

TVEGANJE ZA NAŠE JEZIKE IN IZZIV ZA JEZIKOVNE TEHNOLOGIJE

Priča smo digitalni revoluciji, ki korenito spreminja komunikacijske navade in družbo nasploh. Naj-novejše dosežke na področju digitalnih informacijskih in komunikacijskih tehnologij včasih primerjajo z Gutenbergovim izumom tiskarskega stroja. Toda kaj nam ta primerjava lahko pove o prihodnosti evropske informacijske družbe, zlasti pa o naših jezikih?

Priča smo digitalni revoluciji, ki jo je mogoče primerjati z Gutenbergovim izumom tiskarskega stroja.

Po Gutenbergovem izumu je šele ob dejanjih, kot je bil Lutherjev prevod Biblije v vernakularni jezik, sledil preboj v komunikaciji in izmenjavi znanja. V naslednjih stoletjih so se razvile kulturne tehnike, ki so izpopolnile procesiranje jezika in izmenjavo znanja:

- pravopisna in slovnična standardizacija večjih jezikov je omogočila hitro širjenje novih znanstvenih in intelektualnih idej;
- razvoj uradnih jezikov v okviru določenih (pogosto političnih) meja je njihovim prebivalcem olajšal komuniciranje;
- s poučevanjem jezikov in prevajanjem so bile ustvarjene možnosti za izmenjavo med jeziki;
- nastanek uredniških in bibliografskih smernic je zagotovil kvaliteto in dostopnost tiskanega gradiva;

- nastanek različnih medijev, kot so časopisi, radio, televizija, knjige in ostali formati, je zadovoljil različne komunikacijske potrebe.

V zadnjih dvajsetih letih je informacijska tehnologija prispevala k avtomatizaciji in izboljšanju mnogih od omenjenih procesov:

- programi za namizno založništvo so nadomestili tipkanje in tiskarsko stavljenje;
- Microsoftov PowerPoint je nadomestil grafoskope in prosojnice;
- e-pošta omogoča hitrejšo pošiljanje in prejemanje dokumentov kot faksirni stroj;
- Skype ponuja poceni internetne telefonske pogovore in gosti virtualne sestanke;
- formati za kodiranje avdia in videa omogočajo preprosto izmenjavo multimedijskih vsebin;
- spletni iskalniki zagotavljajo dostopnost spletnih strani na podlagi ključnih besed;
- spletni servisi kot Google Translate ponujajo hitre približne prevode;
- okolja družabnih omrežij, kot so Facebook, Twitter in Google+ lajšajo komunikacijo, sodelovanje in izmenjavo informacij.

Čeprav so ta orodja in programi v veliko pomoč, še vedno niso zmožni podpirati trajnostno naravnane, večjezične evropske družbe, v kateri je vsem pripadnikom omogočen prost pretok informacij in blaga.

2.1 JEZIKOVNE MEJE OVIRAJO EVROPSKO INFORMACIJSKO DRUŽBO

Nemogoče je natančno napovedati, kakšna bo videti bodoča informacijska družba. Vendar je hkrati mogoče reči, da revolucija na področju komunikacijskih tehnologij na nov način združuje ljudi, ki govorijo različne jezike. Posamezniki so s tem izpostavljeni pritisku, da se učijo druge jezike, razvijalci pa v še večji meri temu, da ustvarijo nove tehnološke izdelke, ki omogočajo medsebojno razumevanje in dostop do skupnega znanja. V globalnem gospodarskem in informacijskem prostoru je z novimi vrstami medijev stik med več jeziki, govorci in vsebinami vse hitrejši. Trenutna popularnost družabnih medijev (Wikipedia, Facebook, Twitter, YouTube in v zadnjem času Google+) je le vrh ledene gore.

Zaradi globalnega gospodarskega in informacijskega prostora smo soočeni z vedno več jeziki, govorci in vsebinami.

Danes lahko prenašamo gigabajte besedila okrog sveta v nekaj sekundah, še preden se zavemo, da je besedilo v jeziku, ki ga ne razumemo. Sodeč po nedavni raziskavi Evropske komisije 57 % uporabnikov interneta v Evropi kupuje blago in storitve v jeziku, ki ni njihov materni jezik. (Najbolj pogost tuji jezik je angleščina, ki mu sledijo francoščina, nemščina in španščina.) 55 % uporabnikov bere vsebine v tujem jeziku, a le 35 % uporablja tuji jezik pri pisanju elektronskih sporočil ali spletnih komentarjev [3]. Pred nekaj leti je bila angleščina morda res lingua franca na spletu – velika večina spletnih vsebin je bila v angleščini – toda stanje se je zdaj korenito spremenilo. Količina spletnih vsebin v drugih evropskih jezikih (tudi azijskih in jezikih Srednjega vzhoda) je skokovito narasla.

Začuda ta vseprisotna digitalna vrzel kot posledica jezikovnih meja v javnosti ni zbudila veliko pozornosti; kljub temu pa izpostavlja zelo pereče vprašanje: kateri evropski jeziki bodo v omreženi informacijski družbi znanja dobro uspevali in kateri so obsojeni na izginotje?

2.2 NAŠI JEZIKI SO OGROŽENI

Medtem ko je tiskarski stroj prispeval k povečanju obsega izmenjave informacij v Evropi, je obenem pripomogel tudi k izginotju mnogih evropskih jezikov. Regionalni in manjšinski jeziki so redko prišli do tiskane oblike in jeziki, kot sta kornijski ali dalmatinski, so bili omejeni le na prenos govorne oblike, kar je povzročilo, da so bili rabljeni manj in manj. Bo internet imel enak vpliv na naše jezike?

Raznolikost evropskih jezikov je ena od najbogatejših in najpomembnejših kulturnih dragocenosti Evrope.

Približno 80 evropskih jezikov je ena najbogatejših in najpomembnejših kulturnih dragocenosti Evrope in ključni del njenega edinstvenega družbenega modela [4]. Medtem ko bosta angleščina in španščina zagotovo preživela na nastajajočem digitalnem tržišču, mnogi evropski jeziki v omreženi družbi lahko postanejo nepomembni. To bi ošibilo globalni položaj Evrope in bi bilo v nasprotju s strateškim ciljem zagotavljanja možnosti enakopravne udeležbe za vse državljane ne glede na jezik.

Kot ugotavlja UNESCO v poročilu o večjezičnosti, je jezik osnovni medij uživanja enakih pravic, kot je pravica do političnega izražanja, izobrazbe in sodelovanja pri javnih zadevah [5].

2.3 JEZIKOVNE TEHNOLOGIJE SO KLJUČNE PODPORNE TEHNOLOGIJE

V preteklosti se je investiranje v ohranjanje jezika osredotočalo na poučevanje jezika in prevajanje. V skladu z eno od ocen je bil v letu 2008 evropski trg prevajanja, tolmačenja, lokalizacije programske opreme in globalizacije spletnih strani vreden 8,4 milijarde evrov, pričakuje pa se, da bo letno naraščal za 10 odstotkov [6]. Ta številka pokriva le manjši del sedanjih in prihodnjih potreb po medjezikovnem komuniciranju. Najbolj prepričljiva rešitev, ki bi zagotovila tako obseg kot doseg rabe jezikov v Evropi tudi v prihodnje, je uporaba primerne tehnologije, podobno kot uporabljamo tehnologijo pri transportu in energetiki, ali denimo pri pomoči osebam s posebnimi potrebami.

Evropa potrebuje robustne in dostopne jezikovne tehnologije za vse evropske jezike.

Digitalne jezikovne tehnologije (katerih cilj so vse oblike pisnega in govornega jezika) ljudem pomagajo pri sodelovanju, poslovanju, izmenjavi znanja in udeleževanju v družabnih in političnih razpravah ne glede na jezikovne meje ali obvladovanje računalnika. Tehnologije so pogosto nevidne kot del zapletenih računalniških sistemov, ki nam pomagajo:

- najti informacije s pomočjo spletnih iskalnikov;
- preverjati črkovanje ali slovnično ustreznost v urejevalnikih besedil;
- pregledovati priporočila o izdelkih v spletnih trgovinah;
- poslušati govorna navodila v navigacijskih sistemih v avtomobilu;
- prevajati spletne strani s spletnimi prevajalniki.

Jezikovne tehnologije sestavlja večje število jedrnih aplikacij, ki omogočajo procesiranje jezika v okviru večjih programskih sistemov. Namen zbirke Bela knjiga v projektu META-NET je preverjanje stanja jedrnih tehnologij za vse evropske jezike.

Evropa bo potrebovala jezikovne tehnologije, prilagojene za vse evropske jezike, ki bodo hkrati robustne, dostopne in polno integrirane v ključna programska okolja, če želimo obdržati svoj položaj v prvih vrstah. Brez jezikovnih tehnologij ne bomo mogli doživeti resnično učinkovite interaktivne, večpredstavne in večjezične uporabniške izkušnje v bližnji prihodnosti.

2.4 PRILOŽNOSTI ZA JEZIKOVNE TEHNOLOGIJE

V svetu tiska je hitro razmnoževanje slike besedila (knjižne strani) predstavljalo pravi tehnološki preboj – ob uporabi tiskarskega stroja na primeren pogon. Kljub temu so ljudje še vedno morali opravljati naporno delo pregledovanja, branja, prevajanja ali povzemanja znanja. Treba je bilo počakati na Edisona, da je bilo mogoče shraniti govorjeni jezik – toda njegova tehnologija je omogočala zgolj izdelavo analognih kopij.

Z digitalnimi jezikovnimi tehnologijami je zdaj mogoče avtomatizirati procese prevajanja, ustvarjanja vsebin in upravljanja z znanjem za vse evropske jezike. Z njimi je mogoče tudi opremljati intuitivne tekstovne ali govorne vmesnike v gospodinjskih aparatih, strojih, vozilih, računalnikih in robotih. Dejanska komercialna in industrijska uporaba tehnologij je še v zgodnjih fazah razvoja, toda z raziskovalnimi dosežki se odpira pravo okno priložnosti. Strojno prevajanje, na primer, je na omejenih področjih že zadovoljivo natančno, poskusne aplikacije pa zagotavljajo večjezično upravljanje z informacijami in znanjem za mnoge evropske jezike.

Tako kot pri večini tehnologij so bile prve uporabne aplikacije, kot so govorni uporabniški vmesniki ali sistemi

dialoga, razvite za ozko specializirana področja in njihova uporabnost je bila pogosto dokaj omejena. Tržne priložnosti pa se odpirajo v izobraževalni in zabavni industriji z vključevanjem jezikovnih tehnologij v igre, spletne informacije o kulturni dediščini, zabavne izobraževalne pakete (*edutainment*), knjižnice, simulacijska okolja in programe za usposabljanje. Mobilne informacijske storitve, programska oprema za učenje jezikov s pomočjo računalnika, e-izobraževalna okolja, orodja za samoevalvacijo in programi za odkrivanje plagatorstva so le nekatera od področij, kjer jezikovne tehnologije lahko odigrajo pomembno vlogo. Popularnost družabnih omrežij, npr. Twitterja in Facebooka, kaže, da obstajajo potrebe po zahtevnejših jezikovnih tehnologijah, ki omogočajo spremljanje objav, povzemanje razprav, detekcijo mnenjskih trendov, zaznavanje čustvenih odzivov, prepoznavanje kršenja avtorskih pravic ali spremljanje zlorab.

Jezikovne tehnologije so izjemna priložnost za Evropsko unijo. Pomagajo lahko pri razreševanju zahtevnih vprašanj evropske večjezičnosti – dejstva, da v evropskih poslovnih okoljih, organizacijah in šolah naravno sobivajo različni jeziki. Toda državljani se morajo sporazumevati tudi izven jezikovnih meja, ki prečijo evropski skupni trg, in jezikovne tehnologije lahko pripomorejo pri odstranjevanju te zadnje ovire, pri čemer hkrati tudi podpirajo svobodno in odprto rabo posameznih jezikov.

Jezikovne tehnologije pomagajo pri premagovanju "hendikepiranosti" zaradi jezikovne različnosti.

Če se ozremo celo dlje, bodo inovativne evropske večjezične jezikovne tehnologije postavile merila za naše partnerje po svetu, ko bodo ti začeli vzpostavljati svoje večjezične skupnosti. Jezikovne tehnologije je mogoče dojeti kot eno od oblik "tehnologij za izboljšanje kakovosti

življenja", ki pomagajo pri premagovanju "hendikepiranosti" zaradi jezikovne različnosti in jezikovnim skupnostim omogočajo medsebojni dostop.

Eno od dejavnih področij raziskovanja je nenazadnje tudi uporaba jezikovnih tehnologij pri reševalnih operacijah na prizadetih območjih, kjer uspešno delovanje lahko odloča o življenju in smrti. Bodoči inteligentni roboti s sposobnostjo večjezične komunikacije bodo dejansko lahko reševali tudi življenja.

2.5 IZZIVI ZA JEZIKOVNE TEHNOLOGIJE

Čeprav so jezikovne tehnologije v zadnjih letih precej napredovale, sta tehnološki razvoj in uveljavljanje inovativnih proizvodov prepočasna. Splošno razširjene tehnologije, kot so črkovalniki in slovnični moduli v urejevalnikih besedil, so tipično enojezični in na voljo le za peščico jezikov.

Tehnološki razvoj je trenutno prepočasen.

Spletni strojni prevajalni sistemi se ob zahtevi po natančnih in dokončnih prevodih spopadajo z množico težav, čeprav so uporabni za hitro tvorjenje približne vsebine dokumentov. Zaradi zapletenosti človeškega jezika je računalniško modeliranje jezikov in testiranje v realnih okoliščinah dolg in drag proces, ki potrebuje dolgoročno finančno podporo. Ob soočanju z izzivi večjezične družbe mora Evropa vztrajati pri svoji pionirski vlogi in si zamisliti nove metode pospeševanja razvoja po svojem celotnem zemljevidu. Te metode lahko vključujejo tako napredek znotraj računalništva kot tudi tehnike, kakršna je izkoriščanje moči množice (*crowdsourcing*).

2.6 USVAJANJE JEZIKA PRI LJUDEH IN STROJIH

Da bi prikazali, kako računalniki obvladujejo jezik in zakaj jih je težko sprogramirati tako, da bi ga uspešno uporabljali, najprej pogledajmo, kako ljudje usvajajo prvi ali drugi jezik, potem pa še načine, kako delujejo jezikovnotehnološki sistemi.

Ljudje osvojijo jezikovne spretnosti na dva načina. Dolenjški jezik usvajajo s spremljanjem realne komunikacije med starši, brati in sestrami ali drugimi družinskimi člani. Od drugega leta naprej otroci začnejo izgovarjati prve besede in kratke zveze. To je možno zato, ker so ljudje genetsko nagnjeni k oponašanju in kasnejšemu racionaliziranju tega, kar slišijo.

Učenje drugega jezika v kasnejših letih zahteva več napora, predvsem zato, ker otrok ni del jezikovne skupnosti maternih govorcev. V šoli usvajanje tujih jezikov običajno poteka ob učenju slovničnih struktur, besedišča in izgovorjave, z uporabo vaj, ki opisujejo jezikovna dejstva v obliki abstraktnih pravil, tabel in primerov. Učenje tujih jezikov torej z leti postane vse težje.

Ljudje usvajajo jezikovne spretnosti na dva načina: z učenjem primerov in z učenjem prikritih jezikovnih pravil.

Pri dveh glavnih tipih jezikovnotehnoloških sistemov "usvajanje" jezikovnih zmožnosti poteka na podoben način. Pri statističnem (ali "podatkovnem") pristopu jezikovno znanje izvira iz ogromnih zbirk konkretnih primerov besedil. Za učenje strojnih mehanizmov za potrebe črkovalnikov, na primer, zadostujejo besedila v enem jeziku, za učenje strojnih prevajalnikov pa morajo biti na voljo vzporedna besedila v dveh (ali več) jezikih. Algoritmi strojnega učenja potem "povzamejo" vzorce, s katerimi so prevedene besede, kratke zveze ali celi stavki.

Pri statističnem pristopu moramo imeti na voljo na milijone stavkov, kakovost pa narašča s količino analiziranih besedil. To je eden od razlogov, zakaj ponudniki spletnih iskalnikov skušajo zbrati kolikor je mogoče veliko gradiva. Črkovalniki v urejevalnikih besedil in servisi, kot sta Google Search ali Google Translate, se vsi opirajo na statistični pristop. Velika prednost statistike je ta, da se strojni mehanizmi učijo hitro v ponavljajočih se učnih ciklih, kljub temu da kvaliteta potem lahko variira na nepredvidljiv način.

Drugi pristop k jezikovnim tehnologijam in predvsem strojnemu prevajanju je izdelava sistemov na podlagi pravil (*rule-based systems*). Strokovnjaki s področja jezikoslovja, računalniškega jezikoslovja in računalništva morajo najprej pretvoriti slovnične analize v sistem (pravila prevajanja) in sestaviti spiske besed (leksikone). To zahteva veliko časa in izjemen napor. Nekateri od vodilnih strojnih prevajalnikov na podlagi pravil so v procesu nenehnega dopolnjevanja že več kot dvajset let. Velika prednost sistemov na podlagi pravil pa je ta, da imajo strokovnjaki podrobnejši nadzor nad procesiranjem jezika. To omogoča, da lahko napake v sistemu sistematično odpravljajo in uporabniku ponudijo podrobno povratno informacijo, predvsem kadar se taki sistemi uporabljajo za poučevanje jezika. Zaradi visokih stroškov potrebnega dela pa so bile jezikovne tehnologije, ki temeljijo na pravilih, razvite le za največje jezike.

Ker se prednosti in slabosti pri obeh pristopih, statističnem in pri sistemih na podlagi pravil, dopolnjujejo, se raziskave trenutno osredotočajo na hibridne pristope, ki kombinirajo obe metodologiji. Toda ti pristopi so bili v industrijskih aplikacijah precej manj uspešni kot v raziskovalnih laboratorijih.

Kot smo videli v tem poglavju, mnoge aplikacije, ki jih v današnji informacijski družbi vsakodnevno uporabljamo, ne morejo delovati brez uporabe jezikovnih tehnologij. To zaradi večjezične skupnosti še to-

liko bolj drži za evropski gospodarski in informacijski prostor. Čeprav je bil pri jezikovnih tehnologijah v zadnjih nekaj letih narejen precejšen korak naprej, imajo jezikovnotehnološki sistemi še vedno precejšnje možnosti za izboljšanje kvalitete.

V naslednjem poglavju opisujemo vlogo slovenščine v evropski informacijski družbi in trenutno stanje jezikovnih tehnologij za slovenščino.

The Slovene Language in the Digital Age

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

2012, VI, 79 p. 27 illus., 24 illus. in color., Softcover

ISBN: 978-3-642-30635-8