

# T-(538|725)-MALV, Málvinnsla Setningagreining

Hrafn Loftsson<sup>1</sup> Hannes Högni Vilhjálmsón<sup>1</sup>

<sup>1</sup>Tölvunarfræðideild, Háskólinn í Reykjavík

September 2007

- 1 Setningagreining
- 2 Samhengisfrjáls mállýsing
- 3 Undirflokkun sagna
- 4 Setningafræðileg hlutverk

- 1** Setningagreining
- 2 Samhengisfrjáls mállýsing
- 3 Undirflokkun sagna
- 4 Setningafræðileg hlutverk

# Setningagreining (e. syntactic analysis)

## “Syntax”

- Setningafræði, setningaupbygging.
- Dregið af gríska orðinu “sýntaxis” = arrangement
- Vísar til þess hvernig orðum er skipað saman.

# Tengsl milli orða í setningu

- Ekki er nóg að skoða röð orða í setningu.
- Orðin tengjast mismunandi böndum.
- Þar þarf að skoða þrennt:
  - **Setningarliði** (e. constituents; phrases)
    - Nafnliði, sagnliði o.s.frv.
  - **Setningafræðileg hlutverk**/málfræðileg vensl, (e. syntactic functions/grammatical relations)
    - Frumlag, andlag o.s.frv.
  - **Undirflokkun** (e. sub-categorisation)
    - Sagnir t.d taka ákveðna liði með sér.

Glæra fengin að láni úr námskeiðinu: Inngangur að tungutækni, Eiríkur Rögnvaldsson, HÍ, 2003.

## Hvernig vitum við að orðaruna myndar lið?

- Runan stendur á ákveðnum stöðum í setningu.
- Röð orða innan rununnar er yfirleitt föst.
- Hægt er að færa rununa til sem eina heild.
- Ekki er hægt að stinga neinu inn í rununa.
- Runan myndar merkingarlega heild.

Glæra fengin að láni úr námskeiðinu: Inngangur að tungutækni, Eiríkur Rögnvaldsson, HÍ, 2003.

## Færsla setningarliða

- Ég þekki **gamla manninn** vel en konuna hans ekki.
- **Gamla manninn** þekki ég vel en konuna hans ekki.
- Ég hitti **Jón gamla á Hóli** í morgun.
- **Jón gamla á Hóli** hitti ég í morgun.
- Margir þekkja **gömlu konuna sem missti fölsku tennurnar á gólfið**.
- **Gömlu konuna sem missti fölsku tennurnar á gólfið** þekkja margir.

Höskuldur Þráinsson (1999). *Íslensk setningafræði*, bls. 67.

## Noam Chomsky

- Hið fræga rit Chomsky “Syntactic structures” (1957) byggir á hugtakinu setningarliður.
- [http://en.wikipedia.org/wiki/Syntactic\\_Structures](http://en.wikipedia.org/wiki/Syntactic_Structures)



# Outline

- 1 Setningagreining
- 2 Samhengisfrjáls mállýsing**
- 3 Undirflokkun sagna
- 4 Setningafræðileg hlutverk

## Tilgangur

- e. **context-free grammar (CFG)** eða e. **phrase-structure grammar**.
- Mengi af reglum sem segja til um hvernig tákn í málinu geta verið sett saman til að mynda löglegar setningar.
- Einnig fylgir lýsingunni orðasafn.

# Samhengisfrjáls mállýsing

## Dæmi

Reglur	Orðasafn	
S → NP VP	Det → the	Noun → day
NP → Det Noun	Noun → waiter	Verb → brought
NP → NP PP	Noun → meal	Prep → to
VP → Verb NP	Noun → table	Prep → of
VP → Verb NP PP		
PP → Prep NP		

- Eitt tákn vinstra megin við örina, eitt eða fleiri tákn hægra megin.
- S, NP, VP, PP: **máleiningar** (e. non-terminals)
- Einstök orð: **lokátákn** (e. terminals)

## Samanstendur af

- 1 Mengi af **máleiningum** (e. non-terminals)  $M$
- 2 Mengi af **lokatáknum/tókum** (e. terminals)  $\Sigma$ 
  - Skarast ekki við  $M$
- 3 Mengi af **reglum**  $P$ , á forminu  $A \rightarrow \alpha$ , þar sem  $A$  er máleining og  $\alpha$  er táknastringur úr  $(\Sigma \cup N)^*$
- 4 **Byrjunareiningu**,  $S$

# Samhengisfrjáls mállýsing

## Afleiðsla (e. derivation)

- Í mállýsingunni er  $\rightarrow$  = “rewrite the symbol on the left with the string of symbols on the right”
- Eða  $\rightarrow$  = “leiðir af sér”
- $NP \rightarrow Det \ Noun \rightarrow the \ Noun \rightarrow the \ waiter$
- Strengurinn “the waiter” getur því verið leiddur út af máleiningunni NP.

## Byrjunareining (e. start symbol)

- Málið sem CFG mállýsing stendur fyrir er mengi þeirra strengja sem hægt er að leiða út frá byrjunareiningunni.
- Byrjunareiningin er yfirleitt kölluð S (“the sentence node”).



# Samhengisfrjáls mállýsing

## Afleiðsla (e. derivation)

- Í mállýsingunni er  $\rightarrow$  = “rewrite the symbol on the left with the string of symbols on the right”
- Eða  $\rightarrow$  = “leiðir af sér”
- $NP \rightarrow Det \ Noun \rightarrow the \ Noun \rightarrow the \ waiter$
- Strengurinn “the waiter” getur því verið leiddur út af máleiningunni NP.

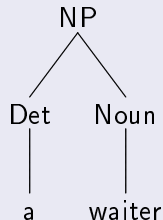
## Byrjunareining (e. start symbol)

- Málið sem CFG mállýsing stendur fyrir er mengi þeirra strengja sem hægt er að leiða út frá byrjunareiningunni.
- Byrjunareiningin er yfirleitt kölluð S (“the sentence node”).



## Þáttunartré (e. parse tree)

Afleiðslu er alltaf hægt að sýna sem þáttunartré:



- Lauf stendur fyrir **lokatakn**, hnútur í trénu stendur fyrir **máleiningu**.
- **Þáttun** (setningagreining; e. parsing): það að varpa streng af orðum yfir í þáttunartré.

# Réttar og rangar setningar

- Líta má á samhengisfrjálsa mállýsingu sem:
  - tæki til að **mynda** (e. generate) setningar.
  - tæki til að gefa setningum **formgerð** (e. structure).
- Málfræðilegar réttar setningar (e. grammatical)
  - setningar sem málfræðin getur myndað.
- Málfræðilega rangar setningar (e. ungrammatical)
  - setningar sem málfræðin getur ekki myndað.

Glæra fengin að láni úr námskeiðinu: Inngangur að tungutækni, Eiríkur Rögnvaldsson, HÍ, 2003.



# Einföld samhengisfrjáls mállýsing fyrir íslensku

Málregla	Skýring
$S \rightarrow NL SL (FL AL)$	NL=nafnliður, SL=sagnliður
$SL \rightarrow so (NL FL)$	FL=forsetningarliður, AL=atviksliður
$NL \rightarrow (LL) (no pfn) (FL)$	LL=lýsingarorðsliður, so=sögn
$LL \rightarrow (AL) lo$	no=nafnorð, pfn=persónufofnafn
$FL \rightarrow fs NL$	fs=forsetning
$AL \rightarrow (AL) ao$	ao=atviksorð

Liður innan sviga er valfrjáls.

- $S \Rightarrow NL SL FL \Rightarrow no SL FL \Rightarrow no so FL \Rightarrow no so fs NL \Rightarrow no so fs no \Rightarrow maðurinn datt í gær$
- $S \Rightarrow NL SL AL \Rightarrow LL no SL AL \Rightarrow lo no SL AL \Rightarrow lo no so AL \Rightarrow lo no so ao \Rightarrow stóri maðurinn hljóp hratt$
- Liðir oft sýndir innan hornklofa:
  - [S [NL **maðurinn**] [SL **datt**] [FL í [NL **gær**]]]
  - [S [NL [LL **stóri**] **maðurinn**] [SL **hljóp**] [AL **hratt**]]

- NL  $\Rightarrow$  (LL) no|pfn (FL)
- Of einfalt – ræður t.d. ekki við:
  - *allir þessir þrír stóru strákar*
    - ófn., áfn., töluorð, LL, no
  - *strákinn sinn*
    - no, efn.
- NL  $\Rightarrow$  (ófn) (áfn) (to) (LL) no|pfn (efn) (FL)
  - [NL Allir þessir þrír [LL stóru LL] strákar NL]
  - [NL strákinn sinn NL]

- NL  $\Rightarrow$  (ófn) (áfn) (to) (LL) no|pfn (efn) (FL)
- Betra núna, en ræður t.d. ekki við nafnliði eins og:
  - [NL Allt NL]
  - [NL Jón Jónsson NL]
  - [NL Mennirnir tveir NL]
  - [NL Hver NL]
  - [NL sjálfan sig NL]
  - [NL þetta allt]
  - [NL landið allt]
  - [NL hinn stóri NL]

$\Rightarrow$  Ekki einfalt að búa til fullkomna samhengisfrjálsa mállýsingu fyrir náttúrulegt tungumál.

# Outline

- 1 Setningagreining
- 2 Samhengisfrjáls mállýsing
- 3 Undirflokkun sagna**
- 4 Setningafræðileg hlutverk

# Undirflokkun sagna (e. subcategorization)

- Sagnir taka mismunandi fylliliði (e. complements)
  - $SL \Rightarrow so (SL \mid (NL) (NL) (FL))$
- Þær hafa mismunandi flokkunarramma (e. subcategorization frames)
  - $[_], [_ NL], [_ NL NL], [_ NL FL], \dots$

Glæra fengin að láni úr námskeiðinu: Inngangur að tungutækni, Eiríkur Rögnvaldsson, HÍ, 2003.

# Undirflokkun sagna (e. subcategorization)

## Dæmi um flokkunarramma

- Ísinn hefur bráðnað [\_]
- Páll hefur saknað þín [\_ NL]
- Þeir hafa fjallað um málið [\_ FL]
- Friðrik hefur verið grannur [\_ LL]
- Þeir munu hafa étið útsæðið [\_ SL]
- Jón hefur lánað Maríu hring [\_ NL NL]
- Ég hef stungið klút í vasann [\_ NL FL]
- Hann vill mála bílinn rauðan [\_ NL LL]

Glæra fengin að láni úr námskeiðinu: Inngangur að tungutækni, Eiríkur Rögnvaldsson, HÍ, 2003.

## Áhrifssagnir og áhrifslausar sagnir

- Sögn sem stýrir falli er **áhrifssögn** (e. transitive verb) og veldur því að fallorð stendur í aukafalli.
  - Fallorðið nefnist **andlag** (e. object)
  - *Ég tek hnífinn (þf.), Ég mætti manni (þgf.), Ég þarfnast næðis (ef.)*
  - *Hann gefur kettinum (þgf.) silung (þf.)* (e. di-transitive verb)
- Sögn sem stýrir ekki falli er **áhrifslaus** (e. intransitive).
  - *Maðurinn hlær* (sögnin að hlæja er **sjálfstæð** sögn).
- Fallorð sem stendur í nefnifalli með áhrifslausri sögn nefnist **sagnfylling** (e. verb complement; predicative nominative).
  - *Konan er góð*



# Outline

- 1 Setningagreining
- 2 Samhengisfrjáls mállýsing
- 3 Undirflokkun sagna
- 4 Setningafræðileg hlutverk

# Setningafræðileg hlutverk

## Frumlag (e. subject)

- Nafnliður sem táknar þann sem gerir, er eða verður það sem umsögnin segir.
- *[NL Gamli hundurinn NL]* beit póstinn.

## Andlag (e. object)

- Nafnliður sem táknar þann sem er (þolandi), eða verður fyrir því sem umsögnin segir.
- *[NL Gamli hundurinn NL]* beit *[NL póstinn NL]*.

# Setningafræðileg hlutverk

## Frumlag (e. subject)

- Nafnliður sem táknar þann sem gerir, er eða verður það sem umsögnin segir.
- *[NL Gamli hundurinn NL]* beit póstinn.

## Andlag (e. object)

- Nafnliður sem táknar þann sem er (þolandi), eða verður fyrir því sem umsögnin segir.
- *[NL Gamli hundurinn NL]* beit *[NL póstinn NL]*.

## Sagnfylling (e. predicative nominative)

- Setningarliður sem stendur með áhrifslausri, ósjálfstæðri sögn (verða, verða, heita, þykja, o.s.frv.) og afmarkar eða fyllir merkingu hennar.
- [NL Hún NL] er [NL mjög góður leikari NL].
- [NL Hann NL] heitir [NL Jón NL].

## Umsögn (e. finite verb)

- Umsögn er sögn í persónuhætti (framsöguhætti, viðtengingar- eða boðhætti). Kjarni eða aðalliður setningar og allir aðrir setningarhlutar tengjast henni á einhvern hátt.
- Hann *skrópaði* í íslensku.

# Setningafræðileg hlutverk

## Sagnfylling (e. predicative nominative)

- Setningarliður sem stendur með áhrifslausri, ósjálfstæðri sögn (verða, verða, heita, þykja, o.s.frv.) og afmarkar eða fyllir merkingu hennar.
- [NL Hún NL] er [NL mjög góður leikari NL].
- [NL Hann NL] heitir [NL Jón NL].

## Umsögn (e. finite verb)

- Umsögn er sögn í persónuhætti (framsöguhætti, viðtengingar- eða boðhætti). Kjarni eða aðalliður setningar og allir aðrir setningarhlutar tengjast henni á einhvern hátt.
- Hann *skrópaði* í íslensku.