

MATEMATICKO-FYZIKÁLNÍ FAKULTA

PRAHA

**ANOTACE NA TEKTOGRAMATICKÉ ROVINĚ.  
DODATKY K ANOTÁTORSKÉ PŘÍRUČCE  
(S OHLEDEM NA ANOTOVÁNÍ PDTSC A PCEDT)**

MARIE MIKULOVÁ

ÚFAL Technical Report  
**TR-2013-51**

ISSN 1214-5521



UNIVERSITAS CAROLINA PRAGENSIS

Copies of ÚFAL Technical Reports can be ordered from:

Institute of Formal and Applied Linguistics (ÚFAL MFF UK)

Faculty of Mathematics and Physics, Charles University

Malostranské nám. 25, CZ-11800 Prague 1

Czech Republic

or can be obtained via the Web: <http://ufal.mff.cuni.cz/techrep>

Anotace na tektogramatické rovině  
Dodatky k anotátorské příručce  
(s ohledem na anotování PDTSC a PCEDT)

Marie Mikulová

## Obsah

Úvod .....	3
Odkazy do analytické roviny .....	4
T-lemma .....	10
Anotace valence u sloves .....	12
Anotace valence u jiných slovních druhů .....	15
Reciprocita .....	17
Elipsy .....	23
Zvláštní případy souřadnosti .....	26
Syntaktické frazeologismy .....	28
Měrové významy .....	29
Výrazy <i>se</i> a <i>si</i> .....	31
Čísla .....	34
Cizí vlastní jména a názvy .....	39
Zkratky .....	41
Nové sekundární předložky .....	45
Různé .....	46
Segmentace vět do stromů a slov do uzlů .....	49
Anotátorská poznámka .....	54
Šablony .....	59
Reference .....	73

## Úvod

Komplexní, sémanticko-syntaktická (tektogramatická) anotace českých textů byla v rámci pražských závislostních korpusů poprvé realizována v Pražském závislostním korpusu 2.0 (Hajič a kol., 2006; zkratka PDT 2.0 z anglického *The Prague Dependency Treebank 2.0*.) Pravidla anotace byla souhrnně popsána v anotačním manuálu *Anotace na tektogramatické rovině. Anotátorská příručka* (Mikulová a kol., 2005). Později vyšla i anglická verze této příručky (Mikulová a kol., 2006a) a též zkrácené referenční verze, opět česky i anglicky (Mikulová a kol., 2006b a 2006c).

Podle těchto pravidel byly anotovány další dva korpusy s českými daty. Nejprve byla tektogramatickou anotací opatřena česká část Pražského česko-anglického závislostního korpusu 2.0 (Hajič a kol., 2011; zkratka PCEDT 2.0 z anglického *The Prague Czech-English Dependency Treebank 2.0*). Korpus PCEDT 2.0 obsahuje články z deníku Wall Street Journal (z roku 1989), které byly pro českou část korpusu manuálně přeloženy do češtiny. Jedná se převážně o texty s ekonomickou tematikou. Následovala tektogramatická anotace Pražského závislostního korpusu mluvené češtiny (připravuje se k vydání v 2014; zkratka PDTSC 2.0 z anglického *The Prague Dependency Treebank of Spoken Czech*). Korpus PDTSC 2.0 obsahuje standardizované přepisy rozhovorů (lehce moderované dialogy s lidmi, kteří přežili holocaust, a konverzace nad osobní sbírkou fotografií jednoho z řečníků).

V korpusech PCEDT 2.0 a PDTSC 2.0 nebyla tektogramatická anotace provedena v takovém rozsahu jako v korpusu PDT 2.0. Byla realizována anotace struktury, funktorů, valence u sloves (a v omezené míře u substantiv), byly anotovány odkazy do a-roviny, gramatická a textová koreference. Oproti PDT 2.0 není realizována anotace gramatémů, typů uzlů, aktuálního členění a některých dalších atributů, např. `quot_type` (více viz Mikulová – Štěpánek, 2009). Toto omezení rozsahu anotace si vyžádalo úpravy dosavadních anotačních pravidel (zejména pravidel týkajících se tektogramatického lemmatu a valence). Anotace nových textů (s novými syntaktickými jevy, s dalšími příklady) pak s sebou přinesla také nové poznatky. Dosavadní pravidla byla zpřesňována, doplňována a pro některé jevy byla přidána pravidla nová.

Poznámka: U každého dodatku je uvedeno, ke které kapitole v původním anotačním manuálu (Mikulová a kol., 2005) se vztahuje. Odkaz začíná zkratkou „MAN“, následuje číslo a název příslušné kapitoly.

## Poděkování

Dokument vznikl za finanční podpory projektu GA ČR P406/2010/0875 a projektu MŠMT LINDAT-Clarín LM2010013.

## Odkazy do analytické roviny

Nahrazuje kapitolu MAN 1.1. Vztah tektogramatické roviny k nižším rovinám.

1	Dva druhy odkazů do a-roviny .....	4
2	Určování a-lex-uzlu a a-aux-uzlů.....	5
2.1	Základní pravidla .....	5
2.2	Specifické případy.....	6
2.2.1	Specifické případy I: T-uzly s víceslovnými t-lematy .....	6
2.2.2	Specifické případy II: Pasivum, dispoziční modalita a reciprocita .....	7
2.2.3	Specifické případy III : Nealfanumerické znaky .....	7
2.3	K odkazům u smíšených souřadných spojení.....	9

Legenda k příkladům: <a-aux-uzel> a-lex-uzel [atribut = hodnota]

Ve špičatých závorkách < > je uzavřen výraz odkazovaný jako a-aux-uzel. Podtržený výraz je odkazovaný jako a-lex-uzel. V každém příkladu jsou pomocí špičatých závorek a podtrženého písma označeny odkazované analytické uzly (a-uzly) vždy pro jeden tektogramatický uzel (t-uzel). V hranatých závorkách [ ] za příkladem mohou být uvedeny hodnoty některých dalších atributů daného tektogramatického uzlu (zejména jeho t-lemma). Složenými závorkami { } je naznačen doplněný tektogramatický uzel.

### 1 Dva druhy odkazů do a-roviny

Tektogramatická rovina je propojena s rovinou analytickou. Vztahy mezi rovinami jsou zachyceny tak, že na každý a-uzel, který nějakým způsobem ovlivňuje hodnotu určitého atributu t-uzlu, vede z daného t-uzlu odkaz. Podle toho, jaký typ atributu a-uzel ovlivňuje, rozlišujeme dva druhy odkazů do a-roviny. Odkazované a-uzly, tj. a-uzly ovlivňující hodnoty atributů t-uzlu, pak označujeme jako:

- **a-lex-uzel** = a-uzel, ze kterého t-uzel získal svůj lexikální význam, případně jeho největší část.  
A-lex-uzel má vliv na hodnotu atributu *t\_lemma*.  
Jedná se zpravidla o a-uzel reprezentující autosémantické slovo.
- **a-aux-uzel** = ostatní odkazované a-uzly.  
A-aux-uzly mají zpravidla vliv na hodnoty atributů jako je *functor*, *subfunctor* a *gram*.  
Jedná se zpravidla o a-uzly reprezentující funkční slova: předložky, spojky, pomocná slovesa, odkazovací slova.

**A-lex-uzel u t-uzlu (pokud existuje) je vždy jen jeden. A-aux-uzlů může být více. Na jeden a-uzel může vést odkaz z více t-uzlů.**

## 2 Určování a-lex-uzlu a a-aux-uzlů

### 2.1 Základní pravidla

#### **Pravidlo A:**

Odkazované a-uzly u t-uzlu reprezentujícího **jméno**:

- **a-lex-uzel**: a-uzel reprezentující jméno (substantivum, adjektivum, číslovku, zájmeno).
- **a-aux-uzly**: a-uzly reprezentující: předložku  
(výraz *se* - u zvrtného deverbativa, viz pravidlo D),  
(tečku za řadovou číslovkou – viz pravidlo Q).

Příklady:

- (1) <na> zahradě [t\_lemma = zahrada]
- (2) Pozor! <přes> dvacet korun [t\_lemma = dvacet]
- (3) <na> jabloni a hrušni [t\_lemma = jabloň]
- (4) <na> jabloni a hrušni [t\_lemma = hrušeň]
- (5) <Ve> <srovnání> <s> Pavlem jsem lepší. [t\_lemma = Pavel]

#### **Pravidlo B:**

Odkazované a-uzly u t-uzlu reprezentujícího **sloveso** (predikát):

- **a-lex-uzel**: a-uzel reprezentující plnovýznamové sloveso.
- **a-aux-uzly**: a-uzly reprezentující: pomocné sloveso,  
modální sloveso,  
odkazovací slovo,  
podřadící spojku,  
(výraz *se* při deagentní diatezi – viz pravidlo J),  
(výraz *se* při dispoziční diatezi – viz pravidlo K),  
(výraz *se* u zvrtného slovesa – viz pravidlo D).

Příklady:

- (6) Zítra <nebudu> dělat nic. [t\_lemma = dělat]
- (7) Udělám to <proto>, <aby> <mohl> přijít. [t\_lemma = přijít]
- (8) <Kdybys> <byl> <býval> přišel, nestalo by se to. [t\_lemma = přijít]
- (9) *Kdybys byl býval přišel, nestalo <by> <se> to.* [t\_lemma = stát\_se]
- (10) Přišel. [t\_lemma = přijít]

#### **Pravidlo C:**

Odkazované a-uzly u **ostatních t-uzlů** (reprezentujících částice, citoslovce, souřadící spojku):

- **a-lex-uzel**: a-uzel reprezentující částici, citoslovce, jednoslovnou souřadící spojku.
- **a-aux-uzly**: žádné.

Příklady:

- (11) Pardon. [t\_lemma = pardon]
- (12) na jabloni a hrušni [t\_lemma = a]

## 2.2 Specifické případy

### 2.2.1 Specifické případy I: T-uzly s víceslovnými t-lematy

#### **Pravidlo D:**

Odkazované a-uzly u t-uzlu reprezentujícího zvrtné jméno nebo sloveso jsou:

- **a-lex-uzel:** a-uzel reprezentující jméno nebo plnovýznamové sloveso.
- **a-aux-uzel:** a-uzel reprezentující zvrtné *se*.

Příklady:

- (13) *smát* <se> [t\_lemma = smát\_se]
- (14) *chovající* <se> [t\_lemma = chovající\_se]

#### **Pravidlo E:**

Odkazované a-uzly u t-uzlu reprezentujícího závislou část frazému (*functor*=DPHR):

- **a-lex-uzel:** a-uzel reprezentující řídící slovo závislé části frazému (slovo je součástí t-lematu). Nelze-li řídící slovo závislé části frazému určit, je a-lex-uzlem první plnovýznamové slovo (podle povrchového pořadí), které náleží závislé části frazému.
- **a-aux-uzly:** a-uzly reprezentující ostatní slova, která náleží závislé části frazému.

Příklady:

- (15) *Běhá mi mráz* <po> <zádech>. [t\_lemma = mráz\_po\_zádech]
- (16) *Bere všechno* <na> <lehkou> *váhu*. [t\_lemma= na\_lehkou\_váhu]

#### **Pravidlo F:**

Odkazované a-uzly u t-uzlu reprezentujícího víceslovný spojovací výraz (*nodetype* = *coop*):

- **a-lex-uzel:** a-uzel reprezentující poslední (podle povrchového pořadí) výraz náležející víceslovnému spojovacímu výrazu.
- **a-aux-uzly:** a-uzly reprezentující ostatní výrazy tvořící spojovací výraz v povrchové podobě věty.

Pozor! Výrazy modifikující (i víceslovný) spojovací výraz jsou reprezentovány samostatnými t-uzly (*functor* = CM).

Příklady:

- (17) <Bud'to> *půjdeme hned, nebo nepůjdeme vůbec*. [t\_lemma = bud'to\_nebo]
- (18) *Udělám to*, <a> *to hned*. [t\_lemma = a\_to]
- (19) *Koupil chleba* <a> *rohlíky a housky*. [t\_lemma = a\_a]

#### **Pravidlo G:**

Odkazované a-uzly u t-uzlu s t-lematem typu „12 hodinový“:

- **a-lex-uzel:** a-uzel reprezentující adjektivum.
- **a-aux-uzly:** a-uzly reprezentující číslo.

Příklad:

- (20) <8> *hodinová pracovní doba* [t\_lemma = 8\_hodinový]

#### **Pravidlo H:**

Odkazované a-uzly u t-uzlu s t-lematem typu „číslo číslo( číslo)<sup>n</sup>“:

- **a-lex-uzel:** a-uzel reprezentující první číslo.
- **a-aux-uzly:** a-uzly reprezentující ostatní čísla.

Příklady:

- (21) *telefon 345* <629> <456> [t\_lemma = 345\_629\_456]
- (22) *linka 28* <40> [t\_lemma = 28\_40]



### Pravidlo I:

Odkazované a-uzly u t-uzlu s t-lematem typu „**van Beethoven**“:

- **a-lex-uzel**: a-uzel reprezentující jméno.
- **a-aux-uzly**: a-uzel reprezentující cizí předložku (případně apostrof).

Dosavadní seznam předložek (podle PDT 2.0): *von, van, O, d, de, da, di, zum*.

Příklad:

(23) *Ludwig* <van> *Beethoven* [*t\_lemma* = *van\_Beethoven*]

## 2.2.2 Specifické případy II: Pasivum, dispoziční diateze a reciprocita

### Pravidlo J:

A-uzel pro výraz *se* přítomný ve větě z důvodu vyjádření **reflexivního deagentivu** je:

- **a-aux-uzel** u efektivního kořene pasivní klauze (u t-uzlu reprezentujícího řídicí sloveso v této diatezi).

Příklady:

(24) *Košile* <se> *pere a kalhoty také*. [*t\_lemma* = *prát*]

(25) *Košile* <se> *pere a kalhoty {prát} také*. [*t\_lemma* = *prát*] – u doplněného uzlu

### Pravidlo K:

A-uzel pro výraz *se* přítomný ve větě z důvodu vyjádření **dispoziční diateze** je:

- **a-aux-uzel** u efektivního kořene klauze v dispoziční diatezi (u uzlu pro řídicí sloveso, které nese dispoziční modalitu, *gram/dispmod* = *dispI*).

Příklad:

(26) *Matematika* <se> *mi studuje dobře*. [*t\_lemma* = *studovat*]

### Pravidlo L:

A-uzly reprezentující formální ukazatele **syntaktické reciprocit** (výraz *se* v různých (předložkových) pádech, *jeden druhý*; jsou zachyceny jako:

- **a-aux-uzly** u doplněného t-uzlu s t-lematem *#Rcp*.

Příklady:

(27) *Pavel a Martin* <na> <sebe> *narazili {#Rcp.PAT}*.

(28) *Pavel a Martin narazili* <jeden> <na> <druhého> {*#Rcp.PAT*}.

(29) *Martin a Radek se* <mezi> <sebou> *domluvili*. {*#Rcp.ADDR*}

## 2.2.3 Specifické případy III : Nealfanumerické znaky

### 2.2.3.1 Specifické případy III : T-uzly reprezentující nealfanumerické znaky

Tato pravidla se týkají t-uzlů, které reprezentují nealfanumerické znaky (nutnou podmínkou je *is\_generated* = 0).

### Pravidlo M:

Odkazované a-uzly u t-uzlu reprezentujícího **jeden nealfanumerický znak**:

- **a-lex-uzel**: a-uzel reprezentující tentýž interpunkční znak.
- **a-aux-uzly**: žádné.

Toto pravidlo se týká t-uzlů se zástupnými t-lematy: *#Comma, #Dash, #Colon, #Slash, #Period, #Semicolon, #Ast, #Amp, #Percnt*.

Příklady:

(30) *Mladost \_ radost*. [*t\_lemma* = *#Dash*; *functor* = *PRED*]

(31) *20 %* [*t\_lemma* = *#Percnt*]

### Pravidlo Na:

A-uzly u kořene souřadné struktury (*nodetype=coap*) reprezentujícího **interpunkční znak jako oddělovač koordinovaných členů** (které nejsou spojeny spojkou):

- **a-lex-uzel:** a-uzel reprezentující poslední z interpunkčních znaků sloužících jako oddělovače koordinovaných členů.
- **a-aux-uzly:** žádné.

Pravidlo se týká uzlů: *#Comma, #Dash, #Colon, #Slash, #Period, #Semicolon*.

Příklad:

(32) *Máme psa, kočku, rybičky, křečka, andulku...* [*t\_lemma = #Comma*]

### Pravidlo Nb:

A-uzly u kořene souřadné struktury (*nodetype=coap*) reprezentujícího **interpunkční znak jako oddělovač aponovaných členů** (které nejsou spojeny spojkou):

- **a-lex-uzel:** a-uzel reprezentující první z interpunkčních znaků sloužících jako oddělovače aponovaných členů.
- **a-aux-uzly:** žádné.

Toto pravidlo se týká t-uzlů se zástupnými t-lematy: *#Comma, #Dash, #Colon, #Slash*.

Příklady:

(33) *Máme rádi Boženu Němcovou, autorku Babičky.* [*t\_lemma = #Comma*]

(34) *Boženu Němcovou, autorku Babičky, máme rádi.* [*t\_lemma = #Comma*]

### Pravidlo O:

Odkazované a-uzly u t-uzlu reprezentujícího **tři tečky** (*t\_lemma=#Period3*):

- **a-lex-uzel:** a-uzel reprezentující první tečku.
- **a-aux-uzly:** a-uzly reprezentující zbývající dvě tečky.

Příklad:

(35) *Máme psa, kočku, rybičky, křečka, andulku.<.><.>* [*t\_lemma = #Period3*]

### Pravidlo P:

Odkazované a-uzly u t-uzlu reprezentujícího **závorku** (*t\_lemma=#Bracket*):

- **a-lex-uzel:** a-uzel reprezentující levou závorku.
- **a-aux-uzly:** žádné.

Příklad:

(36) *ÚFAL (Ústav formální a aplikované lingvistiky)* [*t\_lemma = #Bracket*]

## 2.2.3.2 **Specifické případy III : Nealfanumerické znaky nereprezentované t-uzlem**

### Pravidlo Q:

A-uzel reprezentující **tečku za řadovou číslovkou** je zachycen jako:

- **a-aux-uzel** u t-uzlu reprezentujícího řadovou číslovku.

Příklad:

(37) *2 <.> ledna* [*t\_lemma = 2*]

### Pravidlo R:

A-uzel reprezentující **spojovník v kompozitech** (ve kterých je mezi spojenými slovy determinační vztah) zachycen jako:

- **a-aux-uzel** u t-uzlu reprezentujícího závislé spojené slovo.

Příklad:

(38) *stupnice <->dur* [*t\_lemma = c*]

Na ostatní nealfanumerické znaky nerepresentované t-uzlem nevede ze žádného t-uzlu odkaz. Jedná se o následující a-uzly:

- a-uzly reprezentující **interpunkční znaky oddělující souřadně spojené větné členy** (pokud nejde o typy popsané pravidly: Na a Nb),
- a-uzly reprezentující **čárky (interpunkční znaky), které oddělují závislý člen** (klauzi, volný přívlastek),
- a-uzel reprezentující **dvojtečku uvozující přímou řeč**,
- a-uzel reprezentující **uvozovky ohraničující úsek textu**,
- a-uzel reprezentující **závorku (jiný interpunkční znak) ohraničující vsuvku**,
- a-uzel reprezentující **interpunkční znak značící konec věty**,
- a-uzel reprezentující **tečku za zkratkami**.

### 2.3 K odkazům u smíšených souřadných spojení

Pravidla anotace stanovují, že při „smíšeném“ apozičním spojení má větný člen souřadnosti stejný funktor jako nevětný člen souřadného spojení.

*Váže těžké kovy.PAT\_M, jako.APPS je.PAT\_M plutonium.*

Podobně při koordinačních/apozičních spojeních se vztahným *což*, dostává efektivní kořen věty připojené vztahným *což* stejný funktor jako druhý člen souřadného spojení.

*O práci přijdou všichni.ACT\_M, kdo pracovali, [#Comma.APPS] což je.ACT\_M asi 46 lidí.*

Vzniká otázka, zda v těch případech, kdy je doplnění určující funktor druhého členu souřadného spojení vyjádřené předložkovou formou nebo závislou klauzí uvozenou podřadící spojkou (a tato předložka/spojka má vliv na funktor), zda tato předložka/spojka má být odkázána v atributu *a/aux.rf* i u druhého větného členu souřadného spojení (podobně jako ve standardních případech souřadných spojení: *jel do Prahy a <do> Brna*). Otázka souvisí i s tím, jak jsou v těchto případech nastaveny kontroly valence. A vůbec se nabízí otázka, zda přijatá pravidla anotace (řešit tyto případy jako smíšenou apoziční a koordinaci) jsou dobře navržená.

#### Dosavadní řešení:

Funktor druhého z členů smíšené koordinace nebo apoziční vyjádřeného větně je stejný jako funktor prvního nevětně vyjádřeného členu.

I u druhého členu souřadného spojení (smíšená koordinace, apoziční) je v atributu *a/aux.rf* odkaz na funkční slovo určující funktor souřadně spojených členů.

#### Příklady:

*Koncern si tuto akvizici přál, <protože> společnost A se zabývá.CAUS\_M solárními články, což zapadá.CAUS\_M [a/aux.rf=protože] do odvětví společnosti B.*

*<včetně> gigantů.ACMP\_M, <jako> jsou.ACMP\_M [a/aux.rf=včetně] společnosti A a B*

*Zaměřili se <na> akcie.PAT\_M, <jako> jsou.PAT\_M [a/aux.rf=na]....*

*Akcie stouply <na> 55.PAT\_M, což je.PAT\_M [a/aux.rf=na] 10% z celkové hodnoty*

Poznámka: Manuál ani žádný dodatek nepopisuje „elipsu funkčních slov“, která se odkazují v atributu *a/aux.rf*. Jde o příklady jako:

*jel <do> Prahy.DIR3\_M a Brna.DIR3\_M [a/aux.rf=do]*

*<musel> utírat.PRED\_M nádobí a čistit.PRED\_M [aux: musel] boty*

*<v> Praze.LOC\_M, Brně.LOC\_M [aux: v] a Bratislavě.LOC\_M [aux: v]*

Jde o jasné případy. Nicméně při zpracování problematiky odkazů do a-roviny se na tyto typy nesmí zapomenout.

## T-lemma

Dodatek k MAN 3. Tektogramatické lema (t-lemma)

1	T-lemma u t0-uzlů ( <i>is_generated</i> = 0).....	10
1.1	Případy víceslovných t-lemat.....	10
1.2	Zástupná t-lemata pro nealfanumerické znaky .....	10
1.3	#PersPron.....	11
1.4	Překlep.....	11
1.5	Špatně analyzované základní slovní formy .....	11
1.6	T-lemata cizojazyčných výrazů.....	11
2	T-lemma u t1-uzlů ( <i>is_generated</i> = 1).....	11
2.1	„Zkopírované“ t-uzly.....	11
2.2	Doplněné t-uzly se zástupnými t-lematy .....	11

### 1 T-lemma u t0-uzlů (*is\_generated* = 0)

Za správnou podobu t-lemat t0-uzlů (tj. u uzlů s hodnotou 0 v atributu *is\_generated*) se v první fázi anotace považuje jejich základní slovní forma (nominativ, infinitiv), tedy v zásadě to t-lemma, které už ve stromě je.

(Derivační) změny t-lemat adverbii na adjektivní, změny t-lemat u zájmen apod. (popsané v MAN zejména v kapitole 4.1 Syntaktická a lexikální derivace) se nebudou provádět (a již automaticky provedené změny se nebudou opravovat ani kontrolovat)!

Podobně se nebude anotovat tzv. reprezentativní t-lemma (popsané v MAN 3.3.1).

V první fázi anotace se zachytí konečná podoba t-lematu u následujících případů t0-uzlů.

#### 1.1 Případy víceslovných t-lemat

(Případy jsou popsány v MAN 3.3 Víceslovné t-lemma.)

- **zvrtná slova:** *se* jako součást t-lematu zvrtných slov (t-lemma: *slovo\_se*; například: *smát\_se*, *setkat\_se*, *chovající\_se*)
- **víceslovné spojky** typu *bud'nebo*, *a\_a*: vzhledem k tomu, že nebyly dosud zvoleny teoreticky vhodnější reprezentace souřadících spojek, bude t-lemma sestaveno z přítomných forem spojek ve větě, nebude se řešit tzv. reprezentativní t-lemma.
- **závislá část frazému**, t-lemma uzlů s funktorem DPHR typu *na\_lehkou váhu*.  
Pozor t-lemma je tu poskládáno z forem slov!
- **typ „van\_Beethoven“**
- **typ „číslo\_číslo“:** telefonní čísla, směrovací čísla.

Pokud v těchto případech v t-lematu nějaká část chybí, nejprve se správně oanotují odkazy do a-roviny, pak se opraví t-lemma - pomocí příslušných maker (makro: Alt+i, Alt+L, Alt+F) .

Pokud změnu v t-lematu nebylo možné provést pomocí maker, ale byla provedena ručně, označí se takový uzel výběrem poznámky *T-lemma with underscore*. Stačí vybrat typ poznámky.

#### 1.2 Zástupná t-lemata pro nealfanumerické znaky

Vyřeší se správná podoba t-lemat pro nealfanumerické znaky (vyjádřené v povrchové podobě věty), například: *#Colon* pro dvojtečku apod.

(MAN 3.4. Zástupná t-lemata)

### 1.3 #PersPron

Vyjádřená osobní, přivlastňovací a zvrtná zájmena budou mít přiřazeno t-lemma *#PersPron*.

### 1.4 Překlep

Překlep v t-lematu se ručně opraví. Změna v t-lematu se označí výběrem poznámky *Typo*.

### 1.5 Špatně analyzované základní slovní formy

Při případné špatné morfologické analýze základní slovní formy (například pro *je*-zájmeno se vygeneroval infinitiv *být*), se t-lemma ručně opraví. Změna v t-lematu se označí i výběrem poznámky *M-lemma*. Stačí vybrat typ poznámky.

### 1.6 T-lemata cizojazyčných výrazů

T-lematem cizojazyčných výrazů je morfologická forma, a to i v případě nealfanumerických znaků. Nesprávná t-lemata anotátor opraví pomocí makra (Alt+F). Nelze-li chybné t-lemma opravit pomocí makra, opraví je ručně a změnu t-lematu se označí výběrem poznámky *T-lemma*.

## 2 T-lemma u t1-uzlů (*is\_generated = 1*)

### 2.1 „Zkopírované“ t-uzly

Předpokládá se, že v první fázi anotace budou správně doplněny všechny kopírované uzly. Pozor na odkazy do a-roviny u kopírovaných uzlů!

### 2.2 Doplněné t-uzly se zástupnými t-lematy

V první fázi anotace bude správně určena pro t-rovinu konečná podoba následujících t-lemat:

<i>#AsMuch</i>	<i>#Forn</i>	<i>#Rcp</i>
<i>#Equal</i>	<i>#Idph</i>	<i>#Separ</i>
<i>#EmpNoun</i>	<i>#Neg</i>	<i>#Some</i>
<i>#EmpVerb</i>	<i>#Oblfm</i>	<i>#Total</i>

Pro anotovaná data platí, že všechna tato t-lemata jsou správně.

Anotátorem doplňovaný uzel pro elipsy obligatorních aktantů bude mít jednotné t-lemma: *#NewNode*. Znamená to, že anotátor neřeší, zda doplňovaný uzel má mít t-lemma: *#Cor*, *#QCor*, *#Unsp*, *#PersPron*, nebo *#Gen*. Přiřazení t-lematu *#NewNode* znamená, že daný t1-uzel může mít kterékoli z těchto t-lemat.

T-lemata (*#Cor*, *#QCor*, *#Unsp*, *#PersPron*, *#Gen*) u automaticky vygenerovaných uzlů, se nebudou měnit, ale ověřuje se, zda byl uzel z hlediska úplnosti valenčního rámce správně dogenerován a zda mu byl přidělen správný funktor. V případě nadbytečného vygenerování anotátor uzel smaže.

(V další fázi anotace (při anotaci koreference) se u všech t1-uzlů s t-lematy *#Cor*, *#QCor*, *#Unsp*, *#PersPron* a *#Gen* jejich t-lemma zkontroluje, případně opraví na správnou hodnotu. Též všechny uzly s t-lematem *#NewNode* dostanou konečnou podobu t-lematu.)

## Anotace valence u sloves

Dodatek k MAN 5.2 Valence

1	Krok 1: Kontrola přiřazeného valenčního rámce.....	12
1.1	F (frame): rámec ve vallexu chybí, žádný rámec není přiřazen.....	12
1.2	W (word): slovo ve vallexu chybí, žádný rámec není přiřazen.....	13
1.3	A (alternative): v zápisu formy chybí nějaká alternativa, rámec přiřazen.....	13
1.4	N (new): v rámci chybí neobligatorní aktant, rámec přiřazen.....	13
1.5	X: v alternativní skupině chybí funktor, rámec přiřazen.....	14
1.6	Ostatní případy.....	14
2	Krok 2: Kontrola naplnění valenčního rámce ve stromě.....	14

Valence se určuje u sloves, substantiv, adjektiv i adverbii, pokud valenci mají.

Anotace valence znamená:

1. stanovení a přiřazení valenčního rámce z valenčního slovníku (tj. je vyplněn atribut *val\_frame.rf*)
2. naplnění valenčního rámce v tektogramatickém stromě
  - uzly pro vyjádřená závislá doplnění mají správně vyplněný atribut *functor*
  - do tektogramatického stromu jsou doplněny uzly pro povrchově nevyjádřená obligatorní valenční doplnění

Valenční rámce pro slovesa byly předem ručně zpracovány a automatickou procedurou byly valenční rámce slovesům již také přiřazeny a též byly doplněny uzly pro v povrchové podobě nevyjádřená obligatorní doplnění.

### 1 Krok 1: Kontrola přiřazeného valenčního rámce

Anotátor valenční rámec pro sloveso nevytváří, pouze zkontroluje správnost automaticky vybraného rámce, je-li nesprávný, vybere příslušný správný valenční rámec z již vytvořených valenčních rámců, které ve slovníku jsou.

V žádném případě anotátor v PDT-vallexu nic nemění!

#### Co dělat pokud správný valenční rámec ve vallexu není?

Pokud správný valenční rámec v PDT-vallexu není, může to znamenat několik případů, které popisujeme dále.

#### 1.1 **F (frame): rámec ve vallexu chybí, žádný rámec není přiřazen**

Valenční rámec pro daný význam ve slovníku vůbec není.

Řešení:

Anotátor nepřihádá danému slovesu žádný rámec. (Atribut *val\_frame.rf* zůstane prázdný.) Do textu poznámky typu *New valency frame* zapíše povinně „F“ a návrh nového rámce co nejbližší formě, kterou valenční slovník používá, včetně příkladu (případně poznámky), kterým jednoznačně osvětlí význam. Pro DPHR postačí do závorky zapsat jen „text“ závislé části frazému.

Příklady:

*F ACT(.1) PAT(po+6) Nápoj chutná po třešních.*

*F ACT(.1) PAT(.4) ?ADDR(.3) DPHR(na míru) Šije mu úkoly na míru.*

## **1.2 W (word): slovo ve vallexu chybí, žádný rámeček není přiřazen**

Dané sloveso vůbec není ve valenčním slovníku.

Řešení:

Anotátor nepřidá danému slovesu žádný rámeček. (Atribut *val\_frame.rf* zůstane prázdný.) Do textu poznámky typu *New valency frame* zapíše povinně „W“ a návrh nového rámečku co nejbližší formě, kterou valenční slovník používá, včetně příkladu (případně poznámky), kterým jednoznačně osvětlí význam.

Příklad:

*W ACT(.1) PAT (.4) Pavel candal mléko na podlahu.*

## **1.3 A (alternative): v zápisu formy chybí nějaká alternativa, rámeček přiřazen**

Ve valenčním rámečku pro daný význam slovesa chybí zápis formy aktantu, kterou je příslušný aktant vyjádřen v anotované větě. K tomuto typu patří i případy, kdy ve výčtu jmen u složeného predikátu chybí v zápisu CPHR některé jméno.

Řešení:

Anotátor přiřadí slovesu příslušný rámeček s chybějící formou. Do textu poznámky typu *New valency frame* zapíše povinně písmeno „A“ a uvede pouze aktant, u kterého v přiřazeném rámečku chybí nějaká forma, v zápisu formy u tohoto aktantu uvede pouze chybějící formu. Může uvést i příklad.

Příklady:

*A PAT(zda[.v]) Váhá, zda začít.*

Zápis znamená, že v přiřazeném rámečku chybí u pacientu možnost vyjádřit jej závislou klauzí uvozenou podřadící spojkou *zda*.

*A CPHR({zátěž}) Kladl na něj přílišnou zátěž.*

Zápis znamená, že v přiřazeném rámečku chybí ve výčtu jmen pro jmennou část složeného predikátu jméno *zátěž*.

## **1.4 N (new): v rámci chybí neobligatorní aktant, rámeček přiřazen**

Do valenčního rámečku pro daný význam slovesa se ještě hodí další neobligatorní (!) aktant, který je v anotované větě vyjádřen.

Řešení:

Anotátor přiřadí slovesu příslušný rámeček s chybějícím neobligatorním (tj. fakultativním) aktantem. Do textu poznámky typu *New valency frame* zapíše povinně písmeno „N“ a uvede pouze zápis neobligatorního aktantu, který v přiřazeném rámečku chybí. V zápisu formy u tohoto aktantu uvede formu, která se vyskytla v anotované větě (případně jinou, která je evidentní). Vždy uvede i příklad.

Příklady:

*N ?EFF(na+4) Reorganizuje hypoteční sekci na samostatnou jednotku.*

Zápis znamená, že v přiřazeném rámečku (*ACT(.1) PAT(.4)*) chybí neobligatorní efekt s formou *na+4*.

*N ?PAT(z+2;na+6) Trh prosperuje ze špatných zpráv.*

Zápis znamená, že v přiřazeném rámečku (*ACT(.1)*) chybí neobligatorní PAT s formami *z+6,na+6*.

## 1.5 X: v alternativní skupině chybí funktor, rámec přiřazen

Do alternativní skupiny funktorů (oddělených v rámci svislítkem) ve valenčním rámci pro daný význam slovesa se ještě hodí další funktor, který je vyjádřen v anotované větě. Tento typ je i pro případy, kdy ve valenčním rámci žádná alternativní skupina není, ale chceme ji (dvoučlennou) vytvořit.

Řešení:

Anotátor přiřadí slovesu příslušný rámec s chybějícím funktořem v alternativní skupině. Do textu poznámky typu *New valency frame* zapíše povinně písmeno „X“ a zápis posledního z funktorů alternativní skupiny a za svislítko připojí zápis přidávaného funktoru, který v přiřazeném rámci chybí. V zápisu formy u tohoto funktoru uvede formu, která se vyskytla v anotované větě (případně jinou, která je evidentní). Vždy uvede i příklad.

Příklady:

*X CPR(\*)|MEANS(\*) Vidět svět japonskýma očima.*

Zápis znamená, že do alternativní skupiny v přiřazeném rámci (*ACT(.1) PAT(.4) MANN(\*)/ACMP(\*)/CPR(\*)*) se má doplnit funktor *MEANS(\*)*.

*X LOC(\*)|ACMP(\*) Akcie uzavřely s mírným poklesem.*

Zápis znamená, že ze slotu *LOC(\*)* v přiřazeném rámci (*ACT(.1) LOC(\*)*) se má vytvořit alternativní skupina: *LOC(\*)/ACMP(\*)*.

Obecné poznámky k typům 1.1 až 1.5:

U všech typů 1.1 až 1.5 může anotátor v poznámce typu *Vallex* poznamenat další komentář. Naopak do textu poznámky *New valency frame* se žádné komentáře nepiší, ta má přesně předepsanou formu, kterou je třeba dodržet!

Dávejte si pozor na formu zápisu, na správné zapsání funktoru a formy (zejména pak na překlapy typu *CHRP*). U každého funktoru je vždy v závorkách uvedena forma, z toho plyne, že příklad následuje vždy za zavírací závorkou!

## 1.6 Ostatní případy

- sense needs split in vallex
- word sense(s) hard to distinguish
- frame assigned, but doubts remain
- misspeling in examples
- wrong synonym/antonym/aspect counterpart
- others

Řešení: Anotátor přiřadí valenční rámec, který se mu jeví jako nejvíce vhodný. V textu poznámky typu *Vallex* popíše problém.

## 2 **Krok 2: Kontrola naplnění valenčního rámce ve stromě**

Anotátor zkontroluje přiřazení funktorů/aktantů u uzlů pro vyjádřená doplnění i uzlů pro v povrchové podobě věty nevyjádřená doplnění. U automaticky doplněných uzlů pro v povrchové podobě věty nevyjádřená doplnění nekontroluje t-lemma (více v dodatku *T-lemma*). Chybně vyplněné funktořy opraví, doplní chybějící uzly, smaže nadbytečné.

V případě, že uzlu nebyl přiřazen valenční rámec (výše popsany typ F a W), postupuje anotátor tak, jako by uzlu byl přiřazen rámec, který navrhl v poznámce *New valency frame*.



## Anotace valence u jiných slovních druhů

Dodatek ke kapitole MAN 5.2 Valence

1	Deverbativní substantiva na <i>-ní</i> a <i>-tí</i> .....	15
2	Ostatní substantiva .....	16
3	Adjektiva a adverbia .....	16

### Poznámka:

Valenční rámce pro substantiva, adjektiva a adverbia nejsou v PDT-vallexu doposud systematicky zpracovány. Jejich zpracování si vyžádá ještě hodně času a přemýšlení. V první fázi anotace PCEDT\_cz a PDTSC se proto valence jiných slovních druhů než sloves s výjimkou jejich substantivních derivátů na *-ní/-tí* a deverbativních adjektiv nebude anotovat. Tento dodatek obsahuje „prozatímní“ pravidla, jak zacházet se závislými doplněními „nesloves“ při první fázi anotace PCEDT\_cz a PDTSC. Instrukce vycházejí z toho, že valenční rámce pro substantiva na *-ní* a *-tí* jsou ve valenčním slovníku nejvíce zpracována a máme s nimi největší zkušenost.

### 1 Deverbativní substantiva na *-ní* a *-tí*

U substantiv na *-ní* a *-tí*:

- bude v anotátorské poznámce *New Valency Frame* označeno, zda anotátor chápe význam substantiva jako dějový (P), nebo nedějový (nP).
- v dějovém významu bude mít substantivum přiřazený valenční rámec, pokud anotátor najde ve valenčním slovníku odpovídající (správný) valenční rámec, v opačném případě nebude substantivu přiřazen žádný rámec.

**Anotace.** Anotátor nejprve vyhodnotí, zda význam substantiva je v daném kontextu dějový, či nikoli. Substantiva na *-ní* a *-tí* vyjadřují primárně děj! Předpokládáme, že dějový význam je u nich častější. Pokud anotátor váhá, dává přednost dějovému významu.

#### Význam substantiva je dějový:

Do anotátorské poznámky *New Valency frame* zapíše anotátor písmeno P (process). Komentář lze zapsat do poznámky *Vallex*.

Ve vallexu (pokud tam dané substantivum je) se pokusí vybrat valenční rámec, tj. rámec, který je odvozený od valenčního rámce odpovídajícího základového slovesa (rámce pro dějové významy substantiv mají (mívají) před příklady v závorce tato slovesa uvedena). Najde-li takový valenční rámec, přiřadí ho.

Ve všech ostatních případech (substantivum není ve vallexu, žádný z valenčních rámců není vhodný, s navrženým rámcem pro dějový význam anotátor nesouhlasí) nepřidělí anotátor substantivu žádný rámec. Odpovídající (správný) valenční rámec nikam nezapisuje (to děláme jen u sloves!).

Podle odpovídajícího valenčního rámce (přiřazeného nebo jen myšleného) přidělí funktoři v povrchové podobě věty vyjádřeným doplněním a doplní uzly pro obligatorní valenční doplnění v povrchové podobě věty nevyjádřená (pouze u těchto substantiv anotujeme také důsledně recipocitu).

Shodné adjektivní přívlastky se analyzují, nemají funktoři RSTR, ale odpovídající funktoři pro volná doplnění. Funktoři RSTR náleží shodným adjektivům, jen jde-li:

- o ukazovací zájmena, zdůrazňovací výrazy (*ten, on*)
- o deverbativní adjektiva zastupující závislou klauzi přívlastkovou (*soužití Romů trvajících. RSTR století*)

Vyhodnotí-li význam jako nedějový:

Do anotátorské poznámky *New Valency frame* zapíše zkratku nP (non-process). Komentář lze zapsat do poznámky *Vallex*.

## 2 Ostatní substantiva

U všech ostatních substantiv (včetně substantiv na *-ní* a *-tí*, která nemají dějový význam, tj. v poznámce *New Valency frame* mají zkratku nP) přiděluje anotátor funktoři jen v povrchové podobě věty vyjádřeným doplněním. Valenčním doplněním a všem genitivním doplněním přidělí funktoři RSTR! Do tektogramatického stromu se nedoplňují žádné uzly pro doplnění nevyjádřená v povrchové podobě věty, naopak všechny automaticky doplněné uzly se z tektogramatického stromu vymažou! Shodné adjektivní přívlasky se detailněji neanalyzují, mají funktoři RSTR. Analyzují se pouze volná adverbialní doplnění v předložkových pádech. Valenční rámec anotátor v žádném případě nepřirazuje. U těchto substantiv se také neanotuje reciprocita.

Příklady:

*dálkové ovládání televize*.RSTR (=ovladač)

*balení másla*.PAT *na lince* (děj) / *jedno*.RSTR *balení másla*.RSTR *na regále*.LOC

*myčka nádobí*.RSTR / *myčka na nádobí*.AIM

*obvaz rány*.RSTR *lékařem*.RSTR / *buničítý*.RSTR *obvaz na rány*.AIM

*struhadlo brambor*.RSTR / *struhadlo na brambory*.AIM

*můj*.RSTR *dopis mamince*.RSTR

*výroba tužek*.RSTR / *výrobek ze dřeva*.ORIG

*čekárna nemocných*.RSTR *na vyšetření*.AIM / *čekárna pro matky*.BEN *s dětmi*

*román o válce*.RSTR / *zákon o rodině*.RSTR

*návrh řešení*.RSTR

*žádost o změnu*.RSTR / *prosba o pomoc*.RSTR

*záměna holčičky*.RSTR *za chlapečka*.RSTR

*struktura firmy*.RSTR

*rekord sportovce*.RSTR

*náš*.RSTR *cíl zvítězit*.RSTR

*spor poslanců*.RSTR

*dohoda mezi oběma koncerny*.RSTR

*23*.RSTR *března 2008*.RSTR (ale: rok 1998.ID)

## 3 Adjektiva a adverbia

Závislá doplnění adjektiv analyzujeme, tj. přiřazujeme funktoři aktantů a volných doplnění (nikdy ne RSTR). Valenci těchto slovních druhů anotujeme bez valenčního slovníku, pouze intuitivně (v souladu s instrukcemi v manuálu). Tj. nepřirazují se valenční rámce (případně přiřazené valenční rámce se nekontrolují, neopravují). PDT-vallex je pouze inspirací.

Příklady:

*lidé čekající na vyšetření*.PAT

*muž pracující se dřevem*.PAT

*poslanec zvolený voliči*.ACT

*dopis došlý poštou*.MEANS

*košík plný hub*.PAT

*je blíž umění*.PAT

# Reciprocita

Nahrazuje kapitolu MAN 5.2.4.2 Reciprocita

1	Co je reciprocita? .....	17
1.1	Lexikální reciprocita .....	17
1.2	Syntaktická reciprocita .....	17
2	Reciprocita vyjádřená lexikálně.....	18
2.1	Inherentně reciproční slovesa .....	18
2.1.1	Vlastně inherentně reciproční slovesa.....	18
2.1.2	Inherentně reciproční varianty sloves (odvozená reciproční slovesa) .....	18
2.1.2.1	Odvozená reciproční slovesa od sloves s akuzativní vazbou.....	18
2.1.2.2	Odvozená reciproční slovesa od sloves s dativní vazbou .....	19
2.2	Lexikální ukazatele reciprocit.....	19
3	Reciprocita vyjádřená syntakticky .....	19
3.1	Formální ukazatele reciprocit .....	19
3.2	Zachycení syntaktické reciprocit v tektogramatickém stromě .....	20
4	Reciprocita u jiných slovních druhů .....	21

## 1 Co je reciprocita?

### Reciprocita (vzájemnost)

= vztah mezi dvěma valenčními doplněními (aktanty, popřípadě obligatorními volnými doplněními), která jsou postavena do symetrického vztahu tak, že vztah mezi nimi lze parafrázovat na základě modelu:

*Petr a Pavel se potkali. = Petr potkal Pavla a (zároveň) Pavel potkal Petra.*

Rozlišujeme reciprocitu lexikální a reciprocitu syntaktickou.

### 1.1 Lexikální reciprocita

= vztah reciprocit mezi dvěma doplněními vyplývá z lexikálního významu predikátu.

Vztah reciprocit mezi dvěma valenčními doplněními se u řady sloves předpokládá, je nutný pro vykonávání pojmenované činnosti. Taková slovesa označujeme jako **slovesa inherentně reciproční**.

Příklad: *Petr se setkal s Pavlem.*

### 1.2 Syntaktická reciprocita

= vztah reciprocit mezi dvěma doplněními je vyjádřen syntaktickými (formálními) prostředky. Nad valenčním rámcem predikátu je provedena syntaktická operace, v jejímž důsledku jsou dvě doplnění predikátu uvedena ve vztah reciprocit.

Tuto operaci nazýváme operaci **reciprocikalizace**. Doplnění, na které je operace uplatněna, nazýváme **reciprocikalizovaná doplnění** (případně **reciprocikalizované (valenční) pozice**).

Vstupní podmínky reciprocikalizace:

- lexikální homogenost doplnění, na které bude operace uplatněna.
- doplnění, na které je operace uplatněna, mají stejné aktuální členění.

Výstupní podmínky reciprocikalizace:

- nevyjádření jedné reciprocikalizované pozice (jednoho reciprocikalizovaného valenčního doplnění)
- plurál nebo koordinace v druhé - vyjádřené - reciprocikalizované pozici

V důsledku reciprocikalizace se v povrchové podobě věty ztrácí (nevyjadřuje se) pozice jednoho z reciprocikalizovaných doplnění. V druhé reciprocikalizované pozici jsou obě doplnění v recipročním vztahu vyjádřena najednou, tj. koordinací nebo plurálem.

Syntaktickou operaci reciprocikalizace lze uplatnit jak na slovesa inherentně reciproční, tak na slovesa, která tuto charakteristiku nemají.

Příklady:

Vstup: *Jan se líbá s Marií.*

Výstup: *Jan a Marie se líbají.*

Vstup: *Jan líbá Mariinu ruku.*

Výstup: není splněna podmínka homogenosti doplnění, reciprocikalizaci nelze provést.

## 2 Reciprocita vyjádřená lexikálně

### 2.1 Inherentně reciproční slovesa

Inherentně reciproční slovesa jsou slovesa, u kterých se vztah reciprocity mezi dvěma jeho valenčními doplněními předpokládá. Reciproční význam je však pouze potenciální, kontextem může být vždy zrušen.

Příklad:

*Petr se setkal s Pavlem. - Petr se setkal s nezájmem.*

Inherentně reciproční slovesa dělíme na vlastně inherentně reciproční slovesa a na inherentně reciproční varianty slovesa (odvozená reciproční slovesa).

#### 2.1.1 Vlastně inherentně reciproční slovesa

= neodvozená (primární) slovesa s recipročním významem.

Patří sem:

- **reciproční reflexiva tantum:** *hádat\_se, prát\_se, setkat\_se*, tj. slovesa, u nichž je výraz *se* (slovotvorný formant s významem reciprocity) součástí lemmatu slovesa (naznačeno podtržítkem) a slovesa nelze užít bez tohoto výrazu.

- jiná slovesa. Inherentní význam reciprocity je přítomný i ve významu jiných sloves: *zápasit, obchodovat, diskutovat, polemizovat, soupeřit, soutěžit, sousedit.*

#### 2.1.2 Inherentně reciproční varianty sloves (odvozená reciproční slovesa)

= slovesa s recipročním významem odvozená od sloves s akuzativní nebo dativní vazbou pomocí slovotvorného formantu *se* nebo *si*.

Při odvozování inherentně recipročních variant sloves jde o proces analogický odvozování reflexivních variant sloves (i zde dochází ke změně valenčního rámce, připojení formantu *se* k lemmatu slovesa).

Příklad: *roztrhnout – roztrhnout se*

Valenční rámec slovesa *roztrhnout*: ACT(.1) PAT(.4) *Někdo roztrhl pytlík.*

Valenční rámec odvozeného reflexiva *roztrhnout\_se*: ACT(1) *Pytlík se roztrhl.*

##### 2.1.2.1 Odvozená reciproční slovesa od sloves s akuzativní vazbou

U sloves s doplněními v nominativu a akuzativu lze vztah reciprocity mezi těmito doplněními vyjádřit i lexikálně – odvozením inherentně reciproční varianty původního slovesa, a to následujícím způsobem:

- k základu původního slovesa se připojí slovotvorný formant *se*,
- akuzativní vazba slovesa se nahradí vazbou *s+7*.

### Schematicky:

nominativ	sloveso	akuzativ
↓	↓	↓
nominativ	sloveso_ <i>se</i>	s+instrumentál

### Příklad: *líbat* ⇒ *líbat\_se*

Jan	líbá	Marii.
↓	↓	↓
Jan	se_líbá	s Marií.

Valenční rámec slovesa *líbat*: ACT(.1) PAT(.4)

Valenční rámec odvozeného reciproka *líbat\_se*: ACT(1) PAT (s+7)

Příklady dalších odvozených recipročních sloves: *objímat\_se*, *potkat\_se*, *navštěvovat\_se*, pokud se vyskytne (tj. v datech), tak i: *fotografovat\_se*, *popisovat\_se*, *vidět\_se*, *lčít\_se*.

### **2.1.2.2 Odvozená reciproční slovesa od sloves s dativní vazbou**

U sloves s doplněními v nominativní a dativní a akuzativní formě lze vztah reciprocit mezi těmito doplněními vyjádřit i lexikálně – odvozením inherentně reciproční varianty původního slovesa, a to následujícím způsobem:

- k základu původního slovesa se připojí slovotvorný formant *si*,
- dativní vazba slovesa se nahradí vazbou s+7.

### Schematicky:

nominativ	sloveso	dativ
↓	↓	↓
nominativ	sloveso_ <i>si</i>	s+instrumentál

### Příklad: *telefonovat* ⇒ *telefonovat\_si*

Jan	telefonuje	Marii.
↓	↓	↓
Jan	si_telefonuje	s Marií.

Valenční rámec slovesa *telefonovat*: ACT(.1) PAT(.4/o + 6) ADDR(.3)

Valenční rámec odvozeného reciproka *telefonovat\_si*: ACT(1) PAT(.4/o + 6) ADDR(s+7)

Příklady dalších odvozených recipročních sloves tohoto typu: *telefonovat\_si*, *konkurovat\_si*, *slíbit\_si*, *věřit\_si*, *blahopřát\_si*, *volat\_si*, *sdělovat\_si*, *pomáhat\_si*.

## **2.2 Lexikální ukazatele reciprocit**

Lexikálně se reciprocita vyjadřuje také výrazy (adverbii) vyjadřujícími vzájemnost: *spolu*, *společně*, *vzájemně*, *navzájem*. Tyto výrazy slouží zejména ke zdůraznění recipročního významu při reciprocitě syntaktické.

## **3 Reciprocita vyjádřená syntakticky**

### **3.1 Formální ukazatele reciprocit**

Provedení operace reciprocikalizace má vždy dva základní – povinné - formální ukazatele. V různých případech přibývají k těmto formálním důsledkům, formální ukazatele další.

### Základní formální ukazatele syntaktické reciprocity:

= nastávají ve větě vždy při uplatnění operace reciprocikalizace.

- A. nerealizování jedné reciprocikalizované valenční pozice.
- B. druhá reciprocikalizovaná pozice se formálně realizuje jako:
  - koordinační (i hypotaktické) spojení (*Petr a Pavel*, *Petr s Pavlem*)
  - formální či jen sémantický plurál jména (*kamarádi*, *dvojice*)
  - formou *mezi*+instrumentál (*mezi Petrem a Pavlem*)

### Specifické formální ukazatele syntaktické reciprocity:

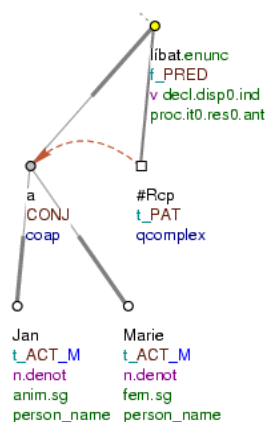
= nejsou ve větě přítomny při každém uplatnění operace reciprocikalizace, ale nastávají s různou mírou pravidelnosti/nutnosti při různých typech reciprocikalizace.

- C. výraz *se* v příslušném tvaru.  
*Se* (pokud je v silném tvaru) je v pádě (i předložkovém) nevyjádřeného reciprocikalizovaného doplnění. Má-li nevyjádřená reciprocikalizovaná pozice formu *s+7*, forma *s sebou* se neuvžívá; může být: *mezi sebou* (Ale: *Jan a Marie se sebou nesouhlasili*). Vylučuje se s formálním ukazatelem D.
- D. výraz *jeden druhý* v příslušném tvaru.  
*Jeden* je v pádě (i předložkovém) vyjádřeného reciprocikalizovaného doplnění, *druhý* je v pádě (i předložkovém) nevyjádřeného reciprocikalizovaného doplnění. Vylučuje se s formálním ukazatelem C.

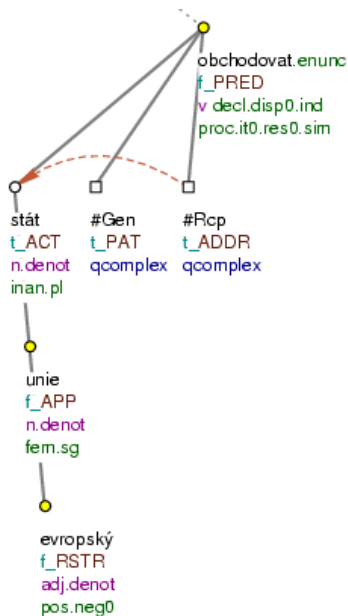
**Pozor!** U recipročních reflexiv tantum a u odvozených recipročních sloves (u recipročních sloves s formantem *se*, *si* v lemmatu) nebývají při reciprocikalizaci kromě základních ukazatelů A a B nutné žádné další formální ukazatele.

## **3.2 Zachycení syntaktické reciprocity v tektogramatickém stromě**

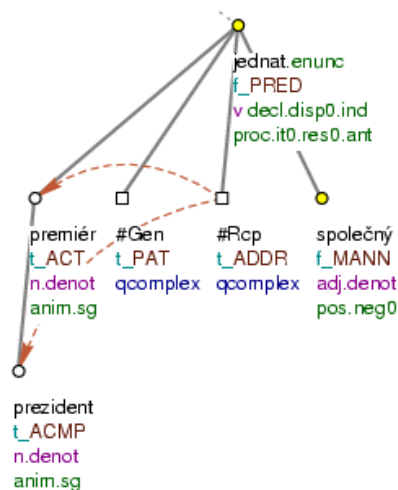
Reciprocitu zachycujeme pomocí nově vytvořeného t-uzlu se zástupným t-lematem *#Rcp*, který je do tektogramatického stromu doplněn na pozici reciprocikalizovaného valenčního doplnění. Doplněný t-uzel (*#Rcp*) má funktor, který této chybějící pozici podle valenčního rámce odpovídá. Mezi doplněným t-uzlem a mezi t-uzlem ve vyjádřené reciprocikalizované pozici je zachycen vztah gramatické koreference. Formální ukazatele reciprocity typu C a D (viz 3.1) nejsou v tektogramatickém stromě reprezentovány samostatným t-uzlem, ale odkazy na odpovídající a-uzly jsou uloženy v atributu *a/aux.rf* u doplněného t-uzlu s t-lematem *#Rcp*. Lexikální ukazatele reciprocity (viz 2.2) jsou naproti tomu reprezentovány samostatnými t-uzly, zpravidla s funktořem MANN.



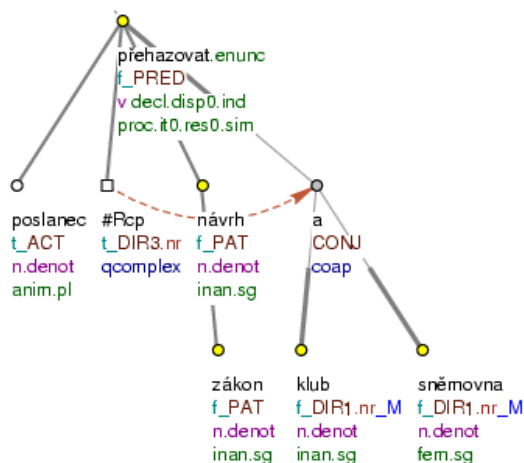
Obr.1: *Jan a Marie libají <jeden druhého>*.



Obr. 2: *Státy Evropské unie <mezi sebou> obchodují*



Obr.3: *Premiér s prezidentem spolu jednali.*



Obr. 4: *Poslanci přehazují návrh zákona mezi klubem a sněmovnou <z jednoho do druhého>.*

#### 4 Reciprocita u jiných slovních druhů

K procesu reciprocikalizace dochází i u (deverbativních) substantiv a adjektiv, srov.:

*boj nepřátelených stran mezi sebou*

*lidé bojující mezi sebou navzájem.*

Reciprocita u jiných slovních druhů než sloves bude v tektogramatických stromech prozatím zachycena důsledně jen:

- u deverbativních substantiv na *-ní* a *-tí*, která mají dějový význam!
- u adjektiv, u kterých je reciprocita vyjádřena některým ze specifických formálních ukazatelů C a D (*mezi sebou, jeden druhý*), tj. kde je třeba něco skrýt!

Příklady:

*jednání Petra.ACT a Pavla.ACT o prodeji domu #Rcp.ADDR*  
*jednání států.ACT <mezi sebou> #Rcp.ADDR*  
*jejich.ACT společně.MANN jednání se protáhlo #Rcp.ADDR*  
*setkání dvou prezidentů.ACT #Rcp.PAT*  
*srovnání dvou nesourodých trhů.PAT #Rcp.EFF*  
*obchodování s ropou mezi státy.ACT #Rcp.ADDR*  
*lidé bojující <mezi sebou> navzájem.MANN #Rcp.ADDR*

V ostatních případech nebude reciprocita zachycena. U ostatních deverbativních substantiv, u kterých je reciprocita vyjádřena některým ze specifických formálních ukazatelů C a D (*mezi sebou, jeden druhý*), nebude reciprocita zachycena. Formální ukazatele reciprocity C a D budou mít uzel ve stromě, ohodnocený funktorem MANN, COMPL.

Příklady:

*boj zneprátelených stran.RSTR mezi sebou.MANN*  
*diskuze mezi čtenáři.RSTR a spisovateli.RSTR*  
*dohoda mezi zastupiteli.RSTR*  
*obchod s ropu.RSTR mezi státy.RSTR*  
*vztah Petra.RSTR a Pavla.RSTR*  
*vzájemný.RSTR vztah Petra.RSTR a Pavla.RSTR*  
*spor poslanců.RSTR*



## Elipsy

Dodatek ke kapitole MAN 5.12 Elipsy

- 1 Aktuální elipsa řídicího substantiva u členských souřadných spojení .....23
- 2 Aktuální elipsa neobligatorního doplnění v souřadném spojení.....24
- 3 Elipsa řídicího substantiva k předložce „*důvody pro*“ .....25

### 1 Aktuální elipsa řídicího substantiva u členských souřadných spojení

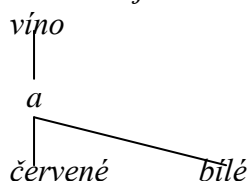
#### Změna! (MAN 5.12.1.2 Elipsa řídicího substantiva)

Aktuální elipsa řídicího substantiva u členských souřadných spojení se nově na tektogramatické rovině neřeší! Platí instrukce: koordinace je co nejnižže ve struktuře! Nedoplňuje se žádný nový kopírovaný uzel!

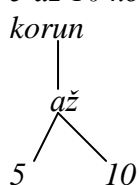
#### 1.1 Typ „červené a bílé víno“

Příklady:

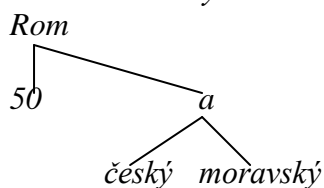
Obr. 1: *Pije červené a bílé víno.*



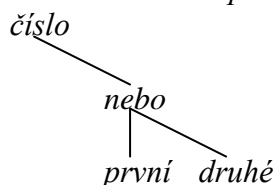
*5 až 10 korun*



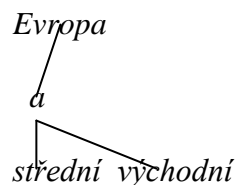
Obr. 3: *50 českých a moravských Romů*



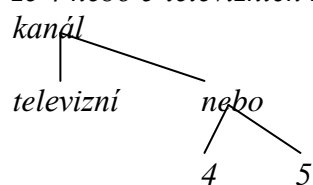
Obr. 5: *volali na první nebo druhé číslo*



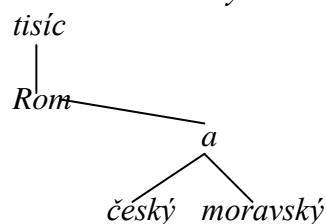
Obr. 2: *střední a východní Evropa*



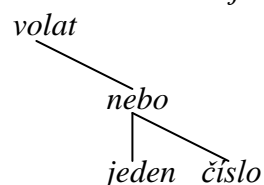
*ze 4 nebo 5 televizních kanálů*



Obr. 4: *tisíc českých a moravských Romů*



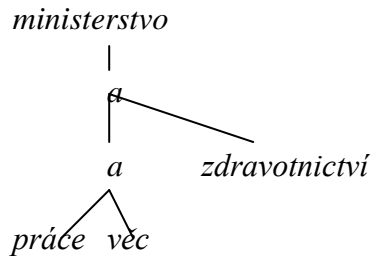
Obr. 6: *volali na jedno nebo druhé číslo*



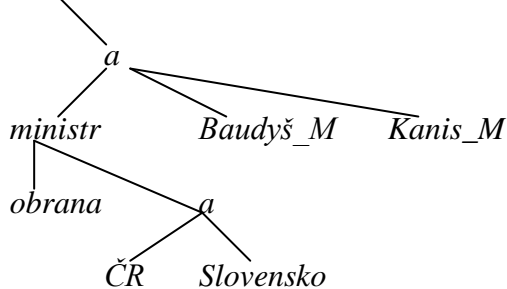
## 1.2 Typ „ministerstvo práce a sociálních věcí a zdravotnictví“

Příklady:

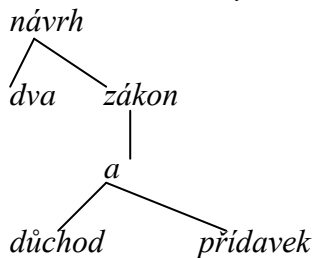
Obr. 7: *ministerstvo práce a sociálních věcí a zdravotnictví*



Obr. 8: *Dohodu podepsali ministři obrany ČR a Slovenska Antonín Baudyš a Pavol Kanis podepsat*



Obr. 9: *dva návrhy zákonů o důchodech a přídavcích na děti*



**Výjimka:** Výjimkou jsou pouze **koordinované názvy**, kde by se při zachycení bez elipsy ztratila informace o tom, zda se jedná o jeden název (s koordinací) nebo více (koordinovaných) názvů. Platí, že nelze koordinovat dva uzly s funktorem ID, jedná-li se o efektivní kořeny dvou různých názvů.

Příklady:

*společnosti Toyota.ID a [společnost] Fuji.ID*

*pohádka Sněhurka.ID a [pohádka] Zlatovláska.ID*

*litr benzínu Super.ID a [benzínu] Speciál.ID*

*Ale: pohádka Bolek.ID a [0] Lolek.ID*

## 2 Aktuální elipsa neobligatorního doplnění v souřadném spojení

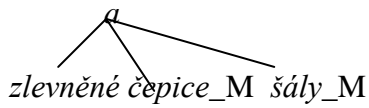
**Změna!** (MAN 5.12.3 Elipsa a princip společného rozvití)

Aktuální elipsa neobligatorních doplnění se pomocí „principu společného rozvití“ bude v souřadných spojeních nadále zachycovat! Nadále platí, že neobligatorní doplnění zachycujeme jako společné rozvití jen v sémanticky zcela jednoznačných případech!

Případy aktuální elipsy neobligatorního doplnění u souřadných spojení je v drtivé většině případů možné zachytit jako společné rozvití. Poté, co neanotujeme valenci substantiv, nevzniká problém ani u typu „*sekretářka, ale zakladatelka firmy*“ (viz obr. 11 a 12).

Příklady:

Obr. 10: *zlevněné čepice a šály*



Obr. 11: *sekretářka, ale zakladatelka firmy*



Obr. 12: *reportér a vydavatel deníku*



### 3 Elipsa řídicího substantiva k předložce „důvody pro“

Níže uvedené případy interpretujeme jako elipsu řídicího substantiva, a na chybějící pozici proto doplňujeme buď zástupný t-uzel s t-lematem #EmpNoun (gramatická elipsa), nebo kopírovaný uzel (aktuální elipsa).

Příklady:

*Všichni byli pro.*

*Důvody pro a proti.*

*Hlasoval proti.*

## Zvláštní případy souřadnosti

Dodatek ke kapitole MAN 5.6.2.1 Specifické konstrukce zachycené jako koordinační nebo apoziční struktura

- 1 Aponovaný člen v závorce .....26
- 2 Apoziční spojení se spojkou „jako“ .....27

### 1 Aponovaný člen v závorce

Za aponovaný člen považujeme i obsah závorky, je-li v ní uvedeno jiným způsobem naprosto totéž, co je uvedeno před závorkou (mezi oba aponované členy lze tudíž vložit výraz „to jest“). Jde o následující případy 1.1 1.2 a 1.3.

**Anotace.** Jako kořen souřadné struktury bude zachycena levá závorka (*#Bracket.APPS\_P*). Výrazy obsažené v závorce i samotný kořen souřadné struktury budou mít *is\_parenthesis = 1*.

#### 1.1 Jmenná fráze ve stejném pádě jako jmenná fráze před závorkou

Příklady:

*Poslanec Byron Dorgan (demokrat z Severní Dakoty) předložil v Kongresu zákon podporovaný také poslancem Lee Hamiltonem (demokratem z Indiany).*

*..už od dnů Alfreda Mahana (důstojníka amerického námořnictva a námořního historika)*

*Námořnictvo má své vlastní letecké síly (flotilu letadlových lodí) a svou vlastní armádu (námořní pěchotu).*

#### 1.2 Zkratka a její přesné rozepsání (a naopak)

Příklady:

*ODS (Občanská demokratická strana)*

*Občanská demokratická strana (ODS)*

*Komerční banka (KB)*

#### 1.3 Převody jednotek

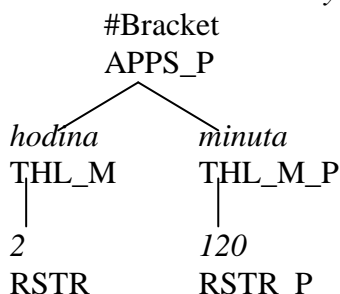
Příklady:

*2 hodiny (120 minut)*

*4 dolary (100 korun)*

*Trvalo to 2 hodiny (120 minut). (Obr.1)*

Obr.1: *Trvalo to 2 hodiny (120 minut).*



## 2 Apoziční spojení se spojkou „jako“

**Typ:** „Váže těžké kovy, jako je plutonium.“

Uzel pro sloveso *být* má stejný funktor jako uzel pro neslovesně vyjádřený terminální člen apozice (pro substantivum).

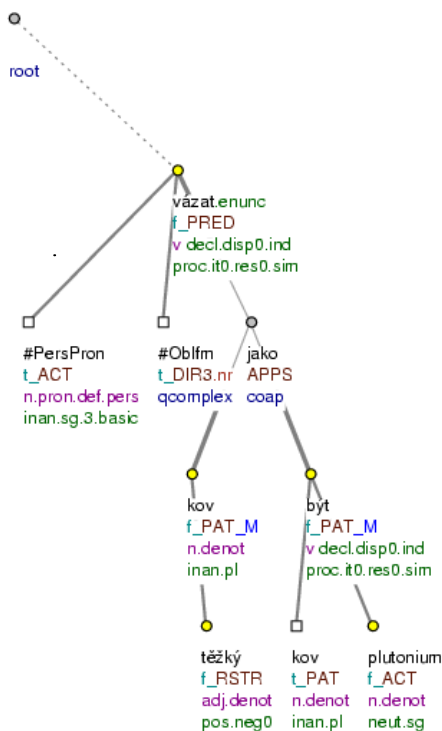
**Výjimka:** Pokud neslovesně vyjádřený terminální člen apozice (substantivum) má funktor DENOM, má uzel reprezentující sloveso *být* funktor PRED.

Příklady:

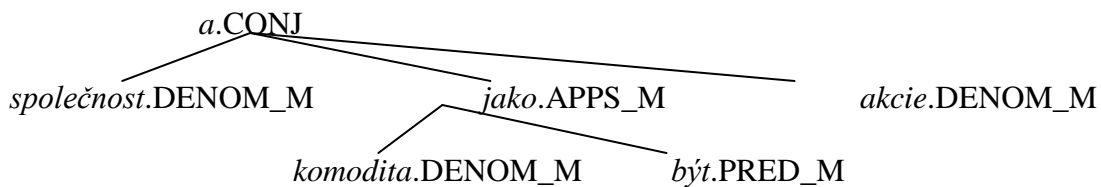
*Váže těžké kovy.PAT\_M, jako je.PAT\_M plutonium.* (Obr. 2)

*Počítačové společnosti, komodity.DENOM\_M podléhající ekonomickým výkyvům, jako jsou.PRED\_M třeba auta, a akcie maloobchodních firem.* (Obr. 3)

Obr.2: *Váže těžké kovy, jako je plutonium.*



Obr.3: *Počítačové společnosti, komodity podléhající ekonomickým výkyvům, jako jsou třeba auta, a akcie maloobchodních firem.*



## Syntaktické frazeologismy

Dodatek ke kapitole MAN 5.8. Frazémy

V manuálu v kap. 5.8. Frazémy je popsáno zachycení frazémů jakožto lexikálních jednotek. Pro zachycování syntaktických frazeologismů (specifické/neobvyklé syntaktické konstrukce pro vyjádření jistého významu, které mohou být různě lexikálně obsazeny) nejsou v manuálu přijata žádná zvláštní pravidla.

Zachycení syntaktických frazeologismů (ale i lexikálních, zejména „jednouzlových“) je tedy třeba ještě vyřešit. V tomto dodatku přinášíme pouze návody na zachycení některých syntaktických frazeologismů podle dostupných možností tektogramatické reprezentace. To, že jde o frazeologickou konstrukci (vyjadřující daný význam až jako celek) však v anotovaných datech není nijak zachyceno.

### ROK CO ROK

měsíc co měsíc, pátek co pátek

*"Rok co rok se budou setkávat s týž soudcem."*

rok co rok = „každý rok“

rok.TWHEN, na tom: rok <aux:co>.THO

### DEN PO DNI

měsíc po měsíci, týden po týdnu, minutu po minutě

*"prodlužování kontraktu den po dni"*

Je to nejspíš o význam "jak často", "v jakých intervalech", znamená to "denně, po jednotlivých dnech, každý den". Tedy:

den.THO po dni.TWHEN

### KUS PO KUSU

balík po balíku, úplatek po úplatku

*"Prodat něco kus po kusu."*

*„Úplatek po úplatku nás vedou po cestě, kterou společnost vyšlapala.“*

Dáváme přednost MANN, před TWHEN:

kus.MANN po kusu.TWHEN

### DEN ODE DNE, PŘÍPAD OD PŘÍPADU

*"justice případ od případu rozšiřuje zákon o zneužívání"*

den.TWHEN, ode dne.THO

Pozor!

ČAS OD ČASU: čas.THO, od\_času.DPHR

BOK PO BOKU: bok.MANN po\_boku.DPHR

RUKU V RUCI: ruku.MANN v\_ruce.DPHR

Ale nejde též o klasické lexikální frazémy.

## Měrové významy

Dodatek k MAN 6.13. Přesnější specifikace významu funktoru

*odliv byl 15 %.PAT*

*odliv byl pod 15 %.PAT*

*odliv byl nad 15 %.PAT*

*odliv byl okolo 15 %.PAT*

*odliv byl přes 15 %.PAT*

Při vyjadřování nepřesné, ale jen přibližné míry/počtu, nebo pokud se míra/počet vyjadřuje (ne/uzavřeným) intervalem, se pravidelně mění obvyklá forma daného doplnění (valenčního i nevalenčního). Teorie předpokládá, že tyto různé odstíny míry se zachytí subfunktory, problém mimo jiné je v tom, že tento jev náleží více funktorům (zatímco subfunktory by měly být specifické vždy jen pro jeden funktor). Typický je tento jev pro funktor EXT (obecný nevalenční význam „jak hodně, kolik“), pro funktor THL (měrový časový funktor s významem „jak dlouho“; měrový prostorový funktor s významem „jak daleko“ neexistuje), ale setkáme se s ním i u aktantů a funktoru RSTR (a možná i jinde). Pracovní přehled viz tabulka na konci tohoto textu.

### Seznam měrových předložek:

*kolem+2*

*do+2*

*na+4*

*okolo+2*

*pod+4*

*po+6*

*přes+4*

*nad+4*

*interval od\_do.OPER*

*od+2*

*k+3*

Prozatímní anotace (do doby přehodnocení funktorů pro vyjadřování významu míry a stanovení způsobu, jakým na tektogramatické rovině míru zachycovat):

**Funktor.** Doplnění nesoucí přídatný význam míry (k významu časovému, prostorovému, k významu patientu atp.) dostane funktor, který mu z nabídky funktorů (a podle současných pravidel přiřazování funktorů) nejlépe náleží.

**Odkazy *a/aux.rf* na měrové předložky.** Měrové předložky by měly být odkázány v atributu *a/aux.rf* u toho uzlu, kterému bude měrový subfunktor nakonec náležet. Na základě toho navrhuje několik pravidel:

(i) Při spojení číslovky a počítaného předmětu náleží odkaz na měrovou předložku číslovce (měrová předložka je v *a/aux.rf* u číslovky, nikoli u počítaného předmětu).

Příklady:

*vydělává 30 000.RSTR korun.PAT*

*vydělává <do> 30 000.RSTR korun.PAT*

*vydělává <od> 30 000.RSTR korun.PAT*

*vydělává <přes> 30 000.RSTR korun.PAT*

*vydělává <pod> 30 000.RSTR korun.PAT*

*vydělává od 30 000.RSTR\_M do 40 000.RSTR\_M korun.PAT (*od\_do.OPER*)*

*Je mu <přes> 40.RSTR let.PAT*

*Byl u společnosti <přes> 20.RSTR let.THL*

*Ani <kolem> deseti.RSTR miliard.ACT nebude stačit.*

*Kupónová sazba ještě nebyla pevně určena, ale pravděpodobně bude stanovena <kolem> 8.RSTR procent.EFF*

(ii) V jiných (složitějších) případech je odkaz na měrovou předložku u řídicího uzlu doplnění.

Příklady:

*poklesnout <pod> hranici.PAT 2,90 marky*

*poklesnout <k> hranici.PAT 2,90 marky*

*Dostal <po> jabličku.PAT*

**Valenční rámec.** Měrové předložky patří k pravidelným změnám v povrchově-syntaktické realizaci, které se nezaznamenávají v případě valenčních doplnění do valenčního rámce.

Příklady:

*Ani <kolem> deseti.RSTR miliard.ACT nebude stačit. [v PDT-vallexu: ACT(.1)]*

*Dostal <po> jabličku.PAT [v PDT-vallexu: PAT(.4)]*

*Kupónová sazba ještě nebyla pevně určena, ale pravděpodobně bude stanovena <kolem> 8.RSTR procent.EFF [ve PDT-vallexu: EFF(na+4)]*

*poklesnout <pod> hranici.PAT 2,90 marky [v PDT-vallexu: PAT(na+4)]*

*nabídnout <nad> 15.RSTR dolarů.PAT [v PDT-vallexu: PAT(.4)]*

			„jak hodně“	„jak dlouho“	„jak daleko“
<b>0</b>	„přesně“		<i>Zaplatil pět korun.</i>	<i>Trvalo to dvě hodiny.</i>	<i>Bydlí tři kilometry odtud.</i>
<b>approx</b>	„přibližně“	<i>kolem+2, okolo+2, asi, přibližně</i>	<i>Zaplatil kolem pěti korun.</i>	<i>Trvalo to kolem dvou hodin.</i>	<i>Bydlí kolem tří kilometrů odtud.</i>
<b>almost</b>	„téměř“	<i>téměř, skoro, bezmála, málem</i>	<i>Zaplatil téměř pět korun.</i>	<i>Trvalo to téměř dvě hodiny.</i>	<i>Bydlí téměř tři kilometry odtud.</i>
<b>less</b>	„méně než“	<i>pod+4, méně než</i>	<i>Zaplatil pod pět korun.</i>	<i>Trvalo to pod dvě hodiny.</i>	<i>Bydlí pod tři kilometry odtud.</i>
<b>more</b>	„více než“	<i>přes+4, nad+4, více než</i>	<i>Zaplatil přes pět korun.</i>	<i>Trvalo to přes dvě hodiny.</i>	<i>Bydlí přes tři kilometry odtud.</i>
<b>lessincl</b>	„méně než včetně“	<i>do+2, maximálně</i>	<i>Zaplatil do pěti korun.</i>	<i>Trvalo to do dvou hodin.</i>	<i>Bydlí do tří kilometrů odtud.</i>
<b>moreincl</b>	„více než včetně“	<i>od+2, minimálně</i>	<i>Zaplatil od pěti korun.</i>	<i>Trvalo to od dvou hodin.</i>	<i>Bydlí od tří kilometrů odtud.</i>

Tab. Vyjadřování měrových významů



## Výrazy *se* a *si*

Dodatek ke kapitole MAN 7.14. Výrazy "se" a "si"

1	Slovotvorný formant.....	31
1.1	Zvratná slovesa (reflexiva/reciproka tantum) .....	31
1.2	Odvozená reflexivní slovesa .....	31
1.3	Odvozená reciproční slovesa .....	32
1.4	Odvozená reflexivně/reciproční slovesa.....	32
1.5	Způsoby slovesného děje.....	32
2	Syntaktický formální prostředek.....	32
2.1	Pasivizace.....	32
2.2	Dispoziční modalita.....	33
3	Větný člen.....	33
3.1	Reflexivita.....	33
3.2	Reciprocita .....	33

Výrazy *se* a *si* mohou mít ve větě různé významy a funkce a v závislosti na těchto významech a funkcích se liší i jejich anotace.

### 1 Slovotvorný formant

Výraz *se/si* je nemotivovanou nebo motivovanou součástí lemmatu slovesa.

**Anotace.** Viceslovné t-lemma slovesa: *sloveso\_se* nebo *sloveso\_si*; výraz *se/si* v *a/aux.rf*.

#### 1.1 Zvratná slovesa (reflexiva/reciproka tantum)

Výraz *se/si* je nemotivovanou součástí slovesa. Slovesa nelze užít bez těchto výrazů. Výraz *se/si* tu nelze nahradit *sebe/sobě*, tj. nejde o reflexivitu.

Příklady: *bát\_se, usmát\_se, snažit\_se, stěžovat\_si, odpočinout\_si, setkat\_se, přát\_se, hádat\_se, utkat\_se, přít\_se, scházet\_se, loučit\_se, domlouvat\_se*

#### 1.2 Odvozená reflexivní slovesa

V daném významu *si* nelze u těchto sloves představit externího původce děje, slovesa vyjadřují bezděčné děje konané na sobě sama. Slovesa lze užít bez výrazů *se/si*, pak ale nemají význam bezděčného děje.

Srov.: *Vlny se šíří prostorem. (šířit\_se) x Rozhlas šíří vlny prostorem. (šířit)*

Příklady: (přírodní děje, děje na neživých předmětech): *zlomit\_se, roztrhnout\_se, rozsypat\_se, šířit\_se, rozlít\_se, rozpadnout\_se, naplnit\_se*

*Větev se zlomila.*

*Pytel se roztrhl*

*Voda se vylila z břehů..*

*Sál se naplnil lidmi.*

Příklady: (děje na člověku, avšak, člověkem nezaviněné (nejde o děje konané člověkem na sobě sama, tj. o reflexivitu): *trápit\_se, zabít\_se, utopit se, zranit\_se*

*Spadl do vody a utopil se.*

*Celý rok se trápil.*

### 1.3 Odvozená reciproční slovesa

Slovesa s recipročním významem odvozená pomocí *se/si* od sloves s akuzativní (*se*) nebo dativní vazbou (*si*). Namísto akuzativu/dativu nastupuje vazba *s+7*.

Derivace u sloves s akuzativní vazbou:

někdo líbá někoho → někdo se líbá s někým

*líbat* ACT(.1) PAT(.4) → *líbat\_se* ACT(.1) PAT(s+7)

Příklady: *líbat\_se, objímat\_se, potkat\_se, pozdravit\_se, vítat\_se, navštěvovat\_se, fotografovat\_se, popisovat\_se, vidět\_se, léčit\_se, natřít\_se.*

Derivace u sloves dativní vazbou:

někdo telefonuje něco někomu → někdo si telefonuje něco s někým

*telefonovat* ACT(.1) PAT(.4) ADDR(3) → *telefonovat\_si* ACT(.1) PAT(.4) ADDR(s+7)

Příklady: *telefonovat\_si, konkurovat\_si, slíbit\_si, věřit\_si, blahopřát\_si, volat\_si, sdělovat\_si, pomáhat\_si*

### 1.4 Odvozená reflexivně/reciproční slovesa

Slovesa s reflexivně/recipročním významem odvozená pomocí *se/si* od sloves s akuzativní vazbou (*se*). Akuzativní valenční pozice je zrušena. U sloves s dativní vazbou nebylo podobné chování zjištěno.

Derivace:

seznámit někoho s někým → „seznámit sebe s někým“ → seznámit se s někým

*seznámit* ACT(.1) ADDR(.4) PAT(s+7) → *seznámit\_se* ACT(.1) PAT(s+7)

Příklady: *seznámit\_se, spojovat\_se, lišit\_se, zaměřit\_se*

### 1.5 Způsoby slovesného děje.

Výrazy *se/si* se uplatňují často ve spojení s dalšími prostředky (s předponami) při utváření specifických významů sloves (způsobů slovesného děje).

Příklady: *natahat\_se, napracovat\_se, zaposlouchat\_se, dočíst\_se, rozepsat\_se, upít\_se, vyběhat\_se, pospat\_si, zalyžovat\_si, zatančit\_si*

## 2 Syntaktický formální prostředek

Výraz *se/si* jako prostředek různých syntaktických operací/transformací nad valenčním rámcem slovesa.

**Anotace.** Výraz *se* v *a/aux.rf* u t-uzlu reprezentujícího sloveso.

### 2.1 Pasivizace

Zevšeobecnění aktora (původce děje).

**Anotace.** #Gen.ACT na pozici aktora; výraz *se* v *a/aux.rf* u t-uzlu reprezentujícího sloveso.”

Příklady:

*Přirozený jazyk se popisuje formálními prostředky.*

*Tancovalo se až do rána.*

*Diskutovalo se o novém objevu.*

## 2.2 Dispoziční diateze

Zvláštní druh modality, který postihuje vztah/postoj agenta k ději. V češtině je nesena specifickým typem konstrukcí: reflexivní forma pasiva, adverbiale způsobu, dativní tvaru s funkcí agentu.

**Anotace.** Dativ má funktor ACT.

Příklady:

*Přirozený jazyk se mi.ACT nepopisuje nejlíp.*

*Matematika se mu.ACT studuje dobře.*

*V téhle troubě se mi.ACT dobře peče.*

## 3 Větný člen

Výraz *se/si* je samostatnou významovou jednotkou, která vyjadřuje totožnost referenta dané pozice s jinou pozicí, a to v reflexivních nebo recipročních konstrukcích.

**Anotace.** T-uzel se zástupným t-lematem, *se/si* v *a/lex.rf* (!); gramatická koreference.

### 3.1 Reflexivita

(Nejbližší) subjekt děje zaměřuje děj na sebe sama.

**Anotace.** Zástupné t-lema: #*PersPron*.

Příklady:

*Nakreslil se.*

*Viděl se v zrcadle.*

*Zaměřuje se jen na sebe.*

*Oholil se.*

*Koupil si auto.*

### 3.2 Reciprocita

Totožnost nevyjádřené pozice s lexikálně vyjádřenou (zrecipročnou) pozicí.

**Anotace.** Zástupné t-lema: #*Rcp*.

Příklady:

*Petr a Pavel se potkali mezi sebou.*

*Petr a Pavel potkali sebe navzájem.*

*Petr a Pavel se od sebe navzájem liší.*

## Čísla

Nahrazuje kapitolu MAN 7.10.2.3. Složené číslovkové výrazy a kapitolu MAN 7.10.2.4. Desetinná čísla a zlomky.

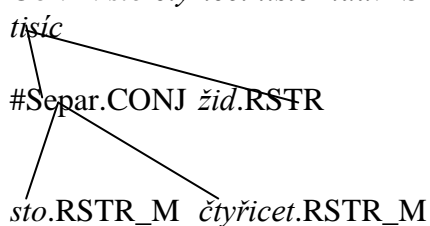
1	Typ <i>sto čtyřicet tisíc lidí</i> .....	34
2	Číslovka zapsaná číslicemi .....	35
3	Desetinné číslo .....	35
4	Typ <i>čas 7:30, byt 2+1, 1/3 vody</i> .....	36
5	Typ <i>tyč dlouhá 2 m 10 cm 4 mm</i> .....	36
6	Typ <i>1 1/3 vody odtekla</i> .....	37
7	Datum <i>3. 2. 2003</i> .....	37
8	Typ <i>40procentní vs. 40 procent a typ 40% vs. 40 %</i> .....	37

### 1 Typ *sto čtyřicet tisíc lidí*

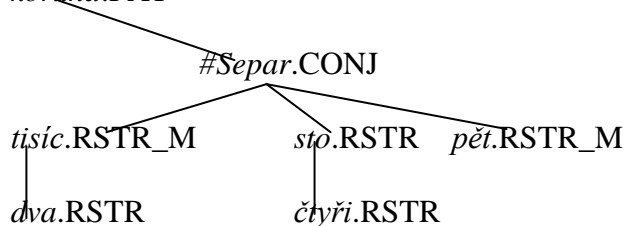
Složené číslovkové výrazy typu *sto čtyřicet tisíc, dva miliony pět set tisíc, třicet osm* zachytíme jako souřadnou strukturu pomocí uzlu *#Separ.CONJ*. Počítáme milióny, tisíce, sta, desítky, jednotky. Jsou-li ve složeném číslovkovém výrazu obsaženy desítky nebo jednotky, nezacházíme s ním jako s kontejnerem, tj. číslovkový výraz (zachycený jako souřadná struktura) rozvíjí počítané jméno jako RSTR.

#### Příklady:

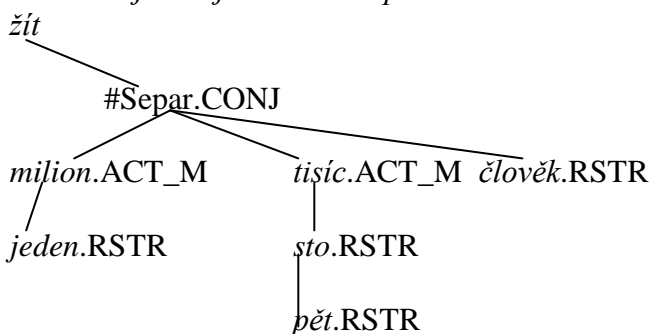
Obr. 1: *sto čtyřicet tisíc židů*.RSTR



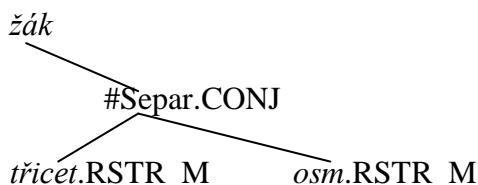
Obr. 2: *Vydělal 2 tisíce čtyři sta pět korun*.PAT



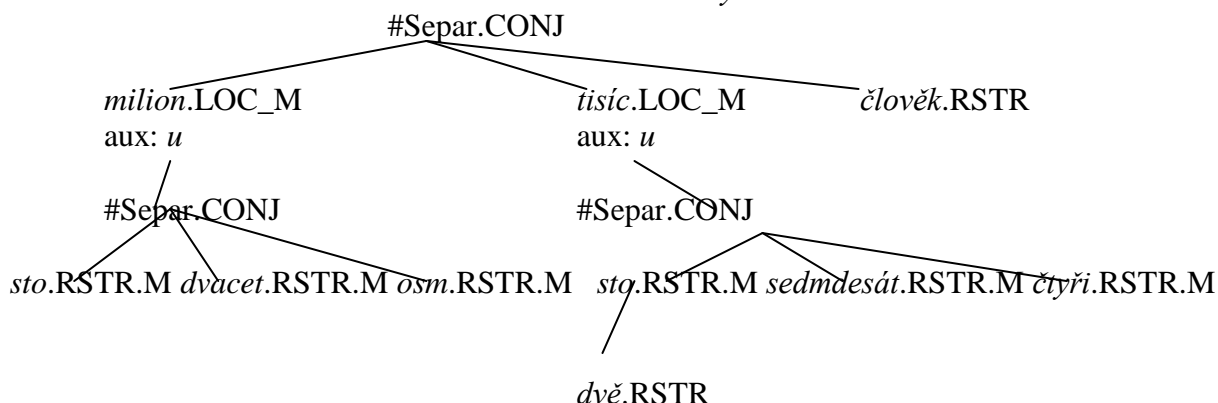
Obr. 3: *Žije zde jeden milion pět set tisíc lidí*.RSTR



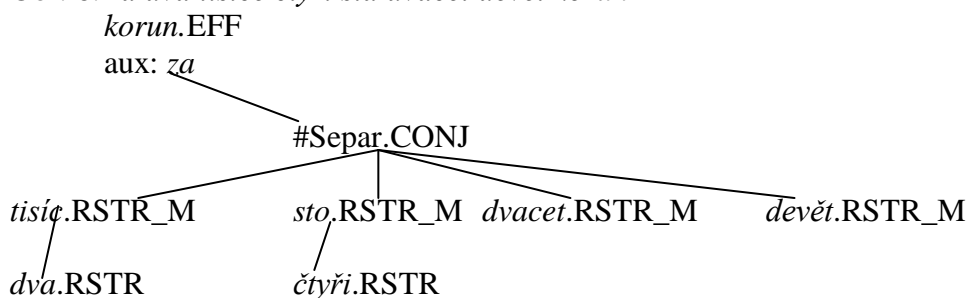
Obr. 4: *třicet osm žáků*



Obr. 5: *u sto dvaceti osmi miliónů dvě stě sedmdesáti čtyř tisíc lidí*



Obr. 6: *za dva tisíce čtyři sta dvacet devět korun*



## 2 Číslovka zapsaná číslicemi

Číslovku zapsanou číslicemi zachytíme jako jeden uzel, který vždy závisí na počítaném předmětu a má funktor RSTR. (Není-li taková číslovka na analytické rovině zachycena jako jeden uzel, spojíme na tektogramatické rovině jednotlivé části do jednoho uzlu, uzel bude mít „podtržítkové“ t-lema!)

### Příklady:

*Mám 38 234.RSTR korun.PAT*

*Mám 38 234.PAT*

*Vydělal 2 405.RSTR korun.PAT*

*Žije zde 1 500 000.RSTR lidí.ACT*

*Vydělává 1 500 000.PAT ročně.*

## 3 Desetinné číslo

Desetinné číslo zachycujeme jako jeden uzel.

### Příklady:

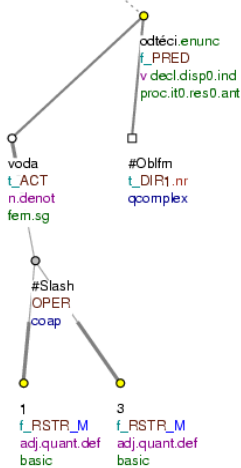
*Výroba se zvýšila o 2,3.RSTR procenta.*

*Výstavu navštívilo 2,5.RSTR tisíc lidí.*

#### 4 Typ čas 7:30, byt 2+1, 1/3 vody

Údaje ve formě matematické operace zachycujeme jako souřadné spojení pomocí funktoru OPER.

Obr. 7: 1 / 3 vody odtekla.



#### 5 Typ tyč dlouhá 2 m 10 cm 4 mm

Číselné údaje složené z kombinace číslovek a fyzikálních jednotek typu 2 m 10 cm 4 mm, 1 h 20 min zachytíme jako souřadnou strukturu pomocí uzlu #Separ.CONJ.

Příklady:

Tyč je dlouhá 2 m.EXT 10 cm.EXT 4 mm.EXT. (Obr. 8)

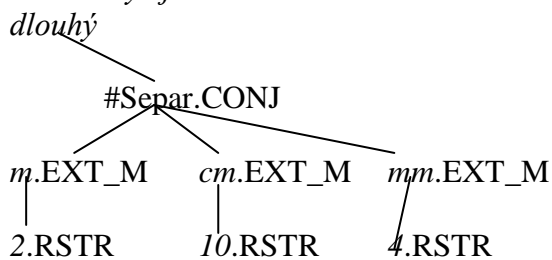
Film začíná ve 2 hod.TWHEN 35 min.TWHEN (Obr. 9)

(Podobně: Film začíná ve 2 hod.TWHEN a 35 min.TWHEN)

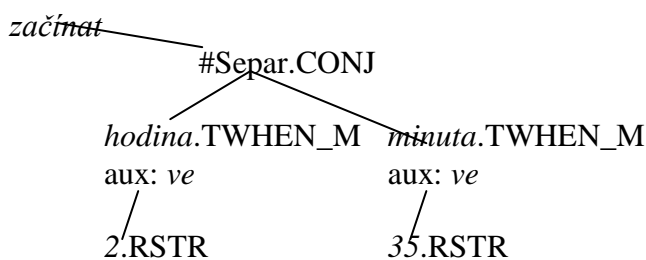
Vážil 53 tun.EXT 15 kg.EXT

(Podobně: Vážil 53 tun.EXT a 15 kg.EXT)

Obr. 8: Tyč je dlouhá 2 m 10 cm 4 mm.



Obr. 9: Film začíná ve 2 hod.TWHEN 35 min.TWHEN

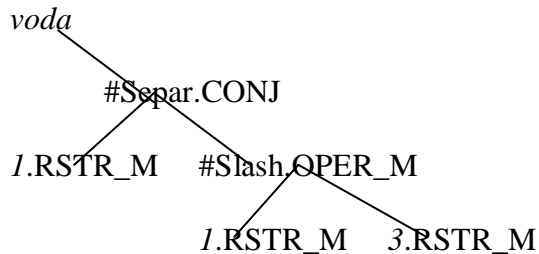


## 6 Typ 1 1/3 vody odtekla

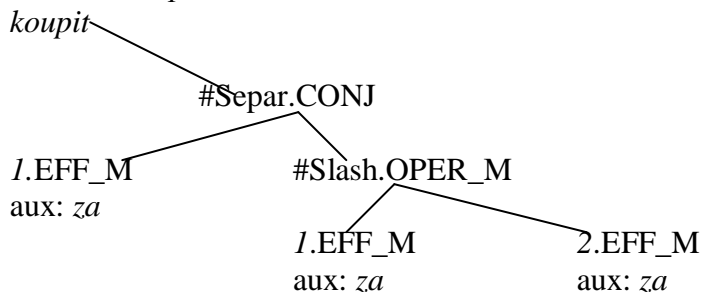
Příklady typu *1 1/3 vody odtekla* zachytíme jako souřadnou strukturu pomocí uzlu *#Separ.CONJ*.

Příklady:

Obr. 10: *1 1/3 vody odtekla*.



Obr. 11: *koupil za 1 1/2*

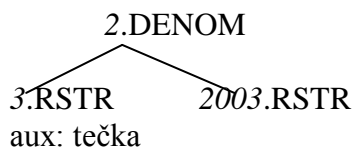


## 7 Datum 3. 2. 2003

Data ve formátu *3. 2. 2003* zachycujeme podle následujícího příkladu-šablony.

Příklad:

Obr. 12: *3. 2. 2003*



## 8 Typ 40procentní vs. 40 procent a typ 40% vs. 40 %

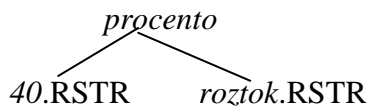
Tokenizer se chová následovně:

- číslo+znak - rozdělí na dva uzly (*40%*)
- číslo+písmeno/a - nerozdělí, 1 uzal (*40procentní*)
- [cokoli+znak+cokoli] - rozdělí na tři uzly (*40-ti*)

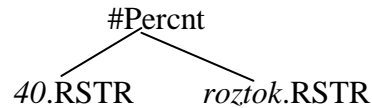
Základní pravidlo anotace: **číslo+slovo zachytíme vždy jako dva uzly**. Nejsou-li to (z důvodu nastavení tokenizace) na a-rovině dva uzly, označíme místo poznámkou *Word Segmentation*.

Příklady (substantiva):

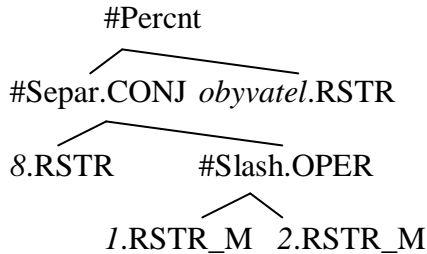
Obr. 13: 40 procent roztoku



Obr. 14: 40 % roztoku

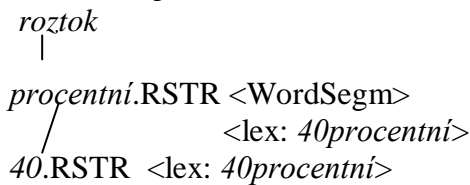


Obr. 15: 8 ½ % obyvatel

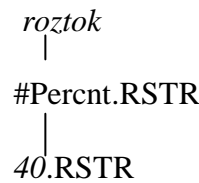


Příklady (adjektiva):

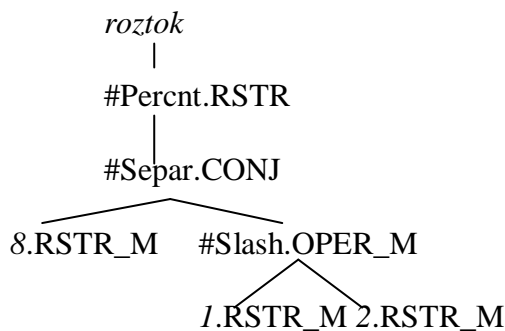
Obr. 16: 40procentní roztok



Obr. 17: 40% roztok



Obr. 18: ½% roztok

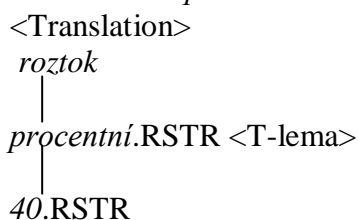


**Podle PČP nesprávné varianty:**

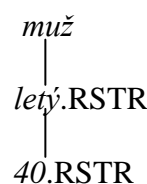
Poznámka: Hrůzy jako 30 – letý, 29 - ti roční, 8 1/8 % - ní oanotujeme analogicky podle výše uvedených pravidel. U adjektivních spojení, kde číslo vyjde odděleně od „slova“ zachováme rozdělení na dva uzly (číslo je RSTR na slově; viz příklady). Každou větu označíme poznámkou *Translation/Reconstruction*.

Příklady:

Obr. 19: 40-tiprocentní roztok



Obr. 20: 40-ti letý muž <Translation>





## Cizí vlastní jména a názvy

Dodatek ke kapitole MAN 7.9. Cizojazyčné výrazy.

1	Co je seznamová struktura (#Forn).....	39
2	Co není seznamová struktura .....	39

### 1 Co je seznamová struktura (#Forn)

Jako seznamovou strukturu, pomocí uzlu s t-lematem #Forn zachycujeme následující případy 1.1, 1.2 a 1.3. Česká předložka je vždy jako a-aux-uzel odkázána u uzlu s t-lematem #Forn. Jako t-lemma uzlů s funktořem FPHR slouží nezměněné tvary cizích slov z textu, tj. forma slov, i vyskoňovaná; interpunkční znaménka, která jsou součástí názvu nemají tektogramatická zástupná t-lemata začínající #! (makro Alt+F). Ve sporných případech (tedy například tehdy, kdy není možné poznat, zda bylo slovo použito sklonně, či nesklonně), dáváme přednost zachycení výrazu jako seznamové struktury.

#### 1.1 Nesklonný cizí název (víceslovný, ale i jednoslovný!)

Příklady:

Z [#Forn.DIR1] *Uyuni.FPFR je to 22 km.* (Obr. 3)

v deníku [#Forn.ID] *Financial.FPFR Times.FPFR*

*Přispíval do [#Forn.DIR3] britských.RSTR Financial.FPFR Times.FPFR* (Obr. 2)

*firma [#Forn.ID] Eagle.FPHR Group.FPHR V.FPHR ..FPHR A.FPHR ..FPHR* (Obr. 1)

#### 1.2 Víceslovný cizí název, u kterého se skloňuje jen nějaká jeho část

Příklady:

do [#Forn.DIR3] *New.FPHR Yorku.FPHR*

v [#Forn.LOC] *Rio.FPHR de.FPHR Janieru.FPHR*

v [#Forn.LOC] *San.FPHR Franciscu.FPHR*

#### 1.3 Víceslovná „neevropská“ vlastní jména osob

Příklad:

*čínský císař [#Forn] Tung.FPHR chun.FPHR Chou.FPHR* (obr.)

### 2 Co není seznamová struktura

Jako seznamová struktura nejsou zachyceny následující případy.

#### 2.1 Skloňované cizí názvy (tj. s českými morfologickými koncovkami)

Příklady:

v *Oaklandu.LOC*

v *Riu.LOC*

*Přispíval do britských.RSTR Timesů.DIR3*

#### 2.2 Cizí vlastní jména osob „evropská“ (George Bush, van Gogh)

#### 2.3 Jednoslovné cizojazyčné názvy v pozici nominativu jmenovacího

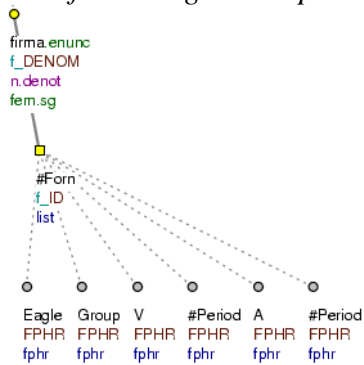
Uzel reprezentující jednoslovný cizojazyčný název dostane v pozici nominativu jmenovacího funktoř ID a žádný nový uzel t-lematem #Forn se již do stromu nedoplňuje.

Příklad:

*časopis Times.ID*

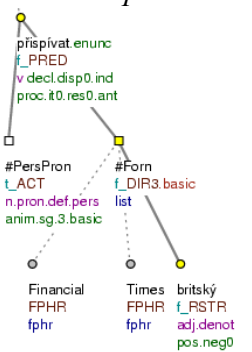
Příklady:

Obr. 1: firma Eagle Group V. A.

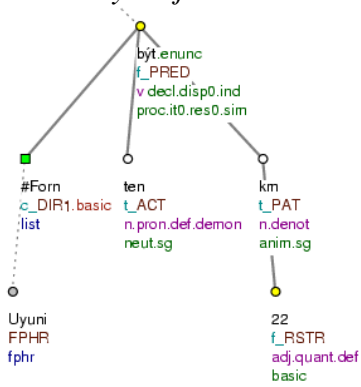


Chyba: místo #Period je v t-lematu „.“

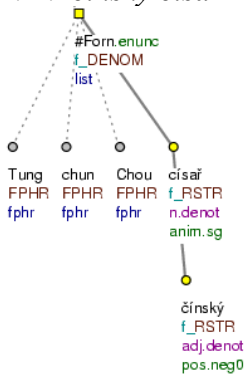
Obr. 2: Přispíval do britských Financial Times.



Obr. 3: Z Uyuni je to 22 km.



Obr. 4: čínský císař Tung chun Chou



## Zkratky

Nahrazuje kapitolu MAN 7.15. Zkratky.

1	Obecné zkratky .....	41
2	Iniciálové zkratky .....	41
3	Zkratková slova .....	42

### 1 Obecné zkratky

= prvních několik písmen zkracovaného slova na konci zpravidla s tečkou; některé jsou ustálené, tj. píšou se vždy stejným způsobem.

Příklady: *p., čís., v. r., př. n. l., a. s., č. j., čtk, atd., pozn. red.*. Patří sem i zkratky vlastních jmen a názvů: *M. Nováková, K. Vary.*

**Anotace.** Pro tyto zkratky nejsou zavedena žádná zvláštní pravidla anotace; pouze zkratky vlastních jmen a názvů podléhají pravidlům o anotaci identifikačních výrazů (MAN 7.8).

Jednotlivé části některých zkratk (s. r. o.; a. s.; hl. m.; pozn. red.) jsou reprezentovány samostatnými uzly a jsou strukturně analyzované; například zkratka *s. r. o.* je reprezentována třemi uzly. Jedním uzlem jsou reprezentovány zkratky jako *čtk, atd.; apod.* T-lemma zůstává buď rozepsané, nebo ve zkrácené podobě; ponecháváme je takové, jaké vygeneroval automatický parser (pokud není špatně). Je-li t-lemma nesprávně rozepsané, opravíme ho tak, aby bylo rozepsané správně. Do budoucna je třeba podobu t-lemat u zkratk sjednotit.

Speciální pravidla v manuálu:

*mj.* – RESTR; MAN 7.6.1 Význam „omezení“

*apod., atd., aj., apod.* – MAN 5.6.2.1 Specifické konstrukce zachycené jako koordinační nebo apoziční struktura

*tj., tzn.* – APPS; MANN Příloha B. Souřadící spojovací výrazy a operátory

### 2 Iniciálové zkratky

= nesklonné zkratky z počátečních písmen psané velkými písmeny.

Příklady: *USA, ZOH, LN, UNICEF, MFF, ČR*

**Anotace.** Iniciálová zkratka je reprezentována jedním uzlem a je zachycena jako identifikační struktura, tj. uzel reprezentující zkratku má funktor ID a závisí na uzlu pro obecné rodové jméno, kterým je zkratka uvozena (*stát, organizace, fakulta*). Pokud obecným rodovým jménem uvozena není, je do stromu doplněn nový uzel s t-lematem *#Idph*.

Příklady:

*V {#Idph.LOC} ČR.ID včera nepršelo.*

*{#Idph.ACT} ČR.ID je malá země.*

*Pojedeme do {#Idph.DIR3} USA.ID*

*Organizace.ACT OSN.ID bojuje proti bezpráví.*

*Studují na {#Idph.LOC} MFF.ID*

*Cyklista leží na oddělení.LOC ARO.ID*

*Cyklista leží na {#Idph.LOC} ARO.ID*

### 3 Zkratková slova

= zkratky, které mají charakter slov, tj. skloňují se, popř. se stávají základem dalšího odvozování.

Příklady: *Čedok, ARO*

**Anotace.** Pokud je zkratkové slovo použito sklonně, dostane funktor podle své pozice ve větě. Je-li použito nesklonně, je zachyceno podle stejných pravidel jako iniciálová zkratka, tj. jako identifikační struktura.

Příklady:

*Na dovolenou jezdí s Čedokem.MEANS*

*„Jak na Nový rok, tak po celý rok,“ povzdechl si pacient na ARU.LOC*

## Spojovník

1	Kompoziční spojení .....	43
2	Jedno slovo .....	44

Dodatek upravuje anotaci spojení: **něco+spojovník+něco** (například: *BMP-1*, *srbsko-černo-horský*, *c-dur*, *Rolls-Royce*).

Na analytické rovině je něco+pomlčka/spojovník+něco vždy reprezentováno jako tři uzly. Na rovině tektogramatické rozlišujeme, zda se jedná o kompoziční spojení nebo o jedno slovo.

### 1 Kompoziční spojení

Spojovníkem/pomlčkou jsou spojena slova (znaky) nesoucí význam i samostatně.

#### 1.1 Slova spojená spojovníkem/pomlčkou jsou ve vztahu souřadnosti

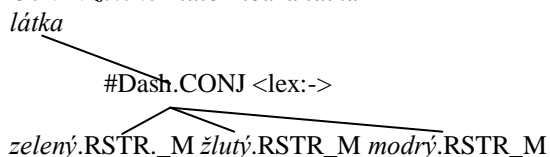
Příklady:

*anglicko-český slovník*, *zeleno-žluto-modrá látka*

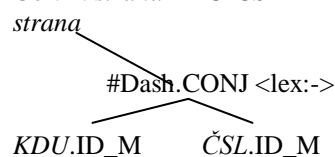
*KDU-ČSL*

**Anotace.** Skupina *něco+spojovník/pomlčka+něco* je reprezentována jako **tři uzly**, uzel pro poslední spojovník/pomlčku je kořenem koordinační struktury (*#Dash.CONJ*). Případné další spojovníky/pomlčky (při vícečetném spojení) nejsou reprezentovány samostatnými uzly, ani nejsou zachyceny jako a-aux-uzly.

Obr. 1: *zeleno-žluto-modrá látka*



Obr. 2: *strana KDU-ČSL*



#### 1.2 Mezi spojenými slovy je vztah determinace

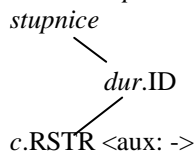
Příklady:

*c-dur*, *RM-systém*, *protein BMP-1*, *C-vitamin*, *alfa-záření*, *Rh-faktor*, *Praha-Smíchov*, *Brno-venkov*, *Havlíček-kritik*.

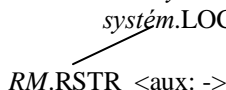
**Anotace.** Skupina *něco+spojovník/pomlčka+něco* je reprezentována jako **dva uzly**: jeden uzel jako řídicí a druhý jako na něm závislý.

Uzel pro pomlčku/spojovník není na t-rovině reprezentován samostatným uzlem, ale je zachycen jako a-aux-uzel u závislého uzlu spojení.

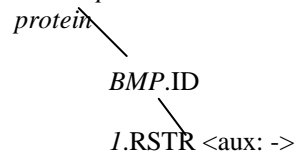
Obr. 3: *stupnice c-dur*



Obr. 4: *v RM-systému*



Obr. 5: *protein BMP-1*



**Pozor!** Speciální pravidla jsou zavedena pro vícedílná vlastní jména osob a měst: *Marie-Anna*, *Sklenářová-Malá*, *Frydek-Místek*! Tato spojení spadají pod typ 1.2 (viz MAN 7.8 Identifikační výrazy).

## 2 Jedno slovo

Spojení není rozložitelné na význam jednotlivých částí (jde zejména o přejatá slova).

Příklady:

*Rolls-Royce*

*ping-pong*

*moucha tse-tse.*

**Anotace.** Spojení zachytíme analogicky podle pravidel pro frazémy (jako DPHR). Řídícím uzlem spojení bude zpravidla druhá skloňovaná část.

Obr. 6: *Na lov jediné Rolls-Roycem.*

*Royce*.MEANS

*Rolls\_*-.DPHR <lex: *Rolls*; aux: ->

Obr. 7: *Končí s ping-pongem.*

*pong*.PAT <aux: *s*>

*ping\_*-.DPHR <lex: *ping*; aux: ->

## Nové sekundární předložky

Dodatek ke kapitole MAN 7.17.1 Sekundární předložky

Seznam schválených sekundárních předložek je přílohou velkého manuálu (MAN Příloha 3). Příloha obsahuje seznam spojení, která mohou být v konkrétním textu sekundární předložkou, spojení obsažená v tomto seznamu však nejsou sekundární předložkou vždy. Slova uvedená jako součást sekundární předložky mohou mít v klauzi svou sémantickou a syntaktickou platnost a pak jsou reprezentována samostatným uzlem ve stromě a tvoří součást předložky.

### Sekundární předložky neobsažené v seznamu:

Jako sekundární předložka však mohou v datech vystupovat i jiná spojení, která nejsou obsažená ve zmíněném seznamu. Pokud má anotátor „podezření“, že nějaký víceslovný výraz plní funkci sekundární předložky, pak postupuje následujícím způsobem:

- víceslovný výraz nebude zachycovat jako sekundární předložku, ale poznámkou typu *Secondary preposition* označí uzel (jeden z uzlů), který by v případě zachycení sekundární předložkou nebyl samostatným t-uzlem. Do textu poznámky zapíše přesné znění sekundární předložky (a nic jiného).
- o potenciálních sekundárních předložkách budeme vést diskuzi a postupně budeme seznam sekundárních předložek rozšiřovat.

## Různé

1	Číslům-nálepším náleží ID .....	46
2	Význam: „jak daleko“ .....	46
3	Cena za něco, něco za nějakou cenu (za+4) .....	47
4	EXT – ID – RSTR.....	47
5	Chybí konec věty .....	48
6	„kontext(y)“ .....	48

## 1 Číslům-nálepším náleží ID

**Změna!** Číslům-nálepším náleží ID (nikoliv RSTR)!

Příklady:

*linka 900.ID*

*telefon 342 434 234.ID*

*PSC 561 64.ID*

*číslo 1.ID*

*Mazda 323.ID*

*v roce 1993.ID*

*kurs 3.ID ku.OPER 5.ID*

## 2 Význam: „jak daleko“

Příklady:

*Odcestoval přes 30 km.farmore na jih.*

*Odcestoval okolo 30 km.farapprox na jih.*

**LOC**

*Umístil se na pátém místě.LOC.far*

*Umístil se kolem pátého místa.LOC.farapprox*

*Umístili se od pátého místa.LOC.farmoreincl*

*Umístil do pátého místa.LOC.farlessincl*

**DIR3**

*Skočil dva metry.DIR3.far*

*Skočil přes dva metry. DIR3.farmore*

*Voda sahá po kolena.DIR3.farlessincl*

**Subfunktory budou zavedeny později.**

Obr. 1: *Vesnice leží 30 km na jih.*

*na jih.DIR3*

*30 km.EXT*

Obr. 2: *Posunuli se dva metry jižněji.*

*jižněji.DIR3*

*dva metry.EXT*

Obr. 3: *chvění bylo pocíťováno až na 200 mil daleko*

*daleko.LOC*

*na 200 mil.EXT/LOC.far*



### 3 Cena za něco, něco za nějakou cenu (za+4)

Valenční (valenci substantiv neanotujeme):

**cena, účet, poplatek za něco.RSTR**

**cena** EXT(kolik) RSTR(za+4)  
*cena 20 korun.EXT za chleba.RSTR*  
RSTR(.2)  
*cena chleba.RSTR*  
*vývozní cena naší oceli.RSTR*  
*ceny mouky.RSTR*  
RSTR(.2;u;od+2) RSTR(.3)  
*(ocenit, medaile) cena sportovcům.RSTR*

**poplatek** EXT(kolik) RSTR(za+4)  
*poplatek 20 korun.EXT za vstupné.RSTR*  
*poplatek za zavolání.RSTR*

**účet** EXT(kolik) RSTR(za+4)  
*účet za tuto knihu.RSTR*

**úrok** EXT(kolik) RSTR(za+4)  
*úrok za půjčku.RSTR*

Nevalenční:

**hodnota za něco.REG**

*6,5 mil. za podíl.REG*  
*koruna za papír.REG*  
*2,10 dolaru za akcii.REG*  
*78 centů za libru.REG*

**něco za hodnotu.EXT**

*podíl za 6,5 mil.EXT*  
*papír za korunu.EXT*  
*akcie za 2,10 dolaru.EXT*  
*libra za 78 centů.EXT*

### 4 EXT – ID – RSTR

*pět.RSTR čísel*  
*číslo pět.ID*  
*páté.RSTR číslo*

Počet „věcí“ je RSTR (bylo by ovšem vhodnější volit zde funktor EXT):

*30 000.RSTR korun*  
*24.RSTR dolarů*  
*dvě.RSTR stanoviska*  
*přečetl jsem přes 100.RSTR stránek.PAT (přes v a/aux.rf u RSTR)*

U sloves se řídíme valenčním slovníkem:

*činí to 30 000 korun.EXT*  
*vydělává 30 000 korun.PAT*  
*získal 30 000 korun.PAT*

Pozor! Pozor! Žádné PAT!

*přijem zboží.RSTR*  
*přijem dvou.RSTR korun.RSTR*  
*přijem dvě.RSTR koruny.EXT*  
*přijem 30 000.RSTR korun.EXT (i jako výplata)*

*zisk pěti.RSTR dolarů.RSTR*  
*zisk pět.RSTR dolarů.EXT*  
*výnos 102.RSTR miliard.EXT*  
*tržba 2.RSTR miliardy.EXT*  
*daň 3 000 korun.EXT*  
*objem 10 kubiků.EXT*  
*objem vody.RSTR 10 kubiků.EXT*  
*objem vody.RSTR*  
*v hodnotě 30 000.RSTR korun.EXT*  
*hodnota vody.RSTR*  
*suma 3 000.EXT*  
*ve výši 30 000.RSTR korun.EXT*  
*délka 3.RSTR metry.EXT*  
*celodenní minimum 2 141,1.EXT*  
*maximum 2 141,1.EXT*

Analogicky neurčité číslovky základní:

*více.RSTR obchodních transakcí*  
*více.RSTR peněz*  
*za kolik.RSTR peněz*

Měrové významy:

*vydělává 30 000.RSTR korun.PAT*  
*vydělává do 30 000.RSTR korun.PAT*  
*vydělává od 30 000.RSTR korun.PAT*  
*vydělává přes 30 000.RSTR korun.PAT*  
*vydělává pod 30 000.RSTR korun.PAT*  
Předložky budou v *a/aux.rf* u RSTR.

## 5 Chybí konec věty

Pokud některé větě chybí její konec (a to i v originále), pak větu oantujeme podle běžných pravidel anotace tak, jako by věta byla nedokončená. V místě chybějící části bude nějaké #Emp\*. Ke kořeni stromu dáme anotátorskou poznámku *Other* s textem: *chybí konec věty*.

## 6 „kontext(y)“

Příklad:

*Typické Glassovo dílo je nejen s otevřeným koncem , je také často mnohoznačné , pokud jde o kontext(y).*

Interpretace: *pokud jde o kontext nebo o kontexty*

Ve větě zachytíme koordinaci mezi *kontext* a zkopírovaným *kontext*, kořenem koordinační struktury bude uzel se zástupným t-lematem #Separ (podle smyslu funktor, v tomto případě DISJ). Zkopírovaný uzel s t-lematem *kontext* bude mít v následujících attributech následující hodnoty:

*is\_generated = 1*

*is\_parenthesis = 1*

*a/lex.rf*: odkaz na *kontext*

*a/aux.rf*: odkazy na: *o ( ) y*

## Segmentace vět do stromů a slov do uzlů

1	Segmentace vět do stromů.....	49
2	Segmentace slov do uzlů .....	49

### 1 Segmentace vět do stromů

#### 1.1 V jednom stromě je více vět

Pokud je v jednom stromě více vět (měly by to tedy být stromy dva, tři atd.), bude takový případ vyřešen tak, že na technický kořen stromu (*nodetype = root*) se zavěsí jako sestry efektivní kořeny obou vět.

Do *anot\_comment* typ *Sentence Segmentation* u technického kořene stromu se zapíše například poznámka: 2 v 1.

Tímto způsobem jsou řešeny pouze evidentní případy spojení více vět do jednoho stromu (tj. v jednom stromě je dvě a více vět ukončených interpunkčním znakem pro konec věty: *Věta první? Věta druhá.* Případy, kdy více větných celků je od sebe odděleno například pomlčkou (*Praha (haš) – Věta.*), zachycujeme jako souřadnou strukturu (koordinace, apozice) a za případ nesprávné segmentace vět je nepovažujeme.

#### 1.2 Jedna věta je rozdělena do více stromů

Pokud je jedna věta rozdělena do více stromů, ve stromech, v kterých má méně uzlů, se všechny její uzly vymažou, a ve stromě, v kterém má nejvíce uzlů, se chybějící uzly znovu vytvoří (tzn. znovu je „vytáhneme“ z analytického stromu, ale budou mít *is\_generated = 0*). V tomto stromě se pak věta oanotuje.

Do *anot\_comment* typ *Sentence Segmentation* u technického kořene toho stromu, ve kterém byla rozdělená věta anotována, se zapíše například poznámka: 1 ve 3 < >. Znaky > nebo < se naznačí, zda zbývající části věty byly v předcházejícím nebo následujícím stromě.

### 2 Segmentace slov do uzlů

#### 2.1 Jeden uzel je více významových jednotek

V případě, že jedním t-uzlem je reprezentováno více významových jednotek (měly by to tedy být uzly dva, tři atd.), vytvoří se ve stromě tolik nových uzlů (budou mít *is\_generated = 0*), kolik je potřeba ke správné anotaci. T-lemata uzlů se opraví ručně.

Je-li chyba i na analytické rovině, vedou z nově vytvořených t-uzlů odkazy na jeden stejný chybný a-uzel.

Příklad: *srbsko-černohorský* je jeden uzel, ale mají to být uzly tři.

Místo označíme poznámkou typu *Word Segmentation* s textem: 3 v 1.

#### 2.2 Jedna významová jednotka je reprezentována více uzly

V případě, že jedna významová jednotka je reprezentována více t-uzly, vybere se jeden t-uzel, který v tektogramatickém stromě zůstane, a ostatní t-uzly se vymažou (připojí jako a-aux/lex-uzly). T-lemma ponechaného t-uzlu opravíme na správnou hodnotu, též správně oanotujeme odkazy na a-uzly.

Příklad: *van Beethoven* jsou dva uzly.

Standardní případy typu *van Beethoven* není třeba poznámkou *Word Segmentation* označovat.

U nestandardních pak poznámku vybíráme a píšeme například: 1 ve 2.

## Chyby v překladu a v rekonstrukci

1	Nevhodnost překladu/rekonstrukce lze odstranit smazáním některých znaků/slov .....	50
2	Nevhodnost překladu nelze odstranit smazáním některých znaků/slov .....	51
2.1	Vhodnější překlad je zřejmý. ....	51
2.2	Vhodnější překlad není zřejmý. ....	51
3	Jiný případ .....	51
3.1	Chyby v interpunkci .....	52
3.2	Morfologické chyby .....	53
3.3	Pravopisné chyby .....	53

Základní pravidlo: **Povinností anotátora není kontrolovat správnost překladu/rekonstrukce.** Pokud je věta dobře česky, lze ji přiřadit tektogramatickou strukturu, anotátor větu oanotuje, originálním anglickým zněním/původním přepisem věty se nemusí zabývat.

**Ale protože nepřesností v překladu/rekonstrukci je poměrně dost, kontroluje anotátor správnost překladu/rekonstrukce při sebemenším podezření, že překlad/rekonstrukce není v pořádku.**

Pokud je překlad/rekonstrukce nesprávný, nepřesný, věta není dobře česky, nelze ji oanotovat, pak anotátor postupuje podle jedné z následujících možností.

### 1 Nevhodnost překladu/rekonstrukce lze odstranit smazáním některých znaků/slov

Smazáním některých znaků/slov (které jsou ve větě navíc) by věta dostala vhodnější znění a bylo by možné ji oanotovat.

Řešení:

Poznámka: **Translation/Reconstruction**

Anotátor vymaže ze stromu t-uzly reprezentující nadbytečné výrazy a větu oanotuje, jako by v ní nadbytečné výrazy nebyly. Do textu poznámky typu *TranslationReconstruction* povinně запиše úseky textu (a úseky znaků), které je třeba z věty vymazat, aby odpovídala provedené anotaci. Poznámka má následující předepsanou formu:

*Vymazat: x y z*

Znamená to: věta bude oanotovaná správně, pokud se z věty vymažou slova *x y* a *z*.

Pokud je slovo *x* (například čárka) ve větě víckrát, pak poznámka vypadá takto:

*Vymazat: {s} x*

a znamená to, že se má vymazat *x* následující po slově *s*.

Příklad:

wsj2103/6: *Společnost Alpine, která prohlásila, že neplánuje dále nabídku rozšiřovat, a prohlásila, že jako odpověď na původní nabídku obdržela nominální hodnotu 615000 dolarů z dluhu.*

Věta má chybnou strukturu – vyšínutí z vazby. „Správnější“ věta se získá vymazáním spojky *a*. Automaticky vygenerovaný t-uzel pro tuto spojku anotátor vymaže a větu oanotuje tak, jako by v ní tato spojka nebyla.

Do poznámky *Translation/Reconstruction* запиše: *Vymazat: a*

## 2 Nevhodnost překladu nelze odstranit smazáním některých znaků/slov

Řešení: Anotátor nebude větu vůbec anotovat. A větu označí poznámkou podle následujících dvou případů.

### 2.1 Vhodnější překlad je zřejmý

Řešení:

Poznámka: *Re-generated*

Pokud je zřejmé, jak má vypadat vhodnější překlad, zapíše anotátor do textu poznámky *Re-generated* přesné nové znění věty. Do poznámky *Not-annotated* (nebo *Translation/Reconstruction*) může zapsat další komentář.

**Pozor!** Do toho typu patří i všechny **případy „chybějících slovních jednotek“**, a to i takových, které na t-rovině nemají svůj vlastní t-uzel, ale jsou reprezentovány jako a-aux-uzly. Výjimkou je chybějící a nesprávná interpunkce (k tomu viz 3.1).

Příklad:

wsj2100/4: *Nově uveřejněný "skutečnost" ukazuje, že "tvrdá reklama" uchovává údaj o preferenci možného vysílacího času pro zítřejší pořad u diváků, kteří ji sledují.*

Originál: *And the new syndicated ``reality" show ``Hard Copy" records viewers' opinions for possible airing on the next day's show.*

Anotátor nebude větu anotovat. Do poznámky *Re-generated* zapíše: *A nově uvedená reality show Tvrdá reklama zaznamená názory diváků, aby je bylo možné druhý den vysílat.*

Příklad (chybějící předložka):

*Ceny akcií uzavřely výš v Singapuru, Tajvanu a Wellingtonu, byly smíšené v Hongkongu, uzavřely níž v Soulu a málo se změnilo i v Sydney.*

V překladu chybí předložka v před *Sydney* (v originále je). Anotátor nebude větu anotovat. Do poznámky *Re-generated* zapíše: *Ceny akcií uzavřely výš v Singapuru, Tajvanu a Wellingtonu, byly smíšené v Hongkongu, uzavřely níž v Soulu a málo se změnilo i v Sydney.*

### 2.2 Vhodnější překlad není zřejmý

Řešení:

Poznámka: *Not-annotated*

Pokud vhodnější překlad není anotátorovi zřejmý (anotátor nedokáže větu vhodněji přeložit, problém je již v originále), zvolí anotátor poznámku *Not-annotated* a v textu poznámky popíše problém.

## 3 Jiný případ

Větu lze za daného překladu oannotovat, ale anotátor by rád uvedl nějaký komentář k překladu.

Řešení:

Poznámka: *Translation/Reconstruction*

Anotátor větu oannotuje, do textu poznámky *Translation/Reconstruction* zapíše komentář.

### 3.1 Chyby v interpunkci

Případy jako:

- chybějící tečka, čárka, interpunkce, která nebude zachycena jako t-uzel
- chybějící tečka, čárka, interpunkce, která bude zachycena jako kořen souřadné struktury
- místo čárky středník
- čárka navíc (= typ 1)

neřešíme jako případ 2.1 (*Re-generated*), ale větu oanotujeme a opatříme poznámkou *Translation/Reconstruction* (případně i jinou).

Příklady (chybějící „neanotovaná“ interpunkce):

*A je to*

*Nevím co máš na mysli.*

Ve větách chybí interpunkční znaménka (tečka, čárka). Anotátor запиše tuto skutečnost do poznámky *Translation/Reconstruction* a větu oanotuje.

Příklady (chybějící interpunkce, kterou potřebujeme pro kořen souřadné struktury):

*Koupil chleba máslo.*

*Jana koupila chleba Hanka koupila máslo.*

Ve větách chybí interpunkční znaménka (tečka, čárka), která potřebujeme pro zachycení koordinace.

Anotátor větu oanotuje. Koordinaci zachytí pomocí uzlu s t-lematem *#Separ*, ale připojí k němu poznámku *T-lemma*, do které uvedeme např. *lépe #Comma, chybí čárka*. Ke kořeni stromu připojí anotátor i poznámku *Translation/Reconstruction* (s textem např.: *chybí interpunkční čárka*).

Příklad (středníky místo čárek):

*Ministerstvo by bylo povinno zastavit vykupování, pokud by šlo k takovému finančnímu oslabení dopravce, že by to ohrozilo bezpečnost; jeho konkurenceschopnost by byla prudce snížena; nebo pokud by transakce vedla k prodeji majetku spojeného s aeroliniemi.*  
Ve větě jsou místo čárek středníky. Anotátor запиše tuto skutečnost do poznámky *Translation/Reconstruction* a větu oanotuje.

Příklad (středník místo čárky jako kořen souřadné struktury):

*Ovšem v posledních dnech Columbia mírně vyšplhala; se ziskem 3/8 se její akcie včera při uzavření trhu obchodovaly dokonce za 5 1/4 dolaru.*

Namísto středníku bychom raději viděli čárku. Větu oanotujeme. Kořenem souřadné struktury bude uzel s t-lematem *#Semicolon*. Uzel bude označen poznámkou typu *T-lemma* (např. *lépe #Comma, opravit středník na čárku*). Celé větě (kořenu stromu) bude přiřazena poznámka *Translation/Reconstruction*.

Příklad („neanotovaná“ čárka navíc):

*Klesající emise lehce přečísly rostoucí emise, 454 ku 451.*

Čárka je v českém překladu zbytečná. Větu oanotujeme tak, jakoby v ní čárka nebyla. Do poznámky *Translation/Reconstruction* napíšeme něco jako: *Vymazat: ,*

### 3.2 Morfologické chyby

Případy jako:

- nesprávně přeložený čas u slovesa, pád u substantiva,
- nespisovná koncovka

neřešíme jako případ 2.1 (*Re-generated*), ale větu oannotujeme.

Pokud jde o t-uzel a t-lema je správně, stačí k t-uzlu přiřadit poznámku *Typo*. V ostatních případech postupujeme v souladu s pravidly pro anotátorskou poznámku, vždy je možné (a vhodné) přiřadit poznámku *Translation/Reconstruction* ke kořeni stromu.

Platí i pro slova, která ve výsledné tektogramatické anotaci zůstanou jen jako a-aux-uzly.

Příklad:

*Mexický prezident Salinas oznámil, že recese jeho země skončila a ekonomika opět rostla.*

Originál: *Mexico's President Salinas said the country's recession had ended and the economy was growing again.*

V překladu je nesprávně *rostla* místo *roste*. Větu oannotujeme a k uzlu *růst* přiřadíme poznámku *Typo* (případně celou větu označíme poznámkou *Translation/Reconstruction*).

Příklad:

*Děti bude plavat.*

V překladu je nesprávně *bude* místo *budou*. Větu oannotujeme a ke kořeni stromu připojíme poznámku *Translation/Reconstruction*.

### 3.3 Pravopisné chyby

Případy jako:

- nesprávné *i/y*
- chyby ve velkých/malých písmenech

neřešíme jako případ 2.1 (*Re-generated*), ale větu oannotujeme a opatříme poznámkou *Typo* nebo *Translation/Reconstruction* (případně oběma).

Platí i pro slova, která ve výsledné tektogramatické anotaci zůstanou jen jako a-aux-uzly.

Příklad:

v překladu: *Mexická vláda*

Správně má být *mexická vláda*. Anotátor t-lema ručně opraví a uzal označí poznámkou *Typo* (případně *Translation/Reconstruction* u kořene stromu).

Příklad (chyba v a-uzlu):

*šel bich*

Správně má být *bych*. Anotátor zapíše chybu do poznámky *Translation/Reconstruction* u kořene stromu.

Poznámka: Poznámkou *Not-annotated* označujeme i věty neoanotované z jiných důvodů než překladových.

## Anotátorská poznámka

1	Nejistota ve struktuře a funkci .....	54
1.1	Structure ?.....	54
1.2	Functor ?.....	55
2	Nesprávné t-lemma .....	55
2.1	T-lemma with underscore ! .....	55
2.2	M-lemma M.....	55
2.3	Typo !.....	55
2.4	T-lemma T .....	55
3	Problémy s valencí .....	55
3.1	New valency frame !.....	55
3.2	Vallex ?.....	56
4	Problémy s překladem, větu nelze oannotovat .....	56
4.1	Translation !.....	56
4.2	Re-generated X.....	56
4.3	Not-annotated X.....	56
5	Nesprávná segmentace .....	57
5.1	Word segmentation w.....	57
5.2	Sentence segmentation s.....	57
6	Chyba v morfologii .....	57
6.1	M-tag t.....	57
7	Sekundární předložka.....	57
7.1	Secondary preposition # .....	57
8	Jiný problém .....	57
8.1	Other ?.....	57

Atribut *annot\_comment* (struktura atributů *type* a *text*) je pracovní atribut, který slouží pro označení problémových míst v datech, ke kterým se bude třeba při závěrečných opravách a kontrolách vrátit a problém vyřešit. Z konečné podoby dat je atribut vymazán.

Určité typy problémů se dopředu předpokládají, byly proto vymezeny typy problémových případů. V pod-atributu *type* anotátor zvolí typ problémového případu. V pod-atributu *text* má možnost uvést komentář.

Pokud je některý jev v datech vymezen jako problémové místo (je zaveden typ poznámky pro tento jev), je povinností anotátora toto místo výběrem správného typu poznámky označit. U některých typů poznámek je pak povinný i text (komentář) poznámky.

(Barvy a znaky v nadpisech odkazují ke způsobu, jakým se anotátorská poznámka zobrazuje v datech.)

### 1 Nejistota ve struktuře a funkci

#### 1.1 Structure ?

Použit kdy? Poznámka označuje místa, kde je nejistota ve struktuře stromu, tj. zejména při nejasné závislosti, anotátor si například není jistý pravidlem, podle kterého větu anotovat.

Text poznámky: povinný. Zapiše se, v čem nejistota spočívá, mezi čím, jakými pravidly anotátor váhá.



## 1.2 Functor ?

Použit kdy? Poznámka označuje uzly, u kterých anotátor váhá ve volbě funktoru a chce k této nejistotě uvést nějaký komentář.

Text poznámky: povinný. Do textu poznámky zapíše anotátor komentář.

Pozor! Alternativní funktoři, mezi kterými váhá, zapíše jako alternativy přímo do atributu *functor*.

## 2 Nesprávné t-lemma

### 2.1 T-lemma with underscore !

Použit kdy? Poznámka označuje uzly, u kterých bylo původní „automatické“ t-lemma změněno na t-lemma víceslovné (na t-lemma s podtržítkem) a při této změně nebylo možné použít zavedených maker (Alt+L; Alt+F), tj. změna musela být provedena ručně.

Text poznámky: NEpovinný.

### 2.2 M-lemma M

Použit kdy? Poznámka označuje uzly, u kterých bylo původní „automatické“ t-lemma (ručně) změněno, protože jeho podoba byla nesprávná z důvodu chybné morfologické analýzy.

Například: *je* má t-lemma *být* místo *#PersPron*.

Neplatí pro uzly typu *fphr* opravené makrem Alt+F.

Text poznámky: NEpovinný.

Výběr tohoto typu poznámky zároveň znamená, že je chyba v morfologickém tagu nebo jiná chyba na a-rovině. Anotátor proto již nezapisuje (nemusí) nic do poznámky typu *M-tag*.

### 2.3 Typo !

Použit kdy? Poznámka označuje uzly, u kterých bylo původní „automatické“ t-lemma ručně změněno z důvodu překlepu v t-lematu.

Text poznámky: NEpovinný.

### 2.4 T-lemma T

Použit kdy? Poznámka označuje uzly, u kterých bylo původní „automatické“ t-lemma ručně změněno z jiných důvodů, než které jsou popsány v 2.1 až 2.3.

Text poznámky: povinný. Popíše se důvod změny v t-lematu.

## 3 Problémy s valencí

### 3.1 New valency frame !

Použit kdy? Poznámka označuje místa, kdy je nutná změna ve vallexu a tato změna je jasně daná; dále slouží k rozlišení dějových a nedějových významů deverbativních substantiv na *-ní* a *-tí*.

Text poznámky: povinný a přesně daný (více viz dodatek *Valence* a dodatek *Anotace valence u „nesloves“*):

F	+návrh nového rámce	(rámec v PDT-vallexu chybí, žádný rámec není přiřazen)
W	+návrh nového rámce	(slovo v PDT-vallexu chybí, žádný rámec není přiřazen)
A	+nová forma	(v zápisu formy chybí nějaká alternativa, rámec přiřazen)
N	+nový funktor	(v rámci chybí neobligatorní aktant, rámec přiřazen)
X	+nový funktor	(v alternativní skupině chybí funktor, rámec přiřazen)
P		(dějový význam deverbativního substantiva na <i>-ní</i> a <i>-tí</i> )
nP		(nedějový význam deverbativního substantiva na <i>-ní</i> a <i>-tí</i> )

### 3.2 Vallex?

Použit kdy? Poznámka slouží jako komentář k vallexu.

Text poznámky: povinný. Uvést komentář.

## 4 Problémy s překladem/rekonstrukcí, větu nelze oannotovat

### 4.1 Translation/Reconstruction!

Použit kdy? Poznámka označuje špatně přeložené/rekonstruované věty (které ovšem byly oannotovány):

- věta byla oannotována, ale je třeba některá slova/znaky (reprezentovaná t-uzly) z věty vymazat (jsou tam navíc). Více viz dodatek *Chyby v překladu/rekonstrukci*.
- jiný problém v překladu.

Poznámku vybereme u kořene stromu (*nodetype = root*).

Text poznámky: povinný. Povinně zapsat, která slova (znaky) je třeba z věty vymazat. Jiné poznámky k překladu.

### 4.2 Re-generated X

Použit kdy? Poznámka označuje špatně přeloženou větu, kterou nelze z tohoto důvodu oannotovat. Je třeba ji přegenerovat v novém znění, které je v textu této poznámky.

Poznámku vybereme u kořene stromu (*nodetype = root*).

Text poznámky: povinný a přesně daný. Obsahuje přesné nové znění věty. Do poznámky nepíšeme nic jiného. Komentář je možné uvést v poznámce *Not-annotated*.

Výběr poznámky typu *Re-generated* zároveň znamená, že věta nebyla vůbec oannotována, anotátor proto již nemusí volit poznámku *Not-annotated*, ani poznámku *Translation/Reconstruction*.

### 4.3 Not-annotated X

Použit kdy? Poznámka označuje věty, které nebyly oannotovány, a to i z jiných důvodů než překladových.

Poznámku vybereme u kořene stromu (*nodetype = root*).

Text poznámky: povinný. Důvod, proč věta nebyla oannotována. Komentář ke znění věty v poznámce *Re-generated*.

## 5 Nesprávná segmentace

### 5.1 Word segmentation w

Použit kdy? Poznámka označuje místa, kde jedním t-uzlem bylo reprezentováno více významových jednotek (měly by to tedy být uzly dva, tři atd.; například: *srbsko-černohorský* byl jeden uzel) a chyba je i na analytické rovině. Poznámku stačí vybrat jen u jednoho z „nových“ t-uzlů. Poznámkou též označujeme nestandardní místa, kdy je více uzlů pro to, co má být zachyceno jedním uzlem.

Více v dodatku *Segmentace vět do stromů a slov do uzlů*.

Text poznámky: povinný a přesně daný. Zapiše se poznámka typu: 3 v 1, 1 ve 2.

### 5.2 Sentence segmentation s

Použit kdy? Poznámka označuje místa, kde v jednom stromě je více vět (měly by to tedy být stromy dva, tři atd.), a místa, kde jedna věta je rozdělena do více stromů.

Poznámku vybereme u kořene stromu (*nodetype = root*).

Více v dodatku *Segmentace vět do stromů a slov do uzlů*.

Text poznámky: povinný a přesně daný. Zapiše se poznámka typu: 2 v 1; 1 ve 2 >.

## 6 Chyba v morfologii

### 6.1 M-tag t

Použit kdy? Poznámka označuje uzly, u kterých byla zjištěna chyba v morfologických kategoriích, v tagu (t-lemma je však správně).

Text poznámky: NEpovinný.

Má-li chyba v morfologickém tagu vliv na podobu t-lematu, vybírá anotátor poznámku *M-lemma* a poznámku *M-tag* již volit nemusí.

## 7 Sekundární předložka

### 7.1 Secondary preposition #

Použit kdy? Poznámka označuje místa s potenciální sekundární předložkou. Náleží jednomu z uzlů, který by v případě, že by šlo o sekundární předložku, nebyl samostatným uzlem stromu.

Text poznámky: povinný a přesně daný. Obsahuje přesné znění potenciální sekundární předložky (a nic jiného).

## 8 Jiný problém

### 8.1 Other ?

Použit kdy? Poznámka označuje místa, kde nastal jiný problémový případ (než vymezený).

Text poznámky: povinný.

**Přehled typů poznámek a informace, zda je/není povinný text poznámky:**

	Typ poznámky	Použit kdy?	Text
<b>M</b>	<b>M-lemma</b>	nesprávné t-lemma kvůli chybě v morfologii	NE
<b>L</b>	<b>T-lemma</b>	jiná chyba v t-lematu	ANO
<b>t</b>	<b>M-tag</b>	chyba v tagu, t-lemma je v pořádku	NE
<b>!</b>	<b>Typo</b>	překlep v t-lematu	NE
<b>!</b>	<b>T-lemma with underscore</b>	ručně opravené t-lemma na víceslovné	NE
<b>!</b>	<b>New valency frame</b>	chybí valenční rámec (návrh nového rámce)	ANO
<b>?</b>	<b>Vallex</b>	problém ve vallexu (jaký)	ANO
<b>w</b>	<b>Word segmentation</b>	chybná segmentace slov do uzlů	ANO
<b>s</b>	<b>Sentence segmentation</b>	chybná segmentace vět do stromů	ANO
<b>!</b>	<b>Translation/Reconstruction</b>	chyba v překladu/rekonstrukci, věta je oantována	ANO
<b>X</b>	<b>Re-generate</b>	chybný překlad (uvést správné znění věty)	ANO
<b>X</b>	<b>Not annotated</b>	větu nelze oantovat (uvést důvod)	ANO
<b>#</b>	<b>Secondary preposition</b>	nová sekundární předložka	ANO
<b>?</b>	<b>Structure</b>	nejistota ve struktuře (uvést proč)	ANO
<b>?</b>	<b>Functor</b>	komentář k nejistotě ve funktoru	ANO
<b>?</b>	<b>Other</b>	jiný problém (uvést jaký)	ANO

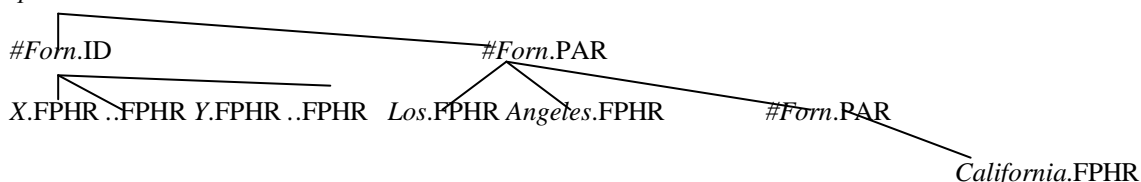
## Šablony

1	Šablony pro konstrukce z „Wall Street Journal“	59
1.1	„Společnost X. Y., Los Angeles, California“	59
1.2	„Zisk se zvýší o více než 20 % oproti 87,7 miliónům dolarů, neboli 1,25 dolarům na akcii za stejné období loňského roku“	60
1.3	„Burzovní seznamy“	61
1.4	„Federální fondy: horní hodnota 8 %, dolní hodnota 7 %, závěrečná nabídka 7,5 %, nabídka 7,9 %“	63
1.5	„Lockheed - 9% dluhopisy v hodnotě 300 milionů dolarů, splatné 15. října 1999, oceněné na 99,90 tak, aby výnos byl 9,39 %.“	64
1.6	„Vkladové certifikáty: 8,09 % jeden měsíc“	65
1.7	„Leon Panetta (demokrat, Kalifornie)“	65
1.8	Bibliografické údaje	66
2	Šablony pro konstrukce z nahrávek Companions	67
2.1	Zahájení rozhovoru	67
2.2	V průběhu rozhovoru	68
2.3	Ukončení rozhovoru	69
2.4	Pozdravy	70
2.5	Zdvořilostní fráze	71
2.6	„Prosím.“	71

## 1 Šablony pro konstrukce z „Wall Street Journal“

### 1.1 „Společnost X. Y., Los Angeles, California“

Obr. 1.1.1: společnost X. Y., Los Angeles, California společnost



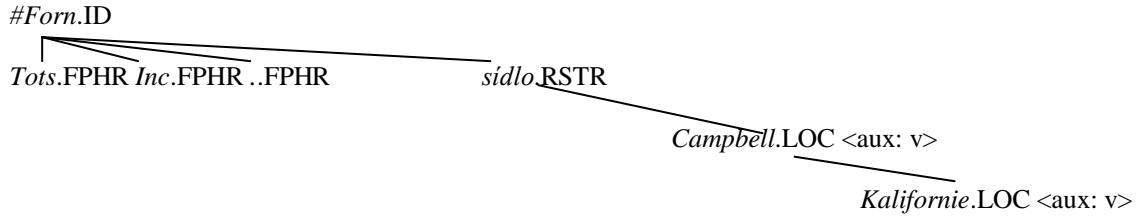
Obr. 1.1.2: společnost X. Y., Los Angeles, Kalifornie společnost



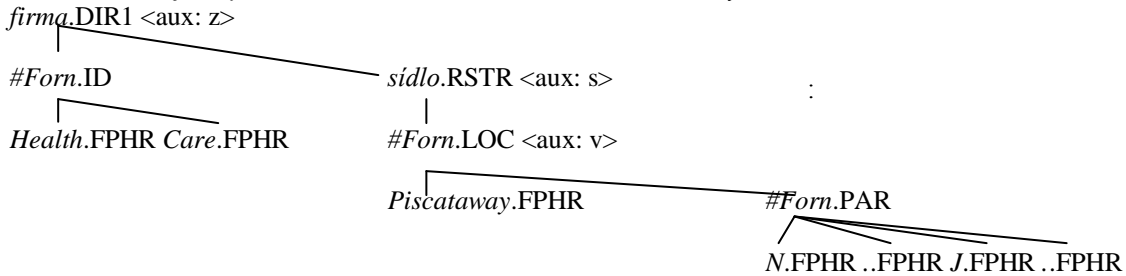
Obr. 1.1.3: v nemocnici Columbia Hospital, Milwaukee, ... nemocnice.LOC <aux: v>



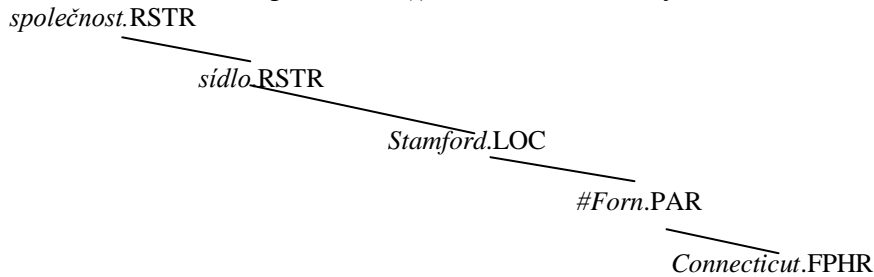
Obr. 1.1.4: *Tiny Tots Inc., sídlem v Campbellu v Kalifornii, ...*



Obr. 1.1.5: *z firmy Health Care(,) se sídlem v Piscataway , N . J.*

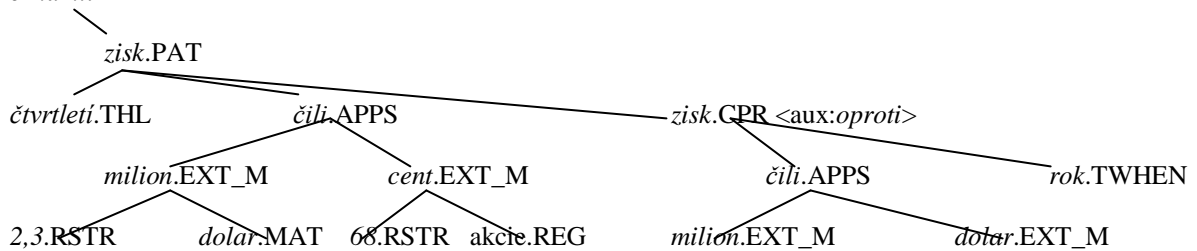


Obr. 1.1.6.: *Mluví společnosti(,) se sídlem ve Stamford, Connecticut, řekl...*

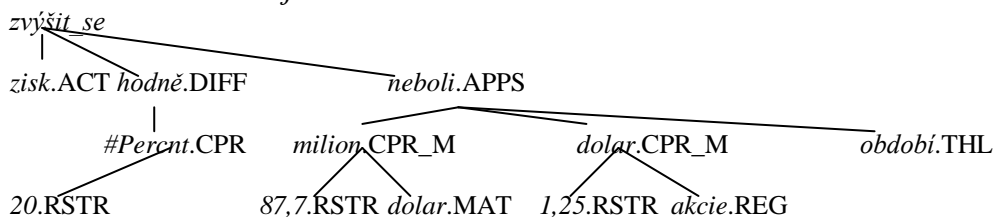


## 1.2 „Zisk se zvýší o více než 20 % oproti 87,7 miliónům dolarů, neboli 1,25 dolarům na akcii za stejné období loňského roku“

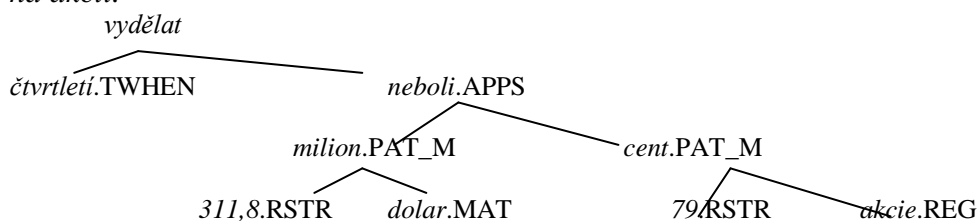
Obr. 1.2.1: *Firma oznámila za třetí čtvrtletí čistý zisk 2,3 milionu dolarů , čili 68 centů na akcii, oproti čistému zisku 5,3 milionu dolarů , čili 1,61 dolaru na akcii, před rokem.*



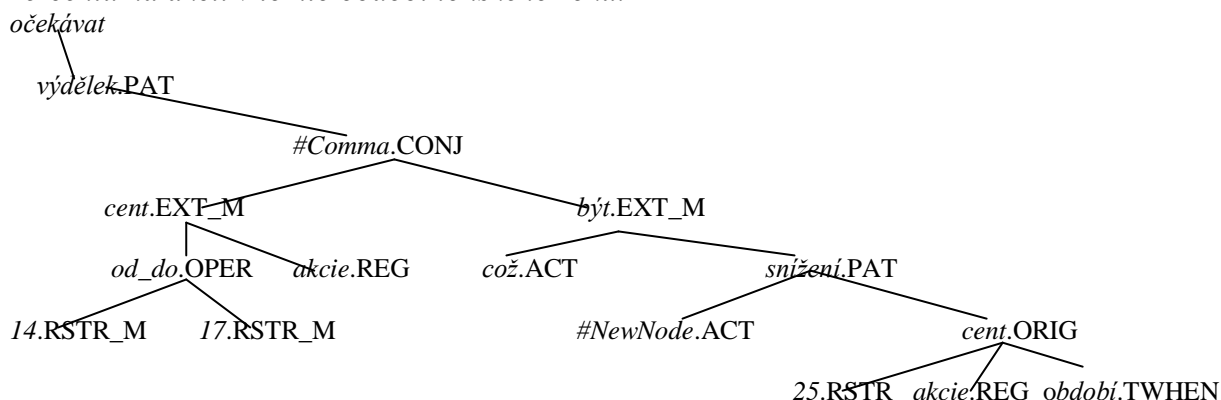
Obr. 1.2.2: *Zisk firmy XYZ se zvýší o více než 20 % oproti 87,7 miliónu dolarů, neboli 1,25 dolaru na akcii za stejné období loňského roku.*



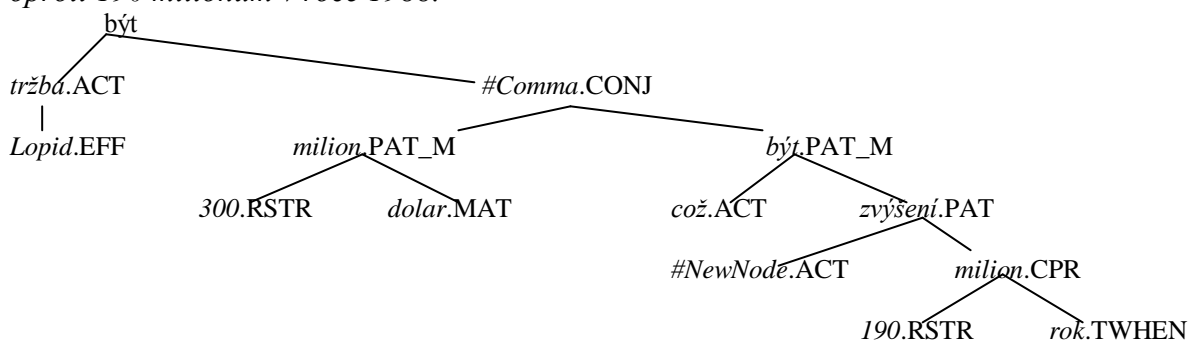
Obr. 1.2.3: Ve třetím čtvrtletí roku 1988 Merck vydělal 311,8 miliónu dolarů, neboli 79 centů na akcii.



Obr. 1.2.4: Společnost uvedla, že očekává výdělek od 14 do 17 centů na akcii, což je snížení z 25 centů na akcii v tomto období loňského roku.



Obr. 1.2.5: Očekává se, že tržby za Lopid budou letos okolo 300 miliónů dolarů, což je zvýšení oproti 190 miliónům v roce 1988.



### 1.3 „Burzovní seznamy“

1.3.1: Columbia Savings & Loan (NYSE ; Symbol : CSV)

1.3.2: Třetí čtvrtletí, 30. září 1989: čistá ztráta 11,527 USD na akcii vs . čistý příjem: 37 centů na akcii

1.3.3: Veřejně obchodované kmenové akcie: 19.6 milionu

1.4.4: H&R Block (Newyorská burza cenných papírů; Symbol: HRB)

1.3.5: Obor podnikání: Daňové poradenství

1.3.6: Rok končící 30. dubna 1989:

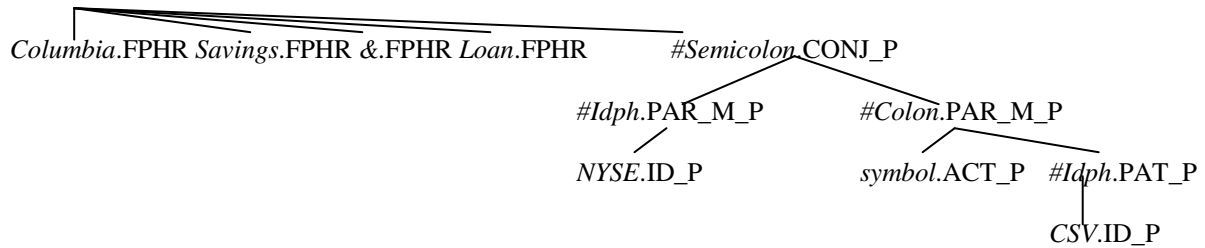
1.3.7: Čistá ztráta: 100,2 milionů dolarů; 1,90 dolaru na akcii

1.3.8: Příjmy na akcii: ztráta 8 centů versus ztráta 9 centů

1.3.9: Pozn.: Do všech údajů "za akcii" byly započítány veškeré výsledky konverzních operací.

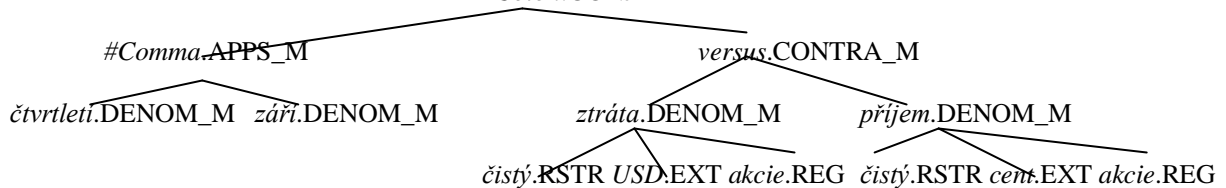
Obr. 1.3.1: *Columbia Savings & Loan (NYSE; Symbol : CSV)*

#Forn.DENOM



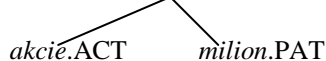
Obr. 1.3.2: *Třetí čtvrtletí, 30. září 1989: čistá ztráta 11,527 USD na akcii vs. čistý příjem: 37 centů na akcii (dvojtečku za „příjem“ v tomto případě zanedbáváme)*

#Colon.CONJ



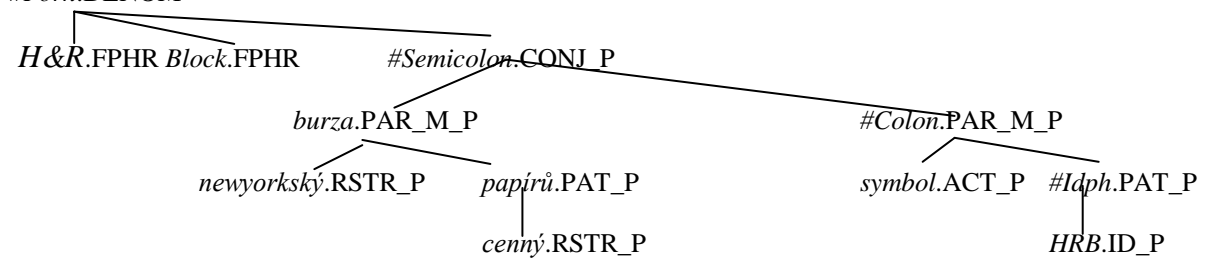
Obr. 1.3.3: *Veřejně obchodované kmenové akcie: 19,6 milionu*

#Colon.PRED



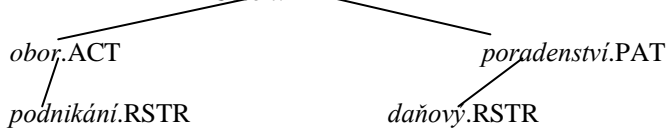
Obr. 1.3.4: *H&R Block (Newyorská burza cenných papírů; Symbol: HRB)*

#Forn.DENOM



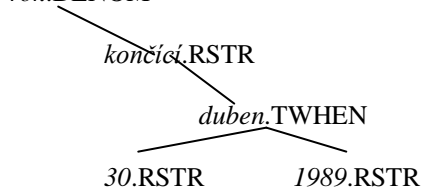
Obr. 1.3.5: *Obor podnikání: Daňové poradenství*

#Colon.PRED



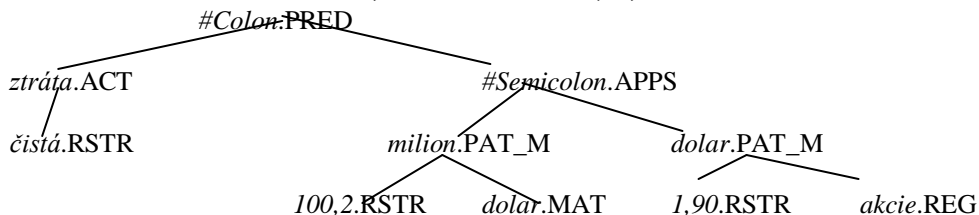
Obr. 1.3.6: *Rok končící 30. dubna 1989:*

rok.DENOM

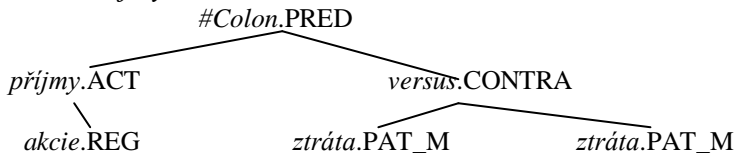




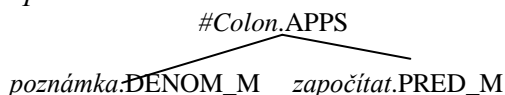
Obr. 1.3.7: Čistá ztráta: 100,2 milionu dolarů; 1,90 dolaru na akcii



1.3.8 Příjmy na akcii: ztráta 8 centů versus ztráta 9 centů

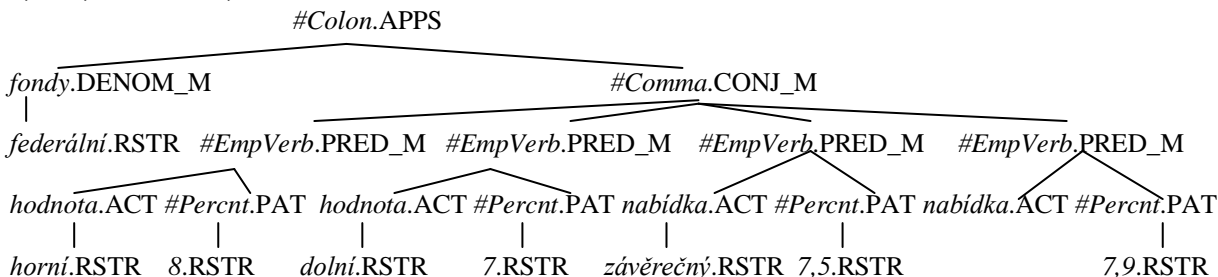


Obr. 1.3.9: Pozn.: Do všech údajů "za akcii" byly započítány veškeré výsledky konverzních operací.

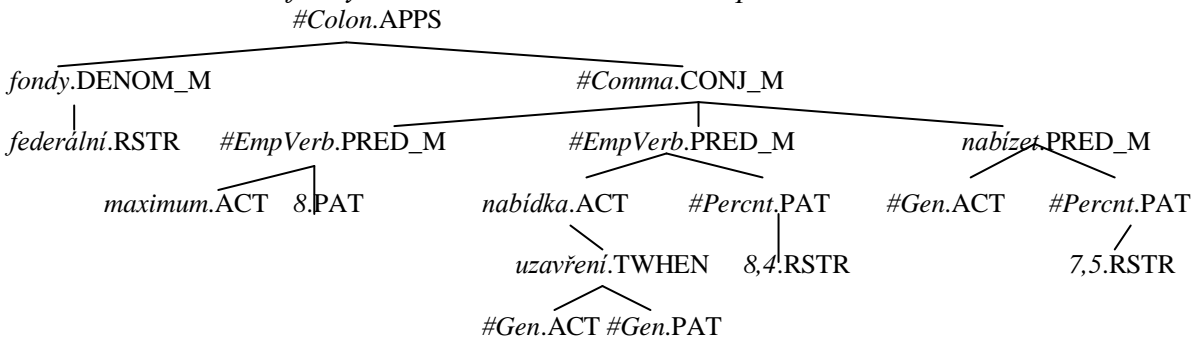


#### 1.4 „Federální fondy: horní hodnota 8 %, dolní hodnota 7 %, závěrečná nabídka 7,5 %, nabídka 7,9 %“

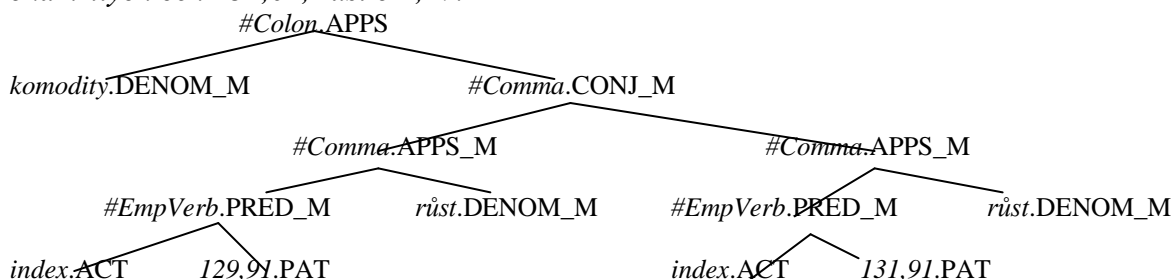
Obr. 1.4.1: FEDERÁLNÍ FONDY: horní hodnota 8 %, dolní hodnota 7 %, závěrečná nabídka 7,5 %, nabídka 7,9 %.



Obr. 1.4.2: Federální fondy: 8 maximum, 8,4 % nabídka před uzavřením, 6 % nabízeno.

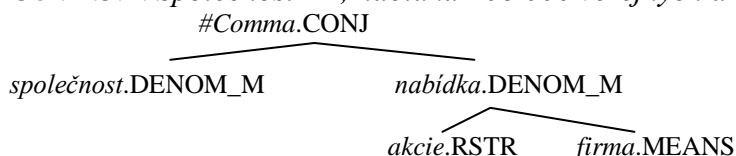


Obr. 1.4.3: *Komodity: Dow-Jonesův index termínových obchodů 129,91, růst o 0,28; index okamžitých cen 131,01, růst o 1,17.*

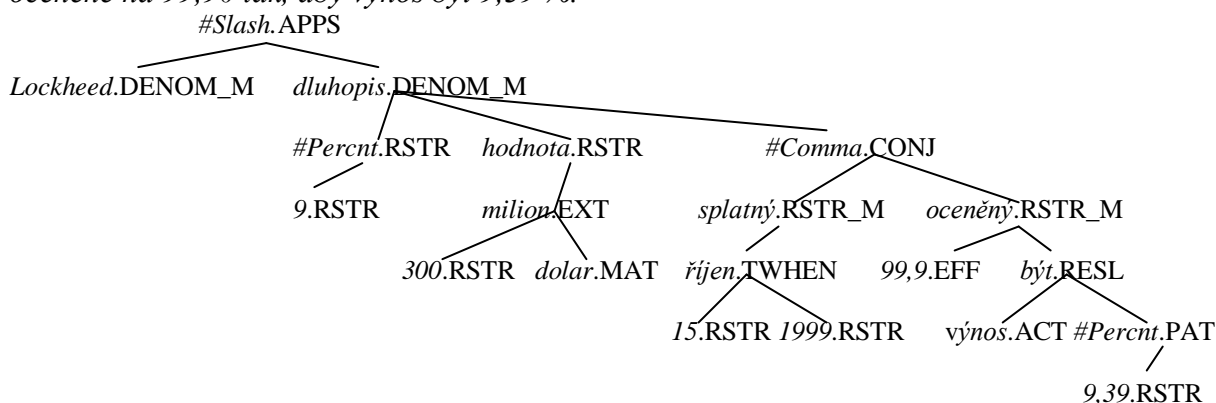


**1.5 „Lockheed - 9% dluhopisy v hodnotě 300 milionů dolarů, splatné 15. října 1999, oceněné na 99,90 tak, aby výnos byl 9,39 %.“**

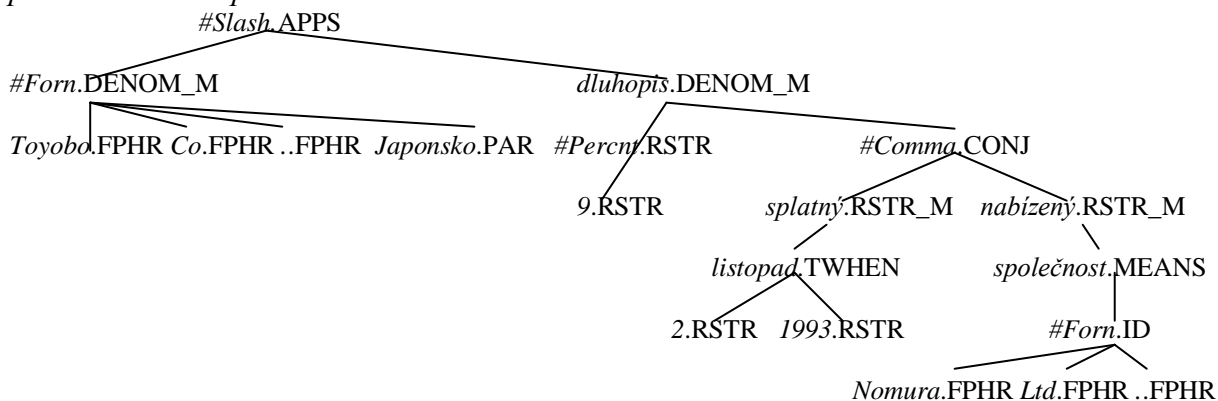
Obr. 1.5.1: *Společnost XY, nabídka 108 000 veřejných akcií(,) prostřednictvím firem AB.*



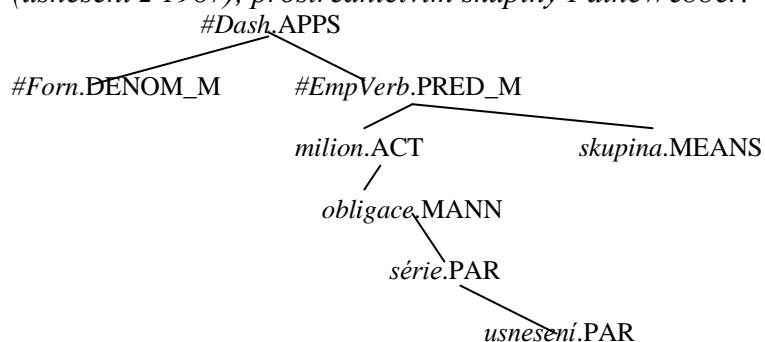
Obr. 1.5.2: *Lockheed - 9% dluhopisy v hodnotě 300 milionů dolarů, splatné 15. října 1999, oceněné na 99,90 tak, aby výnos byl 9,39 %.*



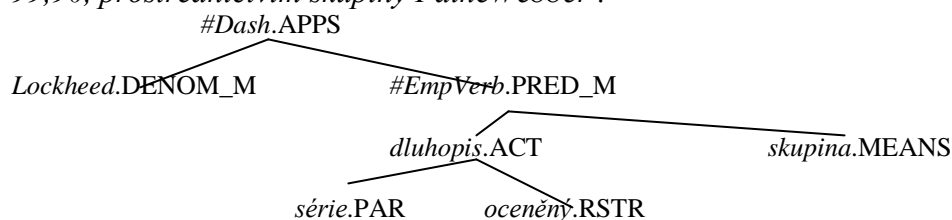
Obr. 1.5.3: *Toyobo Co. (Japan) – 9% dluhopisy splatné 2. listopadu 1993, nabízené prostřednictvím společnosti Nomura Ltd.*



Obr. 1.5.4: *Beverly Hills - 75 milionů dolarů v obecných obligacích z roku 1989, série B (usnesení z 1987), prostřednictvím skupiny PaineWebber.*

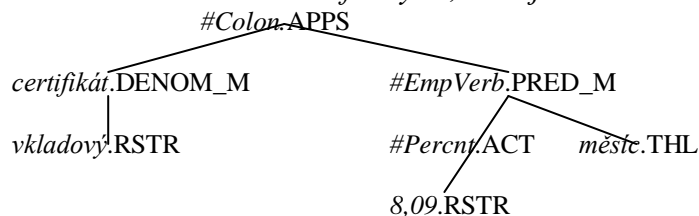


Obr. 1.5.5: *Lockheed - 9% dluhopisy v hodnotě 300 milionů dolarů, série A, oceněné na 99,90, prostřednictvím skupiny PaineWebber.*

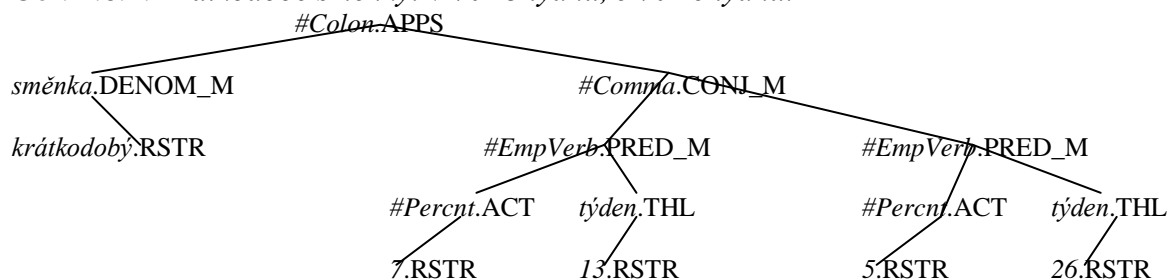


## 1.6 „Vkladové certifikáty: 8,09 % jeden měsíc“

Obr. 1.6.1: *Vkladové certifikáty: 8,09 % jeden měsíc*

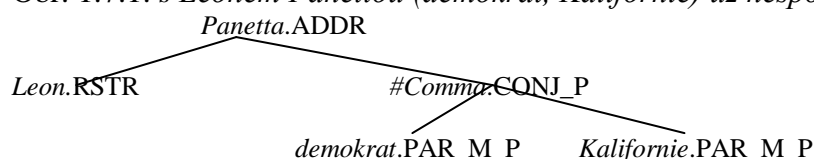


Obr. 1.6.2: *Krátkodobé směnky: 7 % 13 týdnů, 5 % 26 týdnů.*

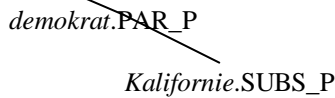


## 1.7 „Leon Panetta (demokrat, Kalifornie)“

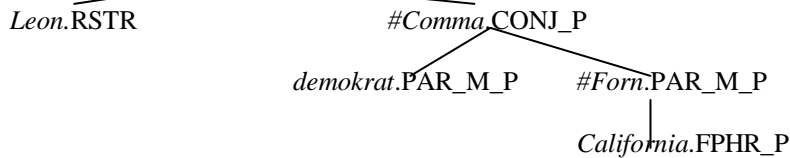
Obr. 1.7.1: *s Leonem Panettou (demokrat, Kalifornie) už nespolupracuji*



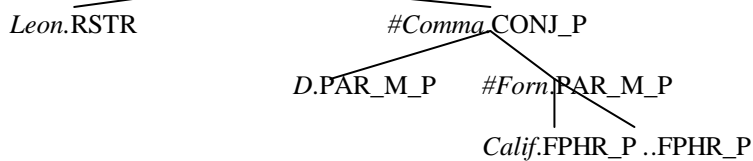
Obr. 1.7.2: s Leonem Panettou (demokrat za Kalifornii) už nespolutracuji  
Panetta.ADDR



Obr. 1.7.3: s Leonem Panettou (demokrat, California) už nespolutracuji  
Panetta.ADDR

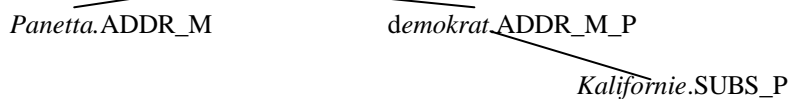


Obr. 1.7.4: s Leonem Panettou (D., Calif.) už nespolutracuji  
Panetta.ADDR



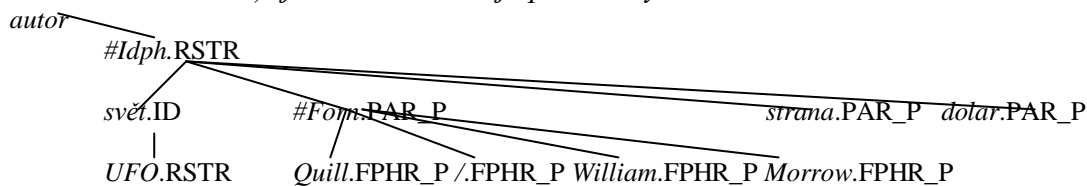
Do poznámky Translation napsat: nepřeložené zkratky.

Obr. 1.7.5: s Leonem Panettou (demokratem za Kalifornii) už nespolutracuji  
#Bracket.APPS\_P

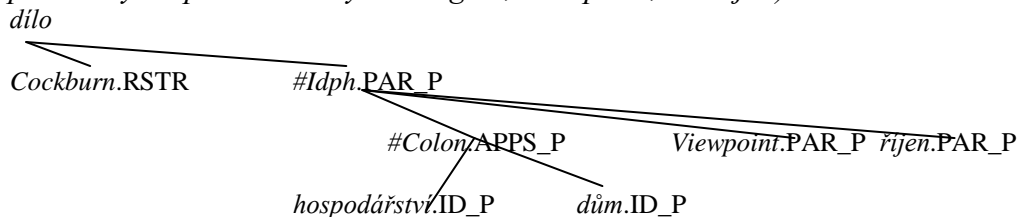


## 1.8 Bibliografické údaje

Obr. 1.8.1: Názor Timothyho Gooda, autora "Do světa UFO" (Quill/William Morrow, 592 stran, 12,95 dolaru), je, že svět se směje příliš brzy.



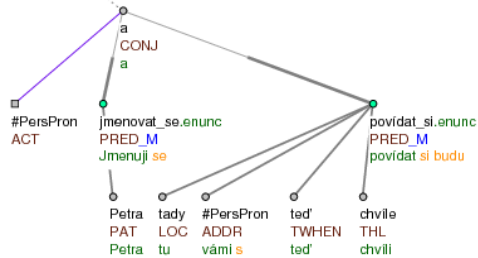
Obr. 1.8.2: Děkuji vtipnému dílu Alexandra Cockburna (Hospodářství Spojených států: Dům postavený na písku rizikových obligací, Viewpoint, 19. října).



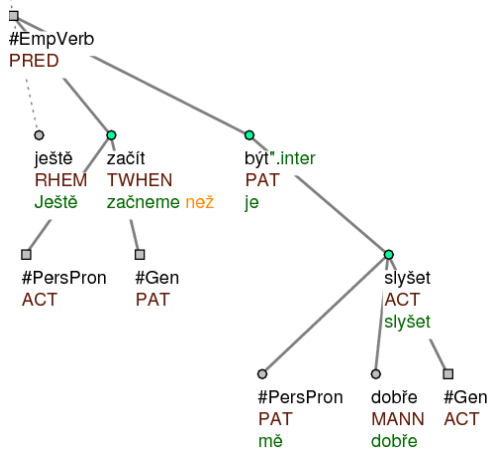
## 2 Šablony pro konstrukce z nahrávek Companions

### 2.1 Zahájení rozhovoru

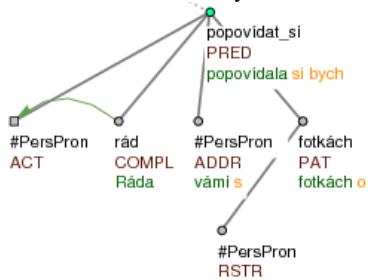
Obr. 2.1.1: *Jmenuji se Petra a budu si tu s vámi teď chvíli povídat.*



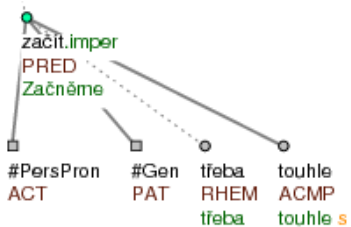
Obr. 2.1.2: *Ještě než začneme, je mě dobře slyšet?*



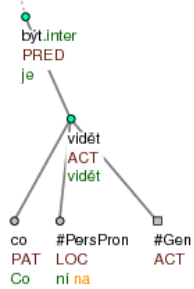
Obr. 2.1.3: *Ráda bych si s vámi popovídala o vašich fotkách.*



Obr. 2.1.4: *Začneme třeba s touhle. / Začneme třeba s touhle.*

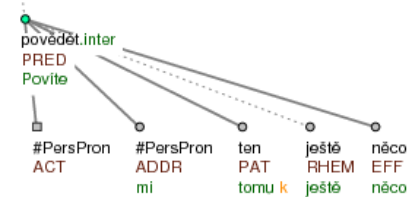


Obr. 2.1.5: *Co je na ní vidět?*

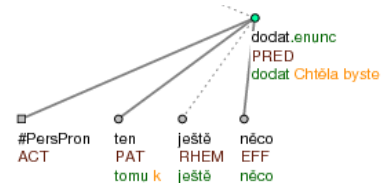


## 2.2 V průběhu rozhovoru

Obr. 2.2.1: *Povíte mi k tomu ještě něco?*

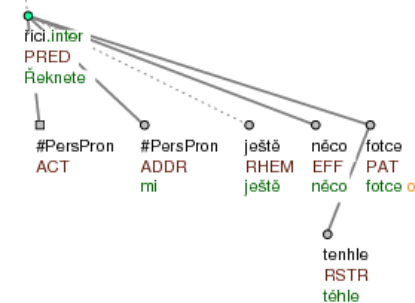


Obr. 2.2.2: *Chtěla byste k tomu ještě něco dodat?*



Obr. 2.2.3:

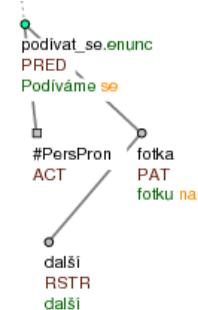
*Řeknete mi ještě něco o téhle fotce.PAT?*  
*Chcete mi říct ještě něco k téhle fotce.REG?* /  
*Co byste mi ještě řekl k této fotce.REG?*



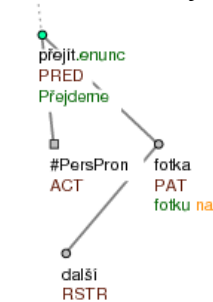
Obr. 2.2.4: *Dobře.*



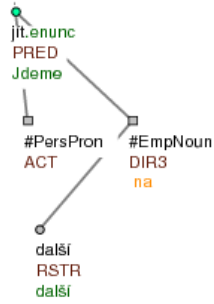
Obr. 2.2.5: *Podíváme se na další fotku.*



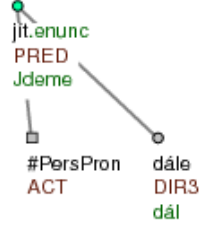
Obr. 2.2.6: *Přejdeme na další fotku.*



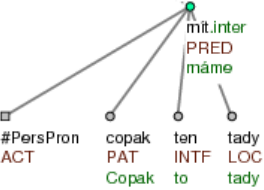
Obr. 2.2.7: *Jdeme na další.*



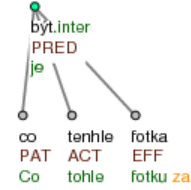
Obr. 2.2.8: *Jdeme dál.*



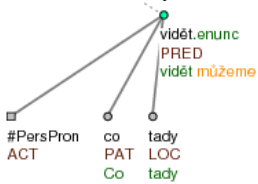
Obr. 2.2.9: *Copak to máme tady?*



Obr. 2.2.10: *Co je tohle za fotku?*

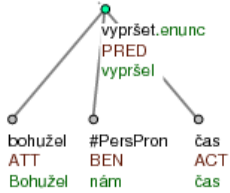


2.2.11 *Co tady můžeme vidět?*



## 2.3 Ukončení rozhovoru

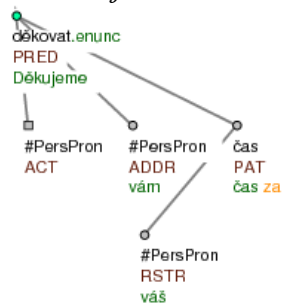
Obr. 2.3.1: *Bohužel nám vypršel čas.*



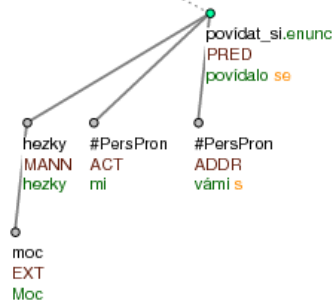
Obr. 2.3.2: *Tohle byla poslední fotka.*



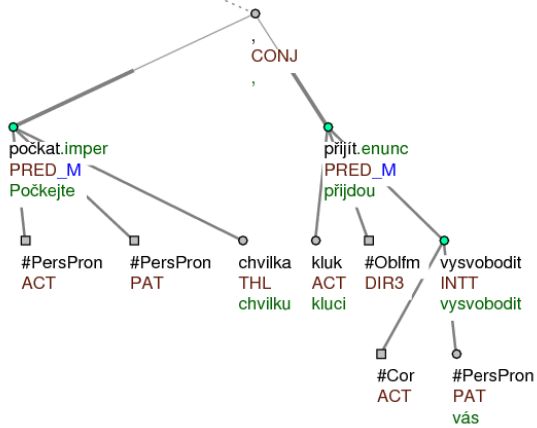
Obr. 2.3.3: *Děkujeme vám za váš čas.*



Obr. 2.3.4: *Moc hezky se mi s vámi povídalo.*



Obr. 2.3.5: *Počkejte chvílku, kluci vás přijdou vysvobodit.*



Obr. 2.3.6: *Konec kazety.*

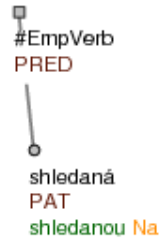


## 2.4 Pozdravy

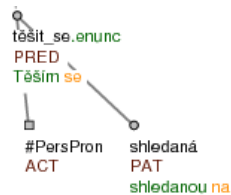
Obr. 2.4.1: *Dobrý den.*



Obr. 2.4.2: *Na shledanou.*



Obr. 2.4.3: *Těším se na shledanou.*



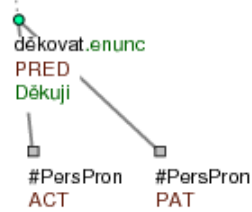


## 2.5 Zdvořilostní fráze

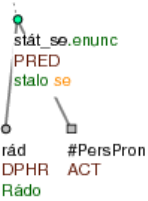
Obr. 2.5.1: *Díky za váš čas.*



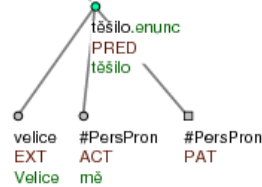
Obr. 2.5.2: *Děkuji.*



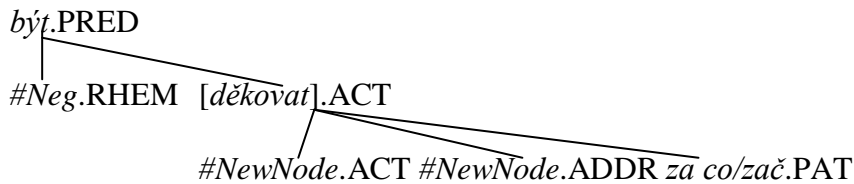
Obr. 2.5.3: *Rádo se stalo.*



Obr. 2.5.4: *Velice mě těšilo.*



Obr. 2.5.5: *Není zač.*



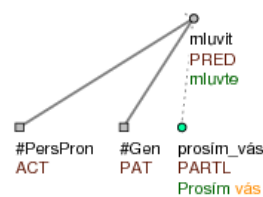
## 2.6 „Prosím.“

2.6.1 (*Děkuji.*) *Prosím.*

Zdvořilost.



Obr. 2.6.2: *Prosím vás, mluvte.*



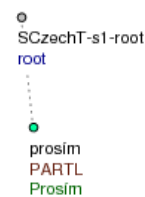
2.6.3: (*Mohu se tě něco zeptat?*) *Prosím.*

Souhlasu, přitakání (= ano).

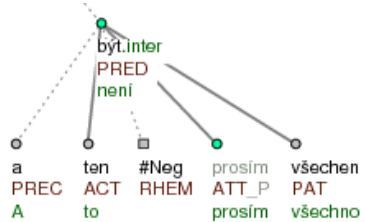


2.6.4: *Prosím? (Nerozuměl jsem.)*

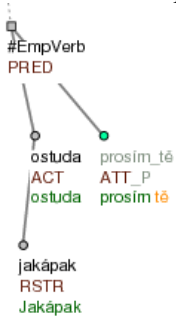
Otázka při nepochopení.



Obr. 2.6.5: *A to není prosím všechno.*  
Zdůraznění.



Obr. 2.6.6: *Jakápak ostuda, prosím tě.*



## Reference

- Hajič a kol. (2011):: *Prague Czech-English Dependency Treebank 2.0*. Data/software, Institute of Formal and Applied Linguistics, Prague, Czech republic, 2011.
- Mikulová, M., Štěpánek, J. (2009): Annotation Procedure in Building the Prague Czech-English Dependency Treebank. In: *Slovko 2009, NLP, Corpus Linguistics, Corpus Based Grammar Research*. Slovenská akadémia vied, Bratislava, Slovakia, ISBN 978-80-7399-875-2, pp. 241-248.
- Hajič a kol. (2006):: *Prague Dependency Treebank 2.0*. Software prototype, Linguistic Data Consortium, Philadelphia, PA, USA, ISBN 1-58563-370-4, www ldc.upenn.edu, 2006.
- Mikulová a kol. (2006a):: *Annotation on the tectogrammatical level in the Prague Dependency Treebank. Annotation manual*. Technical report no. 2006/30, ÚFAL MFF UK, Prague, Czech Rep., 1287 pp., 2006.
- Mikulová a kol. (2006b):: *Annotation on the tectogrammatical level in the Prague Dependency Treebank. Reference book*. Technical report no. 2006/32, ÚFAL MFF UK, Prague, Czech Rep., 193 pp., 2006.
- Mikulová a kol. (2006c): *Anotace na tektogramatické rovině Pražského závislostního korpusu. Referenční příručka*. Technical report no. 2006/31, ÚFAL MFF UK, Prague, Czech Rep., 183 pp., 2006.
- Mikulová a kol. (2005):: *Anotace na tektogramatické rovině Pražského závislostního korpusu. Anotátorská příručka*. Technical report no. 2005/TR-2005-28, ÚFAL MFF UK, Prague, ISSN 1214-5521, 1185 pp., 2005.

## ÚFAL

ÚFAL (Ústav formální a aplikované lingvistiky; <http://ufal.mff.cuni.cz>) is the Institute of Formal and Applied linguistics, at the Faculty of Mathematics and Physics of Charles University, Prague, Czech Republic. The Institute was established in 1990 after the political changes as a continuation of the research work and teaching carried out by the former Laboratory of Algebraic Linguistics since the early 60s at the Faculty of Philosophy and later the Faculty of Mathematics and Physics. Together with the “sister” Institute of Theoretical and Computational Linguistics (Faculty of Arts) we aim at the development of teaching programs and research in the domain of theoretical and computational linguistics at the respective Faculties, collaborating closely with other departments such as the Institute of the Czech National Corpus at the Faculty of Philosophy and the Department of Computer Science at the Faculty of Mathematics and Physics.

## CKL

As of 1 June 2000 the Center for Computational Linguistics (Centrum počítačové lingvistiky; <http://ckl.mff.cuni.cz>) was established as one of the centers of excellence within the governmental program for support of research in the Czech Republic. The center is attached to the Faculty of Mathematics and Physics of Charles University in Prague.

## TECHNICAL REPORTS

The ÚFAL/CKL technical report series has been established with the aim of disseminate topical results of research currently pursued by members, cooperators, or visitors of the Institute. The technical reports published in this Series are results of the research carried out in the research projects supported by the Grant Agency of the Czech Republic, GAČR 405/96/K214 (“Komplexní program”), GAČR 405/96/0198 (Treebank project), grant of the Ministry of Education of the Czech Republic VS 96151, and project of the Ministry of Education of the Czech Republic LN00A063 (Center for Computational Linguistics). Since November 1996, the following reports have been published.

- ÚFAL TR-1996-01** Eva Hajičová, *The Past and Present of Computational Linguistics at Charles University*  
Jan Hajič and Barbora Hladká, *Probabilistic and Rule-Based Tagging of an Inflective Language – A Comparison*
- ÚFAL TR-1997-02** Vladislav Kuboň, Tomáš Holan and Martin Plátek, *A Grammar-Checker for Czech*
- ÚFAL TR-1997-03** Alla Bémová at al., *Anotace na analytické rovině, Návod pro anotátory (in Czech)*
- ÚFAL TR-1997-04** Jan Hajič and Barbora Hladká, *Tagging Inflective Languages: Prediction of Morphological Categories for a Rich, Structural Tagset*
- ÚFAL TR-1998-05** Geert-Jan M. Kruijff, *Basic Dependency-Based Logical Grammar*
- ÚFAL TR-1999-06** Vladislav Kuboň, *A Robust Parser for Czech*
- ÚFAL TR-1999-07** Eva Hajičová, Jarmila Panevová and Petr Sgall, *Manuál pro tektogramatické značkování (in Czech)*
- ÚFAL TR-2000-08** Tomáš Holan, Vladislav Kuboň, Karel Oliva, Martin Plátek, *On Complexity of Word Order*
- ÚFAL/CKL TR-2000-09** Eva Hajičová, Jarmila Panevová and Petr Sgall, *A Manual for Tectogrammatical Tagging of the Prague Dependency Treebank*
- ÚFAL/CKL TR-2001-10** Zdeněk Žabokrtský, *Automatic Functor Assignment in the Prague Dependency Treebank*
- ÚFAL/CKL TR-2001-11** Markéta Straňáková, *Homonymie předložkových skupin v češtině a možnost jejich automatického zpracování*
- ÚFAL/CKL TR-2001-12** Eva Hajičová, Jarmila Panevová and Petr Sgall, *Manuál pro tektogramatické značkování (III. verze)*

- ÚFAL/CKL TR-2002-13 Pavel Pecina and Martin Holub, *Sémanticky signifikantní kolokace*
- ÚFAL/CKL TR-2002-14 Jiří Hana, Hana Hanová, *Manual for Morphological Annotation*
- ÚFAL/CKL TR-2002-15 Markéta Lopatková, Zdeněk Žabokrtský, Karolína Skwarská and Vendula Benešová, *Tektogramaticky anotovaný valenční slovník českých sloves*
- ÚFAL/CKL TR-2002-16 Radu Gramatovici and Martin Plátek, *D-trivial Dependency Grammars with Global Word-Order Restrictions*
- ÚFAL/CKL TR-2003-17 Pavel Květoň, *Language for Grammatical Rules*
- ÚFAL/CKL TR-2003-18 Markéta Lopatková, Zdeněk Žabokrtský, Karolína Skwarska, Václava Benešová, *Valency Lexicon of Czech Verbs VALLEX 1.0*
- ÚFAL/CKL TR-2003-19 Lucie Kučová, Veronika Kolářová, Zdeněk Žabokrtský, Petr Pajas, Oliver Čulo, *Anotování koreference v Pražském závislostním korpusu*
- ÚFAL/CKL TR-2003-20 Kateřina Veselá, Jiří Havelka, *Anotování aktuálního členění věty v Pražském závislostním korpusu*
- ÚFAL/CKL TR-2004-21 Silvie Cinková, *Manuál pro tektogramatickou anotaci angličtiny*
- ÚFAL/CKL TR-2004-22 Daniel Zeman, *Neprojektivity v Pražském závislostním korpusu (PDT)*
- ÚFAL/CKL TR-2004-23 Jan Hajič a kol., *Anotace na analytické rovině, návod pro anotátory*
- ÚFAL/CKL TR-2004-24 Jan Hajič, Zdeňka Uřešová, Alevtina Bémová, Marie Kaplanová, *Anotace na tektogramatické rovině (úroveň 3)*
- ÚFAL/CKL TR-2004-25 Jan Hajič, Zdeňka Uřešová, Alevtina Bémová, Marie Kaplanová, *The Prague Dependency Treebank, Annotation on tectogrammatical level*
- ÚFAL/CKL TR-2004-26 Martin Holub, Jiří Diviš, Jan Pávek, Pavel Pecina, Jiří Semecký, *Topics of Texts. Annotation, Automatic Searching and Indexing*
- ÚFAL/CKL TR-2005-27 Jiří Hana, Daniel Zeman, *Manual for Morphological Annotation (Revision for PDT 2.0)*
- ÚFAL/CKL TR-2005-28 Marie Mikulová a kol., *Pražský závislostní korpus (The Prague Dependency Treebank) Anotace na tektogramatické rovině (úroveň 3)*
- ÚFAL/CKL TR-2005-29 Petr Pajas, Jan Štěpánek, *A Generic XML-Based Format for Structured Linguistic Annotation and Its application to the Prague Dependency Treebank 2.0*
- ÚFAL/CKL TR-2006-30 Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolařová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, Magda Razimová, Petr Sgall, Jan Štěpánek, Zdeňka Uřešová, Kateřina Veselá, Zdeněk Žabokrtský, *Annotation on the tectogrammatical level in the Prague Dependency Treebank (Annotation manual)*
- ÚFAL/CKL TR-2006-31 Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolařová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, Petr Sgall, Magda Ševčíková, Jan Štěpánek, Zdeňka Uřešová, Kateřina Veselá, Zdeněk Žabokrtský, *Anotace na tektogramatické rovině Pražského závislostního korpusu (Referenční příručka)*
- ÚFAL/CKL TR-2006-32 Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolařová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, Petr Sgall, Magda Ševčíková, Jan Štěpánek, Zdeňka Uřešová, Kateřina Veselá, Zdeněk Žabokrtský, *Annotation on the tectogrammatical level in the Prague Dependency Treebank (Reference book)*
- ÚFAL/CKL TR-2006-33 Jan Hajič, Marie Mikulová, Martina Otradovcová, Petr Pajas, Petr Podveský, Zdeňka Uřešová, *Pražský závislostní korpus mluvené češtiny. Rekonstrukce standardizovaného textu z mluvené řeči*
- ÚFAL/CKL TR-2006-34 Markéta Lopatková, Zdeněk Žabokrtský, Václava Benešová (in cooperation with Karolína Skwarska, Klára Hrstková, Michaela Nová, Eduard Bejček, Miroslav Tichý) *Valency Lexicon of Czech Verbs. VALLEX 2.0*
- ÚFAL/CKL TR-2006-35 Silvie Cinková, Jan Hajič, Marie Mikulová, Lucie Mladová, Anja Nedolužko, Petr Pajas, Jarmila Panevová, Jiří Semecký, Jana Šindlerová, Josef Toman, Zdeňka Uřešová, Zdeněk Žabokrtský, *Annotation of English on the tectogrammatical level*
- ÚFAL/CKL TR-2007-36 Magda Ševčíková, Zdeněk Žabokrtský, Oldřich Krůza, *Zpracování pojmenovaných entit v českých textech*
- ÚFAL/CKL TR-2008-37 Silvie Cinková, Marie Mikulová, *Spontaneous speech reconstruction for the syntactic and semantic analysis of the NAP corpus*

- ÚFAL/CKL TR-2008-38 Marie Mikulová, *Rekonstrukce standardizovaného textu z mluvené řeči v Pražském závislostním korpusu mluvené češtiny. Manuál pro anotátory*
- ÚFAL/CKL TR-2008-39 Zdeněk Žabokrtský, Ondřej Bojar, *TectoMT, Developer's Guide*
- ÚFAL/CKL TR-2008-40 Lucie Mladová, *Diskurzni vztahy v češtině a jejich zachycení v Pražském závislostním korpusu 2.0*
- ÚFAL/CKL TR-2009-41 Marie Mikulová, *Pokyny k překladu určené překladatelům, revizorům a korektorům textů z Wall Street Journal pro projekt PCEDT*
- ÚFAL/CKL TR-2011-42 Loganathan Ramasamy, Zdeněk Žabokrtský, *Tamil Dependency Treebank (TamilTB) – 0.1 Annotation Manual*
- ÚFAL/CKL TR-2011-43 Ngųy Giang Linh, Michal Novák, Anna Nedoluzhko, *Coreference Resolution in the Prague Dependency Treebank*
- ÚFAL/CKL TR-2011-44 Anna Nedoluzhko, Jiří Mirovský, *Annotating Extended Textual Coreference and Bridging Relations in the Prague Dependency Treebank*
- ÚFAL/CKL TR-2011-45 David Mareček, Zdeněk Žabokrtský, *Unsupervised Dependency Parsing*
- ÚFAL/CKL TR-2011-46 Martin Majliš, Zdeněk Žabokrtský, *W2C – Large Multilingual Corpus*
- ÚFAL TR-2012-47 Lucie Poláková, Pavlína Jínová, Šárka Zikánová, Zuzanna Bedřichová, Jiří Mirovský, Magdaléna Rysová, Jana Zdeňková, Veronika Pavlíková, Eva Hajičová, *Manual for annotation of discourse relations in the Prague Dependency Treebank*
- ÚFAL TR-2012-48 Nathan Green, Zdeněk Žabokrtský, *Ensemble Parsing and its Effect on Machine Translation*
- ÚFAL TR-2013-49 David Mareček, Martin Popel, Loganathan Ramasamy, Jan Štěpánek, Daniel Zemana, Zdeněk Žabokrtský, Jan Hajič *Cross-language Study on Influence of Coordination Style on Dependency Parsing Performance*
- ÚFAL TR-2013-50 Jan Berka, Ondřej Bojar, Mark Fishel, Maja Popović, Daniel Zeman, *Tools for Machine Translation Quality Inspection*
- ÚFAL TR-2013-51 Marie Mikulová, *Anotace na tektogramatické rovině. Dodatky k anotátorské příručce (s ohledem na anotování PDTSC a PCEDT)*