**Source data**:

* PDT-C .. "almost 2.0" version
  + how big data – total numbers : 175429 sentences
  + problem – data not publically released (where to find them?)
* format conversion (PDT -> Treex). In Treex, there are 3 stages:

1. Building the U-tree from the tectogrammatical tree
2. Process the coreference
3. Adjust special structures (coordination, CONTRD)

**Structural changes:**

* **coordination** … change of the structure to be in compliance with the UMR specification (discourse relations; reification in 2 subtypes)
  + **POZOR,** někde pod "but" víc ARG než 2 (cca 30) případů
* **apposition** … as identity-91
* **coreference-related structures** changed to be in compliance with the UMR specification:
  + re-entrances within a sentence, esp.:
    - anaphor is a personal or possessive pronoun (incl. reflexives): type *Maria – she – the girl*

but also with a nominal anaphor: type *Maria – the girl*

* + - raising and control verbs

(UD: verbs with open clausal complement (xcomp) = verbs with predicative complement)

(incl. cases without overtly expressed anaphor, type *Martin viděl Petra přicházet*)

* + - keep separate nodes if they are further modified!!
  + inverse roles
    - relative clauses

**Nodes labeling:**

**artificial t\_lemmas** (t\_lemma substitutes as, e.g., #Person) … "translated" to abstract concepts where appropriate/possible (or their super-types whenever automatic disambiguation is not possible; e.g., "entity" subsumes both "person" and "thing")

* + !! super-types:
    - entity (subsumes both "person" and "thing")
    - concept (subsumes "entity", "state", "event") in specific constructions, esp. constructions with the meaning of comparison (i.e., two or more events, states or entities are compared)
  + **POZOR**: nejasné, co s qcomplex #Ast, #Period3, #Comma, #Colon, #Dash

**Relations labeling:**

**1. arguments:**

* verb specific conversion for xx% of verb predicates (= frames ≈ rolesets with arguments) (Hajič et al, 2024),
  + exclude být (mít)
  + disambiguation … cca 25 rámců (mail JŠ, 9 May, 2024)
* default conversion table for the rest ((100-xx)% of verb predicates)
  + !! 1 new participant label not covered in UMR (EFF -> effect)

**2. non-arguments:**

* default conversion table (still some problems which need a refinement)
* several new roles to cover PDT-specific annotation:
  + !! new roles
    - comparison (based on CPR; complex structures, should be further inspected and refined)
    - regard (CRIT, REG)
    - result (RESL)
  + !! new discourse role
    - gradation (GRAD)
  + !! new "clausal-marker" role
    - for rhematizers (RHEM), sentence/ linking / modal adverbial expressions (attitude marker ATT, modal marker MOD, discourse marker PREC, conjunction modifier CM)
  + !! new discourse roles for independent clauses
    - PAR, PARTL, VOCAT
* some phenomena captured as relations (edges) in PDT transformed using new concepts (nodes), e.g. XXX ; for their (tentative) "translation" to UMR, several new types are distinguished:
  + !! new concepts for MWE annotation
    - predicative-noun (CPHR; should be refined)
    - part-of-phraseme (DPHR)
  + !! new concepts to cover specific entities
    - contra-entity
    - foreign-phrase (FPHR)
    - math (OPER, intervals, etc.)

**Identification of events:**

**events expressed as "full" predicates:**

* **all verb predicates** are treated in the same way as events (for the time being), disregarding their "packaging" (as there are no clear (formal) criterion for distinguishing, e.g., statives in Czech)

~~e.g., My cat~~ **~~loves~~** ~~wet food.~~

~~-> loving predicate, aspect = state, modal and temporal information acailable in PDT~~

~~e.g., The man, who is tall... (state in reference)~~

~~-> have-mod-91 predicate, aspect = state, modal and temporal inf. available in PDT~~

~~e.g., the same event structure for:~~

~~- My cat~~ **~~loves~~** ~~wet food. She~~ **~~is sitting~~** ~~on my desk. (2 events)~~

~~- My cat, who~~ **~~loves~~** ~~wet food~~**~~, is sitting~~** ~~on my desk. (2 events)~~

- eventive nouns, adjectives, adverbs … not transformed yet

- abstract predicates, reifications … not transformed yet

**UMR attributes**

* ?? **aspect** … based on the aspect grammateme + iter grammateme + diat grammateme (result)
* ?? mode … based on grammatemes: grammateme + deontmod (+ facmode?)
* **?? polarity**
* ?? degree

**UMR enrichments (wrt AMR) for events:**

* **coreference** … see above for intra-sentence
  + ?? inter-sentence coreference
  + bridging anaphora
  + ?? identify coreferential relations among events
* ?? **modality** … modal-strength based on the deontmod grammateme
* ?? **temporal** relations

**NE:**

* identification of NEs:
  + ?? names of persons
  + ?? other types
* ?? NEs anchoring

**TO BE DONE / IMPERFECT CONVERSION / MANUAL ANNOTATION NEEDED**

**Eventive nouns, adjectives, adverbs** … not transformed yet

* **nouns** derived from verbs / nouns with verbal counterparts
  + - ?? -ní/-tí nouns (type *přijíždění*)
    - ?? nominal events (type *příjezd*; type *volby, analýza*; ???)
    - ?? agentive nouns (type *učitel, volič*) (cs: činitelská) -> inverse roles
* adjectives derived from verbs
  + - ?? type *(byl) unavený* (type *unaven* as passive participle, thus verb (MorfFlex))
    - ?? type *přijíždějící*
  + ?? adverbs derived from verbs
* abstract predicates/rolesets:
  + být / mít / other verbs … should be converted to abstract predicates
    - conversion for být
  + identification of other constructions
* semimodals, phase verbs