

KEELETEHNOLOOGIA ehk ARVUTILINGVISTIKA

MIS ON KEELETEHNOLOOGIA EHK ARVUTILINGVISTIKA?

Hübriidala, mis sisaldab keeleteaduse, informaatika, statistika ja matemaatika elemente. Eesmärgiks keelenähtuste (morfoloogia, süntaksi, semantika, pragmaatika) modelleerimine. Termineid *keeletehnoloogia* ja *arvutilingvistika* kasutatakse tänapäeval sünonüümidenä.

Varem on *arvutilingvistikat* defineeritud ka kui arvutiteaduse ja informaatika hübriidala, mille eesmärgiks on töötada välja loomuliku keele automaattöötluks sobivad keele kirjeldus-, analüüsi- ja sünteesivahendid ning need arvutil realiseerida, ja *keeletehnoloogiat* peetud selle rakenduslikuks või tehnoloogiliseks haruks.

Loe ka artiklit „Mis on keeletehnoloogia“ (<https://www.ajakiri.ut.ee/artikkel/3066>).

KESKUSED EESTIS

Tartu Ülikoolis tegeletakse peamiselt kirjaliku teksti automaattöötluks, sh masintõlkega, dialoogsüsteemide modelleerimisega, morfoloogilise ja süntaktilise automaatanalüüsiga, aga ka närvivõrkudel põhineva sünteeskõnega.

Tallinna Tehnikaülikoolis arendatakse kõnetuvastust ning kogutakse kõnekorpust.

Eesti Keele Instituudis arendatakse kõnesünteesi ja sõnastikusüsteeme, samuti keelekorpuse analüüsimise vahendeid.

Eesti Keeleressursside Keskus EKRR (<https://keele-ressursid.ee>) koondab ja teeb kättesaadavaks eesti keele digitaalsed ressursid ja tehnoloogiad.



Arvutilingvistika „maaletoojad“, TÜ emeriitprofessorid Haldur Oim ja Mare Koidu.



TÜ ATI keeletehnoloogia õppetooli juhataja Mark Fišel.

TTÜ kõnetehnoloogia vanemteadur Tanel Alumäe.

EKI arvutilingvist Jelena Kallas.

MIDA ON KEELETEHNOLOOGID-ARVUTILINGVISTID EESTI KEELE HEAKS ÄRA TEINUD?

Laias laastus jaguneb keeletehnoloogide loodu kaheks: lõppkasutajale suunatud rakendused ja loomuliku keele analüüsi moodulid. Viimased on esimeste ehituskivideks ja keeleteadlasele abivahenditeks.

LÕPPKASUTAJALE SUUNATUD RAKENDUSTE NÄITEID:

Masintõlge <https://neurotolge.ee/>. Närvivõrkudel põhinev masintõlge tuvastab keele ning tõlgib teksti eesti, inglise või läti keelde, tulevikus ka vene keelde.

Kõnetuvastus <http://bark.phon.ioc.ee>. Kõnetuvastus on tehnoloogia, mille abil inimkõne muudetakse kirjalikuks tekstiks. Kõnetuvastust saab kasutada nii reaaliajaliselt kui ka veebipõhiselt, loodud on ka eestikeelsel kõnel põhinevad rakendused Androidile. Automaatselt transkribeeritud raadiosaateid saab sirvida <http://bark.phon.ioc.ee/mus30>.

Kõnesüntees teeb tekstist kõne, nt audioraamatute genereerija Vox Populi (<https://heliraamat.eki.ee/vox>), Elisa Raamat (<https://www.elisa.ee/et/raama>), subtiitrite helindaja (ERR subtiitrite asemel sünteeskõne), Eesti Ekspressi ja Postimehe taskuhääling (veebilehel uudiste kuulamine).

Kõne masintõlge <https://live.neurotolge.ee/>. Tegemist on alles arendatava teenusega, milles on kokku pandud kaks tuntuimat keeletehnoloogia rakendust: kõnetuvastus ja masintõlge.

Õigekirjakorrektor e-speller http://www.filosoft.ee/html_speller_et/
EKI **sõnastikusüsteem** (Ekilex, varem EELex eelex.eki.ee ja Termeki term.eki.ee)
EKI **sõnastikuportaal** Sõnaveeb.ee

EESTI KEELE TÖÖTLEMISE MOODULITE NÄITEID:

- eesti keele morfoloogia-analüsaator, vabavara: <https://github.com/Filosoft/vabamorf>
- eesti keele süntaktilise analüüsi vahendid, vabavara: <https://github.com/EstSyntax>
- eesti keele töötlemise Pythoni teekide kogu EstNLTK, vabavara: <https://estnltk.github.io/>

VEEBISÕNASTIKKE KOONDAVAD

Keeleveeb www.keeleveeb.ee
EKI sõnastikuportaal Sõnaveeb.ee

KUS SAAB KEELETEHNOLOOGIAT/ARVUTILINGVISTIKAT ÕPPIDA?

Tartu Ülikooli eesti ja üldkeeleteaduse instituudis õpetatakse arvutilingvistika erialamoodulit bakalaureuseõppes ja arvutilingvistika eriala magistriõppes.

Tartu Ülikooli arvutiteaduse instituudi koosseisus on keeletehnoloogia õppetool ning õpetatakse keeletehnoloogia aineid nii bakalaureuse- kui ka magistritasemel.

Tallinna Tehnikaülikoolis-s saab õppida kõnetehnoloogia ning tekstikaave aineid informaatika magistrantuuris.

MIDA ON KEELETEHNOLOOGIAL PAKKUDA LINGVISTIDELE JA KEELEHUUVILISTELE?

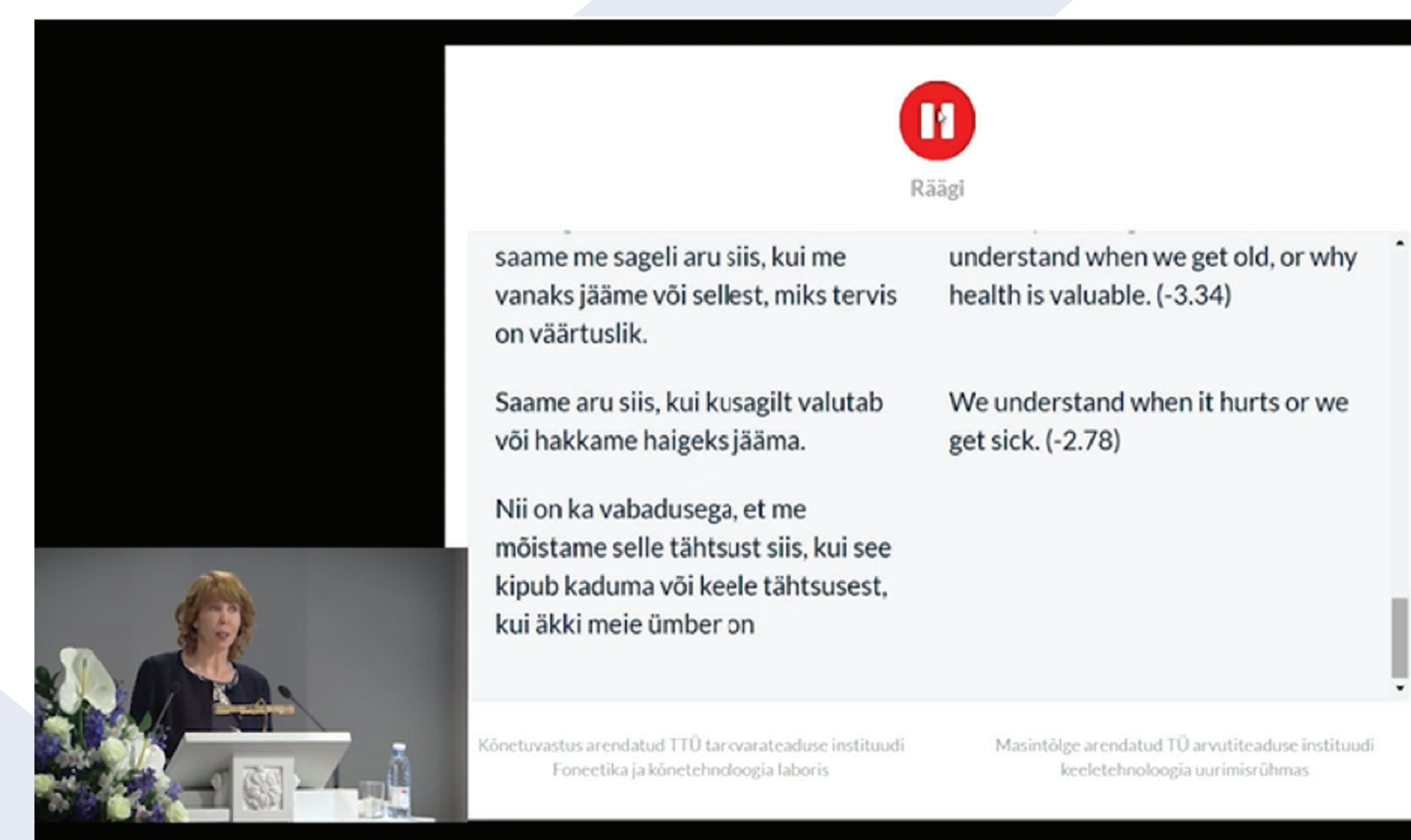
Keelekorpusdest saab ammutada tegeliku keelekasutuse näiteid kasutajaliidestest abil.

Suurtele eesti keele korpusetele saab päringuid esitada nt Keeleveebis www.keeleveeb.ee, KORPi kaudu <https://korp.keeleressursid.ee> või Sketch Engine'i kaudu <https://www.sketchengine.eu/estonian-reference-corpus/>.

Eesti keeletehnoloogia ajaloost loe lähemalt Mare Koidu artiklist „Keeletehnoloogia Eestis. Teekond tänapäevani.“ – Akadeemia 2018, nr 9.

Kirjutusmasin — tõlf.

Londonis äratas vähelepanu ilufugune kirjutusmasin, mis lahest ojaft loos seijab ja nii on ehitatud, et ühel masinal kirjutamise korral, teine kirjutuse automaatselt joovita-
masse keelde tõlgib. Uut aparaati saab tarvi-
tada ka telegraafijaamades vastumõetavate telegrammide tõlkimiseks. Seibujele kuulutataf-
se juurt tuleviffu.



Keeletehnoloogia on vanem kui me arvata oskame: Esmapäev: piltidega nädalleht, nr. 46, 19 november 1923 Allikas: dea.digar.ee

Kõne masintõlge võimaldab tõlkida kõneleja juttu reaajas.

Eesti keeletehnoloogiat toetab riiklik programm. Alates 2018. aastast käivitus järjekorras kolmas programm ning selle tulemused on avalikult tasuta kasutatavad. Haridus- ja Teadusministeeriumi keeletehnoloogia teadus- ja arendustegevuse programm „Eesti keeletehnoloogia 2018–2027“ toetab keeletehnoloogiaalast teadus- ja arendustegevust, et luua uusi eesti keele keeletehnoloogilisi rakendusi, tõsta olemasolevate rakenduste kvaliteeti ning võtta neid kasutusele võimalikult paljudes valdkondades nii era-, avalikus kui ka kolmandas sektoris. (www.keeletehnoloogia.ee)

EKI sõnaraamatute koduleht <http://portaal.eki.ee/sonaraamatud>
Keelevara www.keelevara.ee