M A T E M A T I C K O - F Y Z I K Á L N Í F A K U L T A

P R A H A

UMR 2.0 - Czech: Release Notes

Markéta Lopatková, Eva Fučíková, Federika Gamba, Jan Hajič, Hana Hledíková,
Marie Mikulová, Michal Novák, Jan Štěpánek, Daniel Zeman, Šárka Zikánová

ÚFAL Technical Report
TR-2025-74

ISSN 1214-5521

U N I V E R S I T A S C A R O L I N A P R A G E N S I S

Copies of ÚFAL Technical Reports can be ordered from:

Institute of Formal and Applied Linguistics (ÚFAL MFF UK)
Faculty of Mathematics and Physics, Charles University
Malostranské nám. 25, CZ-11800 Prague 1
Czechia

or can be obtained via the Web: https://ufal.mff.cuni.cz/techrep

UMR 2.0 - Czech: Release Notes

Markéta Lopatková*, Eva Fučíková, Federika Gamba, Jan Hajič, Hana Hledíková, Marie Mikulová, Michal Novák, Jan Štěpánek, Daniel Zeman, Šárka Zikánová
*lopatkova@ufal.mff.cuni.cz

Charles University, Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics

The Czech UMR data have been automatically converted from the Prague Dependency Treebank – Consolidated (PDT-C), version 2.0, (Hajič et al, 2024a; converted on December 21, 2024).

The Czech UMRs can be downloaded as a part of the UMR 2.0 data from the LINDAT-Clarin repository. In total, the data contain 175 424 Czech sentences with partial UMR annotation.

The conversion covers selected phenomena pertaining to the sentence-level annotation (esp. structure of the graph, nodes and relations labeling, and PropBank-like argument structure for verbs). Furthermore, coreference relations are identified, both intra- and inter-sentential ones.

Content

Introduction

General characteristics of the two approaches and their brief comparison are introduced by Lopatková et al. (2024, sect. 2).[1] Basic ideas of the conversion are also sketched there, with stress on the identification of the following phenomena:

- *Phenomena that can be more-or-less directly transferred from the available Czech annotation to the UMR structures,*
- *Phenomena that require specific treatment and detailed analysis but still can be transferred, and*
- *Phenomena that are unavailable in PDT-C and thus necessitate new annotations.*

Both formalisms use (weakly connected) directed graphs - each sentence is represented as a tree-based structure that can be further enriched:
(i) UMR allows for reentrancies: a node variable may repeat in a single graph if there are two parents of the relevant concept, e.g., it serves as an argument for two event nodes),[2]
(ii) PDT contains coreferential links that go beyond the tree structure (when ignoring coreference links, the PDT-MR structures are trees).
Further, PDT and UMR differ in the way they represent discourse relations and coreference relations.

More importantly, the UMR and PDT approaches differ in the level of abstraction—thus, the graph structures represent slightly different phenomena. Here we briefly outline main differences and limitations of the automatic conversion.

1 The full UMR specification is available at the project website.
The PDT documentation is available at the project website.

2 When represented graphically, nodes sharing the same variable are merged; i.e., they are understood as a single node.

I. Sentence-level representation

I.1 Nodes labeling

In UMR, nodes represent concepts—entities and events, but also discourse relations and keywords for, e.g., entity or quantity types). In PDT, nodes stand primarily for content words; further, special types of nodes are reserved for representing paratactic structures (esp. for discourse relations), multiword or foreign expressions, and clausal markers.
As the first step, all PDT nodes are converted to UMR nodes, using the following rules:

- **Content words**

Content words are treated as concepts, albeit this represents a substantial simplification of UMR. Their lemmas, in compliance with the UMR approach, serve as instances of lexicon entries (see also sect. III.1).
In addition, PDT employs artificial lemmas (so-called t_lemma substitutes, as, e.g., #PersPron, #EmpNoun, #EmpVerb, #Oblfm) for unexpressed arguments. These lemmas should be treated in the same way as basic abstract concepts (like "person", "thing", "place", etc.); however, as it is not possible to deduce the correct type from the PDT data automatically, two supertypes are introduced:

- **entity** (subsuming "person" and "thing");
- **concept** (subsuming "entity", "state", "event", "place", "manner"): this supertype is used esp. in constructions with the meaning of comparison (i.e., two or more events, states or entities are compared)

- Nodes for **paratactic structures** (esp. discourse relations in PDT) are converted into nodes for discourse relations (see sect. I.3).

- Nodes for **phrasemes and multiword expressions** remain as in the PDT data at this stage, i.e. their inner structure is represented by special relations (see sect. I.2.2).

- The same is valid for nodes representing roots of **foreign phrases, lists** - inner structure of such phrases is represented by special relations (see sect. I.2.2).

- Nodes for **clausal markers** are kept as special nodes for further refinement (see also sect. I.2.2).

Warnings:

- At this stage, abstract rolesets (as have-91, belong-91, have-mod-91, etc.), required by the UMR specification, are not identified; instead, copula and light verbs are treated as concepts, following the PDT annotation scheme.
- Structured data represented in UMR as special "entities" (e.g., date-entity, further structured with attributes like day, month, year, century, etc.) or "quantities" (e.g., monetary-quantity or temporal-quantity-quantity, both with the attributes quant and unit) are largely not identified in PDT yet.

I.1.1 Alignment

In the PDT data, the graph annotation is aligned with overt words; as a rule, each non-punctuation token is aligned with (at least one) graph node. This applies also to function words (as prepositions,

auxiliary verbs, etc.), which are attached to the lexical words (concepts) they belong to (e.g., prepositions to (syntactic) nouns, auxiliary verbs to lexical verbs, etc.).[3]
The same principle is kept in the UMR conversion. As a consequence, the format specification allows for discontinuous alignment (as there can be additional content inserted between a function word and the concept); for example, in *Putin mu udělil milost kvůli zdravotním důvodům* 'Putin pardoned him for health reasons' (example inspired by ex. 1(f) from the UMR Specification), the preposition *kvůli* 'for' is attached to the *důvod* 'reason' concept - as a result, this concept is annotated as aligned with token 5 (*kvůli*) and 7 (*důvod*), marked as 5-5, 7-7.

I.2 Relations labeling

Both in UMR and PDT, a graph edge represents a relation between a parent node and its child (where UMR allows for two parents of a single node via reentrancy, contrary to PDT, see also sect. I.2.3, I.3.1 and I.4.1 below). Both formalisms use edge labels to determine a type of the relation, UMR uses the term "role" for these labels, PDT calls them "functors".
The PDT to UMR conversion proceeds in two steps, first it translates verb-specific arguments labeling and then converts all remaining labels using the default converting table.

I.2.1 Verb-specific argument labeling

The UMR approach supposes the use of a PropBank-like lexicon, assigning each predicate with a set of arguments (identified as ARG0, ARG1, …), while PDT makes use of its own predicate-argument labeling scheme, as indicated in the PDT-Vallex lexicon (with labels like ACT for actor/bearer, PAT for patient, ADDR for addressee, etc.). Fortunately, one third of the PDT-Vallex rolesets (32% of rolesets, covering 43% of arguments) has been converted to the PropBank style using existing resources (Hajič et al, 2024b).
Thus, for the rolesets with available argument mapping, the PropBank arguments are used in the data.
The arguments in the rolesets without (verb-specific) argument mapping are converted using the default table, see below.

I.2.2 Labeling with the default converting table

Arguments in rolesets without (verb-specific) argument mapping as well as all other functors are translated using the default conversion table, as presented in Appendix A. In most cases, it is possible to find reliable mapping between PDT functors and UMR roles.

- **Arguments**

The verb-specific argument mapping (see above) shows that three out of five PDT-like arguments can be automatically converted—with relatively high reliability—to PropBank-like arguments:

ACT → **:ARG0** (accuracy 82.6%);
PAT → **:ARG1** (accuracy 92.5%);
ADDR → **:ARG2** (accuracy 84.4%).

On the other hand, the verb-specific mapping (see above) for the remaining two PDT arguments does not show any prevailing pattern (due to the different linguistic theory adopted in PDT). Thus, we use the most relevant participant role (as suggested by the guidelines for Stage 0 annotation, see the UMR Specification) for the first one and a new label for the second one:

ORIG → **:source**;
EFF → **:effect** (new label as no UMR participant role covers this PDT-specific annotation).

3 At the same time, each graph node in PDT must be aligned with at least one lexical word (and with as many function words as relevant).

- **Non-arguments**

PDT uses 42 non-argument labels for different deep syntactic relations between a parent and its child. Most of them can be relatively easily mapped onto UMR (participant or non-participant) roles using their semantics (albeit the conversion of some of these roles is still too coarse and will need refinement in the future).
However, several new labels are introduced in the data to cover the PDT-specific annotation:

- **:comparison** … a special relation for complex comparison structures (CPR in PDT);
- **:regard** … role specifying with respect to what something holds (CRIT and REG in PDT).

Further, 3 PDT labels identifying the inner structure of MWEs were introduced (as temporary roles, which will be removed when MWEs are processed):

- **:predicative-noun** … tentative role used in light verb constructions (CPHR in PDT);
- **:part-of-phraseme** … tentative role used for identifying parts of idiomatic expressions (DPHR in PDT; inner structure of such expressions remains unchanged);
- **:FPHR** … tentative role used for identifying parts of foreign expressions (FPHR in PDT; inner structure of such expressions remains unchanged).

- **Discourse relations**

The PDT annotation scheme uses 9 special relation labels to classify paratactic structures. Four of them (ADVS, CONFR, CONJ, DISJ) are translated as discourse relations when converting to UMR (see Appendix A; necessary changes to the graph structure are outlined in sect. I.2.3. and I.3.1). Rather coarse-grained discourse values are chosen from the lattice as the PDT annotation does not specify subtle types of discourse relations. In addition, one new discourse relation is introduced:

- **:gradation** … relation for discourse structures in which each follow-up proposition makes a stronger claim than the previous one, as in *Nemůže se pohnout, natož vstát* 'He can't move, let alone get up' (GRAD in PDT, example from the PDT manual, Mikulová et al., 2006).

One PDT label is reserved for appositions (APPS); this label is converted using the identity-91 predicate, see sect. I.2.3 and I.3.2. Two functors (CSQ, REAS) are treated using the reification mechanism, see sect. I.2.3. The last functor is translated as a special entity (see contra-entity in sect. I.2.3 among Other changes).

- **Others**

PDT also uses 3 relation labels to identify parenthetical clauses, interjections, and vocative clauses, i.e. sentence material that is not syntactically related to the rest of the sentence. These are translated as special relations:

- **:interjection** (PARTL in PDT)
- **:parenthesis** (PAR in PDT)
- **:vocative** (VOCAT in PDT) - in fact, this role appears in the UMR 1.0 data for other languages as well.

Finally, PDT annotates rhematizers (RHEM in PDT), different types of sentence linking expressions (as attitude marker ATT, intensifier INTF, modal marker MOD, discourse marker PREC), or conjunction modifier (CM in PDT). Not to lose this information, we introduce one more relation covering these phenomena:

- **:clausal-marker**.

Warnings:

- Conversion of some functors is too general and requires further refinement.
- As structured data represented in UMR as special "entities" or "quantities" are largely not identified in PDT yet (see sect. I.1 above), they are labeled with roles based on their grammatical structures and/or functions.

I.2.3 Structural changes based on functors

In some cases, the translation of functors requires additional structural changes (not only translation of the label).

- **Abstracting from syntactic form: paratactic vs. hypotactic structures**

PDT-C, following Czech linguistic tradition, applies formal language-specific criteria for distinguishing coordinated and subordinated structures; thus it is endowed with pairs of functors (roles) for such structures. As UMR abstracts from the formal means (applicable in a given language), the conversion identifies such functor pairs and represents them in the same way:

- relations (functors) as discourse relations (see sect. I.3.1):
  - contradiction (CONTRD functor) → contrast-91
    (e.g., both Czech conjunctions *zatímco, kdežto* 'while' express confrontation (pure contrast), the former being classified as subordinating, the latter as coordinating; both of them are transferred as the same discourse relation);
  - concession (CNCS functor) → but-91
    (two conflicting propositions are represented as the but-91 discourse relation);
- relations (functors) as reifications:
  - consequence (CSQ functor) → have-result-91,
  - reason (for)/cause (of) (REAS functor) → have-cause-91
    (e.g., both Czech conjunctions *protože, neboť* 'because' express causative relation, the former being classified as subordinating, the latter as coordinating; they are represented as the :cause relation and its reification have-cause-91, respectively);
- relation (functor) as an abstract predicate:
  - apposition (APPS functor) → identity-91
    (see sect. I.3.2)

- **Dual dependency: predicative complement**

In PDT, a predicative complement is a non-obligatory free modification (adjunct) which has a dual semantic dependency relation: it simultaneously modifies a verb (which can be nominalized) and a noun. The first dependency is represented by an edge (COMPL functor); thus, it is translated as the :manner (or :mod) relation. The dependency on the (semantic) noun is represented by the inverse :mod-of relation (via reentrancy, see also sect. I.3.1 and I.4.1 below).[4] See the (simplified) UMR graph for the sentence *Marie přišla domů spokojená* 'Marie came home happy', interpreted as two predications, (i) the coming event, which is modified by the modifier *happy*, and (ii) the property predication of being happy (at this stage represented as attribution *happy Mary*, which should be modified during the next stages when the non-verbal phrases are processed):

```
(p/ přijít-010 'come-01'
  :ARG0 (m/ Marie  'Marie')
  :ARG4 (d/ domů  'home')
  :manner (s/ spokojená 'happy'
            :mod-of m))
```

- **Other changes**

Several minor changes complete the list of structural changes that are based on functors:

- functors converted to new entities:
  - **contra-entity** (with op menu)
    A relatively fine-grained PDT relation relating two fighting/opposing entities (CONTRA functor) - typically expressed by prepositions *kontra* 'contra' and *versus* (as

4 The UMR data still contain 80 cases with a predicative complement not fully processed: these are labeled with the :COMPL role (with the nominal dependency unmarked); these cases will be refined later.

in *akademie věd kontra vysoké školství* 'Academy of Science contra universities') or dash etc. (as in *utkání Rusko - Švédsko* 'the Russia - Sweden match') - is represented as a new entity in Czech UMR.

- **math-entity** (with op menu)
  Mathematical operations (as addition, multiplication, or proportion/division) and intervals that cannot be analyzed using spatial or temporal functors (as, e.g., *trest od tří do pěti let* 'a sentence from three to five years'), are represented with special label (OPER functor) in PDT. In Czech UMRs, a new entity is used.

I.3 Coordination and discourse relations, apposition

I.3.1 Coordination / discourse relations

In general, representation of paratactic structures (coordination, discourse relations) follows the same principles in PDT-C and UMR: there is a special node in the graph for the whole paratactic structure (assigned with a discourse relation in UMR.) In these cases, the transfer is more-or-less straightforward, dealing mainly with technicalities. Several notes are relevant in this context.

- **Coordinating contents with the same relation to the modified concept**

Typically, the paratactic structures combine more contents with the same relation to the modified concept: e.g. in *John met Carl and Sophia*, both *Carl* and *Sophia* serve as ARG1 of the verb *meet*. Thus, the (non-)participant relation can serve as the parent node common to both/all coordinated members and the discourse concept just "multiply" the same role positions, see the (simplified) UMR representation of the example sentence. The only difference is a technicality, as PDT treats the discourse concept as the head of the structure, as in the left picture (simplified PDT graph of the example sentence; this applies to ADVS, CONFR, CONJ and DISJ functors), while UMR subsumes the discourse relation below the one between the modified and modifying concepts (see the right graph).

```
PDT
(m/ meet-02
  :ARG0 (j/ John)
  :CONJ (a/ and
          :ARG1 (c/ Carl)
          :ARG1 (s/ Sophia)))
```

```
UMR
(m/ meet-02
  :ARG0 (j/ John)
  :ARG1 (a/ and
          :op1 (c/ Carl)
          :op2 (s/ Sophia)))
```

- **Participant and non-participant roles shared by coordinated concepts**

In PDT-C, participant and non-participant roles that are shared by coordinated concepts are represented as children of the discourse relations; see, for example, the left graph below (simplified)[5] representing the sentence *I read a book and listened to music*. When transforming to the UMR scheme, the shared participant is represented using a reentrancy (see also sect. I.2.3 and I.4.1); in particular, the shared concept is displayed as a node dependent on (modifying) the first coordinated member; its relation to the other coordinated concept(s) is identified by repeating its variable in the respective position(s), see the right graph below.

```
(a/ and
  :ARG0 (p/ person
          :refer-person 1st
          :refer-number singular)
  :PRED (r/ read-01
          :ARG1 (b/ book))
```

```
(a/ and
  :op1 (r/ read-01
         :ARG0 (p/ person
                 :refer-person 1st
                 :refer-number singular)
         :ARG1 (b/ book))
```

5 In addition, a special attribute in the PDT graph indicates coordinated nodes (here the predicates *read* and *listen*, as roots representing the respective subgraphs); for simplicity, this attribute is not displayed here.

```
:PRED (l/ listen-01
        :ARG1 (m/ music)))
```

```
:op2 (l/ listen-01
       :ARG0 p
       :ARG1 (m/ music)))
```

- **Abstracting from formal language-specific criteria**

Two types of structures interpreted as relations in PDT are restructured as discourse relations (namely contradiction, concession, see sect. I.2.3). On the other hand, two types of paratactic structures in PDT are represented as reifications (namely consequence and reason (for)/cause (of), see sect. 1.2.3).

- **Negation in coordination**

The attribute of polarity (affirmative/negative) needs a special handling in coordinated structures, see sect. I.5.2.

Warning:

- PDT also allows for cases when two (or even more) different relations are subsumed into a single coordination, as in *Kdy a za jakých podmínek se to stalo?* 'When and under what conditions did it happen?' In these cases, the most frequent role is tentatively used to represent the whole structure. More adequate processing (which involves splitting the material into two propositions (= *Kdy se to stalo a za jakých podmínek (se to stalo)?* 'When did it happen and under what conditions (did it happen)?') is postponed to later stages of the conversion.
- Coordinated structures that are involved in a coreference relation (either as anaphors or as antecedents) need a more detailed analysis, see sect. I.4.2.

I.3.2 Apposition

Apposition is a grammatical construction in which two or more concepts (especially entities) with the same referent stand in the same syntactic relation to the rest of a sentence: One concept identifies the other in a different way. While PDT treats this construction as a paratactic structure, reserving a special node (with the APPS label) for the whole construction, the UMR specification does not cover this phenomenon.

Constructions annotated in PDT as appositions are converted to UMR using the *identity-91* predicate. See, e.g., the PDT-C (left) and UMR (right) structures for the sentence *Charles IV., Holy Roman Emperor resided in Prague* (both graphs are simplified).

```
(r/ reside-01
  :APPS (#Comma
          :ARG0 (n1/ Charles IV)
          :ARG0 (n2/ Holy Roman Emperor)
  :place (p/ Prague)
```

```
(r/ reside-01
  :ARG0 (i/ identity-91
          :ARG1 (n1/ Charles IV)
          :ARG2 (n2/ Holy Roman Emperor)
          :place (p/ Prague))
```

Warning:

- Structures with appositions that are involved in a coreference relation (either as anaphors or as antecedents) need a more detailed analysis, see sect. I.4.2.

I.4 Processing coreference annotation

Different representation of different types of the coreference relation in PDT-C and UMR frameworks significantly affects the overall structure of sentence graphs (Lopatková et al, 2024, Sect. 2 and Sect. 3.3). In PDT, all coreferential expressions are typically represented by separate nodes that are interlinked by a special coreferential relation (coreferential arrows). Designated attributes

identify the type of the coreference (grammatical or text coreference[6]) and the type of reference (e.g., specific or generic). This is applied both to intrasentence and intersentence coreferential relations. On the other hand, UMR introduces several ways to treat phenomena represented by coreferential chains in PDT. For the sentence-level representation, the mechanisms of reentrancy (sect. I.4.1) and inverse roles (sect. I.4.2) are relevant. Coreferences crossing the sentence boundary are treated within the document-level representation (sect. II.1).

I.4.1 Reentrancy (within a sentence)

The UMR specification allows for the so-called reentrancy of a node if the respective concept has several roles in a single sentence (e.g., a participant in a main clause can serve as a participant in a complement clause). The reentrancy[7] is employed esp. for conversion of the following cases where coreference is annotated using coreferential arrows in PDT-C:

- **Anaphor/cataphor is the personal or possessive pronoun (incl. reflexives)**

In PDT, separate nodes for an anaphor and its antecedent are interlinked by a coreferential arrow. The information provided by the arrow is transferred as an identical variable for these nodes: while the node for the antecedent and its lexical content is preserved, the node for anaphor/cataphor loses its lexical content, being represented with the same variable as its antecedent/postcedent (as, e.g., in *Málokdy si dívka dopřeje sklenku vína* 'Very seldom will the girl indulge in a glass of wine', both *dívka* 'girl' and *si* 'self', represented as two nodes with separate concepts (*dívka* and the node for reflexive pronoun interlinked by the coreferential arrow) in PDT, are represented as two nodes sharing their variable *d2* (and thus referring to the same person).

```
(d/ dopřát-001   'indulge'
  :ARG0 (d2/ dívka)
  :ARG1 (s/ sklenka)
  :ARG2 (p/ person)
  :frequency (m/ málokdy))
:coref (d2 :same-entity p)
```

```
(d/ dopřát-001
  :ARG0 (d2/ dívka)
  :ARG1 (s/ sklenka
  :ARG2 d2
  :frequency (m/ málokdy))
```

Similarly for non-reflexive personal or possessive pronouns, as in *Marie přišla domů a tam na ni čekal Marek* 'Marie came home and there Marek was waiting for her', where *Marie* and the personal pronoun *(na) ni* '(for) her' share the same variable.

- **Anaphor/cataphor is a relative pronoun or pronominal adverb**

Relative pronouns (as *kdo, který, jenž* 'who', *co* 'which', *čí* 'whose') and relative pronominal adverbs (as *kde, kudy, kam* 'where', *kdy* 'when') are merged with their antecedents, i.e., they are transferred with the identical variable. In addition, their role changes, as described in section I.4.2 below.

- **Arguments of raising and control verbs**

Nodes for arguments of raising and control verbs (in PDT represented as separate nodes with special labels #Cor, #QCor) are treated in the same way (incl. cases without overtly expressed anaphor, type *Martin viděl Petra přicházet = Martin viděl Petra, jak Petr přichází* 'Martin saw Peter coming = Martin saw Peter as Peter is coming').

In relatively rare cases, separate nodes are kept in UMRs, esp. if an anaphor is further modified, and the coreference relation is indicated in the document-level representation.

---

6 Annotation of bridging anaphora is left aside in this version.

7 Reentrancy is also used for capturing dual dependency (sect. I.2.3 above) and for shared (non-)participants in coordination (sect. I.3.1 above).

I.4.2 Inverse roles

UMR uses the mechanism of inverse roles, among others, for representing relative clauses:[8]

- **Relative clauses**

A relative clause typically contains a relative expression (like a pronoun, as *kdo* 'who', *co* 'what', *jaký, který* 'that, which' or a relative adverb, as *kde* 'where', *kdy* 'when', *jak* 'how'), which is represented as a separate node in PDT-C that is coreferential with the modified expression (in the governing clause). The relative expression typically serves as an argument or adjunct of the predicate of the relative clause. When transferring to UMR, the relative expression is merged with its antecedent (i.e., modified expression); the original parent node of the relative expression is attached to the antecedent with the relation inverse to the original one.
To elucidate this principle, compare the PDT-C like representation (left) and UMR representation (right) of the following example: *Student, který hraje na housle, ...* 'The student, who is playing the violin, …'. The relative *který* 'who' (serving as ARG0 of the verb *hrát* 'play') is merged with its antecedent *student* 'student'; the relative clause is attached to the modified concept *student* 'student' with the inverse (i.e., ARG0-of) relation.

```
(s/ student   'student'                          (s/ student 'student'
  :mod (p/ hrát-001 'play'                             :ARG0-of (p/ hrát-001 'play'
         :ARG0 (w/ kdo)    'who                           :ARG1 (v/ housle 'violin'))
         :ARG1 (v/ housle 'violin')))
:coref (s :same-entity w)
```

The same principle is used also for relative pronominal adverbs (as *kde, kudy, kam* 'where', *kdy* 'when'). For example, in *Oblasti, kde se používal krokydolit , …* 'Areas where the crocidolite was used …', the relative adverb *kde* 'where ' is merged with its antecedent *oblast* 'area'; the whole relative clause is attached using the inverse role (place-of):

```
(o/ oblast 'area'                                  (o/ oblast 'area'
   :mod (p/ používat-001 'use'                        :place-of (p/ používat-001 'use'
            :place (k2/kde) 'where'                                :ARG1 (k/ krokydolit)))
            :ARG1 (k/ krokydolit))) 'crocidolite'
:coref (o :same-entity k2)
```

Warnings:

- UMR does not recognize cataphoric reference (i.e., reference to an expression that appears later in the text). For all such instances, the relation is changed to the anaphoric one (and left for further analysis).
- Coreferential relations between events are only sporadically captured in the PDT-C thus this type is rare (and under-annotated) in the current version of the UMR data.
- The coreference relation between coordinated structures (see sect. I.3.1) needs more detailed analysis. For this version, the UMR annotation reproduces the PDT-C annotation, i.e., the coreferential links may interconnect the whole coordinated structures (not individual coordinated members). For example, in *Marie vzala Vlastu do divadla, kde na ně čekal Marek* 'Marie took Vlasta to the theater, where Marek was waiting for them', the pronoun *(na) ně* '(for) them' refers both to *Marie* and to *Vlasta* (from the PDT manual, Mikulová et al., 2006). However, in many cases, coreference chains include discourse concepts (both in the role of an anaphor/cataphor and antecedent/postcedent). These cases should be analyzed and their transfer refined in the future versions.

8 The mechanism of inverse roles is also used in UMR for embedded interrogative clauses, for nominalization and action nouns, and for some nominal modifications as, e.g., kinship (*my father*). The annotation of such structures is not fully transferred from PDT yet.

- With reciprocal constructions, characterized by a special node label (#Rcp) and a coreferential relation among two (or more) nodes in PDT, the coreference is indicated in the document-level annotation (despite typically interlinking nodes within a single sentence). The conversion deserves a refinement in future versions.
- Though the PDT stores information on the bridging anaphora (as, e.g., part-whole relation), which is relevant for the :subset-of relation, this information is not used at this stage.
- Coreferences crossing the sentence boundary are treated within the document-level representation (sect. II.1).

I.5 UMR attributes

I.5.1 Aspect

The UMR annotation scheme distinguishes rather fine-grained values for the aspect annotation. However, the specification advises to reflect esp. the aspectual distinctions that are grammaticalized and/or obligatory in the language. Thus, the transfer is limited to events expressed as verbs at this stage (the aspect annotation for events expressed as nouns and adjectives are left for further stages, see also sect. III.1 commenting on the event/entity distinction). Further, we stick to the basic values in the lattice (disregarding the possibility of more subtle annotation). The appropriate aspect value is deduced based on the compiled list of statives in Czech and, for verbs not listed there, from their morphological marking:

- Verbs from the (ad hoc) compiled list of stative verbs (ca 400 verbs now) get the "state" value.
- Imperfective verbs that are morphologically marked as iteratives obtain the "habitual" value.
- Imperfective verbs without such marking get the "activity" value (note that this type of verbs often express habitual events; however, this is not captured in the PDT data and thus cannot be distinguished automatically).
- Perfective verbs are marked with the "state" value when they appear in resultative diathesis, otherwise they are identified as "performance".
- Double aspect verbs are annotated as "process" (subsuming "activity" and "performance" classes in the aspect lattice).
- All abstract concepts identified as events (see sect. III.1) get the "state" value.

Warnings:

- One of the aspect values is used for any concept expressed as a verb - this is definitely a simplification to the UMR principles (see sect. III.1 commenting on the event/entity distinction).
- Only those concepts that are recognized as events at this stage get their aspectual value (thus, e.g., eventive nouns are not annotated so far).
- Imperfective verbs can express habitual events. However, as this is not captured in the PDT data, such cases cannot be distinguished automatically. Unfortunately, this ambiguity is not reflected in the UMR aspect lattice and thus all occurrences of imperfective verbs get the "activity" value, albeit inappropriate in these cases.
- The list of stative verbs should be refined and further enriched.

I.5.2 Polarity

The UMR "polarity" attribute serves for identifying morphosyntactic indicators of negation (which do not necessarily signal semantic negation). PDT recognizes two basic types of negation, so-called lexical negation and syntactic negation.

- **Lexical negation**

Lexical negation is annotated using so-called grammatemes, i.e., meaning correlates of morphological categories in PDT. Two grammatemes are relevant here:

- "negation" grammateme for lexical negation with semantic nouns, adjectives and adverbs: these are encoded using (positive) lemmata with the "negation" grammateme set to "neg1", e.g., *nezralost dítěte* 'immaturity of a child ', encoded as *zralost* 'maturity' with negation=neg1] (example from the PDT manual, Mikulová et al., 2006);
- "indeftype" grammateme for lexical negation with pronominals, annotated in the "indeftype" grammateme set to "negat", e.g., *nikdo* 'nobody', or *nikde* 'nowhere' annotated as *kdo* 'who, someone' and *kde* 'where, anywhere' respectively, with indeftype =negat (example from the PDT manual, Mikulová et al., 2006).

- **Syntactic negation**

Syntactic negation is applicable primarily for propositional negation. There are two main subtypes identified in PDT:

- syntactic negation using the **negation morpheme** to create **negative verb form** (*ne-* , as in *nepřijít* 'not to come')[9]: the negative morpheme is represented as a separate child node of the negated expression, marked as #Neg with RHEM or CM functor (depending on type of the modified node);
- syntactic negation with the **negation particles** *ne* 'no', *nikoli/v* 'no/t' in answers or other independent clauses.

All negation types are transferred as the polarity attribute with "-" value of the modified node.

- **Negation in coordinated structures**

A special treatment is necessary when negation combines with coordination. In these cases, the negation scope comprises siblings of the #Neg (not its parent(s)), as, e.g., in *Firmy, které nejsou ochotny uvádět svou celou adresu, ale jen P.O.Box, …* 'Companies that are not willing to provide their full address, but only a P.O. Box, ...', where only *adresa* 'address' is negated (not the whole proposition).

Warning:

- The current version of the Czech UMR data does not use the "modal-strength" attribute for annotation propositional negation. Instead, the attribute "polarity" is used for propositional negation as well. More appropriate conversion is postponed to later stages of the data processing.
- Questions and embedded interrogative clauses (with the UMR polarity value "umr-unknown" and "truth-value") are not processed yet.

I.5.3 Refer-number and refer-person

The value of the "refer-number" and "refer-person" UMR attributes serve esp. for pronominal references, identifying grammatical person information and grammatical number marking, respectively. These attributes can apply to all entity concepts.

- **Refer-number**

9 It is also possible to negate only one of the verbal modifications/constituents using the negation particle (*ne* 'no'): then, in PDT, the negation particle is placed as the left sister node of the negated expression. Then the annotation is preserved in the UMR data, see, e.g., *… a kde se sice to nejkrásnější odehrávalo v hledišti, ne na jevišti …* '... and where the most beautiful things took place in the auditorium, not on stage…', with the negation "-" indicated with *jeviště* 'stage'.

In the PDT data, the "refer-number" attribute corresponds to the grammateme "number". While this grammateme (i.e., a meaning correlate of the morphological category of number), typically reflects the morphological value, it can abstract from surface forms in some cases. Pluralia tantum can serve as an example - while their morphological number is always 'plural' (e.g., *kalhoty* 'trousers', with non-existing **kalhota* '*trouser'), the "number" grammateme discriminates between singular and plural contexts (*jedny kalhoty* 'one pair of trousers' vs. *dvoje kalhoty* 'two pairs of trousers'). However, the annotation of the "number" grammateme is available only in a portion of the PDT-C data (namely the PDT, Faust and partially PDTSC subcorpora). For the portion of the data without the grammateme, the PDT morphological annotation is used for conversion.
The Czech UMR data stick to the basic values in the lattice, namely singular and plural. The very rare cases of dual in Czech are not identified so far (similarly as some other cases with unclear PDT strategy).

- **Refer-person**

The "refer-person" UMR attribute corresponds to the PDT grammateme "person", which distinguishes 3 values, 1 = first person (speaker), 2 = second person (hearer), and 3 = third person (what is talked about).
Similarly as for the number grammateme, person grammateme is available only in a portion of the PDT-C data (namely the PDT, Faust and partially PDTSC subcorpora). For the portion of the data without the grammateme, the PDT morphological annotation is used for conversion.

- **Refer-person and refer number in coreferential chains**

Whenever relevant, the "refer-number" and "refer-person" UMR attributes are propagated through coreferential chains.

Warning:

- In the current conversion, Czech UMR data displays the "refer-number" and "refer-person" attributes whenever corresponding morphological categories appear in the PDT data. This leads to excessive and non-adequate usage of the attributes in UMR structures and should be refined in future versions.

II. Document-level representation

II.1 Coreference

All nodes with a coreferential link in the PDT data that are not processed using reentrancy (sect. I.4.1) or inverse roles (sect. I.4.2) are collected and the respective pairs of coreferring nodes are added to the document-level annotation. The relevant relation between individual pairs is identified, reflecting whether they refer to the same entity or to the same event. As a default, the :same-entity /:same-event relation is used. See Lopatková et al. (2024) for more examples.

Warnings:

- Though the PDT stores information on the bridging anaphora (as, e.g., part-whole relation), which is relevant for the :subset-of relation, this information is not used at this stage.
- Coreferential relations between events are only sporadically captured in the PDT-C thus this type is rare (and under-annotated) in the current version of the UMR data.

III. Main phenomena not covered in the data

III.1 Identification of events

III.1.1 Verb predicates

- **Lexical verbs**

When converting PDT to UMR, all verb predicates (i.e., all lexical verbs, excluding modal and temporal auxiliaries) are treated as events. This means, their "packaging" - whether they are conveyed as a reference, modification, or predication - is disregarded. The reason is simple: there are no clear (formal) criteria for distinguishing, e.g., statives in Czech, no available lists of such verbs. Thus, though this represents a substantial simplification of the UMR principles, we prefer to treat all lexical verbs in the same manner to keep the converted data as consistent as possible.

- **Other types of verbs**

  As the next steps, the following types must be identified:
  - **semimodals** (some of them are treated as modal auxiliaries in PDT, as, e.g. *chtít* 'want')
  - **phase verbs** must be identified (while annotated as lexical verbs in PDT)
    (e.g., UMR: inchoative, completive, and continuative verbs), only inform the aspect value
  - **LVCs** must be identified (based on the CPHR functor)

  This classification should be available as a part of the PDT-Lexicon, which serves as a basis for the PropBank-like Czech lexicon.

III.1.2 Non-verbal predicates

The following types must be identified in PDT:

- **eventive nouns** … derived from verbs / nouns with verbal counterparts
- **eventive adjectives**
- **eventive adverbs**

Selected types of derivational information are stored in several available lexical resources (as MorfFlex, DeriNet, PDT-Vallex, SynSemClass). However, these resources do not cover all information necessary for (semi-)automatic event identification (as exemplified by Lopatková et al, 2024). Thus, various heuristics must be applied and their outputs need careful evaluation.

III.1.3 Abstract predicates/rolesets

- ***být/bývat/bývávat* 'to be'**

PDT distinguishes several types of constructions with *být/bývat/bývávat* 'to be' where the verb is considered lexical:

- existential *být*,
- substitute *být*,
- copula *být*,
- phrasal *být*, and
- *být* in constructions with a single constituent.

This classification is (partially) reflected in the PDT-Vallex lexicon and the conversion should be done for individual verb frames/senses (ca 160 frames). The most frequent cases correspond to UMR abstract predicates; further, some of them can be treated as reifications.
In addition, *být/bývat/bývávat* 'to be' serves as a light verb (CPHR in PDT) and it forms a number of phrasemes (DPHR in PDT).

- ***mít/mívat* 'to have'** and other verbs with similar meaning, as e.g. *patřit* 'belong', *vlastnit* 'own'

Similarly as *být/bývat/bývávat* 'to be', also *mít/mívat* 'to have' is a highly ambiguous verb (ca 218 frames) which calls for manual conversion of the lexicon entries.

In addition, there are other candidate constructions that should be identified, located in the PDT data and converted, as, e.g., *Mariina/její taška*, 'Maria's/her bag'

III.2 UMR attributes

III.2.1 Mode

UMR uses the "mode" attribute to distinguish imperative, interrogative and expressive moods. This attribute should be filled in using the PDT "sentmode" attribute.

Warnings

- The current version of **the Czech UMR data does not cover the "mode" attribute** - the conversion is postponed to later stages of the data processing.

III.2.2 Polite

UMR adopts the "polite" attribute to indicate utterances marked for "deference with respect to the interlocutor". In PDT, this corresponds to the "politeness" grammateme; however, this grammateme is available only in a portion of the PDT-C data (namely the PDT subcorpus).

Warning:

- The current version of **the Czech UMR data does not cover the "polite" attribute** - the conversion is postponed to later stages of the data processing.

III.2.3 Degree

Correct processing of the UMR "degree" attribute requires a list of intensifiers, downtoners, and equals, which has not been compiled for Czech yet.

Warning:

- The current version of **the Czech UMR data does not cover the "degree" attribute** - the conversion is postponed to later stages of the data processing.

III.2.4 Quant

The conversion of the UMR "quant" attribute requires a deeper analysis of the quantitative structures and their representation in the PDT data, which is not available at this stage.

Warning:

- The current version of **the Czech UMR data does not cover the "quant" attribute** - the conversion is postponed to later stages of the data processing.

III.2.5 Modal-strength

The conversion of the UMR "modal-strength" attribute requires a profound analysis of the complex interplay of number of phenomena, such as sentence mode, factual modality (verbal mood), deontic modality (including semimodals), negation, modality expressions), and their treatment in PDT.

Warning:

- The current version of **the Czech UMR data does not cover the "modal-strength" attribut**e - the conversion is postponed to later stages of the data processing.

III.3 Named Entities (NEs)

III.3.1 Identification of NEs

Annotation of named entities in the PDT data is limited to identification of multiword expressions (MWE) and to their (coarse-grained) classification (e.g., person, institution, time, object, number, location, address, bibliographic entity). On top of it, MWE annotation is available only in one of the four PDT-C subcorpora.

Further, there is the NameTag 3 tool available for Czech (among others). This tool identifies proper names in text and classifies them into a set of predefined categories, such as names of persons, locations, organizations, etc., achieving state-of-the-art performance (as of February 2025).

Identification of named entities was postponed to later stages of the conversion.

III.3.2 NEs anchoring

According to the UMR principles, named entities should be anchored in an ontology. While the English UMR data uses English Wikipedia, we lean towards the use of Wikidata, as this represents a language independent resource.
However, only very preliminary and small-scale manual annotations using Wikidata for NE anchoring are available so far.

III.4 Scope for quantification and negation

Scope for quantification and negation is not covered in the Czech UMR data as this information is not fully available in the source data. An in-depth analysis of these phenomena is necessary.

III.5 Temporal relations

Temporal relations are not represented at the document-level annotation so far as only information based on grammatical tense marking is available in PDT. We suppose that the possibility of their automatic conversion is rather limited as it requires understanding of the event temporal structure, which definitely goes beyond the scope of the representation available in PDT.

III.6 Modal dependency

The conversion does not identify modality dependency as it must be based on complex understanding of the event modal structure, taking into account a number of phenomena annotated at the sentence-level (such as sentence mode, factual modality (verbal mood), deontic modality (including semimodals), negation, modality expressions).

References

Bonn Julia, Bonia, Claire, Buchholz Matt, Cheng Hsiao-Jung, Chen Alvin, Chen Ching-wen, Cowell Andrew, Crof, William, Denk Lukas, Elsayed Ahmed, Fučíková Eva, Gamba Federica, Gomez Carlos, Hajič Jan, Hajičová Eva, Havelka Jiří, Havenmeier Loden, Kilgore Ath, Kolářová Veronika, Kučová Lucie, Lai Kenneth, Li Bin, Li Jingyi, Lopatková Markéta, MacGregor Marie, Mikulová Marie, Mírovský Jiří, Nedoluzhko Anna, Myers Skatje, Novák Michal, O'Gorman Tim, Pajas Petr, Palmer Alexis, Palmer Martha, Panevová Jarmila, Post Benét, Pustejovsky James, Sgall Petr, Song Jialin, Song Li, Ševčíková Magda, Štěpánek Jan, Urešová Zdeňka, Sun Haibo, Sun Yao, Vallejos Yopán Rosa, Van Gysel Jens, Vigus Meagan, Wright-Bettner Kristin, Wu Jiawei, Xue Nianwen, Xing Dan, Xu Keer, Xu Zhixing, Yue Liulu, Zeman Daniel, Zhao Jin, Zikánová Šárka, Žabokrtský, Zdeněk (2025): *Uniform Meaning Representation 2.0*. Data/software, LINDAT/CLARIAH-CZ digital library, Charles University, Prague, Czech Republic, http://hdl.handle.net/11234/1-5902, https://umr4nlp.github.io/web/index.html.

Hajič Jan, Bejček Eduard, Bémová Alevtina, Buráňová Eva, Fučíková Eva, Hajičová Eva, Havelka Jiří, Hlaváčová Jaroslava, Homola Petr, Ircing Pavel, Kárník Jiří, Kettnerová Václava, Klyueva Natalia, Kolářová Veronika, Kučová Lucie, Lopatková Markéta, Mareček David, Mikulová Marie, Mírovský Jiří, Nedoluzhko Anna, Novák Michal, Pajas Petr, Panevová Jarmila, Peterek Nino, Poláková Lucie, Popel Martin, Popelka Jan, Romportl Jan, Rysová Magdaléna, Semecký Jiří, Sgall Petr, Spoustová Johanka, Straka Milan, Straňák Pavel, Synková Pavlína, Ševčíková Magda, Šindlerová Jana, Štěpánek Jan, Štěpánková Barbora, Toman Josef, Urešová Zdeňka, Vidová Hladká Barbora, Zeman Daniel, Zikánová Šárka, Žabokrtský Zdeněk (2024a): *Prague Dependency Treebank - Consolidated 2.0 (PDT-C 2.0).* Data/software, LINDAT/CLARIAH-CZ digital library, Charles UniversityPrague, Czech Republic, http://hdl.handle.net/11234/1-5813, https://ufal.mff.cuni.cz/pdt-c.

Hajič Jan, Fučíková Eva, Lopatková Markéta, Urešová Zdeňka (2024b): Mapping Czech Verbal Valency to PropBank Argument Labels. In *Proceedings of the Fifth International Workshop on Designing Meaning Representations (DMR 2024) LREC-COLING 2024*, ELRA Language Resource Association, p. 88-100.

Lopatková Markéta, Fučíková Eva, Gamba Federica, Štěpánek Jan, Zeman Daniel, Zikánová Šárka (2024): Towards a Conversion of the Prague Dependency Treebank Data to the Uniform Meaning Representation. In *Proceedings of the 24th Conference Information Technologies – Applications and Theory (ITAT 2024)*, CEUR-WS.org, Košice, Slovakia, p. 62-76.

Mikulová Marie, Bémová Alevtina, Hajič Jan, Hajičová Eva, Havelka Jiří, Kolářová Veronika, Kučová Lucie, Lopatková Markéta, Pajas Petr, Panevová Jarmila, Razímová Magda, Sgall Petr, Štěpánek Jan, Urešová Zdeňka, Veselá Kateřina, Žabokrtský Zdeněk (2006): *Annotation on the tectogrammatical level in the Prague Dependency Treebank. Annotation manual*. Technical report no. 2006/30, ÚFAL MFF UK, Prague, Czech Rep., 1287 pp., also available online https://ufal.mff.cuni.cz/pdt2.0/doc/manuals/en/t-layer/html/index.html

Uniform Meaning Representation (UMR) 0.9 Specification https://github.com/ufal/umr-guidelines/blob/master/guidelines.md

Appendix A

PDT functor	**UMR label**	**UMR type**	**PDT example** (from Mikulová at al., 2006)
Labels for independent clauses			
DENOM	– ^	–	*Názory*.DENOM *čtenářů.* 'Readers' opinions.'
PAR	parenthesis*	role	*Přijedu 13. prosince (pátek*.PAR *).* 'I will arrive on December 13 (Friday).'
PARTL	interjection*	role	*Pozor*.PARTL*!* 'Attention!' *Ano*.PARTL *to je pravda.* 'Yes, that is true.'
PRED	– ^	–	*Pavel dal*.PRED *kytku Marii.* 'Paul gave a flower to Mary.'
VOCAT	vocative	role	*Milá Jano*.VOCAT*!* 'Dear Jane!'
Labels for arguments			
ACT	ARG0	role	*Její manžel*.ACT *tam však pracuje dál.* 'Her husband still works there, though.' *Byl zabit bleskem*.ACT. 'He was killed by lightning.' *Je mi*.ACT *smutno.* 'I am sad.' *Sklo*.ACT *zůstává nalepené na fólii.* 'The glass stays sticked to the foil.'
PAT	ARG1	role	*Postavili stany*.PAT. 'They pitched the tents.' *Hledal houby*.PAT. 'He was looking for mushrooms.' *Hrál na klavír*.PAT. 'He played the piano.' *Zahrnul sportovce chválou*.PAT. 'He bestowed praise on the sportsmen.' *Dosáhl konce*.PAT. 'He reached the end.' *Bojí se, že bude pršet*.PAT. 'He is afraid that it might rain.' *Kniha patří Janovi*.PAT. 'The book belongs to Jan.' *Házel kamenem*.PAT. 'He was throwing a stone.' *Vyprávěl nám o zájezdu*.PAT *do Tater.* 'He was telling us about his trip to the Tatras.'
ADDR	ARG2	role	*Dal dítěti*.ADDR *hračku.* 'He gave the child a toy.' *Učí děti*.ADDR *angličtinu.* 'He teaches children English.' *Ukradl cizinci*.ADDR *peněženku.* 'He stole a wallet from a foreigner.' *Obrátil se na soud*.ADDR *s problémem.* 'He turned to the court with a problem.'
ORIG	source	role	*Vyráběli nábytek ze dřeva*.ORIG*.* 'They made furniture out of wood.' *Na malých kroužcích*.ORIG *založili novou organizaci.* 'They build a new organization on small groups.' *Získal na dětech*.ORIG *slib.* 'He got a promise from the children.' *Zdražili vstupenky z 500*.ORIG *na 550 Kč.* 'The price of the tickets rose from 500 to 550 CZK.'
EFF	effect*	role	*Považoval Pavla za odborníka*.EFF.

			'He considered Pavel a professional.' *Změnila účes z kudrn na rovné vlasy*.EFF. 'She changed her hairstyle from curly hair to straight hair.' *Bránili město před Švédy*.EFF. 'They defended the town against the Swedes.' *Petr vyprávěl o dovolené zábavné historky*.EFF. 'Petr told us amusing stories about his holiday.'
Labels for temporal (and similar) modifications			
TWHEN	temporal	role	*Přijde za týden.*TWHEN 'He will come in a week.' *Psal/Napsal to za minulého ředitele.*TWHEN 'He wrote it under the last director.' *Sejdeme se 2. února.*TWHEN 'We shall meet on February 2nd.' *Najíme se, až vyjdeme.*TWHEN 'We will eat only when we set off.' *Včera*.TWHEN *přeložil schůzi na pátek*. 'Yesterday he postponed the meeting.'
TFHL	duration	role	*Přijel na tři dni.*TFHL 'He came for three days.' *Je dlouhodobě*.TFHL *nemocen.* 'He is chronically ill.' (= He is ill for a long period)
TFRWH	temporal	role	*Včera přeložil schůzi ze čtvrtka*.TFRWH *na pátek*. 'Yesterday he postponed the meeting from Thursday to Friday.'
THL	duration	role	*Četl půl hodiny.*THL 'He was reading for half an hour.' *Přečetl to za půl hodiny.*THL 'He read it in half an hour.'
THO	frequency	role	*Každé čtyři hodiny*.THO *si musím vzít prášek.* 'I have to take a pill every four hours.'
TOWH	temporal	role	*Včera přeložil schůzi ze čtvrtka na pátek*.TOWH 'Yesterday he postponed the meeting from Thursday to Friday.'
TPAR	temporal	role	*Hraje a přitom*.TPAR *zpívá.* 'He is playing and singing at the same time.' *Zatímco spala*.TPAR*, přemýšlel jsem.* 'While she was sleeping I was thinking about it.'
TSIN	temporal	role	*Od soboty*.TSIN *nepršelo.* 'It was not raining since Saturday.' *Od toho okamžiku*.TSIN *jsem věděl, že je to on.* 'From that moment I knew that it was him.'
TTILL	temporal	role	*Dodnes*.TTILL *nevím, kde je.* 'Till today I do not know where he is'.
Labels for spatial modifications			
DIR1	start	role	*Dovážíme odtud*.DIR1 *potraviny a textil.* 'We import groceries and textile from there. *jeden z chlapců*.DIR1 'one of the boys'
DIR2	path	role	*Šli podél řeky*.DIR2 'They walked along the river. *Rommel ustupoval (údolím řeky Vardaru).*DIR2 'Rommel retreated through the valley of the Vardar river.

DIR3	goal	role	*Polož to doprostřed stolu.*DIR3 'Put it in the middle of the table.' *Kam*.DIR3 *jdete?* 'Where are you going?' *Hleděl (tváří v tvář problému)*.DIR3 'He was facing up to a problem.' *Voda mi sahá po kolena*.DIR3 'Water is reaching my knees.' *Dej to, (kam nedostane).*DIR3 'Put it where he cannot reach.'
LOC	place	role	*Jsme vždy blízko*.LOC *vás.* 'We are always close to you.' *Leží směrem k Národnímu divadlu*.LOC. 'It is located towards the National Theater.' *V oblasti*.LOC *vzdělávání máme velké mezery.* 'We have serious loophopes in the educational area.' *Jeho syn bydlí blízko.*LOC 'His son lives nearby.' *Místy*.LOC *ležel v ulicích ještě sníh.* 'There was still snow in the streets.'
Labels for causal modifications			
AIM	purpose	role	*Jsem tu pro to, abych vám pomohl*.AIM 'I am here to help you.'
CAUS	cause	role	*Nepovím vám to, (protože byste mi stejně nevěřili).*CAUS 'I will not tell you as you would not believe me anyway.' *Díky vaší pomoci*.CAUS *jsme to stihli včas.* 'Thanks to your help we made it on time.'
CNCS	but-91	discourse relation**	*(Ač zemřeli).*CNCS*, ještě mluví.* 'Although they are dead, they still speak.' *Přes své dobré vychování se nezachoval nejlépe.* 'Despite his good behaviour.CNCS he did not act very well.'
COND	condition	role	*(Jestliže Izák zemře).*COND*, komu otec předá tuto víru?* 'If Isaac dies, who will his father give his faith (to)?' *(V případě, že se nedostaví).*COND*, schůzi rozpustíme.* 'If he does not come we shall cancel the meeting.' *Formulář vydává (na telefonické požádání).*COND *zkušební ústav.* 'The conditioning house issues the form on telephonic request.'
INTT	purpose	role	*Šel nakoupit*.INTT*, aby doplnil zásoby.* 'He went shopping to replenish the stock.'
Labels for manner (and similar) modifications			
ACMP	companion	role	*Tatínek s maminkou*.ACMP *šli do divadla.* 'My father with my mother went to the theater.' *Odešel s úsměvem*.ACMP *na tváři.* 'He left with a smile on his face.'
CPR	comparison*	role	*Počínal si hazardérsky.*CPR '*He acted hazardously; i.e. like a daredevil* *Musíme udělat nepochybně menší a snazší manévr, (než byl ten minulý).*CPR 'We have to do a smaller and easier manoeuvre than was the last one.'
CRIT	regard*	role	*Byl odsouzen v souhlase s předpisy*.CRIT 'He was sentenced in compliance with the regulations.'

			Snaží se žít po vzoru velkých osobností.CRIT 'He's trying to live following the example of great personalities.' *Podle našich údajů.*CRIT *vítězí strana ODS.* 'According to our information, ODS is winning.'
DIFF	extent	role	*Nabízejí ho o 100 tisíc*.DIFF *levněji.* 'They offer it cheaper by 100 thousand.' *Zdražili ceny paliva o 50 haléřů.*DIFF 'the fuel prizes went up by 50 heller.'
EXT	extent	role	*Utkání se příliš*.EXT *nevyvedlo.* 'The match wasn't very good.' *Je daleko.*EXT *lepší než já.* 'He is far better than me.' *Jak*.EXT *dlouho to ještě potrvá?* 'How long is it going to take?'
MANN	manner	role	*Pracuje pomalu*.MANN 'He is working slowly.' *Náš vztah k Německu byl tak.*MANN *nadlouho určen.* 'Our relationship to Germany was given by this for a long time.' *Choval jsem se (tak, abych se tam nedostal).*MANN 'I behaved in such a way so that I didn't get there.'
MEANS	instrument	role	*Napsal to na počítači*.MEANS 'He wrote it on the computer.' *Pošli to po Janě.*MEANS 'Send it by Jana.' *Ten na pražské letiště přicestoval letadlem.*MEANS 'He came to the Prague airport by plane.' *Časopisecky*.MEANS *jsem povídky představil již v roce 1965.* 'I introduced the stories in magazines already in 1965.'
REG	regard*	role	*Zevnějškem*.REG *se sobě úplně podobali.* 'As to their external experience, they were similar to each other.' *rozlohou.*REG *malé Slovensko* 'small by area' *specifikace izolačních materiálů z hlediska hořlavosti.*REG 'specification of the materials with respect to their flammability' *Marie, povoláním.REG učitelka* 'Marie, a teacher by profession'
RESL	result	role	*Obarvil vajíčka na zeleno*.RESL 'He painted the eggs green.' *opálená do hněda.*RESL 'tanned to brown' *Mám ruce zmrzlé (ak, že je nenatáhnu).*RESL 'My hands are so numb with cold that I can't stretch them.'
RESTR	subtraction	role	*Kromě Pavla.*RESTR *nepřišel nikdo.* 'Except for Pavel, nobody came.' (= Pavel came, nobody else) *Kromě Pavla.*RESTR *nepřišel ještě Mirek.* 'Apart from Pavel, also Mirek didn't come' (= both Pavel and Mirek came, nobody else)
Labels for other relations			
BEN	affectee	role	*Hraje dětem/pro děti*.BEN *divadlo.* 'He plays theater to/for children.' *Padá mu.*BEN *hlava na prsa.* 'His head is falling on his breast.' *Ten pán vám.*BEN *měl ale fousy!* 'The beard the man had!' *To je další argument (proti existenci mzdové regulace)*.BEN 'That is another argument against the wage regulation.'
CONTRD	contrast-91	discourse relation**	*Zatímco loni prý v premiéře proti Samprasovi hrál*.CONTRD *chaoticky, nyní už měl plán.* 'While he was told to play

			chaotically last year in his premiere against Sampras, now he had a plan.'
HER	source	role	*Pes Blackie zdědil po svém pánovi*.HER *33 tisíc dolarů.* 'The dof Blacky inheritied after his lord 33 thousand dollars.' *Jmenovala se Barbora (podle patronky horníků).*HER 'She was named Barbora after the benefactress of miners.'
OPER	math-entity*	entity concept	*rozměr 4 krát 5 metrů* '4 x.OPER 5 meters' *pět minus dva* 'five minus.OPER two' *Výsledek 5 :.OPER 0 se nám moc zamlouval.* 'We liked the result 5:0.' *věk mezi_a*.OPER *(15) (a 20 lety)* 'the age between 15 and 20' *V jednom místě nakoupím vše od_po*.OPER *(zeleniny) (po mléčné výrobky a drogerii).* 'At one spot I can buy everything from vegetables to dairy products and cosmetics'
SUBS	substitute	role	*Do učeben zasednou otcové (místo svých synů).*SUBS 'The fathers will sit in the classrooms instead of their sons.' *Výměnou za srnku*.SUBS *dostali několik bažantů.* 'In exchange for the deer they got a few feasants.'
Labels for modifications with dual dependency			
COMPL	manner / mod (based on the parent concept)^^	role	*Hosté odcházeli spokojeni*.COMPL 'The guests were leaving satisfied.' *Hráči odcházeli ze hřiště nepřemoženi*.COMPL 'The players were leaving the field undefeated.' *Sledoval ho, (jak se chová k mladším spolužákům).*COMPL 'He watched him how he behaved to the younger classmates.' *Jako odborník*.COMPL *hodnotil situaci jako špatnou.* 'As an expert he evaluated the situation as a bad one.'
	mod-of (reentrancy)	role	
Labels for nominal modifications			
APP	possessor	role	*manžel slavné spisovatelky*.APP 'husband of the famous writer' *moji.*APP *rodiče* 'my parents' *příslušník armády.*APP 'a ember of the army' *tým brankářů*.APP 'a team of goalkeepers' *míra nezaměstnanosti*.APP 'the unemployment rate' *dům mého otce.*APP 'my father's house' *široký pás území*.APP 'a wide stripe of land' *auto roku.*APP 'the car of the year'
AUTH	source	role	*autorova*.AUTH *současná tvorba* 'the author's present work' *básně Vítězslava Nezvala.*AUTH 'poems by Vítězslav Nezval' *dekret (nového ukrajinského prezidenta Leonida Kučmy).*AUTH 'the order of the new president of Ukraine, Leonid Kutchma'
ID	name	subrole	*opera Brundibár*.ID 'opera Brundibár' *nápis (Obětem války).*ID 'the sign "To the victims of war"' *Řekl to v úterý v pořadu Proč*.ID 'He said it on Sunday in the programme "Why".' *pokyn mlčet*.ID 'the instruction to be silent'
MAT	mod	role	*sklenice piva*.MAT 'a glass of beer' *polovina / skupina lidí.*MAT 'half of the / the group of people'

			*jedna porce zmrzliny.*MAT 'one portion of ice cream' *milióny židů.*MAT 'millions of Jews' *množství škodlivých látek.*MAT 'amount of harmful substance' *část Německa.*MAT 'part of Germany'
RSTR	mod	role	*drsné.*RSTR *počasí* 'rough weather' *proti destruktivnímu.*RSTR *způsobu hry* 'against the destructive way of playing' *několik.*RSTR *měsíců* 'a few months' *dvojí.*RSTR *státní občanství* 'double nationality' *Karlova.*RSTR *univerzita* 'Charles University' *Prostřelil libereckého brankáře.*RSTR *Maiera* 'He shot through the Liberec goalman Maier'
Labels for multiword expressions			
CPHR	predicative-noun*	role	*Dostali rozkaz*.CPHR *nevycházet ze stanů.* 'They got a command not to leave their tents.' *Učinil rozhodnutí.*CPHR 'He made a decision.' *Je třeba*.CPHR *odejít.* 'It is necessary to leave.'
DPHR	part-of-phraseme*	role	*Jde mi na nervy*.DPHR 'He gets on my nerves.' *Široko daleko.*DPHR *nikdo nebyl.* 'There was no one far away.'
FPHR	FPHR*	role	*Cizinec zvolal: "This*.FPHR *is*.FPHR *not*.FPHR *true*.FPHR*"* 'The foreigner shouted: "This is not true".'
Labels for coordinated structures			
ADVS	but-91	discourse relation**	*Koupil chleba, ale*.ADVS *ne*.CM *mléko*. 'He bought bread but not milk.' *O výrobek by byl zájem, přesto*.CM *však*.ADVS *ještě*.CM n*emáme výrobce.* 'People would be interested in the product but still there is no producer yet.' *Nebyl to ani*.CM *Petr, ale*.ADVS *též*.CM *ani*.CM *Pavel.* 'It wasn't Petr but it wasn't Pavel either.'
CONFR	contrast-91	discourse relation**	*Bristol je v Anglii, kdežto.*ADVS *Glasgow je ve Skotsku.* 'Bristol is in England, whereas Glasgow is in Scotland.' *Svobodní mládenci mívají nepořádek kolem sebe, kdežto.*ADVS *ženatí naopak.*CM *mívají nepořádek v duši.* 'Bachelors often have a mess all around them whereas married men, on the other hand, have a mess in their souls.'
CONJ	and	discourse relation**	*Mezi smysly patří zrak a.*CONJ *sluch a hmat.* 'Eyesight and hearing and touch belong to the senses.' *žáci i.*CONJ *žákyně* 'male as well as female pupils' *Ve Francii není ani_ani.*CONJ *(vítězů) (poražených).* 'There are neither winners, nor defeated in France.'
CONTRA	contra-entity*	entity concept	*Na veřejnosti je tato otázka vnímána jako spor Klaus versus*.CONTRA *Zieleniec.* 'In public, this issue is perceived as a Klaus versus Zieleniec dispute.' *Utkání Sparta -.CONTRA Slavia bylo zahájeno.* 'The Sparta - Slavia match has started.'

CSQ	have-result-91	reification	*Je to utajeno,*.CSQ *tedy*.CM *chráněno.* 'It is a secret, hence it is protected.' *Pracoval nezodpovědně, a*.CSQ *proto*.CM *také*.CM *dostal výpověď.* 'His was irresponsible, therefore he was fired.' *Byl nemocný, a*.CSQ *tudíž*.CM *nepřišel.* 'He was sick so that's why he didn't come.' *Špatně se učil, načež*.CSQ *propadl u zkoušky.* 'He wasn't learning properly, which is why he failed.'
DISJ	or	discourse relation**	*Mají, či*.DISJ *nemají pravdu?* 'Are they right, or not?' *Vysloví se buď_nebo*.DISJ *(pro) (proti návrhu)*. 'They will be either for or against the proposal.' *Použijí Rakousko,*.DISJ *případně*.CM *i*.CM *Španělsko.* 'They are going to use Austria, possibly Spain, too.'
GRAD	gradation*	discourse relation**	*Nemůže se pohnout, natož*.GRAD *vstát.* 'He can't move, let alone get up.' *Byl v tomto lidu oblíbený, a*.GRAD *navíc*.CM *vynikal krásou.* 'He was popular, moreover he was very handsome.' *To je způsobeno jednak*.CM *dodaným teplem, ale*.GRAD *hlavně*.CM *cenami.* 'This is caused by the heat supply but mainly by the prices.'
REAS	have-cause-91	reification	*Nemohu odejít, neboť.REAS ještě.CM nepřestalo pršet.* 'I can't leave since it hasn't stopped raining yet.' *Úkol splníme, vždyť*.REAS *také*.CM *není obtížný.* 'We'll fulfil the task, for it is not difficult.'
Label for apposition			
APPS	identity-91	abstract predicate	*Labe/*.APPS *Elbe* 'Labe/Elbe' *paviáni, čili*.APPS *africké opice* 'baboons, or African monkeys' *Hobit aneb*.APPS *Cesta tam a zase zpátky* 'The Hobbit or there and back again.' *Právo je souhrnem norem, to jest*.APPS *předpisů, zákazů a sankcí.* 'Law is a collection of norms, i.e. regulations, prohibitions and sanctions.'
Relation roles for clausal markers			
CM	clausal marker*,♮	role	*Rozpočet nejenže*.CM *není přebytkový, ale*.GRAD *dokonce*.CM *je skrytě deficitní.* 'The budget not only isn't surplus, it is even covertly deficit.'
ATT	clausal-marker*	role	*Je to bohudíky*.ATT *za námi* 'Thank God, it is over.' *To je fakticky*.ATT *zlé.* 'That is really bad.' *Jenom*.ATT *se opovaž.* 'Just dare.' *Copak*.ATT *peníze, o ty by nebylo.* 'Well, money, it wouldn't matter' *Dopadne to, doufejme*.ATT *dobře.* 'It will turn out well, let's hope.' *Víte*.ATT *, to je složité.* 'It is complicated, you know'
INTF	INTF*	role	*To*.INTF *Jirka ještě spí.* 'EMPH Jirka is still sleeping.' *Víš, on*.INTF *je náš Baryk docela hodnej.* 'You know, EMPH our Baryk is quite nice.'

MOD	clausal-marker*	role	*Pravděpodobně*.MOD *přijdeme.* 'We will probably come.' *Vím jistě*.MOD *, že Praha mě poznamená.* I know for sure that Prague will affect me.' *Amádní špičky si prý*.MOD *od něj udržují odstup.* 'The army elite are said to keep their distance from him.'
PREC	clausal-marker*	role	*Jsem tedy*.PREC *šťasten.* 'So I am happy.' *Ale.*PREC *to zatím není náš případ.* 'But that is not our case yet.'
RHEM	clausal-marker*,♮	role	*Jen*.RHEM *on o tom nevěděl nic.* 'He just didn't know anything about it.' *Za povážlivou označil Kalvoda i.*RHEM *délku vazby.* 'Kalvoda also described the length of the detention as serious.' *Teprve.*RHEM *před týdnem přestala za prací do Púchova dojíždět.* *'She only stopped commuting to Púchov for work a week ago.'*

* NEW to UMR
^ Functors DENOM and PRED (for roots of independent clauses) remain in several complex patterns (will be fixed in the next data release)
** Formally, discourse relations are represented as special nodes for abstract concepts in UMR graphs.
^^ Functor COMPL remains in several complex patterns (will be fixed in the next data release)
♮ With CM and RHEM functors, special treatment of nodes encoding negative polarity necessary

The ÚFAL Technical Report Series

ÚFAL (Ústav formální a aplikované lingvistiky; https://ufal.mff.cuni.cz) is the Institute of Formal and Applied linguistics, at the Faculty of Mathematics and Physics of Charles University, Prague, Czech Republic. The Institute was established in 1990 after the political changes as a continuation of the research work and teaching carried out by the former Laboratory of Algebraic Linguistics since the early 60s at the Faculty of Philosophy and later the Faculty of Mathematics and Physics. Together with the "sister" Institute of Theoretical and Computational Linguistics (Faculty of Arts) we aim at the development of teaching programs and research in the domain of theoretical and computational linguistics at the respective Faculties, collaborating closely with other departments such as the Institute of the Czech National Corpus at the Faculty of Philosophy and the Department of Computer Science at the Faculty of Mathematics and Physics.

The ÚFAL technical report series has been established with the aim of disseminate topical results of research currently pursued by members, cooperators, or visitors of the Institute. Since November 1996, the following reports have been published.

ÚFAL TR-1996-01 Eva Hajičová, *The Past and Present of Computational Linguistics at Charles University*
Jan Hajič and Barbora Hladká, *Probabilistic and Rule-Based Tagging of an Inflective Language – A Comparison*

ÚFAL TR-1997-02 Vladislav Kuboň, Tomáš Holan and Martin Plátek, *A Grammar-Checker for Czech*

ÚFAL TR-1997-03 Alla Bémová at al., Anotace na analytické rovině, *Návod pro anotátory (in Czech)*

ÚFAL TR-1997-04 Jan Hajič and Barbora Hladká, *Tagging Inflective Languages: Prediction of Morphological Categories for a Rich, Structural Tagset*

ÚFAL TR-1998-05 Geert-Jan M. Kruijff, *Basic Dependency-Based Logical Grammar*

ÚFAL TR-1999-06 Vladislav Kuboň, *A Robust Parser for Czech*

ÚFAL TR-1999-07 Eva Hajičová, Jarmila Panevová and Petr Sgall, *Manuál pro tektogramatické značkování (in Czech)*

ÚFAL TR-2000-08 Tomáš Holan, Vladislav Kuboň, Karel Oliva, Martin Plátek, *On Complexity of Word Order*

ÚFAL/CKL TR-2000-09 Eva Hajičová, Jarmila Panevová and Petr Sgall, *A Manual for Tectogrammatical Tagging of the Prague Dependency Treebank*

ÚFAL/CKL TR-2001-10 Zdeněk Žabokrtský, *Automatic Functor Assignment in the Prague Dependency Treebank*

ÚFAL/CKL TR-2001-11 Markéta Straňáková, *Homonymie předložkových skupin v češtině a možnost jejich automatického zpracování*

ÚFAL/CKL TR-2001-12 Eva Hajičová, Jarmila Panevová and Petr Sgall, *Manuál pro tektogramatické značkování (III. verze)*

ÚFAL/CKL TR-2002-13 Pavel Pecina and Martin Holub, *Sémanticky signifikantní kolokace*

ÚFAL/CKL TR-2002-14 Jiří Hana, Hana Hanová, *Manual for Morphological Annotation*

ÚFAL/CKL TR-2002-15 Markéta Lopatková, Zdeněk Žabokrtský, Karolína Skwarská and Vendula Benešová, *Tektogramaticky anotovaný valenční slovník českých sloves*

ÚFAL/CKL TR-2002-16 Radu Gramatovici and Martin Plátek, *D-trivial Dependency Grammars with Global Word-Order Restrictions*

ÚFAL/CKL TR-2003-17 Pavel Květoň, *Language for Grammatical Rules*

ÚFAL/CKL TR-2003-18 Markéta Lopatková, Zdeněk Žabokrtský, Karolina Skwarska, Václava Benešová, *Valency Lexicon of Czech Verbs VALLEX 1.0*

ÚFAL/CKL TR-2003-19 Lucie Kučová, Veronika Kolářová, Zdeněk Žabokrtský, Petr Pajas, Oliver Čulo, *Anotování koreference v Pražském závislostním korpusu*

ÚFAL/CKL TR-2003-20 Kateřina Veselá, Jiří Havelka, *Anotování aktuálního členění věty v Pražském závislostním korpusu*

ÚFAL/CKL TR-2004-21 Silvie Cinková, *Manuál pro tektogramatickou anotaci angličtiny*

ÚFAL/CKL TR-2004-22 Daniel Zeman, *Neprojektivity v Pražském závislostním korpusu (PDT)*

ÚFAL/CKL TR-2004-23 Jan Hajič a kol., *Anotace na analytické rovině, návod pro anotátory*

ÚFAL/CKL TR-2004-24 Jan Hajič, Zdeňka Urešová, Alevtina Bémová, Marie Kaplanová, *Anotace na tektogramatické rovině (úroveň 3)*

ÚFAL/CKL TR-2004-25 Jan Hajič, Zdeňka Urešová, Alevtina Bémová, Marie Kaplanová, *The Prague Dependency Treebank, Annotation on tectogrammatical level*

ÚFAL/CKL TR-2005-27 Jiří Hana, Daniel Zeman, *Manual for Morphological Annotation (Revision for PDT 2.0)*

ÚFAL/CKL TR-2005-28 Marie Mikulová a kol., *Pražský závislostní korpus (The Prague Dependency Treebank) Anotace na tektogramatické rovině (úroveň 3)*

ÚFAL/CKL TR-2005-29 Petr Pajas, Jan Štěpánek, *A Generic XML-Based Format for Structured Linguistic Annotation and Its application to the Prague Dependency Treebank 2.0*

ÚFAL/CKL TR-2006-30 Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolařová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, Magda Razímová, Petr Sgall, Jan Štěpánek, Zdeňka Urešová, Kateřina Veselá, Zdeněk Žabokrtský, *Annotation on the tectogrammatical level in the Prague Dependency Treebank (Annotation manual)*

ÚFAL/CKL TR-2006-31 Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolařová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, Petr Sgall, Magda Ševčíková, Jan Štěpánek, Zdeňka Urešová, Kateřina Veselá, Zdeněk Žabokrtský, *Anotace na tektogramatické rovině Pražského závislostního korpusu (Referenční příručka)*

ÚFAL/CKL TR-2006-32 Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolařová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, Petr Sgall,Magda Ševčíková, Jan Štěpánek, Zdeňka Urešová, Kateřina Veselá, Zdeněk Žabokrtský, *Annotation on the tectogrammatical level in the Prague Dependency Treebank (Reference book)*

ÚFAL/CKL TR-2006-33 Jan Hajič, Marie Mikulová, Martina Otradovcová, Petr Pajas, Petr Podveský, Zdeňka Urešová, *Pražský závislostní korpus mluvené češtiny. Rekonstrukce standardizovaného textu z mluvené řeči*

ÚFAL/CKL TR-2006-34 Markéta Lopatková, Zdeněk Žabokrtský, Václava Benešová (in cooperation with Karolína Skwarska, Klára Hrstková, Michaela Nová, Eduard Bejček, Miroslav Tichý) *Valency Lexicon of Czech Verbs. VALLEX 2.0*

ÚFAL/CKL TR-2006-35 Silvie Cinková, Jan Hajič, Marie Mikulová, Lucie Mladová, Anja Nedolužko, Petr Pajas, Jarmila Panevová, Jiří Semecký, Jana Šindlerová, Josef Toman, Zdeňka Urešová, Zdeněk Žabokrtský, *Annotation of English on the tectogrammatical level*

ÚFAL/CKL TR-2007-36 Magda Ševčíková, Zdeněk Žabokrtský, Oldřich Krůza, *Zpracování pojmenovaných entit v českých textech*

ÚFAL/CKL TR-2008-37 Silvie Cinková, Marie Mikulová, *Spontaneous speech reconstruction for the syntactic and semantic analysis of the NAP corpus*

ÚFAL/CKL TR-2008-38 Marie Mikulová, *Rekonstrukce standardizovaného textu z mluvené řeči v Pražském závislostním korpusu mluvené češtiny. Manuál pro anotátory*

ÚFAL/CKL TR-2008-39 Zdeněk Žabokrtský, Ondřej Bojar, *TectoMT, Developer's Guide*

ÚFAL/CKL TR-2008-40 Lucie Mladová, *Diskurzní vztahy v češtině a jejich zachycení v Pražském závislostním korpusu 2.0*

ÚFAL/CKL TR-2009-41 Marie Mikulová, *Pokyny k překladu určené překladatelům, revizorům a korektorům textů*

z Wall Street Journal pro projekt PCEDT

ÚFAL/CKL TR-2011-42 Loganathan Ramasamy, Zdeněk Žabokrtský, *Tamil Dependency Treebank (TamilTB) – 0.1 Annotation Manual*

ÚFAL/CKL TR-2011-43 Nguy Giang Linh, Michal Novák, Anna Nedoluzhko, *Coreference Resolution in the Prague Dependency Treebank*

ÚFAL/CKL TR-2011-44 Anna Nedoluzhko, Jiří Mírovský, *Annotating Extended Textual Coreference and*

Bridging Relations in the Prague Dependency Treebank

ÚFAL/CKL TR-2011-45 David Mareček, Zdeněk Žabokrtský, *Unsupervised Dependency Parsing*

ÚFAL/CKL TR-2011-46 Martin Majliš, Zdeněk Žabokrtský, *W2C – Large Multilingual Corpus*

ÚFAL TR-2012-47 Lucie Poláková, Pavlína Jínová, Šárka Zikánová, Zuzanna Bedřichová, Jiří Mírovský,

Magdaléna Rysová, Jana Zdeňková, Veronika Pavlíková, Eva Hajičová,

Manual for annotation of discourse relations in the Prague Dependency Treebank

ÚFAL TR-2012-48 Nathan Green, Zdeněk Žabokrtský, *Ensemble Parsing and its Effect on Machine Translation*

ÚFAL TR-2013-49 David Mareček, Martin Popel, Loganathan Ramasamy, Jan Štěpánek, Daniel Zemana, Zdeněk Žabokrtský, Jan Hajič *Cross-language Study on Influence of Coordination Style on Dependency Parsing Performance*

ÚFAL TR-2013-50 Jan Berka, Ondřej Bojar, Mark Fishel, Maja Popović, Daniel Zeman, *Tools for Machine Translation Quality Inspection*

ÚFAL TR-2013-51 Marie Mikulová, *Anotace na tektogramatické rovině. Dodatky k anotátorské příručce (s ohledem na anotování PDTSC a PCEDT)*

ÚFAL TR-2013-52 Marie Mikulová, *Annotation on the tectogrammatical level. Additions to annotation manual (with respect to PDTSC and PCEDT)*

ÚFAL TR-2013-53 Marie Mikulová, Eduard Bejček, Jiří Mírovský, Anna Nedoluzhko, Jarmila Panevová, Lucie Poláková, Pavel Straňák, Magda Ševčíková, Zdeněk Žabokrtský, *Úpravy a doplňky Pražského závislostního korpusu (Od PDT 2.0 k PDT 3.0)*

ÚFAL TR-2013-54 Marie Mikulová, Eduard Bejček, Jiří Mírovský, Anna Nedoluzhko, Jarmila Panevová, Lucie Poláková, Pavel Straňák, Magda Ševčíková, Zdeněk Žabokrtský, *From PDT 2.0 to PDT 3.0 (Modifications and Complements)*

ÚFAL TR-2014-55 Rudolf Rosa, *Depfix Manual*

ÚFAL TR-2014-56 Veronika Kolářová, *Valence vybraných typů deverbativních substantiv ve valenčním slovníku PDT-Vallex*

ÚFAL TR-2014-57 Anna Nedoluzhko, Eva Fučíková, Jiří Mírovský, Jiří Pergler, Lenka Šíková, *Annotation of coreference in Prague Czech-English Dependency Treebank*

ÚFAL TR-2015-58 Zdeňka Urešová, Eva Fučíková, Jana Šindlerová, *CzEngVallex: Mapping Valency between Languages*

ÚFAL TR-2015-59 Kateřina Rysová, Magdaléna Rysová, Eva Hajičová, *Topic–Focus Articulation in English Texts on the Basis of Functional Generative Description*

ÚFAL TR-2016-60 Kira Droganova, Daniel Zeman, *Conversion of SynTagRus (the Russian dependency treebank) to Universal Dependencies*

ÚFAL TR-2018-61 Lukáš Kyjánek, *Morphological Resources of Derivational Word-Formation Relations*

ÚFAL TR-2019-62 Zdeňka Urešová, Eva Fučíková, Eva Hajičová, *CzEngClass: Contextually-based Synonymy and Valency of Verbs in a Bilingual Setting (CzEngClass: Kontextová synonymie a valence sloves v bilingvním prostředí)*

ÚFAL TR-2019-63 Ján Faryad, *Identifikace derivačních vztahů ve španělštině*

ÚFAL TR-2020-64 Marie Mikulová, Jan Hajič, Jíří Hana, Hana Hanová, Jaroslava Hlaváčová, Emil Jeřábek, Barbora Štěpánková, Barbora Vidová Hladká, Daniel Zeman, *Manual for Morphological Annotation. Revision for Prague Dependency Treebank – Consolidated 2020 release*

ÚFAL TR-2021-65 Rudolf Rosa, *Technická zpráva o vývoji projektu THEaiTRE v roce 2020*

ÚFAL TR-2021-66 Anna Nedoluzhko, Michal Novák, Martin Popel, Zdeněk Žabokrtský, Daniel Zeman *Coreference meets Universal Dependencies – a pilot experiment on harmonizing coreference datasets for 11 languages*

ÚFAL TR-2021-67 THEaiTRobot 1.0, David Košťák, Daniel Hrbek, Rudolf Rosa, Ondřej Dušek, *AI: When a Robot Writes a Play*

ÚFAL TR-2021-68 Markéta Lopatková, Václava Kettnerová, Anna Vernerová, Eduard Bejček, Zdeněk Žabokrtský, *Valenční slovník českých sloves VALLEX*

ÚFAL TR-2021-69 Niyati Bafna, Jan Bodnár, Lukáš Kyjánek, Emil Svoboda, Madga Ševčíková, Jonáš Vidra, Zdeněk Žabokrtský *Towards Universal Segmentations: Survey of Existing Morphosegmentation Resources*

ÚFAL TR-2021-70 Zdeňka Urešová, Eva Fučíková, Jan Hajič, Karolina Zaczynska, *Annotation guidelines for German verbal synonyms included in SynSemClass Lexicon*

ÚFAL TR-2022-71 THEaiTRobot 2.0, Josef Doležal, Klára Vosecká, Tomáš Musil, David Mareček, Rudolf Rosa, *Permeation / Prostoupení*

ÚFAL TR-2022-72 Cristina Fernández-Alcaina, Eva Fučíková, Zdeňka Urešová, *Annotation guidelines for Spanish verbal synonyms in the SynSemClass Lexicon*

ÚFAL TR-2023-73 Zdeňka Urešová, Cristina Fernández-Alcaina, Eva Fučíková, Jan Hajič, *SynSemClass Czech and English Annotation Guidelines*

ÚFAL TR-2025-74 Markéta Lopatková, Eva Fučíková, Federika Gamba, Jan Hajič, Hana Hledíková, Marie Mikulová, Michal Novák, Jan Štěpánek, Daniel Zeman, Šárka Zikánová, *UMR 2.0 - Czech: Release Notes*