



T-Layer of PDT: Coreference

Markéta Lopatková

Institute of Formal and Applied Linguistics, MFF UK

lopatkova@ufal.mff.cuni.cz



Coreference in PDT: Outline

- basic terms
 - grammatical coreference
 - textual coreference
 - bridging anaphora
- } (later: TFA and discourse annotation)

documentation: <http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/t-layer/html/index.html>

Coreference in PDT: Basic Concepts



- reference

- ~ a relation of an language expression to a real world object or situation (referent)

- exophoric r.: referring to a situation or entities outside the text
 - endophoric r.: referring to another expression (within the same text) having the same referent (entity, situation)

⇒ **coreference**

- *coreferring* expression vs. *coreferred* expression:

- anaphoric reference: anaphor vs. antecedent

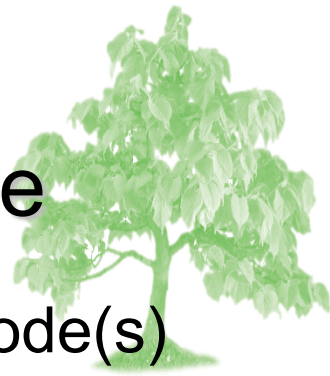
- cataphoric reference: **cataphor** vs. **postcedent**

Near him, John saw a snake.

- controlee vs. controller

- coherence / cohesion (cz návaznost)

Coreference in the Tectogrammatical Tree



- coreference ~ a link between a t-node to another t-node(s)
 - within a single sentence
 - crossing a sentence boundary

- coreferring node:

- the ID of the coreferred node (leaf / root of a subtree)

Vlasta , kterou dobře znáš, šla do divadla. Vlasta , whom you very well know, went to the theater.

- a list of ID(s) (typically in the same subtree) ... for textual coreference

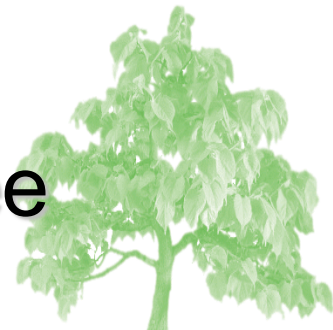
Marie vzala Vlastu do kavárny, tam na ně už čekal Marek.

Marie took Vlasta to the caffè, where Marek already waited for them .

- attributes for coreference

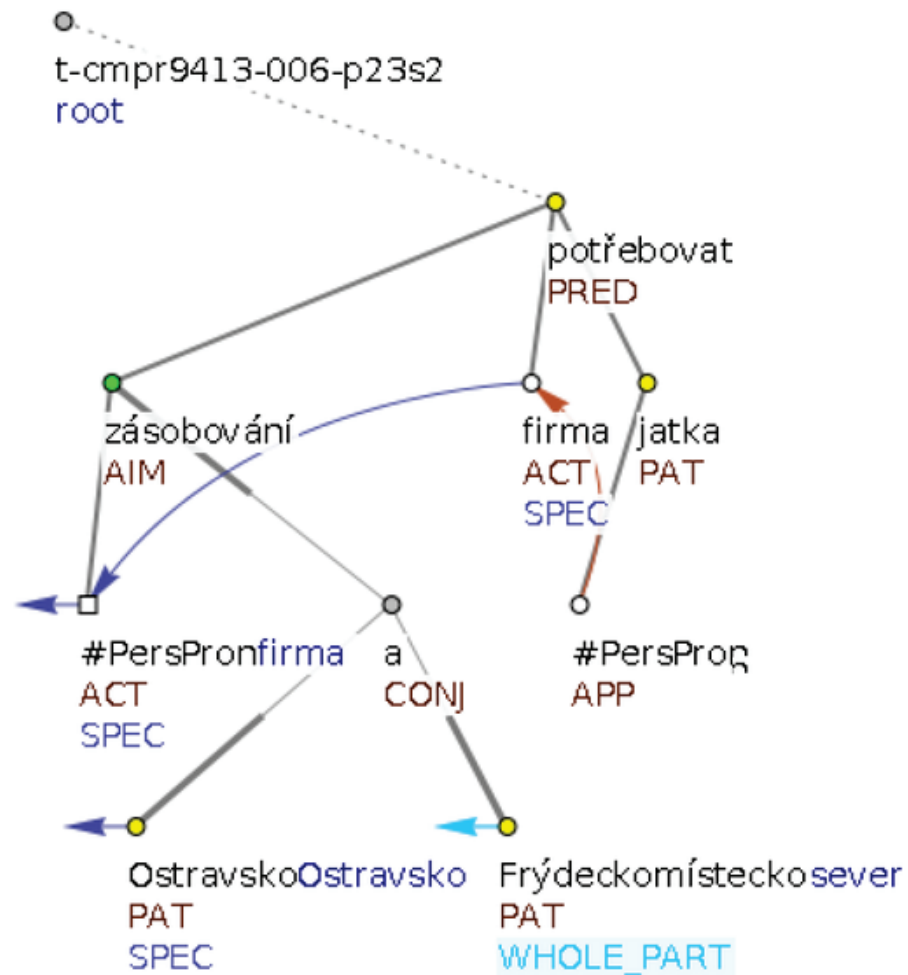
- **coref_gram.rf**
 - **compl.rf**
 - *coref_text*
 - *coref_special* ... special values (special type of a textual coreference, later)
 - *bridging*

Coreference in the Tectogrammatical Tree



brown ... grammatical
green ... compl
dark blue ... textual
light blue ... bridging

*Pro zásobování Ostravska a
Frýdeckomístecka potřebuje firma svá jatka.*
For supplying Ostravsko a Frýdeckomístecko
[the firm] needs its slaughterhouse.



Grammatical vs. Textual Coreference



- grammatical coreference
 - based on grammatical rules
 - typically involves a transfer of morphological info (e.g., agreement)
 - (mostly) within a single sentence
 - may be ambiguous $\Rightarrow$ disambiguated at the t-layer

Grammatical vs. Textual Coreference



- grammatical coreference
 - based on grammatical rules
 - typically involves a transfer of morphological info (e.g., agreement)
 - (mostly) within a single sentence
 - may be ambiguous $\Rightarrow$ disambiguated at the t-layer
- textual coreference
 - not realized by grammatical means alone, but also via context,
 - including references across a sentence boundary
 - vague (indistinct) devices (e.g. personal pronouns)
- "bridging anaphora" as a subtype ... only in PDT 3.0 (e.g., synonyms, generalizing nouns etc.)

Grammatical Coreference – overview



- reflexive pronouns *se/si/svůj*
- relative pronouns and pronominal adverbs
který, jaká, čím, ...; kam, kde, jak, ...; což
- control
Potřebujete poradit {#Cor.ADDR}? *Do you need an advice (lit. to be advised)?*
- quasi-control ... multi-word predicates with a noun with valency requirements
***Karel** podal {#QCor.ACT} stížnost policii.* *Karel filed a complaint to police.*
- complementations with a dual dependency expressed by a verbal form
*Honza zastihl **Hanku**, jak {#Cor.ADDR} běhá kolem rybníka.*
*Honza saw **Hanka** running (lit. how **she** was running) around the lake*
- reciprocity
***Sultáni** se vystřídali {#Rcp.PAT} na trůnu.* *lit. Sultans REFL changed (each other) on the throne.*

Grammatical Coreference



- **reflexive pronouns in Czech ... se / si**
 - does not distinguish gender, number and person
 - 'short' and 'long' forms
 - *se / si* ... (dat, acc):
 - without stress, not with prepositions
(TFA: contextually bound, not contrastive)
 - *sebe, sobě, sebe, sobě, sebou* ... (gen, dat, acc, loc, instr)
stressed and/or in prep. groups
- typically corefer with subject

Pavel vypravoval Martinovi tu historku o sobě / o něm.

Pavel told to Martin a story on himself (on Pavel) / on him (= on Martin).

Marii Karel spatřil nedaleko od sebe / od něho.

Karel saw Marie near himself / near him

passivization:

*Marie byla Karlem spatřena nedaleko od něho / *od sebe*

*Marie was seen (by Karel) near him / * herself*

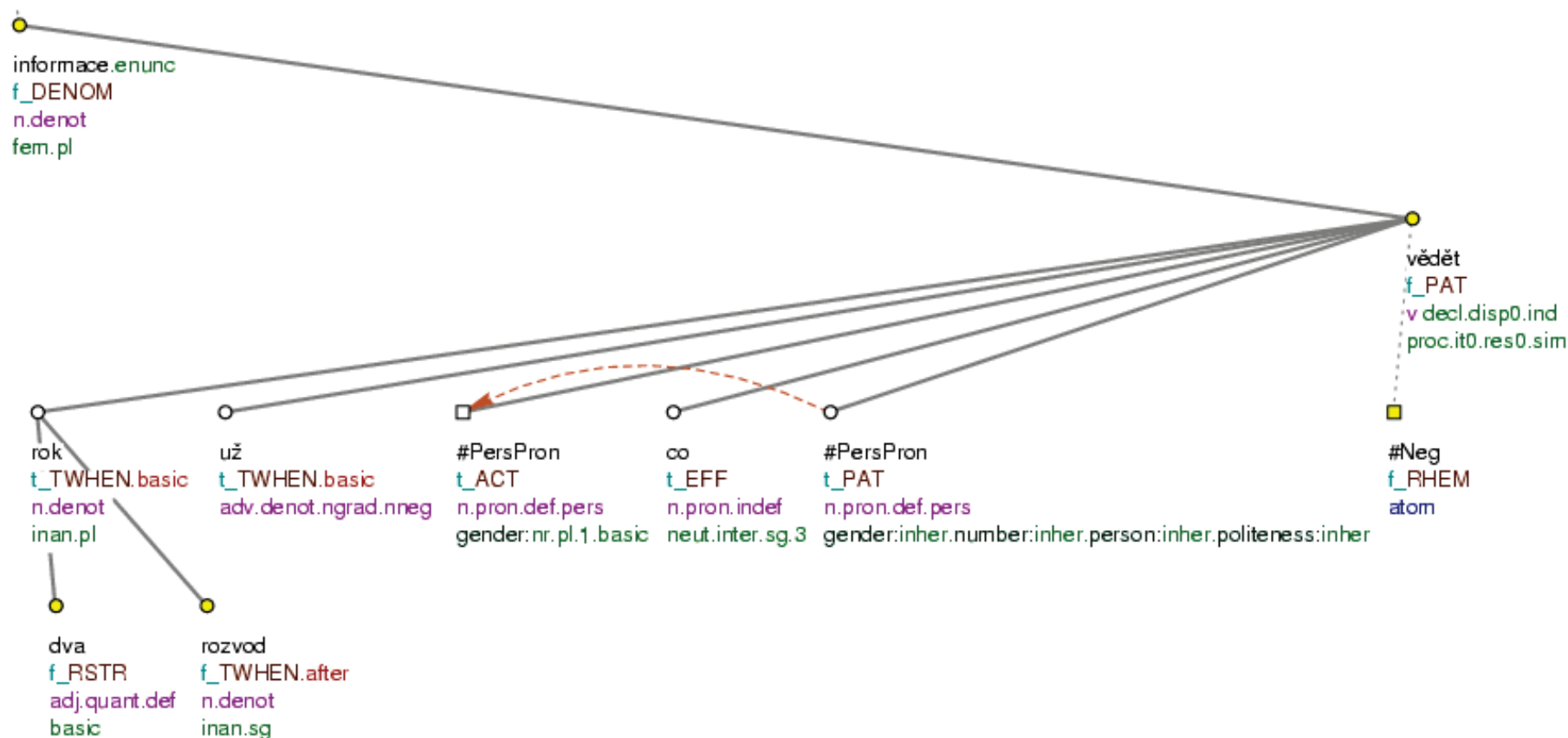


Grammatical Coreference

- reflexive pronouns in Czech ... *se / si*
- typically corefer with subject

*Informace o tom, co o **sobě**, dva roky po rozvodu, už nevíme {#PersPron.ACT}.*

*Information about the things we don't know **about each** other two years after the divorce.*





Grammatical Coreference

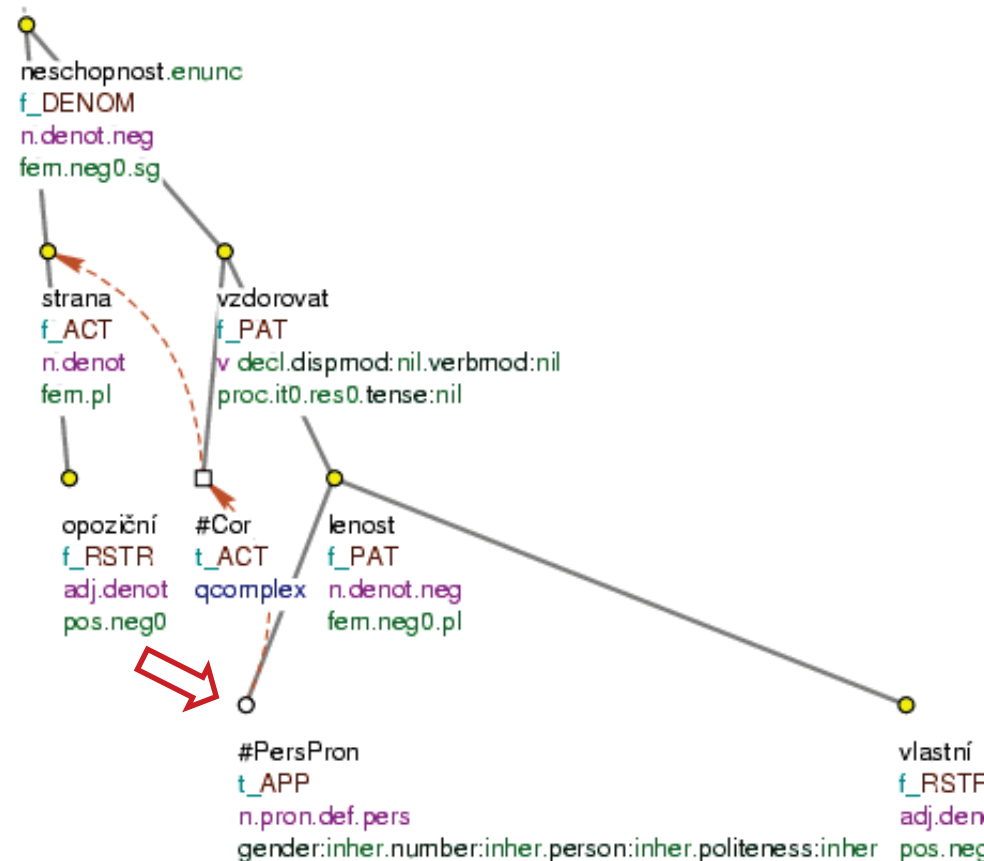
- reflexive possessive pronoun in Czech ... *svůj*
- typically corefer with the subject

Neschopnost opozičních stran

{#Cor.ACT} vzdorovat **své** vlastní lenosti.

An inability of oppositional parties

*to resist **their** own laziness*

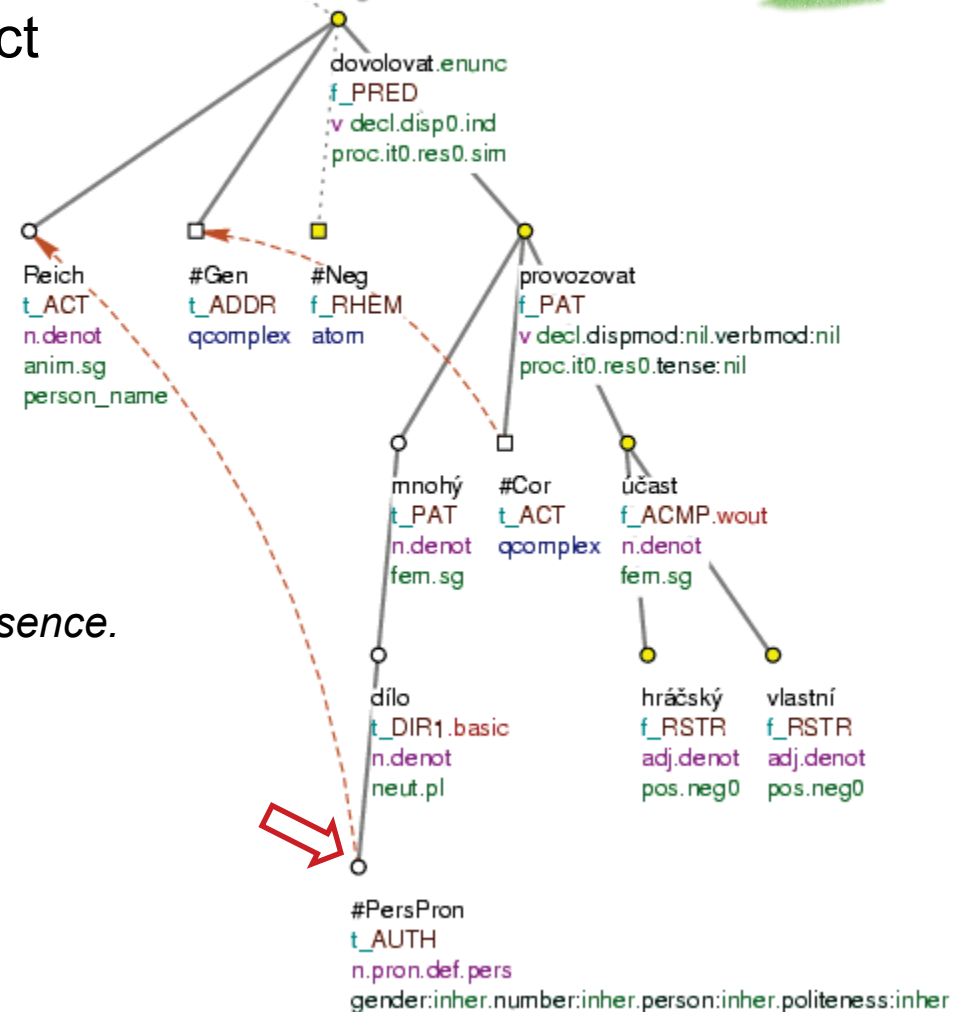


Grammatical Coreference



- reflexive possessive pronoun in Czech ... **svůj**
- typically corefere with the subject
 - BUT more complex
(e.g., subject of the governing clause)

*Mnohá ze **svých** děl Reich nedovoluje
provozovat bez vlastní hráčské účasti.
Many **of his** pieces Reich does not allow
to perform without his (own) player's presence.*



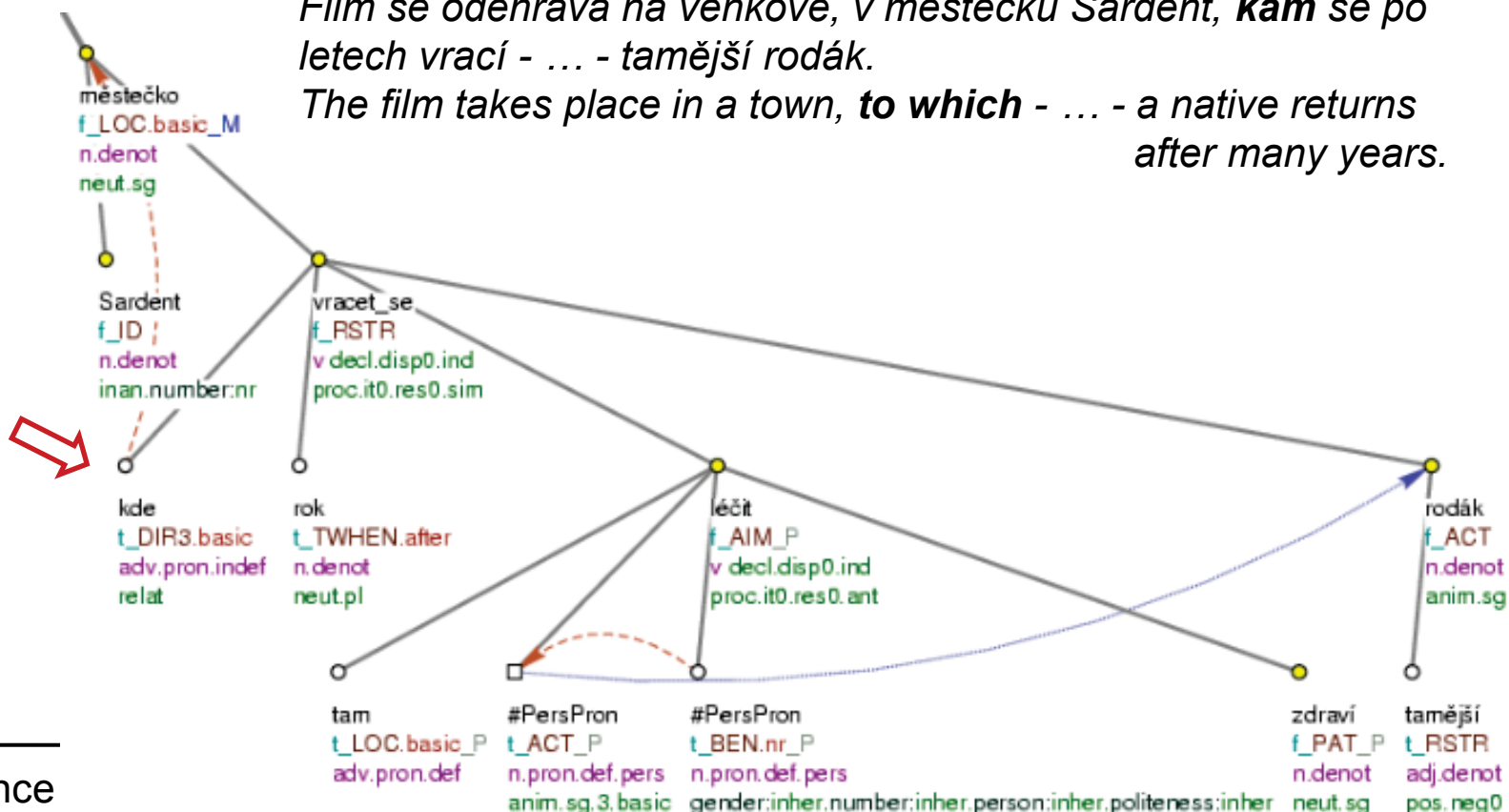
Grammatical Coreference



- **relative pronouns and pronominal adverbs**
který, jaká, čím, ...; kam, kde, jak, ...
- the relative element corefers with the noun modified by the dependent clause

*Film se odehrává na venkově, v městečku Sardent, **kam** se po letech vrací - ... - tamější rodák.*

*The film takes place in a town, **to which** - ... - a native returns after many years.*



Grammatical Coreference



- **control: verbs of control (equi verbs)**
- control ~ the (understood, unexpressed) subject of a given predicate is determined by some expression in context
= superordinate verb "controls" the arguments of a subordinate, nonfinite verb

Susan promised us to help. ... subject control with the obligatory control, predicate *promise*

Fred stopped laughing. ... subject control with the obligatory control, predicate *stop*

We tried to leave. ... subject control with the obligatory control, predicate *try*

Sue asked Bill to stop. ... object control with the obligatory control, predicate *ask*

They told you to support the effort. ... object control with the obligatory control, predicate *tell*

Someone forced him to do it. ... object control with the obligatory control, predicate *force*

Grammatical Coreference



- **control: verbs of control (equi verbs)**
- control ~ the (understood, unexpressed) subject of a given predicate is determined by some expression in context
= superordinate verb "controls" the arguments of a subordinate, nonfinite verb

Susan promised us to help. ... subject control with the obligatory control, predicate *promise*

Fred stopped laughing. ... subject control with the obligatory control, predicate *stop*

We tried to leave. ... subject control with the obligatory control, predicate *try*

Sue asked Bill to stop. ... object control with the obligatory control, predicate *ask*

They told you to support the effort. ... object control with the obligatory control, predicate *tell*

Someone forced him to do it. ... object control with the obligatory control, predicate *force*

- **control predicates** semantically select their arguments

Fred asked **you** to read it. ... *asked* is an object control verb

Jim forced **her** to say it ... *forced* is an object control verb

- **raising predicates** do not semantically select (at least) one of their dependents

Fred expects **you** to read it. ... *expects* is a raising-to-object verb

Jim believed **her** to have said it. ... *believes* is a raising-to-object verb

Grammatical Coreference



- **control: verbs of control (equi verbs)**

- the controller is a member of the valency frame of the governing verb
- the controllee is a member of the valency frame of the infinitive / deverbal noun dependent on the control verb (usually unexpressed subject)
- the infinitive / deverb. noun is a valency complementation of the control verb

Potřebujete poradit {#Cor.ADDR}?

Do you need advice (lit. to be advised)?

PDT-VALLEX

potřebovat

ACT(.1) PAT(.4,.f,aby[.v],.c) v-w4096f1 Used: 183x

nepotřebuje, co vidí; p. se k životu.AIM

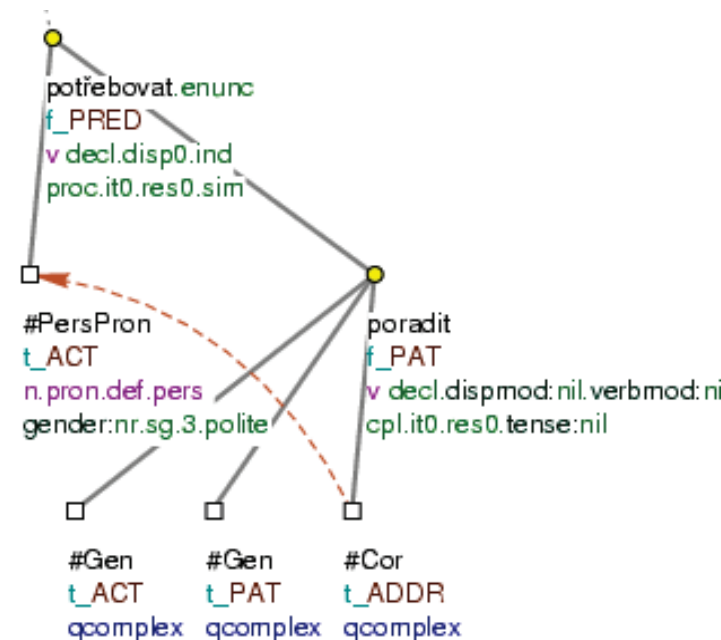
to p. čas; ta věc p. uvážit

poradit

ACT(.1) PAT(.4,.f,že[.v],aby[.v],ať[.v],s-1[.7],.c) ADDR(.3)
v-w3902f1 Used: 19x

poradil Petrovi, aby se myl pravidelně

p. mu se vším; p. mi, podle čeho se mám rozhodnout



Grammatical Coreference



- **control**: subjects of infinitives: possible t-lemmas

#Cor	the subject cannot be expressed	the subject of the infinitive is in a control relation with a modification of the main verb <i>Petr přišel {#Cor.ACT = Petr} pomoci. lit. Petr came to help.</i>
#Gen	the subject is not expressed	the subject of the infinitive is a general argument <i>Ten nápad {#Gen.ACT = sb} založit nadaci se Pavlovi líbí. Pavel likes the idea to found a foundation</i>
#PersPron	the subject is not expressed	it is possible to find the antecedent of the subject but it is not grammatical but rather textual coreference <i>Pavel likes Peter's idea to found a foundation</i>
t-lemma of a noun / #PersPron	the subject is expressed	the subject of the infinitive is expressed by a full noun or personal pronoun; these are the cases of infinitives expressing a condition

Grammatical Coreference



- **quasi-control**

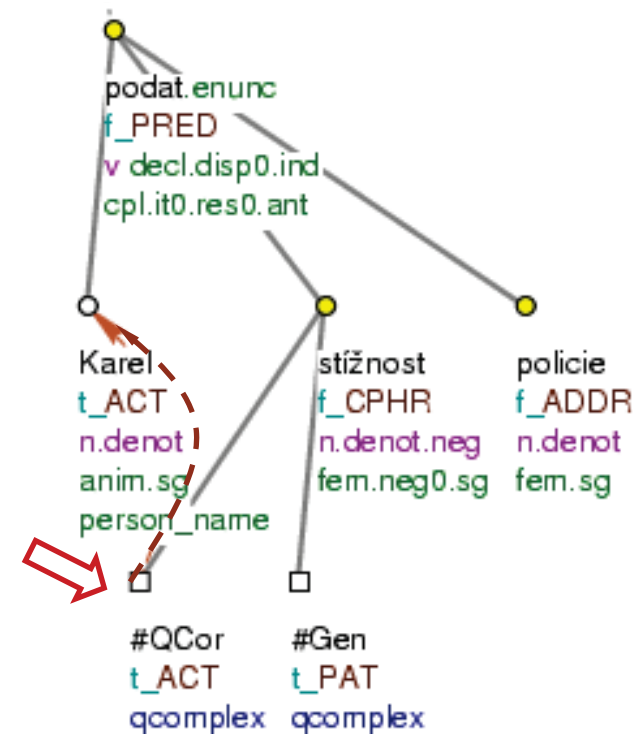
- multi-word predicates
the dependent part of which is a noun with valency requirements
- partially also verbonominal predicates (with the copula "být")

Karel podal {#QCor.ACT} stížnost policii.

Karel filed a complaint to the police.

i.e. Karel si stěžoval (policii)

Karel complained (to the police)



Grammatical Coreference



- **complementations with a dual dependency**

expressed by a verbal form

Honza zastihl Hanku, jak **{#Cor.ACT}** běhá kolem rybníka.

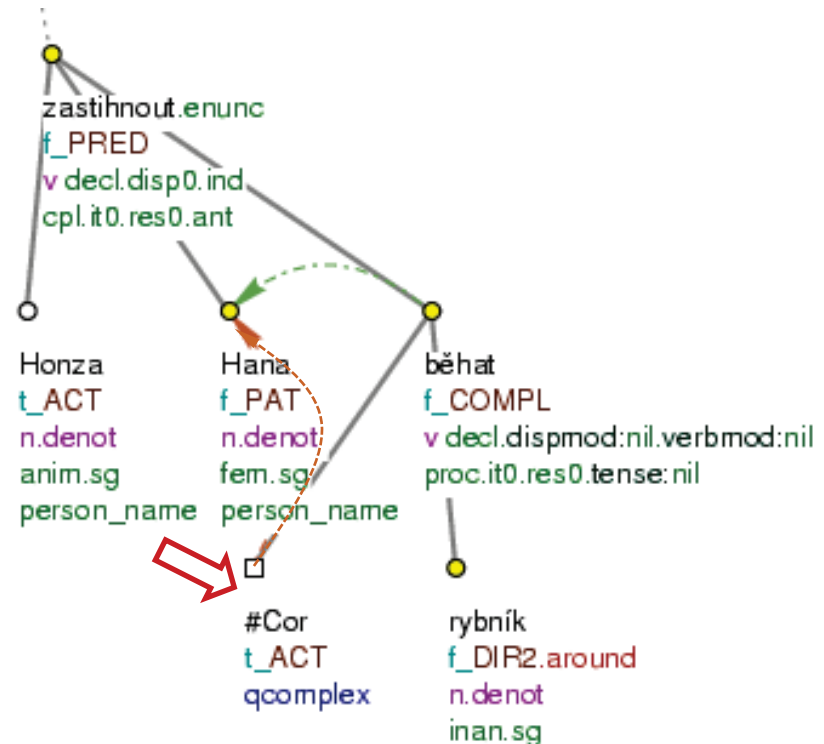
*Honza saw **Hanka** running (lit. how **she** was running) around the lake.*

- verbal forms with dual dependencies

- passive participium
- transgressive
- infinitive
- finite verb form in dependent clause

- function of the verbal form

- COMPLEMENT
- PATient, EFFECT with agreement



Grammatical Coreference



- **reciprocity** (cz vzájemnost)
 - ~ the syntactic operation on valency frames that puts two different valency modifications in a symmetric relation
 - ~ the two valency modifications have to be homogeneous
 - ~ reflexive pronoun (if ACT is involved)

Sultáni se vystřídali {#Rcp.PAT} na trůnu.

lit. *Sultans REFL changed (each other) on the throne.*

Starý sultán a nový sultán se vystřídali {#Rcp.PAT} na trůnu.

lit. *The old sultan and the new sultan REFL changed (each other) on the throne.*

Starý sultán s novým sultánem.ACMP se vystřídali {#Rcp.PAT} na trůnu.

lit. *The old sultan with the new sultan REFL changed (each other) on the throne.*

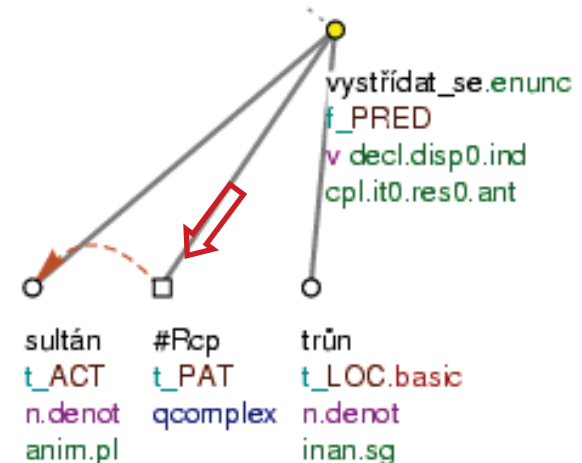
PDT-VALLEX

* vystřídát se

ACT(.1) PAT(*s-I*[.7]) v-w8405f1 Used: 1x

starý sultán se vystřídál s novým sultánem

vlády se v. Rcp. PAT



Grammatical Coreference



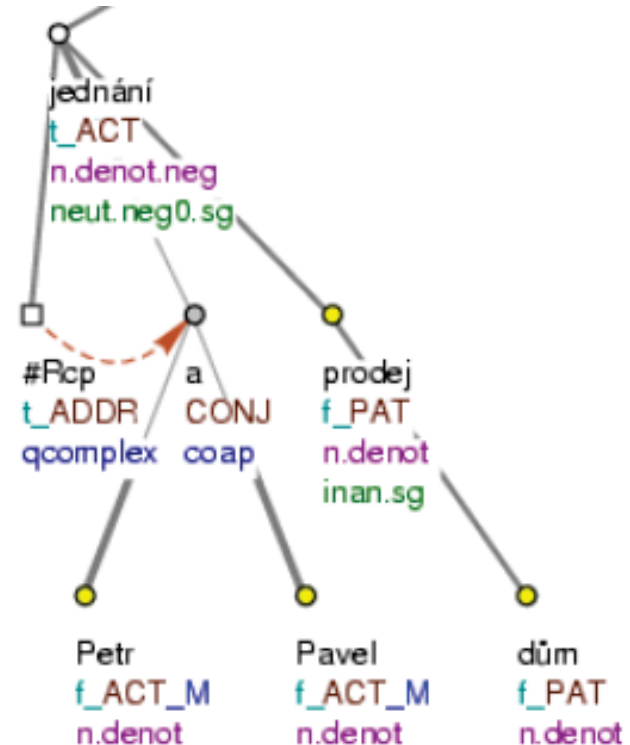
- **reciprocity** (cz vzájemnost)
 - ~ the syntactic operation on valency frames that puts two different valency modifications in a symmetric relation
 - ~ the two valency modifications have to be homogeneous
 - ~ reflexive pronoun (if ACT is involved)

jednání vlády.ACT s prezidentem.ADDR
negotiations (of) government with president
jednání vlády.ACT a prezidenta.ACT {#Rcp.ADDR}
lit. negotiations (of) government and president
jednání mezi vládou.ACT a prezidentem.ACT {#Rcp.ADDR}
lit. negotiations between government and president

PDT-VALLEX

jednání

ACT(.2,.u) **PAT**(o+6,ohledně[.2],věc:/AuxP[v-1,.2],
v-1[věc.6[tento.#]],jestli[.v],aby[.v]) **ADDR**(s+7)





Grammatical Coreference

- t-lemmas of the coreferring nodes

t-lemma	type of coreference	surface
#PersPron	refl. pronouns	<i>se, si, svůj</i>
který, jaký, co kdy, kde co	relative pronouns relative adverbs což	<i>který, jaký, jenž, co kdy, kde, kam, odkud což</i>
#Cor	two dependencies	Ø
#Cor, #GEN, #PersPron	control	Ø
#QCor, #GEN	quasi-control	Ø
#Rcp	reciprocity	<i>se</i> / Ø

References



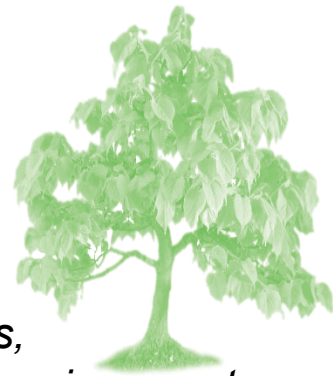
- Manual for Tectogrammatical Annotation
<http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/t-layer/html/index.html>
- Documentation for PDT 3.0
- Kučová, L., Hajičová, E. (2004) Coreferential Relations in the Prague Dependency Treebank. In *Proceedings of the Conference on Discourse Anaphora and Anaphor Resolution*, San Miguel, Azores, pp. 94-102
- Nedoluzhko, A., Mírovský, J. Ocelák, R., Pergler, J. (2009) Extended Coreferential Relations and Bridging Anaphora in the Prague Dependency Treebank. In *Proceedings of DAARC-2009*, Goa/India

Textual Coreference

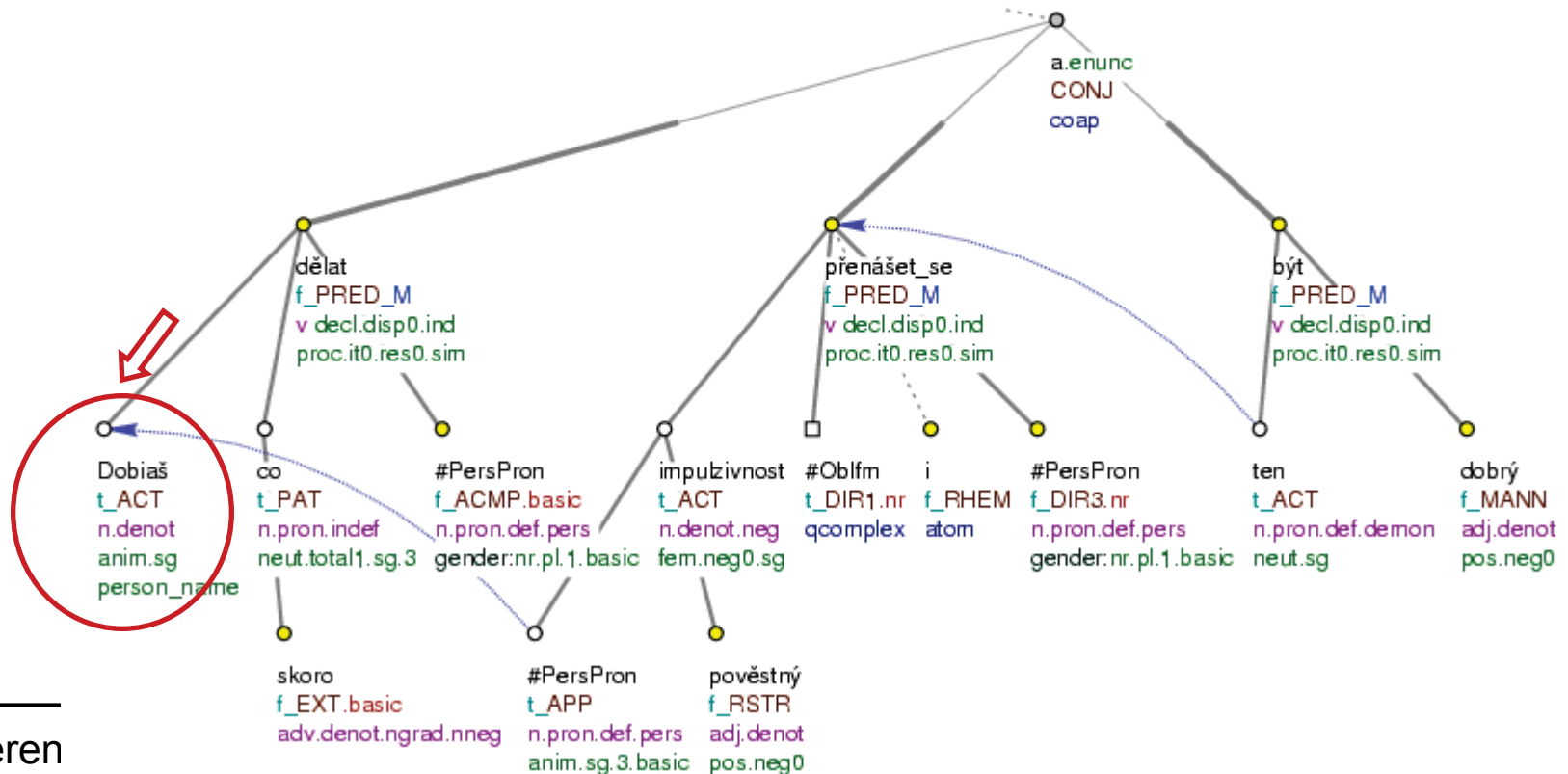


- personal and possessive pronouns, 3rd person
- demonstrative pronouns *ten*, *ta*, *to*
- (actual ellipses, where a new node with the t-lemma substitute #PersPron is added)

Textual Coreference



Dobiaš skoro všechno dělá s námi, **jeho** pověstná impulzivnost se přenáší i na nás, a to je dobře. // **Dobiaš** does almost everything with us; **his** notorious spontaneity carries over to us as well, and that is a good thing.

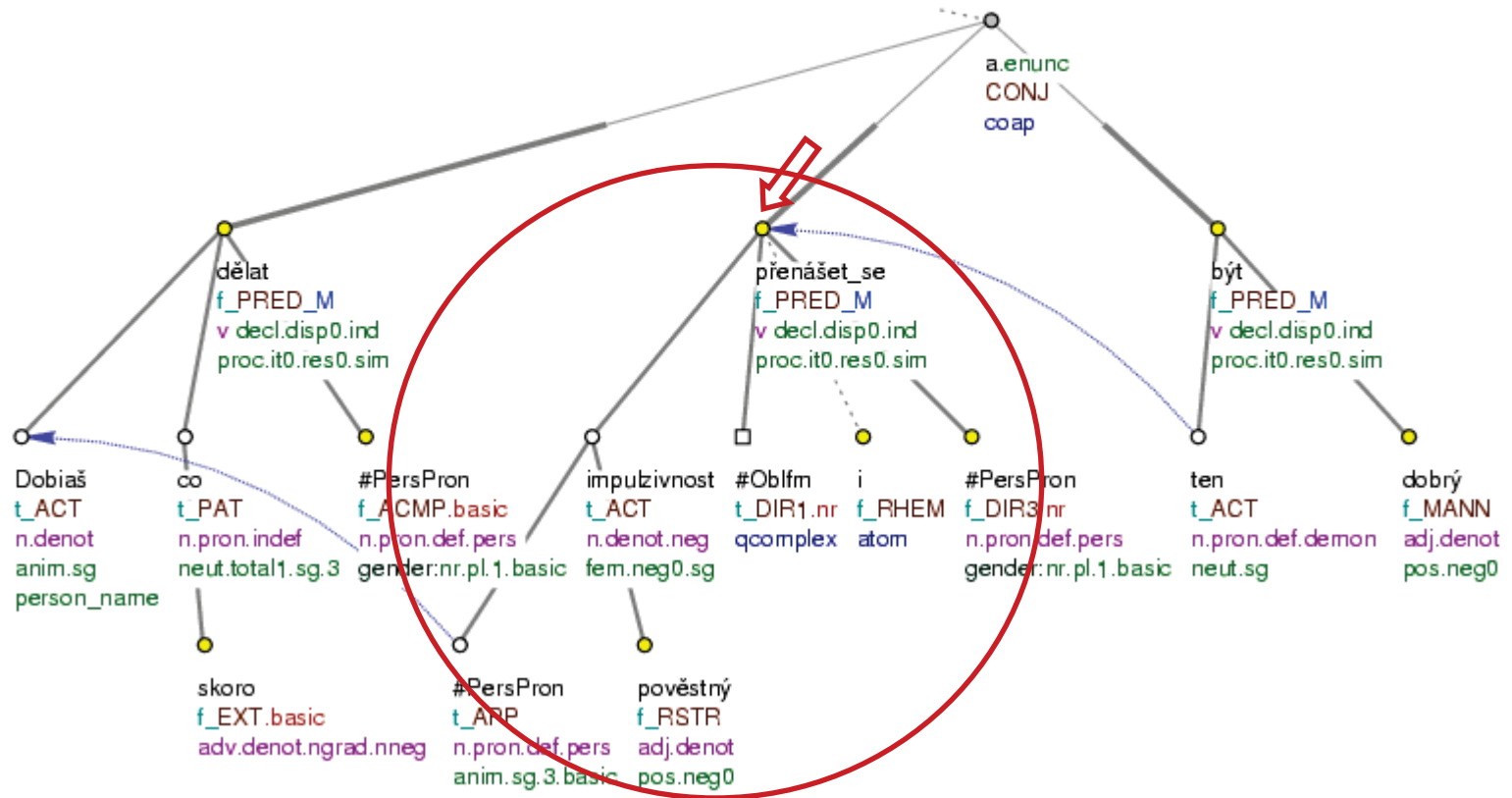


Textual Coreference



Dobiaš skoro všechno dělá s námi, jeho pověstná impulzivnost se přenáší i na nás, a to je dobře. // Dobiaš does almost everything with us; his notorious spontaneity carries over to us as well, and that is a good thing.

*Dobiaš skoro všechno dělá s námi, **jeho pověstná impulzivnost se přenáší i na nás**, a **to** je dobře. // Dobiaš does almost everything with us; his notorious spontaneity carries over to us as well, and that is a good thing.*

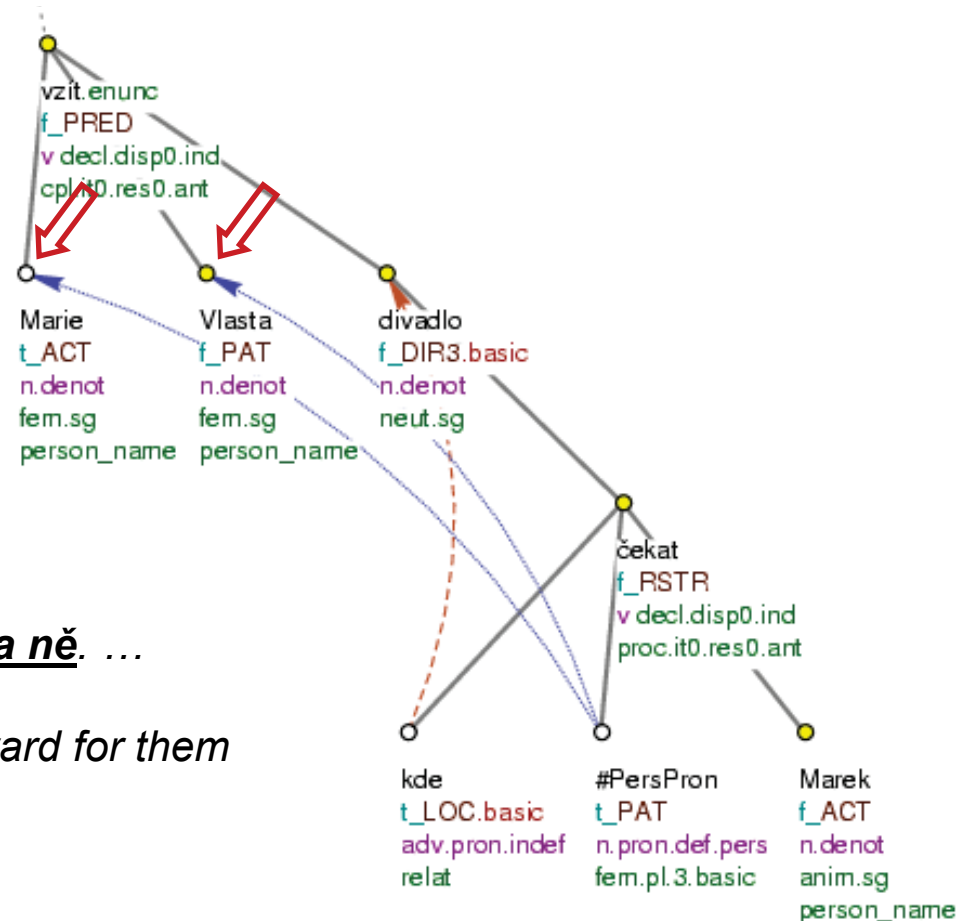


Textual Coreference



Marie vzala **Vlastu** do divadla, kde na ně čekal Marek.

lit. Marie took Vlasta to the theatre, where Marek was waiting for them.



Včera přišli **tatínek s maminkou**, těšili jsme na ně. ...

*lit. Yesterday, Daddy with mama came,
we were looking forward for them*



Special Types of Coreference

- **coref_special**
 - **segm** ... the coreferred element comprises two or more sentences or it may be inferred from them (the segment is not specified)

Rozprava o podobě reformy veřejných financí bude zahájena ve středu. Všechna jednání proběhnou za zavřenými dveřmi. Lidovým novinám **to** sdělil včera ministr financí.

The discussion about the nature of the reform of public finance will begin on Wednesday. All negotiations will take place behind closed doors. Lidové noviny (The People's Daily) was informed of **this** yesterday by the Finance Minister.



Special Types of Coreference

- **coref_special**

- **segm** ... the coreferred element comprises two or more sentences or it may be inferred from them (the segment is not specified)

*Rozprava o podobě reformy veřejných financí bude zahájena ve středu. Všechna jednání proběhnou za zavřenými dveřmi. Lidovým novinám **to** sdělil včera ministr financí.*

*The discussion about the nature of the reform of public finance will begin on Wednesday. All negotiations will take place behind closed doors. Lidové noviny (The People's Daily) was informed of **this** yesterday by the Finance Minister.*

- **exoph** ... a pronoun refers to situations or reality external to the text

*V období vrcholícího léta roku 1939 již málokdo v Evropě mohl uvěřit nadějeplným slovům britského ministerského předsedy Chamberlaina, proneseným z balkonu Buckinghamského paláce po návratu z Mnichova: Myslím, že je **to** mír na celou naši dobu.*

*After the critical summer months of 1939 hardly anyone in Europe could now lend credence to the optimistic words of the British prime minister Chamberlain spoken from the balcony of Buckingham Palace on his return from Munich: I believe **it** is peace in our time.*

Extended Textual Coreference (in PDT 3.0)



- non-pronominal coreference

- full NPs e.g. *Prague – the capital of the Czech Republic*
- anaphoric adverbs e.g. *the capital of the Czech Republic – there*
- numerals e.g. *2010 – that year*
- clauses and sentences if coreferring with NPs
e.g. *They tried to teach him to read – The attempt was not successful*
- some adjectives e.g. *Germany – German*

which have explicit antecedents in previous (ev. subsequent) context

- two types:

- SPEC - coreference of noun phrases with specific reference (Germany – the state)
- GEN - coreference of noun phrases with generic reference

*Droga je tak účinná, že ten, kdo ji. **GEN** užívá, se snadno dostane do „pohody“ kouřením nebo šňupáním.*
The **drug** is so effective that one can easily achieve the state of “coolness” by smoking or snorting **it.GEN**.

Bridging Anaphora (in PDT 3.0)



- bridging ~ a relation between two elements
the second element is interpreted by an inferential process (“bridge”) on the basis of the first one
- PDT 3.0: restriction to relations between nominal expressions



Bridging Anaphora (not in PDT 2.0)

- bridging ~ a relation between two elements
the second element is interpreted by an inferential process (“bridge”) on the basis of the first one
- PDT 3.0: restriction to relations between nominal expressions
- types of relations:
 - **PART_WHOLE // WHOLE_PART** e.g. *room – ceiling; Germany – Bavaria – Munich*
 - **SET_SUB // SUB_SET** e.g. *participants – some participants – one of participants*
 - **P_FUNCT // FUNCT_P** e.g. *government – prime minister*
 - **CONTRAST**... coherence-relevant discourse opposites

Osud ČR bude stále více spojený s Německem ...a osud Slovenska.CONTRAST s Ruskem.

The future of the Czech Republic will become more associated with Germany, while the future of Slovakia.CONTRAST will be more associated with Russia.

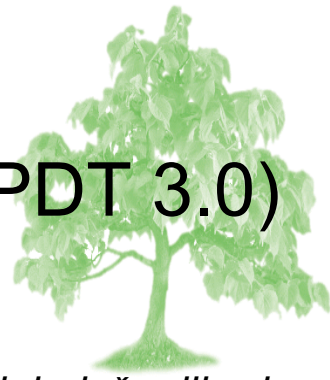
- **ANAF**

Je třeba mít **vysoké cíle** a s **malými [cíli.ANAF]** se nespokojit

It is necessary to have lofty aims and not to be satisfied with small [ones.ANAF].

- **REST** e.g., *location – resident, mother – son, listening – listener, ...*

Coreference and Bridging for pronouns (PDT 3.0)



- 1st and 2nd person

Potřebu dalších investic [#PersPron.ACT] odhaduji do roku dva tisíce na více jak dvě miliardy korun, říká ředitel Nováček.

I estimate the need for further investment in the year two thousand to more than two billion, says the director Nováček.

- generic use:

Slévárně Škoda v Českých Budějovicích dluží plzeňská Škoda 61 miliónů Kčs.

[#PersPron.ACT] Potřebujeme je hned a na stůl. Situace je vážná a z naší strany téměř neřešitelná. Bez finančních prostředků se už [#PersPron.ACT] neobejdeme," řekl včera Milan Fučík.

The Škoda's branch in Pilsen owes the foundry Skoda in České Budejovice 61 million crowns. We need them now, and on the table. The situation is serious and almost unsolvable from our side. We will not manage [to resolve] it without funds," Milan Fucik said yesterday.

References



- Manual for Tectogrammatical Annotation
<http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/t-layer/html/index.html>
- Documentation for PDT 3.0
- Kučová, L., Hajičová, E. (2004) Coreferential Relations in the Prague Dependency Treebank. In *Proceedings of the Conference on Discourse Anaphora and Anaphor Resolution*, San Miguel, Azores, pp. 94-102
- Nedoluzhko, A., Mírovský, J. Ocelák, R., Pergler, J. (2009) Extended Coreferential Relations and Bridging Anaphora in the Prague Dependency Treebank. In *Proceedings of DAARC-2009*, Goa/India