

# Grammatikakorrektor

Süntaksianalüsaator

21. aprill 2006

Kaili Müürisep

# Õigekirjakontroll e spell checker

- Vastavalt Cmarbigde Üiklloios ttuhed unergiuutle ploee olliune, mis jjräekroars theäd sanõs evasined, oinulle on, et enmisee ja vanimie thät oskelid õgeid.
- Milleks siis õigekirjakontrollijat vaja?
  - Näida haritud ja korrektne
  - Vältida valesti mõistmisi: poe-peo, kalale-kakale, paremaks-praemaks, rektsioon-reaktsioon, sihitis-sihtis

# Vigade põhjused

- Sõrmelöökide vale järjekord: paremaks-praemaks
- Vale klahvi tabamine: rääkima-röökima
- Fonoloogilised vead ja teadmatus: hahastama, süia, minija, drenažöör, kokkukirjutamise vead

# Õigekirjakorrektor või grammatikakorrektor

- 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

# Grammatikakorrektorite liigitus

- **Süntaksipõhised:** tekstile tehakse automaatne süntaktiline analüüs ja igale lausele omistatakse puustruktuur. Lause ei ole korrektne, kui analüüs ebaõnnestub.
- **Reeglipõhised:** *need auto kihutavad*
  - see+PL NIMISÕNA+SG VERB+PL → see+PL NIMISÕNA+**PL** VERB+PL
- **Statistilised:** bigramm-mudel
  - Kui eelnev sõna on see mitmuses, siis järgnev sõna on 10% tõenäosusega nimisõna mitmuses ja 0.001% nimisõna ainsuses

# Grammatikavead

- **Strukturaalsed:** parandamiseks tuleb lisada või kustutada sõna või muuta sõnade järjekorda.
- **Mittestruktuuraalsed:** piisab sõna muutmisest.

# Norra CG-põhine GC

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

# Rootsi CG-põhine GC

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