

JEZICI U OPASNOSTI: IZAZOV ZA JEZIČNE TEHNOLOGIJE

U ovome trenutku svjedočimo digitalnoj revoluciji koja korjenito utječe na našu komunikaciju i naše društvo. Najnoviji razvoj digitalnih i mrežnih komunikacijskih tehnologija ponekad se uspoređuju s Gutenbergovim izumom tiska pomičnim slovima. Što nam ta analogija može reći o budućnosti europskoga informacijskoga društva i o našim vlastitim jezicima?

Digitalna revolucija usporediva je s Gutenbergovim izumom tiska pomičnim slovima.

Nakon Gutenbergova izuma pravi su proboji u komunikaciji i razmjeni znanja postignuti pothvatima kao što je Lutherov prijevod Biblije na narodni jezik (ili u hrvatskome slučaju, glagoljički prvotisak Misala iz 1483. kao prve tiskanje knjige na hrvatskome jeziku). U nadolazećim stoljećima razvijeni su razni kulturni postupci koji su omogućili obradbu jezika i razmjenu znanja:

- pravopisno i gramatičko normiranje većih jezika omogućilo je brzu razmjenu novih znanstvenih ideja;
- uspostavljanje službenih jezika omogućilo je građanima komuni-kaciju unutar određenih (često političkih) granica;
- poučavanje jezika i prevođenje omogućilo je razmjenu preko jezičnih granica;
- stvaranje uredničkih i bibliografskih normi osiguralo je kakvoću tiskovina;

- stvaranjem različitih medija kao što su knjige, novine, radio, televizija i drugi, zadovoljavaju se komunikacijske potrebe pučanstva.

U zadnjih je dvadeset godina informacijska tehnologija omogućila olakšavanje i automatizaciju mnogih procesa:

- računalna priprema za tisak zamijenila je tipkanje i grafički slog;
- Microsoft PowerPoint zamijenio je projiciranje s prozirnica;
- e-pošta omogućuje odašiljanje i primanje dokumenata brže od telefaks uređaja;
- Skype nudi jeftine internetske telefonske pozive i održavanje virtualnih sastanaka;
- zajednički formati zapisa zvučnih i vizualnih podataka omogućuju jednostavnu razmjenu multimedij-skih sadržaja;
- tražilice omogućuju pristup www-stranicama na temelju pretrage uporabom ključnih riječi;
- mrežne usluge poput Google prevoditelja nude brze, ali zato približne prijevode;
- društvene mreže kao što su Facebook, Twitter i Google+ pospješuju komunikaciju, omogućuju suradnju i dijeljenje obavijesti.

Premda su takve aplikacije i usluge višestruko korisne, ipak još ne mogu podupirati u cijelosti održivo, više-

jezično europsko društvo u kojem informacije i robe mogu slobodno kolati.

2.1 JEZIČNE GRANICE KOČE EUROPSKO INFORMACIJSKO DRUŠTVO

Ne možemo točno predvidjeti kako će izgledati buduće informacijsko društvo, ali s velikom se vjerojatnošću može očekivati kako će revolucija u komunikacijskim tehnologijama na nove načine zbližiti ljude koji govore različite jezike. To će kod pojedinaca rezultirati potrebom za učenjem novih jezika, a kod razvijatelja aplikacija potrebom za stvaranjem novih tehnoloških aplikacija, ne bi li se osiguralo uzajamno razumijevanje i omogućio pristup razmjenjivome znanju. U globalnome gospodarskom i informacijskome prostoru raste interakcija između različitih jezika, govornika i sadržaja koja se odvija zahvaljujući novim vrstama medija. Trenutačna popularnost društvenih mreža (kao što su Wikipedia, Facebook, Twitter, YouTube i od nedavno Google+) predstavlja samo vršak ledene sante.

Globalizacija gospodarstva i informacijskoga prostora suočava nas sa sve više različitih jezika, govornika i sadržaja.

Danas bez ikakvih prepreka možemo u nekoliko sekunda na drugu stranu svijeta prebaciti gigabajte teksta prije nego što uopće shvatimo kako je on na jeziku koji uopće ne razumijemo. Prema nedavnome izvješću Europske komisije, 57% internetskih korisnika u Europi putem mreže kupuje robu i usluge na jeziku koji nije njihov vlastiti. Engleski je najčešći strani jezik, a slijede ga francuski, njemački i španjolski. 55% korisnika čita sadržaje na stranome jeziku dok ih se samo 35% koristi stranim jezikom za pisanje poruka e-pošte ili ostavljanje

komentara na *www-u* [3]. Prije nekoliko godina engleski je možda bio *lingua franca* *www-a* jer je većina sadržaja na *www-u* tada bila na engleskome, međutim, danas su se prilike u mnogome promijenile. Moglo bi se reći kako je količina sadržaja na drugim jezicima (osobito azijskim i na arapskome) upravo eksplodirala. Iznenadujuće je kako ovaj sveprisutni digitalni jaz prouzrokovao jezičnim preprekama još uvijek nije privukao dovoljno pozornosti u javnim raspravama; pa ipak, upravo nas on navodi na goruće pitanje: „Koji će europski jezici napredovati i održati se u umreženome informacijskome društvu i društvu znanja, a koji će biti osuđeni na izumiranje?“

2.2 OPASNOST ZA NAŠE JEZIKE

Dok je otkriće tiska neizmjereno pridonijelo razmjeni obavijesti u Europi, ono je istodobno dovelo do izumiranja mnogih europskih jezika. Kako se na regionalnim i manjinskim jezicima tiskalo rijetko, mnogi su jezici, npr. *cornwallski* ili *dalmatski*, bili ograničeni samo na govorni oblik komunikacije što je ograničilo doseg njihove uporabe. Hoće li Internet imati isti utjecaj na naše današnje jezike? Osamdesetak europskih jezika najbogatiji je i najvažnijih dio njezina kulturnoga nasljeđa i neizostavni dio jedinstvenoga društvenoga modela [4]. Dok će široko korišteni jezici kao engleski ili španjolski zacijelo održati svoju prisutnost na rastućem tržištu digitalnoga društva, mnogi bi europski jezici mogli biti isključeni iz digitalnih komunikacijskih kanala i postati nevažni za takvo umreženo društvo. Time bi se s jedne strane oslabio globalni položaj Europe, a s druge strane, takav bi razvoj bio u suprotnosti sa strateškim ciljem osiguravanja jednakoga sudjelovanja svakoga građanina EU bez obzira na njegov jezik.

Prema izvješću UNESCO-a o višejezičnosti jezici su ključni medij za ostvarivanje temeljnih ljudskih prava kao što su iskazivanje političkoga stava, obrazovanje i sudjelovanje u društvu [5].

Znatna raznolikost jezika u Europi jedno je od najvažnijih kulturnih dobara i bitan je dio europskoga uspjeha.

2.3 JEZIČNE SU TEHNOLOGIJE KLJUČNE POTPORNE TEHNOLOGIJE

U prethodnim razdobljima ulaganje u jezike usredotočivalo se na učenje jezika i prevođenje. Na primjer, prema nekim procjenama europsko tržište prevođenja, tumačenja, lokalizacije programske podrške i prevođenja www-stranica vrijedilo je 8,4 milijarde eura, a očekivao se njegov rast od 10% godišnje [6]. No čak i uz takve prognoze rasta postojeći kapaciteti nisu dovoljni za zadovoljenje potreba niti sadašnjih, a kamoli budućih potreba. Najuvjerljivije rješenje koje bi osiguralo širinu i dubinu uporabe jezika u sutrašnjoj Europi jest uporaba odgovarajućih tehnologija, upravo kao što rabimo razne tehnologije pri rješavanju npr. svojih transportnih ili energetskih potreba.

Jezične tehnologije usmjerene na sve vrste pisanoga ili govorenoga teksta, pomažu ljudima u suradnji, obavljanju poslova, razmjeni znanja i sudjelovanju u društvenim i političkim raspravama neovisno o stupnju usvojenih jezičnih ili računalnih vještina. One već često djeluju skrivene unutar složenih računalnih sustava koji nam pomažu kad:

- tražimo obavijesti korištenjem internetskih tražilica;
- provjeravamo pravopis ili gramatiku u obradniku teksta;
- gledamo preporuke za proizvode u on-line dućanima;
- slušamo glasovne upute navigacijskoga sustava;

- prevodimo www-stranice uporabom usluge on-line prevođenja.

Jezične tehnologije, o kojima se detaljnije govori u ovoj bijeloj knjizi, čine srž budućih inovativnih aplikacija. Jezične su tehnologije uobičajena potporna tehnologija unutar veće aplikacije kao što su navigacijski sustav ili tražilica. Ove bijele knjige prikazuju stanje osnovnih postignuća u jezičnim tehnologijama za svaki pojedini jezik.

Svi će europski jezici trebati jezične tehnologije koje će biti dostupne i prihvatljive.

Jezične se tehnologije sastoje od niza osnovnih aplikacija koje omogućuju uporabu i obradbu jezika i govora unutar složenijih aplikacijskih sustava. Svrha je ovih META-NET-ovih bijelih knjiga prikazati koliko su te osnovne potporne jezične tehnologije razvijene za svaki od europskih jezika. Europa treba robusne i dostupne jezične tehnologije za sve europske jezike. Kako bi održala svoj položaj globalnoga predvonika u inovacijama, Europa treba jezične tehnologije prilagođene svakome od svojih jezika, a one moraju biti robusne i dostupne ne bi li se što lakše integrirale u šire aplikacijsko okružje. Bez jezičnih tehnologija uskoro više ne ćemo moći postići stvarno interaktivno, multimedijско i višejezično korisničko iskustvo.

2.4 MOGUĆNOSTI JEZIČNIH TEHNOLOGIJA

U svijetu tiska tehnološki je proboj predstavljalo brzo umnožavanje slike teksta uporabom tiskarskoga stroja pomičnim slovima. Međutim, istodobno je ljudima prepušten težak posao traženja, pristupa, prevođenja i sažimanja znanja širenoga i prenošenoga tako umnoženim tekstovima. Morali smo čekati do Edisona koji je otkrio

kako zabilježiti govor, ali ponovno je njegova tehnologija stvarala analogne preslike.

Jezične nam tehnologije danas omogućuju pojednostavljenje i automatizaciju postupaka kao što su strojno prevođenje, stvaranje sadržaja, i upravljanje znanjem na svim europskim jezicima. Jezične tehnologije također stoje u pozadini intuitivnih govornih sučelja za kućansku elektroniku, strojeve, vozila, računala i robote. Premda već postoje mnogi prototipovi, komercijalne i industrijske primjene su još uvijek u ranim stupnjevima razvoja. Međutim, neka su nedavna postignuća u istraživanjima i razvoju otvorila jedinstvene mogućnosti. Na primjer, strojnim prevođenjem već se mogu dobiti prijevodi prihvatljive točnosti unutar posebnih područja, dok istodobno neke eksperimentalne aplikacije već omogućuju dohvat višejezičnih obavijesti i upravljanje znanjem, kao i proizvodnju sadržaja istodobno na mnogim europskim jezicima.

Višejezičnost je pravilo, a ne iznimka.

Kao što je to bio slučaj i s mnogim drugim tehnologijama, prve su jezične aplikacije, kao što su govorna korisnička sučelja i razgovorni sustavi, ponajprije razvijene u visokospecijaliziranim područjima uporabe, ali nerijetko uz ograničenu kakvoću. Pa ipak i za takve aplikacije postoje ogromne tržišne mogućnosti u obrazovanju i zabavnoj industriji s uključivanjem jezičnih tehnologija u računalne igre, obrazovne sustave, knjižnice, simulacijske sustave i sustave za učenje. Mobilne obavijesne usluge, strojno potpomognuti programi učenja jezika, okružja za e-učenje, alati za samoprocjenu i sustavi za pronalaženje plagijata samo su još neki od primjera gdje jezične tehnologije igraju značajnu ulogu. Popularnost društvenih mreža kao što su Twitter ili Facebook nagoviještaju dodatne potrebe za razrađenim jezičnih tehnologijama koje bi mogle nadgledati poruke, sažimati rasprave, predlagati opća kretanja u stavovima

i mišljenjima sudionika, otkrivati emocionalne afinitete, uočavati kršenje autorskih prava ili pratiti zloporabu.

Jezične tehnologije Europskoj uniji pružaju upravo nesagledive ekonomski i kulturno značajne mogućnosti. One mogu pomoći u problemima koje donosi višejezičnost u Europi s obzirom da različiti jezici prirodno supostojе u europskom poslovanju, ustanovama i školama. Građani žele komunicirati onkraj jezičnih granica koje još uvijek postoje na europskome zajedničkom tržištu, a upravo bi jezične tehnologije mogle pomoći u nadilaženju tih preostalih prepreka uz potpomaganje slobodne i otvorene uporabe bilo kojega jezika. Nadalje, inovativne, višejezične jezične tehnologije nama bi Europljanima također pomogle u komunikaciji s našim globalnim partnerima, a njima bi pomogle pri razvoju jezičnih tehnologija u njihovim višejezičnim zajednicama. Jezične tehnologije postaju svojevrstne „potporne“ tehnologije koje omogućuju nadići „prepreke“ jezične raznolikosti i čine različite jezične zajednice međusobno pristupačnijima. Konačno, jedno od aktivnijih područja istraživanja jest uporaba jezičnih tehnologija u spasilačkim operacijama u nesrećenim područjima. U takvim okružjima visoke opasnosti točnost prijevoda može značiti razliku između života i smrti: u budućnosti će inteligentni roboti s višejezičnim sposobnostima moći ljudske spašavati živote.

2.5 IZAZOVI KOJI STOJE PRED JEZIČNIM TEHNOLOGIJAMA

Premda su u nekoliko proteklih godina jezične tehnologije napravile znatan napredak, trenutačan je tempo tehnološkoga napretka i stvaranja novih proizvoda prespor. Jezične tehnologije, koje su već u širokoj uporabi, kao što su provjernici pravopisa ili gramatike u obradnicima teksta, uobičajeno su jednojezične, a dostupne su samo za ograničen broj jezika.

Usluge *on-line* strojnoga prevođenja, premda korisne za stvaranje općega dojma o čemu je u nekom dokumentu riječ, bore se s mnogim poteškoćama kad su nam potrebni visokokvalitetni i potpuni prijevodi. Zahvaljujući složenosti prirodnih jezika, njihovo modeliranje u obliku računalnih programa i provjera u stvarnome životu, dugotrajan je i skup posao koji zahtijeva stalnu financijsku potporu. Europa mora zadržati svoju vodeću ulogu u sučeljavanju s tehnološkim izazovima višjezičnoga društva otkrivanjem novih načina za ubrzavanje razvoja na tom području. To može uključiti i raznorodne pristupe kao što su napredak u računarstvu, ali i tehnike distribuirane ljudske potpore.

Tehnološki se napredak mora ubrzati.

2.6 USVAJANJE JEZIKA KOD LJUDI I STROJEVA

Kako bismo prikazali na koji se način računala nose s prirodnim jezikom i zašto je iznimno težak zadatak programirati ih za obradbu različitih jezika, pogledajmo na kratko kako ljudi usvajaju svoj prvi i ostale jezike, a potom pogledajmo kako djeluju jezičnotehnološki sustavi. Ljudi usvajaju jezične sposobnosti na dva različita načina. Mala djeca usvajaju jezik slušanjem i praćenjem interakcija između svojih roditelja, braće i sestara te ostalih u obitelji. U otprilike dvogodišnjoj dobi sami počinju proizvoditi prve riječi i kratke fraze. To je moguće samo zato jer ljudski rod već ima genetsku predispoziciju za imitiranje i racionalizaciju onoga što čuju.

Učenje drugoga jezika u kasnijoj dobi obično traži više kognitivnoga napora ukoliko dijete nije uronjeno u jezičnu zajednicu izvornih govornika. U školskoj se dobi strani jezici obično usvajaju učenjem njihove gramatičke strukture, rječnika i pravopisa iz knjiga i obrazovnih ma-

terijala koji opisuju jezično znanje u obliku apstraktnih pravila, tablica i primjera.

Ljudi usvajaju jezičnu sposobnost na dva različita načina: učeći na primjerima i učeći temeljna jezičnih pravila.

Kod jezičnotehnoloških sustava dvije su osnovne vrste usvajanja jezične sposobnosti, na sličan način kao i kod ljudi. Statistički (ili podatkovno utemeljeni) pristupi stječu jezično znanje iz golemih zbirki pojedinačnih tekstnih primjera. Dok je dovoljno koristiti tekst na jednome jeziku za treniranje npr. pravopisnoga provjerenika, usporedni tekstovi na dva (ili više) jezika potrebni su za treniranje strojnoprevoditeljskih sustava. Algoritmi strojnoga učenja prepoznaju se obrasci kako se pojedine riječi, kratke fraze ili čitave rečenice prevode s jednoga jezika na drugi.

Međutim, za takve statističke pristupe potrebni su milijuni usporednih rečenica za povećanje kakvoće izvedbe takvih sustava. To je jedan od razloga zašto sastavljači tražilica teže skupiti što je više moguće pisanoga teksta. Provjera pravopisa u obradnicima teksta i usluge kao što su Google Search ili Google Translate počivaju u cijelosti na statističkim pristupima. Prednost statističkih sustava je što strojevi uče brzo i u kontinuiranim ciklusima treniranja premda kakvoća takvih sustava nerijetko oscilira.

Sustavi temeljeni na pravilima predstavljaju drugu osnovnu vrstu jezičnih tehnologija, a time i sustava za strojno prevođenje. Stručnjaci s područja jezikoslovlja, računalnoga jezikoslovlja i računarstva moraju kodirati gramatičke analise (prijevodna pravila) i sastaviti popise riječi (rječnike). Izgradnja sustava temeljenoga na pravilima iznimno je vremenski i poslovno zahtjevnija, a ne može se provesti bez visokospecijaliziranih stručnjaka. Neki od vodećih strojnoprevoditeljskih sustava temeljenih na pravilima u stalnom su razvoju već dvadesetak godina. Prednost sustava temeljenih na pravilima je u tome

što stručnjaci imaju mogućnost istančanijega upravljanja obradbom jezika. Zbog toga je moguće sustavno ispravljati pogreške u programskoj podršci i pružati korisniku detaljne povratne obavijesti, osobito kad se na pravilima temeljeni sustavi koriste za učenje jezika. Na žalost, zbog svoje visoke cijene, jezične tehnologije temeljene na pravilima isplative su samo za jezike s velikim brojem govornika.

Kako su prednosti i nedostaci statističkih pristupa i pristupa temeljenih na pravilima međusobno nadopunjujući, trenutačna se istraživanja usredotočuju na hibridne pristupe koji kombiniraju ova dva pristupa. Međutim, hibridni sustavi do sad su bili znatno manje uspješni u industrijskim uvjetima nego u istraživačkim laboratorijima.

Dvije osnovne vrste jezičnih tehnologija usvajaju jezično znanje na sličan način.

Kao što smo vidjeli u ovome poglavlju, mnoge aplikacije, koje se svakodnevno na široko koriste u današnjem informacijskom društvu, u mnogome ovise o jezičnim tehnologijama, osobito europskome gospodarskom i informacijskom prostoru. Premda su te tehnologije postigle znatan napredak u nekoliko proteklih godina, još uvijek postoje ogromne mogućnosti za poboljšanje kvalitete jezičnotehnoloških sustava. U sljedećem ćemo poglavlju opisati ulogu hrvatskoga jezika u europskom informacijskom društvu i procijeniti trenutačno stanje jezičnih tehnologija za hrvatski jezik.

The Croatian Language in the Digital Age

Rehm, G.; Uszkoreit, H. (Eds.)

2012, VI, 93 p. 38 illus., 28 illus. in color., Softcover

ISBN: 978-3-642-30881-9