

HÆTTUR SEM STEDJA AÐ TUNGUMÁLINU: ÖGRUN FYRIR MÁLTÆKNI

Við verðum um þessar mundir vitni að stafrænni byltingu sem hefur gífurleg áhrif á samskipti og samfélag. Nýleg þróun í stafrænni upplýsinga- og samskiptatækni er stundum borin saman við það þegar Gutenberg fann upp prentverkið. Hvað getur sú samlíking sagt okkur um framtíð evrópsks upplýsingasamfélags og sérstaklega tungumála okkar?

Við verðum um þessar mundir vitni að stafrænni byltingu sem hefur sambærileg áhrif og uppfinning prentverksins á sínum tíma.

Eftir uppfinningu Gutenbergs voru stigin tímamótaskef í samskiptum og deilingu þekkingar með verkum eins og t. d. þýðingu Lúthers á Biblíunni yfir á þjóðtungur. Á þeim öldum sem síðan eru liðnar hafa verið þróaðar menningarbundnar aðferðir til að sinna betur málvinnslu og deilingu þekkingar:

- Stöðlun stafsetningar og málfræðireglna helstu tungumála skapaði möguleika á hraðri útbreiðslu nýrra vísindalegra og vitsmunalegra hugmynda;
- þróun opinberra tungumála gerði fólki kleift að hafa samskipti innan ákveðinna (oft pólitískra) landamerkja;
- tungumálakennsla og þýðingar milli mála gerðu það mögulegt að eiga samskipti þvert á tungumál;
- ritstjórnarreglur og bókfræðileg viðmið tryggðu gæði prentaðs efnis og aðgengi að því;

- tilkoma margvíslegra fjölmiðla, svo sem dagblaða, útvarps, sjónvarps, bóka o.fl. fullnægði mismunandi samskiptaþörfum.

Á síðustu tuttugu árum hefur upplýsingatæknin átt sinn þátt í því að greiða fyrir mörgum ferlum og gera þau sjálfvirk:

- Ritvinnslu- og umbrotskerfi hafa komið í stað vélritunar og setningar;
- Microsoft PowerPoint hefur komið í staðinn fyrir glærur og myndvarpa;
- með tölvupósti eru skjöl send og tekið á móti þeim mun hraðar en með bréfasíma;
- Skype býður upp á ódýr netsímtöl og skapar vettvang fyrir fjarfundi;
- snið hljóð- og myndbandaskráa gerir auðvelt að skiptast á margmiðlunarefni;
- leitarvélar greiða notendum aðgang að vefsíðum með leit byggðri á lykilorðum;
- netþjónusta eins og Google Translate skilar sæmilega réttum þýðingum á svipstundu;
- félagsmiðlar eins og Facebook, Twitter og Google+ greiða fyrir samskiptum, samvinnu og deilingu upplýsinga.

Þrátt fyrir gagnsemi slíkra tóla og búnaðar dugir þetta ekki til að standa undir sjálfbærni margmála evrópsku samfélagsi fyrir alla, með frjálsu flæði upplýsinga og varnings.

2.1 TUNGUMÁLAPRÖSKULDAR STANDA Í VEGI FYRIR EVRÓPSKU UPPLÝSINGASAMFÉLAGI

Við getum ekki vitað nákvæmlega hvernig upplýsingasamfélag framtíðarinnar mun líta út. En miklar líkur eru á því að bylting í samskiptatækni muni skapa nýja tegund tengsla milli fólks sem talar mismunandi tungumál. Þetta setur aukinn þrýsting á fólk að læra ný tungumál og þó sérstaklega á hönnuði að búa til nýjan tæknibúnað sem tryggi gagnkvæman skilning og aðgang að deil-anlegri þekkingu. Í alþjóðasamfélagi viðskipta og upplýsinga tengjast sífellt fleiri tungumál og málnotendur sífellt hraðar með hjálp nýrra miðla. Vinsældir félagsmiðla (Wikipedia, Facebook, Twitter, YouTube og nú nýlega Google+) eru einungis toppurinn á ísjakanum.

Sífellt fleiri tungumál og málnotendur tengjast sífellt hraðar með hjálp nýrra miðla.

Nú á dögum getum við flutt margra gígabæta texta um heiminn þveran og endilangan á örfáum sekúndum áður en við áttum okkur á því að hann er á máli sem við skiljum ekki. Samkvæmt nýrri skýrslu frá framkvæmdastjórn Evrópusambandsins kaupa 57% evrópskra netnotenda vörur og þjónustu með því að nota tungumál önnur en móðurmál sitt. (Enska er algengasta erlenda tungumálið á þessu sviði en þar á eftir koma franska, þýska og spænska.) 55% notenda lesa erlent mál sér til gagns en aðeins 35% nota annað tungumál til þess að skrifa tölvupóst eða gera athugasemdir á vefnum [6]. Fyrir nokkrum árum var enska tungumál netsins – megnið af því efni sem þar var að finna var skrifað á ensku – en þetta hefur nú gerbreyst. Algjör sprenging hefur orðið í textamagni á öðrum Evrópumálum á netinu (og sama gildir um tungumál Asíu og Mið-Austurlanda).

Það sætir furðu að hin altæka stafræna gjá sem munur tungumála skapar skuli ekki hafa fengið mikla athygli í opinberri umfjöllun; samt sem áður vekur hún mjög brýna spurningu: Hvaða Evrópumál munu dafna í netvæddu upplýsinga- og þekkingarsamfélagi og hver eru dæmd til að hverfa?

2.2 TUNGUMÁL OKKAR Í HÆTTU

Þótt prentverkið hraðaði deilingu upplýsinga í Evrópu olli það því einnig að mörg evrópsk tungumál liðu undir lok. Textar á svæðisbundnum málum og minnihlutamálum komust sjaldan á prent og því voru tungumál eins og korníska og dalmatíska eingöngu notuð sem talmál og notkunarsvið þeirra þar með takmarkað. Mun netið hafa sambærileg áhrif á tungumál okkar?

Hin u.þ.b. 80 tungumál Evrópu eru ein ríkulegustu og mikilvægustu menningarverðmæti álfunnar og grundvallarþáttur í hinni einstöku samfélagsgerð hennar [7]. Þótt tungumál eins og enska og spænska muni að öllum líkindum halda stöðu sinni á hinu stafræna markaðstorgi sem er að verða til gætu mörg evrópsk tungumál orðið gagnslaus í netvæddu samfélagi. Slík þróun myndi veikja alþjóðlega stöðu Evrópu og stangast á við markmið um jafna samfélagsþátttöku allra Evrópuþegna óháð tungumáli.

Hin fjölbreyttu tungumál Evrópu eru ein ríkulegustu og mikilvægustu menningarverðmæti álfunnar.

Í skýrslu UNESCO um fjöltyngi er lögð áhersla á að tungumál séu ómissandi tæki til þess að gera mönnum kleift að njóta grundvallarmannréttinda, svo sem tjáningarfrelsis, menntunar og þátttöku í samfélaginu [8].

2.3 MÁLTÆKNI ER GRUNDVALLARSTUÐNINGSTÆKNI

Áður fyrr beindust aðgerðir til að vernda og varðveita tungumál einkum að tungumálakennslu og þýðingum. Giskað hefur verið á að evrópski markaðurinn á sviði þýðinga, túlkunar, staðfærslu hugbúnaðar og alþjóðavæðingar vefsetra hafi velt 8,4 milljörðum evra árið 2008 og er talinn munu vaxa um tíu prósent á ári [9]. Samt sem áður fullnægir þessi upphæð einungis litlum hluta núverandi þarfar og framtíðarþarfa fyrir samskipti milli tungumála. Augljósasta aðferðin til að tryggja breidd og dýpt málnotkunar í Evrópu framtíðarinnar er að nota viðeigandi tækni, rétt eins og við notum tæknina til að leysa þarfir okkar í samgöngum, orku og stuðningi við fatlaða, svo að eitthvað sé nefnt.

Stafræn máltækni sem beinist að öllum myndum ritaðs máls og talsamskipta gerir fólki kleift að vinna saman, stunda viðskipti, deila þekkingu og taka þátt í félagslegum og pólitískum rökræðum óháð tungumáli og tölvufærni. Hún er oft hulin hluti af flóknum hugbúnaði sem við nýtum okkur þegar við:

- öflum upplýsinga með notkun leitarvéla á netinu;
- rýnum stafsetningu og málfræði í ritvinnslukerfi;
- skoðum umsagnir um vörur í netverslun;
- hlustum á talaðar leiðbeiningar leiðsagnarkerfis í bíl;
- þýðum vefsíður með hjálp netþjónustu.

Máltækni felst í ýmsum grundvallarbúnaði sem margvísleg ferli innan stærri hugbúnaðarkerfa byggjast á. Tilgangur hvítbókaðar META-NET er að skerpa sýn okkar á það hversu þroskuð þessi grunntækni sé fyrir hin ýmsu Evrópumál.

Evrópa þarfnast traustar og ódýrrar máltækni fyrir öll tungumál álfunnar.

Til að viðhalda stöðu sinni í fararbroddi nýsköpunar á heimsvísu þarfnast Evrópa máltækni sem er löguð að öllum evrópskum tungumálum og er traust, ódýr og vel samþætтуð helstu hugbúnaðarumhverfum. Án máltækni munum við ekki öðlast gjöfulan margmála reynsluheim, byggðan á gagnvirkni og margmiðlun, í náinni framtíð.

2.4 TÆKIFÆRI MÁLTÆKNINNAR

Í prentheiminum varð stærsta tæknibyltingin þegar farið var að fjölfalda ímynd texta með notkun prentvéla. Menn þurftu áfram að fletta upp þekkingaratriðum, lesa, þýða, og taka saman þekkingu. Það þurfti að bíða eftir Edison með upptökur á talmáli – en sú tækni bjó þó einnig aðeins til afrit.

Stafræn máltækni getur nú gert sjálfvirktt allt ferlið við þýðingu, samningu efnis og þekkingarstjórnun fyrir öll evrópsk tungumál. Hún getur einnig raungert þróun eðlilegs stýrivíðmóts sem byggt er á máli og tali fyrir heimilisraftæki, vélar, bifreiðar, tölvur og vélmenni. Þróun viðskipta- og iðnaðarverkbúnaðar er enn á frumstigi, en áfangar í rannsóknum og þróun á þessu sviði eru þó farnir að opna mikla möguleika. Til dæmis eru vélþýðingar nú þegar sémilega nákvæmar á afmörkuðum sviðum og tilraunabúnaður skilar margmála upplýsingum og sinnir þekkingarstjórnun og samningu efnis á mörgum Evrópumálum.

Eins og oftast er með tækni var fyrsti máltæknibúnaðurinn, svo sem raddstýrð notendaviðmót og samræðukerfi, þróaður með mjög sérhæfða notkun í huga og sýnir því oft takmarkaða hæfni. En geysimikil markaðstækifæri er að finna í menntageiranum og skemmtanaíðnaðinum þar sem hægt væri að nýta máltækni í leikjum, menningarminjasetrum, menntandi skemmtun, bókasöfnum, hermun og æfingaáætlunum. Upplýsingaþjónusta í farsíma, hugbúnaður fyrir tölvustutt tungumálánám, fjarnámsumhverfi, sjálfsmatstól og forrit til að uppgötva ritstuld eru fáein dæmi þar sem máltækni getur

leikið mikilvægt hlutverk. Vinsældir félagsmiðla eins og Twitter og Facebook benda til þess að þörf sé á háþróaðri máltækni sem getur haldið utan um póst, gert útdrætti úr umræðum, bent á hneigð í skoðunum, greint tilfinningar í svörum, bent á brot á höfundarétti eða haft uppi á misnotkun.

Máltækni hjálpar fólki að sigrast á þeirri „fötlun“ sem felst í málfræðilegum fjölbreytileik.

Í máltækni felast gífurleg tækifæri fyrir evrópskt samstarf. Hún getur hjálpað okkur að takast á við hið flókna málumhverfi í Evrópu – þá staðreynd að mismunandi tungumál lifa eðlilegu samlífi í evrópskum viðskiptum, samtökum og skólum. En þegnarnir þurfa að geta haft samskipti yfir þessi tungumálamörk sem skera hinn sameiginlega evrópska markað þvert og endilangt og með aðstoð máltækni má sigrast á þessari hindrun en styðja um leið við óhefta notkun einstakra tungumála.

Ef við horfum enn lengra fram í tímann mun nýskapandi margmála evrópsk máltækni verða viðmiðun fyrir aðra í alþjóðasamfélaginu þegar þeir fara að virkja sín eigin margmála samfélög. Líta má á máltækni sem eins konar „stuðningstækni“ sem aðstoðar okkur við að yfirstíga „fötlunina“ sem fylgir fjölbreytilegu tungumálaumhverfi og gerir málsamfélögin aðgengilegri hvert öðru. Að lokum má nefna virkt rannsóknarsvið innan máltækinnar sem er notkun máltækni við björgunaraðgerðir á hamfarasvæðum, þar sem rétt framkvæmd getur skipt sköpum. Í framtíðinni gætu greind vélmenn búin hæfileikum til margmála málnotkunar bjargað mannlífum.

2.5 ÖGRANIR SEM MÁLTÆKNI STENDUR FRAMMI FYRIR

Þótt töluverðar framfarir hafi orðið í máltækni á síðustu árum er hraði tækni framfara og nýsköpunar í fram-

leiðsluvörum enn of lítill. Sá máltækni búnaður sem mest er notaður, svo sem málfræði- og stafrýnar ritvinnslukerfa, er venjulega einmála og þar að auki einungis til fyrir fáein tungumál.

Núverandi hraði tæknilegra framfara er of lítill.

Þótt vélþýðingar á netinu séu gagnlegar til að fá þokkalega hugmynd um efni skjala glíma þær við alls kyns vandamál þegar þörf er á mjög nákvæmum og fullkomnum þýðingum. Vegna þess hve mannlegt mál er flókið er það bæði langt og dýrt ferli sem krefst langtíma fjármögnunar að skrifa hugbúnað sem líkir eftir mannlegu máli og prófa hann við eðlilegar kringumstæður. Til að halda brautryðjendahlutverki sínu í því að takast á við þær tæknilegu ögranir sem fylgja margmála samfélagi verður Evrópa því að beita nýjum aðferðum til að hraða þróuninni. Hér gæti bæði verið um að ræða framfarir í tölvutækni og aðferðir eins og lýðvirkjun.

2.6 MÁLTAKA MANNA OG VÉLA

Til að útskýra hvernig tölvur fást við tungumál og hvers vegna það er svo erfitt að forrita þær til þess skulum við líta sem snöggvast á það hvernig við tileinkum okkur móðurmálið og önnur mál, og skoða síðan hvernig máltækni kerfin virka.

Mannfólkið öðlast málkunnáttu á tvo mismunandi vegu: Lærir af dæmum og lærir reglurnar sem liggja þar að baki.

Mannfólkið lærir tungumál á tvo mismunandi vegu. Ungbörn læra móðurmál sitt með því að hlusta á samskipti foreldra sinna, systkina og annarra fjölskyldumeðlima. Um það bil tveggja ára gömul fara þau að mynda fyrstu orðin og stuttar setningar. Þetta er því aðeins

mögulegt að börn hafa meðfæddan hæfileika til máls, og til að herma eftir því sem þau heyra og binda það í kerfi. Nám annars máls síðar á ævinni krefst meiri áreynslu, einkum vegna þess að nemandinn er ekki umlukinn málsamfélagi sem hefur málið að móðurmáli. Í skólum eru erlend mál venjulega numin með því að læra málfræðilega formgerð, orðaforða og stafsetningu með mynsturæfingum sem lýsa málfræðilegri kunnáttu í formi óhlutstæðra reglna, tafla og dæma. Nám erlends tungumáls verður erfiðara með aldrinum.

Hinar tvær megingerðir máltækni­kerfa „nema“ tungumál á svipaðan hátt og mennirnir. Tölfræðilegar (eða gagnaknúnar) aðferðir afla málþekkingar úr gífurlega umfangsmiklum textasöfnum. En þótt nægjanlegt sé að nota texta á einu máli til að þjálf­a t. d. stafrýna eru samhliða textar á tveim eða fleiri málum nauðsynlegir þegar kemur að þjálfun vélrænna þýðingarkerfa. Algrím vélræns náms „lærir“ þá mynstur sem sýna hvernig orð, orðasambönd og heilar setningar eru þýdd.

Þessi tölfræðilega nálgun getur krafist milljóna setninga og gæði útkomunnar aukast í réttu hlutfalli við magn greinds texta. Þetta er ein ástæða þess að þeir sem reka leitarvél­ar eru áfjád­ir í að safna eins miklu af rituðu efni og hægt er. Stafrýnar í ritvinnslukerfum og netþjónustur eins og Google Search og Google Translate byggjast á tölfræðilegum aðferðum. Meginkostur tölfræðinálgunarinnar er sá að vélin lærir fljótt í samfelldri röð þjálfunarferla, jafnvel þótt gæðin geti verið með ýmsu móti.

Hin meginaðferðin í máltækni og vélþýðingum er að smíða reglukerfi. Þá þurfa sérfræðingar á sviði mál-

vísinda, tölvumálvísinda og tölvunarfræði fyrst að skrá málfræðigreiningu (þýðingarreglur) og búa til orðalista (orðasöfn). Þetta tekur langan tíma og kostar mikla vinnu. Reglukerfin krefjast einnig sérfræðiþekkingar. Sum helstu reglubýggðu vélþýðingarkerfin hafa verið í stöðugri þróun í meira en tuttugu ár. Meginkosturinn við reglukerfin er að sérfræðingarnir hafa meiri stjórn á málvinnslunni. Þetta gerir það mögulegt að laga kerfisbundið villur í hugbúnaðinum og veita notendum nákvæma endurgjöf, sérstaklega þegar reglukerfin eru notuð í tungumálanámi. En vegna þess hversu kostnaðarsöm þessi vinna er hefur reglubýggð máltækni til þessa einungis verið þróuð fyrir stærstu tungumálin.

Þar sem styrkleikar og veikleikar tölfræðilegu kerfanna og reglubýggðu kerfanna eru á mismunandi sviðum beinast rannsóknir um þessar mundir að blönduðum aðferðum sem tengja þessar tvær gerðir saman. Enn sem komið er hafa slíkar aðferðir þó ekki reynst eins vel í markaðshugbúnaði og á rannsóknarstofunum.

Eins og fram hefur komið í þessum kafla byggist alls kyns búnaður sem notaður er í upplýsingasamfélagi nútímans á máltækni. Í Evrópu á þetta sérstaklega við á sviði viðskipta og upplýsinga vegna þess hversu margmála málumhverfið þar er. En þrátt fyrir að máltækni hafi tekið miklum framförum á síðustu árum eru enn miklir möguleikar á því að auka gæði máltækni­kerfa. Hér á eftir verður hlutverki íslenskunnar í evrópsku upplýsingasamfélagi lýst og mat lagt á stöðu máltækni fyrir íslensku.

The Icelandic Language in the Digital Age
Rehm, G.; Uszkoreit, H. (Eds.)
2012, V, 47 p. 21 illus. in color., Softcover
ISBN: 978-3-642-30173-5