

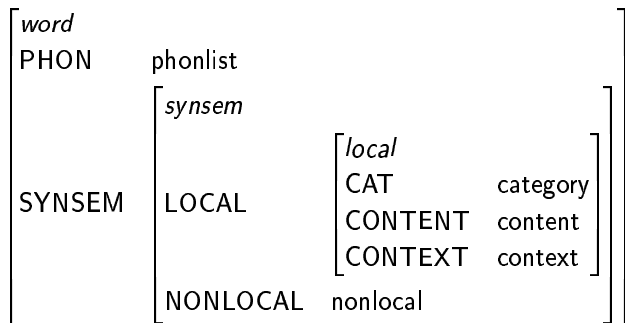
HPSG tutvustus III  
Ühildumine, sõnajärg, leksikon  
Süntaksiteooriad ja -mudelid 2005/06

Kaili Müürisep

ATI

23. märts 2006

## Sõna üldine tunnusstruktuur



**CAT** - informatsioon sõnaliigi ja valentsi kohta

**CONTENT** - sõna semantiline interpretatsioon, eriti viitesuhtega seonduv

**CONTEXT** - kontekstist selguv semantiline informatsioon.

CONTENT väärtused esitavad pigem sõnasõnalist tähendust, CONTEXT väärtused aga tingimusi viidatavale objektile, kasutades taustateadmisi.

# CONTENTist veel

(1) Poiss paitas kassi.

Semantilised rollid: laps on agent, kass on patsient.

«paitas, agent: X, patsient: Y »

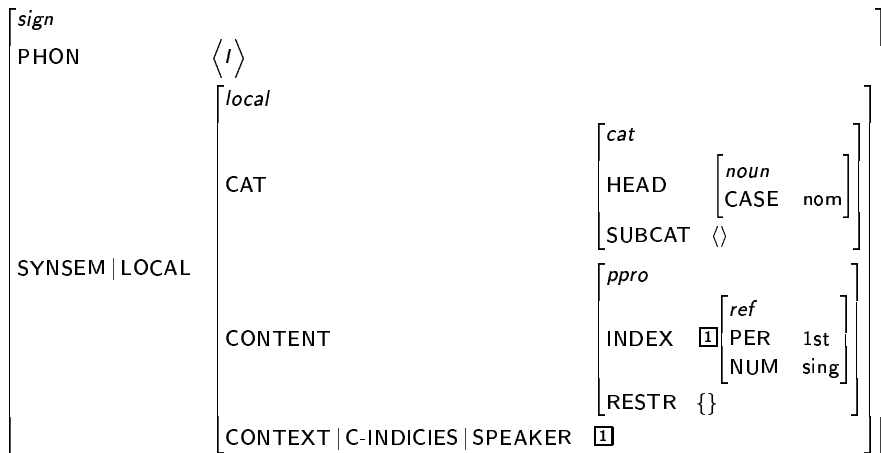
X: «poiss: instance X»

Y: «kass: instance Y»

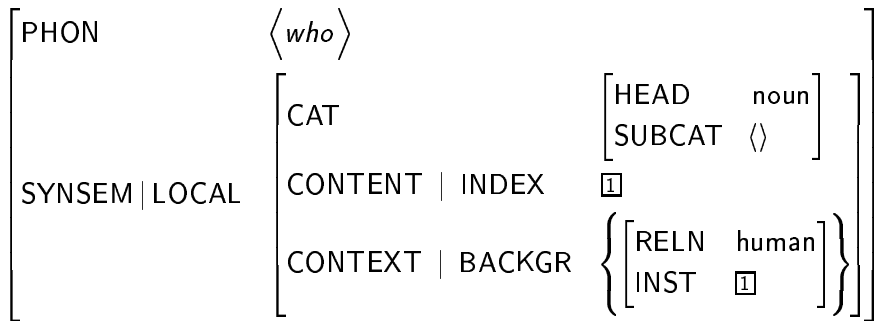
CAT	HEAD	[VFORM fin]					
	SUBCAT	⟨ NP[nom] <sub>[1]</sub> , NP[part] <sub>[2]</sub> ⟩					
CONT	RELN	<b>paitama</b>					
	AGENT	[1]					
	PATIENT	[2]					
INDEX	[1]	PER	1	INDEX	[2]		
		NUM	sg				
		GEN	masc				
RESTR	⟨ RELN		<b>poiss</b>	RESTR	⟨ RELN		<b>kass</b>
	INSTANCE		[1]		INSTANCE		[2]



# Asesõna /



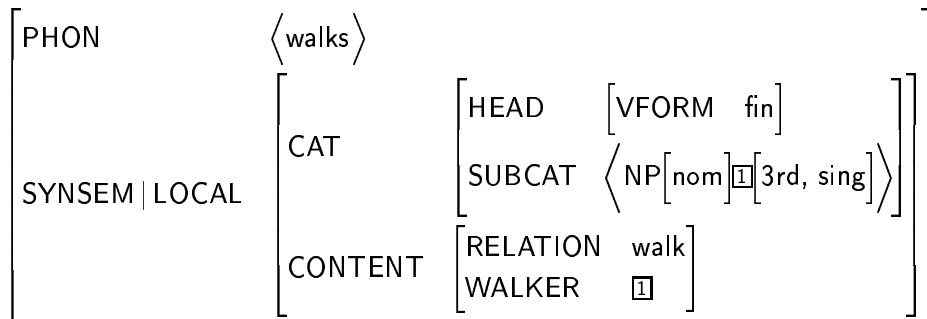
## who/which



(2) \* the boat who I like

(3) the boat which I like

## Verbi ja argumendi ühildumine



Ühildumise spetsifikatsioon on SUBCAT elemendi CONTENT tunnuses, käände valik HEAD tunnuses.

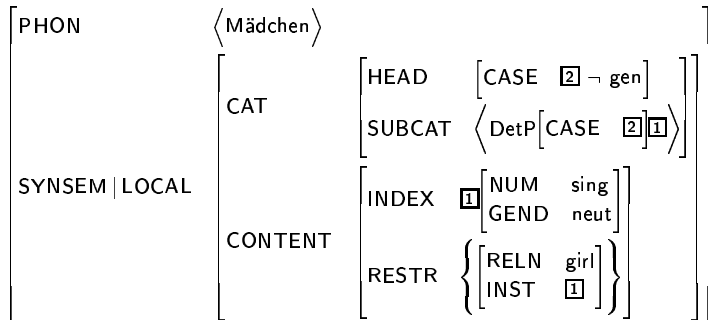
# Determinatiivi ja nimisõna ühildumine

- (4)
- a. every man
  - b. \* every men
  - c. \* all man
  - d. all men

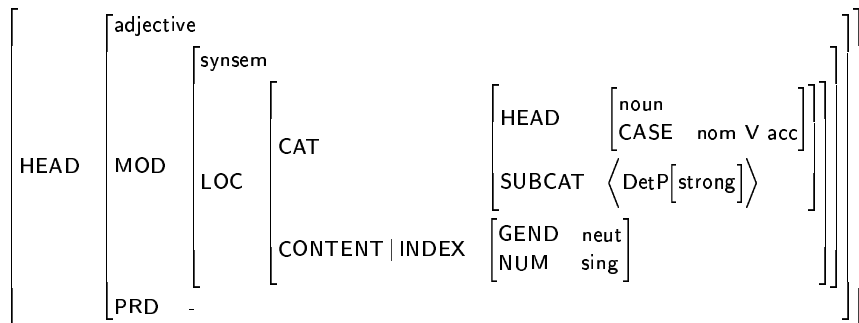
Nimisõnad subkategoriseeruvad determinatiivile ja nende CONTENT väärtused jagavad struktuuri. *Every* tunnus [INDEX | NUMBER sing] ja *all*: [INDEX | NUMBER pl].



# Saksa keele NP



# Atributiivne adjektiiv



Saksa keele omadussõna käänamine sõltub eelnevast artiklist. See on näide nõrgast adjektiivist, mis eeldab, et artikkel on tugev (nt *das*, *dieses*).

- (5) das kluge Mädchen
- (6) \* kluge Mädchen
- (7) \* ein kluge Mädchen

Adjunkti ja põhja konstruksiooni korral toimub struktuuri jagamine adjunkt-tütre MOD väärtuse ja põhjatütre SYNSEM väärtuse vahel:  
*kluge Mädchen* põhja-adjunkti struktuur on selline:

$$\left[ \begin{array}{l} \text{HEAD} \quad \left[ \begin{array}{l} \text{noun} \\ \text{CASE} \quad \boxed{2} \text{ nom V acc} \end{array} \right] \\ \text{SUBCAT} \quad \left\langle \text{DetP}[\text{strong, CASE } \boxed{2}][\text{sing, neut}] \right\rangle \end{array} \right]$$

## Ühildumine HPSG-s

- Derivatsiooniline lähenemine eeldas, et ühildumine on suunatud: on kontrollija ja kontrollitav ning üks määrab teise vormi. Kitsendustepõhine eeldab, et mõlemad osapooled jagavad informatsiooni.
- Ühildumist käsitletakse semantilise nähtusena: ühildumine arvus ja isikus toimub indeks-struktuuris.
- Leidub ka indeksiga mitteseotud ühildumine: süntaktiline ühildumine saksa keeles ja pragmaatiline ühildumine korea keeles.

# Tüübideklaratsiooninäited Matrixis

```
lex-item := sign.           local := avm &
phrase := sign.            [ CAT cat,
                             CONT mrs,
                             ARG-S list ].
synsem := avm &
[ LOCAL local,
  NON-LOCAL non-local ].

cat := avm &
[ HEAD head,
  VAL valence ].

valence :=avm &
[ SUBJ list,
  SPR list,
  COMPS list,
  SPEC list ].
```

## Konstruksioonitüübi näide

```
basic-head-subj-phrase := head-valence-phrase &
                        binary-headed-phrase &
[ SYNSEM.LOCAL.CAT [ VAL [ SUBJ olist,
                        COMPS #comps,
                        SPR #spr ] ],
HEAD-DTR.SYNSEM.LOCAL.CAT
  [ VAL [ SUBJ < #synsem >,
        COMPS #comps,
        SPR #spr ]],
NON-HEAD-DTR.SYNSEM #synsem& canonical-synsem&
  [ LOCAL [ CAT [ VAL [ SUBJ olist,
                    COMPS olist,
                    SPR olist ]]]]].
```

# Kääne

```
noun := head &  
  [ CASE case ].
```

```
case := *top*.  acc := case.
```

```
nom := case.    dat := case.
```

```
poiss := noun-lex &
```

```
  [ STEM < "poiss" > ,
```

```
    SYNSEM [ LOCAL.CAT.HEAD.CASE nom,
```

```
      LKEYS.KEYREL _boy_n_rel ] ].
```

```
trans-verb-lex := verb-lex &
```

```
  [ ...VAL [ SUBJ < [ ...HEAD.CASE nom ] > ,
```

```
    COMPS < [ ...HEAD.CASE acc ] > ]].
```

# PNG

```
png :+ [ PER person,  
        NUM number ].
```

```
person := *top*.  
first  := person.  
second := person.  
third  := person.
```

```
number := *top*.  
sg      := number.  
non-sg  := number.
```



## Aluse ja verbi ühildumine

```
kass := noun-lex &  
[ STEM < "kass" >,  
  SYNSEM [ ...HOOK.INDEX.PNG [ PER third,  
                                NUM sg,  
                                GEND masc ] ],  
  LKEYS.KEYREL _cat_n_rel ] ].  
nurrub := intrans-verb-lex &  
[ STEM < "nurrub" >,  
  SYNSEM [ ...SUBJ < [ ...PNG [ PER third,  
                                NUM sg ] ] >,  
  LKEYS.KEYREL _sneeze_v_rel ] ].
```

## Täiendid

- Täienditel (modifikaatoritel, adjunktidel) on HEAD tunnuses atribuut MOD, mis on sarnane VAL atribuutidega.
- Täiendi-põhjafrasistruktuureegel kontrollib, kas mittepõhjatütre MOD väärtus ja HEAD tütre SYNSEM väärtus langevad kokku

### Adjektiivi ja põhja ühildumine

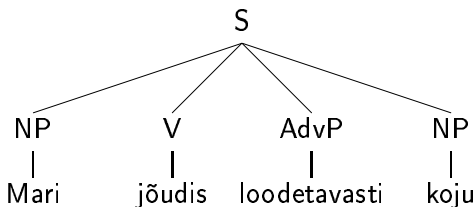
- (8) suur-SG-NOM kass-SG-NOM nurrus.
- (9) väike-SG-NOM poiss-SG-NOM paitas suurt-SG-PART  
kassi-SG-PART

```
kass:=noun-lex &  
[STEM <"kass">,  
  SYNSEM.LOCAL.CAT.HEAD [CASE nom] ].  
suur := adj-lex &  
[STEM <"suur">,  
  SYNSEM.LOCAL.CAT.HEAD [MOD <...HEAD [CASE nom] >]].
```

## Puustruktuur vaba sõnajärje korral

Vaba sõnajärje korral on puustruktuur madal. Nt eesti keele puhul ei liigu verbifraas lauses tervikuna.

(10) Mari jõudis loodetavasti koju



Sama lame on struktuur ka inglise keele küsilauses.

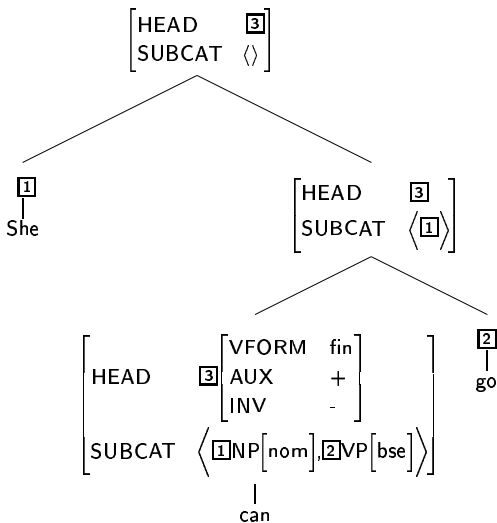
(11) Can she go?

Modaalverbi *can* CAT väärtus:

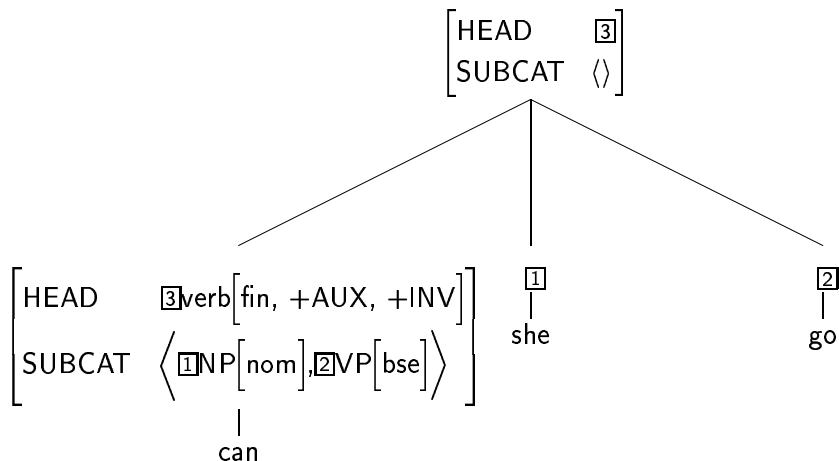
$$\left[ \begin{array}{l} \text{HEAD} \quad \left[ \begin{array}{ll} \text{VFORM} & \text{fin} \\ \text{AUX} & + \end{array} \right] \\ \text{SUBCAT} \quad \left\langle \boxed{\text{I}}\text{NP}[\text{nom}], \text{VP}[\text{bse}] \right\rangle \end{array} \right]$$

Modaalverbe käsitletakse kui tavalisi verbe, ainult nende AUX tunnus on positiivne. Nagu kõik verbid subkategoriseerub *can* alusele ehk nominatiivis nimisõnale ja nagu kõik modaalverbid eeldab põhivormis verbi laiendiks.

She can go.



# Can she go?



# Madala puustruktuuriesituse puudused

- (12)
- a.  $H \rightarrow H[SC\langle A \rangle], A$
  - b.  $H \rightarrow A, H[SC\langle A \rangle]$
  - c.  $H \rightarrow H[SC\langle A \rangle], A, VC$
  - d.  $H \rightarrow H[SC\langle A, B \rangle], A, B, VC$
  - e.  $H \rightarrow H[SC\langle A, B \rangle], B, A, VC$
  - f.  $H \rightarrow A, B, H[SC\langle A, B \rangle]$
  - g.  $H \rightarrow B, A, H[SC\langle A, B \rangle]$
  - h.  $H \rightarrow H[SC\langle A, B \rangle], A, B$
  - i.  $H \rightarrow H[SC\langle A, B \rangle], B, A \dots$

(13)  $2 + \sum_{i=1}^n i! + 2 * (i + 1)!$

n=5 korral 1901 reeglit

## Leksikaliste reeglite kasutamine

Kui võtta kasutusele leksikalised reeglid, mis lubavad SUBCAT listi kõiki permutatsioone, siis reeglite arv väheneb:

- (14)
- a.  $H \rightarrow H[SC\langle A \rangle], A$
  - b.  $H \rightarrow A, H[SC\langle A \rangle]$
  - c.  $H \rightarrow H[SC\langle A \rangle], A, VC$
  - d.  $H \rightarrow H[SC\langle A, B \rangle], A, B, VC$
  - e.  $H \rightarrow A, B, H[SC\langle A, B \rangle]$
  - f.  $H \rightarrow H[SC\langle A, B \rangle], A, B \dots$

5 argumendi korral on reeglite arv 17.

(15)  $2 + 3 * n$

Samas tähendab see seda, et suureneb leksikaalne mitmesus. Näiteks *andma* verbile tekiks 6 leksikalist kirjet. Samuti on keeruline kehtestada reegleid, kui laiendite liikumisel kehtivad siiski mingid piirangud.

Selle asemel kasutatakse: binaarseid puid ja modifitseeritud subkategorisatsiooniprintsiipi või tuuakse sisse uus tunnus Dominance, mis



# Sõnajärjekonfiguratsioonid

```
head-initial := binary-headed-phrase &
  [ HEAD-DTR #head,
    NON-HEAD-DTR #non-head,
    ARGS < #head, #non-head > ].
basic-head-final := binary-headed-phrase &
  [ HEAD-DTR #head,
    NON-HEAD-DTR #non-head,
    ARGS < #non-head, #head > ].
```

## Näide verb3sg reeglist ja aluse ja öeldise ühildumisest

```
estonian.tdl:
```

```
3sg_verb-lex-rule := infl-1tow-rule &  
  [ SYNSEM.LOCAL.CAT.VAL.SUBJ <  
    [ LOCAL.CONT.HOOK.INDEX.PNG [ PER third,  
    NUM sg ] ] >,  
    DTR.SYNSEM.LOCAL.CAT.HEAD verb ] .
```

```
lexicon.tdl: nurru := intr-verb-lex &  
  [STEM <"nurru">, INFLECTED -] .
```

```
ta :=pronoun-lex &  
  [STEM <"ta">,  
    SYNSEM.LOCAL.CONT.HOOK.INDEX.PNG [ PER third,  
    NUM sg ] ]
```

```
irules.tdl
```

```
% (letter-set (!v aeiouõäöü))
```

```
3sg_verb :=
```

```
%suffix (!v !vs)
```

```
3sg_verb-lex-rule.
```