

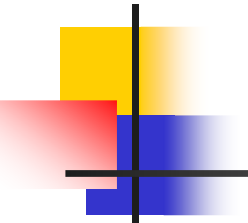
# Automatická předanotace TFA v české části PCEDT



---

[GAP406/10/0875 \(Komputační lingvistika: Explicitní popis jazyka a anotovaná data se zřetelem na češtinu\)](#)

[GAP406/12/0658 \(Koreference, diskurs a aktuální členění v kontrastivním pohledu\)](#)

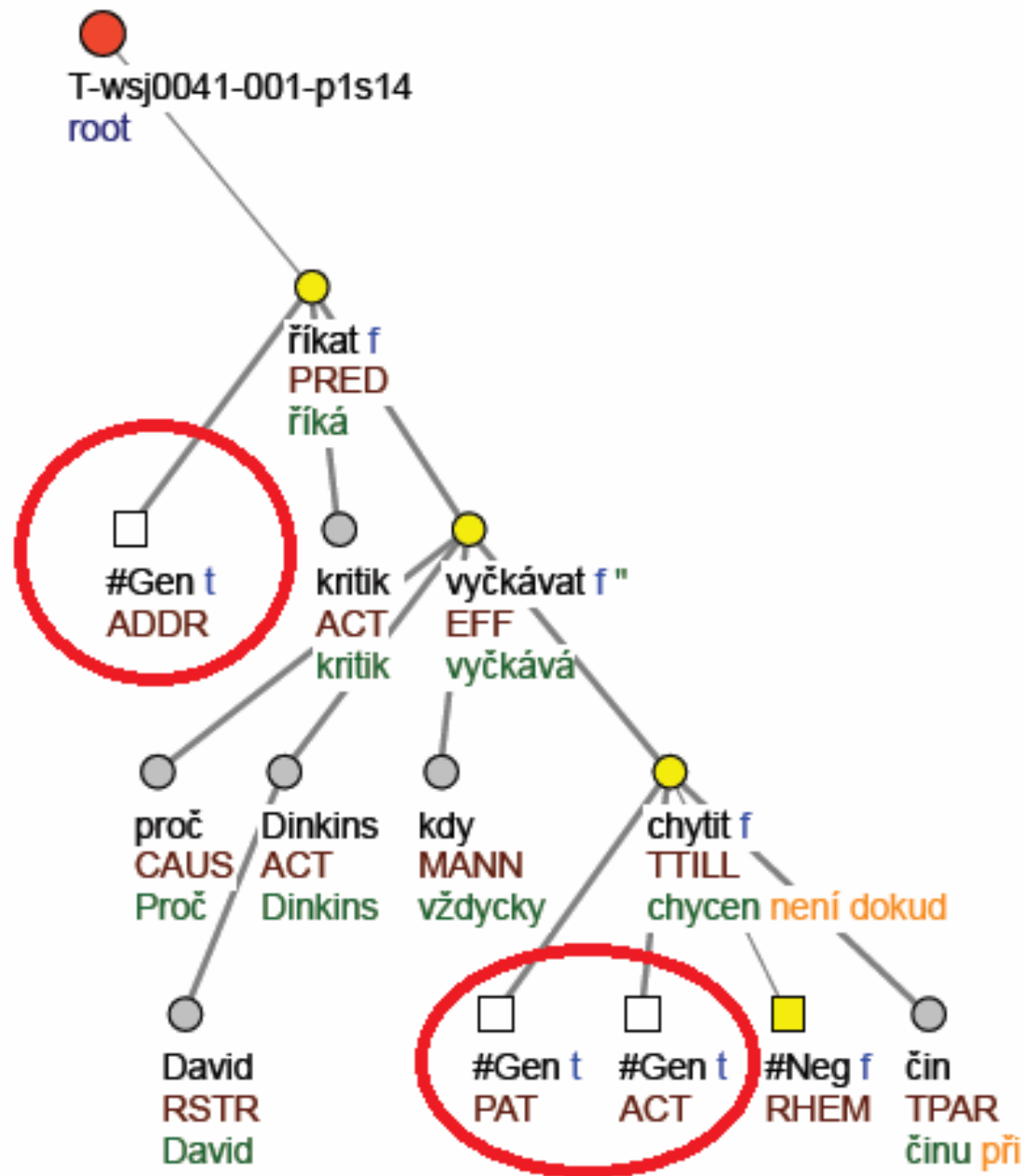
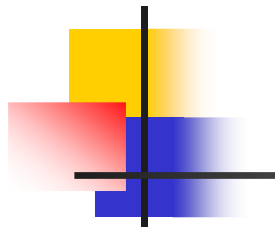
- 
- 
- *Anotace na tektogramatické rovině Pražského závislostního korpusu* (Mikulová a kol. 2005)  
– kapitola „Aktuální členění“
  - Eva Hajičová, Jiří Mírovský, Kateřina Rysová, Magdaléna Rysová
  - <https://wiki.ufal.ms.mff.cuni.cz/external:pcedt:tfa>



# Generované uzly

---

- generované uzly, které nemají analytický protějšek (tj. nejsou to kopírované uzly) a které navíc nejsou RHEM ani #Forn, dostávají automaticky hodnotu  $tfa="t$ ,
- očekávaná chybovost: 0
- *„Proč David Dinkins,“ říká kritik, „vždycky vyčkává, dokud není chycen při činu?“*

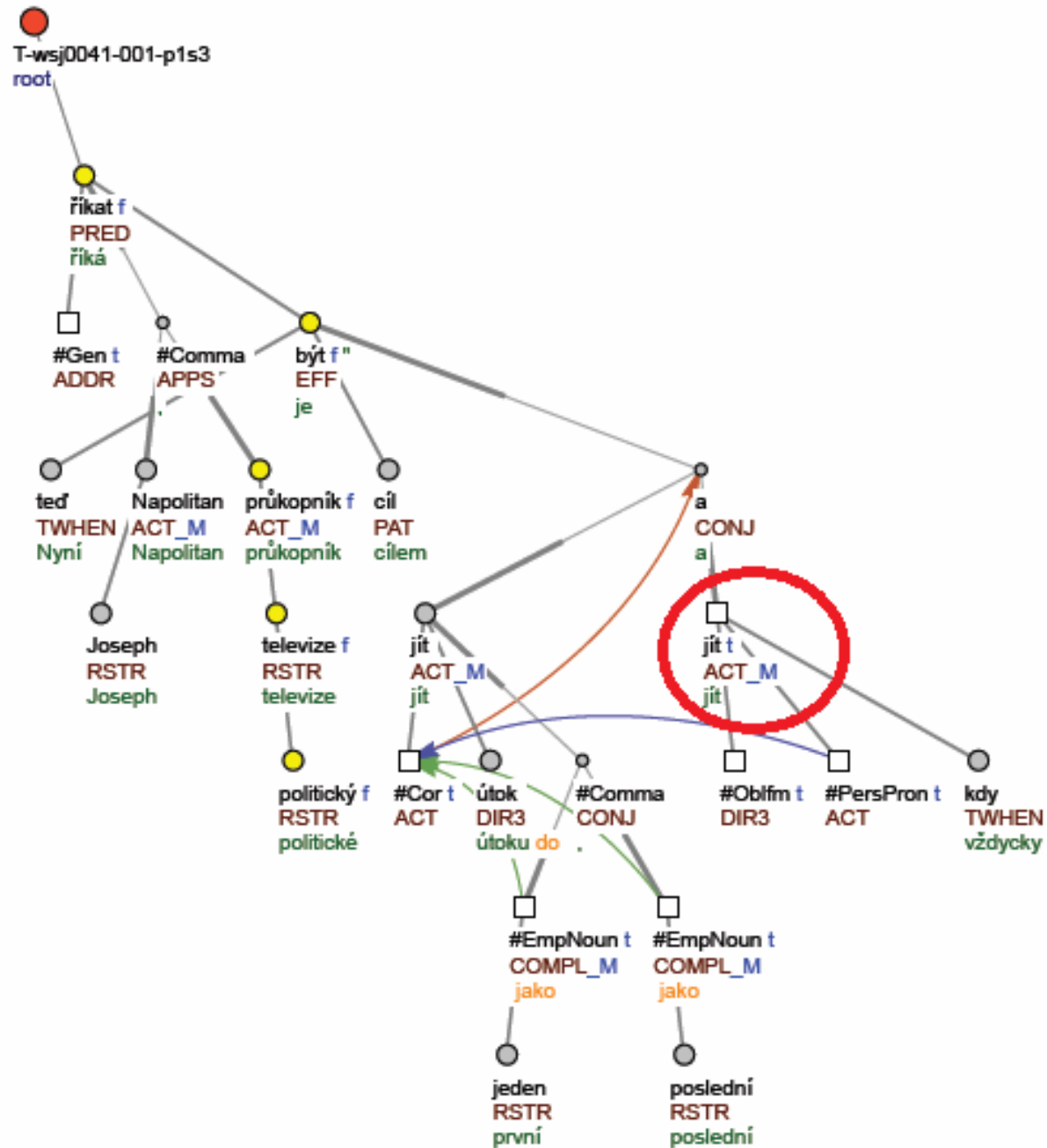
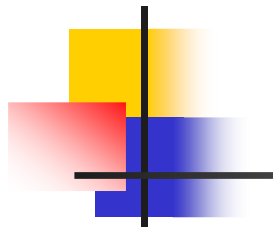




# Generované uzly

---

- generované uzly, které jsou členy koordinace/apozice a mají analytický protějšek (není to tedy např. #Forn), dostávají automaticky hodnotu  $tfa="t,$
- očekávaná chybovost: 0
- *„Nyní,“ říká Joseph Napolitan, průkopník politické televize, „je cílem jít do útoku jako první, poslední a [jít] vždycky.“*

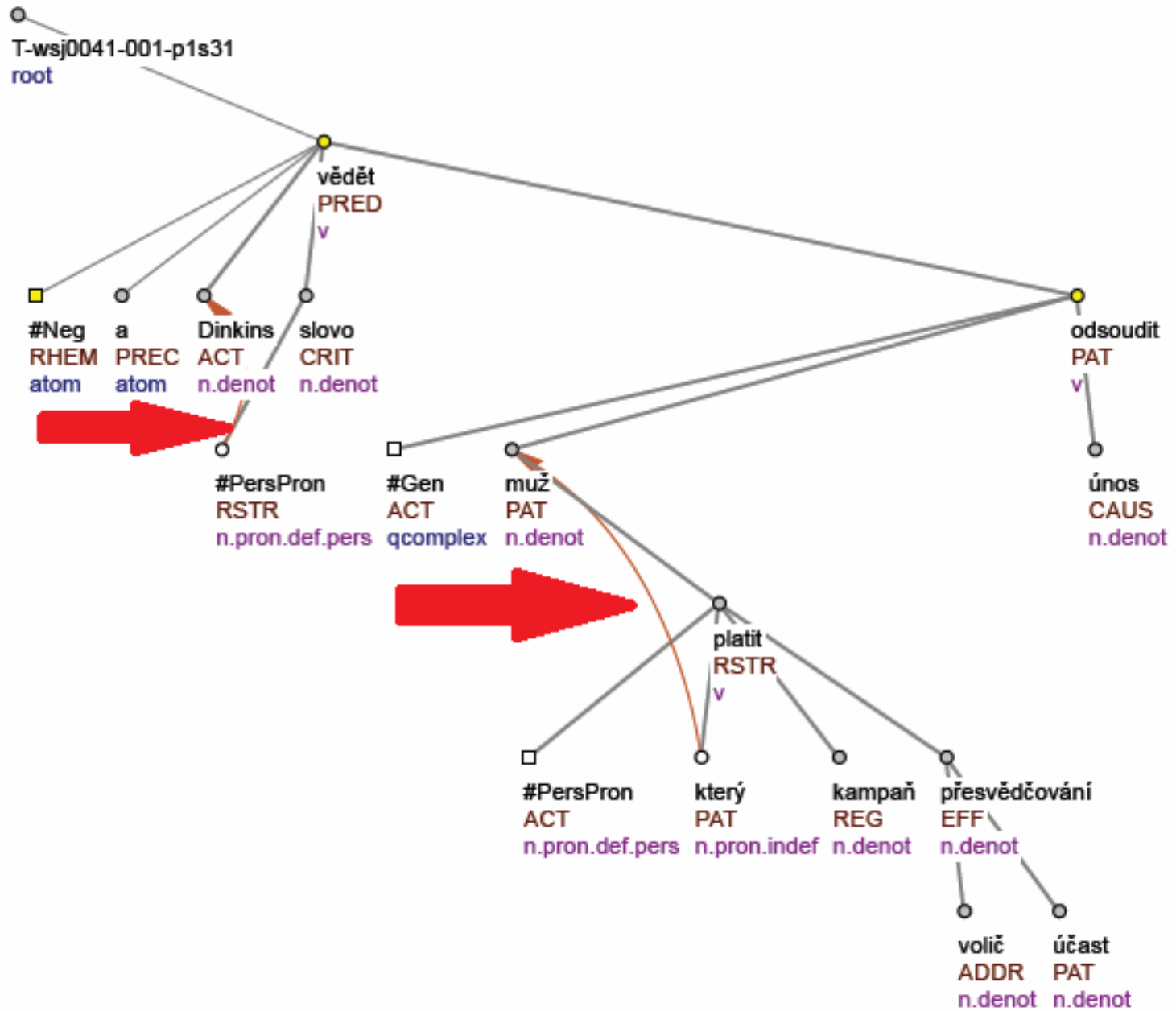
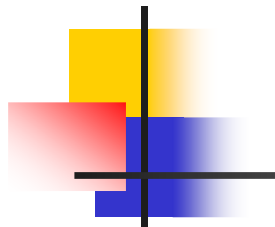




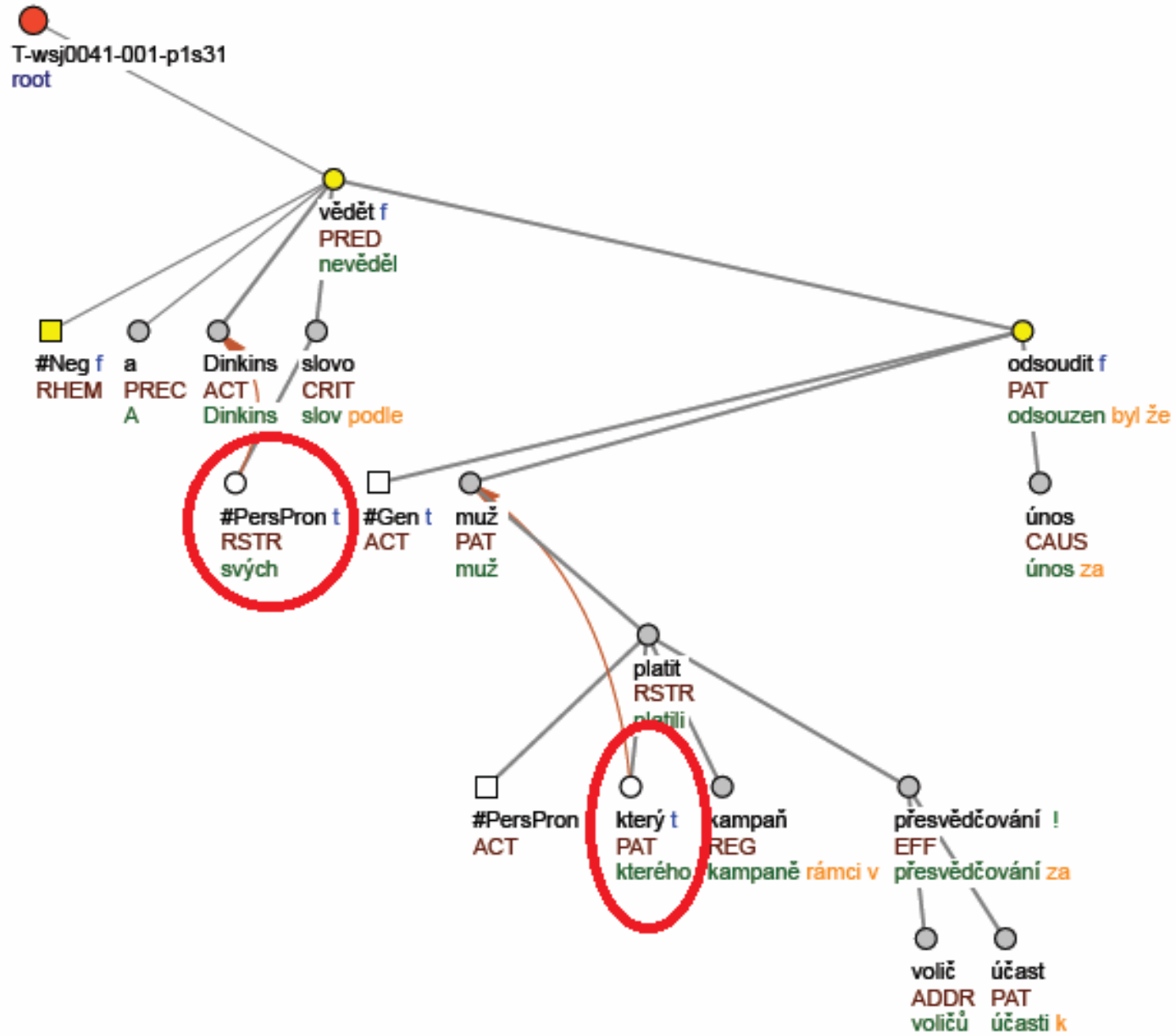
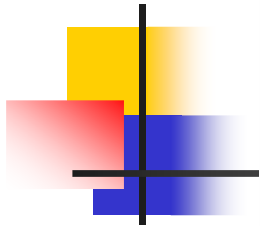
# Koreference

---

- uzly, ze kterých vede gramatická, textová či segmentová koreference, dostávají automaticky hodnotu  $tfa="t"$
- očekávaná chybovost: 1:100
- *A Dinkins podle svých slov nevěděl, že muž, kterého platili v rámci kampaně za přesvědčování voličů k účasti, byl odsouzen za únos.*





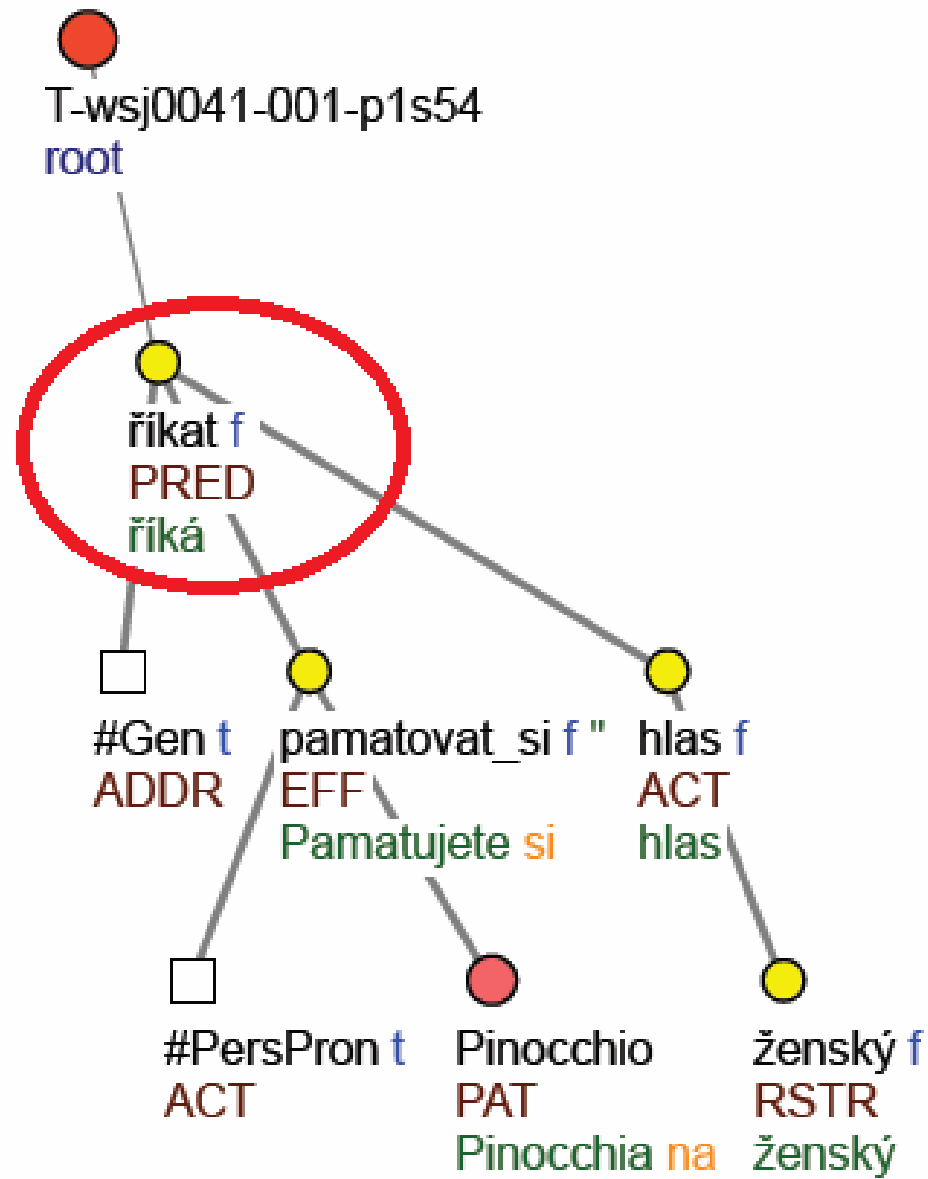
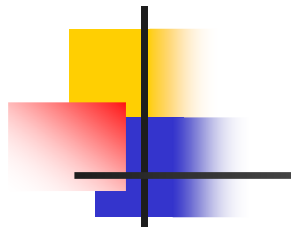




# Uzly PRED

---

- uzly PRED, které nejsou generované a jejichž t\_lemma se nenachází v předchozí větě, dostávají automaticky hodnotu tfa="f"
- očekávaná chybovost: 1:40
- „*Pamatujete si na Pinocchia?*," *říká* ženský hlas.

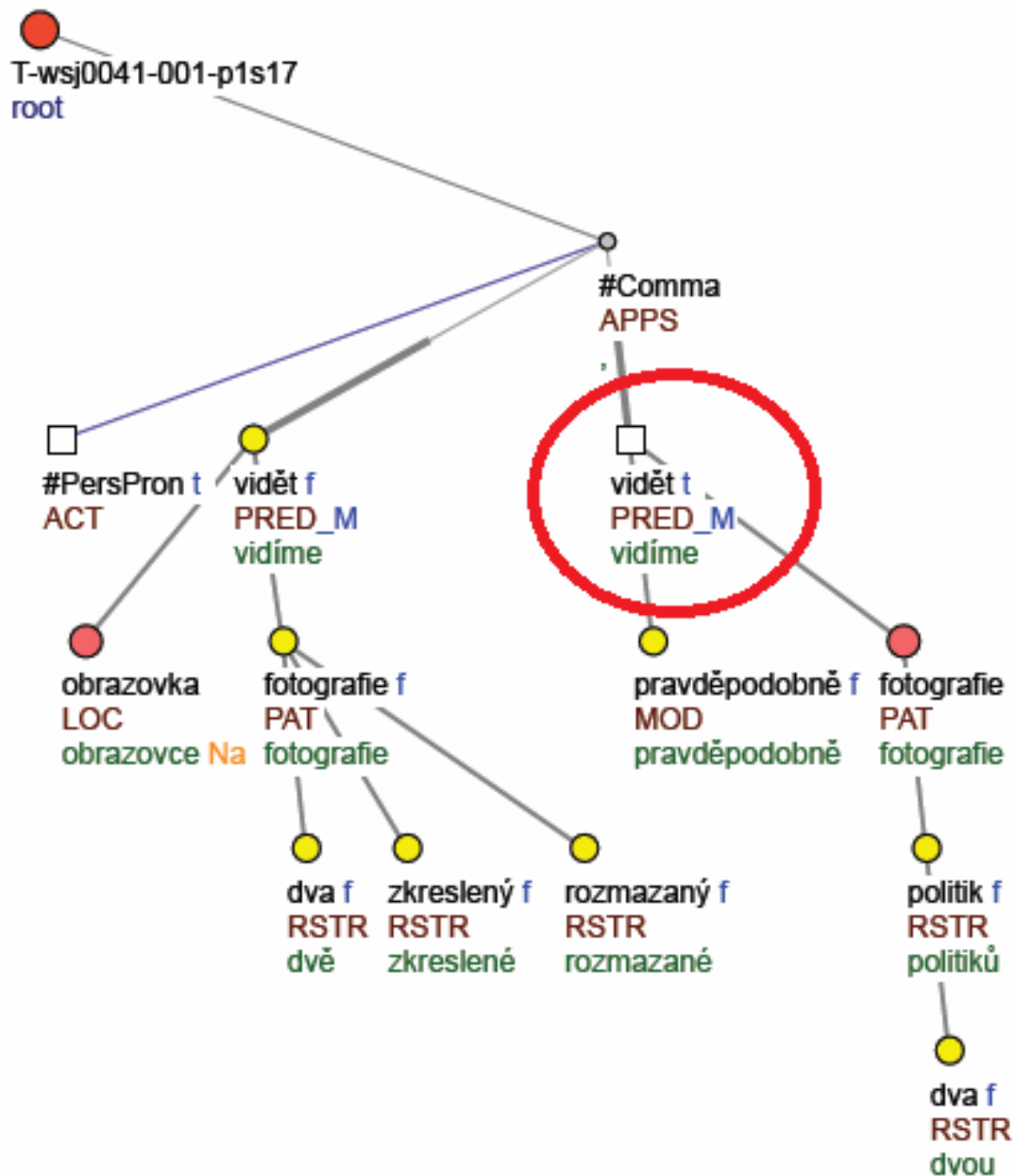
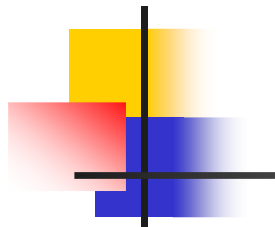




## Uzly PRED

---

- generované uzly PRED dostávají automaticky hodnotu  $tfa="t"$
- očekávaná chybovost: 1:100
- *Na obrazovce vidíme dvě zkreslené rozmazané fotografie, pravděpodobně [vidíme] fotografie dvou politiků.*

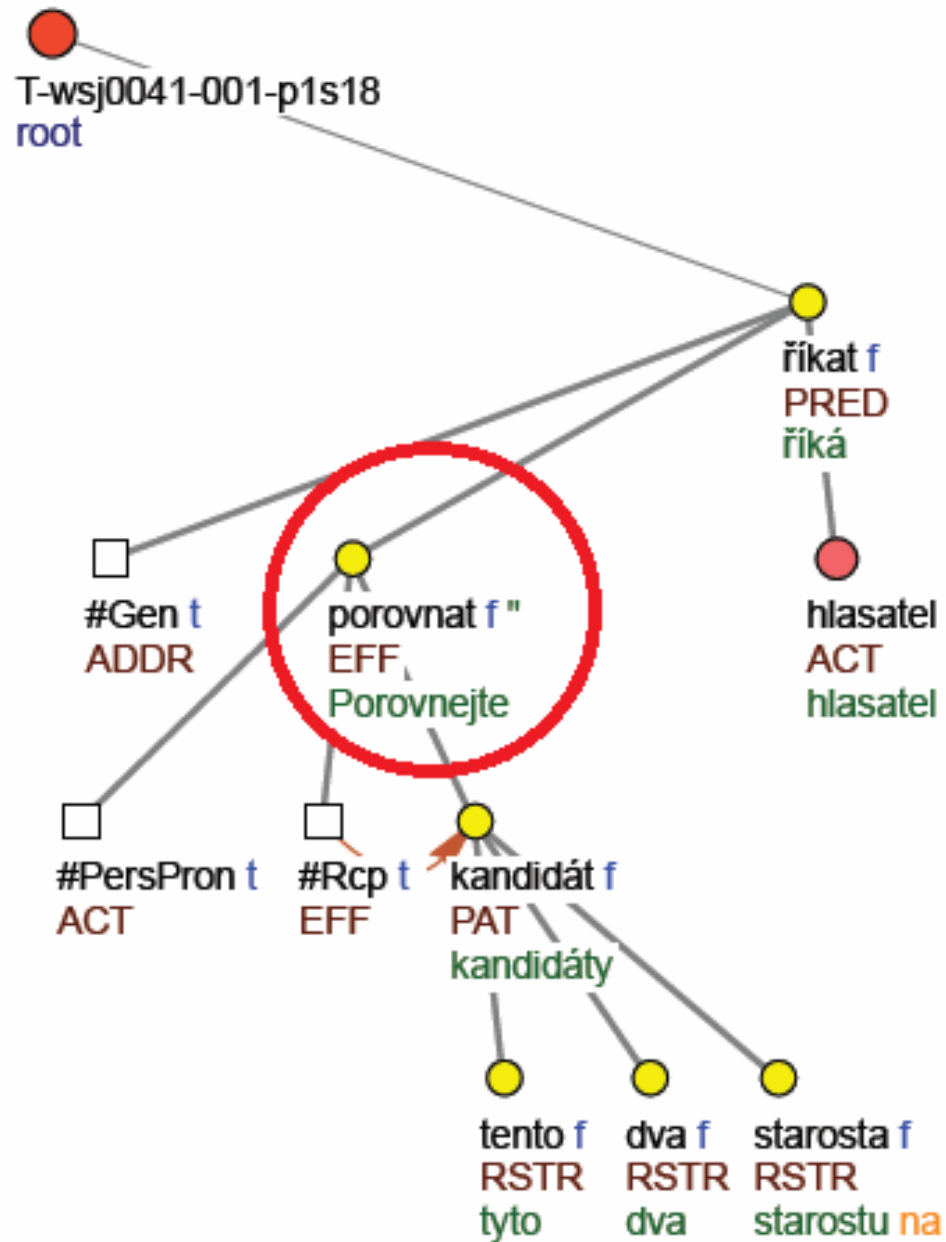
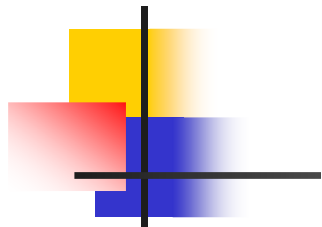




# Ostatní slovesné uzly

---

- ostatní slovesné uzly (gram/sempos="v"), které mají jako funktor jednu z hodnot (ADDR|AIM|CAUS|ACMP|MANN|PAT|EFF|AUT|H|BEN|COMPL|EXT|ORIG|RESL|TFHL|TSIN), dostávají automaticky hodnotu tfa="f"
- očekávaná chybovost: maximálně 1:10
- *„**Porovnejte** tyto dva kandidáty na starostu,“ říká hlasatel.*



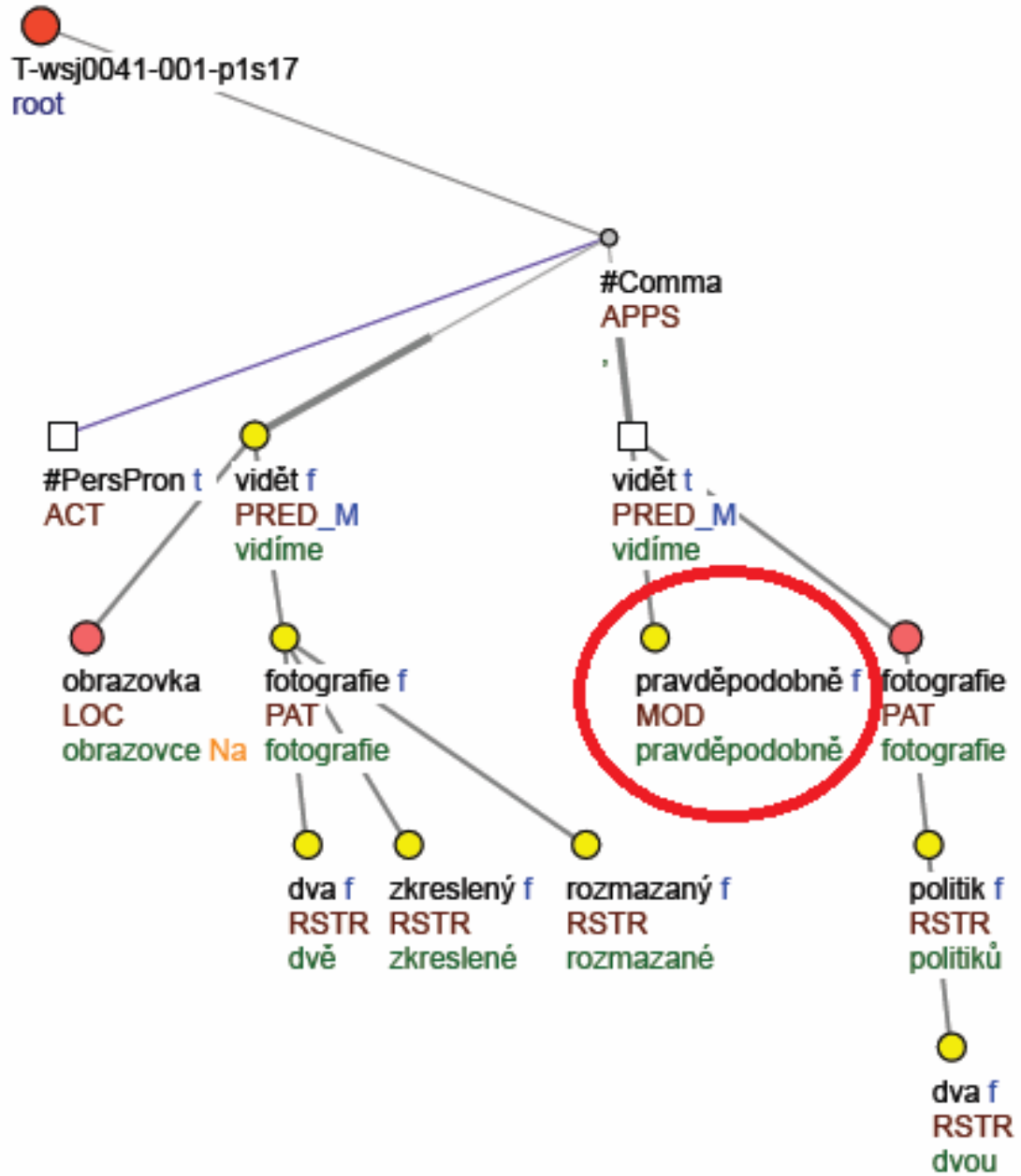
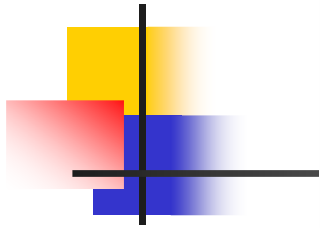


# Uzly PARTL|DENOM|MOD|EXT

---

- uzly, které mají jako funktor jednu z hodnot (PARTL|DENOM|MOD|EXT), dostávají automaticky hodnotu  $tfa="f"$
- očekávaná chybovost: maximálně 1:10
- *Na obrazovce vidíme dvě zkreslené rozmazané fotografie, **pravděpodobně** fotografie dvou politiků.*



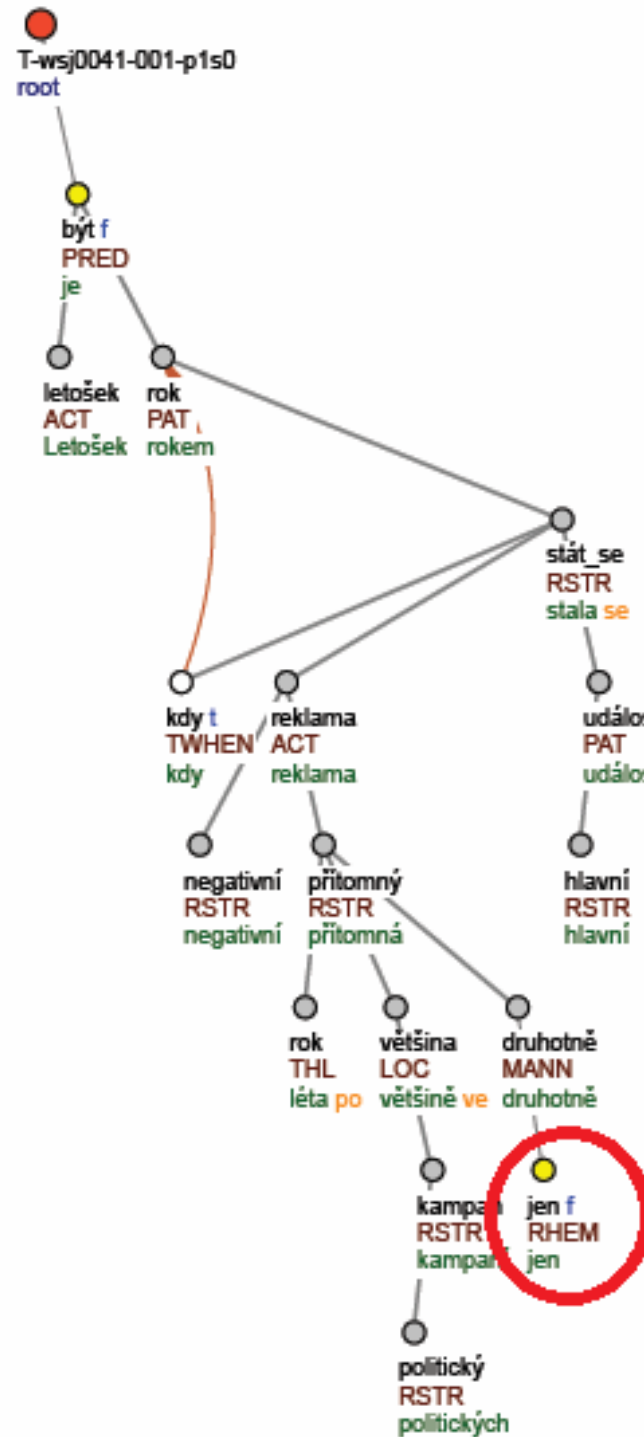




# Uzly RHEM

---

- uzly, které mají jako funktor hodnotu RHEM a nejsou na první pozici ve větě, dostávají automaticky hodnotu  $tfa="f"$
- očekávaná chybovost: 1:10
- *Letošek je rokem, kdy se negativní reklama, po léta přítomná ve většině politických kampaní **jen** druhotně, stala hlavní událostí.*





## „Tady“

- uzly s `t_lemma="tady"` dostávají automaticky hodnotu `tfa="t,,`
- očekávaná chybovost: 1:10
- *Ředitelka Wardová se rozhodla zbavit se „balastu“ v učitelském sboru a obnovit bezpečnost a také **tu** byly další nové faktory, které pracovaly v její prospěch.*

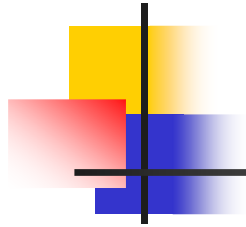




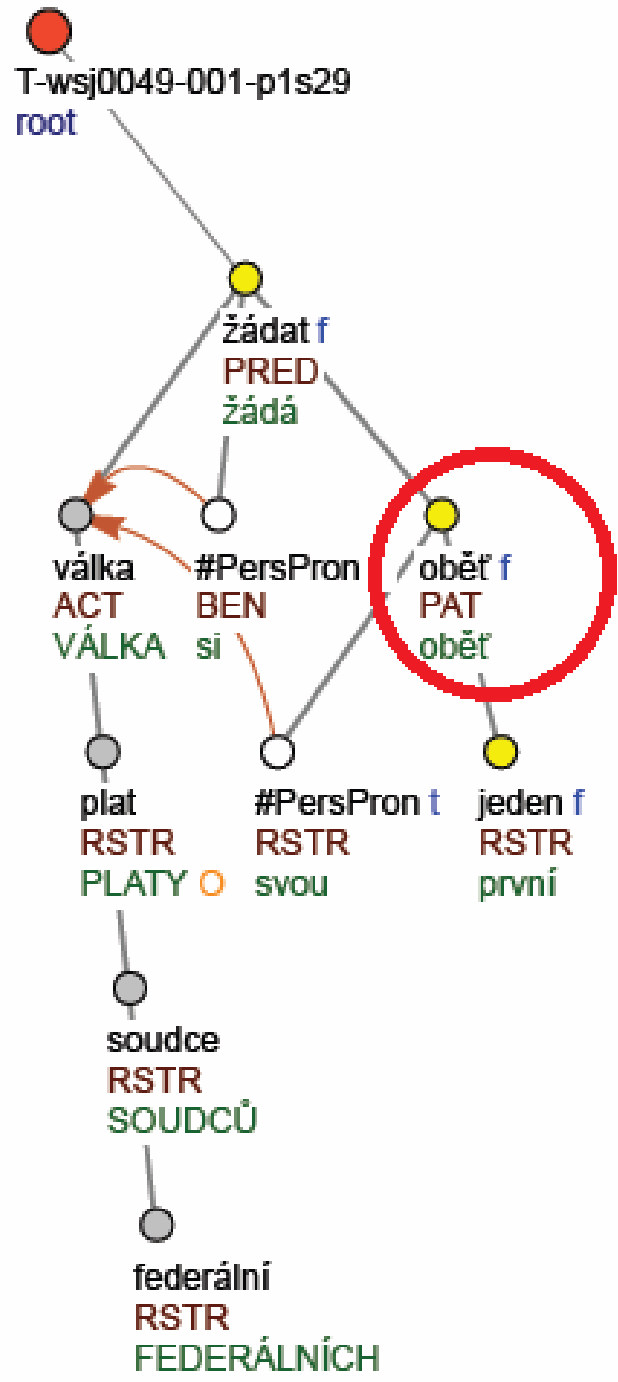
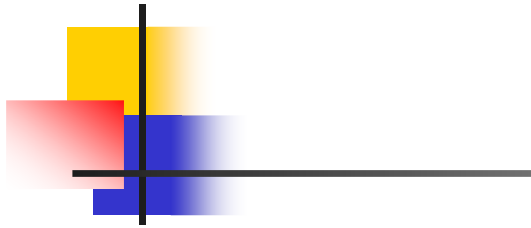
# Neurčitý člen

---

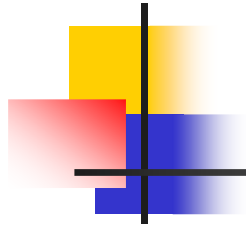
- uzly, které jsou českými protějšky anglických uzlů, které na povrchu stojí za slovesem a mají u sebe neurčitý člen, dostávají automaticky hodnotu `tfa="f"`
- očekávaná chybovost: neznámá



- THE WAR OVER FEDERAL JUDICIAL SALARIES takes a victim.
- VÁLKA O PLATY FEDERÁLNÍCH SOUDCŮ si žádá svou první oběť.







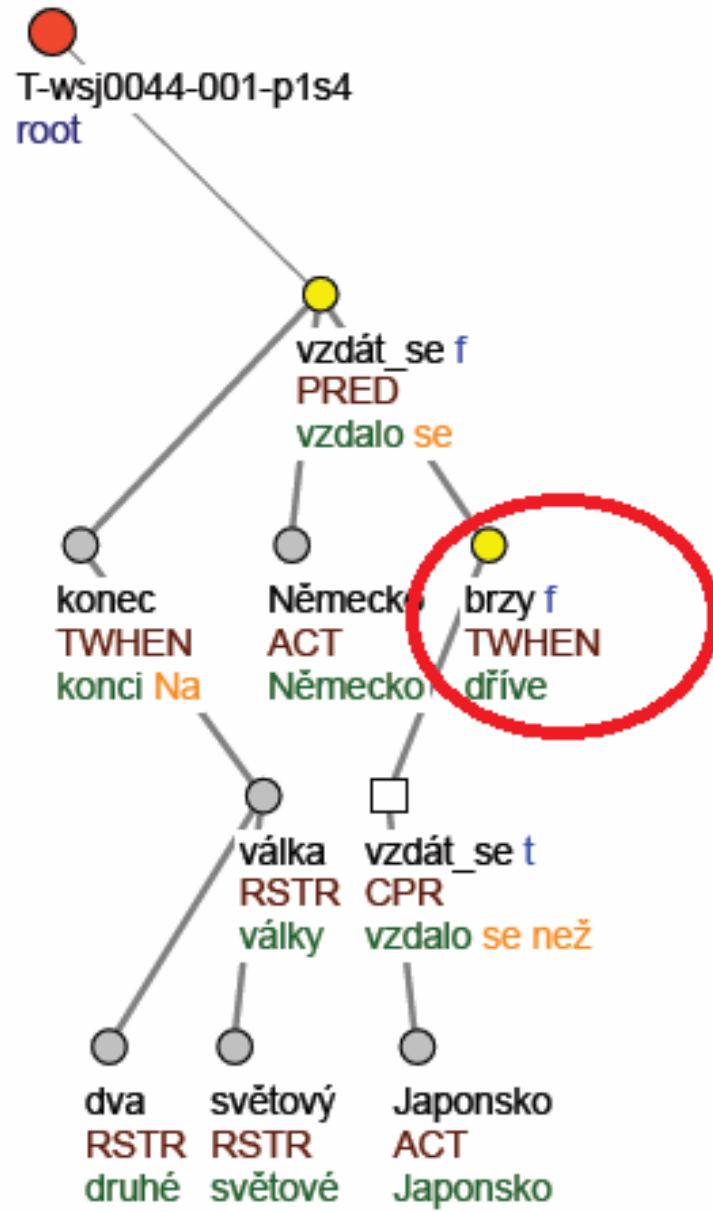
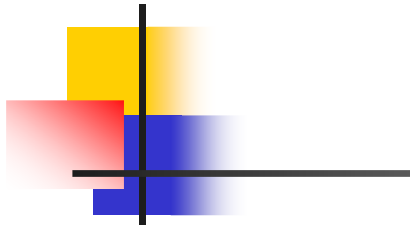
Poté, co jsou aplikovány  
všechny předchozí kroky:



# Synové ohniskového slovesa

---

- synové slovesa, které má  $tfa="f"$  a které ve své klauzi není na první či druhé pozici, dostávají automaticky hodnotu  $tfa="f,,,"$ , pokud stojí v povrchové slovosledu za slovesem
- očekávaná chybovost: neznámá
- *Na konci druhé světové války se Německo vzdalo **dříve** než Japonsko...*

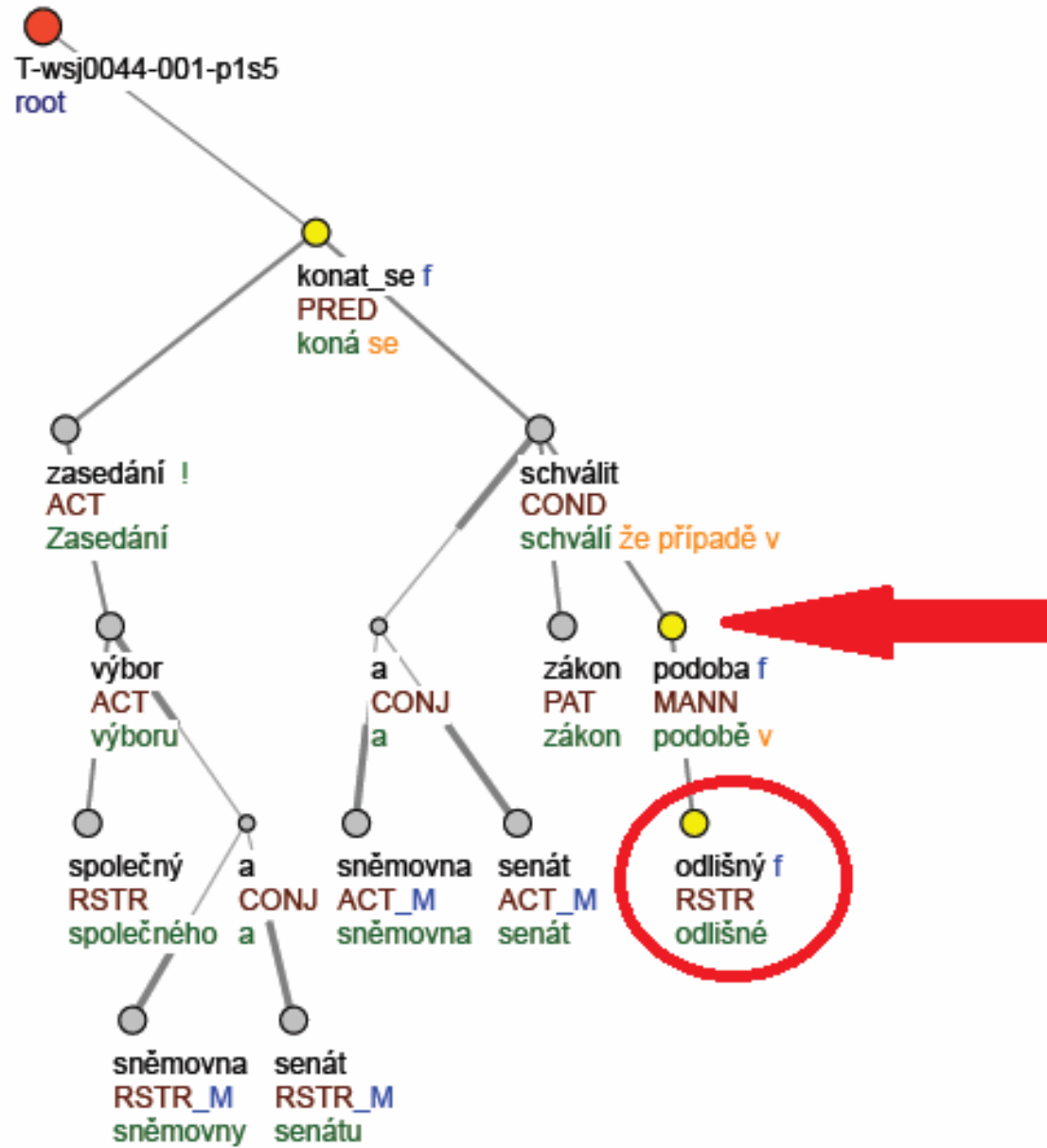
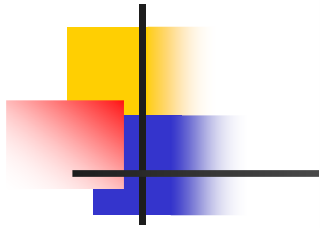


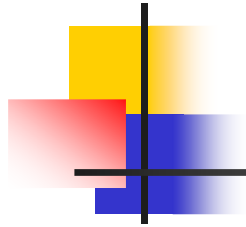


## Uzly RSTR

---

- RSTR synové uzlů, které mají  $tfa="f"$ , dostávají automaticky hodnotu  $tfa="f,,$
- očekávaná chybovost: 1:30
- *Zasedání společného výboru sněmovny a senátu se koná v případě, že sněmovna a senát schválí zákon v **odlišné** podobě.*





Děkujeme za pozornost