



# Prague Dependency Treebank: Annotation of Surface Syntax (Part II.)

Markéta Lopatková

Institute of Formal and Applied Linguistics, MFF UK

[lopatkova@ufal.mff.cuni.cz](mailto:lopatkova@ufal.mff.cuni.cz)

---



---

# PDT: a-layer

## Goal:

- to describe the structure of the sentence and
- to denote the type of relations between "words"

documentation: <http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/a-layer/html/index.html>

## analytical functions (lecture 5):

- predicate ... **Pred, Pnom, AuxV** (, Obj)
  - subject ... **Sb**
  - attribute ... **Atr**
  - object ... **Obj**
  - adverbial ... **Adv**
  - complement ... **Atv, AtvV**
  - auxiliary sentence member ... **AuxC, AuxP, AuxZ, AuxO, AuxT, AuxR, AuxY**
- } combined functions  
... **AtrAtr, AtrObj, ObjAtr, AtrAdv, AdvAtr**

---

# Principles of Annotation: Non-dependency

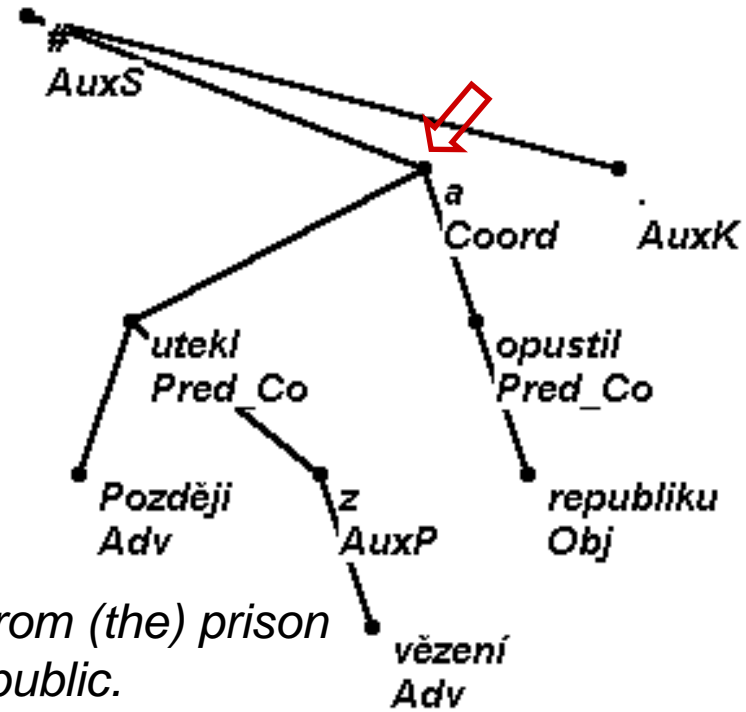
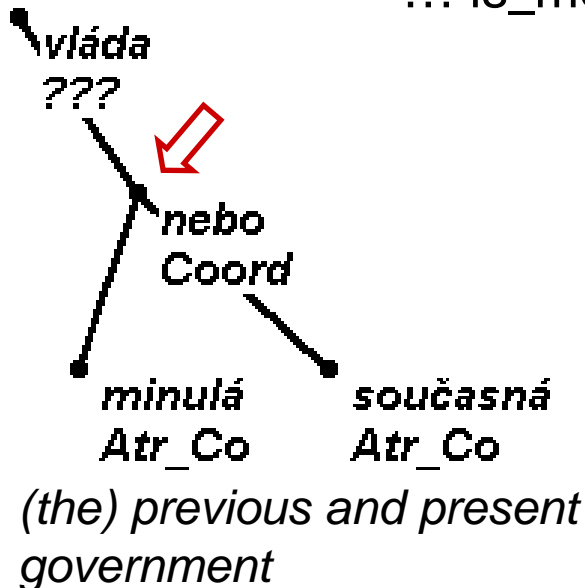


- coordination
- apposition
- parenthesis

# Non-dependency: Coordination



- "multiplication" of a single syntactic position (sentence member / sentence)
- **Coord** ... 'connecting node'
  - primarily:
    - coordinating conjunctions (*a, i, ale, nebo, neboť, proto* ...)
    - connecting expressions (*a tak dále, et cetera, ...*)
    - comma
- visualization ... suffix\_Co vs.  
... is\_member in the data



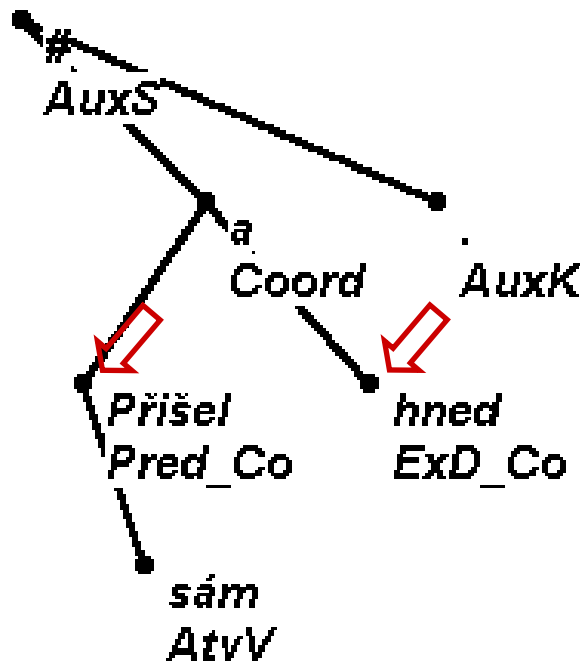
---

# Non-dependency: Coordination (cont.)



all coordinated members ... the **same** analytical function (if not Coord)

- if not possible ... ExD\_Co

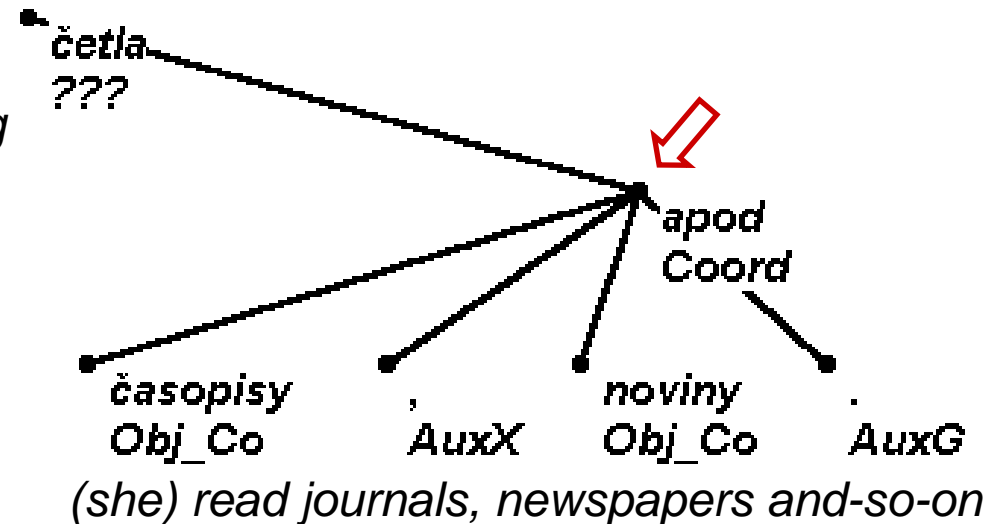
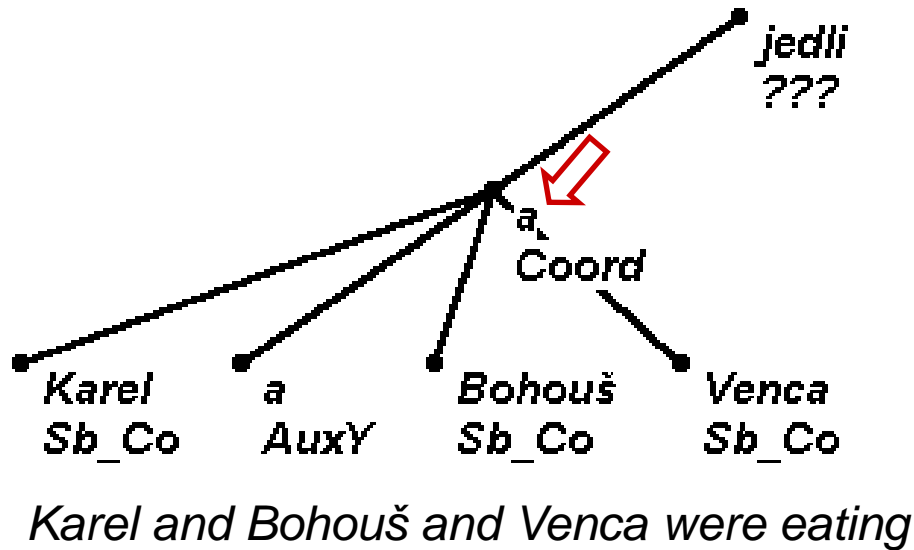


*(He) came alone and immediately*

# Non-dependency: Coordination (cont.)

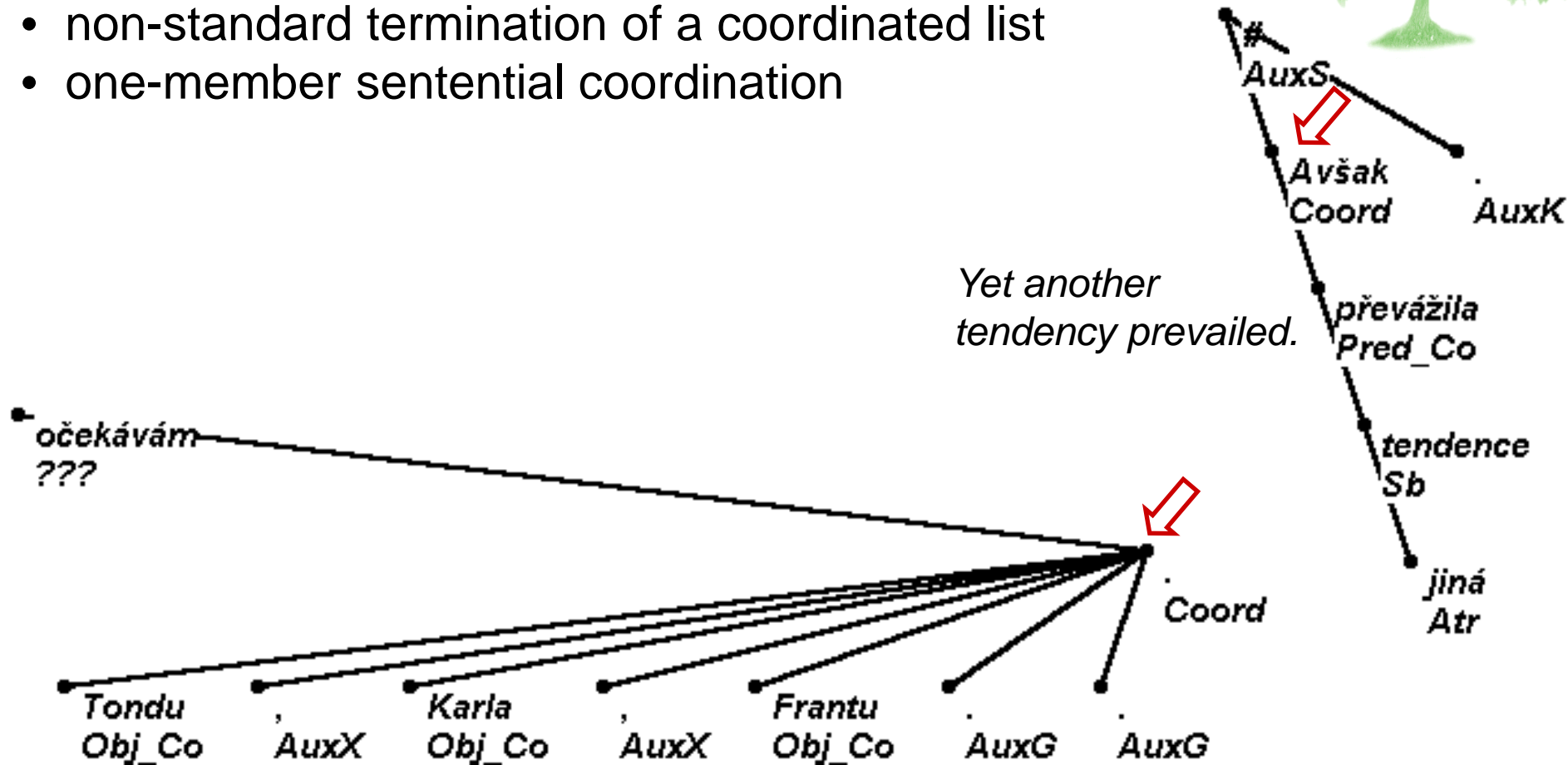


Which node is a 'connecting' node?



# Non-dependency: Coordination (cont.)

- non-standard termination of a coordinated list
- one-member sentential coordination

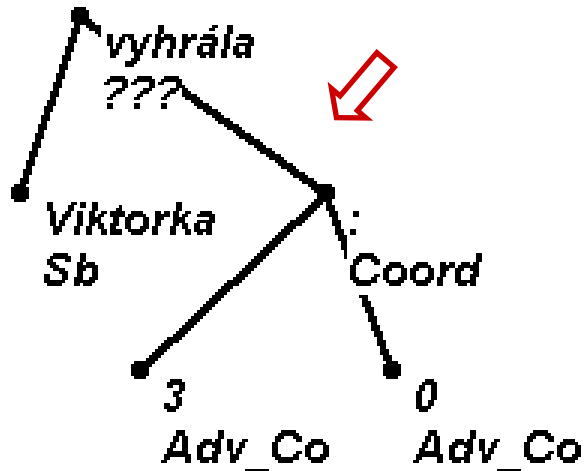


(I) expect Tony, Charles, Frank ...

# Non-dependency: Coordination (cont.)



Specific constructions:



*Viktorka won 3:0*



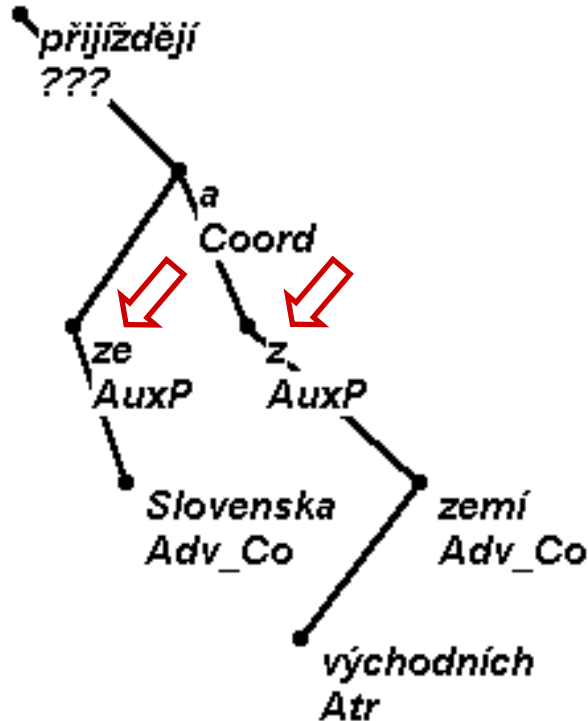
*Czech – German relations*



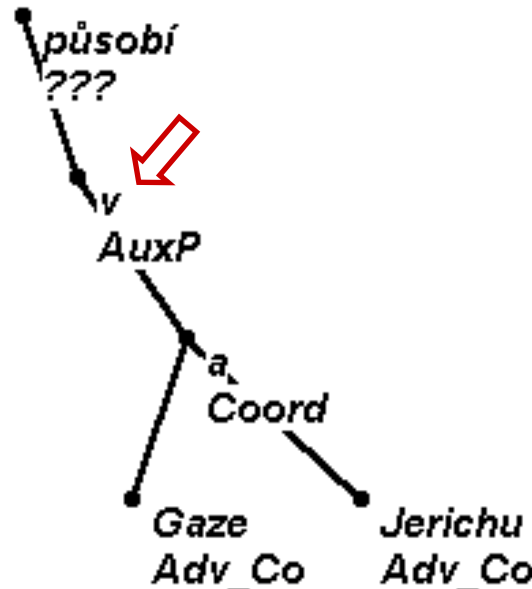
# Non-dependency: Coordination (cont.)



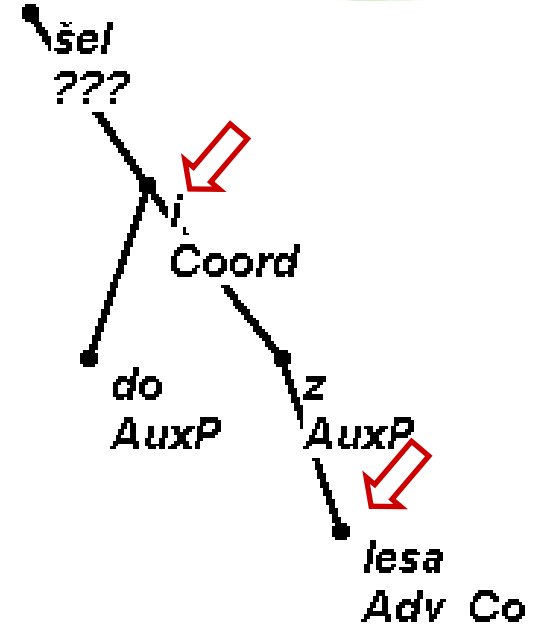
coordinated members with AuxP (or AuxC)



*(they) arrive from Slovakia and from eastern countries*



*(they) operates in Gaza and Jericho*



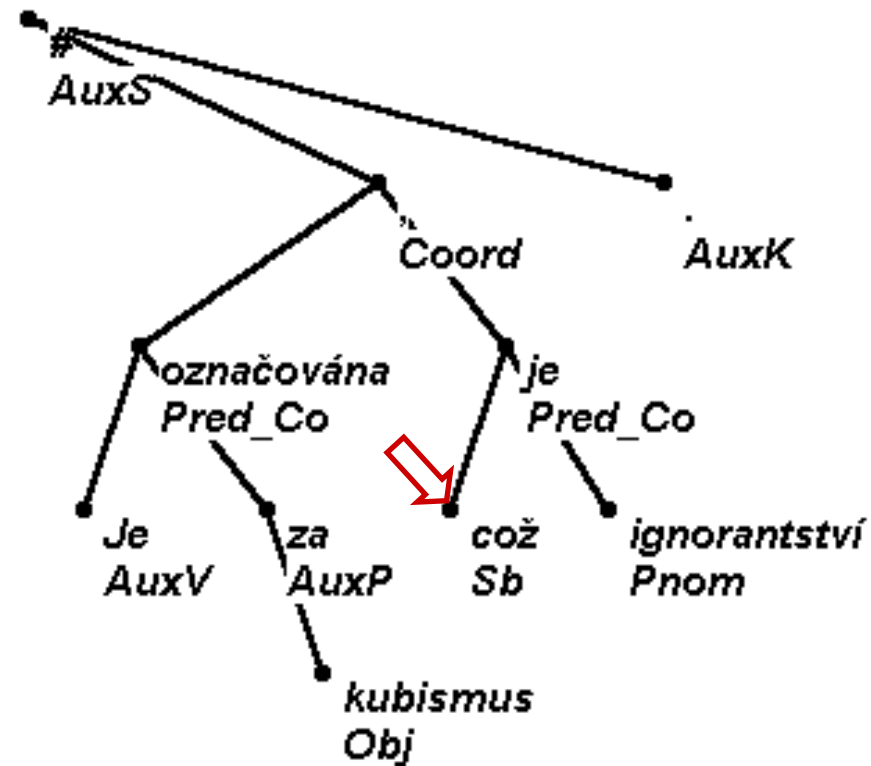
*(he) walked to and from (the) forest*

# Non-dependency: Coordination



coordination vs. apposition

- *což* `which' (incl. inflected forms and forms with prepositions, as e.g. *příčemž*)



(It) is denoted as (a) cubism which is (an) ignorance.

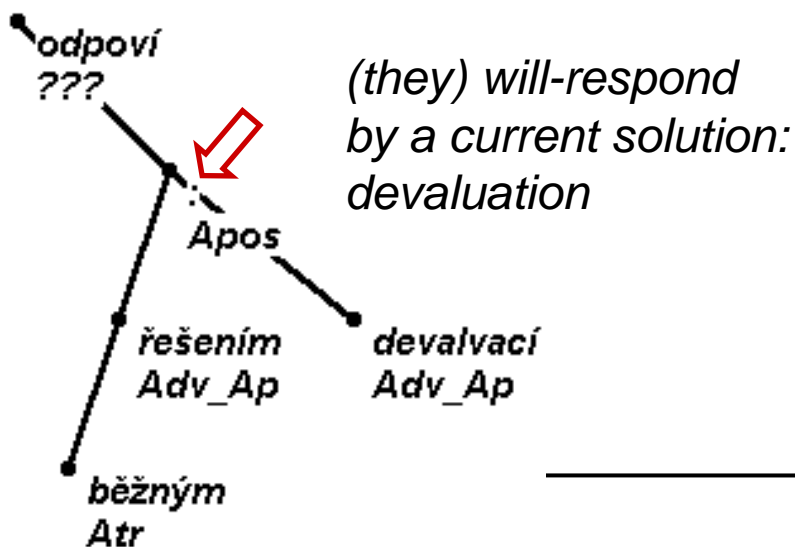


# Non-dependency: Apposition

analytical function: **Apos**

- "multiplication" of a single syntactic position (the same referent)
- members of an Apposition ... interchangeable
  - the **same** function
  - formal condition: typically the same case
- typical connecting expressions

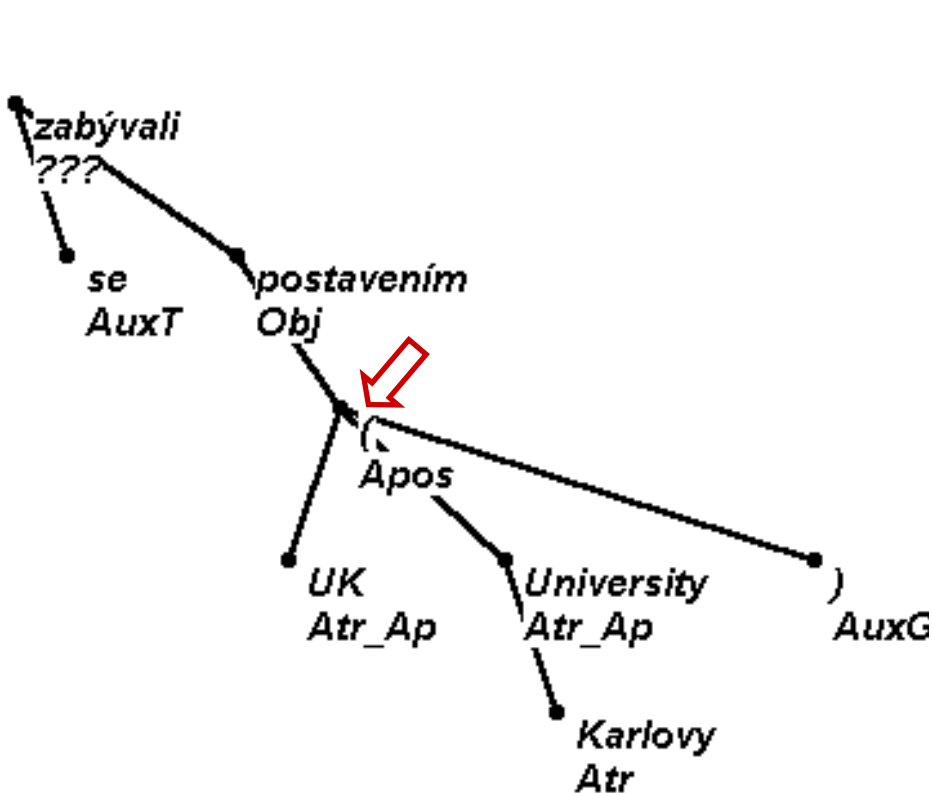
e.g. *jako* [as], *t.j.* [i.e.], *(jako) např.* [e.g.], *ergo*, *tedy* [thus], ...
- visualization suffix **\_Ap** vs. **is\_member** in the data



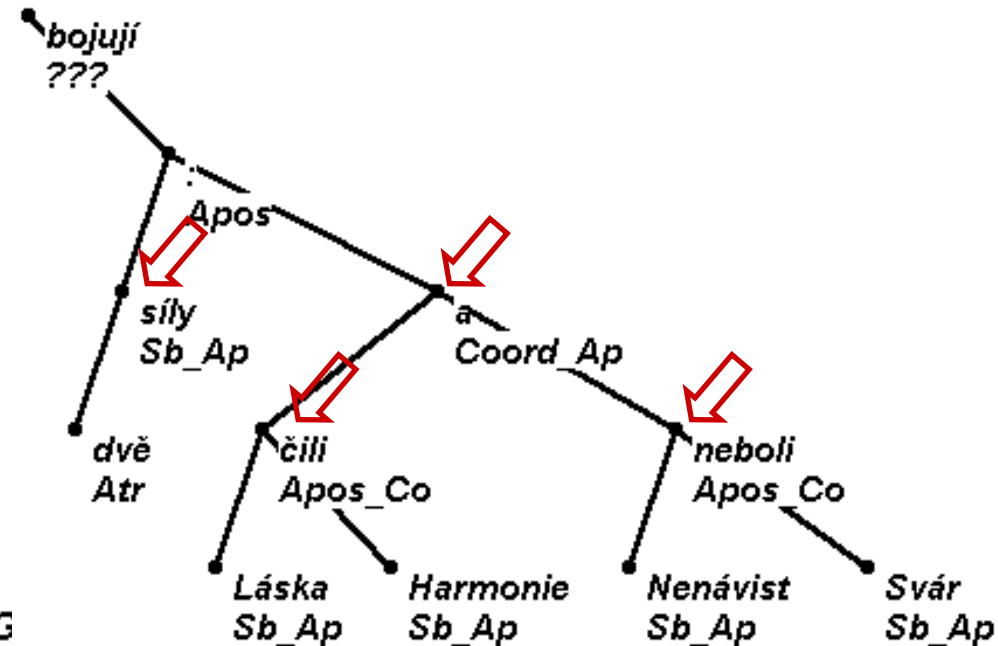
# Non-dependency: Apposition (cont.)



necessarily formally separated, e.g. by comma, brackets, dash ...



*(they) dealt with the position of UK  
(Charles University)*



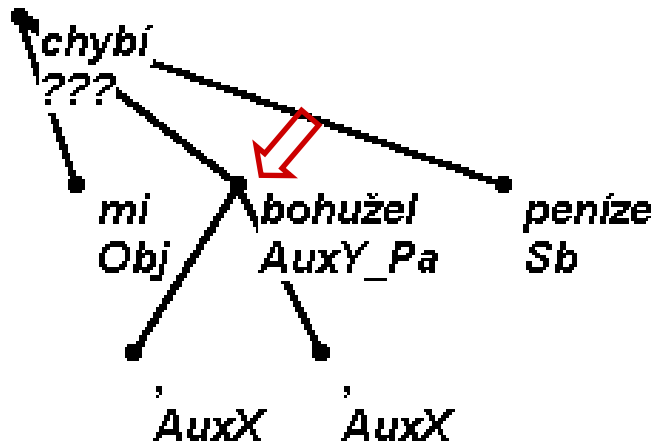
*Two powers fight: Love or Harmony  
and Hatred or Conflict*

# Non-dependency: Parenthesis

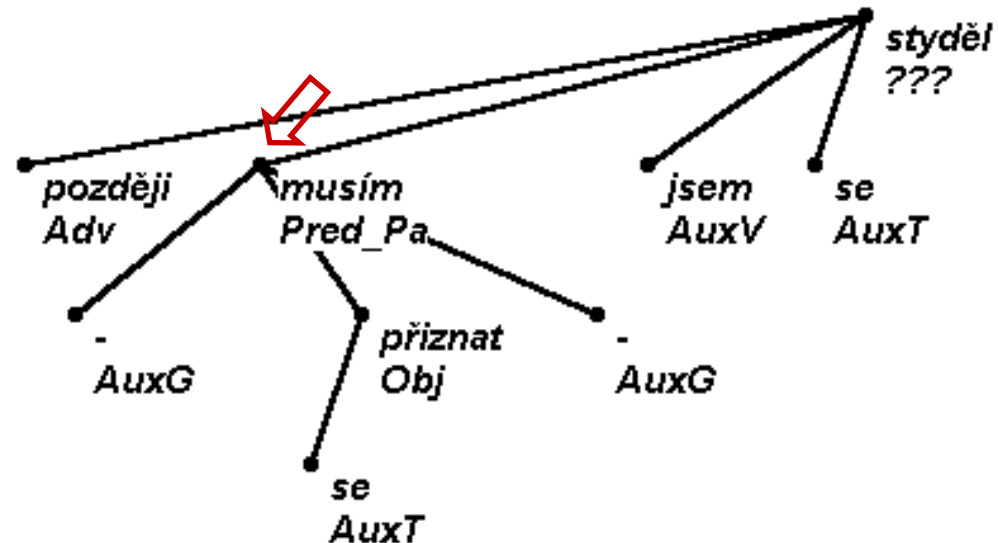


an additional adjunction to the statement in the sentence

- formally marked by some graphic sign(s)
  - bracket, dashes, commas, ...
- visualization: suffix `_Pa` vs. `is_parenthesis_root` in the data



*I am , unfortunately , short of money*



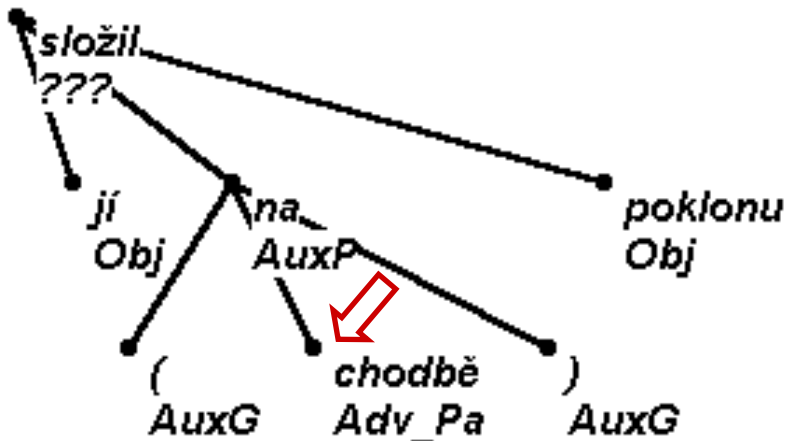
*later - I have to admit – I was ashamed*

# Non-dependency: Parenthesis (cont.)

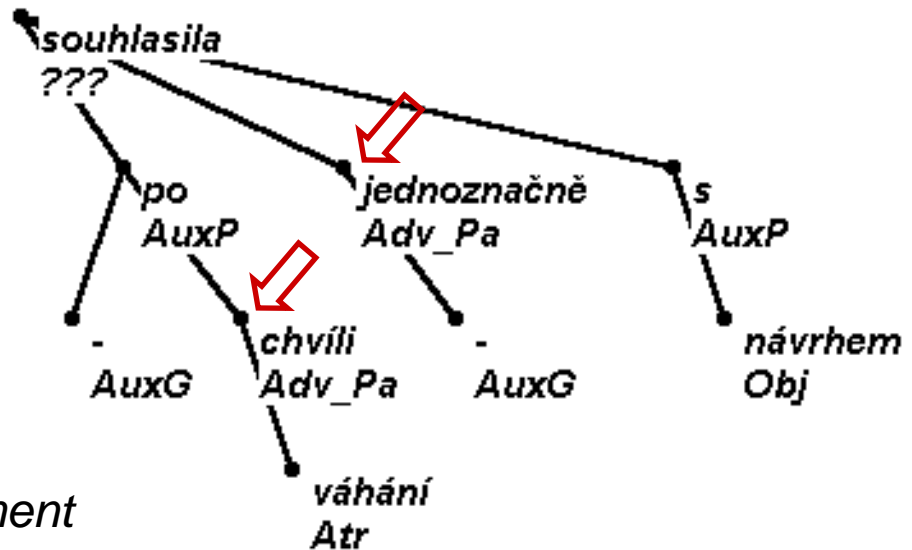


Note:

- AuxP as a 'head' of parenthesis
- parenthesis consisting of several subtrees



*he paid her (on a corridor) a compliment*



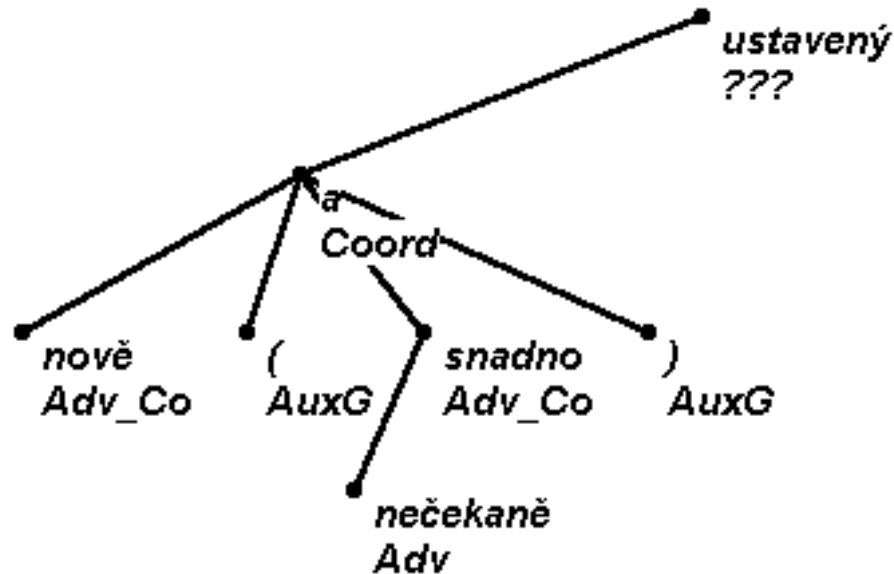
*she agreed – after a moment of  
hesitation expressly – with the proposal*

---

# Non-dependency: Parenthesis (cont.)



'strange' annotation



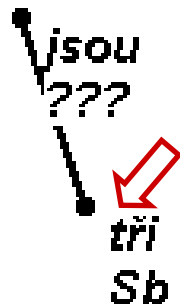
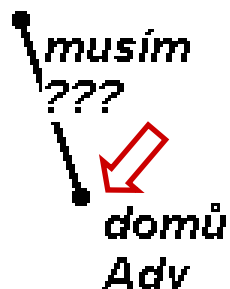
*recently (and unexpectedly easily) appointed*



# Principles of Annotation: Ellipses

**ellipsis** ~ a deletion of an expression that is expected in the given syntactic structure

- **ExD** ... all nodes that would depend on the deleted node
- exception: prepositions (AuxP) and subord. conjunctions (AuxC) connecting nodes for coordination/apposition (Coord, Apos)
- lexicalized ellipses



(I) have to (go) home.

(It) is five (o'clock)

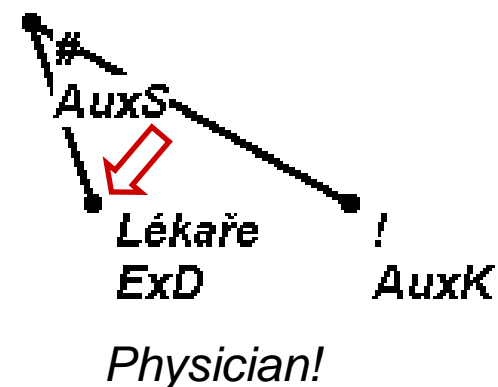
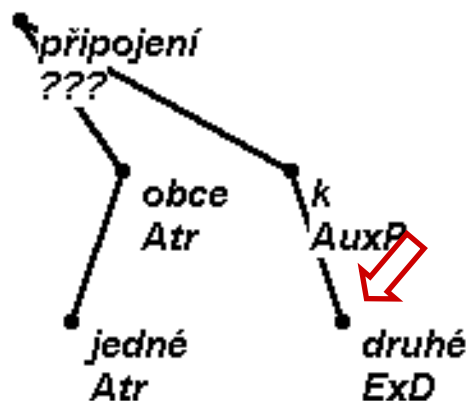




# Principles of Annotation: Ellipses

**ellipsis** ~ a deletion of an expression that is expected in the given syntactic structure

- **ExD** ... all nodes that would depend on the deleted node
- exception: prepositions (AuxP) and subord. conjunctions (AuxC) connecting nodes for coordination/apposition (Coord, Apos)
- lexicalized ellipses vs. actual (textual) ellipses



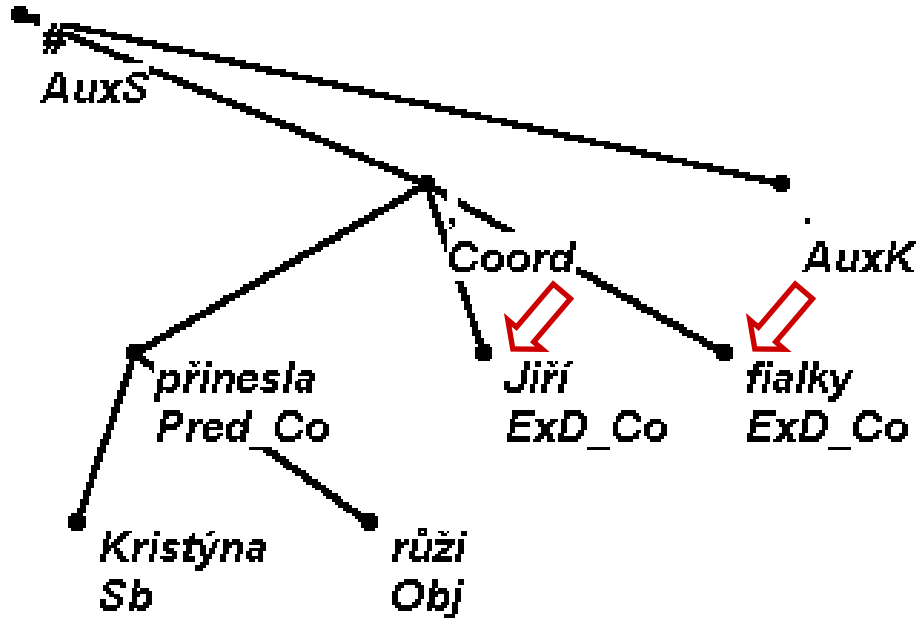
*(the) attachment of one village to another*

# Principles of Annotation: Ellipses (cont.)

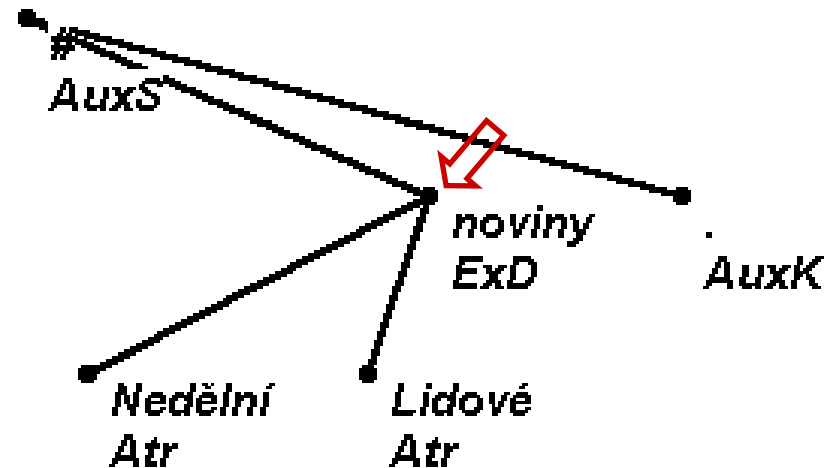


## *ellipsis*

- coordination
- one-member sentences (verb-less)



*Christine brought a rose, George violets.*



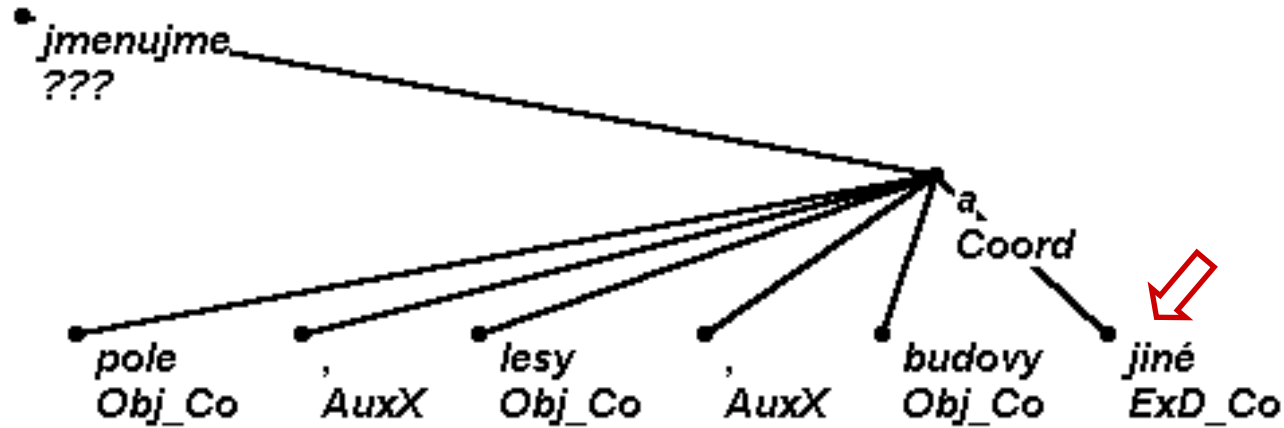
*Sunday Lidové noviny.*

---

# Principles of Annotation: Ellipses (cont.)



ellipsis and coordination:



*let us mention fields, woods, buildings and other (property)*

# Specific Constructions: Numeral Expressions

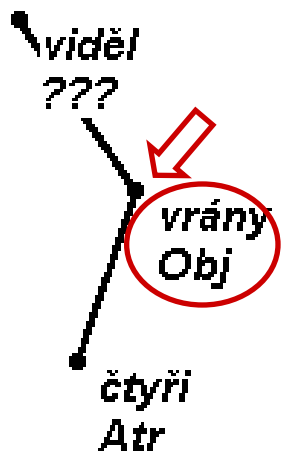


- part expressing the number + part expressing counted objects
- determine the governing and the dependent part ... based on the

**!!! morphological analysis** (→ agreement)

*viděl čtyři<sub>acc</sub> vrány<sub>acc</sub> ~ viděl vrány<sub>acc</sub> BUT viděl pět<sub>acc</sub> vran<sub>gen</sub> (\*viděl vran<sub>gen</sub>)*

... **reduction** principle

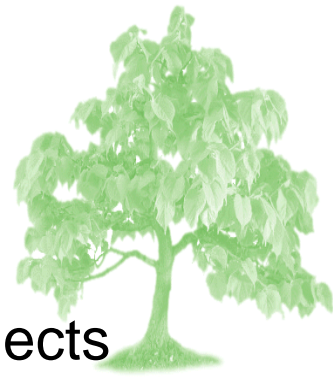


*(he) saw four crows*



*(he) saw five crows*

# Numeral Expressions (cont.)

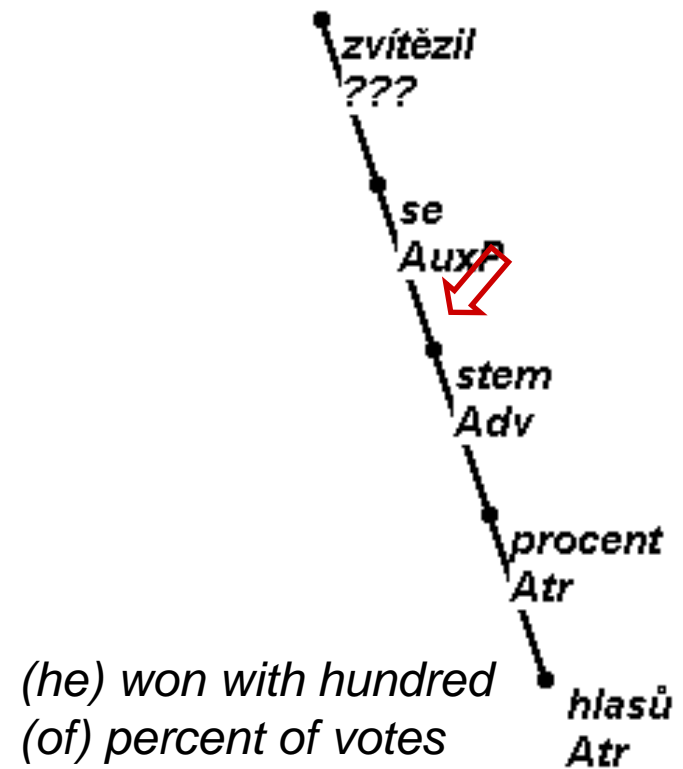
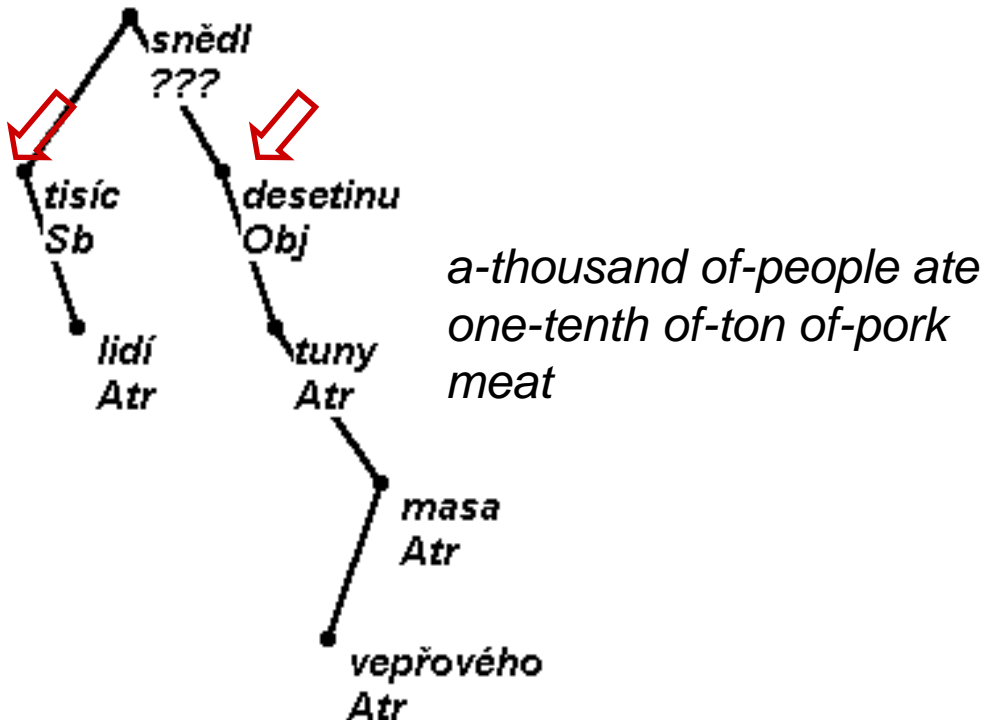


- part expressing the number + part expressing counted objects
- determine the governing and the dependent part ... based on the **!!! morphological analysis** (→ agreement)

*viděl čtyři<sub>acc</sub> vrány<sub>acc</sub> ~ viděl vrány<sub>acc</sub> BUT viděl pět<sub>acc</sub> vran<sub>gen</sub> (\*viděl vran<sub>gen</sub>)*

**!!!** numerals as **sto, tisíc, milion, stovka, setina** ... rendered as **nouns**

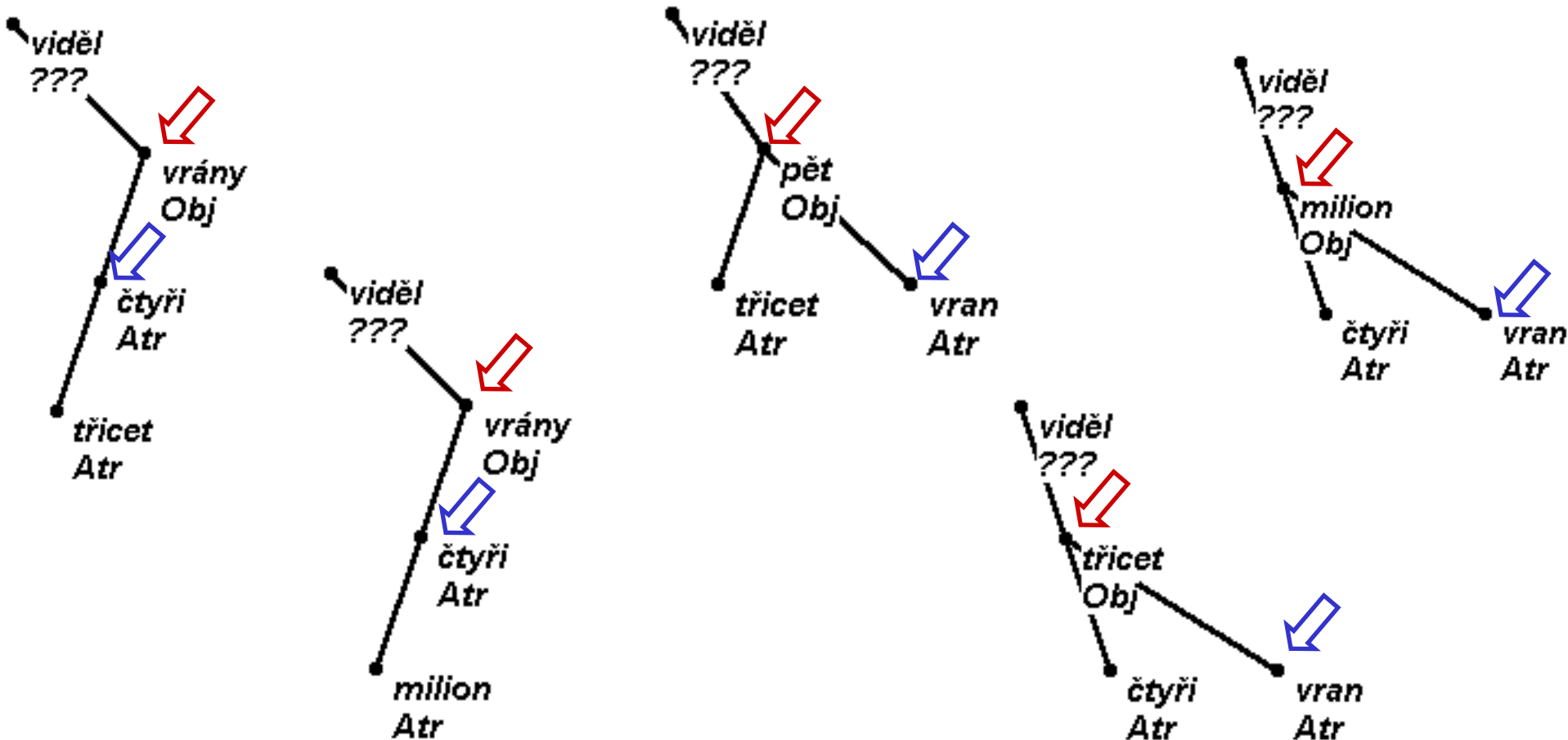
*[hundred, thousand, million, a hundred, a hundredth]*



# Numeral Expressions (cont.)



- structure of the numeral part (technical)  
head: the **lowest** rank (wrt the **agreement**)

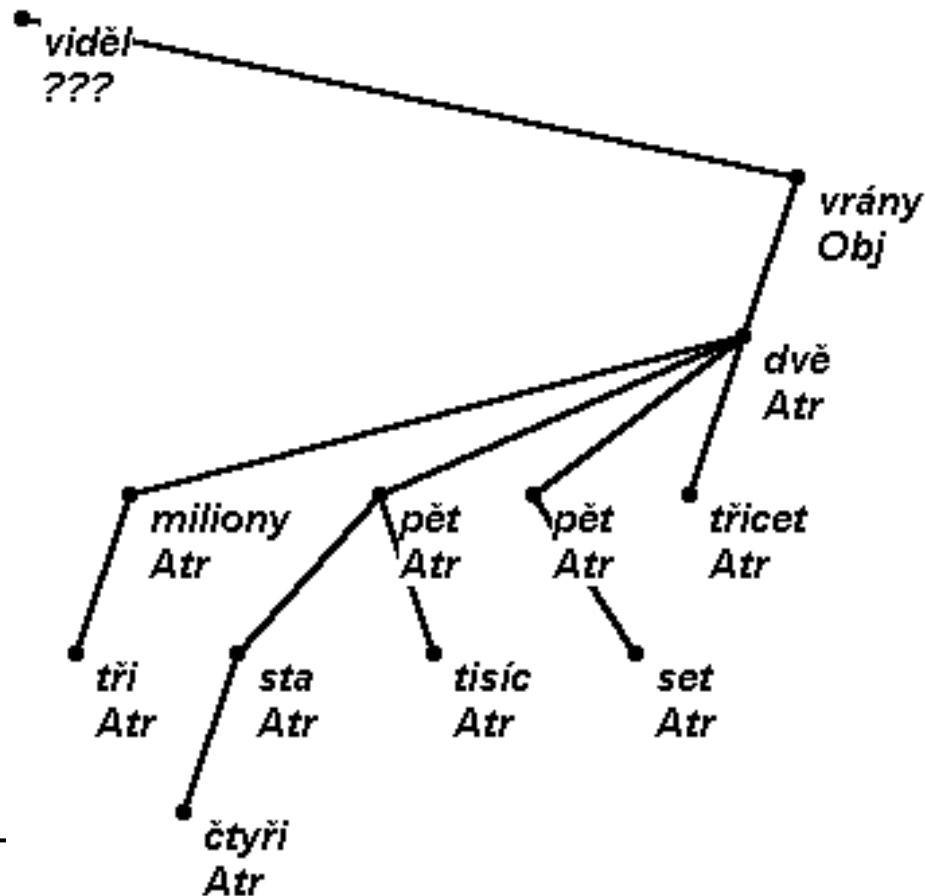


---

# Numeral Expressions (cont.)



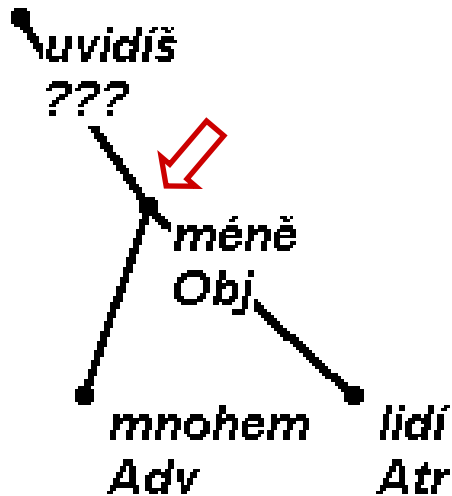
- structure of the numeral part (technical)  
head: the **lowest** rank (wrt the **agreement**)



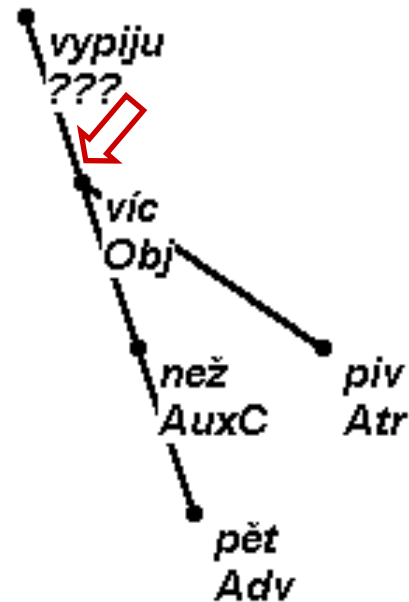
# Numeral Expressions (cont.)



- numerals, type *více* [more], *méně* [less] (not a comparison)



*(you) will-see much less people*



*(I) will-drink more than five beers*

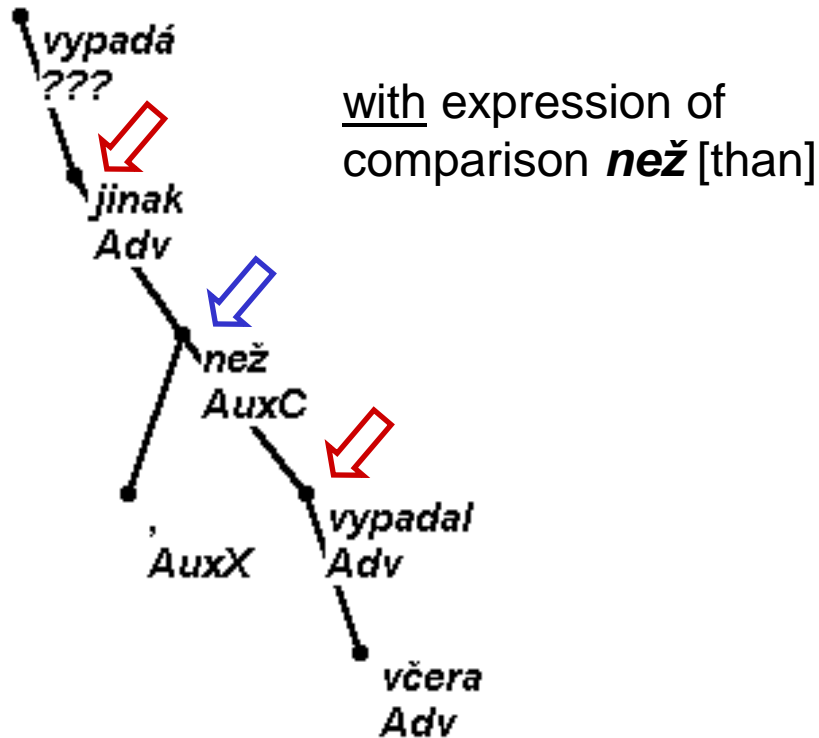


# Specific Constructions: Comparison

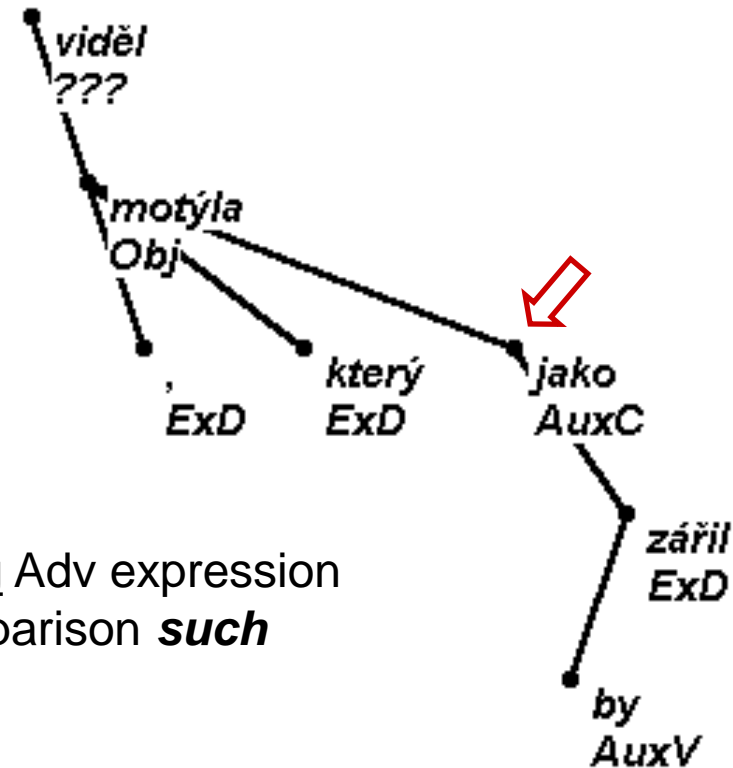


conjunctions *jako* [as, such], *než* [than] ... adverbial

1. full embedded clause



*he looks different from how he  
looked yesterday*



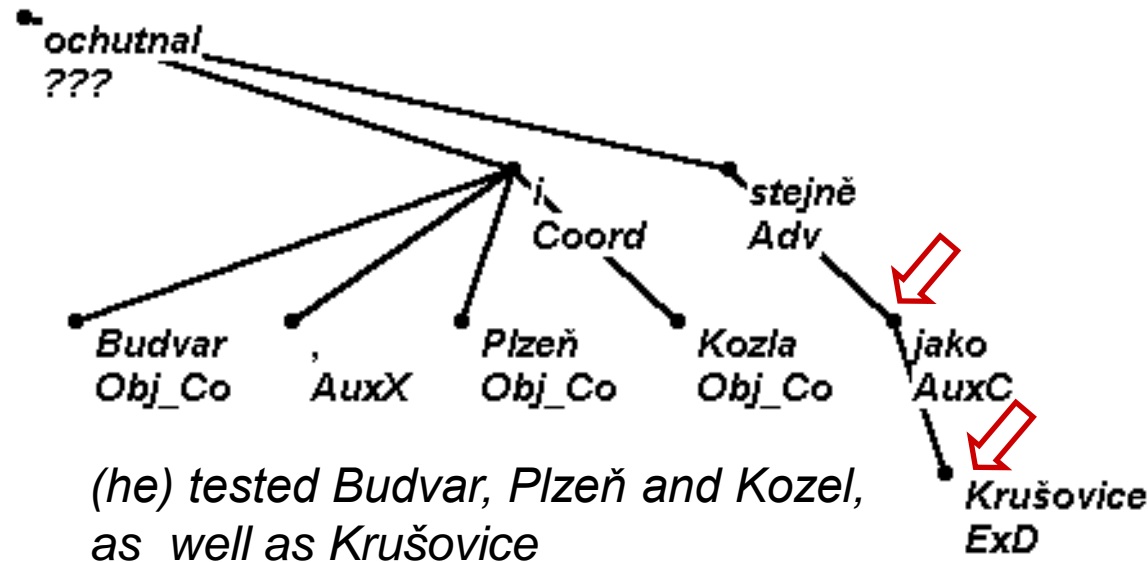
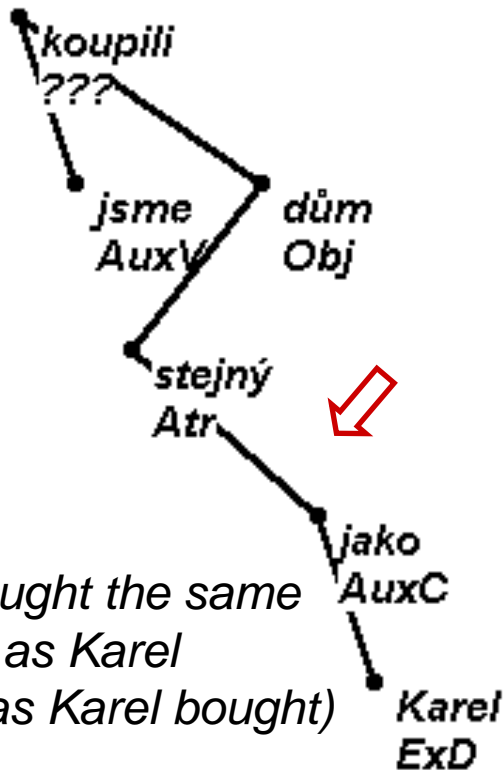
*(he) saw a butterfly that was like shining*

# Specific Constructions: Comparison



conjunctions *jako* [as, such], *než* [than] ... adverbial

1. full embedded clause
2. **ellipsis**

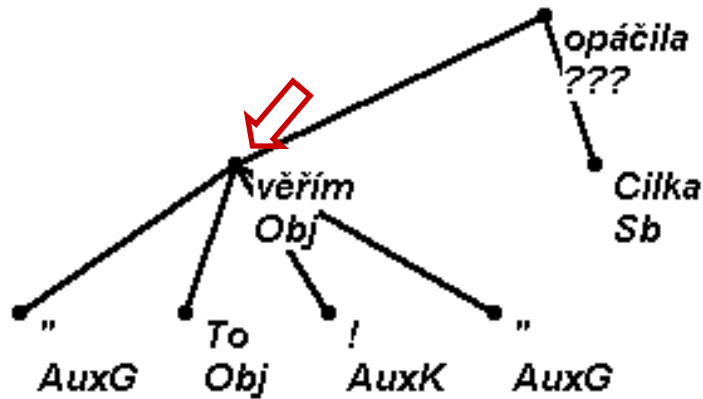


# Specific Constructions: Direct Speech

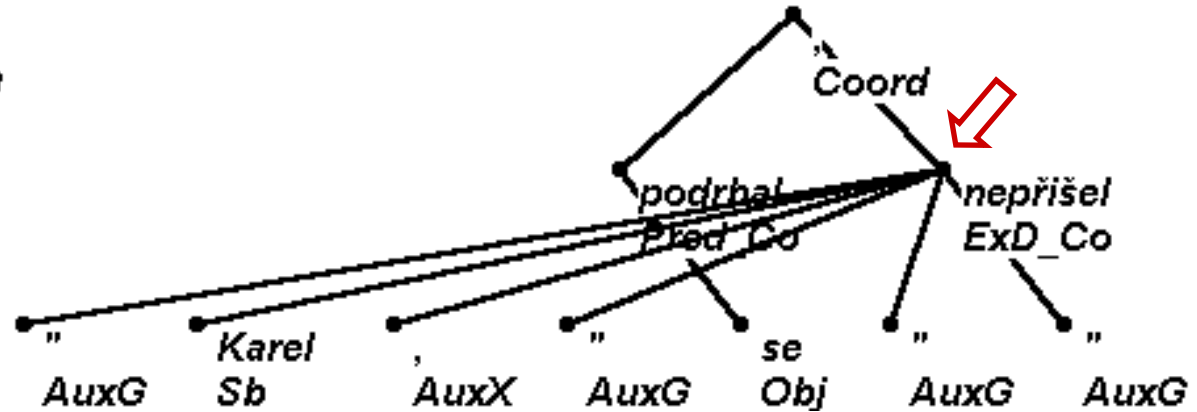


direct, semidirect and indirect speech

1. direct speech as an object
2. direct speech as coordinated sentence  
(i.e., elided communication verb)
3. (exceptions)



*"I believe it!" Cilka replied.*



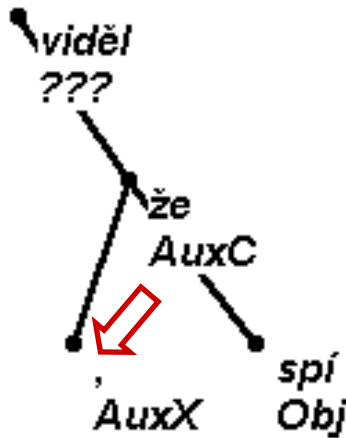
*"Karel," (he) scratched his head, "did not come "  
(= Karel, (he) scratched his head **and said**, did not come )*

# Graphic Symbols, Punctuation

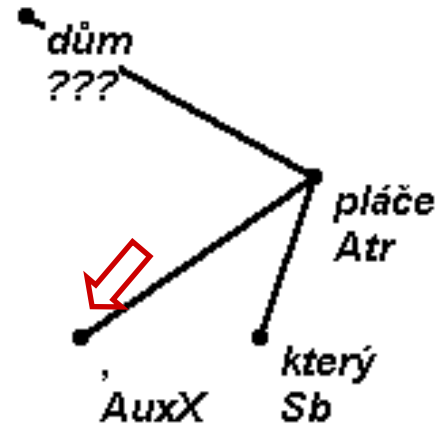


## **AuxX** ... comma

- not for connecting node of coordination and aposition; not for ellipses
  1. comma separating a clause
    - a child of the **head of the subord. clause**



*(he) saw that (he) slept*



*(the) house which is crying*



- not for connecting node of coordination and aposition; not for ellipses
  1. comma separating a clause
    - a child of the head of the subord. clause
  2. comma separating a parenthesis
    - a child of the **head of the parenthesis**



# Graphic Symbols, Punctuation



## AuxX ... comma

- not for connecting node of coordination and aposition; not for ellipses

### 1. comma separating a clause

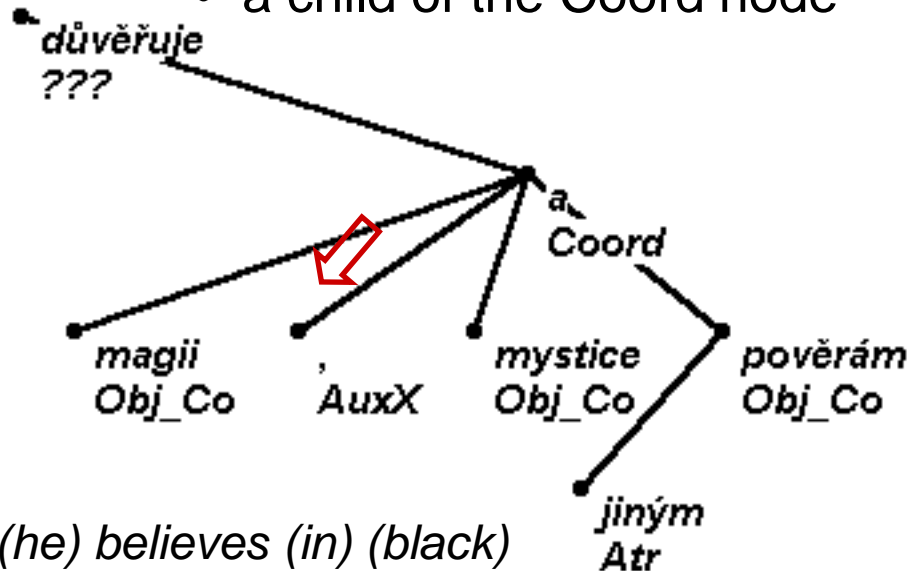
- a child of the head of the subord. clause

### 2. comma separating a parenthesis

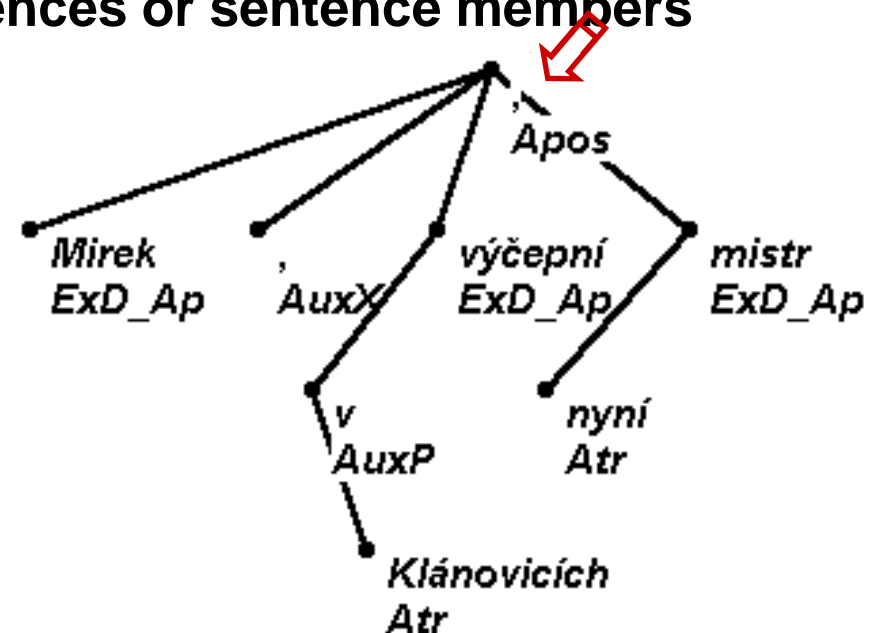
- a child of the head of the parenthesis

## !!! comma separating **coordinated sentences or sentence members**

- a child of the Coord node



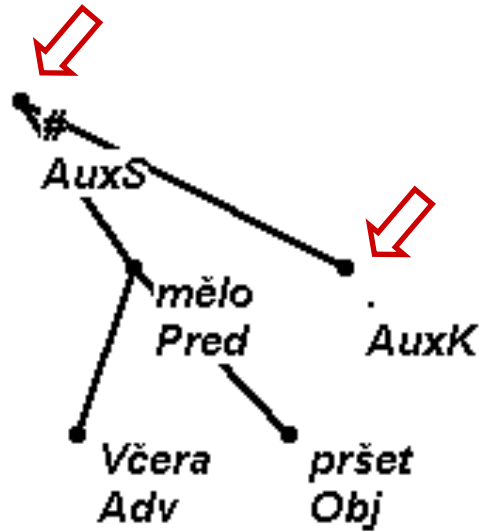
(he) believes (in) (black)  
magic, mystique and superstitions



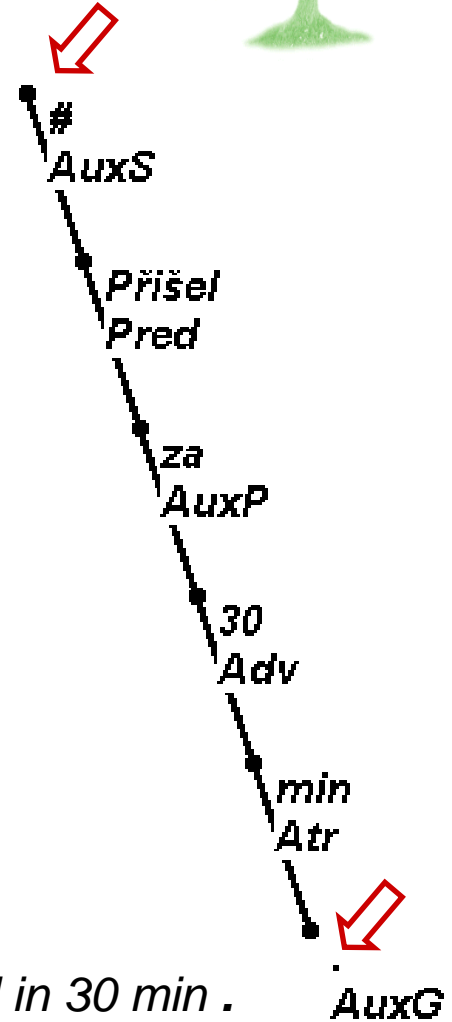
a barman in Klánovice, now a foreman

# Graphic Symbols, Punctuation

- **AuxS** ... root
- **AuxK** ... terminal punctuation mark (all types);
  - a child of the root  
(in case of a multiple function AuxK the lowest priority)



*It should have rained yesterday.*



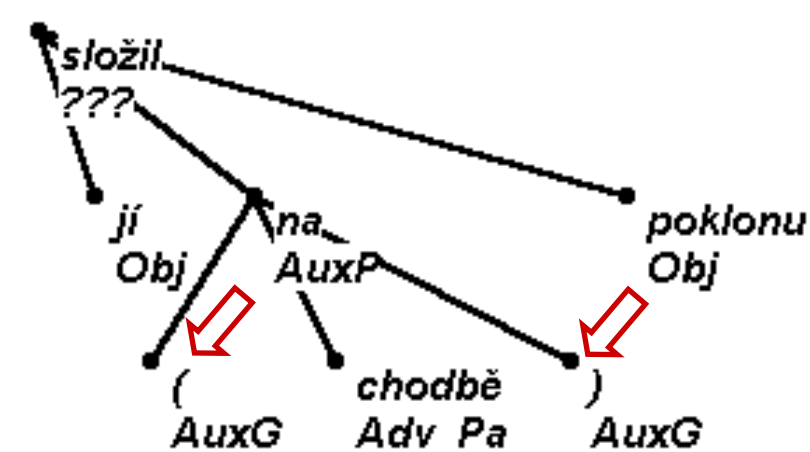
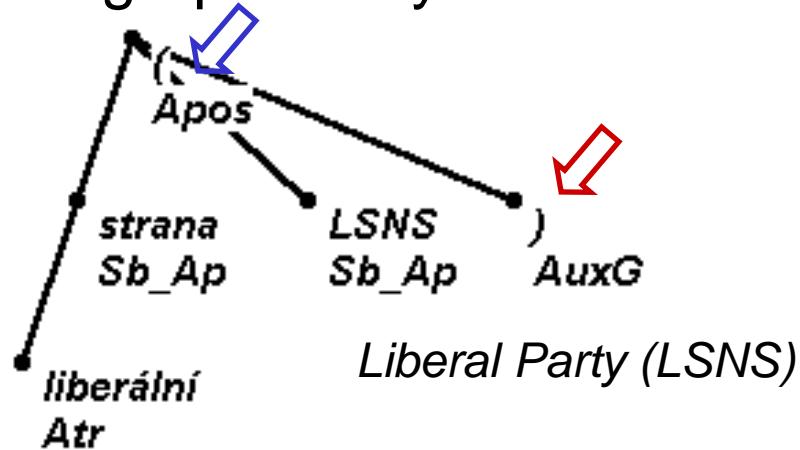
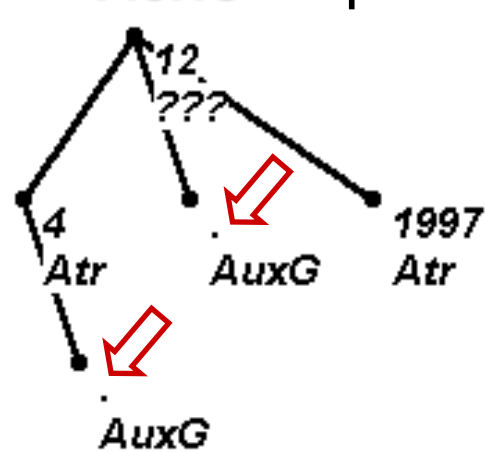
*(He) arrived in 30 min .*



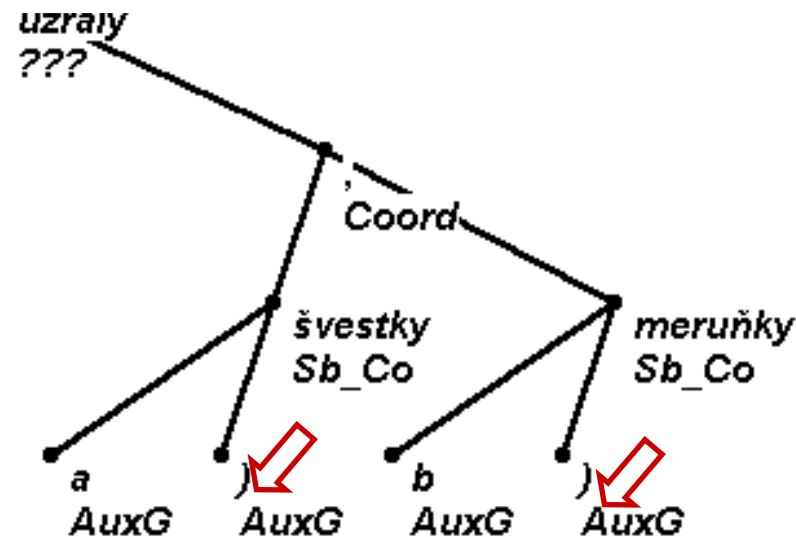
# Graphic Symbols, Punctuation



- **AuxG** ... punctuation, other graphical symbols

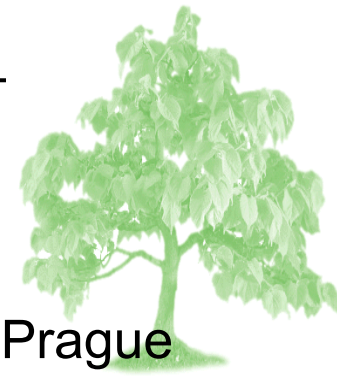


he paid her (on a corridor) a compliment



a) plums, b) apricots (have) ripened





---

# References

- Hajič, J. (1998) Building a Syntactically Annotated Corpus: The Prague Dependency Treebank". In E. Hajičová (ed.): *Issues of Valency and Meaning. Studies in Honour of Jarmila Panevová*, Karolinum, Charles University Press, Prague, Republic, pp. 106-132
- Manual for Analytical Annotation  
<http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/a-layer/html/index.html>