Harmonization of Dependency Trees

Daniel Zeman, Rudolf Rosa

May 4, 2023
Harmonization of Dependency Tree Style

• Universal Dependencies
  • Solved problem?
  • Almost. But not completely.

• Transformations sometimes needed for learnability experiments
• There are still non-UD treebanks that need to be converted
• SUD (Surface-syntactic Universal Dependencies)

• Transformations trivial?
  • Not necessarily…
To kendte russiske historikere

Two known Russian historians
In an authoritarian regime
During celebrations criticized Na oslavách odsoudil
Adpositions

in diesem Jahr kommen

MO

OC

NK

APP R

PDAT

DET

NN

NOUN

VV INF

VERB

det

case

obl

xcomp
Lahore in closed

- PROPN
- ADP
- ADJ

- n
- psp
- adj

- lwg_psp
- jjmod
- nmod

- case
- obl
- amod
Ельцин считал что стране нужен экономист

Yeltsin realized that by-country needed economist
hasta que se fijen los comicios

until that they fix the elections
una sala ha dovuto essere sgomberata

a room has had to-be evacuated
Verb Groups

Based-on can for-example be determined whether

Op_grond_van kan bijvoorbeeld worden vastgesteld of

Prep_N_Prep V Adv V V Conj

ADP NOUN ADP AUX ADV AUX VERB SCONJ

mod vc mod vc cnj

root aux:pass advmod csubj:pass

obl aux
Coordination: Mel’čuk [en]

apples, oranges and lemons
Coordination: Mel’čuk [de]

Äpfel, Orangen und Zitronen
äpplen, apelsiner och citroner
Coordination: Mel’čuk [ja]

リンゴ と オレンジ と レモン
ringo to orenji to remon

conj
punct
cc
conj
Coordination: Mel’čuk [tr]
అపిలల్, నారింజ మరియు నిమమ్పండు
Coordination: Half-Prague [ro]
Coordination: Obligatory Nested Prague [zh]

蘋果，桔子和檸檬

píngguǒ, júzi hé níngméng
ябълки, портокали и лимони
jabălki, portokali i limoni
Coordination: Half-Stanford [da]

æbler, appelsiner og citroner
Coordination: Tesnière [hu]

alma, narancs és citrom
Interaction of Structures in Prague

vlaky do a z Prahy nebo odLiberec
trains to and from Prague or from Liberec

XXX Coord Coord_M AuxP_M Atr AuxP_M Atr AuxP_M

XXX nmod
Interaction of Structures in Prague

vlaky do a z Prahy nebo od Liberce
trains to and from Prague or from Liberec
Interaction of Structures in Prague

vlaky do a z Prahy nebo od Liberec
trains to and from Prague or from Liberec
Interaction of Structures in Prague

vlaky do a z Prahy nebo od Liberec
trains to and from Prague or from Liberec
interaction of structures in Prague

vlaky do a z Prahy nebo od Liberce trains to and from Prague or from Liberec

Coord nmod case conj cc case conj cc case Atr_M AuxP_M Atr
Interaction of Structures in Prague

Trains to and from Prague or from Liberec
Interaction of Structures in Prague

Trains to and from Prague or from Liberec.
Interaction of Structures in Prague

vlaky do a z Prahy nebo od Liberece
trains to and from Prague or from Liberec
Interaction of Structures in Prague

vlaky do a z Prahy nebo od Liberec
trains to and from Prague or from Liberec

XXX

Atr

conj

nmod
Interaction of Structures in Prague

vlaky do a z Prahy nebo od Liberec
trains to and from Prague or from Liberec
Round-Trip: Shared Dependents

Czech students and teachers

Czech students and teachers

Czech students and teachers

Czech students and teachers
Round-Trip: Nested Coordination

- jablka a hrušky nebo citróny
  - apples and pears or lemons

- jablka a hrušky nebo citróny
  - apples and pears or lemons

- jablka a hrušky nebo citróny
  - apples and pears or lemons
Round-Trip: Auxiliary Verbs

he could have seen her

he could have seen her

he could have seen her
Tree-Rewriting Toolkits

• Udapi
  • http://udapi.github.io/ (main page)
  • https://ufal.mff.cuni.cz/~zeman/vyuka/deptreebanks/NPFL075-working-with-UD.pdf (tutorial)

• Grew
  • https://grew.fr/
Phrase vs. Dependency Trees

S

NP

N

Paul

V
gave

NP

N

Peter

C
two

NP

N

pears

VP

NP

nsubj

Peter

iobj
two

root

obj

nummod

gave

pears
Phrase vs. Dependency Trees

- Phrase trees
  - Usually do not mark the head
  - May not mark the function of the constituent in the superordinate constituent
Phrase vs. Dependency Trees

- **Phrase trees**
  - Usually do not mark the **head**
  - May not mark the **function** of the constituent in the superordinate constituent

- **Dependency trees**
  - Do not show **nonterminals** (phrase types)
    - Nor any other phrase-level features
  - Do not show “how the sentence is generated” (order, recursion, **proximity** of constituents)
Example

[Diagram showing the structure of the sentences:]

Sentence 1: John loves Mary
Sentence 2: Mary loves John

[Diagram with tree structures and labels: nsubj, obj]
Dick Darman, call your office.
Prague Tectogrammatical Layer: Ellipsis

he will meet Mary today and (he) (will meet) Clara tomorrow
he will meet Mary today and (he) (will) (meet) Clara tomorrow