Prague Dependency Treebank: Annotation of Surface Syntax (Part II.)

Markéta Lopatková

Institute of Formal and Applied Linguistics, MFF UK

lopatkova@ufal.mff.cuni.cz
Goal:
• to describe the structure of the sentence and
• to denote the type of relations between "words"


analytical functions (lecture 5):
• predicate … Pred, Pnom, AuxV (, Obj)
• subject … Sb
• attribute … Atr
• object … Obj
• adverbial … Adv
• complement … Atv, AtvV
• combined functions … AtrAtr, AtrObj, ObjAtr, AtrAdv, AdvAtr
Principles of Annotation: Non-dependency

• coordination
• apposition
• parenthesis
Non-dependency: Coordination

- "multiplication" of a single syntactic position (sentence member / sentence)
- **Coord** … 'connecting node'
  - primarily:  - coordinating conjunctions (a, i, ale, nebo, nebot’, proto …)
  - connecting expressions (a tak dále, et cetera, …)
  - comma

- visualization … suffix_{Co}
  … is_member in the data

Later (he) escaped from (the) prison and (he) left (the) republic.
"Payments are good right now, but we aren't recommending any long-term lines of credit," says Richard Hastings …
… , despite the large reserves and write-downs and stiffer regulatory requirements that should be in place by year 's end .

… , i přes velké rezervy a odpisy a přísnější regulační požadavky, které by měly nabýt platnosti do konce roku.
Non-dependency: Coordination (cont.)

all coordinated members … the **same** analytical function (if not Coord)
  • if not possible … ExD_Co (see later)

(He) came alone and immediately
Non-dependency: Coordination (cont.)

Which node is a 'connecting' node?

Karel and Bohouš and Venca were eating

(she) read journals, newspapers and so on
Non-dependency: Coordination (cont.)

- non-standard termination of a coordinated list
- one-member sentential coordination

(I) expect Tony, Charles, Frank ...
Non-dependency: Coordination (cont.)

Specific constructions:

Viktorka won 3:0

Czech – German relations
Non-dependency: Coordination (cont.)

coordinated members with AuxP (or AuxC)

(they) arrive from Slovakia and from eastern countries

(they) operate in Gaza and Jericho

(he) walked to and from (the) forest
Non-dependency: Coordination

coordination vs. apposition

- což `which' (incl. inflected forms and forms with prepositions, as e.g. příčemž)

(It) is denoted as (a) cubism which is (an) ignorance.
Non-dependency: Apposition

analytical function:  **Apos**
- "multiplication" of a single syntactic position (the same referent)
- members of an Apposition … interchangable
  - the **same** function
  - formal condition: typically the same case
- typical connecting expressions
  e.g. jako [as], t.j. [i.e.], (jako) např. [e.g.], ergo, tedy [thus], …
- visualization suffix _Ap vs. is_member in the data

(they) will-respond by a current solution: devaluation

BUT: NO Apos notde in English HamleDT
Non-dependency: Apposition (cont.)

necessarily formally separated, e.g. by comma, brackets, dash ...

(they) dealt with the position of UK (Charles University)

Two powers fight: Love or Harmony and Hatred or Conflict
Non-dependency: Parenthesis

an additional adjunction to the statement in the sentence
• formally marked by some graphic sign(s)
  • bracket, dashes, commas, ...
• is_parenthesis_root in the data
• Cz: visualization: suffix _Pa

I am, unfortunately, short of money

later - I have to admit - I was ashamed
The prime minister's insistence on keeping a private coterie of advisers -- including an economic guru ... -- confused the financial community.
Non-dependency: Parenthesis

an additional adjunction to the statement in the sentence
• formally marked by some graphic sign(s)
  • bracket, dashes, commas, …
• visualization: suffix _Pa vs. is_parenthesis_root in the data

BUT: HamleDT English visualization … as ExD_Co (ellipsis, coordination)

Rewards have been suggested – but never adopted – for filters …

PDT: a-layer II.
Non-dependency: Parenthesis (cont.)

Note:
• AuxP as a 'head' of parenthesis
• parenthesis consisting of several subtrees

he paid her (on a corridor) a compliment

she agreed – after a moment of hesitation expressly – with the proposal
Principles of Annotation: Ellipses

**ellipsis** ~ a deletion of an expression that is expected in the given syntactic structure

- **ExD** ... all nodes that would depend on the deleted node
- exception: prepositions (AuxP) and subord. conjunctions (AuxC) connecting nodes for coordination/apposition (Coord, Apos)
- actual (textual) ellipses

(physician, attachment of one village to another)
Principles of Annotation: Ellipses

**ellipsis** ~ a deletion of an expression that is expected in the given syntactic structure

- **ExD** … all nodes that would depend on the deleted node
- exception: prepositions (AuxP) and subord. conjunctions (AuxC) connecting nodes for coordination/apposition (Coord, Apos)
- actual (textual) ellipses vs. lexicalized ellipses

(I) have to (go) home.  
(It) is five (o'clock)
Principles of Annotation: Ellipses (cont.)

**ellipsis**

- coordination
- one-member sentences (verb-less)

*Christine brought a rose, George violets.*  

*Sunday Lidové noviny.*
Principles of Annotation: Ellipses (cont.)

ellipsis and coordination:

let us mention fields, woods, buildings and other (property)
Specific Constructions: Numeral Expressions

- part expressing the number + part expressing counted objects
- determine the governing and the dependent part … based on the ***morphological analysis* (→ agreement)

\[
\text{viděl čtyři}_{\text{acc}} \text{vrány}_{\text{acc}} \sim \text{viděl vrány}_{\text{acc}} \quad \text{BUT} \quad \text{viděl pět}_{\text{acc}} \text{vran}_{\text{gen}} (*\text{viděl vran}_{\text{gen}})
\]

... *reduction* principle

(he) saw four crows           (he) saw five crows
Numeral Expressions (cont.)

- part expressing the number + part expressing counted objects
- determine the governing and the dependent part … based on the **morphological analysis** (→ agreement)

\[
\text{viděl čtyři}_{\text{acc}} \text{ vrány}_{\text{acc}} \sim \text{viděl vrány}_{\text{acc}} \quad \text{BUT} \quad \text{viděl pět}_{\text{acc}} \text{ vran}_{\text{gen}} (*\text{viděl vran}_{\text{gen}} )
\]

**numerals as** sto, tisíc, milion, stovka, setina … rendered as nouns
([hundred, thousand, million, a hundred, a hundredth])

a-thousand of-people ate
one-tenth of-ton of-pork meat

(he) won with hundred (of) percent of votes
Numeral Expressions (cont.)

- structure of the numeral part (technical)
  head: the lowest rank (wrt the agreement)
Numeral Expressions (cont.)

- structure of the numeral part (technical)
  head: the \textbf{lowest} rank (wrt the \textit{agreement})
Numeral Expressions (cont.)

- numerals, type více [more], méně [less] (not a comparison)

(you) will-see much less people

(l) will-drink more than five beers

http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/cz/a-layer/html/ch03s07s09s05.html
Numeral Expressions (cont.)

- English HamleDT

... in favor of two Hees nominees... to pluck the five genes that produced the toxin out of the pertussis bacterium.
Specific Constructions: Comparison

Conjunctions *jako* [as, such], *než* [than] … adverbial

1. Full embedded clause

```
he looks different from how he looked yesterday
```

```
(he) saw a butterfly that was like shining
```
Specific Constructions: Comparison

conjunctions *jako* [as, such], *než* [than] … adverbial

1. full embedded clause
2. **ellipsis**

*We bought the same house as Karel (= as Karel bought)*

*(he) tested Budvar, Plzeň and Kozel, as well as Krušovice*
Specific Constructions: Direct Speech

direct, semidirect and indirect speech
1. direct speech as an object
2. direct speech as coordinated sentence
   (i.e., elided communication verb)

“I believe it!” Cilka replied.

“Karel,” (he) scratched his head, “did not come “
(= Karel, (he) scratched his head and said, did not come)
Specific Constructions: Direct Speech

direct, semidirect and indirect speech
1. direct speech as an object
2. direct speech as coordinated sentence (i.e., elided communication verb)

BUT: HamleDT
Graphic Symbols, Punctuation

**AuxX** ... comma
- not for connecting node of coordination and aposition; not for ellipses
  1. comma separating a clause
    - a child of the **head of the subord. clause**

(he) saw that (he) slept

(the) house which is crying
Graphic Symbols, Punctuation

**AuxX** … comma

- not for connecting node of coordination and aposition; not for ellipses
  1. comma separating a clause
     - a child of the head of the subord. clause
  2. comma separating a parenthesis
     - a child of the **head of the parenthesis**

---

the water, so to speak, grew wise
AuxX … comma

- not for connecting node of coordination and aposition; not for ellipses
  1. comma separating a clause
     - a child of the head of the subord. clause
  2. comma separating a parenthesis
     - a child of the head of the parenthesis

!!! comma separating **coordinated sentences or sentence members**

- a child of the Coord node

(he) believes (in) (black) magic, mystique and superstitions

a barman in Klánovice, now a foreman
Graphic Symbols, Punctuation

- **AuxS** … root
- **AuxK** … terminal punctuation mark (all types);
  - a child of the root
  - (in case of a multiple function AuxK the lowest priority)

```plaintext
mělo
Pred
Včera
Adv
pršet
Obj

(Přišel)
Pred
za
AuxP
30
Adv
min
Atr

(He) arrived in 30 min
```

*It should have rained yesterday.*
he paid her (on a corridor) a compliment

a) plums, b) apricots (have) ripened

*AuxG* … punctuation, other graphical symbols

Liberal Party (LSNS)
References


• Manual for Analytical Annotation