

Süntaktiliste funktsioonide ja sõltuvusseoste märgendamisest

Korpuses on märgendatud süntaktilised funktsioonid ja sõltuvussuhted.

Selles dokumendis ei hakata üle kordama eesti keele lauseanalüüsi (süntaktiliste funktsioonide e lauseliikmete) põhitõdesid, mis on esitatud järgmistes teostes:

- EKG II <http://dspace.utlib.ee/dspace/handle/10062/29437>
- Eesti keele käsiraamat <http://www.eki.ee/books/ekk09/>
- Mati Ereht. Eesti keele lauseõpetus. Sissejuhatus. Öeldis. TÜ eesti keele osakonna preprintid number 4. <http://dspace.utlib.ee/dspace/handle/10062/34069>

Sõltuvussüntaktiline märgendus annab infot iga tekstisõna ülemuse kohta sõltuvuspuus. Süntaktiline sõltuvus on binaarne ebasümmeetriline suhe, st sõltuvussuhe on suhe kahe sõna vahel ja nendest kahest sõnast üks on ülemus ja teine alluv. Sõltuvussüntaksi põhimõtete järgi on osalause kõige kõrgemaks ülemuseks predikaat, millele allub ka subjekt. Kesksete süntaktiliste struktuuride esitamiseks sobib sõltuvusesitus hästi (nt verb + objekt, noomen + atribuut), kuid mõnede struktuuride esitamine sõltuvuste abil on pisut problemaatiline, nende hulgas on tüüpiliselt nimetatud mitmesõnalist predikaati (verbi liitvormid, perifrastilised verbid), kaassõnaühendeid, kvantoriühendeid ning mitmeid muid mitmesõnalisi keelendeid (mh mitmesõnaline arv, nt *kolm tuhat kakssada viiskümmend seitse*; mitmesõnaline nimi, nt *Jaan Juhan Tamm, Wilhelm II* või nimetus, nt *telesaad „Hüppa ja karga“*). Nende märgendamisest täpsemalt allpool.

2.1. Kasutatud märgendid

Korpuse märgendamisel kasutatud süntaktilise funktsiooni märgendid on järgmised

@SUBJ subjekt

@OBJ objekt

@ADVL adverbiaal

@PRD predikatiiv

predikaadi märgendid:

@FCV *olema* liitaegades, modaalverbid, finiiitne vorm (*olen teinud, saan teha*; modaalid on need, mis on üles loetud EKG-s; muud perifrastilise öeldise finiiitsed komponendid saavad märgendi

@FMV; *hakkan tegema*)

@FMV finiiitne põhiverb, ka need perifrastilise öeldise finiiitsed komponendid, mis ei saa

märgendit @IMV (*Laps sööb saia; **hakkan** tegema*)

@ICV *olema* liitaegades, modaalverbid, infiniitne vorm (modaalid on need, mis on üles loetud EKG-s; muud perifrastilise öeldise finiitsed komponendid saavad märgendi @IMV) (*võin **olla** teinud, olen **saanud** teha*)

@IMV infiniitne põhiverb, (*olen **teinud**, hakkan **tegema**, olen **hakanud** tegema*)

@Vpart afiksaaladverb ühendverbi koosseisus (*sai sellest **üle***)

@VpartN afiksaaladverb nominalisatsiooni koosseisus (*tema eesmärgiks on meie **kõrvale** kallutamine*)

@NEG *ei* verbi eitava vormi osana (***ei** taha*)

atribuudi märgendid

@<AN adjektiiv järeltäiendina (ostis diivani, **roheline** ja **lahtitõmmatava**)

@AN> adjektiiv eestäiendina (**roheline** ja **lahtitõmmatav** diivan)

@<NN nimisõna järeltäiendina (foto väikesest **jänesest**)

@NN> nimisõna eestäiendina (väikese **jänese** foto)

@<DN määrsõna järeltäiendina (ehitise toomine liiga **lähedale**)

@DN> määrsõna eestäiendina (**vastupidiselt** ootustele)

@<INFN da-infinitiiv järeltäiendina (soov **õppida**)

@INFN> da-infinitiiv eestäiendina (**õppida** tahtmine)

@<KN kaassõnafraas järeltäiendina (märgend pannakse kaassõna külge) (toetus piimalehma **eest**)

@KN> kaassõnafraas eestäiendina (märgend pannakse kaassõna külge) (piimalehma **eest** lisatoetuse maksmine)

@<P prepositsiooni juurde kuuluv käändsõna (nt *üle **tee***)

@P> postpositsiooni juurde kuuluv (nimi)sõna (nt ***laua** all*)

st kaassõnaühendis on kaassõna (postpositsioon, prepositsioon) ülemus ja tema juurde kuuluv käändsõna alluv.

@<Q kvantori kuurde kuuluv sõna (nt *kolm **meest***)

@Q> kvantori eeslaiend (***inimesi** tulvil*)

st kvantorühendis on kvantor (arv- või hulgasõna) ülemus ja tema juurde kuuluv käändsõna alluv

@J sidend

@I interjektsioon

@??? - morfoloogiaanalüsaatorile tundmatuks jäänud sõna või süntaksianalüsaatorile tundmatu sõnaliik.

@NONE teadustekstides sulgudes sõnad, peamiselt kirjandusviited, nt (Mets, 2002: 48)

@VOC üte

2.3. Mitmesõnaline predikaat: verbi liitvormidest ja perifrastilisest öeldisest

2.3.1. Millised on verbi liitvormis ja liitpredikaadis komponentide omavahelised sõltuvussuhted?

Ülemuseks on nn sisuline predikaat, infiniit, st liitvormis *olen teinud* on selleks *teinud*, ahelverbis *pean tegema* on ülemus *tegema*. Põhjenduseks see, et see nn sisuline predikaat tingib lause argumendistruktuuri, tingib lauses olevad seotud laiendid.

2.3.2. Mitmesõnalise predikaadi määratlemisest

Otsustamisel, millise ühendi puhul on tegu perifrastilise verbiga, millise puhul öeldise ja tema seotud laiendiga, on lähtunud EKG II-st, st perifrastilised verbid on need, mis seal on nii määratletud, modaalverbid on *võima*, *saama*, *pidama*, *tohtima*, *vaja olema*, *vaja minema*, aga mitte *tulema* (*mul tuleb nüüd koju minna* – siin on *minna* alus). Lähemalt EKG II lk 237-260.

Sõltuvussuhete märgendamist mõjutab see niipalju, et perifrastilise verbi, nt modaaliühendi *võin teha* puhul on kogu lause kõige suurem ülemus *teha*, aga verbi ja tema laiendi ühendi puhul on lause kõrgeim ülemus finitiivne verb, st *tuleb teha* puhul on ülemuseks *tuleb*.

Näiteks lauses *Sestap peame püüdlema nii NATOsse kui Euroopa Liitu on püüdlema* IMV-na lause kõrgeim ülemus, millele alluvad vahetult sõnavormid *sestap*, *NATOsse*, *Euroopa Liitu*.

Ühendverbe on käsitsi märgendamisel käsitletud avatud hulgana, st märgendajad ei lähtunud etteantud ühendverbide loendist.

2.3.3. Impersonaal vs passiiv

Selliseid lauseid nagu *Ülesanne oli hästi koostatud* saab käsitleda nii impersonaallausena, st *ülesanne* on objekt ja *koostatud* on @IMV-na lause süntaktiline kese või siis passiivilausena, st *ülesanne* on subjekt, *oli* predikaat (@FMV) ja *koostatud* predikatiiv.

Piiri tõmbamine impersonaali ja passiivi vahel *olema*-verbist ja *tud*-partitsiibist koosneva konstruktsiooni analüüsil on keeruline. Korpuse märgendamisel soovitati kahtluse korral eelistada

impersonaali tõlgendust, nt *Väitekiri, mida (OBJ) on peetud(IMV) trükikoja esimeseks trükiseks.*

Tegelikult on pensionid(OBJ) külmutatud(IMV) küll, neid(OBJ) pole ju teist aastat tõstetud(IMV)

Ka *olema* ja *da-infinitiivis* kognitiivverbide (*teada, näha, märgata, tunda, kuulda, karta, arvata, loota, oodata* jne) ühendite puhul on EKG II: 246 järgi tegu „kas perifrastilise passiivi või impersonaaliga“. Kui see on passiiv, on lauses *Ühtegi kassi polnud näha* sõnavorm *kassi* subjekt, kui impersonaal, siis on *kassi* objekt. Need märgendati korpuses impersonaalina, st *kassi* on objekt. Sellised ühendid märgendati passiiviühenditena (tud-vorm on predikatiiv), kui tud-partitsiip on leksikaliseerunud, nt *Põhjapoolne osa(SUBJ) oli hõredasti asustatud(PRD).*

2.3.4. *saama*-passiiv vs impersonaal

Ka *saama* + tud-partitsiibi konstruktsioonid võivad olla nii passiivi kui impersonaali esindajad, nt järgnev on impersonaalilause: *kui vajalik kokkulepe(OBJ) Tallinna Sadama ja Eesti Meremuuseumi vahel saab sõlmitud(IMV).*

2.4. Subjekt

Sõltuvussüntaksis on verb lause kõige kõrgem ülemus, sh ka subjekti ülemus. Kui öeldiseks on verbi liitvorm või liitpredikaat, siis on ka subjekti ülemuseks infiniitne sisuline predikaat.

2.4.1. Kvantor subjektina

Kvantorina võib subjekti märgendi saada ka määrsõna, nt *Kohale oli tulnud päris palju inimesi; Sinna oli pisut rokka jäänud.*

2.4.2. Kaassõnafraas subjektina

Kaassõnadega *üle* ja *alla* moodustatud kaassõnaühendid võivad olla nii subjekti kui objekti funktsioonis, subjekti märgend pannakse kaassõna kui ülemuse külge:

Alates eelmisest neljapäevast, mil kokkupõrked juutide ja araablaste vahel puhkesid, on surma saanud üle@SUBJ 50@<P inimese@<Q, enamuses palestiinlased.

Flora klubis treenib tema tähelepaneliku pilgu all üle@SUBJ tuhande@<P poisi@<Q .

Praegu on registreerunud üle@SUBJ kolmesaja@<P huvitatu@<Q

Kokku käis katsetamas ilmselt alla@SUBJ saja@<P inimese@<Q.

2.5. Predikatiiv

Vastavalt oma definitsioonile „verbist olema sõltuv omadus- või nimisõna nimetavas või osastavas käändes, ka kvantor või da-infinitiiiv” võib predikatiiv olla ka lauses, kus puudub subjekt (*Siin on nii ilus*) ja ka sellise *olema*-verbi laiend, mis pole lauses öeldiseks (*Noor olla on kevadet rinna sees kanda.*)

vat-tarindi koosseisus võib öeldistäide olla ka omastavas käändes: ... **kelle**(PRD -> *olevat*) *ta väitis enese olevat*

Kui koopulaga lauses (st lause, mille keskmeks on verb *olema*) on kaks nimisõna, mille puhul on keeruline otsustada, kumb neist on subjekt ja kumb predikatiiv, siis lause algusele lähemal olev sõnavorm on analüüsitud subjektiks. Näide.

2.6. Objekt

2.6.1. Kvantor objektina

Kvantiühendite märgendamisel esineb ebaühtlust objekti ja adverbiaali tõlgenduste vahel lausetes nagu *Nende marjade eest sai ta 45 eurot* vs *Talle maksti palka 450 eurot*.

2.6.2. Kaassõnafraas objektina

Kaassõnadega *üle* ja *alla* moodustatud fraasid võivad olla nii subjekti kui objekti funktsioonis ja käsitsi on nad nendeks ka märgendatud. Näiteks:

Iisrael on vahistanud üle@OBJ tuhande@<P palestiinlase@<Q

Kohtunikuna tegutsenud norralane Tor Floysvik, kes kasvatanud üles üle@OBJ kahekümne@<P maailmameistri @<Q, hinnanud üritust kõrgelt.

Sõidab president näiteks välismaale, siis alla@OBJ saja@<P aumärgi@<Q ei tasuks üldse kaasa võtta .

2.7. Verbita lauselühend

Verbita lauselühend (*kepp käes, kott üle õla*) – nendes on see, mille kohta EKK ütleb 'subjektisarnane element' (st *kepp, kott*), subjekt ja lauselühendi kui terviku ülemus. See teine osaline seal konstruktsioonis (*kott, üle õla*) on siis ADVL, mis ripub subjekti küljes.

2.8. Üldlaiend

Üldlaiendit (EKG II lk 99 jj) osad teoreetilised käsitlused peavad ja osad ei pea lauseliikmeks. Küsimust tema kohta esitada ei saa, aga ülemust tavaliselt saab määrata. Üldlaiend on nt *ainult* lausetes *Ma pildistan ainult kasse* ja *Ma ainult pildistan kasse, küllap* lauses *Küllap tuleb tänavu*

varajane kevad, võib-olla lauses Võib-olla ta ootab järgmist rongi.

Üldlaiendina esinevad sõnad *vaid, pigem, ka, enam, juba, ehk, tõenäoliselt, nähtavasti, näiteks, just* jpt.

Üldlaiendil meie süsteemis (praegu) oma märgendit ei ole, sõnaliigiliselt on ta tavaliselt adverb. Märgendatud on ta vastavalt sellele, mis on tema ülemus. Nimisõna alluv on täiend (*Ma pildistan ainult(DN>) kasse*), verbi alluv on määrus (*Ma ainult(ADVL) pildistan kasse*). Omadussõna alluv on samuti määrus (*Mulle meeldivad ainult(ADVL) triibulised kassid*). Asesõna, vähemalt nimisõnalise asesõna alluv on samuti täiend (*Ainult(DN>) sina suudad maailma päästa*).

Üldlaiend laiendab tüüpiliselt talle vahetult järgnevat sõna või selle ülemust või lause põhiverbi.

Üldlaiendite ülemuste määramine käsitsi märgendades on praeguse seisuga toimunud „tunde järgi” ja vajab süstemaatilist lähenemist.

2.9. Atribuut

Atribuut on nimisõna, põhiarvsõna ja asesõna alluv, st iga sõnavorm, mille ülemuseks on üks nimetatud sõnaliikidest, saab igal juhul atribuudi märgendi.

2.10. Ülemus-alluv-suhete erijuhte

2.10.1. Kvantor

Kvantori ja tema juurde kuuluva sõna ühendis on ülemuseks kvantor, st arvsõna või hulgasõna, viimane on avatud hulk: *palju inimesi, rohkesti inimesi, pudel piima, kimp lilli, riiulitais raamatuid, jne.*

Kui kvantor on mitmuses, on ülemus-alluvussuhe vastupidine: ühendis *paljud inimesed* on ülemus *inimesed*, ühendis *kümned inimesed* on ülemus samuti *inimesed*. Samuti ei ole ülemuseks arvsõna *üks*; ühendis *üks kass* on ülemuseks *kass*. Kui kvantori ja nimisõna ühendis on nimisõna seestütlevas käändes, ei ole ta märgendatud <Q vaid <NN: *kolmandik inimestest(<NN)* vs *kolmandik inimesi(<Q)*.

2.10.2. Kaassõna

Kaassõnaühendis on kaassõna ülemus ja tema juurde kuuluv käändsõna alluv. Kaassõnaühendi süntaktilise funktsiooni märgend antakse kaassõnale, näiteks *Juhan on üle(KN>) küla(<P) poiss*.

Kaassõnaühendid ei ole mh *igale poole, teisele poole, igal pool, teisel pool, ühel pool, igalt poolt, omalt poolt, teiselt poolt, ühelt poolt* jms ei ole kaassõnafraasid.

2.10.3. Pealkirjad, nimeüksused – atribuut või ülemus?

Puudutab struktuure nagu ajaleht 'Punane Täht', saade „Hüppa ja karga“, sõiduauto Volkswagen Bora. Kuigi EKK järgi on pealkiri põhi, on korpuses märgendatud teistpidi, st üldnimi on põhi ja pealkirjaks või nimeüksuseks olev sõnaühend on riputatud üldnimelise peasõna külge; eelnevates

näidetes on ülemused *ajaleht, saade, sõiduauto*.

Võõrkeelsed pealkirjad ja nimetused on märgendatud nagu eesti nimisõnafraasid, st ülemuseks on viimane sõna, eelnevad sõnad on ketina üksteise küljes. Näiteks: *Pääses nimekirja oma romaaniga „On (->Chesil) Chesil(->Beach) Beach(<-romaaniga)“*

Eestikeelsetest sõnadest koosnevatesse nimeüksustesse nagu nt *Tartu Majaomanike Ühing* kuuluvaid sõnu tuleb analüüsida nagu harilikke sõnu, st kehtivad tavalised süntaktilised suhted ja morfoloogilised kategooriad, nt *Majaomanike* on mitmuse genitiiv süntaktilise märgendiga @NN>

Isikunimedes on perekonnanimi ülemus ja eesnimi alluv.

2.11. Lühendid

Selliste lühendite nagu *jne jms jt jpt* ülemuseks on eelnev sõna ja märgendatakse kui atribuuti.

Selliste lühenditenagu *nn so* ülemuseks on järgnev sõna ja märgendatakse kui atribuuti.

Ülejäänud lühendite märgendiks on ADVL ja ülemuseks osalause kõrgeim ülemus.

Suurtähelisi akronüüme (*USA, UNESCO* jt) koheldakse kui pärisnimesid.

2.12. Koordinatsioon

Koordineeritud üksusi märgendatakse järgnevalt: esimene koordineeritud üksus on põhi, teine sõltub esimesest ja sidesõna sõltub talle järgnevast koordineeritud elemendist.

Koordineeritud üksuste reegli järgi käib ka *mitte .. vaid, nii .. kui ka* -koordinatsioon, nt lausetes

Kultuuri ei kujunda teadliku tegevusega mitte reakodanik, vaid eliit

*Rahvuslik diskrimineerimine leiaks taunimist nii õiguslike vahendite kui ka riigi ja eraasutuste poolt on sidendid pandud alluma järgnevale koordineeritud elemendile, st *vaid->eliit, nii->vahendite, ka->riigi**

Ka rinnastatud osalused käivad selle reegli järgi: Lauses *Koerad sõid puru ja kassid näksisid konservi* on *näksisid* ülemus *sõid*.

2.14. Osalausevaheliste suhete esitamine

Osalausevaheliste suhete esitamine on korpuse esimeses versioonis lahendatud robustselt: üldreeglina on kõrvallause predikaadi või kõrgeima ülemuse ülemuseks on pealause kõrgeim ülemus.

Erandiks on relatiivkõrvallause, st asesõnaga *kes* või *mis* või nende käändevormidega algav täiendkõrvallause, mis kuulub pealause viimasel positsioonil olevale (pro)noomenile, nt järgmises lauses on kõrvallause predikaadi *kutsuti* ülemus *käänaku*, mitte pealause öeldis *tegi*: *Päris lõpus tegi tänav järsu käänaku, mida enamasti kutsuti Kaptenikäänakuks*.

Relatiivlause võib alata ka sõnaga *kus* või *et*: *olen näinud Maapanga ülekandeid, kus on(<-*

ülekandeid) kirjas ...

Tuli mõte, et saaks(<-mõte) ju teisti

Siinkohal pandagu tähele, et relatiivlausega lause ei ole sama mis korrelaadiga lause. Korrelaadiga lauses on korrelaadil ja kõrvallausel sama süntaktiline funktsioon, kõrvallause ei ole korrelaadi suhtes täiendiks. Relatiivlause jällegi on oma põhisõna suhtes täiendiks. Tüüpiline korrelaat on *see* käändevorm, nt lauses *Ma teadsin seda, et nüüd hakkab vihma sadama* on *seda* objekt, mille ülemus on *teadsin* ja kõrvallause *et nüüd hakkab vihma sadama* on sihitiskõrvallause, mis meie süsteemis ei kajastu, mille ülemus on *teadsin*, mis meie süsteemis kajastub nii, et *sadama* ülemus on *teadsin*.

Koordineeritud osalaused on märgendatud koordinatsiooni üldreegli järgi: järgneva osalause predikaadi ülemuseks on eelneva osalause predikaat, nt *Vihma sajab, udu tõuseb (sajab<-), lombid aina suurenevad (<-tõuseb)*.

Rinnastatud osalaused, millest teine on väljajätteline, nt lause *Koerad sõid puru ja kassid konservi* on märgendatud nii, et teise osalause ülemus on subjekt *kassid*, mille ülemus on omakorda esimese osalause predikaat *sõid*.