



ARENGUSEIRE
KESKUS

PIKKSILM
aprill 2022

Vaimult suureks keele tehnoloogia toel

Meelis Mihkla, Liisi Piits

Artikkel ilmus Arenguseire Keskuse trendiraportis „Pikksilm“.



Keeletehnoloogia üks peamisi eesmärke on võimaldada inimestel masinatega suhelda loomulikus keeles. Tehnika hoogsa arenguga on vajadus sellise suhtluse järele kiirelt kasvanud. Tehisintellekti võidukäik¹ ja üha laiem rakendamine eeldab seda, et nii lihtsamad kodumasinad kui ka uued nutikad abimehed (assistendid, juturobotid, nõuandjad, lepingute sõlmijad) suhtleksid meiega inimkeeles. **Eestikeelsete keeletehnoloogiavahendite kasutamine erinevates digirakendustes on oluline nii emakeelse masin-inimene suhtluse tagamiseks kui ka eesti keele säilimiseks.**

Mõlema eesmärgi täitmiseks on vaja riiklikult jälgida ja suunata keeletehnoloogiliste vahendite arendamist. Eesti keele kõnelejaskond oma miljoni kõnelejaga ei ole keeletehnoloogiliste vahendite väljatöötajatele tulutoov turg, mistõttu riigi abi on vältimatu. Riiklik programm „Eesti keeletehnoloogia 2018–2027“² toetabki keeletehnoloogiaalast teadus- ja arendustegevust, et luua uusi eesti keele keeletehnoloogilisi rakendusi, tõsta olemasolevate rakenduste kvaliteeti ja soodustada keeletehnoloogia laiemat kasutuselevõttu. **Lisaks on muidugi suurepärase, et ka suurkorporatsioon Google on eesti keele arvanud maailma pisut üle saja arenenuma keele hulka, milledele ta on ise loonud kõnesünteesi, kõnetuvastuse ja masintõlke.** Paraku lähtuvad suurfirmit sageli vaid oma huvidest ja ka nt Google'i loodud keeletehnoloogiline tugi on tagatud vaid nende endi rakendustes ja Androidi operatsioonisüsteemi kasutatavates seadmetes. Samuti ei räägi ei Google'i ega teiste suurfirmade virtuaalsed assistendid veel eesti keelt ning me ei saa lootma jääda, et erafirmadel oleks ilmtingimata huvi nii väikese turu jaoks virtuaalseid assistente arendada. Seega on oluline ka edaspidi riiklikult toetada mitmesuguste keeletehnoloogiliste ressurside, baastehnoloogiate ja arendusprojektide valmimist ning kasutuselevõttu.

Keeletehnoloogial on kolm võtmevaldkonda: kõnetuvastus, mis teisendab inimese kõne tekstiks; masintõlge, mis automaatselt tõlgib tekste ühest keelest teise; ja kõnesüntees, mis teisendab kirjaliku teksti kõneks. Nende

valdkondade koostööd ilmestab kõige paremini häältõlge e kõne-kõneks tõlge. Kujutame ette situatsiooni, et teil on Skype'is vaja suhelda oma sõbra või kolleegi Michikoga Jaapanist, aga ühine suhtluskeel, mida te mõlemad hästi valdaksite, paraku puudub. Kõne-kõneks tõlge võimaldab teil mõlemal suhelda emakeeles: eestikeelne kõnetuvastus muudaks teie kõne tekstiks, eesti-jaapani masintõlge teisendaks eestikeelse teksti jaapani kirjas tekstiks ning jaapani keele kõnesüntesaator loeks selle teksti Michikole ette. Ja ideaalis võiks jaapanikeelne kõnesüntesaator olla võimalikult teile sarnase häälega ning teie suu ja huulte liigutused vastaksid jaapanikeelse kõne hääldusele (visuaalne kõnesüntees). Kui Michiko kõneleb, siis käivituks vastupidine protsess ja Michiko mõtted suulisel kõnes jõuaksid teieni sulaselges eesti keeles Michiko sünteeshäälega. Praegu sellist häälega tõlke suhtluse võimalust eesti keele jaoks veel ei ole, aga loodetavasti veel sel kümnendil saab kõne-kõneks tõlge eesti ja paljude teiste keelte vahel võimalikuks. Eeldused selleks on olemas, kuna juba praegu on eesti teadlased ja tarkvaraarendajad välja töötanud kõnesünteesi, kõnetuvastuse ning masintõlke. Aga nii kõik nimetatud võtmevaldkonnad kui ka neid toetavad muud keeletehnoloogiavaldkonnad (morfoloogiline ja süntaktiline analüüs, dialoogisüsteemid jms) vajavad pidevat edasiarendust, et suhtlus toimuks kirjeldatud viisil ja tõlkes midagi olulist kaduma ei läheks.

Seadusandlik alus

Eesti riik ja tema hallatavad asutused peaks eesti keele kasutamisel ja omakeelsete keeletehnoloogiliste vahendite rakendamisel kindlasti eeskujul ja initsiatiivi üles näitama. Naaberriigis Soomes on iseenesest mõistetav, et kõik tehnikaseadmed on soome keele toega ja valdavalt kasutatakse soomekeelset tarbetarkvara. Eestis see nii veel kahjuks ei ole. 2017. aastal läbi viidud keelehoiakute uuringu põhjal pidas vaid 65% Eesti elanikest tähtsaks, et mitmesuguseid tehnikaseadmeid (sh arvuteid, telefone, kodumasinaid) oleks võimalik kasutada eesti keeles³. Ja reaalne eesti keele

kasutus tarbetarkvaras ja tehnikaseadmetes on veelgi madalam. Selles osas võiks riik seadusandlust täiendada, muutes näiteks riigi- ja kohaliku omavalitsuse asutustes ning üldhariduskoolides eestikeelse tarkvara ja eestikeelsete platvormide kasutamise kohustuslikuks. Venekeelsetes koolides eestikeelse tarkvara kasutamine soodustaks ka riigikeele õpet.

Eesti meediaseadus reguleerib eesti keele kasutust massi- ja trükimeedias. Suur hulk kasutajaid on küll tavameediast (raadio, televisioon, paberajakirjandus) üle kolinud Internetti ja sotsiaalvõrgustikku, kuhu seadusesilm ei ulatu, aga ka praegust meediaseadust võiks täiendada eestikeelse juurdepääsu võimaluse avardamisega erivajadustega inimestele. Praegu on vaid Eesti Televisiooni kanalites võimalik subtiitritega varustatud võõrkeelseid saateid ja filme jälgida koos sünteeshäälega ettelooetavate subtiitritega. See lahendus on eelkõige mõeldud vaegnägijatele ja vaeglugejatele (düslektikuid), kes sageli ei suuda või ei jõua subtiitriteid lugeda. Telemeedia kaudse seire põhjal võib nentida, et ligi 12 000 inimest on igapäevaselt Eesti Televisiooni kanalites selle teenuse kasutajad. **Kui meediaseadust täiendada kohustusega võimaldada vaegnägijatele juurdepääs subtiitrites esitatud infole, võiks kasvada telekanalite soov kõnesünteesi subtiitrite ettelugemisel kasutada ja teenusest kasusaajate arv oleks märgatavalt suurem.**

Eesti keele staatus 21. sajandi infoühiskonnas

Maailmas on praegu umbes 6000 erinevat keelt. Kõnelejate arvu poolest on eesti keel esimese veerandtuhande piirimail, kuid kasutuse ulatuselt tunduvalt eespool. Eesti keelel on ka riigikeele staatus maailma ca 200 riigikeele hulgas. Me peame ennast haritud rahvaks ja seda õigustatult. Kõrgharidust saab maailmas umbes 100 keeles, sh Euroopas ligi 30 keeles ja meie keel on üks nendest. Rõõmustav pilt avaneb ka siis, kui vaatleme keeli, millel on keeletehnoloogiline tugi. Omakeelset Windowsi ja Microsofti tarbetarkvara Office saavad praegu

kasutada maailmas veidi üle saja rahva, sh eestlased⁴. Toimiv kõnesüntees on olemas umbes 100 keele ja kõnetuvastus on olemas või valmimas 80 keele jaoks. Google'i tõlkemootor on eesti keele kaasanud pisut üle saja valitud keele hulka. Seega, kui eesti keel on kõnelejate arvult maailmas esimese kolmesaja hulgas, siis hariduse valdkonnas ja infotehnoloogia arengus oleme maailma esimese saja arenenuma keele hulgas.

Tundub, et arvude keeles on kõik hästi ja eesti keele tulevik kindlustatud. Tegelikult asi siiski nii lihtne ei ole ja selleks et ajaga kaasas käia, peab pidevalt n-ö tehnoloogiliselt võidurelvastuma ja aktiivselt tegutsema. **Pessimistlikumate ennustuste kohaselt jääb 21. sajandi lõpuks alles vaid 600–700 keelt ja nimelt just need keeled, millel on olemas korralik keeletehnoloogiline tugi.** Hoiakutes eesti keele suhtes on täheldatavad ka teatud ohumärgid, mis näitavad, et me ei saa loorberitele puhkama jääda ning tuleviku nimel tuleb veel pingutada. Riigipoolsest seadusandlikust mõjutamisest ja ressursside kaasamisest veelgi tähtsamaks tuleb pidada eesti keele maine tähtsustamist. Tuleks veelgi rohkem teadvustada, et eesti keel koos oma keeletehnoloogilise toega kuulub maailma suurte keelte hulka ja eesti keelel on kindel tulevik. Eesti keelt peab igakülgsest saama kõigis valdkondades kasutada.

Meil on küll olemas eestikeelne tarkvara ja keeletehnoloogilised vahendid, kuid nende kasutus pole veel muutunud tavapäraseks normaalsuseks. Miks me ei pööra näiteks sellele tähelepanu, et robottolmuimejad räägivad oma tegemistest ja seisunditest inglise keeles? Ühelt poolt elektroonikafirmade mugavus pakkuda tooteid vaid *lingua franca's*, teiselt poolt meie vähene nõudlikkus, et masinad räägiksid eesti keeles, loovad eeldused selleks, et inglise keel kipub järjest rohkem suhtluskeele tasandile.

Keeletehnoloogia arengust sõltub ka üldine informatsiooni kättesaamine digitaalses maailmas. Kui kunagi paberil ilmunud teoste digimine aitab kaasa sellele, et seal esitatud info on pärast paremini leitav, siis ajal, kui aina

enam infot jagatakse audiovisuaalselt, vajab sel viisil esitatud informatsiooni leidmine ja taaskasutamine kõnetehnoloogiat, mis aitaks kõnelaine tekstiks muuta. Kõnetuvastuse kasutamine päästab suure hulga auditiivset eestikeelset informatsiooni unustusse vajumast ja võimendab nii eesti kultuuris toimuvat.

Keele elujõud 21. sajandi tehnoloogia kiire arengu ajastul sõltub paljudest faktoritest. Keeletehnoloogilise toe ning omakeelse arvuti- ja meediakeskkonna loomine on oluline eesti keele arenguks ja püsijäämiseks.

Seadusandlusega saab riik tehnoloogia arengust tingitud protsesse teadlikult mõjutada ja ka keelehoiakuid kujundada, et tagada eesti keelele kindel positsioon ja igakülgne areng infoühiskonnas. Et saada suureks vaimult – selle Jakob Hurda poolteist sajandit tagasi seatud eesmärgi saavutamisel on ka keeletehnoloogial väärtusliku tööriistana tähtis roll. **Sest vaba vaim saab kasvada ja areneda 21. sajandi infoühiskonnas vaid siis, kui emakeelt on võimalik kõikides valdkondades igakülgsest kasutada.**



Meelis Mihkla
(Eesti Keele Instituut)



Liisi Piits
(Eesti Keele Instituut)

Kasutatud allikad:

¹ https://www.riigikantselei.ee/sites/default/files/riigikantselei/strateegiaburoo/eesti_tehisintellekti_kasutuselevotu_eksperdiruhma_aruanne_final.pdf

² <https://www.keeletehnoloogia.ee/et>

³ <https://www.hm.ee/sites/default/files/aruanne.pdf>

⁴ Windows on olemas 106 keeles/dialektis, vt <https://support.microsoft.com/en-us/windows/language-packs-for-windows-a5094319-a92d-18de-5b53-1cfc697cfca8>. Tarbetarkvaral Office'il on praegu lokaalsed keelemoodulid 102 keelele/dialektile, vt <https://support.microsoft.com/en-us/office/language-accessory-pack-for-office-82ee1236-0f9a-45ee-9c72-05b026ee809f>

