Prague Dependency Treebank: Annotation of Surface Syntax

Markéta Lopatková

Institute of Formal and Applied Linguistics, MFF UK

lopatkova@ufal.mff.cuni.cz

PDT: a-layer

Goal:

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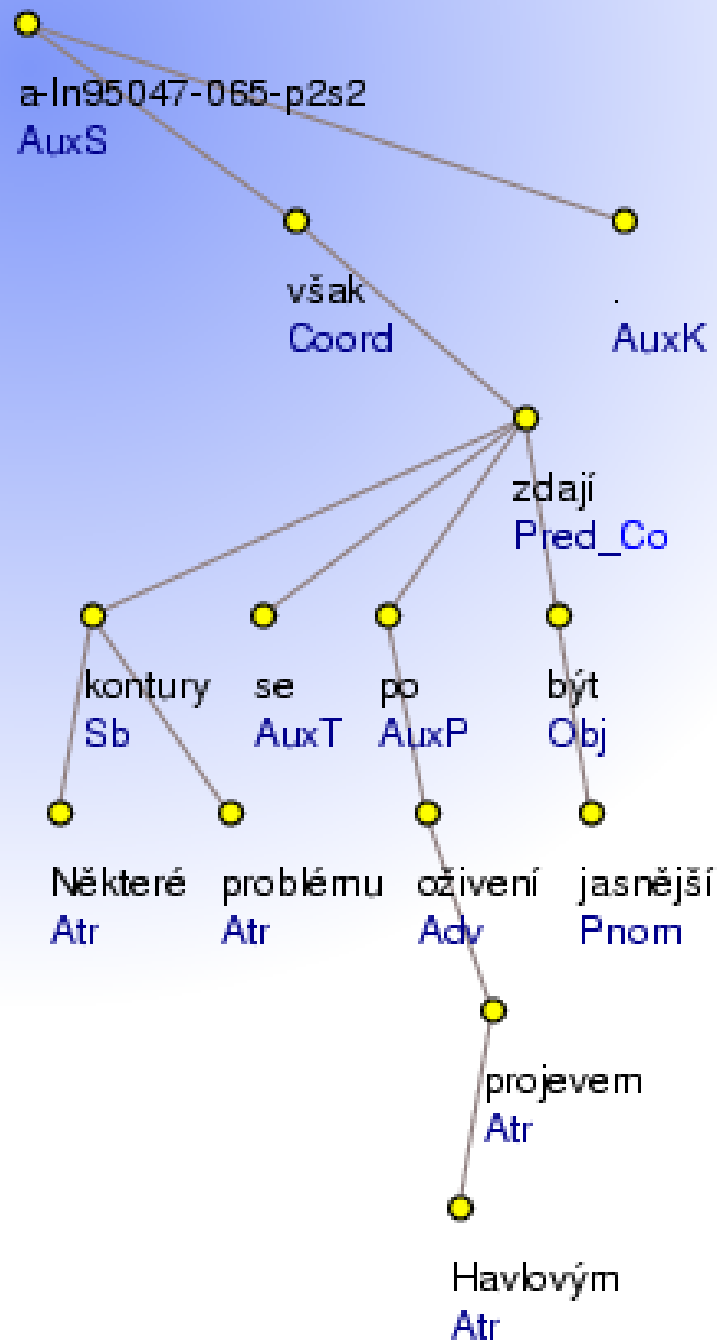- to denote the type of relations between "words"

- dependency tree   !!! tree structure ~ XML structure !!!
- one token from m-layer ~ one node incl. prepositions, punctuation …
  plus technical root
- relations ~ edges
    - dependency, coordination, punctuation, …
    - type of relation: attribute of the child node
    - oriented "upwards", i.e., towards its parent / "governing" node
- linear ordering ~ surface word order

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

PDT: a-layer

Některé kontury problému se však po oživením Havlovým projevem zdají být jasnější .
[Some contours of the problem seem to be clearer after the resurgence by Havel's speech.]
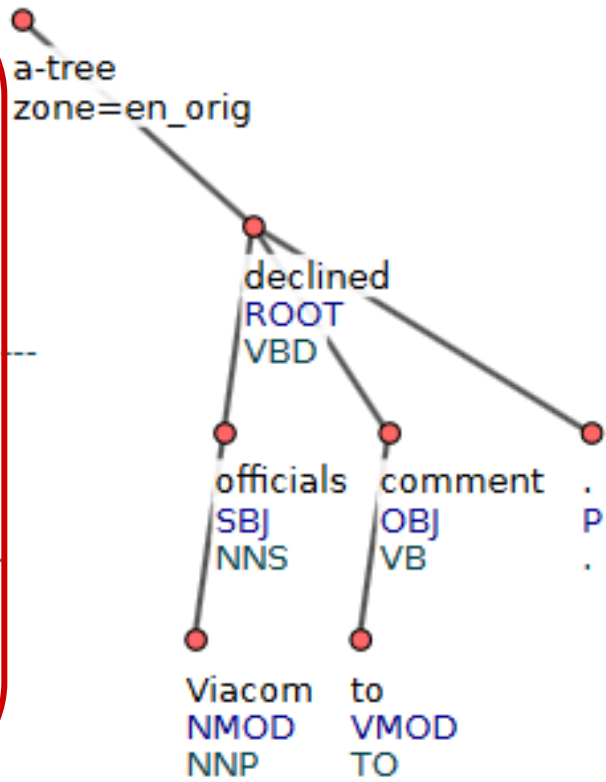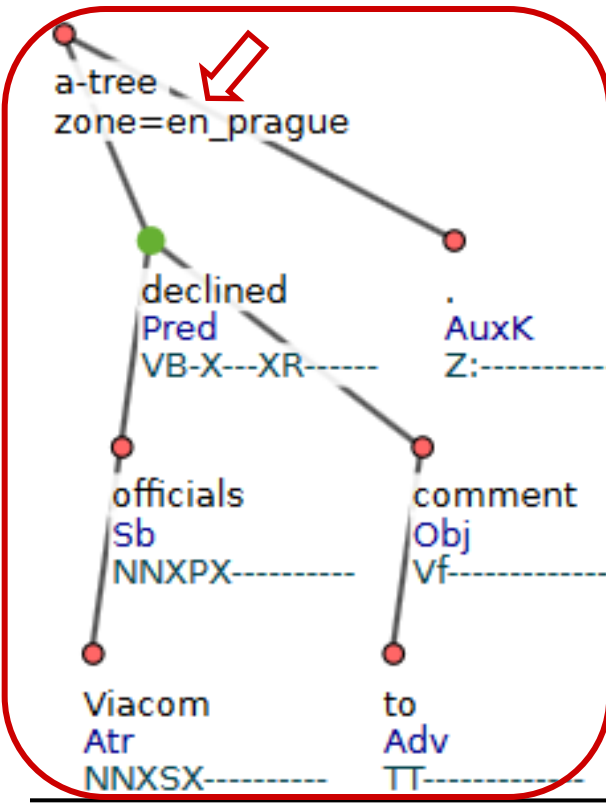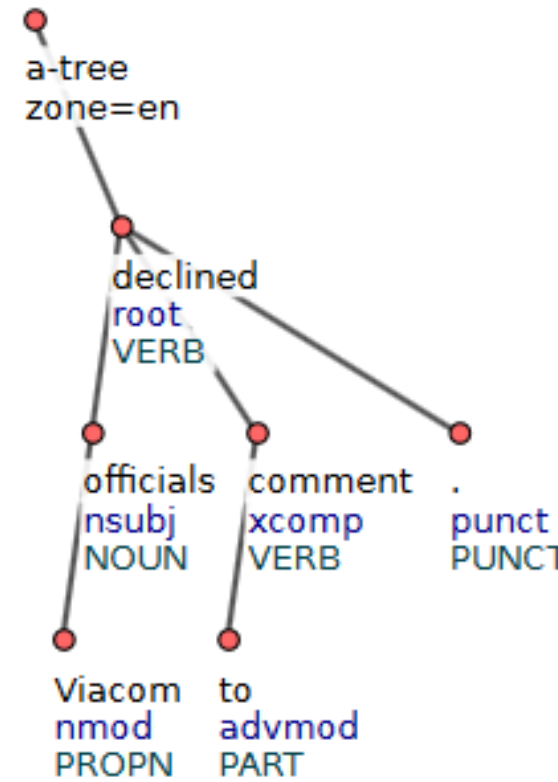
HamleDT: a-layer

(login required for English, free whenever possible)

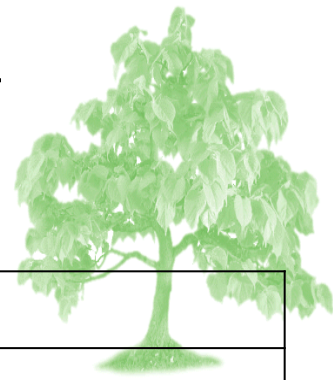PDT style

Universal Dependencies

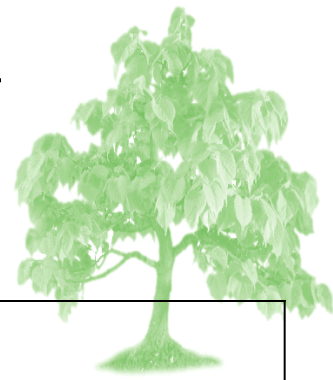PDT: a-layer   Lopatková

A-layer: attributes of a node (non-root)

Attribute Name	Description
id	unique identifier of the node in PDT 2.0
m.rf	PML reference; points to a node on the m-layer (i.e., node m)
afun	**analytical function ~ kind of relation between the node and its parent node**
is_member	0 or 1; denoting members of a coordination or apposition; only children of a node with afun Coord or Apos (disregarding AuxP and AuxC)
is_parenthesis__root	0 or 1; 1 identifies roots of subtrees corresponding to parentheses
ord	positive integer; the (left to right) order of representing the nodes in graphical applications
clause_member	identifies a clause to which the node belongs (for Czech)

A-layer: attributes of a root

Attribute Name	Description
id	unique identifier of the tree in PDT 2.0
s.rf	PML reference (points to a sentence marked s on the m-layer)
afun	AuxS (analytical function)
ord	0 (position in the horizontal ordering of the nodes in a tree)

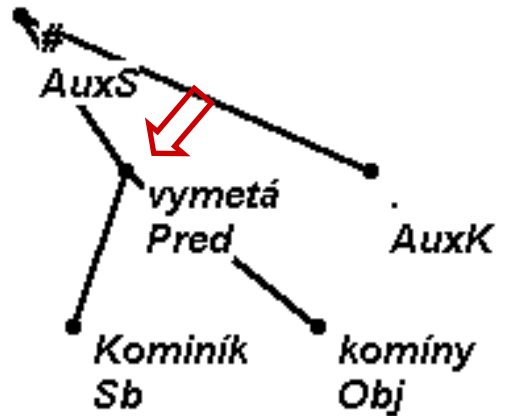# Analytical functions (afun)

A simple sentence

Predicate … analytical functions: **Pred, Pnom, AuxV**

- basic sentence member
- some property/state/change/activity is attributed to the subject
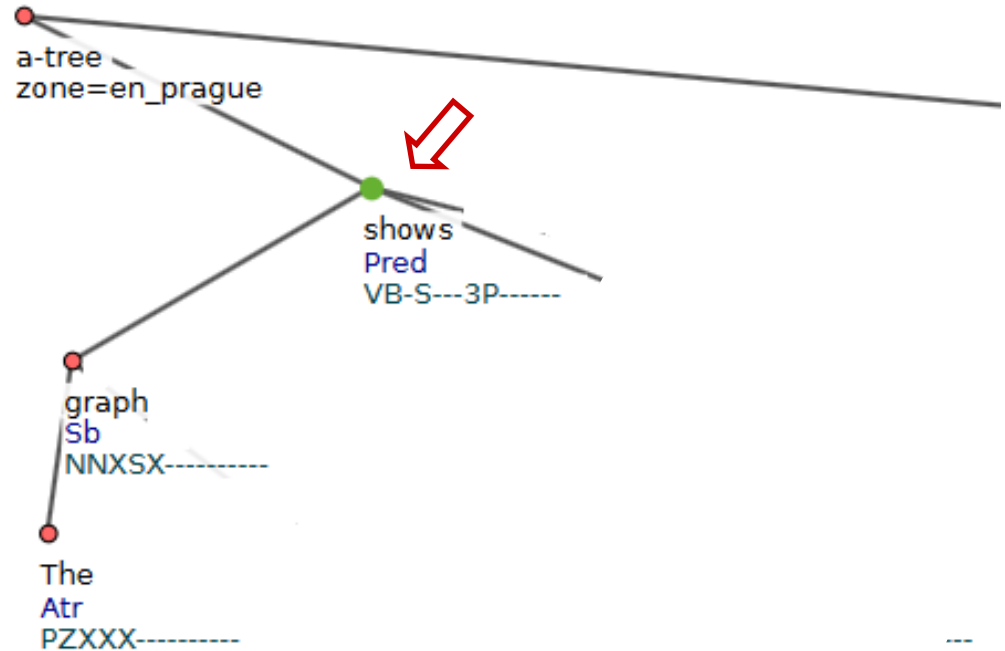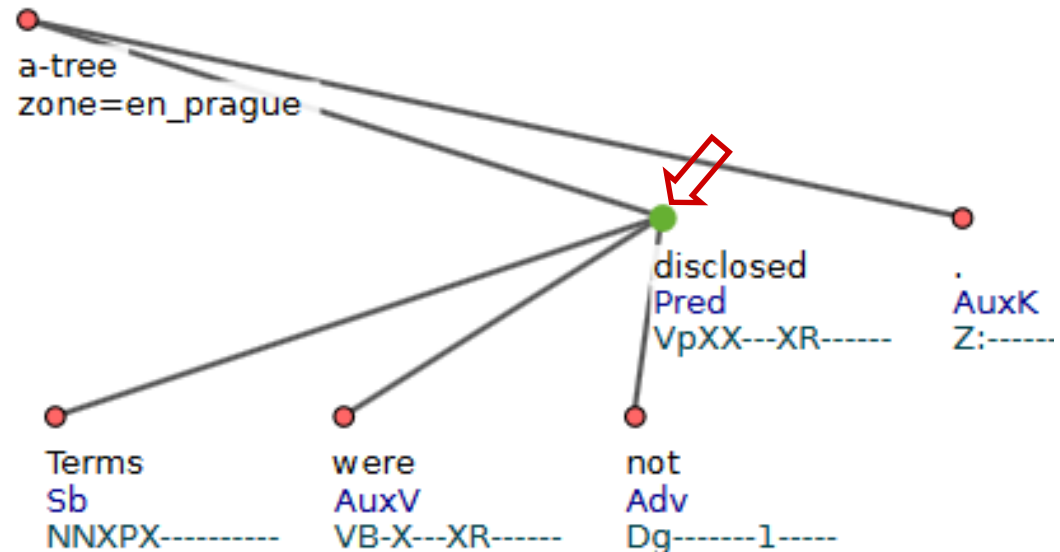- governing node (= head) of its clause

Predicate: Dependency Structure

- predicate of a **main clause**
  - a child of the root, analytical function … **Pred**

AuxS
vymetá
Pred
Kominík
Sb
komíny
Obj
.
AuxK

*Chimneysweep **sweeps** chimneys.*

a-tree
zone=en_prague

shows
Pred
VB-S---3P------

graph
Sb
NNXSX----------

The
Atr
PZXXX----------

The graph shows ...

Predicate: Dependency Structure

- predicate of a **main clause**
  - a child of the root, analytical function … **Pred**

passive sentence

a-tree
zone=en_prague
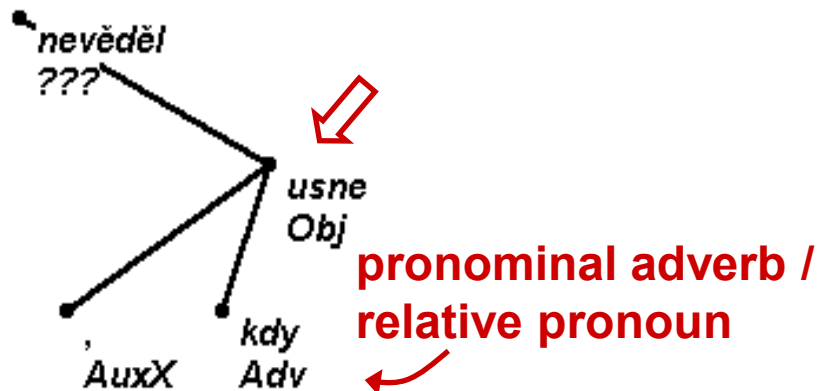
disclosed
Pred
VpXX---XR------

. AuxK
Z:------

Terms
Sb
NNXPX----------

were
AuxV
VB-X---XR------

not
Adv
Dg-------1-----
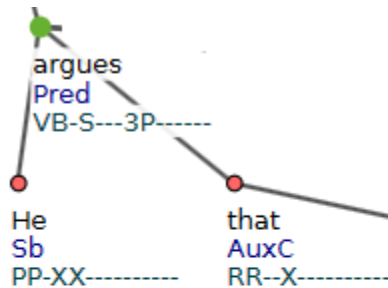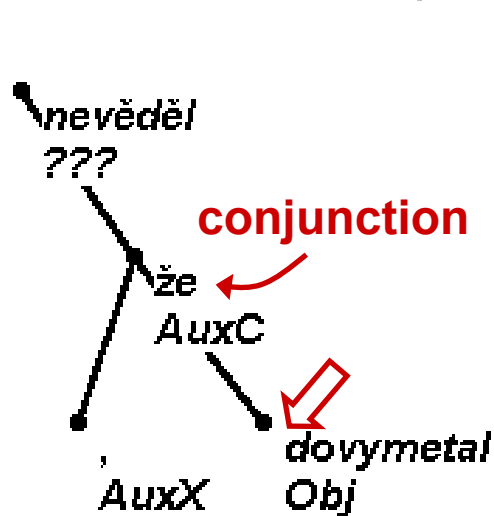
Terms were not disclosed.
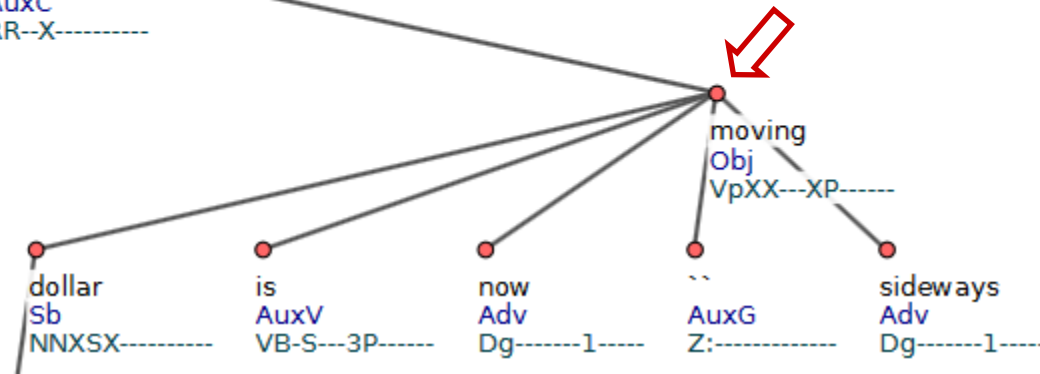
Predicate: Dependency Structure

- predicate of a **main clause**
  - a child of the root, analytical function … **Pred**

BUT:

- predicate of a **subordinate clause**
  - *afun of the respective function*

pronominal adverb / relative pronoun

*did-not-know when **falls-asleep***

a-tree
zone=en_prague

shows
Pred
VB-S---3P------

graph
Sb
NNXSX----------

happens
Obj
VB-S---3P------

The
Atr
PZXXX----------

how
Adv
P4XXX----------

this
Sb
PZXXX----------

Predicate: Dependency Structure

- predicate of a **main clause**
  - a child of the root, analytical function … **Pred**

BUT:

- predicate of a **subordinate clause**
  - afun of the respective function

argues
Pred
VB-S---3P------

He
Sb
PP-XX----------

that
AuxC
RR--X----------

conjunction

neveděl
???

že
AuxC

'
AuxX

dovymetal
Obj

did-not-know that
he-had-finished-sweeping

He argues that the dollar is now ``moving sideways …

moving
Obj
VpXX---XP------

dollar
Sb
NNXSX----------

is
AuxV
VB-S---3P------

now
Adv
Dg-------1-----

``
AuxG
Z:-------------

sideways
Adv
Dg-------1-----

Predicate: Analytical Verb <u>Forms</u>

- auxiliary verb(s) + lexical verb
    - lexical verb … head … **Pred**
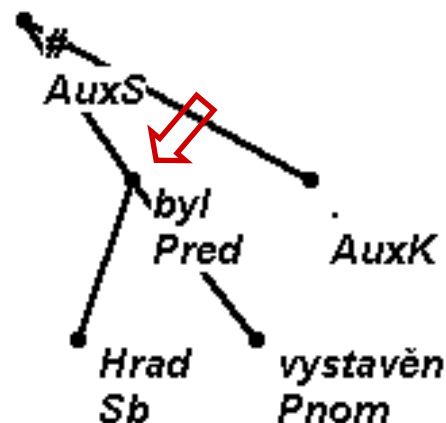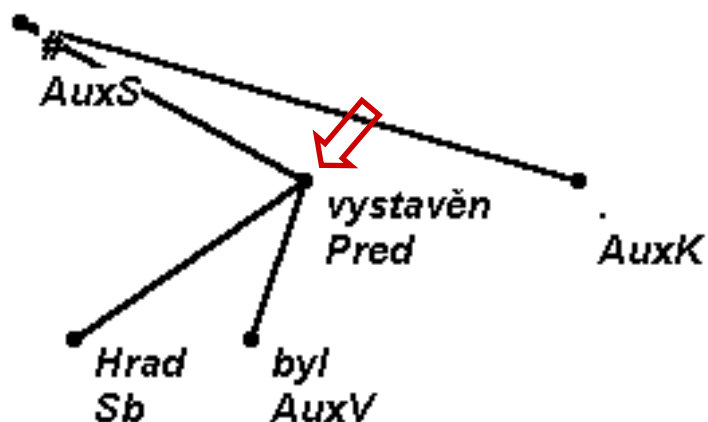   - auxiliary verb(s) … child(ren), afun **AuxV**

a-tree
zone=en_prague

disclosed
Pred
VpXX---XR------

.
AuxK
Z:------

Terms
Sb
NNXPX----------

were
AuxV
VB-X---XR------

not
Adv
Dg-------1-----

He argues that the dollar is now ``moving sideways …

\#
AuxS

sedával
Pred

.
AuxK

Karel
Sb

by
AuxV

byl
AuxV

na
AuxP

židli
Adv

své
Atr

Karel by   byl   sedával   na své židli.
Karel would be$_{Past}$ used-to-sit on his chair.
Charles would have used to sit in his chair.

Predicate: Analytical Verb <u>Forms</u>

- auxiliary verb(s) + lexical verb
  - lexical verb … head … **Pred**
  - auxiliary verb(s) … child(ren), afun   **AuxV**

- Note: **passive participle** (= action)   vs.   **state**
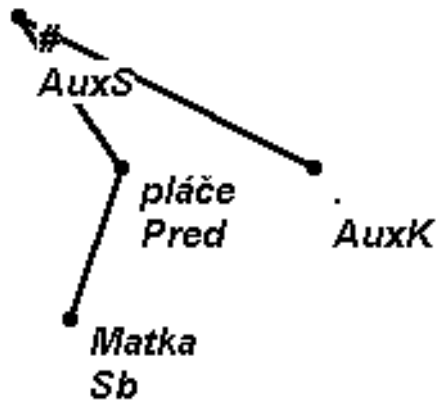

Hrad byl vystavěn.
(The) castle was erected.
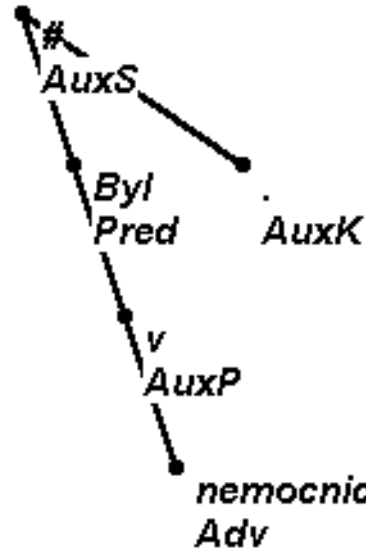
Predicate: Three Different Types

1. simple verbal predicate (incl. analytical verbal forms)
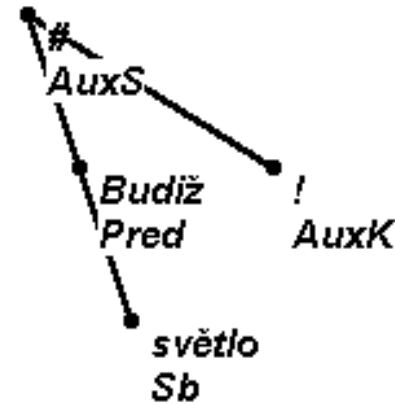   typically a finite verb form … **(AuxV +) Pred**

   lexical or *být [to be]* (existential, substitute), incl. *lze, nelze [is (not) possible]*
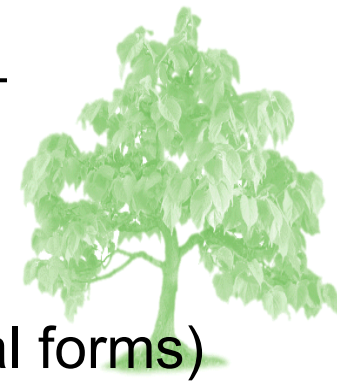

Mother cries.

(He) was in (a) hospital.
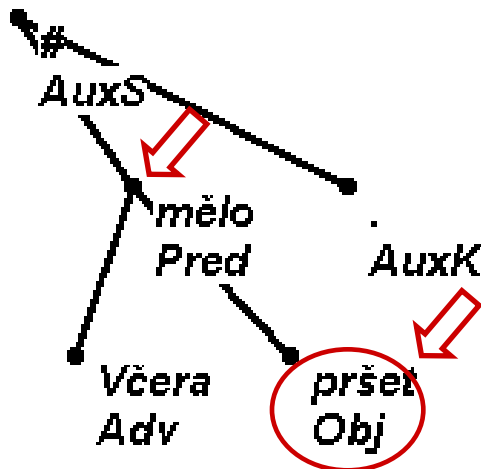
Let there be light.

Hurrah for home!

Predicate: Three Different Types

1. simple verbal predicate (incl. compound/analytical verbal forms)
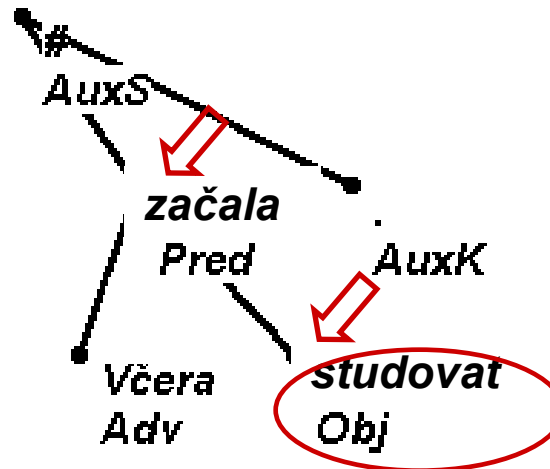2. **compound verbal predicate** … **Pred + Obj**

    finite **modal / phase verb** + infinitive of **lexical verb**
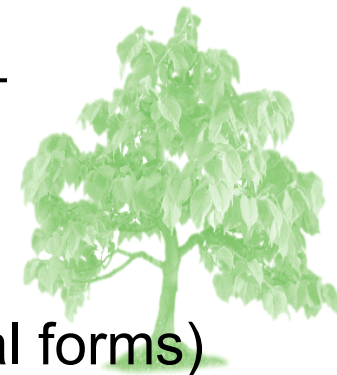
    mít [should], muset [must], moci [can], chtít [to want], dát_se [to be possible], smět [may]; začít [to begin], skončit [to end], …
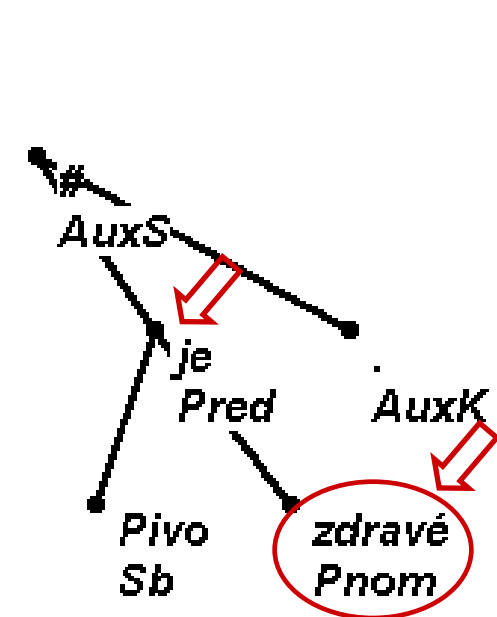

Yesterday (it) should rain.
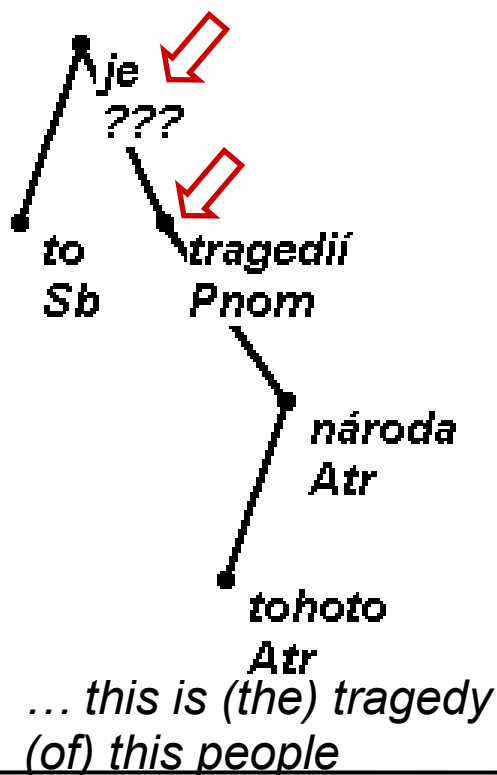
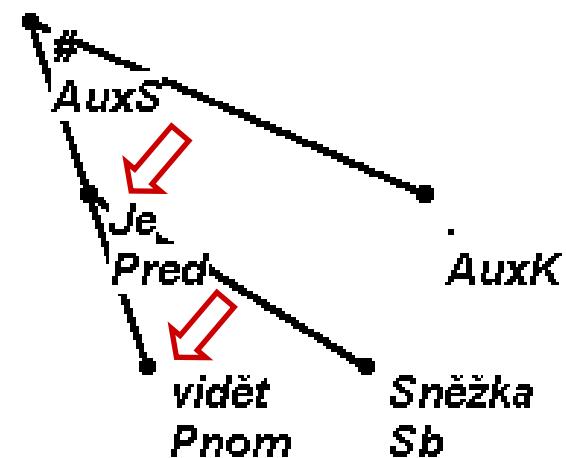*Yesterday (she) began (to) study.*

Predicate: Three Different Types

1. simple verbal predicate (incl. compound/analytical verbal forms)
2. compound verbal predicate
3. **verbal nominal predicate** (přísudek slovesně jmenný) … **Pred + Pnom**
   finite form of copula (spona) *být* + nominal part

Beer is healthy.

… this is (the) tragedy (of) this people

*it-is – to-see – Sněžka
Sněžka can be seen.*

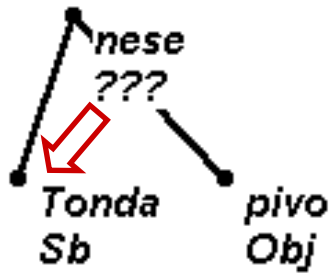# Subject

Subject … analytical function: **Sb**
- any construction answering the question *who (what)*
- dependent on a predicate
- possible forms of subject (see the data)

Subject

Subject … analytical function: **Sb**
- any construction answering the question *who (what)*
- dependent on a predicate
- possible forms of subject (see the data)
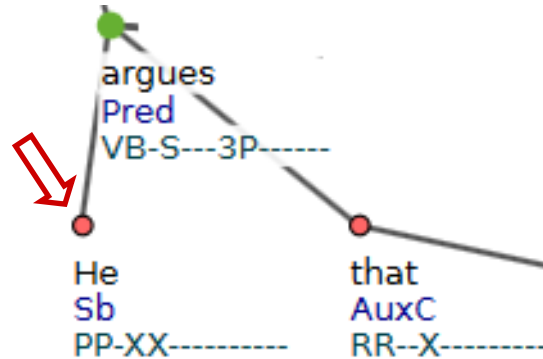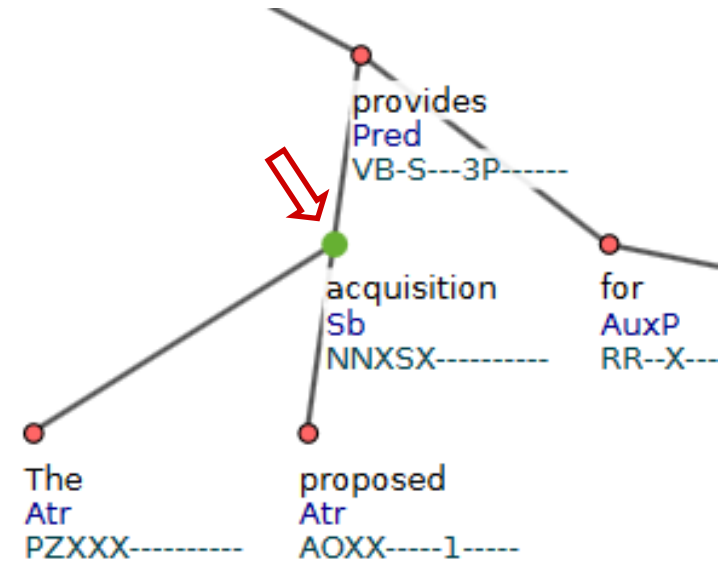
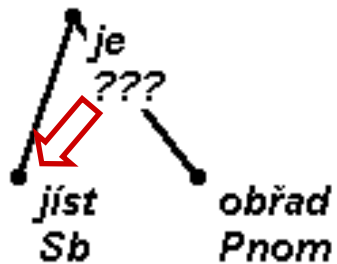

Tonda nese pivo.
Tony – brings – beer

He argues that…

The proposed acquisition provides for …
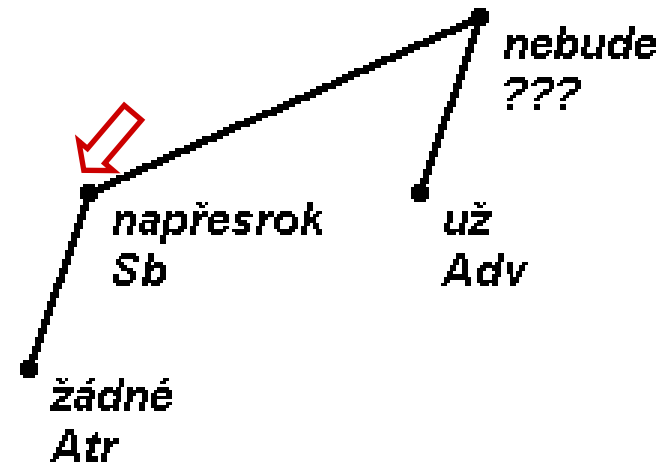
Subject

Subject … analytical function: **Sb**

- any construction answering the question *who (what)*
- dependent on a predicate
- possible forms of subject (see the data)
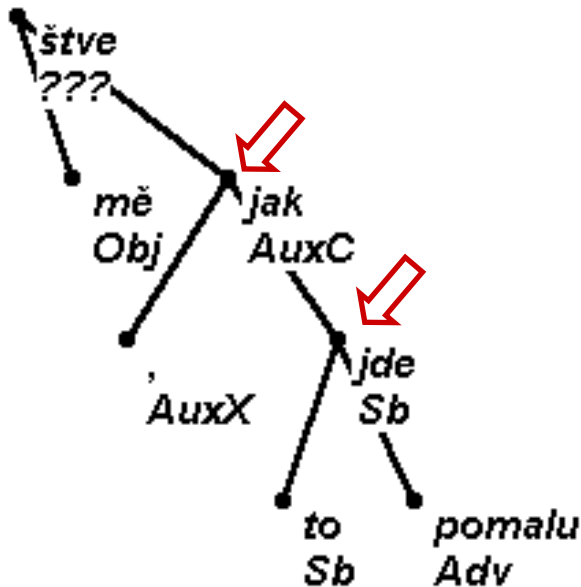

jíst je obřad
to-eat – is – ceremony

žádné napřesrok už nebude
no – next-year – anymore – will-not-be

Subject (cont.)

Subject … analytical function: **Sb**

- any construction answering the question *who (what)*
- dependent on a predicate
- possible forms of subject (see the data)

štve mě, jak to jde pomalu
irritates – me – how – it – proceeds – slowly
it irritates me how slowly it proceeds

Subject (cont.)

Subject … analytical function: **Sb**
- any construction answering the question *who (what)*
- dependent on a predicate
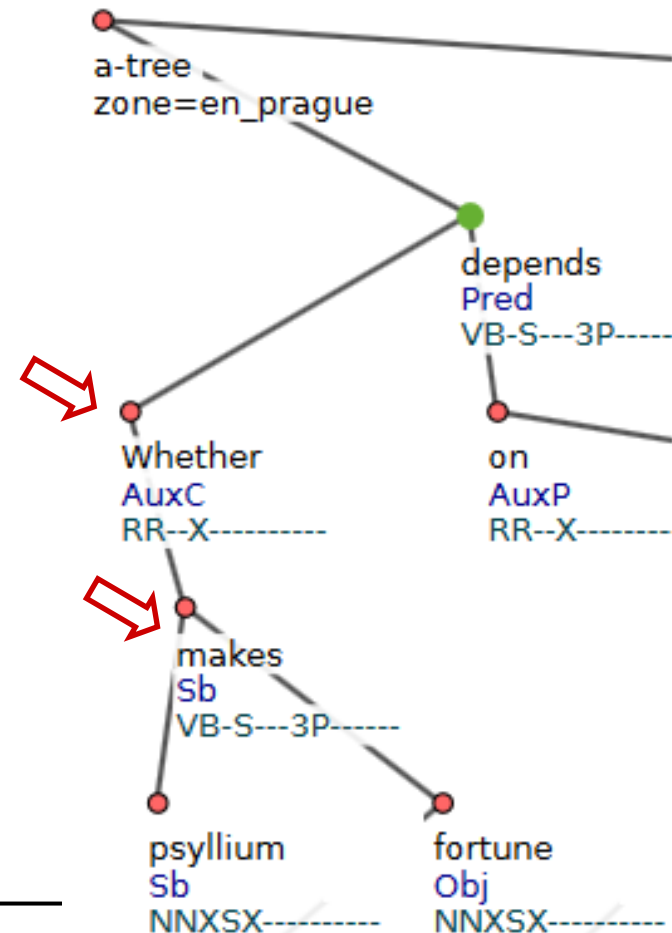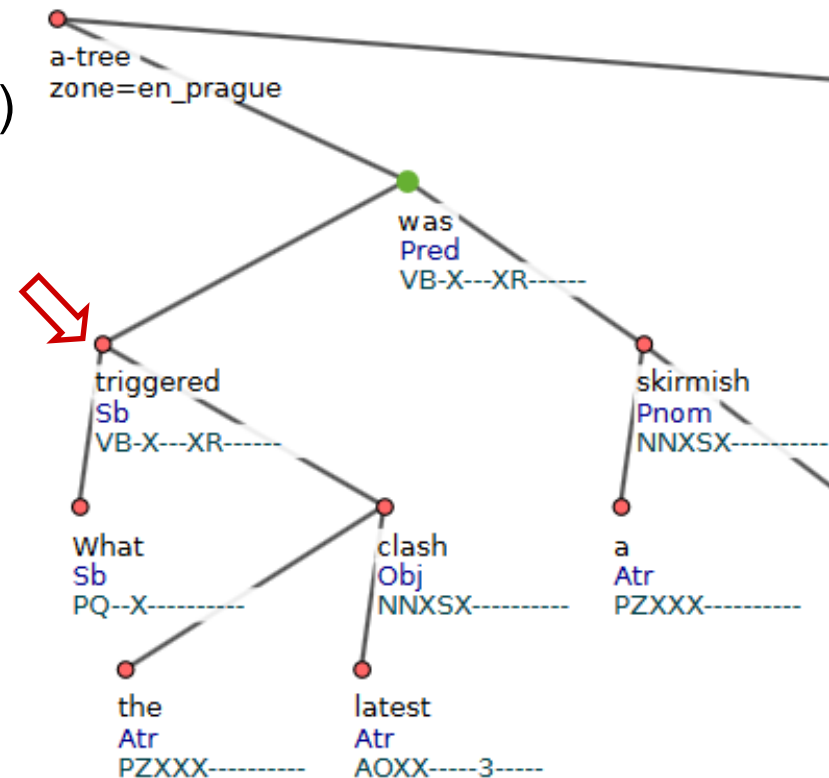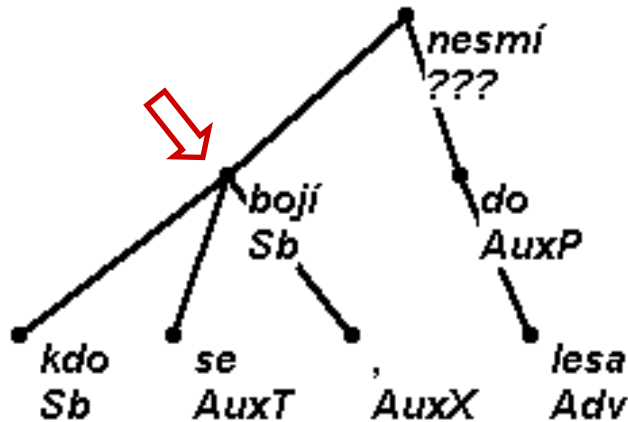- possible forms of subject (see the data)

kdo se bojí, nesmí do lesa
who –refl – is-frightened – must-not – to – wood
whoever is frightened, must not enter a wood

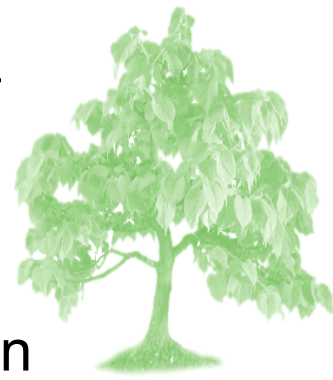What triggered the latest clash was a skirmish …

Attribute

Attribute … analytical functions: **Atr**
- <u>modifies a noun </u>(with any function)
- answers the question … *which*, *what* or *whose* (jaký, který, čí)

Attribute: Agreement

- agreeing attribute … the same case as its governing noun
    - Adj … case, number, gender
    - BUT: *kluku ušatá*

písmena
???

červená
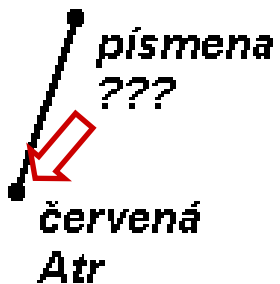*Atr*

red letters

maminka
???

nebožka
Atr

the dead mother

Attribute: Agreement

- agreeing attribute … the same case as its governing noun
    - Adj … case, number, gender
    - BUT: *kluku ušatá*
- non-agreeing attribute

red letters

the dead mother

cottage of our neighbor

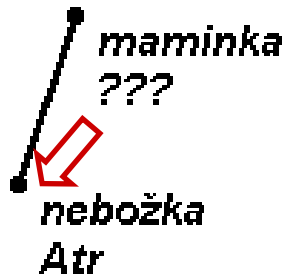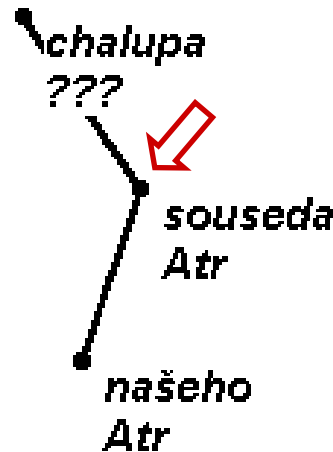*an inscription "Not for sale" was hanging on the door*

Attribute: Agreement

- agreeing attribute … the same case as its governing noun
  - Adj … case, number, gender
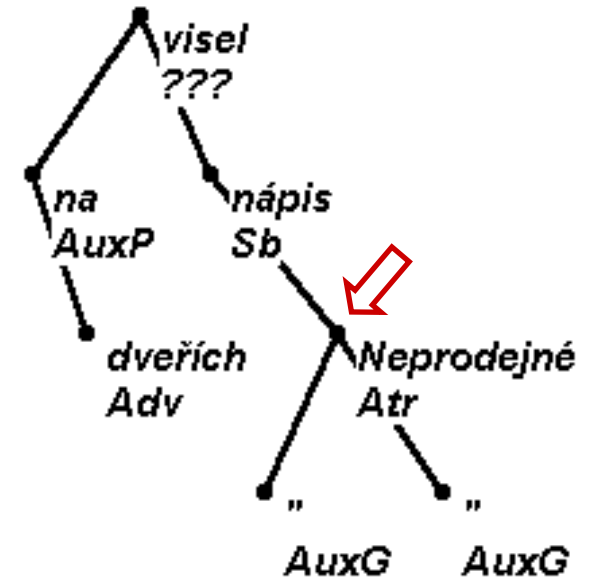 - BUT: *kluku ušatá*
- non-agreeing attribute

passing from defense to attack
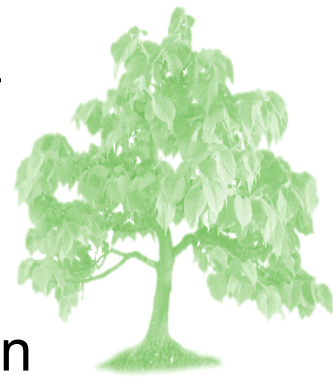
kitchen like (a) palm

knowledge that death does-not-wait

Attribute (cont.)

Atr … ALSO "technical" solution for
- addresses
- names of persons and institutions
- foreign words
- expressions with numerals
- figures

Attribute: Non-projectivity

Czech:

there is a little interest in Letná

Question*: Is there any non-projectivity in English HamleDT (Prague-style) ???*

Object

Object … analytical functions: **Obj**
- object modifies a verb / adjective / adverb
- typically, **a form of an object is prescribed** by its governing word (esp. case; cz term: **rekce**)

Mirek hated (a) sentence analysis

Object

Object … analytical functions: **Obj**
- object modifies a verb / adjective / adverb
- typically, *a form of an object is prescribed* by its governing word (esp. case; cz term: **rekce**)
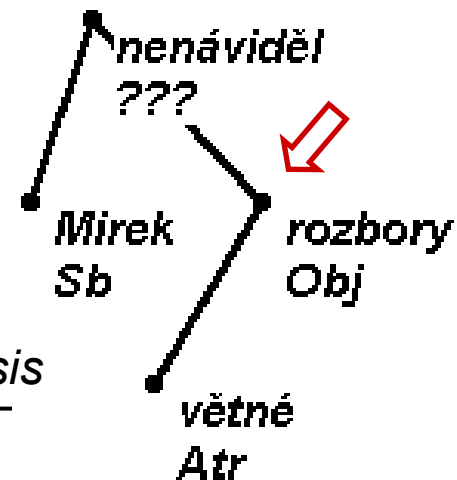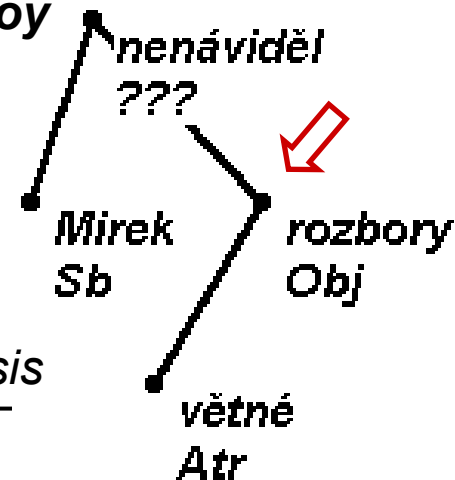
1. direct (accusative), indirect (dative), second object
- **result/effect** of an action,
  e.g. *to write **a letter,** to convert a document from one format **to another**, she was appointed **a special assistant***
- **what is affected**, e.g. *to touch **a table***
- what the action is directly **aiming at**, e.g. *to advise **a boy***
- **an origin**,
  e.g. *to convert a document **from one format** to another*
- (infinitive following a modal or phase verb)

nenáviděl
???

Mirek
Sb

rozbory
Obj

Mirek hated (a) sentence analysis

větné
Atr

Object

Object … analytical functions: **Obj**

- object modifies a verb / adjective / adverb
- typically, **a form of an object is prescribed** by its governing word (esp. case; cz term: **rekce**)

1. direct (accusative), indirect (dative), second object
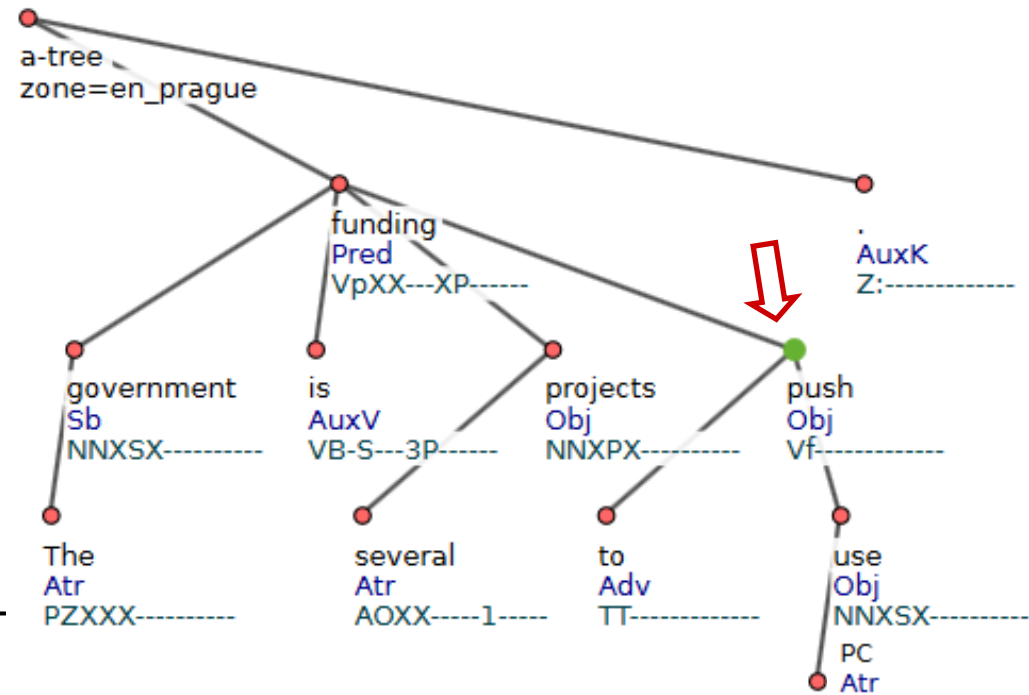- incl. **infinitive constructions**

Object

Object … analytical functions: **Obj**

- object modifies a verb / adjective / adverb
- typically, **a form of an object is prescribed** by its governing word (esp. case; cz term: **rekce**)

2. compound verbal predicate … **Pred + Obj**

   finite **modal / phase verb** + infinitive of **lexical verb**

   mít [should], muset [must], moci [can], chtít [to want], dát_se [to be possible], smět [may]; začít [to begin], skončit [to end], …

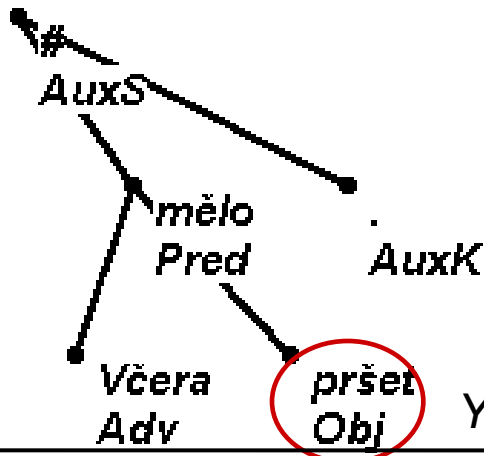*Yesterday (it) should rain.*   *Yesterday (she) began (to) study.*

Object

Object … analytical functions: **Obj**

- object modifies a verb / adjective / adverb
- typically, **a form of an object is prescribed** by its governing word (esp. case; cz term: **rekce**)
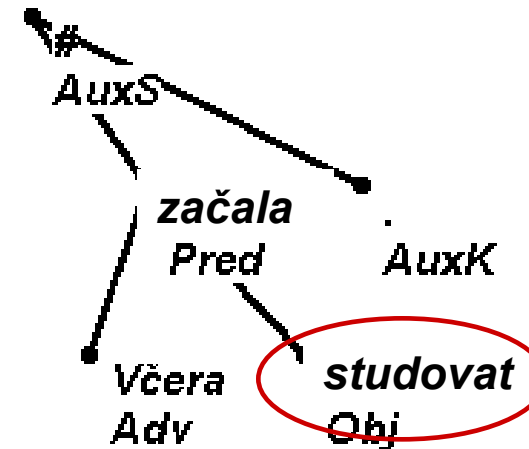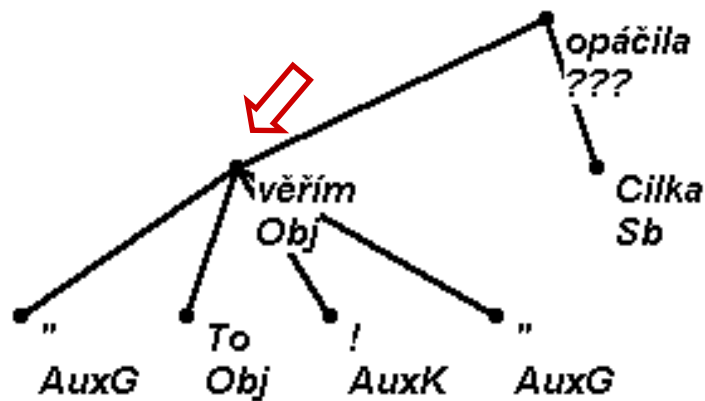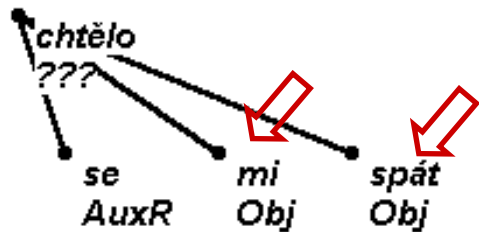
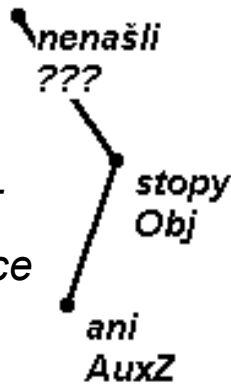**3. direct speech**

"I believe it!" Cilka replied.

Object: Possible Forms

see the data and/or the manual ☺
- typically: noun in accusative/dative/genitive/instrumental case
  preposition case
  infinitive, dependent clause

chtělo
???

se — AuxR
mi — Obj
spát — Obj

it-wanted – Refl – me – to-sleap

… I felt sleepy

nenašli
???

stopy — Obj

… (they) failed-

-to-find any trace

ani — AuxZ

a-tree
zone=en_prague

funding
Pred
VpXX---XP------

government
Sb
NNXSX----------

is
AuxV
VB-S---3P------

projects
Obj
NNXPX----------

push
Obj
Vf------------

.
AuxK
Z:-------------

The
Atr
PZXXX----------

several
Atr
AOXX-----1-----

to
Adv
TT-------------

use
Obj
NNXSX---------

PC
Atr
NNXSX---------

PDT: a-layer

Adverbials

Adverbials … analytical functions: **Adv**

- express the circumstances and relations, such as
     location, time, manner, comparison, extent, means, cause,
     consequence, regard or aim
- can modify a verb, adjective or adverb

Adverbials

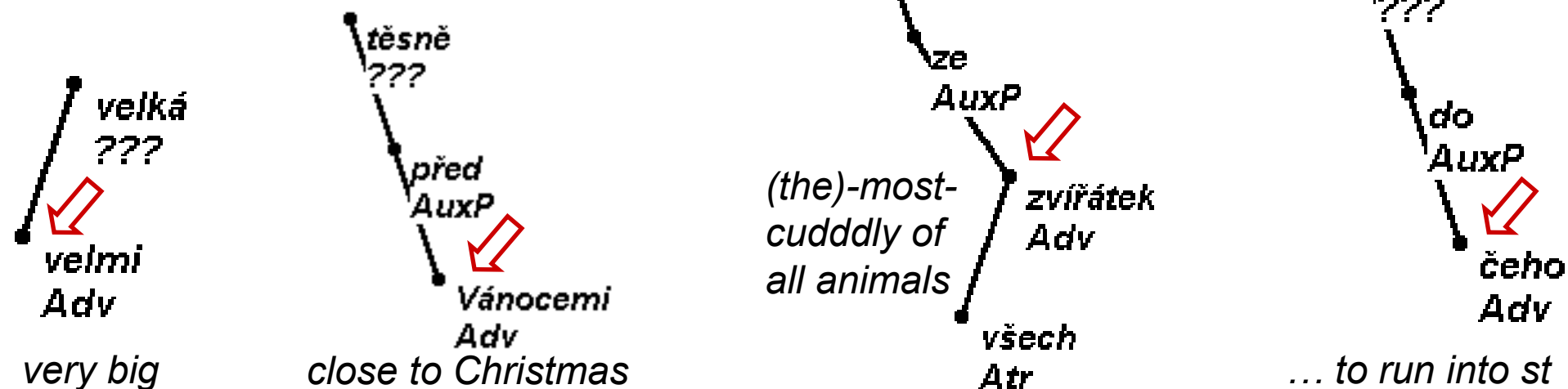**Adverbials** … analytical functions: **Adv**

- express the circumstances and relations, such as
  location, time, manner, comparison, extent, means, cause,
  consequence, regard or aim
- can modify a verb, adjective or adverb
- a **form** of Adv is **not prescribed** by its governing word !
- questions: *where?*, *where to?*, *from where?*, *how long?*, *when?*, *for what purpose?*, *why?*, *how?*
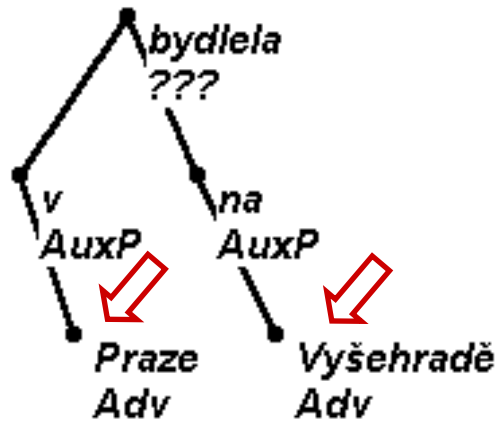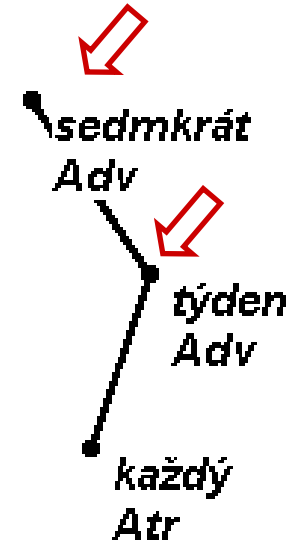- in PDT not further classified

velká
???

velmi
Adv

very big

těsně
???

před
AuxP

Vánocemi
Adv

close to Christmas

nejpřítulnější
???

ze
AuxP

zvířátek
Adv

(the)-most-cudddly of all animals

všech
Atr

narazit
???

do
AuxP

čeho
Adv

… to run into st

Adverbials (cont.)

- several temporal / local adverbials

vstal
???

včera
Adv

ráno
Adv

adverb

brzy
Adv

času
Adv

místního
Atr

(He) got-up yesterday early
morning (of) local time

bydlela
???

v
AuxP

na
AuxP

Praze
Adv

Vyšehradě
Adv

(she) lived in
Prague at
Vyšehrad

sedmkrát
Adv

týden
Adv

každý
Atr
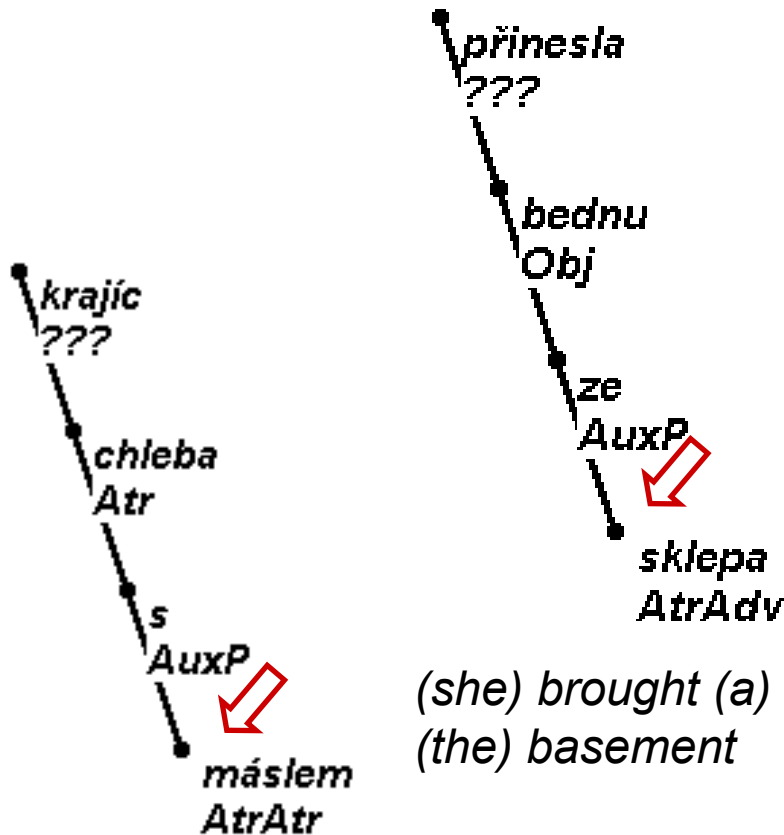
[it usually
happens] seven-
times every week

Combined Functions

analytical functions: **AtrAtr, AtrAdv, AdvAtr, AtrObj, ObjAtr**

krajíc
???

chleba
Atr

s
AuxP

máslem
AtrAtr

slice of bread with butter

přinesla
???

bednu
Obj

ze
AuxP

sklepa
AtrAdv

(she) brought (a) case from (the) basement

nezaplatíme
???

splátku
Obj

mimořádnou
Atr

části
Atr

dluhu
Atr

pojišťovně
AtrObj

naší
Atr

we are not going to pay the extraordinary installment of the part of the debt to our insurance office
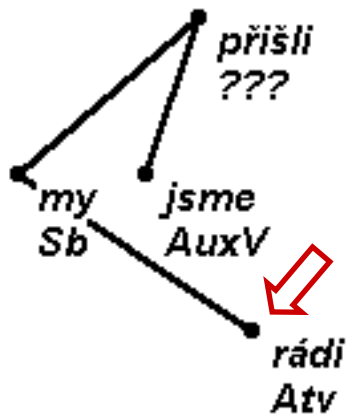
Complement

Complement (verbal attribute) … analytical functions: **Atv, AtvV**
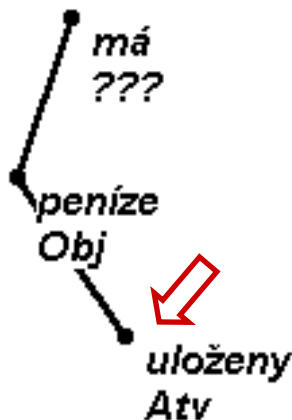
- modifies two sentence members, verb and noun

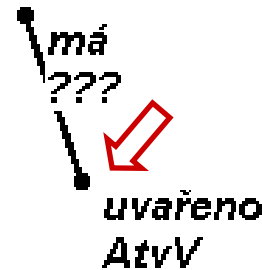*chlapec ležel nemocen, viděl ho nemocného*

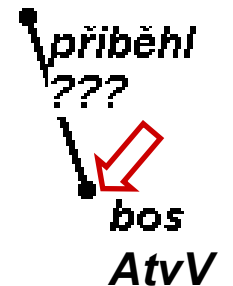*[boy – lay – ill]*   *[he – saw – him – (being) ill]*

we - Aux - came - glad
we were glad to come

money - he-has - deposit
he has put the money on a deposit

she-has – cooked
he has done cooking

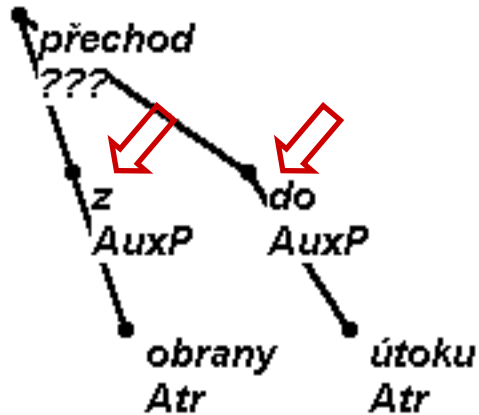*he-arrived - barefooted*

Auxiliary Sentence Members
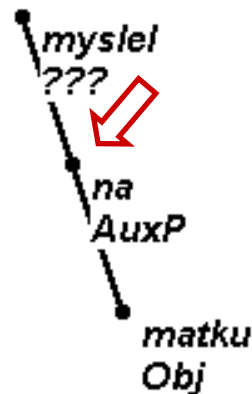
Auxiliary Sentence Members
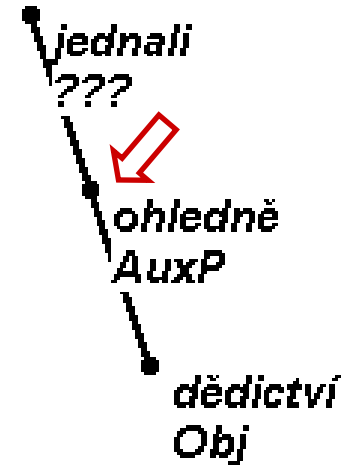
AuxP … prepositions
- parent of a nominal node
- similarly one-word improper prepositions

(the) transition from (a) defense to (an) attack
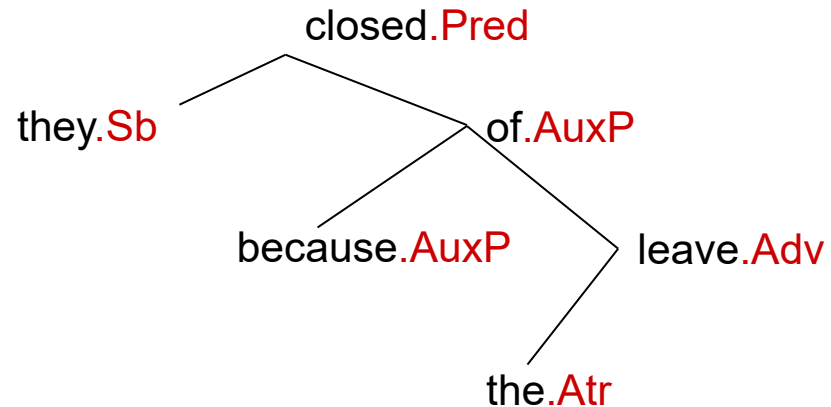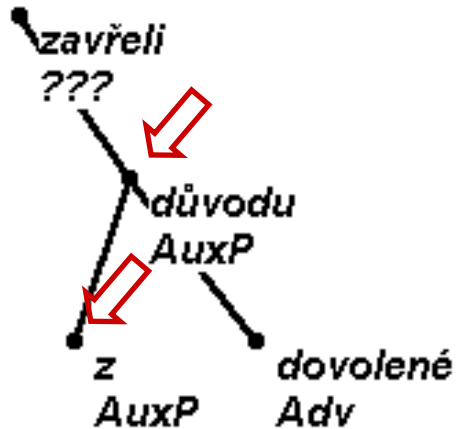
(he) thought of (his) mother

(they) negotiated - about - (the) heritage

Auxiliary Sentence Members

AuxP … secondary prepositions
- consists of several words
- technical solution:
  - the last node … head
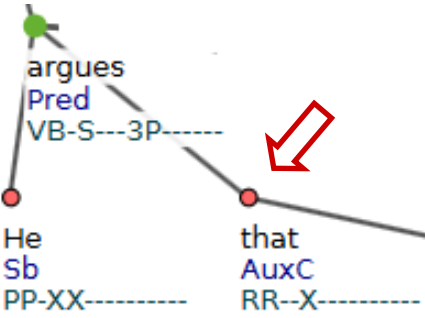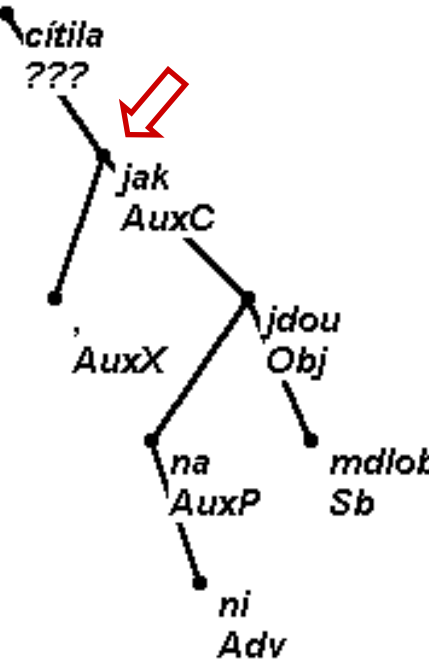 - the remaining words … siblings of the noun governed

zavřeli
???
důvodu
AuxP
z
AuxP
dovolené
Adv

closed.Pred
they.Sb
of.AuxP
because.AuxP
leave.Adv
the.Atr

(they) closed - for - reasons - (of) leave
(they) closed because of the leave

Auxiliary Sentence Members

AuxC … subordinate conjunctions
- subordinate clause introduced by a subordinate conjunction
- conjunction … *root of the subordinate clause*

cítila
???

jak
AuxC

,
AuxX

jdou
Obj

na
AuxP

mdlol
Sb

ni
Adv

… she felt she would go off in a faint

He argues that the dollar is now ``moving sideways …

argues
Pred
VB-S---3P------

He
Sb
PP-XX----------

that
AuxC
RR--X----------

moving
Obj
VpXX---XP------

dollar
Sb
NNXSX----------

is
AuxV
VB-S---3P------

now
Adv
Dg-------1-----

``
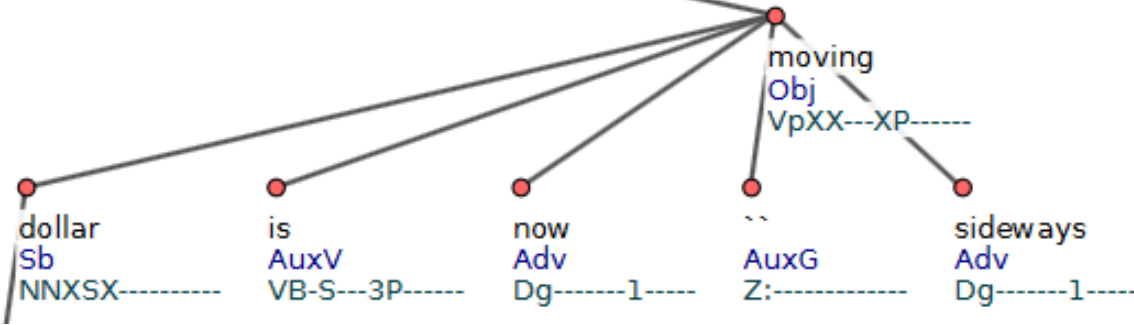AuxG
Z:-------------

sideways
Adv
Dg-------1-----

Auxiliary Sentence Members

AuxC … subordinate conjunctions
- subordinate clause introduced by a subordinate conjunction
- conjunction … *root of the subordinate clause*
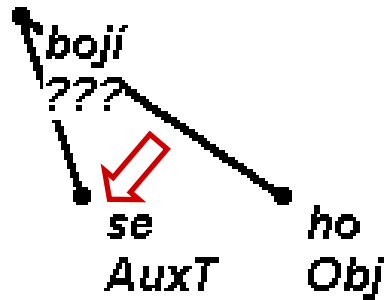 - subordinate conjunction may attach an individual sentence member

```
mluvili
???
   o
   AuxP
              problému
              Obj
   zajímavém
   Atr
           když
           AuxC
   ,          i        důležitém
   AuxX     AuxY      Adv
```

… (they) talked about (an) attractive, even though crucial topic
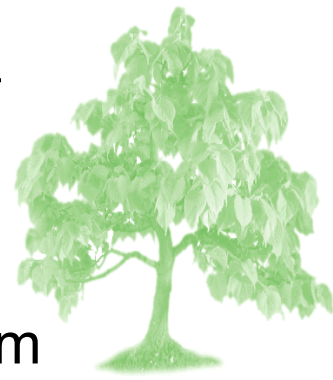
Auxiliary Sentence Members

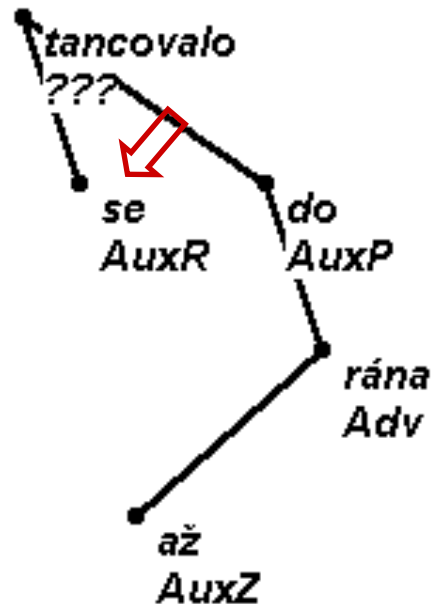***AuxT*** … morphemes ***se, si*** as a part of a reflexive tantum

they are afraid of him

Auxiliary Sentence Members

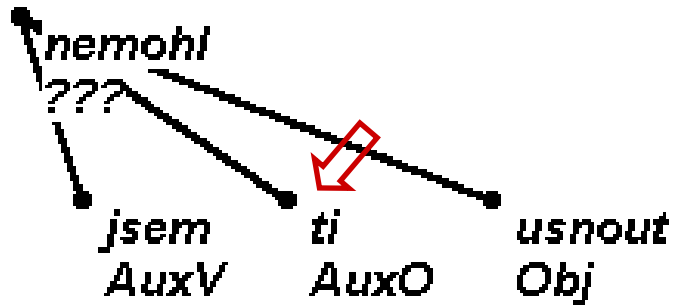***AuxR*** … morphemes ***se*** as a part of a reflexive verbal form

it-danced – Refl – till – to morning
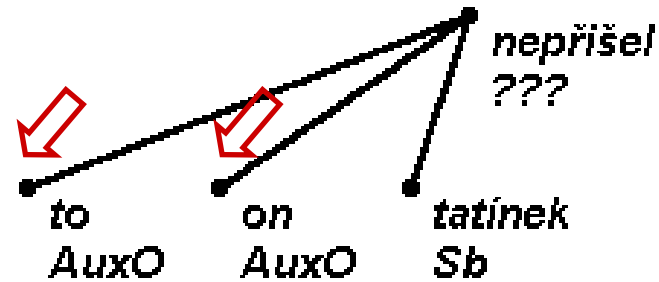*the dancing went on till morning*

Auxiliary Sentence Members

AuxO … emotional particles
mi, vám, si, … to, ono, …

*(I) could not - Aux - **you** - fall asleep*
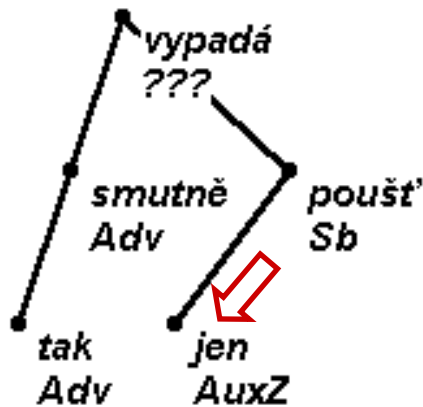*I couldn't fall asleep, you see.*

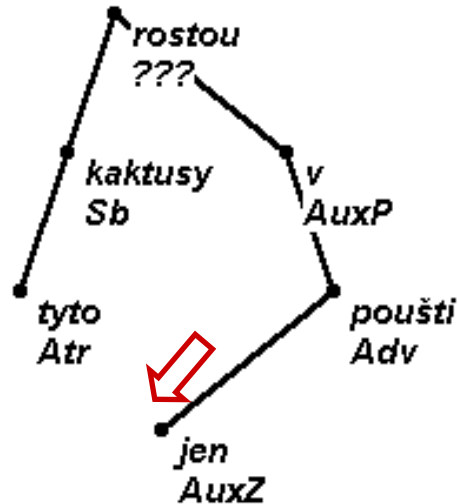*obviously, father failed to come*

Auxiliary Sentence Members
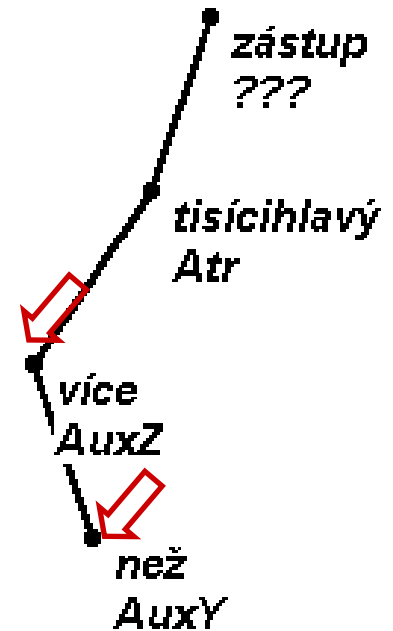
AuxZ … emphasizing words (modifying a sentence member)

e.g., *asi* [maybe, approx.], *dokonce* [even, as far as], *hlavně* [mainly], *ještě [moreover], již* [yet], *leda* [only], *především* [most of all, most probably], *zvláště* [especially], …

vypadá
???

smutně
Adv

poušť
Sb

tak
Adv

jen
AuxZ

*only a desert can look
as sad as that*

rostou
???

kaktusy
Sb

v
AuxP

tyto
Atr

poušti
Adv

jen
AuxZ

*these cactuses grow
only in (a) desert*

zástup
???

tisícihlavý
Atr

více
AuxZ

než
AuxY

*more than thousand-
headed crowd*

Auxiliary Sentence Members

AuxY … particles modifying the whole sentence;

#
AuxS

nepřišel
Pred

.
AuxK

Prý
AuxY

nikdo
Sb

no one has come, they say

vyhrajeme
???

at'
AuxC

už
AuxY

stane
Adv

,
AuxX

se
AuxT

cokoli
Sb

whatever may happen, we shall win

Auxiliary Sentence Members

AuxY … particles modifying the whole sentence
parts of compound conjunctions
comparison
(and others)

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