

HPSG tutvustus III

Ühildumine, sõnajärg, leksikon

Süntaksiteooriad ja -mudelid 2005/06

Kaili Müürisep

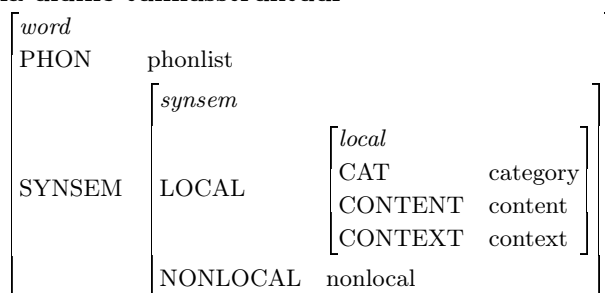
23. märts 2006

Contents

1	Sissejuhatus	1
2	Ühildumine	1
2.1	Inglise keele anafoorne ühildumine	2
2.2	Verbi ja argumendi ühildumine	3
2.3	Determinatiivi ja nimisõna ühildumine	3
2.4	Saksa keele NP ühildumine	4
2.5	Ühildumise esitamine Lingo Grammar Matrixis	5
3	Sõnajärg	7
4	Leksikon	9

1 Sissejuhatus

Sõna üldine tunnusstruktuur



CAT - informatsioon sõnaliigi ja valentsi kohta

CONTENT - sõna semantiline interpretatsioon, eriti viitesuhtega seonduv Noomenite CONTENT sisaldab tunnust INDEX, mis on viitemarker. INDEXil on kolm ühildumistunnust: PERSON, NUMBER ja GENDER. Õeldakse, et kaks noomenit on koindekseeritud, kui nende indeksid jagavad struktuuri: *he shaved himself*. Teine CONTENT'i tunnus RESTRICTION seab INDEXile piiranguid.

CONTEXT - kontekstist selgub semantiline informatsioon. Sisaldab tunnust BACKGROUND, mis kitsendab omakorda INDEXit.

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

CONTENTist veel

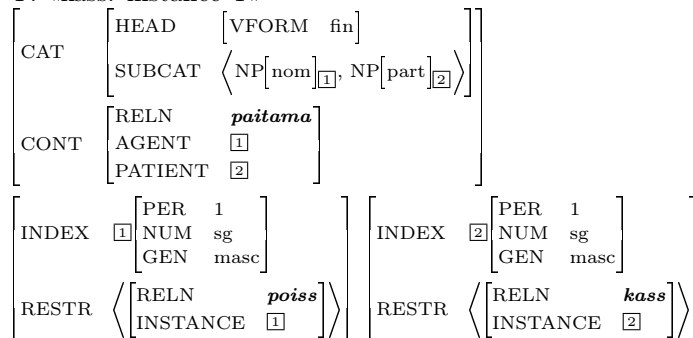
- (1) Poiss paitas kassi.

Semantilised rollid: laps on agent, kass on patsient.

«paitas, agent: X, patsient: Y »

X: «poiss: instance X»

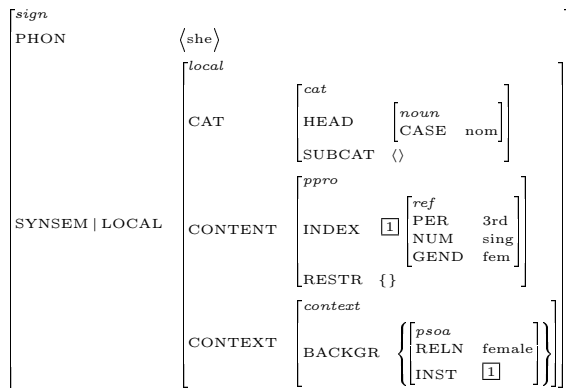
Y: «kass: instance Y»



2 Ühildumine

HPSG ühildumist kontrollivad tunnused

Üldiselt on tavaks lugeda tunnused isik, arv ja sugu süntaktiliste tunnuste hulka. HPSG käsitleb neid aga kui referentidele viitava indeksi struktuuriosa.



2.1 Inglise keele anafoorne ühildumine

Inglise keele asesõna ühildub isikus, arvus ja soos nimisõnaga (referendiga), millele ta viitab/mida ta asendab.

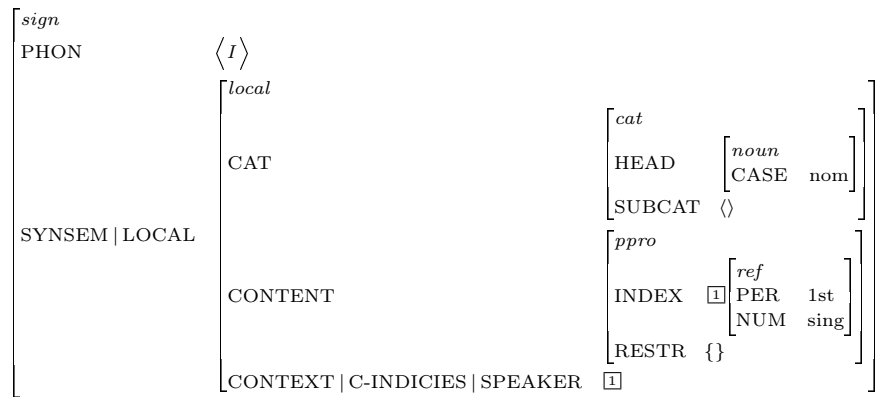
Asesõna *she* CONTENT väärtuse INDEX struktuur näitab, et referendi grammatiline sugu on naissoost. CONTEXT tunnus näitab, et reaalse maailma objekt, millele viidatakse, on naissoost.

Kui loomulik sugu ja grammatiline sugu ei lange kokku (nt saksa keeles), siis CONTENT ja CONTEXT ei jaga sedasi struktuuri.

Asesõna valik sõltub referendi soost:

- (2) My neighbor_{*i*} thinks he_{*i*}/she_{*i*} is smart.

Asesõna *I*



Asesõna *I*. isikus viitab kõnelejale.

who/which

$$\left[\begin{array}{l} \text{PHON} \\ \text{SYNSEM} | \text{LOCAL} \end{array} \left[\begin{array}{l} \langle who \rangle \\ \text{CAT} \\ \text{CONTENT} | \text{INDEX} \\ \text{CONTEXT} | \text{BACKGR} \end{array} \left[\begin{array}{l} \left[\begin{array}{l} \text{HEAD} \quad \text{noun} \\ \text{SUBCAT} \quad \langle \rangle \end{array} \right] \\ \boxed{1} \\ \left\{ \left[\begin{array}{l} \text{RELN} \quad \text{human} \\ \text{INST} \quad \boxed{1} \end{array} \right] \right\} \end{array} \right] \right]$$

(3) * the boat who I like

(4) the boat which I like

Anafoori ühildumine referendiga puudutab ainult isikut, arvu ja sugu. Käände valik sellest ei sõltu

(5) We_i can't stand for people to disagree with us_i.

Ühildumine on arvus ja isikus, kuid mitte käändes.

(6) She_i claims her_i bike has been stolen.

See põhjendab, miks osa ühildumistunnuseid on CAT-tunnuses, osa CONTENTis.

2.2 Verbi ja argumendi ühildumine

$$\left[\begin{array}{l} \text{PHON} \\ \text{SYNSEM} | \text{LOCAL} \end{array} \left[\begin{array}{l} \langle walks \rangle \\ \text{CAT} \\ \text{CONTENT} \end{array} \left[\begin{array}{l} \left[\begin{array}{l} \text{HEAD} \quad \left[\text{VFORM} \quad \text{fin} \right] \\ \text{SUBCAT} \quad \langle \text{NP}[\text{nom}] \boxed{1} [\text{3rd, sing}] \rangle \end{array} \right] \\ \left[\begin{array}{l} \text{RELATION} \quad \text{walk} \\ \text{WALKER} \quad \boxed{1} \end{array} \right] \end{array} \right] \right]$$

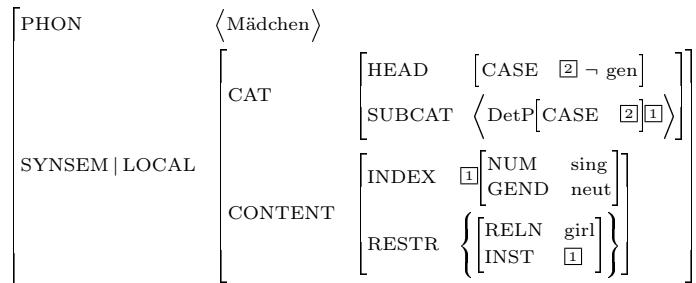
Ühildumise spetsifikatsioon on SUBCAT elemendi CONTENT tunnuses, käände valik HEAD tunnuses.

2.3 Determinatiivi ja nimisõna ühildumine

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

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

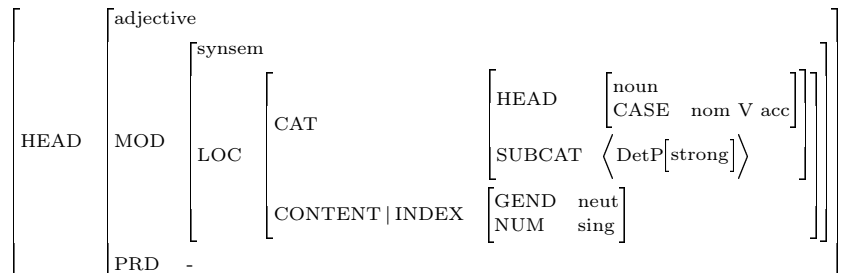
2.4 Saksa keele NP ühildumine



Saksa keele nimisõna ja artikli ühildumisel tuleb kasutada teistsugust lähene- mist kui indeksi ühildumine. See on keelespetsiifiline nähtus, struktuurijagamine peab toimuma nimisõna CASE väärtuse ja determinatiivi või adjektiiviga vahel.

- ein Mädchen
- eines Mädchens
- einem Mädchen
- ein Mädchen

Atributiivne adjektiiv



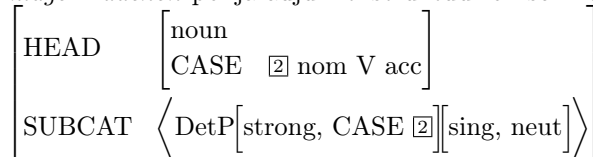
Saksa keele omadussõna käänamine sõltub eelnevast artiklist. See on näide nõrgast adjektiivist, mis eeldab, et artikkel on tugev (nt *das*, *dieses*).

- (8) das kluge Mädchen
- (9) *kluge Mädchen
- (10) *ein kluge Mädchen

Adjektiiv + nimisõna

Adjektiivi ja põhja konstruktsiooni korral toimub struktuuri jagamine adjektiivi MOD väärtuse ja põhjatuure SYNSEM väärtuse vahel:

kluge Mädchen põhja-adjektiivi struktuur on selline:



Nimisõna ja adjektiivi kombineerimine andis järgmised tulemused:

- determinatiiv, millele noomenfraas subkategoriseerub, peab olema tugev

- noomenifraas peab olema nominatiivis või akkusatiivis
- nimisõna ja adjektiivi indeksid on identsed

Predikatiivsel adjektiivil pole käändelõppu, selleks tunnus PRD.

Vahekokkuvõte

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.

2.5 Ühildumise esitamine Lingo Grammar Matrixis

Tüübideklaratsiooninäited Matrixis

```
lex-item := sign.          local := avm &
phrase := sign.          [ CAT cat,
                           CONT mrs,
                           ARG-S list ].
synsem := avm &          ARG-S list ].
[ LOCAL local,
  NON-LOCAL non-local ].
                           valence :=avm &
cat := avm &             [ SUBJ list,
                           HEAD head,    SPR list,
                           VAL valence ]. COMPS list,
                           SPEC list ].
```

Konstruktsioonitüübi 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 ] ] .
```

Tegelikult antakse viimane informatsioon edasi leksikaliste reeglitega ning iga verbivormi eraldi sõnastikku ei kanta.

Täiendid

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

Adjektiivi ja põhja ühildumine

(11) suur-SG-NOM kass-SG-NOM nurrus.

(12) 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] >]].
```

3 Sõnajärg

Puustruktuur vaba sõnajärje korral

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

(13) Mari jõudis loodetavasti koju

S

NP	V	AdvP	NP
Mari	jõudis	loodetavasti	koju

can

Sama lame on struktuur ka inglise keele küsilause.

(14) Can she go?

Modaalverbi *can* CAT väärtus:

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

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.

$$\begin{bmatrix} \text{HEAD} & \boxed{3} \\ \text{SUBCAT} & \langle \rangle \end{bmatrix}$$

$$\boxed{1}$$

 She

$$\begin{bmatrix} \text{HEAD} & \boxed{3} \\ \text{SUBCAT} & \langle \boxed{1} \rangle \end{bmatrix}$$

$$\begin{bmatrix} \text{HEAD} & \boxed{3} & \begin{bmatrix} \text{VFORM} & \text{fin} \\ \text{AUX} & + \\ \text{INV} & - \end{bmatrix} \\ \text{SUBCAT} & \langle \boxed{1}\text{NP}[\text{nom}], \boxed{2}\text{VP}[\text{bse}] \rangle \end{bmatrix}$$

$\boxed{2}$
 go
 |
 can

Can she go?

$$\begin{bmatrix} \text{HEAD} & \boxed{3} \\ \text{SUBCAT} & \langle \rangle \end{bmatrix}$$

$$\begin{bmatrix} \text{HEAD} & \boxed{3}\text{verb}[\text{fin}, +\text{AUX}, +\text{INV}] \\ \text{SUBCAT} & \langle \boxed{1}\text{NP}[\text{nom}], \boxed{2}\text{VP}[\text{bse}] \rangle \end{bmatrix}$$

$\boxed{1}$
 she
 $\boxed{2}$
 go
 |
 can

INV = INVERTED ehk kas modaalsõna on nihutatud teisele positsioonile.

Kui GB-s toimus move- α abil modaalsõna tõstmine, siis HPSG-s näitab kahe lause vahelise seose esinemist nende sõnavara. Sõnad on täpselt samad, ainult muutunud on INV väärtus.

Madala puustruktuuriesituse puudused

Reeglite arv, mida vajatakse on hiigelsuur. Vaja on eraldi reegleid intransitiivsetele, transitiivsetele, ditransitiivsetele jne verbidele. Kui tegemist on

liitverbiga, reeglite arv mitmekordistub. Vabad laiendid võivad paikneda lauses kusiganes ja nende arv ei ole piiratud, seega reeglite arv võib olla VÄGA suur.

- (15) 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. $H \rightarrow H[SC\langle A, B \rangle], B, A \dots$

- (16) a. Magas ta?
 b. Ta magas.
 c. Hat er geschlafen? / Hat er schlafen wollen?
 d. Hat das Kind ihn schlafen lassen?
 e. Hat ihn das Kind schlafen lassen?
 f. dass das Kind ihn schlafen lassen hat. / dass keine Frau diesen Mann liebt.
 g. dass ihn das Kind schlafen lassen hat. / dass diesen Mann keine Frau liebt. ...

$$(17) \quad 2 + \sum_{i=1}^n i! + 2 * (i + 1)!$$

n=5 korral 1901 reeglit

Leksikaliste reeglite kasutamine

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

- (18) 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.

$$(19) \quad 2 + 3 * n$$

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 tunnun Dominance, mis määrab fonoloogiliste üksuste järjekorra.

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 > ].
```

4 Leksikon

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.
```