

Prague Dependency Treebank: Surface Syntactic Annotation

Markéta Lopatková

Institute of Formal and Applied Linguistics, MFF UK lopatkova@ufal.mff.cuni.cz

Prague Dependency Treebank

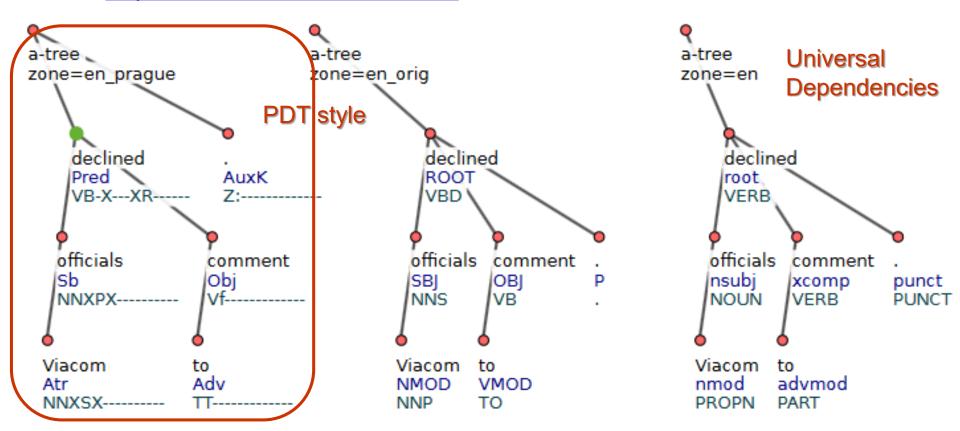
- application of a language theory onto a large set of data
- primarily focused on Czech running texts
- attempt to standardize dependency annotation

- 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),
 36 languages, 42 treebanks in HamleDT 3.0 (2015)

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



Prague Dependency Treebank - scenario

- 4 layers (for texts):
- word layer (w-layer)
- morphological layer (m-layer)
- analytical layer (a-layer)
- tectogrammatical layer (t-layer)

layers of annotation

PDT: a-layer

Goal:

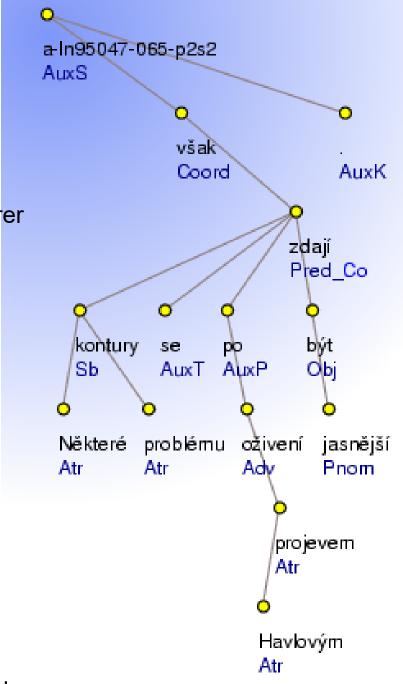
- to describe the structure of the sentence and
- to denote the type of relations between tokens
- dependency tree
- one token from m-layer ~ one node incl. prepositions, punctuation ...
 plus technical root
- relations ~ edges
 - dependency, coordination, punctuation, ...
 - type of relation: attribute 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 <u>oživením</u>
Havlovým projevem zdají být jasnější .

[Some contours of the problem seem to be clearer after the resurgence by Havel's speech.]



Dependency Grammars and Treebanks: PDT a-layer

A-layer: attributes of a root

| Attribute Name | Description | - ded |
|-------------------|---|-------|
| 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) | |

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

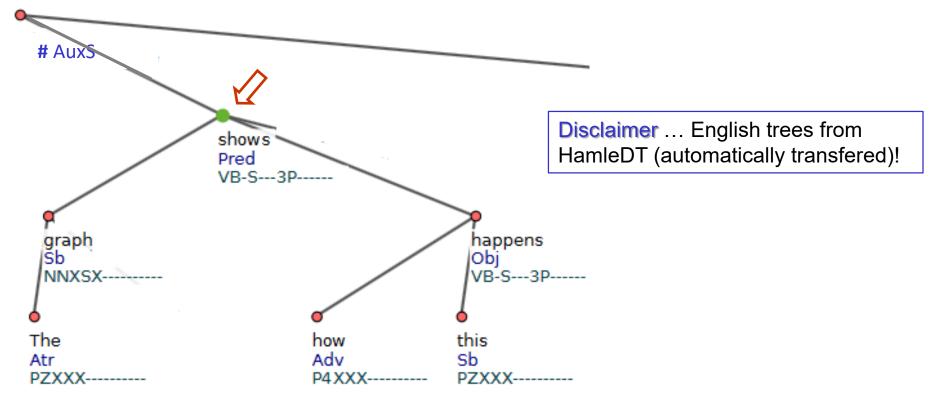
List of analytical functions (afun)

- sentence predicate ... Pred, Pnom, AuxV
- subject ... Sb
- object ... Obj
- adverbial ... Adv
- attribute ... Atr
 - combined functions ... AtrAtr, AtrObj, ObjAtr, AtrAdv, AdvAtr
- secondary predicate... 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 ... analytical functions: Pred, Pnom, AuxV

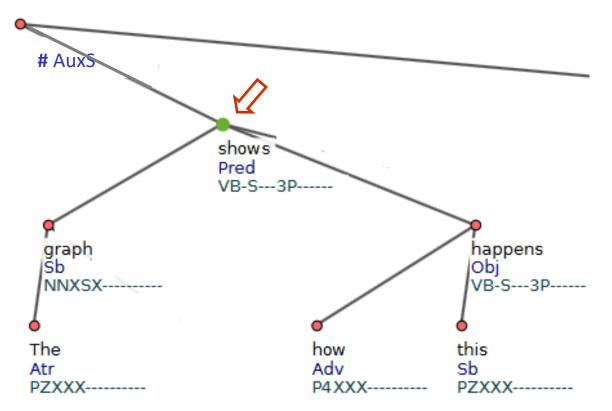
- attributing some property/state/change/activity to the subject
- basic sentence member
- governing node (= head) of its clause

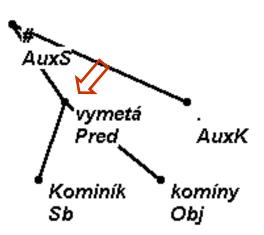
- a child of the technical root
- typically a finite verb form ... Pred



The graph shows how this happens ...

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

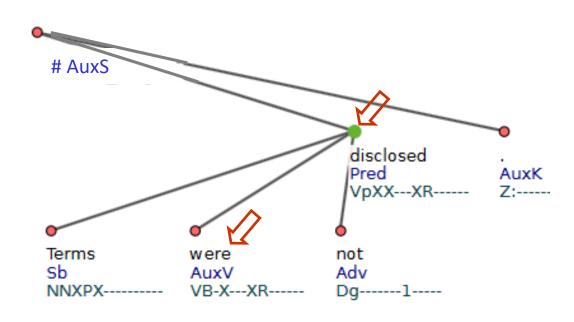




Chimneysweep sweeps chimneys.

The graph shows how this happens ...

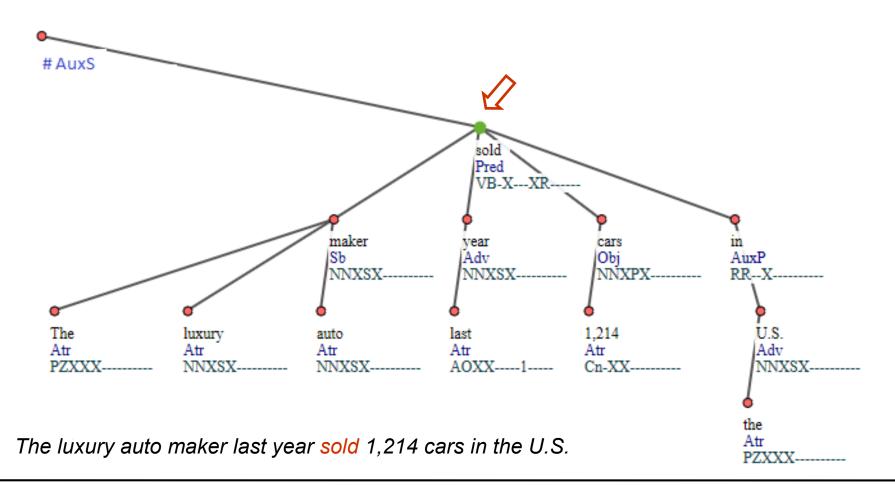
- a child of the technical root
- typically a finite verb form ... Pred
 - analytical verbal forms: auxiliary verbs ... AuxV



Terms were not disclosed.

(1) Simple verbal predicate: content verb

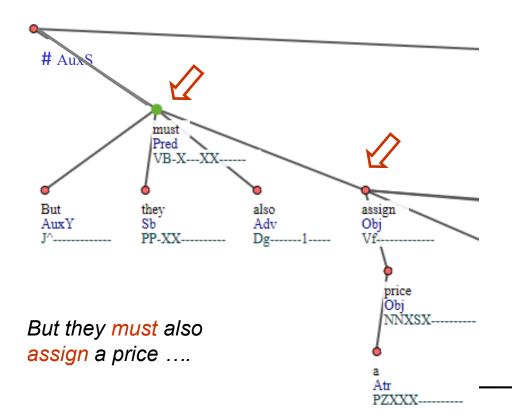


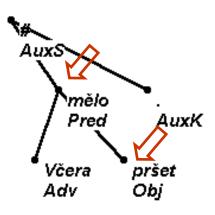


Dependency Grammars and Treebanks: PDT a-layer

(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], ...

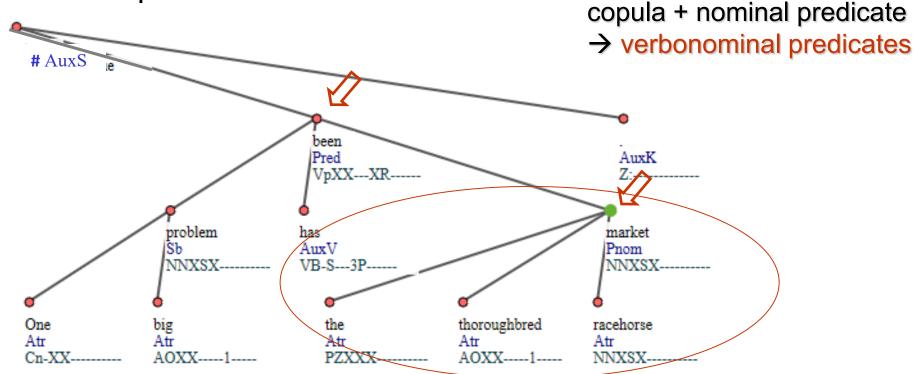




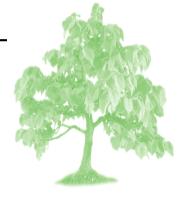
Yesterday (it) should rain.

(3) Copular constructions

- 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

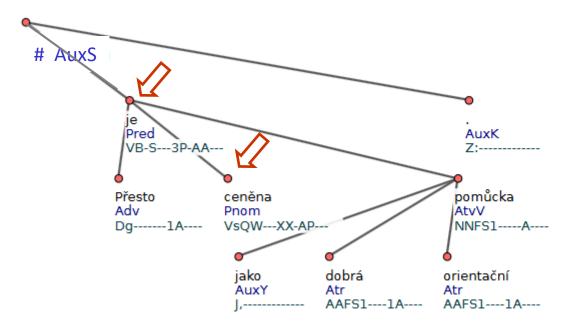


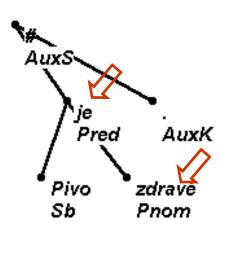
One big problem has been the thoroughbred racehorse market.



(3) Copular constructions

- 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





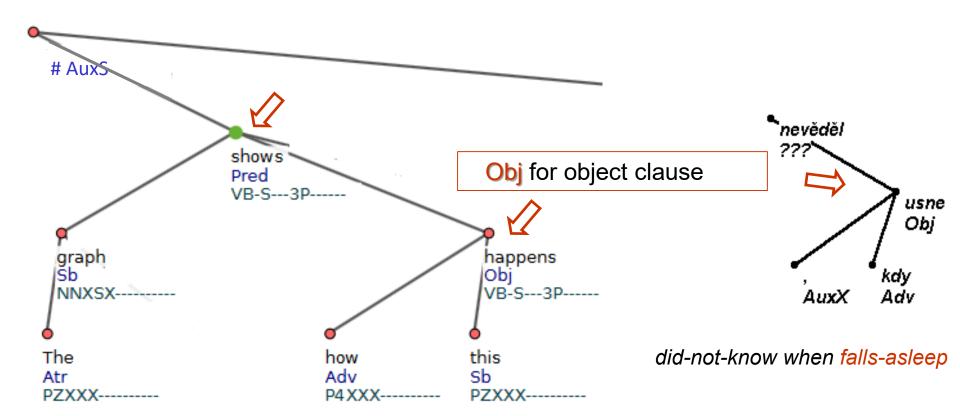
Beer is healthy.

Přesto je ceněna jako dobrá orientační pomůcka.

Dependency Grammars and Treebanks: PDT a-layer

Predicate of a subordinate clause

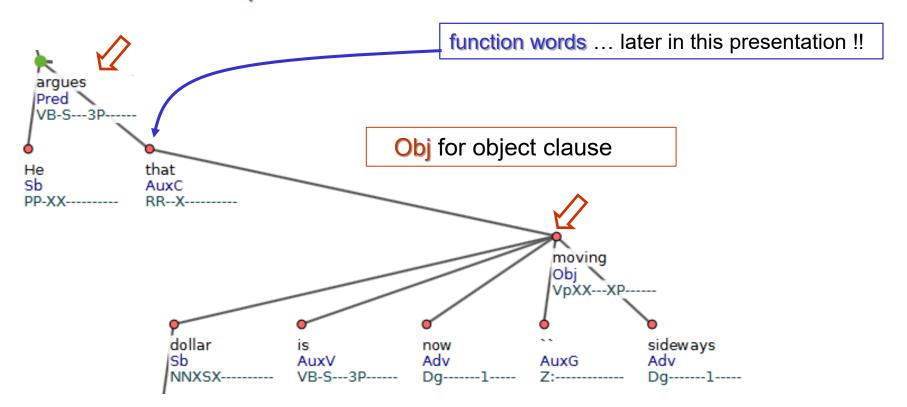
afun of the respective function



The graph shows how this happens ...

Predicate of a subordinate clause

afun of the respective function

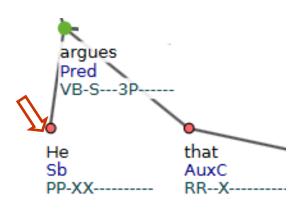


He argues that [...] dollar is now moving "sideways" ...

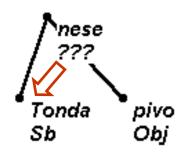
Subject

Subject ... analytical function: Sb

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



He argues that...



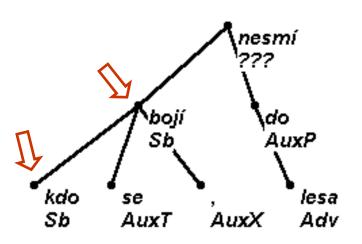
Tonda nese pivo.

Tony – brings – beer

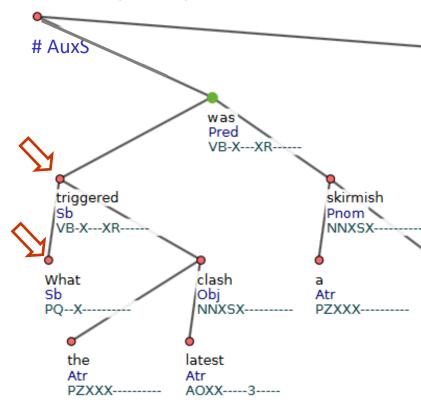
Subject

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

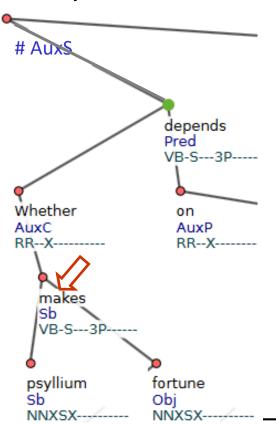


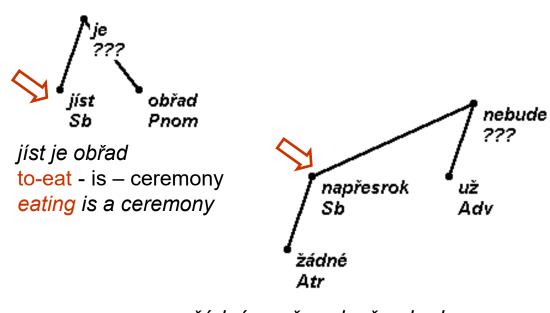
What triggered the latest clash was a skirmish ...

Subject

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

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



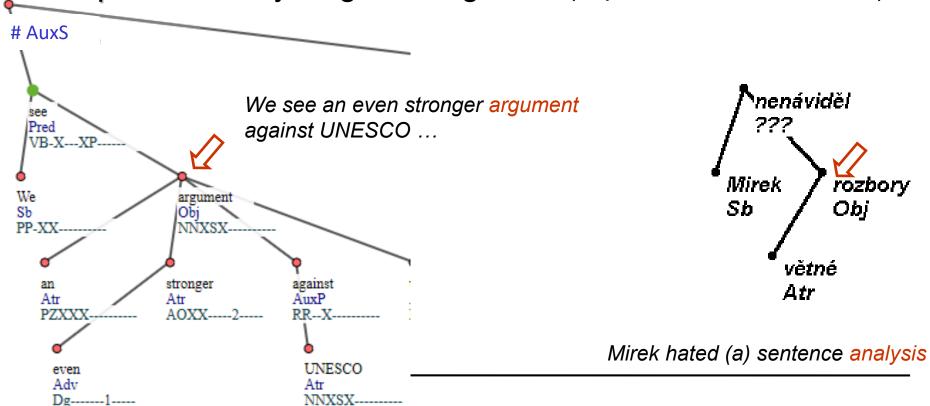


žádné napřesrok už nebude no - next-year - anymore - will-not-be

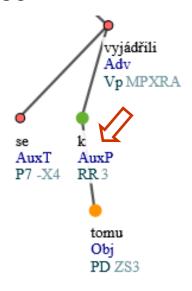
Whether psyllium makes fortune depends on

- **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)



- 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), prepositional case

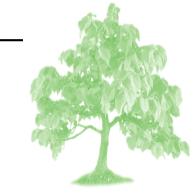


... they commented on it ...

- **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 AuxK Pred VpXX---XP projects push government Obi AuxV Obi NNXSX--The several use Obi PZXXX-----NNXSX PC Atr

The government is funding several projects to push PC use.



NNXSX----

- **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

věřím Cilka Sb To ! " AuxG Obj AuxK AuxG

"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

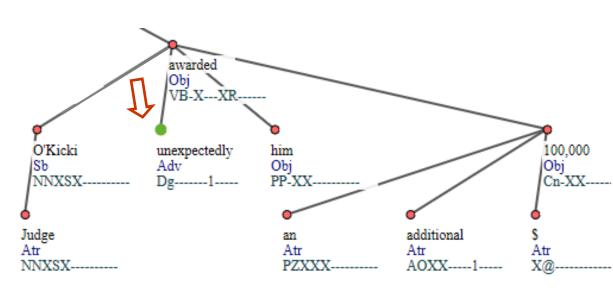
→ questions: where to?, from where?, ...

how long?, when?, ...

for what purpose?, why?, ...

how?, ...

[... he says] Judge O'Kicki unexpectedly awarded him an additional \$ 100,000



Adverbials

Adverbials ... analytical functions: Adv

- express the circumstances and relations, such as location, time, manner, comparison, extent, means, cause, consequence, regard or aim
- → questions: where?, where to?, from where?, ...

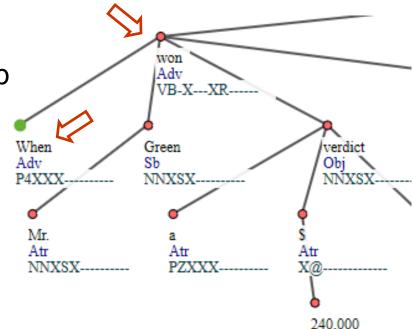
how long?, when?, ...

for what purpose?, why?, ...

how?, ...

can modify a verb, adjective or adverb

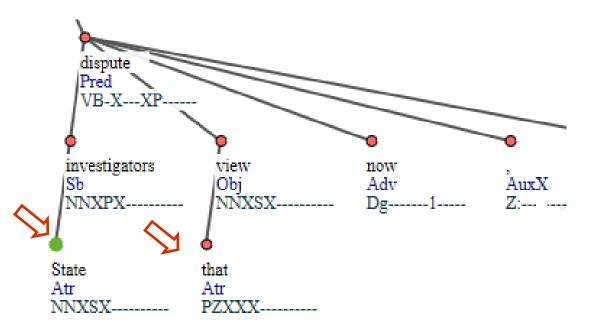
- a form of Adv not prescribed by its governing word!
- in PDT not further classified



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

Attribute

- Attribute ... analytical functions: Atr
- modifies a noun (with any function)
- answers the question ... which, what or whose (jaký, který, čí)



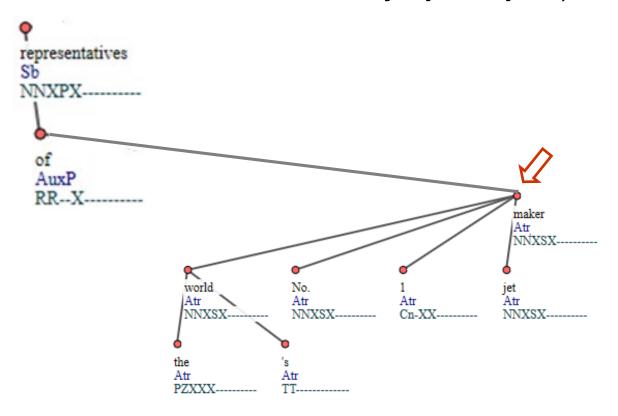
State investigators dispute that view now, ...

Dependency Grammars and Treebanks: PDT a-layer

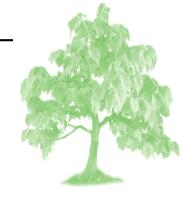


Attribute

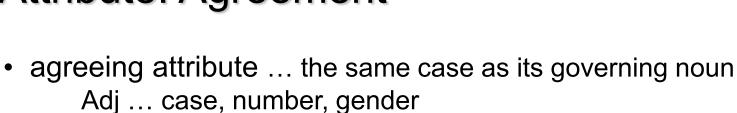
- **Attribute** ... analytical functions: Atr
- modifies a noun (with any function)
- answers the question ... which, what or whose (jaký, který, čí)



... representatives of the world 's No. 1 jet maker ...



Attribute: Agreement



BUT: kluku ušatá

písmena maminka
???

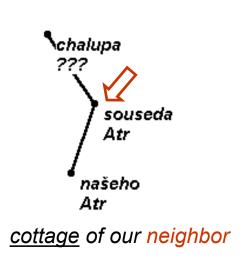
červená nebožka
Atr Atr

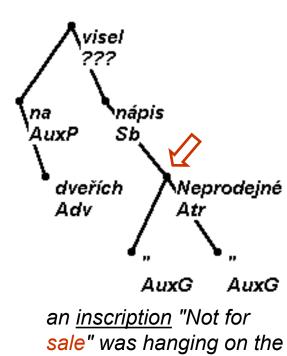
red letters the-dead mother

Dependency Grammars and Treebanks: PDT a-layer

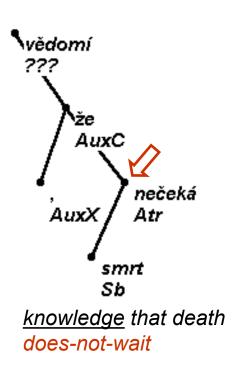
Attribute: Agreement

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





door



Attribute

Atr ... ALSO "technical" solution for

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

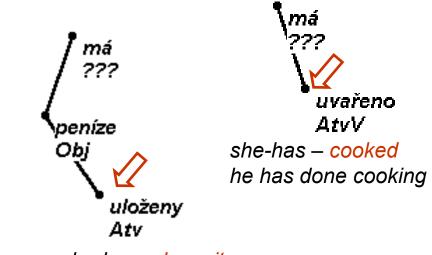


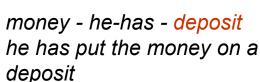
Secondary predication (Cz: doplněk)

Secondary predication (verbal attribute, complement ... analytical functions: Atv, AtvV

 modifies two sentence members, verb and noun She declared the cake beautiful. She entered the room sad.

přišli ???? jsme Sb AuxV rádi Atv we - AuxV - came - glad we were glad to come



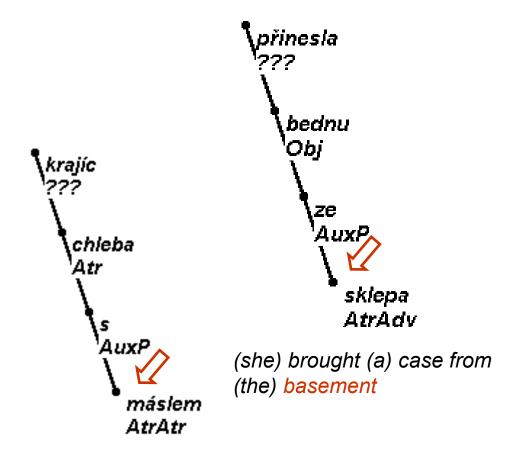


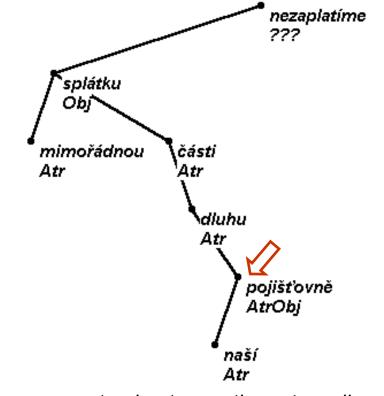


he-arrived - barefooted

Combined Functions

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





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

slice of bread with butter

Dependency Grammars and Treebanks: PDT a-layer

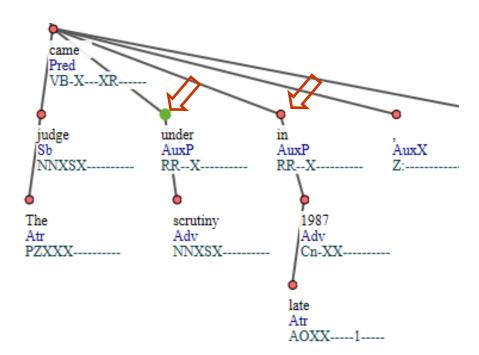
Auxiliary Sentence Members

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



AuxP ... prepositions

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

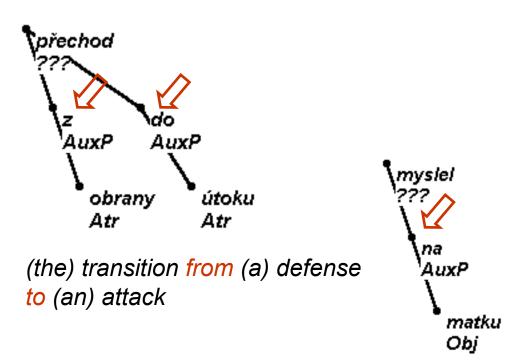


The judge came under scrutiny in late 1987, ...



AuxP ... prepositions

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



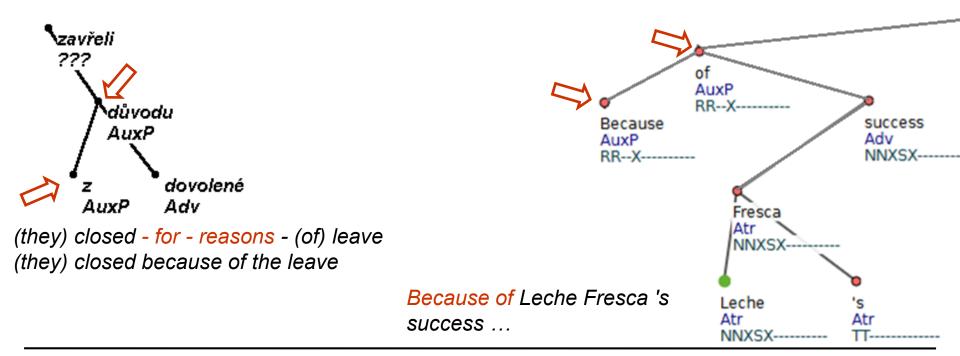


(they) negotiated - about - (the) heritage

(he) thought of (his) mother

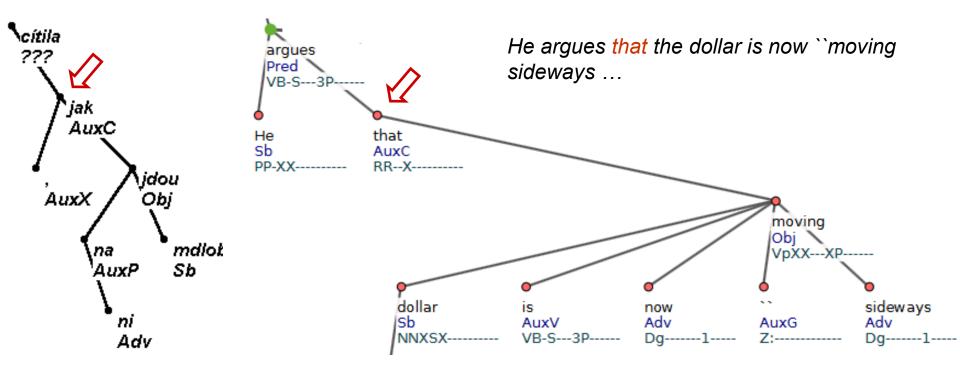
Dependency Grammars and Treebanks: PDT a-layer

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



Dependency Grammars and Treebanks: PDT a-layer

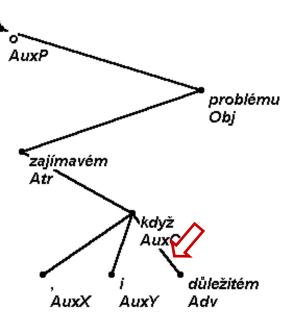
- **AuxC** ... subordinate conjunctions
- subordinate clause introduced by a subordinate conjunction
- conjunction ... root of the subordinate clause



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

AuxC ... subordinate conjunctions

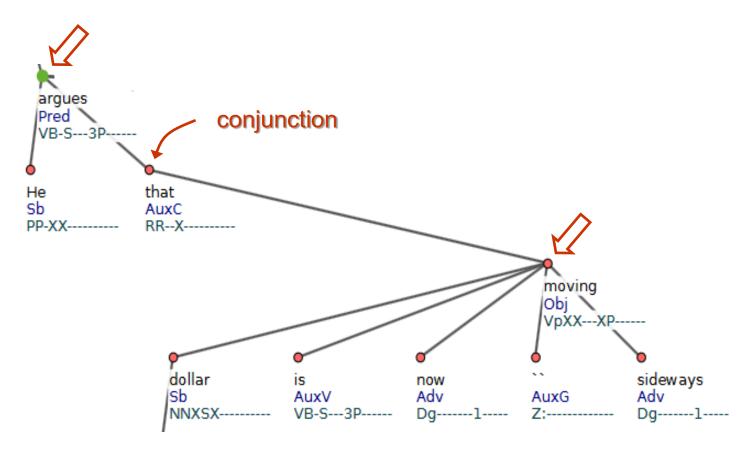
- subordinate clause introduced by a subordinate conjunction
- conjunction ... root of the subordinate clause
- subordinate conjunction may attach an individual sentence member



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



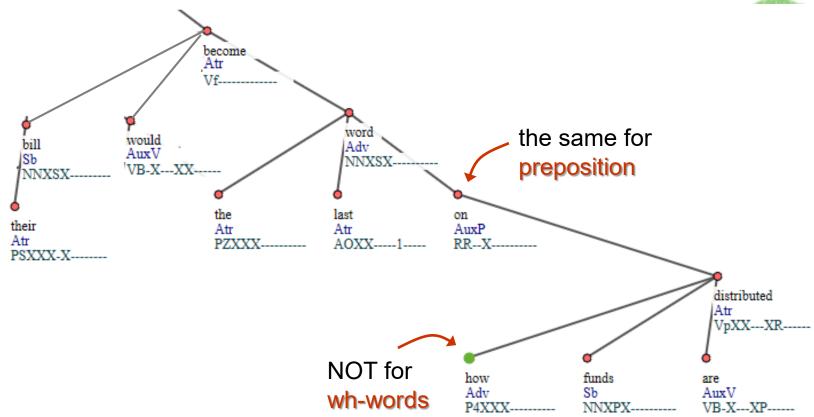




He argues that the dollar is now ``moving sideways ...

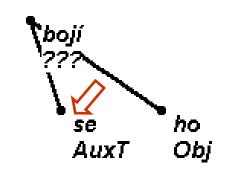
Consequence for subordinated clauses



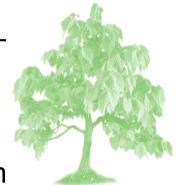


... their bill would become the last word on how funds are distributed .

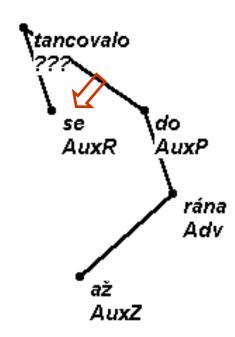
AuxT ... morphemes se, si as a part of a reflexive tantum



(they) - are-affraid - Refl - him they are afraid of him



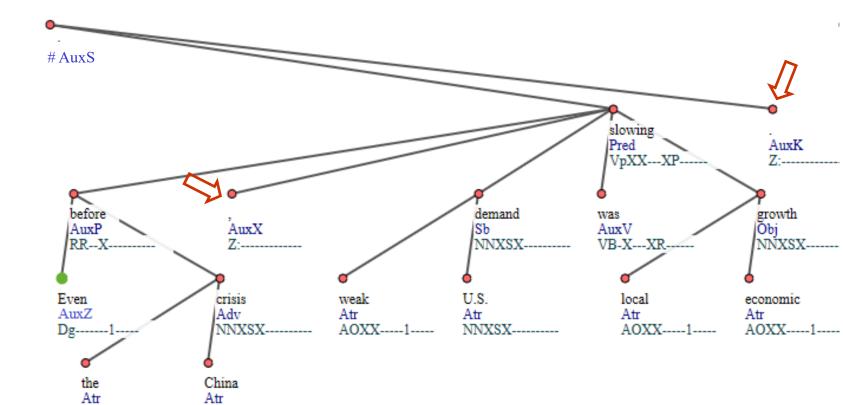
AuxR ... morphemes se as a part of a reflexive verbal form



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

- punctuation ... AuxX
- other graphical symbols ... AyxG
- clause final punctuation ... AuxK

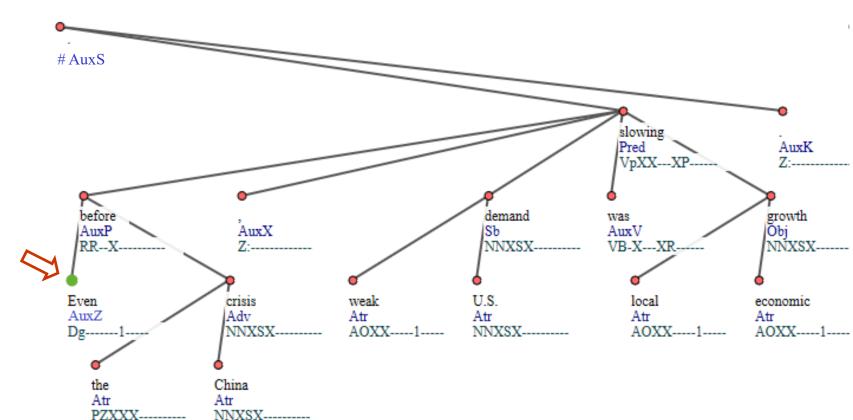
NNXSX-



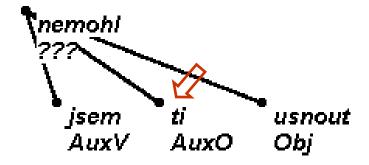


AuxZ ... emphasizing words (modifying a sentence member)

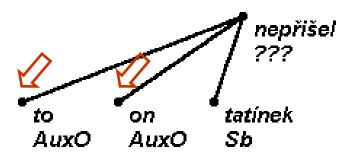
e.g., asi [maybe, approx.], dokonce [even, as far as], hlavně [mainly], ještě [moreover], již [yet], leda [only], především [most of all, most probably], zvláště [especially], ...



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



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



it - he - not-come - papa obviously, papa failed to come

AuxY ... particles modifying the whole sentence;

AuxS



changed Pred VpXX---XR AuxP AuxY AuxV VB-S---3P-----They claimed that the Fed would first give a green light to the Greenspan economy by making credit readily Adv available and then turn on the red NNXSX----and bring growth to a screeching halt. But under Alan Greenspan that Alan Atr has changed. NNXSX----

Dependency Grammars and Treebanks: PDT a-layer

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

Principles of Annotation: Non-dependency

- coordination
- apposition
- parenthesis

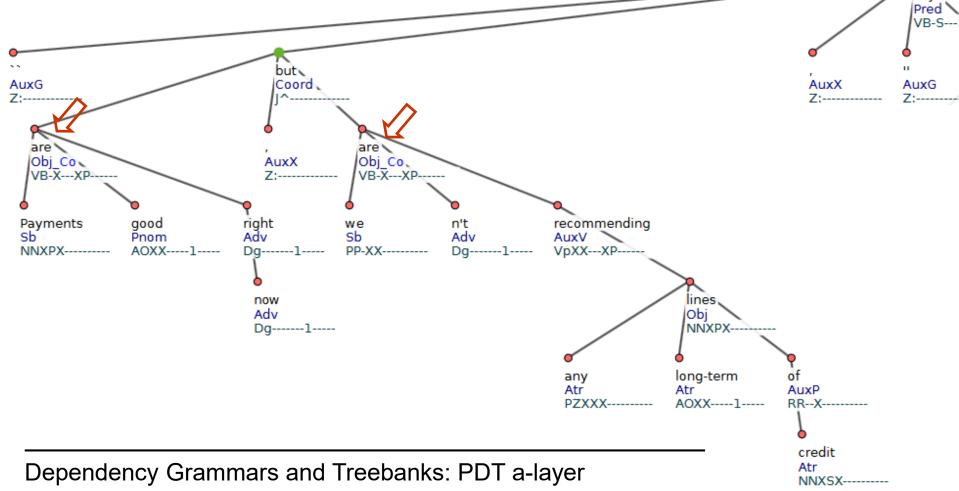
Coordination

- "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



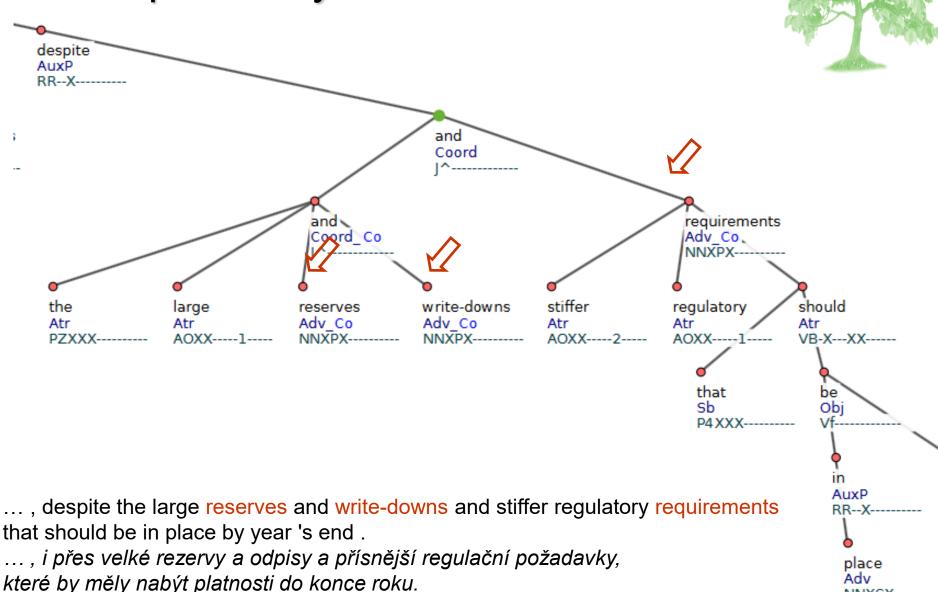
Non-dependency: Coordination

"Payments are good right now, but we are n't recommending any longterm lines of credit," says Richard Hastings ...



says

Non-dependency: Nested Coordination

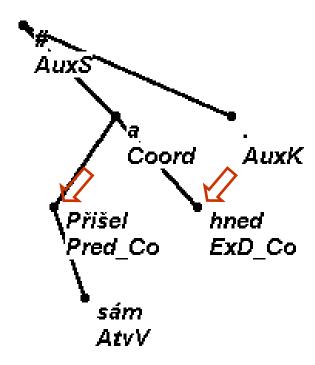


NNXSX---

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

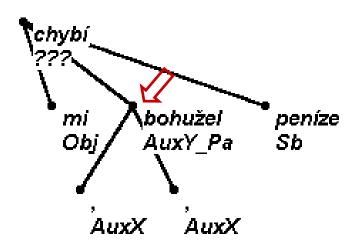
analytical function: Apos

- "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

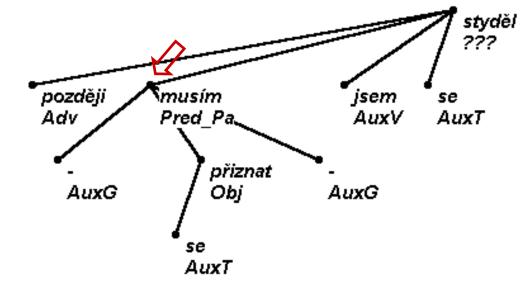
Non-dependency: Parenthesis

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



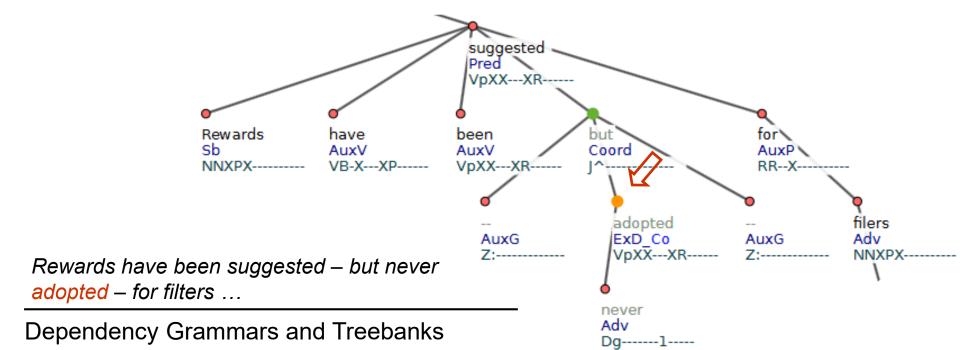
later - I have to admit - I was ashamed

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

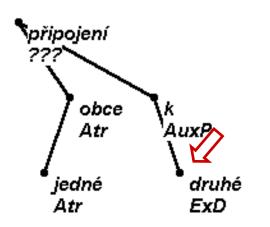
BUT: HamleDT English ... as ExD_Co (ellipsis, coordination)

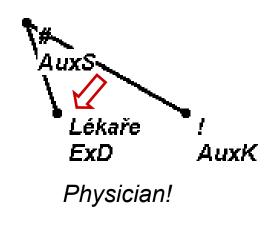


Principles of Annotation: Ellipses

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

- ExD ... all nodes that would depend on the deleted node
- 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



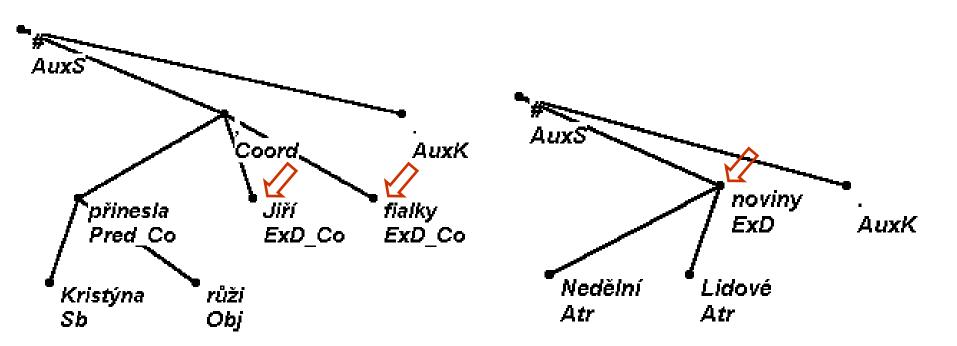
(I) have to (go) home.

(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.

Dependency Grammars and Treebanks: PDT a-layer

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.