

# VÄLJUNDI SELETUS

Väljundis on esitatud iga sõna morfoloogiline analüüs, süntaktilised funktsioonid ja sõltuvussuhted.

## Väljundi näide

"<s>"

"<Mitmekesisus>"

"mitme\_kesisus" L0 S com sg nom cap @SUBJ #1->2

"<on>"

"ole" L0 V main indic pres ps3 sg ps af <FinV> <Intr> @FMV #2->0

"<elu>"

"elu" L0 S com sg gen @NN> #3->5

"<vaieldamatu>"

"vaieldamatu" L0 A pos sg nom @AN> #4->5

"<omapära>"

"oma\_pära" L0 S com sg nom @PRD #5->2

"<.>"

"," Z Fst #6->6

"</s>"

## Väljundi kuju seletus

"<sõna>"

"tüvi" L lõpp sõnaliik morfoloogiline\_info @süntaktiline\_fn #sõna\_jrk\_nr->ülemuse\_jrk\_nr

kus:

- "<sõna>" - tekstis esinenud sõnavorm,
- tüvi - lemma e. algvormi tüvi,
- lõpp - sõna lõpp,
- sõnaliik - ühetäheline sõnaliigi lühend,
- morfoloogiline\_info - noomeni- või verbikategooria täpsemate morfoloogiakategooriate lühendid, nt noomeni korral kääne või verbi korral aeg
- süntaktiline\_fn – süntaktiline funktsioon lauses, nt subjekt,
- sõna\_jrk\_nr - sõna järjekorranumber lauses,
- ülemuse\_jrk\_nr - sõna ülemuse järjekorranumber lauses,
- Sõna eritõlgendused paiknevad eri ridadel.

"<s>" ja "</s>" on vastavalt lause alguse ja lõpu märgendid.

Morfoloogiliseks analüüsiks kasutatakse Filosoofi Vabamorfi <https://github.com/Filosoofi/vabamorfi>, mille väljund teisendatakse süntaksianalüsaatorile sobivale kujule. Põhjalikku morfoloogiliste märgendite kirjeldust tabeli kujul koos näidetega saab leida:

<http://www.cl.ut.ee/korpused/morfliides/seletus>. Selle kirjelduse erinevused süntaksianalüsaatori kasutatavatest märgenditest:

1. sõnaliigi märgend on ilma alakriipsudeta (vt allpool Sõnaliikide lühendid),
2. asesõnadel on lisaks alaliigi märgend (vt allpool Asesõna e pronoomeni (P) alaliik),
3. võivad esineda ka valikulised märgendid (vt allpool Valikulised märgendid).

## Sõnaliikide lühendid

- **A** - omadussõna - (adjektiiv), nii käänduvad kui käändumatud, nt *kallis* või *eht*
- **D** - mäarsõna (adverb), nt *kõrvuti*
- **G** - genitiivatribuut (käändumatu omadussõna), nt *balti*
- **I** - hüüdsõna (interjektsioon), nt *tere*
- **J** - sidesõna (konjunktsioon), nt *ja*
- **K** - kaassõna (pre/postpositsioon ehk adpositsioon), nt *kaudu*
- **N** - arvsõna (numeraal), nt *kaks* või *teine*
- **P** - asesõna (pronoomen), nt *see*
- **S** - nimisõna (substantiiv), nt *asi*
- **V** - tegusõna (verb), nt *lugema*
- **X** - verbi juurde kuuluv sõna, millel eraldi sõnaliigi tähistus puudub, nt *plehku*
- **Y** - lühend, nt *USA*
- **Z** - kirjavahemärk, nt - , . ? !
  
- **T** – morfoloogilise analüsaatori jaoks tundmatu sõna

## Noomeni- e käandsõna kategooriate lühendid

- **Sõnaliik**: nimisõnad ehk substantiivid (**S**), omadussõnad ehk adjektiivid (**A**), arvsõnad ehk numeraalid (**N**), asesõnad ehk pronoomenid (**P**) ja lühendid ning akronüümid (**Y**);

- **Arv**:

**sg** – ainsus e singular,

**pl** – mitmus e plural.

- **Kääne**:

<b>nom</b>	nominatiiv	nimetav
<b>gen</b>	genitiiv	omastav
<b>part</b>	partitiiv	osastav
<b>adit</b>	aditiiv	suunduv (lühike sisseütlev), nt. "majja"
<b>ill</b>	illatiiv	sisseütlev
<b>in</b>	inessiiv	seesütlev
<b>el</b>	elatiiv	seestütlev
<b>all</b>	allatiiv	alaleütlev
<b>ad</b>	adessiiv	alalütlev
<b>abl</b>	ablatiiv	alaltütlev
<b>tr</b>	translatiiv	saav
<b>term</b>	terminatiiv	rajav
<b>es</b>	essiiv	olev
<b>abes</b>	abessiiv	ilmaütlev
<b>kom</b>	komitatiiv	kaasaütlev

- Omadussõna e adjektiiv (A) **võrre**:
  - pos** – algvõrre,
  - comp** – keskvõrre,
  - super** – ülivõrre.
  
- Arvsõna e numeraali (N) **liik**:
  - card** – põhisõna,
  - ord** – järgarvsõna,
  - digit** – (araabia) number,
  - roman** – rooma number.
  
- Asesõna e pronomeni (P) **alaliik**:
  - pers** – personaalpronoomen e isikuline asesõna,
    - personaalpronoomeni e isikulise asesõna **isik**:
      - ps1** – esimene isik,
      - ps2** – teine isik,
      - ps3** – komas isik,
  - dem** – demonstratiivpronoomen e näitav asesõna,
  - indef** – indefiniitpronoomen e umbmäärane asesõna,
  - pos** – possessiivpronoomen e omastav asesõna,
  - inter** – interrogatiivpronoomen e küsiv asesõna,
  - rel** – relatiivpronoomen e siduv asesõna,
  - refl** – refleksiivpronoomen e enesekohane asesõna,
  - rec** – retsiiprookpronoomen e vastastikune asesõna,
  - det** – determinatiivpronoomen e määratlev asesõna.
  
- Lühendi e akronüümi (Y) **liik**:
  - nominal** – nimisõnaline,
  - adjectival** – omadussõnaline,
  - adverbial** – mäarsõnaline,
  - verbal** – tegusõnaline.

## Muutumatu sõnade kategooriate lühendid

- Sidesõna e konjunktsiooni (J) **liik**:
  - crd** – koordineeriv e rinnastav sidesõna,
  - sub** – subordineeriv e e alistav sidesõna;
- Kaassõna e adpositsiooni (K) **liik**:
  - post** – postpositsioon e tagasõna,
  - pre** – prepositsioon e eessõna.

## Verbi- e tegusõna kategooriate lühendid

- **Sõnaliik**: tegusõnad ehk verbid (V);
  
- **Põhiverb, modaalverb, abiverb**:
  - main** – põhiverb,
  - mod** – modaalverb,
  - aux** – abiverb.

- **Arv:**
  - sg** – ainsus e singular,
  - pl** – mitmus e plural.
- **Isik:**
  - ps1** – esimene isik,
  - ps2** – teine isik,
  - ps3** – komas isik.
- **Aeg** (täis- ja enneminevik moodustatakse *olema* abiverbi ja mineviku partitsiibi *past partic* abil):
  - pres** – preesens e olevik,
  - impf** – imperfekt e lihtminevik.
- **Tegumood:**
  - ps** – isikuline,
  - imps** – umbisikuline.
- **Kõneviis:**
  - indic** – kindel,
  - cond** – tingiv,
  - imper** – käskiv,
  - quot** – kaudne.
- **Kõneliik:**
  - af** – jaatav,
  - neg** – eitav.
- Verbi **käändelised vormid:**
  - inf** – *da*-infinitiiv,
  - sup** – supiin:
    - ill** – *ma*-supiin,
    - in** – *mas*-supiin,
    - el** – *mast*-supiin,
    - tr** – *maks*-supiin,
    - abes** – *mata*-supiin,
  - ger** – *des*-vorm ehk gerundiiv,
  - partic pres ps** – *v*-kesksõna,
  - partic pres imps** – *tav*-kesksõna,
  - partic past ps** – *nud*-kesksõna,
  - partic past imps** – *tud*-kesksõna.

## Valikulised märgendid

Morfoloogiliste põhimärgendite järel enne süntaktilist märgendit võib esineda lisamärgendeid:

<cap> - sõna algab suure tähega

Kaassõna reksioon:

<nom> - nominatiivse reksiooniga kaassõna

<gen> - genitiivse reksiooniga kaassõna

<part> - partitiivse reksiooniga kaassõna

<el> - elatiivse reksiooniga kaassõna

<term> - terminatiivse reksiooniga kaassõna

<kom> - komitatiivse reksiooniga kaassõna

<FinV> - verbi pöördeline vorm

partic - adjektiivne partitsiip  
<nud> - adjektiivne nud-partitsiip  
<tud> - adjektiivne tud-partitsiip  
<tav> - tav-partitsiip  
<v> - v-partitsiip  
<mata> - verbi -mata-vorm  
<mine> - mine-liiteline nimisõna

Verbi sihitise rektsioon:

<Intr> - intransitiivne verb (üldjuhul pole sihiline)  
<NGP-P> - sihiline verb, sihitis võib olla nimetavas, omastavas või osastavas käändes  
<NGP> - sihiline verb, sihitis peab olema nimetavas, omastavas või osastavas käändes  
<Part-P> - sihiline verb, sihitis võib olla osastavas käändes  
<Part> - sihiline verb, sihitis peab olema osastavas käändes  
<InfP> - sihiline verb, sihitis võib olla da-infinitiiv

Lisaks verbi rektsioone:

<Ill> - illatiivne  
<In> - inessiivne  
<El> - elatiivne  
<All> - allatiivne  
<Ad> - adessiivne  
<Abl> - ablatiivne  
<Tr> - translatiivne  
<Ter> - terminatiivne  
<Es> - essiivne  
<Kom> - komitatiivne

Kirjavahemärk:

Exc - hüüumärk  
Int - küsimärk  
Fst – punkt  
Ell – kolmpunkt  
Els – kakspunkt  
Com - koma  
Scl - semikoolon  
Col - koolon  
Dsh - mõttekriips  
Dsd – topeltkriips  
Opr - avav ümarsulg  
Cpr - sulgev ümarsulg  
Quo - jutumärk  
Oqu - avav jutumärk  
Cqu - sulgev jutumärk  
Osq - avav kandiline sulg  
Csq - sulgev kandiline sulg  
Sla - kaldkriips

CLB – osalause piir

CLBC – oletatav osalause piir

# Süntaktilised funktsioonid

@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 (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)

@X perifrastilise öeldise koosseisus olev jäänukvorm, mis on morf analüüsil saanud sõnaliigi märgendiks \_X\_ (**mahti** saama, **plehku** panema)

@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-infinitiiiv järeltäiendina (soov **õppida**)

@INFN> da-infinitiiiv 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.

Süntaktiliste funktsioonide määramisel on lähtutud EKG-st:

<http://dspace.utlib.ee/dspace/handle/10062/29437>

## Sõltuvussuhted

Sõltuvussuhted on esitatud # järel, kõigepealt on sõna järjekorranumber lauses ja -> järel selle sõna ülemuse number.