

Learning to Search in Prague Dependency Treebank

Jiří Mírovský, Jarmila Panevová

Charles University in Prague

Institute of Formal and Applied Linguistics

Grammar and Corpora 2007

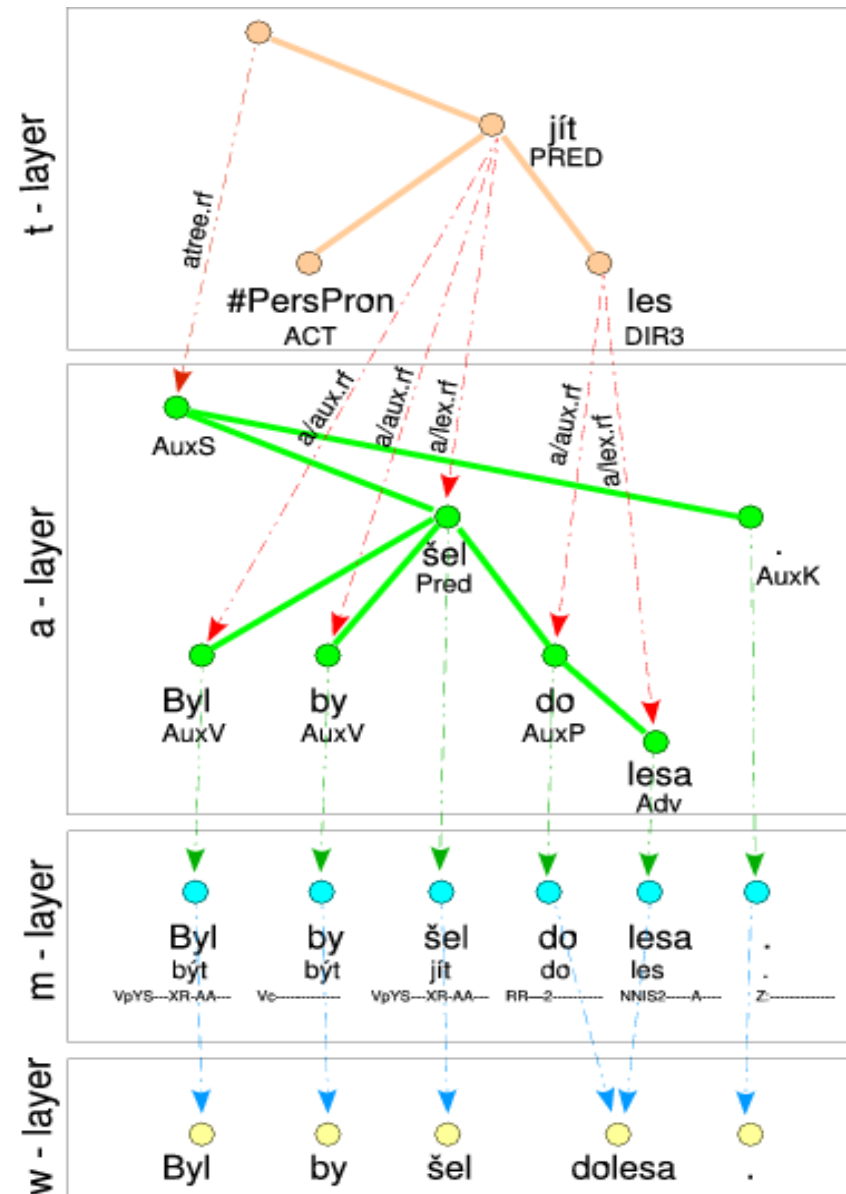
Prague Dependency Treebank 2.0

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**)

Prague Dependency Treebank 2.0



Netgraph – A Tool for Searching in Prague Dependency Treebank 2.0

- Client–server architecture
- Authentication of users
- Subcorpus definition
- Graphic creation of a query
- Searching in the treebank according to the query
- Viewing the result trees
- Basic statistics

A Query Creation

Netgraph 1.83 (30.4.2007)

Soubor Zobrazení Nastavení nápověda

globální hlavička:

atributy:

- atree.rf
- compl.rf
- coref_gram.rf
- coref_special
- coref_text.rf
- deepord
- functor**
- gram/aspect
- gram/degcmp
- gram/deontmod
- gram/dispmod
- gram/gender
- gram/indefunc

možné hodnoty:

- ACMP
- ACT**
- ADDR

reference:

atree.rf

hodnota:

ACT

strom dotazu:

functor=PRED

functor=EFF functor=ACT

továrna:

nový dotaz [] další strom)[] podstrom ([]) bratr ,[] alternativní vrchol [] smaž vrchol

pojmenuj vrchol: N1

zpět zobraz strom dotazu and/or: AND

dotaz:

[functor=PRED]([functor=EFF],[functor=ACT])

historie:

[functor=PRED]([functor=EFF],[functor=ACT])

zastav dotaz negace dotazu stromy dle dotazu nad výsledkem stromy dle dotazu všechny stromy

Soubory Dotazy Stromy Ladění

Načten další výskyt dotazu ve výsledných stromech.

[functor=PRED]([functor=EFF],[functor=ACT])

Viewing the Result

Netgraph 1.83 (30.4.2007)

Soubor Zobrazení Nastavení nápověda

Václav Klaus v rozhlasovém interview řekl, že negativní reakci očekával.

atribut	hodnota
is_mem...	
is_nam...	
is_pare...	
is_state	
nodetype	complex
quot/se...	
quot/type	
sentmod	enunc
subfunc...	
<input checked="" type="checkbox"/> t_lemma	řici
tfa	f
val_fra...	v#v-w5...
sentence	
a/ref_ty...	
a/afun	
a/is_m...	
m/form	
m/lemma	

zobrazované atributy
functor
t_lemma

soubor: /mnt/data, akce zobraz/skryj |<<- <<- <- < [5/64] [5/64/229] > -> ->>

Soubory Dotazy **Stromy** Ladění

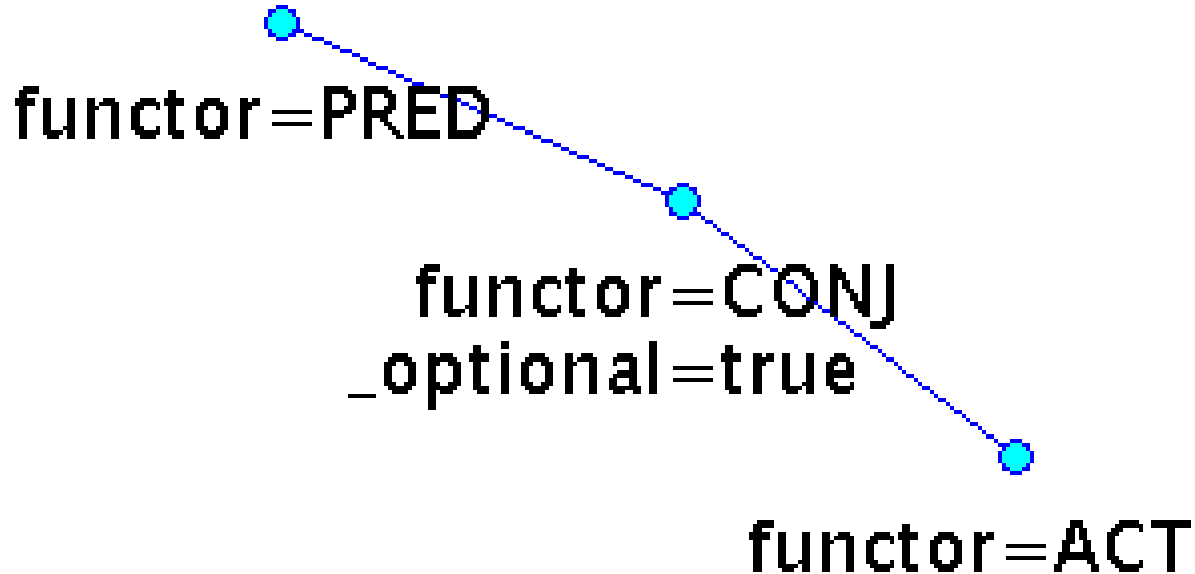
Načten další výskyt dotazu ve výsledných stromech.

- Different order of nodes; additional sons of the **PRED**icate

Meta-attributes

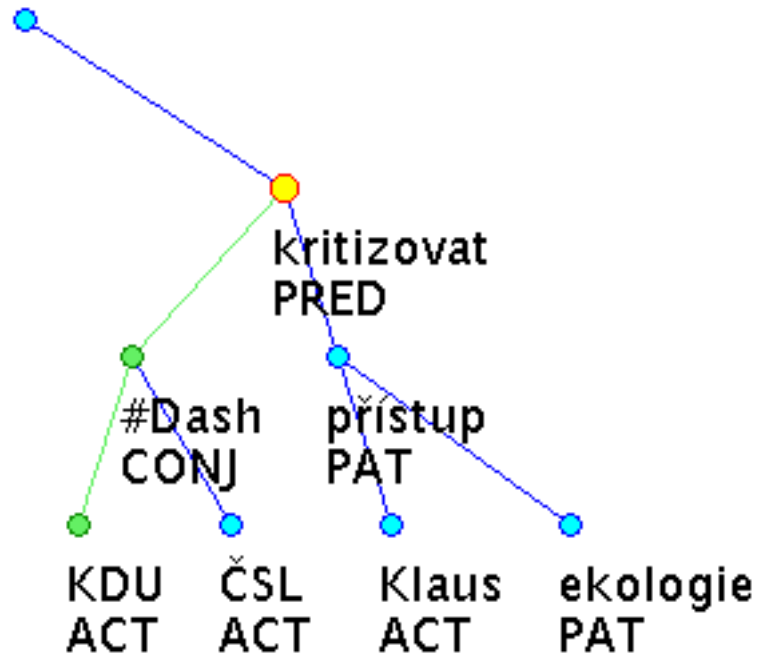
- Additional power to the query language
- Attributes not present in the corpus
- Treated like normal attributes
 - **_optional** (*optional node*)
 - **_#sons** (*exact number of sons*)
 - **_#occurrences** (*number of occurrences of a node*)
 - **_depth** (*distance from the root*)
 - **_sentence** (*linear form of the sentence*)
 - ...

Meta-attribute `_optional`



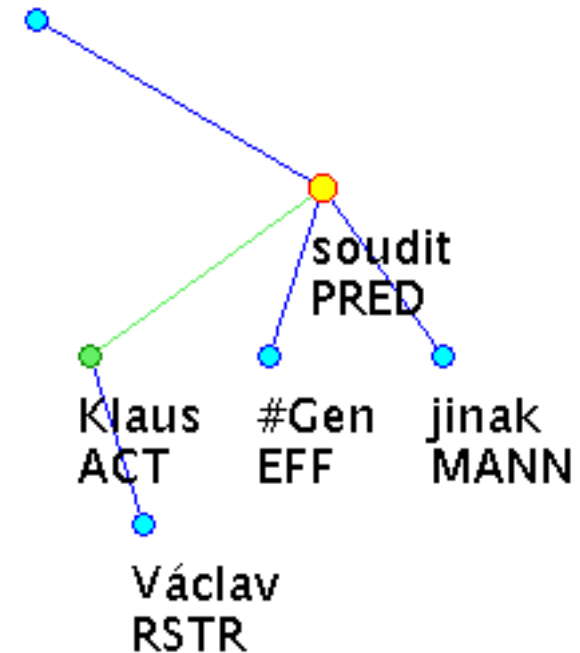
- A query with optional **CONJ**unction node
- Two possible types of result – with and without the optional node

_optional – Result Trees



Czech: **KDU-ČSL** kritizuje
Klausův přístup k ekologii.

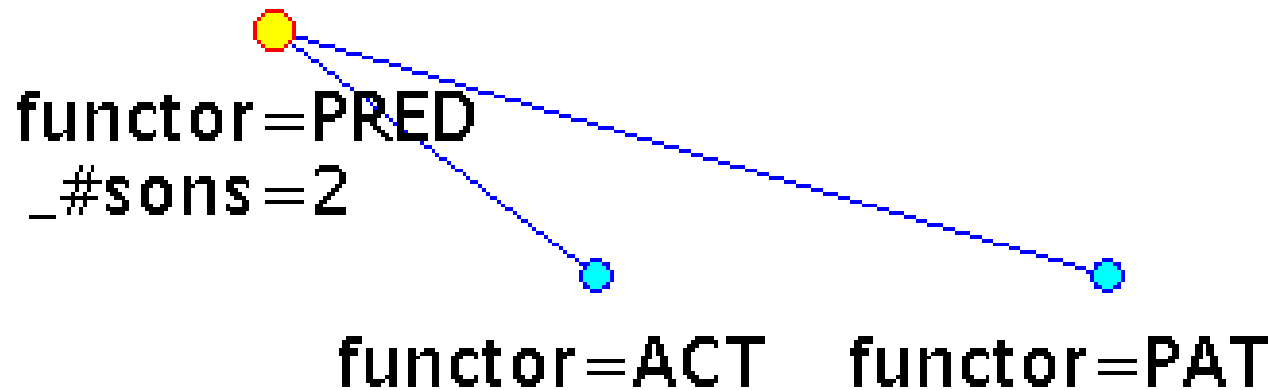
English (lit.): **KDU-ČSL** criticizes
Klaus's attitude towards ecology.



Czech: Václav **Klaus** soudí
jinak.

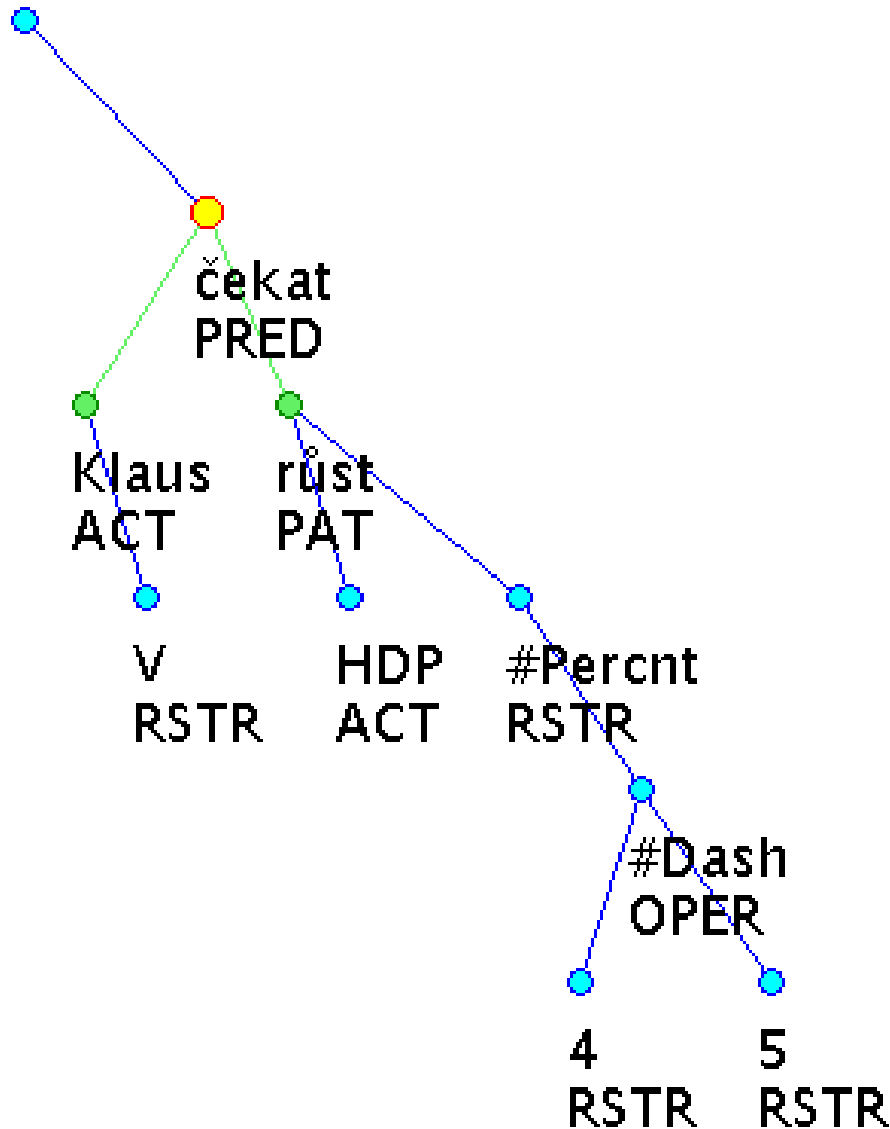
English (lit.): Václav **Klaus**
thinks otherwise.

Meta-attribute `_#sons`



- **PRED**icate with exactly two sons: **ACT**or and **PAT**ient

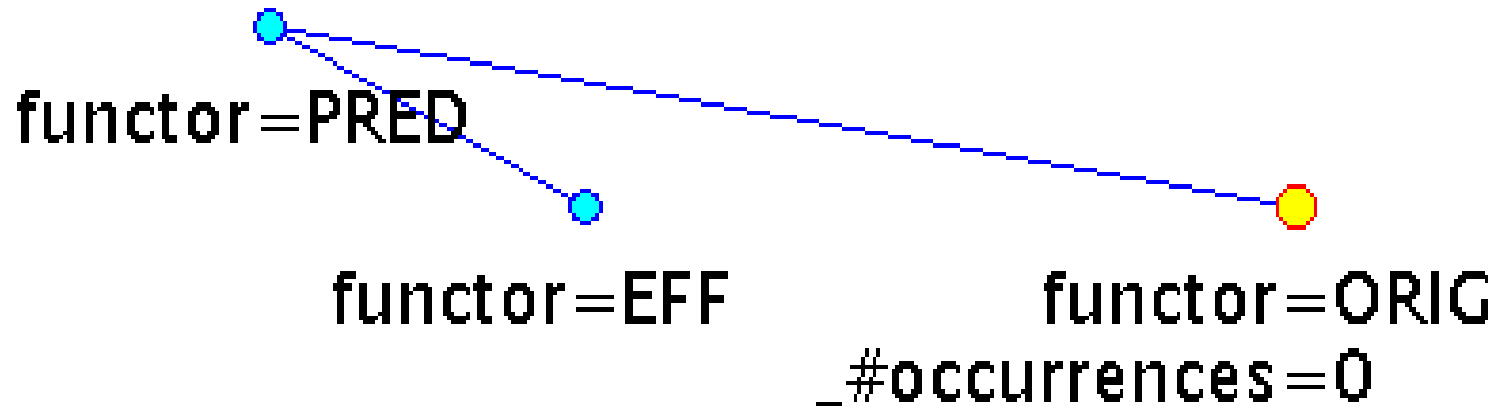
_#sons – A Result Tree



Czech: V. Klaus čeká
4-5% růst HDP.

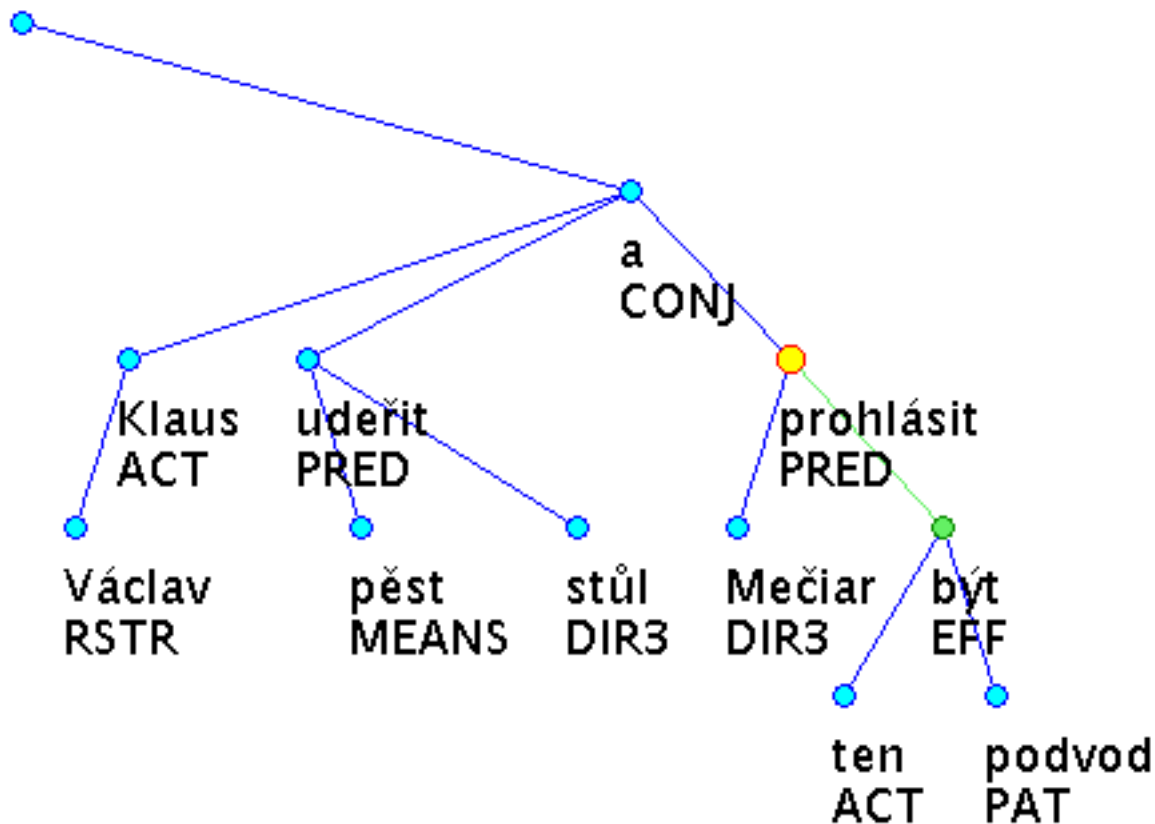
English (lit.):
V. Klaus expects
4-5% increase in GDP.

Meta-attribute `_#occurrences`



- **PRED**icate with at least one son: **EFF**ect, but without an **ORIG**o son

_#occurrences – A Result Tree



Czech: V. Klaus
udeřil pěstí do stolu a
prohlásil směrem k
Mečiarovi: To je
podvod.

English (lit.): V.
Klaus slammed his
fist on the table and
exclaimed towards
Mečiar: It is a deceit.

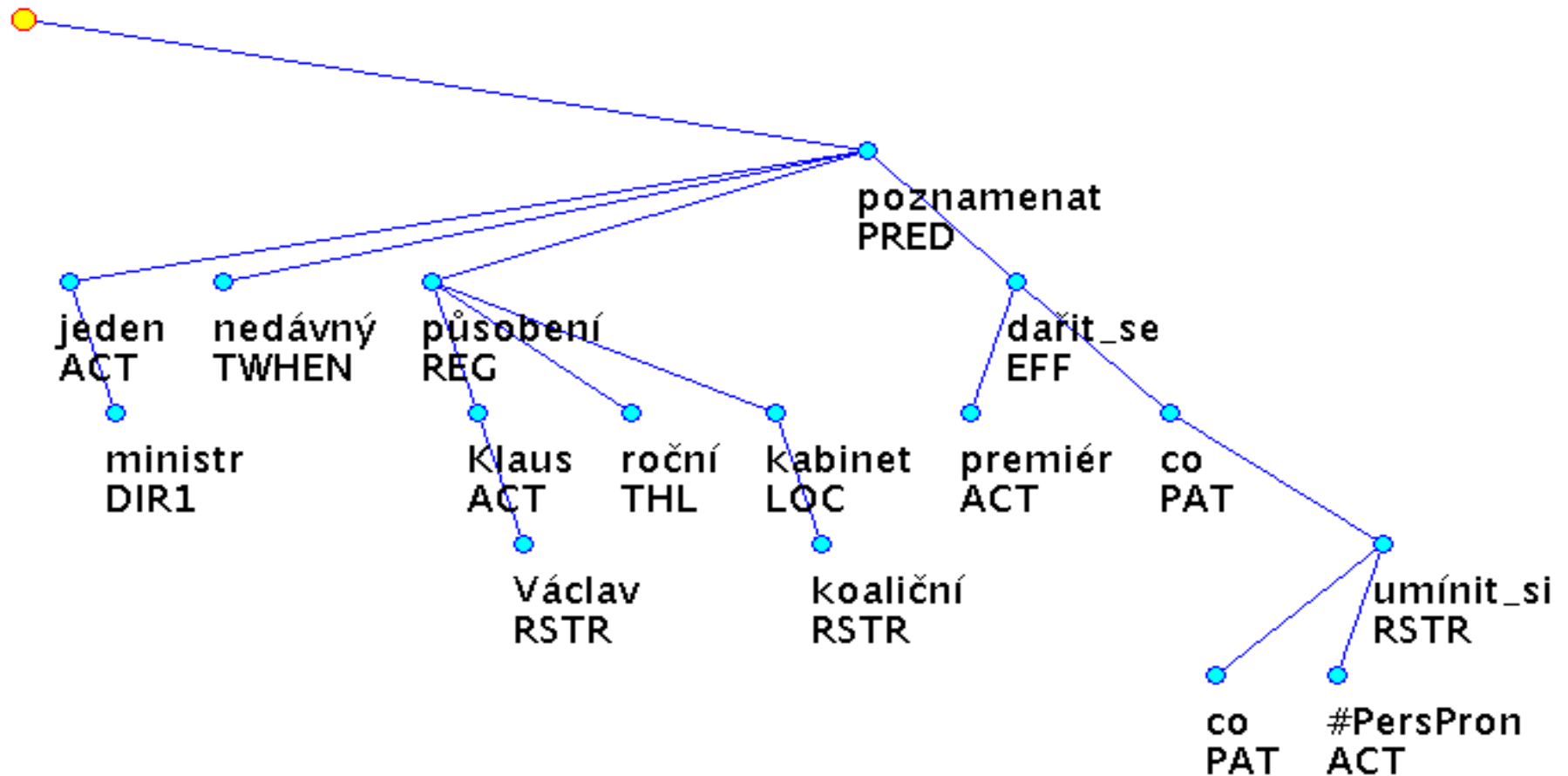
Meta-attribute `_sentence`



```
_depth=0  
_sentence=".*[Vv] souvislosti s .*"
```

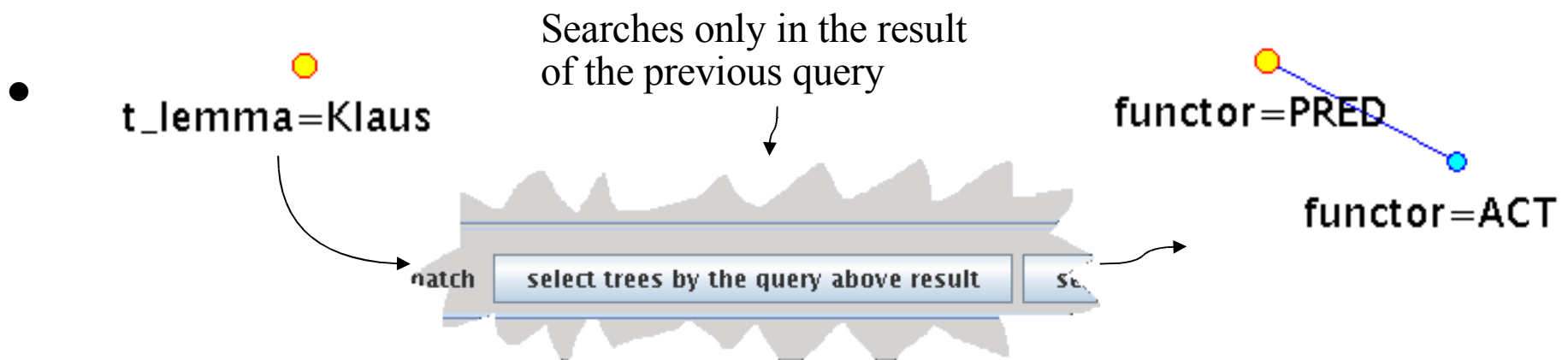
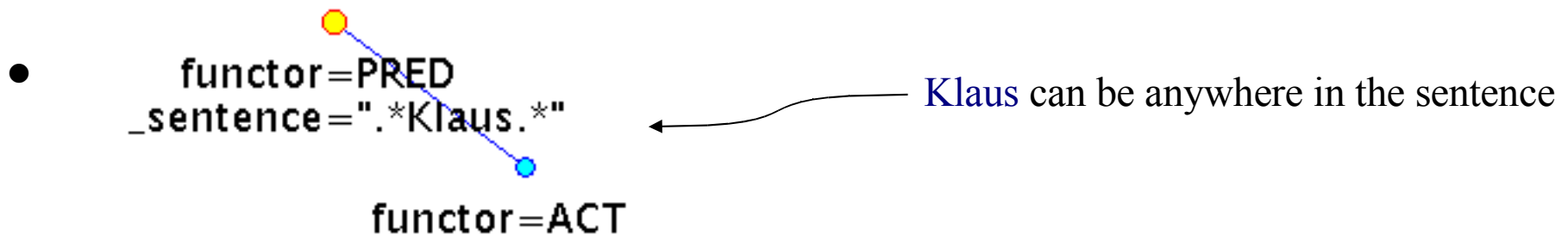
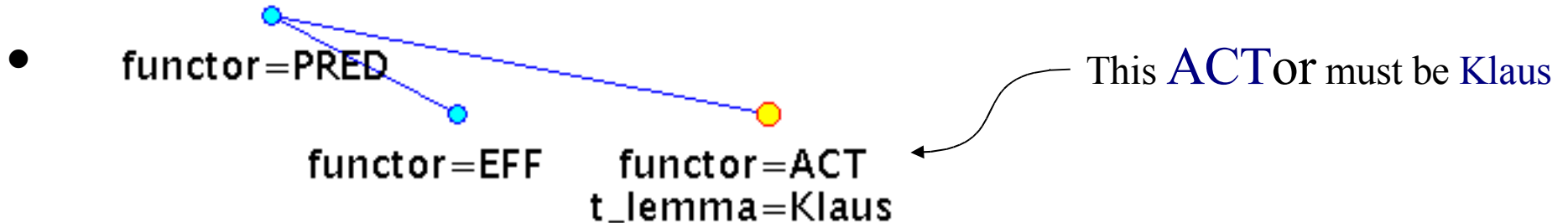
- Searches for sentences containing linear phrase „v souvislosti s“
- Matches all nodes, therefore `_depth=0`

_sentence – A Result Tree



Czech: Jeden z ministrů nedávno **v souvislosti s** ročním působením Václava Klause v čele koaličního kabinetu poznamenal: Premiérovi se daří všechno, co si umínil.

Searching for Klaus



Conclusion

- Netgraph is a simple tool for searching in PDT
- The query language is powerful but comprehensible
- V. Klaus's career provides enough material for linguistic research

References

- Netgraph home page:

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

- Institute of Formal and Applied Linguistics:

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

- Prague Dependency Treebank:

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