Prague Dependency Treebank: Surface Syntactic Annotation

Markéta Lopatková

Institute of Formal and Applied Linguistics, MFF UK

lopatkova@ufal.mff.cuni.cz

Prague Dependency Treebank

- application of a language theory onto a large set of data
- primarily focused on Czech running texts
- attempt to standardize dependency annotation

- language data
- tools
  - TrEd … graphical editor and interface for creating queries (practical lectures) http://ufal.mff.cuni.cz/tred/
- documentation: http://ufal.mff.cuni.cz/pdt3.5/documentation
  - Guide, http://ufal.mff.cuni.cz/pdt2.0/
  - manuals for individual layers
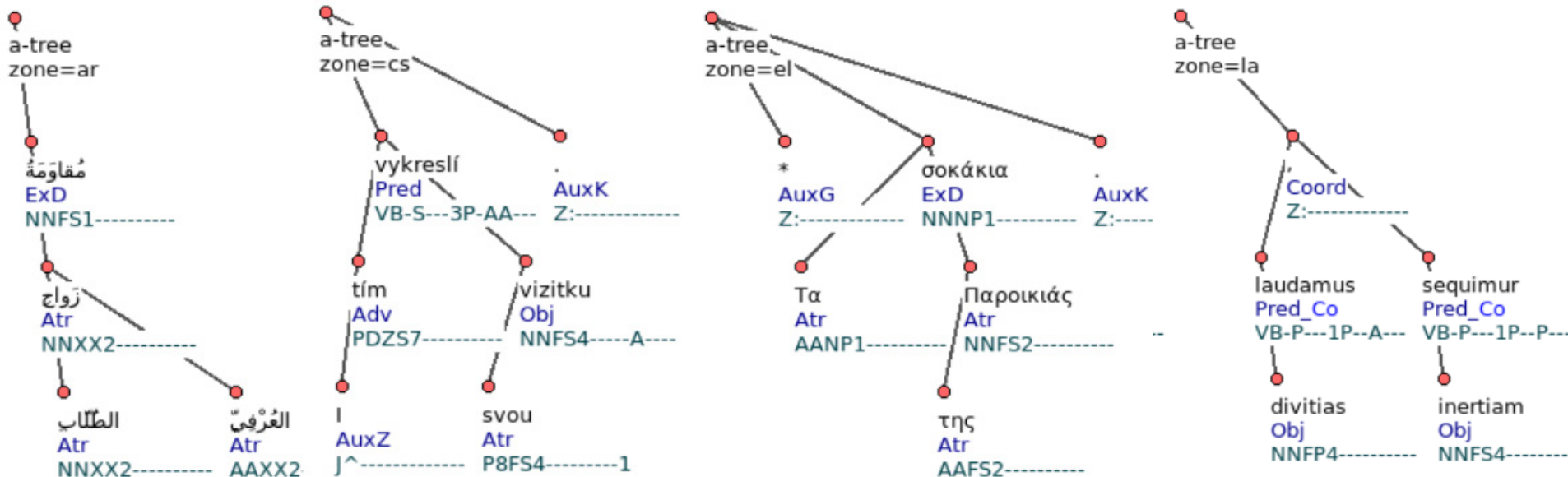 - survey of data formats and tools
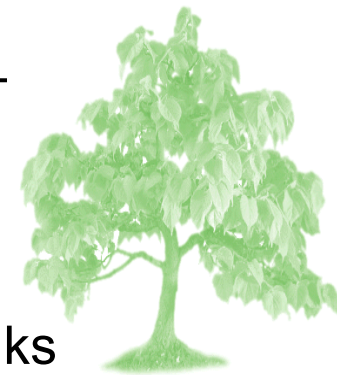
Prague Dependency Treebank

- **HamleDT** ~ a compilation of existing dependency treebanks (or dependency conversions of other treebanks), transformed so that they all conform to the same annotation style … 2012
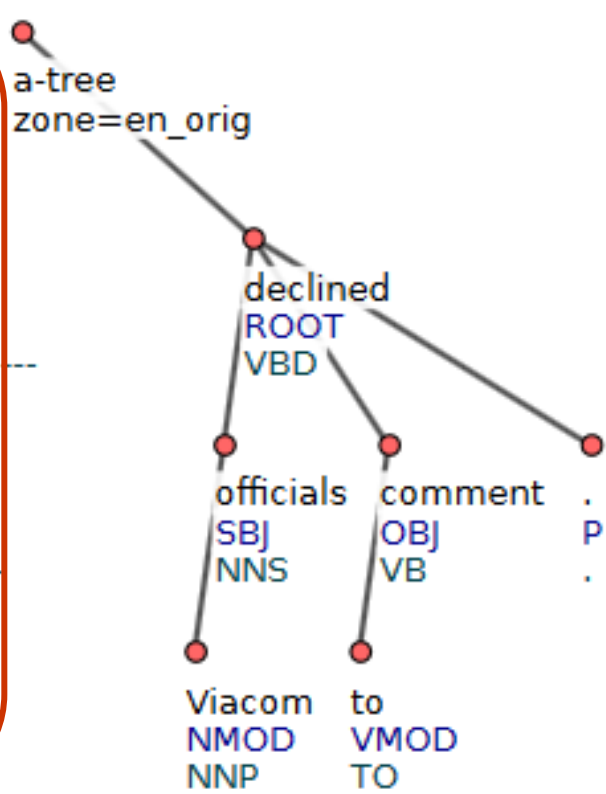
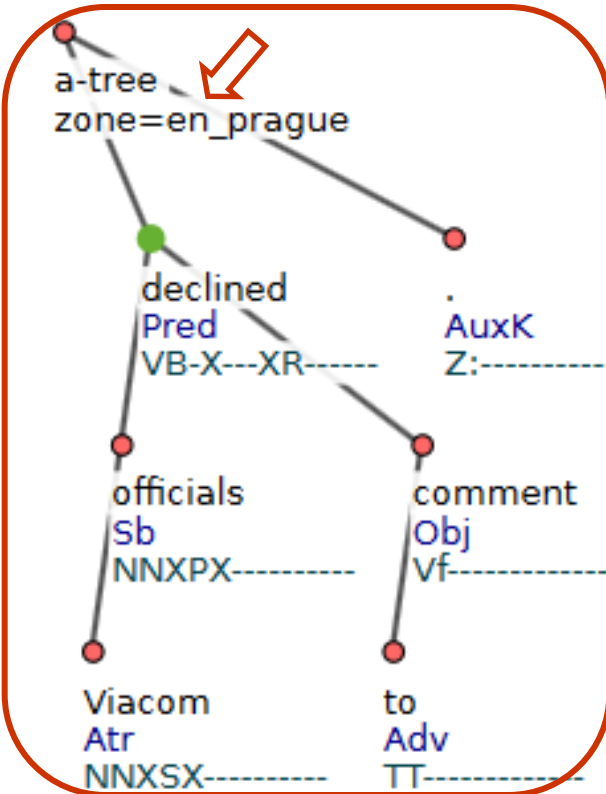 **36 languages**, 42 treebanks in HamleDT 3.0 (2015)

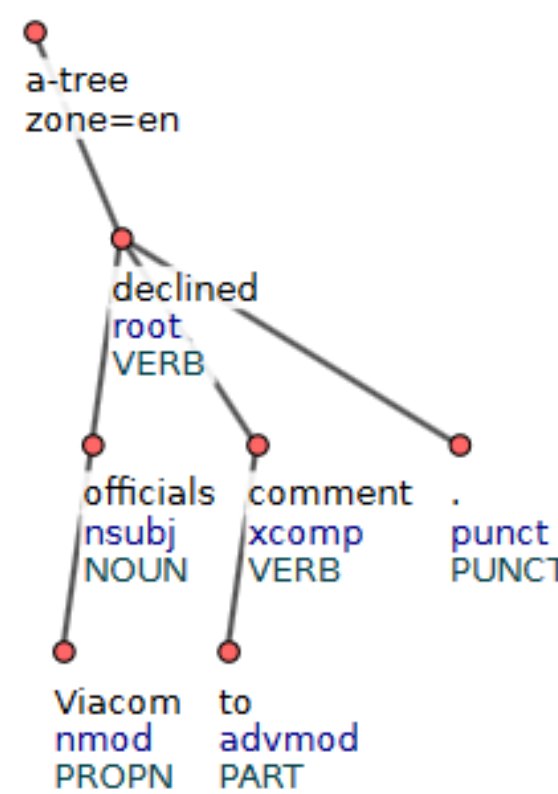 http://ufal.mff.cuni.cz/hamledt/

HamleDT: a-layer

(login required for English, free whenever possible)

PDT style

Universal Dependencies

Dependency Grammars and Treebanks: PDT a-layer

Prague Dependency Treebank

4 layers:

• word layer (w-layer)

• morphological layer (m-layer)

• analytical layer (a-layer)

• tectogrammatical layer (t-layer)

layers of annotation

layers of description	t,a,m-layer				a,m-layer
	train	dtest	etest	total	total
# documents	2 536	316	316	3 168	2 170
# sentences	38 737	5 228	5 477	**49 442**	**38 538**
# tokens	652 700	87 988	92 669	833 357	671 490

PDT: a-layer

Goal:
- to describe the structure of the sentence and
- to denote the type of relations between tokens
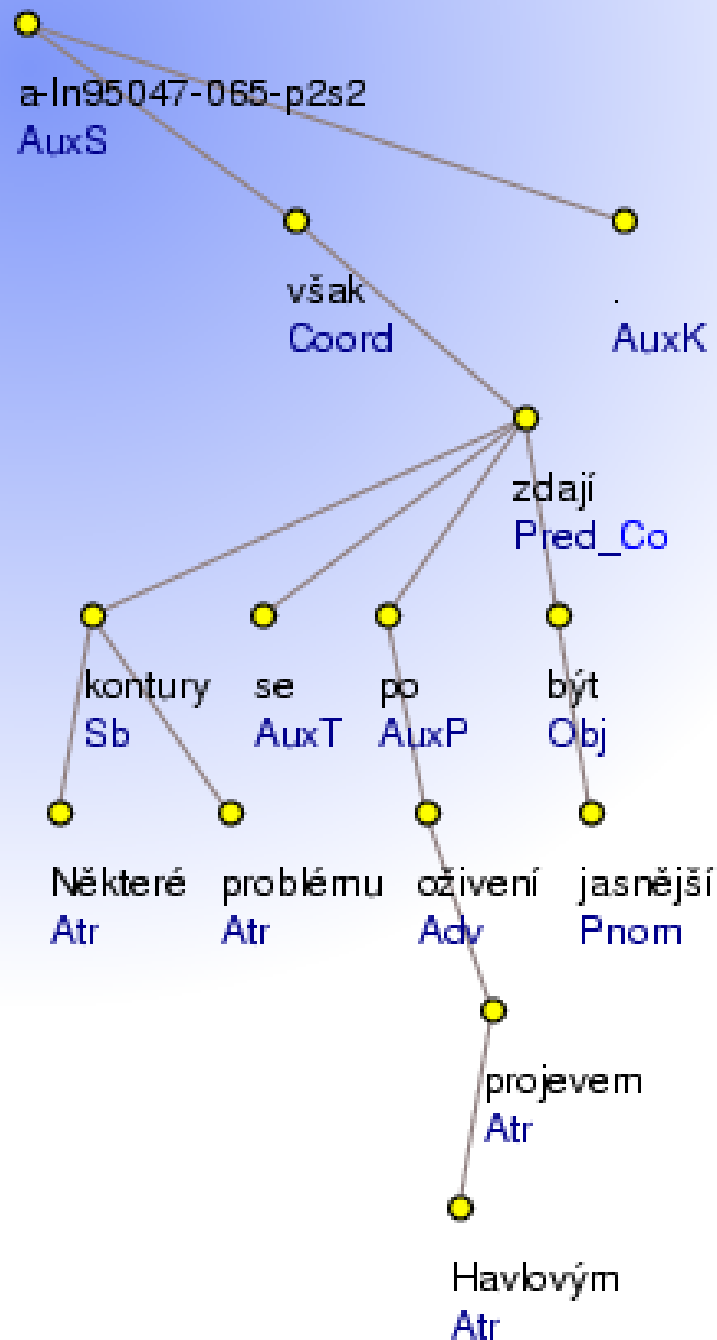
PDT: a-layer

Goal:
- to describe the structure of the sentence and
- to denote the type of relations between tokens

- dependency tree
- 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

!!! tree structure ~ XML structure !!!

PDT: a-layer

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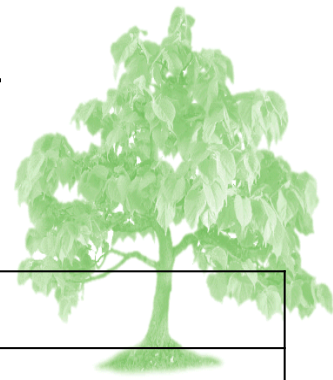[Some contours of the problem seem to be clearer after the resurgence by Havel's speech.]


a-ln95047-065-p2s2
AuxS

však
Coord

.
AuxK

zdají
Pred_Co

kontury
Sb

se
AuxT

po
AuxP

být
Obj

Některé
Atr

problému
Atr

oživení
Adv

jasnější
Pnom

projevem
Atr

Havlovým
Atr

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)

Dependency Grammars and Treebanks: PDT a-layer

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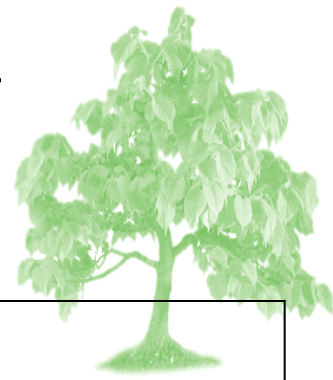
Attribute Name	Description
id	unique identifier of the tree in PDT 2.0
s.rf	PML reference (points to a sentence marked s on the m-layer)
afun	AuxS (analytical function)
ord	0 (position in the horizontal ordering of the nodes in a tree)

List of analytical functions (afun)

- sentence predicate … **Pred, Pnom, AuxV**
- subject   … **Sb**
- object   … **Obj**
- adverbial   … **Adv**
- attribute   … **Atr**
  - combined functions … **AtrAtr, AtrObj, ObjAtr, AtrAdv, AdvAtr**
- complement … **Atv, AtvV**
- auxiliary sentence member

  … **AuxS** (technical root)   **AuxO, AuxY** (emotion, particles)
  … **AuxC, AuxP** (conj ,prep)   **AuxZ**   (rhematizers)
  … **AuxT, AuxR** (reflexives)   **AuxX, AuxG, AuxK** (graphical symbols)
- non-dependency structures … **Coord, Apos**
- ellipsis   … **ExD**
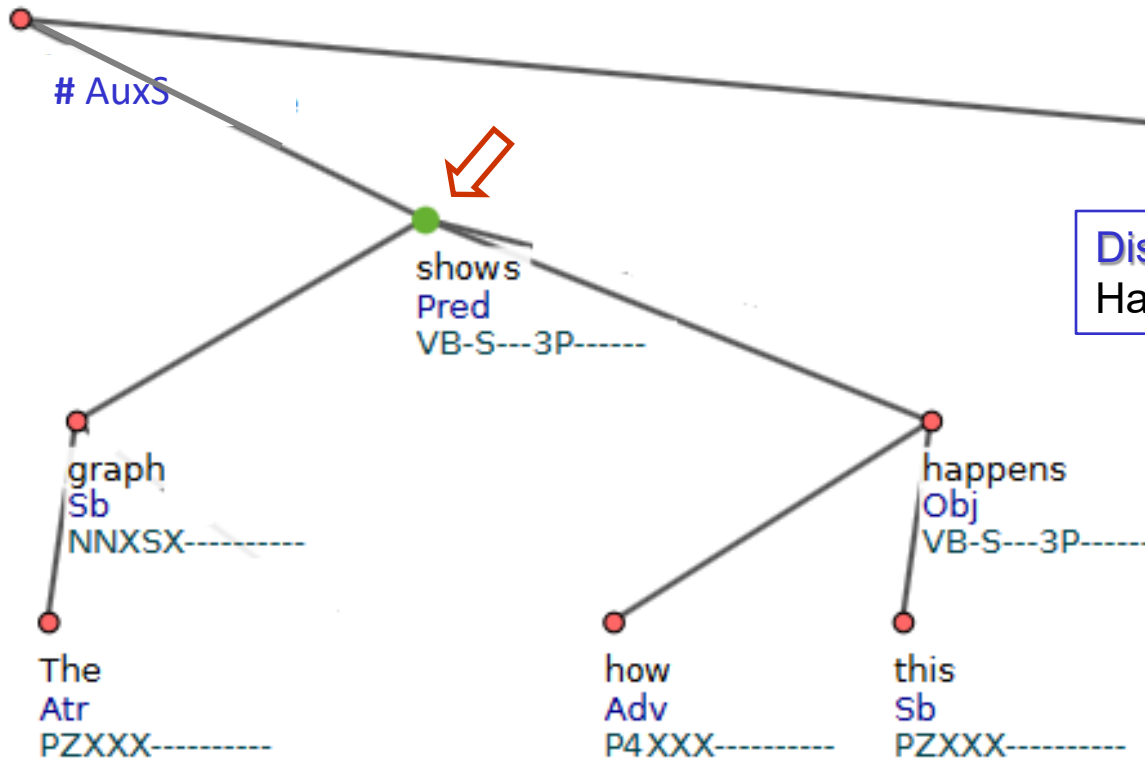
Predicate of a simple sentence / main clause

Predicate … analytical functions: **Pred, Pnom, AuxV**
- attributing some property/state/change/activity to the subject
- basic sentence member
- governing node (= head) of its clause

Predicate of a simple sentence / main clause

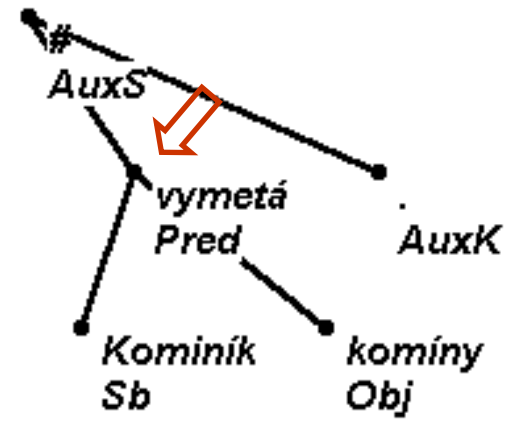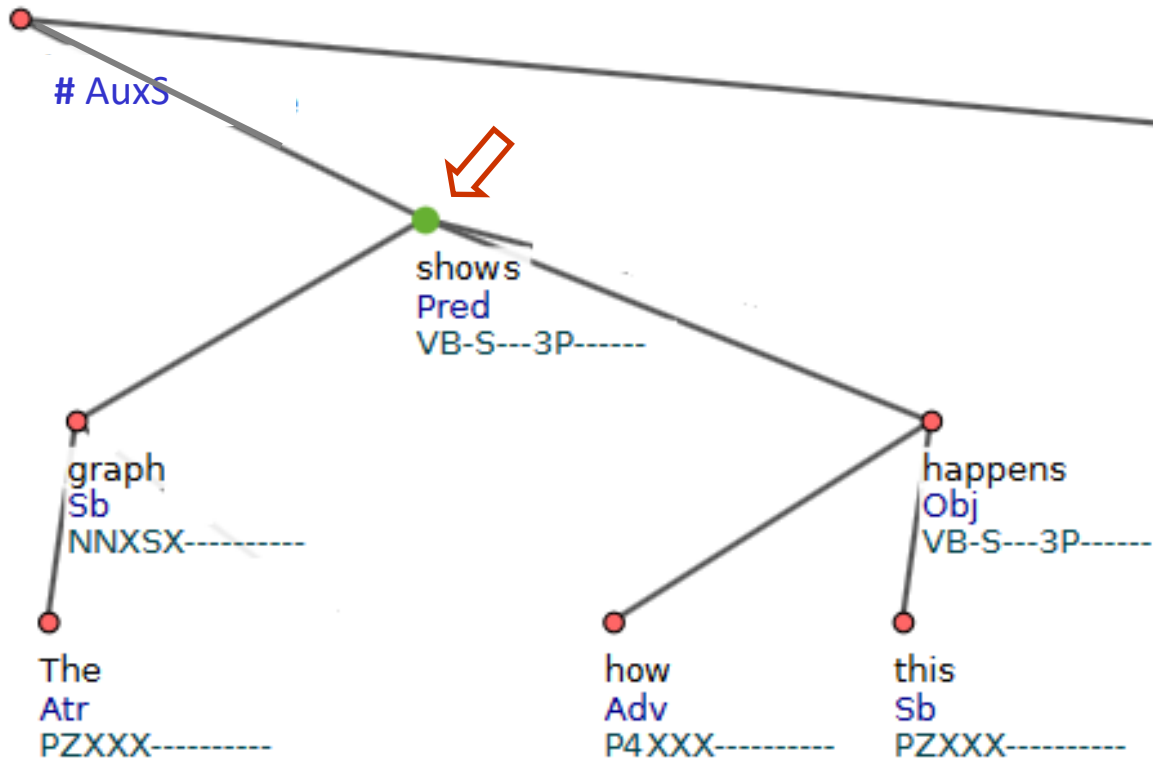- a child of the technical root
- typically a finite verb form … **Pred**

AuxS

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

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

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

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

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

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

Disclaimer … English trees from HamleDT (automatically transfered)!

*The graph *shows* how this happens …*

Predicate of a simple sentence / main clause

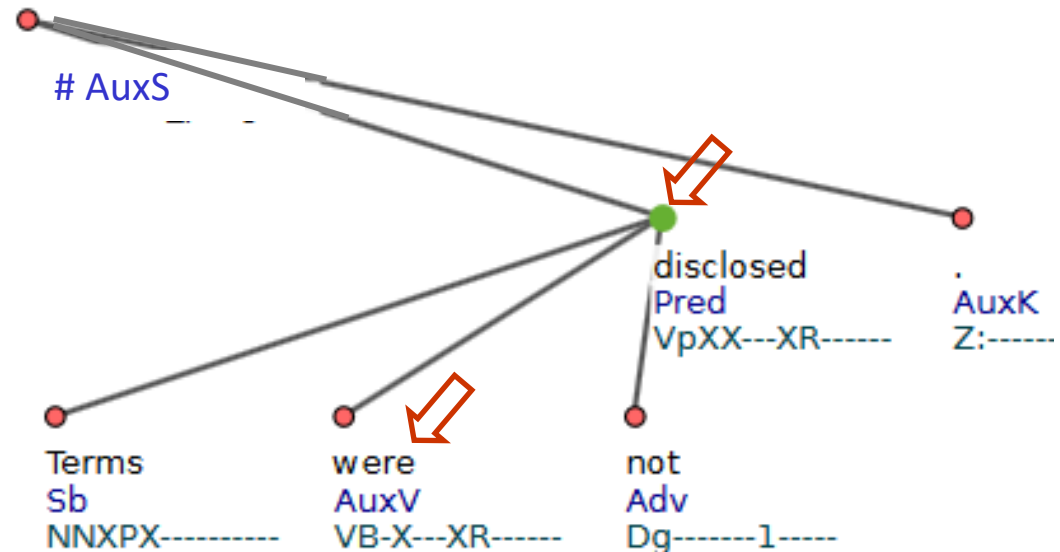- a child of the technical root
- typically a finite verb form … **Pred**

AuxS

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

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

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

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

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

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

The graph shows how this happens …

AuxS

vymetá
Pred

.
AuxK
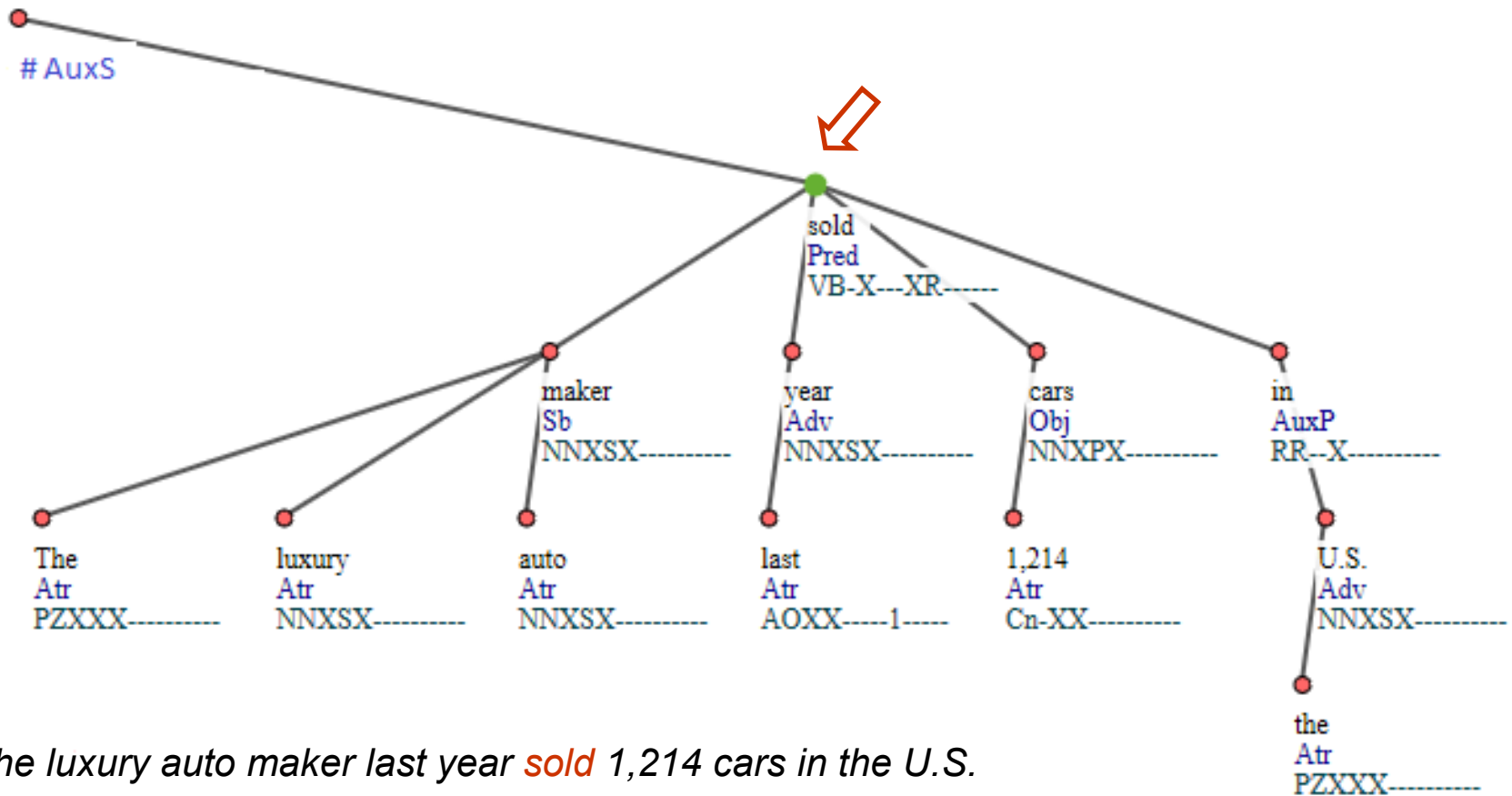
Kominík
Sb

komíny
Obj

Chimneysweep sweeps chimneys.

Predicate of a simple sentence / main clause

- a child of the technical root
- typically a finite verb form … **Pred**
  - analytical verbal forms: auxiliary verbs … **AuxV**

AuxS

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

.
AuxK
Z:-----

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

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

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

Terms were not disclosed.

(1) Simple verbal predicate: lexical verb


```
# AuxS

                                    sold
                                    Pred
                                    VB-X---XR------

        maker           year            cars            in
        Sb              Adv             Obj             AuxP
        NNXSX---------- NNXSX---------- NNXPX---------- RR--X----------

The     luxury  auto    last    1,214           U.S.
Atr     Atr     Atr     Atr     Atr             Adv
PZXXX-- NNXSX-- NNXSX-- AOXX----1---- Cn-XX----- NNXSX----------

                                                        the
                                                        Atr
                                                        PZXXX----------
```
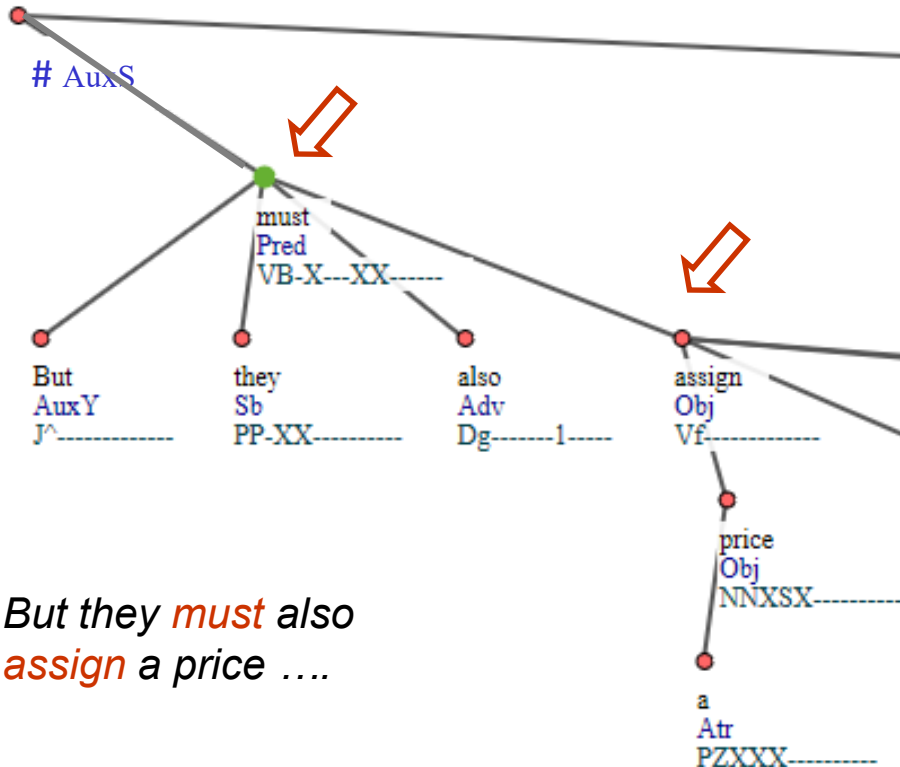
The luxury auto maker last year *sold* *1,214 cars in the U.S.*
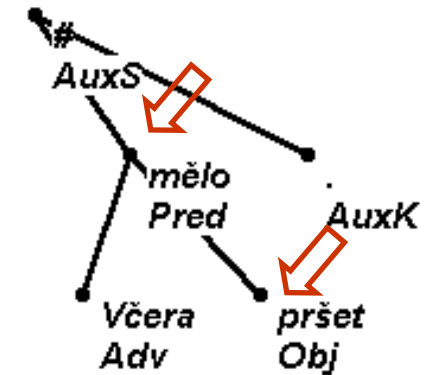
Dependency Grammars and Treebanks: PDT a-layer

(2) Compound verbal predicate

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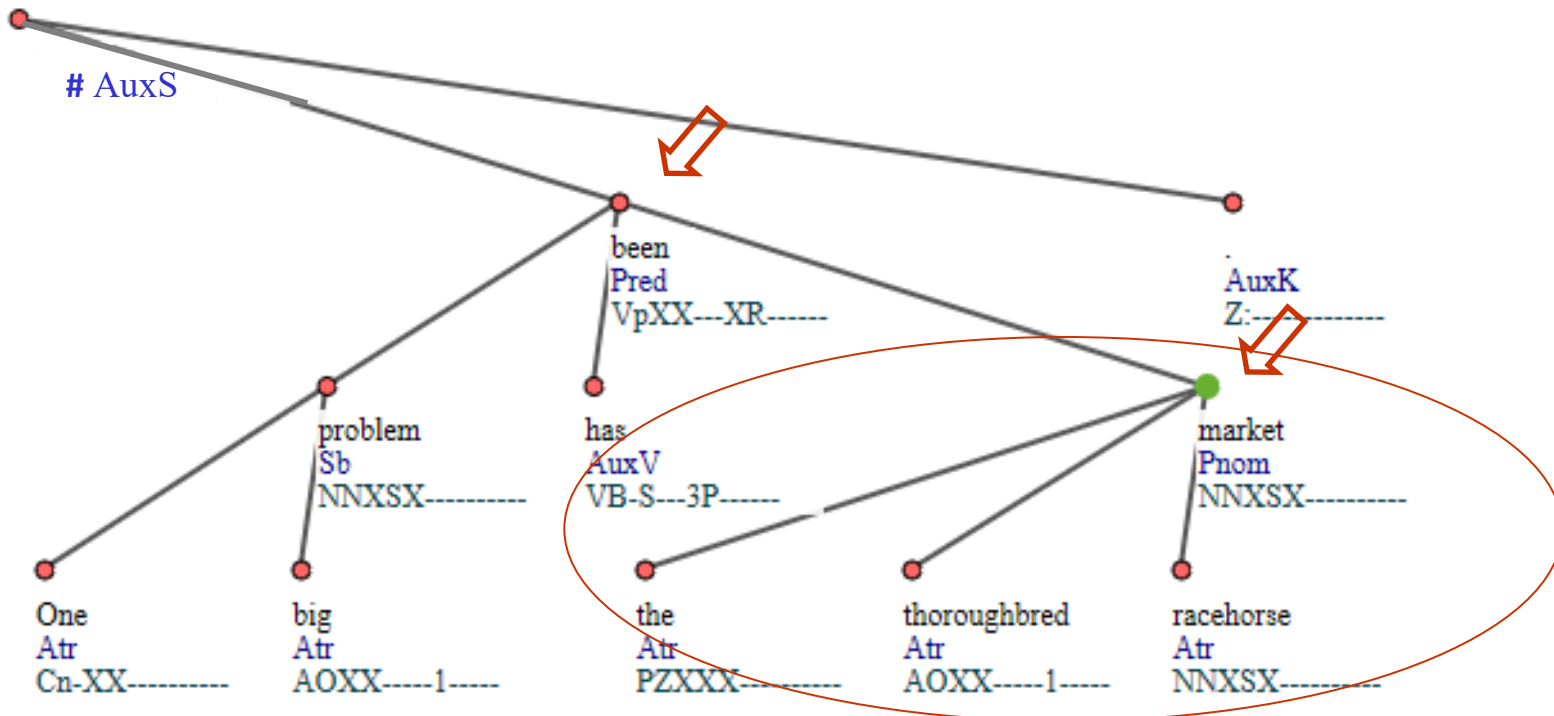

But they must also assign a price ….

Yesterday (it) should rain.

(3) Verbonominal predicates

- copular verb … **Pred**
  - a verb used to link the subject of a sentence with a predicate
  - verb expressing identity, membership in a class, property …
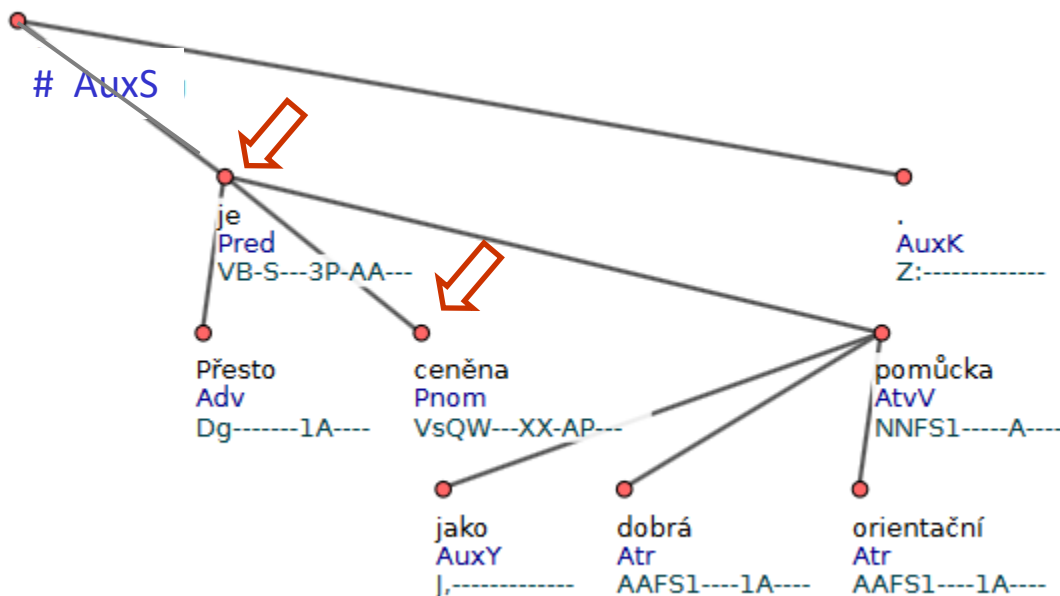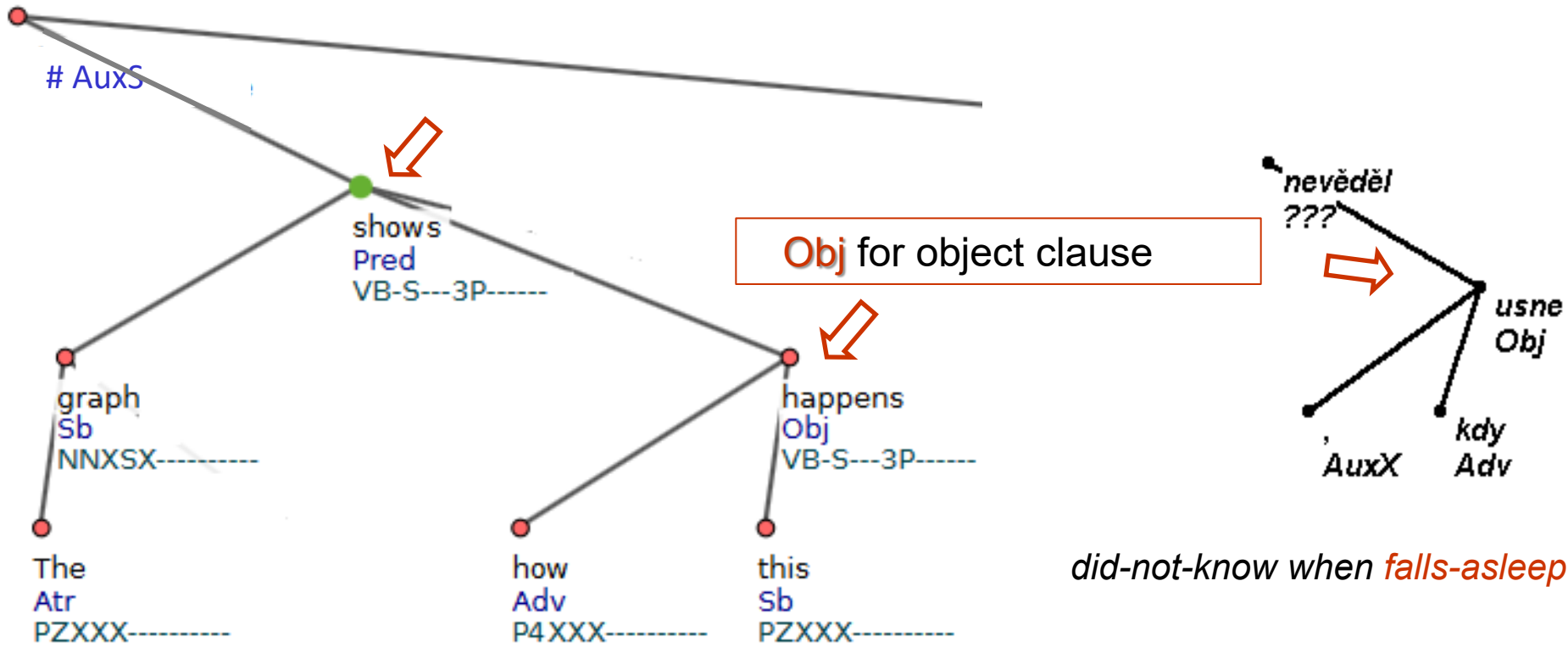- nominal predicate … **Pnom**

# AuxS				

	been		.	
	Pred		AuxK	
	VpXX---XR------		Z:----------	

	problem	has		market
	Sb	AuxV		Pnom
	NNXSX----------	VB-S---3P------		NNXSX----------

One	big	the	thoroughbred	racehorse
Atr	Atr	Atr	Atr	Atr
Cn-XX----------	AOXX-----1-----	PZXXX---------	AOXX-----1-----	NNXSX----------

One big problem has been the thoroughbred racehorse market.

(3) Verbonominal predicates

- copular verb … **Pred**
  - a verb used to link the subject of a sentence with a predicate
  - verb expressing identity, membership in a class, property …
- nominal predicate … **Pnom**

| # AuxS |
| je Pred VB-S---3P-AA--- |
| Přesto Adv Dg-------1A---- |
| ceněna Pnom VsQW---XX-AP--- |
| . AuxK Z:------------ |
| pomůcka AtvV NNFS1-----A---- |
| jako AuxY J,------------ |
| dobrá Atr AAFS1----1A---- |
| orientační Atr AAFS1----1A---- |
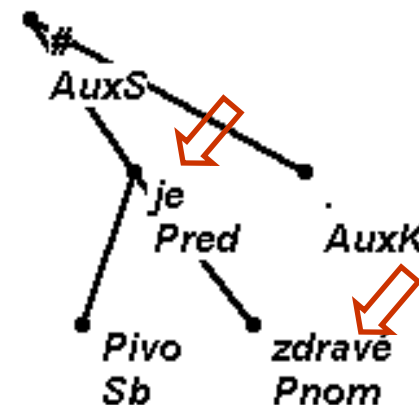
Přesto je ceněna jako dobrá orientační pomůcka.

Beer is healthy.

Predicate of a *subordinate clause*

- afun of the respective function

AuxS

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

Obj for object clause

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

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

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

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

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

The graph shows how this happens …

nevěděl
???

usne
Obj

,
AuxX

kdy
Adv

did-not-know when falls-asleep

Predicate of a *subordinate clause*

- afun of the respective function

function words … later in this presentation !!

Obj for object clause
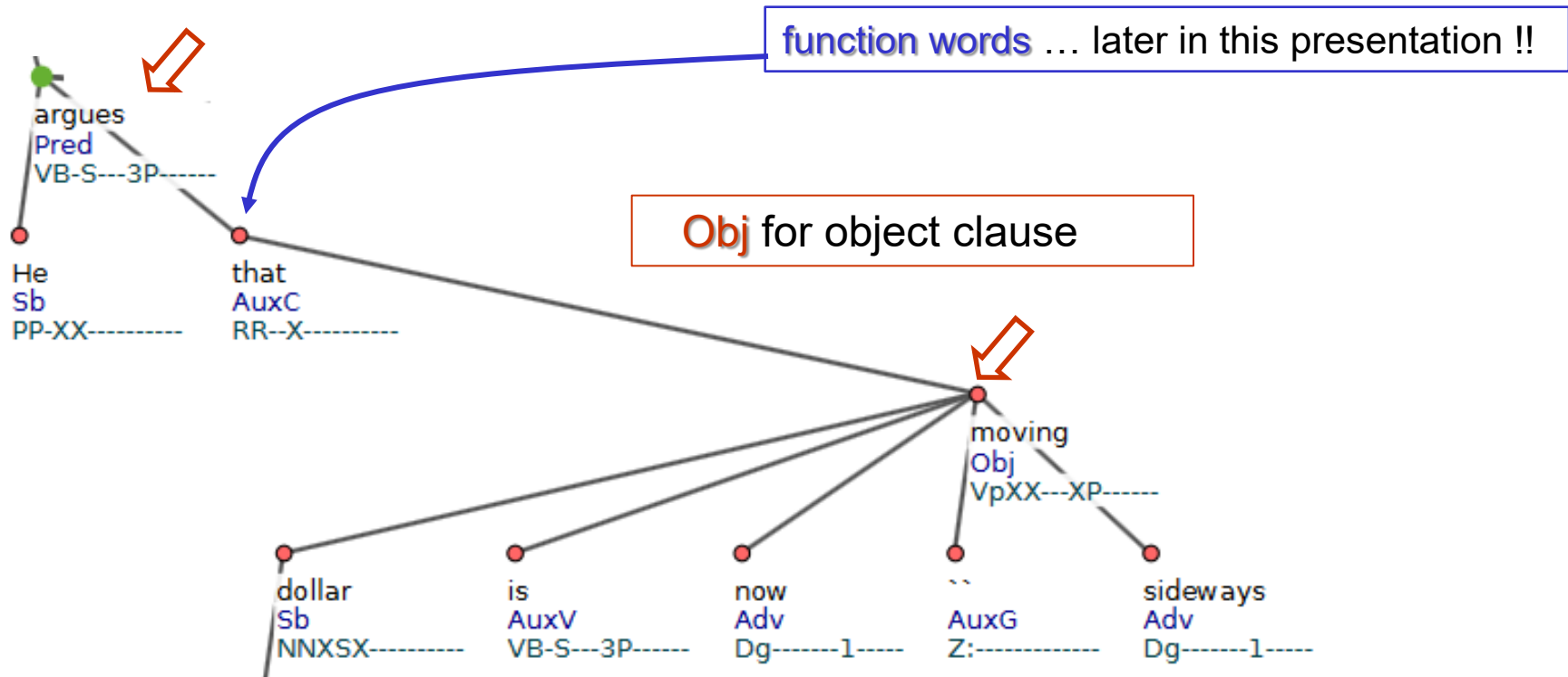
He argues that […] dollar is now moving "sideways" …

Subject

Subject … analytical function: **Sb**
- any construction answering the question *who (what)*
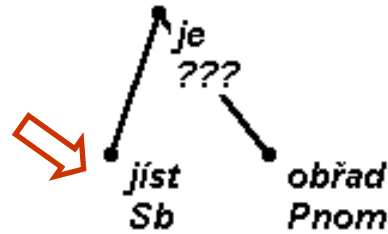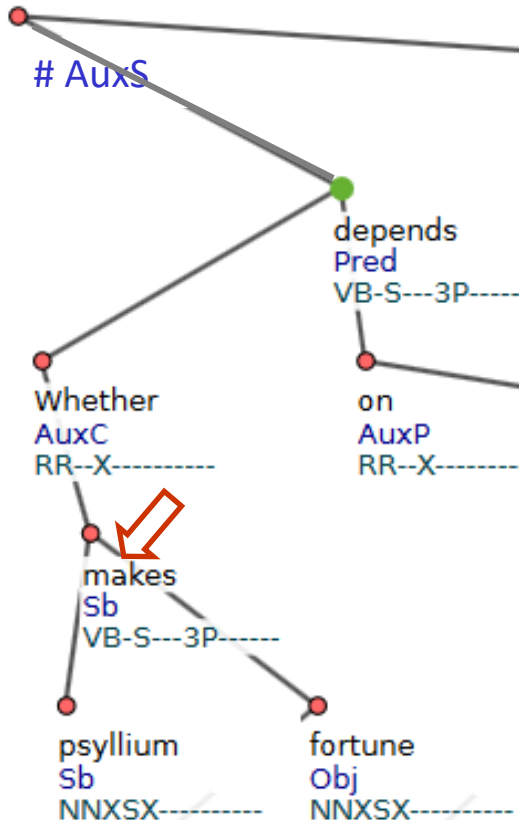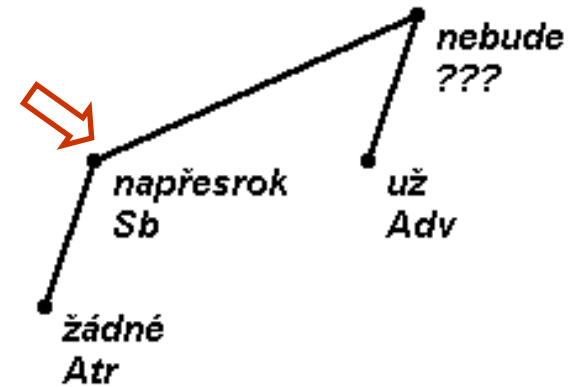- dependent on a clause predicate
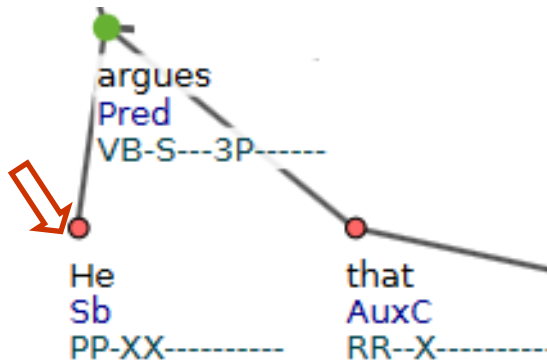
He argues that…
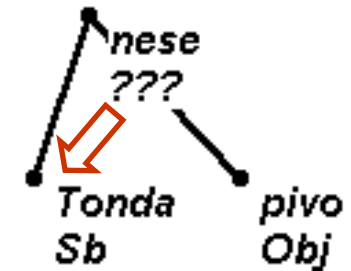
Tonda nese pivo.
Tony – brings – beer

Subject

Subject … analytical function: **Sb**

- any construction answering the question *who (what)*
- dependent on a clause predicate

kdo se bojí, nesmí do lesa
who –refl – is-frightened – must-not – to – wood
whoever is frightened, must not enter a wood
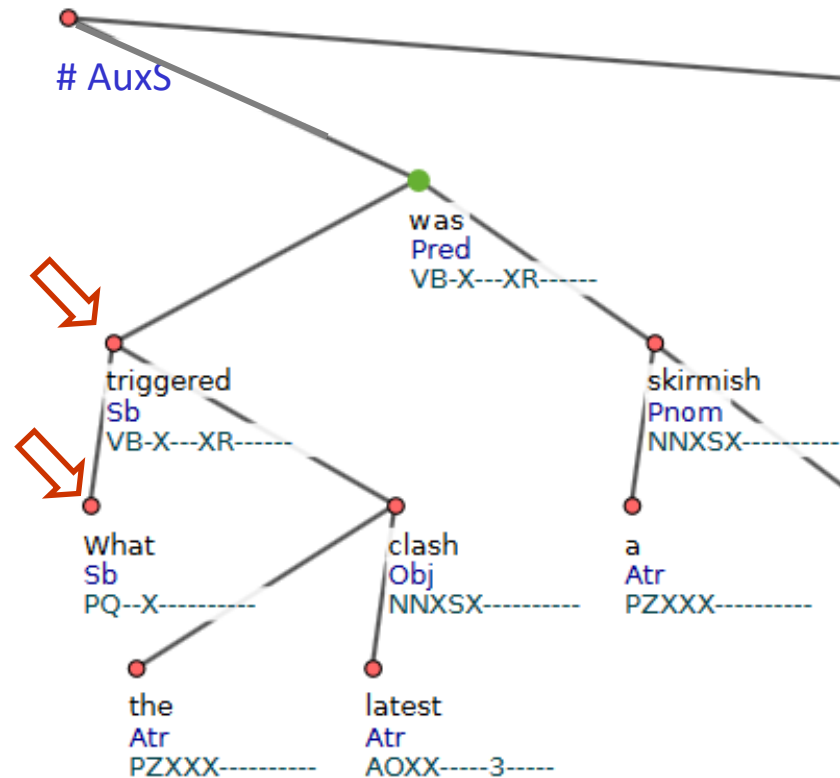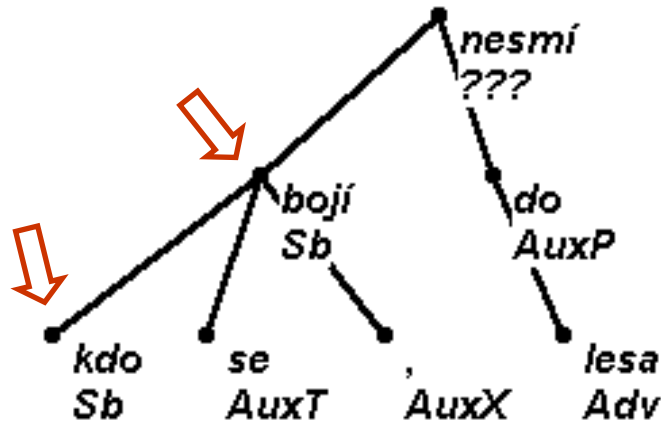
What triggered the latest clash was a skirmish …

Dependency Grammars and Treebanks: PDT a-layer

Subject

Subject … analytical function: **Sb**

- any construction answering the question *who (what)*
- dependent on a clause predicate

AuxS

depends
Pred
VB-S---3P-----

Whether
AuxC
RR--X----------

on
AuxP
RR--X---------

makes
Sb
VB-S---3P------

psyllium
Sb
NNXSX----------

fortune
Obj
NNXSX----------

Whether psyllium fortune depends on …….

je
???

jíst
Sb

obřad
Pnom

jíst je obřad
to-eat - is – ceremony
eating is a ceremony

nebude
???

napřesrok
Sb

už
Adv

žádné
Atr

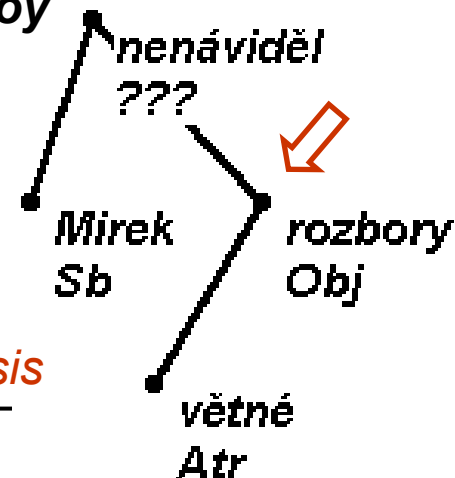*žádné napřesrok už nebude*
no - next-year - anymore - will-not-be

Object

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

AuxS

see
Pred
VB-X---XP------

We
Sb
PP-XX----------

argument
Obj
NNXSX----------

an
Atr
PZXXX----------

stronger
Atr
AOXX-----2-----

against
AuxP
RR--X----------

even
Adv
Dg-------1-----
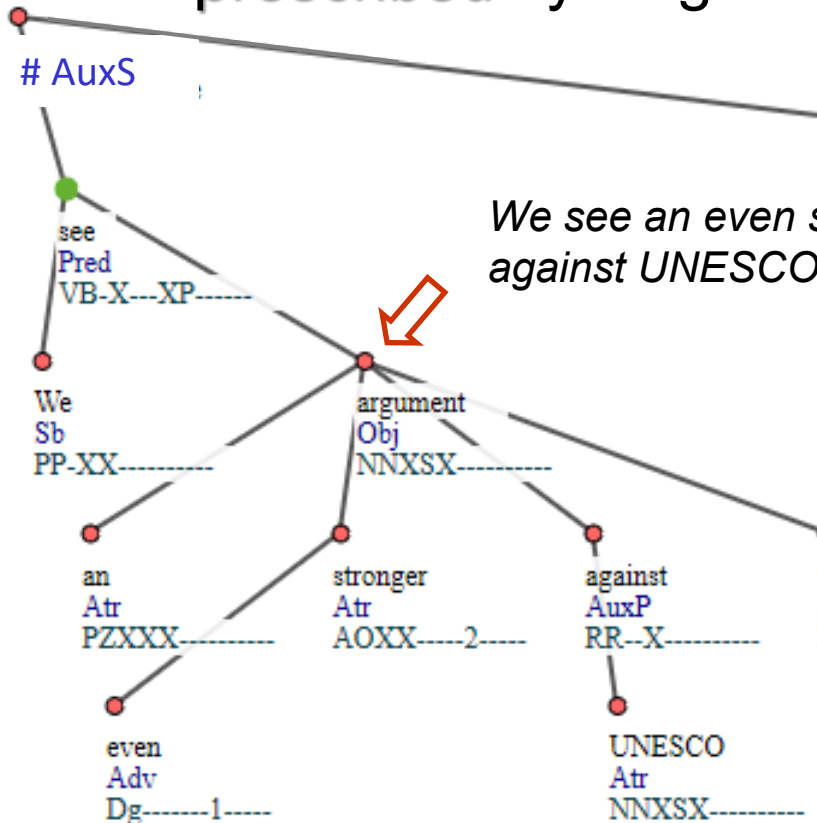
UNESCO
Atr
NNXSX----------

We see an even stronger argument against UNESCO …
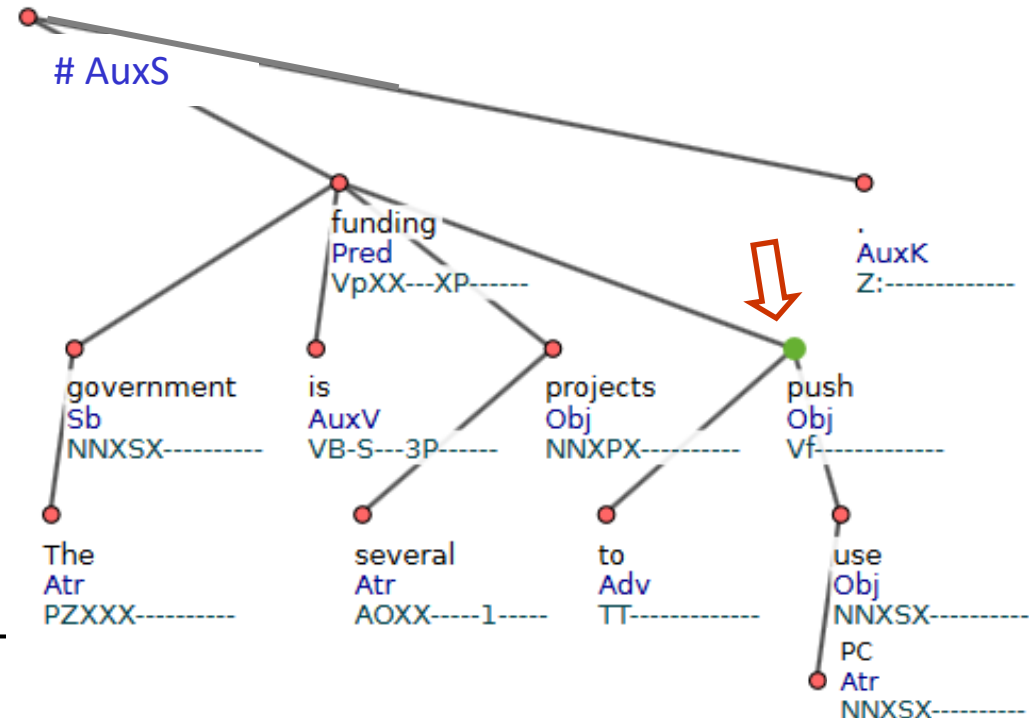
nenáviděl
???

Mirek
Sb

rozbory
Obj

větné
Atr
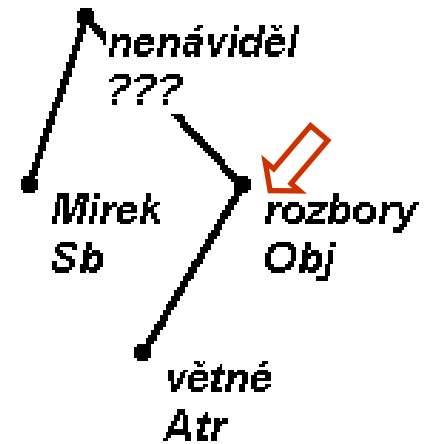
Mirek hated (a) sentence analysis

Object

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

(i) 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

nenáviděl
???

Mirek
Sb

rozbory
Obj

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**)

(i)   direct (accusative), indirect (dative), second object
(ii)   incl. infinitive constructions

AuxS

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----------

The government is funding several projects to push PC use.

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**)

(i)   direct (accusative), indirect (dative), second object
(ii)   incl. infinitive constructions
(iii) compound predicates … **Pred + Obj**

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

(iv) direct speech

"I believe it!" Cilka replied.
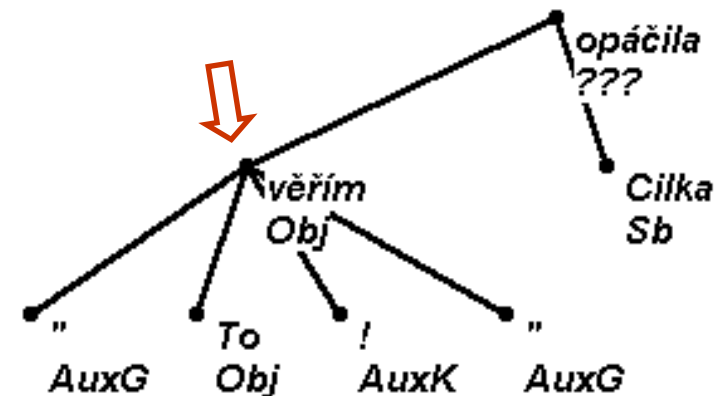
Adverbials

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

[… he says] Judge O'Kicki
unexpectedly awarded him an
additional $ 100,000
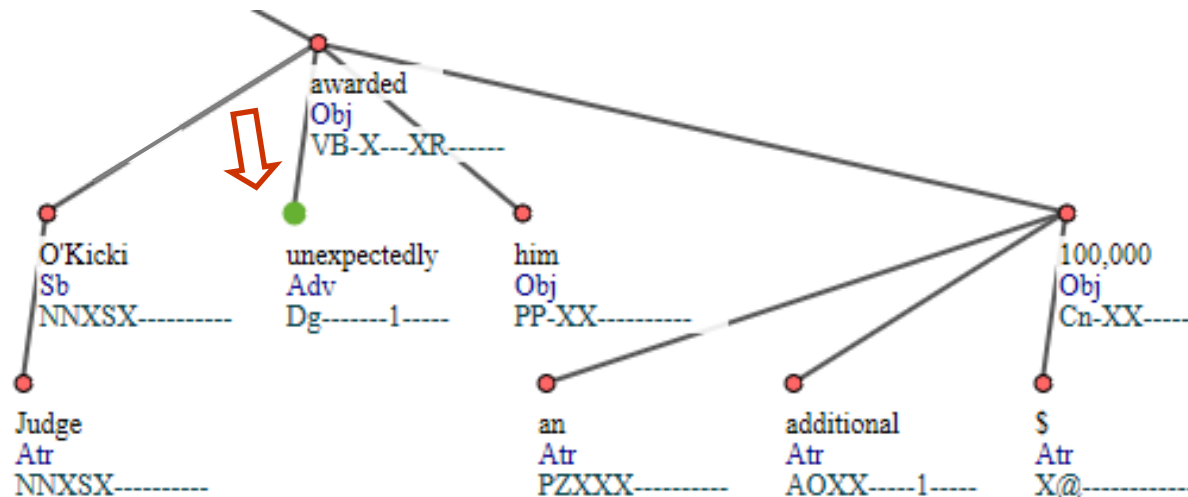
awarded
Obj
VB-X---XR------

O'Kicki
Sb
NNXSX----------

unexpectedly
Adv
Dg-------1-----

him
Obj
PP-XX----------

100,000
Obj
Cn-XX----

Judge
Atr
NNXSX----------

an
Atr
PZXXX----------

additional
Atr
AOXX-----1-----

$
Atr
X@-----------

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

won
Adv
VB-X---XR------

When
Adv
P4XXX----------

Green
Sb
NNXSX----------

verdict
Obj
NNXSX----------

Mr.
Atr
NNXSX----------

a
Atr
PZXXX----------

$
Atr
X@----------
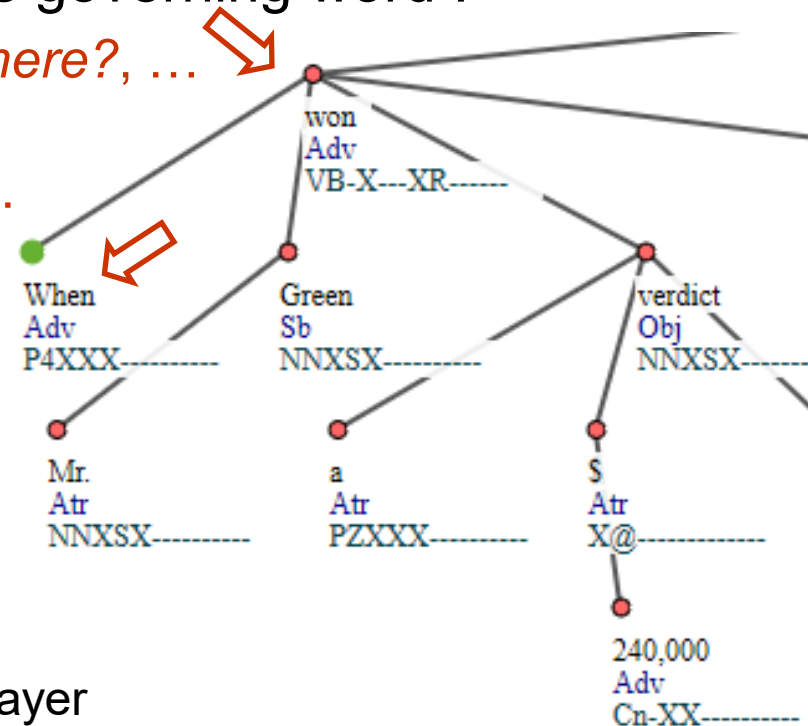
240,000
Adv
Cn-XX----------

When Mr. Green won a $ 240,000 verdict …

Attribute

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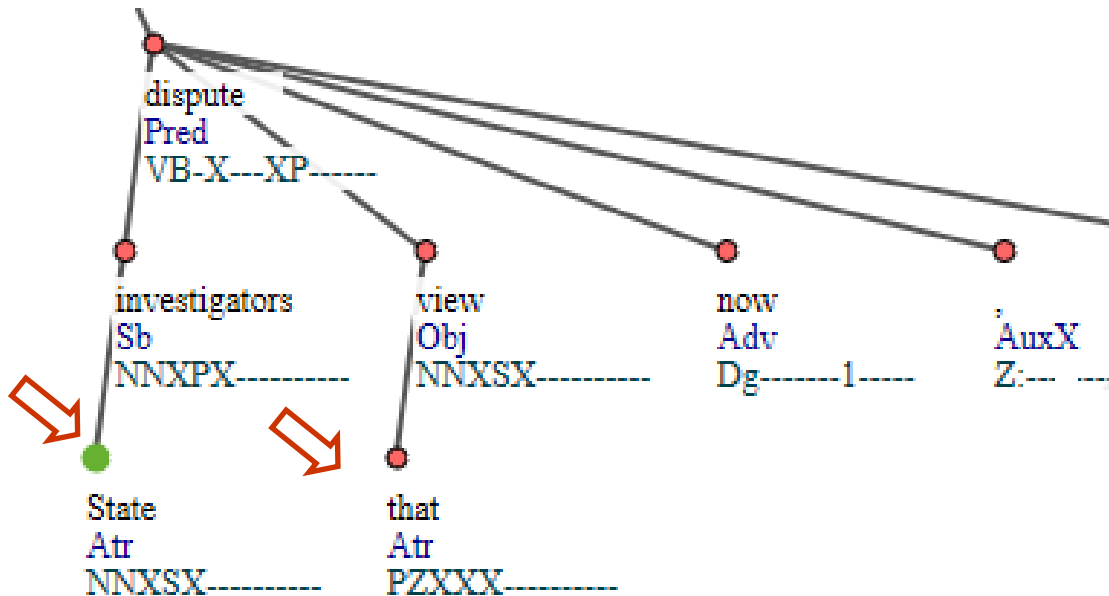
State investigators dispute that view now, …

Dependency Grammars and Treebanks: PDT a-layer

Attribute

Attribute … analytical functions: **Atr**
- <u>modifies a noun </u>(with any function)
- answers the question … *which*, *what* or *whose* (*jaký, který, čí*)
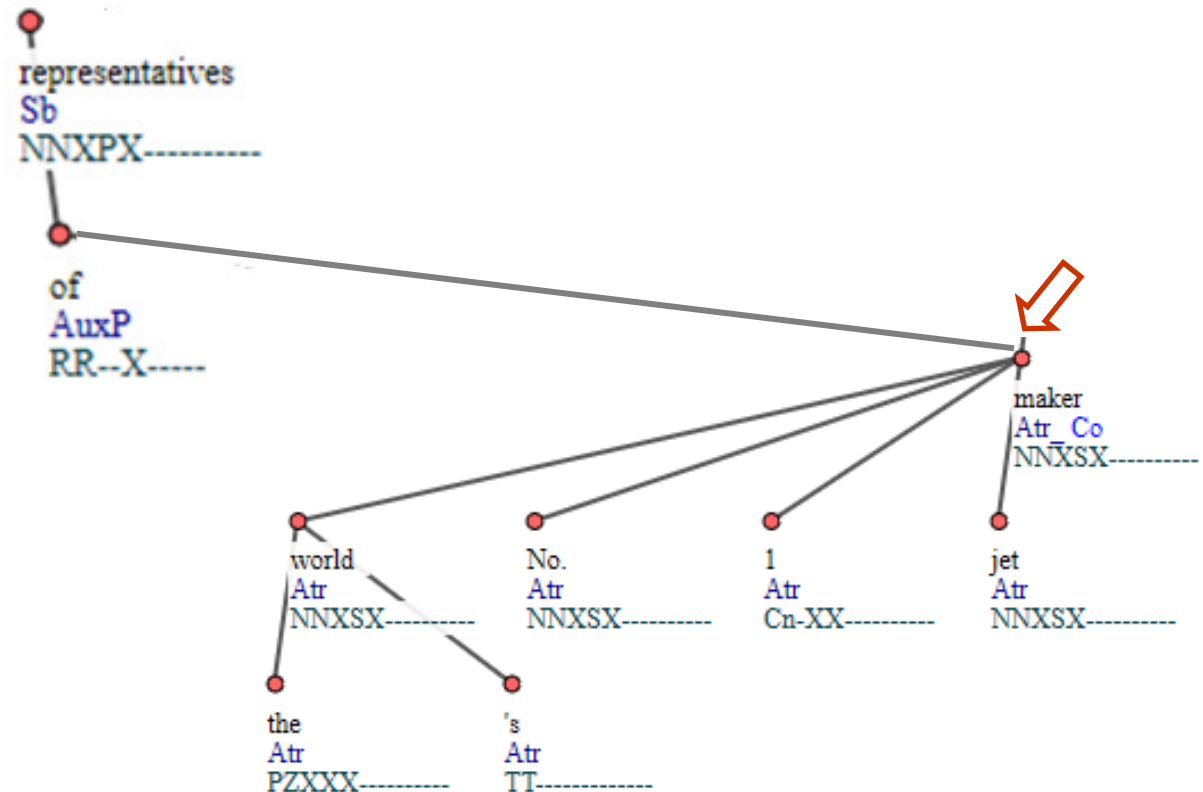
… representatives of the world 's No. 1 jet maker …

Attribute: Agreement

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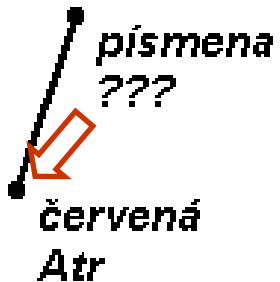
red letters
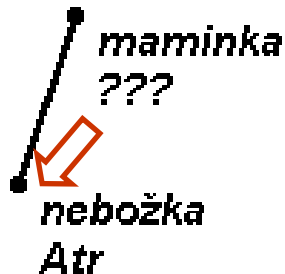
the-dead mother

Attribute: Agreement

- agreeing attribute … the same case as its governing noun
- non-agreeing attribute

chalupa
???

souseda
Atr

našeho
Atr

cottage of our *neighbor*

visel
???

na
AuxP

dveřích
Adv

nápis
Sb

Neprodejné
Atr

"
AuxG

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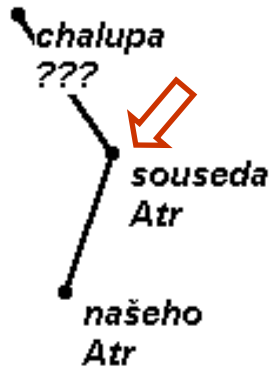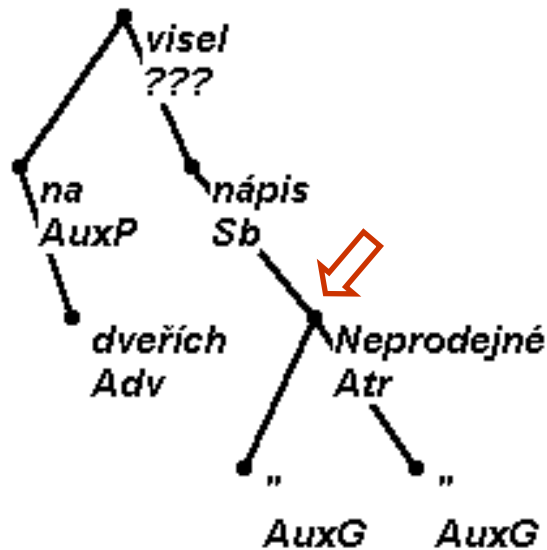
*an inscription "Not for
sale" was hanging on the
door*

vědomí
???

že
AuxC

,
AuxX

nečeká
Atr

smrt
Sb
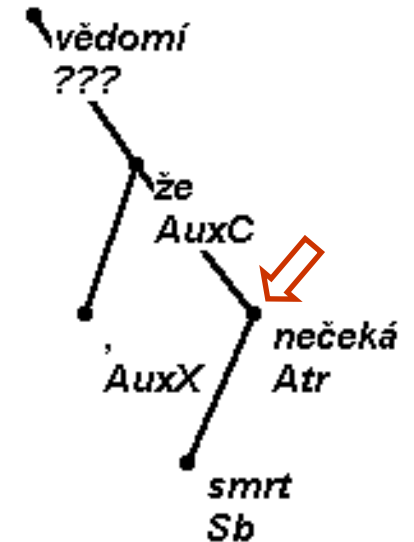
*knowledge that death
does-not-wait*

Attribute

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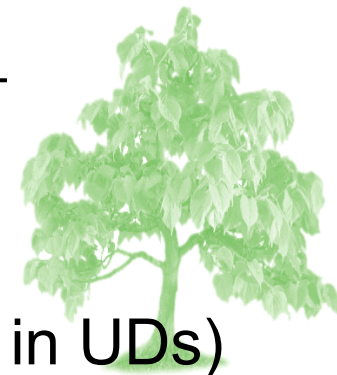
Complement

Complement (verbal attribute, secondary predication in UDs)

… analytical functions: **Atv, AtvV**

- modifies two sentence members, verb and noun

chlapec ležel nemocen, viděl ho nemocného

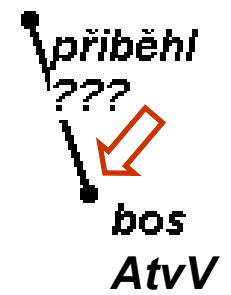*[boy – lay – ill]*   *[he – saw – him – (being) ill]*
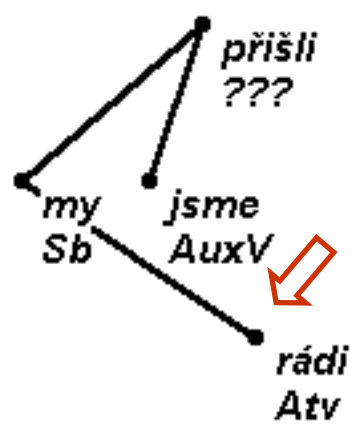
we - AuxV - came - glad
we were glad to come
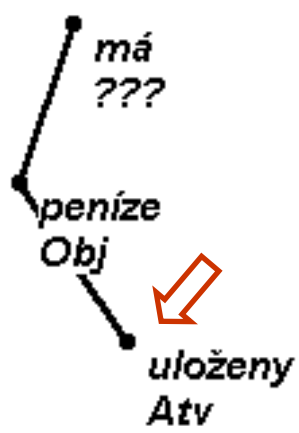
money - he-has - deposit
he has put the money on a deposit
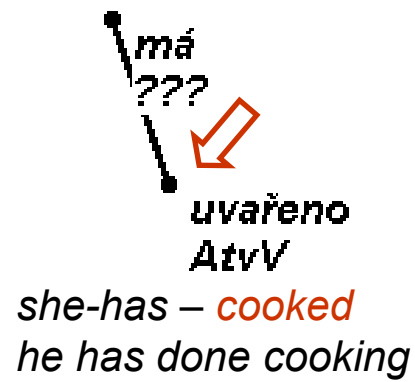
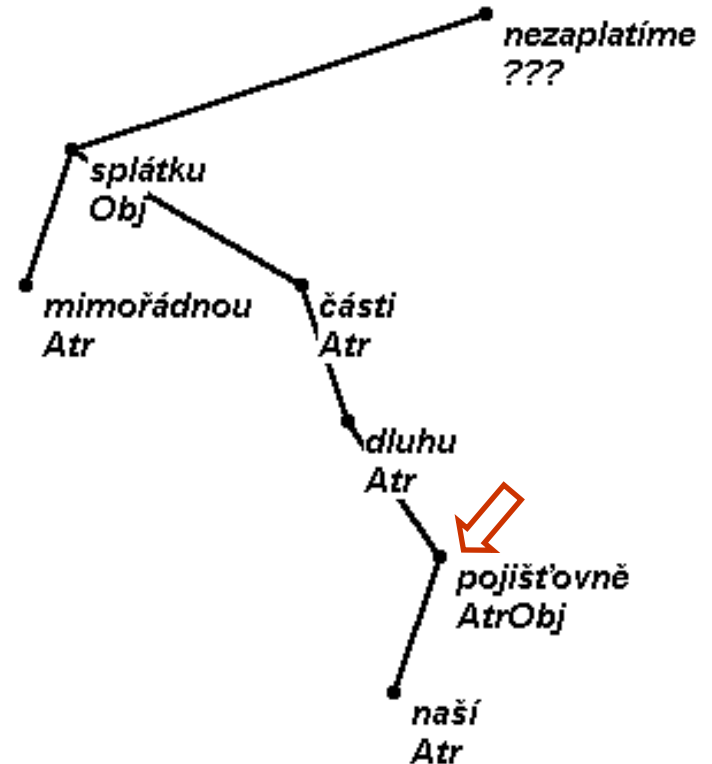*she-has – cooked*
he has done cooking

he-arrived - barefooted

Combined Functions

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

přinesla
???

bednu
Obj

ze
AuxP

sklepa
AtrAdv

krajíc
???

chleba
Atr

s
AuxP

máslem
AtrAtr

(she) brought (a) case from (the) basement
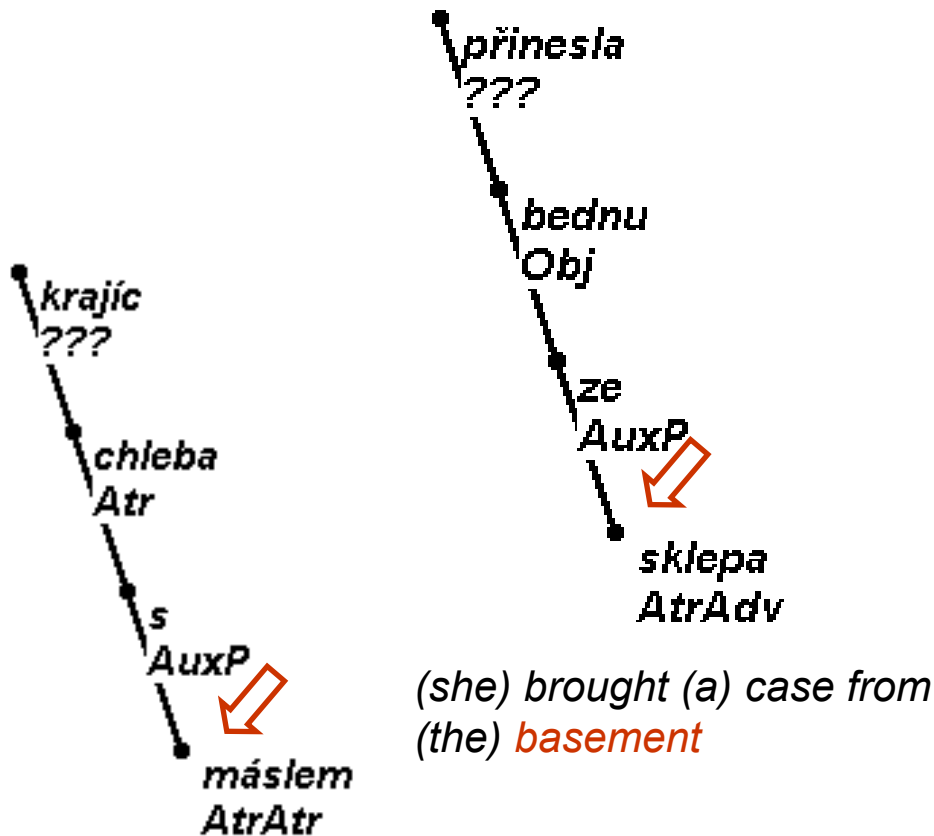
slice of bread with butter

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

Auxiliary Sentence Members

- function words
- punctuation and graphical symbols
- words with unclear syntactic function

Auxiliary Sentence Members

AuxP … prepositions
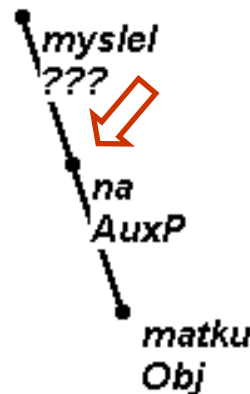- *parent of a nominal node*
- similarly one-word improper prepositions


```
came
Pred
VB-X---XR------

judge          under          in              ,
Sb             AuxP           AuxP            AuxX
NNXSX--------  RR--X--------  RR--X--------   Z:-----------

The            scrutiny       1987
Atr            Adv            Adv
PZXXX--------  NNXSX--------  Cn-XX----------

                              late
                              Atr
                              AOXX-----1-----
```
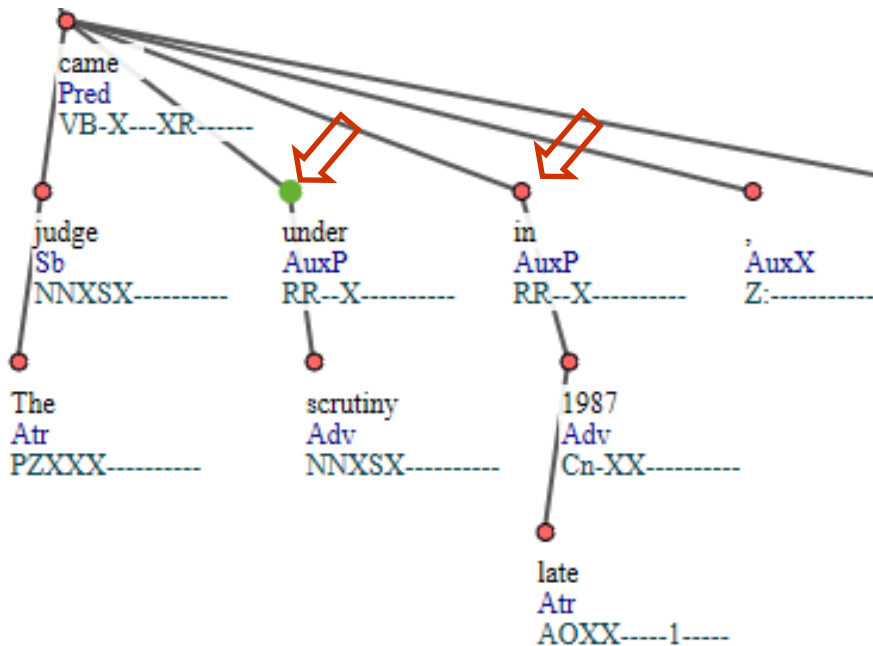
The judge came under scrutiny in late 1987, …

Auxiliary Sentence Members

AuxP … prepositions
- ***parent of a nominal node***
- similarly one-word improper prepositions

přechod
???

z
AuxP

do
AuxP

obrany
Atr

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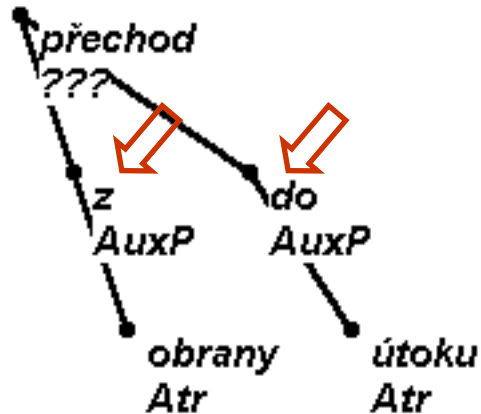
(the) transition from (a) defense to (an) attack

myslel
???

na
AuxP

matku
Obj

(he) thought of (his) mother

jednali
???

ohledně
AuxP

dědictví
Obj
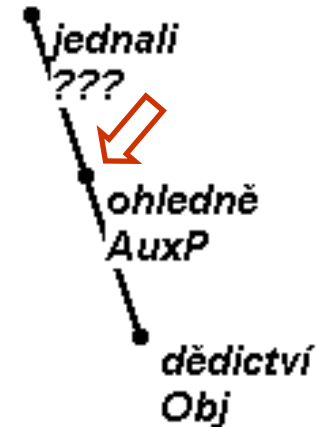
(they) negotiated - about - (the) heritage

Auxiliary Sentence Members

AuxP … secondary prepositions
- consists of several words
- technical solution:
    - the last node … head
    - the remaining words … siblings of the noun governed
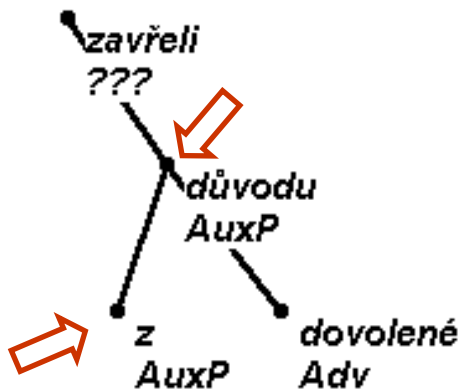
(they) closed *- for - reasons* - (of) leave
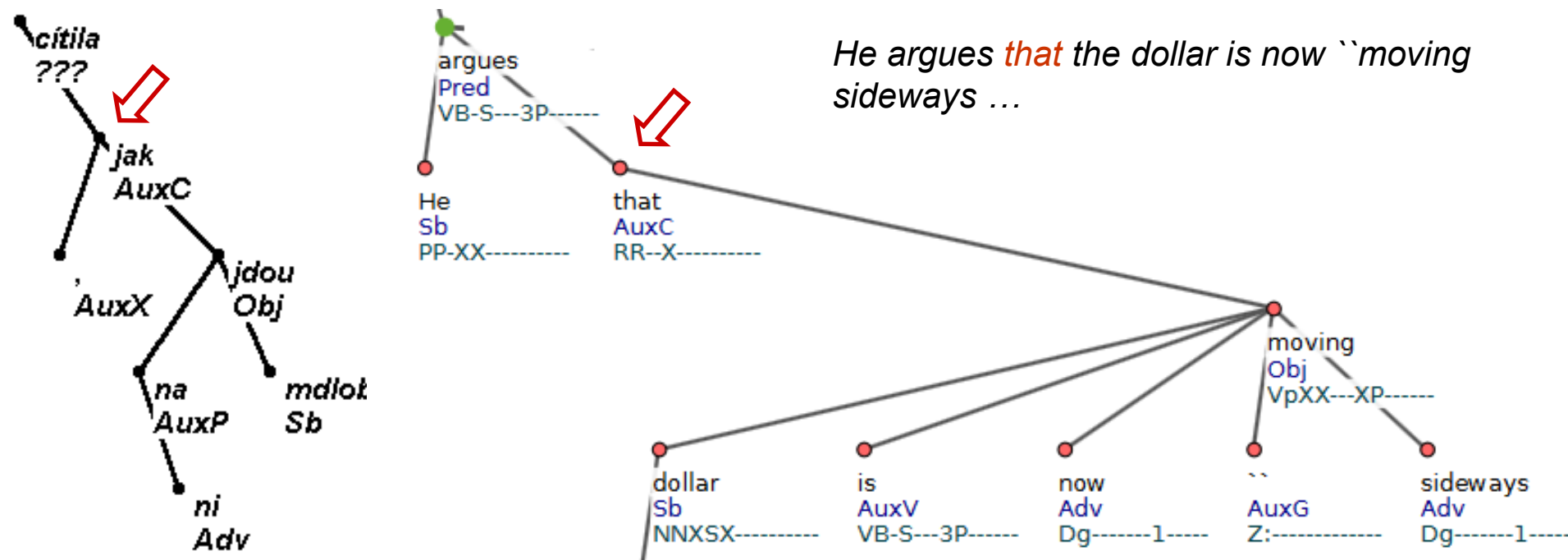(they) closed because of the leave

Because of Leche Fresca 's
success …
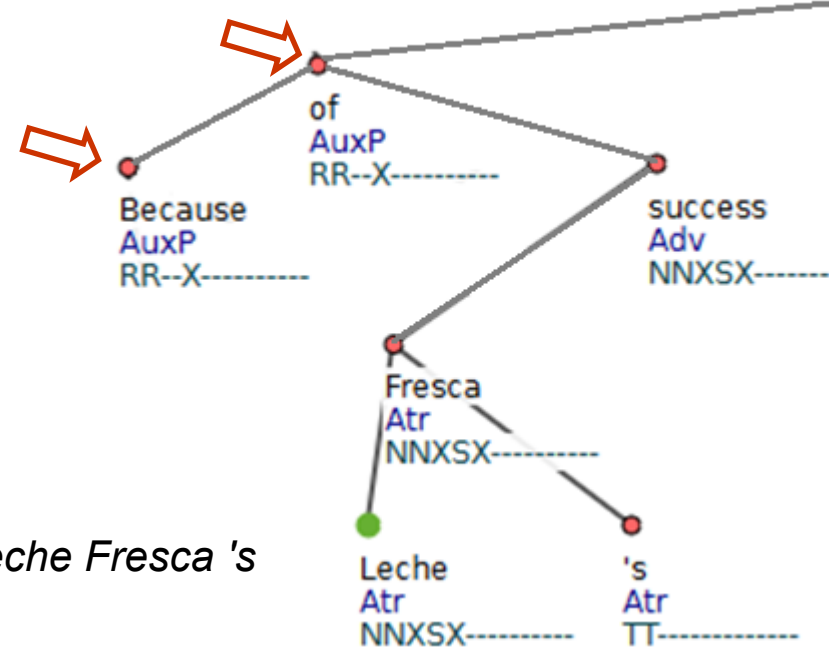
Auxiliary Sentence Members

AuxC … subordinate conjunctions

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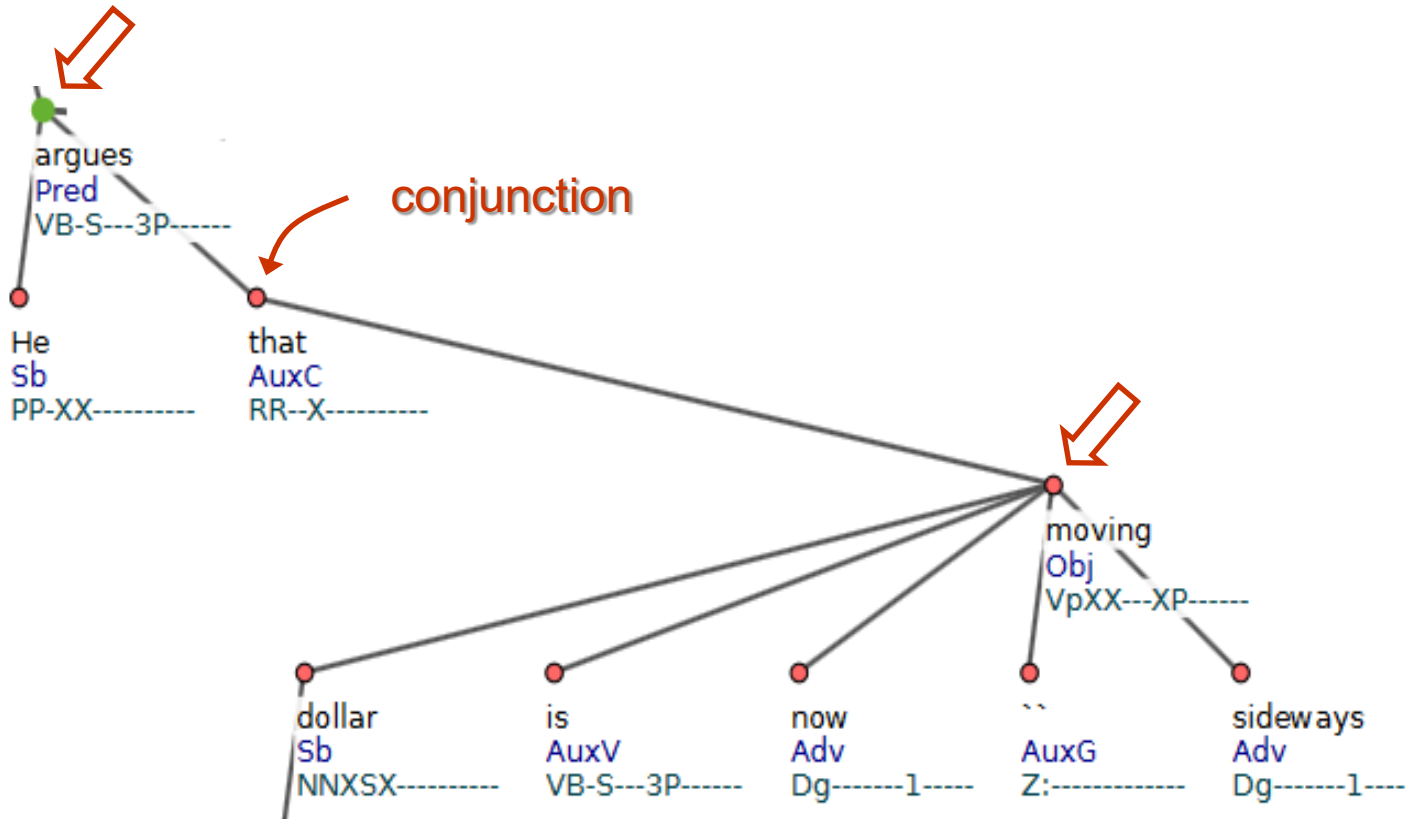- conjunction … *root of the subordinate clause*

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

… she felt as if 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

mluvili
???

AuxP

problému
Obj

zajímavém
Atr

když
AuxC

,
AuxX

i
AuxY

důležitém
Adv
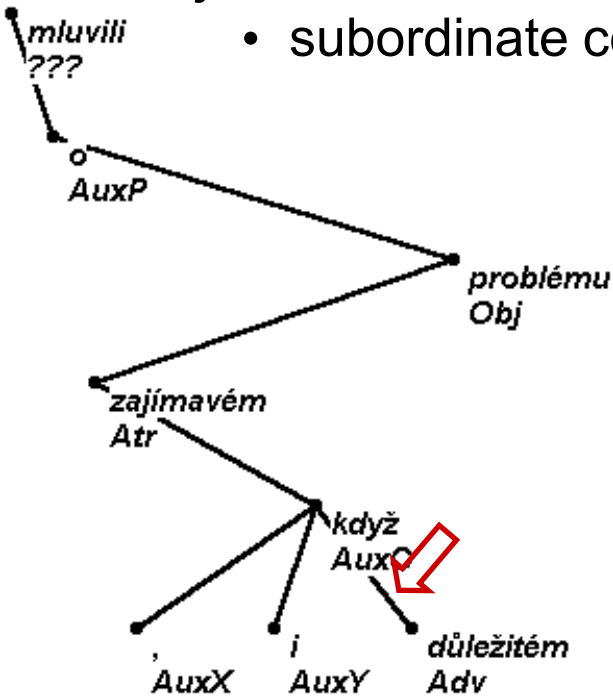
… (they) talked about (an) attractive, even though crucial topic

Consequence for subordinated clauses

conjunction

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-----

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

Consequence for subordinated clauses

the same for **preposition**

NOT for **wh-words**
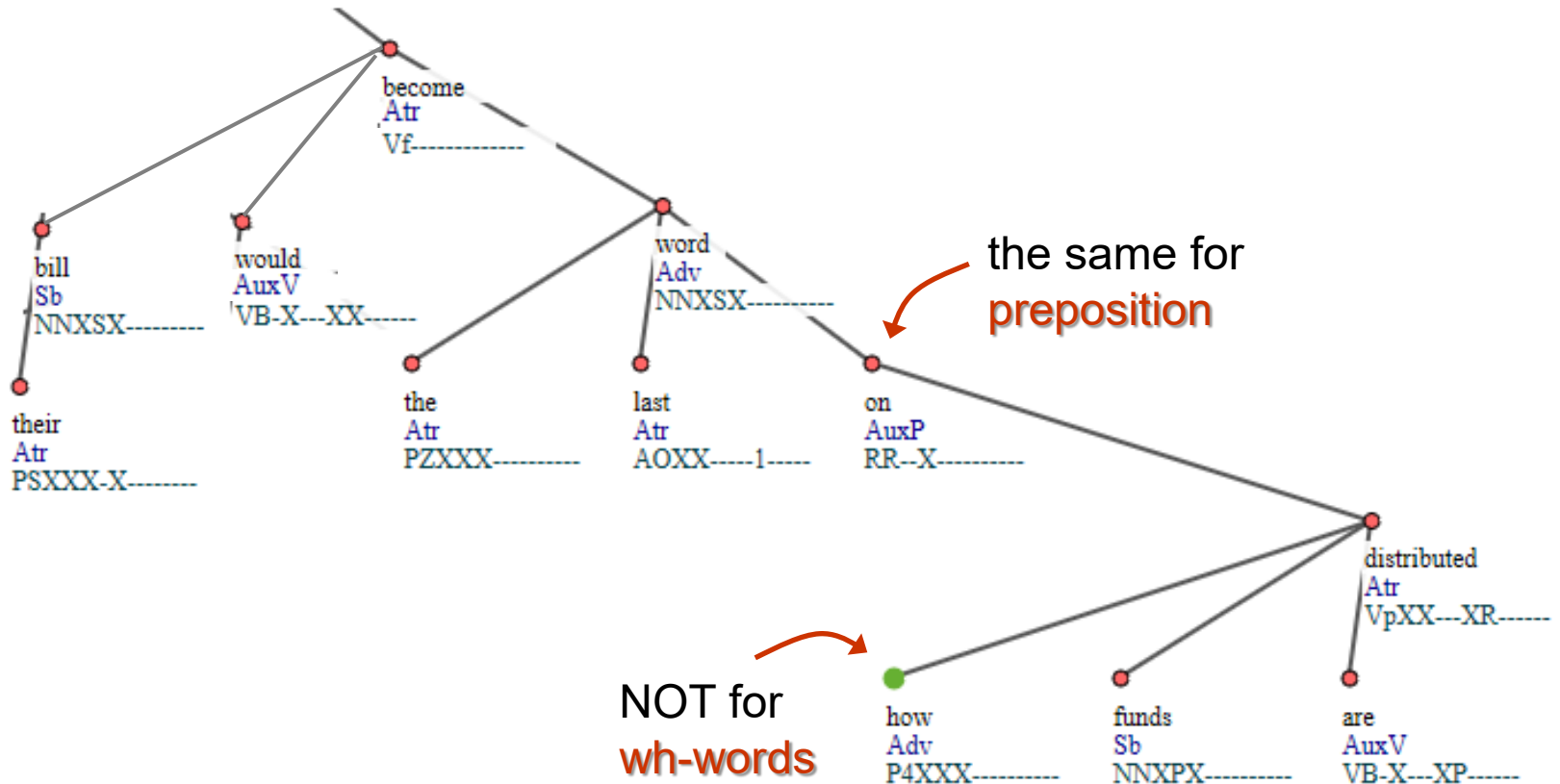
… their bill would become the last word on how funds are distributed .

Auxiliary Sentence Members

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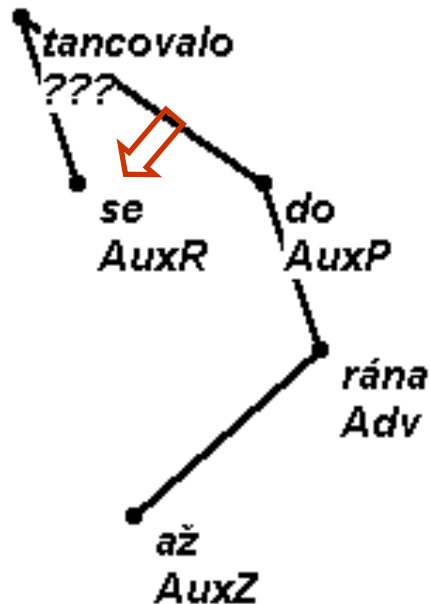
```
 boji
 ???

     se         ho
   AuxT        Obj
```
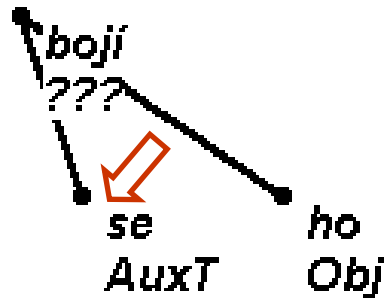
(they) - are-affraid - Refl - him
they are afraid of him

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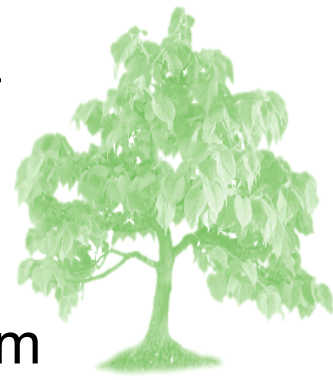
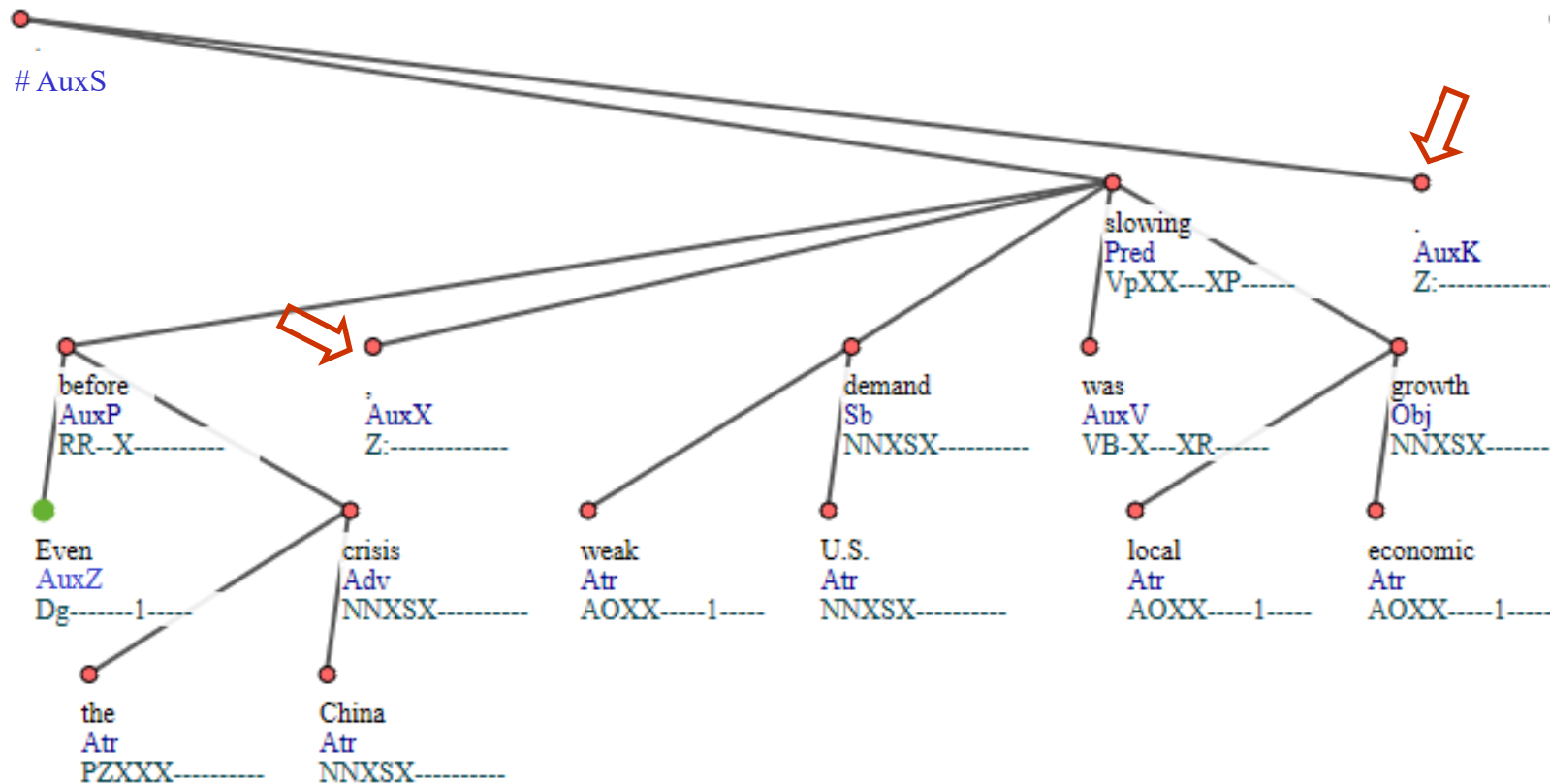*AuxR* … morphemes *se* as a part of a reflexive verbal form

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

Auxiliary Sentence Members

- punctuation … *AuxX*
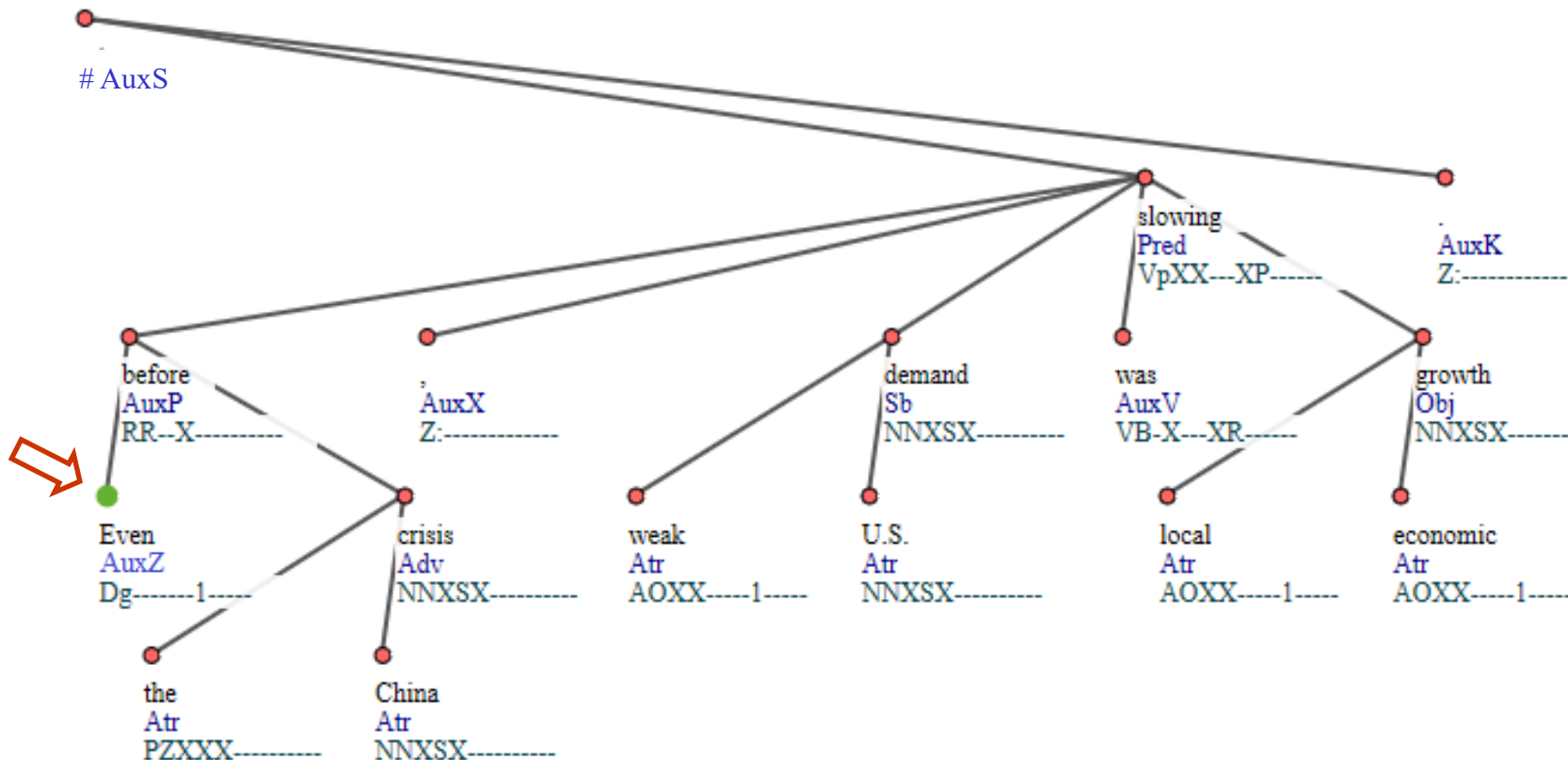- other graphical symbols … *AyxG*
- clause final punctuation … *AuxK*

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], …
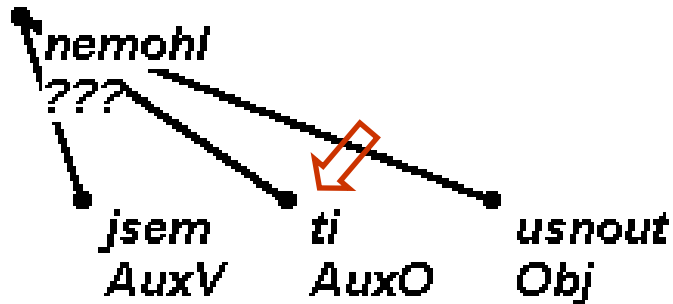
Auxiliary Sentence Members

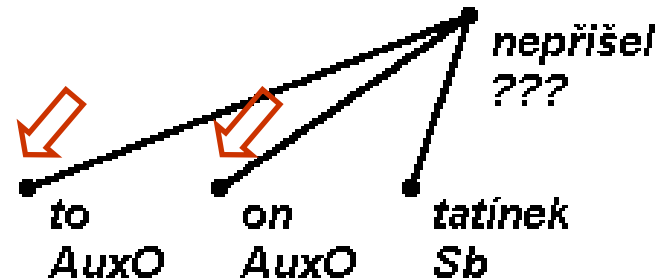***AuxO*** … emotional particles
mi, vám, si, … to, ono, …

nemohl
???

jsem
AuxV

ti
AuxO

usnout
Obj

nepřišel
???

to
AuxO

on
AuxO

tatínek
Sb

(I) could not - Aux - you - fall asleep
I couldn't fall asleep, you see.
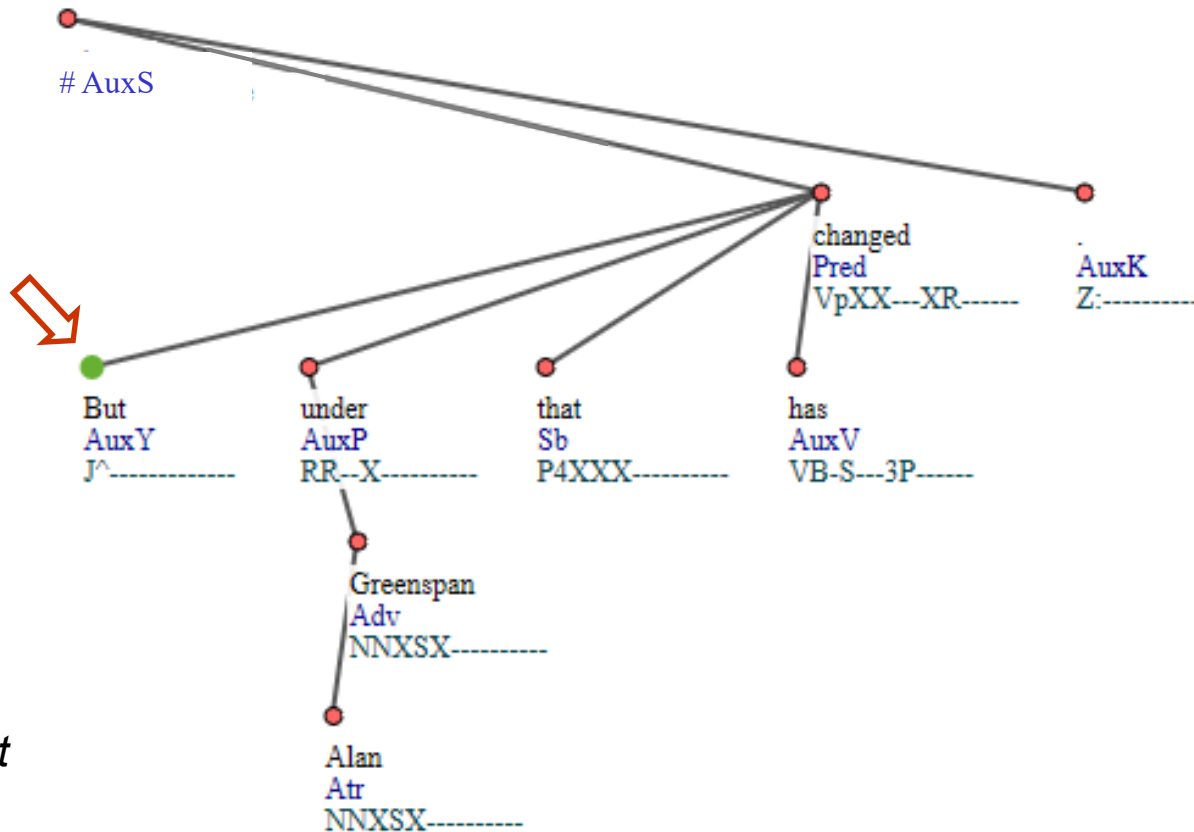
it - he - not-come - papa
obviously, papa failed to come

Auxiliary Sentence Members

AuxY … particles modifying the whole sentence;

AuxS

changed
Pred
VpXX---XR------

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

But
AuxY
J^-------------

under
AuxP
RR--X----------

that
Sb
P4XXX----------

has
AuxV
VB-S---3P------

Greenspan
Adv
NNXSX----------

Alan
Atr
NNXSX----------

They claimed that the Fed would first give a green light to the economy by making credit readily available and then turn on the red and bring growth to a screeching halt. ***But*** *under Alan Greenspan that has changed .*

Auxiliary Sentence Members

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

Principles of Annotation: Non-dependency

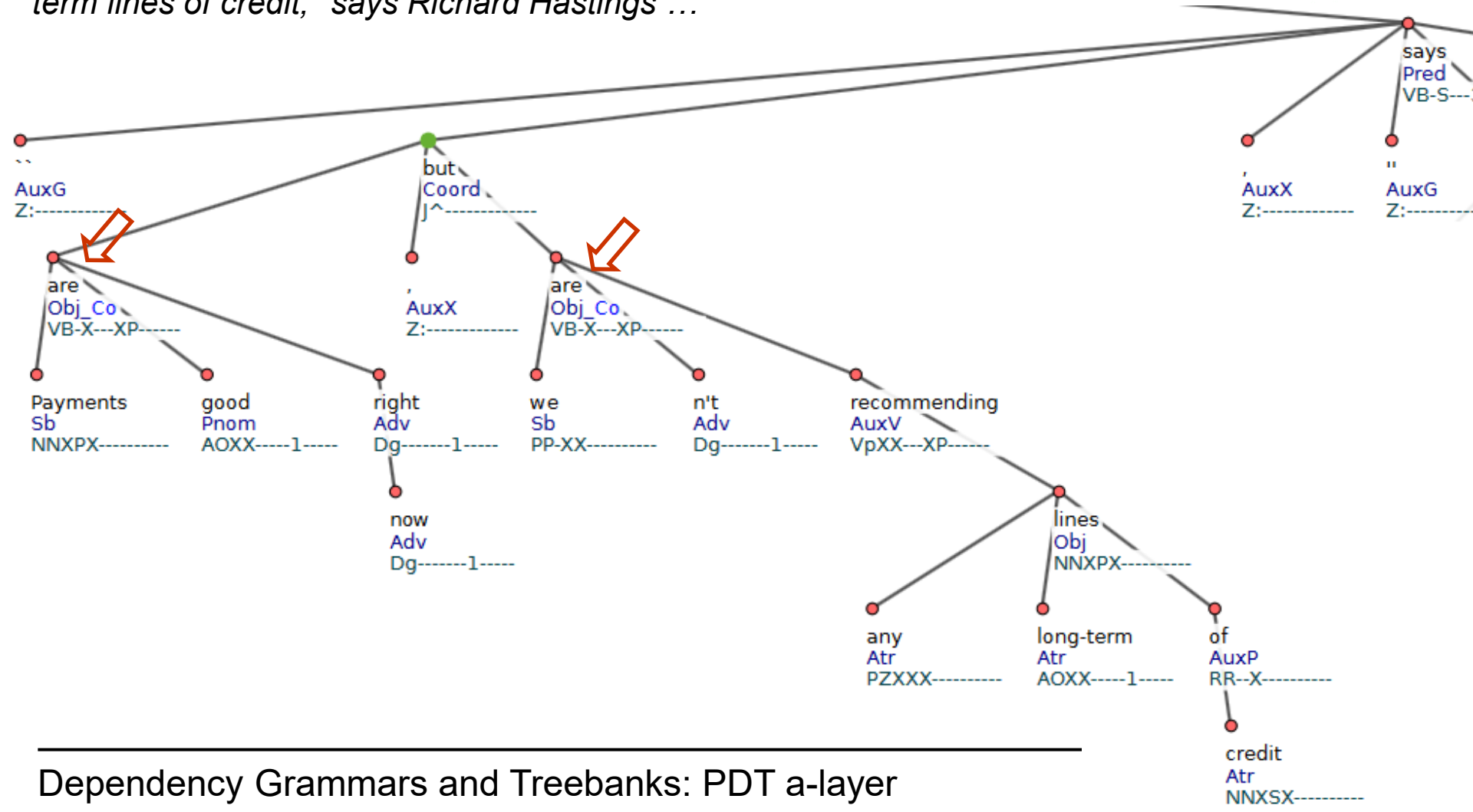- coordination
- apposition
- parenthesis

Coordination

- "multiplication" of a single syntactic position (sentence member / sentence)
- **Coord** … 'connecting node'
  - primarily: - coordinating conjunctions (*and, or, thus, …*)
    - connecting expressions (*and so on, et cetera, …*)
    - comma
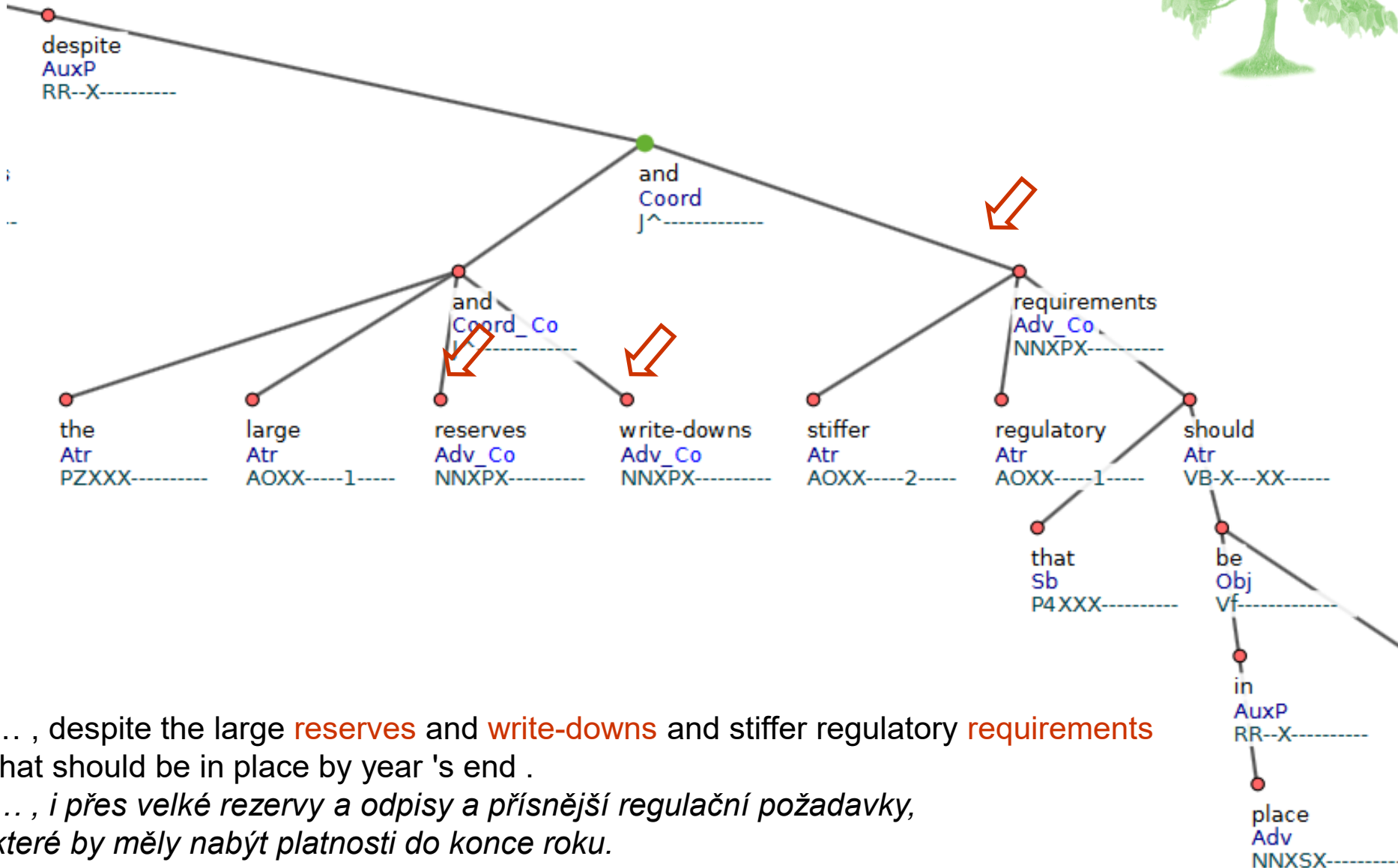- visualization … suffix_**Co**
    … is_member in the data

Non-dependency: Coordination

"Payments are good right now , but we are n't recommending any long-term lines of credit," says Richard Hastings …

Dependency Grammars and Treebanks: PDT a-layer

Non-dependency: Nested Coordination

… , despite the large reserves and write-downs and stiffer regulatory requirements that should be in place by year 's end .

… , i přes velké rezervy a odpisy a přísnější regulační požadavky, které by měly nabýt platnosti do konce roku.
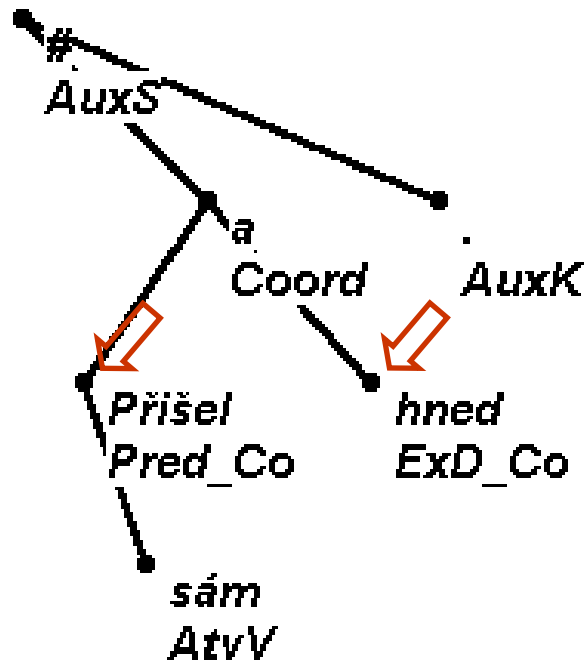
Non-dependency: Coordination (cont.)

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

  • if not possible … **ExD_Co**

(He) came alone and immediately

Non-dependency: Apposition

analytical function: **Apos**

- "multiplication" of a single syntactic position (the same referent)
- members of an Apposition … interchangable
    - 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], …*

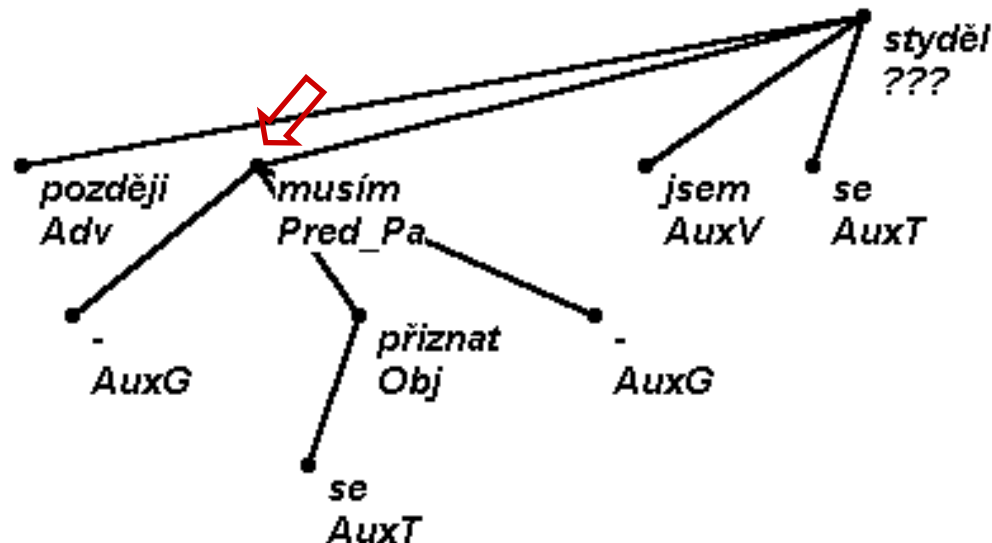- visualization suffix _Ap   vs. is_member in the data

Non-dependency: Parenthesis

an additional adjunction to the statement in the sentence
- formally <u>marked by some graphic sign(s)</u>
    - bracket, dashes, commas, …
- is_parenthesis_root in the data
- Cz: visualization: suffix _Pa
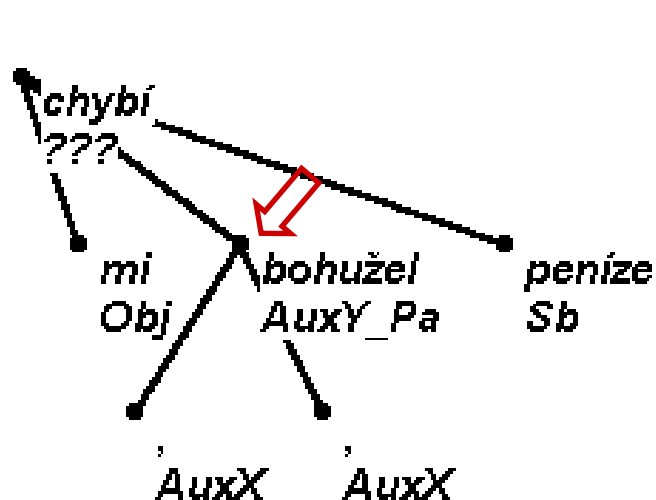
I am, unfortunately, short of money

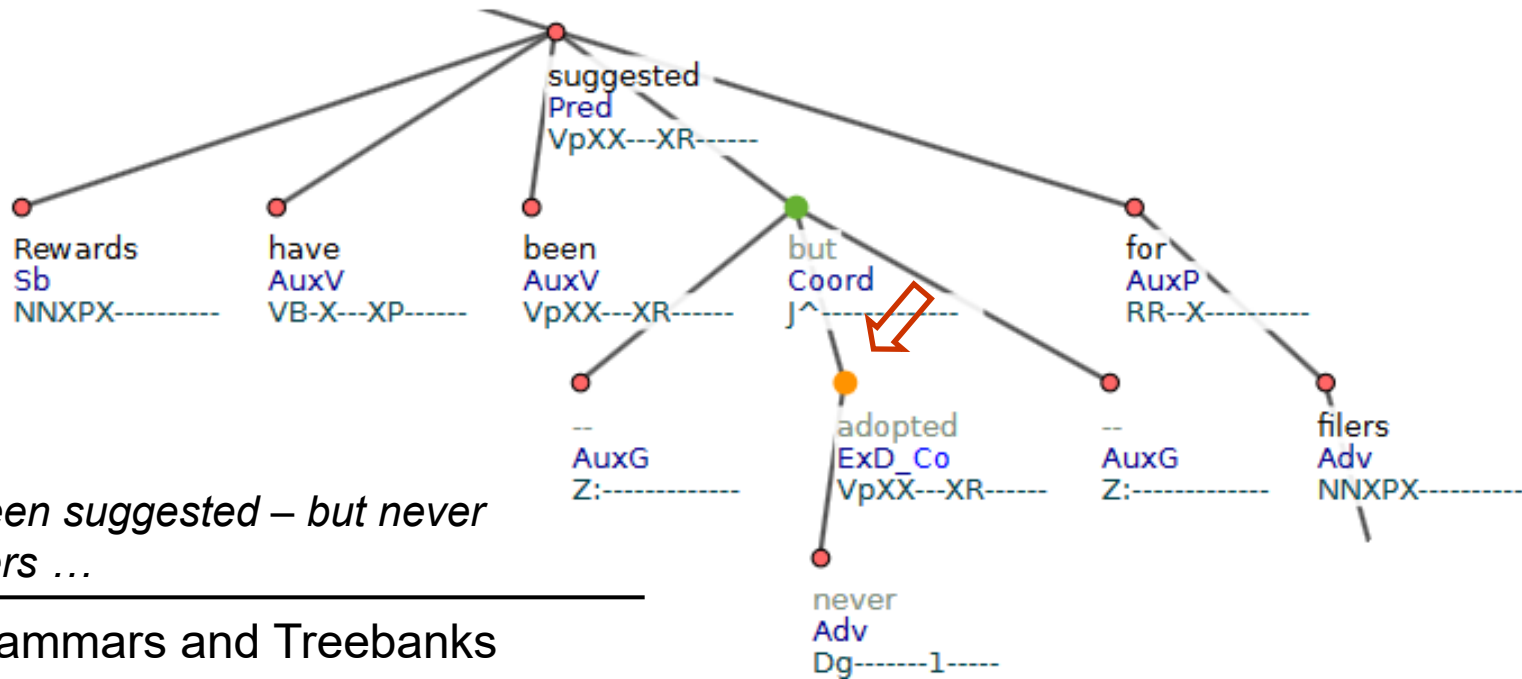*later - I have to admit – I was ashamed*

Non-dependency: Parenthesis

an additional adjunction to the statement in the sentence
- formally <u>marked by some graphic sign(s)</u>
  - bracket, dashes, commas, …
- visualization: suffix _Pa   vs. is_parenthesis_root in the data

BUT: HamleDT English … as ExD_Co (ellipsis, coordination)

Rewards have been suggested – but never
adopted – for filters …

Dependency Grammars and Treebanks

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)

- actual (textual) ellipses
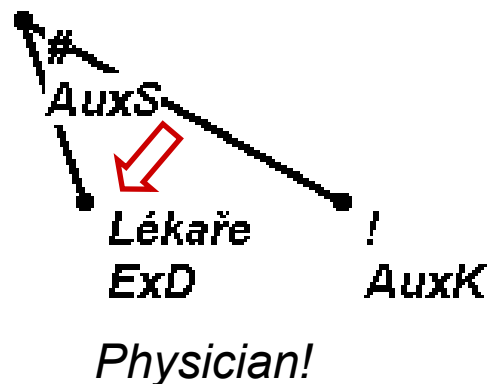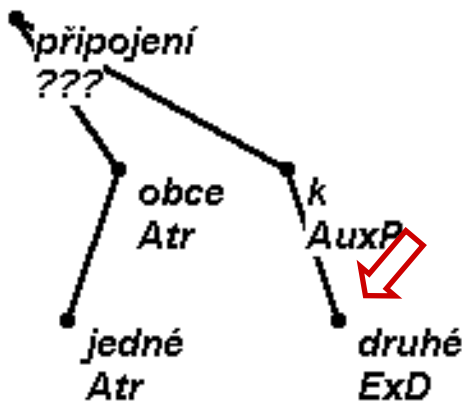
(the) attachment of one village to another (= another village)

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)

- actual (textual) ellipses   vs.   lexicalized ellipses

musím
???
domů
Adv

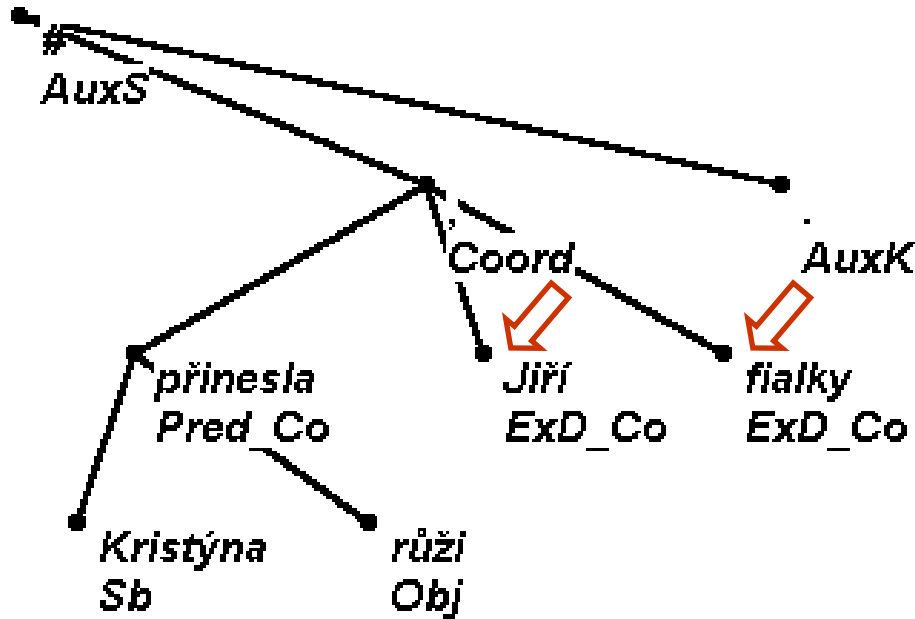*jsou*
???
tři
Sb

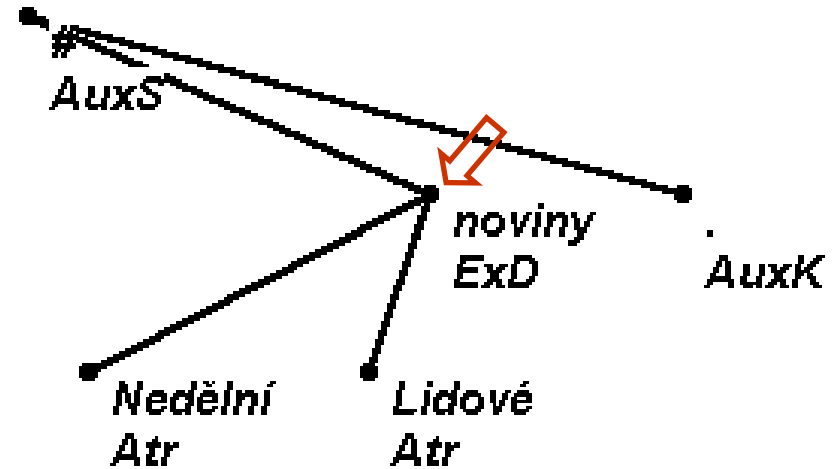*(I) have to (go) home.*   *(It) is five (o'clock)*

Principles of Annotation: Ellipses (cont.)

ellipsis

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

Christine brought a rose, George violets.　　　*Sunday Lidové noviny.*

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

- Štekauer, P., ed. (2000) *Rudiments of English Linguistics*.Slovacontact, Prešov (chapter 4, Syntax)

- Manual for Analytical Annotation http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/a-layer/html/index.html

- Quirk, R., Greenbaum, S., Leech, G., Svartvik, J. (1985) *A Comprehensive Grammar of the English Language*, Longman, London.