

Korpuse formaadi selgitus

Korpuse tekstinäide:

"<s>"

"<Feministid>"

"feminist" Ld S com pl nom cap @SUBJ #1->2

"<tõstavad>"

"tõst" Lvad V main indic pres ps3 pl ps af @FMV #2->0

"<pead>"

"pea" Ld S com sg part @OBJ #3->2

"<ja>"

"ja" L0 J crd CLB @J #4->7

"<mõned>"

"mõni" Ld P pl nom @SUBJ #5->7

"<neist>"

"tema" Lst P pers ps3 pl el @<NN #6->5

"<räägivad>"

"rääki" Lvad V main indic pres ps3 pl ps af @FMV #7->2

"<,>"

"," Z Com CLB #8->8

"<et>"

"et" L0 J sub @J #9->11

"<naistele>"

"naine" Ltele S com pl all @ADVL #10->11

"<pole>"

"ole" L0 V main indic pres ps neg @FMV #11->7

"<mehi>"

"mees" L0 S com pl part @SUBJ #12->11

"<üldse>"

"üldse" L0 D @ADVL #13->11

"<vaja>"

"vaja" L0 D @Vpart #14->11

"<.>"

"," Z Fst #15->15

"<s>"

Näite seletus. Tekst jookseb ülevalt alla, ühel real on „< ja >“ vahel tekstisõna; järgmisel, nelja

tühikuga algaval real jutumärkide vahel algvorm e lemma, L järel morfoloogilised tunnused ja lõpud, seejärel sõnaliigi ning sõnavormis kodeeritud grammatiliste kategooriate märgendid. Morfoloogiliste märgendite loendi leiab siit: <http://www.cl.ut.ee/korpused/morfliides/seletus>

Pangem tähele, et alakriipsud sõnaliigi märgendi ümber selle korpuse märgenduses puuduvad.

Morfoloogilisele märgendile järgneb @-ga algav süntaktilise funktsiooni märgend, mille loendi leiab märgendusjuhendi algusest.

Pindsüntaktilisele märgendile järgneb # järel sõna järjekorranumber lauses ning -> järel tema „ülemuse“, st peasõna number.

<s> tähistab lause algust ja </s> lause lõppu.

Käesolevast korpusest on välja jäetud kogu valentsimärgendus, mis kitsenduste grammatika analüsaatori väljundis paikneb noolsulgude (<>) vahel.