

Estonian Universal Dependencies

30.12.2015

Introduction

The Estonian UD treebank is based on [the Estonian Dependency Treebank](#), created at the University of Tartu. Treebank consists of approximately 400,000 words in 30,000 sentences and covers the text classes of fiction, journalism and scientific texts.

The morphological and syntactic annotation of the Estonian UD treebank is created through a conversion of EDT data.

Tokenization

The tokenization of the Estonian UD treebank is a straightforward whitespace-based tokenization with conventional separation of punctuation. The 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

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
Degree=Cmp	degree: comparative
Degree=Pos	degree: positive
Degree=Sup	degree: superlative
InfForm=Inf	infinite verb form: da-infinitive
InfForm=SupAbe	infinite verb form: abessive case form of supine (mata-form)
InfForm=SupEla	infinite verb form: elative case form of supine(mast-form)
InfForm=SupIll	infinite verb form: illative case form of supine (ma-form)
InfForm=SupIne	infinite verb form: inessive case form of supine (mas-form)
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=Inf verb form: infinite (da-infinitive)
 VerbForm=Part verb form: participle
 VerbForm=Sup verb form: supine (ma-infinitive)
 VerbForm=Trans verb form: transgressive (des-form)
 VerbType=Intr intransitive verb
 VerbType=NGP aspectual verb; a transitive verb that takes direct object in nominative, genitive or partitive case form
 VerbType=Part partitive verb; a transitive verb that takes direct object in partitive case form
 Voice=Act active verb form
 Voice=Pass passive (impersonal) verb form

Estonian Relations

acl: clausal modifier of noun; used only for infinite clauses in this version of Estonian Universal Dependencies

advcl: adverbial clause modifier; used only for comparative constructions and infinite adverbial clauses in this version of Estonian Universal Dependencies

advmod: adverbial modifier; used only for infinite clauses in this version.

advmod: quant: adverbial or nominal quantifier. For numeral quantifiers the type nummod is used.

amod: adjectival modifier

appos: appositional nominal modifier

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

case: case marking; adpositions.

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)

ccomp: clausal complement. Only da-infinitival predicative (subject complement) is labelled as ccomp in this version.

compound: 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: particles in particle verbs

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.

csubj: clausal subject; in this Estonian UD version only da-infinitival clausal subject are annotated using that dependency type.

dep: unspecified dependency. In this Estonian UD version all finite subclauses (i.e. their governors) are labelled dep.

discourse: discourse element; interjections.

dobj: nominal direct object. In Estonian a passive clause does not have a subject, and what in English would be considered the passive subject, is in Estonian the direct object

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: *nagu, kui, siis, miks, kuidas, millal, mil, kus*.

name: name; proper names consisting of more than one word. The last word is the head and the other words are its direct dependents. Person names and multiword place names (e.g. New York) are annotated using that dependency type.

neg: negation modifier; *ei* as part of a negative verb form.

nmod: nominal modifier

nsubj: nominal subject of a non-copular clause

nsubj:cop: nominal subject of a copular clause

nummod: numeric modifier

parataxis: parataxis; in sentences with direct speech, the main verb of reporting clause is annotated using that dependency type.

punct: punctuation

root: root

voc: vocative

xcomp: open clausal complement. In current version:

all infinitival complements of a verb, except da-infinitival predicative (subject complement), which is annotated using the type ccomp.

secondary predicates: nominal or adjectival subject and object complements in translative and essive case forms.