



# Grammatika- korrektor

Süntaksianalüsaator

16. sept 2011

# Speller vs. grammatikakorrektor

I hasn't done nuffin.

- Õigekirjakorrektor kontrollib, kas sõna leidub leksikonis kontekstile tähelepanu pööramata.
  - Avastab ilmsed vead: *krammatika, kogematta* jne
- Grammatikakorrektor vaatab kogu lauset.
  - Leiab (ideaaljuhul) sõnad, mille vorm ei sobi lausesesse (*muutis elu praemaks*),
  - Ühildumis- ja rektsioonivead: *ta võis tegema*
  - Komavead
  - ...

# Kus grammatikakorrektorit kasutada

- Tekstiredaktoris
  - Leiab vead, pakub parandusi
- Keeleõppeprogrammis
  - Ütleb, et lause on vigane, vajadusel annab täpsema veateate
- Kõnetuvastusel
  - Vähema vigade arvuga sisend on tõenäolisem
- Masintõlkes
  - Sisendi automaatne korrektuur enne järgmisele etapile edasi andmist
- ...

# Milliseid vigu leida

- [strukturaalsed/mittestrukтураalsed]
- Ortograafiavead
  - Tänavu pensjone ei tõstets.
- Trükivead
  - Elu muutub kindalasti praemaks.
- Ühilduvusvead
  - Lapsed on õela, lausa häbematu.
- Algustähevead
  - Minu teada õpib mari tartu ülikoolis.

# Milliseid vigu leida

- Kokku-lahkukirjutusvead
  - Minu lemmik loom onhele sinine papagoi.
- Kirjavahemärgivead
  - Mis aitaks praegust kriisi lahendada oleks omavalitsuste poolne abi.
- Copy-paste vead
  - Muul juhul peab koma nõudva sidesõna ees peab olema koma.
- Sõnajärjevead
  - See arve juba kõik sisaldab.
- ...

# Erinevaid lähenemisi

- Süntaksianalüsaator
  - kas parsib või ei parsi?
- Reeglipõhine
  - Näiteks kitsenduste grammatika
- Statistiline
  - N-grammid, masinõppemeetodid

# Statistiline lähenemine

- EdIt (Chen, Chang 2011)
- *He play an important roles to close this deals.*
- *He looks forward to hear you.*



# Edit

- He *play* an important *roles* to close this *deals*.
- He looks forward to *hear* you.
- *Play* – *played*
- *Roles* – *role*
- *To close* – *in closing*
- *To hear* – *to hearing*
- *Hear you* – *hearing from you*



# Edit

- Masin õpib mustrid:
  - Play ~ role in V-ing
  - Look forward to V-ing
- Reaalajas:
  - Märgendab sisendi POS-infoga
  - Võrdleb sisendit mustrite baasiga
  - Valib vähima muutmisvajadusega mustri
- Demo:  
[http://140.114.214.80/theSite/Edit\\_demo2/](http://140.114.214.80/theSite/Edit_demo2/)

# Teised statistilised meetodid

- Juhendamata õppimine
  - Kui midagi ei leidu suures toimetud tekstide korpuses, siis see on tõenäoliselt väär
- Juhendatud õppimine
  - Tekstid on klassifitseeritud vigasteks ja veatuteks
- Viimase aja trend on kasutada grammatilist informiooni
  - POS, bigrams, trigrams, ajaväljendid

# Grammatikakorrektor kitsenduste grammatikas

- Inglise, soome, rootsi, norra, taani, baski, eesti keelel jne.
- Peamine idee: otsi teadaolevalt ebakorrektsaid konstruktsioone, lisa neile veamärgend
- Vigase koha tuvastamine + korrektse paranduse (veateate) näitamine

# Norra

## grammatikakorrektor

- Morfoloogiliselt ühestatud tekstile rakendatakse vigade tuvastamise reegleid.
- NP-sisene ühildumine (definiitsus, sugu, arv)
- Adjektiivide vahel puuduv koma
- Aluse-öeldistäite ühildumine
- Og/a -vead
- Lauses liiga palju verbe või puudub verb
- Sõnajärjevead
- 75% täpsus (positiivsed alarmid/kõik alarmid)

# Rootsi

## grammatikakorrektor

- Morfoloogiliselt ühestatud tekstile rakendatakse veatuvastusreegleid
- NP-sisene ühildumine (sugu, arv, definiitsus)
- Aluse-öeldistäite ühildumine
- Verbiahelate kontroll
- Sõnajärg

# Näide - ühilduvusviga

"<Lapsed>"

"laps" <laps+d> S com pl nom <cap> CLB SUBJ

"<on>"

"ole" <ole+0> V main indic pres ps3 pl ps af <FinV> <Intr> +FMV

"<õela>"

"õel" <õel+0> A pos sg gen AN> @YHIL

"<, >"

"," <, > Z Com

"<lausa>"

"lausa" <lausa+0> D ADVL

"<häbematu>"

"häbematu" <häbematu+0> A pos sg nom AN> @YHIL

"<. >"

"," <. > Z Fst

- “Öeldistäide olgu siin nimetavas käändes”
- “Öeldistäide peab alusega arvus ühilduma”



# Näide - komaviga

"<Mis>"

"mis" <mis+0> P inter rel sg nom <cap> CLB SUBJ

"<aitaks>"

"aita" <aita+ks> V main cond pres ps af <FinV> <Part-P> <InfP> <el>  
+FMV %OK

"<praegust>"

"praegune" <praegune+t> A pos sg part AN>

"<kriisi>"

"kriis" <kriis+0> S com sg part OBJ

"<lahendada>"

"lahenda" <lahenda+da> V main inf <NGP-P> OBJ

"<oleks>"

"ole" <ole+ks> V main cond pres ps af <FinV> <Intr> +FMV @ERR

"<ka>"

"ka" <ka+0> D ADVL

"<omavalitsuste>"

"oma\_valitsus" <oma\_valitsus+te> S com pl gen ADVL

"<poolne>"

"poolne" <poolne+0> A pos sg nom AN>

"<abi>"

"abi" <abi+0> S com sg nom PRD OBJ

- **“Kahe pöördelise verbivormi vahel peab olema koma (või sobiv sidesõna)”**



# Tulemused

- Täpsus - leitud tegelikud vead / kõik alarmid
- Saagis - leitud vead / tekstis leiduvad vead
- Leitud vea puhul vaadata ka seda, kas anti korrektne veateade/parandus.
- Kas korrektor tuleb toime erinevate tekstiliikidega?
- Mis tüüpi vigu antud korrektor üldse leida suudab?