

ОПАСНОСТ ПО НАШЕ ЈЕЗИКЕ И ИЗАЗОВИ ПРЕД ЈЕЗИЧКИМ ТЕХНОЛОГИЈАМА

Сведоци смо дигиталне револуције која драматично утиче на комуникацију и друштво. Најновија достигнућа на подручју дигиталних информационих и комуникационих технологија могу да се пореде са Гутенберговим изумом штампарске пресе. Шта ова аналогија може да нам каже о будућности европског информационог друштва и посебно наших језика?

Дигитална револуција се може упоредити са Гутенберговим изумом штампарске пресе.

После Гутенберговог изума, стварне продоре у комуникацији и размени знања остварила су дела као што је био Лутеров превод Библије. Током наредних векова у култури су развијене технике за бољу обраду језика и размену знања:

- правописна и граматичка стандардизација већих језика омогућила је брзо преношење нових научних и интелектуалних идеја;
- развој званичних језика омогућио је становницима да међусобно комуницирају унутар (често политичких) граница;
- подучавање и превођење језика омогућило је комуникацију која превазилази језичке границе;
- стварање уредничких и библиографских упутстава обезбедило је квалитет и расположивост штампаног материјала;

- појавом различитих медија, као што су новине, радио, телевизија, књиге и други облици, задовољене су различите потребе за комуникацијом.

У последњих двадесет година информационе технологије помогле су да се аутоматизују и поједноставе бројни процеси:

- софтвер за стоно издаваштво је заменио писаћу машину и слагање текста;
- Microsoft PowerPoint је заменио графоскопске фолије;
- документа се шаљу и примају електронском поштом често брже него факс машином;
- Скуре се користи за јефтино телефонирање преко интернета и организовање виртуелних састанака;
- формати аудио и видео записа олакшавају размену мултимедијалних садржаја;
- претраживачке машине обезбеђују приступ веб странама преко кључних речи;
- *онлајн* услуге као што је Google Translate производе брз и приближни превод;
- друштвени медији као што су Facebook, Twitter, и Google+ поједоставаљују комуникацију, сарадњу и размену информација.

Мада су сви ови алати и апликације од велике помоћи, они још нису довољни да остваре одрживо ви-

шејезично европско друштво у коме је свакоме могућ слободан приступ информацијама и роби.

2.1 ЈЕЗИЧКЕ ГРАНИЦЕ ПРЕДСТАВЉАЈУ СМЕТЊУ ЗА ЕВРОПСКО ИНФОРМАЦИОНО ДРУШТВО

Не можемо тачно да предвидимо како ће изгледати будуће информационо друштво. Али постоје велики изгледи да ће револуција у комуникационој технологији повезати на нове начине људе који говоре различитим језицима. То ствара притисак на појединце да уче нове језике и посебно на развојне тимове да стварају нове технолошке производе који ће обезбедити узајамно разумевање и приступ заједничком знању. У глобалном привредном и информационом простору, нови типови медија омогућавају држу размену у којој учествују бројни језици, говорници и садржаји. Актуелна популарност друштвених медија као што је Википедија (Wikipedia), Фејсбук (Facebook), Твитер (Twitter), Јутјуб (YouTube) и, однедавно, Гугл+ (Google+) само је врх леденог брега.

Глобална економија и информациони простор суочавају нас са различитим језицима, говорницима и садржајима.

Данас можемо да допремимо гигабајте текста из целог света за свега неколико секунди, пре него што схватимо да је текст на језику који не разумемо. Према недавном извештају Европске комисије, 57% корисника интернета наручује робу и услуге на језицима који им нису матерњи. (Енглески је најчешћи страни језик а за њим следе француски, немачки и шпански.) 55% корисника чита садржаје на страном језику, док само 35% користи страни језик за писање

електронских порука или коментара на мрежи [3]. До пре неколико година енглески језик је био lingua franca веба – огромна већина садржаја на вебу била је на енглеском језику. Ситуација је данас драстично промењена. Количина онлајн садржаја и на другим европским (као и азијским и средњоисточним) језицима доживела је праву експлозију.

Ова свеprisутна дигитална подељеност као последица језичких граница изазвала је изненађујуће мало пажње у јавности. Ипак, она поставља неодложно питање: „Који ће европски језици напредовати и опстати у умреженом друштву информација и знања, а који су осуђени да нестану?”

2.2 НАШИ ЈЕЗИЦИ СУ УГРОЖЕНИ

Проналазак штампарске пресе допринео је повећању обима размене информација у Европи, али је такође довео до изумирања многих европских језика. На регионалним језицима и језицима мањина ретко се штампало. Ово је довело до тога да су многи језици, као што су корнвалски или далматски, сведени на усмене облике преношења, што је довело до тога да су се све мање користили. Да ли ће интернет довести до истих последица када су наши језици у питању?

Разноврсност језика у Европи јесте једно од њених најдрагоценијих и најзначајнијих културних добара.

Око 80 језика Европе представља једно од њених најбогатијих и најважнијих културних добара и суштински чинилац њеног друштвеног модела [4]. Док ће језици као што су енглески или шпански вероватно преживети на дигиталном тржишту у настајању, многи би европски језици могли да постану не-

битни у умреженом друштву. Овакав развој ствари ослабио би позицију Европе у свету, а то би било и у супротности са стратешким циљем да се обезбеди подједнако учешће за сваког становника Европе без обзира на језик. Према УНЕСКО-вом извештају о вишејезичности, језици су суштински медијум за уживање основних људских права као што су изражавање политичких опредељења, образовање и учествовање у друштву [5].

2.3 ЈЕЗИЧКЕ ТЕХНОЛОГИЈЕ СУ КЉУЧНЕ ПОТПОРНЕ ТЕХНОЛОГИЈЕ

У прошлости, инвестиције у очување језика усмераване су на учење и превођење језика. Према неким проценама, европско тржиште превода, усменог превођења, локализације софтвера и глобализације страница на вебу 2008. год. је износило 8,4 милијарде евра, са очекиваним растом од 10% годишње [6]. Тај износ покрива само мањи део садашњих и будућих потреба у међујезичкој комуникацији. Најубедљивије решење које ће обезбедити дубину и ширину коришћења језика у Европи сутрашњице јесте коришћење одговарајуће технологије на исти начин на који се користе технологије за потребе транспорта, у енергетици или за особе са посебним потребама.

Дигиталне језичке технологије (којима је циљ да овладају свим облицима писаног и говорног језика) помажу људима да сарађују, послују, размењују знање и учествују у политичким и друштвеним дебатама без обзира на језичке баријере или њихове информатичке вештине. Технологије су често невидљиве у сложеним софтверским системима који нам помажу да:

- нађемо информацију помоћу претраживачких машина;

- користимо правописне и граматичке провере у програмима за обраду текста;
- разгледамо препоруке о производима у мрежним продавницама;
- саслушамо гласовна упутства у навигационим системима аутомобила;
- преводимо веб странице помоћу мрежних преводаца.

Језичке технологије чине већи број основних апликација које омогућавају обраду језика у оквирима ширих програмских система. Сврха белих књига, састављених у оквиру МЕТА-НЕТ-а, јесте да се опише степен развоја основних језичких технологија за сваки од европских језика.

Европи су потребне робусне и приступачне језичке технологије за све европске језике.

Да би одржала свој положај у првим редовима светске иновативности, Европи ће бити потребне језичке технологије прилагођене свим европским језицима које ће бити робусне, свима приступачне и потпуно интегрисане у кључна софтверска решења. Без језичких технологија нећемо бити у стању да у скорој будућности остваримо одиста ефикасно интерактивно, мултимедијално и вишејезично корисничко искуство.

2.4 МОГУЋНОСТИ ЈЕЗИЧКИХ ТЕХНОЛОГИЈА

У свету штампане речи, брзо умножавање слике текста (странице) коришћењем штампарске пресе представљало је технолошки пробој. Људима је био препуштен тешки посао прегледања, читања, превођења и апстраховања знања. Требало је да дочекамо Едисона да бисмо снимили говорни језик, при чему је његова технологија производила само аналогне копије.

Дигиталне језичке технологије данас могу да аутоматизују сам процес превођења, генерисање садржаја и управљање знањем за све европске језике. С њима је могуће опремити кориснику блиске, текстуалне или говорне, сумеће (интерфејс) за кућне електричне апарате, машине, возила, рачунаре и роботе. Иако достигнућа истраживања и развоја омогућавају да се наслуते широке могућности, комерцијалне и индустријске примене су још у раној фази развоја. На пример, за многе европске језике машинско превођење достиже задовољавајући ниво тачности у оквиру специфичних домена, а експерименталне апликације обезбеђују вишејезичне информације, управљање знањем и генерисање садржаја.

Језичке технологије помажу да се превазиђе „хендикеп” језичке разноликости.

Као што је то случај са већином технологија, прве језичке апликације, као што су гласовне корисничке сумеће и дијалогски системи, развијене су за уско специјализоване домене, а њихова употребљивост је била често ограничена. Међутим, у образовној индустрији и индустрији забаве леже огромне тржишне могућности за интеграцију језичких технологија у игре, странице везане за културно наслеђе, производе за образовање кроз забаву, библиотеке, симулациона окружења или програме обуке. Мобилне информационе услуге, софтвер за рачунарски потпомогнуто учење језика, окружења за електронско учење, алати за самооцењивање и откривање плагијата само су неки од примера где језичке технологије могу да одиграју важну улогу. Популарност друштвених мрежа као што су Твитер (Twitter) и Фејсбук (Facebook) указује да постоје потребе за језичким технологијама које омогућавају надгледање поште, резимирање дискусија, детекцију трендова у испољеним ставовима, препознавање емотивних реак-

ција, идентификацију повреда ауторских права или праћење злоупотреба.

Језичке технологије представљају огромну прилику за Европску унију. Оне могу да помогну у решавању комплексног питања вишејезичности у Европи – чињенице да различити језици природно коегзистирају у европском пословању, организацијама и школама. Али грађани желе да комуницирају изван језичких граница које још увек постоје на јединственом европском тржишту. Језичке технологије могу да помогну у превазилажењу ове последње препреке својом подршком слободном и отвореном коришћењу појединачних језика. Ако гледамо и корак даље, иновативне европске вишејезичне технологије представљаће узор за наше partnere по свету када они буду почели да обезбеђују ове технологије за своје вишејезичне заједнице. Језичке технологије могу се посматрати као облик „потпорних” технологија које помажу да се превазиђе „хендикеп” језичке разноликости и да језичке заједнице постану блискије. Најзад, једно активно истраживачко подручје јесте коришћење језичких технологија у операцијама спасавања у областима погођеним катастрофама, где успешно деловање може одлучивати о животу или смрти: будући интелигентни работи са вишејезичним способностима у могућности су да спасу животе.

2.5 ИЗАЗОВИ ПРЕД ЈЕЗИЧКИМ ТЕХНОЛОГИЈАМА

Мада су језичке технологије последњих неколико година оствариле значајан напредак, текући темпо технолошког напретка и иновације производа је сувише спор. Језичке технологије које су у широкој употреби, као што су граматичке и правописне провере, по правилу су једнојезичне, и постоје само за мали број језика. Онлајн услуге машинског превођења, мада су корисне да се брзо произведе прихватљи-

ва апроксимација садржаја документа, стварају пуно потешкоћа када је потребно да се произведу високо прецизни и потпуни преводи. Због комплексности природног језика, његово софтверско моделирање и тестирање у реалном свету је дуг и скуп посао, који захтева дугорочно финансирање. Европа мора да одржи своју пионирску улогу у суочавању са технолошким изазовима вишејезичне заједнице изналажењем нових метода којима ће убрзати развој на целој својој територији. У њих би спадале и иновације у области рачунарства и технике које користе потенцијале великог броја учесника (crowdsourcing).

Текући темпо технолошког напретка сувише је спор.

2.6 УСВАЈАЊЕ ЈЕЗИКА КОД ЉУДИ И МАШИНА

Да бисмо илустровали начин на који рачунари поступају са језиком и зашто их је тако тешко испрограмирати да употребљавају језик, бацимо поглед на начин како људи усвајају матерњи и страни језик, а онда погледајмо и како ради језикотехнолошки систем.

Људи стичу језичке вештине на два различита начина. Бебе уче језик слушајући разговоре између родитеља, браће и сестара и других чланова породице. У узрасту од приближно две године деца изговарају своје прве речи или кратке фразе. То је могуће само захваљујући посебној генетској предиспозицији људи да опонашају, а потом и осмисле оно што чују.

Учење страног језика у старијем узрасту захтева више напора, углавном зато што дете не припада језичкој заједници оних којима је тај језик матерњи. У школи

страни језици се обично усвајају учењем граматичких структура, речника и правописа кроз вежбања која описују језичко знање преко апстрактних правила, табела и примера. Учење страног језика са годинама постаје све теже.

Два главна типа система језичких технологија „усвајају“ језичке способности на сличан начин као људи. Статистички приступи (или „приступи вођени подацима“) стичу језичко знање из огромних колекција конкретних примера текстова. За обучавање система за проверу правописа, на пример, довољно је коришћење текстова на једном језику, али су за обучавање машинских преводилаца потребни тзв. паралелни текстови на два (или више) језика. Алгоритам машинског учења затим „учи“ обрасце превођења речи, кратких фраза и комплетних реченица.

Статистички приступи могу да захтевају милионе реченица јер квалитет резултата расте са порастом броја анализираних текстова. То је један од разлога што добављачи претраживачких машина жељно прикупљају што је могуће више писаног материјала. Исправка правописних грешака у програмима за обраду текста и сервиси као што су Google Search и Google Translate, ослањају се на статистички приступ. Велика предност статистике је у томе што машина учи брзо понављајући циклусе обуке, мада квалитет може да варира на непредвидљив начин.

Други приступ језичким технологијама, а посебно машинском превођењу, јесте изградња система заснованих на правилима. Експерти из лингвистике, рачунарске лингвистике и рачунарства морају најпре да граматичку анализу изразе кроз систем правила и да саставе листе речи (лексиконе). То је посао који захтева много времена и велики труд. Неки од водећих система машинског превођења заснованих на правилима у сталном су развоју већ више од двадесет година. Предност система заснованих на правилима је у томе што експерти могу детаљније да контро-

лишу обраду језика. То омогућује да се грешке у софтверу систематски поправљају, а кориснику пружи детаљне повратне информације, пре свега када се такви системи користе за учење језика. Због великих трошкова, језичке технологије засноване на правилима до сада су биле развијане само за велике језике.

Људи стичу језичке вештине на два различита начина: учењем примера и учењем језичких правила.

Како предности и слабости статистичких система и система заснованих на правилима теже да се допуњују, текућа истраживања усмерена су на хибридне приступе који комбинују те две методологије. Па

ипак, ови приступи су до сада били мање успешни у индустријским применама него у лабораторији.

Као што смо видели у овом одељку, многе апликације које су у широкој употреби у данашњем информационом друштву ослањају се у великој мери на језичке технологије. С обзиром на вишејезичност европске заједнице, ово се посебно односи на њен привредни и информациони простор. Мада су језичке технологије значајно напредовале последњих неколико година, још увек постоје огромне могућности за побољшање квалитета језикотехнолошких система. У следећем одељку описаћемо улогу српског језика у европском информационом друштву и даћемо оцену текућег стања језичких технологија за српски језик.

The Serbian Language in the Digital Age

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

2012, VI, 84 p. 32 illus., 24 illus. in color., Softcover

ISBN: 978-3-642-30754-6