

# 3. Výstupy SYNTAKTICKÉ ANALÝZY

Jakub Machura

Masarykova univerzita  
Ústav českého jazyka  
machura@phil.muni.cz

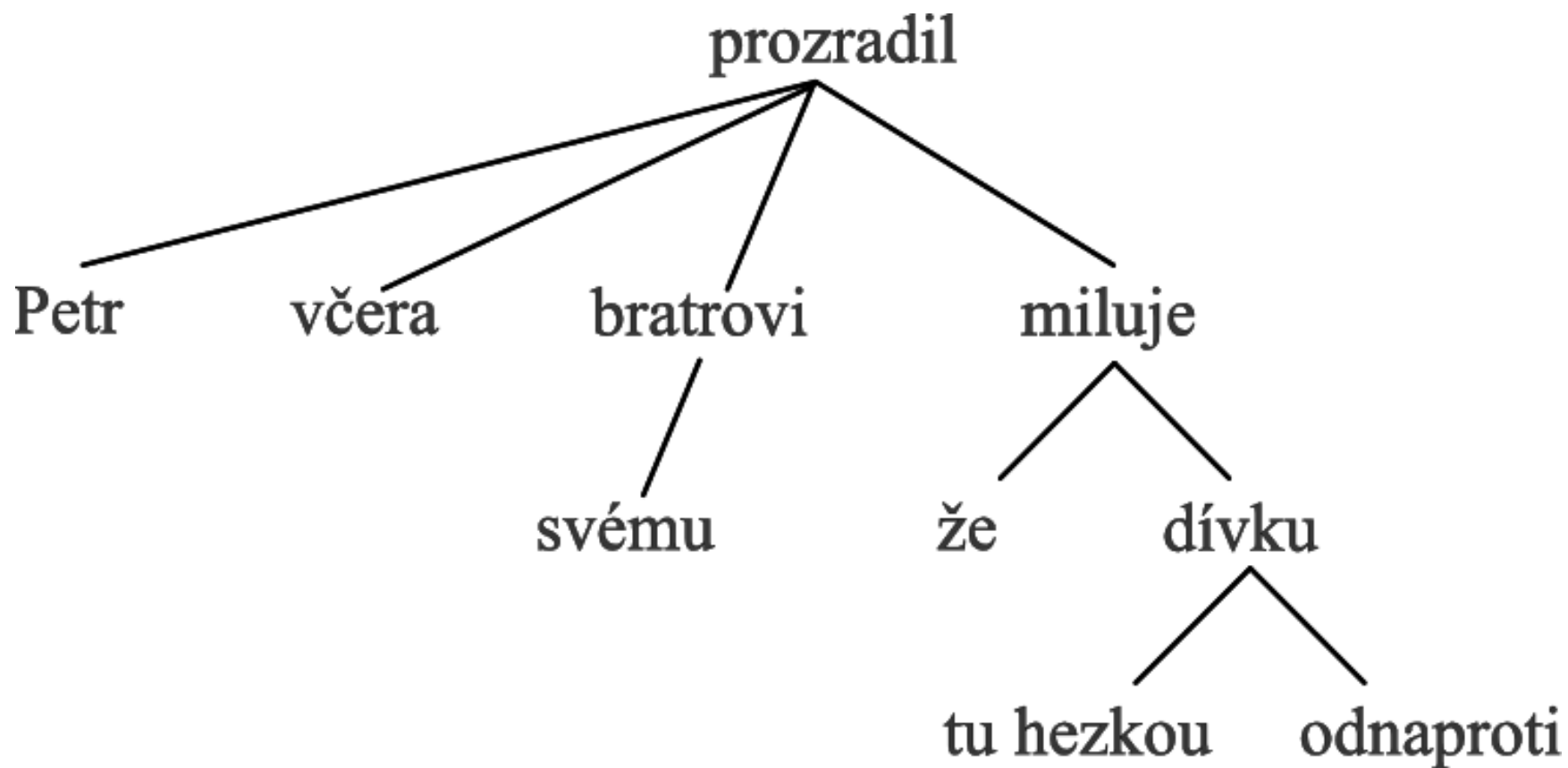
# TEORETICKÁ VÝCHODISKA

---

- L. Tesnière
- navazuje V. Šmilauer a PLK
- vykrytalizovalo v FGP
- binární vztah řídicího a závislého slova
- závislostní strom, treebanky

# ZÁVISLOSTNÍ STROM

---



# ZÁVISLOSTNÍ STROM

---

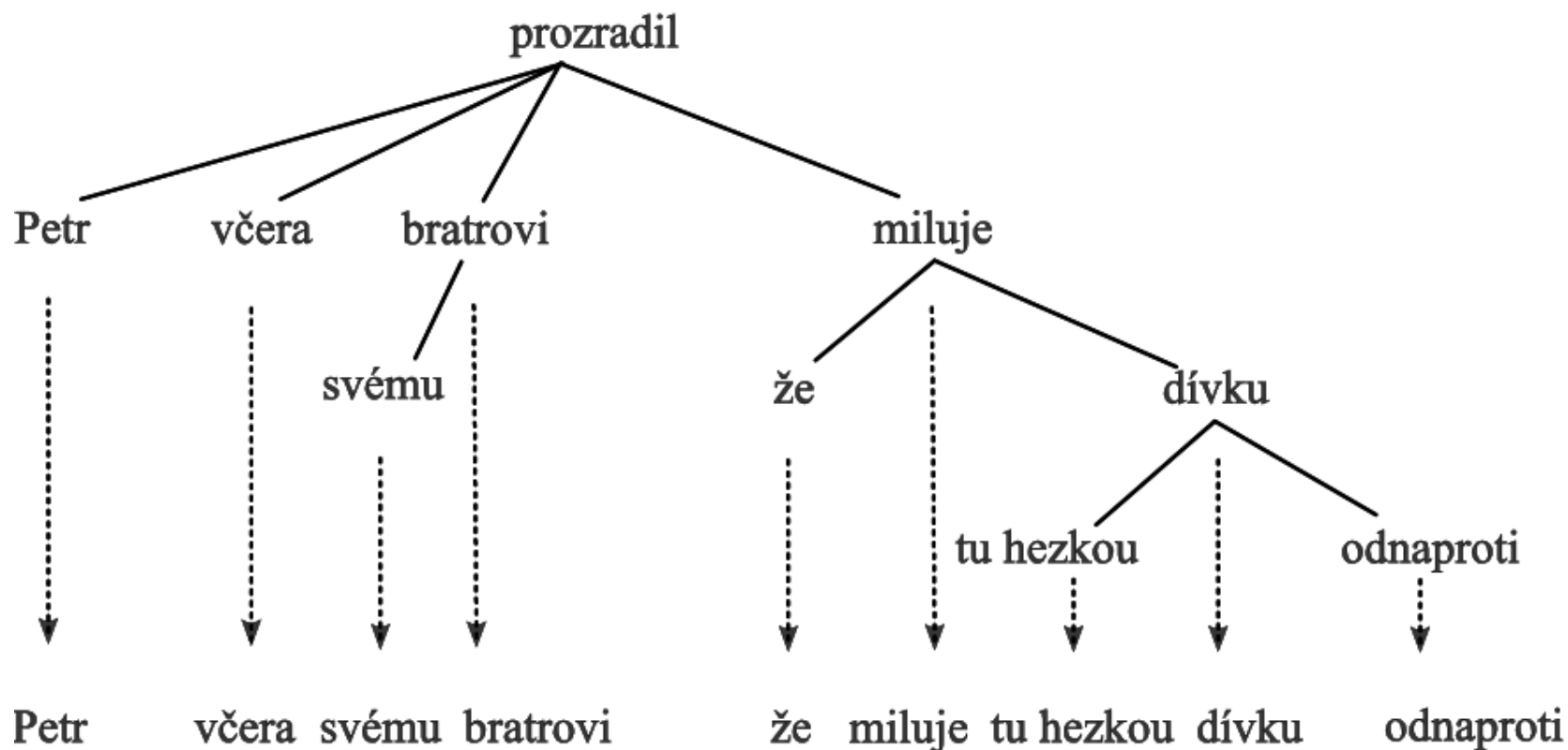
- uzly a hrany
- souvislý graf
- orientovaný graf
- „australské stromy“

# ZÁVISLOSTNÍ STROM

řídící (rozvíjené)	podřízené (rozvíjející)
prozradil	Petr
prozradil	včera
bratrovi	svému
prozradil	bratrovi
miluje	že
prozradil	miluje
dívku	tu
dívku	hezkou
miluje	dívku
dívku	odnaproti

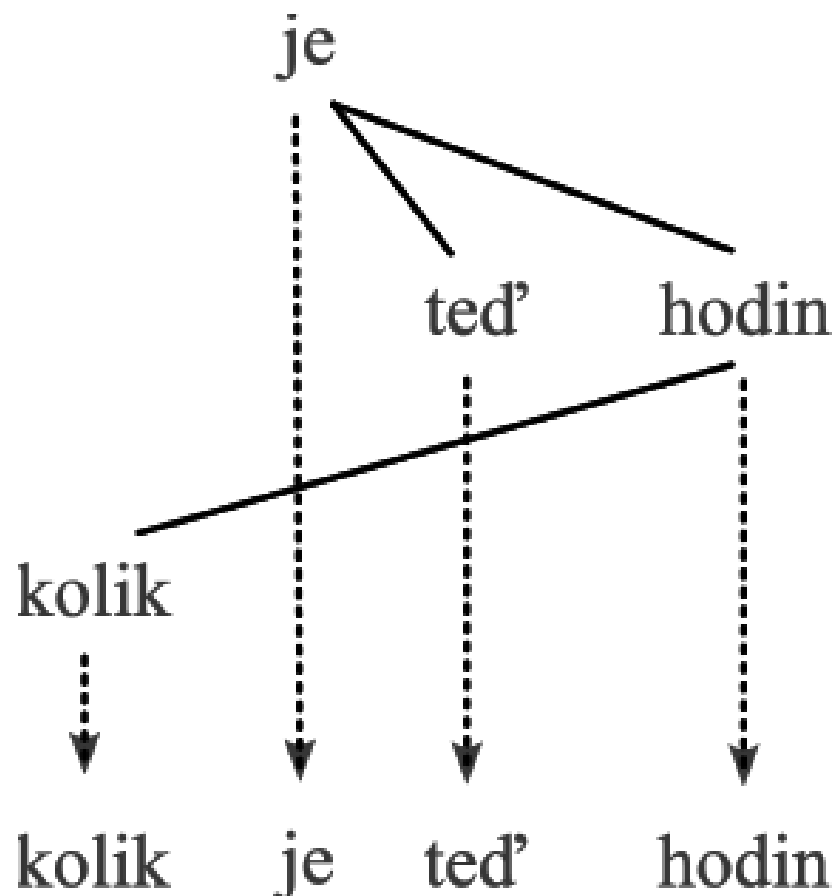
# Závislostní strom

Závislostní strom „**uspořádaný**“



# Závislostní strom

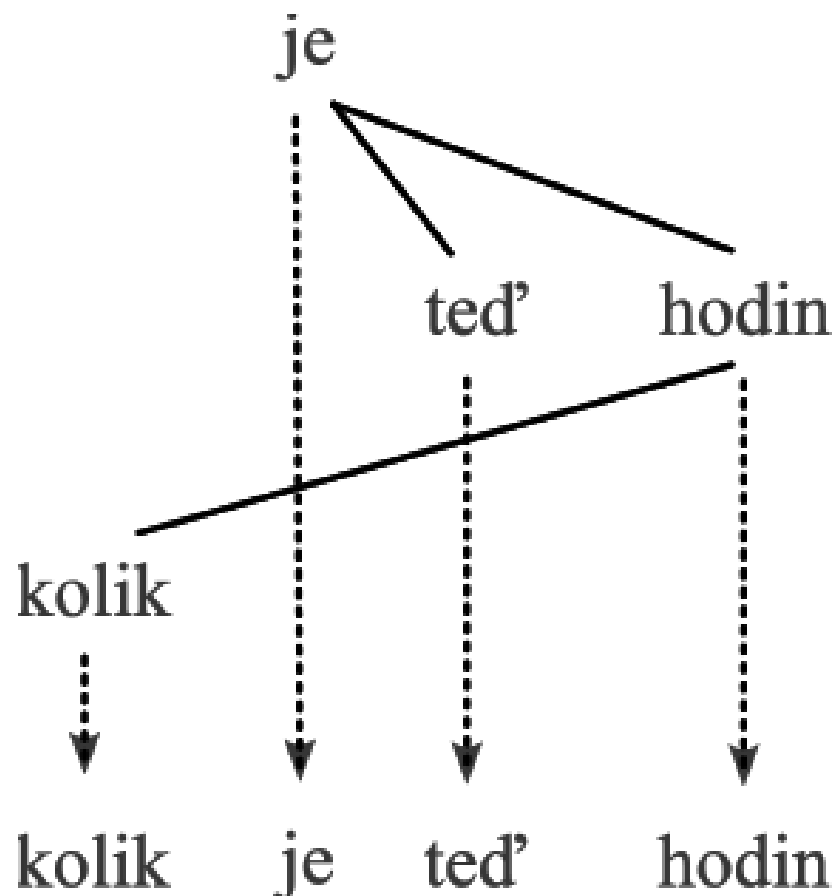
## Neprojektivní konstrukce



# Závislostní strom

## Neprojektivní konstrukce

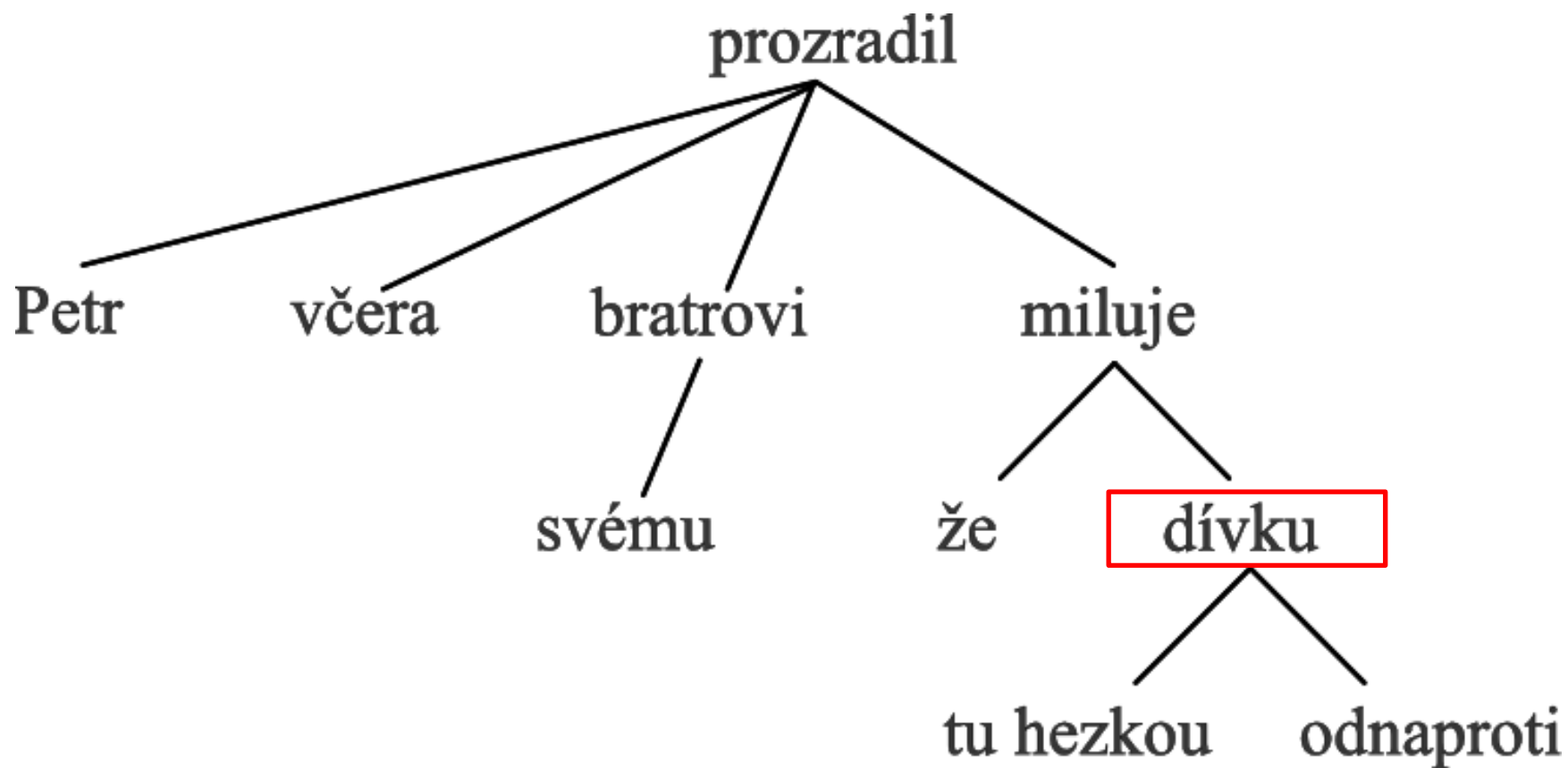
Množina závislostí jednoho uzlu netvoří souvislý úsek.





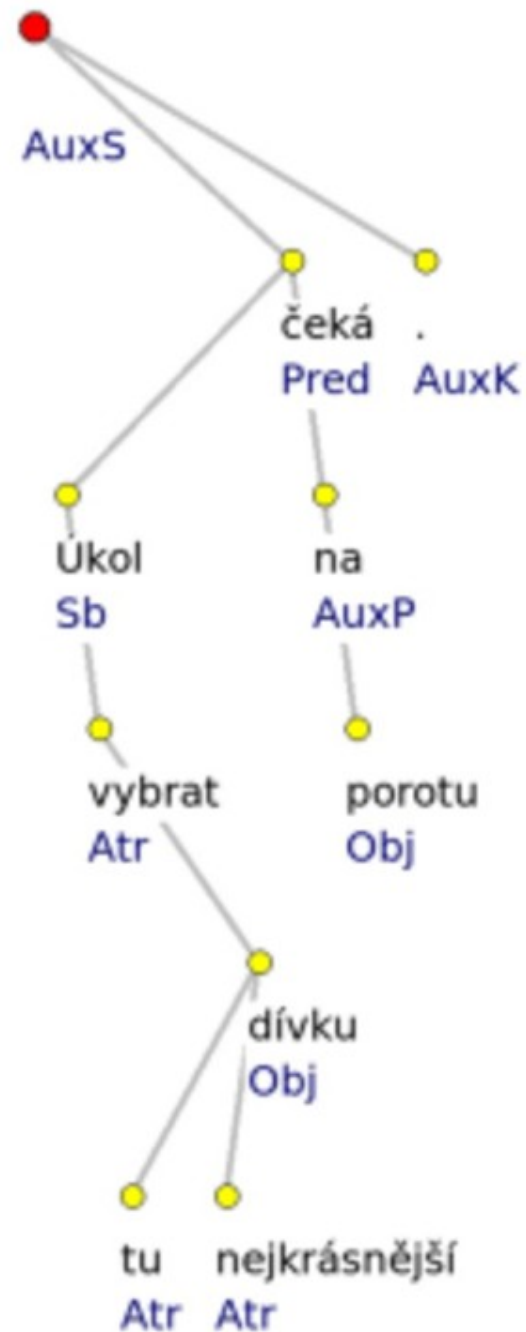
# Závislostní strom

---



# Pražský závislostní korpus

*Úkol vybrat tu nejkrásnější dívku  
čeká na porotu.*



# Závislostní strom

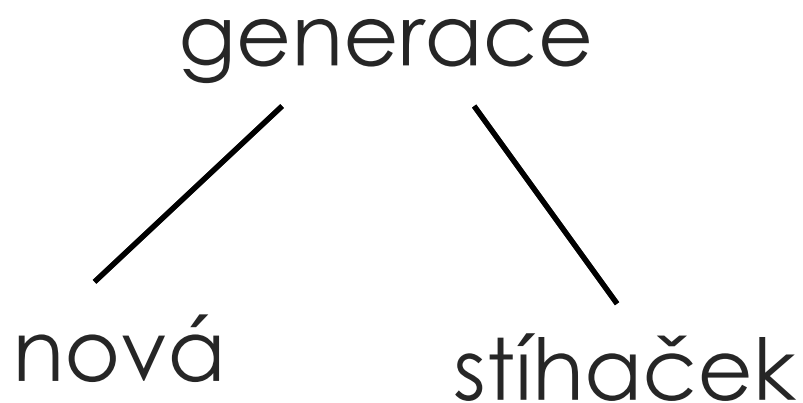
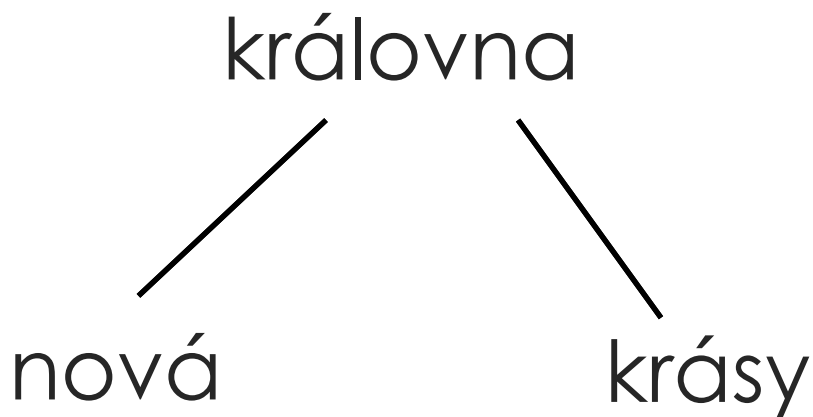
---

- Jak zachytit syntaktické vztahy, které nemají závislostní nebo binární povahu?
- koordinace, apozice, vsuvka, vokativ?, elipsa?, doplněk?

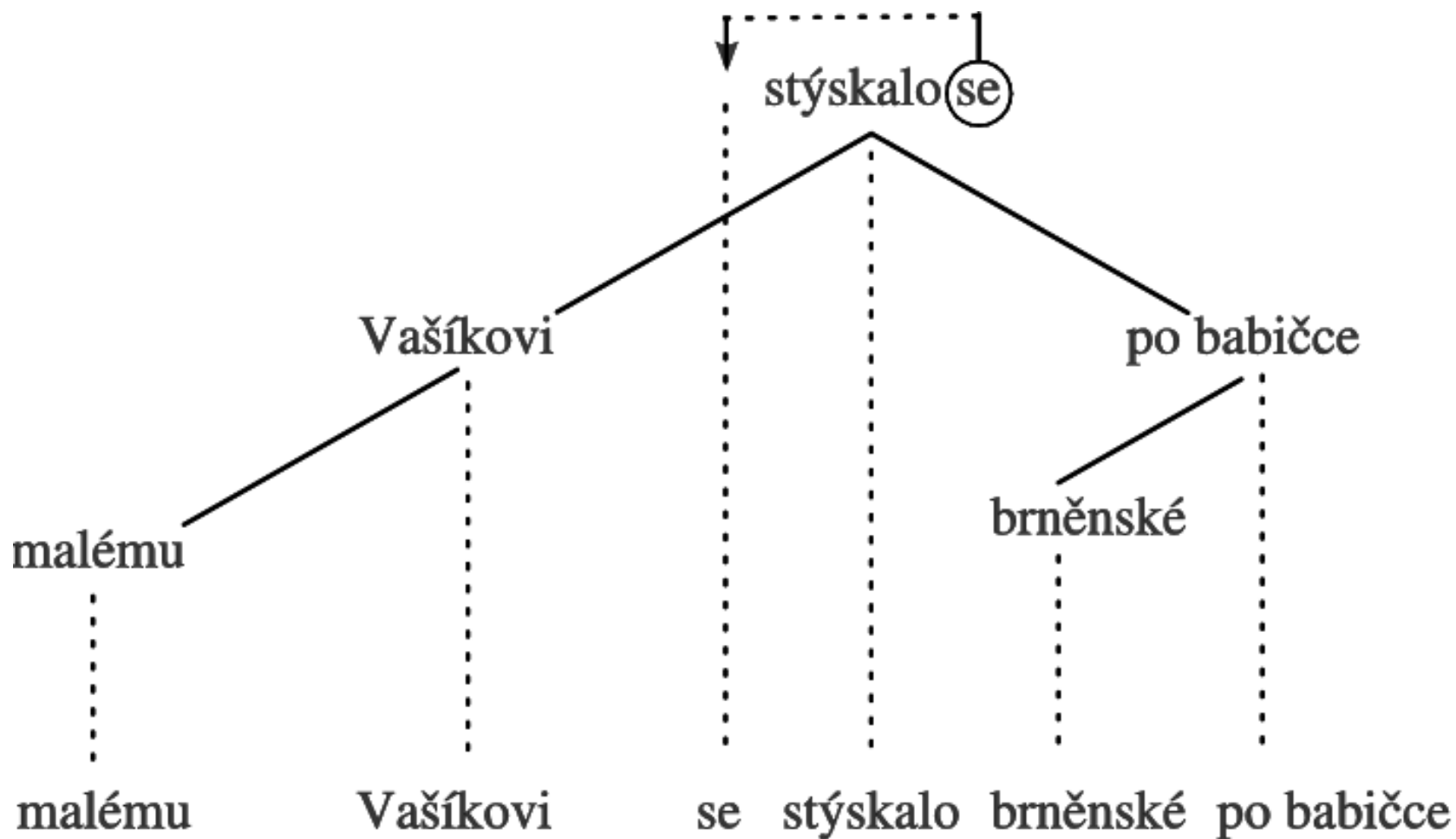
# Sémantická interpretace

---

- hůře se zachytává sémantická interpretace



# TEORETICKÉ PROBLÉMY



# TEORETICKÉ PROBLÉMY

---

Petr tvrdil, že / \*aby / \*zda

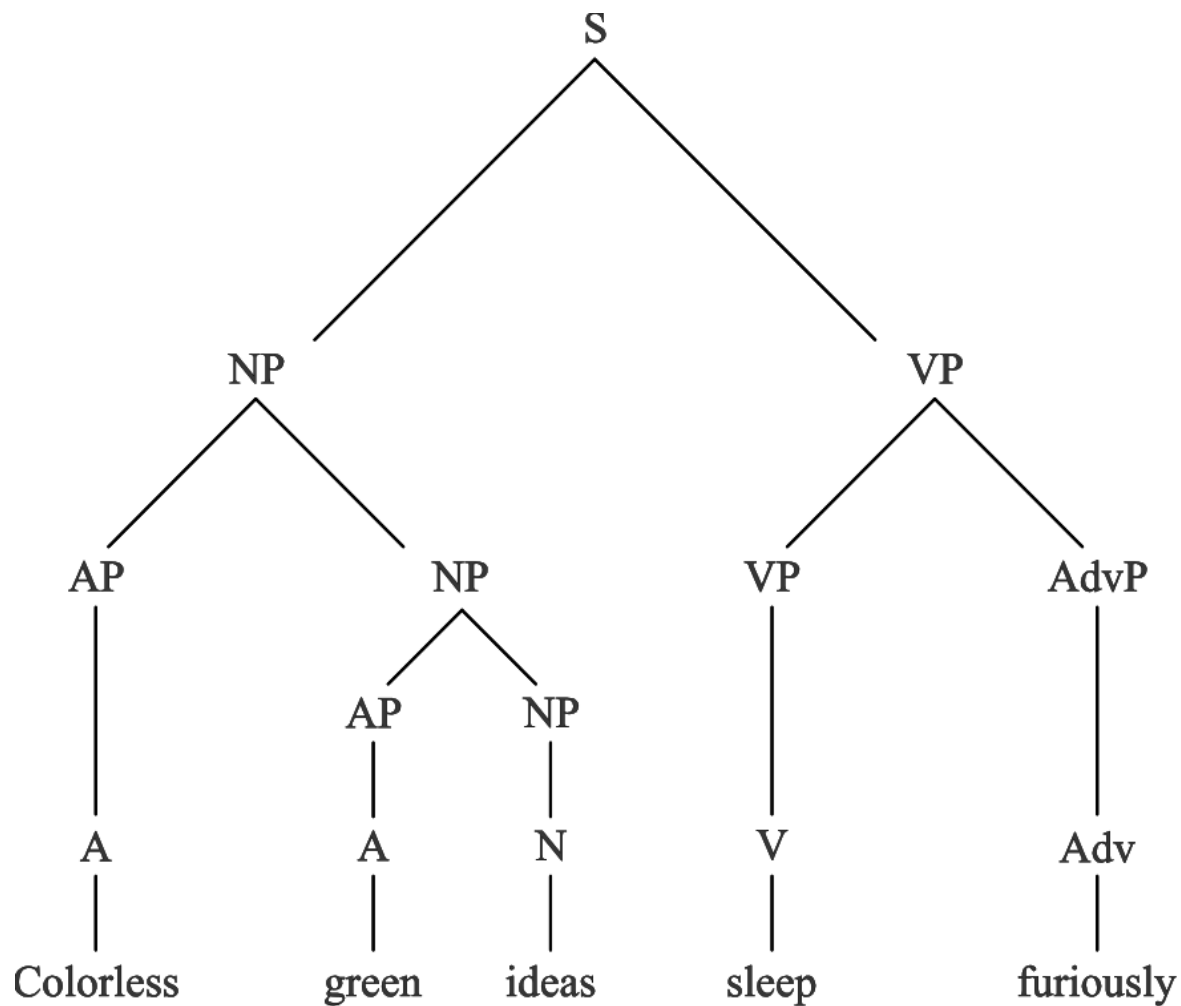
Petr zkoumal, \*že / \*aby / zda

Petr doporučil Pavlovi, \*že / aby / \*zda

Petr doporučil Pavlovi, že + modální sloveso –  
– infinitiv

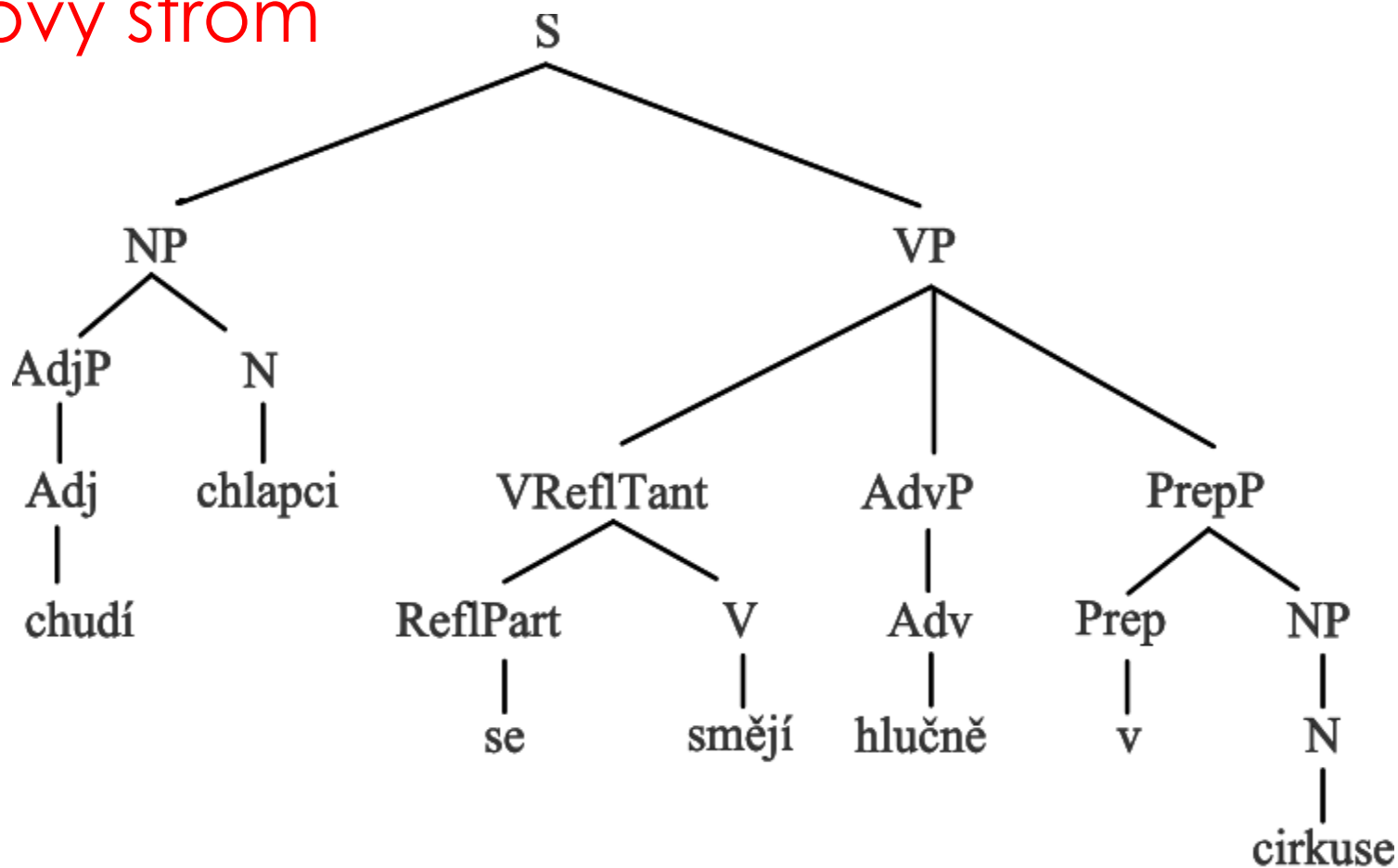
# BEZPROSTŘEDNĚSLOŽKOVÝ STROM

---



# Bezprostředněsložkový strom

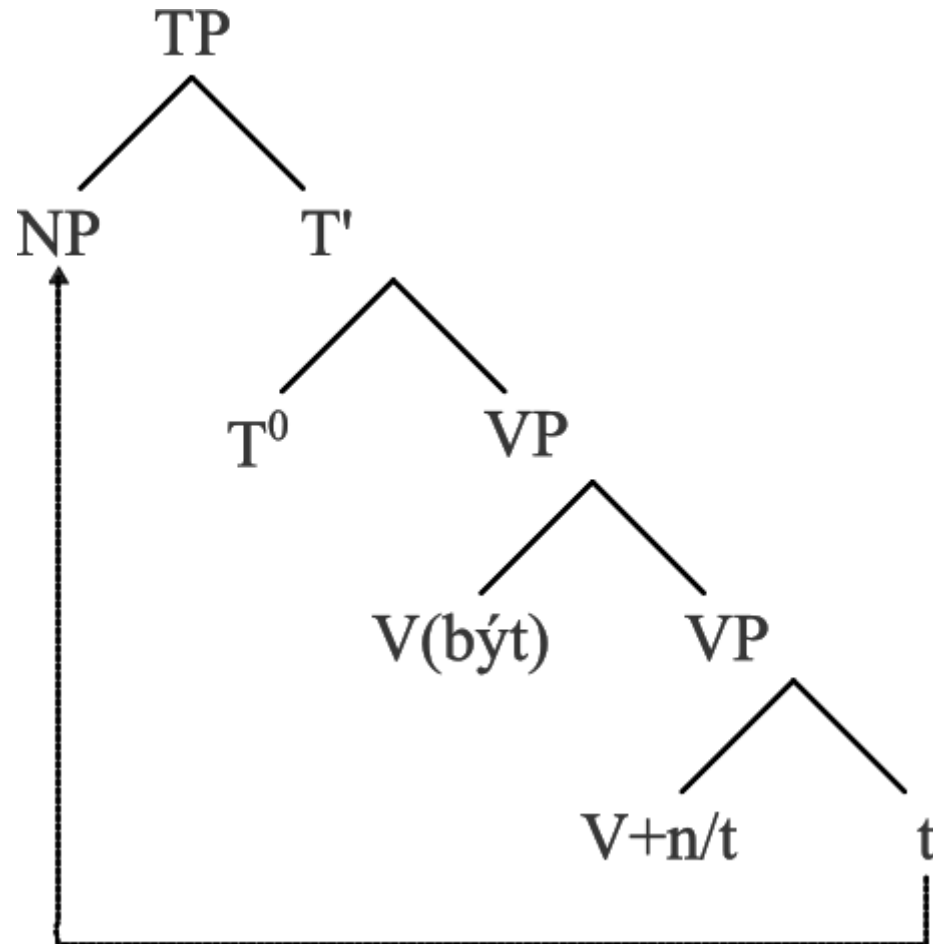
## Složkový strom





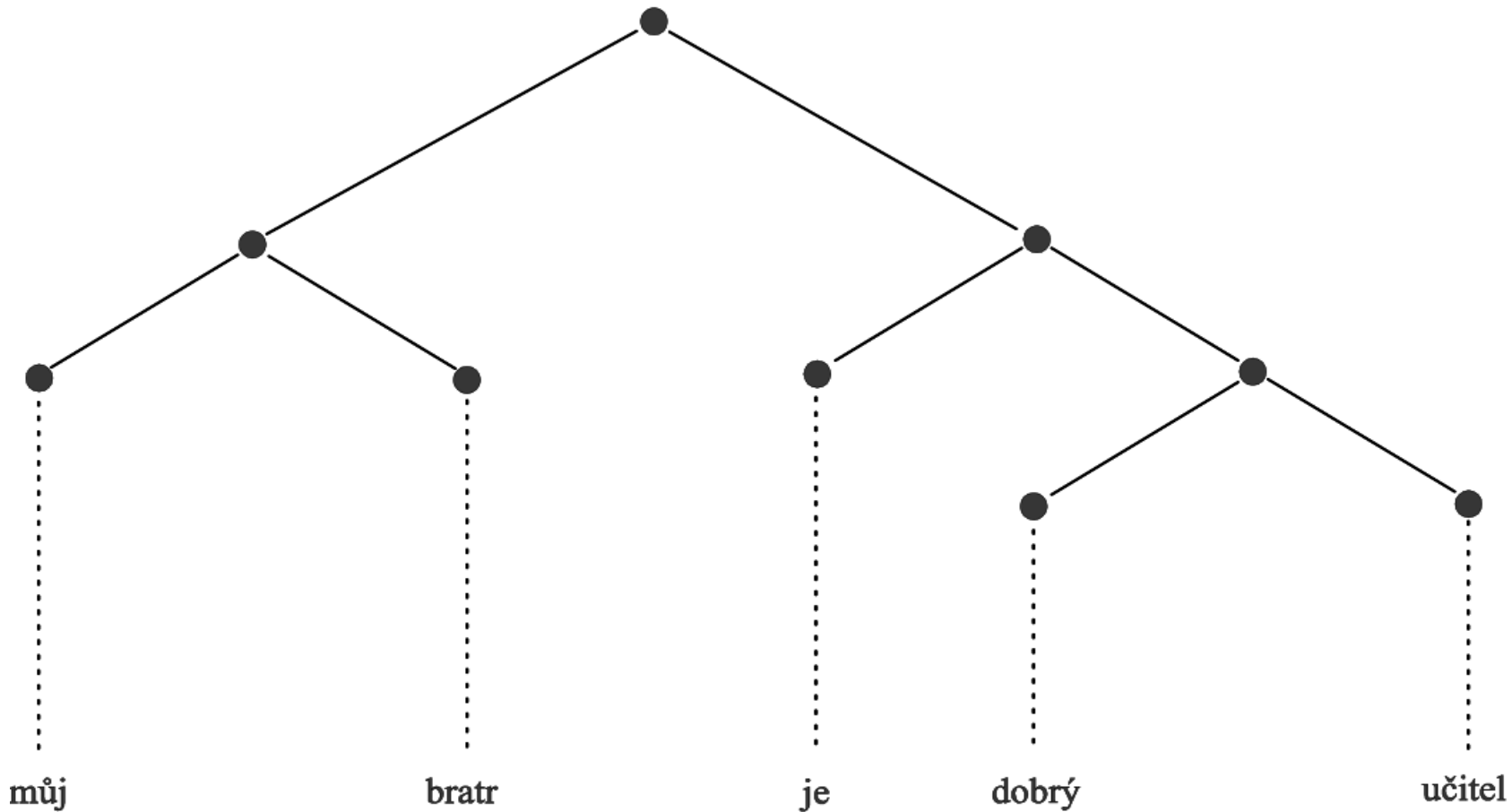
# BEZPROSTŘEDNĚSLOŽKOVÝ STROM

---

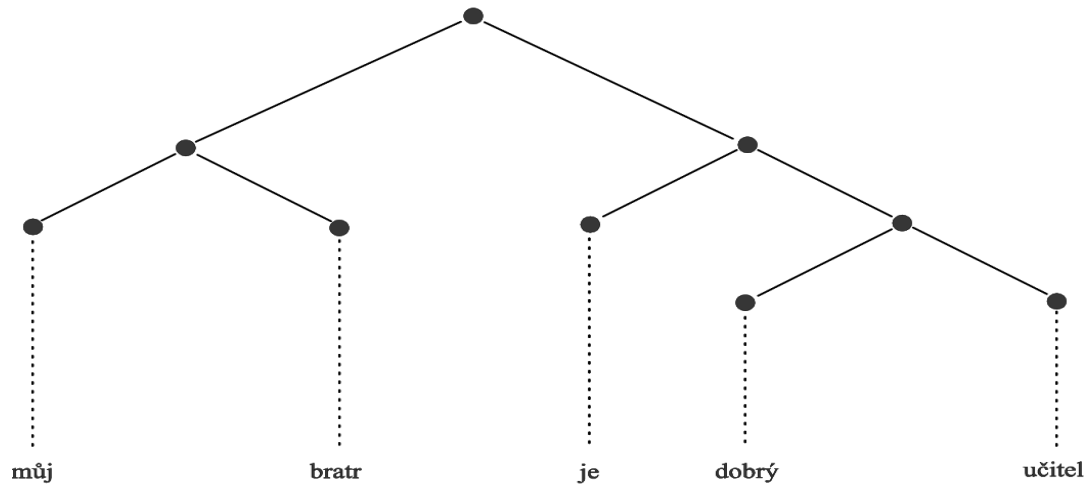
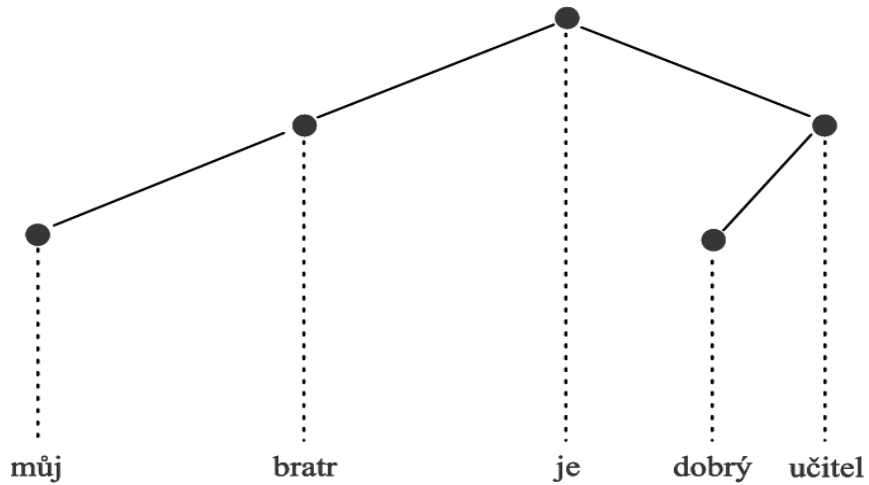


# BEZPROSTŘEDNĚSLOŽKOVÝ STROM

---



# BEZPROSTŘEDNĚSLOŽKOVÝ STROM



# Teoretická východiska

---

## Složkový přístup

- neprojektivita řešena teoreticky  
(např. transformace/pohyby)

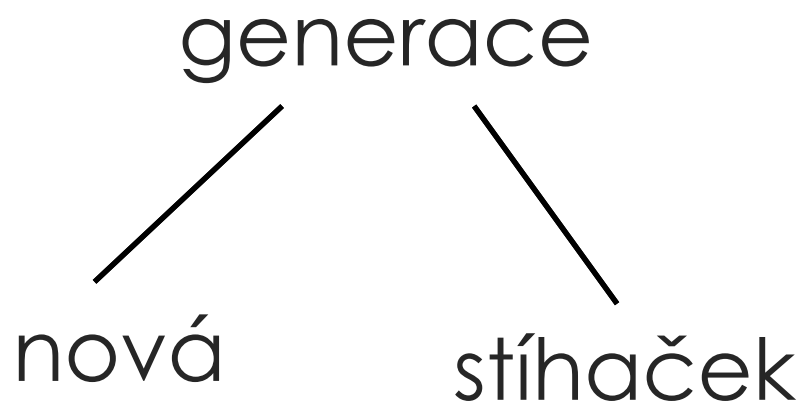
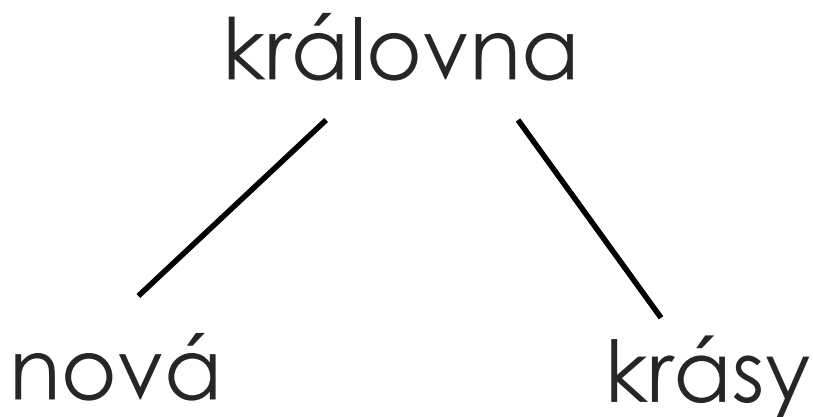
*Ke kolikátému tě předvolali výsledku?*

*Ke kolikátému výsledku tě předvolali?*

# Sémantická interpretace

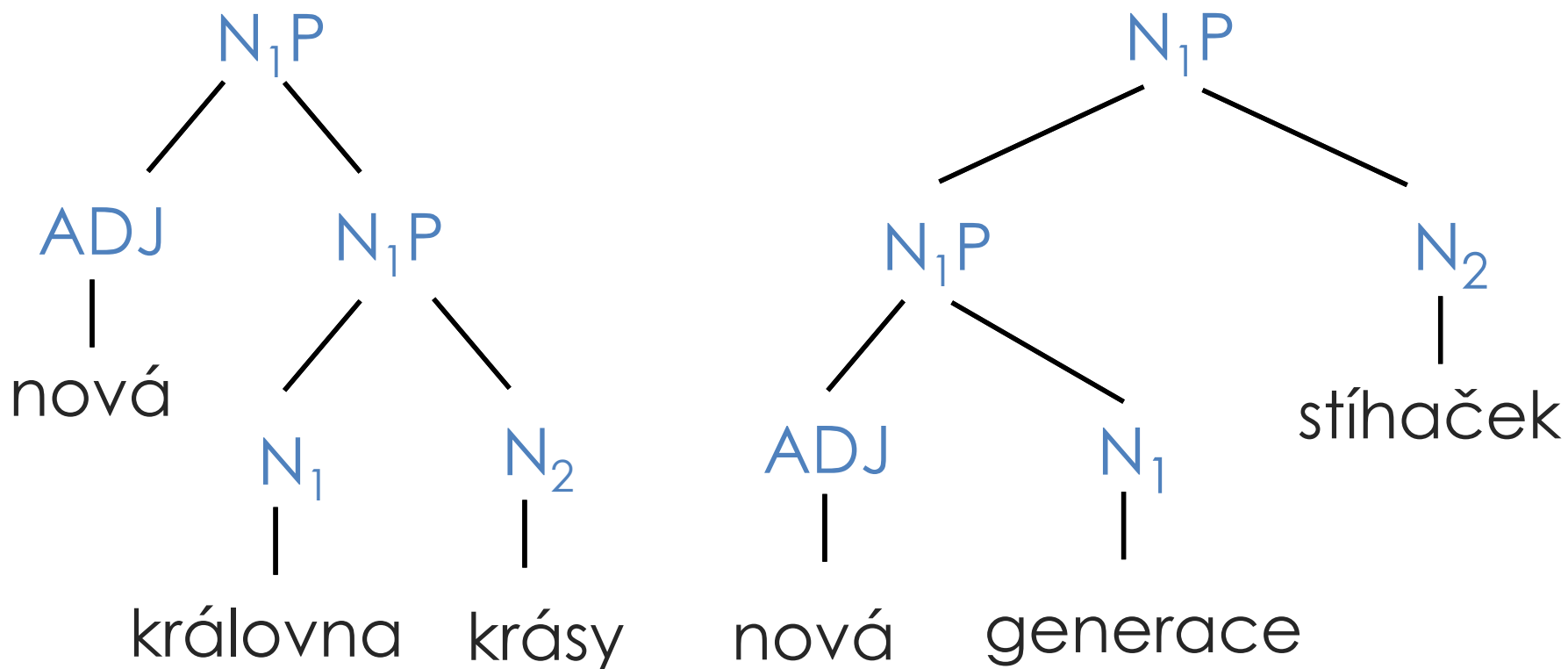
---

- hůře se zachytává sémantická interpretace

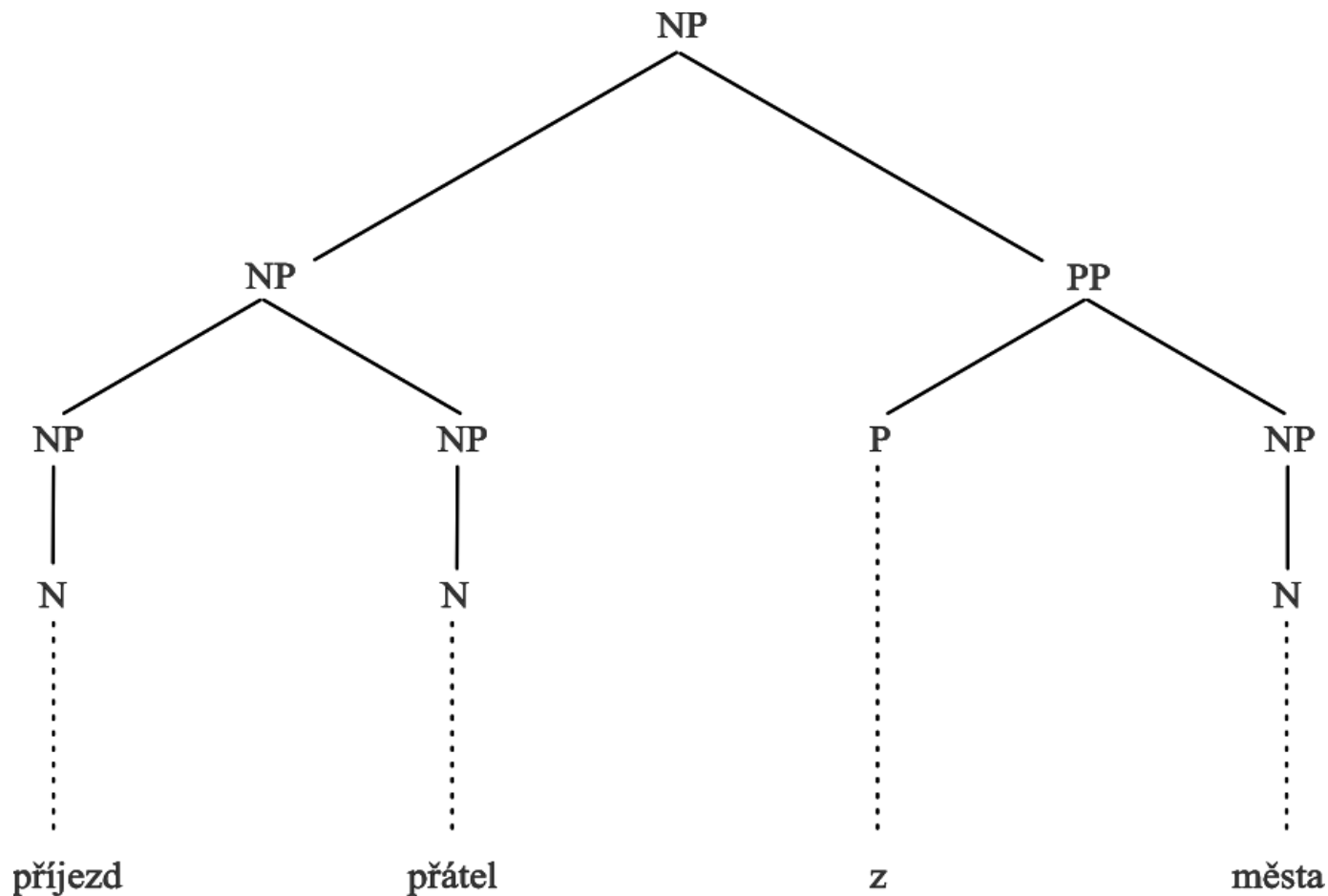


# Sémantická interpretace

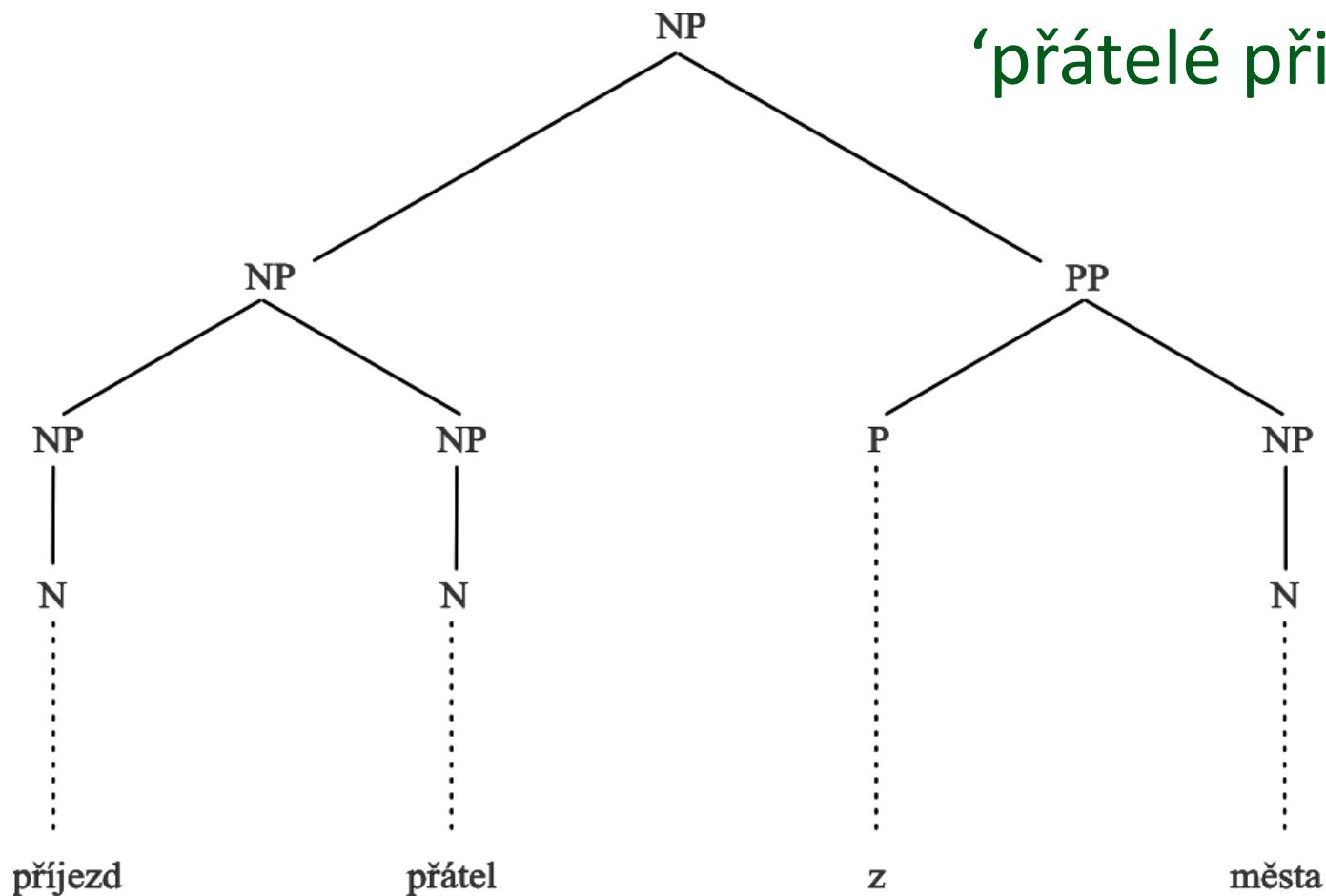
- lépe zachytává sémantickou interpretaci



# BEZPROSTŘEDNĚSLOŽKOVÝ STROM



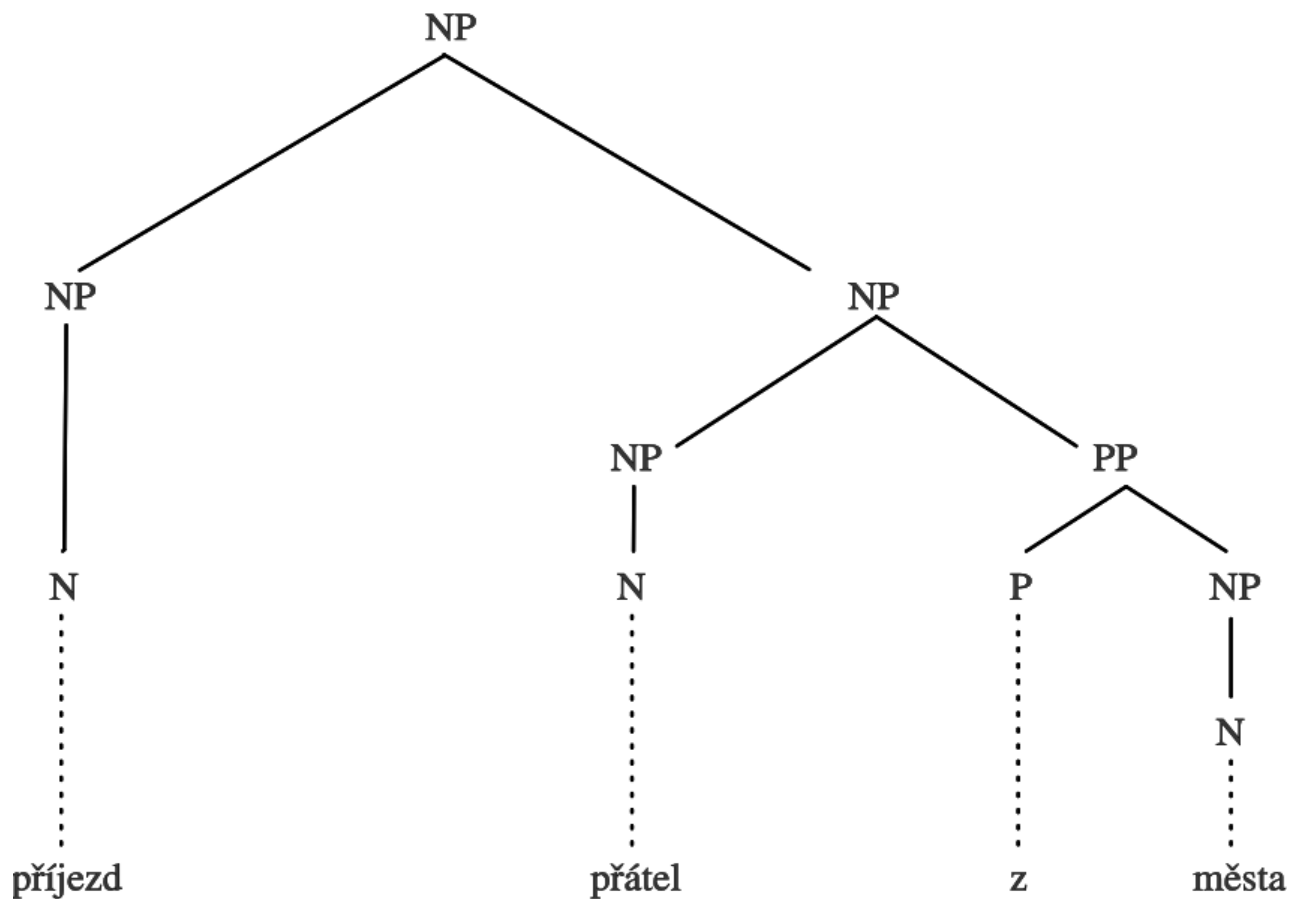
# BEZPROSTŘEDNĚSLOŽKOVÝ STROM



‘přátelé přijeli z města’

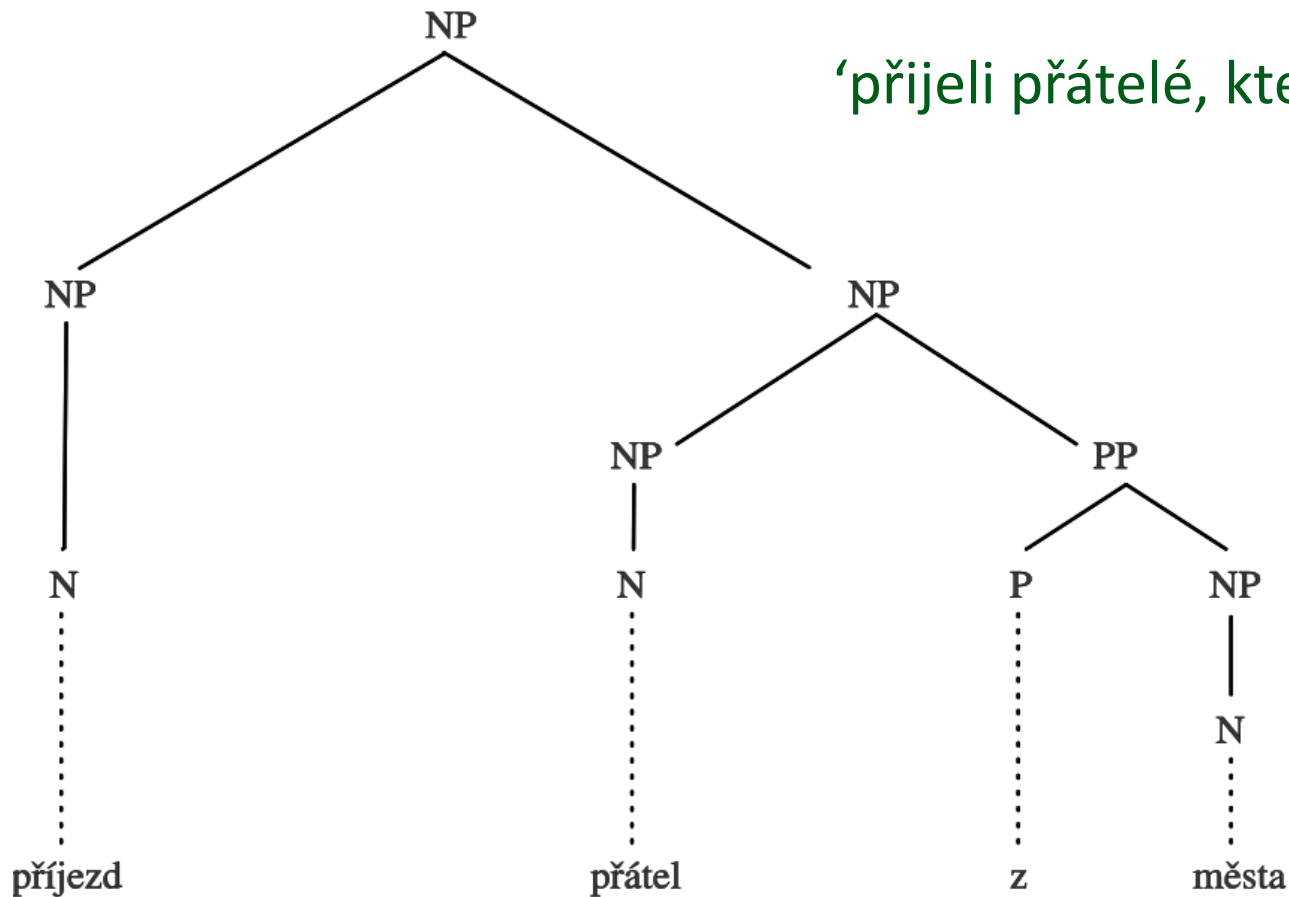


# BEZPROSTŘEDNĚSLOŽKOVÝ STROM

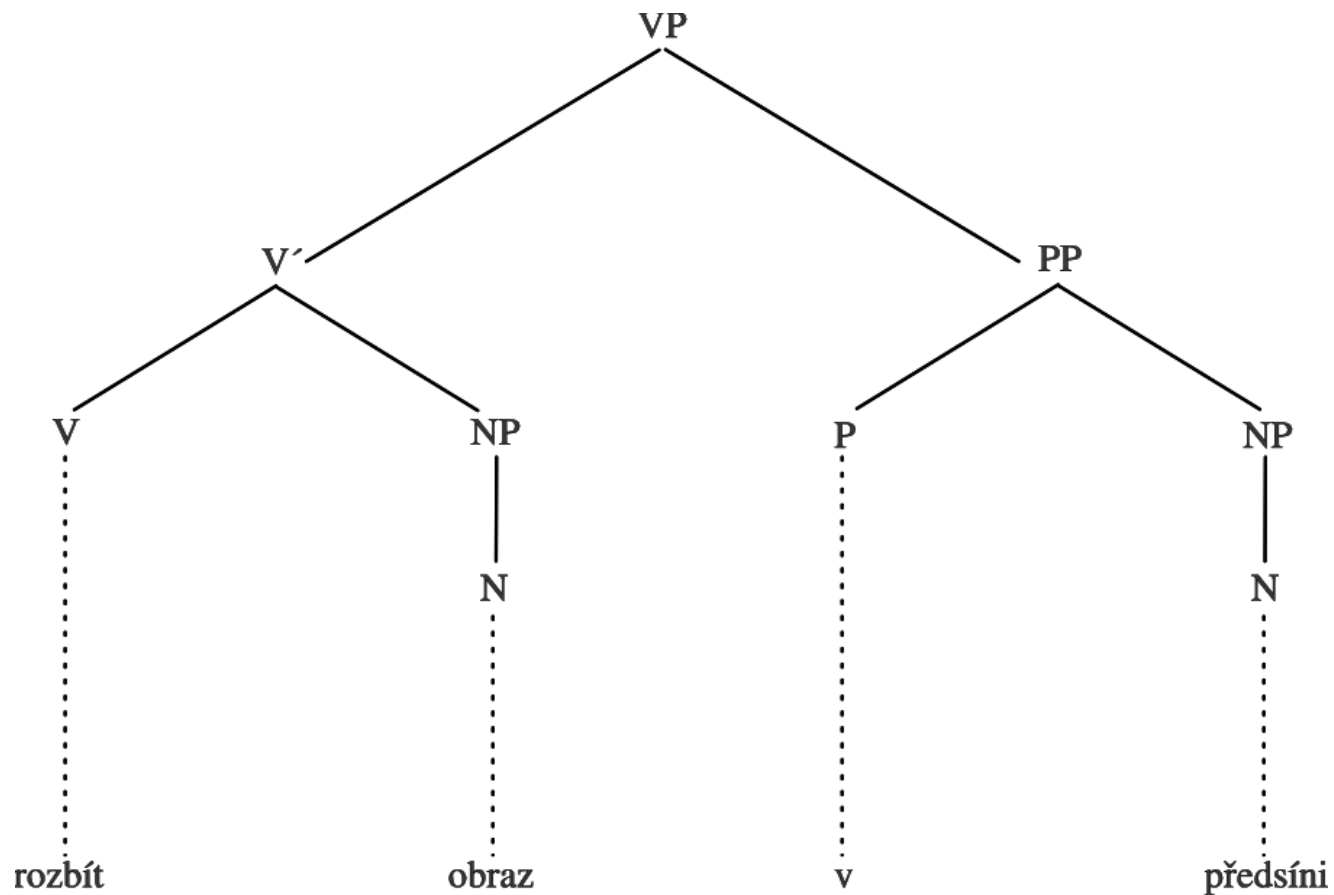


# BEZPROSTŘEDNĚSLOŽKOVÝ STROM

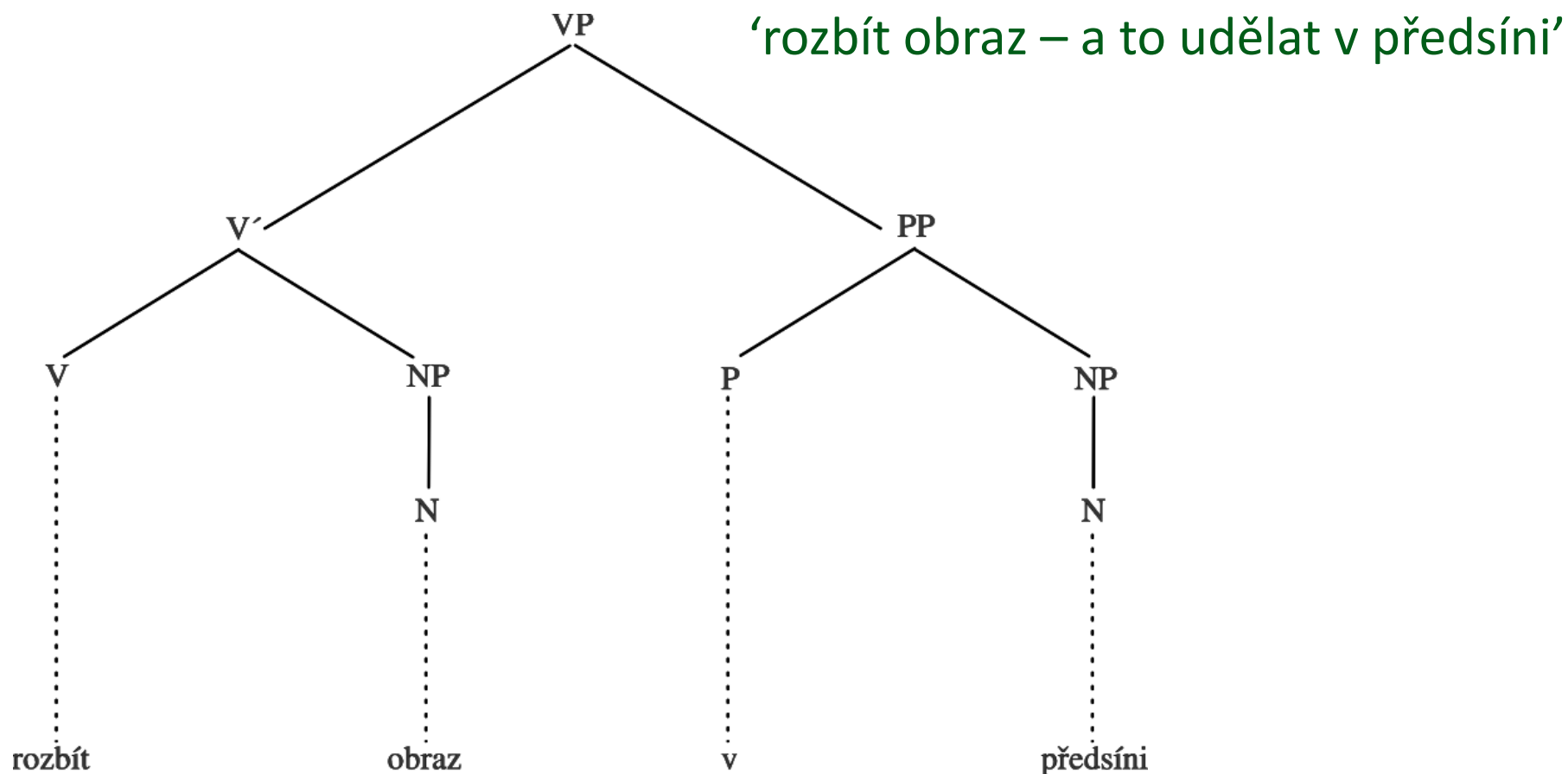
‘přijeli přátelé, kteří žijí ve městě’



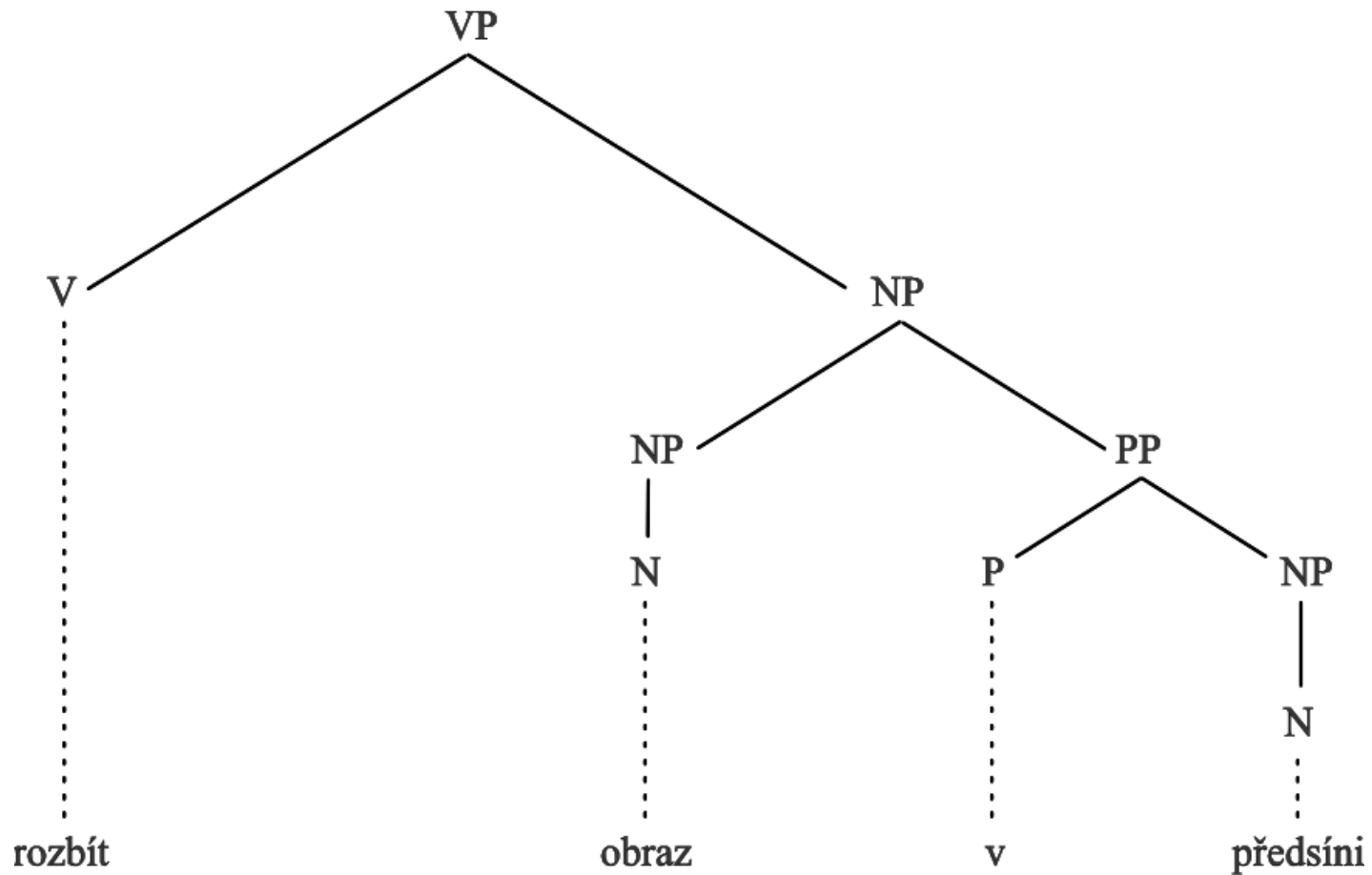
# BEZPROSTŘEDNĚSLOŽKOVÝ STROM



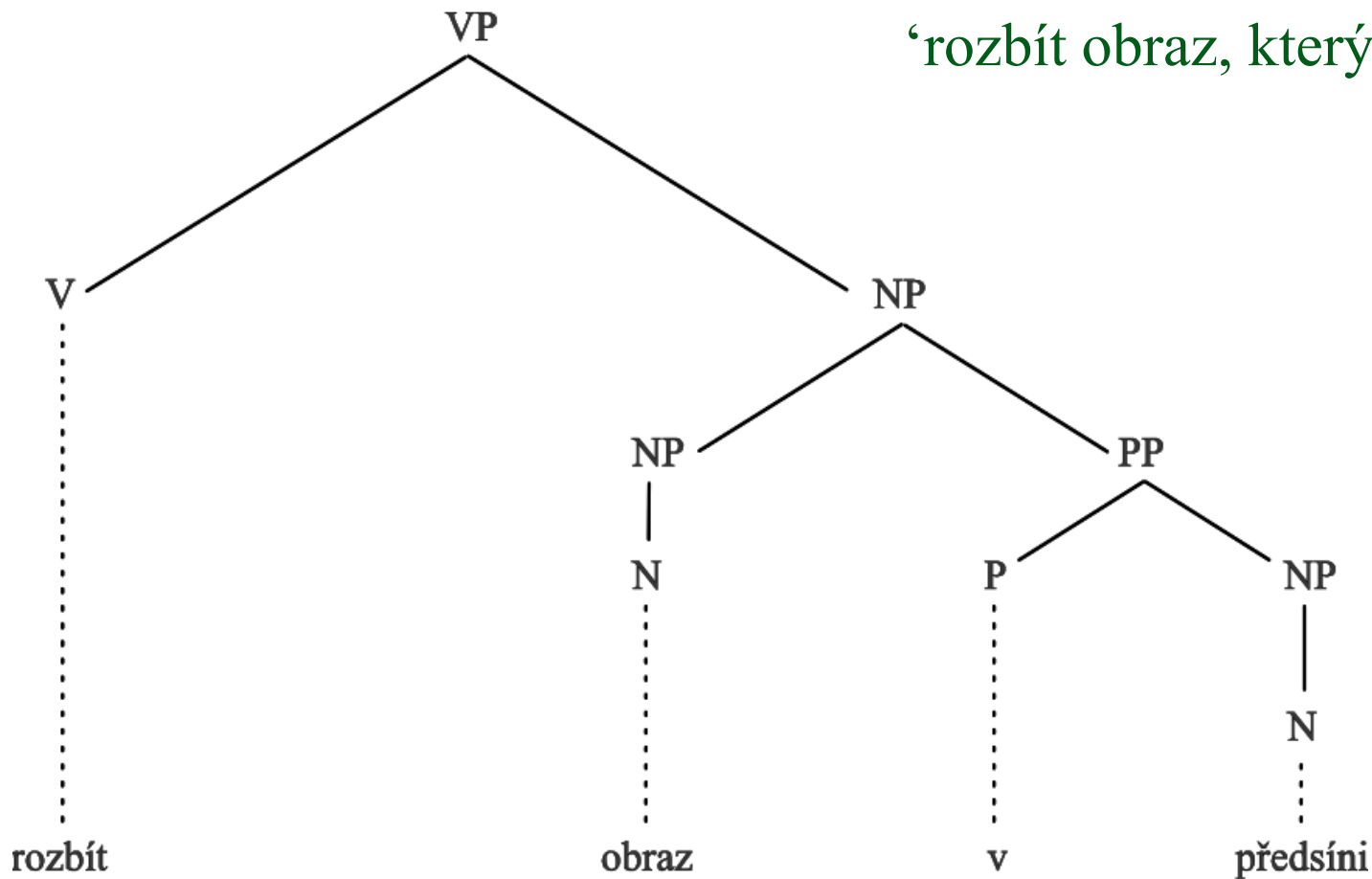
# BEZPROSTŘEDNĚSLOŽKOVÝ STROM



# BEZPROSTŘEDNĚSLOŽKOVÝ STROM



# BEZPROSTŘEDNĚSLOŽKOVÝ STROM



‘rozbít obraz, který se nacházel  
v předsíni’

- 
- v praxi se jen zřídka využívají čistě složkový nebo striktně závislostní přístup

- 
- v praxi se jen zřídka využívají čistě složkový nebo striktně závislostní přístup
  - hybridní stromy



- 
- v praxi se jen zřídka využívají čistě složkový nebo striktně závislostní přístup
  - hybridní stromy
  - parser, který na výstup vypíše jak závislostní, tak složkovou strukturu

- 
- v praxi se jen zřídka využívají čistě složkový nebo striktně závislostní přístup
  - hybridní stromy
  - parser, který na výstup vypíše jak závislostní, tak složkovou strukturu
  - pro čj chybí treebank se složkovými stromy

- 
- v praxi se jen zřídka využívají čistě složkový nebo striktně závislostní přístup
  - hybridní stromy
  - parser, který na výstup vypíše jak závislostní, tak složkovou strukturu
  - pro čj chybí treebank se složkovými stromy
  - Lancaster-Oslo/Bergen

# pravidla × statistika

---

# pravidla × statistika

---

## Statistická analýza

- trénovací (anotovaná) a testovací data
- jazyková univerzálnost
- větší přesnost při vyhodnocování podobnosti stromů
- lépe napodobí anotovaná data než člověk
- časově i finančně náročné

# pravidla × statistika

---

## Analýza založená na pravidlech

- nezávislost na anotovaných datech
- lze snadno upravit podle zadaného úkolu

Je možné kompletně popsat přirozený jazyk pravidly?

# Postup parsingu

---

- „zdola nahoru“ (bottom up)
- „shora dolů“ (top down)
- smíšené (left-corner, head-corner parsers)

„zdola nahoru“ (bottom up)

---

John hits the ball.



# „zdola nahoru“ (bottom up)

---

1. věta se skládá ze jmenné fráze následované slovesnou frází  
( $S \rightarrow NP VP$ ),
2. vlastní jméno samo o sobě je jmennou frází ( $NP \rightarrow \text{„John“}$ ),
3. jmenná fráze může být také determinátor následovaný  
podstatným jménem ( $NP \rightarrow \text{Det N}$ ),
4. slovesná fráze může být sloveso následované jmennou frází  
( $VP \rightarrow V NP$ ),
5. „ball“ je podstatné jméno, „the“ je determinátor a „hit“ je  
sloveso ( $N \rightarrow \text{„ball“}$ ,  $\text{Det} \rightarrow \text{„the“}$ ,  $V \rightarrow \text{„hit“}$ ).

# „zdola nahoru“ (bottom up)

Použité pravidlo	Zpracovávaná část textu	Současná věta
-		John hit the ball
2	„John“ je jmenná fráze sama o sobě	NP hit the ball
5	zbývající slova nahradíme jejich syntaktickými kategoriemi	NP V Det N
3	„the ball“ je jmenná fráze	NP V NP
4	„hit the ball“ je slovesná fráze	NP VP
1	„John hit the ball“ je celá věta	S

# Literatura

---

Nový encyklopedický slovník češtiny online:

<https://www.czechency.org/>

hesla: Závislostní strom, Projektivnost,

Dominance, Bezprostředněsložkový strom,

Složka, Parsing