

Tectogrammatical representation of English

Silvie Cinková

UFAL, May 23, 2014

Outline

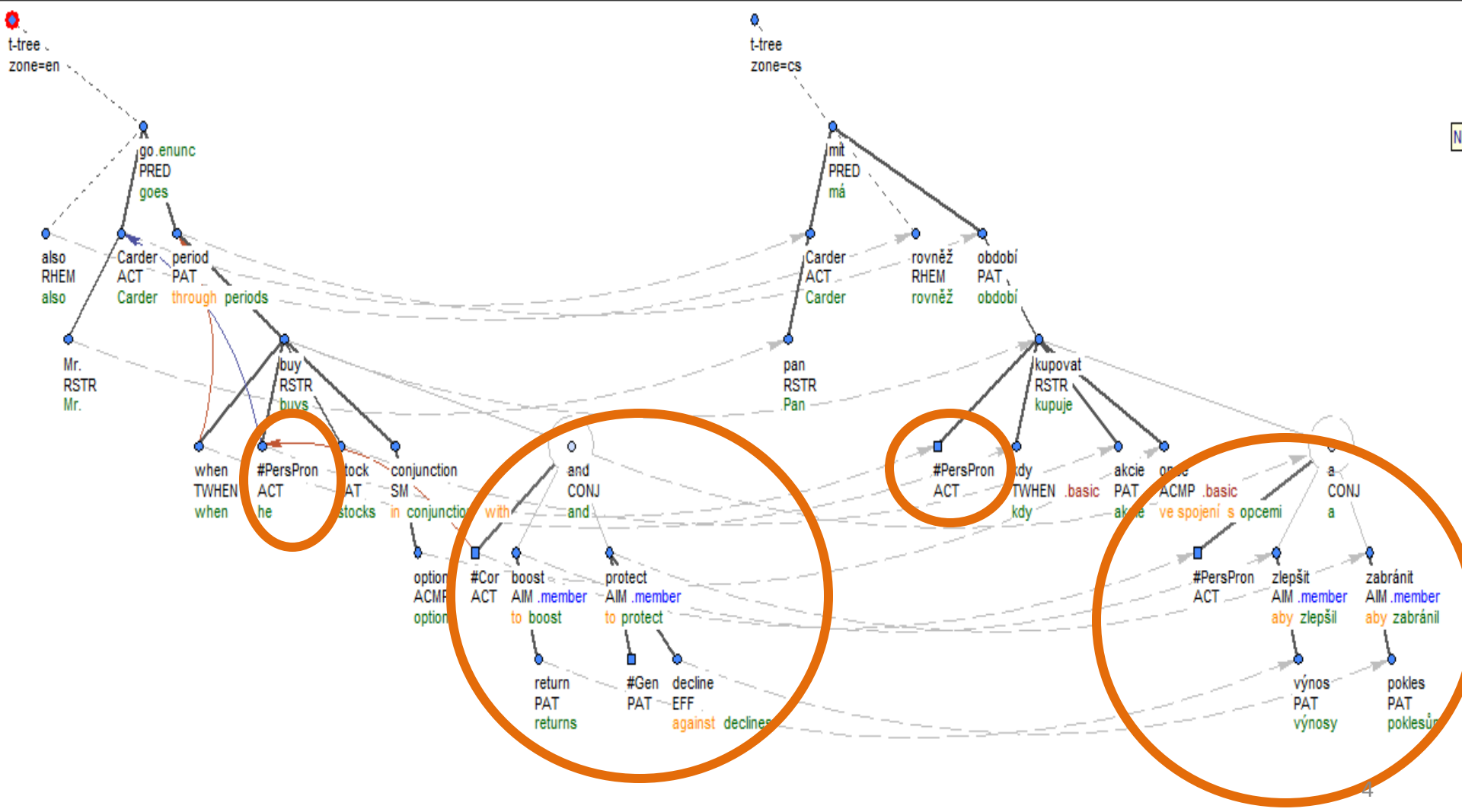
- PEDT, PCEDT-En
- p,a,t layers
- PMLTQ search
- Verb complement
- Cleft and pseudocleft sentences
- Existential constructions
- *do* as pro-form (“dummy”-*do*)
- Complex noun phrase
- Tokenization of hyphenated words at the t and a layer
- Infinitive clause
- Gerundial clause

TR: with structure towards semantics

- Capture the linguistic meaning of each sentence (underlying syntax)
- Syntactic structure
- Semantic labeling
- Argument structure
- Coreference
- ~~TFA~~

Mr. Carder also goes through periods when he buys stocks in conjunction with options to boost returns and protect against declines.

Pan Carder má rovněž období, kdy kupuje akcie ve spojení s opcemi, aby zlepšil výnosy a zabránil poklesům.

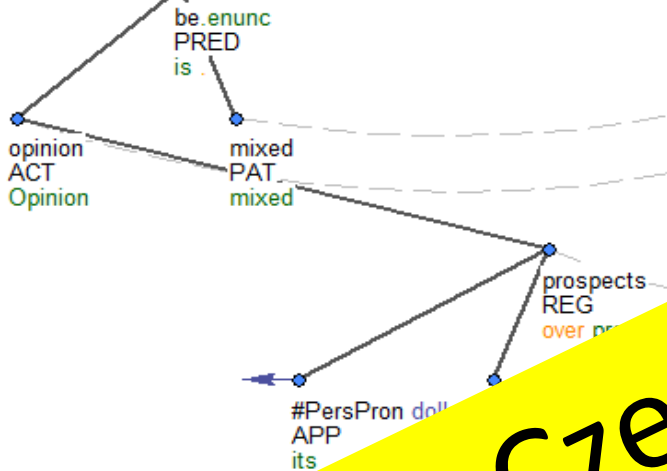


Issues to tackle

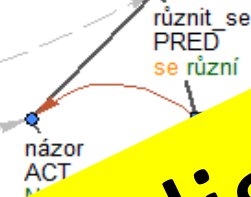
- Languages are different
- Conceptions of grammar are different!
 - missing categories
 - term mismatch
- Give a proof that TR in different languages are similar => synonymy requirement => higher abstraction => more focus on function
- Czech TR: more form-oriented

Prague Czech-English Dependency Treebank

t-tree
zone=en

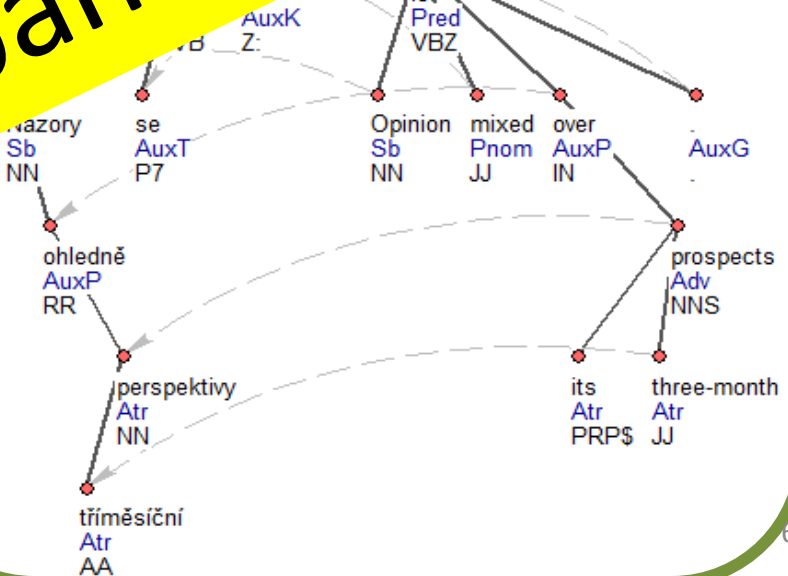


t-tree.
zone=cs



NP-SBJ

Opinion NN is VBZ mixed JJ over IN prospects PRP\$ three-month JJ prospects NNS



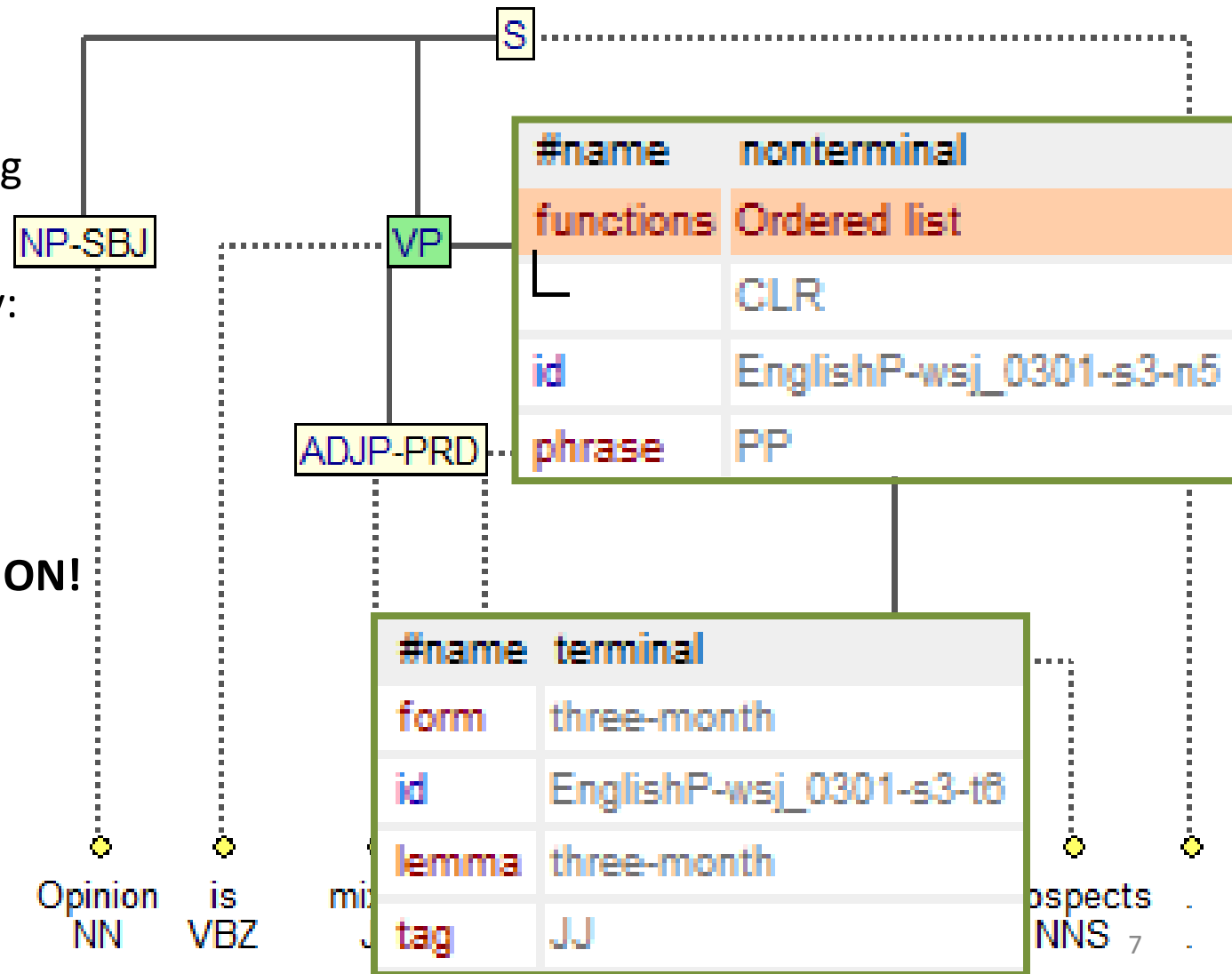
p-layer

- original PTB
- manual tagging
- manual bracketing

added automatically:

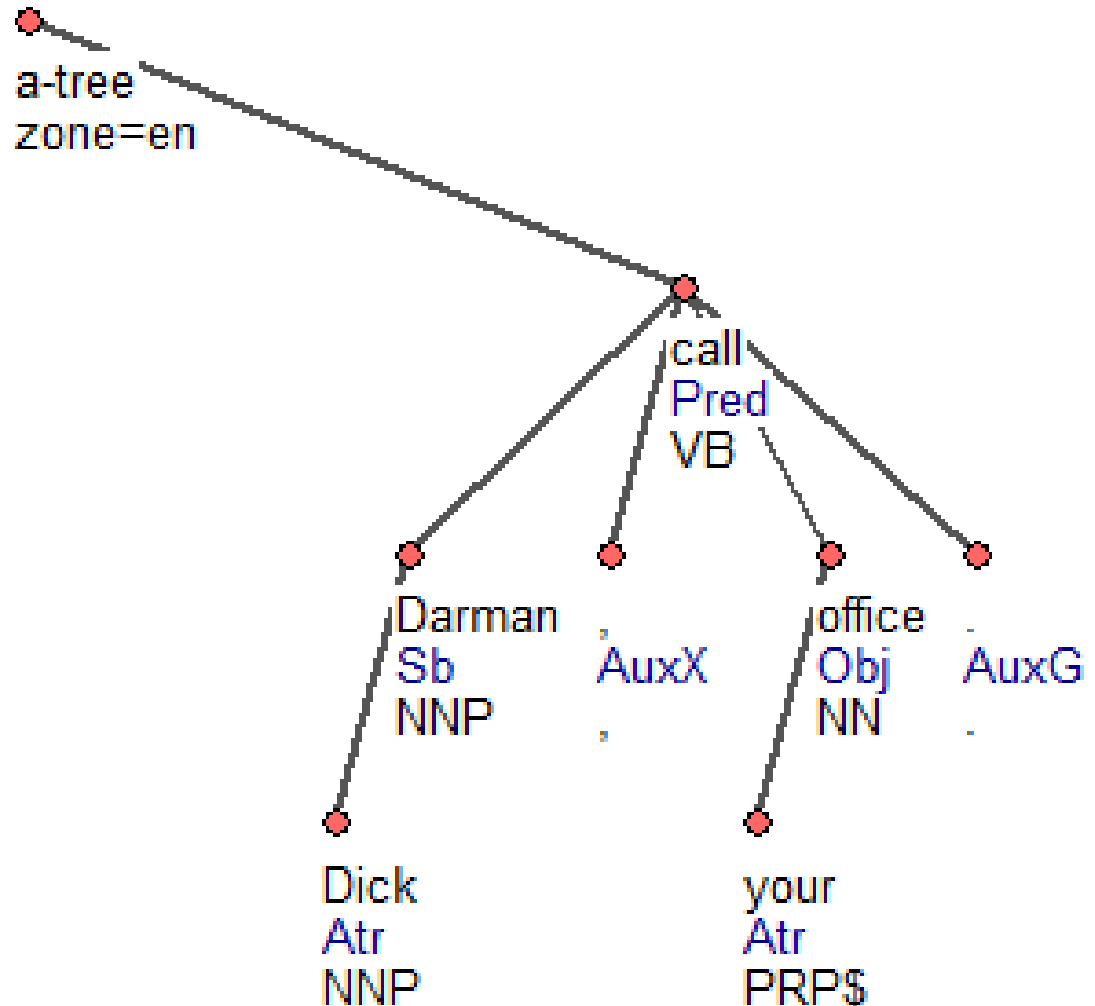
- node id
- lemmatization

NO MANUAL REVISION!



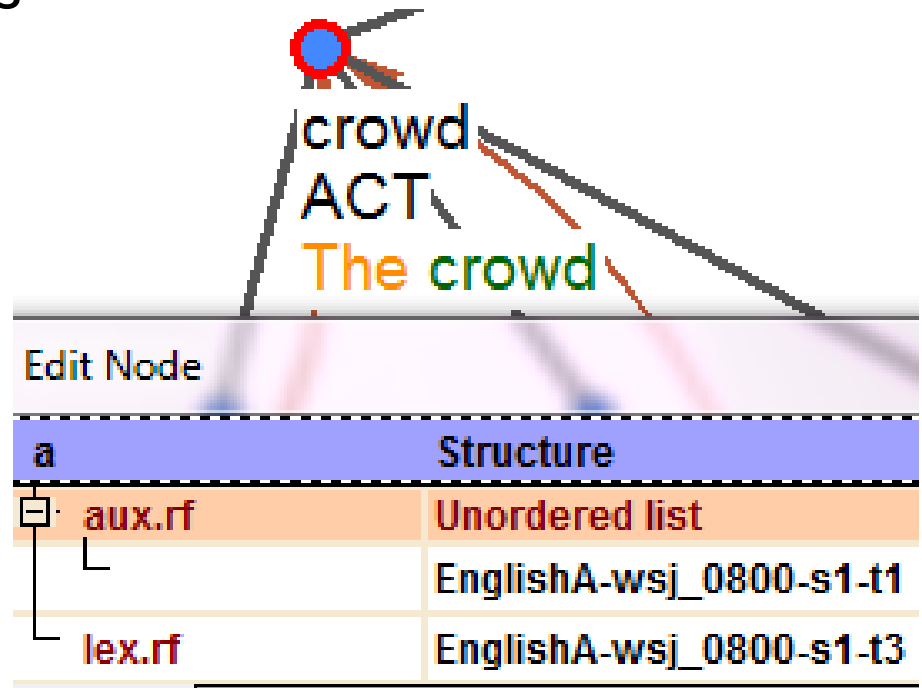
English a-layer (with integrated m-layer)

- **only automatic!!!**
- a Treex script by Martin Popel
 - just a few manual m-tag revisions
- dependency structure
- labeling (afuns)
- tag
- lemma
- p_terminal.rf
- alignment.rf



References to a-layer

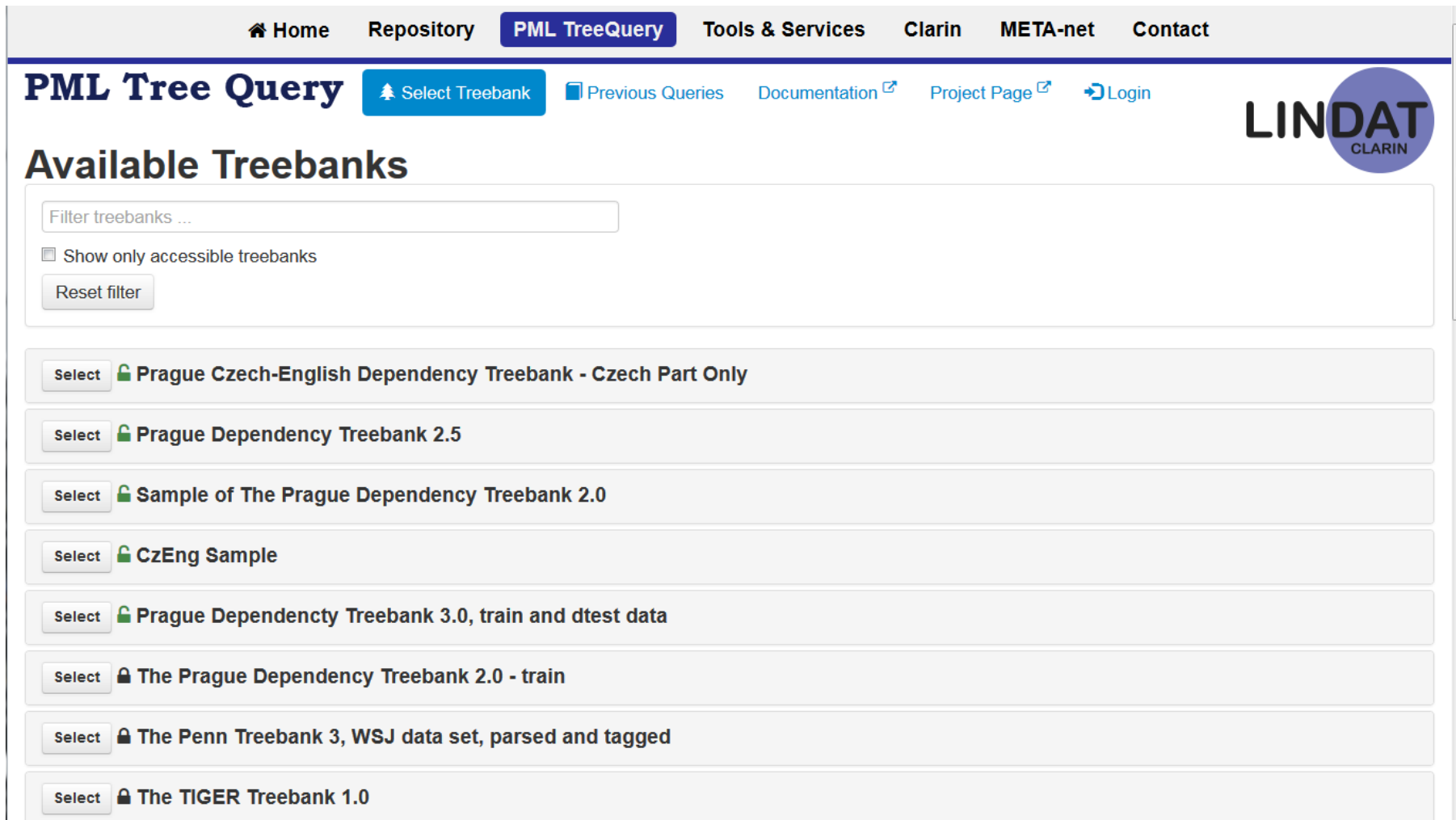
- **a/lex.rf**: content words
- **a/aux.rf**: auxiliary words



CORPUS SEARCH

PMLTQ on the web

- <https://lindat.mff.cuni.cz/services/pmltq/>



The screenshot displays the PML Tree Query web interface. At the top, there is a navigation bar with links for Home, Repository, PML TreeQuery (highlighted), Tools & Services, Clarin, META-net, and Contact. Below the navigation bar, the page title "PML Tree Query" is followed by a "Select Treebank" button and links for Previous Queries, Documentation, Project Page, and Login. The LINDAT CLARIN logo is visible in the top right corner. The main section is titled "Available Treebanks" and features a search filter box, a checkbox for "Show only accessible treebanks", and a "Reset filter" button. A list of treebanks is shown, each with a "select" button and a lock icon:

- Show only accessible treebanks
-
- Prague Czech-English Dependency Treebank - Czech Part Only
- Prague Dependency Treebank 2.5
- Sample of The Prague Dependency Treebank 2.0
- CzEng Sample
- Prague Dependency Treebank 3.0, train and dtest data
- The Prague Dependency Treebank 2.0 - train
- The Penn Treebank 3, WSJ data set, parsed and tagged
- The TIGER Treebank 1.0

Praque Czech-English Dependency Treebank 2.0

Relations ▾ Node Types ▾ Attributes ▾ Operators ▾ Functions ▾

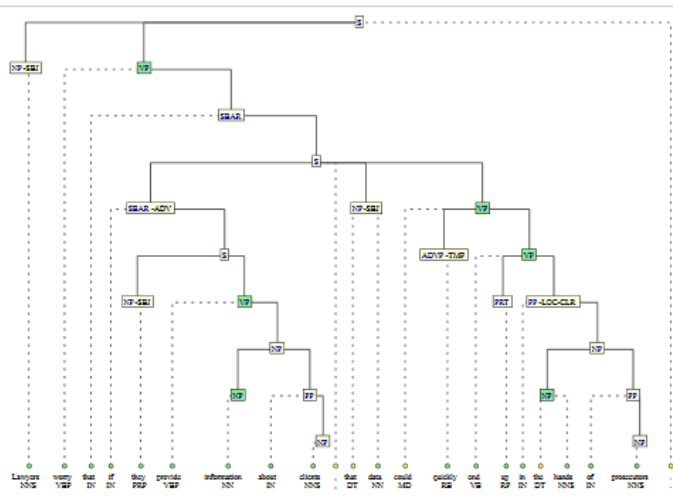
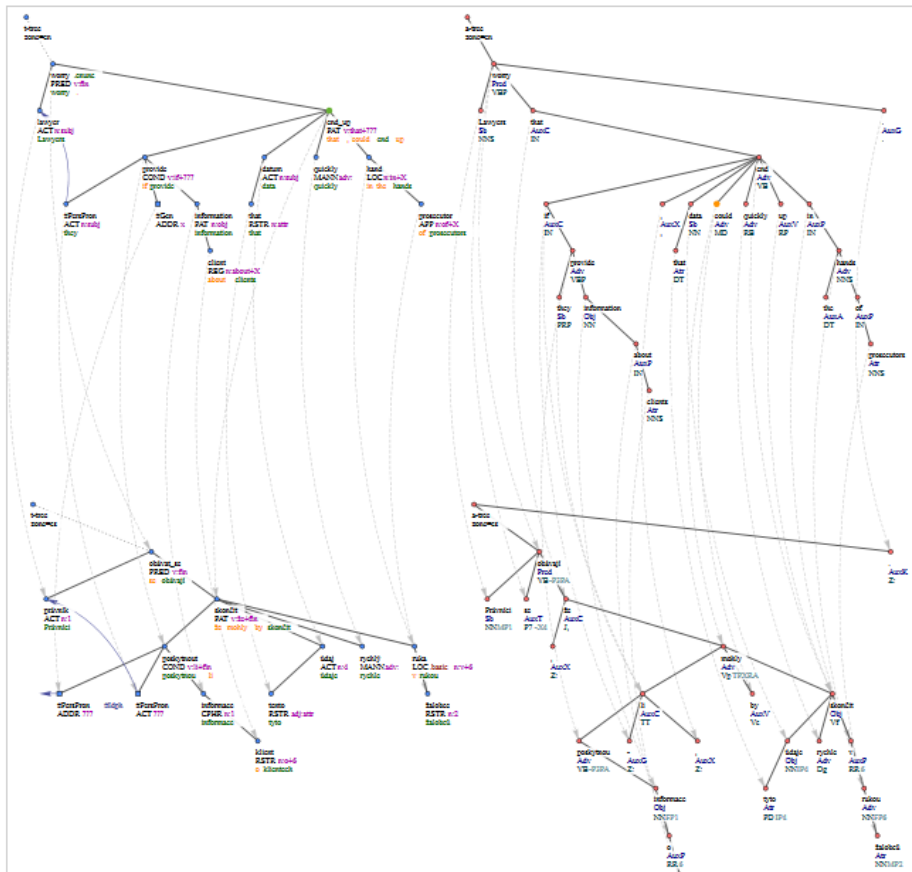
t-node \$a := [a/aux.rf a-node \$b := [tag= "MD"]]

Result: < 7 of 100 >

wsj_0049.treex.gz (25/76)

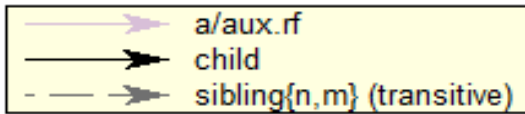
[en] Lawyers worry **that** if they provide information about clients, that data **could** quickly **end up** in the hands of prosecutors.

[cs] Právníci se obávají, **že** poskytnou-li informace o klientech, **mohly by** tyto údaje rychle **skončit** v rukou žalobců.

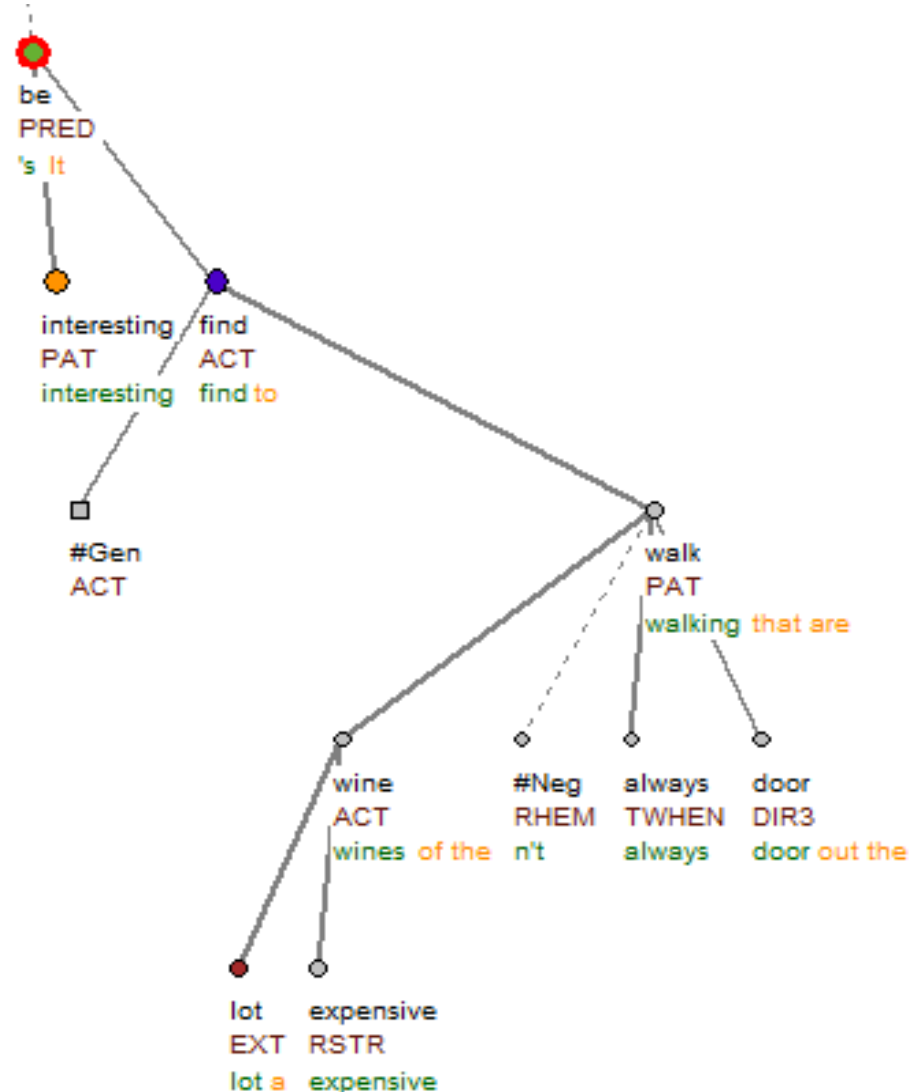
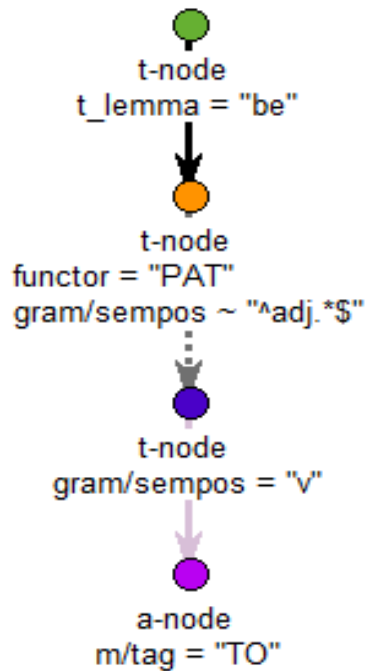


PMLTQ Search

t-node [t_lemma = "be", t-node[functor = "PAT", gram/sempos ~ "^adj.*\$", sibling{1,1} t-node [gram/sempos = "v", a/aux.rf a-node [m/tag = "TO"]]]];



Tree Query
q-12-05-12_151607

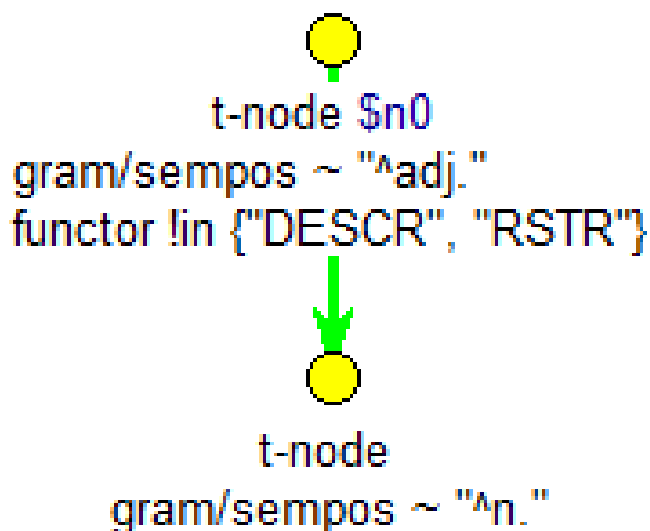
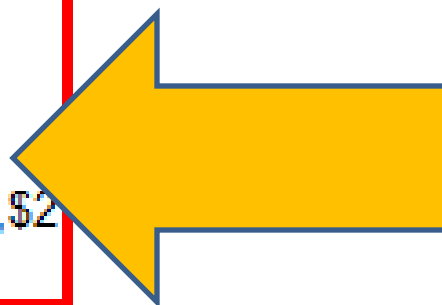


Tree Query

q-12-05-14_095029

Output filters:

```
>> for $n0.t_lemma,$n0.functor
    give distinct $1,count(1 over $1),$2
    sort by $1 asc,$3
```



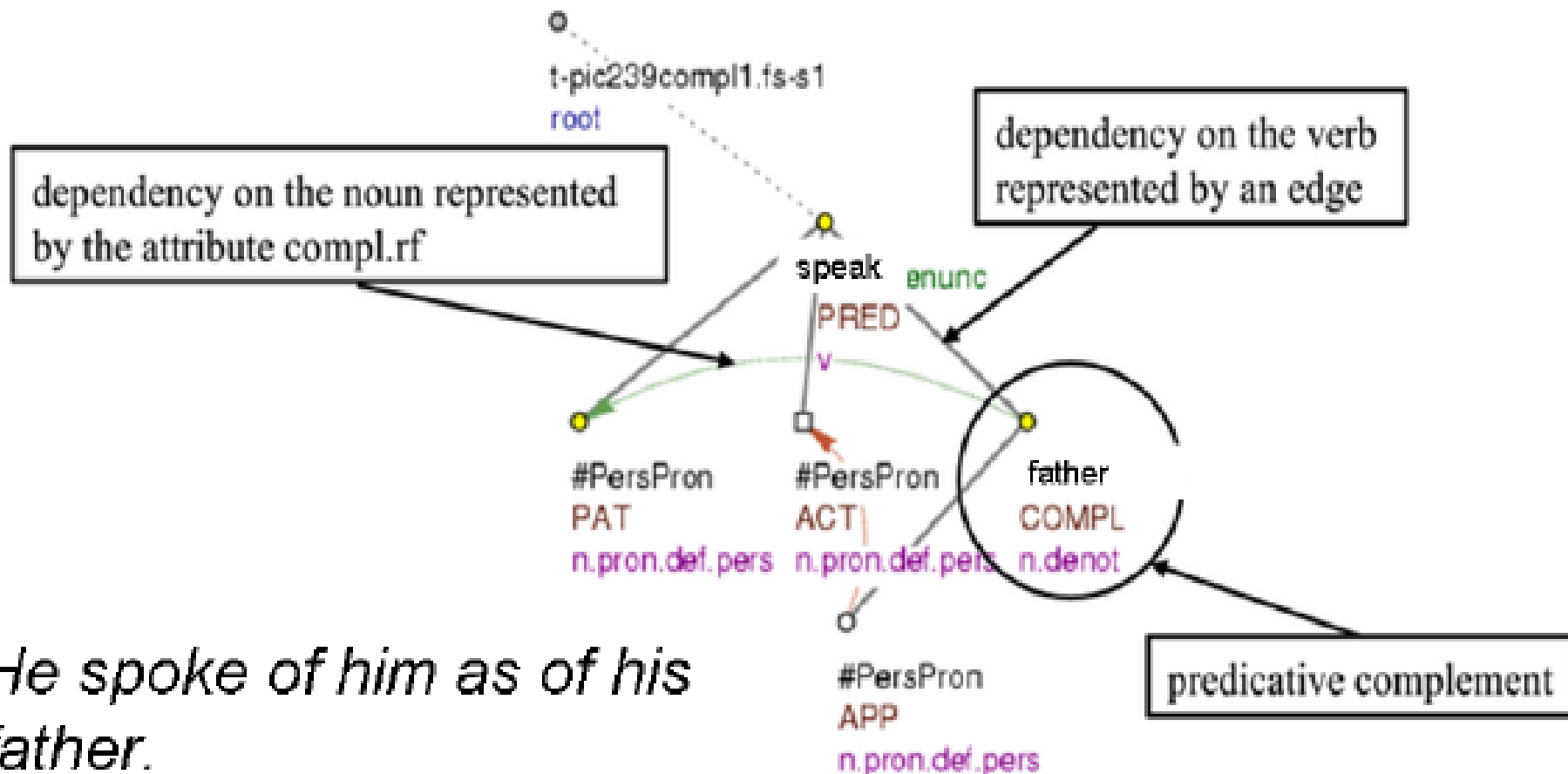
Amazing	1	NE	
ambivalent	1		CNCS
amendment	1		CRIT
american	5		ACT
american	5		APP
american	5		BEN
american	5		LOC
american	5		NE
American	3		APP
American	3		LOC
American	3		NE
amicable	1		MANN
ample	1	EXT	

**A FALSE FRIEND:
VERB COMPLEMENT ≠ DOPLNĚK**

Czech verb complement

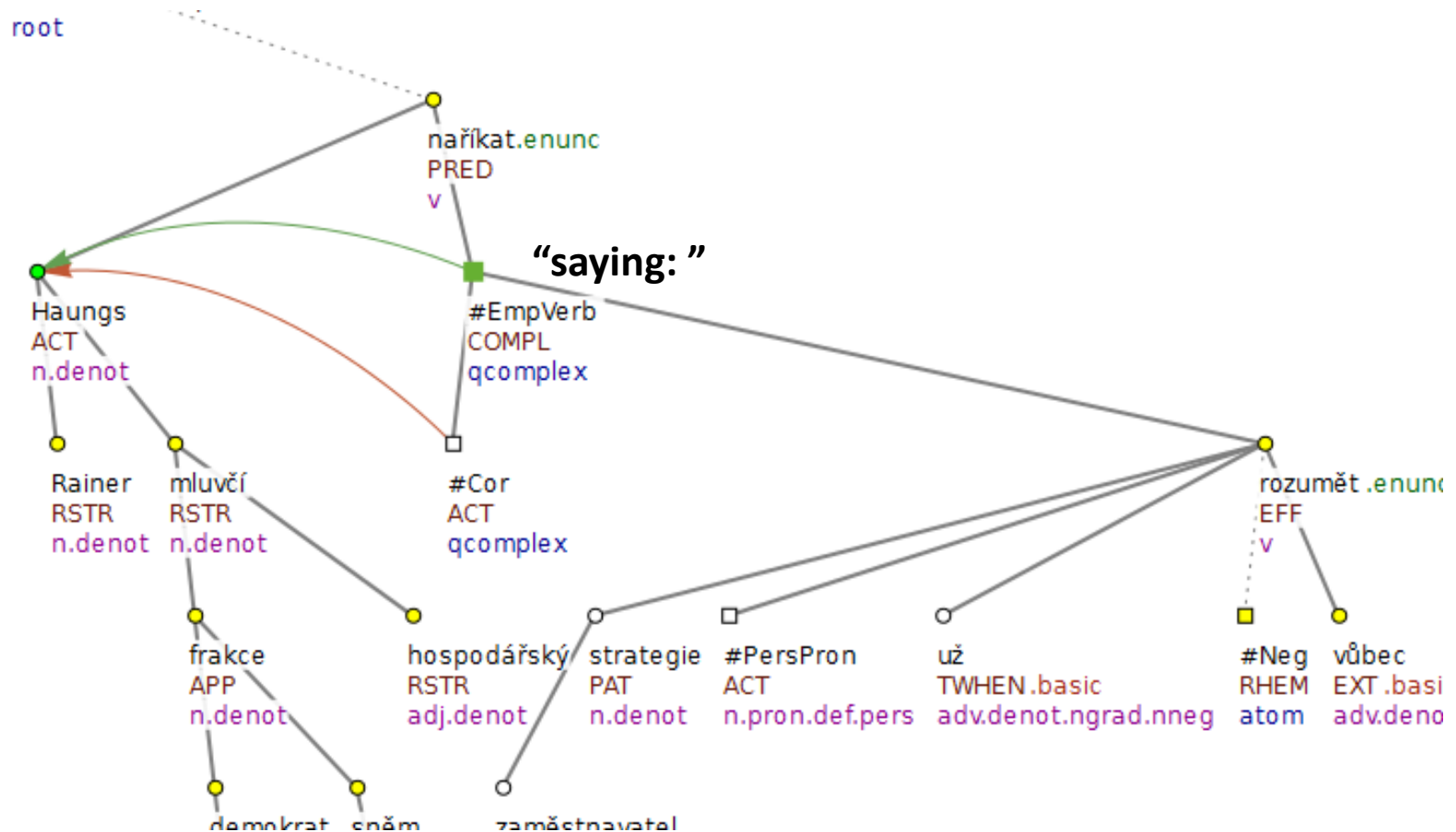
- free modifier of the verb
- related to object or to subject
- *Tady si chlapci sami nakupují a sami si vaří.*
- *K tomu bych rád poznamenal, že ...*
- *Mareš pojednává o nezaměstnanosti jako o sociálním, kulturním a politickém problému.*
- *Pětatřicetiletý sportovec, který první na světě získal všechny zlaté medaile*
- *Do jara jdou Severočeši posílení i po sponzorské stránce*
- *Viděl kočku, jak chytá myš*
- *Odcházela plačíc.*

Verb Complement (COMPL)

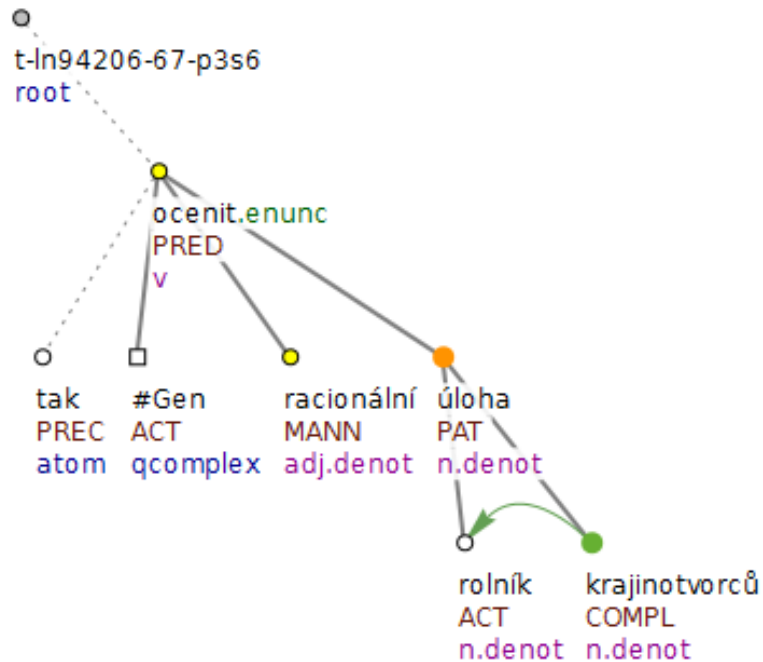


Verbs used as verbs of speaking (CZ)

"Strategii zaměstnavatelů už vůbec nerozumím," naříká hospodářský mluvčí frakce křesťanských demokratů ve Spolkovém sněmu Rainer Haungs.

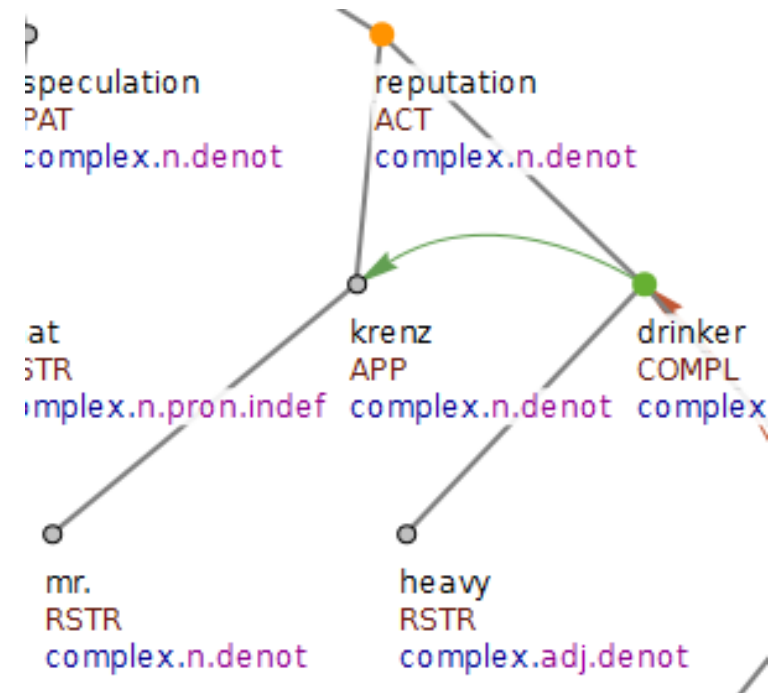


Verb complement dependent on noun



Měla by tak být racionálně oceněna *úloha* rolníků *jako krajinotvorců*.

farmers "are" landscape-shapers



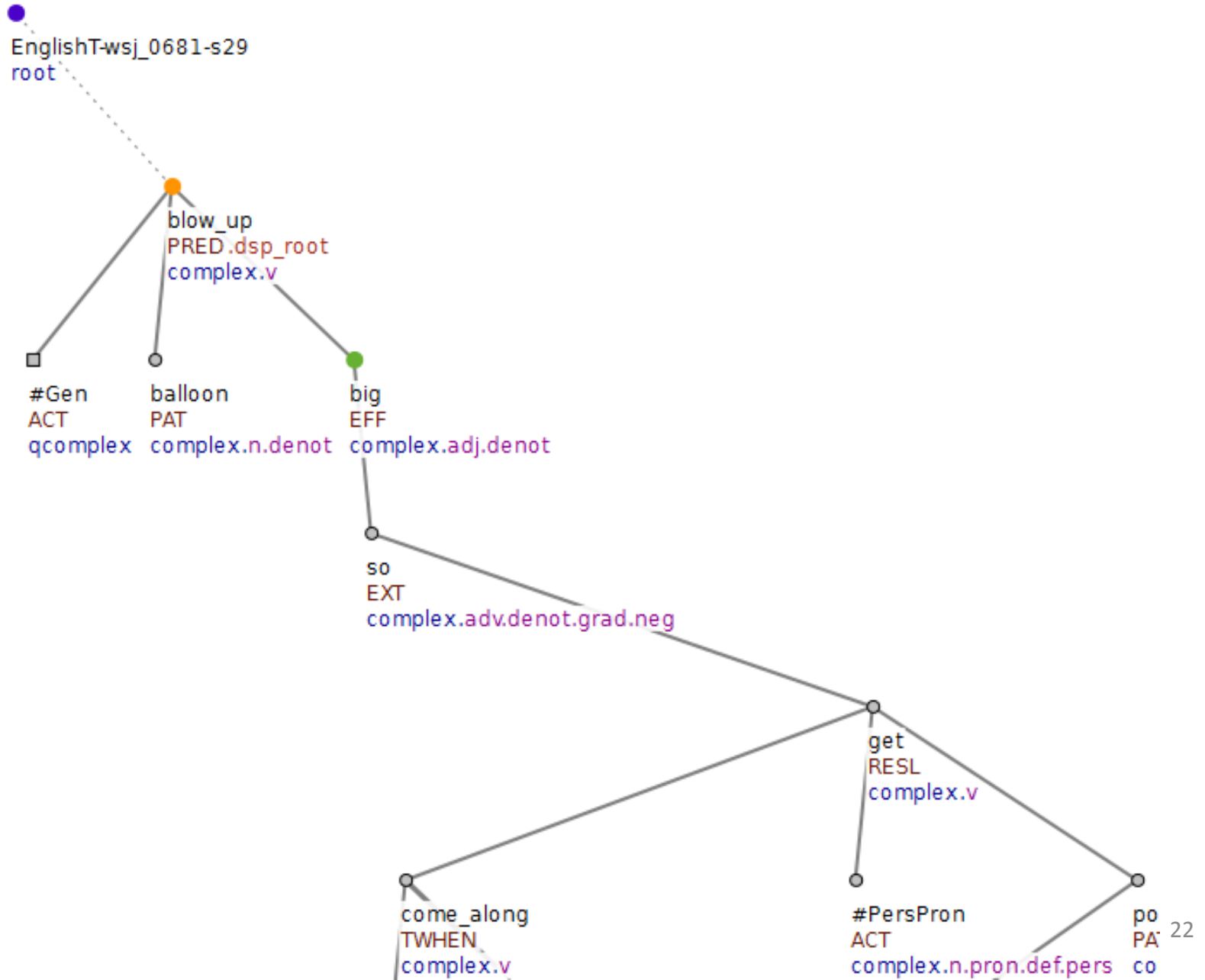
Mr. Krenz's *reputation* *as a heavy drinker*.

Mr. Krenz "is" a heavy drinker

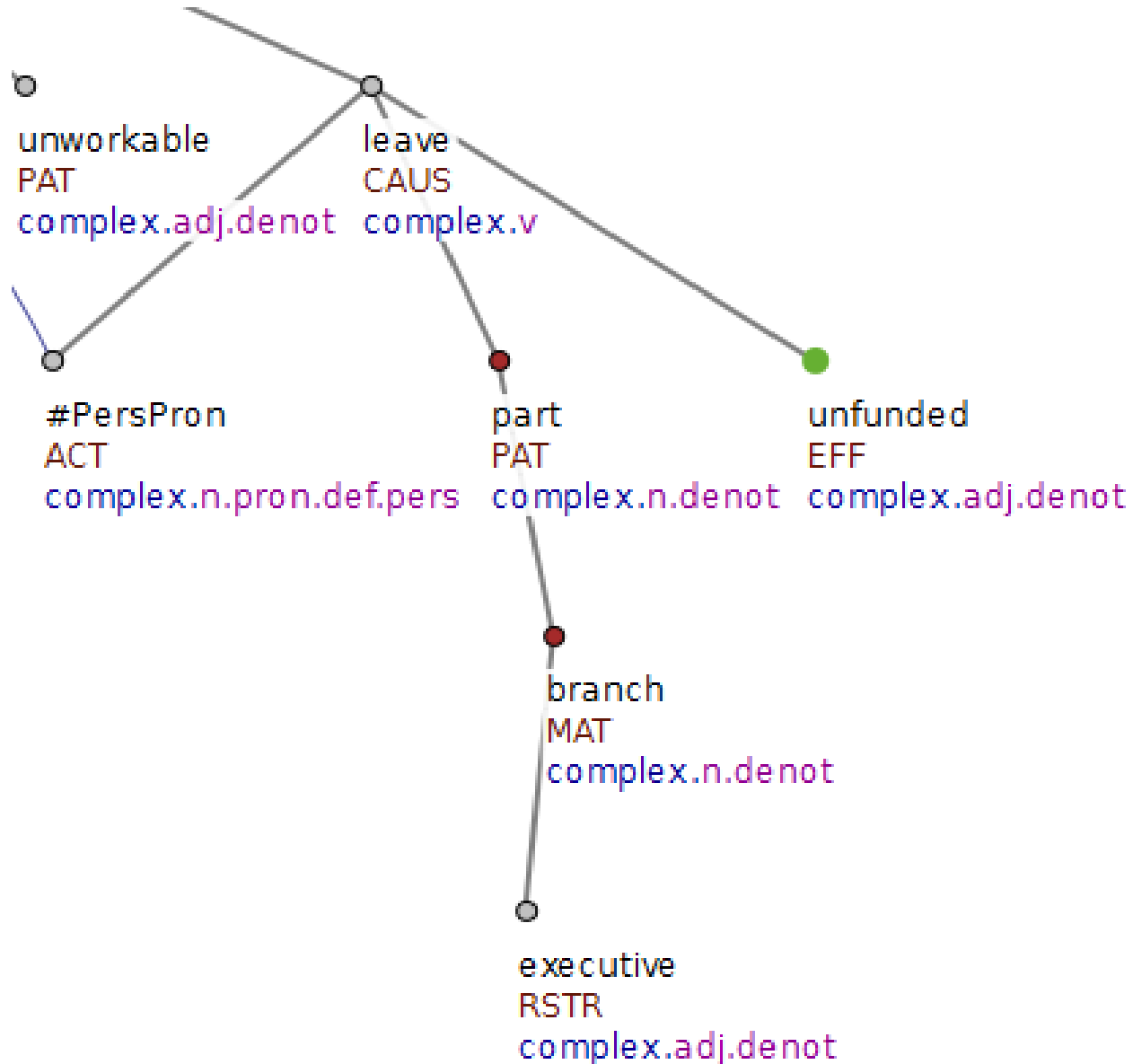
English verb complement

- has a copular relation to the object or the subject (*was X when doing Y*)
- from obligatory to optional
 - *Peter is kind/a kind man.* PAT
 - *We consider Peter a kind man.* EFF
 - *We regard Peter as a kind man.* EFF
 - *I know Peter as a kind man.* EFF
 - *He looks good.* PAT
 - *John, tired, went home.* DESCR
 - *As a kind man, Peter was very popular.* COMPL
 - *Being a kind man, Peter was very popular.* COMPL
 - *We painted the wall green.* RESL/EFF
 - *He stood there, his hands in his pockets.* COMPL- #EmpVerb
 - *He came running.* COMPL

The balloon had been blown up so big that when somebody came along with a pin -- in this case the UAL deal -- we got a little pop.'



it would leave part of the executive branch unfunded



Difficult cases

COMPL

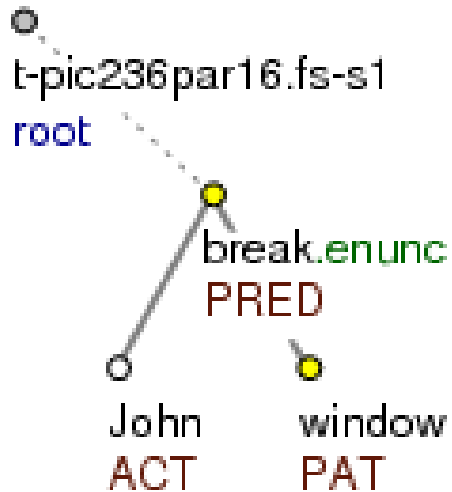
- *The door banged shut.*
- *She married young. COMPL*
- *They sat quiet. COMPL*
- *John met Mary crying. COMPL*
- *Julia, being a nun COMPL, spent much time in prayer.*

Non-COMPL

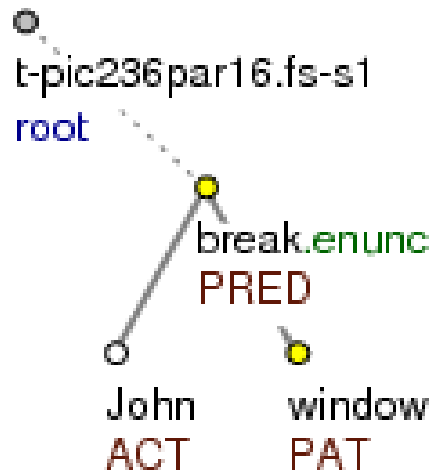
- *The screw worked loose. PAT*
- *She dresses young. MANN*
- *The door fits tight. MANN (adj)*
- *They sat quietly. MANN*
- *John heard Mary crying/cry. PAT*
- *This drives me nuts. DPHR*
- *Julia, who was a nun, DESCR spent much time in prayer.*

MINOR ENGLISH-SPECIFIC PHENOMENA

Cleft Sentences



- *John broke a window. / It was John that broke the window.*

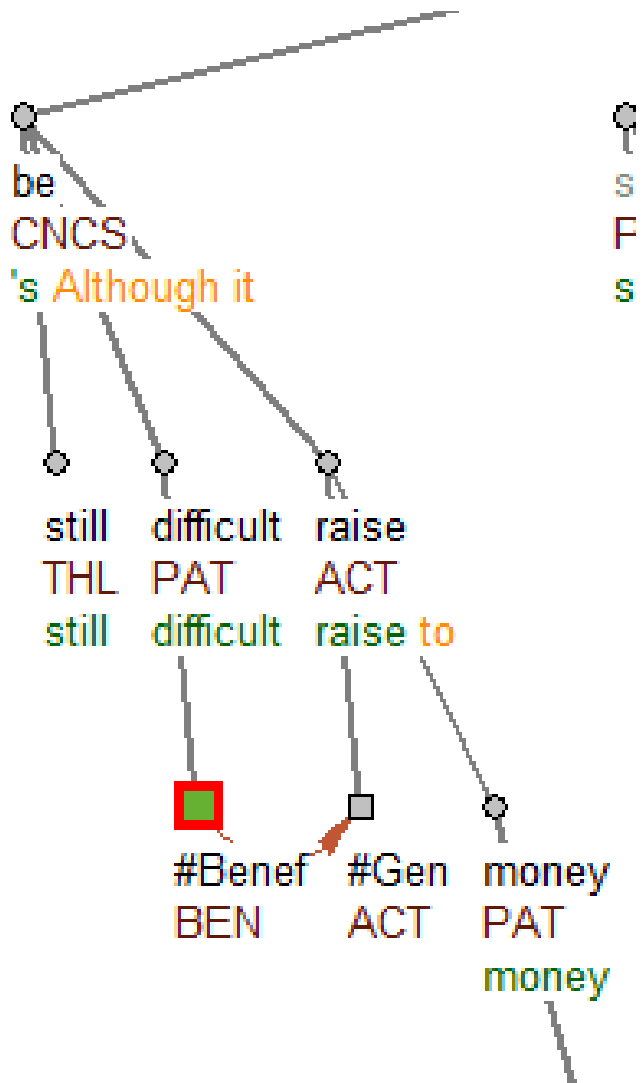


- *It was a window John broke.*

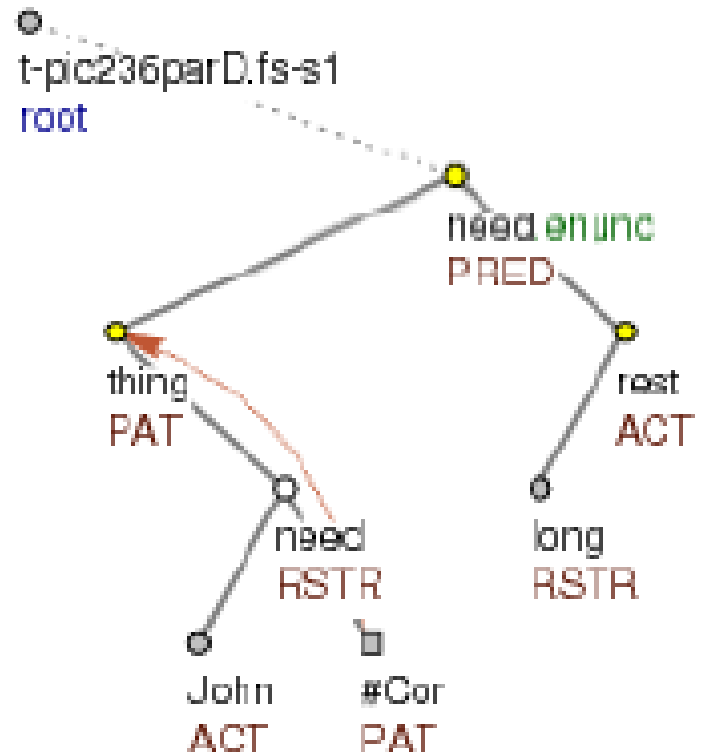
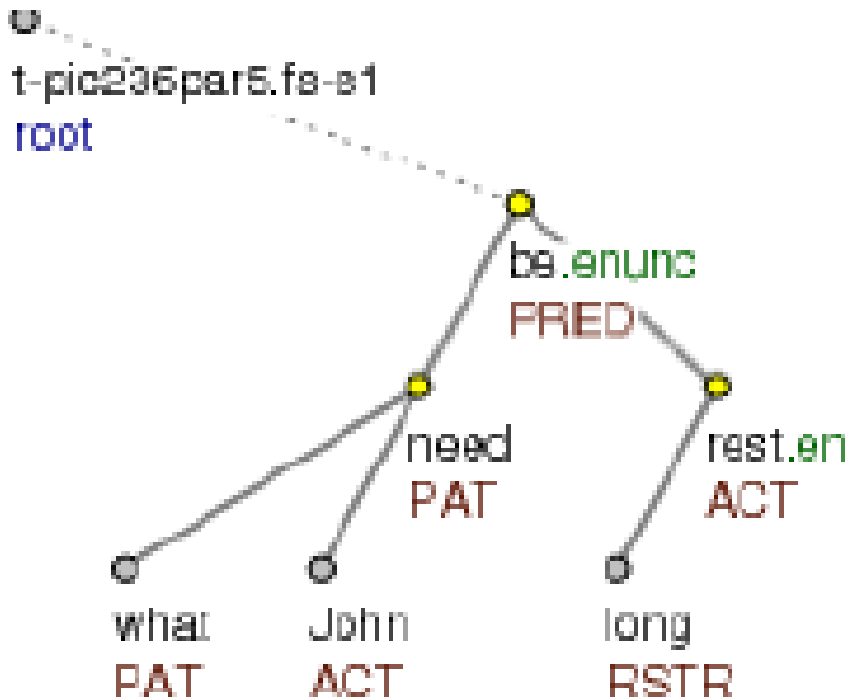
The Pronoun *it*

- **anaphorical** - own TR node:
 - *a hat – it.*
- **deictic/exclamative** - own TR node:
 - It/They is/are my kid(s).
 - It's me!
- **anticipatory/expletive** – auxrf at the predicate verb:
 - *<It> is a pleasure to be here.ACT.*
 - *I find <it> a pleasure.EFF to be here.PAT*
- **prop**– auxrf at the predicate verb:
 - *<It> is late.TWHEN*

Anticipatory/Expletive *it*



Pseudo-cleft sentence



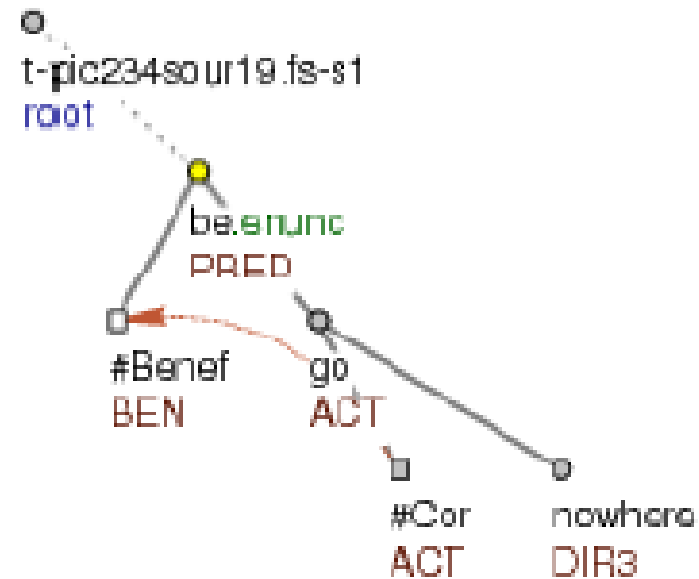
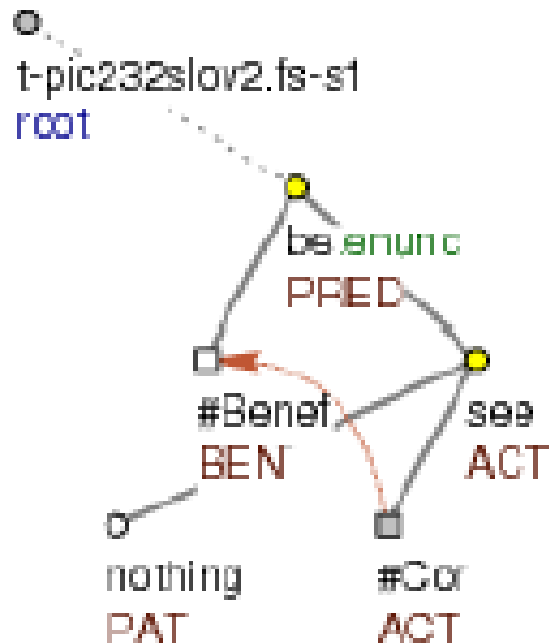
What John needs is a long rest.

The thing John needs is a long rest.

Existential constructions



There is nothing to be seen.



There is nowhere to go.

But there are limits.

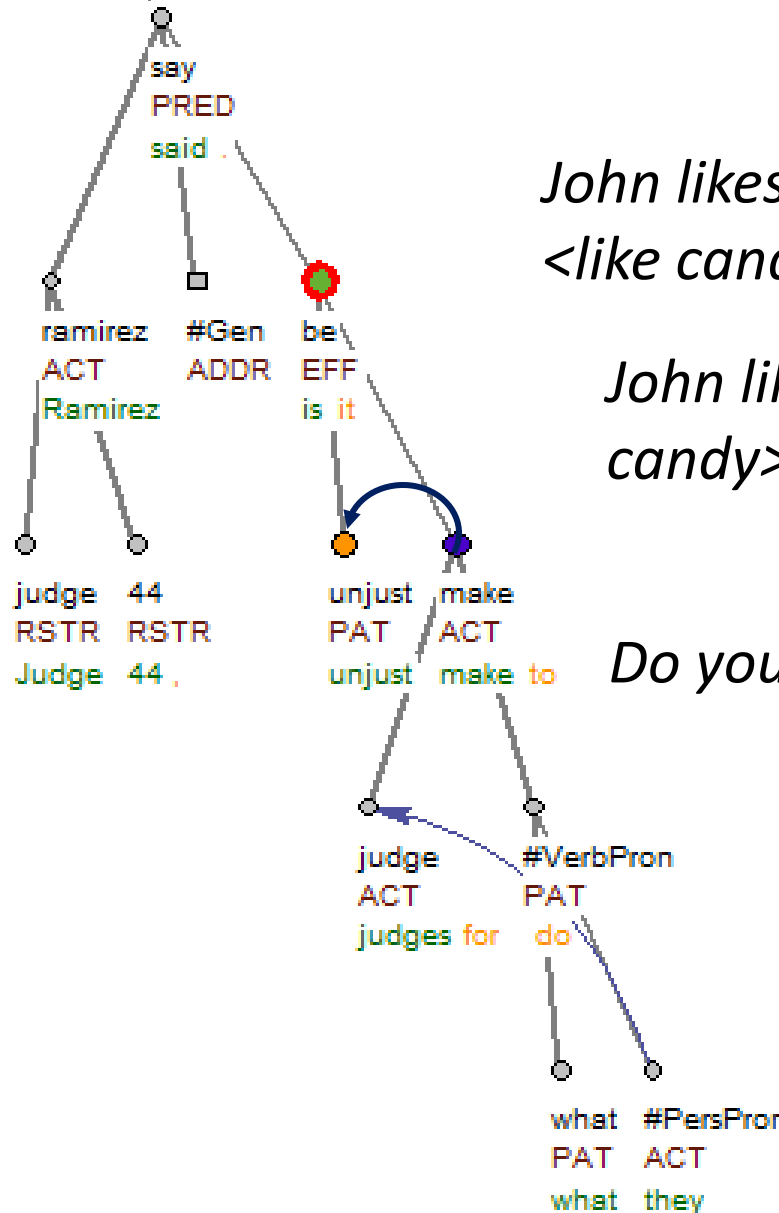
root

Auxiliary *do*

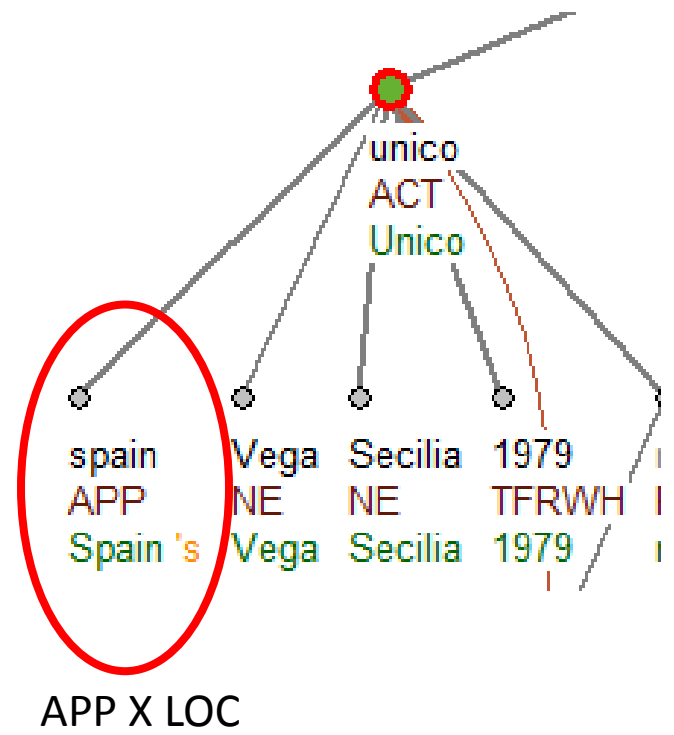
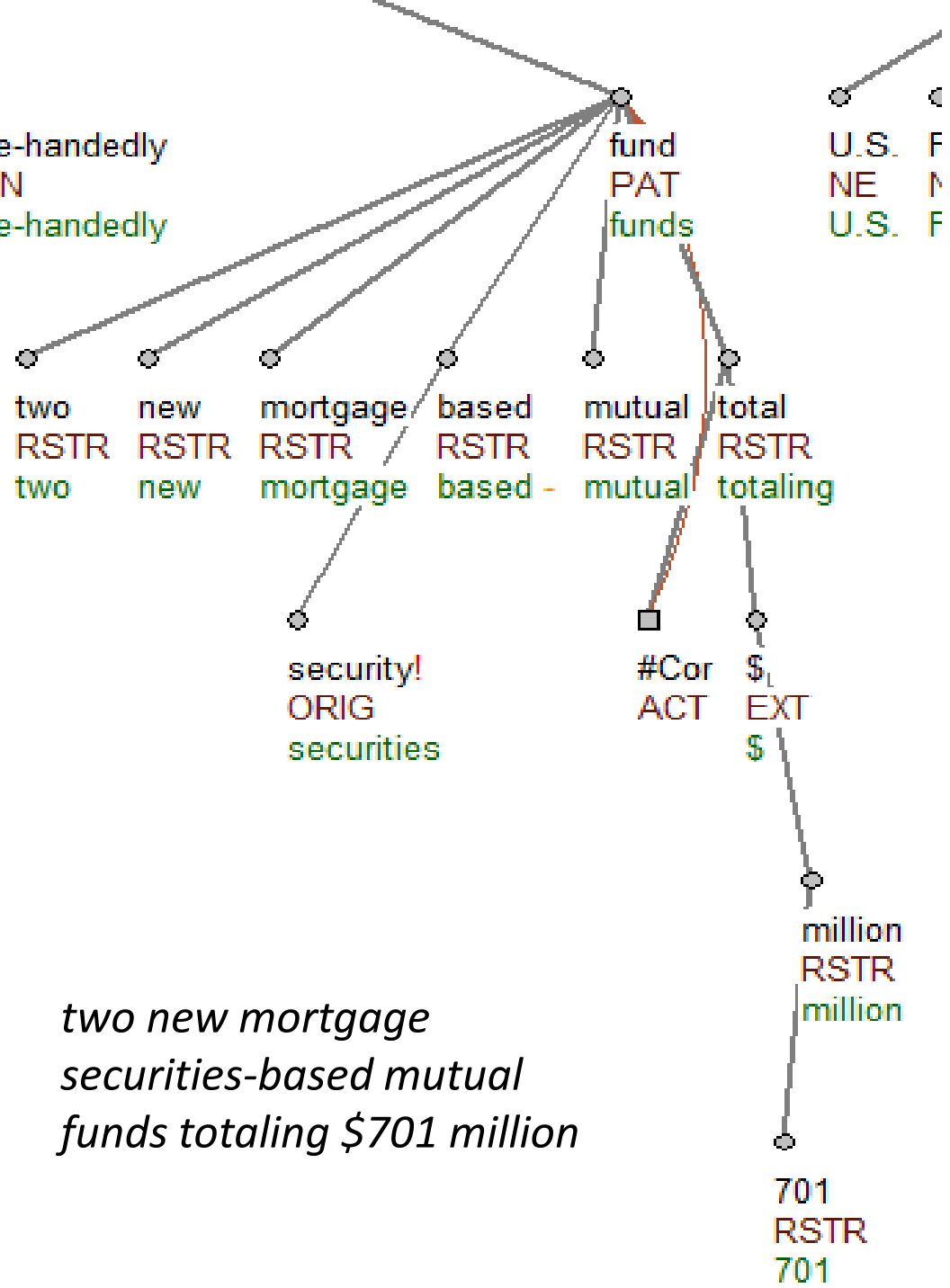
John likes candy but Mary doesn't
 <like candy> = ellipsis

John likes candy. Does Mary <like candy>? = ellipsis

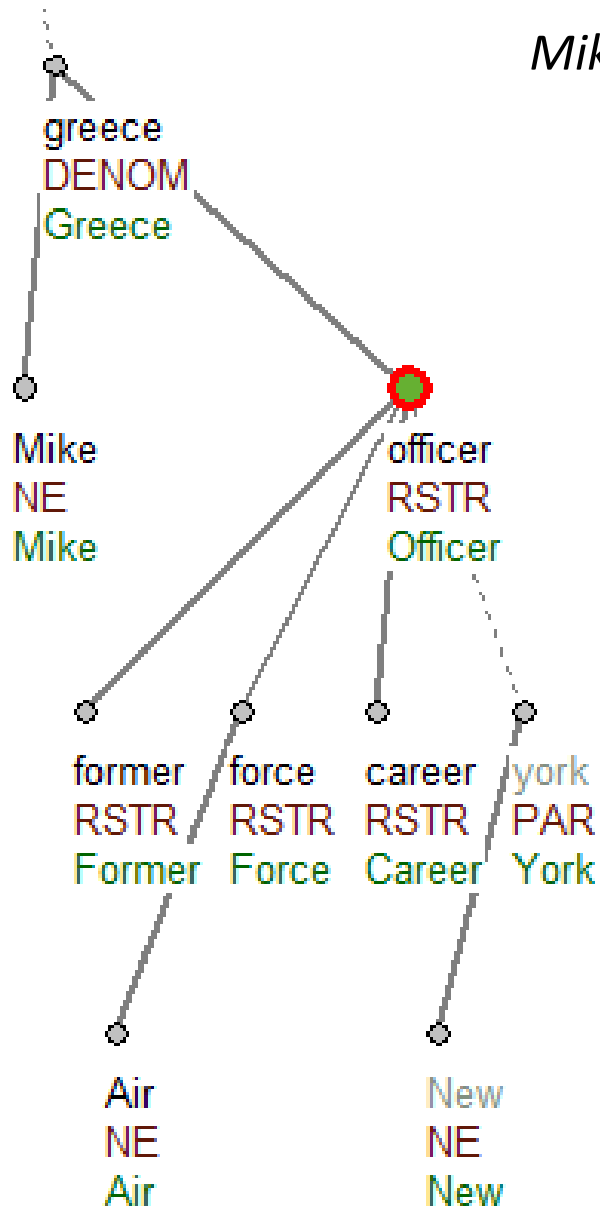
Do you like candy? Yes, I do. ≠ ellipsis



TOTAL UNDERSPECIFICATION: COMPOUND NOUN PHRASES



Mike Greece Former Air Force Career Officer New York



These would deserve multi-word expression annotation.

FUNCTORS IN ADJECTIVES

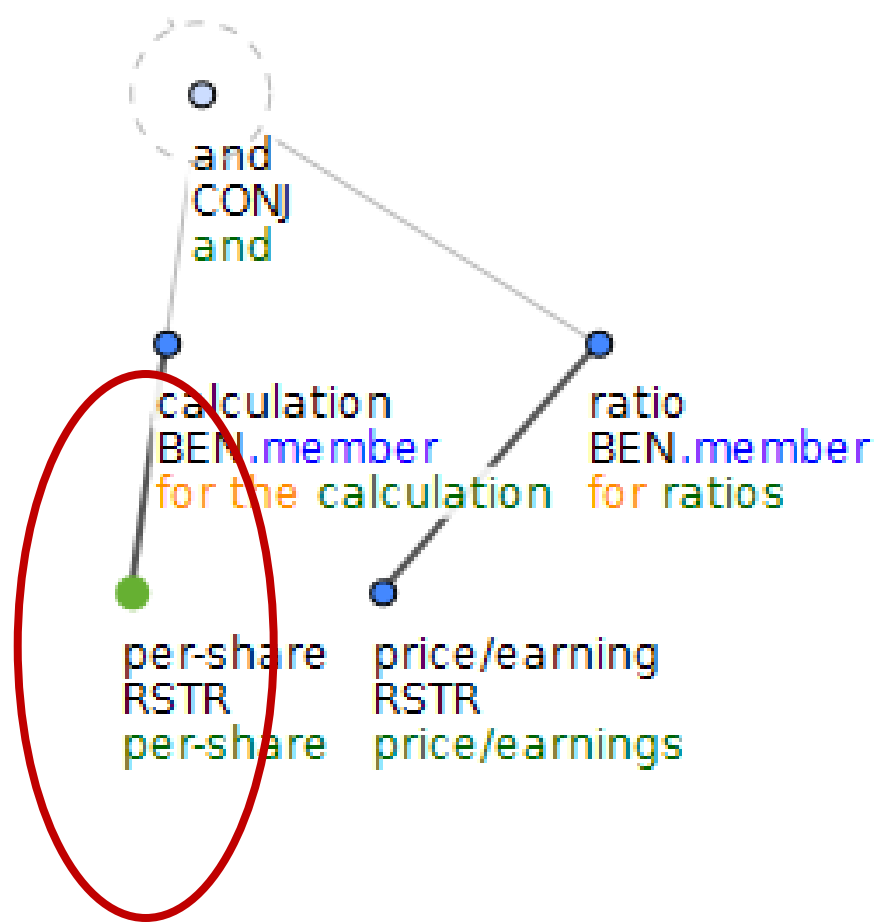
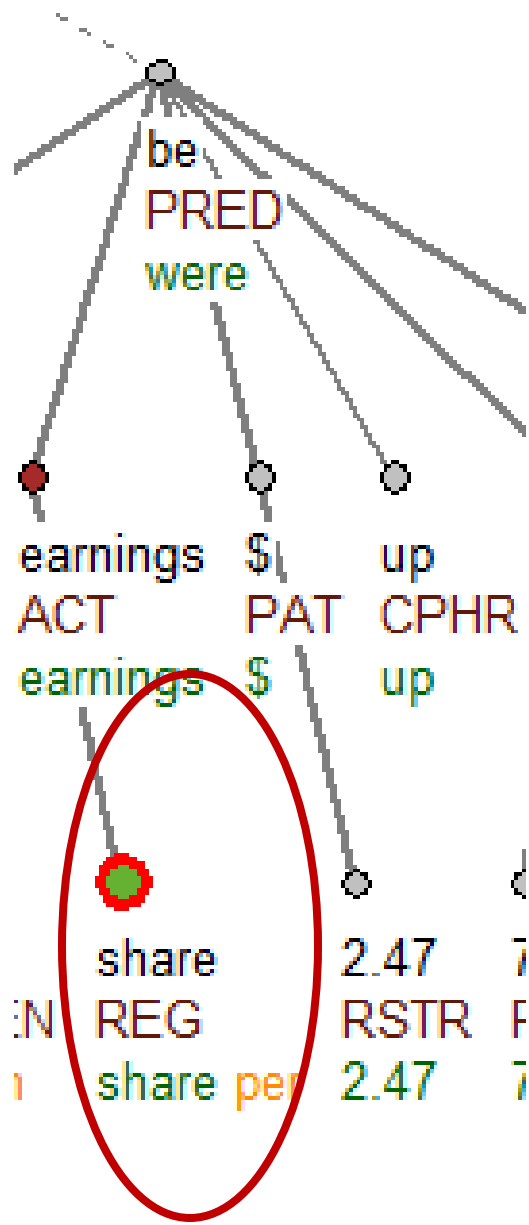
Dependency or own meaning?

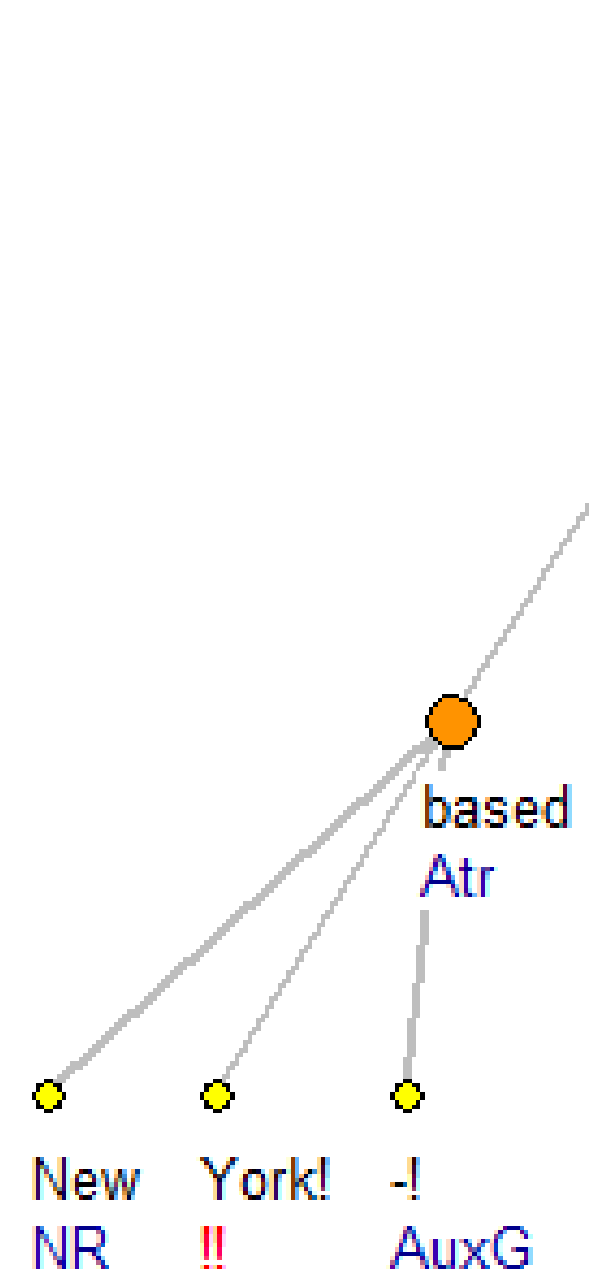
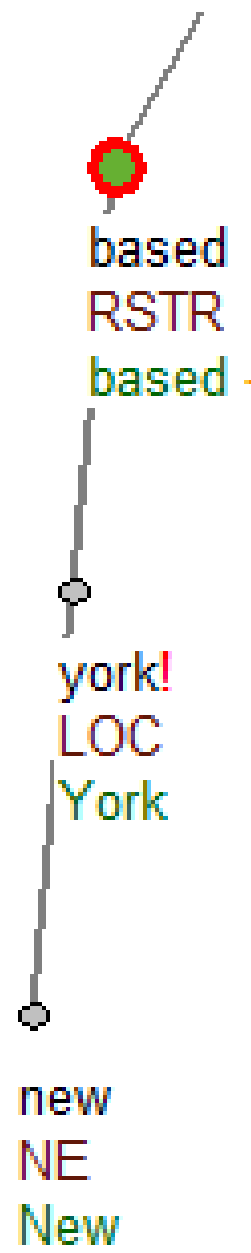
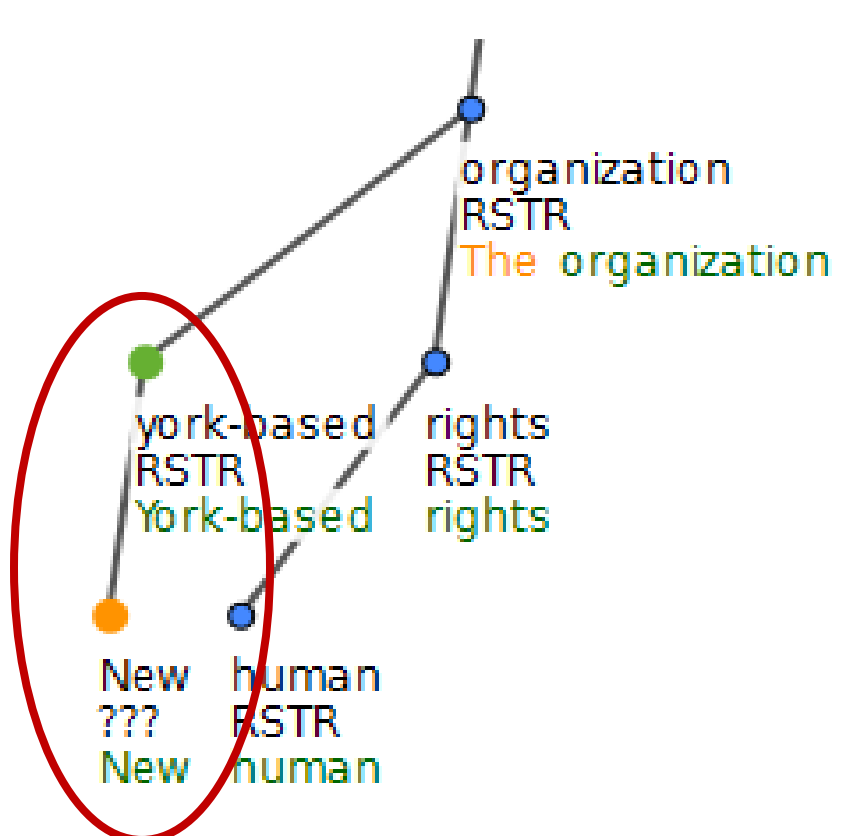
- entity nouns vs. event/property nouns
 - RSTR, DESCR vs. participants and adjuncts
- nominalizations: *-ing, -ment, -ion, -er, -ence...*
 - *American. ACT interference, demand*
 - *American. LOC educators, right, church, singer, showroom*
 - *Reagan consulted the American allies – LOC?*
 - *American share of Japan's market – ACT?*
 - *Critics also say American companies. LOC? [abroad!!!] charge too much.*

Semantically very explicit adjectives

- *past*
- *wooden*
- *unfortunate*
- *old bread* . RSTR X *old scout, old friend* . THL
- *late*. TWHEN *afternoon* X *the late* .RSTR
president
- *plain*. EXT *nonsense, plain girl*. RSTR

PROBLEMATIC TOKENIZATION



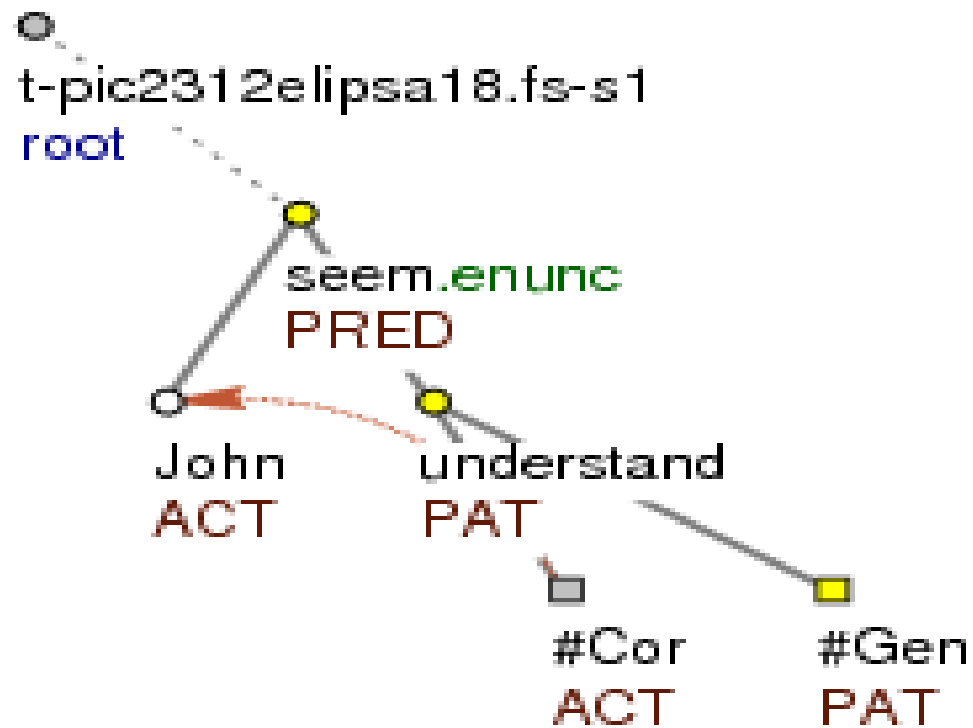


Hyphenated words

- number of t-nodes, t_lemma
- number of a-nodes, m_lemma, m/tag
 - non-compositional: *high-flying*
 - number+noun: *three-day*
 - preposition+noun: *per-share*
 - comparisons: *cornflake-size, lady-like*
 - x+common participle: *LA-based, red-colored*
 - x+uncommon participle: *red-haired*
 - verb+preposition: *hoped-for*
 - Latin preposition: *pre-, anti-, re-, counter-, pro-*
 - *x+only*
 - *non+X*
 - ...

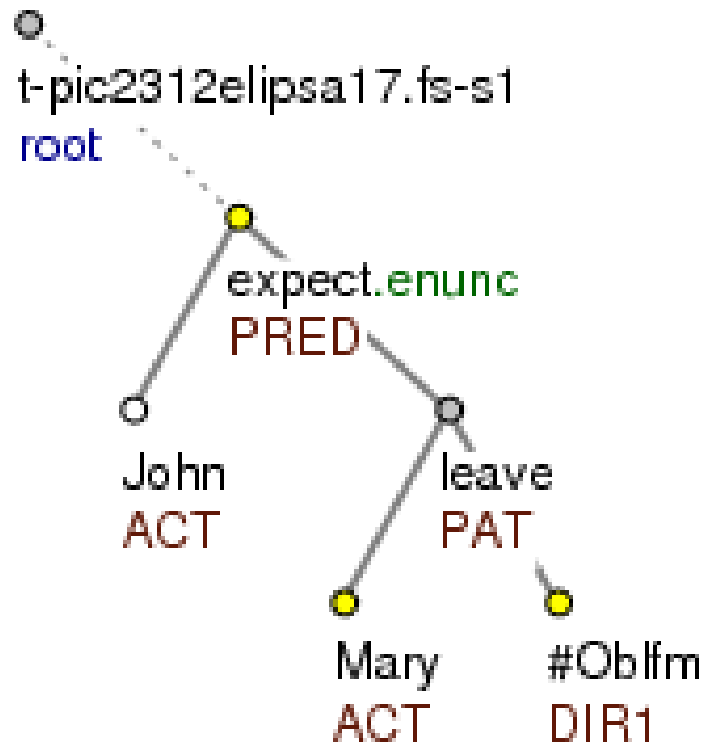
STRUCTURE MISMATCH: ENGLISH NON-FINITE PHRASES

Infinitive – Control



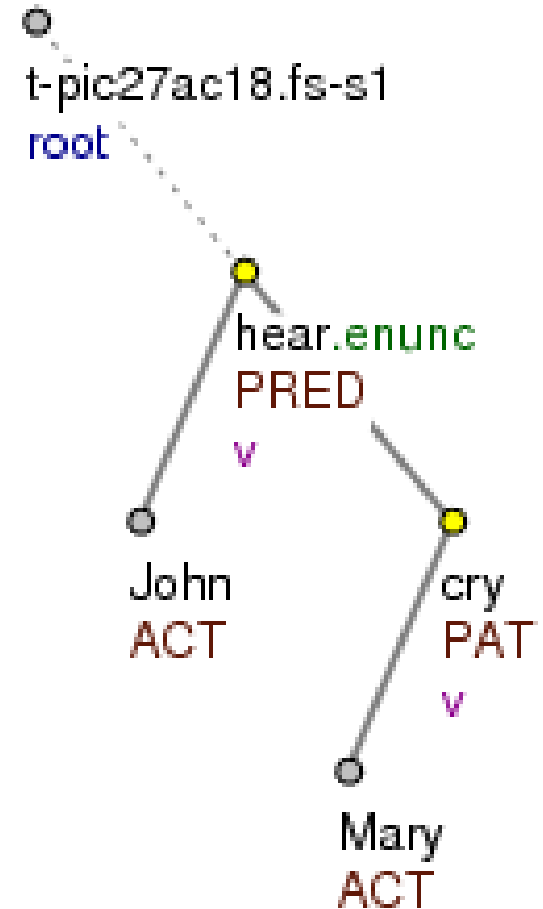
John seems to understand.

Infinitive-Raised Object

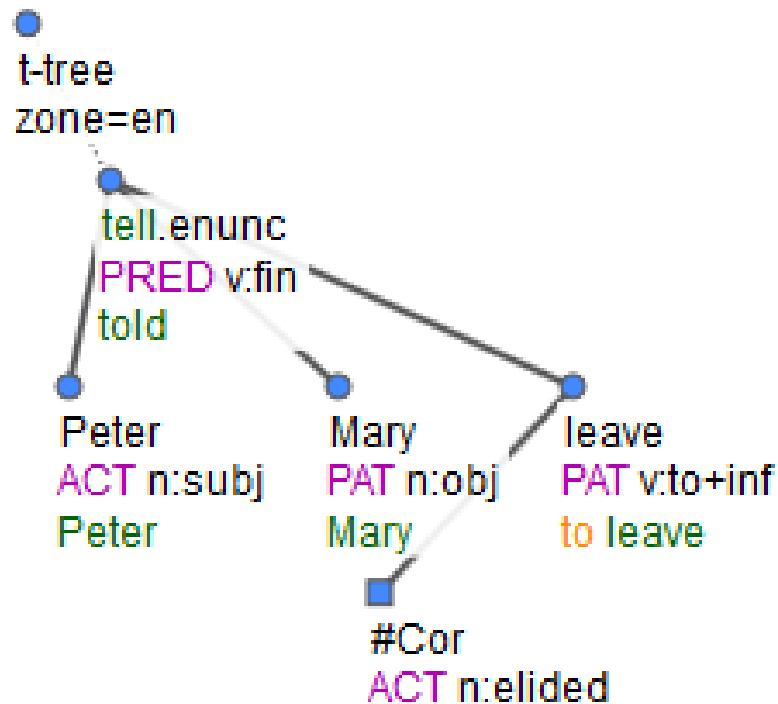


John expects Mary to leave.

John hears Mary cry.

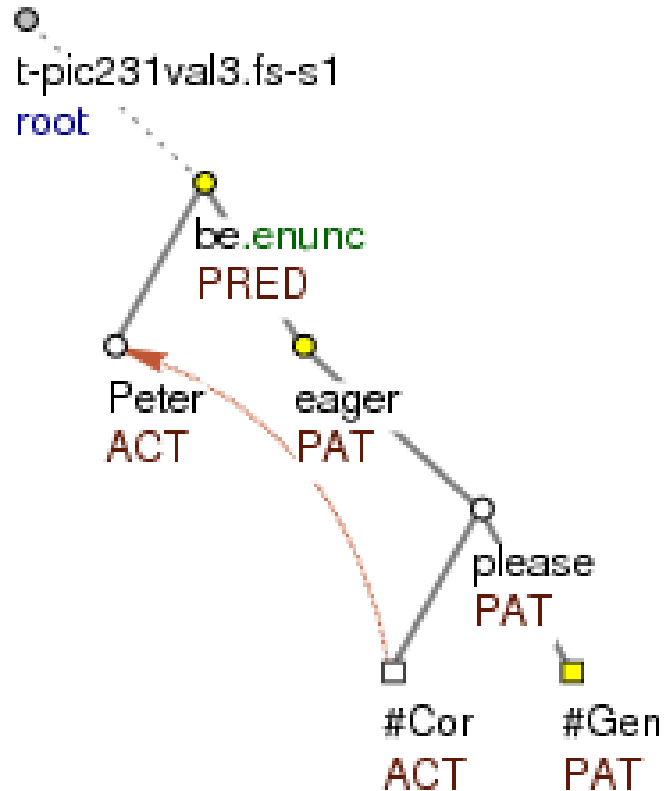


Infinitive – control with object



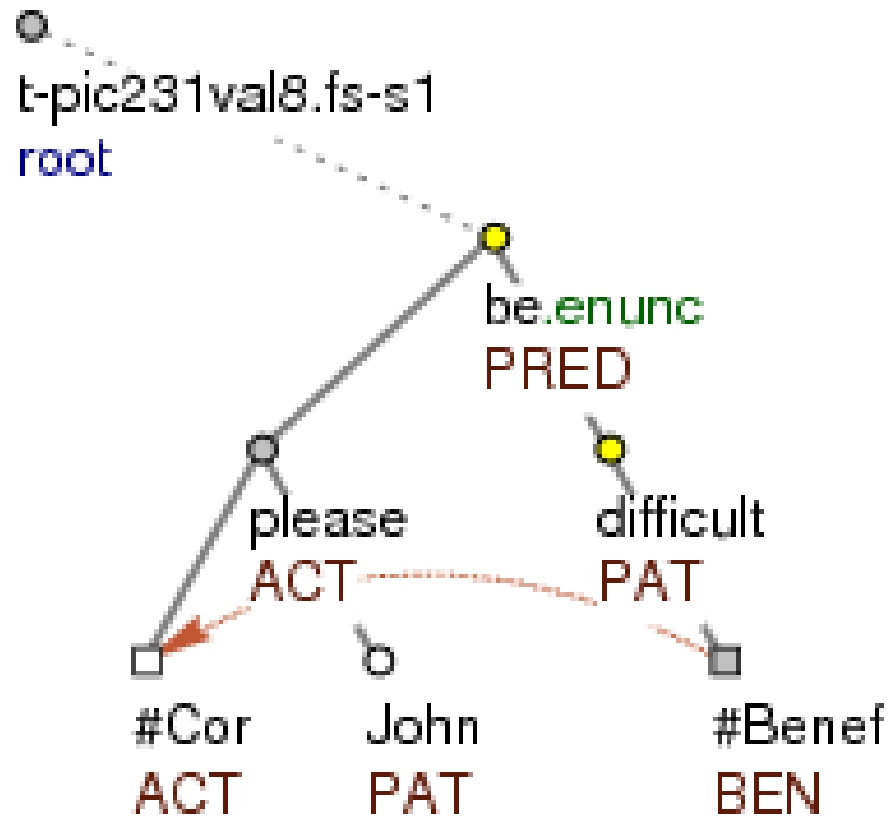
Peter told Mary to leave.

Infinitive Governed by a Predicative Attribute



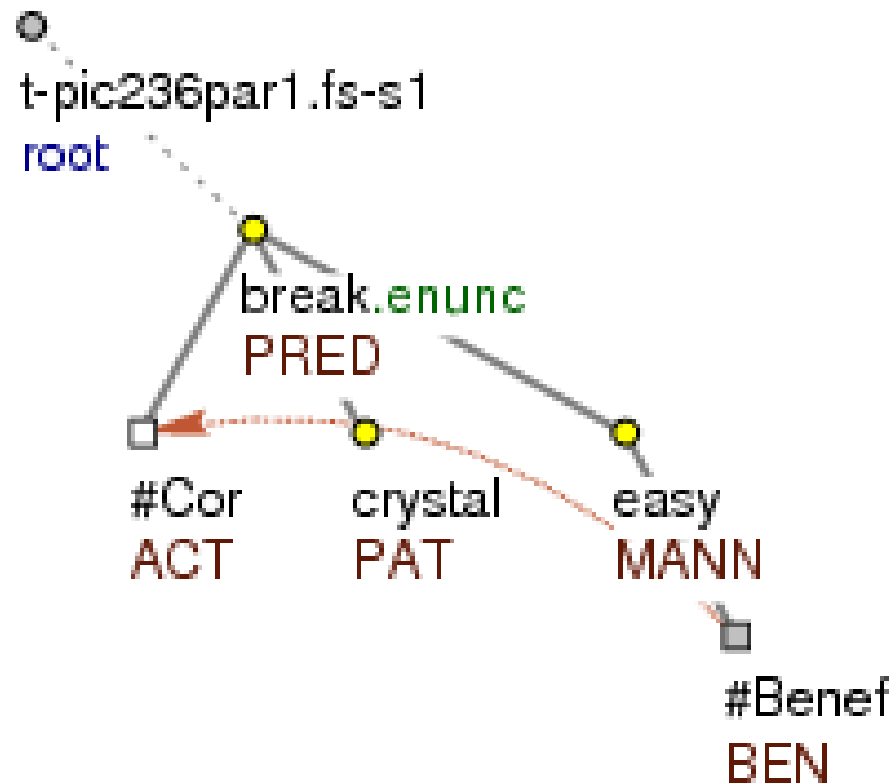
John is eager to please.

Object-Subject Transposition



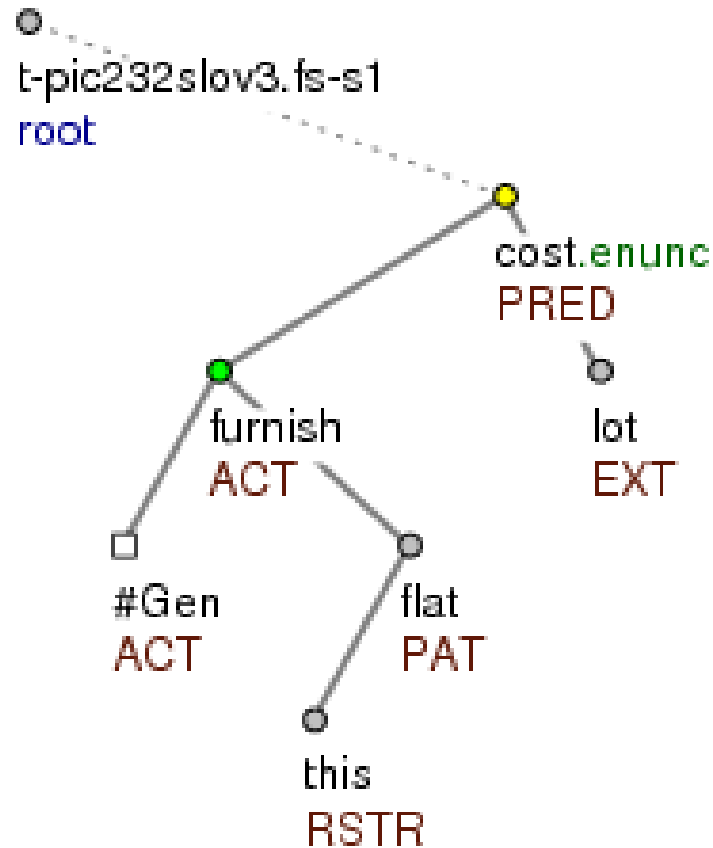
[objsubj=osub3]: *John is difficult to please.*

Object-Subject Transposition



[objsubj=osub1/2] *Crystal breaks easily.* (/at the slightest touch.COND)

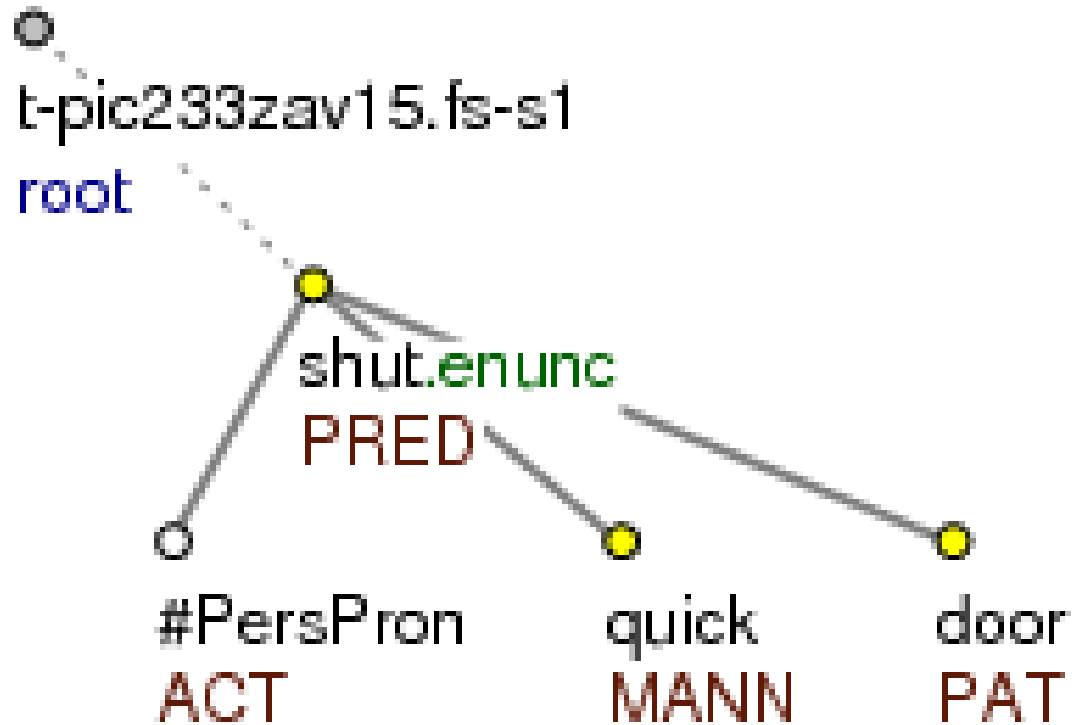
Object-Subject Transposition



[objsubj=osub4] *This flat must have cost a lot to furnish.*

Adverbial adjective

She was quick to shut the door.



Infinitive phrases – coreference in purpose clauses

- *Work began immediately to adapt the existing buildings to accommodate the rows of big glass bolthead flasks that Nicholson favoured for the preparations.*
- *Once these stores are loaded , a process of CICR begins to activate either the IP 3 Rs or RYRs to release calcium.*
- *As was his custom, Sukarno adapted Nehru's actual words to reflect his own hostility to the Dutch and his growing admiration for Japan.*
- *Purchasers of health care have tried to adjust their expected intervention rates to allow for activity in the private sector...*
- *The Romans also followed on these lines but adapted the construction to suit their more complex needs.*
- *She adapted her personal life to suit his wishes, and rationalised her way out of emotional pain caused by Sartre's philandering .*

appoint, employ

- *She used to be with Grand Metropolitan and now, apparently, they've had to appoint men to fill her post.*
- *During most of the sixteenth century , condemnation of women who employed a wet-nurse to suckle their babies was minimal .*
- *but they have to employ more capital to build the new prettily-lit , out-of-town shops that the British want .*
- *Those who employ linguistics to study literature form some kind of social network with distinct practices and a skeletal institutional framework , defined at least by overt inter-disciplinary norms .*

ACT-MEANS alternation

- *John broke the window with a hammer.*
- *The hammer broke the window.*
 - *John used the hammer to broke the window.*

Purpose clause or attributive infinitive?

- *Last autumn the Hypertrophic Cardiomyopathy Association launched a campaign to raise awareness of hypertrophic cardiomyopathy and , together with the Sports Council 's National Sports Medicine Institute , announced a pilot screening programme to detect the condition among young athletes .*
- *In some ranges , you can buy decor panels to fit the front of specially designed built-in appliances, making them match the cupboards .*
- *Letters from home contained frequent reminders to show appreciation.*
- *He accepted the invitation to speak at the conference.*

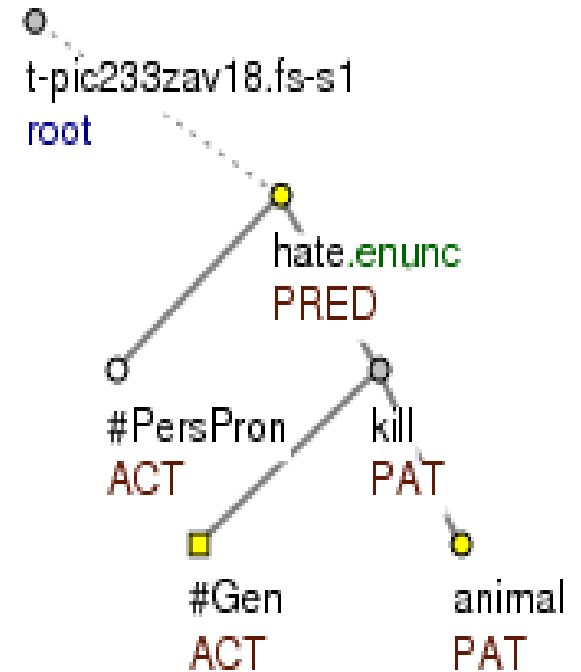
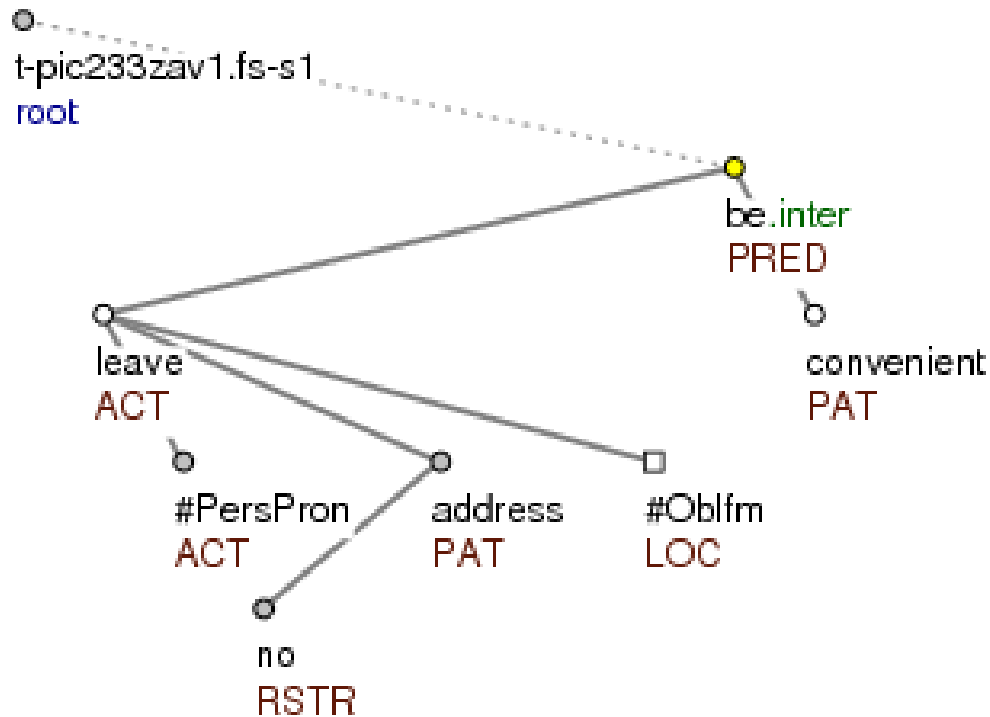
Intransitive = control,
transitive = purpose clause

- *particles begin to vibrate*
- *Duke of Gloucester, who with others began a civil war to remove the King 's favourites from Court.*

Gerund/Present Participle

- can also be used as adjective or noun (see PTB tags)
- modified by
 - an adverb = gerundial clause, verbal frame
 - an adjective = noun, noun frame (not existent yet)
- transitive verbs: direct object without *of* = gerundial clause

Gerundial Clauses



*His leaving no address was most inconvenient.
I hate killing animals.*

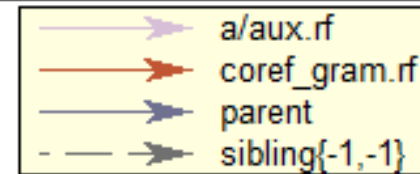
NLP ISSUES

TR parse

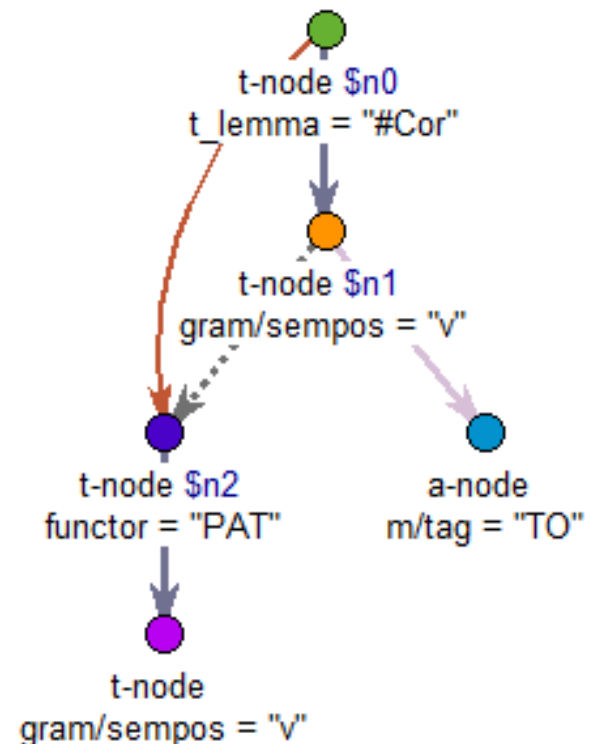
- no TR parser from plain text
- phrase-structure-to a-layer
- a-layer to t-layer
- => never better than phrase-structure parsers
- training data small + too many structures to learn

The biggest problems

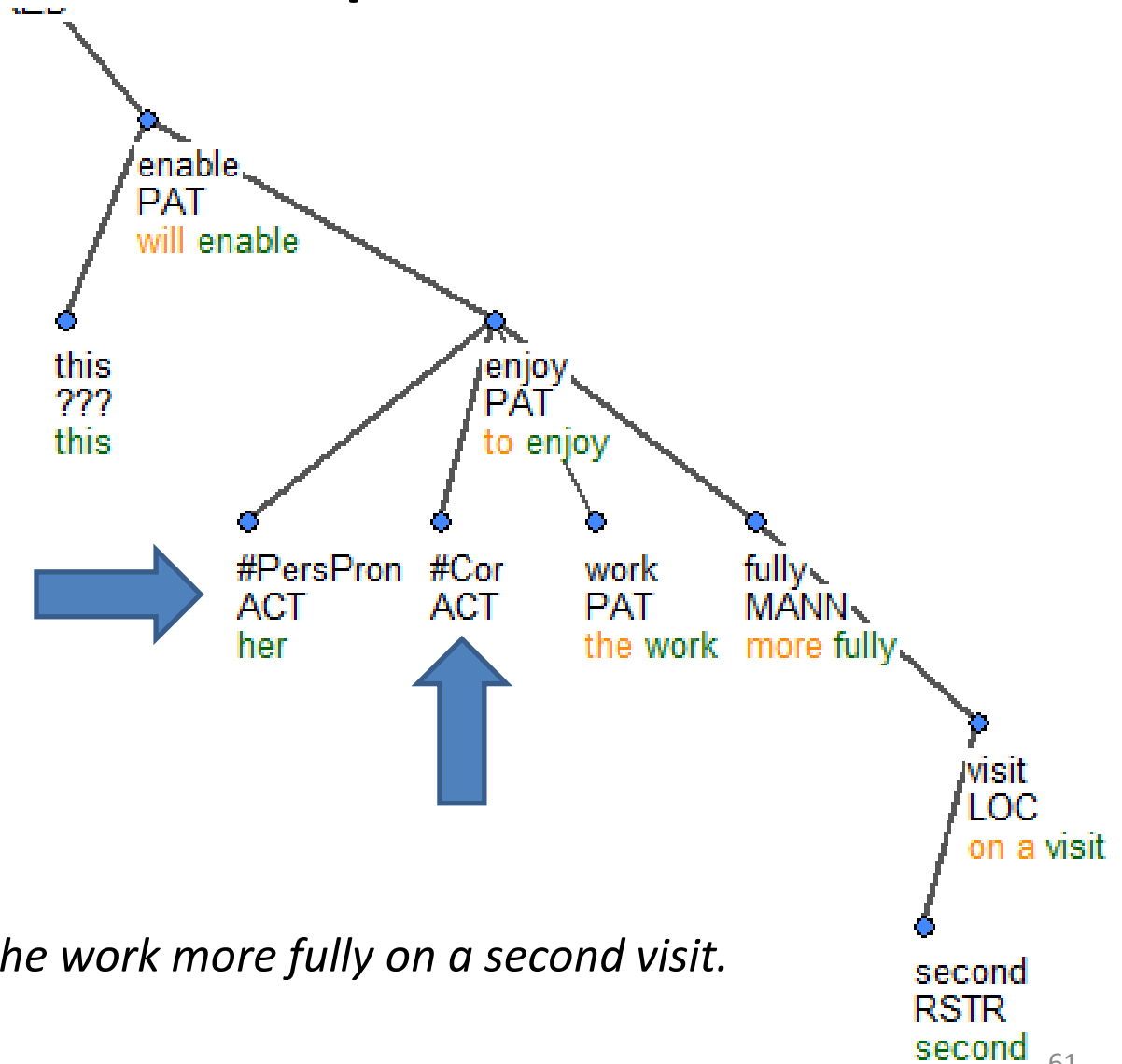
- already at a-layer
- coordinations
- appositions
- several participants of the same type
- no chance to get the sequence *v-obj-infinite clause* correct
- unreliable ellipsis restoration



Tree Query
q-12-05-12_203546



Example



This will enable her to enjoy the work more fully on a second visit.

Explore, exploit, enjoy!



This is only a small sample of the corpus. You need to **order and properly license** the corpus to browse it in its entirety.

● Section: ◀ 00 ▶

● File: ◀ 01 ▶

● Sentence: ◀ 1 ▶

[en] Pierre Vinken, 61 years old, will join the board as a nonexecutive director Nov. 29.

[cs] Jednašedesátiletý Pierre Vinken se připojí ke správní radě jako nevýkonný ředitel dne 29. listopadu.

