

Aasta tagasi ilmunud „Eesti keele arendamise strateegia 2004–2010” seires on öeldud: „Täidetud või täitmisel on kõik EKASis ja riiklikus programmis „Eesti keele tehnoloogiline tugi (2006–2010)” ette nähtud ülesanded” (lk 23). Vaadake ema-keelepäeval üle, milline on seis aasta pärast.

19.–21.11.2007 peeti TTÜ küberneetika instituudis riikliku programmi „Eesti keele keeletehnoloogiline tugi (2006–2010)” konverents.

Vt <http://keeletehnoloogia.cs.ut.ee/konverents/programm> ja kuula Vikerraadio „Keelekõrva” arhiivis [http://www.vikerraadio.ee/index.php?lang=est&main\\_id=848](http://www.vikerraadio.ee/index.php?lang=est&main_id=848)

Riikliku teadusarendusprogrammi „Eesti keele keeletehnoloogiline tugi 2006–2010” (EKKTT) eesmärk on toetada eesti keele tehnoloogia arendamist (täistekst HTMLi kodulehel [www.hm.ee](http://www.hm.ee) -> Teadlasele -> Riiklikud programmid). Peamised kaks alajaotust on keeleressurside ehk korpusete ja keelematerjalide kogumine, süstematiseerimine ja

analüüs ning keeletehnoloogia meetodite edendamine. Keeletehnoloogia vallas rahastatakse nii keele olulisemate aluste nagu morfoloogia, süntaksi ja semantika edendamist, infootsingute tarkvara arendamist, masintõlke-alast uurimistööd, suulise kõne analüüsi ja sünteesi jms. Tarkvara tuleks arendada vähemalt töötavate prototüüpide tasemele, et saaks

usutavalt demonstreerida meetodite kohandavust eesti keelele. Praeguseks on uurimisrühmi rahastatud üle 20 projekti puhul. Rahastamine on planeeritud vähemalt aastani 2010, sealjuures stabiilse tõusuga. Selles on oma positiivne roll nii praegusel valitsusk koalitsioonil kui ühiskonna parteiülesel poliitilisel tahtel. J. V.

## Arvutite eesti keele oskusest

Jaak Vilo, EKKTT juhtkomitee esimees

Arvutid on pöördumatult imunud meie igapäevaelu. Olgu see hall iga-va kujundusega kast laua all juhtme-rägastikuga tolmu kogumas, edev *laptop*, nutitelefon, auto navigatsio-niseade, DVD-mängija või muu riist, mille puhul tavakasutaja ei mõtlegi, et see on tegelikult ennekõike arvuti.

Kuigi eestlased ei ole leiutanud arvutit ise ja enamik tarkvarast ei ole toodetud Eestis, teeme oma tööd üha rohkem arvutiga, suhtleme omavahel emakeeles ning loome uut eestikeelset sisu. Maa ilma arvutipõhine infoühiskond ei tohi tähendada loobumist oma keelest. Kuna eesti keel on meile kõigile jätkuvalt armas – ja mitte ainult tänasel emakeelepäeval –, on meil vaja lahendada oluline küsimus: kuidas tagada, et arvuti teeks võimalikuks eestikeelse suhtlemise ja oskaks toetada eesti keele nüansse?

Asi pole vaid programmide me-nüüdes ja kurikuulsates veateadetes. Kuidas suhelda arvutiga häälega edas-tatud käskude abil? Keeletehnoloogi-listel rakendustel põhinev eesti keele tugi arvutitele on sisuliselt suulise ja kirjaliku keele arvutite „arusaada-vaks” tegemine, arvutite programmeerimine eesmärgiga, et programmid suudaksid keelt selle rikkuses „mõista”, tekitada, vigu parandada jms.

Eesti keele keeletehnoloogiline tugi arvutite tähendab paljusid asju. Ühel päeval loevad arvutid suvalist teksti eesti keele häälduse, intonatsiooni ja rütmiga piisavalt usutavalt ette; kirjanikul on vaja vaid diktofoni, kus kuuldu muudetakse automaatselt kirjalikuks tekstiks; avalikud infokioskid reageerivad kõnele; televisiooni ja raadio heliarhiivist saab otsida sõna, väljendit või nime nagu praegu Google’ist. Interneti-otsing leiab eestikeelse päringu peale üles ka samateemalised inglise, prantsuse ja muukeelsed leheküljed ning pakub automaattõlget. Keeletehnoloogia rakenduseks on muidugi ka vormikontroll ehk speller, et klõbistamisel sisse juhtunud õigekirjavead üles leida ja parandused välja pakkuda. Ka puuetega inimesed saavad mugavalt suhelda – nii arvutiga kui teiste inimestega, arvuti vahendusel.

Siin on segamini soovunelmad ja tegelikkus. Nii mõnigi asi on juba praegu laiatartarkvaras, mõnd alles katsetatakse laboris, miski võib jõuda kasutusse paari, miski ehk alles 10 aasta pärast. Mõni unistus ei täitu ka saja aasta pärast – ei räägi ju keegi sellest, et arvuti peaks tõesti mõistma, mida tekst või kõne tähendab, ja superintelligentust ilmutama.

Selge on aga see, et rahvusvaheline teadmus tuleb rakendada selleks, et ka eesti keelele tekiks suurte keeltega võrreldavad võimalused. Selle töö peavad ära tegema eestlased ise, vähe-

malt seni, kuni masinõppe meetodid ei ole arenenud nii kaugele, et keeletehnoloogia parimad lahendused tulevad inimestelt, kes vastavat keelt üldse ei oskagi. See ei ole tegelikult üldse absurdne mõte, et eesti keele keeletehnoloogia osasid arendataks välja mõne suure elektroonilise arhiivi, näiteks Google’i andmete alusel. Kui andmebaasis on piisavalt eesti keele näiteid koos vastava kõne ja tõlgetega, saab nende andmete pealt treenida masinad keele reegleid „ära õppima”. Kuid kas see kasvatab meie teadmisi meie keele (eba)reeglipärasuse osas? Kas me teame kõigi oma keele sõnade tähendusi ja varjatud tähendusi?

Et keeletehnoloogia saaks arenda, on vaja esiteks mõelda, mida saab arvutite üldse keele kohta „selgeks teha” ja kuidas arvutit selle eesmärgil „õpetada”. Mõned omadused tuleb programmidesse sisse kirjutada tarkvarakoodi osana või reeglina, teised aga masinõppe (ka tehisõpe, ingl *machine learning*) meetoditega alles andmetest tuvastada, kolmandate puhul on vajalik mõni uudne idee ja realiseerida see töötava programmina.

Et rääkida arvutiga, on vaja teada, kuidas eestlased eesti keelt ise kasutavad või vähemalt kuidas seda oleks korrektna kasutada. Kirjalikud ajalehe- ja ajakirjaartiklid, seadusetehtid, ilukirjanduse žanrid, elektronposti ja jututubade keel, uudiste tekstid, suulised raadio- ja telesinemed, peokõned, telefonivestlused jne – kõik need

on eesti keele osad. Tuleb teada sedagi, kuidas kasutavad keelt need, kes seda veel ei oska, s.t lapsed, keeleõppijad, ja arvesse võtta ka aina uusi eriala või suhtlusringkonna alamkeeli ja slängi. Kirjaliku ja suulise kõne näited tuleb kokku koguda, süstematiseerida ja need arvutite arusaadavalt kirjeldada. Selleks koostatakse vastavalt kirjaliku ja suulise kõne andmebaase ehk korpusi, mis on aluseks keeletehnoloogilise tarkvara väljatöötamisel. Suur keeleressurs on kogutud sõnas-

tikesse, mis aitavad defineerida sõnu ja nende tähendusi ning sõnade ja terminite vahelisi seoseid, ka teiste keelte suhtes.

Programmi projektide ([www.keeletehnoloogia.ee](http://www.keeletehnoloogia.ee)) edenemisest sai hea ülevaate 19.–21.11.2007 EKKTT teaduskonverentsil, kus kõik rühmad esitasid oma senise töö tulemusi ja plaanitavat. Loodame, et nii mõnigi projekt annab lahendusi, mida peame peagi loomulikuks infoühiskonna toimimise eelduseks.

### EKKTT projektid <http://www.keeletehnoloogia.ee/projektid>

PROJEKT	PROJEKTIJUHT	ASUTUS
Eestikeelne korpuspõhine kõnesüntees	Meelis Mihkla	EKI
Eesti emotsionaalse kõne korpus	Hille Pajupuu	EKI
Leksikograafi töökeskkond	Ülle Viks	EKI
Korpusepäring keeleveebis	Heiki-Jaan Kaalep	Filosoft OÜ
Eestikeelse kõnetuvastuse meetodite uurimine ja arendamine	Tanel Alumäe	TTÜ
Kõne analüüs ja variatiivsuse mudelid	Einar Meister	TTÜ
Kõnekeele ressursid ja kõnetehnoloogia andmebaasid	Einar Meister	TTÜ
Eesti vanema kirjakeele elektroonilised kogud	Külli Habicht	TÜ
Eesti kõnekeele korpus kogumine ja translitereerimine	Tiit Hennoste	TÜ
Masintõlge I	Heiki-Jaan Kaalep	TÜ
Mitmesõnaliste verbide ja nende kokku-lahku kirjutamise vigade äratundmine eestikeelsetes tekstides	Heiki-Jaan Kaalep	TÜ
Veebipõhine interaktiivne keeleõpe ja selleks vajalikud ressursid	Birute Klaas	TÜ
Eestikeelne infodialoog arvutiga	Mare Koit	TÜ
Eesti keele koondkorpus	Kadri Muischnek	TÜ
TÜ eesti keele tesauruse (eesti wordneti) täiendamine	Heili Orav	TÜ
Süntaksianalüüsil põhinev keeletarkvara ning selle arendamiseks vajalikud keeleressursid	Tiit Roosmaa	TÜ
Eesti keele spontaanset kõne foneetiline korpus	Pire Teras	TÜ
Reeglipõhine keeletarkvara	Jan Villemson	TÜ
Elektrooniliste teatmeteoste kasutajasõbralikud päringusüsteemid	Jaak Vilo	TÜ
Lihlause semantiline analüüs	Haldur Õim	TÜ

Eesti keeletehnoloogia sihtprogramm <http://www.eki.ee/keeletehnoloogia/lingid.html>  
Keeletehnoloogia programmi raames tehtud töid <http://www.eki.ee/keeletehnoloogia/projektid/>

## Kuidas leida nõel heinakuhjast?

Jaak Vilo, TÜ arvutiteaduse instituudi professor

Kõige lihtsam on võtta abiks väga tu-gev magnet, kui heinakuhjas on tõesti üks või vaid mõni metallise. Aga kui kuhjas on palju nõelu, heina ja muud sodi ning tahetakse leida mingi kindla kuju või omadusega ese? Täpselt nii otsime Interneti miljardite veebilehekülgede hulgas (heinakuhi) seda ühte ja õiget, kui kasutame Netit või Google’i otsimootoreid. Kuidas leida Internetist, sõnastikust või teatmeteostest vastus võimas arvutipark (emafirma DEC/Compaq paigutas sinna oma uued suured uhked 64-bitise arhitektuuriga serverid). Vaja läks

aga vaid mõnd uut head ideed ja väga head teostust ning platsi löi puhtaks praegu gigandiks kasvanud Google. Mille poolest oli Stanfordini tudengite loodud otsimootor [google.stanford.edu](http://google.stanford.edu) parem kui teised? See oskas paremini eraldada teri sõkaldest: võeti kasutusele esmapilgul üsna lihtsad relevantsuse hindamise kriteeriumid. Näiteks kui otsida tõelist staari, siis ei ole oluline, et keegi ise end staariks kuulutab, vaid hoopis see, kui tihti võimalikult autoriteetsed leheküljed sellele tiitlile viitavad.

Enamikul info-otsisüsteemidest, olgu need Interneti otsimootorid, andmebaasid või teatmeteosed, on siiski üks viga: otsitav sõna on lihtsalt üks täheühend. Kuidas otsida dokumenti, kui kasutaja on otsitava sõna näiteks kuulmise järgi lihtsalt valesti kirja pannud või kui mõni klahvivajutus on kogemata valesti läinud? Kuidas otsida sõna kõiki esinemisi, kui see võib dokumentides olla mõnes meie paljudest käänevormidest? Käänamise ja pööramise juures on abiks,

kui arvuti teab midagi eesti keele sõnamoodustusest ehk morfoloogiast. Tekste saab morfoloogiliselt analüüsida, et tuvastada sõna algvorm. Või rakendada sünteesi, s.t genereerida kõik võimalikud lubatud käände- ja pöördelõppudega versioonid. Esimese võimaluse puhul saab otsida sõnu nende algvormi järgi, teisel juhul korraga kõiki eri käänevorme. Valesti kirja pandud sõna puhul tuleks leida ka kõige tõenäolisemad ligikaudsed esinemisvormid.

Võõrsõnastiku kasutajaliideses mitme aasta jooksul tehtud pärin-gutes torkas mulle silma, et sagedamini otsitud sõnade hulgas olid „konstanteerima” ja „konstateerima”. Esimest (n-iga variant, mis sobib sõnasse „konstant”) esines küll rohkem, kuid Keeleveeb andis esimese päringu tegijatele tühja vastuse, sest sõnaraamatus sellist sõna ei esine. Kui lubada sõna otsimisel ühe tähe väljajätmist, saadi ka vastus. Keeleveeb toetas küll ka sarnasusel põhinevaid päringuid, kuid olukord polnud siiski kasutaja-

le kõige parem. Sealt tuli idee hakata arendama otsimootorit nii, et kasutaja ei jääks (peaaegu) kunagi tühjade kätega: arvuti peab ise aru saama, mida kasutaja tegelikult otsib, ja vastuse ka korrektselt esitama. Kesken-dume EKKTT programmi projektis „Elektrooniliste teatmeteoste kasutajasõbralikud päringusüsteemid” ennekõike sisestatud sõnaga sarnaste esinemiste otsimisele.

Sarnasust saab mõõta selle järgi, kui mitu tähte sõnas on „kadunud”, juurde lisatud, ära vahetatud või lihtsalt vales järjekorras. Ja kui juba lubada vigu, miks lubada ainult ühe tähe kaupa? On ju näiteks eesti keeleki mõned tähed-häälikud vahetuses. Näiteks ž vs. zh (režissöör vs. rezhissöör), w ja v (kawe vs. kave). Ja kui juba lubada täheühendite asendusi, siis miks mitte lubada päringuid, mis tuvastaksid sõnad restoran vs. PECTOPAH (arvuti jaoks on ladina ja kirillitsa tähed eri koodiga). Ehk miks mitte lubada eestlastel vene- ja ingliskeelseid sõnu häälduse järgi kirja

panna? Ja kui tore oleks veel otsida Eesti isikunimesid korraga nii eestikeelset venekeelsest meediast.

Lähtume ideest, et päringumootoris peab olema võimalik väljendada päringut enam-vähem täpselt ning lubada näiteks vene-eesti sõnastiku kasutust ka siis, kui ei ole venekeelset klaviatuuri ja kasutaja saab sõnu kirja panna ladina tähestikus.

Praegu on olemas kahe ligikaudse otsingu algoritmi esimesed prototüübid. Üks otsib kiiresti nn regulaaraval-disi ja teine lubab treenida näidete pealt tüüpilisi asendusi, mida päringumootori arendaja tahab lubada. Esimese algoritmi realiseerimise autor on **Kristo Tammeoja** ja teisel **Reina Käärrik**, mõlemad Tartu ülikooli arvutiteaduse instituudi doktorandid. Selle aastaga lõpeb projekt ära, need algoritmid on koos kasutusnäidetega valmis ja loodetavasti rakendatud ka esimestes andmebaasides. Kui uudiseid tuleb, levitame neid kindlasti [www.keeletehnoloogia.ee](http://www.keeletehnoloogia.ee) kaudu.

# Keele tugi ja sõnum

Andres Raid

Möödunud aasta novembris toimus küllaltki esinduslik konverents, mis oli pühendatud riikliku programmi „Eesti keele keeletehnoloogiline tugi” hetkeseisule ja suundumustele. Ilma erilise kärata on tehtud küllaltki palju: eesti kõnekeele korpuse kogumine ja transliteerimine, eesti keele spontaanse kõne foneetiline korpus ja palju muudki on n-ö tööstaadiumis ja osa tulemusigi juba „käega” katsuda. Tõsi, semantiline analüüs on veel tulevikumuusika, kuid algus on tehtud. Senine tegevus on olnud paraku projektipõhine ja seetõttu on (ehkki tegijate seas nurinat kuulda pole olnud) nagu ikka projektipõhisuse korral suur osa energiast läinud rahastaja meelitamisele, kauplemisele ja muule administreerivale tegevusele, millel pole loomingu ja otsese uurimistegevusega suurt pistmist.

Omad ohud siin programmis siiski on. Meie igapäevases keelepruugis on suur hulk mõisteid, termineid ja terveid paradigmasid, mis on senistega võrreldes teisenenud. Sama võib muide täheldada ka konverentsi materjale uurides.

Üks üldine häda on meil subjektiga. **Kas subjekti kõrvaldamine keelest on teadlik tegevus või on see juhtunud lihtsalt kogemata?** Raske öelda. Siiani pole ma leidnud näiteks projekti, millel poleks eesmärki – see konstruktsioon on levinud ja kahjuks paljudel juhtudel ka omaks võetud. Mitmed õpetatud pead on aastaid juhtinud tähelepanu tõigale, et eesmärgid saavad olla vaid subjektidel ehk inimestel ning ka vastutada saavad ainult nemad – siiaamaani tulutult. Eesmärgid on meil ikka ülikoolidel ja teaduskondadel, inimese eesmärgile viivast sihikindlust ja -pärastest tegevusest juttu ju olla ei saa! Eesmärgid saavad siiski olla vaid ühes või teises projektis osalejatel või selle koostajatel. Eesmärgid on kindlasti ka nende projektide tulemuste kasutajatel nagu ka vastutus, mida projekti kanda kuidagi panna ei õnnestu. Häda on üleüldine. Probleemid ja ülesanded on koolidel, neid (ülesandeid) ei anta enam ei ametnikele ega teenistujatele, vaid ametkondadele ning ministriumidele. Seadused kohustavad ja ütlevad, liiklusõnnetusi põhjustab libedus või mahasadanud lumi! Kui igapäeva kõnepruugis on see veel talutav, siis ametlike tekstides peaks see olema välistatud. Igasuguste tõrgete või raskuste korral – olgu tegu kooli, vallavalitsuse, toimetuse või linnaga – laseb väljendusde kodeeritud vastutamatus tegelikel vastutajatel end tunda üsna ohutult. Keele arenguloogika järgi võib subjekt kui selline tendentsi jätkudes üldse ära kaduda: kaob kõigepealt sõna, veidi hiljem ka mõiste.

Kui keel sel kombel administratiivse vankri ette rakendatakse eesmärgiga garanteerida ametnike muretus vastutuse pärast, on asi enam kui kriitiline. Pealegi lubab selline konstruktsioon üsna edukalt varjata oma intellektuaalsete ja ehk ka administratiivsete võimete mitte just eriti vaimustavat seis. Ei usu, et emakeelt oleks mõistlik lahterdada kuidagi teisiti, kui seda seni on tehtud: kirja-, kõnekeel, suulised ja

kirjalikud tekstid ja nii edasi. Ometi on püütud mängu tuua lisaks põlualusele kantseliidile veel ka **teaduskeele mõistet**, kuhu (vähemalt reaalteaduste vallas) eesti keelel enam kahjuks asja polegi (Ilmar Tomusk) ja Juhan Parts on ministri-na välja toonud isegi vajaduse rääkida hoopis uues, **turumajanduse keeles**: „Noori tuleb õpetada raha tegema. Kõlab Viru ärika moodi, aga kapitalistlik maailm, turumajanduslik ühiskond – meil tuleb rääkida selles keeles”. Mis keel see selline on ja kas see ka kuskile kõnekorpusse ära mahutatakse, seda ei või teada.

Keeletehnoloogiline tugi võimaldab kindlasti kaasajastamist ja muudatuste sisseviimist, lihtsam on aga kõike seda juba ette arvesse võtta, see omakorda aga peaks eeldama (**tingiv kõneviis on vähemalt poliitilistes tekstides asendanud kindla üsna kindlalt**) mingitki selgust kas või akadeemilistes ringkondades. Seal võib paraku tihtipeale kohata „näimist” „olemise” asemel ja konformsus on kahjuks tõsiteadusliku tegevuse, mille puhul konfliktid on vältimatud, alternatiiv.

See näivus peidab endas terve hulga ohtusid. Kahjuks on tendents toota kõikvõimalikke teadusliku staatusega tekste, mis sisult otsekui Sociali generaatorist tulnud. Täpsustamist ja seletamist nõudvaid küsimusi käsitletakse isikliku kallaletungi ja intrigeerimisena, argumendina kõlbavad imehästi kunagi kuskil võõrkeeles ilmunud kirjutised, mis on ühtlasi teadusliku kompetentsi vaieldamatuks tõendiks. Tundub siiski üsna kahtlane, et meil on sellisel hulgal kas või inglise keelt valdavaid teadlasi, kes küll enda eesti keeles täpse väljendamisega hätta jäävad, seda mängleva kergusega aga mõnes võõrkeeles teevad. Selles, nagu üldse kõiges ametlikus, pole aga kombeks kahelda ja seetõttu võib igaüks, kes jalad kamina poole välja sirutab, üsna kergesti end Giordano Brunoks pidada.

Loomulikult ei asenda keeletehnoloogiline tugi, vaatamata oma võimalustele, elavat keelt, kus sama fraas või sõna võib väljaöelduna panna kas naerma või nutma ning vastu kajama midagi, mis on hing, või jätta selle hoopis puutumata. Keel ise on kõnelejate olemasolu korral imeliselt visa ning „tugi” on küll äärmiselt täpselt oma funktsiooni järgi paika pandud. Raske on praegu ette kujutada, kuidas saab elektroonilisteks märkideks kehakeel, mis saab märkidest ja sümbolitest, nende omavahelisest korrelatsioonist ja koodidest, kuidas teemast hakkab eristuma reema. Viimasega on keeruline juba niigi, ka ilma arvuti toeta, sest tihtipeale ei ole mitmesuguste ametlike tekstide puhul niikuini võimalik aru saada, mida on öieti tahetud öelda või mida silmas peetud. **Selline „udutamine” või „hermeneutiline jukerdamine” võibki olla aga ka eesmärk omaette.** Mõtte- ja väljendusallgus kui vaenlane?! Miks ka mitte!

Tundub siiski, kui keele tehnoloogilise toe juurde tagasi tulla, et tegijad mõistavad ja saavad aru, et keel on elus asi ning selle juurde kuuluv inimlik emotsioon ei pruugi olla arvuti pärusmaa. Loomisel tundub sümptoomil kombel olevat tark ja tõhus abivahend, aga mitte panatsea või koguni asi iseeneses.

## Kõnekeele ressursid ja kõnetehnoloogia andmebaasid

Einar Meister,

TTÜ küberneetika instituudi foneetika ja kõnetehnoloogia labori juhataja

Foneetika ja kõnetehnoloogia labori teadustöö on fokuseeritud põhiliselt suulise eestikeelse kõne eksperimentaalsele uurimisele ja kõnesignaalide analüüsiks, sünteesiks ning tuvastuseks sobivate meetodite arendamiseks.

Erinevalt kirjalikust tekstist ei koosne kõnesignaal lõplikult hulga diskreetsetest elementidest. Igal kirjalikul lausel võib olla lõpmatu hulk akustilisi realisatsioonid, mis sisaldavad lisaks lingvistilisele informatsioonile ka paralingvistilisi (kõnetempo, hääle valjus ja kõrgus, hääldusviis jms) tunnuseid ja ekstralingvistilisi (kõnestiil, emotsioonid jms) nähtusi. Kui inimesele on kõne suur variatiivsus loomulik, siis kõnetehnoloogias on see tõsine probleem: sünteeskõnes on seda liiga vähe – sünteeskõne on masinlik ja monotoonne, inimkõnes liiga palju – kõnelejate eri kõnestiilis kõneldu automaatne tuvastamine on ääretult keeruline.

Kõnevariatsioonide uurimiseks ja modelleerimiseks on vaja suurt hulka mitmekesist ja süstematiseeritud kõnematerjali. Just seda eesmärki teenibki foneetika ja kõnetehnoloogia labori projekt „Kõnekeele ressursid ja kõnetehnoloogia andmebaasid”.

Uurimisülesandest lähtuvalt vajatakse erinevaid kõnekorpusi. Fo-

neetikauuringutes kasutatakse nn laboratoorset kõnet, mis salvestatakse müravabas keskkonnas. Keelejuhtide valikul arvestatakse, et soo- ja vanuserühmad oleksid võrdselt esindatud; tekstid, mida keelejuhid loevad, on spetsiaalselt konstrueeritud ja uuritava nähtuse seisukohalt representatiivsed; keelejuhte on suhteliselt vähe, 10 kuni 100.

Eesti keele foneetikauuringuteks loodud kõnekorpus näiteks on Eesti BABEL, mis kavandati ja salvestati EL BABELi projekti (1994–1997) raames.

Kõnetehnoloogias vajatakse kõnetuvastuse treenimiseks ja testimiseks loomulikke kõnematerjali, mis salvestatakse rakendusomases akustilises keskkonnas (näiteks läbi telefonikanali, kui tuvastust rakendatakse telefoniteenuses) ja mis sisaldab valdavalt rakenduse-spetsiifilist sõnavara. Et tuvastussüsteem oleks kõnelejust sõltumatu, tuleb koguda paljude inimeste kõnenäiteid. Selliste kõnekorpusete puhul on keelejuhte 500 kuni 5000.

Eestikeelse kõnetuvastuse arendamiseks koguti 2000ndate algul Eesti SpeechDati korpus, mille eeskujuks olid EL SpeechDati projekti raames loodud suuremate euroopa keelte kõneandmebaasid.

Praegu on laboris käimas järgmiste korpusete kogumine ja töötlus. **Uudistekorpuse** on vajalik kõnetuvastuse rakenduse loomiseks, et automaatselt teisendada dikteeritud teksti elektroonseks dokumendiks ja

teostada otsinguid audioarhiivides. Korpus sisaldab ca 300 tundi 10 erineva raadiodiktoriga loetud Eesti Raadio lühiuudiseid. **Aktsendikorpusesse** kogutakse eesti keelt võõrkeelena kõnelevate isikute kõnenäited aktsendinähtude akustilis-foneetiliseks uurimiseks ja aktsendiga kõne tuvastamiseks sobivate akustiliste mudelite treenimiseks. Seni on salvestatud prantsuse, vene ja soome emakeelega keelejuhtide kõnenäiteid (igapähealt umbes 20 minutit), eesmärk on koguda kuni 100 isiku vähemalt 20 keele aktsendinäiteid. Kõnetuvastuse üheks rakenduseks on (pool)spontaanse kõne tuvastus, näiteks akadeemiliste loengute, konverentsiettekannete jms salvestuste automaatne teisendamine tekstiks. Tegemist on teksti ettelugemisega võrreldes erineva kõnestiiliga, mistõttu ei anna loetud kõnematerjalil treenitud tuvastussüsteemid spontaanse kõnesisendi korral kuigi häid tulemusi. Lisaks kõnetuvastuse treenimisele on **loengukõne korpus** vajalik ka selle kõnestiili akustilis-foneetilisteks uuringuteks.

Seni on salvestatud ainekursuste loenguid TTÜs (7 keelejuhti, kokku 54 tundi) ja konverentsiettekandeid (3 keelejuhti, kokku 2 tundi), eesmärk on koguda vähemalt 200 tundi kõnematerjali.

**Küberneetika instituudi kõneressursid** <http://www.phon.ioc.ee/dokuwiki/doku.php?id=projects:ressursid:ressursid.et>

## Eestikeelne infodialoog arvutiga

**Reisiagent**

**Küsimus:** Mul oleks vaja ka hotellis ööbimiskohta.  
**Vastus:** Hotellide kohta kahjuks infot ei oska anda. Räägime Tallinnast väljuvatest lendudest.

**Küsimus:**

Kuulan vastust sünteeskõnes

Mare Koit,

TTÜ keeletehnoloogia professor

Inglise ja mõne teise keele jaoks leidub juba arvukalt kõnedialoogsüsteeme osutamaks automaatset telefoniteenust reisi kavandamisel, piletide reserveerimisel, ülikooliastutajate nõustamisel jpm. Eesti keele puhul selline süsteem seni puudub. Seetõttu alustati Tartu ülikoolis projekti „Eestikeelne infodialoog arvutiga”, mille eesmärgiks on luua tarkvara, mis võimaldaks pidada küsimus-vastus-dialoogi arvutiga eesti keeles. Seejuures peaks see jälgendama inimese käitumist, näiteks oskama esitada täpsustavaid küsimusi, anda tagasisidet, tervitada ja tänada nii, nagu inimesed seda teevad. Loodav dialoogsüsteem peab mõistma kasutaja eestikeelseid päringuid, leidma vajaliku info andmebaasist, andma vastuse eesti keeles. Sellise tarkvara loomiseks tarvilik keeleressurs on dialoogikorpus, kus on märgendatud dialoogiaktid (küsimus, info andmi-

ne, tervitus jne). Eesti dialoogikorpus sisaldab eri liiki dialooge, millest mahukaim osa on TÜ eesti suulise keele korpusest valitud inimestevahelised infodialoogid (umbes 1000 transkribeeritud teksti, kokku 178 000 sõnavormi). Lisaks kuulub korpusesse ka mõnikümme inimese ja arvuti dialoogi, mis on kogutud dialoogsüsteemi arendamise käigus, sh simuleeritud dialoogid, kus arvuti rolli täidab teine inimene. Dialoogikorpus märgendatakse tüpoloogialusel, mille loomisel arvestati juba olemas tüpoloogiatega ja inimeste suhtluse uurimise vajadusega. Dialoogiaktide märgendamist hõlbustab programm, mille abil saab aktide puukujulisest hierarhiast valida sobiva akti ja paigutada selle märgendatavas tekstis vajaliku lausungi järele. Dialoogikorpus analüüsiks on loodud veebis kasutatav tarkvara, dialoogikorpus tööpink, mille abil saab korpusest välja valida alamkorpusi ja teha otsinguid eri kriteeriumide alusel, nt otsida mõne osaleja kindlaid dialoogiakte. Tööpingi ka-

sutamine (ja selle kaudu juurdepääs dialoogikorpusse) on paroolipõhine, sest inimestevahelised kõnelused on delikaatne materjal. Korpuse analüüsimise tulemusel on selgitatud dialoogiaktide vormistuslikke iseärasusi eesti keeles ja osalejate suhtluskäitumist. Nii on leitud keelelised märguanded tähtsamate infoaktide (direktiivid, küsimused) ja rituaalsete aktide tuvastamiseks eestikeelsetes suulistes dialoogides. Katsetatud on mitmeid dialoogiaktide automaatse tuvastamise statistilisi meetodeid, sh otsustuspuid ja tehismärguandeid. Siiski pole nende meetodidega seni saavutatud keskmine tuvastustäpsus veel piisav, et automaatset tuvastajat saaks lülitada dialoogsüsteemi koosseisu. Loodava dialoogsüsteemi üheks mooduliks on ajaväljendite tuvastaja, et arvuti saaks aru kasutaja päringutest, milles on väljendeid nagu „homme õhtul” või „aprilli esimesel kolmapäeval”. Selle mooduli aluseks oleva formaalse grammatika loomine kestab. Valminud on dialoogi juhtimise moodul (lõplik automaat), mis kasutaja lausungist tuvastatud dialoogiakti põhjal suunab päringu andmebaasile ja moodustab tulemuse teatamiseks eestikeelse lause. Dialoogiakt tuvastatakse automaatselt praegu märguandesõnade ja -fraaside põhjal. Sõnad taandatakse algvormile eesti keele morfoloogilise analüsaatori toel. Dialoogsüsteemi on seni proovitud siduda kahe andmebaasiga: üks sisaldab lennukite väljumisaegu Tallinna lennujaamast, teine aga Eesti teatrite mängukavasid. Lõppresultaadiks peaks olema eestikeelse telefoniteenuste süsteemi prototüüp, kuhu oleksid loimitud kõnemoodulid, mida arendatakse teiste projektide raames TTÜ ja EKI koostöös.



## KAS arvutitarkvara peab olema eestikeelne? KAS eestlane saab masinaga suhelda eesti keeles?

### VASTAB

Ilmar Tomusk,  
Keeleinspektsiooni peadirektor

Omakeelse tarkvara seisukohalt on eesti keele olukord inglise keele või mitmete väiksemate keeltega võrreldes tunduvalt viletsam, kuid viimastel aastatel on märgata paranemist. Eestikeelse tarkvara osakaal on suurenenud, tulised vaidlusedki tarkvara eesti keelde tõlkimise vajalikkuse ja võimalikkuse teemal on ära vaieldud. „Pisipehmed aknad“ Microsoft Windowsi asemel tundub olevat ilmselge liialdus, kuid lauset „downloadid zipi alla, right-klikiga valid menust directory, kuhu unzippid“ tõrgub eestikeelseks pidamast ka kõige paadunud ingliskeelsete programmide pooldaja. Küsimus, kas tõlkida, on juba ammu asendunud küsimusega, kuidas tõlkida.

Eestikeelse tarkvara arendamise vajadust on rõhutatud ka mitmes strateegias ja arengukavas, mis on keele-elu korraldamise aluseks.

Nii leiame näiteks Vabariigi Valitsuse 2007.–2011. aasta tegevusprogrammist järgmised eesmärgid:

10.3. Soodustab riiklikult eesti keele kasutamist infotehnoloogilistes rakendustes, toetades tarkvara kohandamist ja kasutusliideste tõlkimist.

10.21. Muudab emakeelse tarkvara kasutamise kohustuslikuks avalikule sektorile ja haridusasutustele. Tarkvara riigihangetel sätestatakse

üldnõue, et riigiasutuste ja kohalike omavalitsuste ostetaval avalikus kasutuses olevatel tarkvaradel peab olema eestikeelne kasutusliides.

12.6. Toetab emakeelse õpitarkvara arendamist ja kasutamist ning Interneti-põhiste emakeelsete õpikeskkondade arendamist, millega kaasneb õpetajate koolitus.

Sama keelt räägib ka „Eesti keele arendamise strateegia 2004–2010“ (EKAS), kus sedastatakse: „Strateegia elluviimine tagab, et eesti keelele oleksid väljendusvahendid mis tahes olulises eluvaldkonnas kunstidest ja filosoofiast kuni täppiseadustest ja infotehnoloogiast; eesti keel leiaks kiire rakenduse uuenduslike infotehnoloogilise põhinevate tarbetehnoloogiatena. Strateegia seab ülesandeks sätestada nõue eestikeelse arvutitarkvara kasutamise kohta üldhariduskooli kõikidel astmetel ning motiveerida riigihangetes eestikeelse tarkvara eelistamist, samuti arendada eesti keele infotehnoloogiline tugi tasemeni, mis võimaldab eesti keelel tänapäeva infotehnoloogilises keskkonnas edukalt toimida“.

Mitmesuguste infotehnoloogiliste rakenduste keelevalikuga puutuvad igapäeva elus kokku ka inimesed, kes üliharva arvuti taha satuvad või ei kasuta arvutit üldse. Nii valitsuse tegevusprogrammist kui ka EKASist leiame viite sellele, et lisaks tavapärastele arvutiprogrammidele (nagu operatsioonisüsteem Windows, kontoripakett Office või mitmed õpitarkvaraprogrammid) võiksid eestikeelset olla muudki infotehnoloogilised rakendused. Tavaliselt liigub mõte

kohe sellele, et inimene võiks anda oma kohvimasinale või muusikakeskusele eestikeelseid häälkäsklusi. See pole loomulikult mingi utopia, kuid on mitmeid märksa praktilisemaid valdkondi, kus eesti keel võiks kasutusel olla.

Mobiiltelefonimenüüd, mis veel hiljaaegu olid võõrkeelsed, on viimastel aastatel kõik eesti keelde tõlgitud. Inimesed on harjunud eesti keeles suhtlema ka pangautomaatides ja makseterminalides. Kui mõnikord ilmub eestikeelse teate asemel ingliskeelne *approved* või *declined*, siis suhtub enamik inimesi sellesse mõistvalt: ei ole veel jõutud eestikeelset programmijuppi kirjutada, kuid kui mõne kaupluseketi makseterminal ütleb selges vene keeles, ainult ladina tähtedega, et *udalite kartotshku*, siis see on küll valdkond, millesse keeleinspektsioonil palutakse selgust tuua.

Eestis tegutsevate rahvusvaheliste hotelliketide broneerimissüsteemid on küll paljumeelsete, kuid eesti keelt nende hulgas ei leia. Keeleinspektsioonil on tulnud palju tegelda Eestis registreeritud äriühingute veebilehtedega ja pärast meie ettekirjutusi on eestikeelne teave tavaliselt võõrkeelsele lisatud. Kuid mainitud broneerimissüsteemide puhul, mis on ju samuti tarbijateave ning kus keeleoskamatused põhjustatud vale valiku tegemine võib kliendile kalliks maksma minna, väidetakse, et neile küll eesti keelt lisada ei saa. Tekib küsimus: kui juba seitse keelt on (ja näiteks väike soome keel nende hulgas), miks siis ei saa lisada kaheksandat? Keeleinspektsioonil on tulnud

korduvalt abistada masinate ja mehhanismide operaatoreid, kellele tööandja on küll usaldanud keeruka töövahendi, kuid eestikeelset kasutusjuhendit pole andnud. Viimasel ajal on kasutusjuhend aga paljudel juhtudel kättesaadav vaid tootja veebilehel, kuid eestikeelset nende hulgas ei ole.

Oleme kokku puutunud ka juhtumitega, kus töömehe käsutusse on antud arvutustehnikal rajanev hinnaline seade, mis on töötajale sõna otseses mõttes töökaaslane ja suhtluspartner, kuid millel laiale keelvalikule vaatamata eesti keel puudub, soome ja rootsi keelt on aga võimalik valida. Ja seetõttu tulebki töötajal suhelda masinaga võõrkeelse. Loomulikult võib töötajad kehtestada töötajale võõrkeele oskuse nõude ning maailma praktika näitabki, et näiteks võõrkeeli valdav automehaanik on ükskeelsest loomingulisem ja tema töö tulemuslikum, kuid keeleseaduse kohaselt on igal töötajal õigus eestikeelsele tööalasele teabele. Pealegi käib võõrkeelsete juhendite kasutamise kaasa oht, et selles valdkonnas jääb eestikeelse oskussõnavara areng unarusse.

On palju muudki valdkondi, kus nii riik kui ka erasektor peaksid tegema jõupingutusi, et eestlane saaks end infotehnoloogilises keskkonnas koduselt tunda.

Eesti keel sobib arvutitarkvara arenduseks vahest paremini kui inglise keel, sest häälduse ja kirjaliku kokkuviimise selgitamine „tehisintellekti“ nõuab lisatööd, mis eesti keele puhul suures osas ära langeb.

### KÜSIB Aili Küntler

TLÜ senat kiitis 11.02.2008 Katariina kolledži asutamiskava üksmeelselt heaks. Eesti keelenõukogu arvamuse kohaselt läks algne plaan Eesti keele- ning hariduspoliitika põhisuundadega vastuollu.

KAS nüüd on Eesti sise- poliitilised huvid kaitstud?

### VASTAB

Birute Klaas,  
Eesti keelenõukogu esimees

Tõepoolest, esialgne Katariina kolledži asutamiskava oleks andnud selge märgi ühiskonnale, et Eesti kõrgkoolide seas on koole, kus ollakse valmis loobuma ühtsest keele- ja integratsioonipoliitikast, kuna pakutakse avalik-õiguslikus ülikoolis venekeelset kõrgharidust. Kui kõrgharituks võib saada vene keeles, puuduks muukeelsetel gümnaasiumiõpilasel ka motivatsioon eesti keele koolis korralikult selgeks õppida. Võrreldes selle kavaga, mille TLÜ esitas sügisel ja milles keelenõukogu suhtus väga kriitiliselt, on Katariina kolledži asutamise- ja tegutsemis-eesmärgid läbi teinud suure arengu.

Olen seda meelt, et muukeelsetest koolidest tulnud üliõpilased vajavad eestikeelses kõrghariduses kindlasti tuge, mis keeletuge kui ka juhendamist, kuidas paremini kohaneda Eesti akadeemilise kultuuriga. Katariina kolledž tahab oma üliõpilastele seda ka pakkuda kakskeelse õppe kaudu esimestel aastatel ja järkjärgulise ülemineku- ga eestikeelsele õppele kolmandal aastal. Väga kiiduväärne on see, et üliõpilastele kehtestatakse bakalaureusekraadi saamiseks ka kõrgharidusega spetsialistile vajalik eesti keele oskuse taseme nõue (C1) ja pakutakse keelekursusi, mis aitavad üliõpilastel selle taseme ka saavutada. Jääb üle ainult lootda, et sellistest põhimõtetest ka kinni peetakse.

Jääb siiski arusaamatuks, miks luuakse juurde uued õppekavad, kui TLÜ-l on piisaval hulgal eri valdkondades akrediteeritud õppekavasid, kuhu võiks ju lisada vene- või kakskeelse tugiõppe. Siiski ei soovi ma siin keelenõukogu esimehena võtta sõna *liber art* tüüpi õppekavade otstarbekuse kohta, kuigi selle taustal, et kõik kõrgkoolid kujundavad praegu oma õppekavasid ümber väljundipõhiseks, jäävad selletaoliste õppekavade eesmärgid üsnagi arusaamatuks.

Oma vastuses TLÜ senatile Katariina kolledži asjus tunnustas keelenõukogu ülikooli püüdusi luua tingimused, et venekeelsed noored saaksid kvaliteetset kõrgharidust koos kõrgtasemel riigikeele oskusega. Samas juhtis keelenõukogu senati tähelepanu segregatsiooniohule, mis võib kaasneda omaette kolledži loomisega, mille üliõpilaskonna sihtrühm on määratletud etnilisel alusel.

Endiselt on keelenõukogu seisukohal, et muukeelsetel noortel peavad olema eestikeelsete noortega samasugused võimalused ühiskonnas hakkama saada. Parima sellekohase tulemuse annab aga noorte võimalikult varajane koos õpetamine ja üksteisega kohanemine.

# „Teiste“ tundmine kaotab barjäärid Euroopas

Ján Figel,  
Euroopa Komisjoni volinik

Euroopa kultuuridevahelise dialoogi aastal kutsutakse kõiki Euroopa elanikke uurima meie rikkalikku kultuuripärandit ja õppima vastastiku kultuurilistest traditsioonidest. Mõistmine ja üksteise austamine peab olema Euroopa Liidus kesksel kohal, kui meie ühiskond tahab minna rahumeelselt ja väärikalt läbi 21. sajandi.

Soovimatus teisi tundma õppida on liidu edasise arengu peamine oht. Kultuuridevaheline dialoog on väga oluline teadmatusest ja eelarvamustest inimeste vahele tekkinud barjääride lammutamiseks, et saaksid ilmsiks ühised väärtused, mis ühendavad meid meie mitmekesisuses.

Erinevate väärtuste, harjumuste ja traditsioonidega rahvaste ja institutsioonide kokkuviimine on kahtlemata märkimisväärne kultuuriline ettevõtmine. Jean Monnet', Euroopa Liidu alusepanija memuaaride esimene lause on: „Me ei liida riike, me ühendame inimesi“. Inimeste ühendamine on raskem, kuid ka palju olulisem kui riikide liitude moodustamine!

Praegust maailma iseloomustab üha tugevnev konkurents globaliseeruva teadmispõhise majanduse kontekstis. Kuna esineb arusaama-

tusi teatud gruppide vahel, mis viitab mõnele sügavale juurdunud sotsiaalsetele, religioossetele ja kultuurilistele probleemidele, tuleb keskenduda eelkõige kultuurilistele valdkondadele: haridusele, kultuuridevahelisele dialoogile ja kodakondsusele.

Meie maailmajao multikultuursus on reaalsus ka paljudes ettevõtetes, see nõuab tähelepanu. Ettevõtetes, kus toimib kultuuridevaheline dialoog, saavad kasu avatud ja motiveerivast õhkkonnast, mis tõstab tootlikkust ja kasumlikkust, ent need ettevõtted, kes lubavad töötajate hulgas tekkida kultuurilistest erinevustest tingitud vastuoludel ja vaenulikkusel, kalduvad olema vähem innustavad ja efektiivsed.

Euroopa mitmekesisus ei kujuta endast ohtu, pigem on see vastumürk stagnatsioonile ja tuimusele. See on võimalus tunda huvi, õppida, võrrelda ja mõista. See on vaimundamatult meie kasutuses vahenditest üks väärtuslikumaid, kuna stimuleerib loovust ja innovatiivsust, mis on kaugemas perspektiivis kasvu ja töökohtade mootor.

Seejärel on kultuuridevaheline dialoog iga inimese, aga ka rahvusrühmade ja riikide ning kogu Euroopa Liidu pikaajaline prioriteet. Dialoog ei ole mitme nõrkuse, vaid küpsuse märk. Euroopa Komisjon pühendub eelseisvatel aastatel kultuuridevahelise dialoogi jätkusuutli-

Euroopa kultuuridevahelise dialoogi aasta 2008

Ürituste, uuringute, aasta saadikute jms loetelu Avatud Eesti Fondi kodulehel <http://www.oef.org.ee/et/dialog/> 20080114042032/

Euroopa Liidu 27 liikmesriigis elab ligi 500 miljonit inimest, 3%-l maakerast 7% maailma rahvastikust. Euroopa kultuuridevahelise dialoogi aastat (EKDA) kannab suhtumine, et Euroopa kultuuriline mitmekülsus on meie ainulaadne eelis, mida tuleb tundma õppida. Aasta loosung „Erinevad üheskoos“ tuletab meelde, et kultuure ei või üksteisest eraldada, need peavad saama koos elada ja hingata.

Rahvusvaheline koduleht <http://www.interculturaldialogue2008.eu/> Aasta saadikud aitavad inimestele selgitada kultuuridevahelise dialoogi aspekte, nende seas on näiteks brasiillasest romaaniikirjanik **Paulo Coelho** ja Prantsusmaal elav iraanlanna, filmirežisöör ja paljude koomiksitate autor **Margjane Satrapi**. Toogem siinkohal ära kahe Eestis tuntud inimese seisukohad.

Soome (Eesti) saadik politoloog **Ivi Anna Masso**: „On irooniline, et Euroopa asjatundjad õpetavad sama ideoloogiat eestlastele ajal, mil Lääne-Euroopas sellesse enam ei usuta. Ehkki Eesti probleemid on paljusid läänenaabrite omadest erinevad, on selge, et kogukondade ja keelerühmade eraldi elu ning paralleelseid väärtussüsteeme soosiv poliitika oleks ka Eesti oludes hukatuslik. Ilmselt on akadeemilistes ringkondades halvustatud ameerikalik poliitika „erinev, aga võrdne“, mis nõuab muulastelt kohanemist ja vastutust, kuid samas kohtleb neid võrdsetena võrdsete hulgas, nii rahvusriigi kui ka vähemuste seisukohast parem lahendus. Tegelikult on ka suuremad Euroopa riigid juba hakanud oma integratsioonipoliitikat selles suunas muutma“ (Mittmekultuurilise valupunktid. – Eesti Päevaleht 02.11.2007). Eesti aasta saadik kultuuriajaloolane **David Vseviov**: „Tuleb lõpetada mõtlamine „meie ja nemad“ ning omandada suhtumine „mina ja sina““.

ku strateegia arendamiseks. Eesmärgi, vastastikuse austuse õhkkonna valitsemise tagab iga Euroopa Liidus elava inimese aktiivne osavõtt Euroopa aasta üritustest. Julgustavad on hiljutise üleeuroopalise avaliku arvamuse uuringu tulemused, mille kohaselt arvab 83% eurooplastest, et kultuuridevaheline dialoog on kasulik, eriti nooremale põlvkonnale.

Tugineme sellele positiivsele hoiakule ja võtame liidus elavate

rahvaste vahel säilinud barjääride lammutamise 2008. aastal ette. Ärge eksisteerige koos. Elagem koos, üksteist austades ja mõistes!

**Euroopa kultuuridevahelise dialoogi aastat tähistatakse kohalike, rahvusvaheliste ja üleeuroopaliste üritustega, mida EL rahastab kümne miljoni euro suurusest eelarvest ning millele lisaks eraldab iga liikmesriik lisavahendeid.**

Vt [www.dialogue2008.eu](http://www.dialogue2008.eu).



# Millisest keelest saab aru arvuti?

Vastandus murdekeel-kirjakeel – viimase hulka on arvutatud EKSSis ka kõnekeel – on justkui oma aja ära elanud jaotus. Kas tänapäeval kõneldav on kirjakeel? Kuidas eesti keelt su meelest praegu jaotada annab?

**Tiit Hennoste, suulise keele uurija:** Kindlasti on vastandus murdekeel-kirjakeel oma aja ära elanud, ja juba viiskümmend aastat tagasi, kui murded pea kõikjal Eestis kasutusest kadusid. Mina jaotan eesti keelt nii, nagu sotsiolingvistikas tavaks, kolmelt eri aluselt.

Esimene jaotus on **kohamurded ja sotsiolektid ehk allkeeled inimeste sotsiaalsete omaduste alusel**. Murretest on eesti keelde jäänud vaid aktsendid (nagu saarte ö) ja Võru murre. Sotsiolektid, mille kesksed erijooned on häälduses ja sõnavaras, on aga meil üsna nõrgalt välja kujunenud.

Teine jaotus on **registrid ehk allkeeled kasutusolukorra alusel**. Siin on kesksed eristajad suuline kirjalik ja spontaanne-ettevalmistatud keelekasutus. Registrite erinevused on osalt sõnavaras, näiteks parasiitsõnad ehk partikkelid (noh) ja üneemid (ee) suulises spontaanses jutus. Palju olulisem erinevus on aga kõne ja kirja süntaksi vahel. Igasuguses keelekasutuses teeb kõneleja lauseid umber, aga kirjas ei jõua muutused lugejani, kõnes jõuavad, sest seal ei saa maha tõmmata. Miks inimene ei tee kohe korralikku lauset? Ei suuda. Ja kõneldes on see veel palju raskem, sest kõnelemine on vähemalt kümme korda kiirem kui kirjutamine. Kuigi, kõneldes kasutatakse palju lihtsamat lauseehitust kui kirjas. Nii on lauseparandused meie kõnemesse n-õ sisse ehitatud. Neid on ettevalmistatud jutus vähem kui spontaanses ja avalikus jutus vähem kui argises, aga nad on alati olemas.

Kolmas jaotus on **normikeel/ kirjakeel ja muud allkeeled**.

**Eesti Vabariigi 2006. aasta 7. septembri määrus nr 196 ütleb, et „Kirjakeele norm on määratud õigekirjutuse, häälduse, sõnavara ja grammatika osas Eesti Keele Instituudi uusima õigekeelsussõnaraamatuga, Emakeele Seltsi keeleteoimkonna normingute ja otsustega ning keeleteoimkonna heaks kiidetud ortograafiareegliti, normatiivse käsiraamatu ja grammatikaga”. Kas pole tänapäeval ka kirjutatu pigem kõnekeelele iseloomulike tunnustega?**

Mis on normikeel? See on keele- normingute süstematiseeritud kogu ja erineb sellisena kõigest muudest allkeeltest, milles norminguid ei ole. Kõneldud keeles on allkeeli, mille kaugus kirjakeelele normingutest erineb. Veelgi enam, mõnel juhul ei saagi neid reegleid järgida, näiteks süntaksis. Kirjakeele normingud lähtuvad tegelikult kahest eeldusest, millest käsiraamatutes ei kirjutata: tegemist peab olema kirjaliiku ja redigeeritud keelega.

Ja vastaks kohe ära kolmele karjatusele. Üks: mina küll niimoodi ei kõnele! Kõik kõnelevad. Soovitan kontrollimiseks lihtsalt linnistada oma jutt ja see täpselt üles kirjutada. Teine: vanasti küll nii ei kõnel-



Tiit Hennoste

dud. Kõneldi küll. Suulise keelekasutuse tingimused on samad juba aastatuhandeid ja seega on alati olnud vaja lauseid ümber teha ning parasiitsõnu kasutada. Kolmas: aga normingute järgimine on lihtne nt morfoloogias või häälduses. On. Ja muuseas, morfoloogia ongi eesti kõne väga norminguline (erinevalt nt soome keelest). Häälduse puhul aga pole minu teada eesti keeles mingit hääldusnormingut olemas. Kuid lõpuks tuleks aktsepteerida ka eesti keelemaailmas, et kõneldud keel ja kirjutatud keel on erinevad asjad. Ja et kõneldud keeles saab kirjakeele reegleid järgida põhimõtteliselt vaid osaliselt.

**Oled EKKT projekti „Eesti kõnekeele korpuse kogumine ja transliteerimine” projektijuht. Mille alusel kõnekeele muust eristate? Ja kuidas on allkeelte, slängiga? Sellegi rikkused on pigem ühiskeelde imibunud ja see pole enam nii selgelt piiritletav, kas pole?**

Me ei kasuta sõna „kõnekeel”, vaid kõneleme suulisest keelest. Selle all kogume argiseid kui avalikke, spontaansid kui ettevalmistatud tekste. Ainult mitte päheõpitud või paberilt maha loetud tekste.

Slängi all mõeldakse tavaliselt vaid sõnavaraerinevus. Ja minu arust jaguneb släng kaheks. Üks on mitteametlik erialaterminoloogia (nagu arvutiterminid), mis seostub eeskätt situatsiooni ja registriga ning on kasutusel kogu keelekogukonnas. Teine on sotsiaalset rühma eristav sõnavara (esta, mata jms) ja siin on tänapäeval piiridega suuri raskusi. Vanasti olid suured stabiilsed sotsiaalsed rühmad ja slängi sai piiritleta mingi pisikese rühma eripäraseks ja kiiresti muutuvaks sõnavarajooneks. Nüüd on ühiskond jagunenud paljudeks ja kiiresti muutuvateks pisirühmadeks, kõigil midagi eripärast keeles. Kas need on slängid ja sotsiolektidest pole enam mõtet kõnelda? Või vastupidi: pole enam mõtet kõnelda slängist?

**Oled väitnud „Keelekõrva” raadiovestluses, et eesti keeles sotsiaalne**

**dialekt puudub. Kuidas jääb keelekasutusega „tulääd kaa vää” jms? Huvitav on muidugi see, et „Pygmalion” Eliza keelekasutus ei paista tänapäeval märkivat enam ühiskonnas madalat positsiooni – pigem vastupidi. Tallinna ülikoolis ringi liikudes kohtab üliõpilaste seas nimetatud hääldusmalli üksjagu, ja ilmselt mitte üksnes garderoobis. Ja ülikooliharidusega inimesed on millegipärast ka eesti keele reeglite mitte tundmise üle pigem uhked kui häbenevad seda.**

Nojah. Ma ei taha hakata selle näite põhjal mitte mingeid piire vedama. Selleks, et vastata, peaks uurima. Kui see minu jaoks millegagi seostub, siis eeskätt kooli-ülikoolialaste naissoost inimeste hääldusega. Nii võiks ta ju üks võimalik sotsiolekti marker olla. Aga süntaktiliselt on tegemist tavalise suulise dialoogi lausega, mida kasutavad kõik. Hoiakute ja varandusega pole sel mu jaoks küll mingit pistmist.

**Harrastad ka ise õppejõu ja lektori-na kõnekeelsust ja hoidud liigsest kultiveeritusest. Kus on siis haritud kirjakeele suulise tarvitamise koht ja millise staatusega see praegu ühiskonnas seostub?**

Ma kahjuks ei tea, mis võiks olla haritud kirjakeele suuline tarvitamine. Mu jaoks on eri keeleosades normingute järgimise võimalused väga erinevad, nagu eespool ütlesin. Aga neis osades, kus on võimalik norminguid järgida, võib seda teha igal pool. Arusaamine ei vähene. Aga nii ei tehta, sest keelekasutuse valikutel on peale mõtete edastamise ka muid rolle, üks neist omaenese sotsiaalse oaduste või situatsiooni ühise osutamise ja samal ajal nende loomine keelekasutuse kaudu. Ja see nõuab vastavaid markereid keeles.

**Olen kuulnud arvamusi, et üha sagedamini tuleb abitu ja tüütu lobekeelsuse, laheda familiaaritamise ja labasuste pärast raamat või ajaleht käest, telegas ja raadio kinni või kuhugi mujale klõpsida. Kõnetuvastuse uuringuteks luuakse igasugu kõnekorpusi: spontaanse kõne, valdkondade dialoogide, uudiste lugemise, eri emakeelega inimeste aktsendikorpusi jne. Mis keeles arvuti minuga rääkima hakkab?**

Konverentsil on kaks keeleomandamise eri tahke käsitlevat töötuba. Töötuba „Esimise keele omandamine keeleteadlaste, psühholoogide ja kõneterapeutide vaatevinklist” eettekanded käsitlevad esimese keele omandamist, nii normikohast kui ka mahajäänud keelelist arengut, nii üks- kui ka kakskeelsete laste keeleomandamist. Tööpaja „Teise keele omandamine ja keeleoskuse mõõtmine” eesmärgiks on keskustel keeleoskuse mõtmise ja teise

Aastaid tagasi lähtuti sellest, et arvuti hakkab kõnelema normingulis keeles, kasutab lihtsaid kirjakeele lauseid. Praegu nii ei mõeld. Kõnelevad arvutid hakkavad nagu inimesedki jagunema rühmadesse. Ka arvutite rühmitamisel on kaks olulist alust.

Esimene on teksti mõistetavus neile, kes hakkavad arvutiga suhtlema. Arvuti, mis peab ütlema mulle kõikjal Eestis telefoninumbreid või teatrietenduste pealkirju, hakkab ilmselt kasutama kirjakeele morfoloogiat, lühikesi lauseid ja hääldust, mis on tüüpiline suuremale osale eestlastest. Aga kindlasti ei hakka ta kõnelema nagu paberilt maha lugedes, vaid vabas spontaanses kõnes.

Arvuti, mis vastab neile küsimustele Saaremaal, võib aga hakata kasutama kas või saarte murde aktsenti. See loob kohalikkude koloriiti, kuid on mõistetav ka teistele. Arvatavasti pole mõttekas panna Võrumaa arvutit kõnelema ainult võru keelt, sest see tekitaks kindlasti mittevõrulastele mõistmiskõnki. Nii võiks sealne arvuti olla kakskeelne, nagu pangautomaat.

**Mis siis saab, kui inimene ei taha arvutiga talle peale sunnitud keelepruugisuhelda? Kas variante pakutakse?**

Keelt ei pea valima nupuga, vaid arvuti saab ise aru, mis keeles tema poole pöördutakse. Minu koduse külmakapi või teleri arvuti võib suhelda minuga just minule sobivas allkeeles. Kust ta seda oskab? Suhtlusarvuti ei ole vabrikust tulles lõplikult valmis, vaid õppimisvõimeline programm, kes suheldes ühtlasi ka keelt juurde õpib ja oskab ennast temaga suhtleva inimese keelega kohandada. Ühesõnaga, käitub nagu iga inimese suheldes. Nii et: variandid pole lihtsalt võimalikud, variandid on põhilised. Arvuti, mis kõneleb ainult ühte allkeelt ja pole õppimisvõimeline, on kesine arvuti, nii nagu inimene, kes kõneleb ainult ühte allkeelt ja pole õppimisvõimeline.

**Küsinud Aili Kunstler**

keele omandamise uurimise koostöövõimaluste üle, et selgitada välja, milline on Euroopa keeleoskustasemete lingvistiline sisu. Konverentsi töökeeled on eesti ja inglise keel. Osavõtt on vaba kõigile huvilistele. Esinemissoov registreeritakse teeside põhjal. Teesid saata e-postiga aadressil [ery@eki.ee](mailto:ery@eki.ee) 15. märtsiks, vastused eettekanded sobivuse kohta saadetakse 31. märtsiks. Kogu konverentsiga seotud info on üleval ühingu kodulehel.

## Keeletegu 2007

**Teist korda valitavale aasta keeleteo auhinnale pretenderis mõeldud aasta 11 eesti keelele kasu ja avalikku tähelepanu toonud tegu.**

Jaanuari algusest veebruari lõpuni pakuti tunnustamiseks nii kirjanduse kui kunstiloomingu, keelekorralduse, -õppe, -ajaloo ja -uurimise, õigusloome, integratsioonipoliitika kui keelealaste noorteuurimise valdkonna tegusid. Iseseisvusaegsetest haridus- (ja teadus)ministritest koosnev žüri praeguse ministri Tõnis Lukase juhtimisel pidas mõeldud aasta keeleteoks noorte eestikeelseid laulu- ja tantsupidu „Ilmapuu lüel”. Eestlaste identiteet on küll emakeelekeskne, kuid eesti keele säilimise ühiks vajalikumaks tingimuseks on eestlike väärtuste austamine ja hoidmine. Pidu tunnustati kui tervikut: nii lauljaid kui tantsijaid, nii peajuhte kui korraldajaid, nii kuulajaid kui vaatajaid. Soodustamiseks eestikeelse laulu- ja tantsupeotraditsiooni jätkumist, otsustati auhinna 35 000 krooni anda Eesti Laulu- ja Tantsupeo Sihtasutusele 2011. aasta XI noorte laulu- ja tantsupeo eestikeelse originaalloomingu tellimiseks.

Rahvaauhinna vääriks tunnustasid Internetihääletusest osavõtjad Sulev Iva Võru murret käsitleva doktoritöö „Võru kirjakeele sõnamuutmissüsteem”.

Emakeelepäeva eel tehti Tartus Athena keskkuses keeletegude valimisest kokkuvõtte, austati parimaid ning vaadati Mart Kivastiku näidendit „Sõdur”.

## 20 aastat Wiedemanni keeleauhinda

Tänavu täitub 20 aastat Ferdinand Johann Wiedemanni keeleauhinna avatamist. Sel puhul toimub 16. ja 17. aprillil Väike-Maarjasa rahvusvaheline konverents „Emakeelne Eesti, emakeelne Euroopa”.

Olnut analüüsiv konverents on suunatud tulevikku nii eesti keele kui teiste esindatud maade riigikeele probleemide ja perspektiive silmas pidades. Huna Rätsep räägib akadeemik Wiedemanni, Eevi Ross keeleauhinna saamis- ja arenguloost, Larissa Mukovskaja (Peterburi) Wiedemanni tegevusest Venemaal, Mati Ereli käsitte eesti keele kuuluvust, Urmas Sutrop otsib eesti keele algust, Ilmar Tomusk käsitleb Eesti ja Tõnu Tender Euroopa Liidu keelepoliitilisi suundumusid, Jüri Valge tutvustab keeleteaduse riigiprogrammi. Välismaa kolleegid Raimo Raag (Rootsi), Pirkko Nuolijärvi (Soome), Janina Kurste ja Erika Krautmane (Läti), Eberhard Winkler (Saksamaa) ja János Pusztay (Ungari) räägivad (eesti keeles) oma maa riigikeeltest, nende arengust ja võimalikust tulevikust. Õpilasettekanded Pille-Riini Kriisalt ja Tanel Kuislerilt (Tartu Tamme gümnaasium) ning Alli Alasilt ja Kertu Rebaselt (Väike-Maarja gümnaasium) käsitlevad vastavalt tänapäevast koostööd hõimurahvastega ning Virumaa Teatja pealkirja. Tava kohaselt istutab keeleauhinna laureaat, tänavu Mati Ereli, keeletamme ning austatakse 20 aasta jooksul keeleauhinna pälvinuid. Väike-Maarja gümnaasium ja kohalik muuseum koostavad näituse „Wiedemanni keeleahind 20”.

## Emakeele Selts

(<http://www.emakeeleselts.ee>)

**Emakeele Seltsi õpilaste sõnakogumisvõistus algas 01.02.2008 ja kestab 15.05.2008.**

Esimene kuu jooksul märkisid õpilased üles oma sõprade ja perekonna, koolikaaslaste ja veebiseltisest keelt ning täiendasid võistlus-sõnastikku muu hulgas järgnevalt: *nimps* 'nimi', *totupüü* 'armsalt rumal', *rull* 'raha', *iks idee* 'imelik mõte', *pyhh* 'pühendus', *külliki* 'meeldivali naivne tüdruk', *jaa vä?* 'kas tõesti' ja *rä* 'otsese tähendusega parasiitsõna'. Veebruaris registreerus sõnastiku täiendamiseks 47 inimest, kellest usinamad jõudsid sisestada mitukümmend, tagasihooldikemad ühe sõna. Palju kirjeid lisati ka anonümselt, mistõttu ei saa neid kahjuks auhindamisel arvestada. Auhindamisel arvestab komisjon nii sissekanne- te hulka kui ka kvaliteeti. Viimase määrab kirje põhjalikkus: kas õpilane piirdub ainult sõna ja selle tähendusega või lisab muudki. Näiteks on sõna *puut* 'auto pakiruum' lisaja märkinud üles näitelause („Ole hea, pane need pakid puuti!”) ja tutvustanud sõna päritolu („Austraalias elanud vanaisa sõna. Kasutatavad ka meie tuttavad.”). Võistlus käib kolmes vanuserühmas: 1.–6. klass, 7.–9. klass, 10.–12. klass, lisanduvad eripreemiad. Et sõnad läheksid võistlusel arvesse, tuleks need kirja panna enne 16. maid, ehkki veebisõnastik jääb avatuks hiljemgi. **Sõnastiku aadressil** <http://www.emakeeleselts.ee/sv/> on jätkuvalt väga oodatud kõik sõnad ja väljendid, mis pole seni sõnaraamatutesse jõudnud või millel on seal teine tähendus.

**27.03.2008** kell 14 seltsi aastakoosolek Tallinas EKl saalis (Roosikrantsi 6, III k).

**Väljaandja** SA Kultuurileht, Sird  
**Toimetuse aadress** Väike-Karja 12, 10140 Tallinn  
**Telefon** 640 5775  
**Toimetaja** Aili Kunstler, [aali@sirp.ee](mailto:aali@sirp.ee)

**Eesti Rakenduslingvistika Ühing**  
[www.rakenduslingvistika.ee](http://www.rakenduslingvistika.ee)

**24.-25.04.2008 VII rakenduslingvistika konverents „Keeleteaduse aineestik ja andmekogude kasutamine”.**

Keskendutakse korpusele kui andmekogule ning selle kasutusvõimalustele, rakendustele ning uurimise meetoditele.

**Arutluse alla tuleb:** mis kõik on ja mis võib olla keeleteadusliku uurimise materjaliks (korpused, andmebaasid, küsitlused, kartoteegid, sõnaraamatud, päevikumärkmed, Interneti-materjalid jne); kuidas materjalikogusid hallatakse ja kasutussobivaks tehakse; materjali uurimise meetodid (kvalitatiivsed ja kvantitatiivsed, juhtumi- ja massiuuringud jne); materjali representatiivsuse ja piisavuse problemaatika; uurimisvaldkondade materjalide ja meetodika eripära.