji će potom stvoriti isti tekst na drugom jeziku, koji će izgledati kao da ga je pisao čovjek, predstavlja najveći izazov mašinskog prevođenja.

John Hutchins pod sistemom mašinskog prevođenja podrazumijeva „software for automatic translation, where input units are full sentences of one natural language and the output units are corresponding full sentences of another language“¹ (Hutchins, 2009:4). On navodi i nekoliko tipova ovih softvera s obzirom na njihovu namjenu:

1. *home use* – namijenjen kućnoj upotrebi i osobama koje inače i nemaju posebnog iskustva sa prevođenjem;
2. *for Internet/Web* – namijenjen prevođenju elektronskih dokumenata sa interneta;
3. *professional use* – namijenjen profesionalnim prevodiocima i prevodilačkim agencijama;
4. *enterprise* – namijenjen kompanijskom intranetu kako bi se podržao prevodilački tim;

¹ softver za automatsko prevođenje, gdje su ulazne jedinice pune rečenice jednog prirodnog jezika a izlazne jedinice su korespondirajuće pune rečenice drugog jezika.

-
5. *for Websites* – namijenjen internetskim stranicama kompanija kako bi se omogućilo njihovo prevođenje;
 6. *spoken language* – namijenjen prevođenju govora.

Proces mašinskog prevođenja

Mašinsko prevođenje predstavlja složen i zahtjevan proces koji podrazumijeva obimnu i iscrpnu pripremu. Najprije je potrebno pripremiti jezike za potrebe prevođenja, dajući opise njihovih gramatičkih sistema, te napraviti rječnike u kojima će se dati popis leksema, nastojeći smanjiti mogućnost semantičkih nedoumica.

Sam proces prevođenja obuhvata nekoliko faza. Najprije se tekst treba pripremiti za prevođenje, odnosno unijeti u mašinu. Potom se zadati tekst prebacuje u matematički jezik formula koje u potpunosti odgovaraju strukturi jezika sa kojeg se prevodi. Da bi se ova faza izvela korektno, trebaju biti zadate matematičke formule koje opisuju gramatičku strukturu toga jezika. Nakon ovoga slijedi proces prebacivanja matematičkog koda jezika izvornog teksta u matematički kod jezika na koji se prevodi. Ova faza podrazumijeva prethodno određene mjere kongruentnosti među jezicima i odgovarajuće relacije među njihovim gramatičkim strukturama. U posljednjoj fazi matematičke formule koje odgovaraju strukturi jezika u koji se prevodi pretvaraju se ponovo u ljudski jezik i dobijamo prevedeni tekst. Shematski se proces mašinskog prevođenja može prikazati na slijedeći način:

JEZIK 1 → MATEMATIČKI KOD 1 → MATEMATIČKI KOD 2
→ JEZIK 2

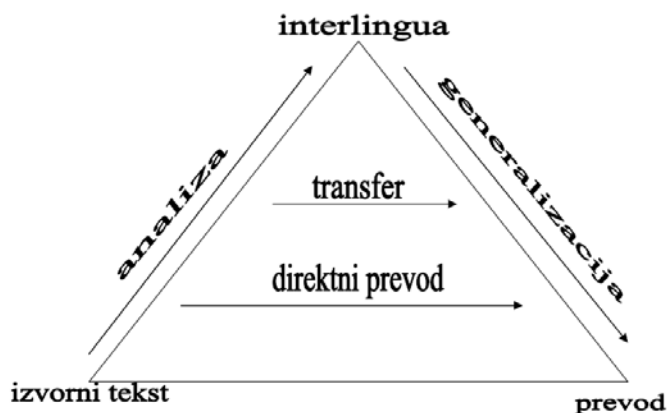
Postoje dva načina procesa mašinskog prevođenja:

1. direktno mašinsko prevođenje i
2. indirektno mašinsko prevođenje.

Direktno mašinsko prevođenje vrši se putem bilingvalnih rječnika sa veoma malom analizom same sintaksičke strukture. Najpoznatiji program ovakvog tipa prevođenja jeste Systran, prvenstveno dizajniran za prevođenje sa ruskog na engleski jezik, a kasnije adaptiran i za englesko-francuske prevode.

S druge strane, indirektno mašinsko prevođenje zahtijeva i neke intermedijalne resurse u procesu prevođenja. Ovakav tip prevođenja može biti interlingvalni, što podrazumijeva upotrebu metajezika (mašinski jezik, interlingua) kao posrednika između dva jezika. Taj metajezik ima svoju standardiziranu sintaksičku

strukturu, koja se sa jedne strane dovodi u vezu sa sintaksičkom strukturom jezika iz kojega sa prevodi (analiza), a sa druge u vezu sa sintaksičkom strukturom jezika u koji se dati tekst prevodi (generalizacija). Indirektno mašinsko prevođenje može uključivati i upotrebu apstraktnih predstavnika oba jezika uz originalnu rečeničnu strukturu i naziva se transfer prevođenje.



Historijat mašinskog prevođenja

Počeci mašinskog prevođenja vežu se za konac četrdesetih godina 20. vijeka. Warren Weaver je gledajući ruski tekst razmišljao na ovaj način: "I have a text in front of me which is written in Russian but I am going to pretend that it is really written in English and that it has been coded in some strange symbols. All I need to do is strip off the code in order to retrieve the information contained in the text."² (www.dialog-21.ru) On je ukazao na četiri ključne tačke mašinskog prevođenja: problem višestrukog značenja, logičke osnove jezika, teorije komunikacije i kriptografske tehnike, te mogućnosti jezičkih univerzalija. Problem višestrukog značenja i dalje je najveći problem mašinskog prevođenja, jer različiti konteksti zahtijevaju upotrebu različitih sinonimnih oblika riječi. S druge strane, pokazalo se da su različiti jezici u svojim osnovnim logičkim strukturama dosta slični, tako da je problem prevođenja uz pomoć kompjutera formalno moguće riješiti. Weaver je vjerovao da je problem mašinskog prevođenja moguće riješiti opsežnijim statističko-semantičkim studijama.

² Ispred sebe imam tekst napisan na ruskom jeziku, ali ću se pretvarati da je, zapravo, pisan na engleskom i kodiran nekim stranim simbolima. Sve što trebam uraditi jeste otkriti taj kod kako bih preuzeo informacije sadržane u tekstu.

Također je smatrao da su jezičke univerzalije ono na čemu treba zasnovati jedan univerzalni jezik iz kojeg će proizilaziti svi drugi jezici. To bi, svakako, uveliko doprinijelo i napretku u mašinskom prevođenju.

Vidi se da je bila izražena potreba za izgrađivanjem jednog jezika koji bi olakšao mašinsko prevođenje, posredujući između izvornog teksta i prevoda. I Louis Hjelmslev je polovinom 20. vijeka naglasio potrebu izgrađivanja jednog metajezika – jezika koji bi bio izgrađen na matematičko-logičkim principima. Težeći „logičkoj gramatici“ koja bi bila jasna i precizna kao algebra, on se približio matematskoj metodi analize jezika. Interesirao se za izgrađivanje opće teorije znakova komunikacije, budući da za sporazumijevanje nije nužna upotreba ljudskog jezika (npr. znakovi u jeziku gluhoonijemih, boje na semaforu, vatra). Hjelmslevljev rad doprinio je tako izgrađivanju metajezika za mašinsko prevođenje, koji predstavlja sistem formula u koji se prebacuje ljudski jezik prilikom prevođenja posredstvom mašine.

Prvi uspješniji pokušaj mašinskog prevođenja predstavljao je eksperiment na Georgetown univerzitetu iz 1954. godine. Tada je prevedeno 49 rečenica sa ruskog na engleski jezik uz rječnik od 250 riječi i samo šest gramatičkih pravila. Godine 1964. uspostavljen je ALPAC (Automatic Language Processing Advisory Committee³) koji je trebao dati ocjenu dotadašnjih dostignuća mašinskog prevođenja. ALPAC-ov izvještaj iz 1966. godine pokazao je da je mašinsko prevođenje bilo sporo, sa dosta propusta i dvostruko skuplje negoli prevođenje koje je radio čovjek. Stoga je preporučeno da se dalje ne ulaže u mašinsko prevođenje nego da se radi na razvoju i unapređenju mašinskih pomagala prilikom prevođenja. Narednih desetak godina nije bilo značajnijih pomaka u procesu mašinskog prevođenja, ali se nije ni odustalo od njega.

Sedamdesetih godina pojavljuje se prva verzija Systran-a za englesko-ruske prevode. Godine 1976. javlja se i verzija za englesko-francuske prevode koja je poslužila i kao uzor za prevođenje sa drugih evropskih jezika. Kako su godine prolazile, prvobitni dizajn se unapređivao. Radilo se, također, i na povećanju kompatibilnosti analize i sinteze komponenata, te na smanjenju troškova.

Rane osamdesete godine obilježila je potreba za nastajanjem novih sistema za mašinsko prevođenje, te povećanje broja zemalja

³ Savjetnički komitet za automatsko procesuiranje jezika

zainteresiranih za ovakav vid prevođenja. Uz veću dostupnost različitih računara i softvera za obradu teksta na tržištu se otvara mogućnost za eksploataciju jeftinijih programa mašinskog prevođenja. Tokom osamdesetih nastavlja se rad na unapređivanju metoda i tehnika mašinskog prevođenja. Dominantna strategija postaje indirektno prevođenje, koje se vrši preko reprezentativnih posrednika uključujući i morfološku i sintaksičku i semantičku analizu i sintezu, a katkad čak i vanlingvistička znanja.

Devedesete godine donose nove prekretnice. Značajna novina je da se ni sintaksička ni semantička pravila više ne koriste u analizi tekstova i odabiru leksičkih ekvivalenata. Najveća inovacija u mašinskom prevođenju uopće jeste početak istraživanja prevođenja govora, što uključuje raspoznavanje govora, njegovu sintezu i module prevođenja.

Danas programi za mašinsko prevođenje postaju potreba širih narodnih masa. I dalje se radi na razvijanju novih komponenti takvih programa, ali se unapređuju i oni prvobitni programi. Prevođenje postaje gotovo nezamislivo bez programa koji, ako ništa drugo, onda barem daju smjernice prevodiocu za bolji i kvalitetniji prevod.

Literatura

Ivić. Milka, *Pravci u lingvistici* (četvrto izdanje), Državna založba Slovenije, Ljubljana, 1978.

Международная конференция по компьютерной лингвистике,
www.dialog-21.ru

Hutchins, John, *Publications on the history of machine translation*,
www.hutchinsweb.me.uk

MACHINE TRANSLATION

Summary

The term machine translation (MT) refers to computerized system responsible for the production of translation with or without human assistance. Although the ideal goal of MT systems may be to produce high-quality translation without human intervention at any stage, in practice this is not possible except in highly constrained situations.

There is now a multitude of translation software products on the market offering a wide range of systems, for mainframes, for workstations, and for personal computers; for the use of professional translators, and for casual use by the general public; for the production of high-quality documents, for the production of low quality 'gists' (for information analysis or review), for translating electronic mail, for translating Web pages, and so forth.