

HPSG tutvustus I

Süntaksiteooriad ja -mudelid 2005/06

Kaili Müürisep

ATI

9. märts 2006

Ülevaade

- Kontekstivaba grammatika
- Tunnuste struktuurid
- Tüübihierarhia
- Unifikatsioon
- Leksikon oluline
- Konstruktsioonid

KVG puudused

- Väga palju reegleid
- Reeglitel pole kitsendusi $NP \rightarrow Adv AP$
- Pole üldiseid reegleid

Tunnuste struktuurid

- Fraasistruktuurireeglite tipud on tunnus-väärtus paaride kogumid.

DCG

$np(K,A) \rightarrow ap(K,A), np(K,A).$

See on samaväärne:

$np(\text{kääne: } K, \text{ arv: } A) \rightarrow ap(\text{kääne: } K, \text{ arv: } A), np(\text{kääne: } K, \text{ arv: } A).$

Tunnuste struktuurid

- Fraasistruktuurireeglite tipud on tunnus-väärtus paaride kogumid.
- Tunnused võivad olla ka keerukamad

slash-tunnus DCG-s

$s \rightarrow$

$np(\text{pers}:P, \text{numb}:N, \text{sex}:S, \text{case}:C),$

$s(\text{slash}:np(\text{pers}:P, \text{numb}:N, \text{sex}:S, \text{case}:C)).$

$s \rightarrow$

np	,	s	.
$\left[\begin{array}{l} \text{pers} : P \\ \text{numb} : N \\ \text{sex} : S \\ \text{case} : C \end{array} \right]$		$\left[\begin{array}{l} \text{slash} : \\ \quad np \\ \quad \left[\begin{array}{l} \text{pers} : P, \\ \text{numb} : N, \\ \text{sex} : S, \\ \text{case} : C \end{array} \right] \end{array} \right]$	

Tunnuste struktuurid

- Fraasistruktuurireeglite tipud on tunnus-väärtus paaride kogumid.
- Tunnused võivad olla ka keerukamad
- Väärtustatakse ainult need tunnused, mis olulised:

DCG

$np(\text{kääne: } K, \text{ arv: } A) \rightarrow np(\text{kääne: } gen, \text{ arv: } _), np(\text{kääne: } K, \text{ arv: } A).$

Tunnuste struktuurid

- Fraasistruktuurireeglite tipud on tunnus-väärtus paaride kogumid.
- Tunnused võivad olla ka keerukamad
- Väärtustatakse ainult need tunnused, mis olulised:
- Tunnused võimaldavad üldistada: kõik verbid on tunnusega [*HEAD : verb*] sõltumata nt nende transitiivsusest ja pöördest.

Tunnuste struktuurid

- Fraasistruktuurireeglite tipud on tunnus-väärtus paaride kogumid.
- Tunnused võivad olla ka keerukamad
- Väärtustatakse ainult need tunnused, mis olulised:
- Tunnused võimaldavad üldistada: kõik verbid on tunnusega [*HEAD : verb*] sõltumata nt nende transitiivsusest ja pöördest.
- Tunnusstruktuurid võivad olla rekursiivsed:

Genealoogia

[isik		
EESNIMI	ANU	
PEREKONANIMI	TAMM	
SÜNNIAEG	09.03.2006	
ISA	[isik	
	EESNIMI	JAAN
	PEREKONANIMI	TAMM
	SÜNNIAEG	09.03.1966
	ISA	...

Veel tunnuste struktuuridest: subsumptsioon

(1) FS1: [*CAT np*]

(2) FS2: []

(3) FS3: $\left[\begin{array}{l} \text{CAT } np \\ \text{ARG } \left[\begin{array}{l} \text{NUM } sg \\ \text{PERS } 3rd \end{array} \right] \end{array} \right]$

(4) FS4: $\left[\begin{array}{l} \text{CAT } np \\ \text{ARG } [\text{NUM } sg] \end{array} \right]$

Veel tunnuste struktuuridest: subsumptsioon

- (1) FS1: [*CAT np*]
- (2) FS2: []
- (3) FS3: $\left[\begin{array}{l} \text{CAT } np \\ \text{ARG} \left[\begin{array}{l} \text{NUM } sg \\ \text{PERS } 3rd \end{array} \right] \end{array} \right]$
- (4) FS4: $\left[\begin{array}{l} \text{CAT } np \\ \text{ARG} [\text{NUM } sg] \end{array} \right]$

FS2 \sqsubseteq FS1 \sqsubseteq FS4 \sqsubseteq FS3

Veel tunnuste struktuuridest: subsumptsioon

- (1) FS1: [*CAT np*]
- (2) FS2: []
- (3) FS3: [*CAT np*
ARG [*NUM sg*
PERS 3rd]]
- (4) FS4: [*CAT np*
ARG [*NUM sg*]]

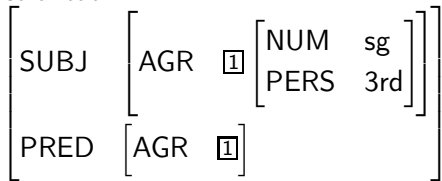
$FS2 \sqsubseteq FS1 \sqsubseteq FS4 \sqsubseteq FS3$

Definition

Tunnuste struktuur B subsumeerib tunnuste struktuuri A ($B \sqsubseteq A$), kui B sisaldab informatsiooni, mis on alamhulk A informatsioonist.

Struktuuri jagamine

Kaks või enam erinevat atribuuti võivad omada üht ja sama tunnuste struktuuri



Tüübihierarhia

- Tüübid võimaldavad täpselt deklareerida, milliseid operatsioone seda tüüpi objektiga saab teha, millised on tema argumendid ja nende väärtused.
- Tüüpide deklareerimine hõlbustab keeruliste andmestruktuuride korral tagada grammatika (programmi) korrektsust.
- HPSG signatuur defineerib, millist tüüpi objekte eksisteerib, mis on nende atribuudid ja mis on nende atribuutide võimalikud väärtused.
- Tüübid on struktureeritud hierarhiatesse, mis võimaldab vastavaid objekte jagada klassidesse.

ERGi 3559 tüüpi

Unifitseerimine

- Tunnuste struktuurid võivad sisaldada erinevat, kuid siiski ühilduvat informatsiooni

(5) FS5: $\left[\begin{array}{l} CAT \textit{ np} \\ ARG [NUM \textit{ sg}] \end{array} \right]$

(6) FS6: $\left[\begin{array}{l} CAT \textit{ np} \\ ARG [PERS \textit{ 3rd}] \end{array} \right]$

Unifitseerimine

- Tunnuste struktuurid võivad sisaldada erinevat, kuid siiski ühilduvat informatsiooni
- Aga ka erinevat ja mitteühilduvat informatsiooni

(7) FS7: $\left[\begin{array}{l} CAT \textit{ np} \\ ARG [NUM \textit{ sg}] \end{array} \right]$

(8) FS8: $\left[\begin{array}{l} CAT \textit{ np} \\ ARG [NUM \textit{ pl}] \end{array} \right]$

Unifitseerimine

- Tunnuste struktuurid võivad sisaldada erinevat, kuid siiski ühilduvat informatsiooni

(5) FS5: $\left[\begin{array}{l} CAT \textit{ np} \\ ARG [NUM \textit{ sg}] \end{array} \right]$

(6) FS6: $\left[\begin{array}{l} CAT \textit{ np} \\ ARG [PERS \textit{ 3rd}] \end{array} \right]$

- Aga ka erinevat ja mitteühilduvat informatsiooni
- Kui kaks struktuuri ühilduvad, siis leidub kolmas, mida nad mõlemad laiendavad.

(9) FS9: $\left[\begin{array}{l} CAT \textit{ np} \\ ARG \left[\begin{array}{l} NUM \textit{ sg} \\ PERS \textit{ 3rd} \end{array} \right] \end{array} \right]$

Unifitseerimine

- Tunnuste struktuurid võivad sisaldada erinevat, kuid siiski ühilduvat informatsiooni

(5) FS5: $\begin{bmatrix} CAT\ np \\ ARG\ [NUM\ sg] \end{bmatrix}$

(6) FS6: $\begin{bmatrix} CAT\ np \\ ARG\ [PERS\ 3rd] \end{bmatrix}$

- Aga ka erinevat ja mitteühilduvat informatsiooni
- Kui kaks struktuuri ühilduvad, siis leidub kolmas, mida nad mõlemad laiendavad.

(9) FS9: $\begin{bmatrix} CAT\ np \\ ARG\ \begin{bmatrix} NUM\ sg \\ PERS\ 3rd \end{bmatrix} \end{bmatrix}$

Definition

Kahe tunnuste struktuuri A ja B *unifikatsiooniks* nimetatakse tunnuste struktuuri C, nii et $A \sqsubseteq C$ ja $B \sqsubseteq C$.

- HPSG-s on enamik informatsiooni leksikonis
- Leksikoni kirjed sisaldavad informatsiooni ortograafia, sõnaliigi, ühildumise, valentsi, semantika ja argumentide liitmise kohta.
- Kasutatakse rohkesti supertüüpe, nii et tegelik kirje määrab vaid sõna tüübi, ortograafia ja võtmerelatsioonid
- Leksikaalsed reeglid seovad leksikoni baaskirjed teiste kirjetega (mitmus, passiivivorm jne)

Konstruksioonid

Väga üldised fraasistruktuurireeglid teevad enamuse tööst:

- põhi-määratleja e *head-specifier*
- põhi-argument e *head-complement*
- põhi-alus e *head-subject*
- põhi-täitja e *head-filler*
- põhi-laiend e *head-modifier*

Üksikud konstruksioonid vajavad eraldi reegleid:

- Relatiivlused
- Noun-noun compounds
- Lisand

ERGi on 105 reeglit

- Sarnasused

- ▶ Kontekstivabad grammatikad
- ▶ Mõlemad kasutavad üldisi printsiipe
- ▶ Reeglite osakaal väike
- ▶ Leksikonil suur osakaal
- ▶ Mitmed printsiibid on väga sarnased

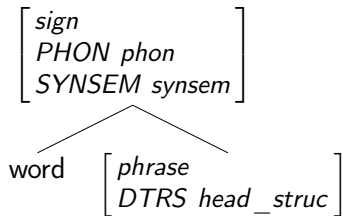
- Erinevused

- ▶ Puudub igasugune transformatsiooni mõiste
- ▶ Tasandid on teisiti jaotatud
- ▶ 'radical nonautonomy' - ükski grammatilise teadmiste tasand pole privilegeeritud seisuses
- ▶ HPSG on andmetele orienteeritud, GB teooriale orienteeritud.

Märgi struktuur

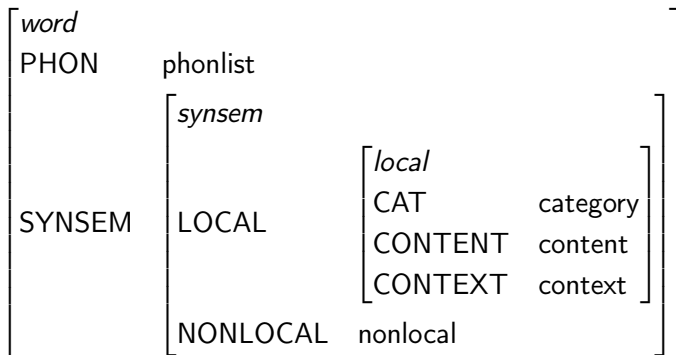
- Märk *Sign* - lause, sõna, fraas, mitmelauselised diskursused (analüüsiühik).
- Märk - vormi ja tähenduse paar *ehk*
- Märk on struktuur, mis sisaldab fonoloogilist, süntaktilist, semantilist, diskursuse ja fraasistruktuuri informatsiooni.
- Märgil on vähemalt kaks atribuuti:
 - ▶ PHON - kirjeldab märgi foneetilist sisu;
 - ▶ SYNSEM - kirjeldab süntaktilist ja semantilist informatsiooni. Jaguneb omakorda:
 - ★ LOCAL (LOC)-CATEGORY, CONTENT, CONTEXT
 - ★ NONLOCAL (NONLOC) -kasutatakse määramata liikumiste kirjeldamisel

Märgi struktuurist veel



Kõik objektid tüüpi *sign* peavad olema kas tüüpi *word* või tüüpi *phrase*.
Kõik objektid tüüpi *sign* on atribuutidega PHON ja SYNSEM, ja lisaks on tüüpi *phrase* objektidel atribuut DTRS.

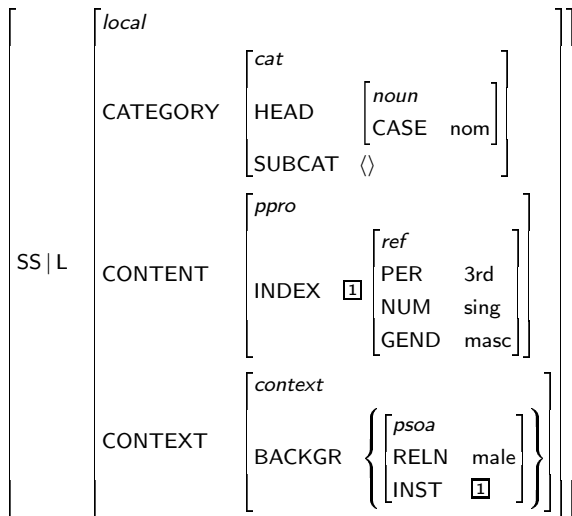
Üldine pilt



LOCAL

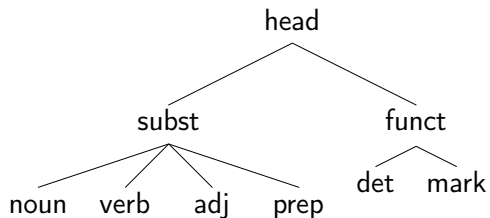
- CATEGORY tunnusel on 2 atribuuti: HEAD ja SUBCAT
- SUBCATi asemel kasutatakse ka kolme valentsitunnust: SUBJ (subject), COMPS (complements) ja SPR (specifier).
- CONTENT kannab semantilist informatsiooni
- CONTEXT sisaldab kontekstitundlikku lingvistilist informatsiooni

LOCAL informatsioon asesõnale *he*



CATEGORY

HEAD tunnus näitab sõnaliiki. Tüübistruktuur:

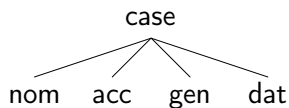


NOUN

[*CASE case*]

Käände tüübi jagunemine (inglise k)

HEAD informatsioon asesõnale *he*



PREPOSITION

[*PFORM pform*] Inglise keele eessõna jagunemine:

pform

to for by of ...

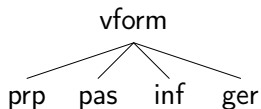
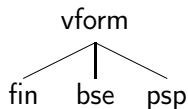
```
graph TD; pform --> to; pform --> for; pform --> by; pform --> of; pform --> dots[...]
```

Verb *depend* vajab eessõna *on*:

[*PFORM on*]

VERB

$\left[\begin{array}{l} \text{VFORM } \textit{vform} \\ \text{AUX } \textit{boolean} \\ \text{INV } \textit{boolean} \end{array} \right]$



Verbivorm *gives*

$\left[\text{HEAD } \left[\begin{array}{l} \textit{verb} \\ \text{VFORM } \textit{fin} \\ \text{AUX } - \end{array} \right] \right]$

Verbivorm *has*

$\left[\text{HEAD } \left[\begin{array}{l} \textit{verb} \\ \text{VFORM } \textit{fin} \\ \text{AUX } + \\ \text{INV } \textit{boolean} \end{array} \right] \right]$

SUBCAT

- SUBCAT on tunnus CATEGORY väärtuses
- subkategorisatsioon on spetsifikatsioon, mis näitab selle märgiga seotud teiste elementide arvu ja vormi. Eesti keele grammatikatradiitsioonis rektsioon + ...

magab: SUBCAT \langle NP[nom] \rangle

armastab: SUBCAT \langle NP[nom], NP[part] \rangle

mõtleb SUBCAT \langle NP[nom], NP[elat] | PP[üle] | S[et] \rangle

SUBCAT

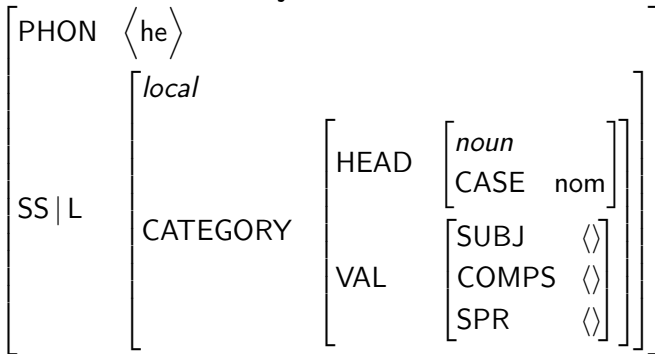
- SUBCAT on tunnus CATEGORY väärtuses
- subkategorisatsioon on spetsifikatsioon, mis näitab selle märgiga seotud teiste elementide arvu ja vormi. Eesti keele grammatikatraditsioonis rektsioon + ...
- Elementide järjekord freimis väljendab grammatiliste seoste olulisust, mitte pindmist laiendite järjekorda

SUBJECT \Rightarrow *DIRECTOBJECT* \Rightarrow *INDIRECTOBJECT* \Rightarrow
OBLIQUES \Rightarrow *GENITIVES* \Rightarrow *OBJECTSOFCOMPARISON*

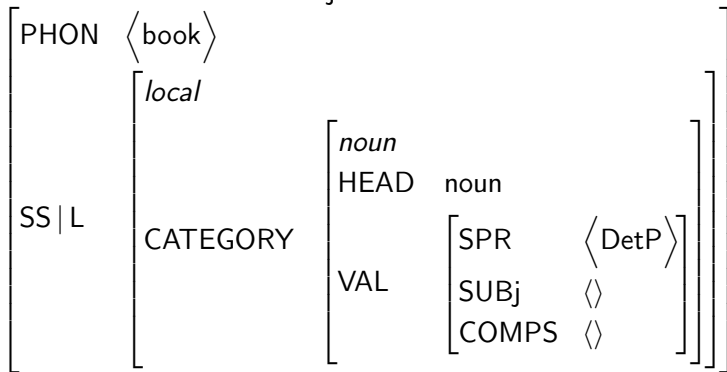
SUBCAT

- SUBCAT on tunnus CATEGORY väärtuses
- subkategorisatsioon on spetsifikatsioon, mis näitab selle märgiga seotud teiste elementide arvu ja vormi. Eesti keele grammatikatradsioonis rektsioon + ...
- Elementide järjekord freimis väljendab grammatiliste seoste olulisust, mitte pindmist laiendite järjekorda
- Hiljem SUBCAT tunnus asendati kolme valentsi tunnusega SUBJ, COMP ja SPR. See lubab paremini kirjeldada kategooriaid, mis vajavad laiendit, kuid mitte alust. Määratlejad eristatakse alusest, sest nad võivad mõlemad esineda sama põhja juures (öeldistäitega lausetes).

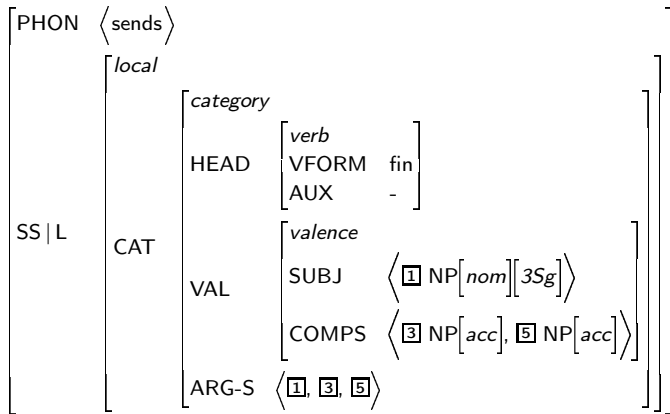
Asesõna *he* leksikonikirje



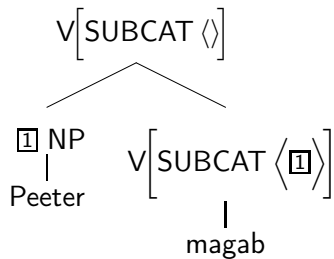
Nimisõna *book* leksikonikirje



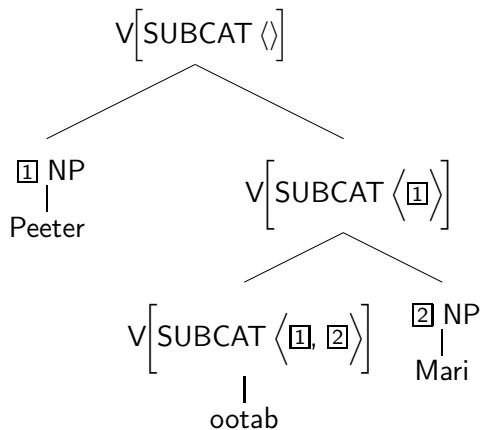
Verb *sends*



Eestikeelsed näited



Transitiivne verb



Valentsireeglid HPSGs

- Spetsiifiline reegel verbi põhi-argumentseose kohta:

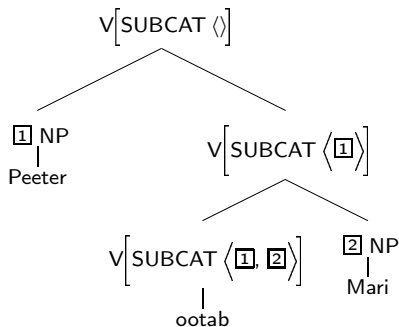
$$V[\text{SUBCAT } \boxed{A}] \rightarrow \boxed{A}, V[\text{SUBCAT } \boxed{A} \oplus \langle \boxed{B} \rangle]$$

- kus \oplus on otsekorrutis:

$$(a) \oplus (b) = (a, b)$$

$$$(a, b) \oplus () = (a, b)$$$

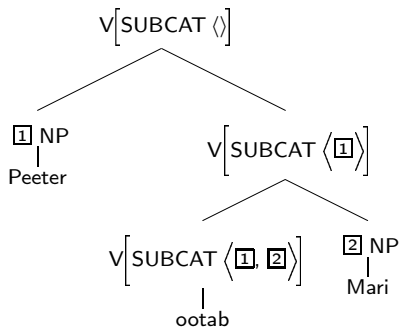
Põhi-argumentreeglite näide



VP2- \rightarrow NP,VP

$$V[\text{SUBCAT } \boxed{A}] \rightarrow \boxed{A}, V[\text{SUBCAT } \boxed{A} \oplus \langle \boxed{B} \rangle] \quad \boxed{A} = (), \boxed{B} = 1$$

Põhi-argumentreeglite näide



VP- \rightarrow V NP

$$V[\text{SUBCAT } \boxed{A}] \rightarrow \boxed{A}, V[\text{SUBCAT } \boxed{A} \oplus \langle \boxed{B} \rangle] \quad \boxed{A} = \boxed{1}, \boxed{B} = \boxed{2}$$

Põhi-argumentreeglite üldistamine

$$\begin{aligned}V[SUBCAT \boxed{A}] &\rightarrow \boxed{A}, V[SUBCAT \boxed{A} \oplus \langle \boxed{B} \rangle] \\A[SUBCAT \boxed{A}] &\rightarrow \boxed{A}, A[SUBCAT \boxed{A} \oplus \langle \boxed{B} \rangle] \\N[SUBCAT \boxed{A}] &\rightarrow \boxed{A}, N[SUBCAT \boxed{A} \oplus \langle \boxed{B} \rangle] \\P[SUBCAT \boxed{A}] &\rightarrow P[SUBCAT \boxed{A} \oplus \langle \boxed{B} \rangle], \boxed{A} \\H[SUBCAT \boxed{A}] &\rightarrow \begin{array}{c} \downarrow \\ H[SUBCAT \boxed{A} \oplus \langle \boxed{B} \rangle], \boxed{A} \end{array}\end{aligned}$$

Fraasid

Fraaside tunnusstruktuuril on atribuut DAUGHTERS (DTRS), mille väärtuseks on struktuur tüübist *constituent-structure* (*con-struct*), mis esitab fraasi vahetuid allujaid.

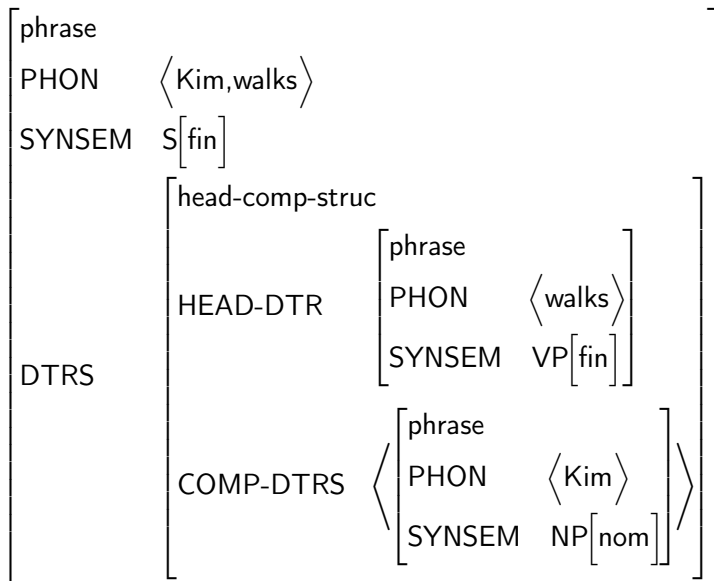
con-struct tüübil on palju alamtüüpe erinevat liiki tütartippude kirjeldamiseks.

Olulisim neist on *headed-structure* (*head-struct*), mida kasutatakse kõigis põhjaga konstruktsioonides.

- **HEAD-DAUGHTER** (HEAD-DTR)
- **COMPLEMENT-DAUGHTERS** (COMP-DTRS)
- ADJUNCT-DAUGHTER
- FILLER-DAUGHTER
- MARKER-DAUGHTER

[head-struct]
	HEAD-DTR (a sign)	
	COMP-DTRS (a list of signs)	

Lause näide



Subkategorisatsiooni printsiip

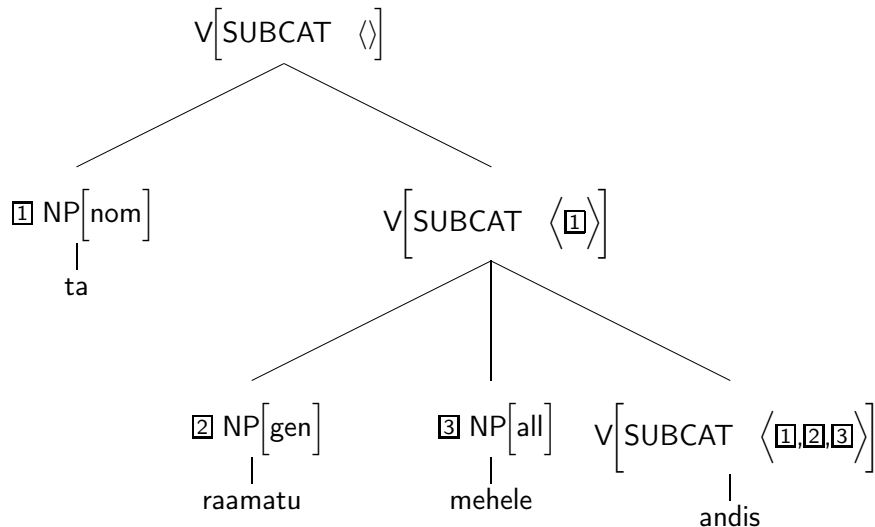
Printsiip

In headed phrase (i.e. a phrasal sign whose DTRs value is of sort head-struct), the SUBCAT value of the head daughter is the concatenation of the phrase's SUBCAT list with the list of SYNSEM values of the complement daughters.

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

$$\left[\begin{array}{l} \text{SUBCAT} \\ \text{HEAD-DTR} \\ \text{COMPS} \end{array} \right. \left. \begin{array}{l} \boxed{1} \\ \left[\text{SUBCAT } \boxed{1} \oplus \langle \boxed{2} \rangle \right] \\ \left[\text{SUBCAT } \langle \boxed{2} \rangle \right] \end{array} \right]$$

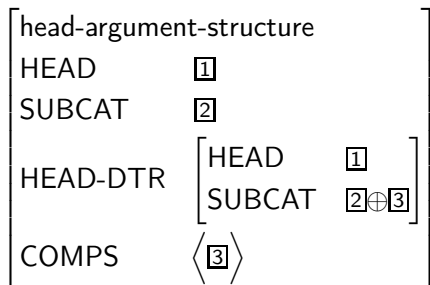
Ta andis mehele raamatu



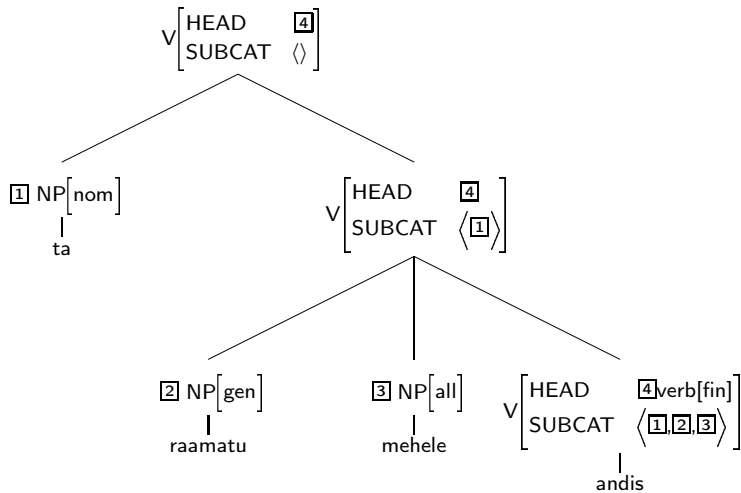
Põhja tunnuse printsiip (HFP)

Printsiip

The HEAD value of any headed phrase is structure-shared with the HEAD value of the head daughter.



Näide põhjaprintsiibist ja subkategorisatsioonist



Kokkuvõte

- KVG: Kogu info, mis ühes sõlmes vaja oli, tulenes otseselt reeglist.
- FSG: Nii reeglid, leksikon kui printsiibid kirjeldavad grammatikat osaliselt
- Tunnusstruktuurid on keeleüksuste mudelid, mis kirjeldavad täielikult antud üksust. Konkreetne kirjeldus võib olla poolik.
- Grammatika on hea, kui tunnusstruktuurid vastavad reaalse maailma üksustele.
- Esitusviisid: tunnusväärtusmaatriksid (AVM), kastdiagramm
- Grammatika ülesanne: kuidas leksikoni üksused on seotud sõna tunnusstruktuuriga; kuidas väiksematest üksustest saab kokku panna suuremaid
- Põhja, alamkategoriate ja tunnusstruktuuri abil saadud grammatika on võrreldes KVGga abstraktsem ja üldistavam.