Prague Dependency Treebank: Annotation of Surface Syntax

Markéta Lopatková

Institute of Formal and Applied Linguistics, MFF UK

lopatkova@ufal.mff.cuni.cz
PDT: a-layer

Goal:
• to describe the structure of the sentence and
• to denote the type of relations between "words"

• dependency tree  !!! tree structure ~ XML structure !!!
• 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

Některé kontury problému se však po oživením Havlovým projevem zdají být jasnější.
[Some contours of the problem seem to be clearer after the resurgence by Havel's speech.]
HamleDT: a-layer

https://lindat.mff.cuni.cz/services/pmltg/#!/treebank/hamledt_en/help
(login required for English, free whenever possible)
## 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><strong>analytical function ~ kind of relation between the node and its parent node</strong></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>
Analytical functions (afun)
A simple sentence

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

- basic sentence member
- some property/state/change/activity is attributed to the subject
- governing node (= head) of its clause
Predicate: Dependency Structure

- predicate of a **main clause**
  - a child of the root, analytical function … **Pred**

*Chimneysweep sweeps chimneys.*

The graph shows …
Predicate: Dependency Structure

- predicate of a **main clause**
  - a child of the root, analytical function … **Pred**

Terms were not disclosed.

passive sentence

Terms were not disclosed.
Predicate: Dependency Structure

- predicate of a **main clause**
  - a child of the root, analytical function \( \text{Pred} \)

**BUT:**

- predicate of a **subordinate clause**
  - *afun of the respective function*

---

*nevěděl*

???

- usne
  - Obj
- kdy
  - Adv
- \( \text{AuxX} \)

**Pronominal adverb / relative pronoun**

*did-not-know when* **falls-asleep**
Predicate: Dependency Structure

• predicate of a **main clause**
  • a child of the root, analytical function … **Pred**

**BUT:**

• predicate of a **subordinate clause**
  • a fun of the respective function

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

did-not-know that
he-had-finished-sweeping
He argues that the dollar is now ``moving sideways ... 

Karel by byl sedával na své židli. Karel would be$_{Past}$ used-to-sit on his chair. Charles would have used to sit in his chair.
Predicate: Analytical Verb Forms

- auxiliary verb(s) + lexical verb
  - lexical verb … head … Pred
  - auxiliary verb(s) … child(ren), afun AuxV

- Note: passive participle (= action) vs. state

Hrad byl vystavěn.
(The) castle was erected.
Predicate: Three Different Types

1. simple verbal predicate (incl. analytical verbal forms) typically a finite verb form … *(AuxV +)* Pred

lexical or být [to be] (existential, substitute), incl. lze, nelze [is (not) possible]

Mother cries.

Let there be light.

*(He) was in (a) hospital.*

Hurrah for home!
Predicate: Three Different Types

1. simple verbal predicate (incl. compound/analytical verbal forms)
2. compound verbal predicate … \( \text{Pred} + \text{Obj} \)
   finite modal / phase verb + infinitive of lexical verb
   \( \text{mít} \) [should], \( \text{muset} \) [must], \( \text{moci} \) [can], \( \text{chtít} \) [to want], \( \text{dát_se} \) [to be possible], \( \text{smět} \) [may]; \( \text{začít} \) [to begin], \( \text{skončit} \) [to end], …

Yesterday (it) should rain.

Yesterday (she) began (to) study.

PDT: a-layer

Lopatková
Predicate: Three Different Types

1. simple verbal predicate (incl. compound/analytical verbal forms)
2. compound verbal predicate
3. verbal nominal predicate (přísudek slovesně jmenný) … Pred + Pnom
   finite form of copula (spona) být + nominal part

Beer is healthy.

…it is (the) tragedy (of) this people

it-is – to-see – Sněžka
Sněžka can be seen.
Subject

Subject ... analytical function: Sb
- any construction answering the question who (what)
- dependent on a predicate
- possible forms of subject (see the data)
Subject

**Subject** ... analytical function: \( Sb \)
- any construction answering the question *who (what)*
- dependent on a predicate
- possible forms of subject (see the data)

*Tonda nese pivo.*
Tony – brings – beer

He argues that...

The proposed acquisition provides for ...

PDT: a-layer  Lopatková
Subject

Subject … analytical function: **Sb**
- any construction answering the question *who (what)*
- dependent on a predicate
- possible forms of subject (see the data)

**jíst je obřad**
to-eat – is – ceremony

**žádné napřesrok už nebude**
no – next-year – anymore – will-not-be
Subject (cont.)

**Subject** ... analytical function: **Sb**
- any construction answering the question *who (what)*
- dependent on a predicate
- possible forms of subject (see the data)

štve mě, jak to jde pomalu
irritates – me – how – it – proceeds – slowly
*it irritates me how slowly it proceeds*
Subject (cont.)

Subject ... analytical function: Sb
- any construction answering the question who (what)
- dependent on a predicate
- possible forms of subject (see the data)

What triggered the latest clash was a skirmish ...

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 ...
Attribute

... analytical functions:  **Attr**

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

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

písmena
???
červená
Atr
red letters

maminka
???
nebožka
Atr
the dead mother
Attribute: Agreement

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

- písmena ???
  červená Atr
  red letters
- maminka ???
  nebožka Atr
  the dead mother
- souseda Atr
  cottage of our neighbor
- chalupa ???
- našeho Atr
- visel ???
  dveřích Adv
  an inscription "Not for sale" was hanging on the door
  na Aup
  nápis Sb
  the door
- nепроданнě Atr
  Neprodejné

PDT: a-layer
Attribute: Agreement

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

- passing from defense to attack
- kitchen like (a) palm
- knowledge that death does-not-wait

PDT: a-layer
Attribute (cont.)

Atr … ALSO "technical" solution for
- addresses
- names of persons and institutions
- foreign words
- expressions with numerals
- figures
Attribute: Non-projectivity

Czech:

```
je  ???
teď'  Adv
zájem Sb
o AuxP
malý Atr
Letnou Atr
```

there is a little interest in Letná

Question: Is there any non-projectivity in English HamleDT (Prague-style) ???
**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*)

*Mirek hated (a) sentence analysis*
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*)

1. 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*
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)

1. direct (accusative), indirect (dative), second object
   - incl. infinitive constructions
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)

2. compound verbal predicate ... Pred + Obj
finite modal / phase verb + infinitive of lexical verb
mit [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.

Yesterday (she) began (to) study.
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)

3. direct speech

“I believe it!” Cilka replied.
Object: Possible Forms

see the data and/or the manual 😊

- typically: noun in accusative/dative/genitive/instrumental case
- preposition case
- infinitive, dependent clause

… (they) failed-to-find any trace

… it-wanted – Refl – me – to-sleep

… I felt sleepy
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
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
- 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

very big  close to Christmas  … to run into st

(17x459) Adverbials
Adverbials (cont.)

- several temporal / local adverbials

(He) got-up yesterday early morning (of) local time

(she) lived in Prague at Vyšehrad

[it usually happens] seven-times every week
Combined Functions

analytical functions: **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
**Complement** (verbal attribute) ... analytical functions:

- modifies two sentence members, verb and noun
  
  *chlapec ležel nemocen, viděl ho nemocného*
  
  [boy – lay – ill]   [he – saw – him – (being) ill]

- **Atv, AtvV**

- **Atv** modifies the verb
  
  - *přišli***
  
  [we - Aux - came - glad]
  
  we were glad to come

- **AtvV** modifies the noun
  
  - *příběhl***
  
  [he-arrived - barefooted]

- *má***
  
  we - Aux - has

  [she-has – cooked]
  
  he has done cooking

- *jsme***
  
  [we -Aux - came - glad]
  
  we were glad to come

- *my***

  [my]

- *rádi***

  [he - Aux - has]

  he has put the money on a deposit

- *peníze***

  [money - he-has - deposit]

  he has put the money on a deposit

- *uloženy***

  [money - he-has - deposit]

  he has put the money on a deposit

- *uvařeno***

  [money - he-has - deposit]

  he has put the money on a deposit

- *má***

  [she-has – cooked]

  he has done cooking

- *uvařeno***

  [she-has – cooked]

  he has done cooking

- *přišli***

  [we - Aux - came - glad]

  we were glad to come

- *jsme***

  [we -Aux - came - glad]

  we were glad to come

- *my***

  [my]

- *rádi***

  [he - Aux - has]

  he has put the money on a deposit

- *peníze***

  [money - he-has - deposit]

  he has put the money on a deposit

- *uloženy***

  [money - he-has - deposit]

  he has put the money on a deposit

- *má***

  [she-has – cooked]

  he has done cooking

- *uvařeno***

  [she-has – cooked]

  he has done cooking

- *přišli***

  [we - Aux - came - glad]

  we were glad to come

- *jsme***

  [we -Aux - came - glad]

  we were glad to come

- *my***

  [my]
Auxiliary Sentence Members
**Auxiliary Sentence Members**

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

- *(the) transition from (a) defense to (an) attack*
- *(he) thought of (his) mother*
- *(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
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 ..."
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
Auxiliary Sentence Members

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

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*
(I) could not - Aux - you - fall asleep
I couldn't fall asleep, you see.

obviously, father failed to come
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], ...

---

only a desert can look as sad as that

* these cactuses grow only in (a) desert
* more than thousand-headed crowd
Auxiliary Sentence Members

**AuxY** ... particles modifying the whole sentence;

no one has come, they say

whatever may happen, we shall win
Auxiliary Sentence Members

**AuxY** … particles modifying the whole sentence parts of compound conjunctions comparison (and others)
References


• Manual for Analytical Annotation