

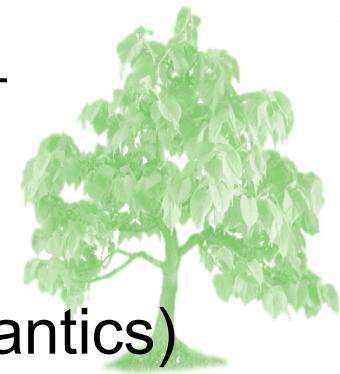


PDT – Tectogrammatical Layer (Intro)

Markéta Lopatková

Institute of Formal and Applied Linguistics, MFF UK

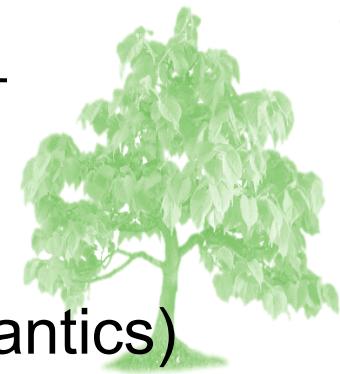
lopatkova@ufal.mff.cuni.cz



PDT: t-layer

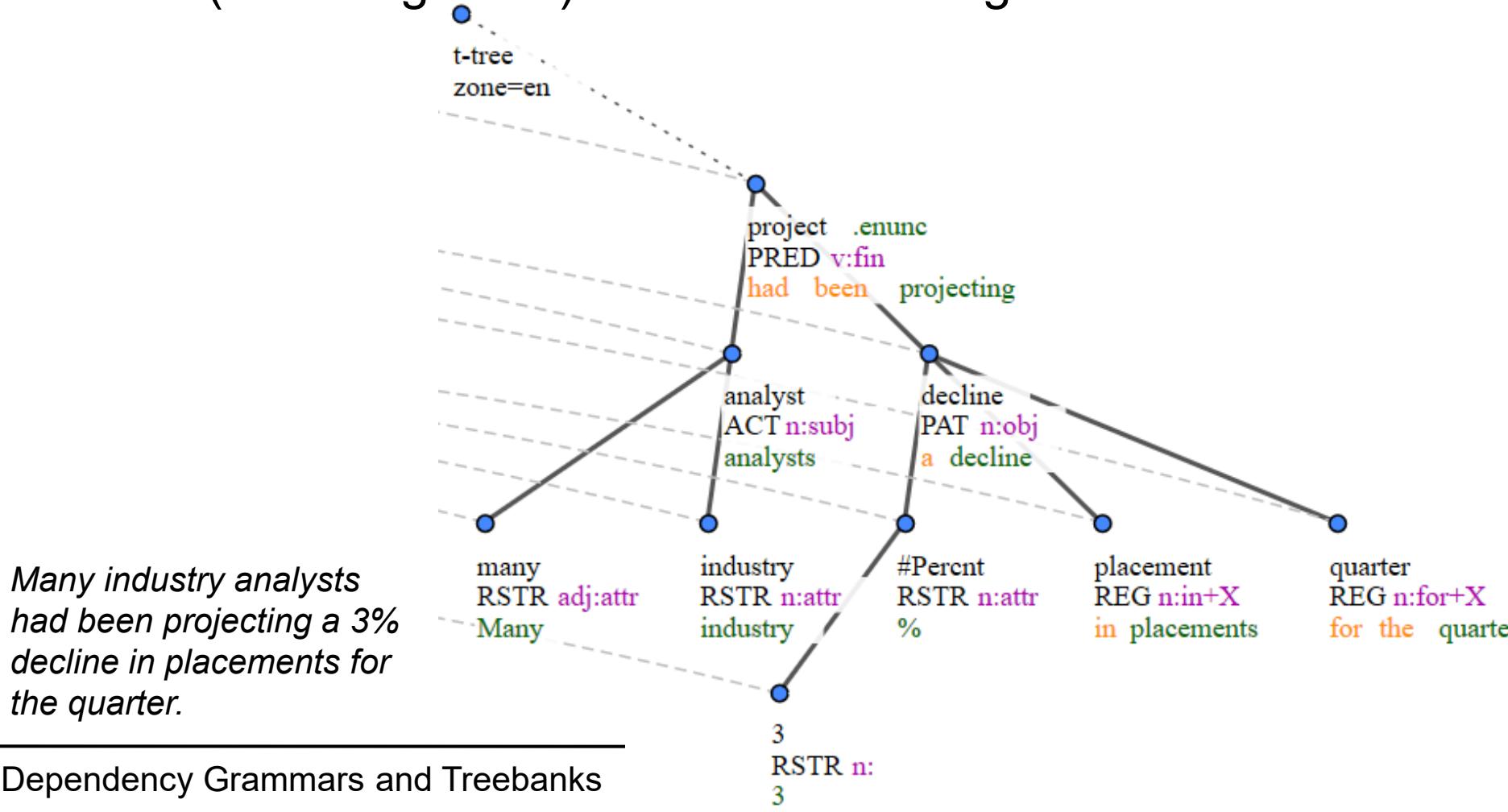
- deep syntactic structure of a sentence (coarse semantics)
 - ~ (disambiguated) sentence meaning
- dependency-based 'trees'
 - with (phrase-structure like) 'connecting' constructions for coordination
 - with coreference
- disambiguated ... synonymy vs. ambiguity
- 'completed' ... elided words added / restored / copied
 - 'missing' argument
 - coordination: stripping and gapping
- only lexical words as separate nodes
 - auxiliary and modal verbs
 - prepositions, subordinating conjunctions, ...
 - BUT coordinating conjunctions
 - plus technical root

documentation: <http://ufal.mff.cuni.cz/pdt3.5>



PDT: t-layer

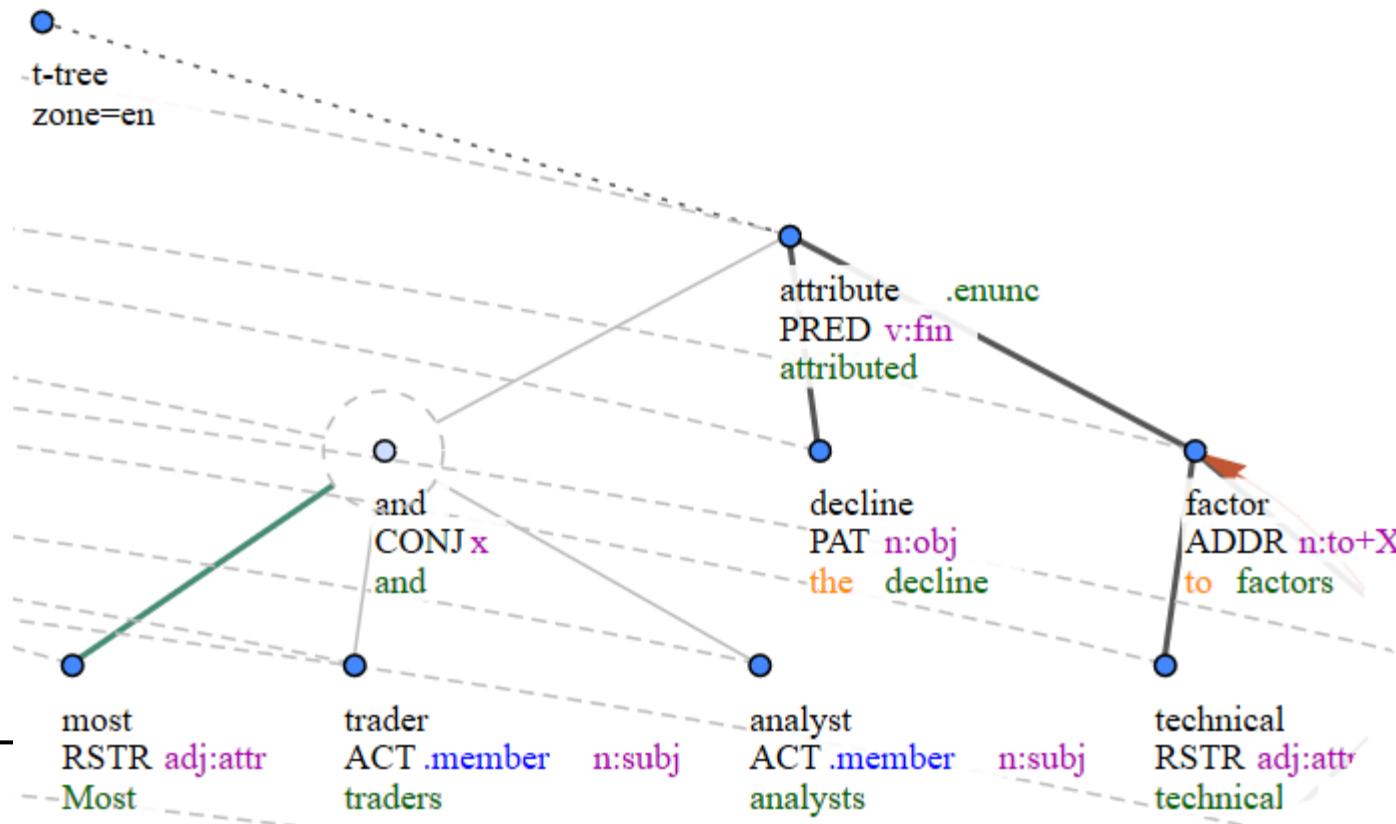
- deep syntactic structure of a sentence (coarse semantics)
~ (disambiguated) sentence meaning





PDT: t-layer

- dependency-based 'trees'
 - with (phrase-structure like) 'connecting' constructions for coordination

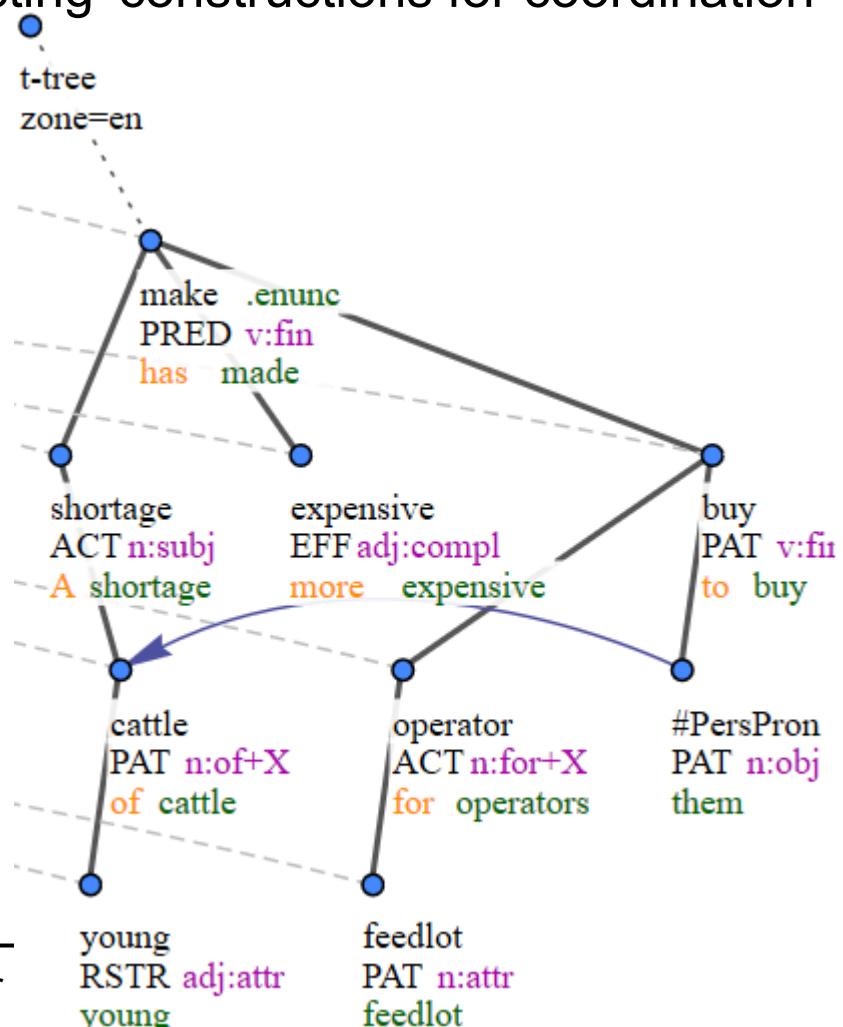


Most traders and analysts attributed the decline to technical factors ...



PDT: t-layer

- dependency-based 'trees'
 - with (phrase-structure like) 'connecting' constructions for coordination
 - with coreference

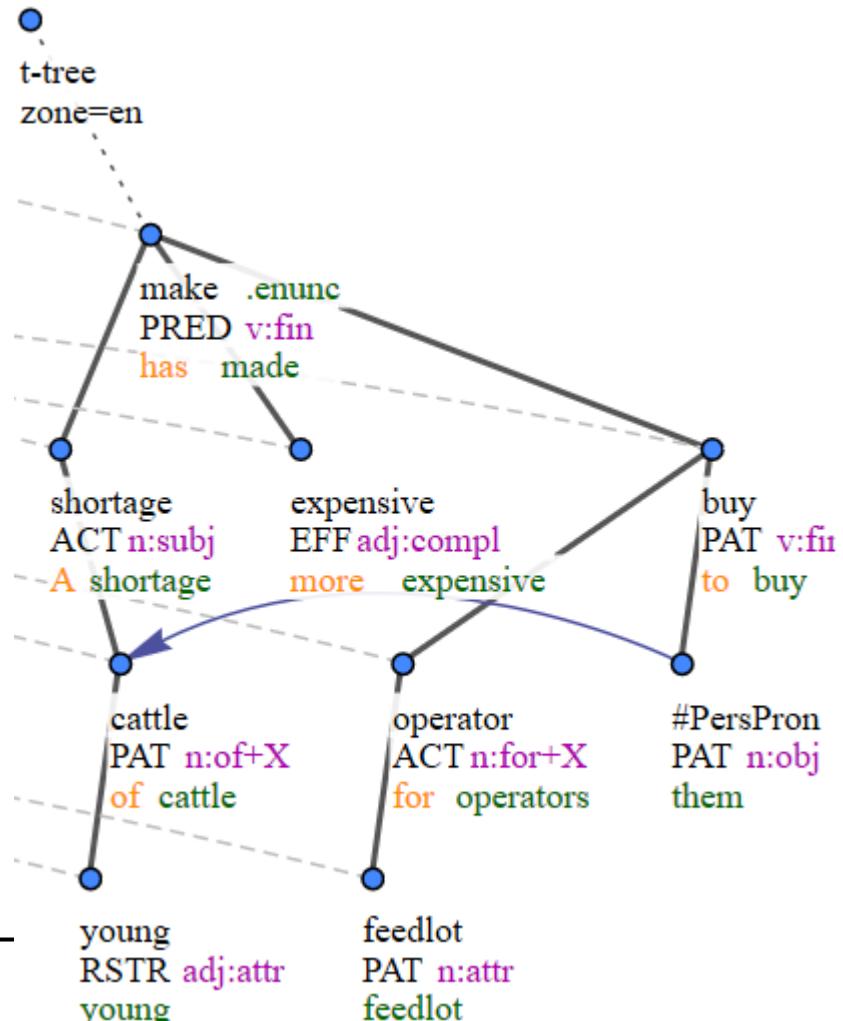


A shortage of young cattle has made them more expensive for feedlot operators to buy (= to buy it = to buy the cattle) .



PDT: t-layer

- 'completed' structure
 - 'missing' argument:
pro-drop subjects, objects, ...

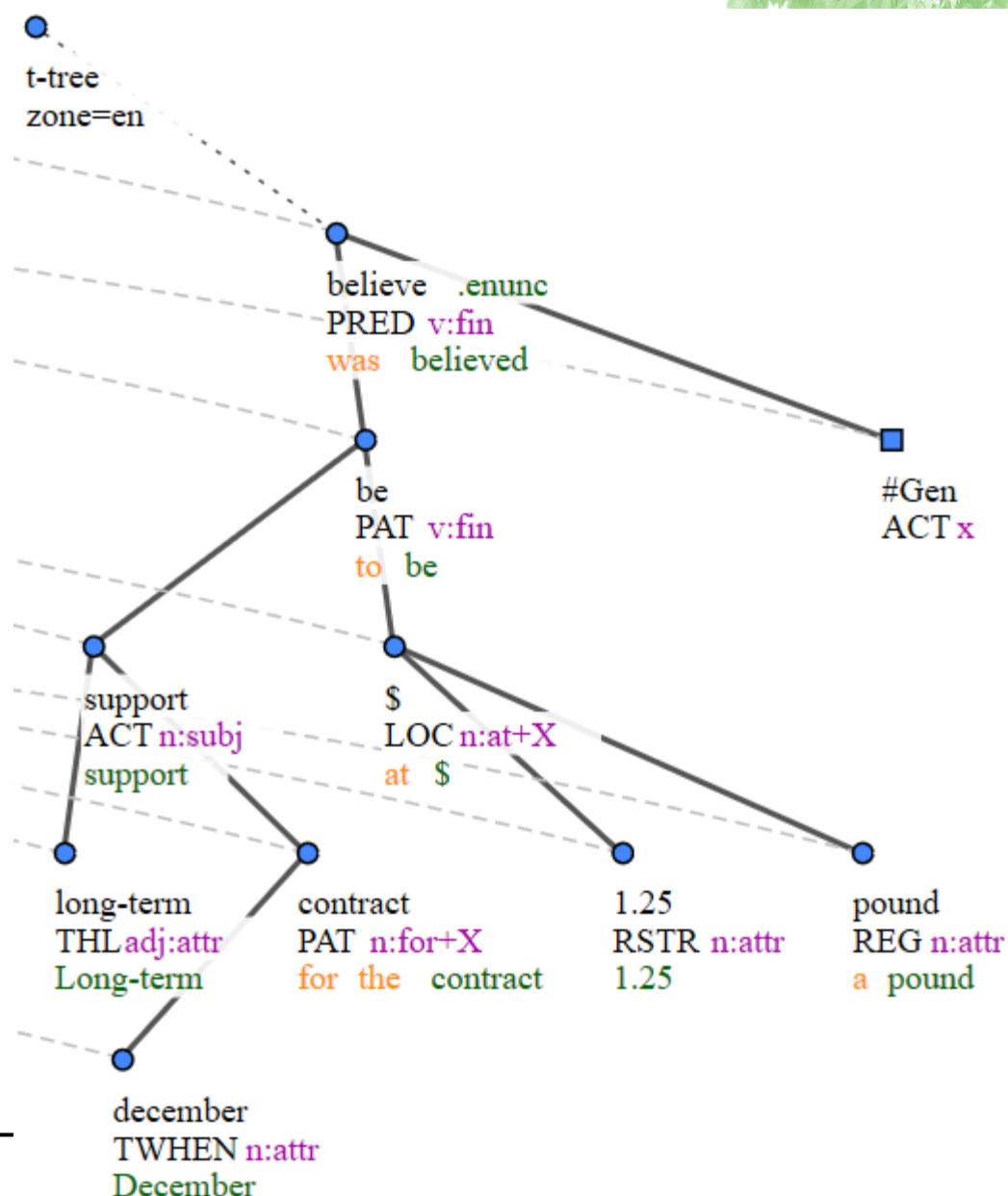


A shortage of young cattle has made them more expensive for feedlot operators to buy (= to buy it = to buy the cattle).



PDT: t-layer

- 'completed' structure
 - 'missing' argument:
pro-drop subjects
suppressed Agent in passive



Long-term support for the December contract was believed to be at \$1.25 a pound. (ACT = who believed)

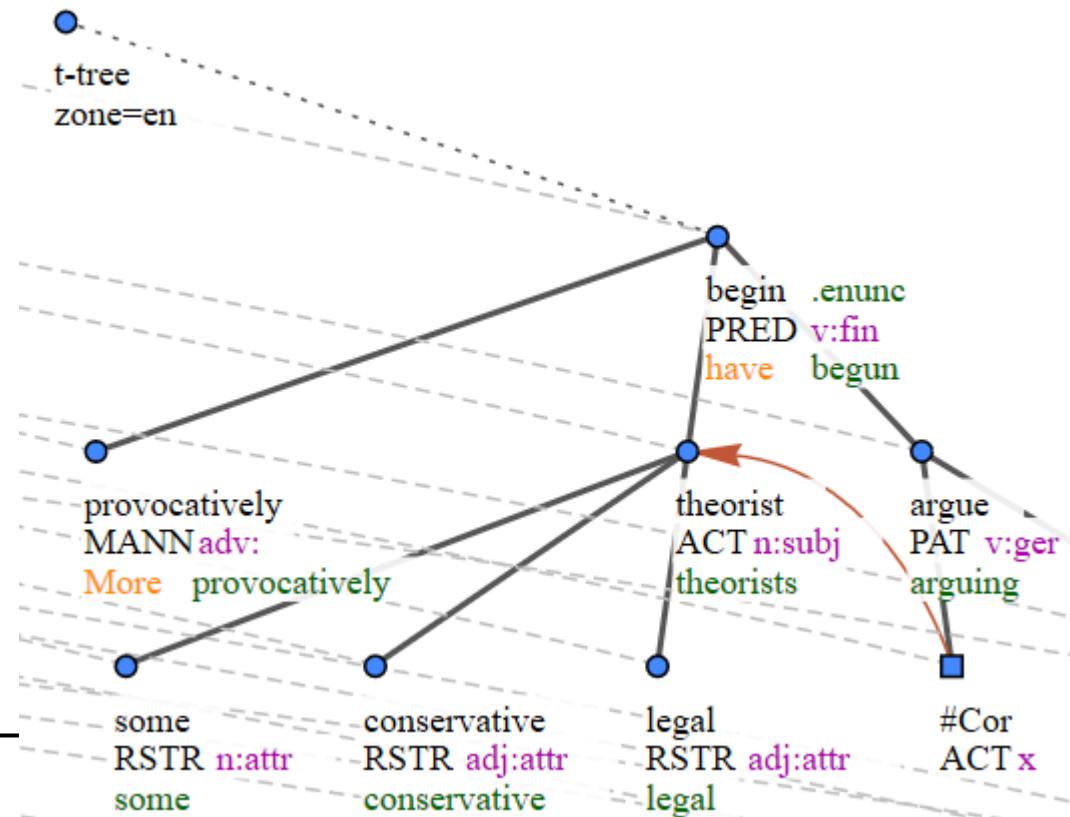


PDT: t-layer

- 'completed' structure
 - 'missing' argument:
 - pro-drop subjects
 - suppressed Agent in passive
 - control constructions

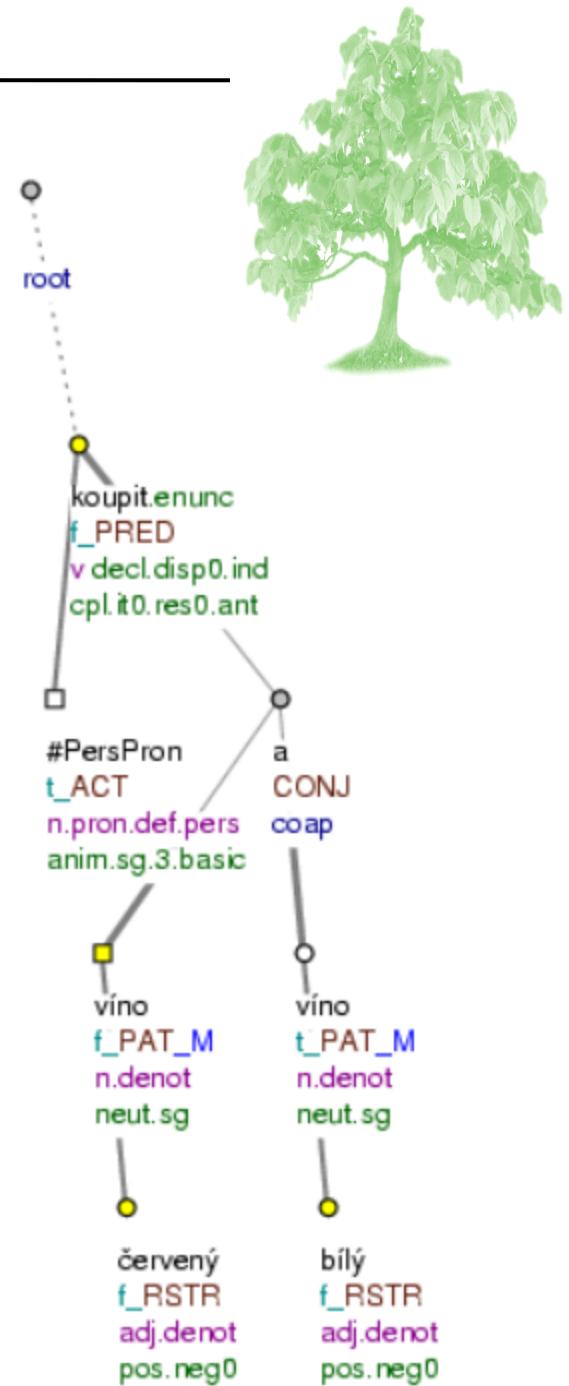
...

More provocatively, some conservative legal theorists have begun arguing that ... (= the theorists argue)



PDT: t-layer

- 'completed' structure
 - coordination:
stripping and gapping

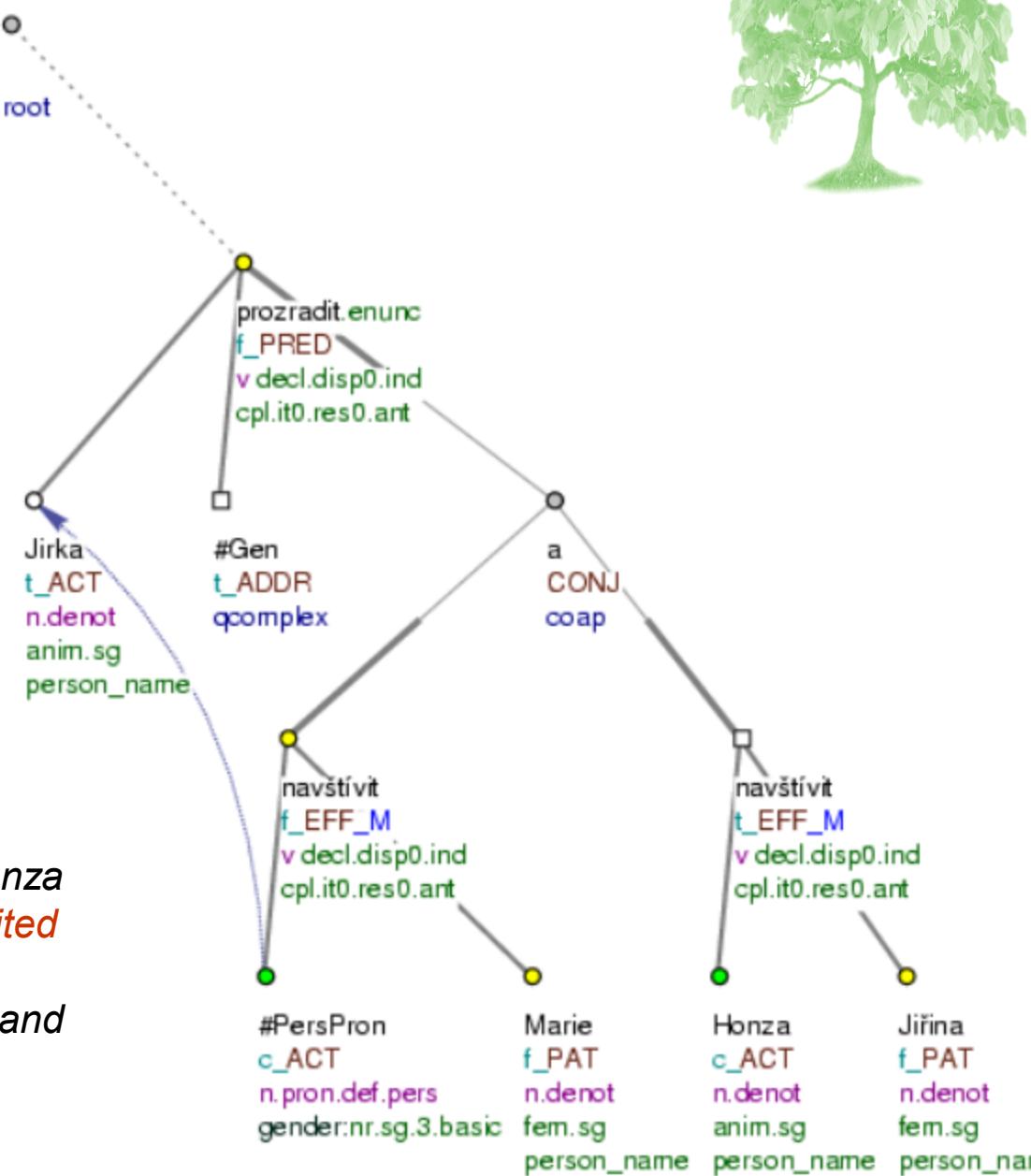


Koupil červené a bílé víno. (=lit. (He) bought red and white wine.)
= He bought red wine and white wine.



PDT: t-layer

- 'completed' structure
 - coordination:
stripping and gapping



*Jirka prozradil, že on navštívil Marii a Honza Jiřinu. (=lit. George disclosed that *he* visited Mary and John Henriette.)*

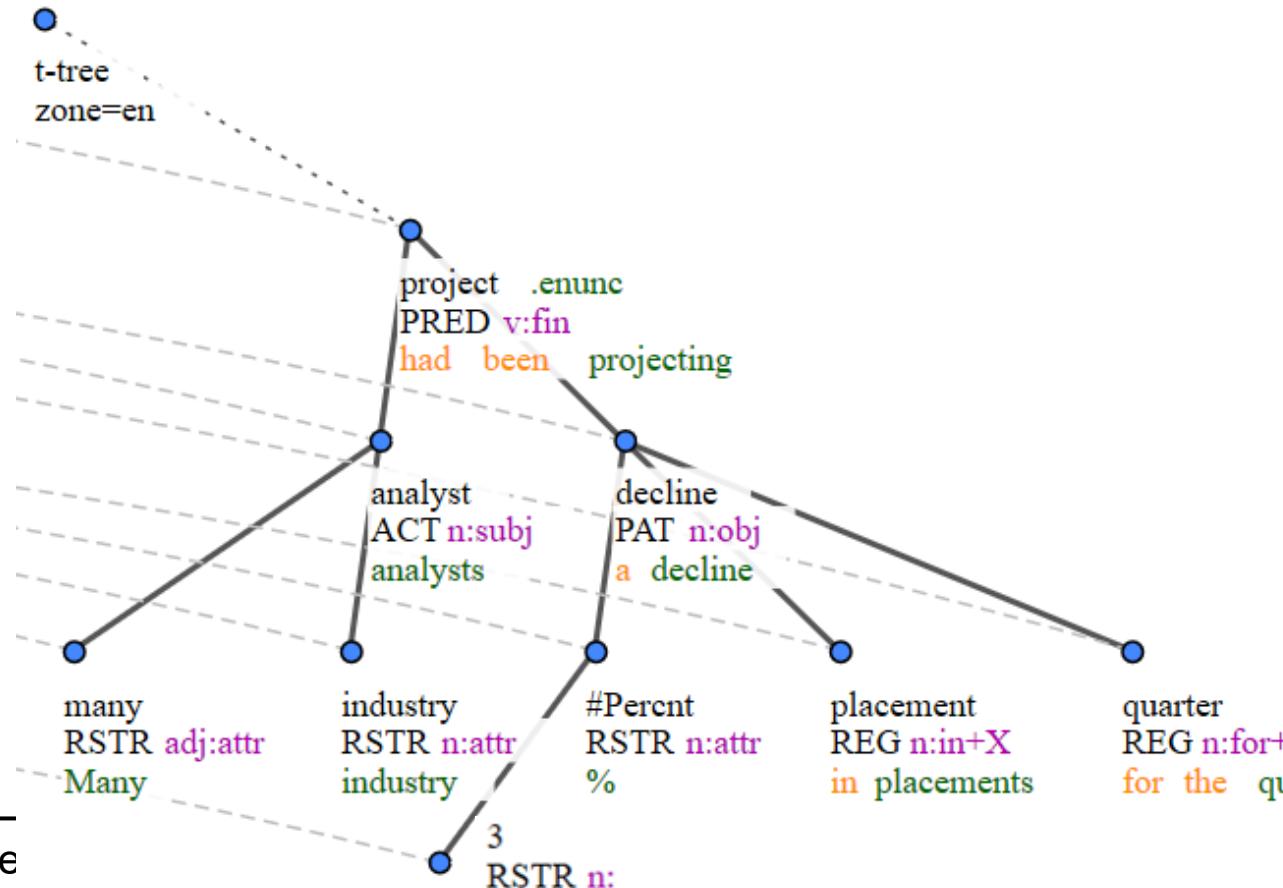
= George disclosed that he *visited* Mary and John *visited* Henriette.



PDT: t-layer

- only lexical words as separate nodes
 - auxiliary and modal verbs
 - prepositions, subordinating conjunctions, ...
 - BUT coordinating conjunctions
 - no punctuation
 - plus technical root

*Many industry analysts
had been projecting a 3%
decline in placements for
the quarter .*

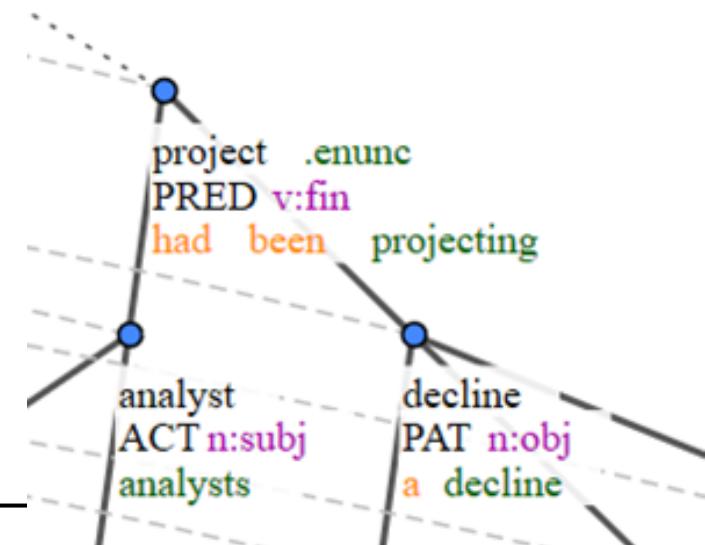




PDT: t-layer

- lexical words (cf. semantemes in FGD)
 - 1) lexical items: **t-lemma**
link to the PDT-Vallex / EngVallex dictionary for verbs and selected nouns
 - 2) meaning of morphological categories: **grammatemes**
(a pair of) trousers ... singular
denominating (pojmenovávací) vs. *correlating* (usouvztažňující) categories
 - 3) deep syntactic structure: **functors + subfunctors**
valency as a key property → a sentence syntactic structure

*Many industry analysts
had been projecting a 3%
decline ...*





PDT: t-layer

- lexical words (cf. semantemes in FGD)
 - 1) lexical items: ***t-lemma***
link to the PDT-Vallex / EngVallex dictionary for verbs and selected nouns
 - 2) meaning of morphological categories: ***grammatemes***
(a pair of) trousers ... singular
denominating (pojmenovávací) vs. *correlating* (usouvzažňující) categories
 - 3) deep syntactic structure: ***functors + subfunctors***
valency as a key property → a sentence syntactic structure
- additional info:
 - coreferential links
 - 'word order': increasing communicative dynamism
topic focus articulation (cz: aktuální členění)
- additional annotation in PDT 3.5:
 - multiwords expressions, extended coreference and discourse relations, genre specification, ...

T-node attributes and their values in PDT 2.0

Notation:

attribute – attribute name
value – attribute value

A. Lexical content

t_lemma – tectogrammatical lemma
val_frame.rf – valency frame (reference to PDT-VALLEX)

B. Semantic roles and other structural relations

functor – role of the node within the t-tree structure

Functors for independent clauses:

1. **PRED** – predicate clause
2. **DENOM** – denominative clause
3. **VOCAT** – vocative clause
4. **PARTL** – interjectional clause
5. **PAR** – parenthetical clause

Actants:

6. **ACT** – actor
7. **PAT** – patient
8. **ADDR** – addressee
9. **ORIG** – origin
10. **EFF** – effect

Temporal modifiers:

11. **TWHEN** – when
12. **TFHL** – for how long
13. **TFRWH** – from when
14. **THL** – how long
15. **THO** – how often
16. **TOWH** – to when
17. **TPAR** – temporal parallel
18. **TSIN** – since when

19. **TTILL** – till when

Spatial modifiers:

20. **LOC** – where
21. **DIR1** – from where
22. **DIR2** – through where
23. **DIR3** – to where

Implicational/causal modifiers:

24. **AIM** – aim
25. **CAUS** – cause
26. **CNCS** – concession
27. **COND** – condition
28. **INTT** – intention

Various types of manner:

29. **ACMP** – accompaniment
30. **CPR** – comparison
31. **CRIT** – criterion
32. **DIFF** – difference
33. **EXT** – extent
34. **MANN** – manner
35. **MEANS** – means
36. **REG** – regard
37. **RESL** – result
38. **RESTR** – restriction

Specific adnominal modifiers:

39. **RSTR** – attribute
40. **APP** – appurtenance
41. **AUTH** – author
42. **MAT** – material
43. **ID** – identity

Paratactic structures:

44. **ADVS** – adversative
 45. **CONFR** – confrontation
 46. **CONJ** – conjunction
 47. **CONTRA** – contrariety
 48. **CSQ** – consequence
 49. **DISJ** – disjunction
 50. **GRAD** – gradation
 51. **REAS** – reason
 52. **APPS** – apposition
 53. **CM** – coordination modifier
- Multiword lexical units:
54. **CPHR** – part of complex predicate
 55. **DPHR** – dependent part of an idiomatic expression

Other:

56. **COMPL** – predicative complement
57. **BEN** – benefactor
58. **CONTRD** – contradiction
59. **HER** – heritage
60. **RHEM** – rhematizer
61. **SUBS** – substitution
62. **ATT** – attitude
63. **INTF** – intensifier
64. **MOD** – modality
65. **PREC** – reference to preceding text
66. **FPHR** – foreign language expression

subfunctor – more detailed functor specification

1. **basic** – basic value (prototypical for the given functor)
2. **nr** – not recognized

Values specific to spatial functors:

3. **abstr** – in abstract space
4. **along** – along
5. **around** – around
6. **above** – above
7. **behind** – behind
8. **below** – below
9. **betw** – between
10. **elsew** – elsewhere
11. **ext** – extent
12. **front** – in front of
13. **near** – near
14. **opp** – opposite
15. **target** – target
16. **to** – to
17. **across** – across

Values specific to ACMP:

18. **circ** – circumstance
19. **incl** – inclusion
20. **wout** – negative accompaniment (without someone)

Values specific to CPR:

21. **than** – difference

22. **wrt** – with respect to

Values specific to BEN:

23. **agst** – against
24. **approx** – approximately
25. **less** – less
26. **more** – more

Values specific to TWHEN:

27. **after** – after
28. **approx** – approximately
29. **before** – before
30. **begin** – at the beginning of
31. **betw** – between
32. **end** – at the end of
33. **flow** – in the course of
34. **mid** – in the middle of

is_member – distinction

between members of paratactic structures and shared modifiers

1. **0** – non-member
2. **1** – member

is_parenthesis

1. **0** – unmarked value
2. **1** – part of parenthesis

is_state

1. **0** – unmarked value
2. **1** – modifier expressing being in certain state

C. Communicative dynamism

tfa – topic/focus articulation

1. **t** – non-contrastive contextually bound expression
2. **f** – contextually non-bound expression
3. **c** – contrastive contextually bound expression

deepord – non-negative integer representing deep word order

D. Coreference and predicative complement

coref_gram.rf – (list of) reference(s) to antecedent(s) in the sense of grammatical coreference

coref_text.rf – (list of) reference(s) to antecedent(s) in the sense of textual coreference

coref_special – special types of coreference (without obvious t-node antecedent)

1. **segm** – coreference with a sequence of preceding sentences, without more explicit limitations
2. **exoph** – antecedent not present in the text at all

compl.rf – reference to "secondary" parent t-node (in the case of "dual" complement dependency)

E. Types of t-nodes

nodetype – basic node classification

1. **root** – technical root
2. **complex** – complex node
3. **qcomplex** – quasi-complex node
4. **atom** – atomic node
5. **coap** – paratactic structure root (coordination or apposition)
6. **dphr** – dependent part of an idiomatic expression
7. **fphr** – part of a foreign-language expression
8. **list** – root node of a list structure

- sempos** – semantic part of speech (further subdivision of complex nodes)
1. **n.denot** – denotative semantic noun
 2. **n.denot.neg** – denotative semantic noun with separately represented negation
 3. **n.pron.def.demon** – demonstrative definite pronominal semantic noun
 4. **n.pron.def.pers** – personal definite pronominal semantic noun
 5. **n.pron.indef** – indefinite pronominal semantic noun
 6. **n.quant.def** – definite quantificational semantic noun
 7. **adj.denot** – denotative semantic adjective
 8. **adj.pron.def.demon** – demonstrative definite pronominal semantic adjective
 9. **adj.pron.indef** – indefinite pronominal semantic adjective
 10. **adj.quant.def** – definite quantificational semantic adjective
 11. **adj.quant.indef** – indefinite quantificational semantic adjective
 12. **adj.quant.grad** – gradable quantificational semantic adjective
 13. **adv.denot.ngrad.nneg** – non-gradable denotative semantic adverb, impossible to negate
 14. **adv.denot.ngrad.neg** – non-gradable denotative semantic adverb, possible to negate
 15. **adv.denot.grad.nneg** – gradable denotative semantic adverb, impossible to negate
 16. **adv.denot.grad.neg** – gradable denotative semantic adverb, possible to negate
 17. **adv.pron.def** – definite pronominal semantic adverb

18. **adv.pron.indef** – indefinite pronominal semantic adverb
19. **v** – semantic verb

F. Grammatemes

- sentmod** – sentence modality
1. **enunc** – indicative mood
 2. **excl** – exclamatory mood
 3. **desid** – desiderative mood
 4. **imper** – imperative mood
 5. **inter** – interrogative mood

value applicable to all following grammatemes:

nr – not recognized

gram/aspect

- aspect
1. **proc** – processual (counterpart to imperfective)
 2. **cpl** – complex (counterpart to perfective)

gram/degcmp

- degree of comparison
1. **pos** – positive
 2. **comp** – comparative
 3. **acomp** – absolute comparative
 4. **sup** – superlative

gram/deontmod

- deontic modality
1. **deb** – necessary
 2. **hrt** – obligatory
 3. **vol** – wanted/intended
 4. **poss** – possible
 5. **perm** – permitted
 6. **fac** – ability to do something
 7. **decl** – unmarked

gram/dispmod

- dispositional modality
1. **disp0** – dispositional modality absent

2. **disp1** – dispositional modality present
3. **nil** – not applicable (with infinitive)

gram/gender

- gender
1. **anim** – masculine animate
 2. **inan** – masculine inanimate
 3. **fem** – feminine
 4. **neut** – neuter
 5. **inher** – "inherited" from antecedent

gram/indeftype

- type of (pronoun) indefiniteness
1. **relat** – relative
 2. **inter** – interrogative
 3. **negat** – negative
 - 4.-10. **indef1** – **indef6** – other types of indefiniteness
 - 11.-12. **total1**, **total2** – totalizers

gram/iterativeness

- iterativeness
1. **it0** – non-iterative verb
 2. **it1** – iterative verb

gram/negation

- negation
1. **neg0** – affirmative
 2. **neg1** – negative

gram/number

- number
1. **sg** – singular
 2. **pl** – plural
 3. **inher** – "inherited" from antecedent

gram/numbertype

- type of numeral expression
1. **basic** – basic numeral
 2. **frac** – fractional numeral
 3. **kind** – sort numeral

4. **ord** – ordinal numeral
5. **set** – set numeral

gram/person

- person
1. **1** – first person
 2. **2** – second person
 3. **3** – third person
 4. **inher** – "inherited" from antecedent

gram/politeness

- politeness
1. **basic** – common use
 2. **polite** – polite form

gram/resultative

- resultative
1. **res0** – non-resultative
 2. **res1** – resultative

gram/tense

- verb tense
1. **sim** – simultaneous
 2. **ant** – preceding (anterior)
 3. **post** – subsequent (posterior)
 4. **nil** – not applicable (with infinitive)

gram/verbmod

- verb modality
1. **ind** – indicative
 2. **imp** – imperative
 3. **cdn** – conditional
 4. **nil** – not applicable (with infinitive)

G. Links to a-layer

- atree.rt** – reference to the corresponding a-tree technical root (only with technical t-tree root)

- a/lex.rt** – reference to (identifier of) the corresponding "autosemantic" a-node

- a/aux.rf** – (list of) reference(s) to the corresponding auxiliary a-node(s)

- is_generated** – distinction between nodes expressed/unexpressed in the surface form
1. **0** – surface counterpart exists
 2. **1** – newly created (or "copied") node

H. Quotation and direct speech

- quot/type** – type of quoted expression

1. **citation** – citation
2. **dsp** – direct speech
3. **meta** – "meta" use
4. **title** – title
5. **other** – other type

- quot/set_id** – id dedicated for co-indexing all nodes within a quoted expression

- is_dsp_root** – root of direct speech

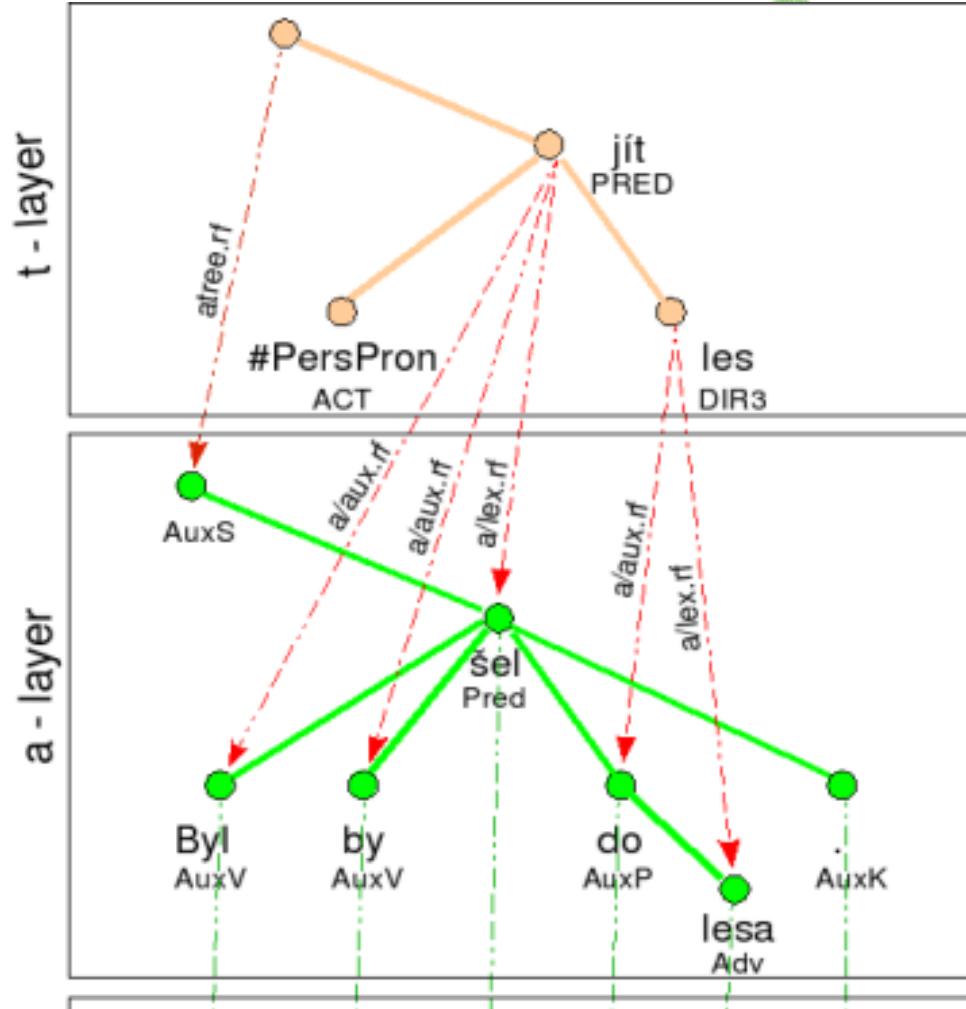
1. **0** – unmarked value
2. **1** – root of subtree representing direct speech

I. Other

- id** – node identifier

- is_name_of_person** – personal proper name
1. **0** – unmarked value
 2. **1** – proper name of a person

PDT: relation between t-layer and a-layer



(He) would go to the forest.



PDT: relation between t-layer and a-layer

each t-node ... PML reference to a-layer

- non-root t-node... attribute *a* consisting of 2 attributes:
 - *lex.rf* → id of a-node from which the t-node got its lexical meaning
 - *aux.rf* → list of ids of all other a-nodes related to the t-node
- technical t-root ... *atree.rf* → id of a root of correspond. anal. tree



PDT: relation between t-layer and a-layer

each t-node ... PML reference to a-layer

- non-root t-node... attribute *a* consisting of 2 attributes:
 - *lex.rf* → id of a-node from which the t-node got its lexical meaning
 - *aux.rf* → list of ids of all other a-nodes related to the t-node
- technical t-root ... *atree.rf* → id of a root of correspond. anal. tree
- t-node with **no analytical counterpart**: lex.rf and aux.rf **empty**

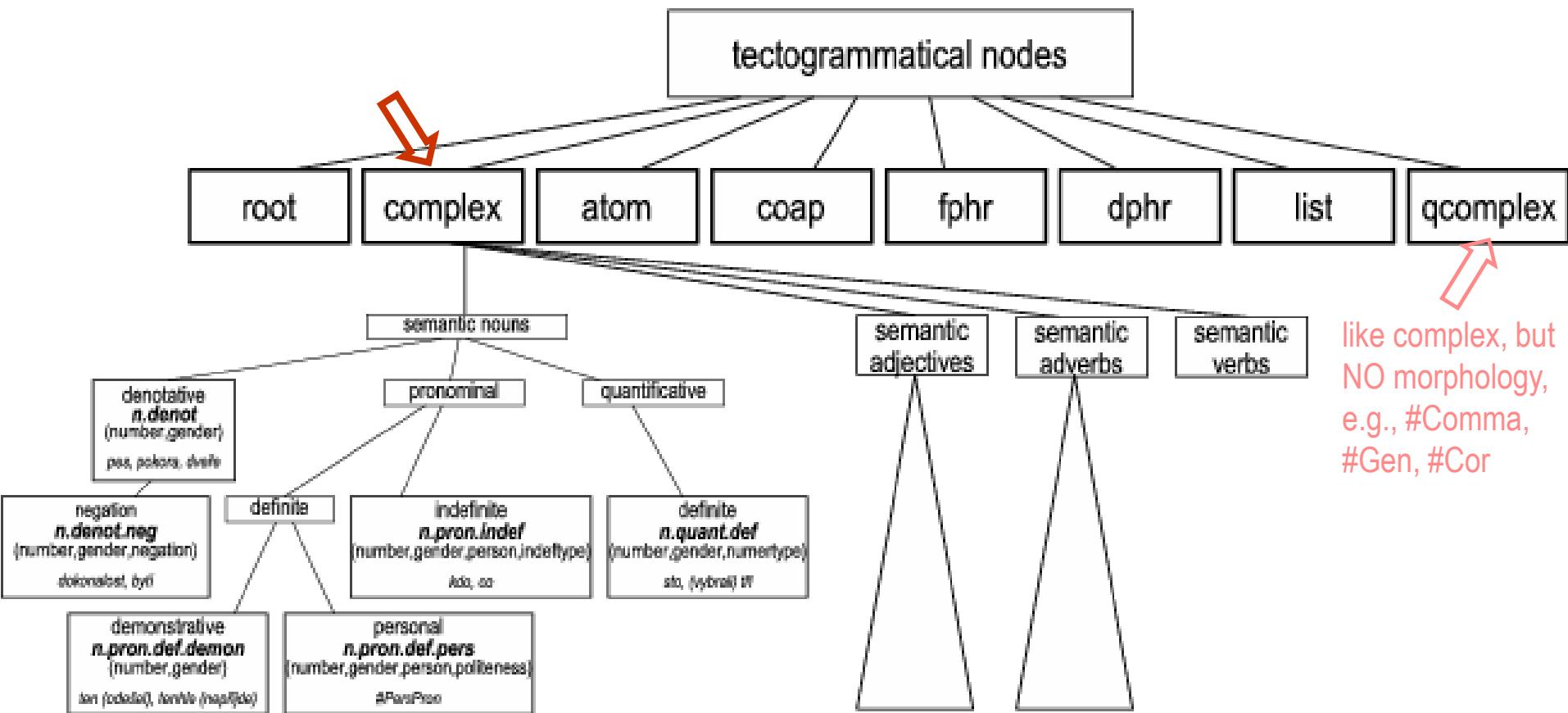
Dovolil mu odejít. {#Cor.ACT}

- **copied nodes:**
 - lexical items with several occurrences at the t-layer
but expressed only once in a surface sentence
(e.g., *červené a bílé víno* = *červené víno a bílé víno*)
→ all attributes a/lex.rf, a/aux.rf
 - ... id(s) of the corresponding a-node(s)



PDT: node types

- eight node types ... attribute *nodetype*
- defined on the basis of a *t-lemma* and/or a *functor*





References

- Hajič, J., Hajičová, E., Mikulová, M., Mírovský, J.: Prague Dependency Treebank. Chapter in Ide, N., Pustejovsky, J. (eds.) *Handbook of Linguistic Annotation*, Springer, Berlin, pp. 555-594, 2017
- Manual for Tectogrammatical Annotation
<http://ufal.mff.cuni.cz/pdt3.0/documentation>
- and its shortened version <http://ufal.mff.cuni.cz/techrep/tr35.pdf> (recommended!)
- Cinková, S., Toman, J., Hajič, J. et al.: Tectogrammatical Annotation of the Wall Street Journal. *The Prague Bulletin of Mathematical Linguistics* 92, pp. 85-104, 2009
- PDT 3.5 <https://ufal.mff.cuni.cz/pdt3.5>
- PCEDT 2.0 <http://ufal.mff.cuni.cz/pcedt2.0/>