**Meeting Sep 4, 2023 (Markéta, Zdeňka, Federica, Šárka, J. Hajič, Míša, Cristina)**

**Next meetings:**

* Monday September, 11 at **14:00**, lecture hall S1 – the same day SIGDIAL is held at hotel Pyramida
* Hejnice (September, 17-19): an unofficial UMR meeting; no presentation there
* Monday September 25, at 13:00, lecture hall S1 (absent: Zdeňka, Markéta, JH)

**Míša:**

* short introduction of the new release of the UMR data, as described in the email.
* Visualization based on this corpus. TASK any comments and suggestions are welcome – how to distinguish any relations, concepts etc.
* Another project will be a conversion of Czech files to UMR
* We can let the UMR team know this tool exists. It is not publishable yet.

Michal’s report about the tool in the e-mail:

Hi all,

I adjusted the AMR Reader tool (<https://github.com/panx27/amr-reader>) to be able to read UMR files and visualize them. Following our discussion two weeks ago, I also changed the visualization so that atomic values are displayed as nodes' attributes, which makes the final visualization less messy even for large graphs. I will publish UMRGraphViz on UFAL GitHub soon.

I used the tool to visualize the documents from the recent UMR 1.0 release. You can browse the documents here: <https://ufallab.ms.mff.cuni.cz/~mnovak/umr/graphs/umr-v1.0/> (open visualization.html for a given language and document to visualize it).

Known issues:

- it displays only sentence graphs. Alignment and document-level annotation are ignored.

- original sentence in its surface form is not displayed for some languages (e.g. Arapaho, Navajo). The reason is that the format of representing this information has slightly changed in the official release (compared to <https://github.com/cu-clear/UMR-Annotation/tree/main>).

- there is no visualization for arapaho\_umr-0001.txt and other Arapaho documents may be corrupted, too. Parsing of that file failed from yet unknown reason.

Best

M.

* Plan: small adjustments, then release

Markéta: eventive concepts

* What about words like *válka, příchod*? (Either entities – conservative approach, or events – based on the Vallex, they have a valency frame. Problem: not consistent, some similar words in PDT Vallex are not assigned a valency frame.)
* JH: it would be good to use our sources. Nominal valency does not depend on the distinction between states, events etc. E.g., a book can have an author as a position in a valency frame, nevertheless, it does not mean that a book is an event.

Thus, even if the NP has a frame, it does not mean it is an event.

Markéta: we need to know what this difference means for the annotation

JH: let us see real examples from Míša’s database

Míša: <https://ufallab.ms.mff.cuni.cz/~mnovak/umr/graphs/umr-v1.0/>

0001.txt, sentence 7 (roofing as a material on the roof)

Markéta: with some words, there can be connection to the Wiki (as entities) as well as to a valency frame (e.g., WW2). Can we connect one concept with both features (valency frame and Wiki)?

JH: what can prevent us to do that?

Markéta: We have not seen such an example in the manual.

JH: Most events do not have Wikipedia entries, it is rather entities.

Do we want to connect a general entry of war (a type) to Wikipedia, as well as special occurrences (WW2)? Do we want to have both connections with one phrase in our text?

JH: maybe it would be good to make a differences (and mark it) between a type and an occurrence (individuum). Maybe, we should always have a type (or valency frame, or later on Synsemclass, Propbank…).

Zdeňka: mé psaní (my writing) – how to interpret it? Either I am the author, or just an owner.

JH: when it is a letter, it is an entity.

JH: in UMR, there is just an unclear border between entities and events. But we have to follow the instructions. It would be good to put “type” everywhere where it is possible.

The primary database for us is “Wikidata”.

Example in UMR data: file 0001, sentence 4: Eyewitness is a person who witnesses.

Míša: should the types be linked to the Wikidata?

Federica: working on a file in Latin, still technical and scientific issues. Trying to prepare a PDT-like file that could be later converted to UMR-like file.

Šárka: can we annotate the Estonian file in the Míša’s tool?

Markéta: yes, she tried that. Dan did it too, in his way. Markéta proposes us to try the annotation separately. TASK try (before Hejnice) to annotate at least a part of Estonian part. In Hejnice, we can compare it. In a text format, then Míša can converse it to his visualization.

**Meeting Aug 28, 2023 (Markéta, Dan, Zdeňka, Federica, Šárka, Eva F., J. Hajič)**

**Markéta:**

Eventive concept document – TASK comment it

One or two concepts? – not ready yet for comments

Functors to AMR labels:

* Markéra: Table: conversion between functors and AMR labels (update) – not all functors are covered yet. Questions esp. about coordinative functors, they are solved in a different way in AMR.
* Eva F.: document – verb-specific mappings between functors and argument roles (from SynSemClass)

Based on these tables, Michal can prepare a conversion from PDT to AMR

**Plans:**

Dan – revise the annotation of the Estonian file

Šárka – meet Markéta, discuss what is needed now

Federica – merging two versions of Vallex

**J. Hajič:**

Martha Palmer published small data – English (300 sentences), Arapau – official UMR.

It will soon be in Lindat. Examples for us. The release is being prepared by Jin, communicating with Ondřej Košarko.

Probably the same version of data is on the UMR Github, or maybe it is smaller?

**Markéta:**

PropBank webpage – there were some abstract predicates there, but they somehow disappeared.

**Next meetings:**

* Monday September, 4 at 13:00, lecture hall S1
* Monday September, 11 at **14:00**, lecture hall S1 – the same day SIGDIAL is held at hotel Pyramida
* Hejnice (September, 17-19): an unofficial UMR meeting; no presentation there
* Monday September 25, at 13:00, lecture hall S1 (absent: Zdeňka, Markéta, JH)

**Reservations for October:** 510 – ask Jirka Mírovský for reservation when it is open; or later we can try it ourselves.

# Next meetings

28/8 – 13:00 Dan +, Federica +, JH +, Markéta +, Míša +, Zdeňka +;

Cristina –, Eva +?, Šárka ?

# Meeting 2023-08-21 (Dan, Eva, Federica, Markéta, Michal)

**(0) cu-clear/UMR-Annotation repository**

invitation by Julia Bonn (July 26)

<https://github.com/cu-clear/UMR-Annotation>

Dan did not have an access to the repository – he should send an email to JH (he will ask Julia to add Dan as a member).

**(1) updated AMR viewer for UMR annotations**

Michal: visualization of the Eng data from the UMR-Annotation repository

just sentence-level annotation

problem: too complicated graph

relations displayed as edges between concepts, repeating mentions as single nodes … OK

BUT attribute-value pairs displayed as relations as well (inherited from the original viewer) … KO

todo (Miša):

attributes (with pre-defined set of values) should be represented as attributes of individual nodes

**(2) entities vs. events**

solution from the mast meeting (2023-07-31) discussed

todo (Markéta): update the Eventive concept notes accordingly

<https://github.com/ufal/UMR/blob/main/doc/eventive-concepts.md>

**(3) functors 🡪 UMR relations/attributes conversion**

Markéta: 2014 table updated and commited to the github repository (still provisional)

<https://github.com/ufal/UMR/tree/main/tecto2umr>

todo (Markéta): further refinements needed

**(4) verb specific functors 🡪 UMR relations conversion**

Eva: extracted conversions for individual SSC classes

todo (Eva): conversions for individual verb frames

# Next meetings

21/8 – 13:00 Cristina +, Dan +, Federica + (zoom); Markéta +, Míša + (?zoom);

JH in China – zoom +?, Zdeňka +?, Eva ?, Šárka ?

28/8 – 13:00 JH +, Dan +, Federica +, Markéta +, Zdeňka +

Eva ?, Šárka ?

Cristina –

# Meeting 2023-07-31 (JH, Cris, Eva, Markéta, Michal, Zdeňka)

**(0) cu-clear/UMR-Annotation repository**

invitation by Julia Bonn (July 26)

Does it arrive to everybody? If not, send an email to JH, he will ask her to add you as a member.

**(1) convertor from AMR**:

Michal:

update and plans:

* defined first draft of u-layer
* work on vizualization in Tred, now it works (with Dan)

**(2) coreference / co-occurences – visualization in TrEd:**

… keep both nodes (and merge them on demand / when converting u-layer to UMR

??? both for inter-sentence and intra-sentence relations

*Peter likes himself.*

UMR annotation: and its visualization in TrEd:

#Ref

l/like

p/Peter

:ARG0

:ARG1

l/like

:ARG0 (p/person

:name ( n/name :op1 "Peter" ))

:ARG1 (p)

**(3) relations vs. attributes:**

* relation between two concepts (two UMR nodes)
* attributes: only for one concept/UNR node, with (?close] set of values (like sg/pl or numerals)

BUT there are some labels defined as attributes in UMR guidelines which are later treated as relations (as well)

e.g., :quant in ex. 1 (1) *… more than eight months* :quant as attribute (with value 8)

:range (m/ more-than

:op1 (t3/ temporal-quantity :quant **8**

:unit (m2/ month)))

BUT 1 (6) … *the next several days* :quant as relation (*several* as a child)

(t/ temporal-quantity

:quant (**s2/ several**)

:unit (d/ day

:mod (n/ next)))

***provisional solution for u-layer***:

* all UMR attributes that appear also as a relation provisionally represent as relations
* UMR attributes with only predefined (close) set of values represent as attributes

list of attributes/relations <https://github.com/ufal/UMR/blob/main/doc/relations-attributes.md>

todo: update whenever found a new attribute/relation (or change their description there)

**(4) functors** 🡪 **UMR relations/attributes** **conversion** … todo

**(5) entities vs. events** … unclear cases

**rule 1**: Whenever a concept has a roleset 🡪 represent it as event

(link to the lexicon with its roleset)

**rule 2**: JH: ***Be conservative (at least for the time being)!***

i.e.: We do not want to add a big number of nouns / adjectives to the lexicon OR represent them as verbs, unless we have a clear case of a deverbal noun (ending with -ní/-tí) / deverbal adjective (??) … YES - morphological criterion!!!

🡪 ***verbs:***

**rule 1a**: JH: consider them as eventive concepts (both action and state/stative verbs)

***!! in conflict with some UMR examples !!***

TODO: to be changed/stressed in the internal guidelines <https://github.com/ufal/UMR/blob/main/doc/eventive-concepts.md>

🡪 ***nouns:***

e.g., *jídlo* … should not be represented as ARG1of *jíst* but as an entity

(as it can be related also to other eventive concepts as, e.g., *vařit*)

BUT agentive nouns like *teacher*, *driver* … as ARG0 of the respective verb (*teach-01*, *drive-01*) even in context different than teaching, driving (intention of the speaker to use just these nouns) ***!! in conflict with some UMR examples !!***

TODO: to be changed in the internal guidelines <https://github.com/ufal/UMR/blob/main/doc/eventive-concepts.md>

🡪 ***adjectives:***

**rule 1b**: clear deverbal cases should be represented as events

(mapped onto respective verbs in the lexicon, as ARGx-of …)

e.g., *podobný* ‘resembing’ 🡪 *podobat se* ‘resembe’(morphology!)

**rule 2a**: primary adjectives (like *chytrý, vysoký, clever, high*) as non-eventive concepts

compare with UMR examples

3-1-1-3 (2a) The tall man...

3-1-1-3 (2b) The man, who is tall...

(m/ man

:mod(t/ tall)) (the UMR graph by ML, guidlines only says that this is not an event)

3-3-1-3 (1b)The doctor is tall.

(h/ have-mod-91

:ARG1 (d/ doctor

:ref-number Singular)

:ARG2 (t/ tall))

Compare also to AMR guidelines, sect. Adjectives that invoke predicates

<https://github.com/amrisi/amr-guidelines/blob/master/amr.md#adjectives-that-invoke-predicates>

(i)*the attractive man* = the man who is ARG0-of attract-01 … **OK for CZ** (morphology!)

the same solution for Eng. adjectives ending with -ed (acquainted)

the same solution for other types of endings, like -able (edible) -ful,

(ii) to be+adjectives … often exist natural corresponding verbal predicates … **NOT for CZ ??**

*The soldier was aware of the battle.* … *be aware (of X)* 🡪 *realize-01*

(iii) adjectives without natural verbal predicates … create predicate … **NOT for CZ ??**

*be responsible (for X)* 🡪 *responsible-01* (cause), *responsible-02* (trustworthy), *responsible-03* (duty)

*be nervous (about X)* 🡪 *nervous-01*

*be serious (about X)* 🡪 *serious-01* (no kidding), *serious-02* (grave)

(iv) adjectives like *sad, white, and free* … as predicates if there is an implied event or proces

… **NOT for Czech ??**

general question: How much inferencing we want to include in the annotation?

JH: Be conservative with ***reification*** (reification = converting a role into a concept is called reification)!

Reification is used, e.g., for the change of TFA in AMR guidelines (“One reason is to make a relation the focus of an AMR fragment.”)

more advanced solution: create more general predicates like *color-01* … **?? not now for CZ**

“ideal” solution: JH: all adjectives as predicates (like the proposal by Dan), but **not now for CZ**

TODO: if agreed, then to be changed in the internal guidelines <https://github.com/ufal/UMR/blob/main/doc/eventive-concepts.md>

# Meeting 2023-07-24 (Dan, Michal, Eva, Markéta, Federica)

**AMR guidelines**

We should consult also ***AMR*** guidelines!

<https://github.com/amrisi/amr-guidelines/blob/master/amr.md>

Frame Files … PropBank 3.4 should be used

<http://propbank.github.io/v3.4.0/frames/index.html>

**Convertor from AMR**:

Michal:

***relation*** … relation between two concepts (UMR nodes)

as, e.g., :ARG0 is understood as a relation between the predicate and its participant

However, in UMR guidelines, attributes and their values are also sometimes mentioned as relations

as, e.g., :polarity

🡪 consequences for TrEd

update and plans:

* work on PDT-to-UMR conversion, now it works :-)
* now work on vizualization in Tred (with Dan)
* plan: focus on coreference (as it leads to structural changes)
* plan: start with verbs
* work on functors conversion (incomplete list of functors from 2014), based on AMR guidelines (and on sample English annotation, where necessary) (Markéta:)

Problem repeated throughout the UMR (and AMR) guidelines – uneasy to extract **an exhaustive list** of relations (between nodes/concepts) and **full list of values for individual attributes**.

**coreference**

Míša: 3 different ways in UMR

* inversed participant roles
* reference to an already specified concept variable (only within a single sentence)
* :coref attribute (definitely inter-sentential relations)

Suggestion: Use the :coref attribute for two (or more) mentions of the same entity within a single sentence only when necessary (i.e., when some of their values differ)

**ellipses**

Estonsko:

Dan suggests to annotate such fragments using the abstract predicate have-location-91

(h/ have-location-91

:ARG1 (e/ event)

:ARG2 (c/ country

:wiki "Q191"

:name (n/ name :op1 "Estonsko")))

How should coreference look like? All events in the article annotated as subset-of event?

**quantities**

:quant

* typically an attribute with a numeric value
* but sometimes understood as a relation … approximate cardinalities (*more than, most*),

How to annotate percentage?

How to annotate **indefinite quantity**?

[la] *paululum commoratus* 'having waited for a while'

**English concepts in Czech annotation?**

where to use Czech concept words and where to stick to English ones

e.g., more-than

* abstract events … OK (distinguished by the \_91 suffix)
* named entities … governing organization (and as a node Ministerstvo … zemědělství)

?? Where is the borderline – where else we want to use Eng

* more than 3 annotated as the more-than relation … and for Czech?
* the abstract concept percentage-entity with the numeric attribute :value

**Lattices for individual attributes?**

Make use of values on different levels?

e.g., for number, we want to use *singular*, *dual* (***only when annotated as such in PDT***) and *plural*, even if they appear on different levels of the lattice (and plural in not a leaf node there)

The singular/plural values seem to be in compliance with English examples.

<https://github.com/umr4nlp/umr-guidelines/blob/master/Guidelines_figures/Number%20Lattice.jpg>

**UMR validator**

We will definitely need an UMR validator!

# Meeting 2023-07-17 (Federica, Cris, Zdeňka, Eva F., JH, Dan, Šárka, Michal)

**UMR Writer:**

Summary from the last meetings: the interface does not seem to be developed enough yet, for now, we are not going to use it.

JH: somebody (everybody) in our group should be able to work with it. The American group will develop the tool later and we yet need to stay in contact.

JH: UMR-writer 2.0 – [Sijia Ge](https://aclanthology.org/people/s/sijia-ge/) from ACL meeting (Boulder) is going to develop it. It should be better. This guy is from China, employed just til December 2023. JH has his e-mail address.

Paper on UMR writer 2.0. <https://aclanthology.org/2023.law-1.21/>

**Using TrEd**

JH: It would be good to work in TrEd. (edges in the graphs, based on the nodes; possibility to see the tree as a whole).

But we should be still able to edit it in UMR Writer

**Convertor from AMR**

Míša:

Three types how coreference is represented in UMR (described in the documentation on Github)

The original instructions from UMR about coreference are unclear.

JH: whenever we have a question, we should write to Julia – we need to set a frequent communication.

Dan: write it as an issue to Julia (official [UMR github](https://github.com/umr4nlp/umr-guidelines/issues), not our github)

**Events (processes) and entities**

Dan: Named events (named entities): should they have the abstract concept of entity?

Basically: event (ARG0, ARG1…), and then adding attributes which are usual for entities (name, link to Wikipedia)

Dan: **Basic question: why do we need to discern states and entities? Why are states not understood as events (with copulas)?**

JH: AMR classic problem: entity vs. type. It is clear with named entities (the type does not show up). IBM – name, type: company-entity.

JH: UMR is about the structure of the graph. They talk about the types only.

ŠZ: What is the difference between a concept and a type?

JH: more or less, they are the same. The most important difference is between specific cases (my running at Marathon) and concepts (running itself)

JH: *a clever man – he is clever*:

* In the ideal representation by JH it would be the same (predicate, eventive concept)
* In the UMR representation it would be different (modification – event / predicate)

Dan: still, what does it change in the annotation? Ask by Julia?

**Inferences**

JH: we want to make immediate inferences (if you listen to me, what you remember that I said), not more

**How far should our annotation be based on the word formation?**

Cf. CZ *jídlo* (food, from *jíst,* annotated as a part of the verb *jíst – to eat*) and EN *food* (not connected with *eat*) – different annotation?

Zdeňka: **How far should we match UMR?**

JH: within “UMR Light-Contact project” we should be as close as possible.

In other projects, we need UMR as a referential annotation – our next annotation can show what is wrong there and make it better.

For now, do it in the spirit of the guidelines.

Zdeňka: could we join internal Bolder etc. meeting?

JH: Julia said there are some files annotated for English. Could we get an access to some file? – JH can ask Julia.

Dan: Solution: until further notice, we will do it this way…

JH: We are preparing an example file; the proper annotation will follow.

Dan: **How is SynsemClass connected with our annotation?**

JH: Not yet, we will use rather PDT-Vallex for Czech.

**Next steps**

JH: try to annotate a short text

Federica: tried the annotation for Latin, comparing it to the Czech valency files on github (frames)

ŠZ: proposed points of the meetings:

* Presentation of a part of guidelines (as before)
* Presentation of a part of the annotation (Estonian file, Latin by Federica…)
* Discussion on the problems of the annotation

# Meeting 2023-07-10 (Markéta, Dan, Zdeňka, Federica, Cris, Šárka)

1. **Meeting with Julia – summary:**

* Dan: Julia is not using the UMR writer either
* Markéta: the structures disappear there
* Summary: we are not going to master the UMR writer, for the time being

1. **Eventive concepts (a document by Markéta) – comments on the document on Github**

* Markéta: not found any references to semantic type / information packaging
* Dan: information packaging – typological term, very general definition
* Markéta: there are books on information packaging by Bill Croft; Federica has the book Radical Construction Grammar
* The man who is tall = the tall man
* Please, go through the document and comment on the examples
* References about basic terms: Jackendoff, William Croft
* Zdeňka: we can ask Mirjam Fried about construction grammar
* Dan: concentrate on data, not so much on the theories

1. **Entities (a document by Dan) – comments on the document on Github**

* Dan: there are categorial named entities (name of a product – many things); what about the Estonians – nationality in UMR; Novákovi – family in UMR;
* Dan: proper noun – morphological term Dan is using
* When NE should be used?
  + Estonian
  + A general affiliation to a country rather than nationality etc.

If possible, please, go through the document and comment on the examples

* Is connection to a valency lexicon AND Wikipedia possible (like a specific war – a fight would be in a valency lexicon, the name of the war, as a named entity, would be in Wikipedia)

**In general:** we need to collect examples in Czech and other languages and ask Julia

Find a nice sentence with the WW2 (classification event / entity) and ask JH, Julia…