

# OHT MEIE KEELTELE JA VÄLJAKUTSE KEELETEHNOLOOGIALE

Oleme tunnistajateks digirevolutsioonile, mis avaldab tohutut mõju meie suhtlusele ja ühiskonnale. Viimast arengut digitaalses info- ja kommunikatsioonitehnoloogias võrreldakse Gutenbergi trükipressi leiutamise mõjuga. Mida ütleb see analoogia meile Euroopa infühiskonna, täpsemalt meie keelte tuleviku kohta?

---

Me oleme tunnistajaks digitaalsele revolutsioonile, mis on võrreldav Gutenbergi trükipressi leiutamisega.

---

Pärast Gutenbergi leiutist toimus tõeline läbimurre kommunikatsioonis ja teadmiste jagamises, näiteks tõlkis Luther Piibli rahvakeelde. Sellele järgnenud sajanditel on arendatud kultuuritehnoloogiaid keeletöötuse ja teadmistevahetuse edendamiseks:

- suuremate keelte õigekirja ja grammatika standardiseerimine tegi võimalikuks teaduse ja ideede kiire leviku;
- ametlike keelte areng võimaldas kodanikel teatud (sageli poliitiliste) piiride raames suhelda;
- keelte õpetamine ja tõlkimine tegi võimalikuks keelteülese suhtluse;
- kirjutiste toimetamise ja bibliograafiaalaste juhtnõude loomine kindlustas trükimaterjalide kvaliteedi ja kättesaadavuse;
- uut liiki meedia – ajalehtede, raadio, televisiooni, raamatute ja muude formaatide – teke rahuldab erinevaid kommunikatsioonivajadusi;

Viimase kahekümne aasta jooksul on infotehnoloogia aidanud kaasa mitme protsessi automatiseerimisele ja lihtsustamisele, nt:

- kirjastustarkvara on asendanud masinakirja ja trüki-  
ladumise;
- Microsoft PowerPoint on asendanud lüümikud ja  
grafoprojektorid;
- meilidega saadetakse ja saadakse dokumente kiire-  
mini kui faksi teel;
- Skype annab võimaluse odavateks internetikõne-  
deks ja virtuaalsete koosolekute pidamiseks;
- audio- ja videokodeeringud lihtsustavad multime-  
dia jagamist;
- otsingumootorid lubavad veebilehtedeni jõuda  
märksõnade kaudu;
- veebiteenused, nagu näiteks Google Translate, anna-  
vad kiireid ligikaudseid tõlkeid;
- sotsiaalmeedia platvormid, näiteks Facebook, Twit-  
ter ja Google+, lihtsustavad suhtlust, koostööd ja in-  
fovahetust.

Kuigi neist tööriistadest ja rakendustest on abi, ei suuda need veel toetada jätkusuutlikku mitmekeelset Euroopa ühiskonda, kus info ja kaup liiguksid vabalt.

## 2.1 KEELEPIIRID TÕKESTAVAD EUROOPA INFOÜHISKONDA

Me ei oska täpselt ennustada, milline näeb välja tuleviku infoühiskond. Kuid on väga tõenäoline, et kommunikatsioonitehnoloogia revolutsioon ühendab uuel moel eri keeli kõnelevaid inimesi. See paneb inimesed uusi keeli õppima ja arendajad looma uusi rakendusi, mis aitaksid kaasa üksteisemõistmisele ja võimaldaksid juurdepääsu jagatud teadmisele. Uued meedialiigid seovad üha rohkem keeli, kõnelejaid ja teavet, mis liigub ülemaailmses majandus- ja infosfääris. Sotsiaalmeedia (Wikipedia, Facebook, Twitter, YouTube, viimasel ajal ka Google+) praegune populaarsus on vaid jäämäe tipp.

Tänapäeval saame saata gigabaitides teksti ümber maailma kõigest paari sekundiga, enne kui taipame, et see oli kirjutatud keeles, mida me ei mõista. Euroopa Komisjoni hiljutise uuringu kohaselt ostab 57% internetikasutajatest Euroopas tooteid ja teenuseid keeltes, mis ei ole nende emakeel. Kõige levinum võõrkeel on inglise keel, sellele järgnevad prantsuse, saksa ja hispaania keel. 55% kasutajatest loeb võõrkeelseid materjale, samas kui vaid 35% kasutab teist keelt ise meilide kirjutamisel või veebikommentaare postitamisel [4]. Mõned aastad tagasi oli inglise keel interneti *lingua franca* – valdav enamus veebist oli inglisekeelne – ent praeguseks on olukord drastiliselt muutunud. Teistes Euroopa keeltes (aga ka Aasia ja Lähis-Ida keeltes) oleva materjali maht on internetis plahvatuslikult kasvanud.

Üllataval kombel pole see keelepiiridest tulenev üldlevinud digitaalne lõhe pärvinud kuigi suurt avalikkuse tähelepanu. Samas tõstatab see pakilise küsimuse: milliseid Euroopa keeli saadab võrgupõhises info- ja teadmusühiskonnas edu ja millised on määratud kaduma?

---

Maailmamajandus ja inforuum seavad meid vastamisi erinevate keelte, kõnelejate ja sisuga.

---

## 2.2 MEIE KEELED ON OHUS

Kuigi trükipress aitas kaasa Euroopasisese infovahetuse kiirenemisele, viis see ka paljud Euroopa keeled väljasuremiseni. Piirkondlikke ja vähemuskeeli trükiti harva, nii säilisid näiteks korni ja dalmaatsia keel vaid suulisel kujul, see omakorda piiras oluliselt nende kasutusvaldkonda. Kas interneti mõju meie keeltele on samasugune?

Euroopa ligi 80 keelt on üks tema väärtuslikumaid ja tähtsamaid kultuuriväärtusi ning eluline osa tema ainulaadsest ühiskonnamudelist [5]. Samal ajal kui inglise või hispaania keelel pole tõenäoliselt probleeme tekkival digitaalsel turul ellujäämisega, võivad mitmed Euroopa keeled võrguühiskonnas vähetähtsaks jääda. See omakorda aga nõrgestaks kogu Euroopa positsiooni maailmas ja oleks vastuolus meie strateegilise eesmärgiga kindlustada võrdsed võimalused kõigile Euroopa kodanikele, olenemata nende emakeelest.

---

Euroopa keeleline mitmekesisus on meie üks rikkamaid ja olulisimaid kultuurivarasid.

---

UNESCO mitmekeelsuse raporti järgi on keeled hädavajalik vahend oma põhiõiguste, näiteks poliitilise väljendusvabaduse, hariduse ja ühiskonnas osalemise tagamiseks [6].

## 2.3 KEELETEHNOLOOGIA ON VÕTMEHNOLOOGIA

Varem tähendas keele säilitamine keeleõppele ja tõlkele keskendumist. Arvatakse, et 2008. aastal oli tõlkimise, tarkvara lokaliseerimise ja veebilehtede globaliseerimise turuosa Euroopas 8,4 miljardit eurot, ning ennustatakse, et see kasvab 10% aastas [7]. Samas katab see summa vaid väikese osa praegusest ja tulevast keeltevahelisest kommunikatsioonivajadusest. Ahvatlev la-

hendus tagamaks tuleviku Euroopas keelekasutuse laia katvust ja head kvaliteeti oleks keeletehnoloogia kasutamine, samamoodi nagu me kasutame tehnoloogiat transpordi- ja energiavajaduste rahuldamiseks.

Digitaalne keeletehnoloogia hõlmab kõiki kirjaliku ja suulise keele suhtluse vorme. Seega soodustab ta koostööd, äritegevust, teadmiste jagamist ning ühiskondlikus ja poliitilises diskussioonis osalemist, sõltumata seejuures kasutaja võimalikust keelebarjäärist ja arvutikasutamise oskuse tasemest. Sageli on keeletehnoloogia juba keerulistes süsteemidesse lõimitud ja see aitab meil:

- otsimootori abil veebist informatsiooni leida;
- tekstiredaktoriga õigekirja ja grammatikat kontrollida;
- veebipoes tootesoovitusi näha;
- auto navisüsteemi hääljuhiseid kuulda;
- internetiteenuste abil veebilehti tõlkida.

Keeletehnoloogia koosneb mitmetest kesketest rakendustest, mis suuremas rakenduste raamistikus on vajalikud teiste programmide tööks. META-NETi keele raportite eesmärgiks välja selgitada iga Euroopa keele tuumikrakenduste tase.

---

Euroopa vajab veakindlat ja kättesaadavat keeletehnoloogiat kõigi Euroopa keelte jaoks.

---

Jätkuvalt ülemaailmselt innovatiivseks eeskujuks olemiseks vajab Euroopa kõigile oma keeltele kohandatud keeletehnoloogiat, mis oleks nii robustne (veakindel) kui taskukohane ja samas olulisematesse IT-süsteemidesse tihedalt lõimitud. Lähitulevikus ei jõuta ilma keeletehnoloogiata mitmekeelse ning tõeliselt efektiivse ja interaktiivse multimeediapõhise kasutajakogemuseni.

## 2.4 KEELETEHNOLOOGIA VÕIMALUSED

Trükitehnika läbimurdeks oli võimalus teksti (lehekülge) trükipressi abil kiiresti kopeerida. Teadmiste otsimise, lugemise, tõlkimise ja kokkuvõtmise raske töö jäi inimestele. Kõne salvestamiseks tuli oodata Edisoni – ja ka tema tehnoloogia suutis luua kõigest analoogkoopiaid. Kaasaegne keeletehnoloogia võimaldab automatiseerida kõigis Euroopa keeltes tõlkimise, sisutootmise ja teadmushalduse. Tänu sellele on võimalik luua kodulektroonikale, masinatele, sõidukitele, arvutitele ja robotitele intuiitiivseid keelel ja kõnel põhinevaid kasutajaliideseid. Reaalselt kasutatavad äri- ja tööstusrakendused on praegu alles arendamise algusjärgus. Kuid saavutused teadusvallas on tekitanud rakenduste loomiseks uusi võimalusi. Nii näiteks töötab masintõlge kindla valdkonna raames juba mõistliku täpsusega ning on olemas eksperimentaalseid rakendusi, mis pakuvad mitmekeelset infot, teadmushaldust ning sisutootmist paljudes Euroopa keeltes.

Nagu teistegi tehnoloogiatega, loodi ka esimesed keeletehnoloogia rakendused (kõnepõhised kasutajaliideseid ja dialoogisüsteemid) kindlatele valdkondadele ning seetõttu oli nende efektiivsus sageli piiratud. Tohtu turupotentsiaaliga on haridus- ja meelelahutustööstus. Keeletehnoloogiat lõimitakse mängudesse, harivasse meelelahutusse, raamatukogudesse, simulatsioonidesse ja treeningprogrammidesse. Keeletehnoloogia mängib olulist rolli mobiilsetes infoteenustes, arvutipõhises keeleõppetarkvaras, e-õppe keskkonnas, enesehindamisprogrammides, plagiaatide tuvastamise tarkvaras ning paljudes teistes rakendusvaldkondades. Twitteri- ja Facebookilaadsete sotsiaalmeediarakenduste populaarsusega kaasneb suurenenud vajadus keeletehnoloogia järele, mis peaks jälgima postitusi, võtma kokku arutelusid, hindama arvamustrende, leidma emotsionaalseid vastuseid, tuvastama ja jälitama autoriõiguse rikkumisi ja väärkasutust.

Keeletehnoloogia loob Euroopa Liidule tohutuid võimalusi. See aitab lahendada keerulisi mitmekeelsuse probleeme, mis tekivad Euroopa ettevõtetes, asutustes ja koolides erinevate keelte koos kasutamise tõttu. Keeletehnoloogia võimaldab kodanike suhtlust Euroopa ühisturul, kõrvaldades takistavad keelebarjäärid, ent samas toetades üksikute keelte vaba kasutust.

---

Keeletehnoloogia aitab saada üle keelelise mitmekesisuse “puudest”.

---

Tulevikus on Euroopa innovaatiline mitmekeelne keeletehnoloogia eeskujuks meie ülemaailmsetele partneritele, kui nad alustavad oma mitmekeelsete kogukondade toetamisega. Keeletehnoloogiat võib pidada tugitehnoloogiaks, mis aitab jagu saada keelelise mitmekesisuse “puudest” ja muudab keelekogukonnad üksteisele lihtsamini ligipääsetavateks.

Lõpuks veel ühest aktuaalsest uurimisvaldkonnast – keeletehnoloogia kasutamisest katastroofiirkondade päästeoperatsioonidel. Kriisiolukorras tegutsemine võib olla elu ja surma küsimus, seega keelest sõltumata oskustega intelligentsed robotid suudaksid päästa elusid.

## 2.5 KEELETEHNOLOOGIA VÄJAKUTSED

Kuigi viimastel aastatel on keeletehnoloogia märkimisväärselt arenenud, on praegune tehnoloogiline edasiminek ja tooteinnovatsioon siiski liiga aeglased. Laialdaselt kasutatavad tehnoloogiad, nagu tekstiredaktorite spellerid ja grammatikakorrektorid, on tüüpiliselt ükskeelsed ja saadaval vaid loetud keeltele.

---

Praegune tehnoloogilise arengu tempo on liiga aeglane.

---

Veebipõhised masintõlketeenused on küll kasulikud dokumendi sisust kiire ülevaate saamiseks, ent nad jäävad hätta täpse ja täieliku tõlkega. Inimkeele keerukus raskendab loomuliku keele modelleerimist tarkvaras ning rakenduse tegelikus elukeskkonnas testimine on pikk ja kulukas protsess, mis vajab järjepidevat rahalist toetust. Selleks, et Euroopa oleks endiselt mitmekeelse kogukonna tehnoloogia teerajaja rollis, tuleb leiutada uusi meetodeid arengu kiirendamiseks. Need hõlmavad nii tarkvaralisi uuendusi kui *crowdsourcingu* stiilis tehnikaid.

## 2.6 KUIDAS INIMESED JA MASINAD KEELT OMANDAVAD

Et näitlikustada, kuidas arvutid keelt käsitlevad ja miks on nii raske arvuteid loomuliku keele kasutamiseks programmeerida, anname lühikese ülevaate sellest, kuidas inimesed keelt omandavad ning kuidas keeletehnoloogiasüsteemid töötavad.

---

Inimesed omandavad keeleoskuse kahel viisil: õppides näidetest ja õppides keelereegleid.

---

Inimesed omandavad keeli kahel erineval viisil. Väikelapsed omandavad emakeele vanemate, õdede-vendade ja teiste pereliikmete vahelist suhtlust kuulates. Umbes teisel eluaastal lausuvad lapsed oma esimesi sõnu ja lühikesi fraase. Keeleõpe on võimalik ränu inimeste geneetilisele soodumusele kuuldut imiteerida ja mõtestada.

Vanemas eas nõuab teise keele omandamine suuremat pingutust, peamiselt seetõttu, et õppija ei kuulu emakeelel kõnelejate kogukonda. Koolis õpitakse võõrkeletundides tavaliselt selgeks keele grammatiline struktuur, sõnavara ja õigekiri. Õppimiseks kasutatakse harjutusi, mis kirjeldavad keelelist teadmust abstraktsete reeglite,

tabelite ja näidete abil. Vanemaks saades muutub võõrkeele omandamine raskemaks.

Kaks peamist keeletehnoloogiasüsteemides kasutatavat meetodit “omandavad” keelelised oskused sarnasel viisil. Statistilised ehk andmejuhitud meetodid omandavad keelelise teadmuse suurtest näidistekstide kogudest. Kui näiteks spelleri treenimiseks piisab ükskeelsetest tekstidest, siis masintõlkesüsteemi treenimiseks läheb vaja paralleeltekste kahes või enamas keeles. Treeningtekstidest “õpib” masintõlkealgoritm sõnade, fraaside ja lausete tõlkimiseks mustreid.

Selline statistiline lähenemine vajab toimimiseks miljonid lauseid. Mida rohkem näitetekste analüüsitakse, seda parem tõlketulemus saadakse. Tekstiredaktorites olev speller ning näiteks Google'i otsingumootor ja tõlge kasutavad statistilist lähenemist. Andmejuhitud meetodi eeliseks on see, et masin õpib järjestikustes treeningtsüklites kiiresti, kuigi tulemuse kvaliteet võib oluliselt varieeruda.

Teine meetod, mida keeletehnoloogias ja kitsamalt ka masintõlkes kasutatakse, on reeglipõhiste süsteemide loomine. Keeleteaduse, arvutuslingvistika ja arvutiteaduse valdkonna eksperdid kodeerivad esmalt grammatilised analüüsid (tõlkereeglid) ja koostavad sõnade nimestikud (leksikonid). See on vägagi aeganõudev ja tömahukas tegevus. Mõnda juhtivat tõlkesüsteemi on pidevalt arendatud juba üle kahekümne aasta. Reeglipõhiste süsteemide suureks eeliseks on asjaolu, et eksperti-

del on keele töötamise üle täpsem kontroll. See teeb võimalikuks tarkvaras leiduvate vigade süstemaatilise parandamise ja kasutajale täpsema tagasiside andmise, seda eriti siis, kui reeglipõhised süsteemid on kasutuses keeleõppe abina. Kõrge kulu tõttu on seni reeglipõhiseid süsteeme arendatud üksnes suuremate keelte jaoks.

---

### Keeletehnoloogiasüsteemide kaks peamist tüüpi omandavad keelt samal viisil.

---

Kuna statistiliste ja reeglipõhiste süsteemide plussid ja miinused kalduvad teineteist täiendama, siis uuemad uurimused keskenduvad neid lähenemisi kombineerivatele hübriidsüsteemidele. Kahjuks pole need süsteemid seni tööstusrakendustes sama edukad olnud kui teaduslaborites.

Käesolevast peatükist selgus, et paljud tänapäeva infoühiskonnas laialt levinud rakendused on tihedalt seotud keeletehnoloogiaga. Võttes arvesse meie mitmekeelset kogukonda, kehtib see väide iseäranis selgelt Euroopa majandus- ja infosfääri puhul. Kuigi keeletehnoloogia on viimastel aastatel märkimisväärselt arenenud, on veel kõvasti arenguruumi süsteemide kvaliteedi parandamise osas.

Järgnevalt toome välja eesti keele rolli Euroopa infoühiskonnas ja hindame eesti keele keeletehnoloogilise toe praegust seisut.

The Estonian Language in the Digital Age

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

2012, VI, 75 p. 24 illus. in color., Softcover

ISBN: 978-3-642-30784-3