

Süntaksiteooriad ja -mudelid

MTAT.06.031

5. loeng

Kaili Müürisep
kaili.muurisep@ut.ee

ATI

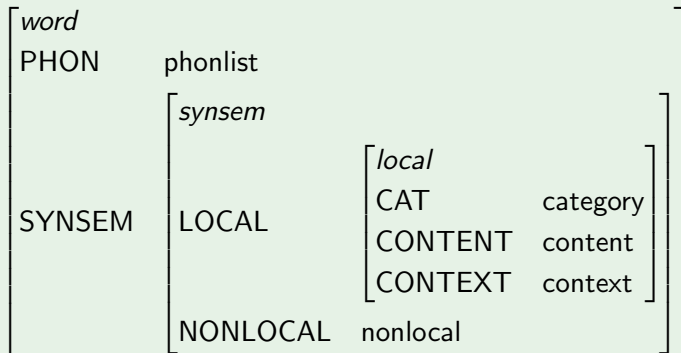
13. märts 2008

- 1 Kordamine
 - Tunnusstruktuur
 - Põhja tunnuse printsiip
 - Alamkategoriad
- 2 Semantika HPSG-s
- 3 Põhi-laiend-skeem
- 4 Ühildumine
 - Ülevaade
 - Probleemid kontrollija ja kontrollitavaga
 - Klassikaline HPSG ühildumine
 - Verbi-argumendi ühildumine
 - Determinatiivi-nimisõna ühildumine
- 5 Põhi-adjunkt-skeem

Tunnusstruktuur

- Kogu info antakse edasi tunnusstruktuuridega

Tunnusstruktuur



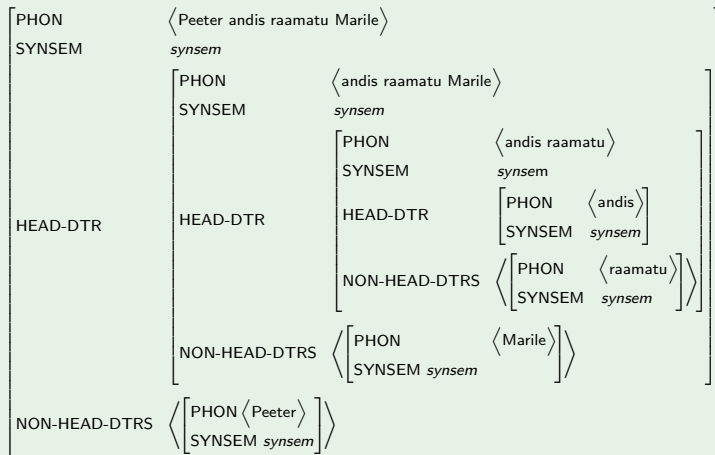
Tunnusstruktuur

- Kogu info antakse edasi tunnusstruktuuridega
- Tunnusstruktuuridel on tüübid, mis määravad, millised argumendid ja väärtused konkreetsel struktuuril olla võivad. Tüübid on jagatud hierarhiatesse. Alamtüüp omab üldisema tüübi kõiki omadusi ning lisaks veel spetsiifilisi kitsendusi.

Tunnusstruktuur

- Kogu info antakse edasi tunnusstruktuuridega
- Tunnusstruktuuridel on tüübid, mis määravad, millised argumendid ja väärtused konkreetsel struktuuril olla võivad. Tüübid on jagatud hierarhiatesse. Alamtüüp omab üldisema tüübi kõiki omadusi ning lisaks veel spetsiifilisi kitsendusi.
- Fraasstruktuur esitatakse tunnusstruktuurina

Fraasistruktuur



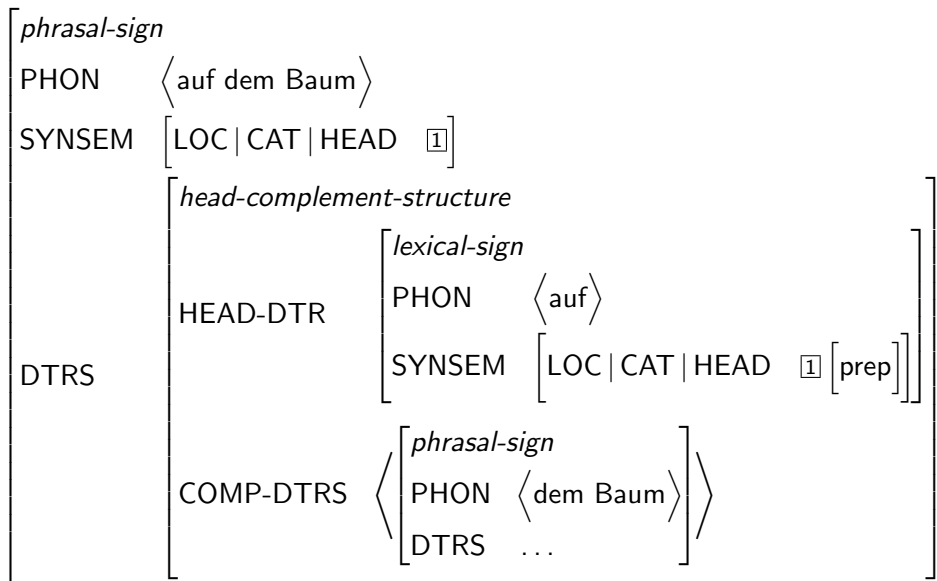
Põhja tunnuse printsiip (HFP)

Printsiip

Põhjaga fraasi HEAD atribuudi väärtus jagab struktuuri tuumtütre HEAD atribuudi väärtusega.

$$\left[\begin{array}{l} \textit{head-argument-structure} \\ \text{SYNSEM.LOCAL.CAT} \left[\text{HEAD} \boxed{1} \right] \\ \text{HEAD-DTR} \left[\text{SYNSEM.LOCAL.CAT} \left[\text{HEAD} \boxed{1} \right] \right] \end{array} \right]$$

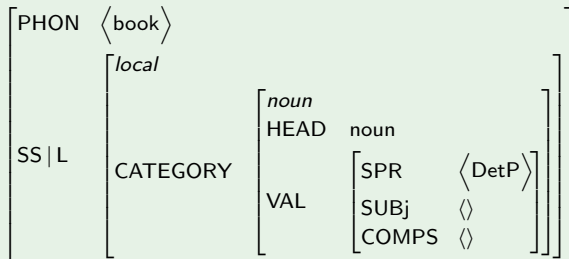
Näide



Alamkategoriad

Subkategorisatsioon on spetsifikatsioon, mis näitab selle märgiga seotud teiste elementide arvu ja vormi.

Sõna *book* valentsiinfo



Alamkategoriad

Subkategorisatsioon on spetsifikatsioon, mis näitab selle märgiga seotud teiste elementide arvu ja vormi.

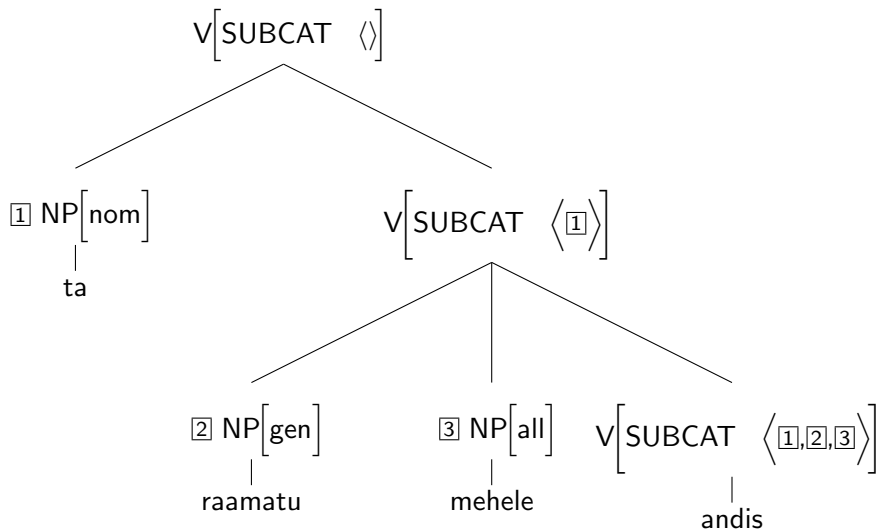
Printsiip

Tuumtütre SUBCAT atribuudi väärtuseks on fraasi SUBCAT listi ja laiendtütarde SYNSEM-väärtuste listi konkatenatsioon.

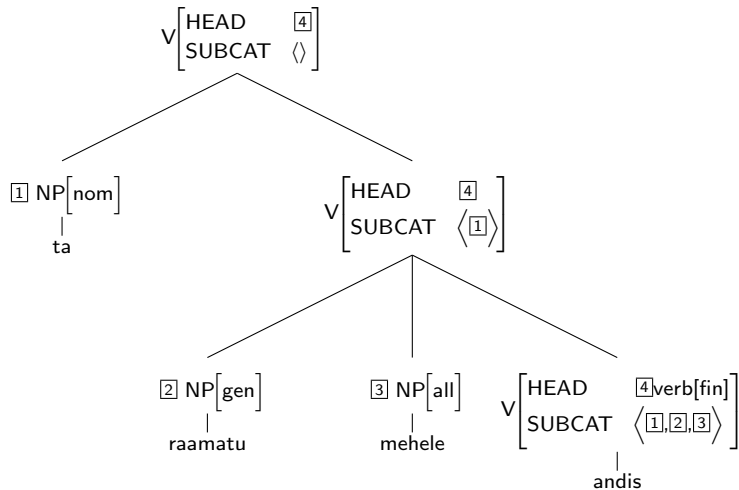
$$H[SUBCAT \boxed{A}] \rightarrow H[SUBCAT \boxed{A} \oplus \langle \boxed{B} \rangle], \boxed{B}$$

$$\left[\begin{array}{l} SYNSEM \quad [LOC | CAT | SUBCAT \quad \boxed{1}] \\ DTRS \quad \left[\begin{array}{l} HEAD-DTR \quad [SYNSEM | LOC | CAT | SUBCAT \boxed{1} \oplus \langle \boxed{2} \rangle] \\ COMPS-DTRS \quad \boxed{2} \end{array} \right] \end{array} \right]$$

Ta andis mehele raamatu



Näide põhjaprintsiibist ja subkategorisatsioonist



Semantilisest notatsioonist

Maaailm koosneb indiviididest, esemest, kontseptsioonidest ja relatsioonidest nende vahel. Relatsioonid võivad olla erineva aarsusega:

- 0 sajab
- 1 magab, sureb, niheleb
- 2 armastab
- 3 annab
- 4 ostab

Argumentidel on üldised nimed: agent, patsient, kogeja, allikas, eesmärk, teema, koht, instrument ...

θ -teooria

- Chapman killed Lennon.
- *kill*: verb; 1 2
NP NP

Chapman on **AGENT**; Lennon on **PATIENT**.

θ -teooria

θ -rollid

AGENT - keegi, kes alustab predikaadiga tähistatud tegevust.

PATSIENT - predikaadiga tähistatud tegevuse objekt.

TEEMA - asi või olend, keda liigutati tegevuse käigus.

KOGEJA - olend, kes koges tegevuse käigus.

BENEFITSIENT - olend, kes sai kasu tegevuse käigus.

SIHT - asi või olend või koht, mille suunas käib tegevus.

ALLIKAS - asi või koht, kust midagi liigutati.

KOHT - koht, kus tegevus toimub.

θ -teooria

θ -rollid

AGENT - keegi, kes alustab predikaadiga tähistatud tegevust.

PATSIENT - predikaadiga tähistatud tegevuse objekt.

TEEMA - asi või olend, keda liigutati tegevuse käigus.

KOGEJA - olend, kes koges tegevuse käigus.

BENEFITSIENT - olend, kes sai kasu tegevuse käigus.

SIHT - asi või olend või koht, mille suunas käib tegevus.

ALLIKAS - asi või koht, kust midagi liigutati.

KOHT - koht, kus tegevus toimub.

Näited

- 1 Constance rolled the ball towards Poirot.
- 2 The ball rolled towards the pigsty.
- 3 Madame Maigret had been cold all day.
- 4 Maigret likes love stories.

Semantilisest notatsioonist

«Kolmik: relatsioon, roll:referent;polaarsus»

```
<<schlafen, thema:karl;1>>  
<<mann, instance:maria;0>>  
<<glauben, patient:karl,  
    proposition:<<lieben, agent:maria,patient:karl;1>>;1>>  
<<scheinen,proposition:<<schlafen,thema:karl;0>>;1>>
```

Semantilisest notatsioonist

«Kolmik: relatsioon, roll:referent;polaarsus»

```
<<schlafen, thema:karl;1>>  
<<mann, instance:maria;0>>  
<<glauben, patient:karl,  
    proposition:<<lieben, agent:maria,patient:karl;1>>;1>>  
<<scheinen,proposition:<<schlafen,thema:karl;0>>;1>>
```

Parametriseeritud suhted:

```
<<lieben,agent:X,patsient:Y;1>>  
X|<<mann,instance:X;1>>,Y|<<kind,instance:Y;0>>
```

Indeksid

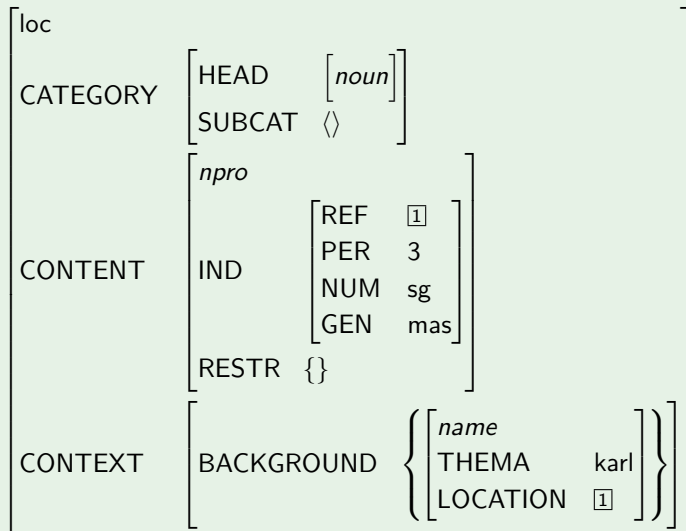
- HPSG-s on semantilised suhted kirjeldatud tunnuses SYNSEM.LOC.CONT.
- Kui väärtused on parametrizeeritud, ja parameeter sõltub kontekstist. Seda haldab SYNSEM.LOC.CONTEXT.IND tunnus.
- Sel on isiku, arvu, soo ja referendi tunnused. INDEX esineb koos RESTRICTION tunnusega, mis seab referendile piiranguid
- Referendiks võib olla nimisõna või asesõna, selle tunnused jagunevad omakorda sõltuvalt sellest, kas tegemist on isikulise asesõnaga või refleksiivse või retsiprooksega.

Karl

| | | | | | | | | | | | | |
|---|--|---|-------------|-----|------|---|-------------|---|-------------|---|------|---|
| <i>npro</i> | | | | | | | | | | | | |
| IND | <table><tr><td>REF</td><td>1</td></tr><tr><td>PER</td><td>3</td></tr><tr><td>NUM</td><td>sg</td></tr><tr><td>GEN</td><td>mas</td></tr></table> | REF | 1 | PER | 3 | NUM | sg | GEN | mas | | | |
| REF | 1 | | | | | | | | | | | |
| PER | 3 | | | | | | | | | | | |
| NUM | sg | | | | | | | | | | | |
| GEN | mas | | | | | | | | | | | |
| RESTR | <table><tr><td><table><tr><td><i>mann</i></td><td></td></tr><tr><td>INST</td><td>1</td></tr></table></td><td>,</td><td><table><tr><td><i>dumm</i></td><td></td></tr><tr><td>INST</td><td>1</td></tr></table></td></tr></table> | <table><tr><td><i>mann</i></td><td></td></tr><tr><td>INST</td><td>1</td></tr></table> | <i>mann</i> | | INST | 1 | , | <table><tr><td><i>dumm</i></td><td></td></tr><tr><td>INST</td><td>1</td></tr></table> | <i>dumm</i> | | INST | 1 |
| <table><tr><td><i>mann</i></td><td></td></tr><tr><td>INST</td><td>1</td></tr></table> | <i>mann</i> | | INST | 1 | , | <table><tr><td><i>dumm</i></td><td></td></tr><tr><td>INST</td><td>1</td></tr></table> | <i>dumm</i> | | INST | 1 | | |
| <i>mann</i> | | | | | | | | | | | | |
| INST | 1 | | | | | | | | | | | |
| <i>dumm</i> | | | | | | | | | | | | |
| INST | 1 | | | | | | | | | | | |

Background

Karli taust



Näide

gibt

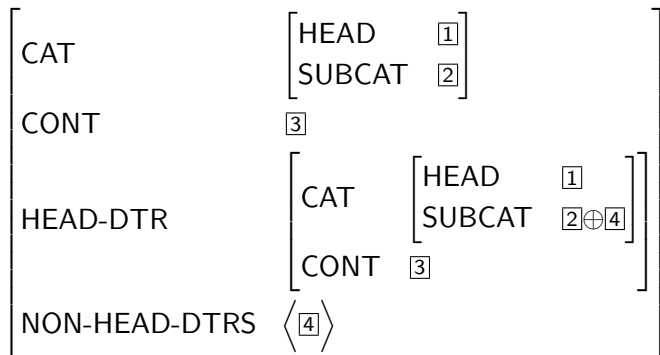
| | | |
|------|---|---|
| CAT | HEAD | $\begin{bmatrix} \textit{verb} \\ \text{VFORM} \quad \textit{fin} \end{bmatrix}$ |
| | SUBCAT | $\langle \text{NP}[\textit{nom}]_{\boxed{1}}, \text{NP}[\textit{acc}]_{\boxed{2}}, \text{NP}[\textit{dat}]_{\boxed{3}} \rangle$ |
| CONT | $\begin{bmatrix} \textit{geben} \\ \text{AGENT} \quad \boxed{1} \\ \text{THEMA} \quad \boxed{2} \\ \text{GOAL} \quad \boxed{3} \end{bmatrix}$ | |

Semantikaprintsiip

[SYNSEM | LOC | CONT 1]
[DTRS | HEAD-DTR | SYNSEM | LOC | CONT 1]

- Ei kehti Põhi-adjunkt seose korral.

Põhi-laiend-skeem



Ühildumine

Ühildumise ehk kongruentsi all mõeldakse sõna teatud vormitunnuse tingitust teise elemendi (sõna või moodustaja) vastavatest vormitunnustest või semantilistest omadustest. Kaks erinevat lähenemist ühildumisele:

- Tuletusepõhine lähenemine: ühildumine on ühesuunaline, üks pool (ühildumise kontroller) dikteerib ühildatava vormi (verb ühildub alusega).
- Kitsendusepõhine lähenemine: kumbki osapool kannab osalist informatsiooni, see informatsioon peab ühilduma.

Ühildumine

Ühildumise ehk kongruentsi all mõeldakse sõna teatud vormitunnuse tingitust teise elemendi (sõna või moodustaja) vastavatest vormitunnustest või semantilistest omadustest. Tavalisemad ühildumiskategooriad

- sugu - *novõi dom, novaja kniga, novoe taksi*
- kääne - *selle+s maitsva+s viineri+s,*
- arv - *nen+de+s maitsva+te+s viineri+te+s*
- määratus e definiitsus - rootsi keele artikkel
- omamine e possessiivsus - *minun kirja+ni, sinun kirja+si*
- isik - *mina tule+n, sina tule+d ...,*
- elusus/elutus - kes - mis

Ühildumine

Ühildumise ehk kongruentsi all mõeldakse sõna teatud vormitunnuse tingitust teise elemendi (sõna või moodustaja) vastavatest vormitunnustest või semantilistest omadustest.

- (1) illarum duarum bonarum
see-GEN.FEM.PL kaks-GEN.FEM.PL kena-GEN.FEM.PL
feminarum
naine-GEN.FEM.PL
nende kahe kena naise

Suahiili keele aluse ja verbi arvuühildumine

- (2) a. Kikapu kikubwa kimoja kilianguka
korv suur üks kukkus
Üks suur korv kukkus.
- b. Vikapu vikubwa vitatu vilianguka
korvid suured kolm kukkusid
Kolm suurt korvi kukkusid.

Probleemid kontrollija ja kontrollitavaga

Possessiivsus- ja isikuühildumise puhul pole iseenesest selge, mis on ühildumise kontrollija. Nt on tavaline, et ainsuse ja mitmuse 1. ja 2. isiku asesõnaline subjekt võib ära jääda. *Tule+n* (pro-drop).

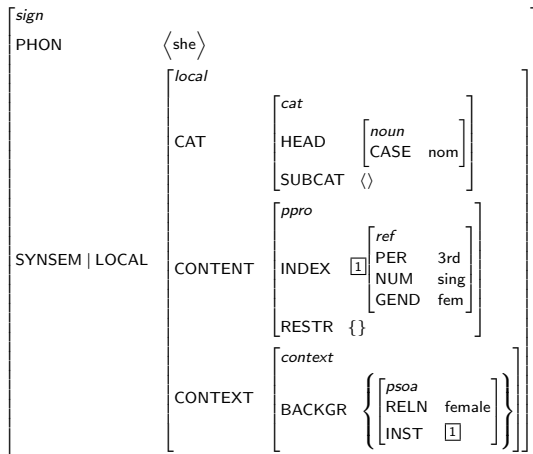
Kasutatakse foneetiliselt tühja subjekti, et ühildumist saavutada.

Kuna subjekt on kontrollija, siis tuletusliku lähenemise korral tuleb luua nii mitu tühja subjekti, kui palju on erinevaid verbi infleksioonivorme.

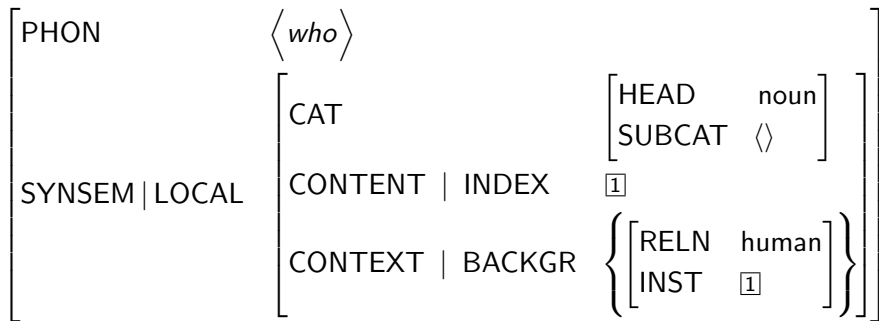
Kitsendusepõhise lähenemise korral toimub verbi ja subjekti (mille ühildumistunnused on määramata) vastavate tunnuste unifitseerimine ja kunstlikku mitmesust pole vaja luua.

Klassikaline HPSG ühildumine

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



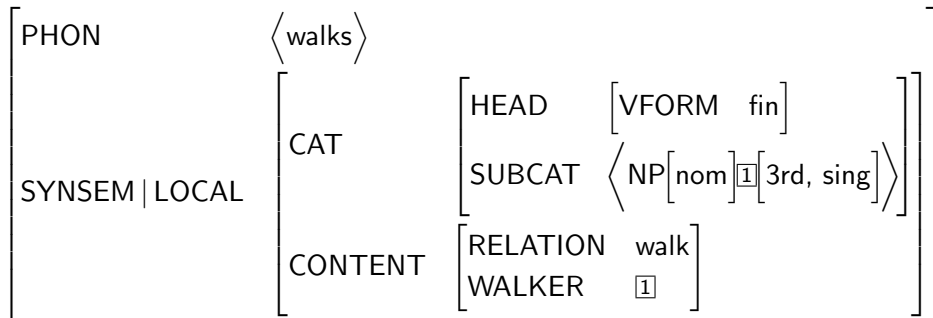
who/which



(3) * the boat who I like

(4) the boat which I like

verbi-argumendi ühildumine

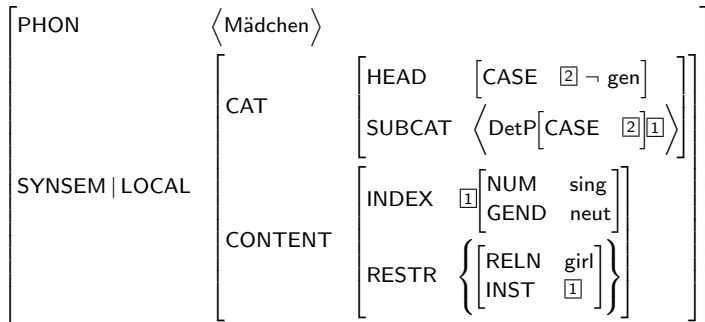


Determinatiivi-nimisõna ühildumine

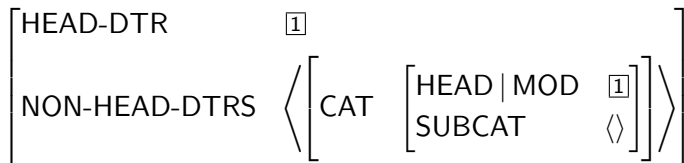
- (5)
- a. every man
 - b. * every men
 - c. * all man
 - d. all men

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

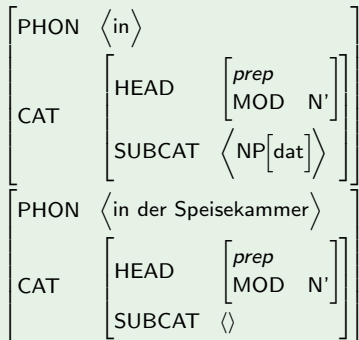
Saksa keele NP



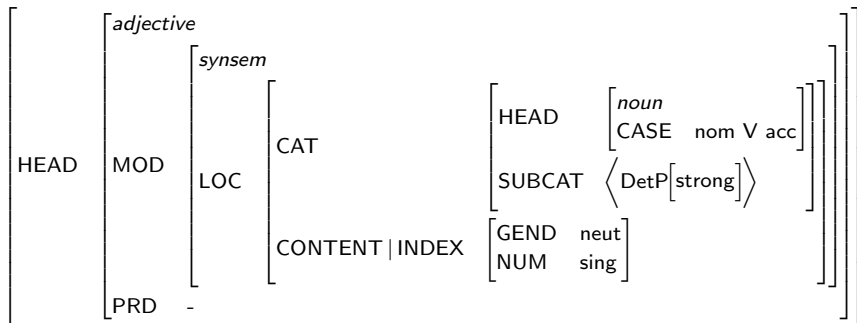
Põhi-adjunkt-skeem



die Wurst in der Speisekammer



Atributiivne adjektiiv



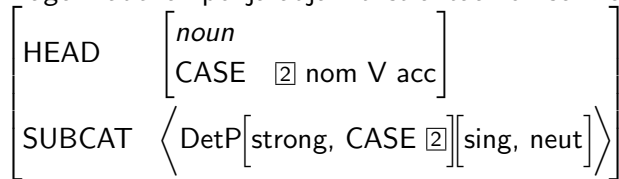
Saksa keele omadussõna käänamine sõltub eelnevast artiklist. See on näide nõrgast adjektiivist, mis eeldab, et artikkel on tugev (nt *das*, *dieses*).

- (6) das kluge Mädchen
- (7) * kluge Mädchen
- (8) * ein kluge Mädchen

Adjektiiv+nimisõna

Adjunkti ja põhja konstruktsiooni korral toimub struktuuri jagamine adjunkt-tütre MOD väärtuse ja põhjatütre SYNSEM väärtuse vahel:

kluge Mädchen põhja-adjunkti struktuur on selline:



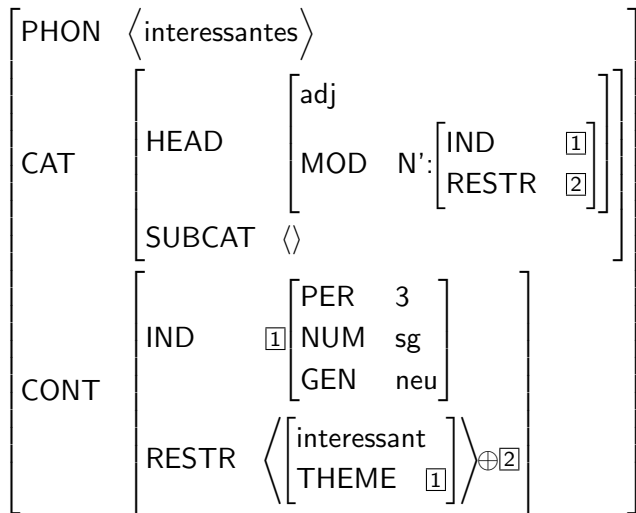
Tähenduse kombineerimisest

- huvitav(X), raamat(Y), $X=Y$

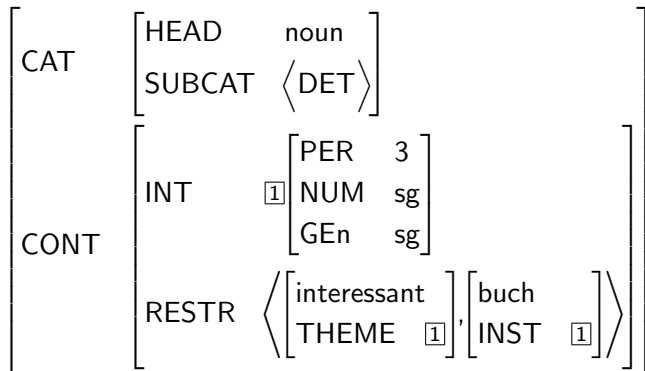
Tähenduse kombineerimisest

- huvitav(X), raamat(Y), $X=Y$
- potentsiaalne(X), mõrvar(Y), $X \neq Y \Rightarrow$ Neid tuleb kombineerida!

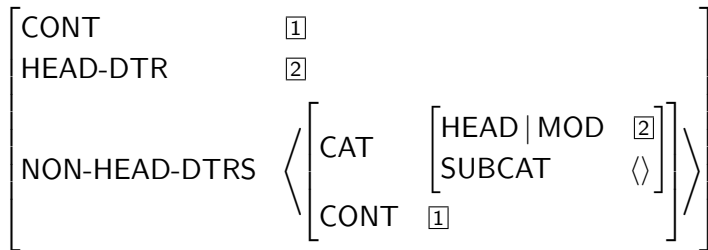
Adjektiivi tähenduse tunnusstruktuur



Fraasi tunnusstruktuur



Põhi-adjunkt-skeem + tähendus



Semantikaprintsiip

- Põhi-mitteadjunkt $\left[\begin{array}{l} \text{CONT} \quad \boxed{1} \\ \text{HEAD-DTR} \mid \text{CONT} \quad \boxed{1} \end{array} \right]$
- Põhi-adjunkt $\left[\begin{array}{l} \text{CONT} \quad \boxed{1} \\ \text{NONHEAD-DTRS} \quad \langle \left[\text{CONT} \quad \boxed{1} \right] \rangle \end{array} \right]$

Ühildumine HPSG-s

- Derivatsiooniline lähenemine eeldas, et ühildumine on suunatud: on kontrollija ja kontrollitav ning üks määrab teise vormi.
- 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.

- LKB - Linguistic Knowledge Builder
- PET - ilma silumisvahenditeta optimiseeritud parser
- Grammatikad
 - ▶ LinGO English Resource Grammar (ERG)
 - ▶ La Grenouille
 - ▶ JaCY Grammar
 - ▶ Korean Resource Grammar
 - ▶ Modern Greek Resource Grammar
 - ▶ NorSource Grammar
 - ▶ Spanish Resource Grammar
- Lingo Grammar Matrix