Prague Dependency Treebank: Surface Syntactic Annotation

Markéta Lopatková

Institute of Formal and Applied Linguistics, MFF UK

lopatkova@ufal.mff.cuni.cz
Prague Dependency Treebank

- application of a language theory onto a large set of data
- primarily focused on Czech running texts
- attempt to standardize dependency annotation

- language data
- tools
  - TrEd ... graphical editor and interface for creating queries (practical lectures) [http://ufal.mff.cuni.cz/tred/](http://ufal.mff.cuni.cz/tred/)
  - manuals for individual layers
  - survey of data formats and tools
Prague Dependency Treebank

- **HamleDT** ~ a compilation of existing dependency treebanks (or dependency conversions of other treebanks), transformed so that they all conform to the same annotation style … 2012
  36 languages, 42 treebanks in HamleDT 3.0 (2015)

Dependency Grammars and Treebanks: PDT a-layer
HamleDT: a-layer

https://lindat.mff.cuni.cz/services/pmltg/#!/treebanks
(login required for English, free whenever possible)
Prague Dependency Treebank

4 layers:
- word layer (w-layer)
- morphological layer (m-layer)
- **analytical layer (a-layer)**
- tectogrammatical layer (t-layer)

<table>
<thead>
<tr>
<th>layers of description</th>
<th>t,a,m-layer</th>
<th>a,m-layer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>train</td>
<td>dtest</td>
</tr>
<tr>
<td># documents</td>
<td>2 536</td>
<td>316</td>
</tr>
<tr>
<td># sentences</td>
<td>38 737</td>
<td>5 228</td>
</tr>
<tr>
<td># tokens</td>
<td>652 700</td>
<td>87 988</td>
</tr>
</tbody>
</table>
PDT: a-layer

Goal:
- to describe the structure of the sentence and
- to denote the type of relations between tokens
PDT: a-layer

Goal:
- to describe the structure of the sentence and
- to denote the type of relations between tokens

- dependency tree
- one token from m-layer ~ one node incl. prepositions, punctuation … plus technical root
- relations ~ edges
  - dependency, coordination, punctuation, …
  - type of relation: attribute of the child node
  - oriented "upwards", i.e., towards its parent / "governing" node
- linear ordering ~ surface word order

!!! tree structure ~ XML structure !!!
Some contours of the problem seem to be clearer after the resurgence by Havel's speech.
# A-layer: attributes of a node (non-root)

<table>
<thead>
<tr>
<th>Attribute Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>unique identifier of the node in PDT 2.0</td>
</tr>
<tr>
<td>m.rf</td>
<td>PML reference; points to a node on the m-layer (i.e., node m)</td>
</tr>
<tr>
<td>afun</td>
<td>analytical function ~ kind of relation between the node and its parent node</td>
</tr>
<tr>
<td>is_member</td>
<td>0 or 1; denoting members of a coordination or apposition; only children of a node with afun Coord or Apos (disregarding AuxP and AuxC)</td>
</tr>
<tr>
<td>is_parenthesis_root</td>
<td>0 or 1; 1 identifies roots of subtrees corresponding to parentheses</td>
</tr>
<tr>
<td>ord</td>
<td>positive integer; the (left to right) order of representing the nodes in graphical applications</td>
</tr>
<tr>
<td>clause_member</td>
<td>identifies a clause to which the node belongs (for Czech)</td>
</tr>
</tbody>
</table>
## A-layer: attributes of a root

<table>
<thead>
<tr>
<th>Attribute Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>unique identifier of the tree in PDT 2.0</td>
</tr>
<tr>
<td>s.rf</td>
<td>PML reference (points to a sentence marked s on the m-layer)</td>
</tr>
<tr>
<td>afun</td>
<td>AuxS (analytical function)</td>
</tr>
<tr>
<td>ord</td>
<td>0 (position in the horizontal ordering of the nodes in a tree)</td>
</tr>
</tbody>
</table>
List of analytical functions (afun)

- sentence predicate … Pred, Pnom, AuxV
- subject … Sb
- object … Obj
- adverbial … Adv
- attribute … Atr
  - combined functions … AtrAtr, AtrObj, ObjAtr, AtrAdv, AdvAtr
- complement … Atv, AtvV
- auxiliary sentence member
  - … AuxS (technical root) AuxO, AuxY (emotion, particles)
  - … AuxC, AuxP (conj, prep) AuxZ (rhematizers)
  - … AuxT, AuxR (reflexives) AuxX, AuxG, AuxK (graphical symbols)
- non-dependency structures … Coord, Apos
- ellipsis … ExD
Predicate of a simple sentence / main clause

**Predicate** … analytical functions:  *Pred, Pnom, AuxV*

- attributing some property/state/change/activity to the subject
- basic sentence member
- governing node (= head) of its clause
Predicate of a simple sentence / main clause

- A child of the technical root
- Typically a finite verb form … Pred

The graph shows how this happens …

Disclaimer ... English trees from HamleDT (automatically transferred)!
Predicate of a simple sentence / main clause

- a child of the technical root
- typically a finite verb form … Pred

The graph shows how this happens …

The graph shows how this happens …

Dependency Grammars and Treebanks: PDT a-layer
Predicate of a simple sentence / main clause

- a child of the technical root
- typically a finite verb form … Pred
  - analytical verbal forms: auxiliary verbs … AuxV

Terms were not disclosed.
The luxury auto maker last year sold 1,214 cars in the U.S.
(2) Compound verbal predicate

- finite modal / phase verb **Pred** + infinitive of lexical verb **Obj**
  - mít [should], muset [must], moci [can], chtít [to want], dát_se [to be possible], smět [may], ...
  - začít [to begin], skončit [to end], ...

Yesterday (it) **should** rain.

But they **must** also **assign** a price ....
One big problem has been the thoroughbred racehorse market.
(3) Verbonominal predicates

- copular verb … **Pred**
  - a verb used to link the subject of a sentence with a predicate
  - verb expressing identity, membership in a class, property …
- nominal predicate … **Pnom**

---

**Přesto je ceněna jako dobrá orientační pomůcka.**

Beer *is healthy.*
Predicate of a *subordinate clause*

- afun of the *respective function*

The graph shows how this *happens* ...
He argues that [...] dollar is now moving “sideways” ...
Subject

Subject ... analytical function: Sb
• any construction answering the question who (what)
• dependent on a clause predicate

He argues that...
Subject

**Subject** … analytical function: **Sb**

- any construction answering the question *who (what)*
- dependent on a clause predicate

*kdo se bojí, nesmí do lesa*

who –refl – is-frightened – must-not – to – wood

*whoever is frightened, must not enter a wood*

*What triggered the latest clash was a skirmish …*
Subject

*Subject* ... analytical function: **Sb**
- any construction answering the question *who (what)*
- dependent on a clause predicate

Whether *psyllium* fortune depends on .......
Object

Object … analytical functions: Obj
• object modifies a verb / adjective / adverb
• Czech … a form of an object is typically prescribed by its governing word (esp. case; cz term: rekce)

We see an even stronger argument against UNESCO …

Mirek hated (a) sentence analysis
Object

Object ... analytical functions: Obj

- object modifies a verb / adjective / adverb
- Czech ... a form of an object is typically prescribed by its governing word (esp. case; cz term: rekce)

(i) direct (accusative), indirect (dative), second object
  - result/effect of an action,
    e.g. to write a letter, to convert a document from one format to another, she was appointed a special assistant
  - what is affected, e.g. to touch a table
  - what the action is directly aiming at, e.g. to advise a boy
  - an origin,
    e.g. to convert a document from one format to another
  - (infinitive following a modal or phase verb)

Mirek hated (a) sentence analysis

Dependency Grammars and Treebanks: PDT a-layer
Object

Object … analytical functions: Obj
• object modifies a verb / adjective / adverb
• typically, a form of an object is prescribed by its governing word (esp. case; cz term: rekce)
  (i) direct (accusative), indirect (dative), second object
  (ii) incl. infinitive constructions

The government is funding several projects to push PC use.
Object

Object ... analytical functions: Obj
• object modifies a verb / adjective / adverb
• typically, a form of an object is prescribed by its governing word (esp. case; cz term: rekce)
(i) direct (accusative), indirect (dative), second object
(ii) incl. infinitive constructions
(iii) compound predicates ... Pred + Obj
    mít [should], muset [must], moci [can], chtít [to want], dát_se [to be possible], smět [may];
    začít [to begin], skončit [to end], ... 
(iv) direct speech

“I believe it!” Cilka replied.

Dependency Grammars and Treebanks: PDT a-layer
Adverbials

Adverbials … analytical functions: Adv

• express the circumstances and relations, such as location, time, manner, comparison, extent, means, cause, consequence, regard or aim
• can modify a verb, adjective or adverb

[... he says] Judge O'Kicki unexpectedly awarded him an additional $100,000
Adverbials

... analytical functions: Adv

- express the circumstances and relations, such as
  location, time, manner, comparison, extent, means, cause,
  consequence, regard or aim
- can modify a verb, adjective or adverb
- a form of Adv is not prescribed by its governing word!
- questions: where?, where to?, from where?, …
  how long?, when?, …
  for what purpose?, why?, …
  how?, …
- in PDT not further classified

When Mr. Green won a $240,000 verdict …
Attribute

**Attribute** ... analytical functions: **Atr**
- modifies a noun (with any function)
- answers the question ... *which*, *what* or *whose* (*jaký, který, čí*)

*State* investigators dispute *that* view now, ...
Attribute

... analytical functions: \textbf{Atr}

- modifies a noun (with any function)
- answers the question ... \textit{which, what} or \textit{whose} (jaký, který, čí)

... representatives of the world 's No. 1 jet maker ...
Attribute: Agreement

- agreeing attribute … the same case as its governing noun
  Adj … case, number, gender
  BUT: *kluku ušatá*

 dependency grammars and treebanks: PDT a-layer
Attribute: Agreement

- agreeing attribute … the same case as its governing noun
- non-agreeing attribute

- cottage of our neighbor

- an inscription "Not for sale" was hanging on the door

- knowledge that death does-not-wait
Atr ... ALSO "technical" solution for

- addresses
- names of persons and institutions
- foreign words
- expressions with numerals
- figures
Complement (verbal attribute, secondary predication in UD) … analytical functions: Atv, AtvV

- modifies two sentence members, verb and noun

\[\text{chlapec ležel nemocen, viděl ho nemocného}\]

\[[\text{boy – lay – ill}] \quad [\text{he – saw – him – (being) ill}]\]

\[\begin{align*}
\text{we - AuxV - came - glad} \\
\text{we were glad to come}
\end{align*}\]

\[\begin{align*}
\text{money - he-has - deposit} \\
\text{he has put the money on a deposit}
\end{align*}\]

\[\begin{align*}
\text{přiběhl} \\
\text{he-arrived - barefooted}
\end{align*}\]
Combined Functions

analytical functions: \textbf{AtrAtr, AtrAdv, AdvAtr, AtrObj, ObjAtr}

(slice of bread with butter)

(she) brought (a) case from (the) basement

we are not going to pay the extraordinary installment of the part of the debt to our insurance office

Dependency Grammars and Treebanks: PDT a-layer
Auxiliary Sentence Members

- function words
- punctuation and graphical symbols
- words with unclear syntactic function
Auxiliary Sentence Members

**AuxP** … prepositions

- **parent of a nominal node**
- similarly one-word improper prepositions

The judge came under scrutiny in late 1987, …

Dependency Grammars and Treebanks: PDT a-layer
Auxiliary Sentence Members

**AuxP** … prepositions

- **parent of a nominal node**
- similarly one-word improper prepositions

(he) thought of (his) mother

(the) transition from (a) defense to (an) attack

(they) negotiated - about - (the) heritage
Auxiliary Sentence Members

**AuxP** ... secondary prepositions
- consists of several words
- technical solution:
  - the last node ... head
  - the remaining words ... siblings of the noun governed

(they) closed - for - reasons - (of) leave
(they) closed because of the leave

Because of Leche Fresca's success ...
Auxiliary Sentence Members

**AuxC** … subordinate conjunctions
- subordinate clause introduced by a subordinate conjunction
- conjunction … *root of the subordinate clause*

He argues *that* the dollar is now ``moving sideways …

... she felt *as if* she would go off in a faint
Auxiliary Sentence Members

**AuxC** … subordinate conjunctions
- subordinate clause introduced by a subordinate conjunction
- conjunction ... *root of the subordinate clause*
  - subordinate conjunction may attach an individual sentence member

… (they) talked about (an) attractive, *even though crucial topic*
He argues *that the dollar is now `moving sideways* …
Consequence for subordinated clauses

… their bill would become the last word on how funds are distributed.

Dependency Grammars and Treebanks: PDT a-layer
Auxiliary Sentence Members

**AuxT** ... morphemes *se, si* as a part of a reflexive tantum

(they) - are- afraid - *Refl* - him
they are afraid of him
Auxiliary Sentence Members

AuxR ... morphemes se as a part of a reflexive verbal form

it-danced – Refl – till – to morning
the dancing went on till morning

Dependency Grammars and Treebanks: PDT a-layer
Auxiliary Sentence Members

- punctuation … **AuxX**
- other graphical symbols … **AyxG**
- clause final punctuation … **AuxK**
Auxiliary Sentence Members

**AuxZ** … emphasizing words (modifying a sentence member)

e.g., *asi [maybe, approx.], dokonce [even, as far as], hlavně [mainly], ještě [moreover], již [yet], leda [only], především [most of all, most probably], zvláště [especially], …

[Diagram showing syntactic relationships and example words]
Auxiliary Sentence Members

**AuxO** … emotional particles

*mi, vám, si, … to, ono, …*

(I) could not - Aux - you - fall asleep

I couldn't fall asleep, you see.

it - he - not-come - papa

obviously, papa failed to come

Dependency Grammars and Treebanks: PDT a-layer
They claimed that the Fed would first give a green light to the economy by making credit readily available and then turn on the red and bring growth to a screeching halt. *But under Alan Greenspan that has changed.*
Auxiliary Sentence Members

**AuxY** ... particles modifying the whole sentence parts of compound conjunctions comparison (and others)
Principles of Annotation: Non-dependency

- coordination
- apposition
- parenthesis
Coordination

• "multiplication" of a single syntactic position (sentence member / sentence)
• Coord … 'connecting node'
  • primarily: - coordinating conjunctions (and, or, thus, …)
    - connecting expressions (and so on, et cetera, …)
    - comma
• visualization … suffix_Co
  … is_member in the data
" Payments are good right now, but we are n'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

(He) came alone and immediately
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

Dependency Grammars and Treebanks: PDT a-layer
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
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 ... as ExD_Co (ellipsis, coordination)

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

Dependency Grammars and Treebanks
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

(\(\text{připojení} \, ??? \text{obce} \, Atr \rightarrow \text{AuxP} \rightarrow \text{k} \, \text{druhé} \, ExD \rightarrow \text{jedně} \, Atr\))

\[\begin{array}{c}
\text{AuxS} \\
\text{Lékaře} \\
\text{ExD} \\
\text{!} \\
\text{AuxK} \\
\end{array}\]

(\(\text{(the) attachment of one village to another} (= \text{another village})\))

Dependency Grammars and Treebanks: PDT a-layer
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

(l) have to (go) home. (Lt) 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.

Dependency Grammars and Treebanks: PDT a-layer
References


• Manual for Analytical Annotation