

HPSG tutvustus I

Süntaksiteooriad ja -mudelid 2005/06

Kaili Müürisep

9. märts 2006

"there are more people working on and with implemented head-driven phrase structure grammars than with any other linguistic grammar model" (Uzkoreit, 1996).

Contents

1	Ülevaade	1
1.1	Kontekstivabade grammatikate puudused	1
1.2	Tunnusstruktuurid	1
1.3	Tüübihierarhia	3
1.4	Unifitseerimine	3
1.5	Suur leksikoni osatähtsus	4
1.6	Konstruksioonid	4
2	Võrdlus GB-ga	4
3	Märgi struktuur	5
3.1	PHON ja SYNSEM	5
3.2	Sõna ja fraas	5
3.3	LOCAL	6
3.4	CATEGORY	6
3.5	SUBCAT	7
3.6	Näited VAL kohta leksikonist: he, book, sends	8
4	Lause esitus HPSG-s	10
4.1	Fraasid	10
4.2	Lause näide: Kim walks.	11
4.3	Subkategorisatsiooni printsiip	11
4.4	Põhja tunnuse printsiip	12
5	Kokkuvõte	13

1 Ülevaade

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

1.1 Kontekstivabade grammatikate puudused

- Väga palju reegleid
- Reeglitel pole kitsendusi $NP \rightarrow Adv AP$
- Pole üldiseid reegleid (Mida ühist on V-intrans-sg ja V-tr-pl?)

1.2 Tunnusstruktuurid

KVG-s on iga sümbol atomaarne, ei saa eristada selle sisemist struktuuri. Et seda eristada, kasutatakse tunnusstruktuure.

Tunnusstruktuur:

Intuitiivselt annab edasi grammatilist infot

Formaalselt tunnuste hulk, kus iga tunnuse juures on antud tema väärtus

Praktiliselt kirjutatakse tunnusstruktuurid üles kandiliste sulgude vahel

$$\left[\begin{array}{ll} \text{TUNNUS}_1 & \text{VÄÄRTUS}_1 \\ \text{TUNNUS}_2 & \text{VÄÄRTUS}_2 \end{array} \right]$$

- Fraasstruktuurireeglite tipud on tunnus-väärtus paaride kogumid.

DCG

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

See on samaväärne:

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

- Tunnused võivad olla ka keerukamad

slash-tunnus DCG-s

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

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

- Väärtustatakse ainult need tunnused, mis olulised:

DCG

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

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

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

Definition 1. 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

SUBJ	[AGR	[1	[NUM	sg
			[PERS	3rd]
PRED	[AGR	[1]		

1.3 Tüübihierarhia

Igal kategoorial on omad relevantssed tunnused, selle alusel defineeritakse grammatikas tüübid (i.k. *sorts*) Nt. verbil on pöördelõpp, isik ja arv, nimisõnal kääne ja arv.

- 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

1.4 Unifitseerimine

- Tunnuste struktuurid võivad sisaldada erinevat, kuid siiski ühilduvat informatsiooni

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

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

- Aga ka erinevat ja mitteühilduvat informatsiooni

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

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

- Kui kaks struktuuri ühilduvad, siis leidub kolmas, mida nad mõlemad laiendavad.

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

•

Definition 2. Kahe tunnuste struktuuri A ja B *unifikatsiooniks* nimetatakse tunnuste struktuuri C, nii et $A \sqsubseteq C$ ja $B \sqsubseteq C$.

Kategooriad on unifitseeruvad, kui neil pole vastukäivaid tunnuseid ega väärtusi.

1.5 Suur leksikoni osatähtsus

- 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)

1.6 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:

- Relatiivlaused
- Noun-noun compounds
- Lisand

ERGis on 105 reeglit

2 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 monostratal e ühetasandiline, pole derivatsioone, mis teisendaks ühe grammatikastruktuuri teiseks
 - Tasandid on teisiti jaotatud (PHON ja DTRS vastavad PF ja s-struktuurile, SYNSEMile ühest vastet pole)

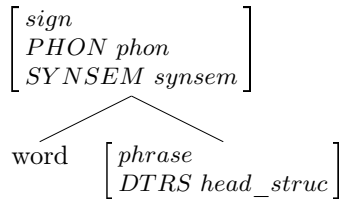
- 'radical nonautonomy' - ükski grammatilise teadmiste tasand pole privilegeeritud seisuses GB-s 'autonomy of syntax'. HPSGs pööratakse suurt tähelepanu grammatilistele liidestele, mille abil erineva grammatilised tasandid üksteist kitsendavad, ilma et ükski neist domineeriks.
- HPSG on andmetele orienteeritud, GB teooriale orienteeritud.

3 Märgi struktuur

3.1 PHON ja SYNSEM

- Märk *Sign* - lause, sõna, fraas, mitmelauseelised diskursused (analüüsiühik).
- Märk - vormi ja tähenduse paar *chk*
- 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

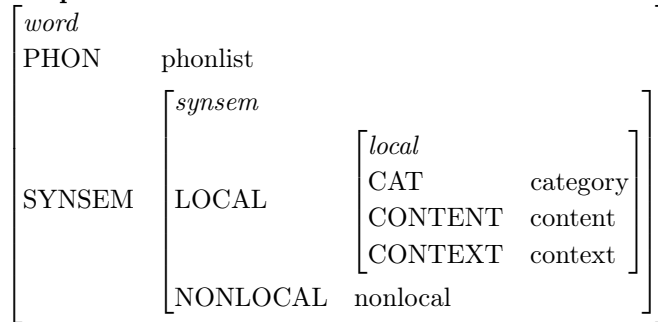
3.2 Sõna ja fraas



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.

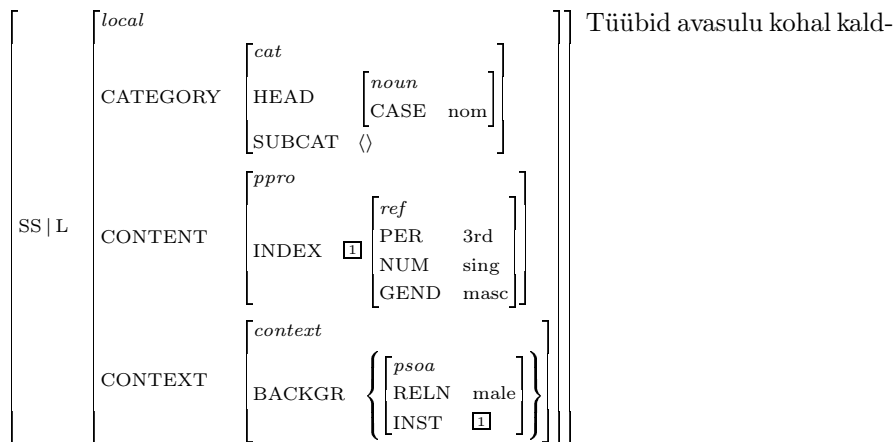
Üldine pilt



3.3 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*



kirjas. Struktuuri jagamise sümbol on number kastis. Hulgad on loogelistes sulgudes, listid kolmnurksulgudes.

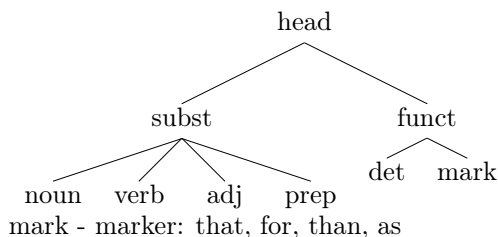
CATEGORY tunnust vaatame edaspidi.

CONTENT: nimisõnadel tunnus INDEX mis näitab referentsust, ppro - personal pronoun

CONTEXT: psoa (parametrized state of affairs) seab piiranguid, kus see (nimi)sõna võib esineda.

3.4 CATEGORY

HEAD tunnus näitab sõnaliiki. Tüübistruktuur:

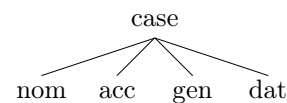


NOUN

[CASE case]

Käände tüübi jagunemine (inglise k)

HEAD informatsioon asesõnale *he*

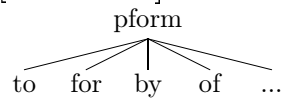


PREPOSITION

[PFORM *pform*]

Verb *depend* vajab eessõna *on*:

[PFORM *on*]



VERB

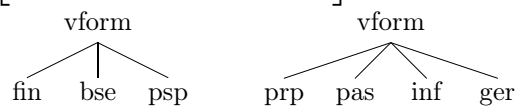
[VFORM *vform*]
[AUX *boolean*]
[INV *boolean*]

Verbivorm *gives*

HEAD [*verb*
VFORM *fin*
AUX *-*]

Verbivorm *has*

HEAD [*verb*
VFORM *fin*
AUX *+*
INV *boolean*]



3.5 SUBCAT

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

magab: SUBCAT⟨NP[nom]⟩

armastab: SUBCAT⟨NP[nom], NP[part]⟩

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

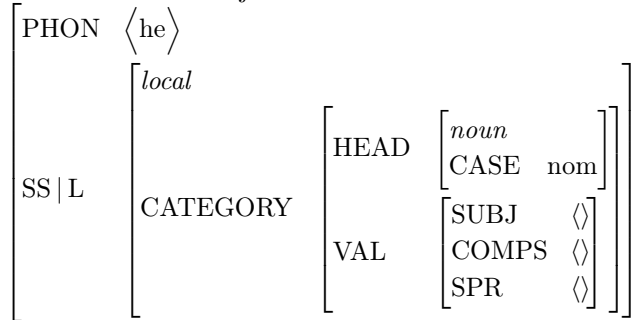
- Elementide järjekord freimis väljendab grammatiliste seoste olulisust, mitte pindmist laiendite järjekorda

SUBJECT ⇒ *DIRECTOBJECT* ⇒ *INDIRECTOBJECT* ⇒
OBLIQUES ⇒ *GENITIVES* ⇒ *OBJECTSOFCOMPARISON*

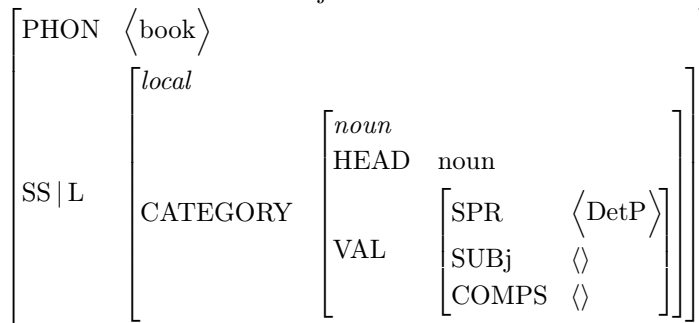
- Hiljem SUBCAT tunnus asendati kolme valentsi tunnusega SUBJ, COMP ja SPR. See lubab paremini kirjeldada kategooriaid, mis vajavad laiendit, kuid mitte alust. Määratlejad eristatakse alusest, sest nad võivad mõlemad esineda sama põhja juures (öeldistäitega lausetes).

3.6 Näited VAL kohta leksikonist: he, book, sends

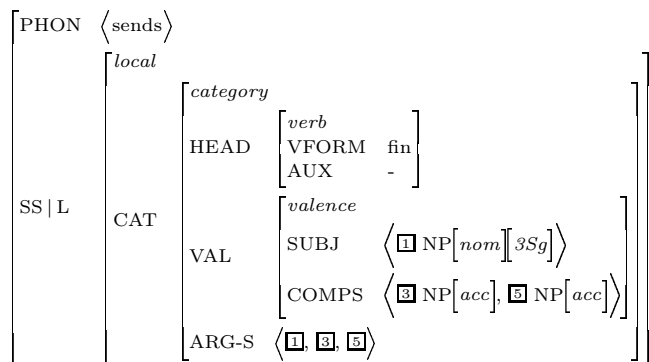
Asesõna *he* leksikonikirje



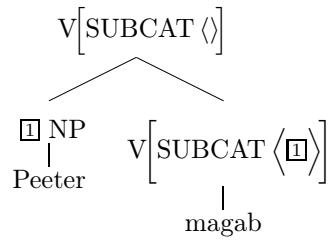
Nimisõna *book* leksikonikirje



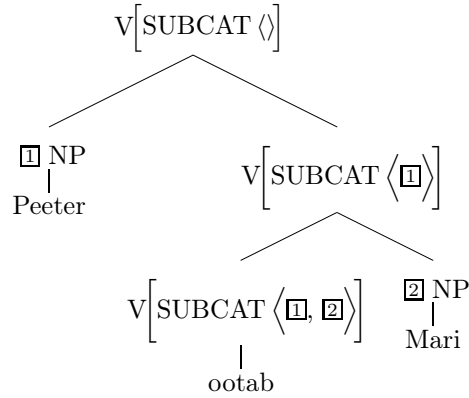
Verb *sends*



Eestikeelsed näited



Transitiivne verb



Valentsireeglid HPSGs

- Spetsiifiline reegel verbi põhi-argumentseose kohta:

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

- kus \oplus on otsekorrutis: $(a) \oplus (b) = (a, b)$ $() \oplus (a, b) = (a, b)$ $(a, b) \oplus () = (a, b)$

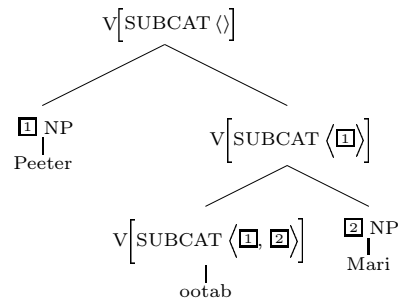
Põhi-argumentreeglite näide

$VP_2 \rightarrow NP, VP$

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

$VP \rightarrow V NP$

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



Põhi-argumentreeglite üldistamine

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

4 Lause esitus HPSG-s

4.1 Fraasid

Fraasid

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

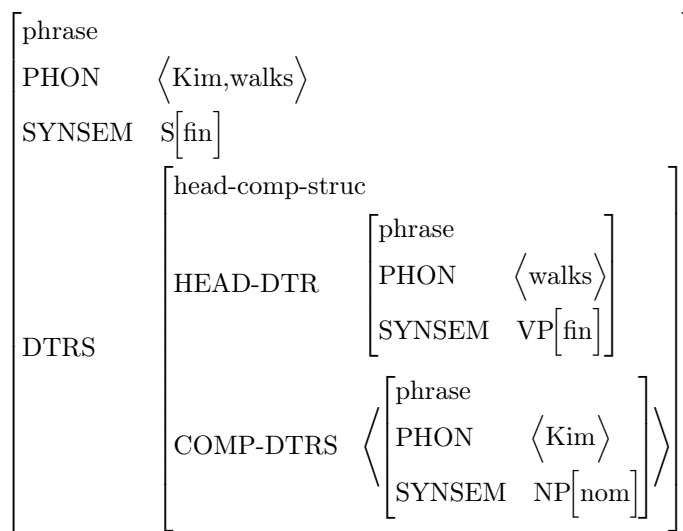
con-struct tüübil on palju alamtüüpe erinevat liiki tütarlippude kirjeldamiseks.

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**

$$\left[\begin{array}{l} \text{head-struct} \\ \text{HEAD-DTR} \quad (\text{a sign}) \\ \text{COMP-DTRS} \quad (\text{a list of signs}) \end{array} \right]$$

4.2 Lause näide: Kim walks.



4.3 Subkategorisatsiooni printiiip

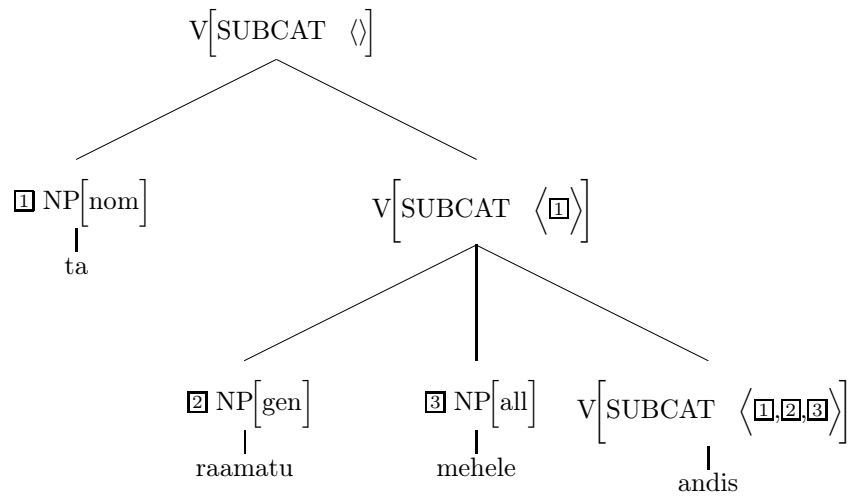
Printsiip. *In headed phrase (i.e. a phrasal sign whose DTRs value is of sort head-struct), the SUBCAT value of the head daughter is the concatenation of the phrase's SUBCAT list with the list of SYNSEM values of the complement daughters.*

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

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

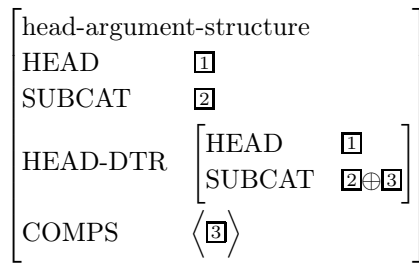
Tuumtütre SUBCAT on ema SUBCATi ja õdede SYNSEMide konkatenatsioon. Kui reeglis pole teisiti märgitud, siis peavad ema SUBCAT väärtused unifitseeruma tuumtütre vastavate tunnustega.

Ta andis mehele raamatu



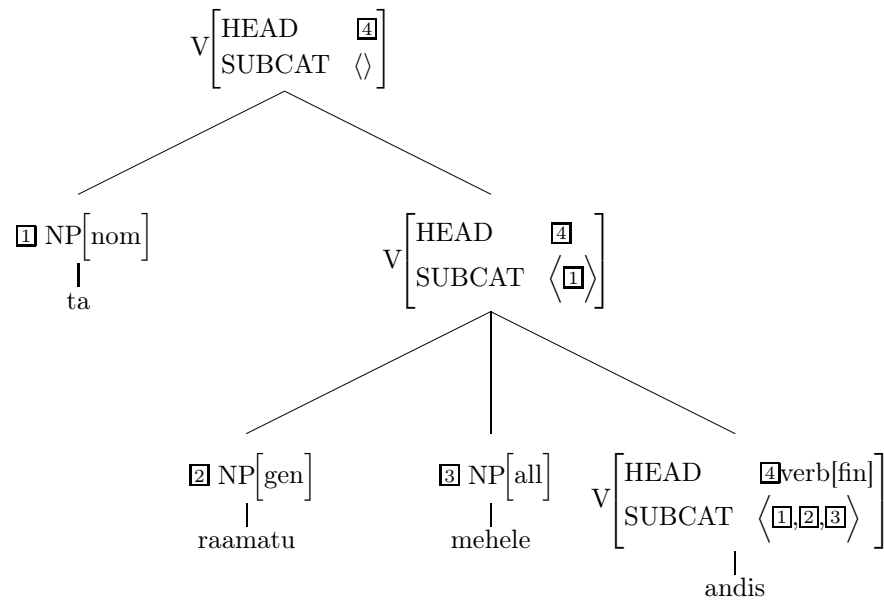
4.4 Põhja tunnuse printsiip

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



Iga põhjaga fraasi kirjelduses peavad ema ja põhitütre HEAD-tunnused unifikseeruma.

Näide põhjaprintsiibist ja subkategorisatsioonist



5 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.