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

**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.

**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, 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 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

**Next meetings:**

24/7 13:00 (Markéta +, Dan +, Federica +, Cris +, Míša +, Eva F +, Zdeňka ? – zoom?, JH -, no zoom)

31/7 14:00 (Zdeňka+, JH +, Cris+, Markéta +?, Míša +, Eva +, Dan -, Šárka -, Federica –

21/8 – 13:00 (Federica – zoom, Cris +, Zdeňka +?, JH in China – zoom +?, Eva ?, Dan +, Míša +/zoom, Šárka ?)

28/8 – 13:00 (Zdeňka +, JH +, Cris -, Dan +, Federica +, Eva ?, Šárka ?)

**Meeting 10/7/2023 (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 VW2 (classification event / entity) and ask JH, Julia…