PDT – Tectogrammatical Layer
( Intro )

Markéta Lopatková

Institute of Formal and Applied Linguistics, MFF UK

lopatkova@ufal.mff.cuni.cz
PDT: t-layer

- deep syntactic structure of a sentence (coarse semantics)
  ~ (disambiguated) sentence meaning
- dependency-based 'trees'
  - with (phrase-structure like) 'connecting' constructions for coordination
  - with coreference
- disambiguated … synonymy vs. ambiguity
- 'completed' … elided words added / restored / copied
  - 'missing' argument
  - coordination: stripping and gapping
- only lexical words as separate nodes
  - auxiliary and modal verbs
  - prepositions, subordinating conjunctions, …
  - BUT coordinating conjunctions
  - plus technical root

documentation: http://ufal.mff.cuni.cz/pdt3.5
Many industry analysts had been projecting a 3% decline in placements for the quarter.
PDT: t-layer

- dependency-based 'trees'
  - with (phrase-structure like) 'connecting' constructions for coordination

Most traders and analysts attributed the decline to technical factors …
PDT: t-layer

- dependency-based 'trees'
  - with (phrase-structure like) 'connecting' constructions for coordination
  - with coreference

A shortage of young cattle has made them more expensive for feedlot operators to buy (= to buy it = to buy the cattle) .
A shortage of young cattle has made them more expensive for feedlot operators to buy (= to buy it = to buy the cattle).
Long-term support for the December contract was believed to be at $1.25 a pound. (ACT = who believed)
More provocatively, some conservative legal theorists have begun arguing that ... ( = the theorists argue)
PDT: t-layer

- 'completed' structure
  - coordination: stripping and gapping

Koupil červené a bílé víno. (=lit. (He) bought red and white wine.)
= He bought red wine and white wine.
PDT: t-layer

- 'completed' structure
  - coordination: stripping and gapping

Jirka prozradil, že on navštívil Marii a Honzu Jiřinu. (=lit. George disclosed that he visited Mary and John Henriette.)
= George disclosed that he visited Mary and John visited Henriette.

Dependency Grammars
Many industry analysts had been projecting a 3% decline in placements for the quarter.
PDT: t-layer

• lexical words (cf. semantemes in FGD)

1) lexical items: **t-lemma**
link to the PDT-Vallex / EngVallex dictionary for verbs and selected nouns

2) meaning of morphological categories: **grammatemes**
(a pair of) trousers ... singular
*denominating* (pojmenovávací) vs. *correlating* (usouvztažňující) categories

3) deep syntactic structure: **functors + subfunctors**
*valency* as a key property → a sentence syntactic structure

Many industry analysts had been projecting a 3% decline …
PDT: t-layer

• lexical words (cf. semantemes in FGD)
  1) lexical items: *t-lemma*
     link to the PDT-Vallex / EngVallex dictionary for verbs and selected nouns
  2) meaning of morphological categories: *grammatemes*
     (a pair of) trousers … singular
     denominating (pojmenovávací) vs. correlating (usouvztažňující) categories
  3) deep syntactic structure: *functors + subfunctors*
     valency as a key property → a sentence syntactic structure

• additional info:
  • coreferential links
  • 'word order': increasing communicative dynamism
    topic focus articulation (cz: aktuální členění)

• additional annotation in PDT 3.5:
  • multiwords expressions, extended coreference and discourse relations, genre specification, …
### T-node attributes and their values in PDT 2.0

**Notation:**
- **attribute** – attribute name
- **value** – attribute value

#### A. Lexical content
- **t_lemma** – textgrammatical lemma
- **val_frame** – valency frame (reference to PDT-VALLEX)

#### B. Semantic roles and other structural relations
- **functor** – role of the node within the t-tree structure
- **act** – act
- **pat** – patient
- **addr** – addressee
- **orig** – origin
- **eff** – effect
- **temporal modifiers:**
  - **twhen** – when
  - **tfhll** – for how long
  - **tfwh** – from when
  - **thl** – how long
  - **tho** – how often
  - **towh** – to when
  - **tpar** – temporal parallel
  - **tsin** – since when

#### Other:
- **compl** – predicative complement
- **ben** – benefactor
- **contrd** – contradiction
- **rc** – heritage
- **rhem** – reformatio
- **subs** – substitution
- **att** – attitude
- **intf** – intensifier
- **mod** – modality
- **prec** – reference to preceding text
- **fphr** – foreign language expression

#### Subfunctor
- **abstr** – abstract space
- **along** – along
- **around** – around
- **above** – above
- **behind** – behind
- **below** – below
- **betw** – between
- **elsew** – elsewhere
- **ext** – extent
- **front** – in front of
- **near** – near
- **opp** – opposite
- **target** – target
- **to** – to
- **across** – across

#### Values specific to ACMP:
- **circ** – circumstance
- **incl** – inclusion
- **wout** – negative accompaniment
- **depord** – non-negative integer representing deep word order

#### Values specific to CPR:
- **than** – difference

#### Values specific to TWHEN:
- **after** – after
- **approx** – approximately
- **before** – before
- **begin** – at the beginning
- **betw** – between
- **end** – at the end
- **flow** – in the course of
- **mid** – in the middle of

#### D. Coreference and predicative complement
- **coref_gram** – (list of) reference(s) to antecedent(s) in the sense of grammatical coreference
- **coref_text** – (list of) reference(s) to antecedent(s) in the sense of textual coreference
- **coref_special** – special types of coreference (without obvious t-node antecedent)

#### E. Types of t-nodes
- **nodetype** – basic node classification
- **root** – technical root
- **complex** – complex node
- **qcomplex** – quasi-complex node
- **atom** – atomic node
- **coap** – paratactic structure root (coordination or apposition)
- **depord** – dependent part of an idiomatic expression
- **fphr** – part of a foreign-language expression
- **list** – root node of a list structure

#### Communicative dynamism
- **tfa** – topic/focus articulation
  - **1** – non-contrastive contextually bound expression
  - **2** – contextually non-bound expression
  - **3** – contrastive contextually bound expression

- **deepord** – non-negative integer representing deep word order
PDT: relation between t-layer and a-layer

(He) would go to the forest.
PDT: relation between t-layer and a-layer

each t-node … PML reference to a-layer

• non-root t-node… attribute a consisting of 2 attributes:
  • lex.rf → id of a-node from which the t-node got its lexical meaning
  • aux.rf → list of ids of all other a-nodes related to the t-node

• technical t-root … atree.rf → id of a root of correspond. anal. tree
PDT: relation between t-layer and a-layer

each t-node … PML reference to a-layer

• non-root t-node… attribute \( a \) consisting of 2 attributes:
  • \( \text{lex.rf} \) \( \rightarrow \) id of a-node from which the t-node got its lexical meaning
  • \( \text{aux.rf} \) \( \rightarrow \) list of ids of all other a-nodes related to the t-node

• technical t-root … \( \text{atree.rf} \) \( \rightarrow \) id of a root of correspond. anal. tree

• t-node with no analytical counterpart: lex.rf and aux.rf empty

  Dovolil mu odejít. \{\#Cor.ACT\}

• copied nodes:
  • lexical items with several occurrences at the t-layer
    but expressed only once in a surface sentence
    (e.g., červené a bílé víno = červené víno a bílé víno )

  \( \Rightarrow \)  all attributes a/lex.rf, a(aux.rf)

  … id(s) of the corresponding a-node(s)
PDT: node types

- eight node types ... attribute *nodetype*
- defined on the basis of a *t-lemma* and/or a *functor*
References


• Manual for Tectogrammatical Annotation http://ufal.mff.cuni.cz/pdt3.0/documentation

• and its shortened version http://ufal.mff.cuni.cz/techrep/tr35.pdf (recommended!)


• PDT 3.5 https://ufal.mff.cuni.cz/pdt3.5

• PCEDT 2.0 http://ufal.mff.cuni.cz/pcedt2.0/