



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:

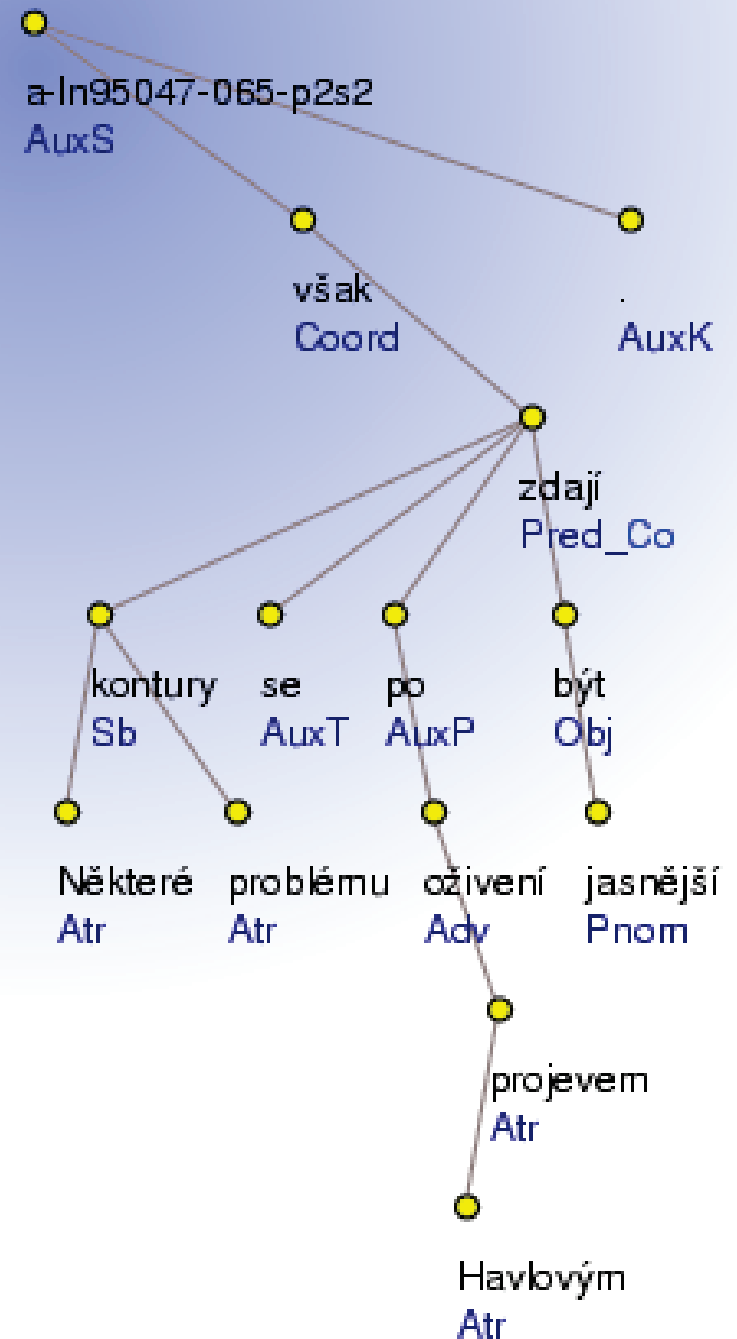
- to describe the structure of the sentence and
- to denote the type of relations between "words"
- dependency tree **!!! tree structure ~ XML structure !!!**
- one token from m-layer ~ one node incl. prepositions, punctuation ... plus technical root
- relations ~ edges
 - dependency, coordination, punctuation, ...
 - 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.]



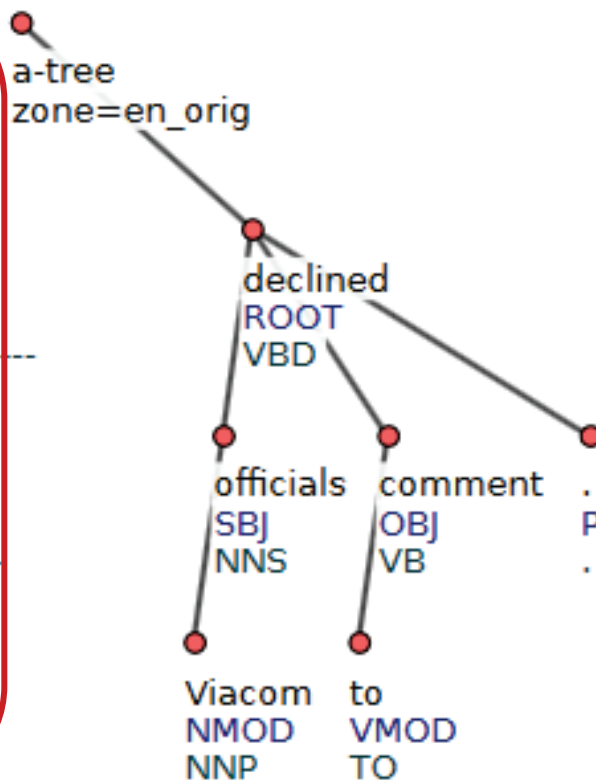
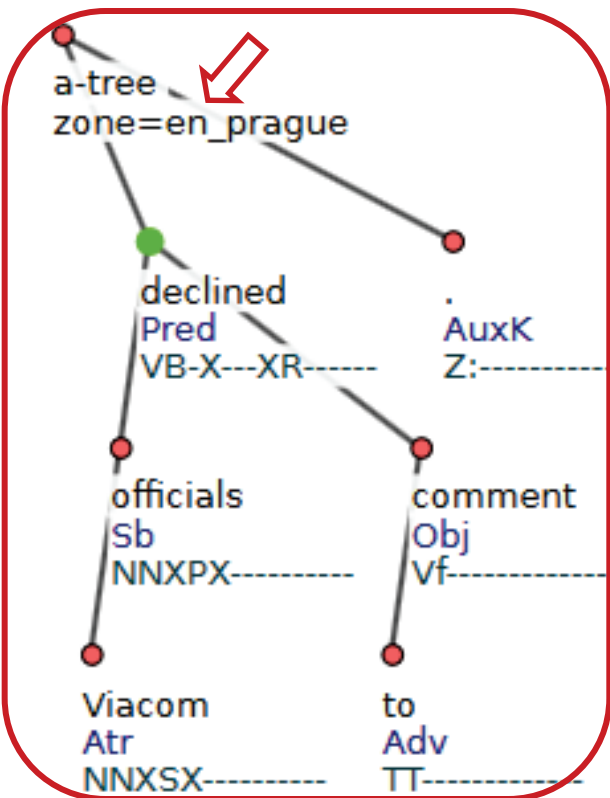


HamleDT: a-layer

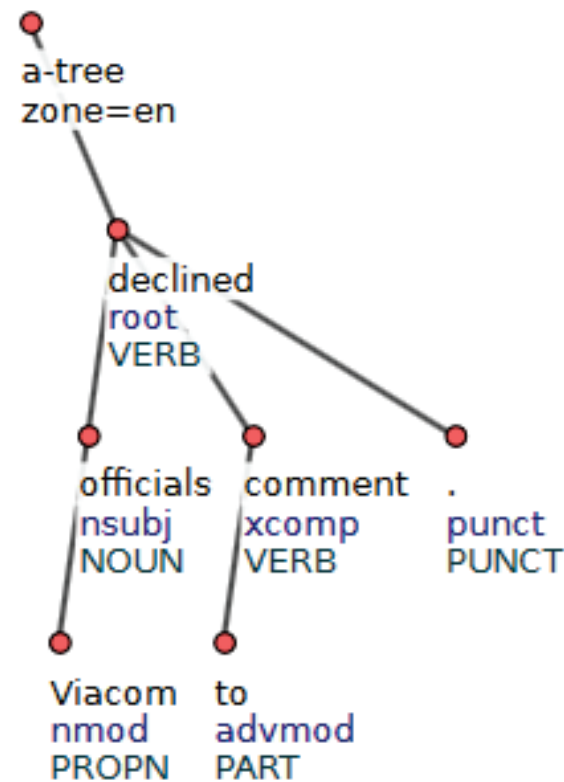
https://lindat.mff.cuni.cz/services/pmltq/#!/treebank/hamledt_en/help

(login required for English, free whenever possible)

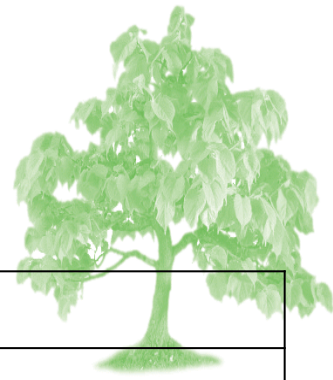
PDT style



Universal Dependencies

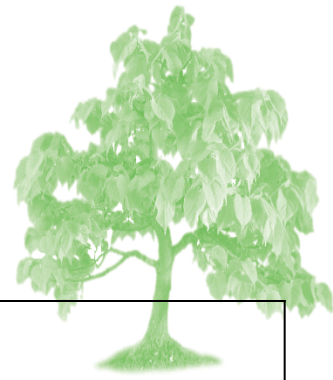


A-layer: attributes of a node (non-root)



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



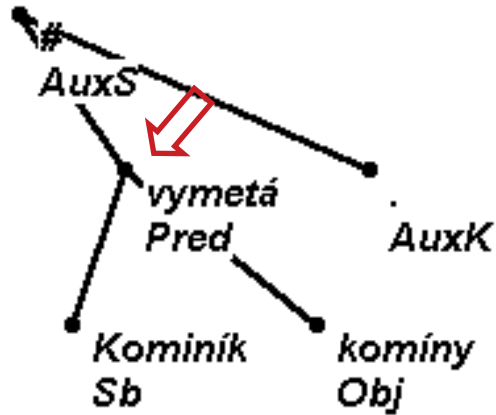
Predicate ... analytical functions: **Pred, Pnom, AuxV**

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

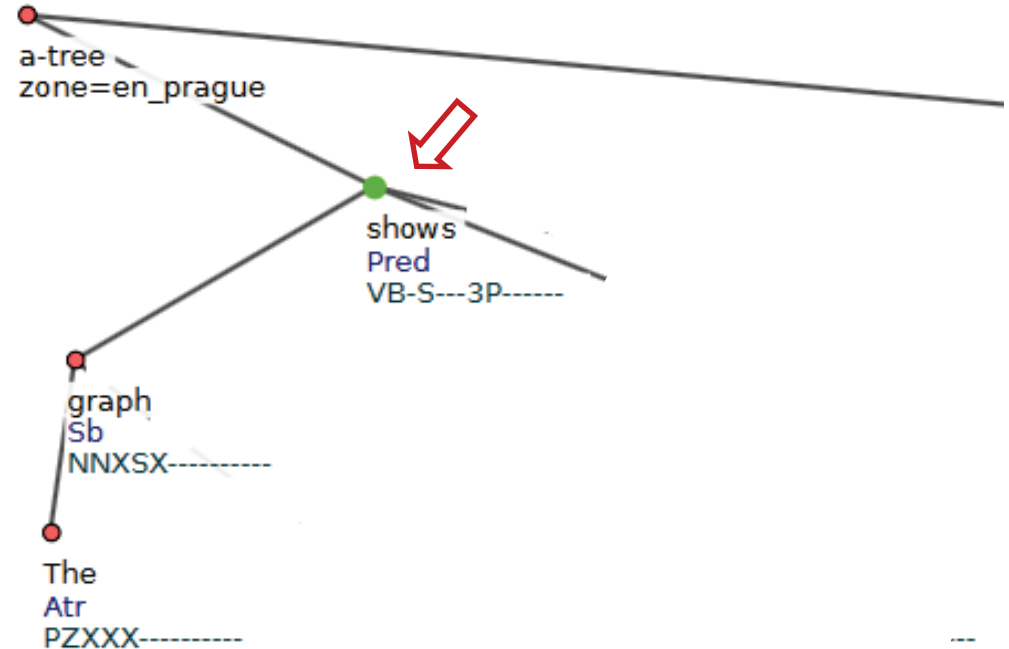


Predicate: Dependency Structure

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



Chimneysweep sweeps chimneys.



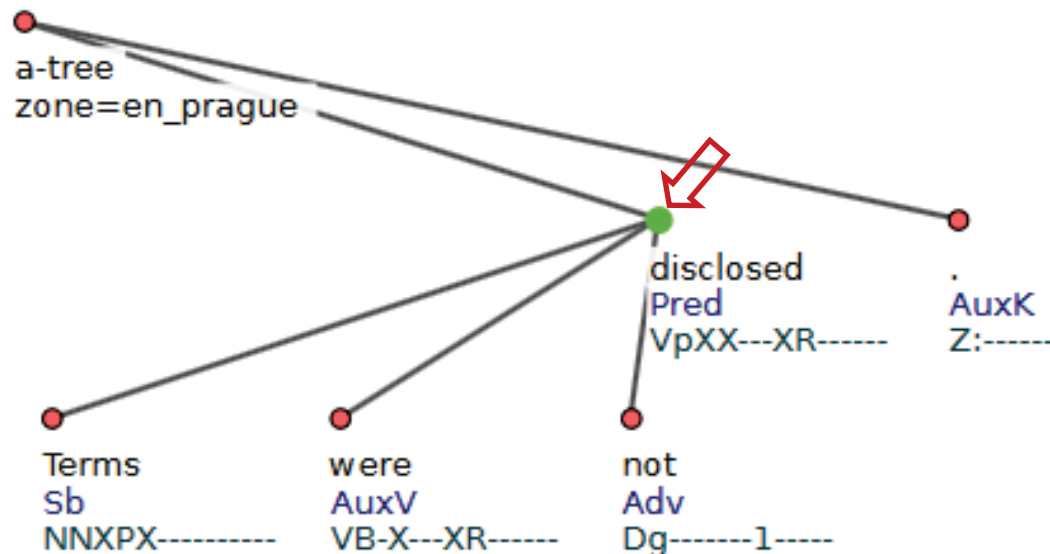
The graph shows ...



Predicate: Dependency Structure

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

passive sentence



Terms were not disclosed.

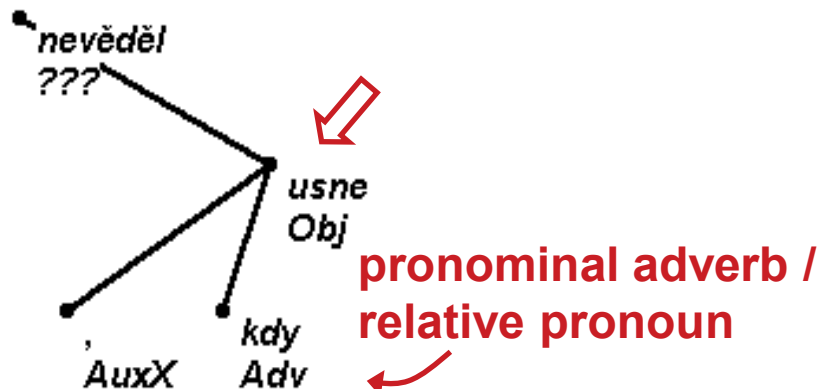


Predicate: Dependency Structure

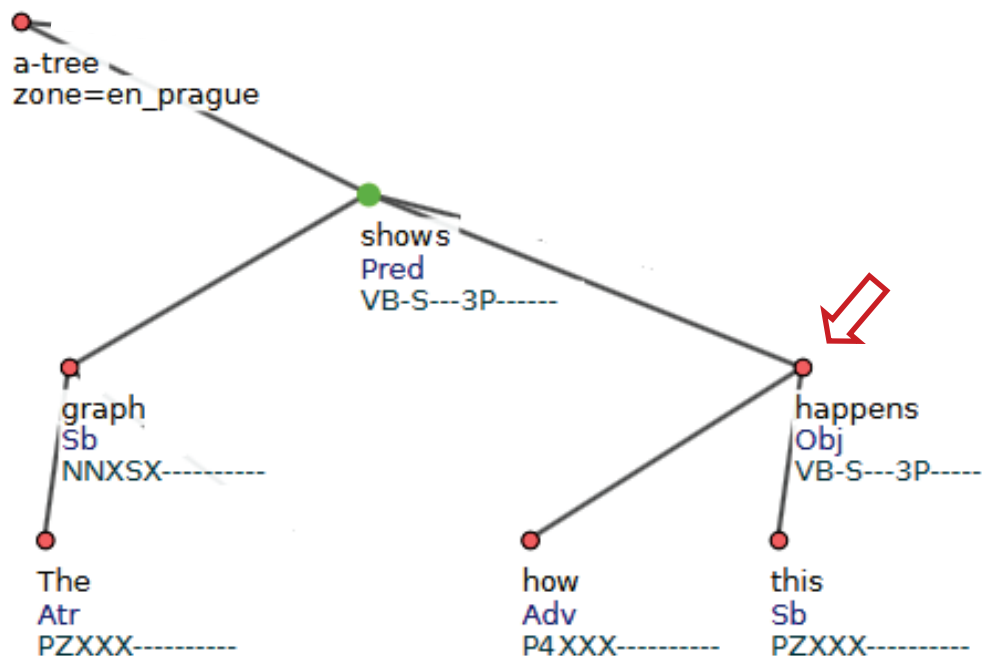
- predicate of a **main clause**
 - a child of the root, analytical function ... **Pred**

BUT:

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



did-not-know when falls-asleep





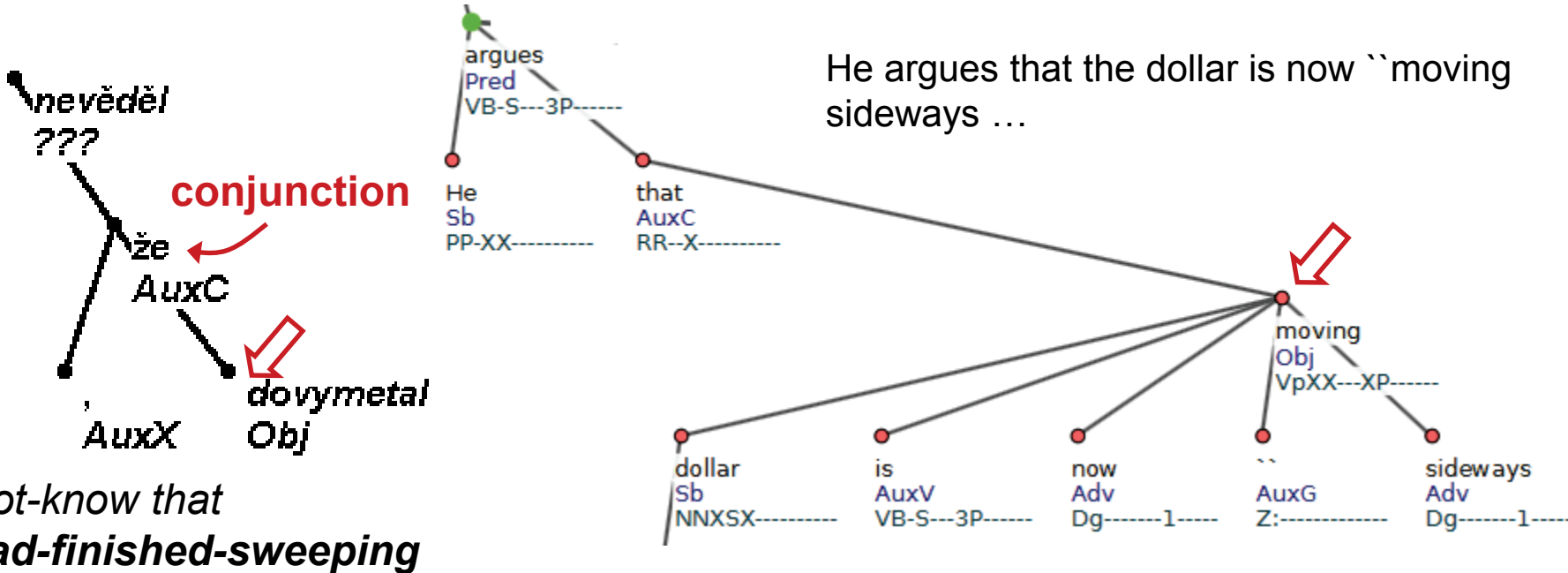
Predicate: Dependency Structure

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

BUT:

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

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

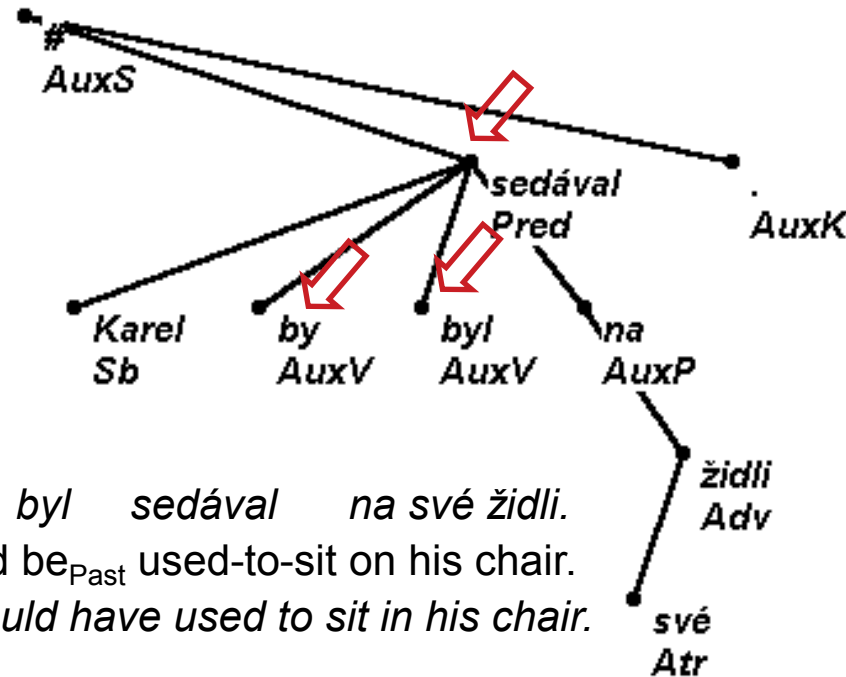
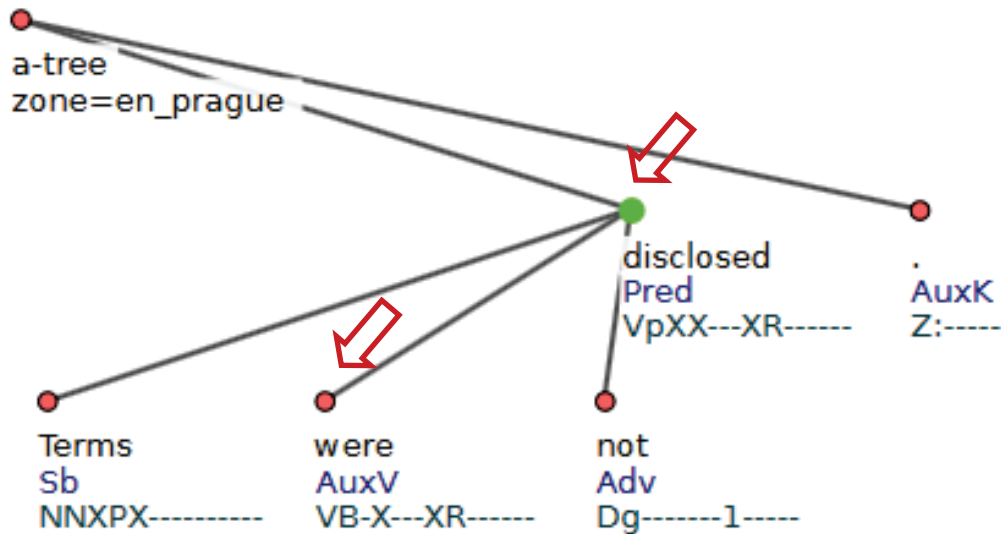


did-not-know that he-had-finished-sweeping



Predicate: Analytical Verb Forms

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



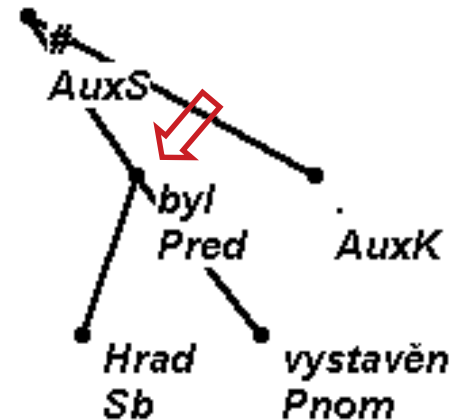
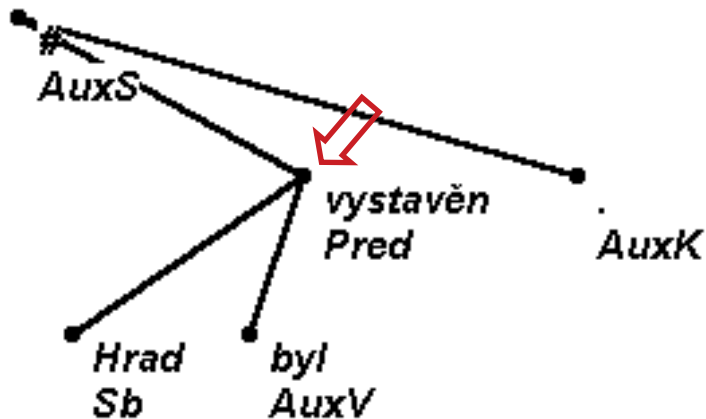
He argues that the dollar is now ``moving sideways ...

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 Forms

- 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.*

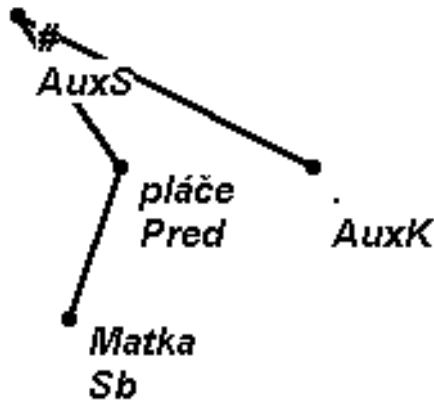


Predicate: Three Different Types

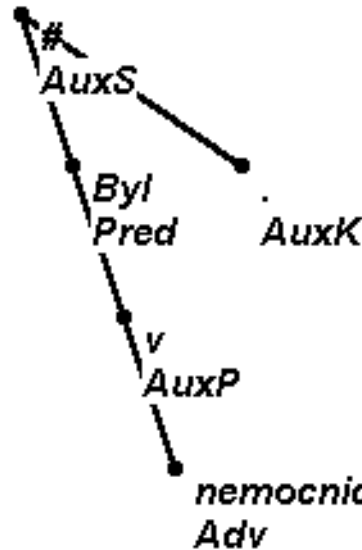
1. simple verbal predicate (incl. analytical verbal forms)

typically a finite verb form ... **(AuxV +) Pred**

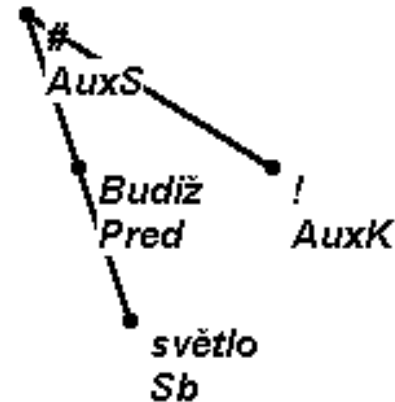
lexical or *být* [to be] (existential, substitute), incl. *lze, nelze* [is (not) possible]



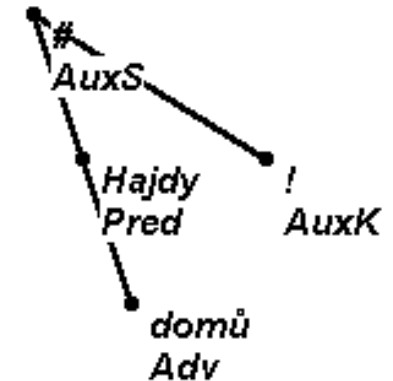
Mother cries.



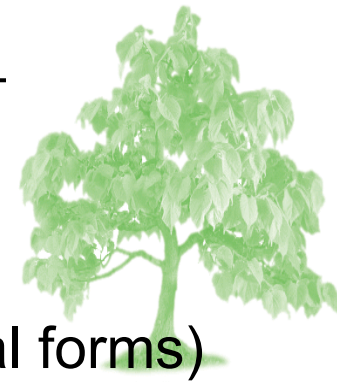
(He) was in (a) hospital.



Let there be light.



Hurrah for home!



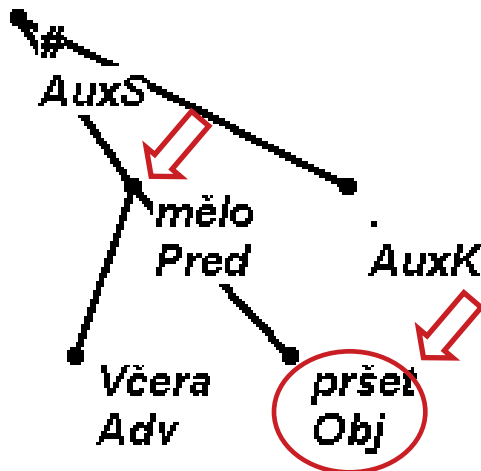
Predicate: Three Different Types

1. simple verbal predicate (incl. compound/analytical verbal forms)

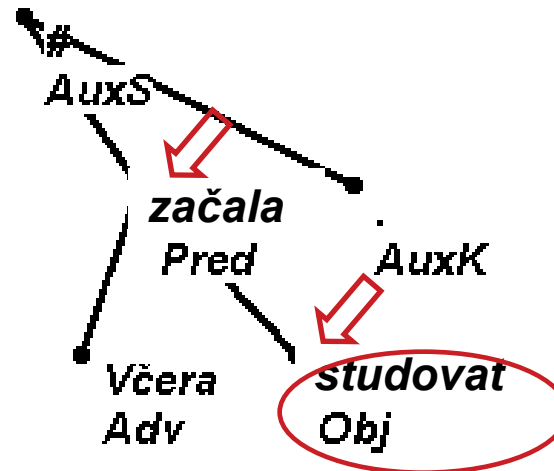
2. **compound verbal predicate ... Pred + Obj**

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

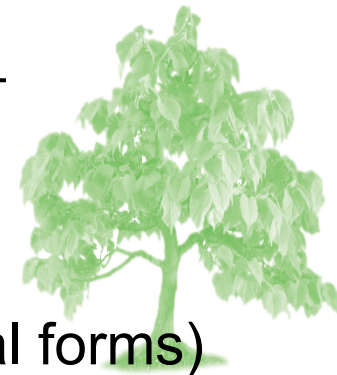
mít [should], muset [must], moci [can], chtít [to want], dát_se [to be possible], smět [may]; začít [to begin], skončit [to end], ...



Yesterday (it) should rain.

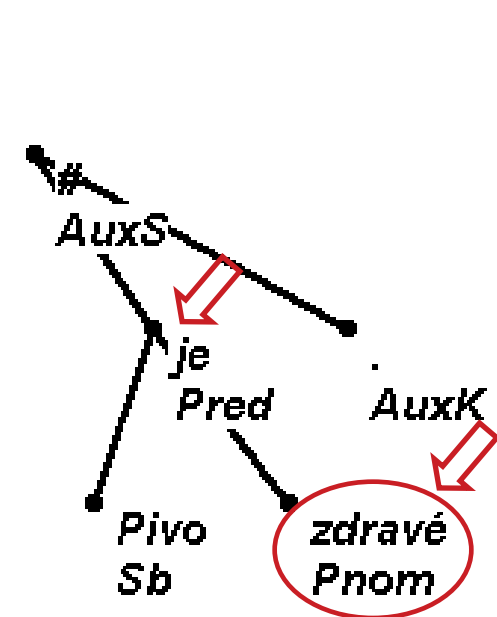


Yesterday (she) began (to) study.

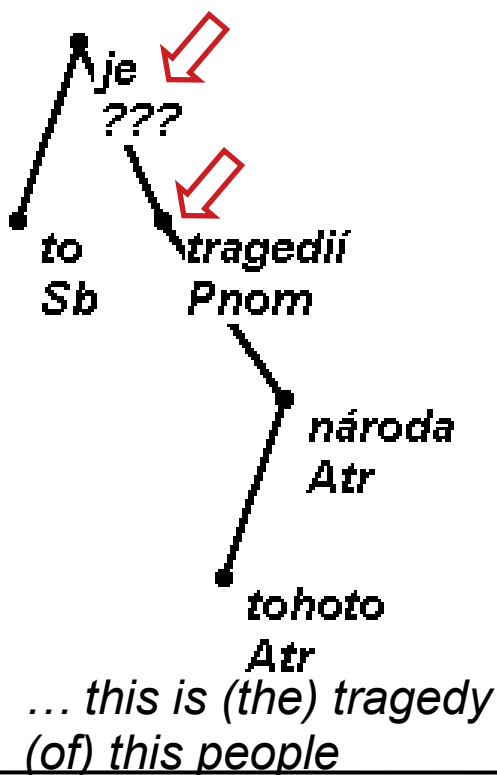


Predicate: Three Different Types

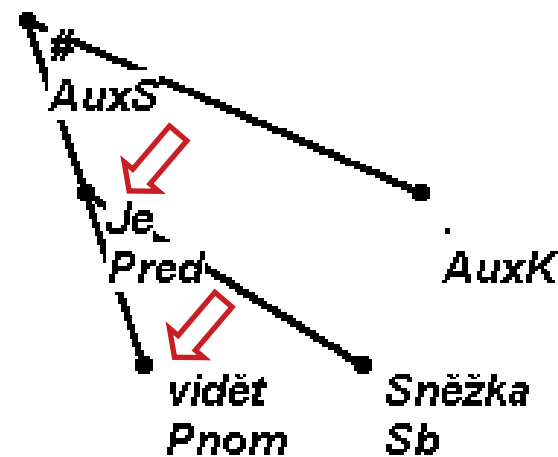
1. simple verbal predicate (incl. compound/analytical verbal forms)
2. compound verbal predicate
3. **verbal nominal predicate** (přísudek slovesně jmenný) ... **Pred + Pnom**
finite form of copula (spona) *být* + nominal part



Beer is healthy.



*... this is (the) tragedy
(of) this people*



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

Subject



Subject ... analytical function: **Sb**

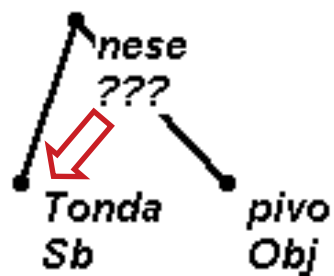
- any construction answering the question *who (what)*
- dependent on a predicate
- possible forms of subject (see the data)



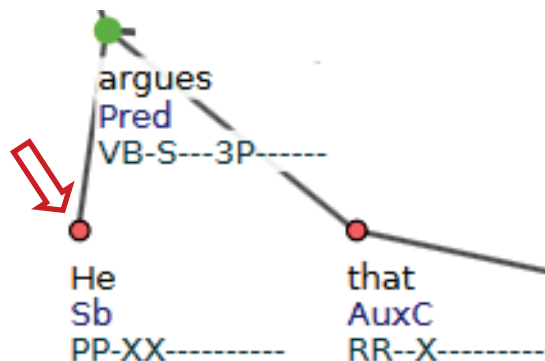
Subject

Subject ... analytical function: **Sb**

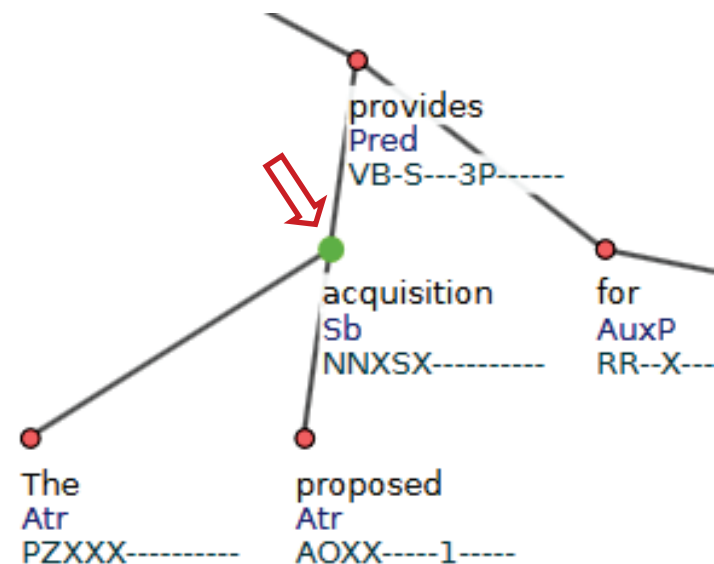
- any construction answering the question *who (what)*
- dependent on a predicate
- possible forms of subject (see the data)



Tonda nese pivo.
 Tony – brings – beer



He argues that...



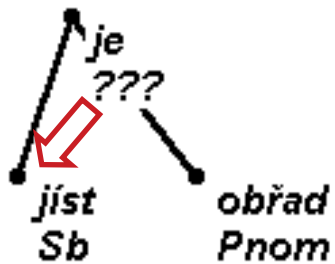
The proposed acquisition provides for ...

Subject

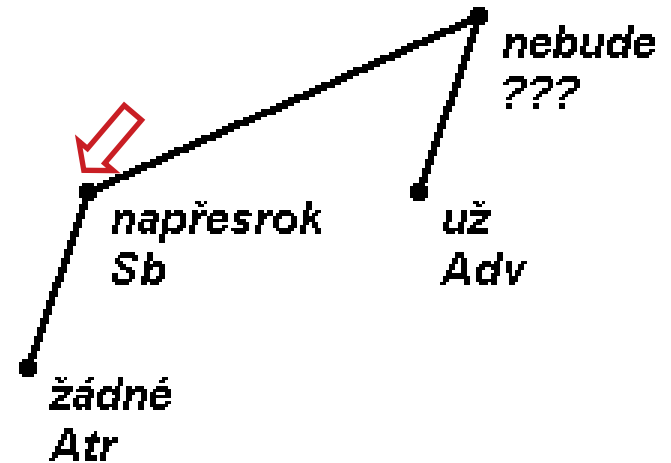


Subject ... analytical function: **Sb**

- 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



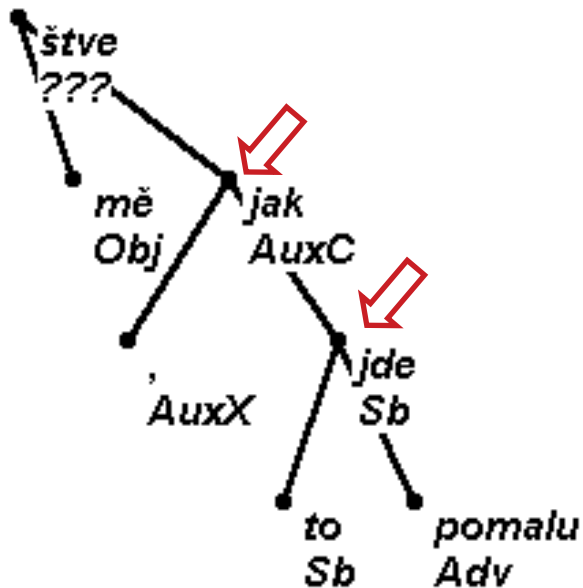
žádné napřesrok už nebude
no – next-year – anymore – will-not-be

Subject (cont.)



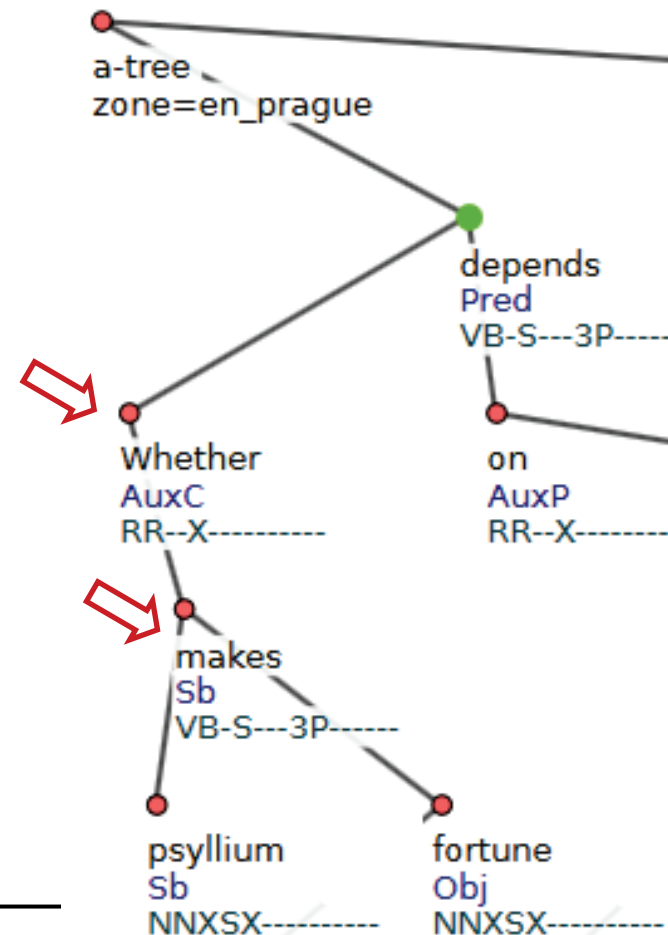
Subject ... analytical function: **Sb**

- 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

PDT: a-layer

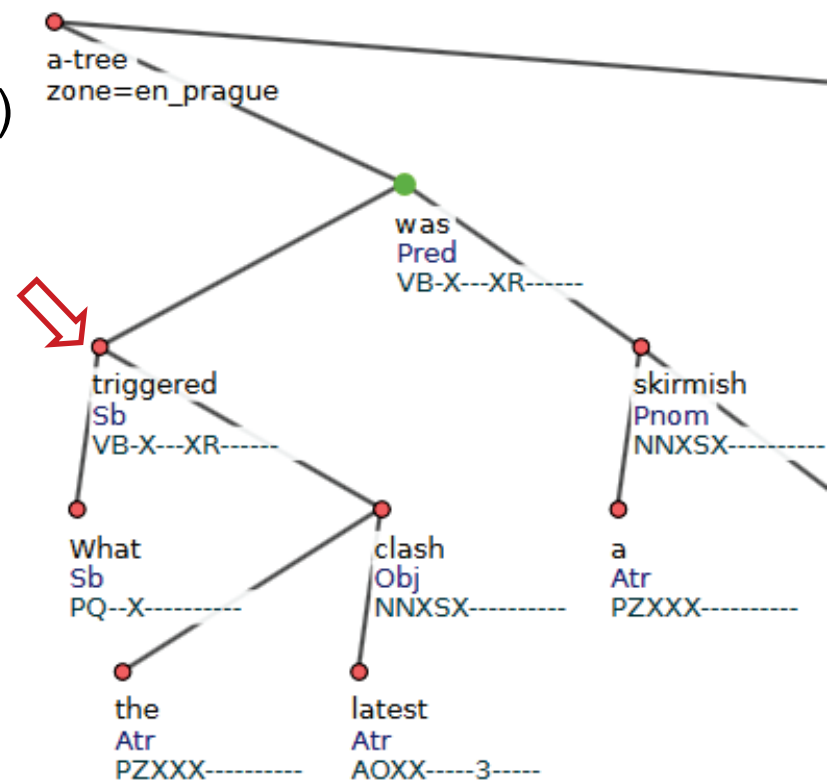
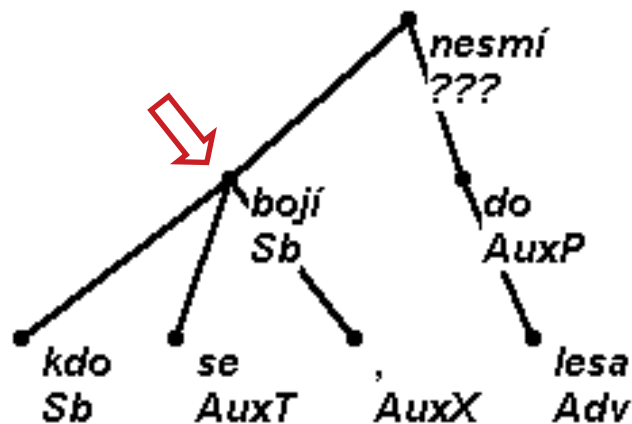




Subject (cont.)

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 ...

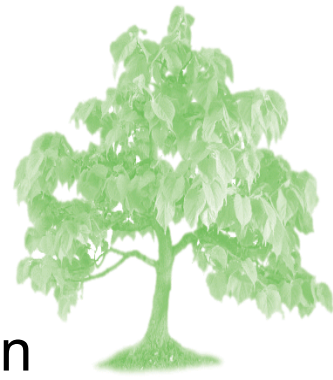
Attribute



Attribute ... analytical functions: **Atr**

- modifies a noun (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

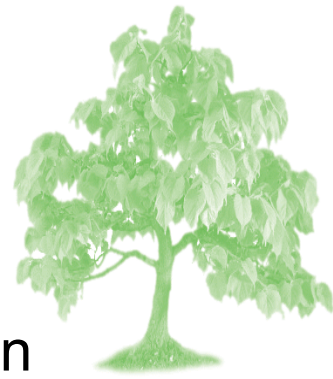
red letters

maminka
???

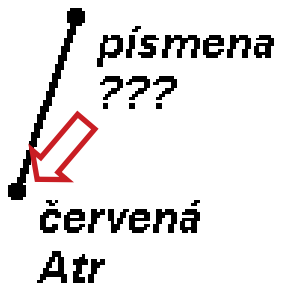
nebožka
Atr

the dead mother

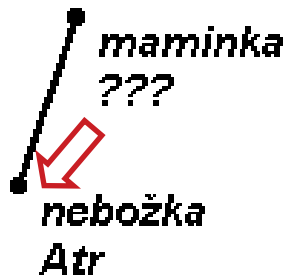
Attribute: Agreement



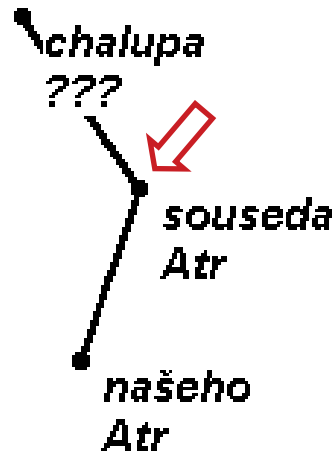
- agreeing attribute ... the same case as its governing noun
Adj ... case, number, gender
BUT: *kluku ušatá*
- non-agreeing attribute



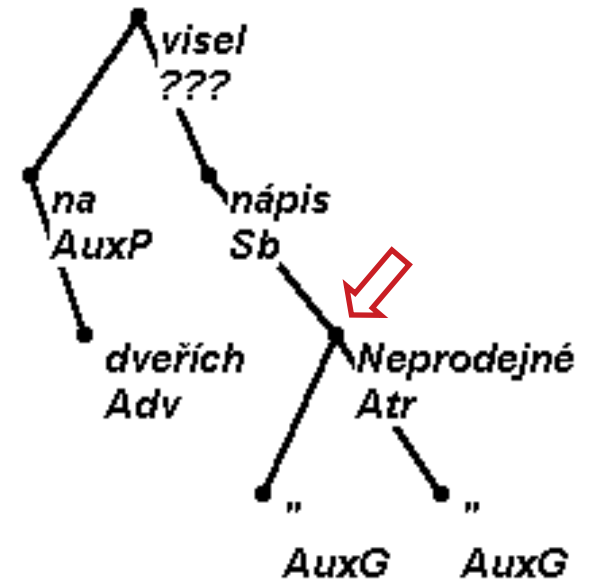
red letters



the dead mother

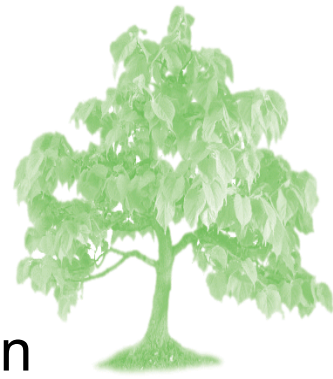


cottage of our neighbor

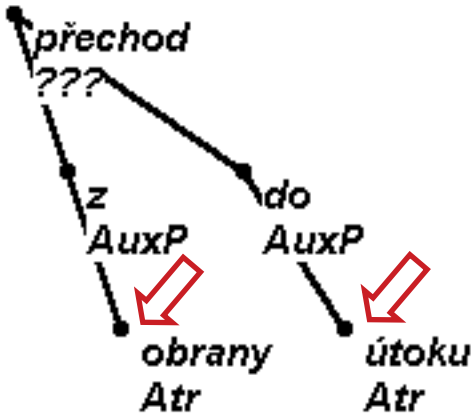


an inscription "Not for sale" was hanging on the door

Attribute: Agreement



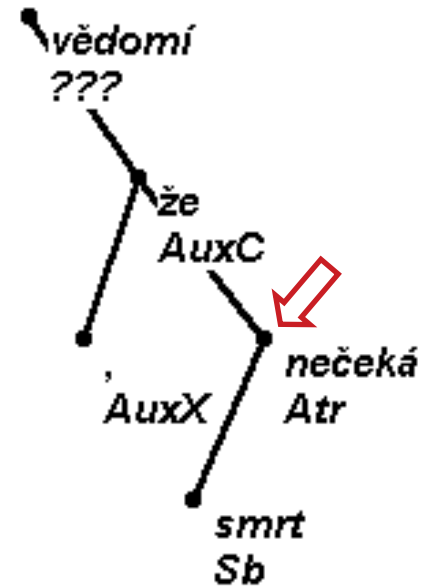
- agreeing attribute ... the same case as its governing noun
Adj ... case, number, gender
BUT: *kluku ušatá*
- non-agreeing attribute



passing from defense to attack



kitchen like (a) palm



knowledge that death does-not-wait

Attribute (cont.)

Atr ... ALSO "technical" solution for

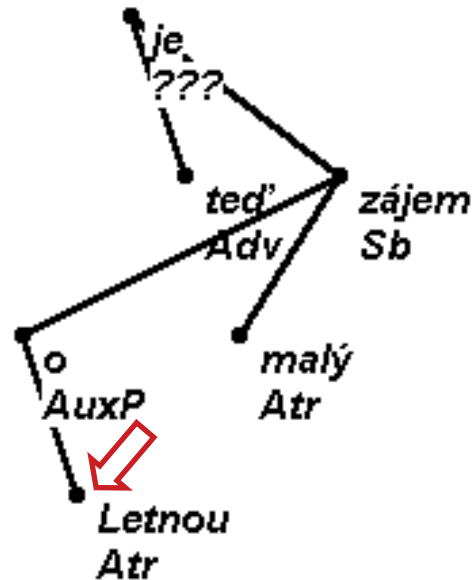
- addresses
- names of persons and institutions
- foreign words
- expressions with numerals
- figures



Attribute: Non-projectivity



Czech:



there is a little interest in Letná

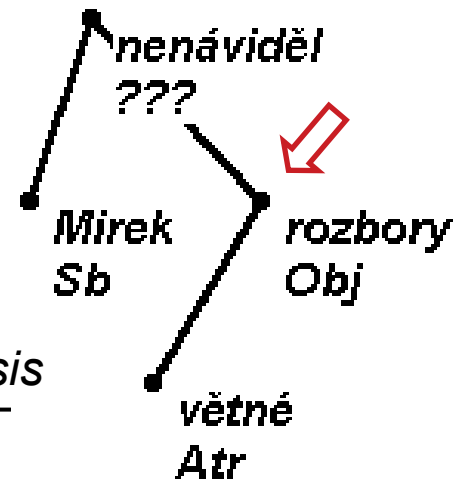
Question: *Is there any non-projectivity in English HamleDT (Prague-style) ???*

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



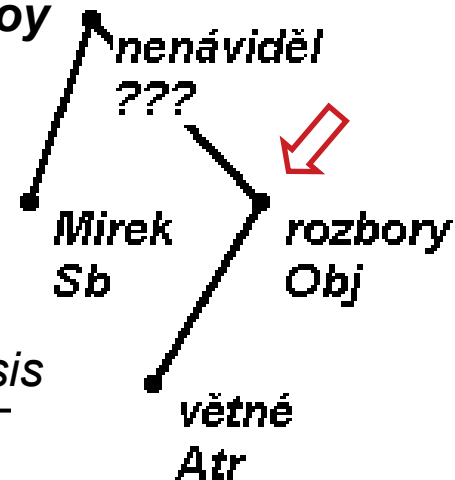
Object

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)



Mirek hated (a) sentence analysis



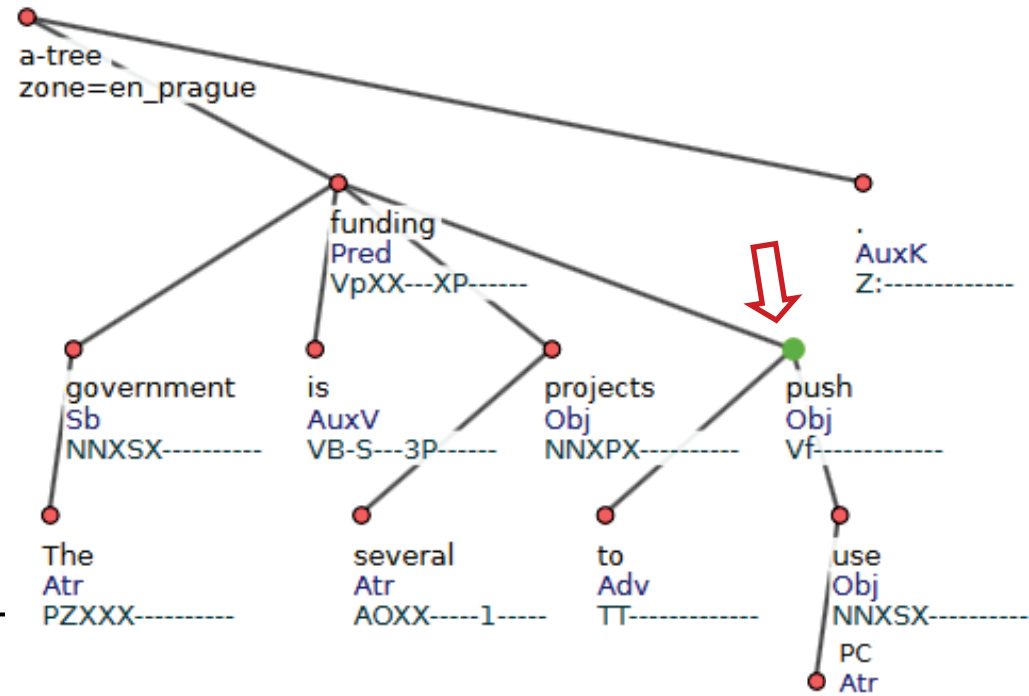
Object

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

- incl. **infinitive constructions**





Object

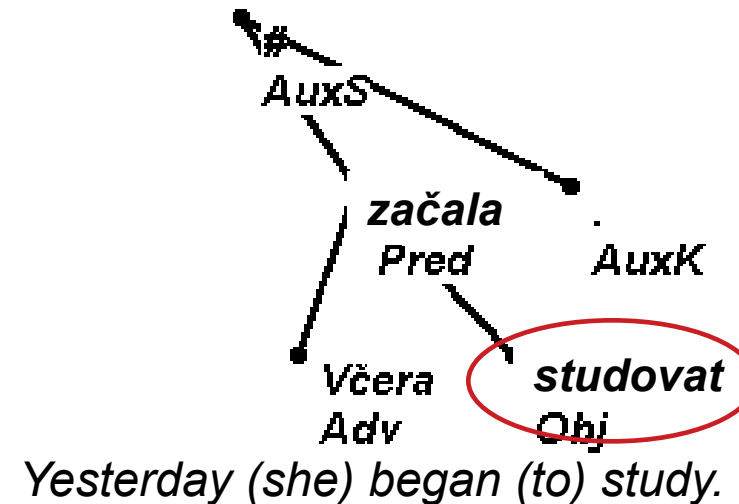
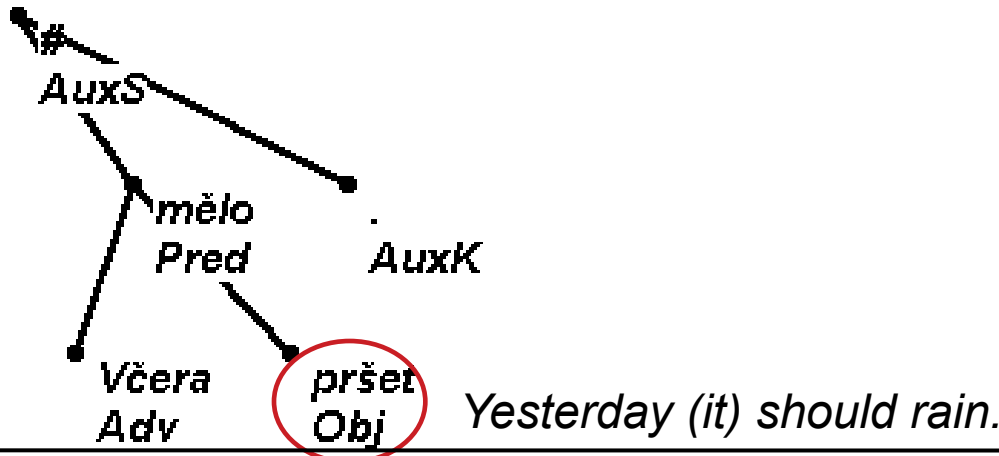
Object ... analytical functions: **Obj**

- 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], ...



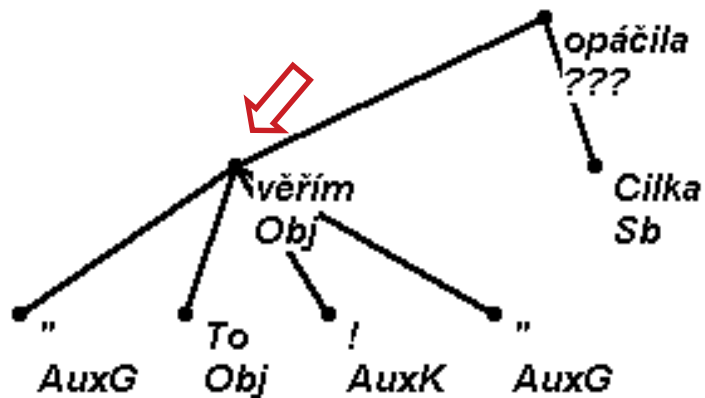
Object



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

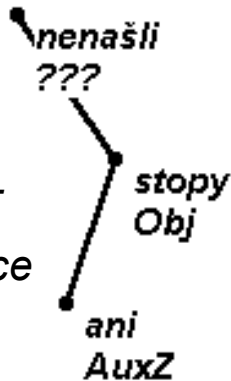


see the data and/or the manual 😊

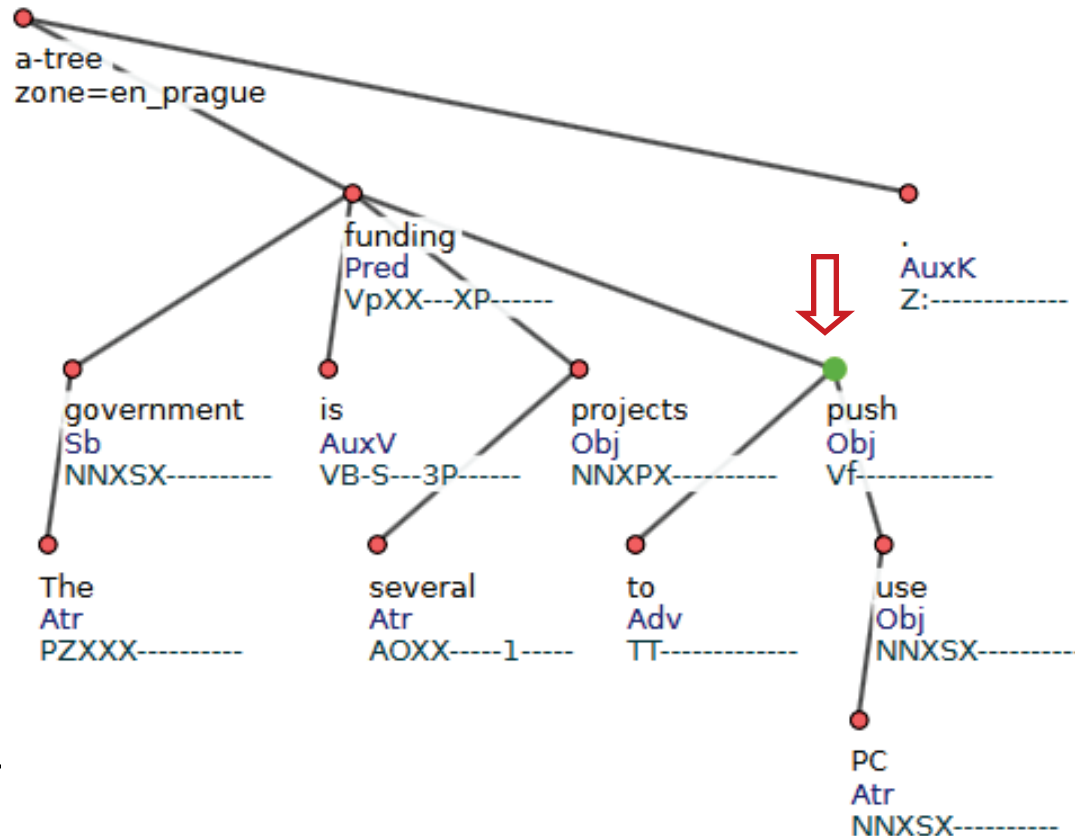
- typically: noun in accusative/dative/genitive/instrumental case
- preposition case
- infinitive, dependent clause



it-wanted – Refl – me – to-sleep
... I felt sleepy



... (they) failed-
-to-find any trace



Adverbials



Adverbials ... analytical functions: **Adv**

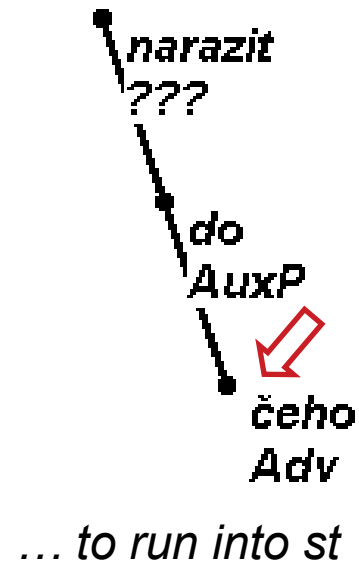
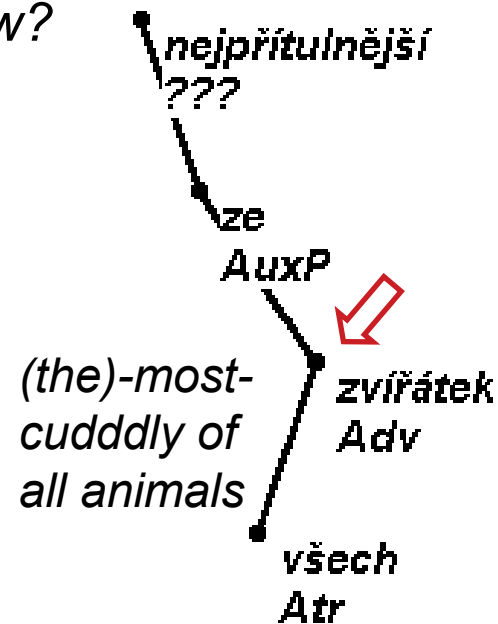
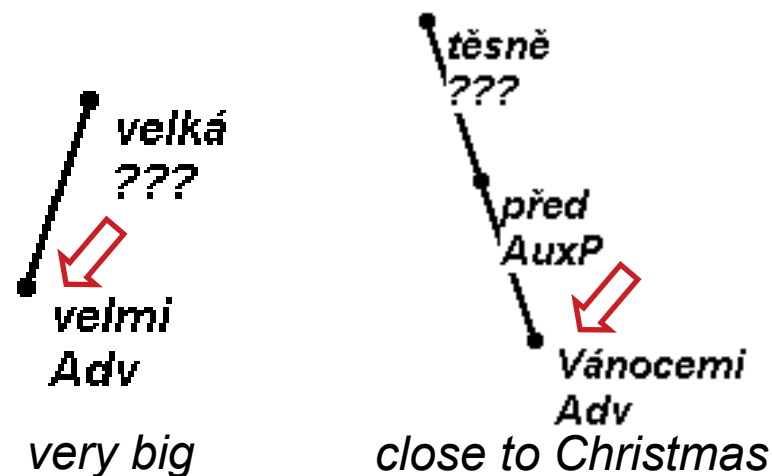
- express the circumstances and relations, such as location, time, manner, comparison, extent, means, cause, consequence, regard or aim
- can modify a verb, adjective or adverb

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



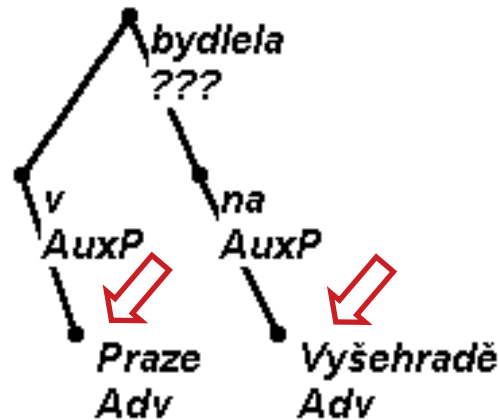
Adverbials (cont.)



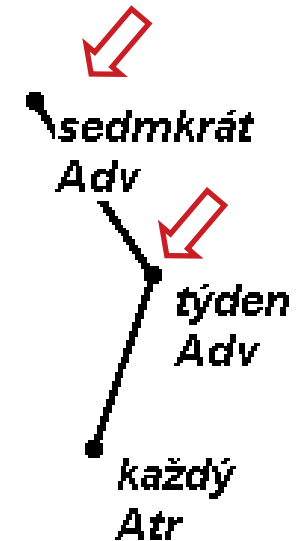
- several temporal / local adverbials



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



(she) lived in Prague at Vyšehrad

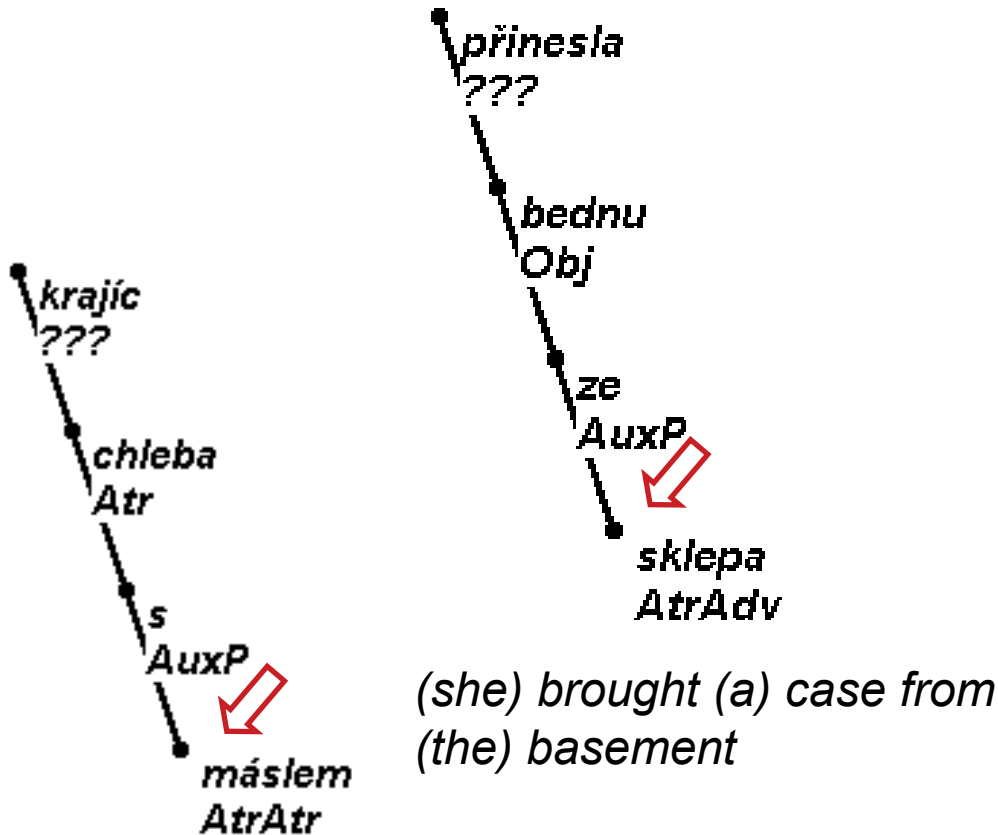


[it usually happens] seven-times every week

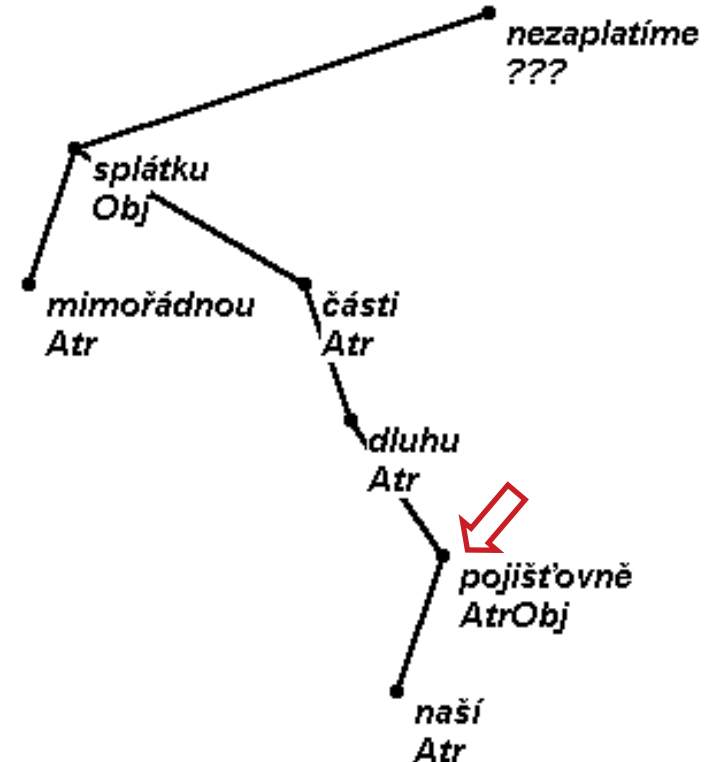
Combined Functions



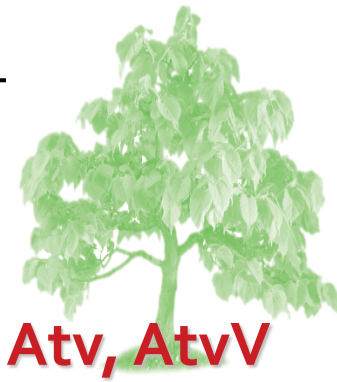
analytical functions: **AtrAtr, AtrAdv, AdvAtr, AtrObj, ObjAtr**



slice of bread with butter



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



Complement

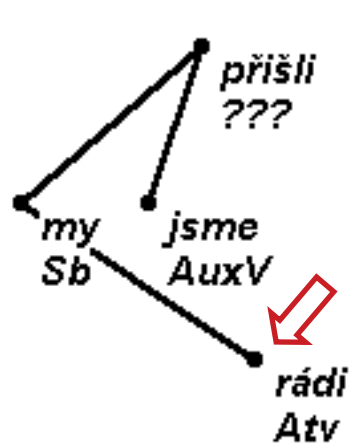
Complement (verbal attribute) ... analytical functions: **Atv, AtvV**

- modifies two sentence members, verb and noun

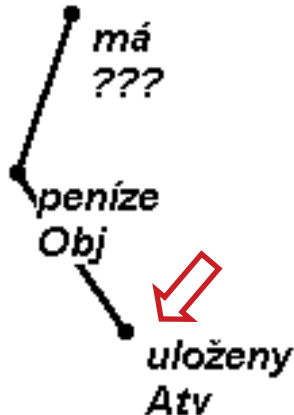
chlapec ležel nemocen, viděl ho nemocného

[boy – lay – ill]

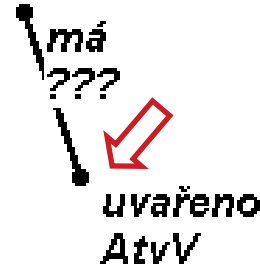
[he – saw – him – (being) ill]



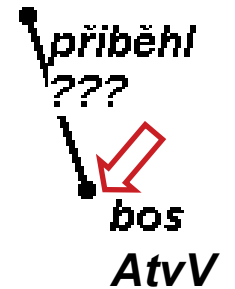
we - Aux - came - glad
we were glad to come



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

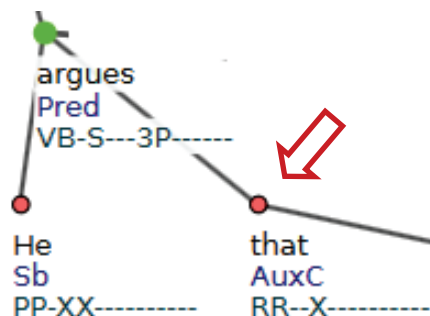
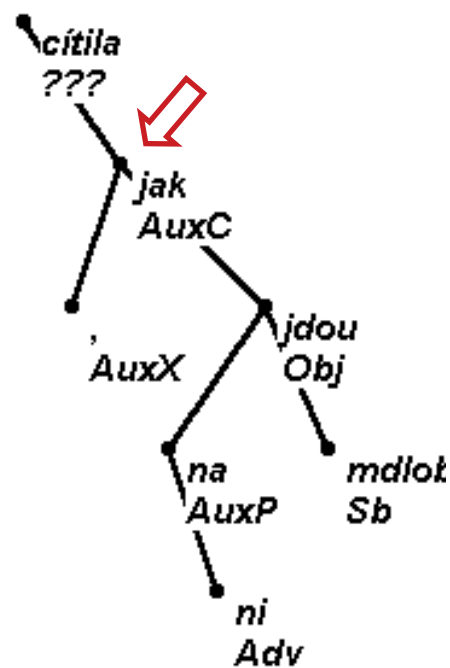




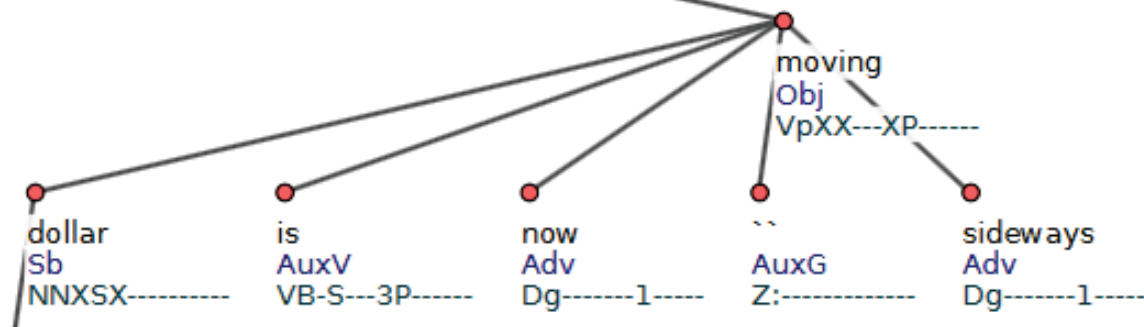
Auxiliary Sentence Members

AuxC ... subordinate conjunctions

- subordinate clause introduced by a subordinate conjunction
- conjunction ... *root of the subordinate clause*



He argues that the dollar is now "moving sideways ...



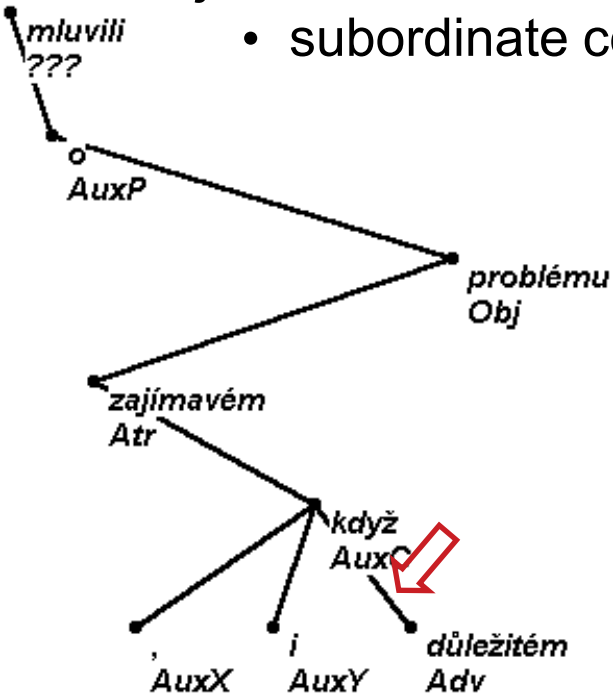
... she felt she would go off in a faint

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



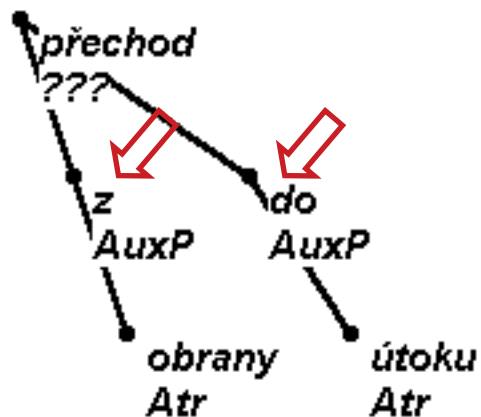
... (they) talked about (an) attractive, even though crucial topic



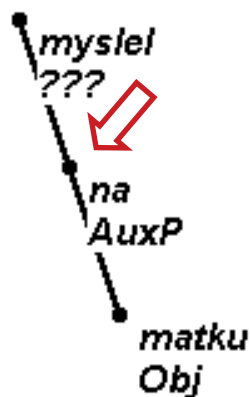
Auxiliary Sentence Members

AuxP ... prepositions

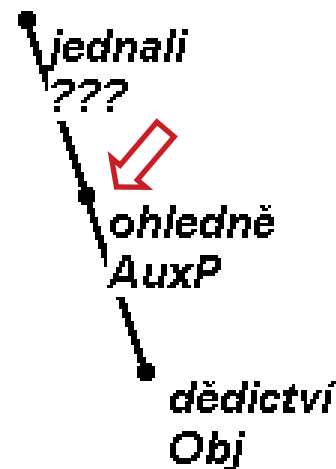
- parent of a nominal node
- similarly one-word improper prepositions



(the) transition from (a) defense to (an) attack



(he) thought of (his) mother



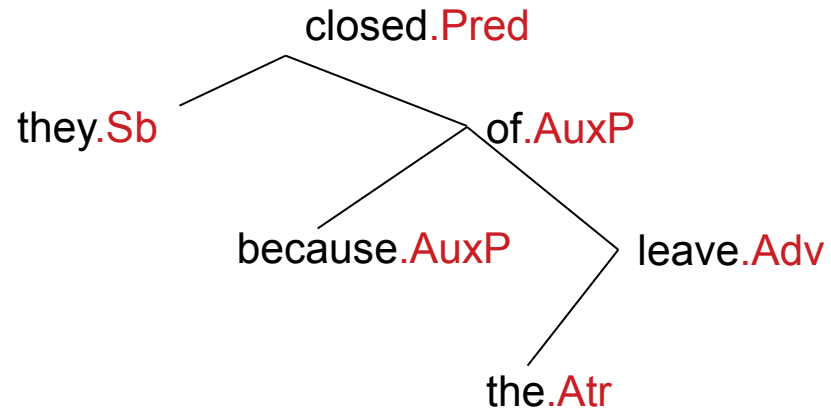
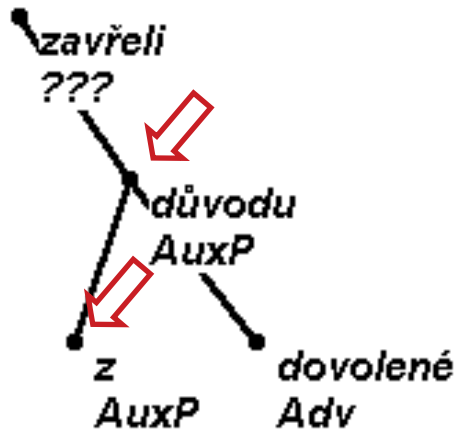
(they) negotiated - about - (the) heritage

Auxiliary Sentence Members



AuxP ... secondary prepositions

- consists of several words
- technical solution:
 - the last node ... head
 - the remaining words ... siblings of the noun governed



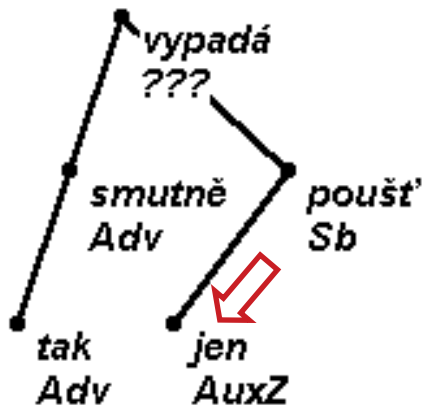
(they) closed - for - reasons - (of) leave
(they) closed because of the leave

Auxiliary Sentence Members

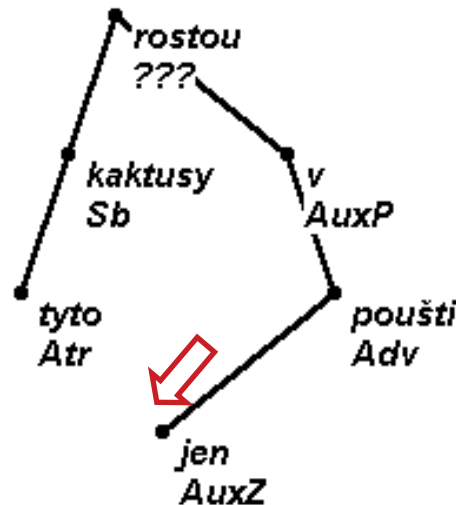


AuxZ ... emphasizing words (modifying a sentence member)

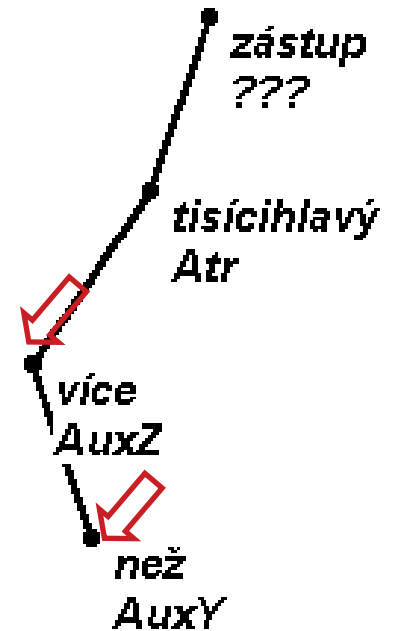
e.g., *asi* [maybe, approx.], *dokonce* [even, as far as], *hlavně* [mainly], *ještě* [moreover], *již* [yet], *leda* [only], *především* [most of all, most probably], *zvláště* [especially], ...



only a desert can look as sad as that



these cactuses grow only in (a) desert



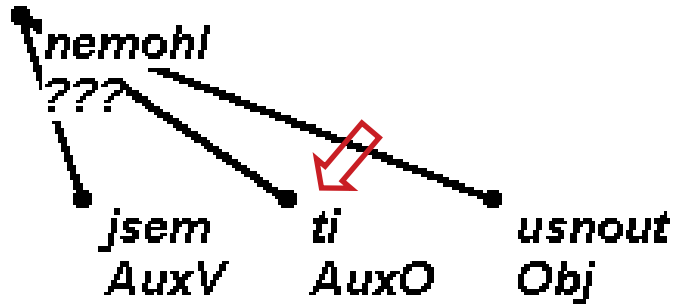
more than thousand-headed crowd

Auxiliary Sentence Members

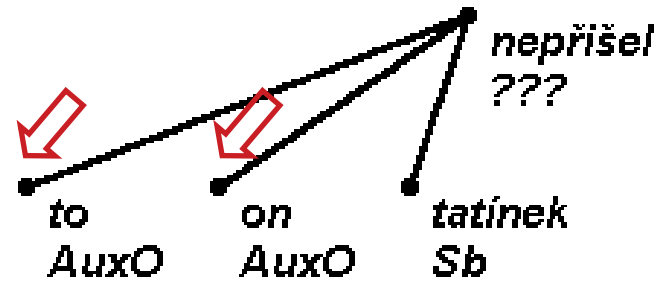


AuxO ... emotional particles

mi, vám, si, ... to, ono, ...



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

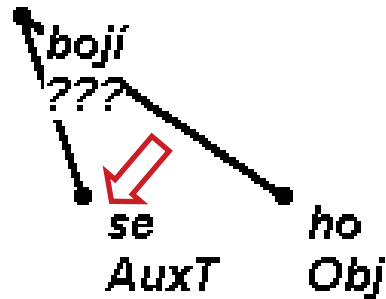


obviously, father failed to come

Auxiliary Sentence Members

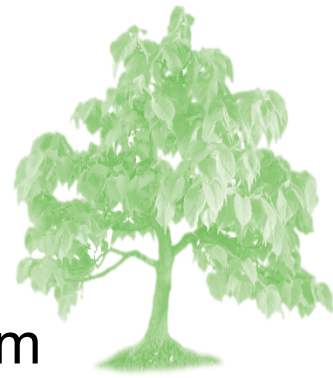


AuxT ... morphemes **se**, **si** as a part of a reflexive tantum

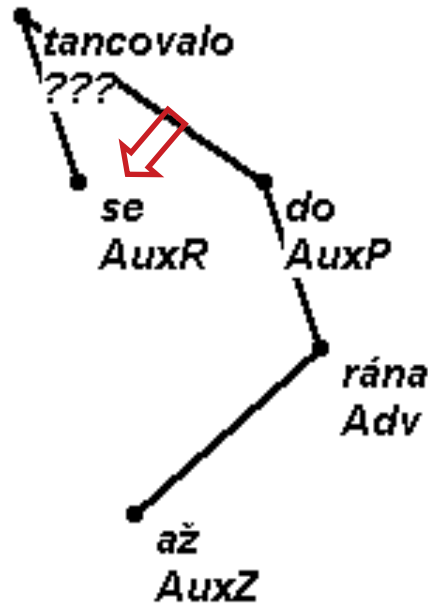


they are afraid of him

Auxiliary Sentence Members



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

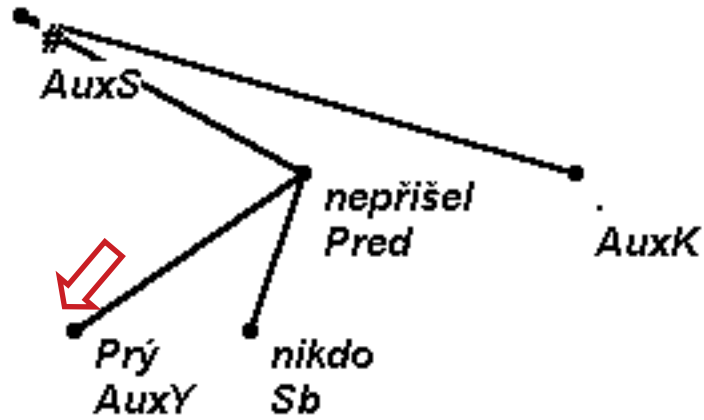


it-danced – Refl – till – to morning
the dancing went on till morning

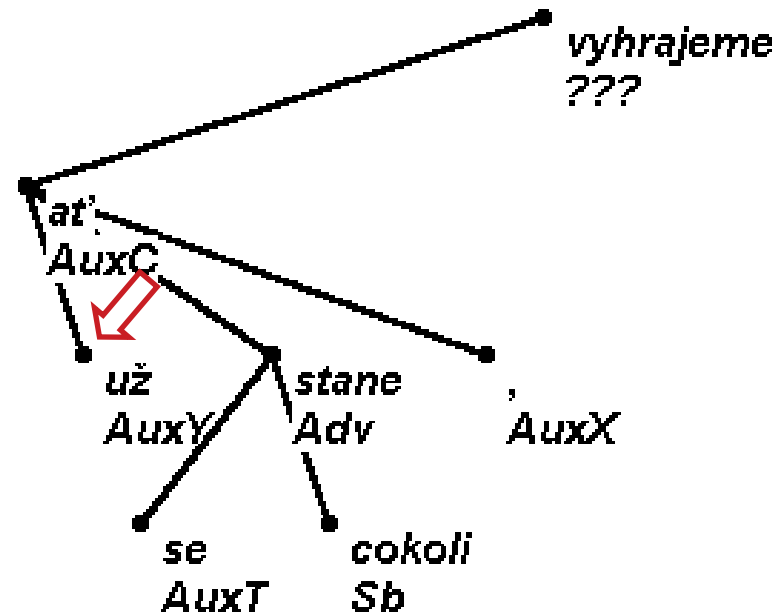
Auxiliary Sentence Members



AuxY ... particles modifying the whole sentence;



no one has come, they say

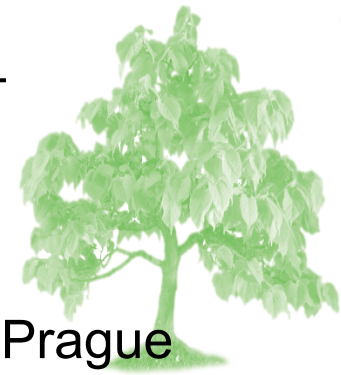


whatever may happen, we shall win

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>