

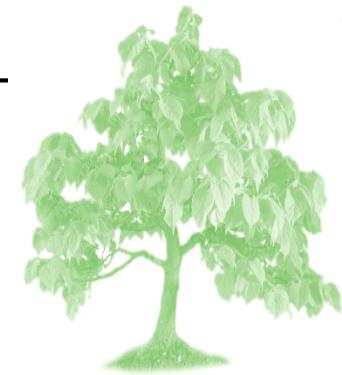


Lexical information in the PDT family (t-layer): t-lemma

Markéta Lopatková

Institute of Formal and Applied Linguistics, MFF UK

lopatkova@ufal.mff.cuni.cz



PDT family: t-layer

Goal:

- to describe deep, semantic structure of a sentence
 - ~ a sentence meaning
- disambiguated
- dependency 'tree'
- information on
 - lexical items: *t-lemma* (primarily)
 - meaning of morphological categories: *grammatemes*
 - relations between lexical words ~ **deep structure**:
functors + subfunctors
 - coreferential links
 - topic focus articulation: linear order + tfa attribute
(cz: aktuální členění)
- PDT 3.5: multiwords expressions, discourse relations, genre specif.
documentation: <http://ufal.mff.cuni.cz/pdt3.0/documentation>



Digression: Syntactic and lexical derivation

- "traditional" part of speech classification (PoS)
 - morphological tag
 - 8 basic classes (English) vs. 10 basic classes (Czech)



Digression: Syntactic and lexical derivation



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



Digression: Syntactic and lexical derivation

- "traditional" part of speech classification (PoS)
 - morphological tag
 - 8 basic classes (English) vs. 10 basic classes (Czech)
- 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)



Digression: Syntactic and lexical derivation

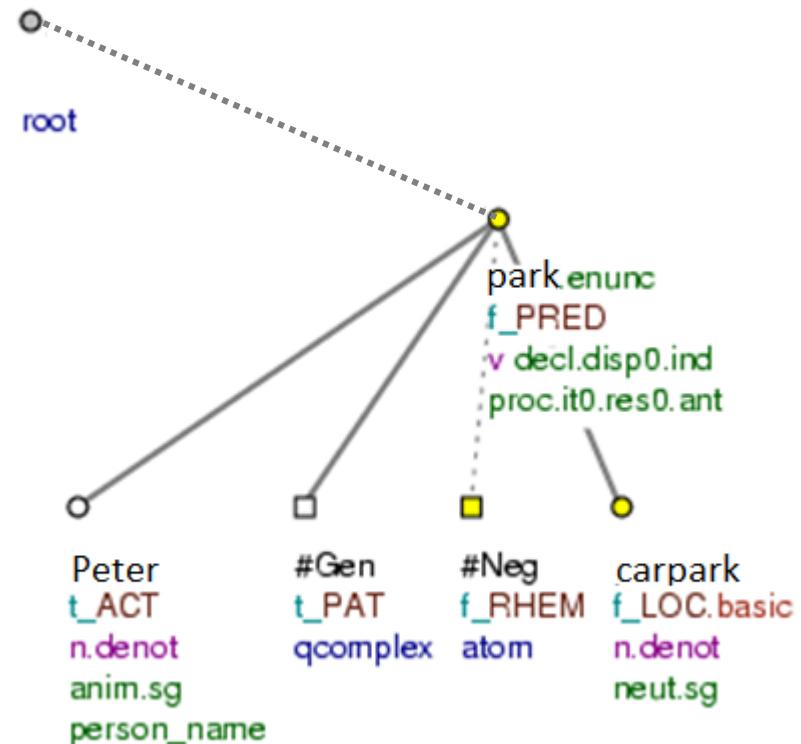
- FGD theory (**but NOT fully implemented in PDT**):
derived words represented by the t-lemma of the original word
- syntactic derivation
 - new syntactic function (change of PoS)
 - the same semantics
 - e.g. *přicházet* → *přicházení*; *to arrive* → *arriving* (not in PDT)
 - přicházení* → *příchod*; *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)

Lexical information in PDT family: t-lemma



- typically: *t-lemma = m-lemma*

Petr neparkuje na parkovišti.
Peter is not parking in a car-park.





Lexical information in PDT family: t-lemma

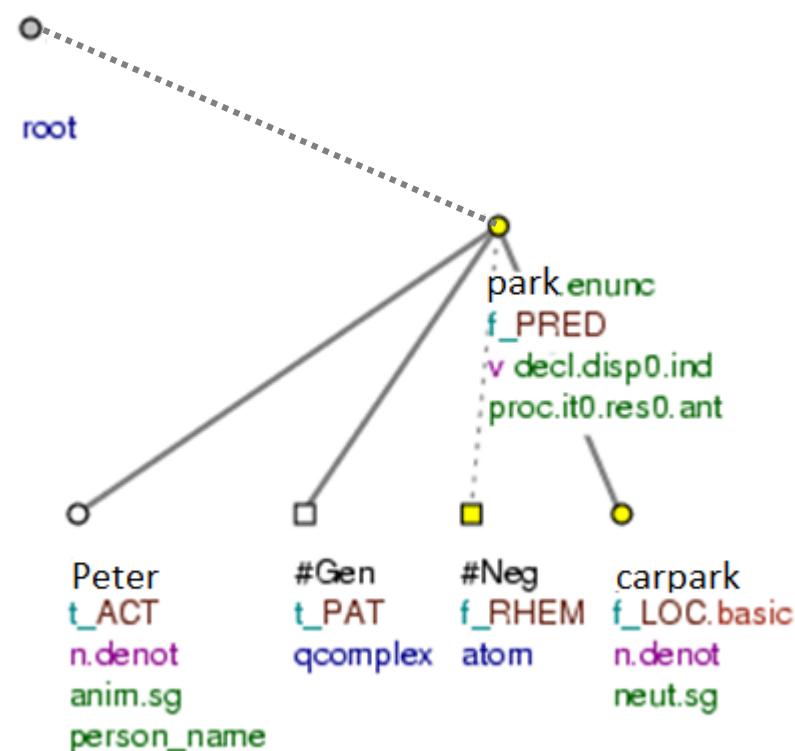
- typically: *t-lemma = m-lemma*

BUT:

- 1) t-lemma substitutes ... starting with #

- personal and possessive pronouns: **#PersPron**
- newly established words (not copied)
#Gen, #Rcp, #Cor ... (ellipses)
#Forn, #Idph, ...
- negation: **#Neg**
- punctuation: **#Comma, #Dash, #Slash, ...**

Petr neparkuje na parkovišti.
Peter is not parking in a car-park.



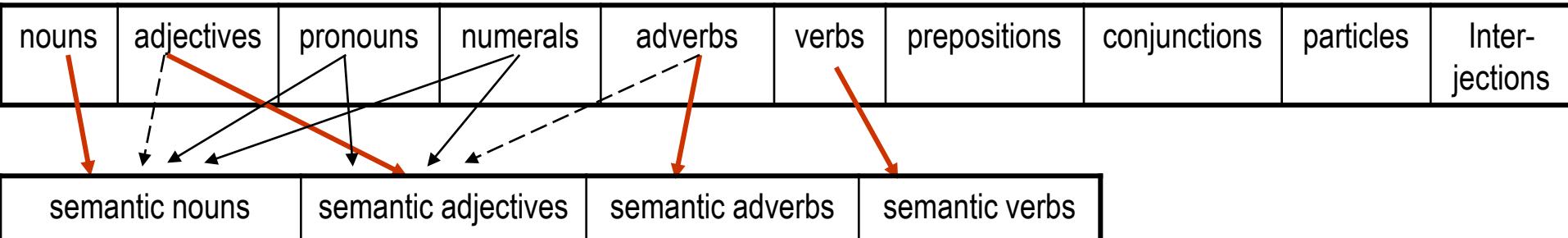


Lexical information in PDT family: t-lemma

- typically: *t-lemma = m-lemma*

BUT:

- 1) t-lemma substitutes ... starting with #
- 2) derived words represented by the *t-lemma of the original word*





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 adverbs)

e.g. *tudy* [this way] → *tady* [here]; *kudy* [which way] → *kde* [where]

- temporal adverbs (→ answering *When?*)

e.g. *doted'* [until now] → *ted'* [now]; *dokdy* [till when] → *kdy* [when]

- short forms of adjectives

e.g. *zklamán* [disappointed] → *zklamany*

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
- grammatememe **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
- grammateeme *indeftype*

e.g. *někdo* [somebody] → *kdo* [who]

nikdo [nobody] → *kdo*

kdokoliv [anybody] → *kdo*

nic [nothing] → *co* [what]

několik [several] → *kolik* [how many]

všechn [all] → *co* [what]

žádný [no] → *který* [which]

lexical derivation:

m-lemma ← t-lemma + *indeftype*



PDT: t-lemma for multi-word expressions

- inherent reflexive verbs (reflexivum tantum)
 - e.g. *smát se* [to laugh Refl] → *smát_se*
setkat se [to meet] → *setkat_se*
- complex conjunctions and conjunction pairs, operators
 - e.g. *bud' ... nebo* [either ... or] → *bud'_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/pdt3.0/documentation>