

LREC 2022: Uniform Meaning Representation, a Cross-lingual Annotation Framework for Document-level Semantics

Jens Van Gysel, Meagan Vigus, Jin Zhao, Nianwen Xue

January 14, 2022

1 Abstract of the tutorial

Uniform Meaning Representation (UMR) is a general purpose cross-lingual meaning representation that we have developed under an NSF-supported project entitled “Building a Uniform Meaning Representation for Natural Language Processing”. UMR is based on Abstract Meaning Representation (AMR), but substantially extends and adapts AMR in a number of ways. While AMR is a sentence-level representation that focuses on the predicate-argument structure, UMR is a document-level representation that also captures semantic relations that go beyond sentence boundaries. At the sentence level, UMR adds *aspect*, *person* and *number*, as well as *quantifier scope*. At the document level, UMR adds temporal and modal dependencies as well as reference relations. UMR also adapts AMR to make it cross-linguistically applicable, particularly to morphologically complex low-resource languages. UMR achieves this by defining a set of language-independent participant roles for languages that do not have a lexicon of PropBank-style predicate-specific semantic roles, and by arranging UMR semantic concepts and relations hierarchically in lattices to accommodate variability across languages, so that UMR developers for individual languages can select a level of granularity that is most appropriate for a language based on its grammaticalization patterns.

2 Outline of the tutorial

- Overview of UMR
 - Background: Abstract Meaning Representation
 - Overview of new sentence-level UMR additions to AMR: aspect, person, number, and quantification scope
 - Overview of UMR document-level representation: temporal and modal dependencies, coreference
 - Multi-word concepts and multi-concept words in UMR
 - How UMR addresses uniformity vs variability among languages
 - UMR-Writer
- UMR sentence-level representation
 - Typologically based representation of meanings in various semantic domains to ensure the structure of UMR is not biased towards English / high-resource languages
 - * Concept identification and concept-word mismatches – multi-word expressions vs. multi-concept words
 - Event identification

- Abstract, cross-linguistically comparable concepts for non-verbal predication
- Mapping of morphologically complex words as a whole to multiple concepts, i.e. without requiring morphological decomposition
 - ⇒ verbal argument indexation
 - ⇒ noun incorporation
 - ⇒ valency-changing operations
 - ⇒ TAM auxiliaries and morphology
 - ⇒ associated motion
- Participant roles
 - * Roadmap stages:
 - PropBank-style frame files as final stage of annotation
 - General semantic roles to make the annotation scheme much more accessible for low-resource languages
 - * Typological justification for the type and definition of the semantic roles selected
- Organization of annotation values in many semantic domains in lattices
 - * aspect
 - * modality
 - * person-number
 - * modification
 - * discourse relations
- UMR Document-level structure
 - Modal dependency:
 - * Justification for representation as a dependency structure compared to a single modal value for each event
 - * Specifics of the dependency structure (conceivers, events, epistemic strength links)
 - * Examples of complicated modal relations (conditionals, nested modals, evidentials)
 - * Lattice of epistemic strength values based on Boye’s work
 - Temporal dependency
- Hands-on tutorial on UMR-Writer
 - Getting to know the interface of UMR-Writer, including importing resources, annotation interface and exporting annotations.
 - * Making UMR appeal to field linguists working on low- or no-resource languages by allowing them to re-use existing materials
 - Frequently used data formats for annotated texts – texts exported from FLEx and ELAN with morphological segmentation, glossing, and free translations
 - Import lexicons created in FLEx / Toolbox and enhance them with argument structure information during annotation
 - Allow users to start a lexicon from scratch if they so desire, but make sure export can be imported into FLEx
 - Live demo of annotating sentence-level UMR graphs, including annotating lexicalized or abstract concepts, semantic relations, and UMR attributes, token-concept alignment feature.
 - Live demo of annotating document-level UMR graphs, including annotating temporal, modal and coreference annotation.
 - Live demo of cross-lingual annotation examples, including lexicon-building feature.