**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 (discourse relations)

**POZOR,** někde pod "but" víc ARG než 2 (cca 30) případů

* OK roznásobeni spol. rozvití
* ?? apposition
* coreference in PDT is treated as:
  + 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
  + ??

**relations labeling:**

* 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)
* free modifications ≈ adjuncts:
  + default conversion table (still some problems which need a refinement)
  + !! 2 new roles (CRIT, NORM -> according-to; RESL -> result),
  + !! 1 new general abstract concept for intervals (UMR: date-interval, value-interval, between, slash)
* 1 new keyword (contra)
* ?? #Neg -> polarity … RHEM functer, where else ?? (detailed sempos, grammateme)
* ?? CPR
* ?? CPHR / DPHR / FPHR
* ?? CM (conjunction modifier) and functors for rhematizers AND sentence/ linking / modal adverbial expressions

**for "events":**

* identification of events:
  + **all verb predicates** are treated in the same way for the time being (disregarding their "packaging")

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)

* + **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

**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**