



# Annotation of Inter-Sentential Relations in PDT

Jiří Mírovský, Lucie Mladová, Pavlína Jínová, Anja Nedoluzhko, Eva Hajičová, Šárka Zikánová,  
Veronika Pavlíková, Jana Zdeňková, Magdaléna Rysová, Kateřina Rysová, Jiří Pergler, Radek Ocelák  
a další



# Topics of the Annotation

- Extended Textual Coreference
- Bridging Anaphora
- Discourse Relations
  - inter-sentential with expl. connectives
  - intra-sentential (in prep.)

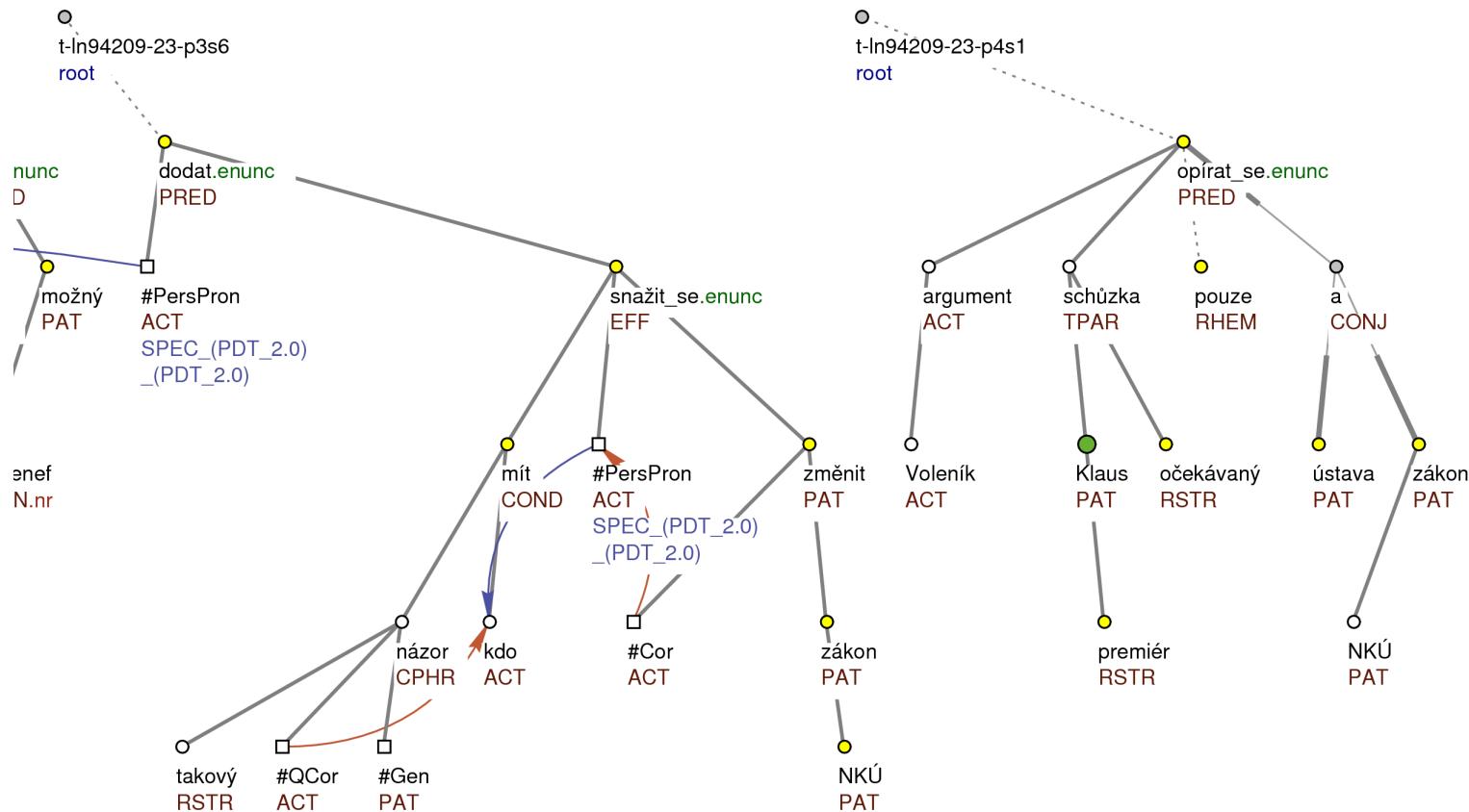


# A Few Sentences from PDT

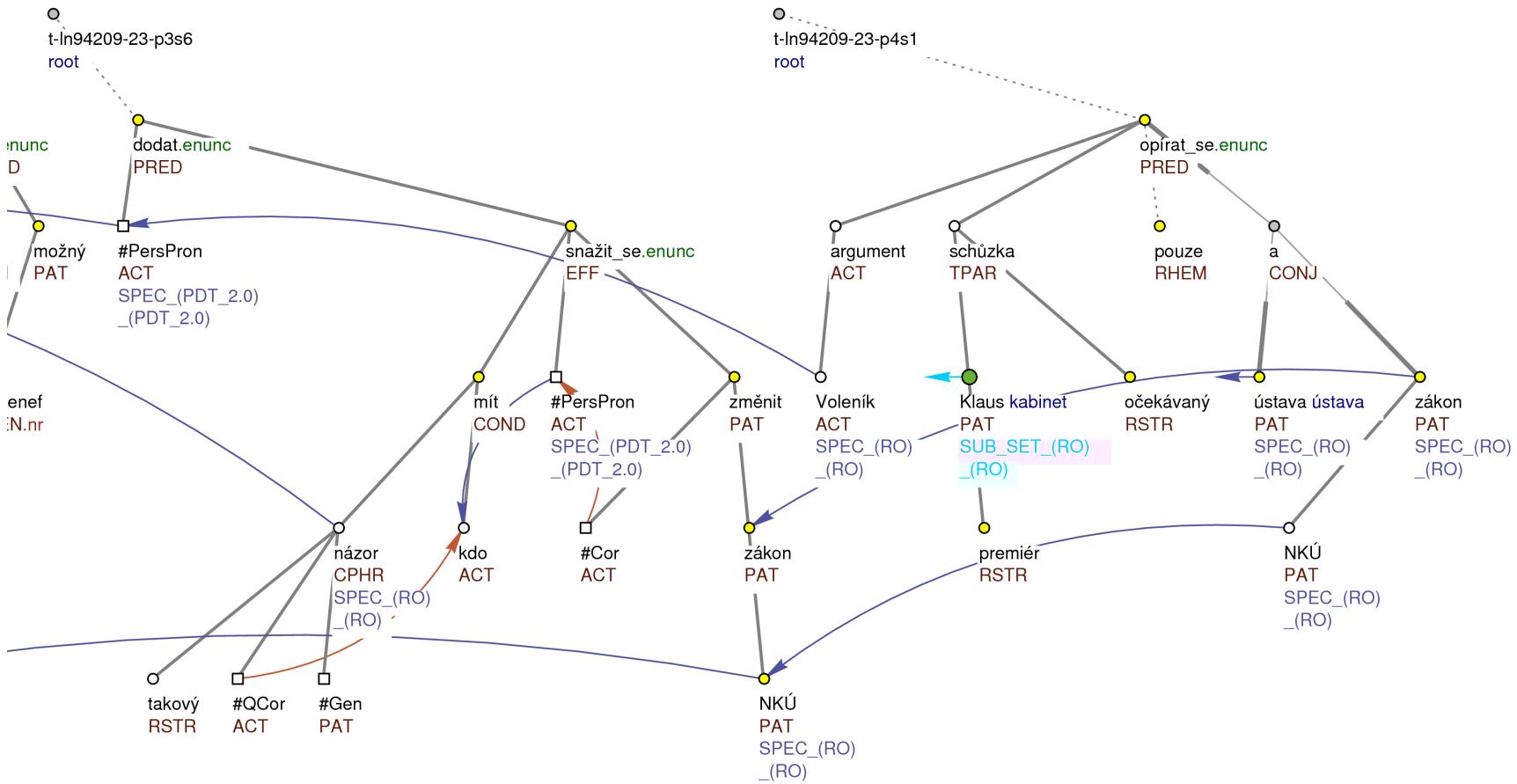
Pokud by někdo takový názor měl,  
musel by se snažit změnit zákon o  
NKÚ, dodal.

Při očekávané schůzce s premiérem  
Klausem se budou Voleníkovy  
argumenty opírat pouze o ústavu a  
zákon o NKÚ.

# Before the Annotation



# After the Annotation





# Before the Annotation

Pokud by někdo<sub>1</sub> takový názor měl,  
[#PersPron<sub>1</sub>] musel by se snažit změnit  
zákon o NKÚ, [#PersPron<sub>2</sub>] dodal.

Při očekávané schůzce s premiérem  
Klausem se budou Voleníkovy  
argumenty opírat pouze o ústavu a  
zákon o NKÚ.



# After the Annotation

Pokud by někdo<sub>1</sub> takový názor<sub>6</sub> měl,  
[#PersPron<sub>1</sub>] musel by se snažit změnit  
zákon<sub>4</sub> o NKÚ<sub>3</sub>, [#PersPron<sub>2</sub>] dodal.

Při očekávané schůzce s premiérem  
Klausem [->kabinet] se budou Voleníkovy<sub>2</sub>  
argumenty opírat pouze o ústavu<sub>5</sub> a  
zákon<sub>4</sub> o NKÚ<sub>3</sub>.



# In Numbers



# E.T. Coref. & Bridging An.

number of annotated documents	3,165
total number of sentences	49,431
total number of words (tokens)	833,195
total number of tectogrammatical nodes (excl. the technical root)	674,965
number of pronominal coref-text links originally in PDT 2.0	20,547
number of newly annotated links	97,973
- number of newly annotated coref-text links	65,802
- number of newly annotated bridging anaphora links	32,171
number of coref-text and bridging links	118,520
% of co-referring nodes (percentage of nodes where a link starts, counting all text-coref and bridging)	17%

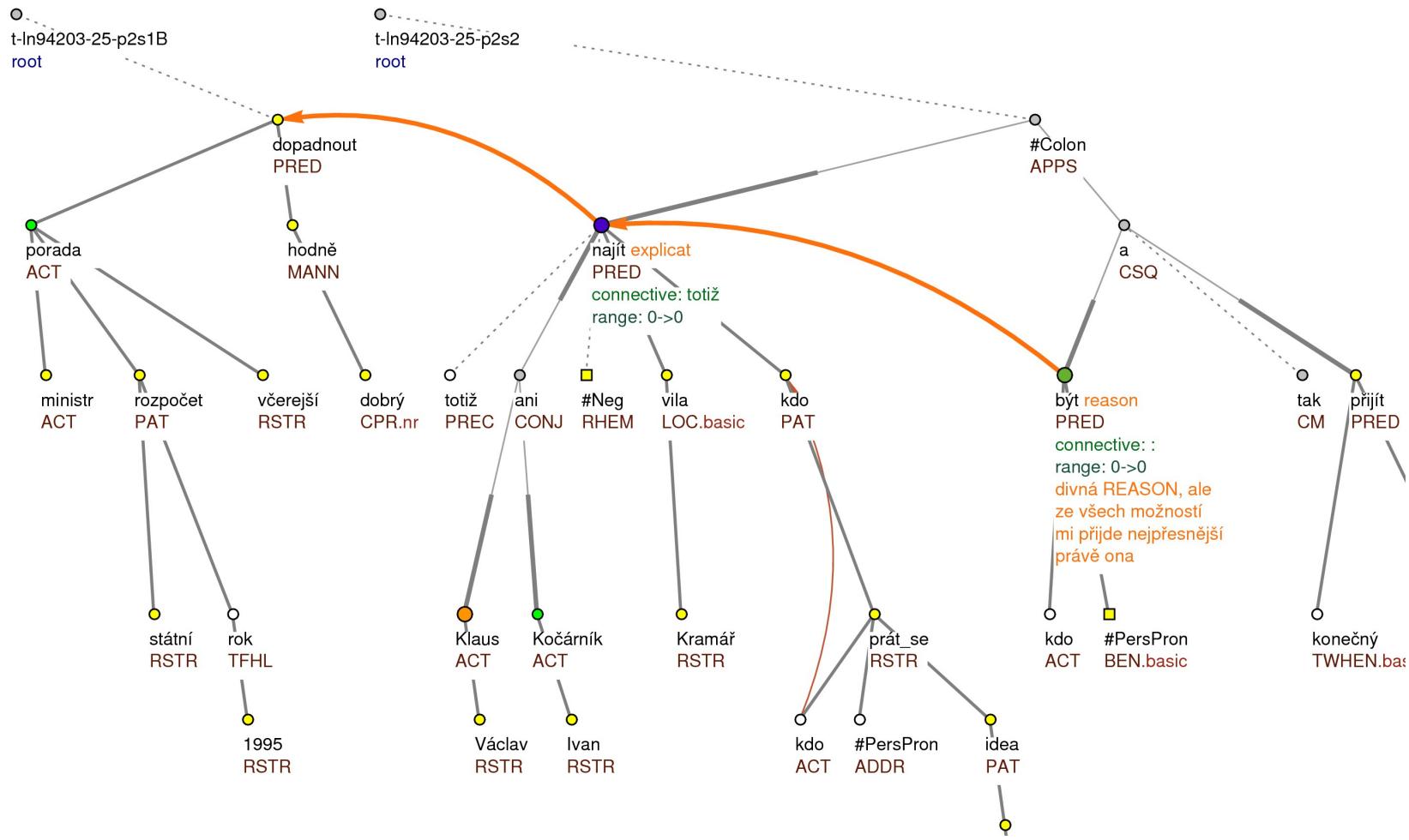


# A Few Sentences from PDT

Včerejší porada ministrů o státním rozpočtu na rok 1995 dopadla víc než dobře.

Václav Klaus ani Ivan Kočárník **totiž** nenašli v Kramářově vile nikoho, kdo by se s nimi chtěl prát o ideu vyrovnaného rozpočtu: všichni byli pro, a tak...

# The Annotation





# Inter-Sent. Discourse Rel.

number of annotated documents	3,165
total number of sentences	49,431
total number of words (tokens)	833,195
total number of tectogrammatical nodes (excl. the technical root)	674,965
number of discourse links originally in PDT 2.0	0
number of newly annotated (inter-sentential) links	6,571
- number of newly annotated discourse links	6,214
- number of newly annotated 'list' links	357



Now Working on...



# E.T. Coref. & Bridging An.

- Transformation to PDT 2.5 (almost done)
- Annotation of 1<sup>st</sup> and 2<sup>nd</sup> person
- Annotation of English data - English part of PCEDT (seems to have priority)



# Discourse Relations

- Transformation to PDT 2.5 (in progress)
- Annotation of intra-sentential relations (in prep.)
- Extensive checks of the annotated data (in progress)



# Thank you!