

Süntaksiteooriad ja -mudelid

MTAT.06.031

4. loeng

Kaili Müürisep
kaili.muurisep@ut.ee

ATI

6. märts 2008

- 1 Kontekstivabade grammatikate kokkuvõte
- 2 Tunnusstruktuurid
 - Unifitseerimine
- 3 HPSG
 - Suur leksikoni osatähtsus
 - Konstruktsioonid
- 4 Võrdlus GB-ga
- 5 Märgi struktuur
 - PHON ja SYNSEM
 - Sõna ja fraas
 - LOCAL
 - CATEGORY
 - SUBCAT
 - Näited VAL kohta leksikonist: he, book, sends
- 6 Lause esitus HPSG-s
 - Fraasid
 - Lause näide: Kim walks.
 - Subkategorisatsiooni printsiip
 - Põhja tunnuse printsiip

KVG probleemid

- 1 Reeglitel ei ole kitsendusi. $NP \rightarrow Adv AP$
- 2 Pole üldiseid reegleid
- 3 Ühildumist ja subkategorisatsiooni pole võimalik mõistlikult edasi anda.

$S \rightarrow NP-SG VP-SG;$

$VP-SG \rightarrow TV-SG NP-SG;$

$VP-SG \rightarrow TV-SG NP-PL;$

...

Head küljed

- Lihtne formalism, mis suudab genereerida lõpmatut hulka keeli ja omistada neile mõistliku lingvistilise struktuuri.
- Lingvistilised nähtused, mida ei saa KVG abil kirjeldada, on haruldased.
- Arvutuslikult lihtne ja head meetodid on välja töötatud.

Atomaarsed vrs strukturaarsed kategooriad

Siiani esitati grammatilist kategooriat atomaarsena, millel puudus igasugune sisemine struktuur.

Erand oli DCG:

Atomaarsed vrs strukturaarsed kategooriad

Siiani esitati grammatilist kategooriat atomaarsena, millel puudus igasugune sisemine struktuur.

Erand oli DCG:

DCG

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

See on samaväärne:

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

Tunnusstruktuurid

Tunnusstruktuur koosneb tunnus-väärtus-paaridest.

Tunnusstruktuur

$$\left[\begin{array}{ll} \text{FEATURE}_1 & \text{VALUE}_1 \\ \text{FEATURE}_2 & \text{VALUE}_2 \\ \dots & \\ \text{FEATURE}_N & \text{VALUE}_N \end{array} \right]$$

Näide

$$\left\langle \text{lind}, \left[\begin{array}{ll} \text{POS} & \text{noun} \\ \text{NUM} & \text{sg} \end{array} \right] \right\rangle$$

Tunnuste tüübid

Kui me vaatame keelt analüüsidest mõisteid fraas, sõna, kategooria jne, siis tüübid kirjeldavad neile vastavaid tunnusstruktuure.

Tunnuste tüübid

Kui me vaatame keelt analüüsidest mõisteid fraas, sõna, kategooria jne, siis tüübid kirjeldavad neile vastavaid tunnusstruktuure.

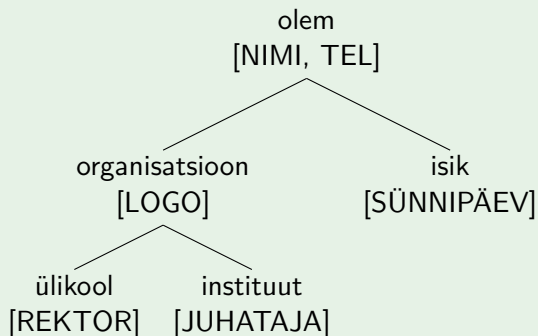
Tartu ülikool

- [NIMI TARTU ÜLIKOOL
TEL 737 5100]
- [NIMI ARVUTITEADUSE INSTITUUT
TEL 737 5]
- [NIMI KAILI MÜÜRISSEP
TEL 737 5 ...]

Tunnuste tüübid

Kui me vaatame keelt analüüsidest mõisteid fraas, sõna, kategooria jne, siis tüübid kirjeldavad neile vastavaid tunnustruktuure.

Tüübihierarhia



Tunnuste tüübid

Kui me vaatame keelt analüüsidest mõisteid fraas, sõna, kategooria jne, siis tüübid kirjeldavaid neile vastavaid tunnusstruktuure.

Tüübid

- [ülikool
NIMI TARTU ÜLIKOOL
TEL 737 5100
REKTOR ALAR KARIS]
- [instituut
NIMI ARVUTITEADUSE INSTITUUT
TEL 737 5445
JUHATAJA TIIT ROOSMAA]
- [isik
NIMI KAILI MÜÜRISSEP
TEL 737 5 ...
SÜNNIPÄEV 06.03.1986]

Tunnuste tüübid

Kui me vaatame keelt analüüsidest mõisteid fraas, sõna, kategooria jne, siis tüübid kirjeldavad neile vastavaid tunnustruktuure.

ülikool		
NIMI	TARTU ÜLIKOOL	
TEL	737 5100	
REKTOR	[isik	
	NIMI	ALAR KARIS
	TEL	...
	SÜNNIPÄEV	...

Rekursiivsed tunnusstruktuurid

Genealoogia

isik		
EESNIMI	ANU	
PEREKONNANIMI	TAMM	
SÜNNIAEG	06.03.08	
ISA		
	isik	
	EESNIMI	JAAN
	PEREKONNANIMI	TAMM
	SÜNNIAEG	06.03.68
	ISA	...

Tunnusstruktuurid grammatikas

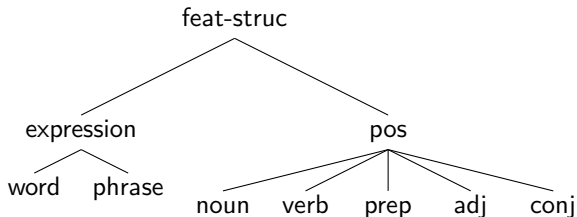
Fraas ja Sõnad leksikonis

$\left[\begin{array}{l} \text{phrase} \\ \text{HEAD} \quad \text{noun} \end{array} \right] \left\langle \text{bird}, \left[\begin{array}{l} \text{word} \\ \text{HEAD} \quad \text{noun} \end{array} \right] \right\rangle$

Tunnusstruktuurid grammatikas

Fraas ja Sõnad leksikonis

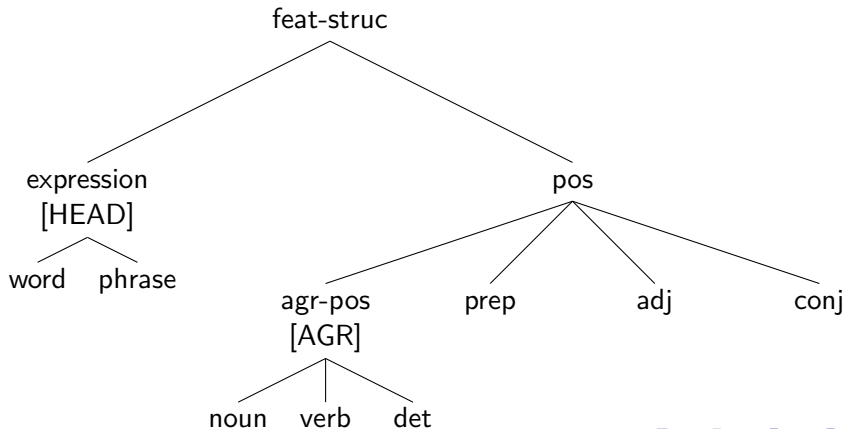
$\left[\begin{array}{l} \text{phrase} \\ \text{HEAD} \quad \text{noun} \end{array} \right] \left\langle \text{bird}, \left[\begin{array}{l} \text{word} \\ \text{HEAD} \quad \text{noun} \end{array} \right] \right\rangle$



Tunnusstruktuurid grammatikas

Fraas ja Sõnad leksikonis

$\left[\begin{array}{l} \text{phrase} \\ \text{HEAD} \quad \text{noun} \end{array} \right] \left\langle \text{bird}, \left[\begin{array}{l} \text{word} \\ \text{HEAD} \quad \text{noun} \end{array} \right] \right\rangle$



Veel tunnuste struktuuridest: subsumptsioon

- FS1: $\left[\text{CAT} \quad \text{np} \right]$
- FS2: $\left[\right]$
- FS3: $\left[\begin{array}{l} \text{CAT} \quad \text{np} \\ \text{ARG} \quad \left[\begin{array}{l} \text{NUM} \quad \text{sg} \\ \text{PERS} \quad \text{3rd} \end{array} \right] \end{array} \right]$
- FS4: $\left[\begin{array}{l} \text{CAT} \quad \text{np} \\ \text{ARG} \quad \left[\text{NUM} \quad \text{sg} \right] \end{array} \right]$

Veel tunnuste struktuuridest: subsumptsioon

- FS1: $\left[\begin{array}{l} \text{CAT} \quad \text{np} \end{array} \right]$
- FS2: $\left[\right]$
- FS3: $\left[\begin{array}{l} \text{CAT} \quad \text{np} \\ \text{ARG} \quad \left[\begin{array}{l} \text{NUM} \quad \text{sg} \\ \text{PERS} \quad \text{3rd} \end{array} \right] \end{array} \right]$
- FS4: $\left[\begin{array}{l} \text{CAT} \quad \text{np} \\ \text{ARG} \quad \left[\text{NUM} \quad \text{sg} \right] \end{array} \right]$

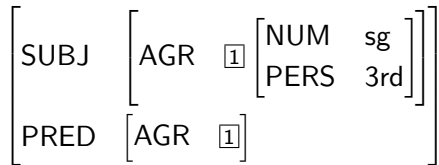
$FS2 \sqsubseteq FS1 \sqsubseteq FS4 \sqsubseteq FS3$

Definitsioon

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



Unifitseerimine

- Tunnuste struktuurid võivad sisaldada erinevat, kuid siiski ühilduvat informatsiooni

- ▶ FS5: $\left[\begin{array}{l} \text{CAT} \qquad \text{np} \\ \text{ARG} \left[\text{NUM} \quad \text{sg} \right] \end{array} \right]$
- ▶ FS6: $\left[\begin{array}{l} \text{CAT} \qquad \text{np} \\ \text{ARG} \left[\text{PERS} \quad \text{3rd} \right] \end{array} \right]$

Unifitseerimine

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

▶ FS7: $\left[\begin{array}{ll} \text{CAT} & \text{np} \\ \text{ARG} & \left[\text{NUM sg} \right] \end{array} \right]$

▶ FS8: $\left[\begin{array}{ll} \text{CAT} & \text{np} \\ \text{ARG} & \left[\text{NUM pl} \right] \end{array} \right]$

Unifitseerimine

- Tunnuste struktuurid võivad sisaldada erinevat, kuid siiski ühilduvat informatsiooni

▶ FS5: $\left[\begin{array}{l} \text{CAT} \quad \text{np} \\ \text{ARG} \left[\text{NUM} \quad \text{sg} \right] \end{array} \right]$

▶ FS6: $\left[\begin{array}{l} \text{CAT} \quad \text{np} \\ \text{ARG} \left[\text{PERS} \quad \text{3rd} \right] \end{array} \right]$

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

▶ FS9: $\left[\begin{array}{l} \text{CAT} \quad \text{np} \\ \text{ARG} \left[\begin{array}{l} \text{NUM} \quad \text{sg} \\ \text{PERS} \quad \text{3rd} \end{array} \right] \end{array} \right]$

Unifitseerimine

- Tunnuste struktuurid võivad sisaldada erinevat, kuid siiski ühilduvat informatsiooni

▶ FS5: $\left[\begin{array}{l} \text{CAT} \quad \text{np} \\ \text{ARG} \left[\begin{array}{l} \text{NUM} \quad \text{sg} \end{array} \right] \end{array} \right]$

▶ FS6: $\left[\begin{array}{l} \text{CAT} \quad \text{np} \\ \text{ARG} \left[\begin{array}{l} \text{PERS} \quad \text{3rd} \end{array} \right] \end{array} \right]$

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

▶ FS9: $\left[\begin{array}{l} \text{CAT} \quad \text{np} \\ \text{ARG} \left[\begin{array}{l} \text{NUM} \quad \text{sg} \\ \text{PERS} \quad \text{3rd} \end{array} \right] \end{array} \right]$

Head-driven Phrase Structure Grammar

- Pollard and Sag 1987, 1994
- Sag and Wasow 1999

- Leksikaliseeritud
- Pindmised struktuurid
- Ühetasandiline
- Hierarhiline organiseeritus
- Tüpiseeritud tunnuste struktuurid (leksikon, morfoloogia, fraasid, printsiibid)
- Unifitseerimine/Pärimine
- Fonoloogia, süntaks, semantika ühendatud

- 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 konstruktsioonid vajavad eraldi reegleid:

- Relatiivlauseid
- Noun-noun compounds
- Lisand

ERGi on 105 reeglit

Võrdlus GB-ga

- Sarnasused

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

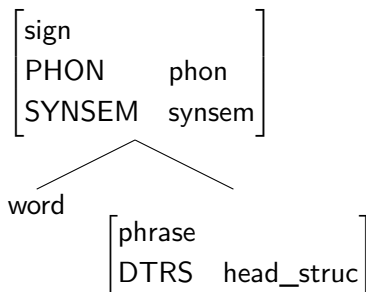
- Erinevused

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

Märgi struktuur

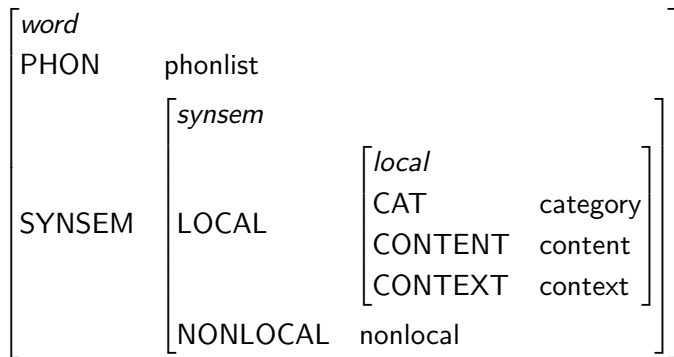
- Märk *Sign* - lause, sõna, fraas, mitmelauselised diskursused (analüüsiühik).
- Märk - vormi ja tähenduse paar *ehk*
- Märk on struktuur, mis sisaldab fonoloogilist, süntaktilist, semantilist, diskursuse ja fraasistruktuuri informatsiooni.
- Märgil on vähemalt kaks atribuuti:
 - ▶ PHON - kirjeldab 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.

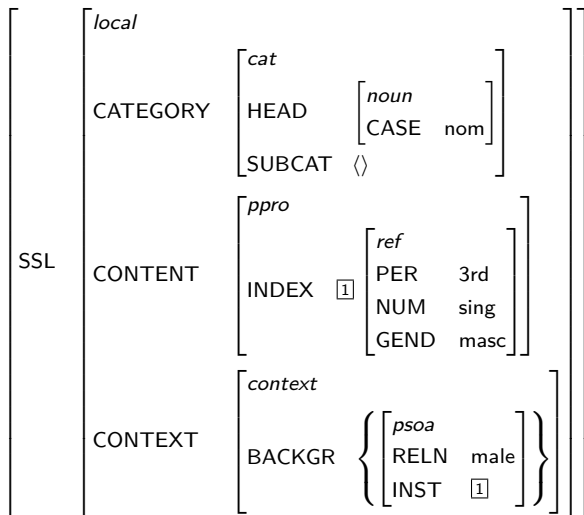
Üldine pilt



LOCAL

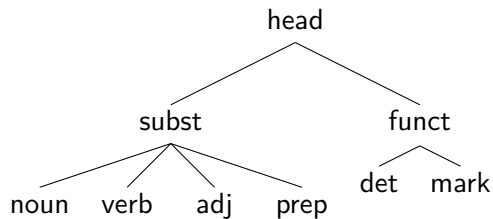
- CATEGORY tunnusel on 2 atribuuti: HEAD ja SUBCAT
- SUBCATi asemel kasutatakse ka kolme valentsitunnust: SUBJ (subject), COMPS (complements) ja SPR (specifier).
- SUBCAT on uuemal ajal ARG-S
- CONTENT kannab semantilist informatsiooni
- CONTEXT sisaldab kontekstitundlikku lingvistilist informatsiooni

LOCAL informatsioon asesõnale *he*



CATEGORY

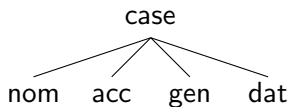
HEAD tunnus näitab sõnaliiki. Tüübistruktuur:



NOUN

[CASE case]

Käände tüübi jagunemine (inglise k)



HEAD informatsioon asesõnale *he*

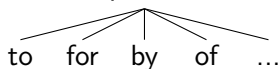


PREPOSITION

[PFORM pform]

Inglise keele eessõna jagunemine:

pform

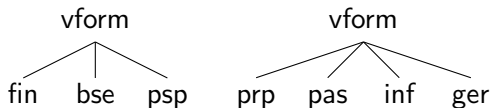


Verb *depend* vajab eessõna *on*:

[PFORM on]

VERB

VFORM	vform
AUX	boolean
INV	boolean



Verbivorm *gives*

HEAD	<table><tr><td><i>verb</i></td></tr><tr><td>VFORM</td><td>fin</td></tr><tr><td>AUX</td><td>-</td></tr></table>	<i>verb</i>	VFORM	fin	AUX	-
<i>verb</i>						
VFORM	fin					
AUX	-					

Verbivorm *has*

HEAD	<table><tr><td><i>verb</i></td></tr><tr><td>VFORM</td><td>fin</td></tr><tr><td>AUX</td><td>+</td></tr><tr><td>INV</td><td>boolean</td></tr></table>	<i>verb</i>	VFORM	fin	AUX	+	INV	boolean
<i>verb</i>								
VFORM	fin							
AUX	+							
INV	boolean							

SUBCAT

- SUBCAT on tunnus CATEGORY väärtuses

SUBCAT

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

magab: SUBCAT⟨NP[nom]⟩

armastab: SUBCAT⟨NP[nom], NP[part]⟩

mõtleb: SUBCAT⟨NP[nom], NP[elat] | PP [üle] | S [et]⟩

SUBCAT

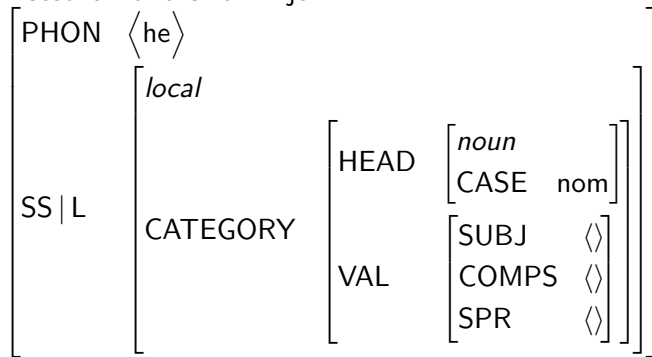
- SUBCAT on tunnus CATEGORY väärtuses
- subkategorisatsioon on spetsifikatsioon, mis näitab selle märgiga seotud teiste elementide arvu ja vormi. Eesti keele grammatikatraditsioonis rektsioon + ...
- Elementide järjekord freimis väljendab grammatiliste seoste olulisust, mitte pindmist laiendite järjekorda

SUBJECT ⇒ DIRECT OBJECT ⇒ INDIRECT OBJECT ⇒ OBLIQUES ⇒ GENITIVES ⇒ OBJECTS OF COMPARISON

SUBCAT

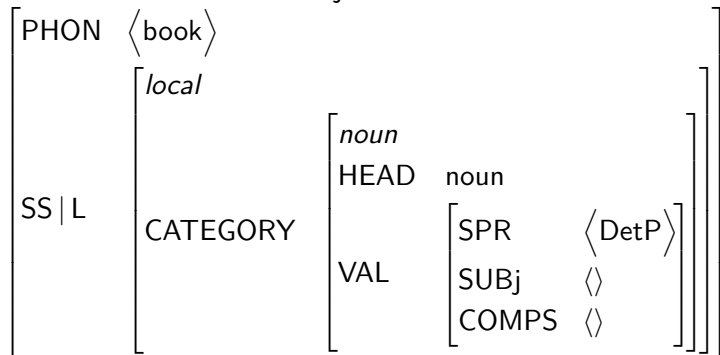
- SUBCAT on tunnus CATEGORY väärtuses
- subkategorisatsioon on spetsifikatsioon, mis näitab selle märgiga seotud teiste elementide arvu ja vormi. Eesti keele grammatikaraditsioonis rektsioon + ...
- Elementide järjekord freimis väljendab grammatiliste seoste olulisust, mitte pindmist laiendite järjekorda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

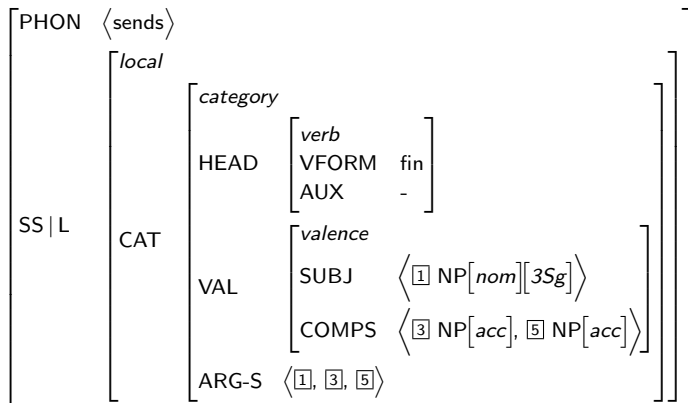


book

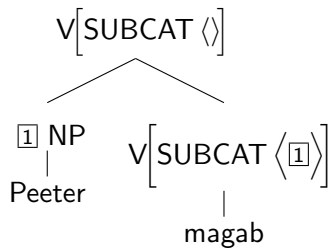
Nimisõna *book* leksikonikirje



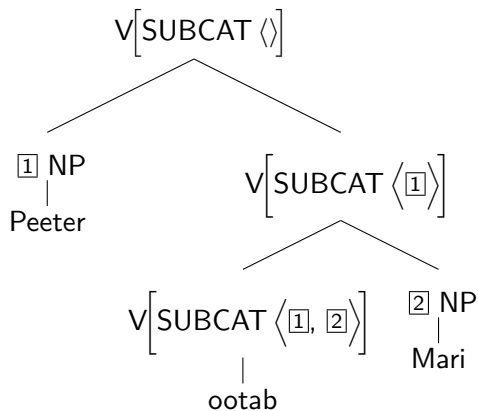
Verb *sends*



Eestikeelsed näited



Transitiivne verb



Valentsireeglid HPSGs

- Spetsiifiline reegel verbi põhi-argumentseose kohta:

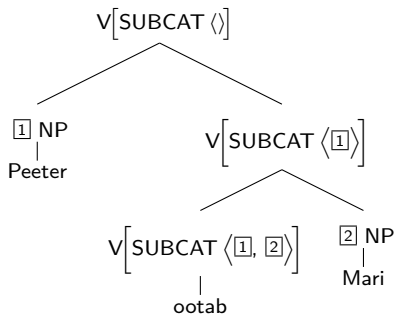
$$V[\text{SUBCAT } \boxed{A}] \rightarrow \boxed{A}, V[\text{SUBCAT } \boxed{A} \oplus \langle \boxed{B} \rangle]$$

- kus \oplus on otsekorrutis:

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

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

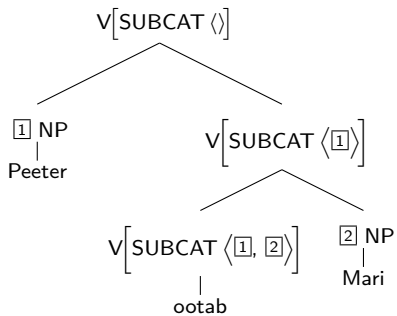
Põhi-argumentreeglite näide



$VP2 \Rightarrow NP, VP$

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

Põhi-argumentreeglite näide



VP->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

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

Fraasid

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

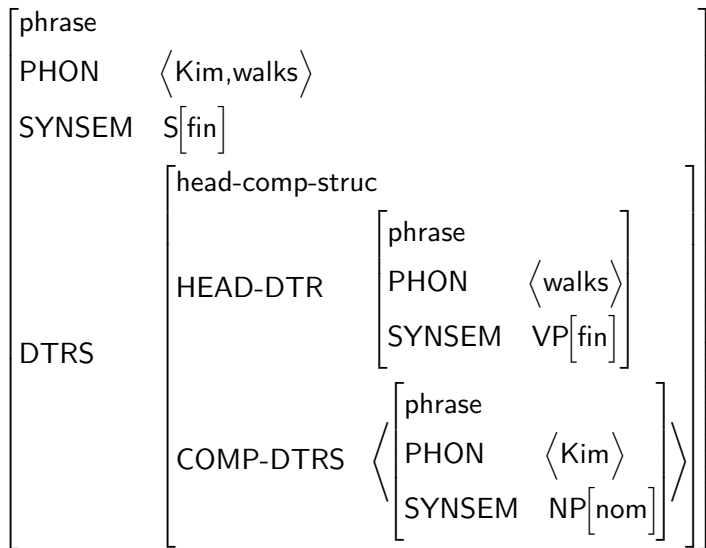
con-struct tüübil on palju alamtüüpe erinevat liiki tütartippude kirjeldamiseks.

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

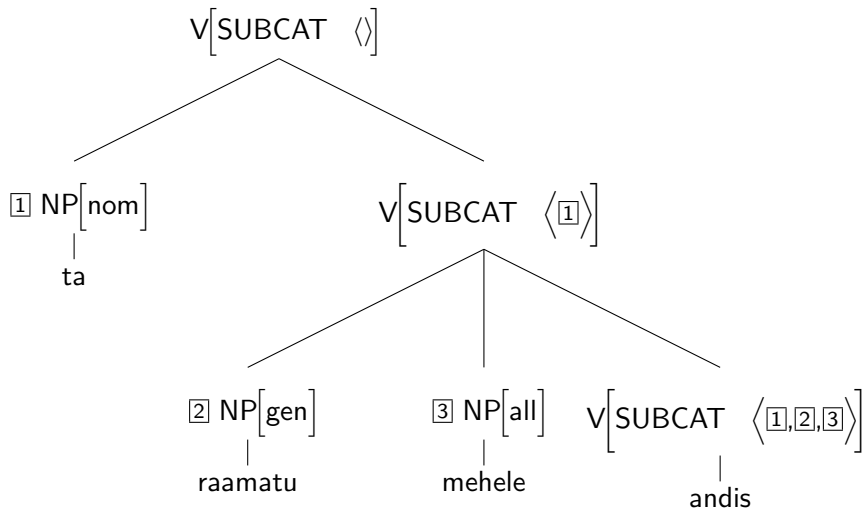
Printsiip

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.

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

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

Ta andis mehele raamatu



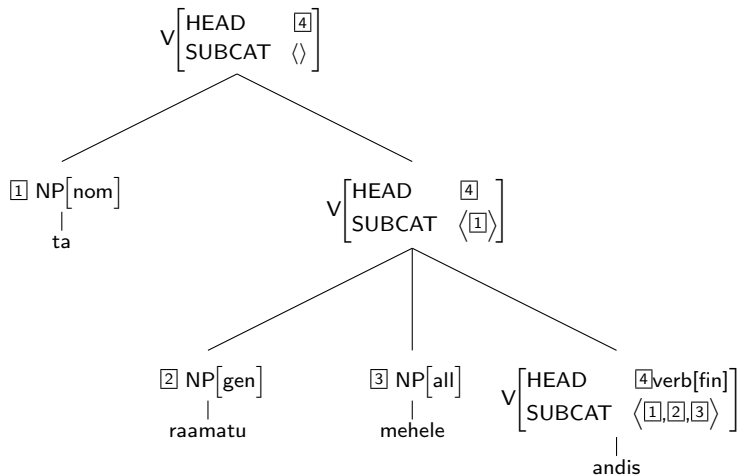
Põhja tunnuse printsiip (HFP)

Printsiip

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

head-argument-structure					
HEAD	[1]				
SUBCAT	[2]				
HEAD-DTR	[<table><tr><td>HEAD</td><td>[1]</td></tr><tr><td>SUBCAT</td><td>[2 ⊕ 3]</td></tr></table>]	HEAD	[1]	SUBCAT	[2 ⊕ 3]
HEAD	[1]				
SUBCAT	[2 ⊕ 3]				
COMPS	⟨[3]⟩				

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.