

Abimaterjal ja sisend keeletehnoloogia näidetest Digitaalse maailmapildi ainesse

Palun aidake oma valdkonna kõige olulisemad/stabiilsemad lingid ja sisendid siia sisse kirjutada. Siis saan neid kasutada loengumaterjalis ja lisaks ka koduülesannete andmisel või online testide tegemisel.

Kaili Müürisepp
Heiki-Jaan Kaalep
Mark Fišel
Heiki-Jaan Kaalep
Heili Orav
Liisa Rätsep
Andre Tättar

Tekstikorpused:

Mitmed eripärased korpused, nt. vigaste lausete paralleelkorpus, silbitatud korpus, morfoloogiliselt märgendatud korpus (sh. mõnikord märgendatud käsitsi; vt kirjeldusi)<https://cl.ut.ee/korpused/>

Kõige uuemad ja suuremad eesti keele korpused on SketchEngine's

<https://www.sketchengine.eu/>

Ülikoolipere liikmele tasuta kasutatav (institutional login) kuni aprillini 2022, edasi plaanib HVEE osta institutsionaalse litsentsi. Kõige uuem ja suurem on seal eesti keele ühendkorpus 2019, mida on kirjeldatud ka siin

https://metashare.ut.ee/repository/browse/eesti-keele-uhendkorpus-2019/cd9633fab22e11ea_a6e4fa163e9d4547b71a2df64d1f43f1ac26dbd8508ea951/

2021. aasta lõpus peaks veebist kraabitama uus eesti keele veebikorpus EKI juhtimisel.

Sõnade sagedusloend

<https://cl.ut.ee/ressursid/sagedused1/>

Eriti vt loendit, kus on sõnakaupa koos tema kõik korpuses esinenud sõnavormide sagedused (ja nende summa):<https://cl.ut.ee/korpused/>

https://cl.ut.ee/ressursid/sagedused1/failid/lemma_sonavorm_alfabeetilisest.txt

Eesti sõnad (väike, keskmine, suur)
Loend
Sagedusloend

Kirjanik X vs Kirjanik Y - tekstifailid

Eesmärk -- lihtsalt näidata kuidas võrrelda kahe dok sõnade sagedusi

Kus oleks lihtsad tekstifailid autorite ja teoste kaupa.

Luule näited luuletaja X vs luulateja Y või **Eesti luuletajad vs tõkeluule ?**

Kristel Uibo on võrrelnud kirjandusteoseid siin

<http://www.tekstikaeve.ee/blog/2017-10-25-kirjandusteoste-automaatanalyys/>

Tegelikult on ka muud Kristeli tööd <http://www.tekstikaeve.ee/> asjakohased.

Homonüümia - sama sõna täiesti erinev mõte - näited

Morfoloogia

<https://www.filosoft.ee/>

Süntaksipuu - kust saab häid näiteid ja kus aitab süntaksipuu "lahendada"?

Süntaksipuu online:

<https://lindat.mff.cuni.cz/services/udpipe/> on hea näide selle kohta, et kui luua puudepank st süntaktiliselt märgendatud korpus samade märgenduspõhimõtete järgi erinevates keeltes, siis saab lüüa palju kärbsed ühe hoobiga.

Süntaktilise märgendamise ja puudepankade loomise initsiatiiv on siin

<https://universaldependencies.org/>

See on hetkel süntaktilise märgendamise de facto standard maailmas.

Nende puudepankade peal on treenitud mh Stanza <https://stanfordnlp.github.io/stanza/> ; mudelid, sh ka eesti keele jaoks ja SpaCy eesti keele mudel

<https://github.com/EstSyntax/EstSpaCy>

EstNLTK-s on muide veel paremad Stanza mudelid.

Eestikeelse lause süntaksipuid online saab ka siit:

siin on reeglipõhise parseri online demo (cgparser):

<https://korpused.keeleressursid.ee/syntaks/>

ja vanem statistiline (malt) seal kõrval

← → ↺ 🏠 korpused.keeleressursid.ee/syntaks/index.php?keel=ee

Süntaksianalüsaator
 Dependency parser of Estonian



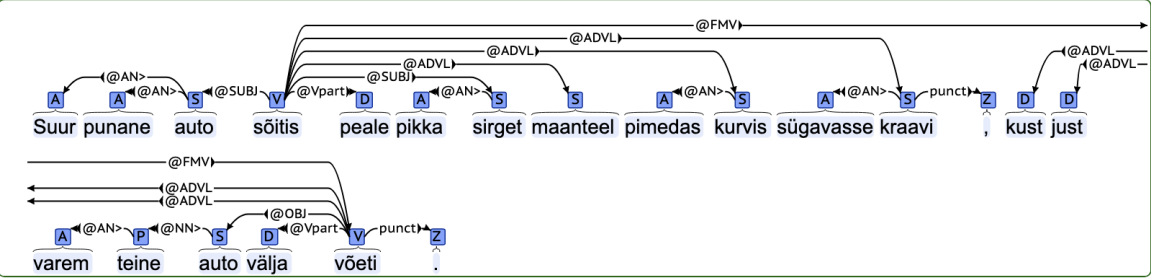
Eesti keeles
 In English

Süntaksianalüüs ☒ Cgparser ☐ Maltparser
 Väljund ☒ Cgparser ☐ Maltparser

Faili analüüs

Suur punane auto sõitis peale pikka sirget maanteel pimedas kurvis sügavasse kraavi , kust just varem teine auto välja võeti .

[edit](#)



Kõnestuvastus online

<https://tekstiks.ee/>

Kõnesüntees online

<https://neurokone.ee/>

<https://eki.ee/heli>

Masintõlge

<https://neurotolge.ee/> + <https://soome-ugri.neurotolge.ee/>

<https://translate.google.com/>

<https://translate.tilde.com/>

Kõne -> tekst -> tõlge

(puudub)

Sõnaraamatud

EKI sõnastikud

Keeleveeb

Keelevara

EKI on ühendamas oma sõnaraamatuid ühtseks andmebaasiks nimega Sõnaveeb

<https://sonaveeb.ee/>

wordnet-tüüpi sõnastik. Esimene Princetoni WordNet. Praeguseks väga paljude keelte jaoks,

vt <http://globalwordnet.org/resources/wordnets-in-the-world/>

Eesti Wordnet: ligi 92000 mõistet. Tähenduste eristamine ja semantiliste seoste võrgustik. Selle põhjal tehtud ka arvuti Alias <https://www.keeleressursid.ee/alias/>

Ing. k.

- GPT-3

*Boonus: <https://ner.tartunlp.ai/>

Kus on **RegExp** toetatud praegu otsingutes?

<https://cl.ut.ee/korpused/kasutajaliides/>

Eesti keelest:

- Kiirlugemine: <https://kiirlugemine.keeleressursid.ee/>
- Õpime käändeid: <http://prog.keeleressursid.ee/opimekaandeid/>
- Keelemäng: <http://167.71.3.184/>
- Eesti Wordneti visualiseerija: <http://prog.keeleressursid.ee/estwnvis/>
- Lauseehituse lihtsustamine: <http://prog.keeleressursid.ee/lihtlauseks/>
- Eesti viipekeele õppeprogramm: <http://prog.keeleressursid.ee/viipekeel/>

Muusikast:

- Muusikateose kirjeldamine loomulikus eesti keeles:
<https://muusik-veeb.herokuapp.com/>
- Notatsiooniprogramm vaegnägijatele (klaviatuuri abil noodid, väljundiks notatsiooniprogrammidele loetav muscxml-fail):
<http://prog.keeleressursid.ee/viipekeel/>
- Notatsiooni faili teisendamine pimedate punktkirja:
<http://prog.keeleressursid.ee/musicdots/>