Multilingual Natural Language Processing: Week 10

Daniel Zeman, Rudolf Rosa, Ondřej Bojar
zeman@ufal.mff.cuni.cz
http://ufal.mff.cuni.cz/courses/npfl120
Harmonization of Dependency Tree Style

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

- Transformations sometimes needed for learnability experiments

- Transformations trivial?
  - Not necessarily…
To know the Russian historians.

- **subj**: To
- **nobj**: Two
- **mod**: known
- **mod**: Russian
- **nummod**: historikere
- **amod**: AC
- **amod**: AN
- **amod**: AN
- **nsubj**: Danish Dependency Treebank
Danish Dependency Treebank

I et autoritært regime
In an authoritarian regime
During celebrations, Na odsoudil.
Adpositions

in this year come

in diesem Jahr kommen

APPR PDAT NN VVINF

MO WK OC

NK

NK

det case

obl

xcomp
Adpositions

Lahore in closed

Adpositions
Eльцин считал что стране нужен экономист

Yeltsin realized that by-country needed economist

S PROPN V VERB CONJ SCONJ S NOUN A ADJ S NOUN
Subordinate Clauses

hasta que se fijen los comicios

Daniel Zeman (ÚFAL MFF UK)
Verb Groups

Verb Group Diagram:

- det
- aux
- modal
- aux:pass
- nsubj:pass
- root

Verbs:
- una
- sala
- ha
- dovuto
- essere
- sgomberata

Nouns:
- a
- room

Auxiliaries:
- has
- had

Root:
- evacuated

Categories:
- DET
- NOUN
- AUX
- VERB
Verb Groups

Op_grond_van kan worden vastgesteld of bijvoorbeeld can be determined whether Based-on can for-example be determined whether Prep_N_Prep V Adv V Con conj ADP NOUN ADP V AUX ADV V AUX VERB SCONJ

Daniel Zeman (ÚFAL MFF UK)  Multilingual NLP  NPFL120-10
Coordination: Mel’čuk [en]
Coordination: Mel’čuk [de]
Coordination: Mel’čuk [sv]

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
Coordination: Mel’čuk [tr]

ehma, portakal ve limon
Coordination: Prague [cs]

jablka, pomeranče a citróny
Coordination: Prague [te]
Coordination: Half-Prague [ro]
Coordination: Obligatorily Nested Prague [zh]

蘋果 píngguǒ

，

桔子 júzi

和 hé

檸檬 níngméng

XXX

Daniel Zeman (ÚFAL MFF UK) Multilingual NLP NPFL120-10 20 / 33
Coordination: Stanford [bg]
Coordination: Half-Stanford [da]

æbler, appelsiner og citroner
Coordination: Tesnière [hu]

```
 Coordination: Tesnière [hu]

 alma, narancs és citrom
```

Diagram:

```
XXX PUNCT XXX CONJ XXX
  |
  |   |
  |   |
  |   |
alma  ,  narancs    és  citrom
  |
  |
  |
  |
XXX conj XXX XXX
```

Notes:
- The diagram represents the coordination structure of the sentence in Hungarian: "alma, narancs és citrom" which means "almond, orange and lemon".
- The nodes represent the words and the edges represent the relationships between them.
- The labels on the edges indicate the type of coordination: PUNCT for punctuation, CONJ for conjunction, and cc for coordinating conjunction (választozó kötőkép).
Interaction of Structures in Prague

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

Coord M

AuxP_M

Atr

XXX

-case

conj

cc

nmod
Interaction of Structures in Prague

dvaky
trains
do
to
a
and
z
from
Prahy
nebo
or
Liberce
Liberrec
and
from
nmod
case
cc
conj
AuxP_M
Coord
Coord_M
Atr
XXX
XXX
XXX
Interaction of Structures in Prague

vlaky trains

d do a and z from Prahy Prague

d nebo od from Liberec Liberec

Coord

Coord_M

Atr

conj

cc

conj

case

nmod
Interaction of Structures in Prague

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

Grammar tree:

- Coord
- Coord_M
- Atr
- conj
- cc
- case
- nmod

Daniel Zeman (ÚFAL MFF UK)
Interaction of Structures in Prague

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

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

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

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

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

vlaky (trains) do a and z Prahy (Prague) nebo (or) od Liberec (Liberec)

Daniel Zeman (ÚFAL MFF UK)
Round-Trip: Shared Dependents

čeští studenti a učitelé
Czech students and teachers

čeští studenti a učitelé
Czech students and teachers

Atr

Coord

XXX_M

amod

conj

cc

Atr

Coord

XXX_M

XXX_M

Čeští studenti a učitelé
Czech students and teachers

Daniel Zeman (ÚFAL MFF UK)
Round-Trip: Nested Coordination

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

Daniel Zeman (ÚFAL MFF UK)  Multilingual NLP  NPFL120-10  27 / 33
Phrase vs. Dependency Trees

\[
S \quad \rightarrow \quad NP \quad VP
\]

\[
NP \quad N \quad V \quad NP \quad NP
\]

\[
NP \quad NP \quad N \quad C \quad N
\]

Paul\quad gave\quad Peter\quad two\quad pears

\[
root \quad obj \quad nsubj \quad iobj \quad nummod
\]

Paul\quad gave\quad Peter\quad two\quad pears

Daniel Zeman (ÚFAL MFF UK)  Multilingual NLP  NPFL120-10  28 / 33
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

```
  S
     NP   VP
       N   V   NP
       John loves N
       Mary

  S
     VP
       N   V   N
       John loves Mary

John loves Mary
```
Prague Tectogrammatical Layer

Dick Darman, call your office.
he will meet Mary today and (he) (will meet) Clara tomorrow
Basic vs. Enhanced Universal Dependencies

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