



Overview

- * **Annotation of textual phenomena in the Prague Dependency Treebank (April 27)**
- * Coreferential expressions in English and Czech (April 28)
- * Coreference in Czech and cross-lingually - ideas and perspectives (April 30)



“made in ÚFAL” Institute of Formal and Applied Linguistics, Charles University in Prague, Faculty of Mathematics and Physics)



HUMAN CAPITAL
NATIONAL COHESION STRATEGY

EUROPEAN UNION
EUROPEAN
SOCIAL FUND



Annotation of textual phenomena in the Prague Dependency Treebank

Anna Nedoluzhko,
Warszawa, 27.4.2015





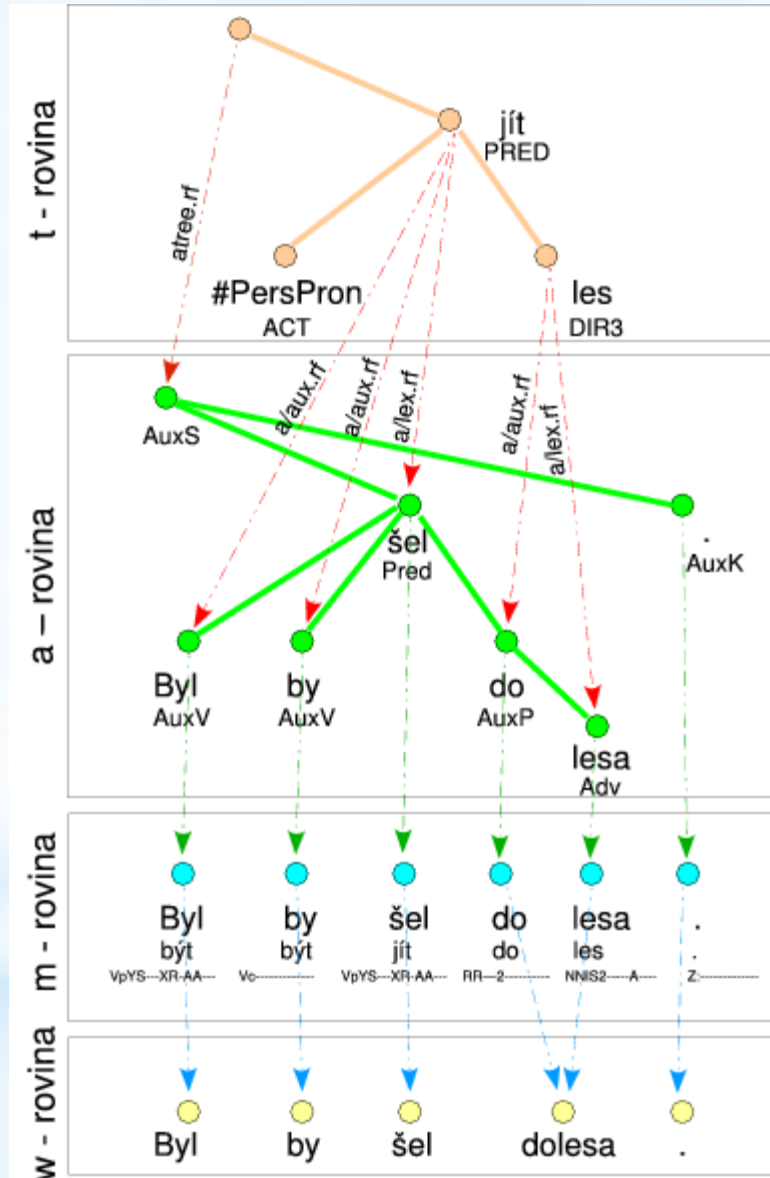
Prague Dependency Treebank (PDT 3.0)

- * The Prague Dependency Treebank - linguistically annotated corpus, Czech newspaper texts (<http://ufal.mff.cuni.cz/pdt2.0>), data themselves available at LDC (No. LDC2006T01)
- * The three PDT layers - capture grammatical information: morphological, surface shape (analytical) and underlying syntactic (tectogrammatical)
- * 3165 documents (= 49431 sentences = 833195 tokens)
- * multi-purpose
- * for other corpora see <http://ufal.mff.cuni.cz/data>



Byl by šel dolesa.

(He) would go to the forest.





Coherence-related annotations

The tectogramatical layer of PDT 3.0 includes annotation of linguistic phenomena from the perspective of discourse structure and coherence.

Four coherence-related subprojects:

- * annotation of **coreference**,
- * annotation of **bridging relations**,
- * ellipsis
- * contextual boundness (tfa)
- * **discourse connectives**, discourse units linked by them and semantic relations between these units



Our “roots”

- * Functional Generative Description (Sgall et al., 1986)
- * Coreference
 - * Ontonotes (Pradhan et al. 2007), Poesio (MATE, GNOME, VENEX, ARRAU), AnCora (Recasens), PoCoS (Chiarchos-Krasavina) et al. but on dependency trees
- * Discourse
 - * Penn Discourse Treebank (for English, PDTB, Prasad et. al., 2008) - identification of discourse markers and relations they express



State of coherence-related annotations in Prague corpora

	PDT 3.0	PCEDT (Prague Czech-English Dependency Treebank)	
		English	Czech
grammatical and pronoun (zeros + #PersPron) coreference	YES	YES	YES
textual full-NP coreference	YES	YES	YES
bridging	YES	no (split ante)	no (split ante)
Ellipses reconstruction	YES	YES	YES
discourse relations	YES	no	no
Information structure	YES	planned	work in progress



HUMAN CAPITAL
NATIONAL COHESION STRATEGY

EUROPEAN UNION
EUROPEAN
SOCIAL FUND



Coreference



Coreference

- * **grammatical coreference** -
 - * grammar-driven by the use of pronouns,
 - * mostly possible to identify the antecedent on the basis of grammatical rules
 - * within one sentence
- * **textual coreference** -
 - * not restricted to grammatical means alone
 - * can be realised by pronominalisation, grammatical agreement, repetitions, synonyms, paraphrasing, hyponyms/hyperonyms, etc.
 - * often occurs between entities in different sentences



Grammatical coreference

- * arguments in constructions with verbs of control
 - *John asked **Mary** to [#Cor.ACT] come.*
 - ***John** submitted a [#QCor.ACT] complaint to the police.*
- * reflexive pronouns
 - ***John** shaved **himself**.*
- * relative pronouns
 - ***John, who** came late, apologized.*
- * coreference with verbal modifications that have dual dependency
 - *Honza saw **Hanka** [#Cor.ACT] run around the lake.*
- * reciprocity
 - *John and Mary kissed [#Rcp.PAT] .*



Textual coreference

- * personal and possessive pronouns (*Peter* returned. He forgot his hat),
- * demonstrative pronouns *ten, ta, to* (=that)
- * with textual ellipsis (zeros), where a new node with the t-lemma substitute #PersPron is added to the textogrammatical tree
- * nouns (*Helena* asked her mother to wait for her, but mother ignored her daughter's wish.)
- * local adverbs (*Helen* asked mother to come to Prague but she decided not to go there.)
- * some adjectives (*He* came to Prague and found the Prague atmosphere quite casual.)
- * reference to events (*Helena* asked her mother to wait for her, but mother ignored her daughter's wish).

If antecedent is a whole segment of (previous) text larger than one sentence (phrase) – special type of textual coreference segm(ent) without explicitly marked antecedent:

The custom union will formally function for twelve more months, but in fact the relations will be of a kind of international trade. The bureaucracy will go up. The latest steps of the Slovak government confirm this direction.



Textual coreference - types

- * **SPEC(ific)** for coreference of NPs with specific reference
(*Helen_a asked her_a mother_b to wait_c for her_a, but mother_b ignored her_b daughter_a's wish_c*)
- * **GEN(eric)** for coreference of NPs with generic reference, e.g.

Your comments implied we had discovered that the principal cause of homelessness is to be found in the large numbers of mentally ill and substance-abusing people in the homeless population. [...] The study shows that nearly 40% of the homeless population is made up of women and children and that only 25% of the homeless exhibits some combination of drug, alcohol and mental problems



Exophora

An expression refers to situations or reality external to the text - special attribute coref_special, type exoph

- * *Dokončeny by měly být [...] na sídlišti Barrandov v těchto dnech. (= It should be finished [...] in Barrandov district in these days [meaning, in the recent days])*
- * *A tu se dostáváme zpět k počátku tohoto textu . (= With this, we come back to the beginning of this text)*
- * *Informace v tomto přehledu jsou bezplatnou službou podnikatelům. (=The information in this report is a free service to businessmen.)*



HUMAN CAPITAL
NATIONAL COHESION STRATEGY

EUROPEAN UNION
EUROPEAN
SOCIAL FUND



Bridging Relations



HUMAN CAPITAL
NATIONAL COHESION STRATEGY

EUROPEAN UNION
EUROPEAN
SOCIAL FUND



Bridging Relations

- **part – whole** (PART_WHOLE and WHOLE_PART)

Germany – Bavaria - Munich



Bridging Relations

- **part – whole** (PART_WHOLE and WHOLE_PART)
Germany – Bavaria - Munich
- **set — subset/element of set** (SET_SUB and SUB_SET)
students – some students – a student



Bridging Relations

- **part – whole** (PART_WHOLE and WHOLE_PART)

Germany – Bavaria - Munich

- **set — subset/element of set** (SET_SUB and SUB_SET)

students – some students – a student

- **function - object** (P_FUNCT and FUNCT_P)

prime-minister – government, trainer – football team



Bridging Relations

- **part – whole** (PART_WHOLE and WHOLE_PART)

Germany – Bavaria - Munich

- **set — subset/element of set** (SET_SUB and SUB_SET)

students – some students – a student

- **function - object** (P_FUNCT and FUNCT_P)

prime-minister – government, trainer – football team

- **semantic contrast** (CONTRAST)

*A přesvědčen jsem ještě o jednom - je třeba mít **vysoké cíle** a s **malými [cíli]** se nespokojit. (= And I am sure about one thing: it is necessary to have **lofty aims** and not to be satisfied with **small (ones).**)*



Bridging Relations

- **part – whole** (PART_WHOLE and WHOLE_PART)

Germany – Bavaria - Munich

- **set — subset/element of set** (SET_SUB and SUB_SET)

students – some students – a student

- **function - object** (P_FUNCT and FUNCT_P)

prime-minister – government, trainer – football team

- **semantic contrast** (CONTRAST)

*A přesvědčen jsem ještě o jednom - je třeba mít **vysoké cíle** a s **malými [cíli]** se nespokojit. (= And I am sure about one thing: it is necessary to have **lofty aims** and not to be satisfied with **small (ones).**)*

- **explicit anaphora without coreference** (ANAF)

*"**Duha?**" Kněz přiloží prst **k tomu slovu**, aby nezapomněl, kde skončil. - "A **rainbow?**" The priest pointed to **the word** ...*

*Jak se Vám zamlouvala **Pragobanka Cup?** - **Takovou/podobnou/stejnou akci** bychom také uvítali - How did you like **the Pragobanka Cup?** We would welcome **a similar event** ...*



Bridging Relations

- **part – whole** (PART_WHOLE and WHOLE_PART)

Germany – Bavaria - Munich

- **set — subset/element of set** (SET_SUB and SUB_SET)

students – some students – a student

- **function - object** (P_FUNCT and FUNCT_P)

prime-minister – government, trainer – football team

- **semantic contrast** (CONTRAST)

*A přesvědčen jsem ještě o jednom - je třeba mít **vysoké cíle** a s **malými [cíli]** se nespokojit. (= And I am sure about one thing: it is necessary to have **lofty aims** and not to be satisfied with **small (ones).**)*

- **explicit anaphora without coreference** (ANAF)

*"**Duha?**" Kněz přiloží prst **k tomu slovu**, aby nezapomněl, kde skončil. - "A **rainbow?**" The priest pointed to **the word** ...*

*Jak se Vám zamlouvala **Pragobanka Cup?** - **Takovou/podobnou/stejnou akci** bychom také uvítali - How did you like **the Pragobanka Cup?** We would welcome **a similar event** ...*

- **other** (REST)

*family (grandfather - grandson), place – inhabitant, author – work, same denomination to support cohesion of the text (**a chance helped – another chance entered the game** ...) and event – participant of the event (**enterprise - entrepreneur**)*



Coreference and Bridging statistics

type of relation	absolute numbers
all textual coreference links	86,349
textual coreference (specific NPs)	20,243 (pronouns)+50,593 (nouns) = 70,836
textual coreference (generic NPs)	3,095(pronouns)+12,418(nouns) = 15,513
all bridging links	32,171
bridging SUBSET	5,820(SUB_SET) +12,580(SET_SUB) = 18,400
bridging PART	1,982(PART_WHOLE)+4,372(WHOLE_PART) = 6,354
bridging FUNCT	1,719(P_FUNCT)+418(FUNCT_P) = 2,137
bridging CONTRAST	2,238
bridging ANAF	802
bridging REST	2,212
percentage of nodes where a link starts, counting all text-coref and bridging	17.6%



Tree Editor TrEd

- * PML (Prague Markup Language) - XML-based format, designed for annotation of treebanks
- ↓
- * fully customizable tree editor TrEd (Pajas & Štěpánek 2008)
- ↓
- * extensions, included as modules
- ↓
- * the extensions for Bridging and Coreference, for discourse, etc.



The Annotation Tool - coreference and bridging module

- * pre-annotation of the data with highly probable coreference relations
- * supporting features implemented into the annotation tool - help **during** the annotation process



Pre-annotation

list of pairs of words that with a high probability form a coreferential pair in texts (cca. 6,000 couples)

Praha (noun) - pražský (adj.) (Prague - Prague)

He arrived in Prague and found the Prague atmosphere quite casual

USA - United States of Amerika



Annotation

* *manual pre-annotation for nodes with the same t_lemma*

- (1) Generál Jiří Nekvasil: V české armádě se hrozné věci nedějí
- (7) Náčelník generálního štábu Jiří Nekvasil nám k tomu řekl, že nepořádek v armádě nikdy nezastíral, popírá však, že se v ní dějí hrozné věci.
- (8) Poměry v útvech podle něj odpovídají atmosféře ve společnosti.
- (9) Armáda jde krok za krokem k lepšímu, říká Nekvasil.
- (10) Za nepořádky náčelník generálního štábu považuje nízkou kázeň vojáků na veřejnosti, při výcviku nebo strážní službě.
- (11) Starosti mu dělá nedodržování bezpečnostních zásad a neoprávněná manipulace se zbraněmi.
- (14) Nekázeň se podle generála týká útvarů, jež transformace teprve zasáhne, zejména když mají být zrušeny.
- (17) O panujícím strachu v armádě Nekvasil neví.
- (18) Po mnoha letech podle něj dochází k narovnání vztahů mezi veliteli a podřízenými.
- (19) Generál kromě toho připravuje nařízení, podle něhož se na něj budou moci obrátit všichni, kteří se domnívají, že se jim děje bezpráví.
- (21) I proto se Nekvasil zabývá jednotlivými případy mladých důstojníků, kteří se rozhodli odejít do zálohy.
- (22) Šéf generálního štábu považuje lustrace za uzavřenou záležitost.
- (26) Jestliže by při jejich penzionování porušila zákon, soud by podle Nekvasila určitě prohrála.



Annotation

* manual pre-annotation for nodes with the same t_lemma

- (1) Generál Jiří Nekvasil: V české armádě se hrozné věci nedějí
- (7) Náčelník generálního štábu Jiří Nekvasil nám k tomu řekl, že nepořádek v armádě nikdy nezastíral, popírá však, že se v ní dějí hrozné věci.
- (8) Poměry v útvech podle něj odpovídají atmosféře ve společnosti.
- (9) Armáda jde krok za krokem k lepšímu, říká Nekvasil.
- (10) Za nepořádky náčelník generálního štábu považuje nízkou kázeň vojáků na veřejnosti, při výcviku nebo strážní službě.
- (11) Starosti mu dělá nedodržování bezpečnostních zásad a neoprávněná manipulace se zbraněmi.
- (14) Nekázeň se podle generála týká útvarů, jež transformace teprve zasáhne, zejména když mají být zrušeny.
- (17) O panujícím strachu v armádě Nekvasil neví.
- (18) Po mnoha letech podle něj dochází k narovnání vztahů mezi veliteli a podřízenými.
- (19) Generál kromě toho připravuje nařízení, podle něhož se na něj budou moci obrátit všichni, kteří se domnívají, že se jim děje bezpráví.
- (21) I proto se Nekvasil zabývá jednotlivými případy mladých důstojníků, kteří se rozhodli odejít do zálohy.
- (22) Šéf generálního štábu považuje lustrace za uzavřenou záležitost.
- (26) Jestliže by při jejich penzionování porušila zákon, soud by podle Nekvasila určitě prohrála.



Annotation

- (1) Generál Jiří **Nekvasil**: V české armádě se hrozné věci nedějí
- (7) Náčelník generálního štábu Jiří **Nekvasil** nám k tomu řekl, že nepořádek v armádě nikdy nezastíral, popírá však, že se v ní dějí hrozné věci.
- (8) Poměry v útvarech **podle něj** odpovídají atmosféře ve společnosti.
- (9) Armáda jde krok za krokem k lepšímu, říká **Nekvasil**.
- (10) Za nepořádky **náčelník generálního štábu** považuje nízkou kázeň vojáků na veřejnosti, při výcviku nebo strážní službě.
- (11) Starosti **mu** dělá nedodržování bezpečnostních zásad a neoprávněná manipulace se zbraněmi.
- (14) Nekázeň se podle **generála** týká útvarů, jež transformace teprve zasáhne, zejména když mají být zrušeny.
- (17) O panujícím strachu v armádě **Nekvasil** neví.
- (18) Po mnoha letech **podle něj** dochází k narovnání vztahů mezi veliteli a podřízenými.
- (19) **Generál** kromě toho připravuje nařízení, podle něhož se **na něj** budou moci obrátit všichni, kteří se domnívají, že se jim děje bezpráví.
- (21) I proto se **Nekvasil** zabývá jednotlivými případy mladých důstojníků, kteří se rozhodli odejít do zálohy.
- (22) **Šéf generálního štábu** považuje lustrace za uzavřenou záležitost.
- (26) Jestliže by při jejich penzionování porušila zákon, soud by podle **Nekvasila** určitě prohrála.

File Node Tree View Macros Setup Help

Mode: PML_T_Bridging

Style: PML_T_Bridging

New query Import Suggest Connect Configure Edit query Edit node Edit subtree Filters Cut Copy Paste Paste new tree (Un)Expand (Un)Expand all

Add node NOT AND OR Equality Regexp Name Type Relation Add rel Optional

Generál Jiří Nekvasil: [V české armádě](#) se hrozně věci nedějí

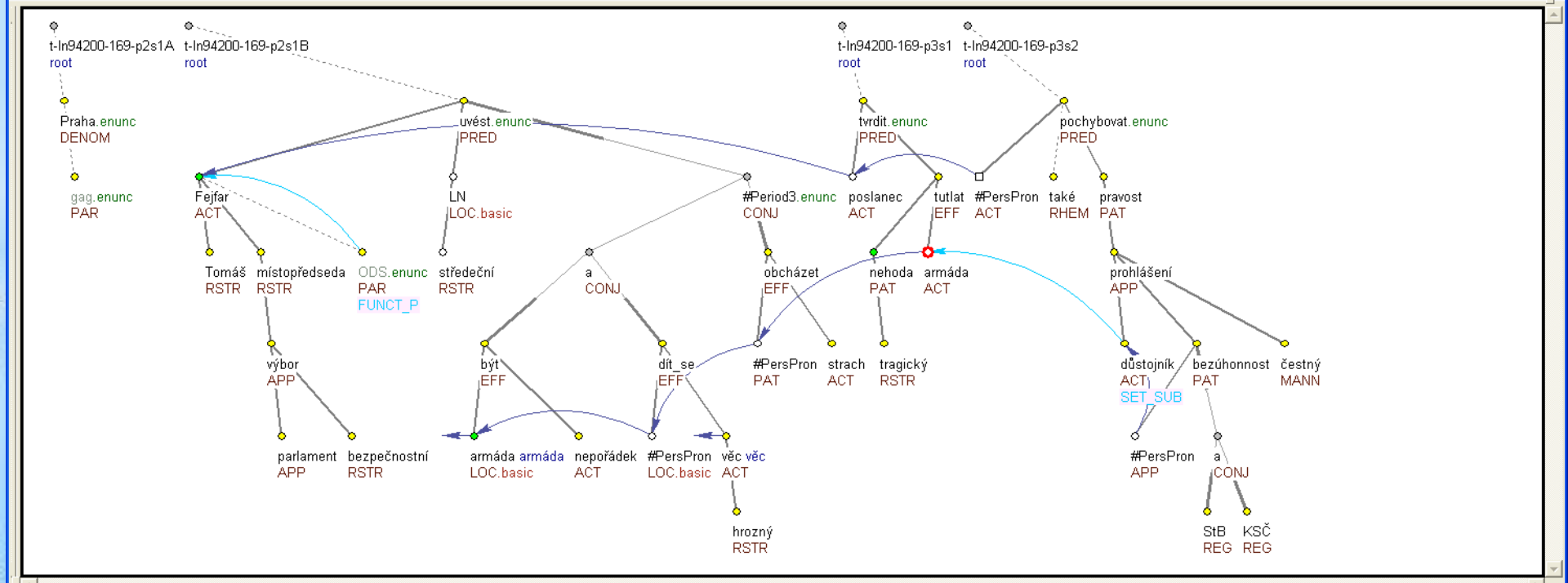
Praha (gag) -

[V armádě](#) je nepořádek a dějí se v ní hrozně věci... obchází ji strach, uvedl ve středečních LN místopředseda bezpečnostního výboru parlamentu Tomáš Fejfar (ODS).

--> Poslanec tvrdí, že tragické nehody [armáda](#) tutlá.
Pochybuje také o pravosti čestných prohlášení [důstojníků](#) o jejich bezúhonnosti ohledně StB a KSČ.
Tito lidé podle něj obklíčili [armádu](#) a vyřizují si při snižování počtů účty s bezúhonnými veliteli.

Náčelník generálního štábu Jiří Nekvasil nám k tomu řekl, že nepořádek [v armádě](#) nikdy nezastíral, popírá však, že se v ní dějí hrozně věci.
Poměry [v útvech podle něj](#) odpovídají atmosféře ve společnosti.
[Armáda](#) jde krok za krokem k lepšímu, říká Nekvasil.
Za nepořádky náčelník generálního štábu považuje nízkou kázeň [vojáků](#) na veřejnosti, při výcviku nebo strážní službě.
Starostí mu dělá nedodržování bezpečnostních zásad a neoprávněná manipulace se zbraněmi.
Jen letos zahynulo k 11. červenci ve službě 6 základáků (většinou ve strážní službě) a 1 profesionál.
V době volna či dovolených 19 základáků (zpravidla dopravní nehody) a 8 profesionálů (sebevraždy).
Nekázeň se podle generála týká útvarů, jež transformace teprve zasáhne, zejména když mají být zrušeny.
Nejistota z budoucnosti prý snižuje svědomitost velitelů.
Jednotky, které se již stabilizovaly, odvádějí naopak dobré výsledky.

O panujícím strachu [v armádě](#) Nekvasil neví.
Po mnoha letech podle něj dochází k narovnání vztahů [mezi veliteli a podřízenými](#).
Generál kromě toho připravuje nařízení, podle něhož se na něj budou moci obrátit všichni, kteří se domnívají, že se jim děje bezpráví.

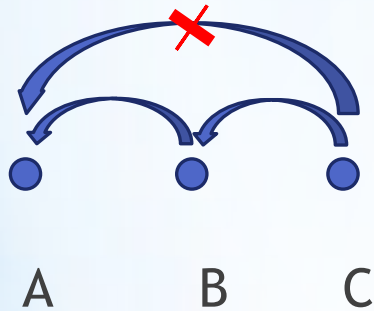




Annotation

* finding the nearest antecedent

automatic redirection of a newly created coreferential arrow to the nearest antecedent

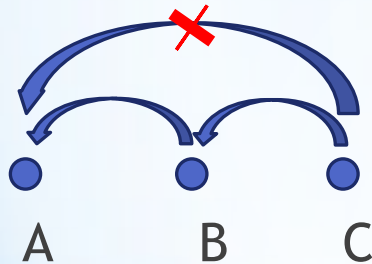




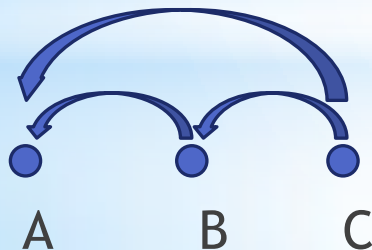
Annotation

* finding the nearest antecedent

automatic redirection of a newly created coreferential arrow to the nearest antecedent



* automatic redirection during annotation





Annotation

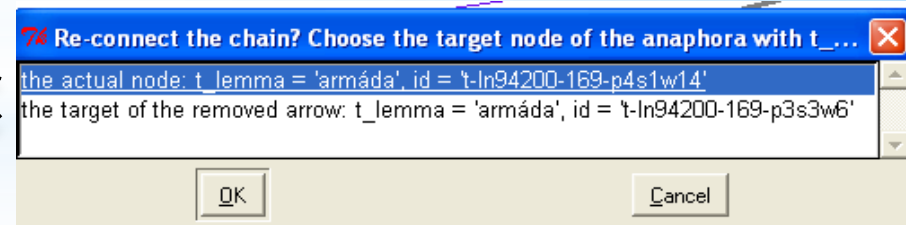
*preserving the coreferential chain

annotator removes an arrow and interrupts the chain :

the tool can reconnect the chain or leave in interrupted



A B C D



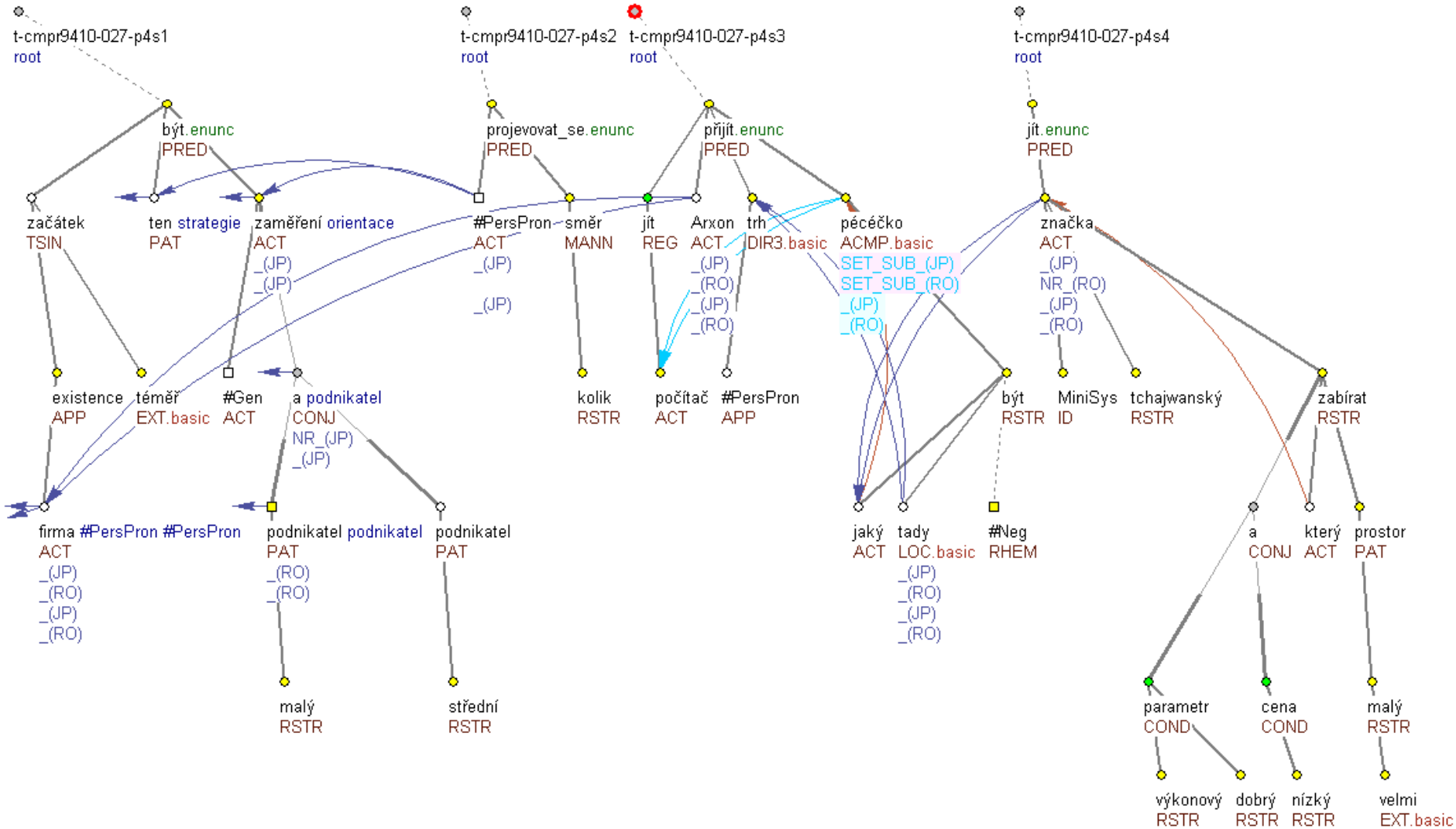


Comparing different annotations

- * visual comparison of different annotations of the same data, e.g. annotations from different annotators in the inter-coder agreement measurement



Comparing different annotations





HUMAN CAPITAL
NATIONAL COHESION STRATEGY

Zeros

EUROPEAN UNION
EUROPEAN
SOCIAL FUND





Zeros with coreference

t-lemma	description	example
#PersPron	assigned to a node representing personal or possessive pronouns; applies both to newly established nodes and to those present at the surface level	<i>(Firma měla doručit zboží zákazníkovi.) Doručení {#PersPron.ACT} {#PersPron.PAT} {#PersPron.ADDR} se však neuskutečnilo. (=lit. (The_company was_to deliver the_goods to_the_customer.) The_delivery {#PersPron.ACT} {#PersPron.PAT} {#PersPron.ADDR} however did_not_happen.)</i>
#Cor	assigned to a newly established node representing the (usually inexpressible) controllee in control constructions.	<i>Helen decided to {#Cor.ACT} answer him.</i>
#QCor	lemma assigned to newly established nodes representing a (usually inexpressible) valency modification in quasi-control constructions.	<i>He offered Jan {#QCor} protection</i>
#Rcp	assigned to newly established nodes representing participants that are left out as a result of reciprocation.	<i>The lovers kissed {#Rcp.PAT}.</i>
inserted full lemma	In case of textual ellipsis, a copy of the node representing the same lexical unit as the omitted element is inserted into the appropriate position.	<i>Clients of insurance companies which shut down will automatically return to the General {insurance company}.</i>



Zeros without coreference

t-lemma	description	example
#Gen	assigned to a newly established node representing a general participant absent at the surface level	<i>Houses are built {#Gen.ACT} from bricks.</i>
#Unsp	assigned to newly established nodes representing valency modifications not present at the surface level the semantic content of which is very vague (non-specific)	<i>U Nováků {#Unsp.ACT} dobře vaří. (=They cook well at Nováks'.)</i>
#EmpNoun	assigned to newly established nodes representing non-expressed nouns governing syntactic adjectives, not the case of textual ellipsis	<i>Přišli jen {#EmpNoun.ACT} mladší. (=lit. Came only {#EmpNoun} younger.)</i>



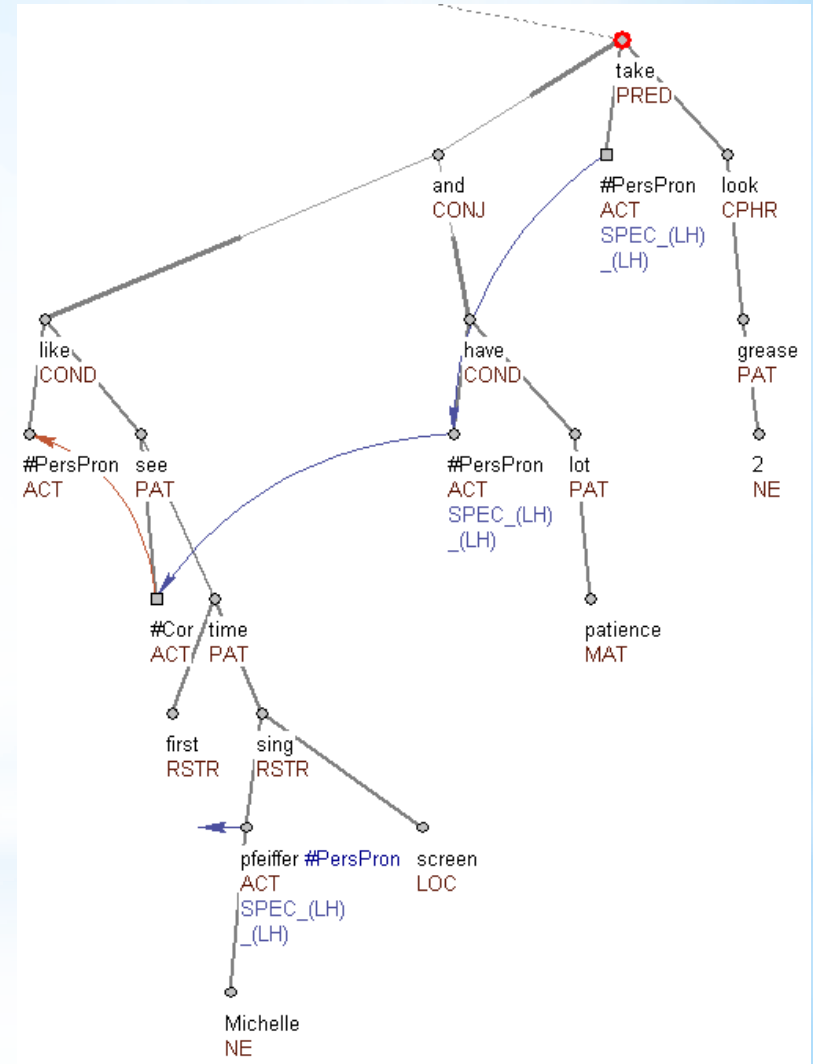
Zeros without coreference

t-lemma	description	example
#Oblfm	assigned to a newly established node representing an obligatory adjunct absent at the surface level.	Ta vypadá. {#Oblfm.MANN} (=lit. That.fem looks; meaning: She looks awful/so strange...)
#Equal	assigned to newly established nodes used in comparative constructions	Udělal to jako Tonda. (=lit. (He) did it like Tonda)
#Some	lemma assigned to newly established nodes representing the nominal part of verbonominal predicates (not present at the surface level), used mainly in comparative constructions	Není ovšem ford jako ford. - lit. However, Ford is not like Ford (meaning Not all Fords are the same)



#PersPron, #Cor

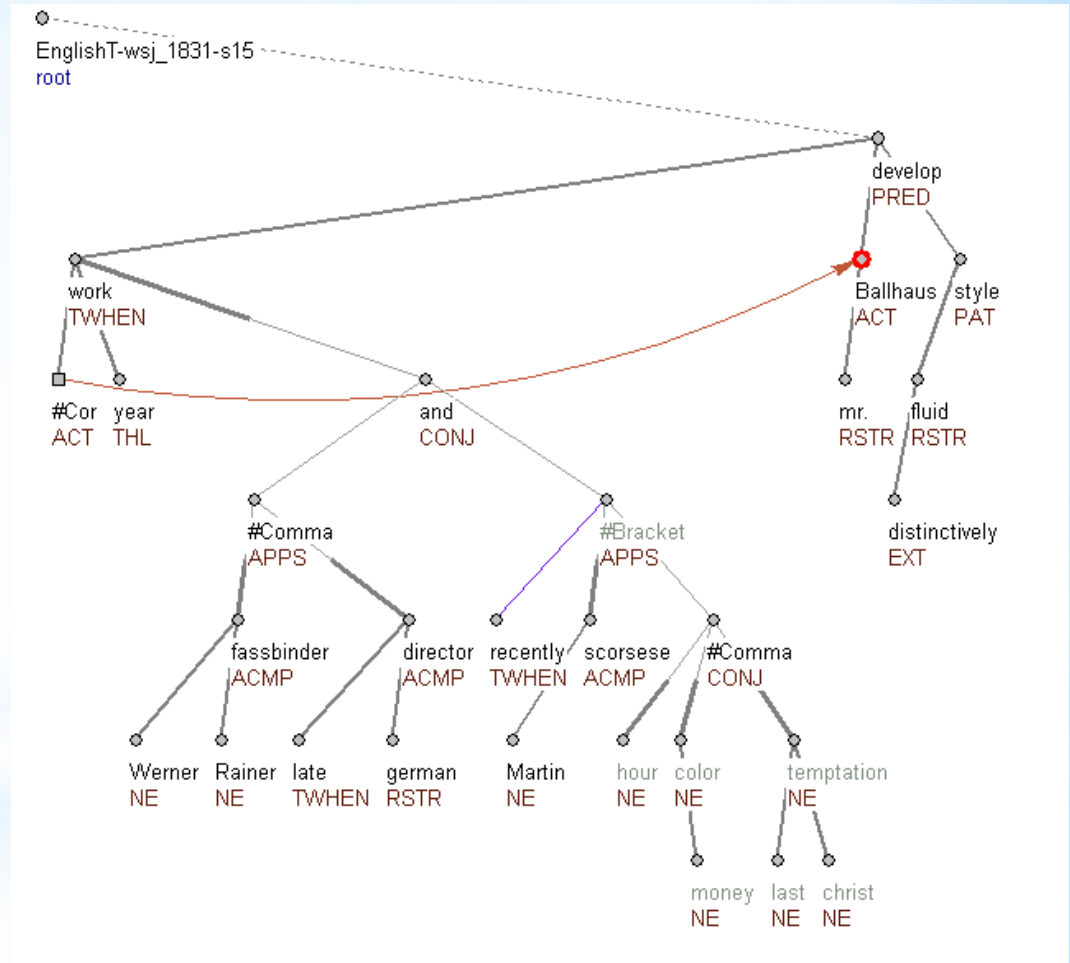
If you'd like {#Cor.ACT} to see the first time Michelle Pfeiffer sang on screen, and you have a lot of patience, #PersPron take a look at “Grease 2”.





#Cor

After {#Cor.ACT} working for years with Werner Rainer Fassbinder, the late German director, and more recently with Martin Scorsese "After Hours", "The Color of Money", "The Last Temptation of Christ", Mr. Ballhaus has developed a distinctively fluid style.



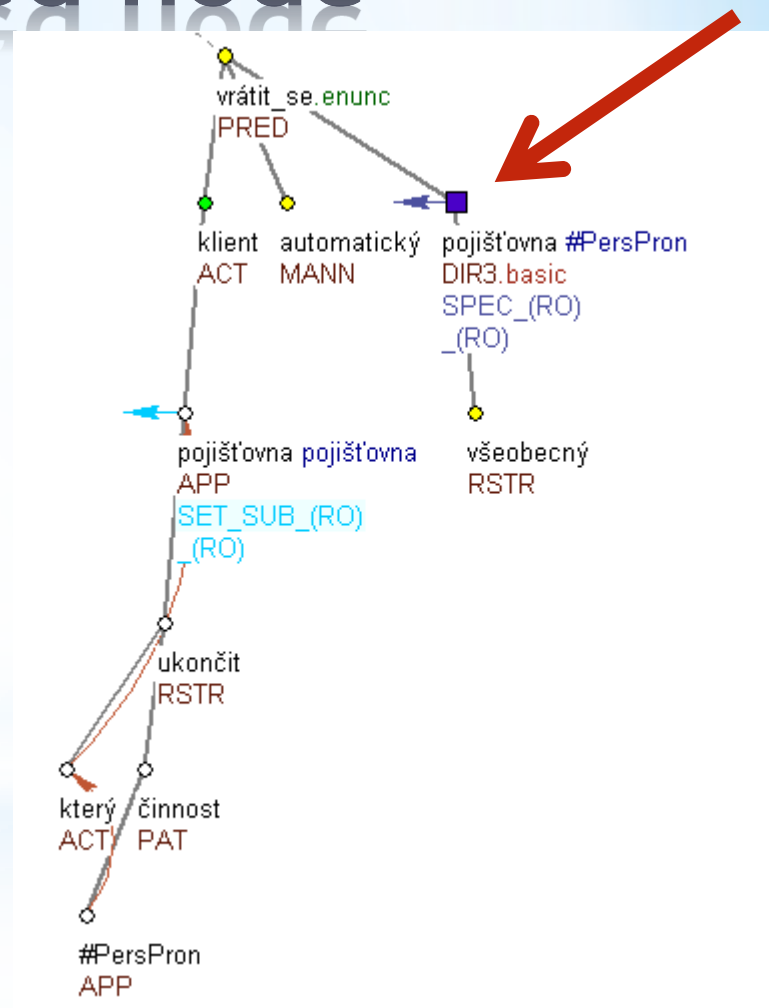
Reconstruction of the omitted node



If it is clear (and possible to find in the text) which noun has been omitted in the surface structure of the sentence (the case of textual ellipsis), a copy of the node representing the same lexical unit as the omitted element is inserted into the appropriate position.

cz. *Klienti pojišťoven, které ukončí svou činnost, se automaticky vrátí k Všeobecné.*

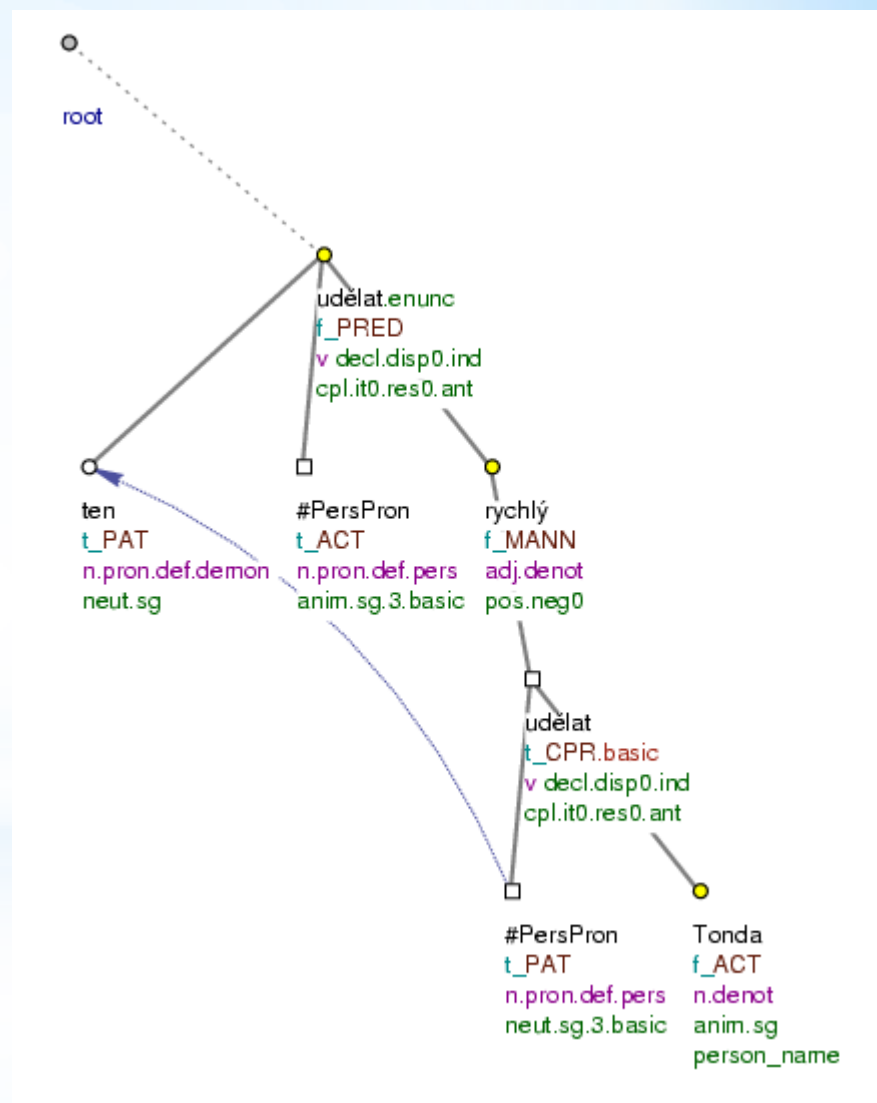
lit. *Clients of insurance companies which shut down will automatically return to the General {one}.*





Reconstruction of a syntactic construction

*Udělal to rychle jako Tonda.
(=lit. (He) did it fast like
Tonda)*





HUMAN CAPITAL
NATIONAL COHESION STRATEGY

EUROPEAN UNION
EUROPEAN
SOCIAL FUND



Contextual boundness (TFA)



Contextual boundness (TFA)

- * Topic-focus articulation (information structure): reflects the cognitively based ,given‘ - ,new‘ strategy BUT belongs to the systems of individual languages rather than to the domain of cognition
- * the description of TFA is an integral part of the representation of the (literal) meaning of the sentence (Sgall et al. 1986) = **TFA is semantically relevant**
- * a declarative sentence asserts that its **Focus** holds about its **Topic: F(T)**



Contextual boundness (TFA)

a specific TFA attribute

values:

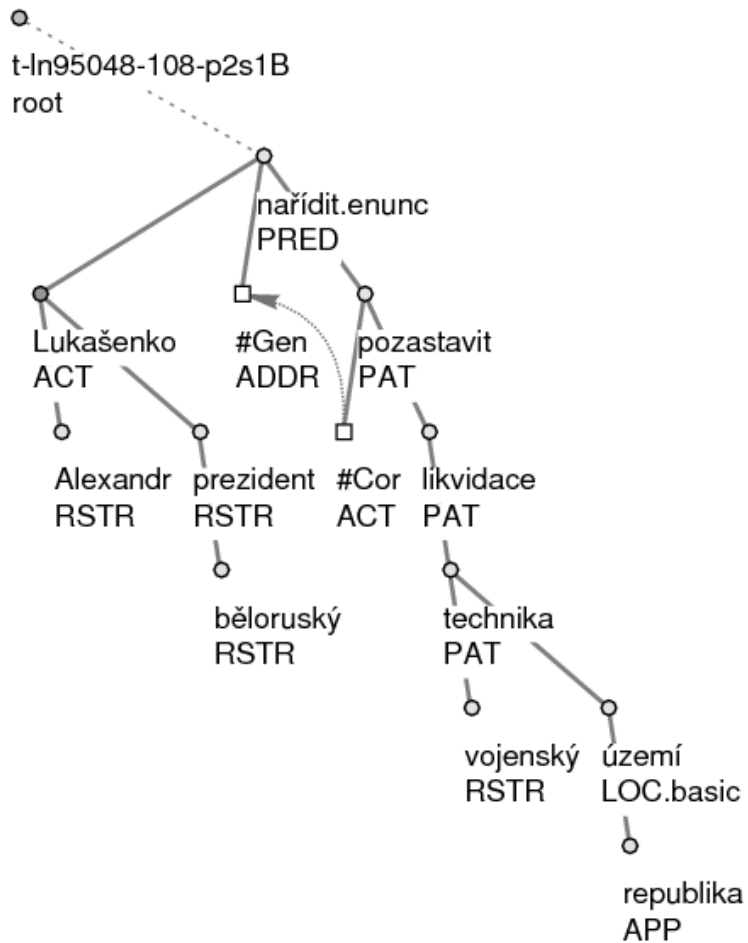
t for a non-contrastive **contextually bound node**

c for a **contrastive contextually bound node**

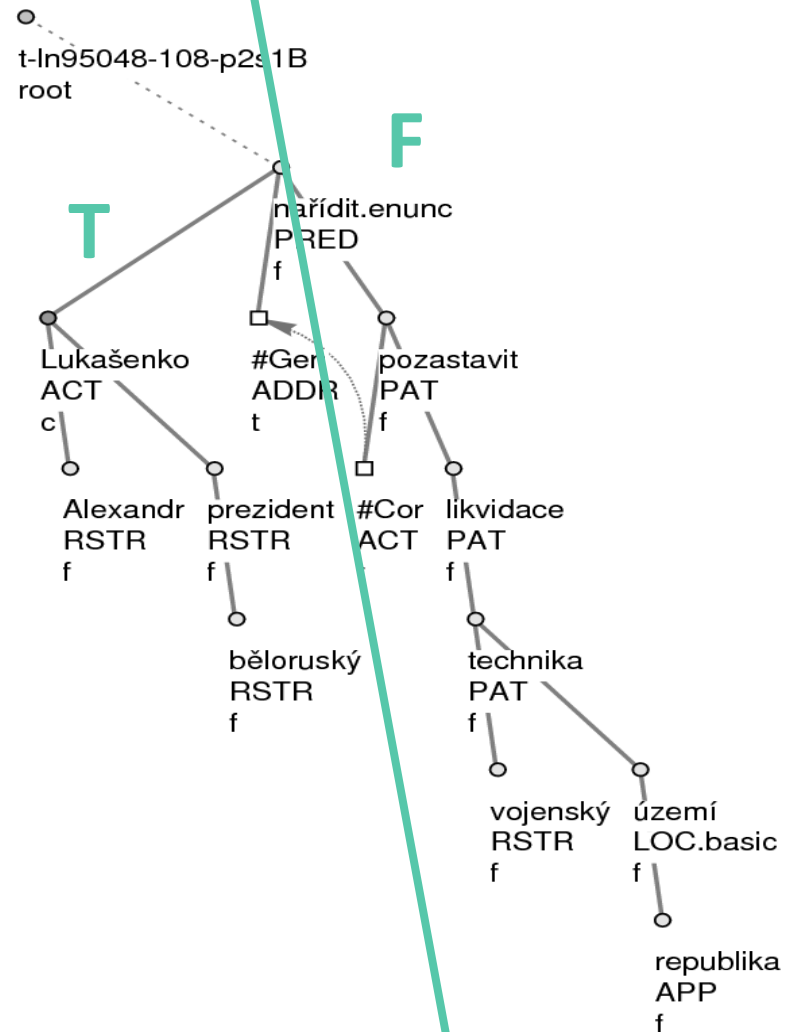
f for a contextually **non-bound node**

Cz.: Běloruský president Alexandr Lukašenko nařídil pozastavit likvidaci vojenské techniky na území republiky.

Engl.: (lit.) Belorussian president Alexandr Lukašenko ordered to stop liquidation (of) military technique on territory of republic.



Běloruský prezident Alexandr Lukašenko nařídil pozastavit likvidaci vojenské techniky na území republiky.



Běloruský prezident Alexandr Lukašenko nařídil pozastavit likvidaci vojenské techniky na území republiky.



TFA -present programme

- * Annotation of TFA in an English parallel text based on the same theory
- * => preparation of a manual for such an annotation
- * Comparison of results
- * → material for contrastive studies on TFA



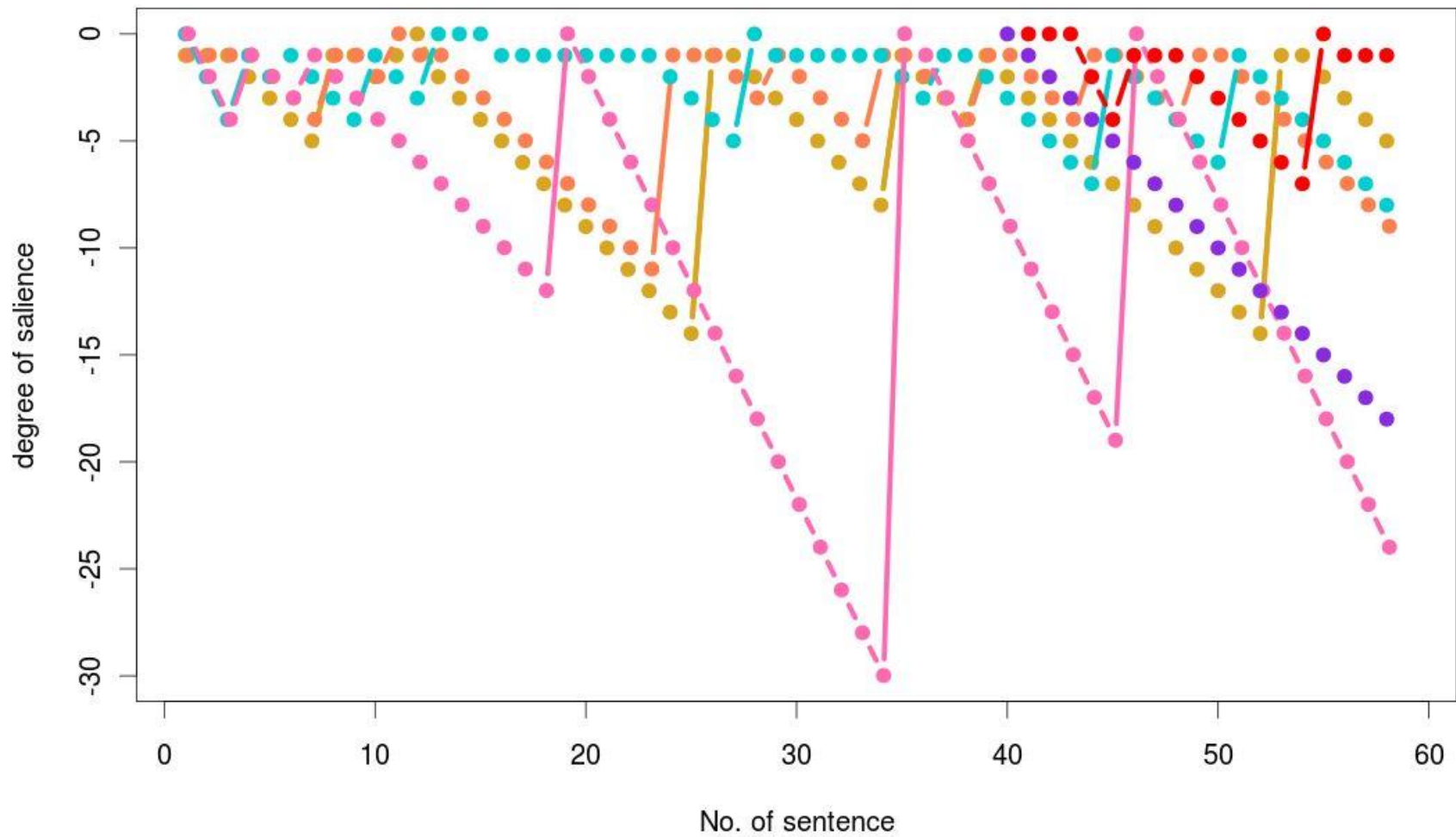
TFA salience -tentative “rules”

- (i) if r is expressed by a weak pronoun (or zero, i.e. deleted in the surface shape) in a sentence, it retains its salience degree after this sentence is uttered: $dg_i(r) := dg_{i-1}(r)$;
- (ii) if r is expressed by a noun (group) carrying nb , then $dg_i(r) = 0$;
- (iii) if r is expressed by a noun (group) carrying cb , then $dg_i(r) = 1$;
- (iv) $dg_i(q) := dg_i(r) + 2$ obtains for every referent q that is not itself referred to in S_i , but is immediately associated with an item present here;
- (v) if r neither is included in S_i , nor refers to an associated object, and has been mentioned in the focus of the preceding sentence, then $dg_i(r) := dg_{i-1}(r) + 2$.
- (vi) if r neither is included in S_i , nor refers to an associated object, and has been mentioned in the topic of the preceding sentence, then $dg_i(r) := dg_{i-1}(r) + 1$.



Salience -Text

1. Across the river **they** could now see a fire with **two figures** beside it. 2. When they moved closer, 3. they could make out two white horses against the background of the dark bushes. 4. Then he recognized them. 5. The pale blue **buggy**. 6. Two hours ago, **the beauty** from Chicago had sat on the seat. 7. While the **black man** in livery had gone into Kapino's for beer. 8. They stopped ... 9. and looked across the river. 10. The young lady in the white dress was biting into a chicken leg. ... 11. **He** looked at **Magda**. 12. The child's eyes, wide in amazement, stared across the river at this fairytale banquet. ... 13. He looked at the straw hat. 14. Yes, beside it in the grass a pair of white shoes had been casually tossed 15. and beside them lay a crumpled white pile. ... 31. and ... she had lifted the skirt over her head, 32. slipped out of it 33. and stood there in nothing but white knee-length knickers ... 34. He couldn't take his eyes off her. ... 35. From downstream they could hear a banjo playing. 36. A pleasant baritone voice sang: „...“ 37. The girl let her hands drop... 38. Cautiously, she stepped into the water. 39. On their side of the river, ..., something creaked. 40. Looking towards the sound, he could barely distinguish the outline of a small **rowboat** 41. and, in it, **someone's dark silhouette**. 42. The moonlight fell on the head, the white whiskers, the hair in disarray. 43. The Master! 44. He looked quickly across the stream 45. and saw the Rusalka up to her waist in the water. „Borne like a vapour ...“ 46. The Master's head turned in profile towards the velvet baritone. 47. He doesn't see; 48. he only hears, 49. he thought. 50. He himself saw. ... 51. The Rusalka was slowly lowering herself into the water, ... 52. Finally, all that remained on the water was a burning waterlily. 53. Suddenly the child saw too 54. and shrieked, 55. „Papa!“ 56. The Master started, 57. looked around 58. and then saw.





HUMAN CAPITAL
NATIONAL COHESION STRATEGY

EUROPEAN UNION
EUROPEAN
SOCIAL FUND



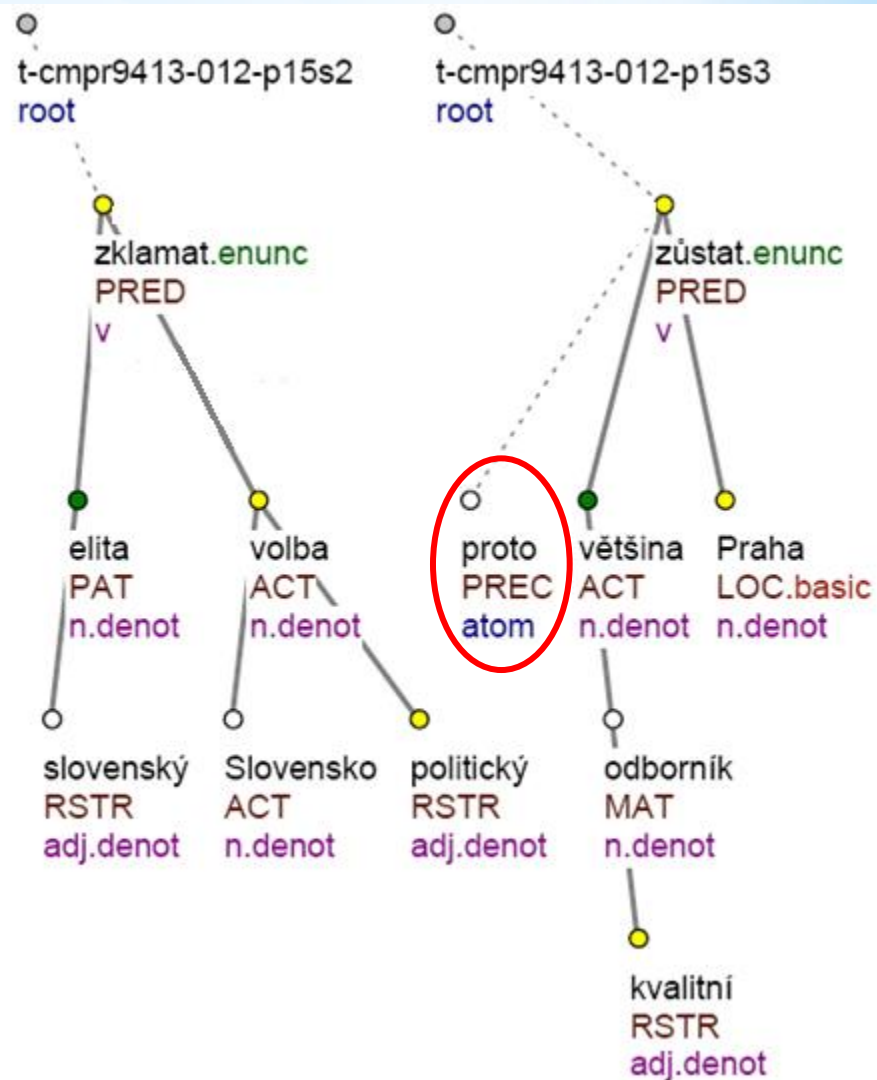
Discourse relations - Connectives, arguments, Annotation on Trees

*** *"Slovak elite was disappointed by the political choice of Slovakia. Therefore, the majority of quality specialists remained in Prague."***

* ***"Slovak elite was disappointed by the political choice of Slovakia.
Therefore, the majority of quality specialists remained in Prague."***

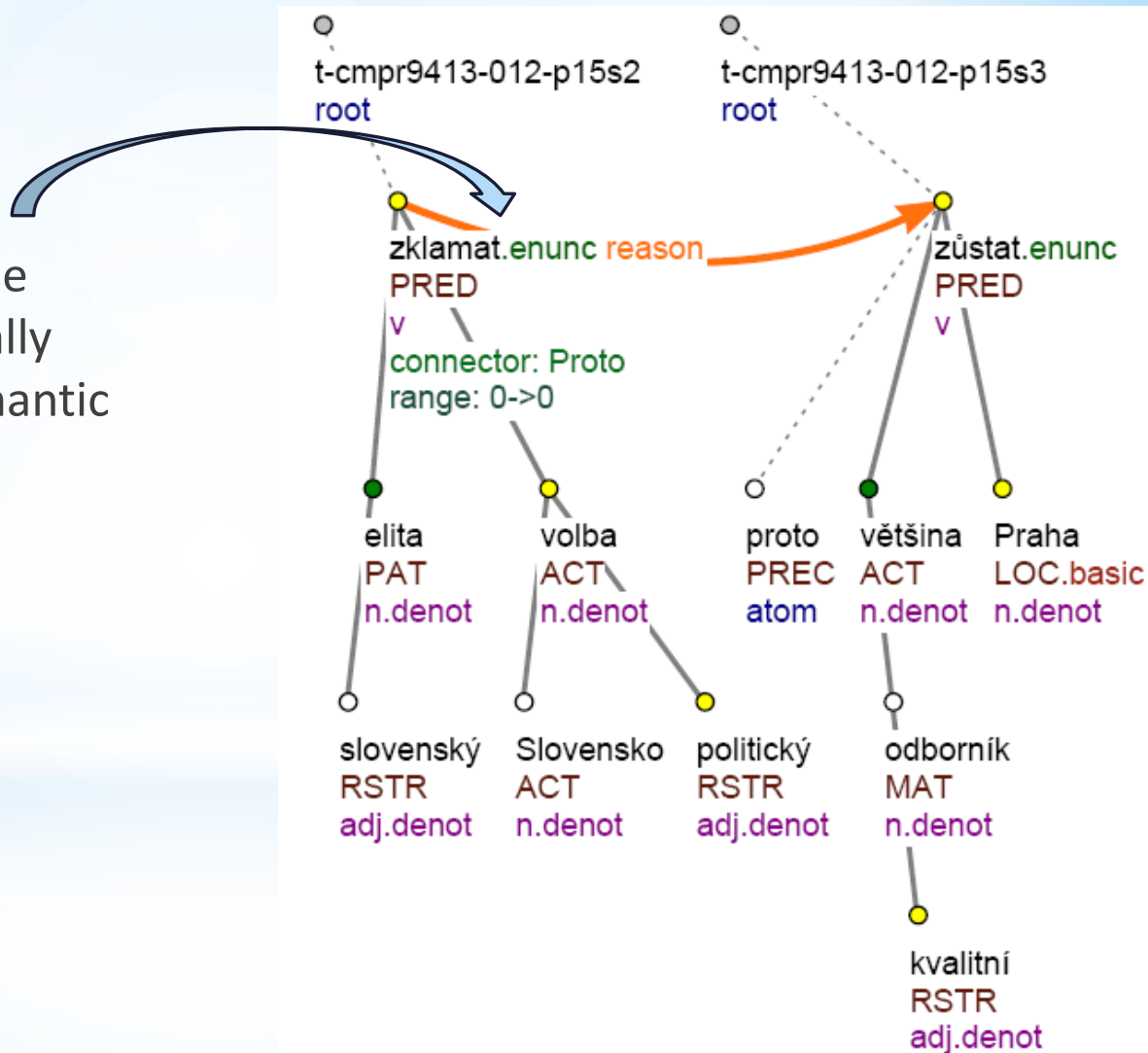
* *"Slovak elite was disappointed by the political choice of Slovakia.
Therefore, the majority of quality specialists remained in Prague."*

* *"Slovak elite was disappointed by the political choice of Slovakia. Therefore, the majority of quality specialists remained in Prague."*



* *"Slovak elite was disappointed by the political choice of Slovakia.
 Therefore, the majority of quality specialists remained in Prague."*

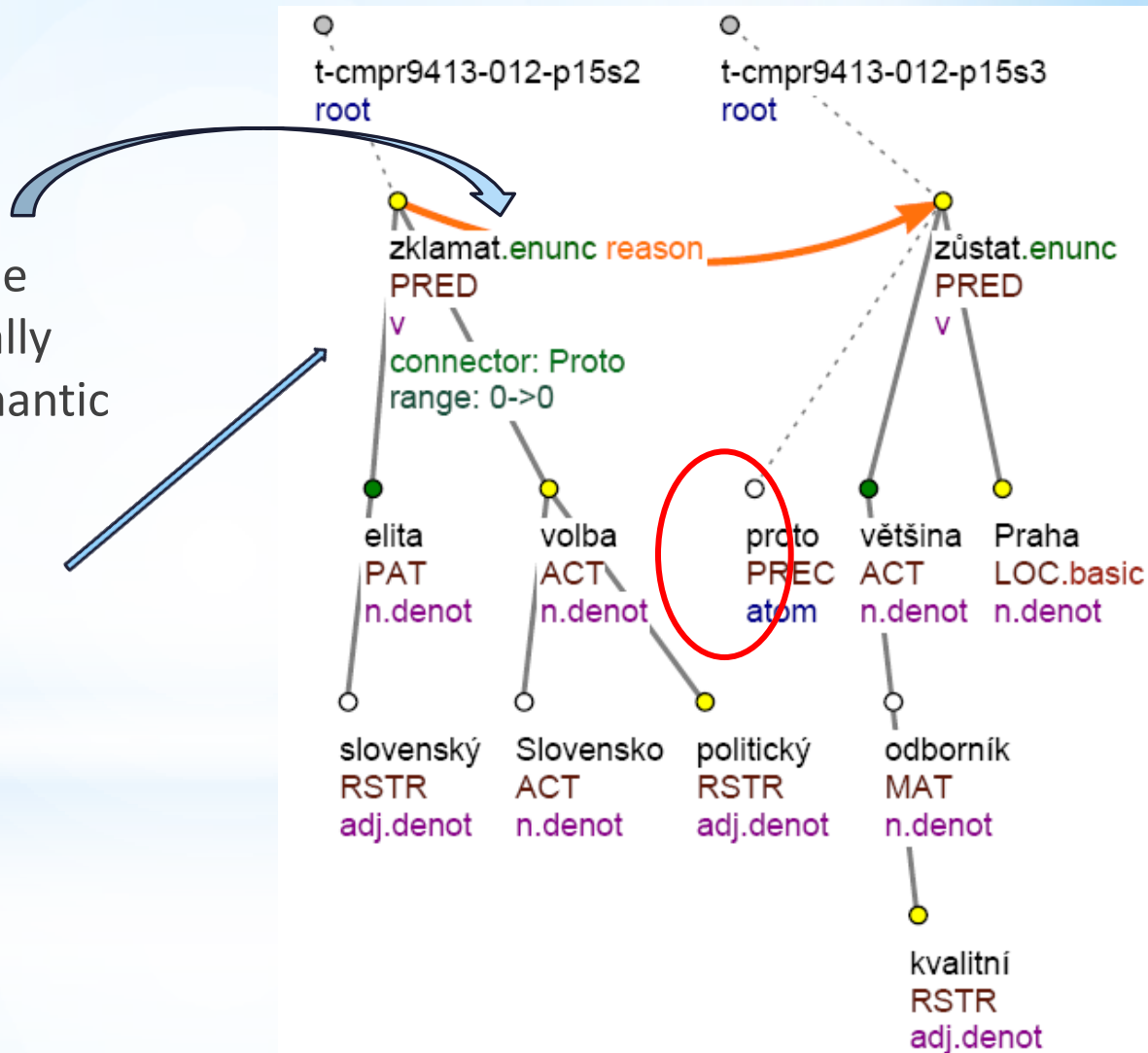
* linking of the arguments of the relation by arrow (automatically connected with choice of semantic interpretation)



* *"Slovak elite was disappointed by the political choice of Slovakia. Therefore, the majority of quality specialists remained in Prague."*

* linking of the arguments of the relation by arrow (automatically connected with choice of semantic interpretation)

* adding of the connective

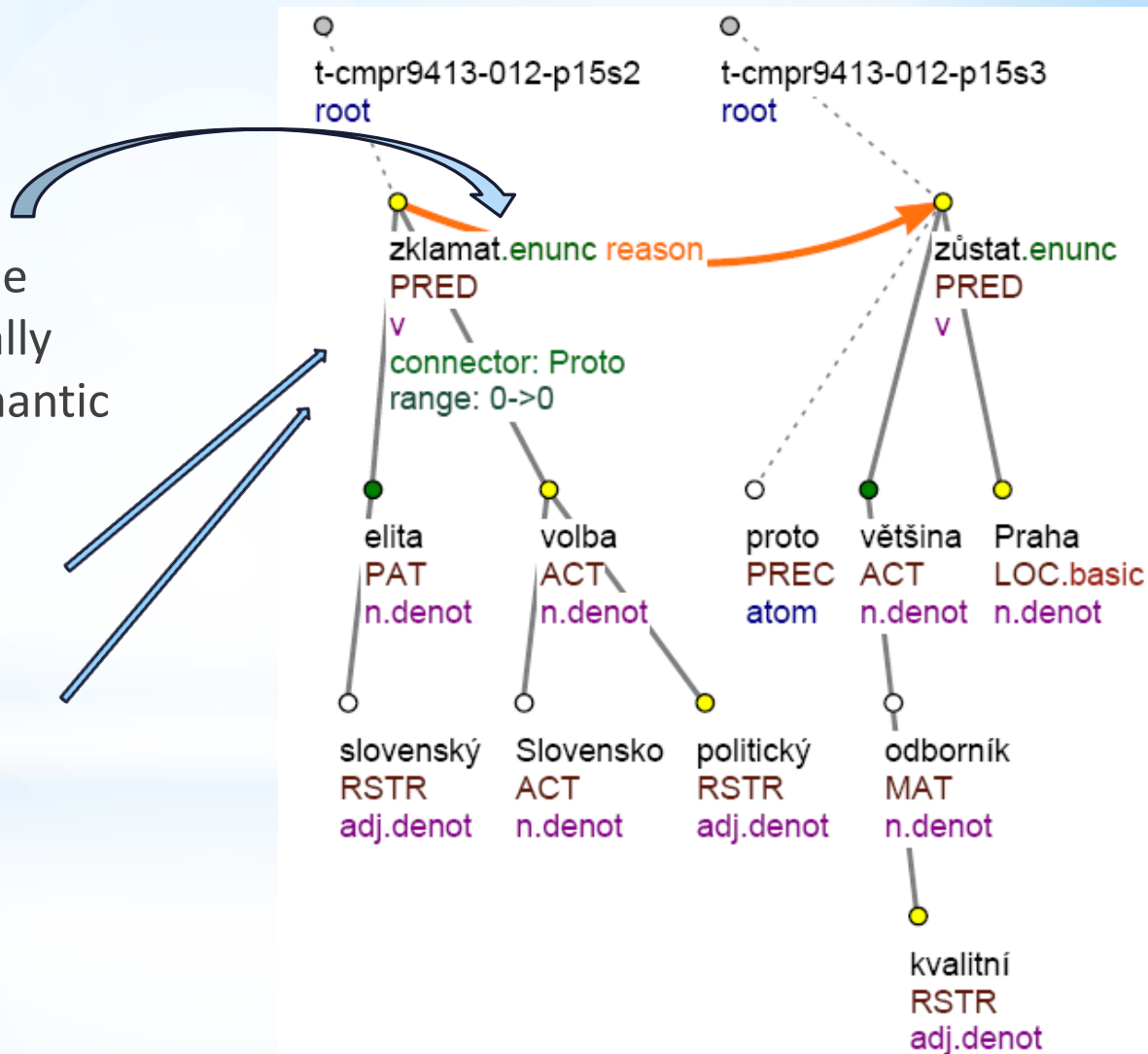


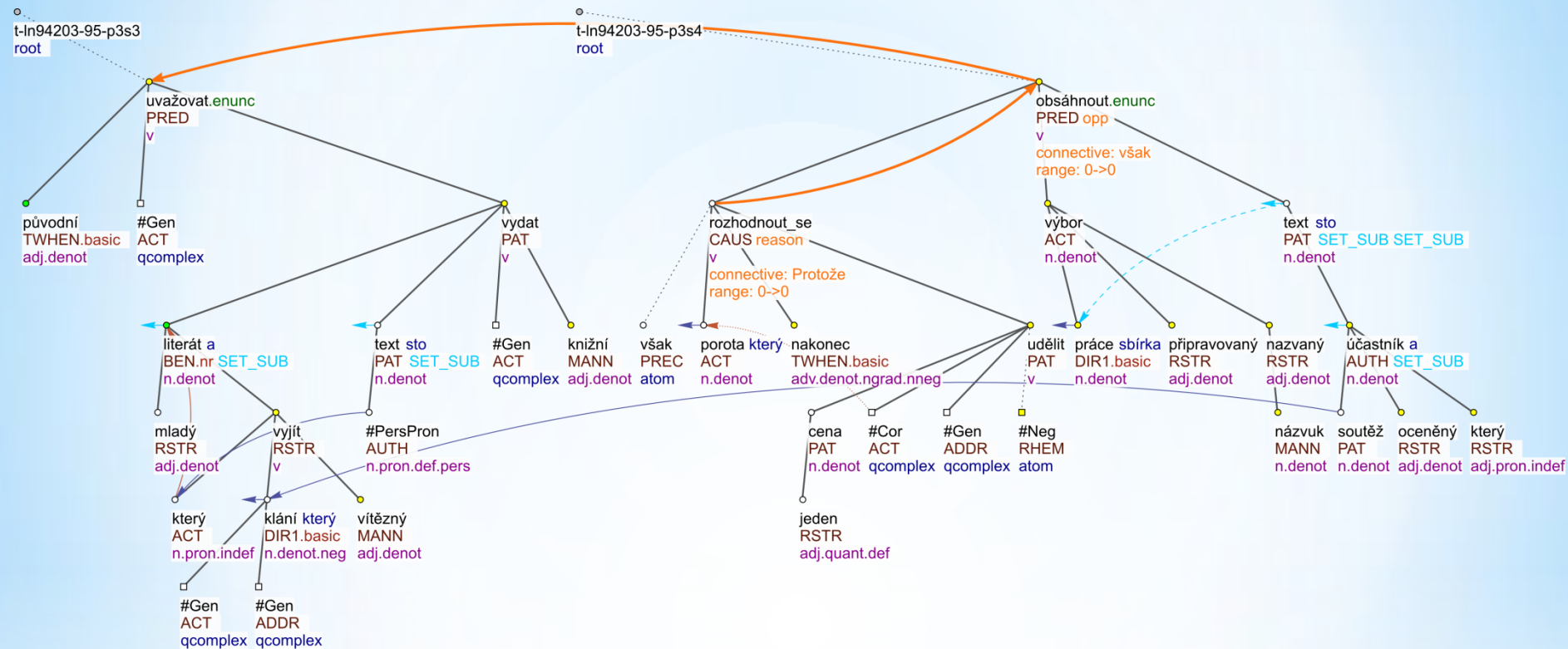
* *"Slovak elite was disappointed by the political choice of Slovakia.
 Therefore, the majority of quality specialists remained in Prague."*

* linking of the arguments of the relation by arrow (automatically connected with choice of semantic interpretation)

* adding of the connective

* marking of the span of the arguments





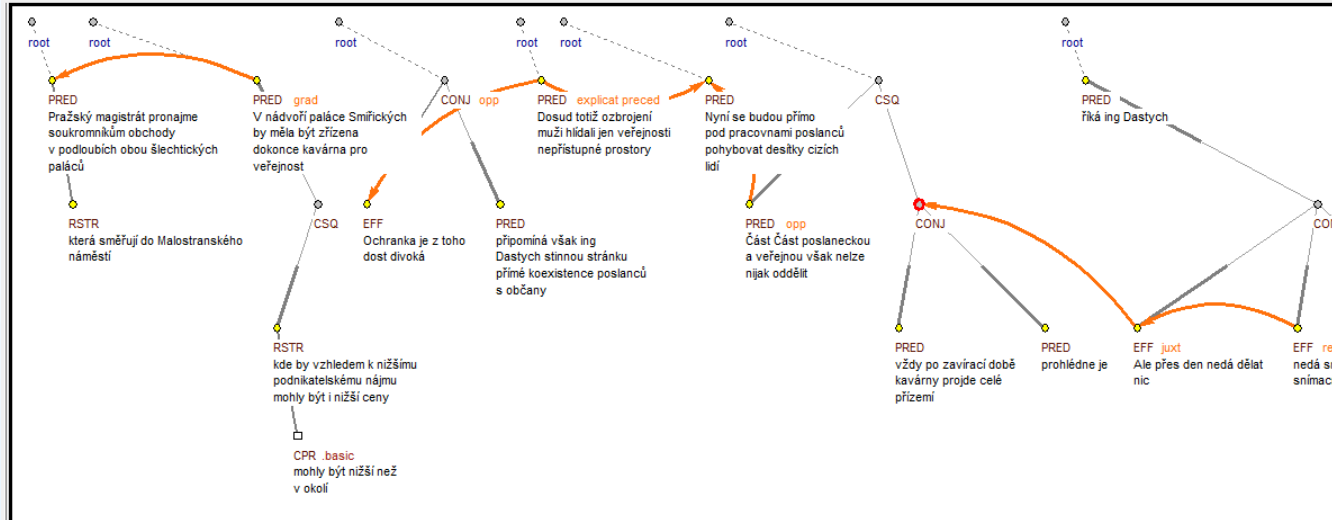
Původně se uvažovalo o tom, že mladému literátovi, jenž by vyšel z klání vítězně, budou jeho texty vydány knižně. Protože se však porota rozhodla nakonec první cenu neudělit, obsáhne připravovaný výbor z prací nazvaný Názvuky texty všech oceněných účastníků soutěže.

Initially, the book of the winner was supposed to be published. However, as the jury decided not to give the first price, the collected works of all authors will be published.



Ochranka je z toho dost divoká, připomíná však ing. Dastych stinnou stránku přímé koexistence poslanců s občany.
 Dosud totiž ozbrojení muži hlídali jen veřejnosti nepřístupné prostory.
 Nyní se budou přímo pod pracovními poslanců pohybovat desítky cizích lidí.
 -> Část poslaneckou a veřejnou však nelze nijak oddělit, proto vždy po zavírací době kavárny projde ochranka se psem celé přízemí a prohlédne je.
 Ale přes den se nedá dělat nic, snad jen instalovat snímáči kamery a služba je bude muset nějak ohlídat, říká ing. Dastych.

Tunel, nebo most?



* Discourse Representation - Compact View

Discourse Semantic Labels

TEMPORAL	CONTINGENCY	COMPARISON (CONTRAST)	EXPANSION
precedence - succesion	reason - result	confrontation	conjunction
synchronous	pragmatic reason - result	opposition	instantiation
	purpose	pragmatic contrast	specification
	explication	restrictive opposition + exception	equivalence
	condition	concession	generalization
	pragmatic condition	correction (replacement)	conjunctive alternative
		gradation	disjunctive alternative



Additional Information to Connective Annotation

Also annotated:

- * List structures - quite typical way of text composition
- * Headings
- * Alternative lexical expressions of the connectives, for example:
Z toho vyplývá... = tedy, takže...
(*This implies... = so, therefore...*)
- * Double-sense relations
- * The so-called collections

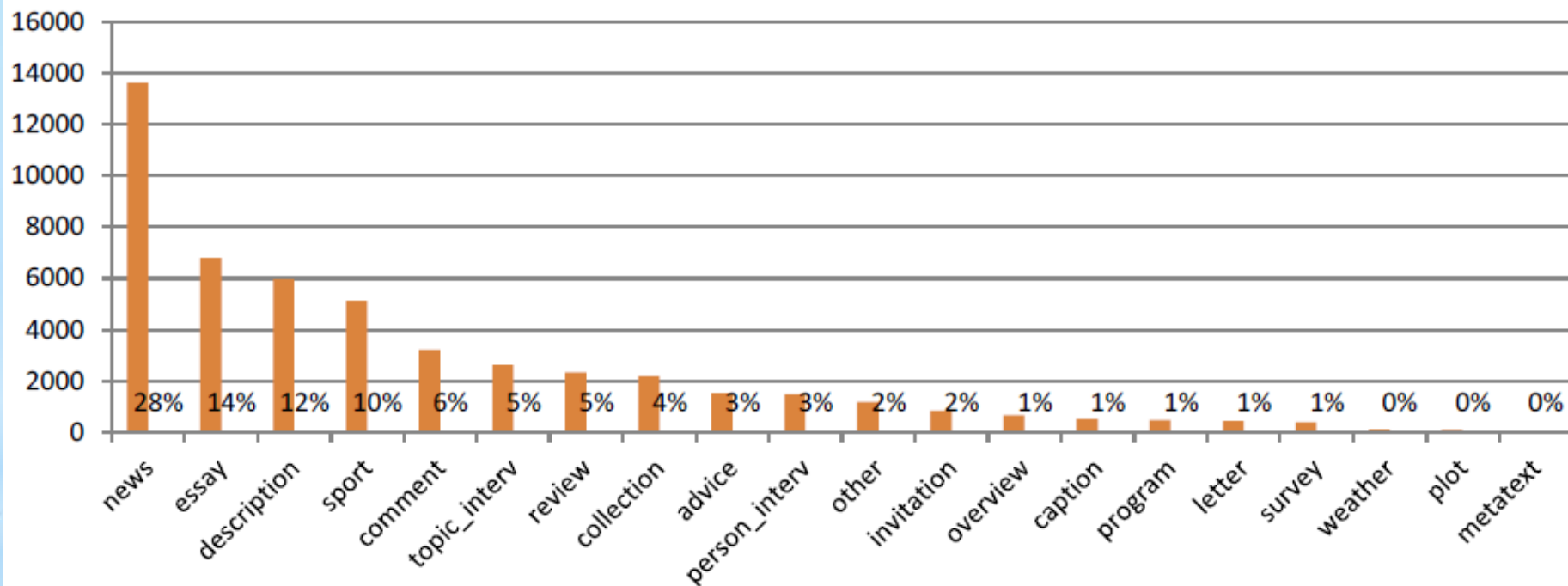
Not annotated so far:

- * Implicit connectives and their arguments and senses (in comparison to Penn Discourse TreeBank 2.0)



Genre classification

* manually annotated (Poláková et al. 2014, LREC)





Linguistic research based on the annotated data

* **coreference:**

- * coreference of generic NPs (Nedoluzhko 2013)
- * translation of *it* in Czech using dependency structure (Novák - Nedoluzhko 2013)
- * comparison of coreferential expressions in Czech and English (Novák - Nedoluzhko 2014)
- * premodifiers

* **discourse:**

- * problematic connectors: Explication and Cause (Rysová 2011), cause (Jínová 2011)
- * Sentence Structure and Discourse Structure: Possible Parallels (Jínová et al. 2011)

* **coreference + discourse:**

- * problem of contrast in tfa, discourse and coreference annotation (Zikánová 2011)
- * Interplay of Coreference and Discourse Relations (Poláková et al. 2012)
- * tectogrammatcs and discourse (Mírovský et al. 2012), tectogrammatcs and coreference (Nedoluzhko - Mírovský 2013)

* **coreference + discourse + TFA** (salience, etc.) - running grant project

- * TFA in Czech, English and German



* Acknowledgements

- * The presentation of the results is co-financed by the European Union from resources of the European Social Fund
- * The research was supported from the Grant Agency of the Czech Republic (grant P406/12/0658 Coreference, discourse relations and information structure in a contrastive perspective), GAUK 3389/2015, EU (grant FP7-ICT-2013-10-610516 – QTLeap) and SVV project number 260 224. This work has been using language resources developed, stored, and distributed by the LINDAT/CLARIN project of the Ministry of Education, Youth and Sports of the Czech Republic (project LM2010013).
- * I gratefully acknowledge my colleagues Eva Hajičová, Šárka Zikánová, Lucie Poláková, Jiří Mírovský, Kateřina Rysová et al.





HUMAN CAPITAL
NATIONAL COHESION STRATEGY

EUROPEAN UNION
EUROPEAN
SOCIAL FUND



Thank you!

<http://ufal.mff.cuni.cz/discourse>

