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/
- documentation: http://ufal.mff.cuni.cz/pdt3.5/documentation
  - 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), 36 languages, 42 treebanks in HamleDT 3.0 (2015)
  

**PDT style**

**Universal Dependencies**
Prague Dependency Treebank - scenario

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

\{ layers of annotation \}
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 !!!

Dependency Grammars and Treebanks: PDT a-layer
Some contours of the problem seem to be clearer after the resurgence by Havel's speech.

Některé kontury problému se však po oživením Havlovým projevem zdají být jasnější.
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>
## 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>
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
- secondary predicate… 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 …

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 … \textit{Pred}
  - analytical verbal forms: auxiliary verbs … \textit{AuxV}

\textit{Terms were not disclosed.}

Dependency Grammars and Treebanks: PDT a-layer
The luxury auto maker last year sold 1,214 cars in the U.S.

Dependency Grammars and Treebanks: PDT a-layer
(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], …

But they **must** also **assign** a price ….

Yesterday (it) **should** rain.
(3) Copular constructions

- 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**

One big problem has **been** the thoroughbred racehorse **market**.
(3) Copular constructions

- 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.

Dependency Grammars and Treebanks: PDT a-layer
Predicate of a *subordinate clause*

- a fun of the respective function

*The graph shows how this happens ...*
Predicate of a *subordinate clause*

- *afun of the respective function*

He argues that [...] dollar is now *moving* “sideways” [...]
Subject … analytical function: **Sb**

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

*He argues that...*

Dependency Grammars and Treebanks: PDT a-layer
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** … analytical function: **Sb**
- any construction answering the question *who (what)*
- dependent on a clause predicate

---

Whether psyllium makes 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), prepositional case

... they commented on it ...
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*)
  1. direct (accusative), indirect (dative), second object
  2. incl. infinitive constructions
  3. 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], …
  4. direct speech

“*I believe it!*” Cilka replied.
Adverbials

**Adverbials** … analytical functions: **Adv**
- express the circumstances and relations, such as location, time, manner, comparison, extent, means, cause, consequence, regard or aim
- questions: *where?*, *where to?*, *from where?*, …
  *how long?*, *when?*, …
  *for what purpose?*, *why?*, …
  *how?*, …

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

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

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

Dependency Grammars and Treebanks: PDT a-layer
Attribute

**Attribute** ... analytical functions:  **Atr**

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

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

Atr … ALSO "technical" solution for
  • addresses
  • names of persons and institutions
  • foreign words
  • expressions with numerals
  • figures
Secondary predication (Cz: doplněk)

**Secondary predication** (verbal attribute, complement ... analytical functions: $Atv$, $AtvV$

- modifies two sentence members, verb and noun

  - She declared the cake *beautiful*.
  - She entered the room *sad*.

---

Dependency Grammars and Treebanks: PDT a-layer
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

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

(they) negotiated - about - (the) heritage

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

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

Dependency Grammars and Treebanks: PDT a-layer
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*
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], ...
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 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**

(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
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

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, …
• visualization: suffix _Pa vs. is_parenthesis_root in the data

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

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

(the) attachment of one village to another (= 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.  (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.*

Dependency Grammars and Treebanks: PDT a-layer
References


• Manual for Analytical Annotation

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

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"


analytical functions (lecture 5):
• predicate … Pred, Pnom, AuxV (, Obj)
• subject … Sb
• attribute … Atr
• object … Obj
• adverbial … Adv … AtrAtr, AtrObj, ObjAtr, AtrAdv, AdvAtr
• complement … Atv, AtvV
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.

(the) previous and present government

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
Non-dependency: Coordination

coordination vs. apposition

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

(Ilt) 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
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
Non-dependency: Parenthesis

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

*(the) 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)

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

... *reduction* principle

\[(he) \text{ saw four crows} \quad (he) \text{ saw five crows}\]
Numeral Expressions (cont.)

- part expressing the number + part expressing counted objects
- determine the governing and the dependent part … based on the

!!! morphological analysis (→ agreement)

viděl čtyří acc vrány acc ~ viděl vrány acc  BUT  viděl pět acc vrany gen (*viděl vran 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**)

PDT: a-layer II.

Lopatková
Numeral Expressions (cont.)

- structure of the numeral part (technical)
  head: the lowest rank (wrt the 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

PDT: a-layer II.
Lopatková
... to pluck the five genes that produced the toxin out of the pertussis bacterium.

... in favor of two Hees nominees

Numeral Expressions (cont.)

- English HamleDT
Specific Constructions: Comparison

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

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

(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

Most of the growth, he said, will come in …
**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
(he) believes (in) (black) magic, mystique and superstitions

Mirek, 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)

It should have rained yesterday.

(He) arrived in 30 min.
Graphic Symbols, Punctuation

- **AuxG** … punctuation, other graphical symbols

he paid her (on a corridor) a compliment

a) plums, b) apricots (have) ripened
References


- Manual for Analytical Annotation
Prague Dependency Treebank: Annotation of Surface Syntax

Markéta Lopatková

Institute of Formal and Applied Linguistics, MFF UK
lopatkova@ufal.mff.cuni.cz
Copula

• a word used to link the subject of a sentence with a predicate
• semanticaly empty, but syntactic center

  Mary and John are my friends.
  The sky was blue.
  I am taller than most people.
  The birds and the beasts were there.
Copula

• a word used to link the subject of a sentence with a predicate
• semantically empty, but syntactic center
  
  *Mary and John are my friends.*
  *The sky was blue.*
  *I am taller than most people.*
  *The birds and the beasts were there.*

• it expresses
  
  • identity ... *The Morning Star is the Evening Star.*
  • membership of a class or a subset relationship ... *She was a nurse.*
  • express some property, relation or position, permanent or temporary:
    *The trees are green.*
    *I am your boss.*
    *John was in England.*
    *The children are confused.*

• main/principal copula .... typically (equivalent of) "to be"
  
  *plus "verbs" equivalent in function to the principal copula, the verb "be"
  → no clear cut between copular verbs and full verbs*
<table>
<thead>
<tr>
<th>morphology</th>
<th>John is preparing his homework.</th>
<th>Byl by přišel včas.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>John has been attending the school since 2011.</td>
<td>Bude chodit včas.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>syntactic core of (surface) sentence</th>
<th>copula</th>
<th>light verbs</th>
<th>modal verbs</th>
<th>phase verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>afun (main clause)</td>
<td>Pred</td>
<td>Pred</td>
<td>Pred</td>
<td>Pred</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>semantic core</th>
<th>content verbs</th>
<th>Pred</th>
</tr>
</thead>
</table>

PDT: a-layer

Lopatková
Copula

Randolph Quirk: *A comprehensive grammar of the English Language*

“A verb is said to have a copular complementation when it is followed by a *subject complement* or a *predicational adjunct*,
and when this complement cannot be dropped without changing the meaning of the verb.
The verb in such a clause is a copular (or linking) verb, and is equivalent in function to the principal copula, the verb *be*.”

**current copulas** (typically stative):

(i) be (friendly) (my friend)
(ii) appear (happy) (the only solution)
feel (annoyed) (a fool)
look (pretty) (a fine day)
seem (very restless) (a genius)
smell (sweet)
sound (surprised) (a reasonable idea)
taste (bitter)

(iii) remain (uncertain) (good friends)
keep (silent)
stay (motionless)

**resulting copulas**:

(iv) become (older) (an expert)
come (true)
end up (happy) (her slave)
got (ready)
go (sour)
grow (tired)
prove (rather useful) (his equal)
turn (cold) (traitor)
turn out (fortunate) (a success)
winds up (drunk) (a millionaire)
Copula in Czech

Czech ling. tradition:
**Sponová slovesa** se využívají v přísudku slovesně-jmenném (přísudek jmenný se sponou), kde jsou doplněna jménem.
Mezi sponová slovesa patří *být, bývat, stát se, stávat se* a v záporu *nebýt, nebývat, nestat se a nestávat se*.

-Marta je nemocná.
-Simona bývala učitelkou.
-Petr se stal lékařem.
(jinde též mít, zůstat, zdát se)
Complex sentences

• **main clause** (= independent clause) vs. **non-main clause**

• **relative clause** … syntactic function: Atribute
  
  *The Ferrari [which I can't afford].*
  
  *I imagined the house where I want to live.*
  
  • starting with a relative pronoun or relative adverb

• **dependent clause** … either modify the main clause of a sentence or serve as a component of it (plus recursivity!!)

  … syntactic function: Object or Adverbial

  • starting with a subordinating conjunction (that, if, whether, in order that)
    
    *I know (that) he is here.*
  
  • starting with an interrogative pronoun or relative adverb
    
    *I know who said that.*
    
    *I know where he wanted to go.*
References


• Manual for Analytical Annotation

• PML-TQ and repository
  http://lindat.mff.cuni.cz/services/pmltq/#!/home
Copula in data (Czech and English)

LINDAT/CLARIN … provides technical background and assistance … to share, create and modernise their tools and data used for research in linguistics … also provides an open digital repository and archive open to all academics

http://lindat.mff.cuni.cz/services/pmltq/#/home
Copula in **HamleDT** for Czech (PDT vs. UDs, version 1)

PML-TQ query:

```pml
a-node $pred :=
  [id ~ "prague", depth() = "1",
   1x a-node [ afun = "Pnom",
   ]
  ]
>> for $pred.lemma give $1, count()
>> sort by $2 desc
```

http://lindat.mff.cuni.cz/services/pmltq/#!/home

PDT: a-layer

Lopatková
Copula in **HamleDT** for Czech (PDT vs. UDs, version 1)

http://lindat.mff.cuni.cz/services/pmltq/#!/home

<table>
<thead>
<tr>
<th>lemma</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>být</td>
<td>7974</td>
</tr>
<tr>
<td>bývat</td>
<td>41</td>
</tr>
<tr>
<td>:</td>
<td>37</td>
</tr>
<tr>
<td>a</td>
<td>10</td>
</tr>
<tr>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td>či</td>
<td>2</td>
</tr>
<tr>
<td>=</td>
<td>2</td>
</tr>
<tr>
<td>bývávat</td>
<td>2</td>
</tr>
<tr>
<td>respektive</td>
<td>1</td>
</tr>
<tr>
<td>,</td>
<td>1</td>
</tr>
<tr>
<td>totiž</td>
<td>1</td>
</tr>
</tbody>
</table>
Copula for English in UDs, version 1

PML-TQ query:
```sql
a-node $cop := [ deprel = "cop" ] >> for $cop.lemma give $1, count() >> sort by $2 desc
```

http://lindat.mff.cuni.cz/services/pmltq/#/home
Copula for English in UDs, version 1

http://lindat.mff.cuni.cz/services/pmltq/#!/home

<table>
<thead>
<tr>
<th>copula</th>
<th>count</th>
<th>example</th>
</tr>
</thead>
<tbody>
<tr>
<td>be</td>
<td>5598</td>
<td>This is one thought-provoking film.</td>
</tr>
<tr>
<td>'s</td>
<td>8</td>
<td>“Yo Mama’s so fat ....”</td>
</tr>
<tr>
<td>'</td>
<td>1</td>
<td>“so, enemy, what's your point?”</td>
</tr>
<tr>
<td>b</td>
<td>1</td>
<td>“o and the cheaper the better (we are trying to save money now, sooo..thatd b helpful) thanks”</td>
</tr>
</tbody>
</table>
Copula for Czech in UDs, version 1

<table>
<thead>
<tr>
<th>lemma</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>být</td>
<td>20 480</td>
</tr>
<tr>
<td>bývat</td>
<td>110</td>
</tr>
<tr>
<td>stát</td>
<td>3</td>
</tr>
<tr>
<td>bývávat</td>
<td>3</td>
</tr>
<tr>
<td>moci</td>
<td>1</td>
</tr>
<tr>
<td>mít</td>
<td>1</td>
</tr>
</tbody>
</table>
"Doplněk" ~ secondary predication
"Doplněk" ~ secondary predication

**Complement** (verbal attribute) ... analytical functions: Atv, AtvV

- modifies two sentence members, verb and noun

  *chlapec ležel nemocen, viděl ho nemocného*
  
  [boy – lay – ill]  
  [he – saw – him – (being) ill]
"Doplněk" in HamleDT for Czech (PDT vs. UDs, version 1)
"Doplněk" in HamleDT for Czech (PDT vs. UD, version 1)
"Doplněk" in HamleDT for Czech (PDT vs. UDs, version 1)