LVTB un UD lomu atbilsme

Atkarības

ko mums darīt?
kā man būt?
visiem pieceļties!
labi un lēti ir nesavienojamas lietas

nomens šeit ir n.*, a.*, m.*, v..pd.*, p.*, x.*, y[npa].*
palīgvārdi šeit šobrīd ir saikļi un izsauksmes vārdi
ja reducētam verbam (ne divdabim) nav zināma kārta, pieņem darāmo

***saulei** lecot, devāmies ceļā*
*spuldzes, **sensoru** iedarbinātas, uzreiz iedegās*
*viņš, **pats** darīdams, izdarīja labāk*
*viņš, **pats** salijis, neļāva liedēties citiem*

dzīve ir skaista un *dzīvot ir mīlēt*
ir darāmā kārta

iebraukt ir aizliegts

Akad. gram. tiepīgi saka, ka *pētījuma mērķis ir noteikt, kā...*
teik. pr. ir *noteikt*

pārliecība ir, ka jāgādā programmas

Saitiņa + *tāds/tas/viss*

Problēma:
Kā LVTB automātiski identificēt nenoteiksmes subjektus?

nomens šeit ir n.*, a.*, v..pd.*, p.*, x.*, y[npa].*

riņķiši zem acīm, atsauce uz direktīvu

*boss ieteica **padoties***
*viņš lika bērniem **ēst***

es atgriezos peldēt
dzīvot ir mācīties mīlēt
censties pabeigt skolu šobrīd ir svarīgākais
dzīvot ir tiekties sasniegt mērķus
līdzšinējā pieredze viņam ļaujot secināt...
zīmēt gribošais bērns

viņam piemīt spēja ātri rīkoties
nomens šeit ir n.*, a.*, p.*, x.*, y[npa].*

Aizķerot trolejbusa vadus, cirka ēka var sabrukt
ir valodas kļūda, bet bieža.

Vajag kokā pārvietot līmeni augstāk.

viņa, dūšīgi paēdusi, atgriezās pie darba
suns, svaigi sarāts, noslēpās zem galda
redzēju tevi, ļoti starojošu

nomens šeit ir n.*, a.*, v..pd.*, p.*

kontroles atklātie pārkāpumi
***vienam** cīnīties ir grūti* *Var būt problēmas ar x.* un y.* locījuma "minēšanu".*

panākumi zinātnēs, seminārs Rīgā (kad lok. nav apstāklis)
*šie ļaudis **paši** bija pārdodamo zemeņu audzētāji*

Adjektīviem, lok. divdabjiem, skaļa v., ya.*
būtu jābūt pie atbalsta vārda.

mātes** grāmata, viedoklis **par to

Nākotnei:
Varētu pārskatīt ar xx.* un xf.*
izteiktos attr un novērtēt, kas
skaitliski tie biežāk ir.

kaut kas notiks 22.00
pantu numuri juridiskos dokumentos

xSimile: *vairāk kā/nekā [pieci]*

nomens šeit ir n.*, a.*, v..pd.*, p.*

kontroles atklātie pārkāpumi
***vienam** cīnīties ir grūti* *Var būt problēmas ar x.* un y.* locījuma "minēšanu".*

"Washington (CNN)" nevaram identificēt

Pēc līdzības ar daļu *etc.* angļu valodā un daļu *o.v.* vācu valodā

advmod:emph nav obligāts,
Izvēlēts pēc līdzības ar PDT

Vispārējie izteicēju principi:
* fāzes un modālie (sākšana, beigšana, varēšana) izteicēji ir izteicējs + xcomp, kur izteicējs = sākšana/beigšana/varēšana, bet xcomp = basElem.
* saitiņas funkcijā UD lieto tikai *(ne)būt*, pārējos gadījumos izteicējs + xcomp.
* ciešamā kārta ir *(ne)tikt/tapt/būt* + -ts divdabis.
* darāmā kārta ir -is divdabis.
* -ošs, -ams divdabis saliktos laikus neveido, -ts divdabis var veidot nominālu izteicēju.
* reizēm ir grūti atšķirt -ts divdabja izteicējus – ciešamā kārta vai sastata – tad lieto tagu.

Te ciešamā kārta nemēdz būt.

Sk. *LV-UD-annotations.doc*, bet AUX ir *(ne)būt*, *(ne)tikt*, *(ne)tapt* ar tagiem vc.* un va.*

Izņēmumi, ko nedala sīkāk:
xn, xo
u.t.jpr., u.c., u.tml., v.tml.

Diskusija: jāuzmanās, jo UD goeswith ir domāts tikai kļūdaini sadalītiem vārdiem

Aprakstīts LV-UD-annotations.docx

Situācija, kad cilvēks pārlabo iepriekš teikto. LVTB nav šādu piemēru, mūsu modelī nav izdomāts, kā marķēt.

Izteicēja atkarība no mākslīgas koka virsvirsotnes. Šāda loma kokā var būt tikai viena un, ja citādi galīgi nesanāk, tad UD vadlīnijas liek kreisāko no tiem mākslīgi padarīt par vienīgo *root*.

Kad visi striķi trūkst: kaut ko nevar konvertēt vai parseris devis nepilnu rezultātu.

Universal Dependencies

Core arguments

- Nominals
  - nsubj
    - nsubj:pass
  - obj
  - iobj
- Clauses
  - csubj
    - csubj:pass
  - ccomp
  - xcomp

Nominal dependents

- Nominals
  - nmod
  - appos
  - nummod
- Clauses
  - acl
- Modifier words
  - amod
- Function words
  - det
  - clf *(liekas, ka nav latviešu valodā?)*
  - case

Non-core dependents

- Nominals
  - obl
  - vocative
  - expl *(nav latviešu valodā)*
  - dislocated
- Clauses
  - advcl
- Modifier words
  - advmod
    - advmod:emph
    - advmod:neg
  - discourse
- Function words
  - aux
    - aux:pass
  - cop
  - mark

Frāzes loma

Not a dependency in the narrow sense

- Coordination
  - conj
  - cc
- MWE
  - fixed
  - flat
    - flat:foreign
    - flat:name
  - compound
- Loose
  - list *(nav LVTB marķēts)*
  - parataxis
- Special
  - orphan
  - goeswith
  - reparandum
- Other
  - punct
  - root
  - dep

subj

nomens, apst. v., palīgvārds
pie
pred vai verbāla *spc* ne cieš.k-ā (nenot., divd., saitiņizt.);
nomens, apst. v., palīgvārds pie *basElem* dar.k-ā

nomens, apst. v., palīgvārds
pie
cieš. k-as *pred* vai cieš. k-as *basElem*, vai cieš. k-as *spc*

nenoteiksme pie *pred* ne cieš. k-ā;
nenoteiksme pie *basElem* dar. k-ā

nenoteiksme pie *pred* cieš. k-ā vai *basElem* cieš. k-ā

obj

jebkurš locījums pie a.*

akuzatīvs;
nominatīvs pie vajadzības izt.

cits locījums vai *xPrep*

subjCl

pie *xPred* ne cieš. kārtā vai *pred* dar. kārtā

pie cieš. k-as *pred* vai cieš. k-as *xPred*

pie *subj*

predCl

pakārtots visam izteicējam

pakārtots *xPred* daļai

objCl

spc

par + nomens pie darb. v

xPrep pie darb. v., kas nav lok. divd., vai *xPred*

xPrep pie nomena

(±*pmc*) *xSimile* pie lietv., pie vietniekv., pie y[np].*

(±*pmc*) citi xSimile

(±*pmc*) nenot. pie darb. v., kas nav lok. divd., vai *xPred*, nenot. subj. ≠ izt. subj.;
esošais marķējums neļauj nošķirt xcomp

(±pmc) nenot. pie darb. v., kas nav lok. dvid., vai xPred, nenot. subj. = izt. subj.;
(±*pmc*) nenot. pie lok. divdabja

(±*pmc*) nenoteiksme pie nomena, kas nav lok. divd.

(±*pmc*) -dams divdabji

(±*pmc*) -am, -ot divdabji

spcPmc ar *xPrep basElem* pie nomena

spcPmc ar *xPrep basElem* citur

spcPmc ar apst. v., yr.* *basElem*

spcPmc ar lietvārdu, y[np].* *basElem* pie darb. v., kas nav lok. divd.

spcPmc ar locījumā sask. lietvārdu, y[np].* *basElem* pie (n|pp|y[np]).* **pēc** tā

spcPmc ar lietvārdu, y[np].* *basElem* pie vietniekv. **pirms** tā

spcPmc ar lok. divd. *basElem* pie darb. v., kas nav lok. divd., vai *xPred*

spcPmc ar augstāk neaprakstītu lietvārdu, y[np].* *basElem*;
spcPmc ar īp. v., lok. divd., ya.* *basElem*

vietniekv. ar *cits*, *cita* pie *xPrep* ar vai pie plika vietniekv. ar *cits*, *cita* vai pie *xPrep*;
[mp].* ar *viens*, *viena* pie *xPrep* ar vai pie plika [mp].* ar *otrs*, *otra* vai pie *xPrep*;

attr

nomens ģen.;
rez., y[npa].* pie cieš. k. pag. divd.

nomens pie lietvārda, vietniekvārda, y[np].*;
vietniekv. pie *xPred* ar tipu *subst*

vietniekv. pie verba vai divd., vai cita *xPred*;
lietv. ne nominatīvā, y[np].* pie verba vai divd., vai *xPred*

lietv. nominatīvā pie verba vai divd., vai *xPred*;
īpašības v., lok. divd., skaitļa v., ya.* pie verba vai divd., vai *xPred*

xPrep; lietvārdi; y[np].*; %; xu.*; pp.* (valid. dēļ)
xf.* *(nevar atšķirt adjektīvus)*
vietniekv. pēc vecāka *(nevar atšķirt dzeju)*

skaitļa vārdi; skaitļi

adjektīvi; divdabji; ya.*;
kārtas skaitļi; kārtas skaitļa vārdi;
*xx.** *(varbūt ka te iekrīt kāds* nmod*)*

vietniekvārdi;
manējais u.c.; *daudzi; vairāki*

adverbi; yr.*

attrCl

appCl

adv

mc.*, xn.*, iekaitot attiecīgus *xPrep*

lietvārdi; vietniekv.; y[npa].*;
kārtas skaitļi; xo.*; *ne visi* xf.*;
xPrep

adverbi; yr.*; %, gram. xSimile, *daļa xf.**

partikulas; yd.*

sit

det

placeCl , timeCl
manCl , degCl
causCl , purpCl
condCl , cnsecCl
cncesCl , motivCl
compCl

quasiCl

ins

basElem vai pats ir z.*

ar *pred* iekš *pmc;*
ar *crdClauses* iekš vienīgā *basElem* iekš *pmc*;
punct ir iekava ()[]

ar citu *basElem* iekš *pmc*; pārējie bez *pmc*

dirSp

nc

sastāv no *address*

utt., *u.t.jpr.*, *u.c.*, *u.tml.*, *v.tml.*

partikulas pie lietv., pie vietniekv., pie y[np].*

citas partikulas, izs. v.; xx.*; citi yd.*; *interj*

repeat

xPred

Vispārējie izteicēju principi:
* fāzes un modālie (sākšana, beigšana, varēšana) izteicēji ir izteicējs + xcomp, kur izteicējs = sākšana/beigšana/varēšana, bet xcomp = basElem.
* saitiņas funkcijā UD lieto tikai *(ne)būt*, pārējos gadījumos izteicējs + xcomp.
* ciešamā kārta ir *(ne)tikt/tapt/būt* + -ts divdabis.
* darāmā kārta ir -is divdabis.
* -ošs, -ams divdabis saliktos laikus neveido, -ts divdabis var veidot nominālu izteicēju.
* reizēm ir grūti atšķirt -ts divdabja izteicējus – ciešamā kārta vai sastata – tad lieto tagu.

xPred ar mod

- mod
- basElem

xPred bez mod

- auxVerb+ — AUX
- auxVerb — nav AUX; AUX, xPred = dar. kārta; AUX, xPred = cieš. kārta; AUX, xPred = sastata izt.
- basElem — vienīgais auxVerb nav AUX; Visi auxVerb ir AUX

Vārdi ar atstarpēm

Vārdforma satur ' '

- *pirmais*
- *otrais+* — koptags xu.*, xx.*; koptags xf.*; pārējie gadījumi

UD: *fixed* ir īsi, gramatikalizējušies elementi ar fiksētu struktūru, savukārt *flat* ir "semi-fixed" vai "headless (exocentric)". *compound* ir "headed (endocentric)" struktūra. Savukārt, salīdzinot *compound* ar *nmod – compound* ir stabils (?).

Vārdi (bez xPred)

xApp

- basElem+
- basElem

xPrep

- prep
- basElem

phrasElem

- basElem
- basElem+

subrAnal *vv*

- basElem
- basElem+

subrAnal *ipv*

- basElem — *pēdējais a.*, ya.*, v..pd.**; *vietniekv.*

subrAnal *skv*

- basElem — *pirmais p.*, ja p.*[skv.*];*
pirmais (mc|xn)., ja (mc|xn).*[skv.*]*; *mc.*, xn.**; *p.**

subrAnal *set*

- basElem
- *xPrep* — *lietvārds, vietniekv., y[cn].**; *mc.*, xn.**; *īpašības v., mo.*, ya.*, xo.*, v..pd.**

subrAnal *sal*

- *xSimile*
- basElem

xFunctor

- basElem
- basElem+

coordAnal

- basElem
- basElem+

xNum

- basElem+
- basElem

xParticle

- no — *[aff.*]*; *[neg.*]*
- basElem

xSimile

- conj — gramatizējies *xSimile*; *xPred*, *utter* sastāvdaļa; spc lomā vai *spcPmc*, *insPmc* sastāvdaļa
- basElem

namedEnt

- basElem
- basElem+ — z.*; citi

unstruct

- basElem
- basElem+ — z.*; citi, ja frāze ir xf.*; citi

Koordinācija

crdClauses

- crdPart
- crdPart+ — pirms šīs daļas ir ';'; pirms šīs daļas nav ';'
- punct
- conj

crdParts

- crdPart
- crdPart+
- punct
- conj

PMC

utter

- basElem — pirmais + utter ir koka saknē; pirmais + utter nav koka saknē; nākamais, atdalīts ar *punct*
- conj — cs.*; cc.*
- no — sastāv no *adress*; n...v.*; partikula; izsauksmes v.; *interj*
- punct

sent, quot

- pred / basElem — *sent*, *quot* ir koka saknē; *sent*, *quot* nav koka saknē
- conj — cs.*; cc.*
- no — sastāv no *adress*; n...v.*; partikula; izsauksmes v.; *interj*
- punct

mainCl
subrCl
insPmc
dirSpPmc

- pred / basElem
- conj — cs.* c.* iekš *subrCl*; cc.*
- punct
- no — sastāv no *adress*; n...v.*; q.*; i.*.; *interj*

spcPmc
particle
addressp

- basElem
- punct

interj

- basElem
- basElem+
- punct

Pārdomas:
Nez vai šis ir adekvāti arī vientuļiem sakārtojuma saikļiem teikuma vidū?

Apzīmējumi

- LVTB neaktuāla loma
- Loma, kas satopama vairākās vietās
- loma — nosacījums; neatkarīgs nosacījums; nosac-a daļa, vai cita daļa; atkal neatk. nosac-s
- Diskusija
- Komentārs, *piemērs*

padomnieks leitnants Priedītis

tāds priecīgs

abi divi, viens pats, visi trīs kādi pieci

Problēma:
Mūsu frāzes ir gramatiskā līmenī salīdzinoši stipri fiksējušās vai tomēr brīvi lokāmas un papildināmas ar citiem vārdiem? Atkarībā no tā, tās marķējamas kā flat, fixed vai compound.

tādas kā balsis
tāds kā priecīgs
xSimile ir gramatizējies

Gramatizējušamies xSimile – pirmais vārds saņem frāzes lomu, pārējie ir fixed. Pieņēmums, ka gram. xSimile pirmais vienmēr ir basElem

nejūties ne kā bērns, ne kā filozofs

Virsteikumi, kas atdalīti ar komatiem - kā vienlīdzīgi izteicēji.
Virsteikumi, kas atdalīti ar semikolu – **parataxis**.

Problēma:
Vai parataxis ir tikai šie gadījumi?