



Staðan í íslenskri máltækni

Starfsdagur Þýðingamiðstöðvar utanríkisráðuneytis
15. október 2015

Eiríkur Rögnvaldsson



Úlfar Erlingsson - september 2015

„Við erum í raun og veru í miðjunni á byltingu á því hvernig tölvur geta gert það sem hingað til hefur aðeins verið á færi manneskja,“ segir Úlfar í samtali við Kjarnann. Hlusta má á samtalið í heild sinni í spilaranum hér að ofan. Hann útskýrir þessa byltingu þannig að fyrir fjórum árum hafi vísindamenn farið að vinna með svokölluð tauganet til þess að vinna úr gríðarlegu magni upplýsinga sem tauganetin geta svo lært af. „Þá geta þau leyst verkefni sem eru aðeins á færi manneskja. Til dæmis horft á mynd og skilið strax hvað stendur á myndinni, hlustað á einhvern tala og ekki aðeins gripið það að þarna sé verið að tala eitthvað tungumál heldur þýtt merkingu þess á annað tungumál.“

„Það sem er mekilegast í þessu er að það er sami hugbúnaðurinn sem er bak við allar þessar þjónustur,“ útskýrir Úlfar og vísar þar í nýjungar í Google Photos, Google Translate og þar fram eftir götunum. „Það er sama vélin sem er notuð til að túlka myndir yfir í orð eins og köttur, sólarlag eða barn. Sama tækni er notuð til að þýða úr þýsku yfir á kínversku.“



Ný tækni er handan við hornið

- Gerbylting í samskiptum er á næsta leiti
 - eins og fram kemur í viðtalinu við Úlfar Erlingsson
- Ef við fylgjumst ekki með erum við illa stödd
 - eins og auðvelt er að sýna fram á
- Nýlega var gerð könnun á stöðu 30 Evrópumála
 - hvað varðar máltækni og rafrænan stuðning
- Gerð á vegum Evrópuverkefnisins [META-NET](#)
 - 70-80 síðna skýrsla (hvítbók) fyrir hvert mál
 - tvímála skýrslur — á viðkomandi máli og ensku



Sterk staða málsins

- Staða íslensku er að mörgu leyti sterk
 - sterkari en ýmissa mála sem fleiri tala
 - en eru „aukatungur“ þjóða
 - eða tungur minnihlutahópa í stærri þjóðfélögum
 - t.d. bretónska í Frakklandi, velska í Wales
- Slík mál eiga mjög í vök að verjast
 - vegna þess að notkunarsvið þeirra er takmarkað
 - málnotendur eiga val



Málið í samfélaginu

- Íslenska er notuð á öllum sviðum samfélagsins
 - í stjórnkerfinu
 - í menntakerfinu
 - í viðskiptalífinu
 - í fjölmiðlum
 - í öllum daglegum samskiptum fólks
- Þó eru blikur á lofti
 - vísbendingar um veikari stöðu málsins
 - einkum meðal ungs fólks



Gunnar Smári Egilsson – mars 2015

Íslenskan er að deyja og samfélagið með

7. mars 2015 • Gunnar Smári Egilsson • Skrifa athugasemd

Þegar maður þælir í því er augljóst að íslenskan mun deyja út á næstu 50 til 70 árum. Alla vega sem fullburða tungumál í fjölþættum samskipum fólks. Íslenskan varð að sjálfstæðri tunga vegna einangrunar og eftir því sem einangruninni léttir hverfa forsendur þess að fólkið á Íslandi tali sérstakt tungumál. Æ minni hluti þess sem Íslendingar heyra og lesa er á íslensku og Íslendingar eru of fáir til að geta dregið sístækkandi og æ margbreytilegri veröld inn í íslenska tungu. Og jafnvel þótt Íslendingar væru fleiri þá er vafasamt hvort slíkur hugsanaflutningur væri til bóta.



Linda Björk Markúsardóttir – apríl 2015

„Ég kann þetta ekkert á íslensku“

MENNTUN



Linda Björk Markúsardóttir talmeina- og íslenskufraeðingur

Starf mitt sem talmeinafræðingur felst, meðal annars, í því að meta orðaforða og málkunnáttu barna og unglinga. Undanfarið hefur það hefur færst í aukana að ég fái til mín alíslensk börn sem kunna ekki íslensku nema að litlu og yfirborðskenndu leyti. Þeim sýni ég myndir af algengum hlutum og bið þau að segja mér hvað þeir heita. Oft fæ ég svör á borð við: „Ég veit alveg hvað þetta er sko, ég bara kann þetta ekkert á íslensku.“

Sé þeim boðið að nefna hlutinn á ensku lætur svarið ekki á sér standa. Hvernig má þetta vera? Hvernig geta börn og unglingar sem eiga íslenska foreldra, eru fædd og uppalin hér og hafa alla tíð gengið í íslenska skóla ekki búið yfir íslenskum grunnorðaforða? Sé ekki um eiginlega málþroskaröskun að ræða er svarið við þessum spurningum einfalt: Tölvur og tækni. Nú fylgir tæknin móðurmjólkinni og börn og unglingar lifa og hrærast í heimi rafrænna samskipta sem fara nánast öll fram á ensku. Tölvur og símar „tala“ mjög takmarkaða íslensku og til hvers að brúka móðurmál skitinna 320.000 hræða þegar enskan opnar dyr allra samskipta upp á gátt?

Eiríkur Rögnvaldsson, prófessor í íslenskri málfræði, hefur löngum verið ótull talsmaður íslenskrar máltækni og

→ Hvað er hægt að gera? Eigum við sem þjóð að halla okkur aftur í tölvustólunum, skerpa á enskunni og bíða þess sem verða vill?

mikilvægis uppbyggingar hennar fyrir íslenska tungu. Í þættinum Orðbragð, sem RÚV hefur haft til sýninga, lýsti Eiríkur því yfir að raunveruleg hættu væri á því að íslenska verði ekki til eftir 100 ár. Hann vísaði í skýrslu um stafræna stöðu íslenskunnar máli sínu til stuðnings en hún kom út árið 2012.

Á þeim tímapunkti hélt ég að þetta væri nú óþarfa vænisyki í manninum, að ekkert fengi grandað mínu ástkæra ylhýra. Nú er annað bersýnilega að koma í ljós og ekki er lengur hægt að loka augunum fyrir hættunni á því að íslenskan deyi dauða sínum. Hvað er hægt að gera? Eigum við sem þjóð að halla okkur aftur í tölvustólunum, skerpa á enskunni og bíða þess sem verða vill? Eða eigum við ef til vill að beita okkur fyrir því að auknum fjármunum verði veitt í að efla stafræna stöðu íslenskunnar? Eigum við kannski að tala meira við börnin okkar, lesa fyrir þau og gera þeim grein fyrir mikilvægi þess fyrir okkur sem þjóð að íslenskan haldist á lífi?



GIDS-kvarði um lífvænleik tungumála

Fishman's Graded Intergenerational Disruption Scale

Stage 8: most vestigial users of Xish are socially isolated old folks and Xish needs to be re-assembled from their mouths and memories and taught to demographically unconcentrated adults

Stage 7: most users of Xish are a socially integrated and ethnolinguistically active population but they are beyond child-bearing age

Stage 6: the attainment of intergenerational informal oralcy and its demographic concentration and institutional reinforcement

State 5: Xish literacy in home, school and community, but without taking on extra-communal reinforcement of such literacy

Stage 4: Xish in lower education (types a and b) that meets the requirements of compulsory education laws

Stage 3: use of Xish in the lower work sphere (outside of the Xish neighborhood/community) involving interaction between Xmen and Ymen

Stage 2: Xish in lower government services and mass media but not in the higher spheres of either

Stage 1: some use of Xish in higher level educational, occupational, governmental and media efforts (but without the additional safety provided by political independence)



Viðmið UNESCO um lífvænleik tungumála





Þrjú teikn um að tungumál sé í hættu

- A language may not be completely dead until the death of its last speaker, but there are three clear signs of imminent death observable well in advance.
 - First, there is loss of function, seen whenever other languages take over entire functional areas such as commerce.
 - Next, there is loss of prestige, especially clearly reflected in the attitudes of the younger generation.
 - Finally, there is loss of competence, manifested by the emergence of ‘semi-speakers’ who still understand the older generation, but adopt a drastically simplified (reanalyzed) version of the grammar.



Niðurstöður rannsóknar META-NET

- 2/3 Evrópumála (21 af 30) eru í hættu
 - ná ekki að fylgja hraðri þróun tölvutækninnar
 - verða því ekki nothæf á mikilvægum sviðum þjóðlífsins í náinni framtíð ef svo fer sem horfir
- Vitundarvakning er nauðsynleg
 - til að vekja athygli stjórnvalda og almennings á þeirri hættu sem tungumálin standa frammi fyrir
 - og benda á möguleika máltækni til að auðvelda samskipti þjóða, og milli manns og tölvu



Ályktanir META-NET skýrslunnar

- Íslenska er ekki í bráðri hættu, þrátt fyrir yfirburði enskunnar í máltækni og tölvumálvísindum.
- Á hinn bóginn gæti staðan gerbreyst á svipstundu þegar ný kynslóð tækninnar fer fyrir alvöru að ráða við mannlegt mál á skilvirkan hátt.
- Eigi íslenska að vera lífvænleg þjóðtunga í þróuðum heimi verður hún að geta staðið undir kröfum upplýsingatækninnar.
- Fjárfesting í máltækni verður því að vera grunnþáttur í framkvæmd íslenskrar málstefnu.



Þrengt notkunar svið íslensku

- Hvað ef íslenskan yrði ekki gjaldgeng
 - á sviði sem væri mikilvægt í daglegu lífi almennings?
- Hvað ef hún yrði ekki nothæf
 - til að tala og skrifa um áhugamál ungs fólks
 - í nýsköpun á sviði bókmennta og lista
 - í nýrri tækni og öðru sem væri nýtt og spennandi
 - á öðrum sviðum þar sem nýsköpun ætti sér stað
 - og þar sem ný atvinnutækifæri byðust?



Egill Helgason – september 2015

Hvernig segjum við börnum að læra íslensku?

Ég hef svolítið velt fyrir mér hvernig við segjum börnum og unglingum að gott sé fyrir þau að læra íslensku – og læra hana vel.

Ljóð með línunum eins og „land, þjóð og tunga, þrenning sönn og ein“ sannfæra engan. Rökin að með því að kunna íslensku geti þau lesið Kiljan og Íslendingasögurnar duga heldur ekki langt og ekki að með því að halda í íslenskuna varðveitum við sérstakan menningarheim, eins og Guðmundur Andri Thorsson orðar það í ágætri grein þar sem hann fjallar um nauðsyn þess að tölvur séu mæltar á íslensku.

Og varla dugir heldur að segja að með því að læra íslenskuna séu þau sérstök – að þau tilheyri fámennum hópi sem á þetta tungumál sem enginn annar kann eða skilur.

Þetta síðastnefnda kann reyndar frekar að virka eins og rök gegn því að læra íslensku í nútímanum. Það hljómar fremur eins og sjálfskipuð einangrun.

Þjóðerniskenndin ristir heldur ekki djúpt. Hvernig má það líka vera þegar möguleikarnir til að komast áfram í lífinu eru fyrst og fremst í alþjóðlegum greinum sem þekkja engin landamæri, viðskiptum, hugbúnaði, ferðaþjónustu?

Börn og unglingar vaxa nú upp við tvítýngi. Sjónvarps- og tölvuefni sem þau nota er mestanpart á ensku. Þau horfa ekki á það sem er kallað línulaga dagskrá í sjónvarpi og þau lesa ekki blöð. Það er mjög auðvelt að vera íslenskt barn og sjá aldrei né heyra menningarefni á íslensku.

Þetta er altént eitthvað sem við megum velta fyrir okkur. Mér finnst eins og að rökin sem voru notuð við mína kynslóð dugi ekkert sérlega vel lengur – en við lifðum líka í menningarheimi sem var miklu íslenskari en nú er.



Íslensk málstefna 2009 – meginmarkmið

- Að íslensk tunga verði nothæf – og notuð – á öllum þeim sviðum innan tölvu- og upplýsingatækninnar sem varða daglegt líf alls almennings
 - Þetta merkir í fyrsta lagi að viðmót algengs hugbúnaðar þarf að vera íslenskt; í öðru lagi að til þarf að vera ýmiss konar hugbúnaður sem liðsinnir og leiðbeinir notendum við notkun íslensks máls; og í þriðja lagi að unnt á að vera að nota íslensku sem samskiptamál við ýmiss konar tölvu- og tæknibúnað



Nefnd um íslensku í tölvuheiminum

- Nefnd til að fylgja málstefnunni eftir
 - skipuð 2010, lauk störfum 2012
- Ýmsar tillögur um
 - íslenskt notendaviðmót
 - hugbúnaðarþýðingar, stuðning við íðorðastarf o.fl.
 - máltækni
 - leiðréttingarforrit, talgreining
 - rafrænt námsefni fyrir spjaldtölvur
- Engu hefur verið hrint í framkvæmd



Þingsályktun, 12. maí 2014

- Notkun íslensku í stafrænni upplýsingatækni
 - Alþingi ályktar að fela mennta- og menningarmálaráðherra að skipa nefnd sérfræðinga í málvísindum og upplýsingatækni sem geri áætlun um aðgerðir er miði að því að gera íslensku gjaldgenga í stafrænni upplýsingatækni og stuðla að notkun hennar á þeim vettvangi. Áætlunin feli í sér tímasett yfirlit um aðgerðir og áfanga, kostnaðarmat og fjármögnun. Nefndin leggi áætlun sína fram í síðasta lagi 1. september 2014.



Tillögur nefndar

- Nefnd skipuð í byrjun september 2014
 - skilaði álitum í desember
- Tillögur eru þrískiptar:
 - tækniþróunarverkefni
 - s.s. talgreining, málfarsleiðrétting, vélrænar þýðingar
 - innviðaverkefni
 - s.s. uppbygging og viðhald mállegra gagnasafna
 - styrkir til doktorsnema
- Heildarkostnaður: 990 milljónir á 10 árum



Innviðir

- [Almannarómur](#)
 - sjálfseignarstofnun stofnuð 2014 með 5,75 m.kr.
 - stofnaðilar félagasamtök, fyrirtæki og stofnanir
 - mun standa að smíði máltæknilausna fyrir íslensku
- [Máltækni Jóður](#)
 - stofnaður vorið 2015 með 15 m.kr. í fjárlögum
 - áætlað framlag 30 m.kr. í fjárlagafrumvarpi 2016
 - ætlað að stuðla að því að íslenska sé gjaldgeng í stafrænni upplýsingatækni



Verkefni sem nú er unnið að

- Talgreining
 - í samvinnu við röntgenlækna á Landspítalanum
 - í samvinnu við Alþingi
- Talgervill
 - [verkefni Google](#)
- Ýmis smærri verkefni
 - sem meistaranemar vinna í námi sínu
- Þátttaka í [ELRC](#)



European Language Resource Coordination

What is the European Language Resource Coordination (ELRC) action?

The Automated Translation platform of the Connecting Europe Facility (CEF.AT) will be based on statistical data-driven machine translation (MT) technology, which learns how to translate from existing human translations. To create MT engines that can deliver useful results for European public administrations and for users of public online services, appropriate **language resources** are needed.

In each of the 30 participating countries, local **workshops** will be organized to raise awareness about the importance of data sets, thus preparing the ground for collecting the language resources needed for CEF.AT.

To answer questions that may arise on the technical or legal aspects related to the use, production, collection, or sharing of language resources, ELRC also provides a **Technical and Legal Helpdesk**. The Helpdesk is available free of charge and will provide an answer to your question within a working day.

A **conference** will be organized in the second quarter of 2016, in order to inform users within European public administrations on the progress of the project and to inject and reinvigorate the pan-European momentum.

Also, a **Language Resources Board (LRB)** has been set up as the governance body of this initiative, composed of representatives of all of the countries involved.



Takk fyrir áheyrnina!

eirikur@hi.is