

Keeletehnoloogia kujunemisest & keeletehnoloogilistest ressurssidest

15.11.2019

Pille Eslon

Mõisteid

- Arvutilingvistika (*computational linguistics*)
- Keeletehnoloogia (*language technology*)
 - inimkeeletehnoloogia (*human language technology*)
 - loomuliku keele töötlus (*natural language processing – NLP*)
- Keeletehnoloogia = rakenduslik arvutilingvistika (*applied computational linguistics*)

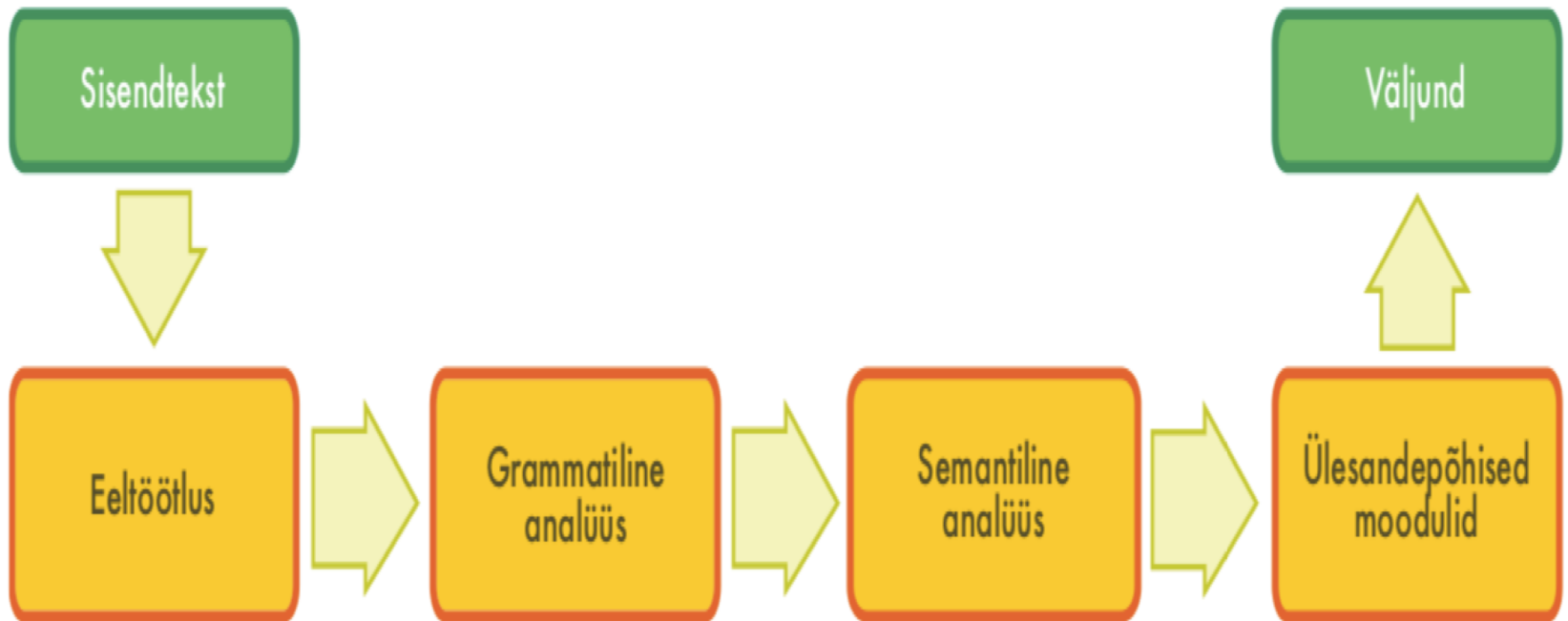
Tegevusi

- Keelealaste teadmiste rakendamine paremate arvutisüsteemide loomiseks ↔ arvutisüsteemide rakendamine inimsuhtluses, teadus- ja arendustöös
 - Tegevused
 - võimaldada inimeste masinsuhtlust erinevatel kommunikatiivsetel vajadustel (Cole jt 1998)
 - tõhustada inimese ja arvuti vahelist suhtlust
 - luua infootsingu automatiseeritud süsteeme, laiendada teksti ja keele automaatse analüüsi ja sünteesi võimalusi
 - arendada ja rakendada uusi tõhusamaid tehnoloogiaid inimkeele automaatseks töötlemiseks, ammutada uusi teadmisi keele olemuse, mõistmise ja produtseerimise kohta (Kaalep 1999)

Tulemusi

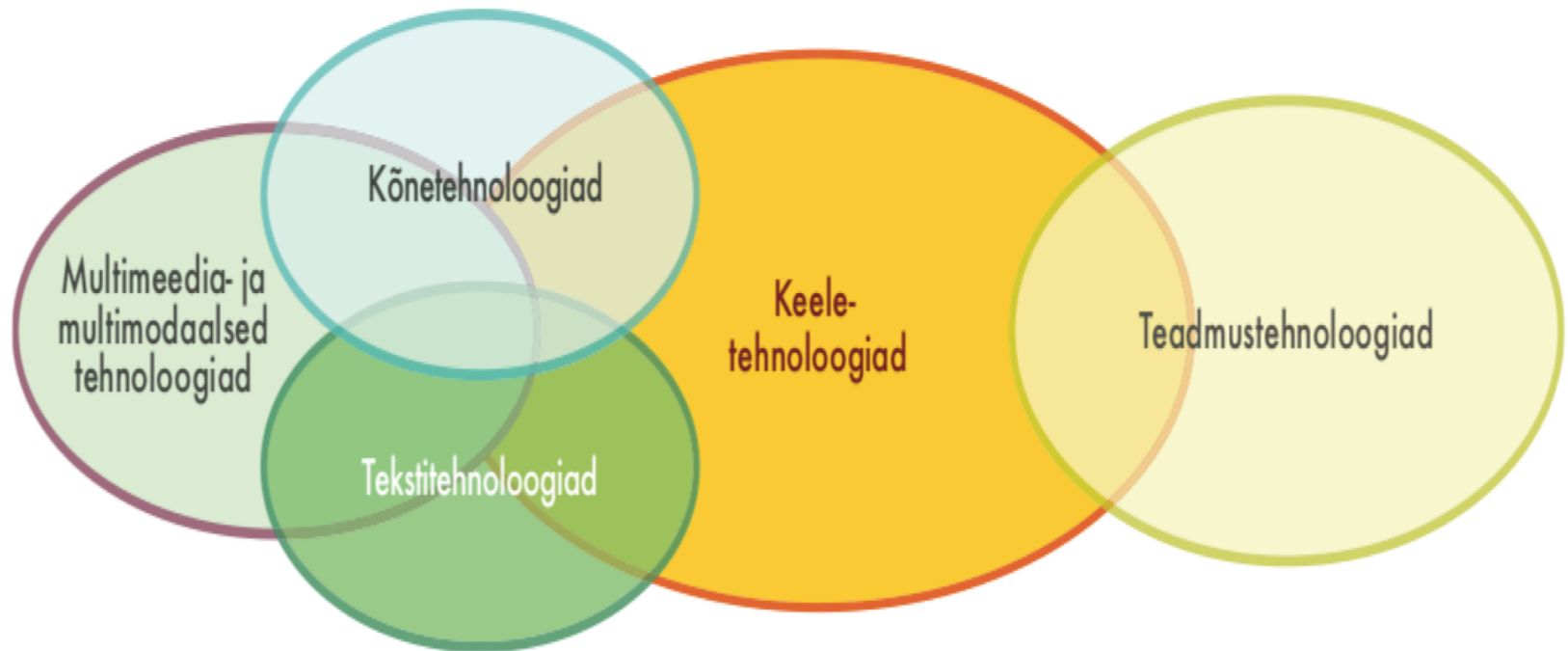
- **Arvutikasutus lihtsustub:** võimalus suhelda arvutiga tehnilistest oskustest olenemata ja erivajaduse korral
- **Infotöötlus muutub kordades efektiivsemaks:** tohutud andmehulgad, vormistatud enamjaolt keelelisel kujul
- **Keelebarjäärid kaovad:** võimalus sujuvaks äri-, poliitika- jm suhtluseks ka siis, kui ei räägita samas keeles
- **T&A-tegevuse tulemusel rakenduvad tõenduspõhiselt välja töötatud süsteemid ja mudelid:** nt keele õpetamise ja hindamise automaatsed rakendused, e-sõnastikud, tesaurused, iga keeleoskustaseme aktiivse sõnavara ja grammatika omandamisele suunatud õppe korraldamine, uudsete meetodite kasutamine keeleõppe tõhustamiseks jne

Keeletöötlaste üldine skeem



2: Tüüpiline keeletöötlaste arhitektuur

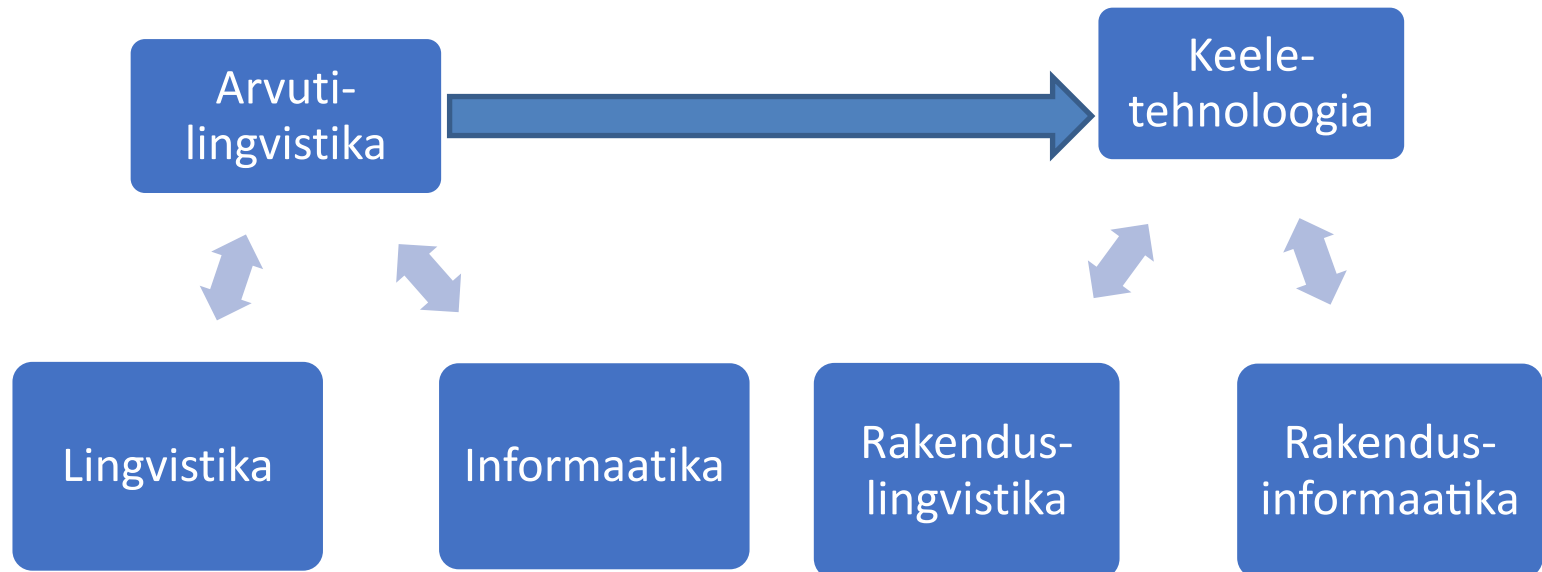
Keeletehnoloogia infotehnoloogia kontekstis



1: Keeletehnoloogia infotehnoloogia kontekstis

Keeletehnoloogia piirialad

- Keeletehnoloogia = rakenduslik arvutilingvistika



Lähtekohti 1:

XX sajandi kvantitatiivne lingvistika

- Morris Swadesh (1951; 1952; 1955): Põhisõnavara = universaalsed üldmõisted, mille koosseis on diakrooniliselt muutumatu; sõnade kokkulangevus erinevates keeltes 81%–86%
- Valter Tauli (1952-1953): esimene eesti keele sõnavarastatistiline töö vabatahtlike abiga, käsitsi (A. Mälk, Tee kaevule I–II. Lund: Eesti Kirjanike Kooperatiiv, 1952–1953).
- Juhan Tuldava *alias* Artur Johan Haman (1987; 1995) kvantitatiivne sõnavarauurimus
- Aet Lees (2004): eesti- ja soomekeelsete originaal- ja tõlketekstide võrdlev-statistiline analüüs objektikäänete varieerumise uuringus

Lähtekohti 2:

kombinatoorne statistiline analüüs

- Keelestruktuuride kombinatoorne analüüs → tulemuste kasutamine masintõlkes ja aluskeele rekonstrueerimisel (Nikolai Andrejev 1965; 1967)
- Eesti morfoloogia statistilis-kombinatoorne mudel (Helje Holm 1965)
- Distributiivse analüüsi sidumine statistilise analüüsiga (Anatoli Šajkevitš 1976)
 - algoritmiliste protseduuride jada, mis määrab elementide jaotumise tekstis (nt tähtede järgnevused sõnas, st kahe tühiku vahel, sõnade järgnevused kahe kirjavahemärgi vahel – kui sageli samalaadne järgnevus tekstis kordub)

Lähtekohti 3: empirism arvutilingvistikas

- Alusepanija: vene matemaatik **Andrei Markov** (1856-1922)
 - Markovi ahelad, st statistiline mudel, mis koosneb sõlmedest ja neid ühendavatest suunatud kaartest, kus iga kaar on varustatud tema läbimise tõenäosusega
 - 1913 ennustas, kas Puškini „Jevgeni Onegini“ tekstis järgnev täht on vokaal või konsonant, arvestades ühte või kahte eelnevat tähte, st kasutades andmekaeve bigramm- ja trigramm-mudeleid
- 1947 – ameerika matemaatik Warren Waveri küsimus küberneetika rajajale Norbert Wienerile
 - kas arvuteid saaks kasutada tekstide tõlkimisel ühest keelest teise
 - masintõlge = salakirja „lahti muukimine“
- Aktiivne tegevus masintõlke rindel

- 1954 – firma IBM arvuti tõlkis vene keelest 200-sõna teksti inglise keelde
 - Sõna-sõnalised tõlkevasted (paralleel: reaalne tõlge)
- 1949 – Claude Shannon, informatsiooniteooria rajaja genereeris inglise keele n-gramm-sõnamudelid
 - Statistiline keelemudel, kus sõna esinemise tõenäosus sõltub n-1 talle eelnevast sõnast
- 1957 – Frank Rosenblatti lihtne närvivõrgu matemaatiline mudel (ingl perceptron ‘tajur’)
 - Esitatakse komponentide võrgustikuna
 - Avanes keeleandmete paralleeltöötlus

- 1960-ndad – esimesed arvutikorpused
 - Standardised miljon sõnet sisaldavad Browni, London-Lundi, Lancaster-Oslo-Bergeni korpus
- Statistilisi keelemudeleid hakati treenima elektroonses vormis ette antud tekstidel
- Markovi ahela arendus - Markovi peitmudel
 - Vaadeldavate olekute põhjal määratakse peitolekud (nt sõnade jäljendi alusel sõnaliigid)
- 1967 – Viterbi algoritm, et määrata sõnajärjend kõige tõenäolisemate sõnaliikide järjendina
 - Loodi tõenäosuslikud grammatikad, kus iga reegel on varustatud tema rakendamise tõenäosusega

Lähtekohti 4: ratsionalism arvutilingvistikas

- Chomsky 1956 – generatiivne grammatika
 - Andis keele automaattöötusele formalismid ja mudelid
 - $G = (T, N, P, S)$
 - **T** on terminaalide (nt sõnavormid), **N** mitteterminaalide (nt grammatilised kategooriad) ja
 - **P** ümberkirjutusreeglite hulk kujul $x \rightarrow y$
 - **S** (Sentence 'lause') on üks erilises rollis mitteterminaal
 - ... *terminaalsete järjendite hulka, mida saab sümbolist S lähtudes ja ümberkirjutusreegleid rakendades tuletada, nimetatakse selle grammatikaga genereeritavaks keeleks* (Mare Koit 2006)

Ümberkirjutusreeglite näide

Mari kirjutab

Jüri kirjutab

Väike Mari loeb hästi

Väike tubli Jüri kirjutab hästi meelsasti

G = (T, N, P, S) ümberkirjutusreeglid

T = {väike: Ad], tubli:Adj, Mari:N, Jüri:N, kirjutab:V.
laulab. V. hästi:Adv, meelsasti:Adv}

eesti keele sõna(vormi)d **N** = {S, VP, NP, N, V, Adj,
Adv}

grammatiliste kategooriate nimetused **P** = {S -> NP
VP, NP Adj NP, NP -> N, VP VP Adv, VP -> V}

Ümberkirjutusreeglite kuju

Chomsky hierarhias

Grammatika / keele tüüp	Ümberkirjutusreeglite kuju
0 (piiranguteta)	$x \rightarrow y,$ kus $x \in (\mathbf{T} \cup \mathbf{N})^+$ $y \in (\mathbf{T} \cup \mathbf{N})^*$
1 (kontekstisidus, ingl <i>context sensitive</i>)	$zAw \rightarrow ztw,$ kus $A \in \mathbf{N},$ $t \in (\mathbf{T} \cup \mathbf{N})^+$ $z, w \in (\mathbf{T} \cup \mathbf{N})^*$
2 (kontekstivaba, <i>context free</i>)	$A \rightarrow y,$ kus $A \in \mathbf{N},$ $y \in (\mathbf{T} \cup \mathbf{N})^*$
3 (regulaarne, <i>regular</i>)	$A \rightarrow a, A \rightarrow Ba$ (vasaklineaarne), kus $A, B \in \mathbf{N},$ $a \in \mathbf{T}$

Chomsky grammatikate ja keelte hierarhia

CV: Haldur Õim | ETIS - Eesti keeleteadusliku mõtte | teoreetiline_keeleteadus_2_ocr.pdf

Not secure | dspace.ut.ee/bitstream/handle/10062/33926/teoreetiline_keeleteadus_2_ocr.pdf

Paused

Joonis 2. Chomsky grammatikate ja keelte hierarhia

Ka Tartu ülikoolis alustas 1960. aastate keskpaigas keeleteadlase Huno

Start | PDF | File Explorer | Word | Edge | Chrome | PowerPoint | ET | 12:21 13.02.2019

- 1960-ndate keskel – Huno Rätsepa seminar, hilisema nimetusega Generatiivse Grammatika Grupp (GGG)
 - Kogunesid keele arvutitööstlusest huvitatud lingvistika- ja matemaatikaüliõpilased ning –õppejõud
 - Plaan luua eesti keele generatiivne grammatika
- Töötati välja ja võeti kasutusele mitmesuguseid reeglipõhiseid formalisme
 - transformatsioonigrammatika (Chomsky 1965)
 - käändegrammatika (Fillmore 1967)
 - laiendatud üleminekuvõrk (ATN – Augmented Transition Network, Woods 1970)
 - määravate osalauseite grammatika (DCG – Definite Clause Grammar, Colmerauer 1978)

- 1980-ndatest peale
 - Unifikatsioonigrammatikaid (FUG – Functional Unification Grammar, M. Kay 1979), süntaktiliste kategooriate põhistele reeglitele lisandusid keeruliste tunnusestruktuuridega reeglid
 - Leksikaal-funktsionaalne grammatika (LFG – Lexical Functional Grammar, J. Bresnan & R. Kaplan 1982)
 - Üldine fraasistruktuurigrammatika (GPSG – Generalized Phrase Structure Grammar, G. Gazdar 1985)
 - Peajuhitav fraasistruktuurigrammatika (HPSG – Head driven Phrase Structure Grammar, C. Pollard & I. Sag 1987)
 - nt inglise, saksa, vene, tšehhi ja bulgaaria keele analüsaatorite ning generaatorite loomisel

- Lõplik (oleku)automaat ~ morfoloogiamudel
 - Alus: Markovi matemaatiline mudel, mis koosneb sõlmedest (olekutest) ja neid ühendavatest suunatud kaartest (üleminekutest)
 - Lõplikku automaati võib esitada olekudiagrammina
 - Lause (sõna) analüüs tähendab diagrammi läbimist algolekust lõppolekuni
 - Liikumine toimub mööda olekuid ühendavaid kaari (märgendatud lauses esinevate sõnavormidega või vastavuses sõnas esinevate morfeemidega)

- Kaheastmeline morfoloogiamudel:
 - Süvatasandi mudel
 - Sõnastik – sisaldab morfeemide süvakujusid
 - Pindtasandi ehk pindmine mudel
 - Reeglite ja sõnastike vaheliste viitade abil genereeritakse kõik realselt kasutatavad sõnavormid
 - Eesti keele morfoloogiline analüüs ja süntees (Uibo 2006)

Lähtekohti 5:

uus empirismi periood arvutilingvistikas

- Eeldused:
 - Horning 1969: tõenäoliselt ligikaudu korrektne õppimine positiivsete näidete varal (ingl probably approximately correct)
 - Tõenäosuslikus grammatikas on iga reegel varustatud tema rakendamise tõenäosusega, mis on arvutatud korpuse põhjal
 - Olekuautomaadile võib lisada ühest olekust teise üleminekute tõenäosused – saame Markovi mudeli
 - 1976: tõenäosuslike mudelite kasutamine kõnetuvastuses
 - 1985: kõne sünteesimisel määrati pärisnimede päritolu – kasutati statistilisi trigramm-mudeleid (K. Churchil)
 - Statistiliste keeletöötlusmeetodite kasvav populaarsus
 - 1990: statistiliste keeletöötlusmeetodite võidukäik tänaseni, reeglipõhised meetodid jäid kõrvale

- Põhjused:
 - Statistilised mudelid töötavad hästi mittetäieliku teadmuse puhul
 - neid on edukalt kasutatud kõnetuvastuses ja kõnesünteesis
 - Konneksionistlikud keelemudelid (nt tehishärvivõrgud) töötavad samuti hästi
 - koosneb suurest hulgast omavahel seotud lihtsatest mittelineaarsetest komponentidest, mis töötavad paralleelselt (mitte järjestikku nagu lõplik automaat)
 - kasutatakse samuti nagu statistilise mudeli puhul treeningandmeid, mille põhjal mudel õpib, kuid keeletöötlussüsteemi arhitektuur on keerukam ja tänu sõlmede dubleerimisele on tehishärvivõrk töökindlam (Koit 2006)

Eesti keele automaattöötuse meetodeid

CV: Haldur Öim | ETIS - Eesti keeleteadusliku mõtte | teoreetiline_keeleteadus_2_ocr.pdf

Not secure | dspace.ut.ee/bitstream/handle/10062/33926/teoreetiline_keeleteadus_2_ocr.pdf

Tabel 2. Eesti keele automaattöötusel kasutatud meetodid

<i>Keeletöötuse tase</i>	<i>Kus tegeldakse</i>	<i>Meetodid</i>
SUULINE KEEL kõne => tekst, tekst => kõne	TTÜ Kübi + EKI	andmepõhised
KIRJALIK KEEL (tekst)		
<i>Morfoloogia</i>	Filosoft + TÜ EKI	reeglipõhine + andmepõhine reeglipõhine (avatud morfoloogiamudel)
<i>Süntaks</i>	TÜ TÜ	reeglipõhine (lõplikud automaad reeglipõhine (kitsenduste grammatika)
<i>Semantika</i> sõnatähenduste	TÜ	andmepõhine + reeglipõhine

Start | PDF | File Explorer | Word | Edge | Chrome | Photoshop | ET | 12:31 13.02.2019

tekst => kõne

KIRJALIK KEEL (tekst)

Morfoloogia

Filosoof + TÜ
EKI

reeglipõhine + andmepõhine
reeglipõhine

(avatud morfoloogiamudel)

Süntaks

TÜ
TÜ

reeglipõhine (lõplikud automaad
reeglipõhine
(kitsenduste grammatika)

Semantika

sõnatähenduste
ühestamine

TÜ

andmepõhine + reeglipõhine

Pragmaatika

dialoogiaktide
tuvastamine

TÜ

andmepõhised
(tehisnärvivõrgud, otsustuspuud)

- **Statistilised** ehk **andmejuhitud** süsteemid
 - aluseks suured tekstikogud, millest masintõlkealgoritm õpib mustreid eristama
 - + masin õpib kiiresti
 - – kvaliteet võib oluliselt varieeruda
- **Reeglipõhised** süsteemid
 - grammatilised analüüsid kodeeritakse reegliteks
 - + täpsem kontroll keele töötamise üle
 - – väga aja- ja töömahukas
- **Hübriidsed** süsteemid
 - kombineerivad statistilisi ja reeglipõhiseid süsteeme

Keeletehnoloogia harud

- **Keele vormi põhjal**

- **Kõnetehnoloogiad**

- Nt kõnetuvastus, kõnesüntees, kõnelejatuvastus

- **Tekstitehnoloogiad**

- Nt veebiotsing, õigekirjakontroll, lemmatiseerimine, silbitamine, automaatne märgendus (sõnaliigid, morfoloogia, süntaks, semantika jm), keelekasutusmustrite (nt n-grammide) alusel autorluse tuvastamine, individuaalsete või sotsiaalsete hoiakute esiletoomine (võimudiskursus), emotsioonidetektor, stiili- ja žanri keeleliste tunnuste leidmine jm

- **Funktsiooni järgi**

- **Keeleanalüüs**

- Nt lemmatiseerimine, indekseerimine, sõnaliikide määramine, morfo- ja süntaksianalüüs, semantiline analüüs, õigekirja- ja grammatikavigade parandamine, keeleoskuse hindamine

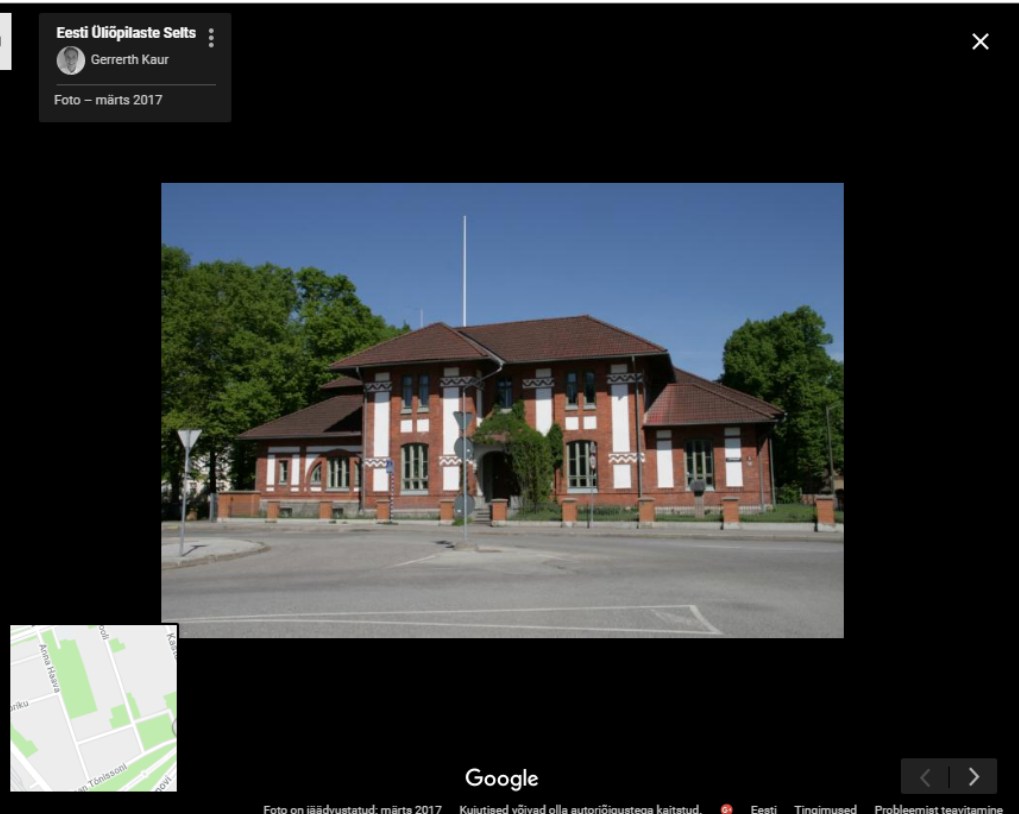
- **Keeleloome** ehk **-süntees**

- Nt sisukokkuvõtja, morfoloogiasüntesaator, kõnesüntesaator, dialoogisüsteemid, masintõlge

- **Dialoogisüsteemid** ja **masintõlge** kasutavad nii keeleanalüüsi kui ka -sünteesi

Eesti keeletehnoloogia kujunemisest

- 1940-ndad – ilmusid esimesed arvutid
- 1950-ndate lõpus – vene-eesti masintõlke projekt
 - Matemaatik Ülo Kaasik koos üliõpilastega
 - Töötati välja venekeelse matemaatilise teksti morfoloogilise analüüsi reeglid ja koostati programm arvutile Ural (Kaasik, Korjus 1959; Palm 1962)
- 1959-1972 end. Eesti Üliõpilaste Seltsi majas Tartu Riikliku Ülikooli arvutuskeskus (Ural-1, Ural-4)
 - 2009 – avati arvutimuuseum J. Liivi tn 2 TÜ Arvutiteaduste Instituudi majas



- **Keeletehnoloogia kui rakenduslik arvutilingvistika**
 - Levis Eestis alates 1990-ndatest
 - Seotud üldise keeletehnoloogilise arendusstrateegiaga Euroopas
 - Eesmärk: luua Euroopa keeletehnoloogiline infrastruktuur
 - Euroopa riikide tarkvarasüsteemide ja e-ressursside integreerumine ühtses võrgustikus
 - Integreeruvate riikide kultuuride ja keelte säilitamine

Loe:

Human Language Technologies – The Baltic Perspective (2012). Toim. Arvi Tavast, Kadri Muischnek, Mare Koit. Täpsemalt: Kadri Vider, Krista Liin, Neeme Kahusk. Strategic Importance of Language Technology in Estonia (pp. 273-279).

<https://pdfs.semanticscholar.org/f2fb/f536db9a0144057500391a8dd33b6d8cff9d.pdf>

Võrdle:

Uszkoreit et al. (1997), vt <http://www.coli.uni-saarland.de/publikationen/softcopies/Uszkoreit:1997:OFT.pdf>

- 1992 – Danzini raport: väikeste keelte üleminek arvutipõhisele keelekasutusele ei toimu iseenesest
 - Vaja keeletöötlusvahendeid (arvutiprogramme)
- Keeletehnoloogilise arvutitoe loomine

Tarkvara	Keeleressursid
Õigekirja kontroll	Elektroonsed sõnastikud
Terminite otsimine tekstist	Formaliseeritud grammatikakirjeldused
Automaatne morfo- ja süntaksianalüüs	Terminoloogia andmebaasid
Kõnetuvastus jne	Tekstikorpused jm
INIMKEELE PAREM MÕISTMINE INIMKEELE AUTOMAATNE TÖÖTLEMINE	

- 1997 – Eesti Informaatikakeskus > rahvusliku infosüsteemi loomine
- 1998 – Tartu Ülikoolis uus eriala – arvutilingvistika
- 1999 – Eesti keeletehnoloogia arengukava
 - Olulised valdkonnad
 - Automaatne kõnetuvastus
 - Grammatikakorrektor
 - Tõlkeprogrammid
 - Infootsingu programmid
 - Interaktiivsed keeleõppe programmid jm

Haldur Õim 2001: *Kui viime ükskõik missuguse keelevaldkonna kirjelduse sellisele formaliseerituse tasandile, et see on arvutisse viidav, siis avastame probleeme, mida me seni ei ole näinud, ja see sunnib meid olema oma lahendustes ja kirjeldustes täpsed ja ühesed.*

- Keeletehnoloogia arendamine on riiklik strateegia
 - Riiklikult rahastatud programmid
 - Eesti keeletehnoloogia programmid 1998-2000
 - Eesti keelenõukogu koostatud Eesti keele arendamise strateegia 2004-2010
 - Eesti keele arengukava 2011-2017 (keele areng, õpetamine, uurimine, kaitse) jne
 - Eesti Keele Keeletehnoloogiline Tugi 2006-2010
 - Jätkuprogramm 2011-2017 (vt <http://vana.keeletehnoloogia.ee>)
 - Käimas 2018-2027 (vt <https://www.keeletehnoloogia.ee/et>)

- Eesti keeletehnoloogia infrastruktuuri loomine
 - täites Euroopa Nõukogu strateegilisi otsuseid keeleressursside kaardistamise, litsenseerimise, tarkvara arendamise, litsenseerimise ja jagamise osas
 - Eesti Keeleressursside Keskus
 - Kuulub Euroopa CLARIN-i infrastruktuuri ja võrgustikku
 - META-SHARE (ressursside kirjeldus ja jagamine)
 - TEI anoteerimismärgenduse standard
 - Multext-East projekt – lingvistilise märgenduse standard
 - Võrgustikud META-NET, META-NORD (Põhjamaade ja Baltimaade keeled)
 - European Language Technology infrastructures, sustainability of linguistic data
 - Digital Humanities, Text Technology, Markup Languages, Linked Open Data, Semantic Web, Information Modelling, Big Data etc.

Multext-East (4. versioon)

Erjavac_ME_2010.pdf - Adobe Acrobat Reader DC

File Edit View Window Help

Home Tools FAIA247-0273.pdf Erjavac_ME_2010.pdf x

2 / 4

Share

Language	Family	MSD Specs	MSD Lexicon	1984 MSD	1984 Alignment	1984 Corpus	Comparable Corpus	Speech corpus
English	Germanic	V1	V1	V1	N/A	V1	V1	-
Romanian	Romance	V1	V1	V1	V1	V1	V1	V1
Polish	West Slavic	V4	V4	V4	-	-	-	-
Czech	West Slavic	V1	V1	V1	V1	V1	V1	-
Slovak	West Slavic	V4	V4	V4	-	-	-	-
Slovene	South West Slavic	V1/V4	V1/V4	V1/V4	V1	V1	V1	V1
Resian	dialect of Slovene	V3/V4	V4	-	-	-	-	-
Croatian	South West Slavic	V2	-	-	-	-	-	-
Serbian	South West Slavic	V2	V2/V4	V3	V3	V2	-	-
Russian	East Slavic	V4	V4	-	-	V2	-	-
Ukrainian	East Slavic	V4	V4	-	-	-	-	-
Macedonian	South East Slavic	V4	V4	V4*	V4	-	-	-
Bulgarian	South East Slavic	V1	V1	V1 [†]	V1	V1	V1	-
Persian	Indo-Iranian	V4	V4	V4	-	-	-	-
Estonian	Finno-Ugric	V1	V1	V1	V1	V1	V1	V1
Hungarian	Finno-Ugric	V1/V4	V1/V4	V1	V1	V1	V1	V1

Table 1: The MULTEXT-East resources by language, and the version when the resources were made / substantially updated. **MSD specs** = morphosyntactic specifications; **MSD lexicon** = morphosyntactic lexicon (entry = wordform+lemma+MSD); **1984 MSD** = MSD and lemma annotated “1984” corpus (* = non-disambiguated; † = automatically tagged with reduced tagset only); **1984 alignment** = sentence alignments with English; **1984 corpus** = “1984”, extensively annotated with structural information (poems, quoted speech, etc.) 100,000 words per language; **Comparable corpus** = fiction and newspaper structurally annotated corpus, 2 x 100,000 words per language; **Speech corpus** = parallel speech corpus, 200 sentences per language, spoken + text.

Start

ET 1:30 18.02.2019

Linke

- Eesti Keeleressursside Keskus:
<https://keeleressursid.ee/et/>
- Keeleveeb: <http://www.keeleveeb.ee/>
- Sõnaveeb: <https://sonaveeb.ee/>
- TÜ arvutilingvistika uurimisrühm, vt
<https://www.cl.ut.ee/index.php?lang=et>
- Github: access the tools and events they need to shape the next generation of software development,
<https://education.github.com/>
- Gitlab: open source software to collaborate on code, nt
https://ained.ttu.ee/pydoc/git_ttu.html

- Keele alkeemia https://alchemy-language-demo.mybluemix.net/?cm_mc_uid=84202806432014811944632&cm_mc_sid_50200000=1481194463
- Eestikeelse teksti sisukokkuvõtja 2013
<http://lepo.it.da.ut.ee/~kaili/Syntax/estsum.html>
- EKI emotsioonidetektor
<http://peeter.eki.ee:5000/valence>

- EKK = Eesti keele käsiraamat, keelenõuanne jm, vt <http://portaal.eki.ee/>
- ES = Emakeele Selts, vt [http://www.emakeeleselts.ee/keeletoimkond.
htm](http://www.emakeeleselts.ee/keeletoimkond.htm)

Kirjandust

- 1. Chomsky (1965) = Хомский Ноам (1965).** Логические основы лингвистической теории. – Новое в лингвистике. Вып. IV. Москва: Прогресс, 465–576. Перевод И. А. Мельчука. Allikas: **Noam Chomsky (1962).** The Logical Basis of Linguistic Theory. – Preprints of Papers for the Ninth International Congress of Linguists. Cambridge, Mass.
- 2. Kaasik, Ülo, Ain Korjus (1959).** Automaatses tõlkimisest. - Keel ja Kirjandus 11,663-673.
- 3. Koit, Mare (2006).** Ratsionalism ja empirism keeletöötluses: Vastasseis või koostöö? – Toim. Ilona Tragel ja Haldur Õim. Teoreetiline keeleteadus Eestis II (Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 7). Tartu: Tartu Ülikooli Kirjastus, 41–54.
- 4. Palm 1962 = Пальм Рээдик (1962).** О морфологическом анализе русской фразы. – Сообщения по машинному переводу 1. Таллин, 59-83.
- 5. Tauli, Valter (1964).** Word Index to August Mälk's Tee kaevule I. August Mälgu Tee kaevule I. Sõnaloend. Uppsala: Soome-Ugri Keelte Instituut.
- 6. Viks, Ülle (2000).** Eesti keele avatud morfoloogiamudel. – Arvutuslingvistikalt inimesele. (TÜ üldkeeleteaduse õppetooli toimetised 1.) Toim. Tiit Hennoste. Tartu, 9-36.

7. Weitzmann, John Hendrik, Georg Rehm, Hans Uszkoreit (2010). Licensing and Sharing Language Resources: An Approach Inspired by Creative Commons and Open Science Data Movements. – Eds. Khalid Choukri, Denise DiPersio, Marc Kupietz, Valérie Mapelli. Proceedings of the LREC 2010 Workshop Legal Issues for Sharing Language Resources: Constraints and Best Practices, Malta.

8. Õim, Haldur (2006). Hargnemisi teoreetilises keeleteaduses. – Toim. Ilona Tragel ja Haldur Õim. Teoreetiline keeleteadus Eestis II (Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 7). Tartu: Tartu Ülikooli Kirjastus, 2–20.

9. Õim, Haldur (2001). Keeletehnoloogiast ja eesti keelest. – Keel ja Kirjandus, 7, 499–501.