

# PDT 2.0 Requirements on a Query Language

---

Jiří Mírovský

Charles University in Prague

Institute of Formal and Applied Linguistics

Czech Republic

ACL 2008, Columbus, Ohio, June 16, 2008

# Abstract

---

- ➔ Introduction to the Prague Dependency Treebank 2.0
- ➔ PDT 2.0 requirements on the query language + examples of queries in Netgraph
- ➔ Summary of the query language features

# Abstract

---

- ➔ Introduction to the Prague Dependency Treebank 2.0
- ➔ PDT 2.0 requirements on the query language + examples of queries in Netgraph
- ➔ Summary of the query language features

# Introduction to the Prague Dependency Treebank 2.0

## *Layers in PDT 2.0*

---

Czech: Byl by šel do lesa.

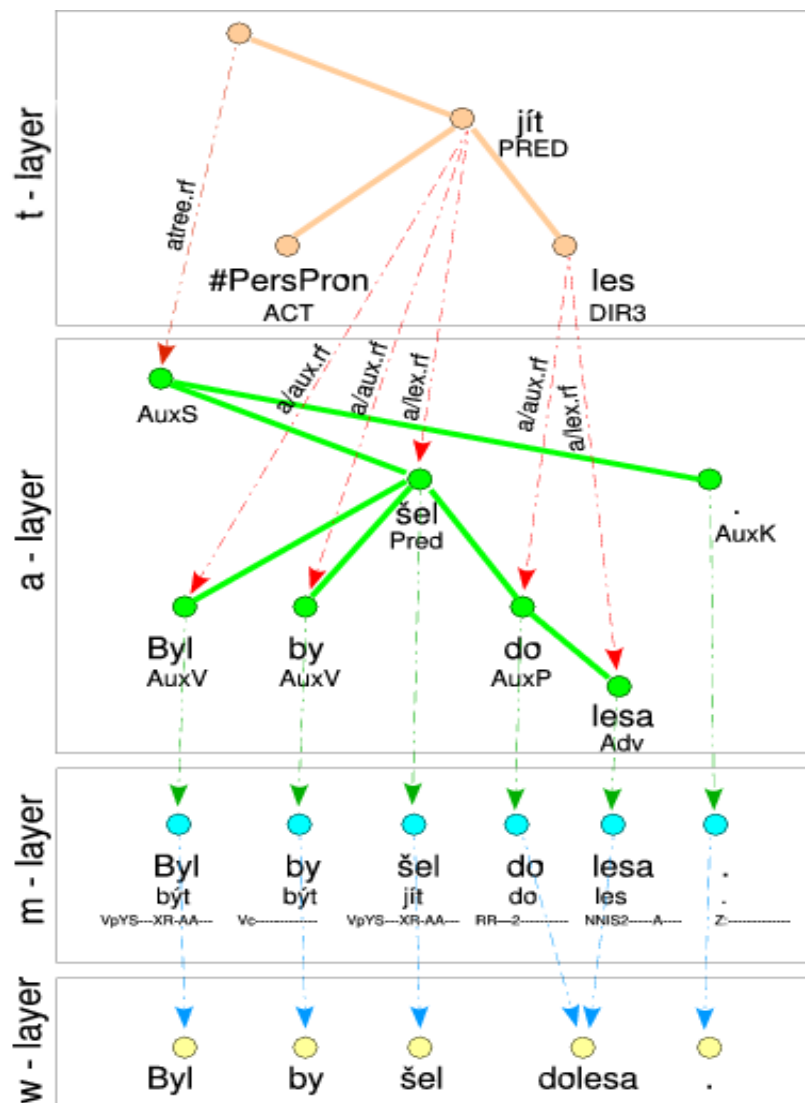
English: He would have gone to the forest.

Three layers of annotation:

- ➡ Morphological layer (NNIS2-----A-----)
- ➡ Analytical layer (Adv)
- ➡ Tectogrammatical layer (DIR3)

# Introduction to the Prague Dependency Treebank 2.0

## *Layers in PDT 2.0*



# Abstract

---

- ➔ Introduction to the Prague Dependency Treebank 2.0
- ➔ **PDT 2.0 requirements on the query language + examples of queries in Netgraph**
- ➔ Summary of the query language features

# Linguistic Requirements

## *Valency*

---

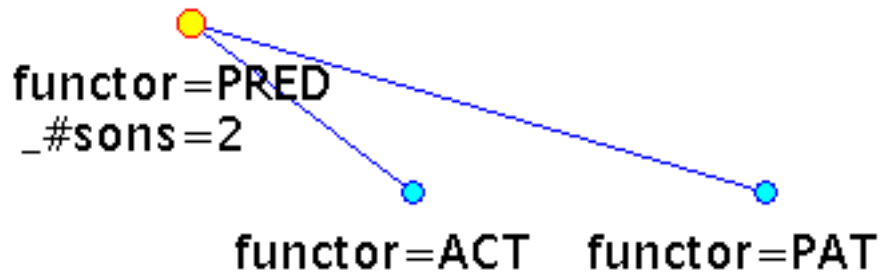
To study valency, the query language should be able to:

- ➔ control a presence of a particular type of son
- ➔ control a non-presence of a son
- ➔ control number of sons

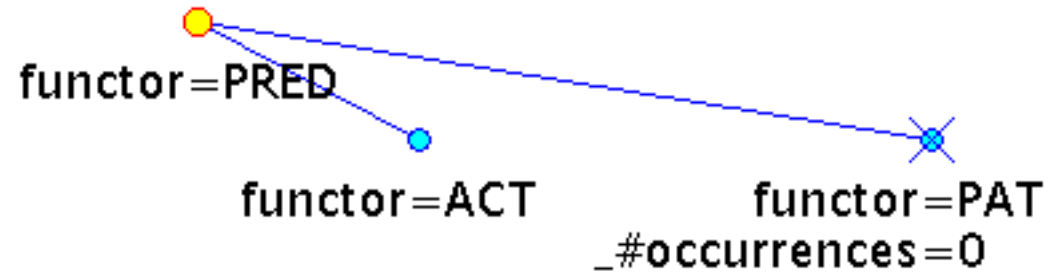
# Linguistic Requirements

## *Valency*

---



one **ACT**or, one **PAT**ient,  
no other sons



at least one son (**ACT**or),  
no **PAT**ient



# Linguistic Requirements

## *Coordination etc.*

---

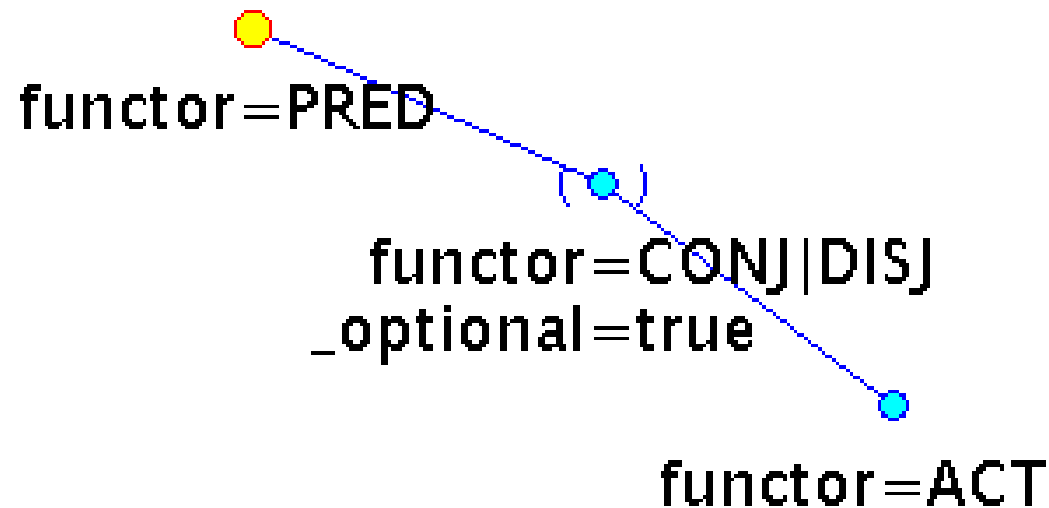
Tree dependency is not always linguistic dependency. We need:

 to skip a node (etc. coordination, apposition)

# Linguistic Requirements

## *Coordination etc.*

---



The **CONJ**unction/**DISJ**unction becomes optional.

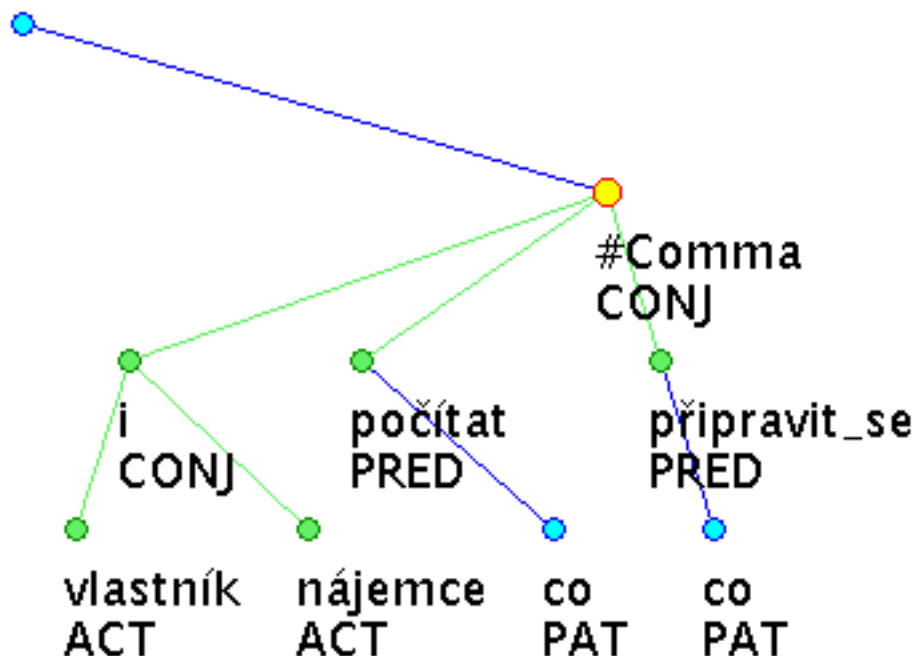
# Linguistic Requirements

## *Complex Example of Coordination*

---

Czech: S čím mohou **vlastníci i nájemci počítat**, na co by se měli **připravit**?

English (lit.): What can **owners and tenants expect**, what they should **get ready** for?



# Linguistic Requirements

## *Coordination etc.*

---

Tree dependency is not always linguistic dependency. We need:

➡ to skip a node (etc. coordination, apposition)

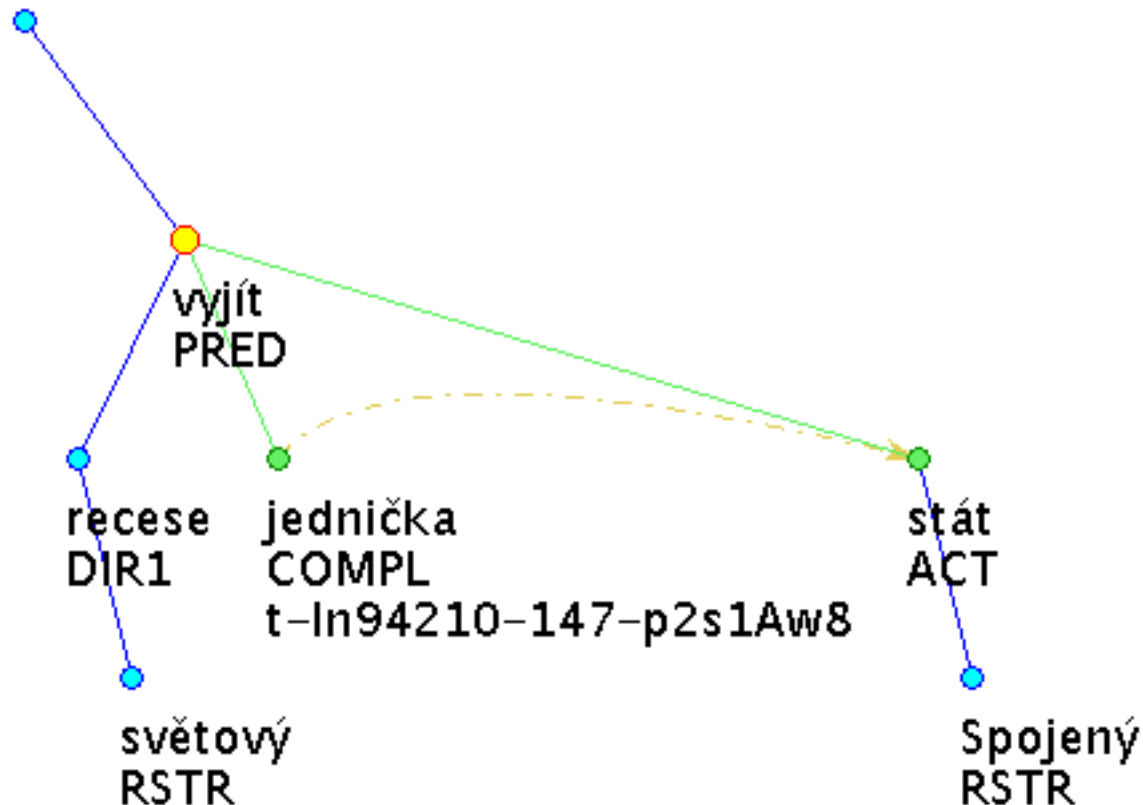
➡ even better: to set a linguistic dependency

# Linguistic Requirements

## *Predicative Complement*

---

Czech: Ze světové recese vyšly jako jednička Spojené státy.  
English: The United States emerged from the world recession as number one.

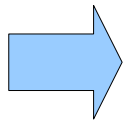


# Linguistic Requirements

## *Predicative Complement*

---

The dual dependency is represented by means of a **reference** to another node (attributes `compl.rf` and `id`). We need:



to match values unknown at the time of creating the query

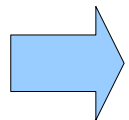
# Linguistic Requirements

## *Coreferences (Grammatical and Textual)*

---

Represented by means of **references**  
(attributes `coref_gram.rf` and  
`coref_text.rf` (and `id`)).

Again, we need:

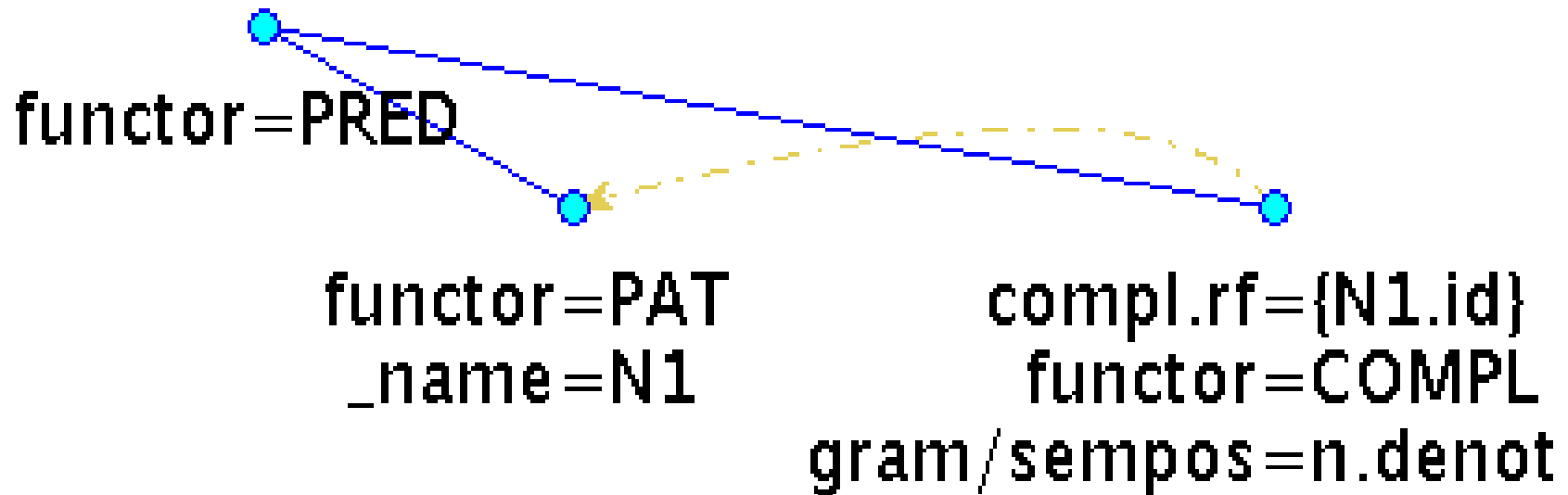


to match values unknown at the time  
of creating the query

# Linguistic Requirements

## *Predicative Complement*

---



a nominal predicative **COMPL**ement  
with second dependency on a **PAT**ient



# Linguistic Requirements

## *...other phenomena*

---

- ➔ *Focus proper* – combination of references, non-existence of a node and transitive closure of dependency
- ➔ *Rhematizers* – closest left son, closest left brother
- ➔ *(Non-)projectivity* – multiple-tree query to combine several one-tree queries representing different orientations of non-projective edges

# Linguistic Requirements

## *...other phenomena*

---

- ➔ *Idioms etc.* – searching in the linear form of the sentence (with regular expressions)
- ➔ *Agreement* – reference to only a part of a value of an attribute of another node (e.g. the fifth position of the morphological tag for case)
- ➔ *Word order* – measuring the horizontal distance between words

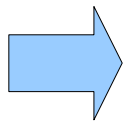
# Linguistic Requirements

## *Accessing Lower Layers*

---

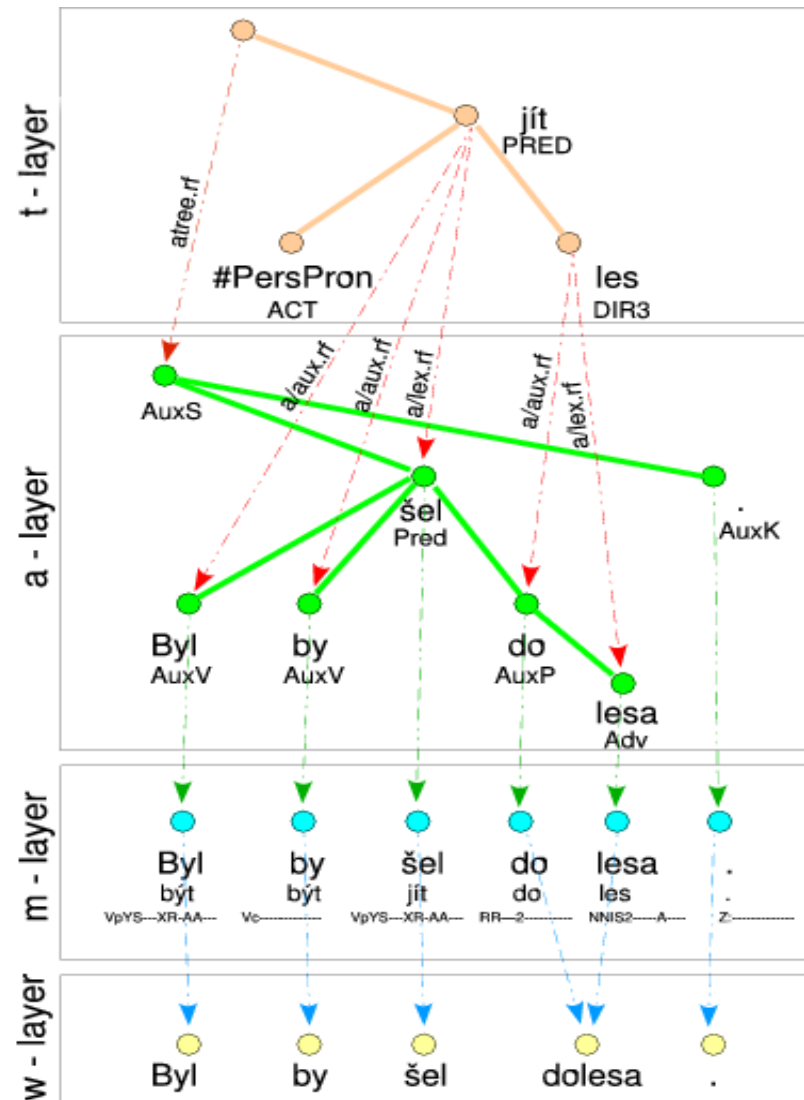
Queries across the layers of annotation:

- A **PAT**ient expressed with a preposition **k** and a noun in the dative
- A **PAT**ient less dynamic than an **ACT**or but on the left side from it in the sentence



We need to have means of accessing the lower layers.

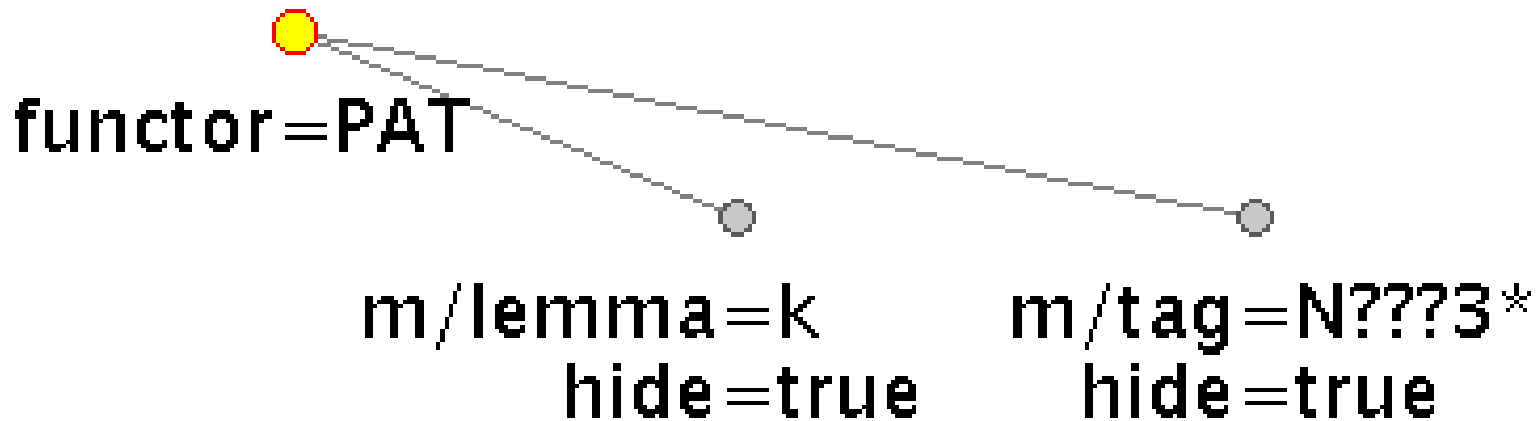
# Linguistic Requirements *Layers in PDT 2.0*



# Linguistic Requirements

## *Accessing Lower Layers – Hidden Nodes*

---



a **PAT**ient expressed with a preposition **k**  
and a noun (**N???3\***) in the dative (**N???3\***)

# Abstract

---

- ➔ Introduction to the Prague Dependency Treebank 2.0
- ➔ PDT 2.0 requirements on the query language + examples of queries in Netgraph
- ➔ **Summary of the query language features**

# Linguistic Requirements

## *Summary*

---

### *Evaluation of a node*

- multiple attributes evaluation
- alternative values
- alternative nodes (alternative evaluation of the whole set of attributes)
- wild cards (regular expressions)
- negation, relations other than “equal to”

# Linguistic Requirements

## *Summary*

---

### *Dependencies between nodes* *(vertical relations)*

- direct, transitive (existence, non-existence)
- vertical distance (from root, from one another)
- number of sons (zero for leaves)



# Linguistic Requirements

## *Summary*

---

### *Horizontal relations*

- precedence, immediate precedence
- negation of it
- horizontal distance

### *Secondary relations*

- secondary dependencies, coreferences

# Linguistic Requirements

## *Summary*

---

### *Other features*

- multiple-tree queries
- accessing several layers of annotation at the same time
- searching in the linear form of the sentence

# References

---

## *Prague Dependency Treebank*

- <http://ufal.mff.cuni.cz/pdt>

## *Netgraph home page*

- <http://quest.ms.mff.cuni.cz/netgraph>