

Süntaksiteooriad ja -mudelid
MTAT.06.031

4. loeng

Kaili Müürisep
kaili.muurisep@ut.ee

6. märts 2008

Tänud Øyvind Rangøyle konspekti täiendamise eest.

Tänane loeng

Contents

1	Kontekstivabade grammatikate kokkuvõte	1
2	Tunnusstruktuurid	2
2.1	Unifitseerimine	6
3	HPSG	7
3.1	Suur leksikoni osatähtsus	7
3.2	Konstruksioonid	8
4	Võrdlus GB-ga	8
5	Märgi struktuur	9
5.1	PHON ja SYNSEM	9
5.2	Sõna ja fraas	10
5.3	LOCAL	10
5.4	CATEGORY	11
5.5	SUBCAT	12
5.6	Näited VAL kohta leksikonist: he, book, sends	13
6	Lause esitus HPSG-s	16
6.1	Fraasid	16
6.2	Lause näide: Kim walks.	17
6.3	Subkategorisatsiooni printsiip	17
6.4	Põhja tunnuse printsiip	18

1 Kontekstivabade grammatikate kokkuvõte

KVG probleemid

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

Kontekstivabade grammatikate tugevaks küljeks on ühest küljest lihtsad reeglid, kuid reeglite lihtsus tekitab omakorda uusi probleeme. KVGs ei ole reeglitel kitsendusi, mistõttu kontekstivaba fraasistruktuurigrammatika reegel ei kajasta näiteks ühilduvusnähtustega seotud piiranguid. Teisest küljest ei ole reeglid ka piisavalt üldised, kuna grammatiliste nähtuste sarnasused - ehk kategooriate jagamine - võib olla täiesti olemas, aga raskesti kirjeldatav kontekstivaba fraasistruktuurigrammatika reeglitega. Lõpuks on KVG probleemiks reeglite suur hulk, mis on loogiline tulemus ülalmainitud probleemidest.

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 struktuursed 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:

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

2 Tunnusstruktuurid

Üheks lahenduseks on tunnuste ja tunnusstruktuuride kasutuselevõtt. Tunnusstruktuurid võivad KVG-le lisada võimalusi sümbolite sisemise struktuuri kirjeldamiseks, mis omakorda võimaldab kasutada sümbolite sisemise struktuuri sarnasusi reeglites.

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

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

Tunnusstruktuurid on tunnuste hulk, mis koosneb tunnustest ja väärtustest. Tunnusstruktuure kirjeldatakse nurksulgudega - tüüpiline notatsioon on tunnuste ja väärtuste maatriks (AVM). Tunnusstruktuur võib olla lihtne - näiteks ainult ühe tunnuse ja väärtusega - või keerukama hierarhilise ülesehitusega. Tunnusstruktuuri hierarhilisus tuleneb sellest, et üks tunnusstruktuur võib olla suurema struktuuri tunnuse väärtuseks.

KVG kontekstis võimaldavad just hierarhilised tunnusstruktuurid üldistuste tegemist. Näiteks jagavad kõik verbid tunnuse väärtust [HEAD: verb], sellest sõltumata, mida täpsustavad teised tunnused verbi omadustest.

Tunnuste tüübid

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

Tartu ülikool

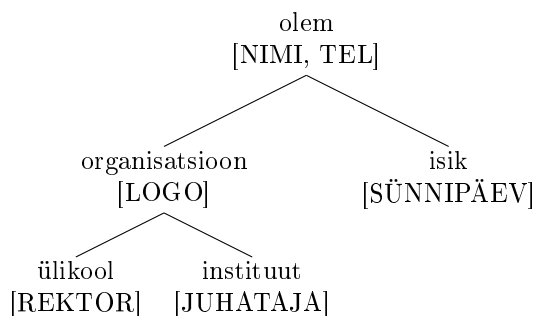
- $\left[\begin{array}{ll} \text{NIMI} & \text{TARTU ÜLIKOOL} \\ \text{TEL} & 737\ 5100 \end{array} \right]$
- $\left[\begin{array}{ll} \text{NIMI} & \text{ARVUTITEADUSE INSTITUUT} \\ \text{TEL} & 737\ 5 \end{array} \right]$
- $\left[\begin{array}{ll} \text{NIMI} & \text{KAILI MÜÜRISSEP} \\ \text{TEL} & 737\ 5 \dots \end{array} \right]$

Tüübierarhia

Tüübid

- $\left[\begin{array}{ll} \text{ülikool} & \\ \text{NIMI} & \text{TARTU ÜLIKOOL} \\ \text{TEL} & 737\ 5100 \\ \text{REKTOR} & \text{ALAR KARIS} \end{array} \right]$
- $\left[\begin{array}{ll} \text{instit uut} & \\ \text{NIMI} & \text{ARVUTITEADUSE INSTITUUT} \\ \text{TEL} & 737\ 5445 \\ \text{JUHATAJA} & \text{TIIT ROOSMAA} \end{array} \right]$
- $\left[\begin{array}{ll} \text{isik} & \\ \text{NIMI} & \text{KAILI MÜÜRISSEP} \\ \text{TEL} & 737\ 5 \dots \\ \text{SÜNNIPÄEV} & 06.03.1986 \end{array} \right]$

$$\left[\begin{array}{ll} \text{ülikool} & \\ \text{NIMI} & \text{TARTU ÜLIKOOL} \\ \text{TEL} & 737\ 5100 \\ \text{REKTOR} & \left[\begin{array}{ll} \text{isik} & \\ \text{NIMI} & \text{ALAR KARIS} \\ \text{TEL} & \dots \\ \text{SÜNNIPÄEV} & \dots \end{array} \right] \end{array} \right]$$

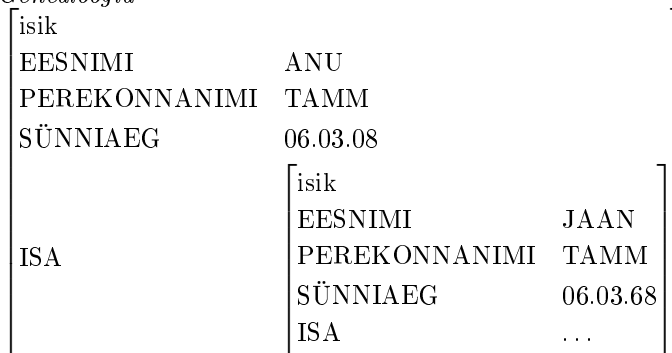


Tüübid defineeritakse relevantsete tunnuste alusel. Tüüpide abil saab täpselt määrata, milliseid operatsioone objektiga on võimalik teha, ning millised on nõutavad argumendid ja nende väärtused.

Tüübid aitavad tagada järjekindlust ja seega grammatika korrektsust. HPSG-s defineerib signatuur grammatika tüüpe koos atribuutide ja võimalike atribuutiväärtusega. Tüüpe saab hierarhiliselt klassideks jagada.

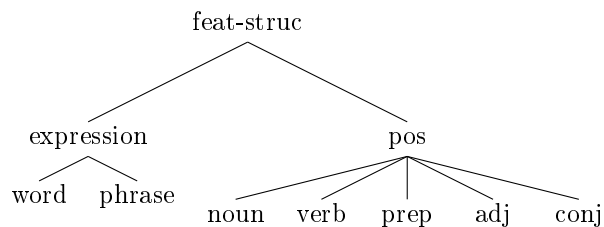
Rekursiivsed tunnusstruktuurid

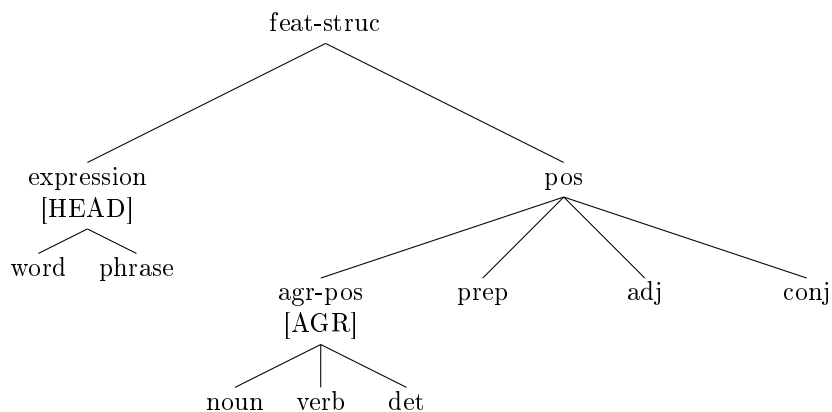
Genealoogia



Tunnusstruktuurid grammatikas

Fraas ja Sõnad leksikonis





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

Oluline mõiste tunnuste struktuuride koha pealt on subsumptsioon. Kui tunnuste struktuur $T1$ subsumeerib struktuuri $T2$, siis tähendab see seda, et struktuuris $T1$ sisalduv informatsioon on alamhulk struktuuri $T2$ informatsioonist - ehk teisisõnu, et $T1$ on üldisem kui (spetsiifilisem) struktuur, mida ta subsumeerib.

Struktuuri jagamine

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

$$\left[\begin{array}{l} \text{SUBJ} \quad \left[\begin{array}{l} \text{AGR} \quad \boxed{1} \quad \left[\begin{array}{l} \text{NUM} \quad \text{sg} \\ \text{PERS} \quad \text{3rd} \end{array} \right] \\ \text{PRED} \quad \left[\begin{array}{l} \text{AGR} \quad \boxed{1} \end{array} \right] \end{array} \right] \end{array} \right]$$

Teine oluline mõiste on tunnuste struktuuri jagamine. Sama tunnuste struktuur võib subsumeerida kahte või enam atribuuti. Sel juhul ei jaga need

atribuudid mitte ainult teatud struktuurilisi ühisjooi - tegemist on täiesti identsete alamstruktuuridega.

2.1 Unifitseerimine

Unifitseerimine

- Tunnuste struktuurid võivad sisaldada erinevat, kuid siiski ühilduvat informatsiooni

$$\begin{array}{l}
 \text{– FS5: } \left[\begin{array}{cc} \text{CAT} & \text{np} \\ \text{ARG} & \left[\begin{array}{cc} \text{NUM} & \text{sg} \end{array} \right] \end{array} \right] \\
 \text{– FS6: } \left[\begin{array}{cc} \text{CAT} & \text{np} \\ \text{ARG} & \left[\begin{array}{cc} \text{PERS} & \text{3rd} \end{array} \right] \end{array} \right]
 \end{array}$$

- Aga ka erinevat ja mitteühilduvat informatsiooni

$$\begin{array}{l}
 \text{– FS7: } \left[\begin{array}{cc} \text{CAT} & \text{np} \\ \text{ARG} & \left[\begin{array}{cc} \text{NUM} & \text{sg} \end{array} \right] \end{array} \right] \\
 \text{– FS8: } \left[\begin{array}{cc} \text{CAT} & \text{np} \\ \text{ARG} & \left[\begin{array}{cc} \text{NUM} & \text{pl} \end{array} \right] \end{array} \right]
 \end{array}$$

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

$$\text{– FS9: } \left[\begin{array}{cc} \text{CAT} & \text{np} \\ \text{ARG} & \left[\begin{array}{cc} \text{NUM} & \text{sg} \\ \text{PERS} & \text{3rd} \end{array} \right] \end{array} \right]$$

-

Definitsioon 2.2. *Kahe tunnuste struktuuri A ja B unifikatsiooniks nimetatakse tunnuste struktuuri C, nii et $A \sqsubseteq C$ ja $B \sqsubseteq C$.*

Kui kaks erinevat tunnuste struktuuri ühilduvad, kuna nad sisaldavad erinevat, kui mitte vastukäivat informatsiooni, siis leidub alati kolmas struktuur, mida mõlemad struktuurid laiendavad.

Kui struktuuri C subsumeerib nii struktuur A kui ka struktuur B (ehk kui nii struktuur A kui ka struktuur B on struktuuri C alamhulgad), siis ütleme, et C on A ja B *unifikatsioon*.

3 HPSG

Head-driven Phrase Structure Grammar

- Pollard and Sag 1987, 1994
- Sag and Wasow 1999
- Leksikaliseeritud
- Pindmised struktuurid
- Ühetasandiline
- Hierarhiline organiseeritus
- Tüpiiseeritud tunnuste struktuurid (leksikon, morfoloogia, fraasid, printsi-
ibid)
- Unifitseerimine/Pärimine
- Fonoloogia, süntaks, semantika ühendatud

3.1 Suur leksikoni osatähtsus

- HPSG-s on enamik informatsiooni leksikonis
- Leksikoni kirjed sisaldavad informatsiooni ortograafia, sõnaliigi, ühildu-
mise, 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)

Klassikalises kontekstivabas fraasistruktuurigrammatikas on esmatähtsad reeglid. Leksikon on sekundaarne ja väljendub eelkõige reeglite terminaalsete sümbolitega. HPSG on perspektiiv teine: leksikoni definitsioon on väga olulisel kohal, kuna just leksikon sisaldab info, mille põhjal toimub põhiosa grammatika tööst. Leksikoni kirjed sisaldavad põhimõtteliselt väga palju infot - ortograafia/fonoloogia, valents ja ühildumine, semantika ja argumentide liitumise põhimõtteid. Siiski ei ole tegemist tarbetu informatsiooni liiasusega, kuna võtmemõisteks on just juba tutvustatud struktuuride jagamine - informatsioon, mis ühendab leksikoni kirjete tüüpe on antud jagatud struktuuride abil, nii et ühe sõna leksikaalne kirje peab lisaks sõna tüübile ainult täpsustama seda, mis on konkreetse sõna puhul teiste sama tüübi sõnade suhtes unikaalne - ehk sõna ortograafiline kuju ja mõned võtmerelatsioonid. Lisaks kirjetega on ka antud leksikaalsed reeglid, mis seovad baaskirjeid teiste kirjetega, näiteks noomenite mitmuse vormide või tegusõna passiivi vormide puhul.

Kui nii suur osa informatsioonist on määratletud leksikonis, ei ole enam vaja palju spetsiifilisi fraasistruktuurireegleid, nagu KVG-s oli. Suurema osa töö tegemisest piisab vaid vähestest, üldistest fraasistruktuurireeglitest: põhi-määratleja, põhi-argument, põhi-alus, põhi-täitja ja põhi-laiend. Siiski nõuavad

mõned konstruktsioonid ka spetsiifilisemaid reegleid, nagu relatiivlausete ja lisandite puhul.

ERG (English Resource Grammar) on inglise grammatika, mis on juurutatud HPSG põhimõtete järgi Stanfordi ülikooli projekti Linguist Grammars Online (LinGO) raames. HPSG tüüpide ja reeglite osatähtsust grammatika kirjeldamises illustreerib seik, et ERGis on 3559 tüüpi ja 105 reeglit.

3.2 Konstruktsioonid

Konstruktsioonid

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

ERGis on 105 reeglit

4 Võrdlus GB-ga

HPSG-l on palju ühist GB (Government and Binding) grammatikatega. Mõlema puhul on lähtepunktiks kontekstivaba grammatika ja mõlemas on põhiprintsiibid väga üldised, leksikoni osakaal suur ja reeglite osakaal väike. Ka mitmed tööpõhimõtted on sarnased.

Põhimõtteline erinevus on siiski see, et HPSG ei arvesta mingisuguste transformatsioonidega.

HPSG-s struktuure ei teisendata. Edasi on struktuurihierarhia tasandid HPSGs võrreldes GB-ga erinevalt jaotatud. Põhimõtteline erinevus on ka see, HPSG ei üleenda süntaksit privilegeeritud seisusse võrreldes teiste grammatikatasanditega - kõik grammatiliste teadmiste tasandid on võrdselt koheldud ja eri tasandid kitsendavad teineteist, üksi tasand ei domineeri. Viimaks on HPSG põhiliselt andmetele orienteeritud, GB-s on olulisemal kohal teooria.

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 privilegeeritud seisuses
 - HPSG on andmetele orienteeritud, GB teooriale orienteeritud.

5 Märki struktuur

5.1 PHON ja SYNSEM

Märki struktuur

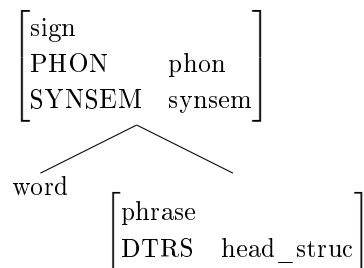
- Märk *Sign* - lause, sõna, fraas, mitmelauseelised diskursused (analüüsiühik).
- Märk - vormi ja tähenduse paar *ehk*
- Märk on struktuur, mis sisaldab fonoloogilist, süntaktilist, semantilist, diskursuse ja fraasistruktuuri informatsiooni.
- Märkil on vähemalt kaks atribuuti:
 - PHON - kirjeldab märki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

HPSGs on tunnuste hierarhias võtmemõiste märk ehk inglise keeles sign. Märk võib olla lause, fraas või sõna, samuti mitmelauseelised diskursused kui need on analüüsiühikuks. Märk sisaldab struktuurina informatsiooni iga grammatikatasandi kohta - nii fonoloogia, süntaksi, semantika, diskursuse kui fraasistruktuuri kohta. Minimaalselt on märkil atribuudid PHON ja SYNSEM.

Esimene kirjeldab märki foneetilist sisu ehk avaldumist kõnes, teine sisaldab süntaktilist ja semantilist informatsiooni, ehk teisisõnu - kui märk on sõna, siis täieliku sünonüümia korral erineksid kahe sõna märgistruktuuris PHON, aga SYNSEM oleks jagatud. SYNSEM jaguneb atribuutideks LOCAL ja NONLOCAL, neist viimane kasutatakse määramata liikumiste kirjeldamisel ja võib ka olla tühi - tavaliselt on olulisemal kohal LOCAL (LOC), mis jaguneb atribuutideks CATEGORY, CONTENT ja CONTEXT.

Märk - ehk objekt tüübist *sign* peab alati olema tüüpi *word* või *phrase*. Igal märkil on atribuudid PHON ja SYNSEM, phrase tüüpi objektidel on lisaks tütreid ehk atribuut DTRS (DAUGHTERS).

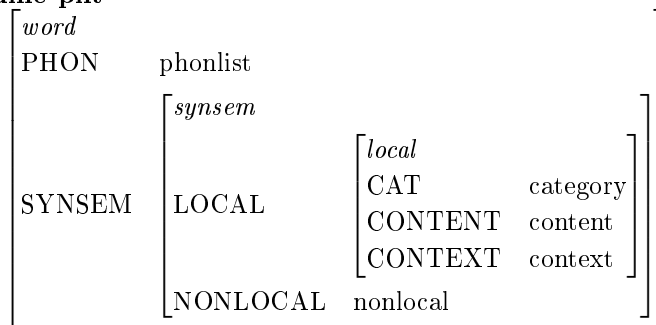
5.2 Sõna ja fraas



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



5.3 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

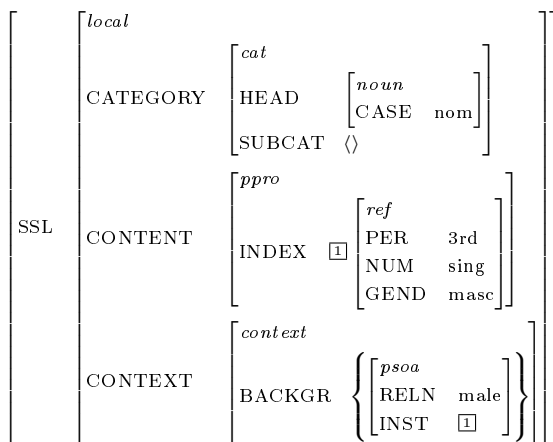
SYNSEMi tunnusel CATEGORY on 2 atribuuti: HEAD ja SUBCAT. SUBCATi asemel võidakse ka kasutada kolme valentsitunnust SUBJ, COMPS ja SPR.

Tunnus CONTENT kannab semantilist informatsiooni. Noomenitel on CONTENTil tunnus INDEX, mis näitab referentsust. CONTENTi väärtus võib näiteks olla tüüpi *ppro* - isikuline asesõna.

Inglise keele meessoost kolmande pöörde ainsuse isikulise asesõna *he* puhul on atribuudil INDEX väärtus tüübist *ref*, millele on kolm atribuuti: PER (isik ehk pööre) väärtusega 3rd, NUM (arv) väärtusega sing, ja GEND (sugu) väärtusega masc.

CONTEXT sisaldab kontekstitundlikku informatsiooni. Konteksti atribuudil BACKGR on väärtuseks hulk objekti tüübist *psoa* ehk “parametrized states of affairs” mis seab piiranguid kontekstidele, kus sõna võib esineda.

LOCAL informatsioon asesõnale *he*

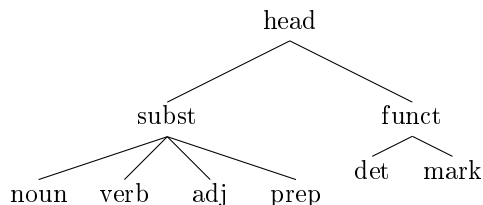


Tunnuste ja väärtuse maatriksis näidatakse struktuuride jagamist märgendite (tags) abil, milleks on numbrid kastis. Hulgad loogelistes sulgudes { }, listid kolmnurksulgudes. < >.

5.4 CATEGORY

CATEGORY

HEAD tunnus näitab sõnaliiki. Tüübistruktuur:



CATEGORY väärtuses olev atribuut HEAD näitab sõnaliiki. Sõnaliike jagatakse tüübistruktuuris kaheks põhirühmaks, sisulised ja funktsionaalsed ehk subst ja funct. Sisuliste hulgas on nimisõnad (noun), tegusõnad (verb), omadussõnad (adj) ja eessõnad (prep). Funktsionaalsed sõnad on eelkõige seotud määratusega (det) ja muude markeritega (mark). that, for, than, as.

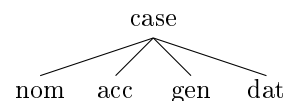
NOUN

[CASE case]

Käände tüübi jagunemine (inglise k)

HEAD informatsioon asesõnale *he*

[HEAD [noun
CASE nom]]

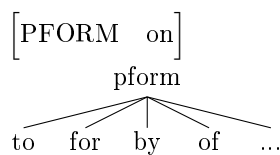


Juhul, kui sõnaliigi (HEAD) atribuudi väärtus on tüüp noun, on selle atribuudiks CASE, mille võimalikud väärtused inglise keeles näiteks võivad olla nominatiiv nom, akusatiiv acc, genitiiv gen või daativ dat.

PREPOSITION

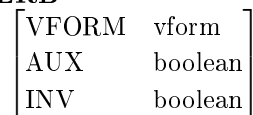
[PFORM pform]

Verb *depend* vajab eessõna *on*:

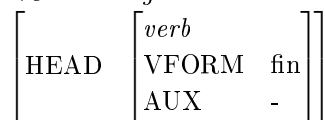


Eessõnade puhul on vastavaks atribuudiks PFORM, ehk konkreetne eessõna, mille väärtuseks peab olema üks võimalikest eessõnadest.

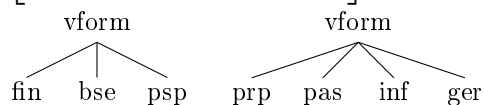
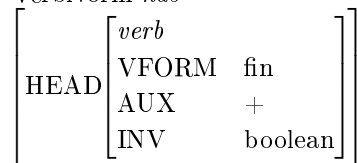
VERB



Verbivorm *gives*



Verbivorm *has*



Kui tüübiks on verb on atribuudiks verbi vorm, VFORM, ning lisaks kaks boole'i väärtust AUX ja INV. Verbi vormiks võib näiteks olla finitne (pöörde-line) vorm fin või infinitiiv inf. AUX on tõene juhul, kui verb toimib abiverbina.

5.5 SUBCAT

SUBCAT

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

magab: $\text{SUBCAT} \langle \text{NP}[\text{nom}] \rangle$

armastab: $\text{SUBCAT} \langle \text{NP}[\text{nom}], \text{NP}[\text{part}] \rangle$

mõtleb: $\text{SUBCAT} \langle \text{NP}[\text{nom}], \text{NP}[\text{elat}] \mid \text{PP} [\text{üle}] \mid \text{S} [\text{et}] \rangle$

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

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

- Hiljem SUBCAT tunnus asendati kolme valentsi tunnusega SUBJ, COMP ja SPR. See lubab paremini kirjeldada kategooriaid, mis vajavad laiendit, kuid mitte alust.

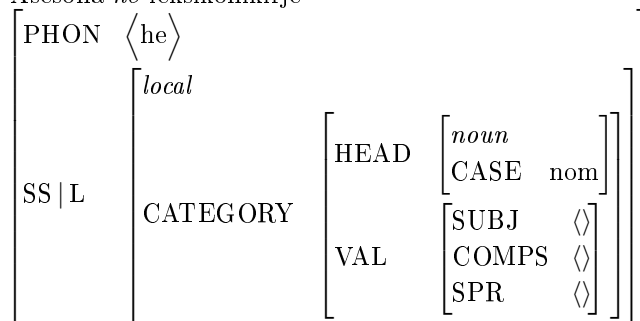
Määratlejad eristatakse alusest sest nad võivad mõlemad esineda sama põhja juures (öeldistäitega lauses).

Edasi on atribuudi CATEGORY väärtuses ka tunnus SUBCAT, mille väärtuseks on loend. Selles subkategorisatsioonis määratakse märgiga seotud elementide arvu ja vormi, mille abil näiteks võidakse kirjeldada rektsiooninähtusi. Loendis eraldatakse komaga erinevaid kohti, püstkriipsuga alternatiive. Elementide järjekorda tingib seoste olulisus, mitte järjekord pindstruktuuris. SUBCATi väärtus oli olemas HPSG esituses raamatus Pollard & Sag (1994), kuid asendati hiljem kolme valentsi (VAL) tunnusega: SUBJ, COMP ja SPR, kõik eraldi loendid. Niimoodi eraldati alust määratlejateks, mille tõttu on kergem kirjeldada kategooriaid, mis küll vajavad laiendit, kuid mitte alust.

5.6 Näited VAL kohta leksikonist: he, book, sends

he

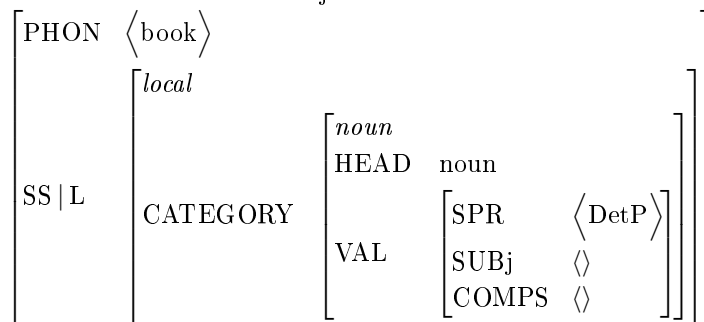
Asesõna *he* leksikonikirje



Inglise keele isikulise asesõna *he* puhul on kõigi VAL väärtuses olevate atribuutide väärtuste tühjad loendid.

book

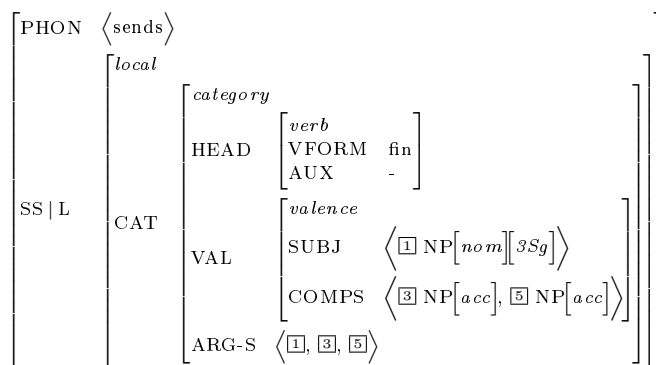
Nimisõna *book* leksikonikirje



Nimisõna *book* puhul on SPR väärtuses Detp, SUBJ ja COMPS on endiselt tühjad.

sends

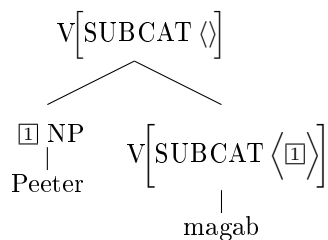
Verb *sends*



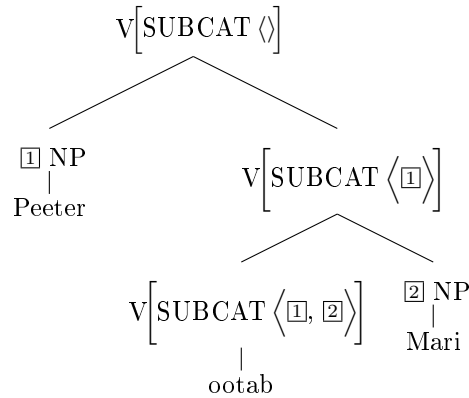
Verbi leksikonikirjes läheb rohkem informatsiooni vaja. Inglise verbil (pöördelisel verbivormil) *sends* peab olema alus. Seetõttu ei ole VAL|SUBJ siin tühi - väärtuses on noomenifraas NP, kuid noomenifraasil on lisaks omad ka oma tunnuste väärtused - *nom* ja *3Sg*. Edasi on eraldi verbi valentsis täpsustatud VAL|COMPS, siin võib olla kaks erinevat laiendit, kuna inglise verbil võib olla nii otsene kui kaudne sihitis - need on loendis eraldatud komaga, kuigi nende väärtused on samad: Nii otse- kui kaudsihitis on inglise keeles NP[acc], kuna inglise keeles ei ole eraldi kaudsihitise markeerimiseks eraldi käändevormi.

Kogu valentsi (VAL) tunnusstruktuur kirjeldab mõnes mõttes verbi argumente. Seda asjaolu peegeldab notatsioonis atribuut CAT|ARG-S, mis numbrimärgendite abil viitab VAL tunnusstruktuuris olevate lauseliikmete spetsifikatsioonidele.

Eestikeelsed näited



Transitiivne verb



Eesti keele lihtlause *Peeter magab* süntaksipuud analüüsidest võiksime HPSG notatsiooni kasutades numbrimärgenditega viidata sellele, et verb *magab* nõuab subjektiks NP, mille struktuur on jagatud verbi SUBCAT ja aluse noomenifraasi CAT vahel. Samuti oleks verbi SUBCATis transitiivse verbiga lause *Peeter ootab Mari* puhul numbrimärgenditega viidatavat struktuuride jagamist nii aluse kui ka sihituse noomenifraasidega.

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

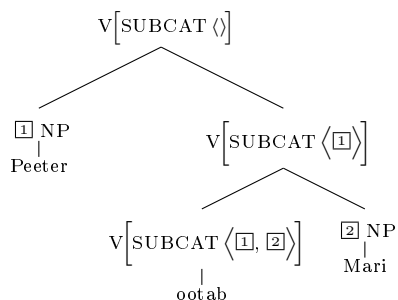
Valentsireeglite sõnastamisel saavad selgemaks HPSG eelised klassikalise kontekstivaba fraasistruktuurigrammatikaga võrreldes. Üldised reeglid võimaldavad ühe reeglina hõlmata fraase, mis KVGs nõuaksid erinevaid reegleid. See on võimalik selle tõttu, et ka siis, kui osa tunnusstruktuurides olevaid argumente on tühjad, toimib reegel edasi nende struktuuride osade põhjal, mis on relevantset. Formaalselt on see võimalik otsekorrutise abi, mis ka võimaldab põhja ja argumentide seoste reegleid üldistada ($[A]$ ja $[B]$ on viited, õiges notatsioonis kastides): $H[\text{SUBCAT } [A]] \Rightarrow H[\text{SUBCAT } [A] \oplus \langle [B] \rangle], [A]$.

Kuna tegemist on otsekorrutisega, on reegel rakendatav sellest sõltumata, kas $[A]$ on tühi või mitte. Seega asendab selline kuju ühe kujuga näiteks $VP_2 \Rightarrow NP$ VP ja $VP \Rightarrow V$ NP .

Põhi-argumentreeglite näide

$$VP_2 \Rightarrow NP, VP \\ V[\text{SUBCAT } \boxed{A}] \rightarrow \boxed{A}, V[\text{SUBCAT } \boxed{A} \oplus \langle \boxed{B} \rangle] \quad \boxed{A} = (), \boxed{B} = \boxed{1}$$

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



Põhi-argumentreeglite üldistamine

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

6 Lause esitus HPSG-s

6.1 Fraasid

Fraasid

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

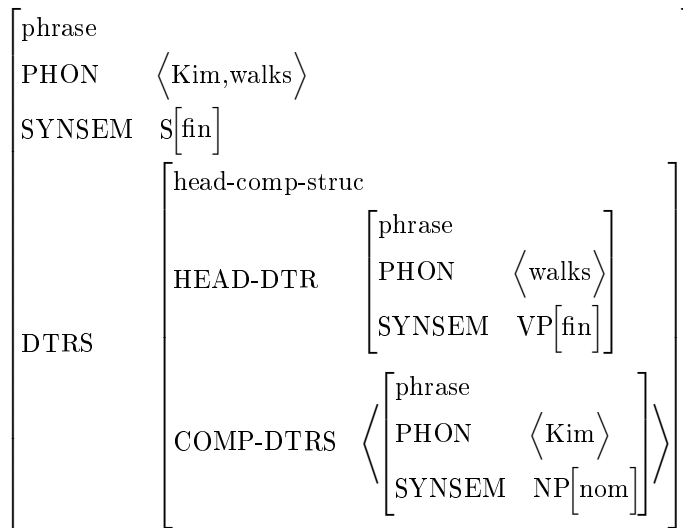
con-struct tüübil on palju alamtüüpe erinevat liiki tütarlippude kirjeldamiseks.

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**

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

6.2 Lause näide: Kim walks.



Kuigi suhteliselt kesksel kohal on leksikon, on HPSG - nagu ka nimi ütleb - ikka fraasistruktuurigrammatika. Fraasil on oma tunnusstruktuur, mis kirjeldab struktuuri atribuudi DAUGHTERS (DTRS) kaudu. Tütarde atribuudi väärtus on moodustajate struktuur ehk struktuur tüübist constituent-structure (con-struc). Moodustajastruktuuri tüübil on palju alamtüüpe, neist olulisim on headed-structure (head-struc), mida kasutatakse põhjaga konstruktsioonides - ja nagu HPSG nimigi viitab, on tegemiste põhjajuhitava grammatikaga. Põhjaga struktuuril on atribuute nagu HEAD-DAUGHTER (HEAD-DTR), COMPLEMENT-DAUGHTERS (COMP-DTRS), ADJUNCT-DAUGHTER, FILLER-DAUGHTER, MARKER-DAUGHTER. Tuumtütre väärtuseks on üks märk, täiendavate tütarde väärtuseks märkide loend.

Näiteks on inglisekeelne fraas (ja lause) ‘*Kim walks.*’ kirjeldatav tütarde kaudu, ning põhjaks on siis verb (verbifraas), mida esindab verb *walks*. Tervel fraasil on oma foneetiline struktuur PHON, mida esindab konkreetsete sõnavormide loend $\langle \text{Kim, walks} \rangle$. Fraasil on atribuudi SYNSEM (süntaktiline ja semantiline informatsioon) $\text{S}[\text{fin}]$, mis viitab et tegemist on finitse lausega. Lisaks on fraasil atribuut DTRS, mille väärtuses on atribuudid HEAD-DTR, mille väärtuses on üks fraas ja COMP-DTRS, mis on täiendavate fraaside loend. HEAD-DTR on SYNSEMi väärtusega $\text{VP}[\text{fin}]$ ehk finitne verbifraas, täiendavata tütarde loendi ainsal liikmel on SYNSEMi väärtus $\text{NP}[\text{nom}]$.

6.3 Subkategorisatsiooni printiiip

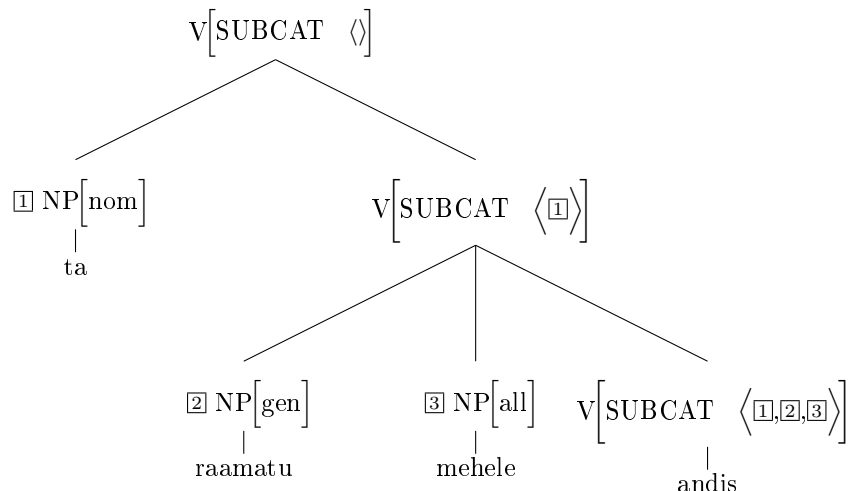
Subkategorisatsiooni printsiip

Printsiip 6.1. *In headed phrase (i.e. a phrasal sign whose DTRS value is of sort head-struc), 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.*

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

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

Ta andis mehele raamatu



HPSG üks põhiprintsiipe on subkategorisatsiooni printsiip. See ütleb, et põhjaga fraasis on tuumtütre SUBCATi väärtus fraasi (ema) ja õdede (emafraasi täiendavate tütarde) konkatenatsioon. See tähendab üldjuhul, et fraasi ja tuumtütre SUBCATi väärtused unifitseeruvad. Seega on fraasi SUBCATi väärtus fraasi tuumtütre SUBCATi väärtuse alamhulk.

6.4 Põhja tunnuse printsiip

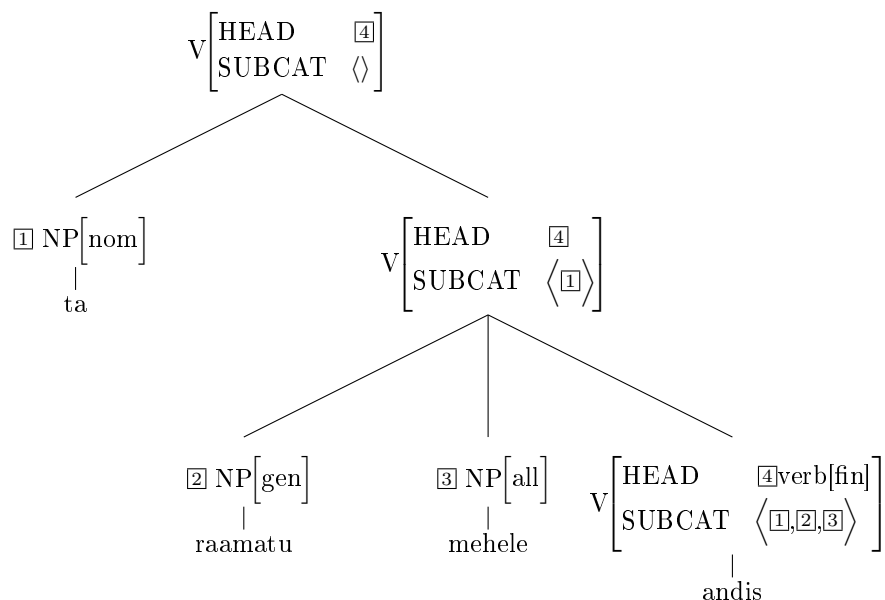
Põhja tunnuse printsiip (HFP)

Printsiip 6.2. *The HEAD value of any headed phrase is structure-shared with the HEAD value of the head daughter.*

$$\begin{bmatrix} \text{head-argument-structure} \\ \text{HEAD} & \boxed{1} \\ \text{SUBCAT} & \boxed{2} \\ \text{HEAD-DTR} & \left[\text{HEAD} \boxed{1} \right. \\ & \left. \text{SUBCAT} \boxed{2} \oplus \boxed{3} \right] \\ \text{COMPS} & \langle \boxed{3} \rangle \end{bmatrix}$$

Teine HPSG põhiprintsiipidest on põhja tunnuse printsiip, mis ütleb et HEAD väärtus jagab põhjaga fraasis tuumtütre HEAD väärtusega struktuuri - ehk et põhjaga fraasis peavad alati ema ja tuumtütre HEAD-tunnused unifitseeruma. Subkategorisatsiooni ja põhja tunnuse printsiipide koosmõju on suures osas HPSG mootoriks - fraasistruktuuri moodustajad täpsustavad põhistruktuuri, mis oli juba fraasil endal olemas.

Näide põhjaprintsiibist ja subkategorisatsioonist



HPSG on seega fraasistruktuurigrammatika, kus nii reeglid, leksikon kui ka üldised printsiibid kõik osaliselt kirjeldavad grammatikat - erinevalt KVG grammatikast, kus iga sõlm kasutab ainult ühes reeglis sisaldavat infot ja mitte midagi muud. HPSG-s on erinevate osade ja tasandite “koostöö” võimalik tunnusstruktuuride tõttu. Tunnusstruktuur on ideaalselt üksuse täielik kirjeldus, kuid konkreetsed kirjeldused võivad olla osalised tunnusstruktuurid. Sel juhul subsumeerib alati osaline struktuur täielikku tunnusstruktuuri. Tunnusstruktuuride abil kirjeldatud grammatika on hea siis, kui struktuurid vastavad tegelikkusele ja kirjeldavad ammendavalt reaalsel olemas nähtusi. Tunnusstruktuure saab esitada tunnuste ja väärtuste maatriksiga.

Tervikuna peab grammatika kirjeldama kuidas leksikon on seotud tunnusstruktuuridega ja kuidas väiksemad üksused kombineeruvad suuremateks üksusteks. Kuna HPSG kasutab hierarhilise sisemise struktuuriga tunnuste kirjeldusi, on grammatika võrreldes KVG-ga tunduvalt üldisem ja abstraktsem ja üritab jagatud struktuuride abil ilma informatsiooni kaotamisega ka liiasust vähendada.

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.