PDT – Tectogrammatical Layer

Introduction and T-lemma

Markéta Lopatková

Institute of Formal and Applied Linguistics, MFF UK

lopatkova@ufal.mff.cuni.cz

PDT: t-layer

- Intro
- Relation between t-layer and a-layer
- T-lemma

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

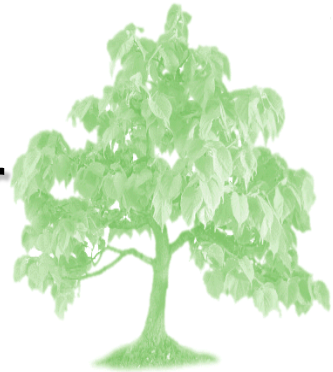
PDT: t-layer

Goal:

- to describe deep, semantic structure of a sentence
  ~ a sentence meaning
- disambiguated
- dependency 'tree'
- information on
  - lexical items: ***t-lemma*** (primarily)
  - relations between lexical words ~ **deep structure**: ***functors + subfunctors***
  - ***grammatemes***
  - coreferential links
  - topic focus articulation: linear order + tfa attribute
    (cz: aktuální členění)

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

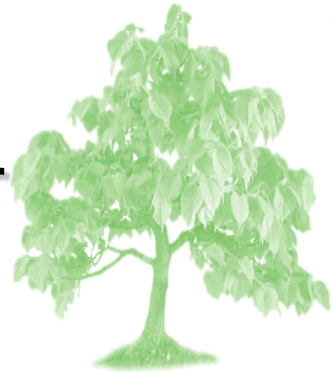
PDT: relation between t-layer and a-layer



each t-node … PML reference to a-layer

- technical t-root … ***atree.rf*** → id of a root of correspond. anal. tree
- 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

PDT: relation between t-layer and a-layer



each t-node … PML reference to a-layer

- technical t-root … ***atree.rf*** → id of a root of correspond. anal. tree
- 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

- 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*)

    ⇨ all attributes a/lex.rf, a/aux.rf

    … id(s) of the corresponding a-node(s)

PDT: t-lemma

two types of nodes wrt t-lemma

- individual lexical units (present at surface or ‘restored’)

  ⇨ prototypically t-lemma = m-lemma (suffixes are ignored)

  BUT *lexical and syntactic derivation*

  multi-word expressions

  frozen verbal forms (e.g. *myslím, soudě*) } t-lemma = m-form

  foreign-language expressions

PDT: t-lemma

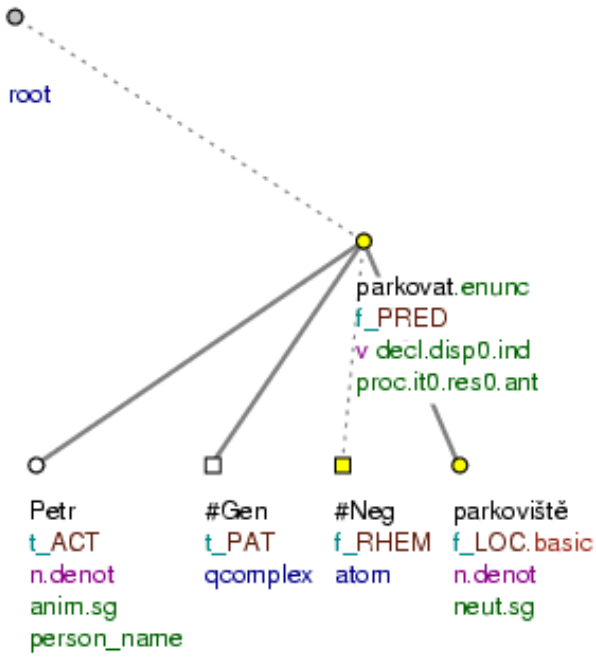
two types of nodes wrt t-lemma

- individual lexical units (present at surface or ‘restored’)

  ⇨ prototypically t-lemma = m-lemma (suffixes are ignored)

  BUT *lexical and syntactic derivation*
  multi-word expressions
  frozen verbal forms (e.g. *myslím, soudě*)
  foreign-language expressions
  } t-lemma = m-form

- t-lemma substitutes … starting with #
  - personal and possessive pronouns: #PersPron
  - newly established words (not copied)
    #Gen, #Rcp, #Cor … (diff. type of ellipses)
    #Forn, #Idph, …
  - negation: #Neg
  - punctuation: #Comma, #Dash, #Slash, #Bracket, ..



Syntactic and lexical derivation

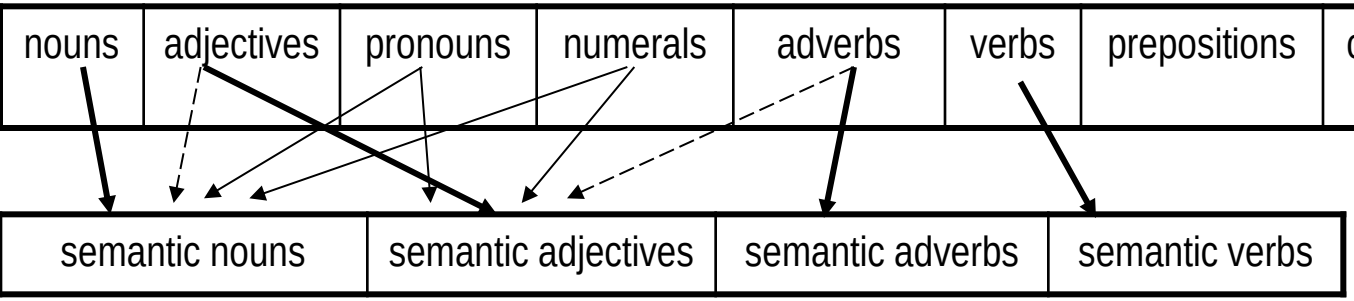
- "traditional" part of speech classification (PoS)
  - morphological tag
  - 10 basic classes

Syntactic and lexical derivation

- "traditional" part of speech classification (PoS)
  - morphological tag
  - 10 basic classes
- syntactic part of speech classification
  - syntactic nouns, adjectives, adverbs, verbs
  - e.g., Šmilauer "skladebné podstatné jméno"

Syntactic and lexical derivation

- "traditional" part of speech classification (PoS)
  - morphological tag
  - 10 basic classes
- syntactic part of speech classification
  - syntactic nouns, adjectives, adverbs, verbs
  - e.g., Šmilauer "skladebné podstatné jméno"
- semantic part of speech classification
  - syntactic vs. lexical derivation (Kuryłowicz)
  - attribute ***sempos***



Syntactic and lexical derivation (cont.)

- FGD theory (but NOT fully implemented in PDT):
  derived words represented by the t-lemma of the original word
- syntactic derivation
  - new syntactic function (change of PoS)
  - the same semantics

    e.g. *přicházet* → *přicházení*; *to arrive* → *arriving* (not in PDT)

    přicházení → *příchod*; *to arrive* → *arrival* (not in PDT)

    pěkně [nicely] → *pěkný* [nice]
- lexical derivation
  - new syntactic function (change of PoS)
  - change in semantics

    e.g. *učit* → *učitel*; *to teach* → *teacher* (not in PDT)

    učit → *učebna* [classroom] (not in PDT)

PDT: t-lemma for derived words

Implemented in PDT:

- personal and possessive pronouns
  - e.g. *já, mi, tobě, sebe, je* [I, me, you, myself, them] → #PersPron
    tvé, jejich, svoje [your, their, refl] → #PersPron
- possessive adjectives
  - e.g. *matčin* [mother's] → *matka* [mother]
    Pavlova [Pavel's] → *Pavel*
- deadjectival adverbs
  - e.g. *pěkně* [nicely] → *pěkný* [nice]
- directional adverbs (→ locative)
  - e.g. *tudy* [this way] → *tady* [here]; *kudy* [which way] → *kde* [where]
- temporal adverbs (→ answering *When?*)
  - e.g. *doteď* [until now] → *teď* [now]; *dokdy* [till when] → *kdy* [when]
- short forms of adjectives
  - e.g. *zklamán* [disappointed] → *zklamaný*

  NOT for passive participles: *pozván* [invited] → *pozvat* [to invite]

syntactic derivation:
m-lemma ⇐ t-lemma + functor

PDT: t-lemma for derived words

- ***numerals***

  ordinal, sort/kind, set and fraction numerals

  ⇨ derived from the cardinal numerals

  - t-lemma of the cardinal number
  - grammateme numertype

    ord (cz řadové) … *třetí* [the_third] → *tři* [three]

    kind (cz druhové) … *trojí* [three_kinds_of] → *tři*

    set (cz souborové)… *troje* [three_sets/pairs/…_of] → *tři*

    frac (cz dílové) … *třetina* [(one) third] → *tři*

lexical derivation:

m-lemma ⇐ t-lemma + numertype

PDT: t-lemma for derived words

- ***pronouns***, pronominal numerals and pronominal adverbs

  relative, indefinite, interrogative, negative and totalizing

  ⇨ derived from the correspond. interrogative or relative pronoun / numeral / adverb
    - t-lemma
    - grammateme indeftype

      e.g. *někdo* [somebody] → *kdo* [who]
      nikdo [nobody] → *kdo*
      kdokoliv [anybody] → *kdo*
      nic [nothing] → *co* [what]
      několik [several] → *kolik* [how many]
      všechen [all] → *co* [what]
      žádný [no] → *který* [which]

lexical derivation:

m-lemma ⇐ t-lemma + indeftype

PDT: t-lemma for multi-word expressions

- reflexiva tantum
  - e.g. *smát se* [to laugh Refl] → *smát_se*
    setkat se [to meet] → *setkat_se*
- complex conjunctions and conjunction pairs, operators
  - e.g. *buď … nebo* [either … or] → *buď_nebo*
    od … přes … do [from … via … to] → *od_přes_do*
    a nebo [or] → *a_přes_do*
- numeral expressions
  - e.g. *278 11* → *278_11*
    41 letý [forty-one_years_old] → *41_letý*
- idioms
  - e.g. *nohy na ramena* [legs on shoulders] → *nohy_na_ramena*
- etc.

PDT: t-lemma for multi-word expressions

- reflexiva tantum
  - e.g. *smát se* [to laugh Refl] → *smát_se*
    setkat se [to meet] → *setkat_se*
- complex conjunctions and conjunction pairs, operators
  - e.g. *buď … nebo* [either … or] → *buď_nebo*
    od … přes … do [from … via … to] → *od_přes_do*
    a nebo [or] → *a_přes_do*
- numeral expressions
  - e.g. *278 11* → *278_11*
    41 letý [forty-one_years_old] → *41_letý*
- idioms
  - e.g. *nohy na ramena* [legs on shoulders] → *nohy_na_ramena*
- etc.
- grammatemes
  - e.g. *chtít přijít* [to want to come] → *přijít* [to come] + volitive (deontic modality)
- special functors
  - e.g. CPHR: *mít dojem* [to have the impression]

References

- Manual for Tectogrammatical Annotation
  http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/t-layer/html/index.html
- Kuryłowicz, J. (1936). Dérivation lexicale et dérivation syntaxique. *Bulletin de la Société de liguistique de Paris*, 37, s. 79–92.
  Český překlad in: *Principy strukturní syntaxe I*. Praha, Univerzita Karlova, s. 87–94.