

# Keeletehnoloogia

## Süntaksianalüsaator ja selle rakendused

Kaili Müürisep

ATI

19. märts 2008

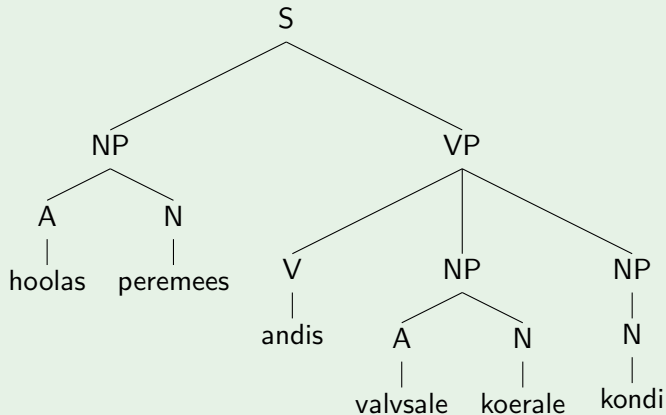
# Täna loengus:

- 1 Sissejuhatus
- 2 Morfoloogiline ühestamine
- 3 Süntakiline analüüs
  - Fraasistruktuurianalüüs
  - Sõltuvusstruktuurianalüüs
  - Pindmine süntakiline analüüs
- 4 Rakendused
  - Grammatikakorrektor
  - Sisukokkuvõtja
    - EstSum
  - Teised rakendused

- Süntaksi** ehk lauseõpetuse uurimisobjekt on lausete ehitus: see, millisest väiksematest osadest laused koosnevad, missugused on osade ülesanded ja omavahelised suhted, kuidas lauseosi saab omavahel ühendada.
- Parser** e süntaksianalüsaator on programm, mis leiab lause struktuuri.

# Näide struktuurist

## Fraasistruktuur



# Näide struktuurist

## Sõltuvusstruktuur

Peremees andis koerale maitstva kondi

```
graph LR; A[Peremees] --> B[andis]; B --> C[koerale]; B --> D[maitstva]; B --> E[kondi];
```

## Näited lausetest

- Kirjalikus keeles peetakse lauseks ühest või mitmest osalausest koosnevat üksust, mis lõpeb punkti, hüüumärgi või küsimärgiga.

### Seadus

Kui tahteavaldus, mis pidi tahteavalduse saajani jõudma kindla ajavahemiku jooksul, jõuab tahteavalduse saajani hilinenult, loetakse tahteavaldus õigeaegselt kättesaaduks, kui tahteavaldus ei jõudnud tahteavalduse saajani õigeaegselt asjaolude tõttu, mille esinemise riisikot kannab tahteavalduse saaja.

*/Tsiivielseadustiku üldosa seadus §69 (4)/*

- Suulise keele põhiüksuseks peetakse lausungit, mis on intonatsiooniline tervik.

# Näited lausetest

- Kirjalikus keeles peetakse lauseks ühest või mitmest osalausest koosnevat üksust, mis lõpeb punkti, hüüumärgi või küsimärgiga.
- Suulise keele põhiüksuseks peetakse lausungit, mis on intonatsiooniline tervik.

## Dialog

A: see oli kõik ee ausatel eesmärkidel et perele auto

B: ei no loomulikult

A: saada. aga lissalt mai saand seda autot

# Analüüsi etapid

- 1 Teksti tükeldamine sõnadeks.

maitea

Liivi 2-339

it's wouldn't ...



# Analüüsi etapid

- 1 Teksti tükeldamine sõnadeks.
- 2 Lausepiiride määramine

Lause algab suure tähega ja lõpeb punktiga. 19. märtsil peab K. Müürisep loengut Liivi t. 2 aud. 207. T. Roosmaa on konverentsil.

# Analüüsi etapid

- 1 Teksti tükeldamine sõnadeks.
- 2 Lausepiiride määramine
- 3 Morfoloogiline analüüs

```
siia  
siia+0 //_D_ //  
siig+0 //_S_ com sg gen //
```

# Analüüsi etapid

- 1 Teksti tükeldamine sõnadeks.
- 2 Lausepiiride määramine
- 3 Morfoloogiline analüüs
- 4 Süntaktiline analüüs
  - 1 Morfosüntaktiline ühestamine

# Analüüsi etapid

- 1 Teksti tükeldamine sõnadeks.
- 2 Lausepiiride määramine
- 3 Morfoloogiline analüüs
- 4 Süntaktiline analüüs
  - 1 Morfosüntaktiline ühestamine
  - 2 Pindsüntaktiline analüüs: fraasipiirid ja ühetasandiline funktsionaalne märgendus

## Fraasipiirid

[Teadmata kadunuks tunnistatud isiku taasilmumisel] või [tema viibimiskoha kindlakstegemisel] tühistab [kohus] [isiku enda või huvitatud isiku nõudel] [tema teadmata kadunuks tunnistamise] .

# Analüüsi etapid

- 1 Teksti tükeldamine sõnadeks.
- 2 Lausepiiride määramine
- 3 Morfoloogiline analüüs
- 4 Süntaktiline analüüs
  - 1 Morfosüntaktiline ühestamine
  - 2 Pindsüntaktiline analüüs: fraasipiirid ja ühetasandiline funktsionaalne märgendus

## Fraasipiirid

[Teadmata kadunuks tunnistatud isiku taasilmumisel] või [tema viibimiskoha kindlakstegemisel] tühistab [kohus] [isiku enda või huvitatud isiku nõudel] [tema teadmata kadunuks tunnistamise] .

- 3 Sügavam süntaktiline analüüs

# Sõnaliikide määramine

- They can fish.
- Time flies like an arrow.
- ...
- Otsi vead üles!
- Keel on võimas realiteet
- Jälgi koolimatemaatika arenguteel
- Täpselt **neid** asju muidugi ei tea, kuid tasuks uurida.
- Võib-olla oleks siis juba mõttekas pakkida sigaretid näiteks äravisatud kooretopsikutesse vms. ja müüa **neid** pruugitud prügikonteineritest.

# Eesti keele sõnaliikide mitmesus

- 1 määrsõna ja sidesõna: *aga, kui, siis, nagu*
- 2 määrsõna ja "ei"
- 3 nud-partitsiibid: *surnud, alustanud ...*
- 4 määrsõna ja alaltütlevas käändes omadussõna: *kindlalt, tugevalt*
- 5 määrsõna ja kaassõna: *mööda, üle, vastu*
- 6 määrsõna ja nimisõna: *siin, kord, siia, lõpuni*
- 7 "oma"
- 8 tud-partitsiibid: *liigutatud, kaotatud*
- 9 nimisõna nimetavas ja verb 3. pöördes: *kõrgus, muutus, tuli, käis, teadis*
- 10 nimisõna seesütlevas ja verb 3. pöördes: *ajas, piisas, mälus*
- 11 nimisõna alalültevas ja määrsõna: *veel*
- 12 nimisõna, määrsõna, sidesõna, verb: *või*
- 13 nimisõna sisseütlevas, määrsõna: *ikka, maha, kokku*

# Käändemitmesused

- maailma-OMAST juhtivad majandusriigid
- maailma-OSAST juhtivad majandusriigid
- maailma-SISSE juhtivad majandusriigid



# Kitsenduste grammatika

Fred Karlsson, 1995. a. morfoloogiliseks ühestamiseks ja pindmiseks süntaktiliseks analüüsiks.

Morfoloogiaanalüsaator lisab kõik märgendid, ühestaja eemaldab need, mis konteksti ei sobi.

## Kitsenduste grammatika

Fred Karlsson, 1995. a. morfoloogiliseks ühestamiseks ja pindmiseks süntaktiliseks analüüsiks.

Morfoloogiaanalüsaator lisab kõik märgendid, ühestaja eemaldab need, mis konteksti ei sobi.

Kahte liiki reegleid:

- vali üks tõlgendus ja kustuta teised

### SELECT

SELECT (V) (-1 (“<ei>”));

*ei pea, ei vea, ei kohta.*

## Kitsenduste grammatika

Fred Karlsson, 1995. a. morfoloogiliseks ühestamiseks ja pindmiseks süntaktiliseks analüüsiks.

Morfoloogiaanalüsaator lisab kõik märgendid, ühestaja eemaldab need, mis konteksti ei sobi.

Kahte liiki reegleid:

- vali üks tõlgendus ja kustuta teised
- kustuta üks tõlgendus

### REMOVE

REMOVE (K prep) (NOT \*1 (S) OR (P) OR (N) BARRIER (V));

*Ta sõitis õigest tänavavaotsast **mööda**.*

# Kitsenduste grammatika

Fred Karlsson, 1995. a. morfoloogiliseks ühestamiseks ja pindmiseks süntaktiliseks analüüsiks.

Morfoloogiaanalüsaator lisab kõik märgendid, ühestaja eemaldab need, mis konteksti ei sobi.

Kahte liiki reegleid:

- vali üks tõlgendus ja kustuta teised
- kustuta üks tõlgendus

Jätab alles mitu tõlgendust, kui ei oska või kahtleb.

Alati jääb alles vähemalt üks tõlgendus.

Reeglid on käsitsi koostatud, regulaaravaldisel sarnased, kontrollivad konteksti, sõnajärge ja sõna enda morfoloogilist informatsiooni

# Näide

```
Peeter      peeter+0 //_S_ prop sg nom //
ei          ei+0  //_V_ aux neg //
pea        pida+0 //_V_ main pres neg //
           pea+0 //_S_ com sg nom //
           pea+0 //_S_ com sg gen //
seda       see+0 //_P_ dem sg part //
tänavat    tänav+0 //_S_ com sg part //
mööda      mööda+0 //_D_ //
           mööda+0 //_K_ prep //
           mööda+0 //_K_ post //
minema     mine+ma //_V_ sup ill //
.          . //_Z_ Fst //
```

## Näide

```
Peeter      peeter+0 //_S_ prop sg nom //
ei          ei+0 //_V_ aux neg //
pea        pida+0 //_V_ main pres neg //
           pea+0 //_S_ com sg nom //
           pea+0 //_S_ com sg gen //
```

```
SELECT(_V_) (-1 ('ei'));
```

```
seda       see+0 //_P_ dem sg part //
tänavat    tänav+0 //_S_ com sg part //
mööda      mööda+0 //_D_ //
           mööda+0 //_K_ prep //
           mööda+0 //_K_ post //
minema     mine+ma //_V_ sup ill //
.          . //_Z_ Fst //
```

# Näide

```
Peeter      peeter+0 //_S_ prop sg nom //  
ei          ei+0  //_V_ aux neg //  
pea        pida+0 //_V_ main pres neg //  
seda       see+0 //_P_ dem sg part //  
tänavat    tänav+0 //_S_ com sg part //
```

```
REMOVE (_K_ prep) (NOT *1 (_S_) BARRIER (_V_))
```

```
mööda      mööda+0 //_D_ //  
           mööda+0 //_K_ prep //  
           mööda+0 //_K_ post //  
minema     mine+ma //_V_ sup ill //  
.          . //_Z_ Fst //
```

# Näide

```
Peeter      peeter+0 //_S_ prop sg nom //  
ei          ei+0  //_V_ aux neg //  
pea        pida+0 //_V_ main pres neg //  
seda       see+0 //_P_ dem sg part //  
tänavat    tänav+0 //_S_ com sg part //  
mööda      mööda+0 //_D_ //  
           mööda+0 //_K_ post //  
minema     mine+ma //_V_ sup ill //  
.          . //_Z_ Fst //
```

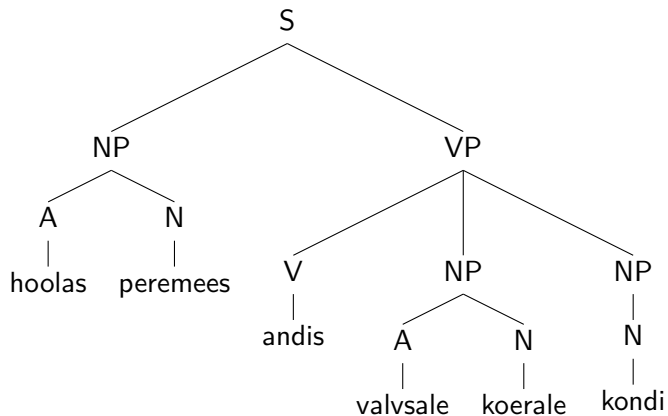
words 8, removed 3, kept 9.



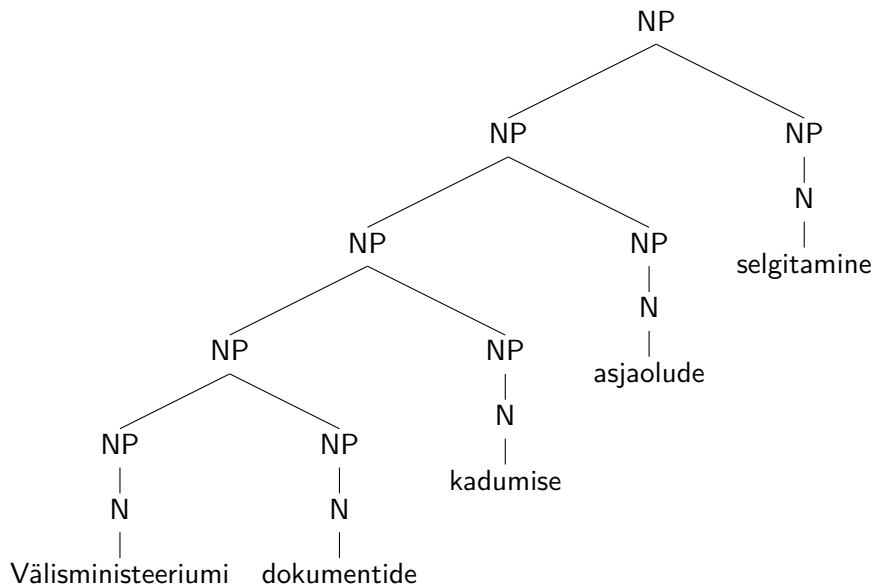
# Tulemused

- 86-92 % sõnadest saavad ühese morfoloogilise tõlgenduse.
- Vigu on vähem kui 2 %, neist kolmandik on eksimused käände valikul.

# Fraasistruktuuripuu



# Fraasid



# Teised fraasid

AP üsna vana koer

PP üle minu laiba

PP vana maja taga

ADVP väga hilja

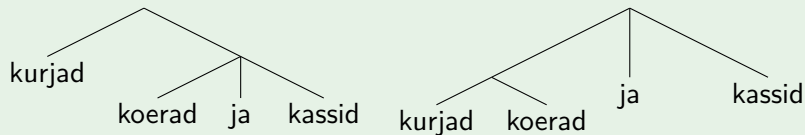
QP palju koeri

VP on tahtnud

VP oleks võinud tahta ujuma minna

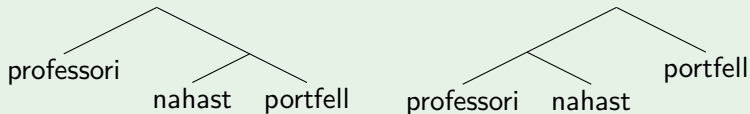
# Struktuurilised mitmetitõlgendatavadused

## Koordinatsioon



# Struktuurilised mitmetitõlgendatavused

## Modifikatsioon



*I saw a man in the hill with a telescope.*

# Fraasistruktuurigrammatika

Asendusreeglid

$S \rightarrow NP VP$

$NP \rightarrow AP NP$

$NP \rightarrow N$

$AP \rightarrow A$

$VP \rightarrow V NP$

$VP \rightarrow V NP NP$

Leksikon

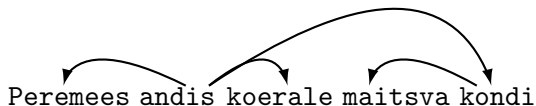
$N \rightarrow \{\textit{peremees, koerale, kondi}\}$

$A \rightarrow \{\textit{hoolas, valvsale}\}$

$V \rightarrow \{\textit{andis}\}$

# Sõltuvusgrammatikad

- Vaatab sõnadevahelisi seoseid.
- Analüüsi keskmeks on verb ja tema rektsioonid.
- Ei kasutata grupeerimist





# Traditsiooniline lauseliigendus

Traditsioonilises grammatikas kirjeldatakse lausestruktuuri lauseliikme mõistet kasutades. Lauses esinevad sõnad jagunevad kõigepealt oma staatuse alusel põhisõnadeks ja laienditeks.

Alusrühm		Öeldisrühm		
Atribuut	Subjekt	Predikaat	Adverbiaal	Objekt
e täiend	e alus	e öeldis	e määrus	e sihitis
tähelepanelik	tudeng	uuris	pingsalt	konspekti
adj	n	v	adv	n

# Eesti keele lauseliikmed

**Predikaat** e öeldis - tegevus

**Subjekt** e alus - tegija

**Objekt** e sihtis - millele tegevus on suunatud, tulemus

**Predikatiiv** e öeldistäide - omadus

**Adverbiaal** e määrus - välised detailid

**Atribuut** e täiend - omadus

Süntaktilised funktsioonid tuleb määrata tekstilausete morfosüntaktiliste omaduste abil.

- grammatiliste morfeemide, eriti käändetunnuste ning partiklite ja adpositsioonide kasutus
- kongruents ja rektsioon
- sõnajärg

# Eesti keele kitsenduste grammatika

- Grammatikas on kasutusel 26 märgendit (alus, öeldise liigid, sihitis, määrus, täiendid jne)
- Märgendus ei ole päris ühene, nt täiendi märgend ei näita, millise sõna juurde ta täpselt kuulub.  
professori (@NN>) nahast (@NN>) portfelli
- Hetkel on grammatikas üle 1400 süntaksireegli.

# Probleemid

- Määrused või määruslikud täiendid:

- ▶ Ta võttis **praeahjust** panni AGA Ta võttis **vasest** ahjuroobi
- ▶ Mees sai siiski pidada ühendust **mobiiltelefoniga** (@ADV L @NN > @ < NN) Kosovos sõdivate poegadega.

- Omastavas käändes sihitis või eestäiend

- ▶ Ta asetab **mantli** (@OBJ @NN >) **tooli** (@OBJ @NN >) seljatoele

- Alus ja sihitis

- ▶ **Eesti rahuvalvajak** vahetasid Bosnias välja **lätlased**.

- Alus ja määrus

- ▶ See **kord** ilmus, **kord** kadus.

# Tulemused

- Korrektsus 96,5 - 98,5%
- Üheseid analüüse 83-90%

Mitmesus tingitud peamiselt semantikast ja lause keerulisest struktuurist

# Demo

# Soome keele sõltuvusgrammatika

<http://www.connexor.eu/technology/machinese/demo/>

# Süntaksianalüsaatori rakendused

- Praktilised rakendused:
  - ▶ grammatikakorrektor
  - ▶ masintõlge, tõlkemälu
  - ▶ teised “sisu mõistvad” rakendused:
    - ★ sisukokkuvõtja,
    - ★ dialoogsüsteemid
    - ★ õpiprogrammid
- Teoreetilised rakendused:
- süntaktiliselt märgendatud korpused ja puudepangad
- sisend semantilisele analüüsile



# Õigekirjakontroll e spell checker

## Mis see on?

Vsativlat Cmarbigde Üilkloios ttuhed unergiuutle ploee olliune, mis jjräekroars theäd sanõs evasined, oinulle on, et enmisee ja vanimie thät oskelid õgeid.

# Õigekirjakontroll e spell checker

## Mis see on?

Vsativlat Cmarbigde Üilkloios 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.
- Mittestrukturaalsed: piisab sõna muutmisest.

# Eesti keele grammatikakorrektor

- Piirduda esialgu komavigadega
- Kasutades kitsenduste grammatika analüsaatori mootorit märgendada kahtlased sõnad märgendiga “korrektne” või “vigane”

# Eesti keele grammatikakorrektor

- Piirduda esialgu komavigadega
- Kasutades kitsenduste grammatika analüsaatori mootorit märgendada kahtlased sõnad märgendiga “korrektne” või “vigane”

```
"<Soovitan>" "soovita" <soovita+n> V main indic pres ps1 sg ps +FMV
"<kõikidel>" "kõik" <kõik+del> P det pl ad ADVL
"<kes>" "kes" <kes+0> P inter rel pl nom **CLBC SUBJ OBJ @ERR
"<sellist>" "selline" <selline+t> P dem sg part NN>
"<teed>" "tee" <tee+d> S com sg part OBJ
"<näinudki>" "näge" <näge+nudki> V main partic past ps FMV
"<pole>" "ole" <ole+0> V aux indic pres ps neg .FinV .Intr +FCV
"<, >" ", " <, > Z Com
```



# Mis on sisukokkuvõtja?

- Programm, mis teeb olemasolevast tekstist lühema versiooni, esitades kasutajale ainult vajalikku infot.
- Sisukokkuvõtte põhieesmärk on esitada teksti peamised ideed väiksemas mahus.

# Sisukokkuvõtete liigitus

- Väljavõte - sisukokkuvõtte laused on originaaltekstist välja valitud
- Ülevaade - sisukokkuvõtte laused on automaatselt genereeritud
- Indikatiivne sisukokkuvõtte peab andma arusaama, millest on dokumendis juttu, ilma detailidesse laskumata.
- Informatiivsed sisukokkuvõtted peavad edastama lühidalt kogu olulise informatsiooni.
- Üldised sisukokkuvõtted
- Teemaatilised sisukokkuvõtted

# Sisend

- Valdkond
- Sisendi pikkus
- Meedia
- Žanr
- Sisendite arv
- Keel

# Kokkuvõtte pikkusest

- Tavaliselt 1%-30%
- Artiklite puhul 10%
- Mitme dokumendi sisukokkuvõte on ettemääratud pikkusega
- Võib olla ette määratud füüsilise kontekstiga - üks ekraanitäis.
- Ühikuks võib olla lõik, lause, sõna, sümbol.

- Ühekeelne
- Mitmekeelne
- Keeltevaheline

# Meetodid

- Pindmised
  - ▶ Sõnasagedused
  - ▶ Asukoht
  - ▶ Pealkirja sõnad
  - ▶ Märkusõnad

# Meetodid

- Pindmised
- Üksuse tasandil

Üksuse tasandil meetodid analüüsivad üksuste vahelisi seoseid:

- ▶ Sarnasus
- ▶ Kaugus
- ▶ Kohesioon ehk ühendatavus

# Meetodid

- Pindmised
- Üksuse tasandil
- Diskursuse tasandil

Diskursuse tasandil meetodid analüüsivad kogu dokumendi struktuuri:

- ▶ Formaat: peatükkide struktuur, hüpertekstide märgendus
- ▶ Teemade lõimed
- ▶ Retooriline struktuur



# Estsumist

- Genereerib väljavõtte
- Väga pindmine meetod ja lihtne algoritm
- Perli-programm
- Eelkõige ajaleheartiklid (uudised)

# Arhitektuur

- 1 HTML-konverter
- 2 Lausestaja
- 3 Väljavõtete tegija

## Postimees

```
<font face="Verdana, Arial" size="3">
<b>Looduskaitstjad uinutasid küla hirmutanud karu </b></font>
<font face="Verdana, Arial" size="2" color="#CC0000"></font>
<!<
img hspace=6 vspace=6 border="" src="http://pm.inspiral.net/post...
>
<br><font face="Verdana, Arial" size="2"><table width=170 align="right">
<tr><td><br><font size=1> Alles viies annus uimasteid ja lihaseid halvavat
  ainet sundis visalt ärkvel püsinud karu alla andma </td></tr></table>
<B>Viljandimaa keskkonnateenistuse ja Nigula looduskaitseala töötajatel
kulus neljapäeval tunde, et kinni püüda ja Pärnumaa metsasügavustesse viia
karu, kes oli päev varem Marna küla elanikke hirmutanud.<P>
</B><P>
```

## Postimees

```
<div1 type='unknown'><head>Looduskaitsjad uinutasid küla hirmutanud karu</head>
<p>Pildi allkiri: Alles viies annus uimasteid ja lihaseid halvavat ainet sundis
visalt ärkvel püsinud karu alla andma</p>
<p><hi rend='bold'>Viljandimaa keskkonnateenistuse ja Nigula looduskaitseala
töötajatel kulus neljapäeval tunde, et kinni püüda ja Pärnumaa metsasügavustesse
viia karu, kes oli päev varem Marna küla elanikke hirmutanud.</hi></p>
```

# Lausestaja

```
<div1 type='unknown'>
<head><s>Looduskaitstjad uinutasid küla hirmutanud karu</s></head>
<p>
<s>Pildi allkiri: Alles viies annus uimasteid ja lihaseid halvavat ainet sundis
visalt ärkvel püsinud karu alla andma</s>
</p>
<p>
<s><hi rend='bold'>Viljandimaa keskkonnateenistuse ja Nigula looduskaitseala
töötajatel kulus neljapäeval tunde, et kinni püüda ja Pärnumaa metsasügavustesse
viia karu, kes oli päev varem Marna küla elanikke hirmutanud.</hi></s>
</p>
<p>
<s>Marna küla Lohu talu peremees Vahur rääkis, et eelmisel õhtul küla piiranud
karu liikus öö läbi samas ringi ja loom leiti hommikul siitsamast metsast.</s>
</p>
```

## Lause skoor

$$W(s) = \alpha P(s) + \beta F(s) + \gamma K(s)$$

- $P(s)$  - positsiooniskoor
  - ▶ Teksti esimene lause
  - ▶ Lõigu esimene lause
  - ▶ Lõigu teine ja kolmas lause
- $F(s)$  - formaadiskoor
  - ▶ Rasvane ja kaldkiri
  - ▶ Hüüu- ja küsimärgid
  - ▶ Jutumärgid
- $K(s)$  - Võtmesõnade skoor

# Ekstraktor - Näide

```
<div1 type='unknown'><head><s>Looduskaitssjad uinutasid küla hirmutanud
karu</s></head>
Min score. 4.14064398
#####
<p>
2. p=9.523810 f=12.745098 w=4.957299 s=8.984648 <s><hi rend='bold'>Viljandimaa
keskkonnateenistuse ja Nigula looduskaitseala töötajatel kulus neljapäeval tunde,
et kinni püüda ja Pärnumaa metsasügavustesse viia karu, kes oli päev varem Marna
küla elanikke hirmutanud.</hi></s>
</p>
<p>
3. p=4.761905 f=4.901961 w=4.628001 s=4.716316 <s>Marna küla Lohu talu peremees
Vahur rääkis, et eelmisel õhtul küla piiranud karu liikus öö läbi samas ringi ja
loom leiti hommikul siitsamast metsast.</s>
</p>
<p>
4. p=4.761905 f=4.901961 w=3.012578 s=4.183227 <s>Hommikul üritati teda metsast
välja ajada ja tehti tema pihta esimene uinutilask.</s>
```

- Kuidas hinnata sisukokkuvõtte headust? Mis teeb ühe sisukokkuvõtte heaks ja teise halvaks?



# Tulemused

- Kuidas hinnata sisukokkuvõtte headust? Mis teeb ühe sisukokkuvõtte heaks ja teise halvaks?
- Kahe inimese poolt koostatud väljavõtetes kattub ainult 70% lausetest.

- Kuidas hinnata sisukokkuvõtte headust? Mis teeb ühe sisukokkuvõtte heaks ja teise halvaks?
- Kahe inimese poolt koostatud väljavõtetes kattub ainult 70% lausetest.
- EstSumi poolt valitud laused kattusid 60% ulatuses inimese poolt valitud lausetega. Parimal juhul oli samu lauseid 85% ja halvimal juhul ei kattunud üksi.

# Probleemid

- Liiga väike andmehulk nii õppimiseks kui hindamiseks

# Probleemid

- Liiga väike andmehulk nii õppimiseks kui hindamiseks
- Lingvistilise mooduli puudumine

# Probleemid

- Liiga väike andmehulk nii õppimiseks kui hindamiseks
- Lingvistilise mooduli puudumine
- Asesõnade lahendamine

## Anafoor

Hoobilt Tartu korvpallipubliku südamed võitnud 20aastane grusiin **Georgi Tsintsadze** ihkas Eestis leida hüppelauda Euroopasse ja korralikku gruusia toitu. Kuid mitut head ei saa, **ta** peab Taaralinnas leppima itaalia köögiga. SL Õhtulehe katse vaadata, mida **Tsintsadze** valib söögiks gruusia restoranis, kukub haledalt läbi. **Ta** teatab nimelt, et Tartus asuv säärane asutus jääb tegelikust gruusia köögist kaugele. **Tema** proovis, aga jalga sinna rohkem tõsta ei taha.

# Probleemid

- Liiga väike andmehulk nii õppimiseks kui hindamiseks
- Lingvistilise mooduli puudumine
- Asesõnade lahendamine
- Sidususe probleemid

# Demo

`http://math.ut.ee/~kaili/estsum/estsumframe.cgi`

# Keele- ja grammatikaõpe

<http://beta.visl.sdu.dk/visl/et/edutainment/games/>



# Näide reeglipõhisest masintõlkest

`http://gramtrans.com/`

`http://sproget.dk/`