Estonian Universal Dependencies v 1.4

December 2016

Introduction

This dependency treebank is part of the *Universal Dependencies* initiative
http://universaldependencies.org/
NB! This version of Estonian UD follows the *Universal Dependencies* v1 guidelines. *Universal Dependencies* v2 annotation scheme was published in December 2016 and Estonian UD will be converted to it during 2017.
The Estonian UD treebank is based on the Estonian Dependency Treebank
https://github.com/EstSyntax/EDT, created at the University of Tartu.

The morphological and syntactic annotation of the Estonian UD treebank is created through automatic conversion of EDT data tu UD format and manual post-editing of the outcome.

Tokenization

The tokenization of the Estonian UD treebank is a straightforward whitespace-based tokenization with conventional separation of punctuation. The current version of Estonian UD treebank does not contain multiword tokens.

Estonian POS tags

ADJ adjective
ADP adposition
ADV adverb
AUX auxiliary verb
CONJ coordinating conjunction
INTJ interjection
NOUN noun
NUM numeral
PRON pronoun
PROPN proper noun
PUNCT punctuation
SCONJ subordinating conjunction
SYM symbol
VERB verb
X other

Estonian Features

Abbr=Yes abbreviation

AdpType=Post adposition: postposition
AdpType=Prep adposition: preposition Abbr=Yes abbreviation
Case=Abe case: abessive
Case=Abl case: ablative
Case=Add case: aditive (short illative)
Case=Ade case: adessive
Case=All case: allative
Case=Com case: comitative
Case=Ela case: elative

Case=Ess case: essive
Case=Gen case: genitive
Case=Ill case: illative
Case=Ine case: inessive
Case=Nom case: nominative
Case=Par case: partitive
Case=Ter case: terminative
Case=Tra case: translative
Connegative=Yes verb as a part of a negative verb form; e.g. *tee* in *ei tee.*

Degree=Cmp degree: comparative
Degree=Pos degree: positive
Degree=Sup degree: superlative
Foreign=Yes
Hyph=Yes
Mood=Cnd mood: conditional
Mood=Imp mood: imperative
Mood=Ind mood: indicative
Mood=Qot mood: quotative
Negative=Neg negation
Number=Plur number: plural
Number=Sing number: singular
NumForm=Digit numeral form: written as digits
NumForm=Letter numera forml: written as words NumForm=Roman numeral form: written as Roman numbers
NumType=Card numeral type: cardinal
NumType=Ord numeral type: ordinal
Person=1 person: 1
Person=2 person: 2
Person=3 person: 3
Poss=Yes possessive (for pronouns only)
PronType=Dempronoun: demonstrative
PronType=Ind pronoun: indefinite
PronType=Int pronoun: interrogative
PronType=Prs pronoun: personal
PronType=Rcp pronoun: reciprocal
PronType=Rel pronoun: relative
PronType=Tot pronoun: total or collective
Reflex=Yes reflexive (for pronouns only)
Tense=Past tense: past
Tense=Pres tense: present
VerbForm=Fin verb form: finite
VerbForm=Ger des-form

VerbForm=Inf verb form: infinite (da-infinitive)
VerbForm=Part verb form: participle
VerbForm=Sup verb form: supine (ma-infinitive)
Voice=Act active verb form
Voice=Pass passive (impersonal) verb form

Estonian Relations

`root`: root

Core dependents of clausal predicates

`nsubj`: nominal subject of a non-copular clause: *Kass nägi koera.*
`nsubj:cop`: nominal subject of a copular clause: *Kass on triibuline.*
`csubj:` clausal subject: *Tüdrukule meeldib tantsida. Tundus, et oleme asjast aru saanud.*
`csubj:cop:` clausal subject in copular sentences: *Laenu on kerge võtta. Tema sõnul on väheusutav, et vaatajate arv edaspidi tõuseks.*
`dobj:` nominal direct object: *Kass nägi koera.*
Note that in Estonian an impersonal passive clause does not have a subject, and what in English would be considered as passive subject, is in Estonian a direct object: *Kook söödi ära. 'The cake was eaten up'*
`xcomp:` open clausal complement. In the current version of Estonian UD:
- all infinitival complements of a verb, except da-infinitival predicative (subject complement): *hakkan tegema, jäi magama, ajab nutma, tahan teha*
- secondary predicates: nominal or adjectival subject and object complements in translative and essive case forms: *President nimetas Juhani ministriks; Ta tahtis saada rikkaks; Need majad on luksusliku eluviisi võrdpildiks; See tundus meile olulisena.*
`ccomp:` – clausal complement, mainly complementing verbs of communication (e.g. *ütlema*) or mental activity (e.g. *arvama*). Most common complementizer is *et*.
Clauses complementing nouns are labelled `acl:relcl`.
- da-infinitival predicative: *Mõlema hobi on kassipilte netti riputada.*

Non-core dependents of clausal predicates nmod advcl advmod neg
`advcl` adverbial clause modifier
`advmod` adverbial modifier; also question particle *kas*.
`advmod:quant` adverbial or nominal quantifier. For numeral quantifiers the type nummod is used.
`neg` *ei, ära, ärge, ärgem, ärgu* as part of a negative verb form
`nmod` nominal modifier

Noun dependents
`acl` clausal modifier of noun
`acl:relcl` relative clauses, also clauses complementing nouns
`amod` adjectival modifier
`appos` appositional nominal modifier
`nmod` nominal modifier
`nummod` numeric modifier
`det` determiner, in this version of Estonian UD words like s*ee, too, ise, oma, kõik, esimene* (as pronoun) , *teine* (as pronoun), *kolmas* (in plural, e.g. *kolmandad riigid*), *miski, nihuke, sihuke, siuke, teistsugune, minusugune, meiesugune, temasugune* are annotated using that relation.

Adpositions
`case` adpositions

Compounding and unanalyzed
`compound` numerical expressions consisting of multiple tokens. The last word of the numerical expression is the governor, and the other numbers are its dependents.
`compound:prt` particle component of a particle verb
`name` name; proper names constisting of more than one word. The last word is the head and the other words are its direct dependents.

`foreign` foreign words (not loanwords or even direct loans)

Special clause dependents

`aux` auxiliary; *olema* 'be' in compound tense forms; *saama* ' can', *võima* 'may', *pidama* 'must' in modal constructions.

`cop` copula; verb *olema* 'be' in copular clauses with nominal or adjectival predicative.

discourse: discourse element; interjections.

`mark` marker; subordinating conjunctions, comparative conjunctions, interrogative pronouns. In this version following words in the beginning of a clause or in comparative construction are marked with this dependency type: *ehkki , et , justkui , kuhu , kui , kuidas , kuigi , kuivõrd , kuna , kuni , kus , kusjuures, kust , kustkohast , miks, mil , millal , milleks , nagu , otsekui , selmet , sest* and also abbreviation *st* in the beginning of a clause.

`punct` punctuation

`vocative` vocative

Coordination

`cc` coordinating conjunction; the preceding coordinated word is the head.

`cc:preconj` preconjunction marks the first part of those two-part coordinating conjunctions where the two parts are separated by coordinated elements. In the current version following words are annotated using that type: *nii | niihästi* | niivõrd (conjunct: kui); kas (või); küll (küll); nii | sellepärast (et); selle asemel | vaatamata | hoolimata | enam (et); siis | samal ajal (kui); nii (nagu)

`conj:` conjunct; first coordinated element is the governor and the rest its dependents. cop: copula; verb olema in copular clauses with nominal or adjectival predicative.

Loose joining relations

`list` numbers or letters introducing list elements

`parataxis` in sentences with direct speech, the main verb of reporting clause is annotated using that dependency type. Also sentences containing reported speech, but having no complementizer word are annotated using `parataxis`.