



# (Před-)anotace aktuálního členění v české (a anglické) části PCEDT

Magdaléna a Kateřina Rysovy, Jiří Mírovský

GAČR P406/12/0658 (Koreference, diskurs a aktuální členění v kontrastivním pohledu)  
GAČR P406/2010/0875 (Komputační lingvistika: Explicitní popis jazyka a anotovaná data se zřetelem na češtinu)

# Anotace TFA v PCEDT

## Česká část

(od ledna 2013, Magdaléna Rysová)

- *anotační prostředí*
  - pro češtinu hotovo
- *předanotace*
  - pro češtinu hotovo
- *anotace 5 tis. vět + 1 tis. v anglické části*
  - do června 2013

# Anotace TFA v PCEDT

## *Anglická část*

*(od dubna 2013, Kateřina Rysová)*

- *anotační prostředí*
  - *v dubnu*
- *předanotace*
  - *průběžně podle českých dat*
- *anotace 5 tis. vět + 1 tis. v české části*
  - *do července 2013*



# Anotační prostředí (pro češtinu)

*součást rozšíření TrEdu*

*Wall Street Journal  
Annotation (wsj-anot)*

*mód PML\_Cz\_T\_TFA*

# Anotační prostředí

TEd ver. 2.4982 Default (1/1): /home/mirovsky/wsJ0003.cz.t

File Node Tree View Macros Setup Help

Mgde: PML\_Cz\_T\_TFA

Style: PML\_T\_TFA

"Není pochyb, že se někteří z těchto dělníků a manažerů nakazili nemocemi spojenými s azbestem," řekl Darrell Phillips, viceprezident společnosti Hollingsworth & Vose pro lidské zdroje. 28/30

Node Attributes

Hide empty values

atree.rf	a#a-wsj0003-001-p1s27
deepord	0
eng_sentence	``There's no question that some of t
id	T-wsj0003-001-p1s27
nodetype	root

T-wsj0003-001-p1s27  
root

Current File List  
Browse File System  
List of Macros

a:

Scale: 100%

# Anotační prostředí

*TFA-relevantní uzly  
všechny až na:*

- *root*
- *nodetype=coap* nebo *fphr*,
- *functor=CM*
- *pro jistotu i functor=FPHR*

# Anotační prostředí

TrEd ver. 2.4982 Default(1/1): /ha/home/mirovsky/ufal/Majda\_Rysova/anotace\_TFA/data\_3/ws0034.cz.t

File Node Tree View Macros Setup Help Mode: PML\_Cz\_T\_TFA

Style: PML\_T\_TFA

A aby toho dosáhly, rozhazují firmy své síť široko daleko. 16/39

**Node Attributes**

Hide empty values

atree.rf	a#a-wsj0034-001-p1s15
deepord	0
eng_sentence	And the firms are stretching their
id	T-wsj0034-001-p1s15
nodetype	root

Current File List  
Browse File System  
List of Macros

a:

T-wsj0034-001-p1s15  
root

rozhazovat f  
PREL  
rozhazují

dosáhnout f  
AIM  
dosáhly aby

firma  
ACT  
firmy

sít f  
PAT  
sítě

široko  
LOC  
široko

a  
PREC  
A

ten t  
PAT  
toho

#PersPron t  
ACT  
najít

#PersPron t  
RSTR  
své

#Cor  
PAT  
široko

daleko  
DPHR  
daleko

Scale: 100%

# Anotační prostředí

TrEd ver. 2.4982 Default(1/1): /ha/home/mirovsky/ufal/Majda\_Rysova/annotace\_TFA/data\_3/ws0034.cz.t

File Node Tree View Macros Setup Help Mode: PML\_Cz\_T\_TFA

Style: PML\_T\_TFA

A aby toho dosáhly, rozhazují firmy své sítě široko daleko. 16/39

**Node Attributes**

Hide empty values

a	Structure
lex.rf	a#a-wsj0034-001-p1s15W1-Ta
annot_comment	Unordered list
-	Structure
text	M_1t
type	TFA_log
deepord	1
functor	PREC
id	T-wsj0034-001-p1s15a1
nodetype	atom
t_lemma	a
tfa	t

Current File List  
Browse File System  
List of Macros

a: [A](#)

Scale: 100%

```
graph TD
    Root((T-wsj0034-001-p1s15 root)) --- Node1((rozhazovat f  
PREC))
    Node1 --- Node2((dosáhnout f  
AIM))
    Node1 --- Node3((firma  
ACT))
    Node1 --- Node4((sít f  
PAT))
    Node1 --- Node5((široko  
LOC))
    Node2 --- Node6((dosáhly aby  
AIM))
    Node2 --- Node7((ten t najít  
ACT))
    Node2 --- Node8((#PersPron t  
ACT))
    Node6 --- Node9((a t  
PREC))
    Node6 --- Node10((toho  
PAT))
    Node3 --- Node11((firmy  
ACT))
    Node4 --- Node12((sítě  
PAT))
    Node5 --- Node13((široko  
LOC))
    Node5 --- Node14((daleko  
DPHR))
```



# Anotační prostředí

TrEd ver. 2.4982 Default(1/1): /ha/home/mirovsky/ufal/Majda\_Rysova/anotace\_TFA/data\_3/wsj0034.cz.t

File Node Tree View Macros Setup Help Mode: PML\_Cz\_T\_TFA

A aby toho dosáhly, rozhazují firmy své sítě široko daleko.

16/39

**Node Attributes**

Hide empty values

a	Structure
lex.rf	a#a-wsj0034-001-p1s15W11-Ta
annot_comment	Unordered list
text	M_1f
type	TFA_log
deepord	10
functor	DPHR
gram	Structure
sempos	adv.denot.grad.neg
id	T-wsj0034-001-p1s15a10
nodetype	dphr
t_lemma	daleko
tfa	f

Current File List  
Browse File System  
List of Macros

a: [daleko](#)

Scale: 100%

# Anotační prostředí

TrEd ver. 2.4982 Default(1/1): /ha/home/mirovsky/ufal/Majda\_Rysova/annotace\_TFA/data\_3/ws0034.cz.t

File Node Tree View Macros Setup Help

Mgde: PML\_Cz\_T\_TFA

Style: PML\_T\_TFA

16/39

A aby toho dosáhly, rozhazují firmy své síť široko daleko.

**Node Attributes**

Hide empty values

a	Structure
└ aux.rf	Unordered list
└ lex.rf	Unordered list
└ annot_comment	Unordered list
└ text	A1_vAIM_f@A2_mssr@M_1c@A2_mssl
└ type	TFA_log
deepord	5
functor	AIM
gram	Structure
└ sempos	v
id	T-wsj0034-001-p1s15a4
nodetype	complex
t_lemma	dosáhnout
tfa	c
val_frame.rf	v#v-w714f1

Current File List  
Browse File System  
List of Macros

frame: Valency lexicon not found or not loaded! a: dosáhly, aby

Scale: 100%

```
graph TD
    Root["T-wsj0034-001-p1s15 root"] --- Node1(( ))
    Node1 --- Node2["rozhazovat f  
PRED  
rozhazují"]
    Node1 --- Node3(( ))
    Node1 --- Node4(( ))
    Node2 --- Node5["firma t  
ACT  
firmy"]
    Node2 --- Node6["dosáhnout c  
AIM  
dosáhly aby"]
    Node2 --- Node7["sít f  
PAT  
síť"]
    Node2 --- Node8["široko f  
LOC  
široko"]
    Node6 --- Node9["a t  
PREC  
A"]
    Node6 --- Node10["ten t najít  
PAT  
toho"]
    Node6 --- Node11["#PersPron t #Cor  
ACT"]
    Node7 --- Node12["#PersPron t #Cor  
RSTR  
své"]
    Node8 --- Node13["daleko f  
DPHR  
daleko"]
```

# Anotační prostředí záznam akcí

*A1\_vAIM\_f@A2\_mssr@M\_1c@A2\_mssl*

- **A1\_vAIM\_f** – AIM syn slovesa během předanotace nastaven na 'f'
- **A2\_mssr** – automatický posun vpravo během anotace
- **M\_1c** – ruční nastavení hodnoty 'c'
- **A2\_mssl** – automatický posun vlevo během anotace

# Anotační prostředí omezení

- *převěšovat lze pouze uzly s  
funktoem **RHEM**.*
- *povolen **přesun uzlů či  
podstromů** vpravo či vlevo (o  
uzel či o celý podstrom)*
- ***funktor lze měnit jen**  
RHEM ↔ vybraná skupina funktorů*



# Anotační prostředí podpora

- **automatické přerovnání** synů uzlu podle jejich hodnoty *tfa*
- speciální podmínky přerovnání **u slovesných a jmenných frází**
- **coap uzel** přerovnán, pokud všichni jeho synové mají stejnou hodnotu *tfa*

# Předanotace TFA

## Česká část

### **Využívá:**

- *anglické části PCEDT*
- *převažujících hodnot tfa u funktorů, lemmat apod. v PDT*

*Vždy očekávaná úspěšnost **≥ 90%***

- *až na využití angl. části (kde nevíme)*



# Předanotace TFA

**1096** vět, **25 342** tokeny

TFA-relevantní uzly: **21 566**

Předanotováno uzlů: **8 687**

Pohyby uzlů: **1 284** doprava, **377** doleva

# Předanotace TFA

*Anglická část*

***Chceme využít:***

- *české, již anotované části PCEDT*

*Očekávaná úspěšnost*

- *zatím nevíme*





**Děkuji za pozornost!**