

# Suulise dialoogi süntaktiline analüüs

*Parsing Spoken Dialogue*  
*Lauri Carlson*

Margus Treumuth

# Sissejuhatus

- Autori arvamus:
  - suuline keel on kirjalikule keelele väga lähedal
  - väikesed grammatilised erinevused
- Põhiidee:
  - Dialoogi süntaktilises analüüsis tuleks vaadelda pikemat lõiku kui üks voor
  - Üksikutest fraasidest tuleks rekonstrueerida lause

# Lause rekonstrueerimine (1)

- Süntaktilises analüüsis püütakse rekonstrueerida ühe osaleja mõtte üle mitme vooru.

A: mul on

B: mida?

A: vaja ära minna

# Lause rekonstrueerimine (2)

fragment

vaja

@+FMV

ära

@ADV L

minna

@-FMV @ADV L

tervik

mul

@ADV L

on

@+FMV

vaja

@ADV L

ära

@ADV L

minna

@SUBJ

# Kaks vaadet suulisele keelele

- Suulises keeles on teatud erilised protsessid
- Suuline keel on kõrvalekalle kirjakeelest
- Eesmärk: Neid teadmisi rakendades laiendada kirjakeele süntaktilise analüsaatori grammatikat, et see sobiks suulisele keelele

# Suulise keele protsessid

- 1) Ellipsis - fraasi alguse ja/või lõpu kaotamine, kaotatakse ebaoluline osa
- 2) Aposiopesis - lause lõpetamata jätmine, kaotatakse oluline osa
- 3) Asyndeton - ilma sidesõnadeta rääkimine “tulin, nägin, võitsin”
- 4) Anakolouthon - lause algab ühtmoodi, lõpeb teistmoodi
- 5) Apokoinou - sama, mis eelmine, aga keskel on mingi ühisosa
- 6) Repetition - kordus
- 7) Correction - parandus

<http://humanities.byu.edu/rhetoric/>

# Suuline keel kui kõrvalekalle kirjakeelest

*Janne Bondi Johannessen and Fredrik Jørgensen*

Kõrvalekalded

- lexical corrections - sõnaparandused
- syntactic corrections - grammatika parandused
- missing material - kadunud tekst
- interruption - katkestused
- blend – kokkuhääldamine
- false start - valestart
- repetition - kordus
- exclamation – hüüatus

# Kõrvalekallete sagedused

- Katkestusi 50%
- Puuduolevat teksti 30%
- Kui süntaktilises analüüsis käsitleti kõrvalekaldeid, siis suulise keele süntaktiline analüüs kirjakeele vahenditega andis tulemuseks 75% korrektsuse



# Grammatika laiendamine

Analüüsitav lause “This is a sentence and – ”

Kategoriaalne võrrand

$$x \Rightarrow y = x/z \Rightarrow y/z$$

$x/z$  tähendab “x paremas kontekstis on z”

Trahv rakendatakse sõltuvalt sellest kui raske on leida fraas z.

Grammatika

$$S \Rightarrow S \text{ and } S$$

$$S/S \Rightarrow S \text{ and } S/S$$

$$S/S \Rightarrow S \text{ and}$$

# Kategoriaalne grammatika

Intransitiivne verb:  $S \backslash NP$

{koer} haugub

Transitiivne verb:  $(S \backslash NP) / NP$

{laps} sööb {õuna}

# Ühestamine

- Analüüsi leidmine ei ole probleem
- Probleem on korrektse analüüsi leidmine
- Lähenemised:
  - Püütakse leida konstrueeritava lause tähendus
  - Püütakse vaadelda foneetilisi/prosoodilisi tunnuseid
- Näited:
  - pausi järgi on võimalik ennustada järgneva fraasi pikkust
  - väljahingamise järel soovitakse voor üle anda
- Väide: Prosoodilised tunnused on isegi paremad kui kirjavahemärgid kirjakeeles

# Suulise keele märgendamine ja süntaktiline analüüs

Janne Bondi Johannessen and  
Fredrik Jorgensen

# Kava

- NoTa - suulise norra keele korpus
- Kuidas suuline keel erineb kirjakeelest süntaktiliselt
- Soovitused suulise keele süntaktiliseks analüüsiks

# NoTa korpus

- intervjuu iga osalejaga 10 min
- pooletunnine vestlus osalejate paari vahel
- kokku 144 osalejat
- osalejate jaotus (sugu, piirkond, vanus, ...)
- kõne transkribeeritud
- veebiliides otsinguks
- saadaval ka salvestised
- transkriptsioon on aga tavapärase ortograafia alusel (erinevus eesti dialoogikorpussega)
- käsitsi tehtud süntaksipuud

# Suulise keele ja kirjakeele erinevused

- kirjakeeles teatud normid, kus peamiseks on kirjakeeles lause moodustamine, kõnekeeles ei ole see peamine
- sõna parandamine ilma lauset rikkumata
- mitme sõna parandamine, rikutakse ka lause
- alguste/lõppude ärajätmine (ennustatavad algused/lõpud)
- lauseosade ärajätmine, lause poolik
- kaks lauset jagavad ühte elementi ja ei ole selge millise lause osa see on
- valestardid, peale sidesõna jääb lause pooleli
- kordused

# Kõrvalekallete sagedused

- Lauseosad puudu 53%
- Algus/lõpp puudu 26%
- Valestardid 7%



# Soovitused suulise keele analüüsiks (1)

- Lähenemised
  - Ignoreeri kõiki suulise keele tunnuseid
  - Ignoreeri osaliselt
  - Ära ignoreeri ühtegi tunnust
- Soovitatakse kasutada viimast lähenemist

# Soovitused suulise keele analüüsiks (2)

- Fragmentidest tuleks konstrueerida laused
- 75% fragmentidest on juba esialgsel kujul lausena analüüsitavad
- Need, mis ei ole laused, on peamiselt sellised, kus lauseosa/algus/lõpp on puudu
- Puuduolev osa on üldjuhul ennustatav