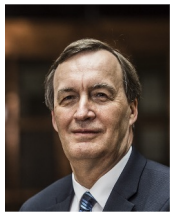
Meaning Representations for Natural Languages Tutorial Part 2

Common Meaning Representations

Julia Bonn, Jeffrey Flanigan, **Jan Hajič**, Ishan Jindal, and **Nianwen Xue**



Representation Roadmap

Meaning Representations for Natural Languages Tutorial Part 2

Common Meaning Representations

- AMR Format & Basics
- Some Details & Design Decisions
- Practice - Walking through a few AMRs
- Multi-sentence AMRs
- Relation to Other Formalisms
- **UMR**
- Open Questions in Representation

Outline

- Background
  - Do we need a new meaning representation? What's wrong with existing meaning representations?
- Aspects of Uniform Meaning Representation (UMR)
  - UMR starts with AMR but made a number of enrichments
  - UMR is a document-level meaning representation that represents temporal dependencies, modal dependencies, and coreference
  - UMR is a cross-lingual meaning representation that separates aspects of meaning that are shared across languages *language-independent* from those that are idiosyncratic to individual languages (*language-specific*)
- UMR-Writer -- a tool for annotating UMRs

Why aren't existing meaning representations sufficient?

- Existing meaning representations vary a great deal in their focus and perspective
  - Formal semantic representations aimed at supporting *logical inference* focus on the proper representation of quantification, negation, tense, and modality (e.g., Minimal Recursion Semantics (MRS) and Discourse Representation Theory (DRT).
  - Lexical semantic representations focus on the proper representation of core predicate-argument structures, word sense, named entities and relations between them, coreference (e.g., Tectogrammatical Representation (TR), AMR).
- The semantic ontology they use also differ a great deal. For example, MRS doesn't have a classification of named entities at all, while AMR has over 100 types of named entities

UMR uses AMR as a starting point

- Our starting point is AMR, which has a number of attractive properties:
  - Easy to read,
  - scalable (can be directly annotated without relying on syntactic structures),
  - has information that is important to downstream applications (e.g., semantic roles, named entities and coreference),
  - represented in a well-defined mathematical structure (a single-rooted, directed, acylical graph)
- Our general strategy is to augment AMR with meaning components that are missing and adapt it to cross-lingual settings

Participants of the UMR project

- UMR stands for Uniform Meaning Representation, and it is an NSF funded collaborative project between Brandeis University, University of Colorado, and University of New Mexico, with a number of partners outside these institions

Faculty

Students

Partners

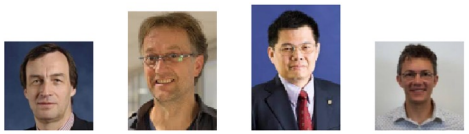


Table: The UMR team

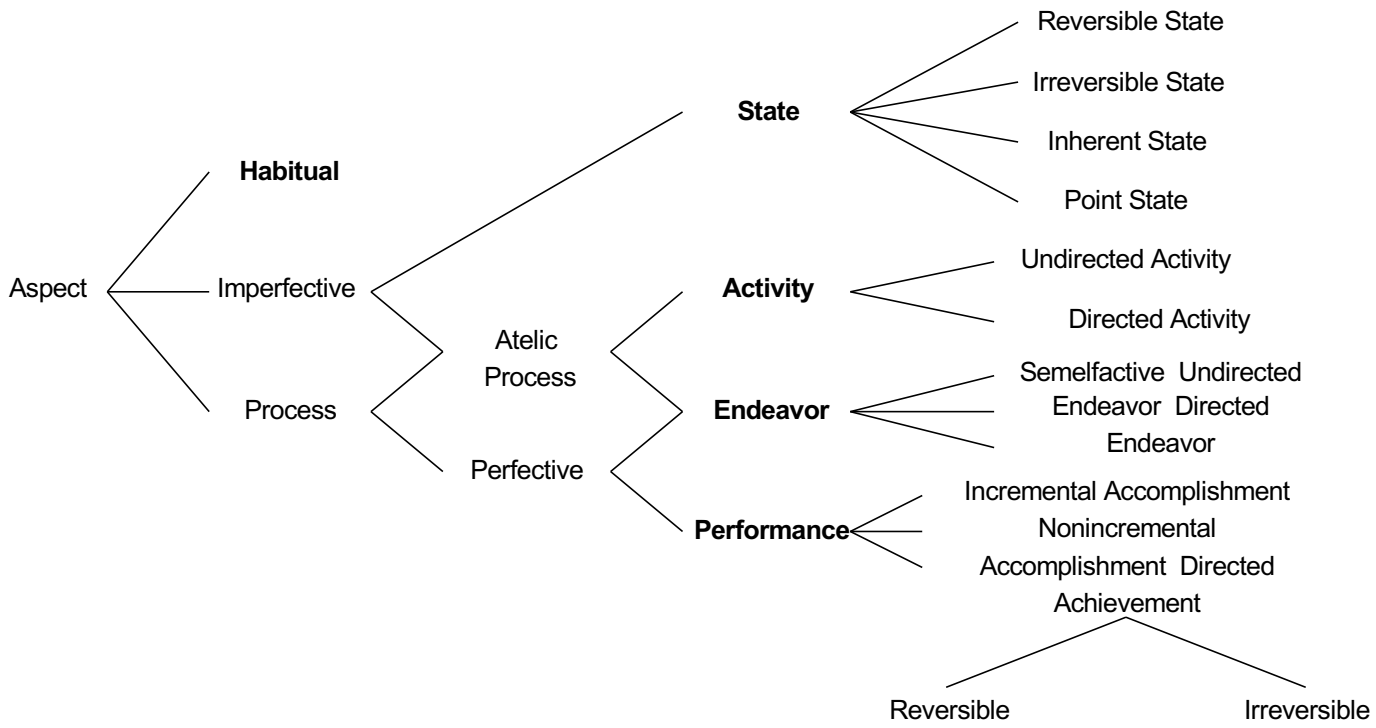
From AMR to UMR Gysel et al. (2021)

- At the sentence level, UMR adds:
  - An *aspect* attribute to eventive concepts
  - *Person* and *number* attributes for pronouns and other nominal expressions
  - Quantification scope between quantified expressions
- At the document level UMR adds:
  - Temporal dependencies in lieu of tense
  - Modal dependencies in lieu of modality
  - Coreference relations beyond sentence boundaries
- To make UMR cross-linguistically applicable, UMR
  - defines a set of language-independent abstract concepts and participant roles,
  - uses lattices to accommodate linguistic variability
  - designs specifications for complicated mappings between words and UMR concepts.

UMR sentence-level additions

- An *Aspect* attribute to event concepts
  - *Aspect* refers to the internal constituency of events - their temporal and qualitative boundedness
- *Person* and *number* attributes for pronouns and other nominal expressions
- A set of concepts and relations for discourse relations between clauses
- Quantification scope between quantified expressions to facilitate translation of UMR to logical expressions

UMR attribute: aspect



UMR attribute: coarse-grained aspect

- *State*: unspecified type of state
- *Habitual*: an event that occurs regularly in the past or present, including generic statements
- *Activity*: an event that has not necessarily ended and may be ongoing at Document Creation Time (DCT).
- *Endeavor*: a process that ends without reaching completion (i.e., termination)
- *Performance*: a process that reaches a completed result state

Coarse-grained Aspect as an UMR attribute

He wants to travel to Albuquerque.

(w / want
    :aspect State)

She rides her bike to work.

(r / ride
    :aspect Habitual)

He was writing his paper yesterday.

(w / write
    :aspect Activity)

Mary mowed the lawn for thirty minutes.

(m / mow
    :aspect Endeavor)

Fine-grained Aspect as an UMR attribute

My cat is hungry.

(h / have-mod-91
:aspect Reversible state)

The wine glass is shattered.

(h / have-mod-91
:aspect Irreversible state)

My cat is black and white.

(h / have-mod-91
:aspect Inherent state)

It is 2:30pm.

(h / have-mod-91
:aspect Point state)

AMR vs UMR on how pronouns are represented

- In AMR, pronouns are treated as unanalyzable concepts
- However, pronouns differ from language to language, so UMR decomposes them into person and number attributes
- These attributes can be applied to nominal expressions too

AMR:

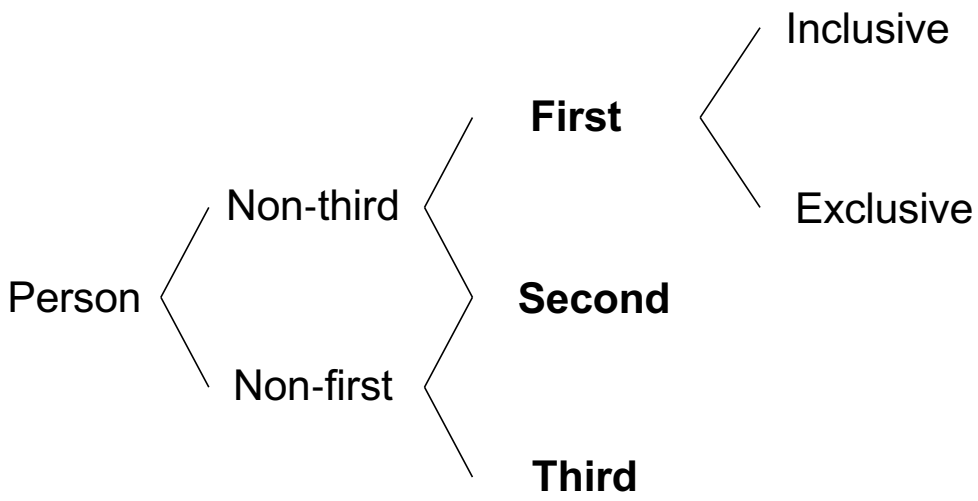
```
(s / see-01
    :ARG0 (h/ he)
    :ARG1 (b/ bird
              :mod (r/ rare)))
```

"He saw rare birds today."

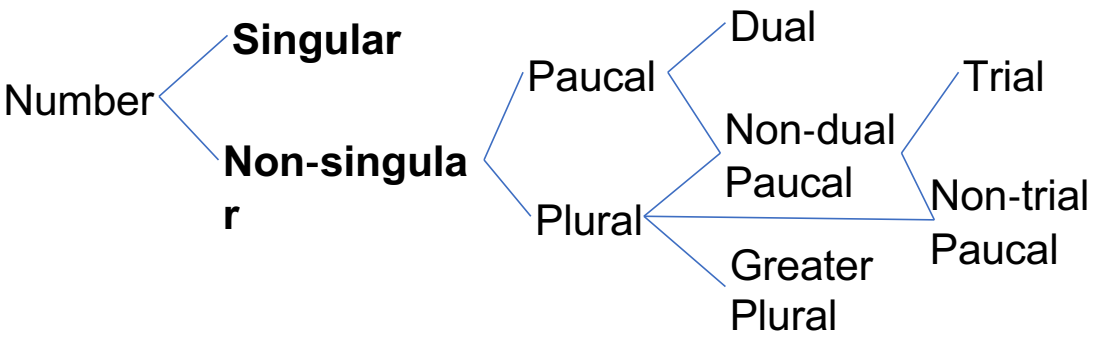
UMR:

```
(s / see-01
    :ARG0 (p / person
              :ref-person 3rd
              :ref-number Sing.)
    :ARG1 (b / bird
              :mod (r/ rare)
              :ref-number Plural))
```

UMR attributes: Person



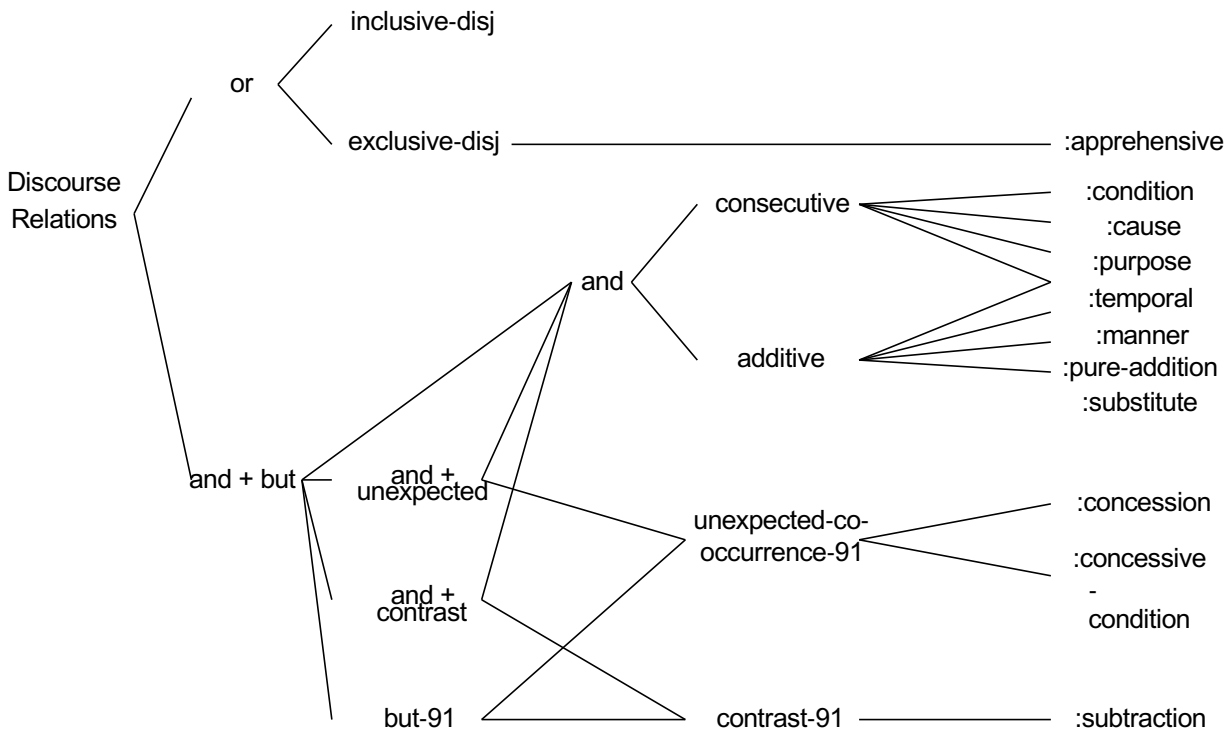
UMR attributes: number



Discourse relations in UMR

- In AMR, there is a minimal system for indicating relationships between clauses - specifically coordination:
  - *and* concept and *:opX* relations for addition
  - *or*/*either*/*neither* concepts and *:opX* relations for disjunction
  - *contrast-01* and its participant roles for contrast
- Many subordinated relationships are represented through participant roles, e.g.:
  - :manner
  - :purpose
  - :condition
- UMR makes explicit the semantic relations between (more general) “coordination” semantics and (more specific) “subordination” semantics

Discourse relations in UMR



Disambiguation of quantification scope in UMR

"Someone didn't answer all the questions"

```
(a / answer-01
   :ARG0 (p / person)
   :ARG1 (q / question :quant All :polarity -)
   :pred-of (s / scope :ARG0 p :ARG1 q))
```

$\exists p(\text{person}(p) \wedge \neg \forall q(\text{question}(q) \rightarrow$
$\exists a(\text{answer-01}(a) \wedge \text{ARG1}(a, q) \wedge \text{ARG0}(a, p))))$

Quantification scope annotation

- Scope will not be annotated for summation readings, nor is it annotated where a distributive or collective reading can be predictably derived from the lexical semantics.
  - The linguistics students ran 5 kilometers to raise money for charity (distributive).
  - The linguistics students carried a piano into the theater. (collective)
  - Ten hurricanes hit six states over the weekend. (summative)
- The scope annotation only comes into play when some overt linguistic element forces an interpretation that diverges from the lexical default
  - The linguistics students together ran 200 kilometers to raise money for charity.
  - The bodybuilders each carried a piano into the theater.
  - Ten hurricanes each hit six states over the weekend.

From AMR to UMR Gysel et al. (2021)

- At the sentence level, UMR adds:
  - An *aspect* attribute to eventive concepts
  - *Person* and *number* attributes for pronouns and other nominal expressions
  - Quantification scope between quantified expressions
- At the document level UMR adds:
  - Temporal dependencies in lieu of tense
  - Modal dependencies in lieu of modality
  - Coreference relations beyond sentence boundaries
- To make UMR cross-linguistically applicable, UMR
  - defines a set of language-independent abstract concepts and participant roles,
  - uses lattices to accommodate linguistic variability
  - designs specifications for complicated mappings between words and UMR concepts.

UMR is a document-level representation

- Temporal relations are added to UMR graphs as temporal dependencies
- Modal relations are also added to UMR graphs as modal dependencies
- Coreference is added to UMR graphs as identity or subset relations between named entities or events
- UMR favors relations over attributes where possible

UMR represents temporal relations in a document as temporal dependency structures (TDS)

- The temporal dependency structure annotation involves identifying the most specific reference time for each event
- Time expressions and other events are normally the most specific reference times
- In some cases, an event may require two reference times in order to make its temporal location as specific as possible

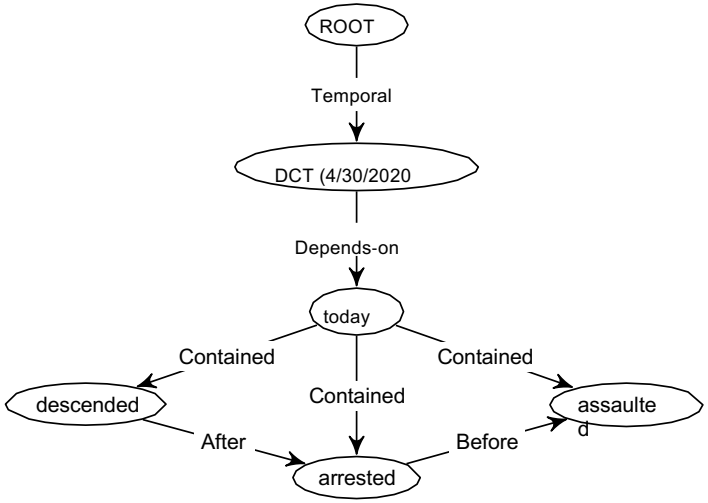
Zhang and Xue (2018); Yao et al. (2020)

TDS Annotation

- If an event is not clearly linked temporally to either a time expression or another event, then it can be linked to the DCT or tense metanodes
  - Tense metanodes capture vague stretches of time that correspond to grammatical tense
    - Past_Ref, Present_Ref, Future_Ref
  - DCT is a more specific reference time than a tense metanode

Temporal dependency Structure (TDS)

- If we identify a reference time for every event and time expression in a document, the result will be a Temporal Dependency Graph.



"700 people descended on the state Capitol today, according to Michigan State Police. State Police made one arrest, where one protester had assaulted another, Lt. Brian Oleksyk said."

Genre in TDS Annotation

- Temporal relations function differently depending on the genre of the text (e.g., Smith 2003)
- Certain genres proceed in temporal sequence from one clause to the next
- While other genres involve generally non-sequenced events
- News stories are a special type
  - many events are temporally sequenced
  - temporal sequence does not match with sequencing in the text

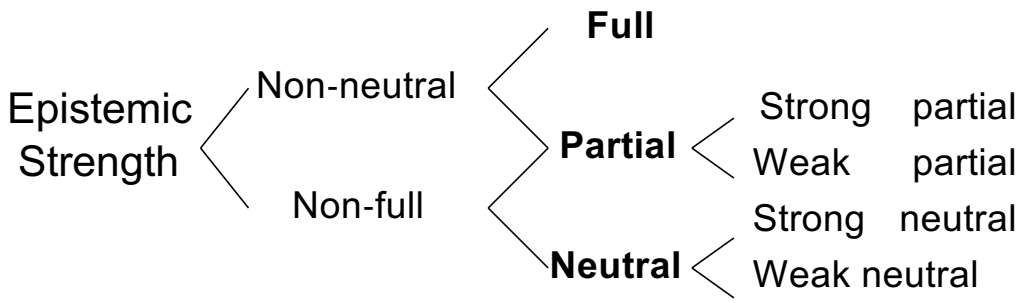
Modality in AMR

- Modality characterizes the reality status of events, without which the meaning representation of a text is incomplete
- AMR has six concepts that represent modality:
  - *possible-01*, e.g., “The boy can go.”
  - *obligate-01*, e.g., “The boy must go.”
  - *permit-01*, e.g., “The boy may go.”
  - *recommend-01*, e.g., “The boy should go.”
  - *likely-01*, e.g., “The boy is likely to go.”
  - *prefer-01*, e.g., “They boy would rather go.”
- Modality in AMR is represented as senses of an English verb or adjective.
- However, the same exact concepts for modality may not apply to other languages

Modal dependency structure

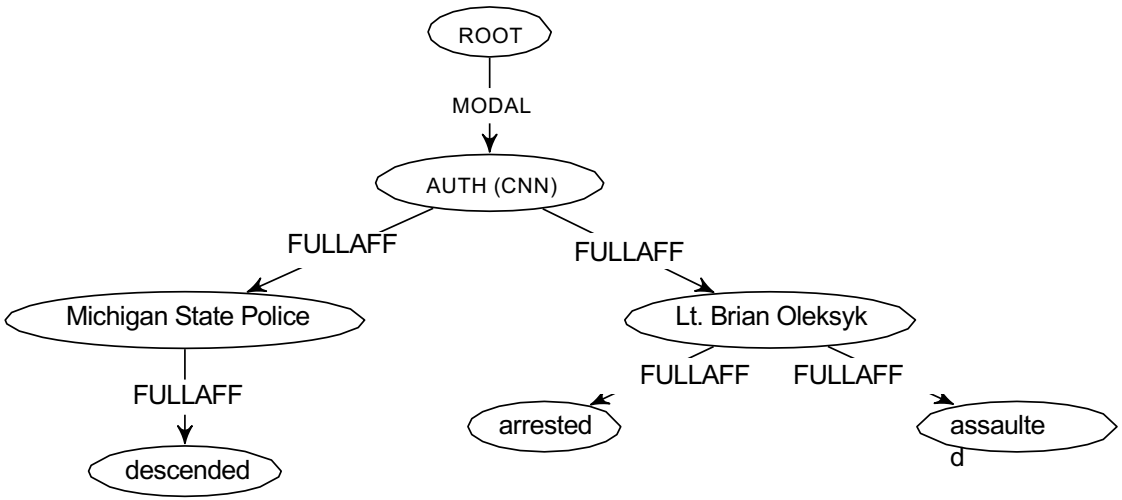
- There are two types of nodes in the modal dependency structure: **events** and **conceivers**
- Conceivers
  - Mental-level entities whose perspective is modelled in the text
  - Each text has an author node (or nodes)
  - All other conceivers are children of the AUTH node
  - Conceivers may be nested under other conceivers
- Mary said that Henry wants...

Epistemic strength lattice



Full: The dog barked.
Partial: The dog probably barked.
Neutral: The dog might have barked.

Modal dependency structure (MDS) (Vigus et al., 2019; Yao et al., 2021):



"700 people descended on the state Capitol today, according to Michigan State Police. State Police made one arrest, where one protester had assaulted another, Lt. Brian Oleksyk said."

Entity Coreference in UMR

- same-entity:
  1. Edmund Pope tasted freedom today for the first time in more than eight months.
  2. He denied any wrongdoing.
- subset:
  1. He is very possesive and controlling but he has no right to be as we are not together.

Event coreference in UMR

- same-event
  1. El-Shater and Malek's property was confiscated and is believed to be worth millions of dollars.
  2. Abdel-Maksoud stated the confiscation will affect the Brotherhood's financial bases.
- same-event
  1. The Three Gorges project on the Yangtze River has recently introduced the first foreign capital.
  2. The loan , a sum of 12.5 million US dollars , is an export credit provided to the Three Gorges project by the Canadian government , which will be used mainly for the management system of the Three Gorges project .
- subset:
  1. 1 arrest took place in the Netherlands and another in Germany.
  2. The arrests were ordered by anti-terrorism judge fragnoli.

An UMR example with coreference

He is controlling but he has no right to be as we are not together.

```
(s4c / but-91
        :ARG1 (s4c3 / control-01
                        :ARG0 (s4p2 / person
                                        :ref-person 3rd
                                        :ref-number Singular))
        :ARG2 (s4r / right-05
                :ARG1 s4p2
                :ARG1-of (s4c2 / cause-01
                                :ARG0 (s4h / have-mod-91
                                        :ARG0 (s4p3 / person
                                                        :ref-person 1st
                                                        :ref-number Plural)
                                        :ARG1 (s4t/ together)
                                        :aspect State
                                        :modstr FullNeg))
                :modstr FullNeg))
(s / sentence
   :coref ((s4p2 :subset-of s4p3)))
```

The challenge: Integration of different meaning components into one graph

- How do we represent all this information in a unified structure that is still easy to read and scalable?
- UMR pairs a **sentence-level** representation (a modified form of AMR) with a **document-level** representation.
- We assume that a text will still have to be processed sentence by sentence, so each sentence will have a fragment of the document-level super-structure.

Integrated UMR representation

1. Edmund Pope **tasted** freedom today for the first time in more than eight months.
2. Pope is the American businessman who was **convicted** last week on spying charges and **sentenced** to 20 years in a Russian prison.
3. He **denied** any wrongdoing.

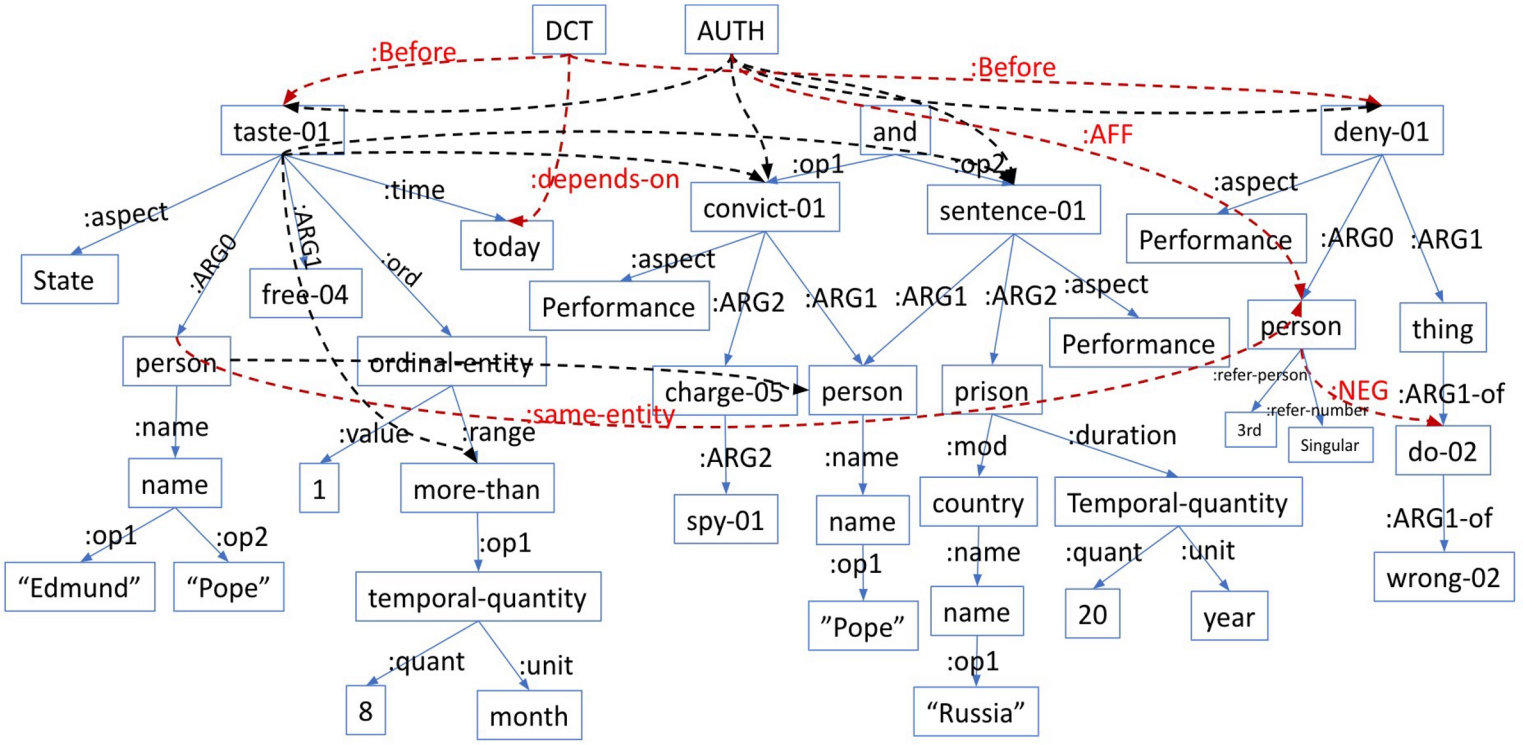
Sentence-level representation vs document-level representation

Edmund Pope **tasted** freedom today for the first time in more than eight months.

```
(s1t2 / taste-01
      :Aspect Performance
      :ARG0 (s1p / person
                  :name (s1n2 / name
                              :op1 “Edmund”
                              :op2 “Pope”))
      :ARG1 (s1f / free-04 :ARG1 s1p)
      :time (s1t3 / today)
      :ord (s1o3 / ordinal-entity
                 :value 1
                 :range (s1m / more-than
                             :op1 (s1t / temporal-quantity
                                       :quant 8
                                       :unit (s1m2 / month)))))
```

```
(s1 / sentence
    :temporal ((DCT :before s1t2)
    (s1t3 :contained s1t2)
    (DCT :depends-on s1t3))
    :modal ((ROOT :MODAL AUTH)
    (AUTH :FullAff s1t2)))
```

UMR graph



“Edmund Pope tasted freedom today for the first time in eight months.”

“Pope was convicted on spying charges and sentenced to 20 years in a Russian prison.”

”He denied any wrong-doing.”

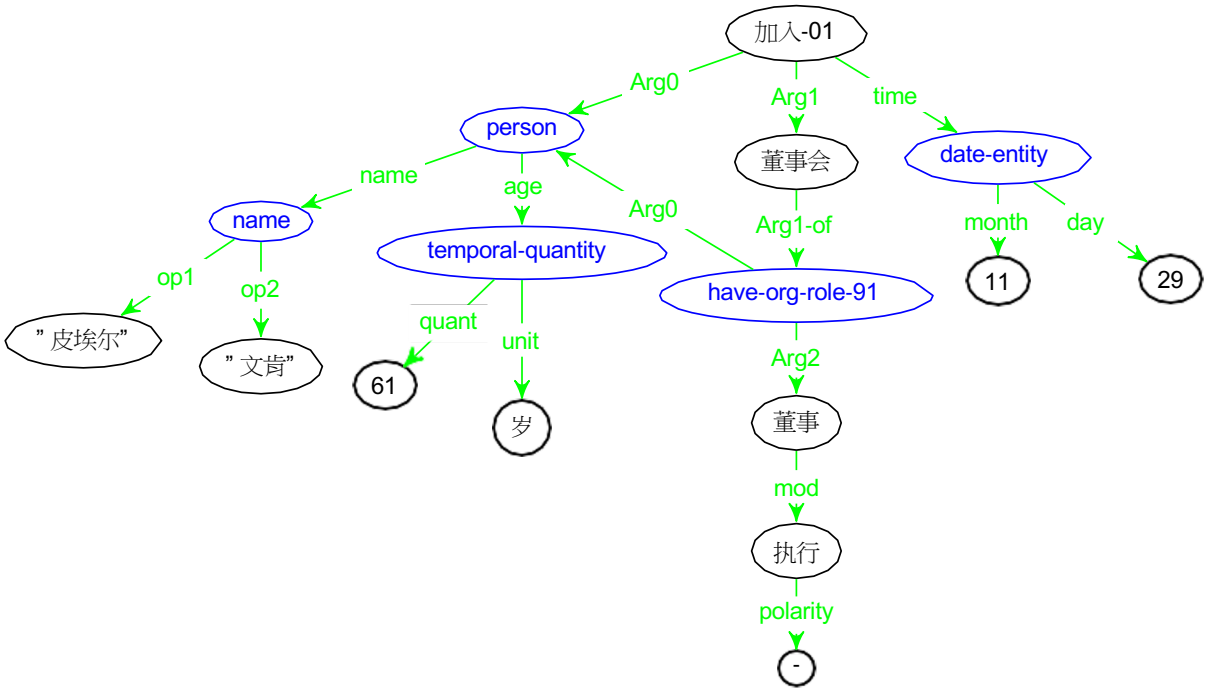
From AMR to UMR Gysel et al. (2021)

- At the sentence level, UMR adds:
  - An *aspect* attribute to eventive concepts
  - *Person* and *number* attributes for pronouns and other nominal expressions
  - Quantification scope between quantified expressions
- At the document level UMR adds:
  - Temporal dependencies in lieu of tense
  - Modal dependencies in lieu of modality
  - Coreference relations beyond sentence boundaries
- To make UMR cross-linguistically applicable, UMR
  - defines a set of language-independent abstract concepts and participant roles,
  - uses lattices to accommodate linguistic variability
  - designs specifications for complicated mappings between words and UMR concepts.

Elements of AMR are already cross-linguistically applicable

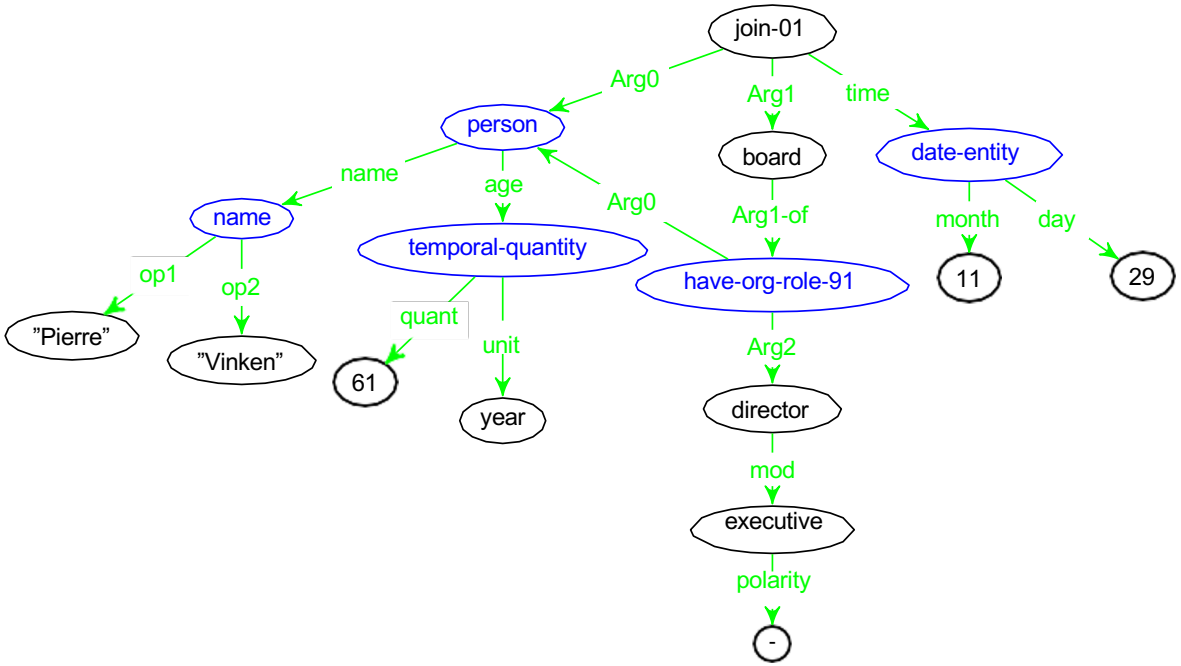
- Abstract concepts (e.g., *person, thing, have-org-role-91*):
  - Abstract concepts are concepts that do not have explicit lexical support but can be inferred from context
- Some semantic relations (e.g., :manner, :purpose, :time) are also cross-linguistically applicable

Language-independent vs language-specific aspects of AMR



“61 岁的 Pierre Vinken 将于 11 月 29 日加入董事会，担任 非执行董事。”

Language-independent vs language-specific aspects of AMR



"Pierre Vinken , 61 years old , will join the board as a nonexecutive director Nov. 29 ."

Abstract concepts in UMR

- Abstract concepts inherited from AMR:
  - Standardization of quantities, dates etc.: *have-name-91, have-frequency-91, have-quant-91, temporal-quantity, date-entity...*
- New concepts for abstract events: “non-verbal” predication.
- New concepts for abstract entities: entity types are annotated for named entities and implicit arguments.
- Scope: *scope* concept to disambiguate scope ambiguity to facilitate translation of UMR to logical expressions (see sentence-level structure).
- Discourse relations: concepts to capture sentence-internal discourse relations (see sentence-level structure).

Sample abstract events

Clause Type	UMR Predicates	Arg0	Arg1	Arg2
Thetic/presentational possession	have-91	possessor	possessum	
Predicative possession	belong-91	possessum	possessor	
Thetic/presentational location	exist-91	location	theme	
Predicative location	have-location-91	theme	location	
property-predication	have-mod-91		theme	property
Object predication	have-role-91	theme	Ref point	Object category
Equational	identity-91	theme	equated referent	

Language-independent vs language-specific participant roles

- Core participant roles are defined in a set of frame files (valency lexicon, see Palmer et al. 2005). The semantic roles for each sense of a predicate are defined:
  - E.g. boil-01: *apply heat to water*
    ARG0-PAG: *applier of heat* ARG1-PPT:
    water
- Most languages do not have frame files
  - But see e.g. Hindi (Bhat et al. 2014), Chinese (Xue 2006)
- UMR defines language-independent participant roles
  - Based on ValPaL data on co-expression patterns of different micro-roles (Hartmann et al., 2013)

Language-independent roles: an incomplete list

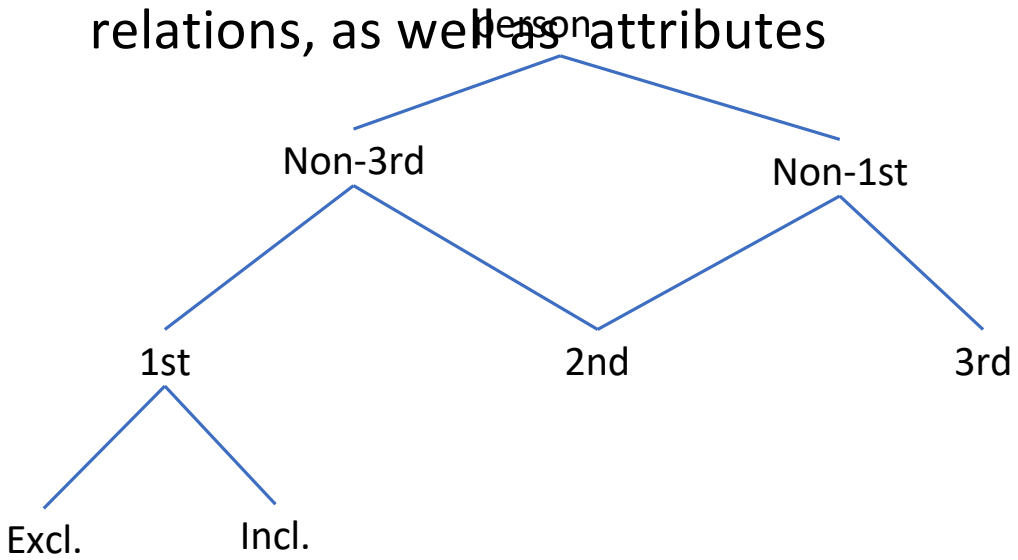
UMR Annotation	Definition
Actor	animate entity that initiates the action
Undergoer	entity (animate or inanimate) that is affected by the action
theme	entity (animate or inanimate) that moves from one entity to another entity, either spatially or metaphorically
Recipient	animate entity that gains possession (or at least temporary control) of another entity
force Causer	inanimate entity that initiates the action
causer	animate entity that acts on another animate entity to initiate the action
experiencer	animate entity that acts on another animate entity to initiate the action
stimulus	animate entity that cognitively or sensorily experiences a stimulus

How UMR accommodates cross-linguistic variability

- Not all languages grammaticalize/overtly express the same meaning contrasts:
  - English: *I* (1SG) vs. *you* (2SG) vs. *she/he* (3SG)
  - Sanapaná: *as-* (1SG) vs. *an-/ap-* (2/3SG)
- However, there are typological patterns in how semantic domains get subdivided:
  - A 1/3SG person category would be much more surprising than a 2/3SG one
- UMR uses lattices for abstract concepts, attribute values, and relations to accommodate variability across languages.
  - Languages with overt grammatical distinctions can choose to use more fine-grained categories

Lattices

- Semantic categories are organized in “lattices” to achieve cross-lingual compatibility while accommodating variability.
- We have lattices for abstract concepts, relations, as well as attributes



Wordhood vs concepthood across languages

- The mapping between words and concepts in languages is not one-to-one: UMR designs specifications for complicated mappings between words and concepts.
  - Multiple words can map to one concept (e.g., multi-word expressions)
  - One word can map to multiple concepts (morphological complexity)

Multiple words can map to a single (discontinuous) concept

(x0/帮忙-01
:aspect Performance
:arg0 (x1/地理学)
:affectee (x2/我)
:degree (x3/大))

地理学帮 了我很大的忙。
“Geography has helped me a lot”

(w / want-01
:Aspect State
:ARG0 (p / person)
:ref-person 3rd
:ref-number Singular
:ARG1 (g / give-up-07
:ARG0 h
:ARG1 (t / that)
:aspect Performance
:modpred w)
:ARG1-of (c / cause-01
:ARG0 (a / umr-unknown))
:aspect State)
“Why would he want to ~~give~~ that ~~up~~?”

One word maps to multiple UMR concepts

- One word containing predicate and arguments

Arapaho:

he'ih'iixooxookbixoh'oekoohuutoono' he'ih'ii-xoo-xook-
bixoh'oekoohuutoo-no'
NARR.PST.IPFV-REDUP-through-make.hand.appear.quickly-PL
``They were sticking their hands right through them [the ghosts] to the other side.''

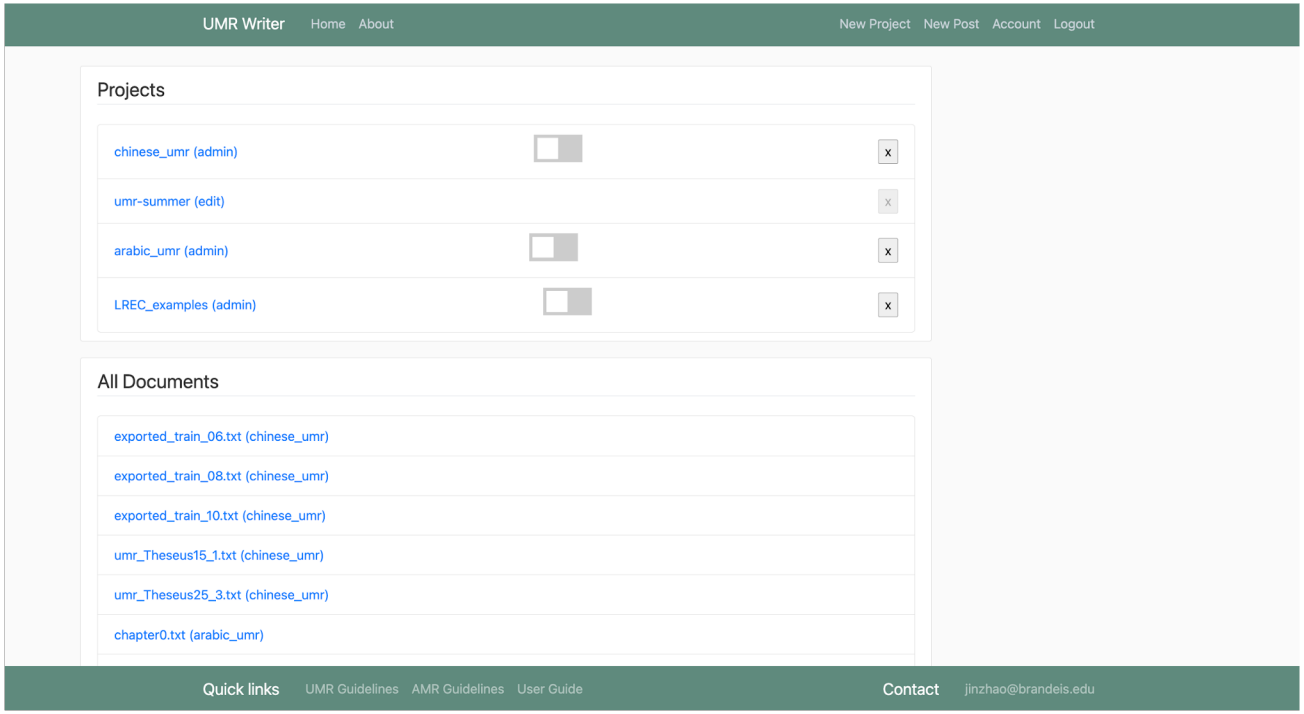
(b/ bixoh'oekoohuutoo `stick hands through'
    :actor (p/ person :ref-person 3rd :ref-number Plural)
    :theme (h/ hands)
    :undergoer (g/ [ghosts])
    :aspect Endeavor
    :modstr FullAff)

- Noun Incorporation (less grammaticalized): identify predicate and argument concept

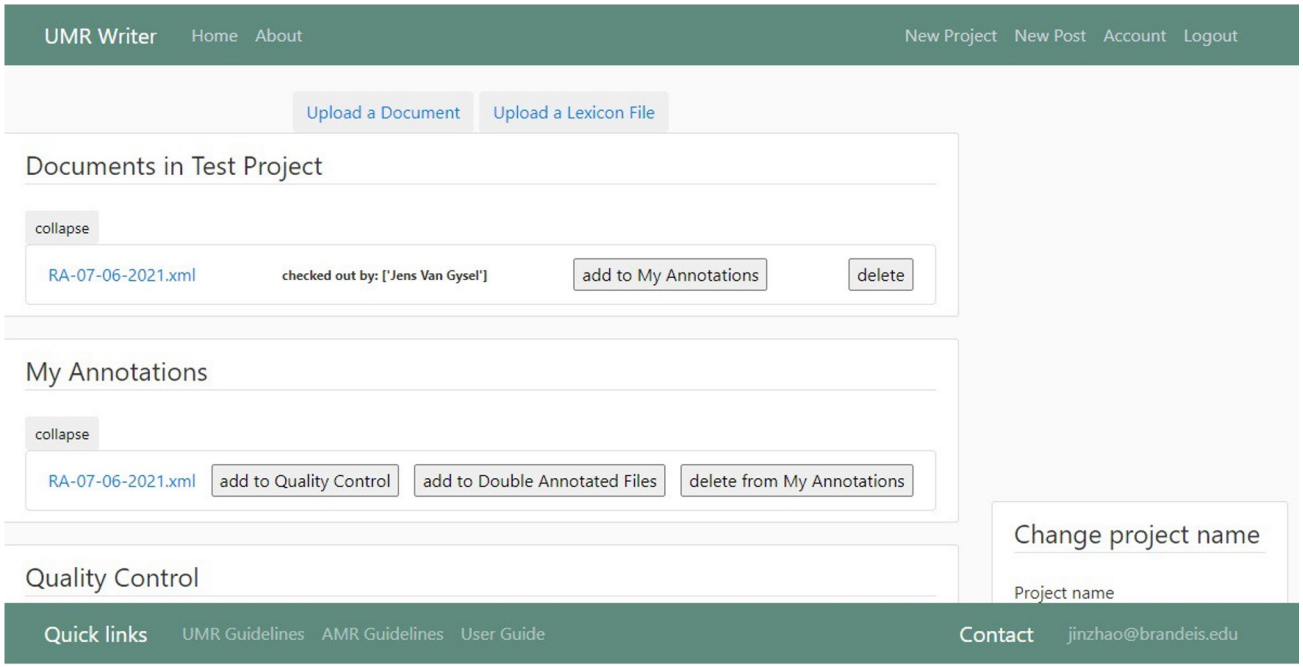
UMR-Writer

- The annotation interface we use for UMR annotation is called UMR-Writer
- UMR-Writer includes interfaces for project management, sentence-level and document-level annotation, as well as lexicon (frame file) creation.
- UMR-Writer has both keyboard-based and click-based interfaces to accommodate the annotation habits of different anntotators.
- UMR-Writer is web-based and supports UMR annotation for a variety of languages and formats. So far it supports Arabic, Arapaho, Chinese, English,Kukama Navajo, and Sanapana. It can easily extended to more languages.

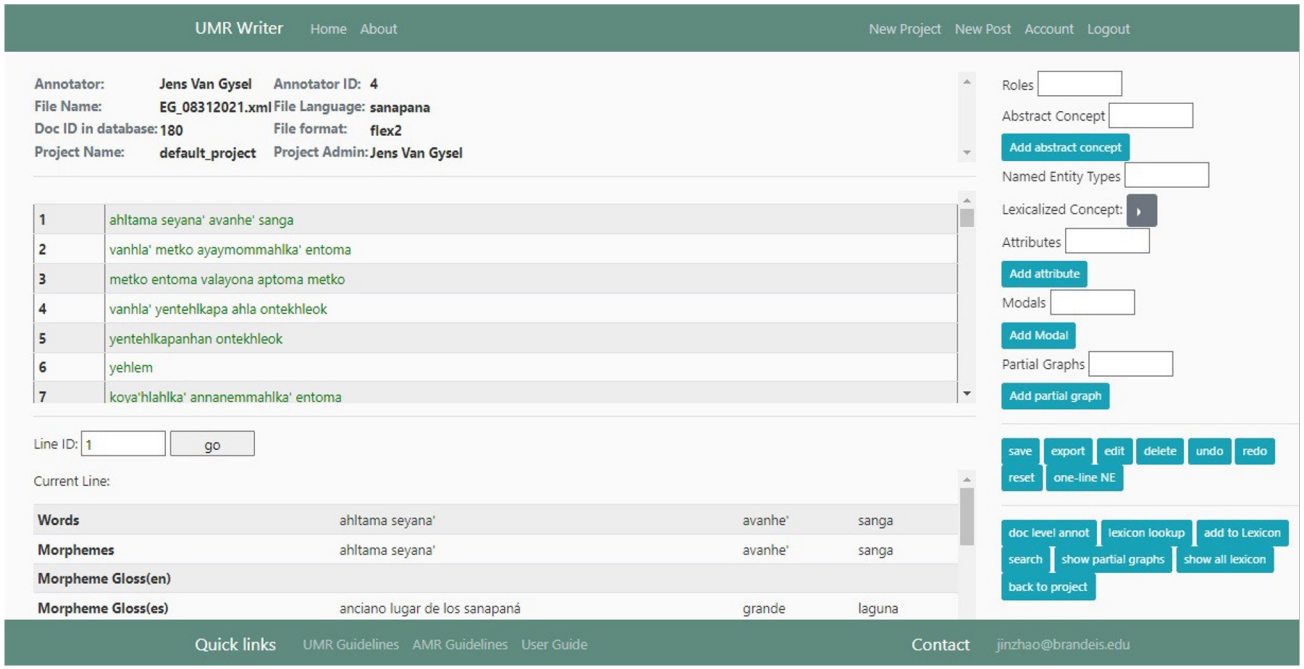
UMR writer: Project management



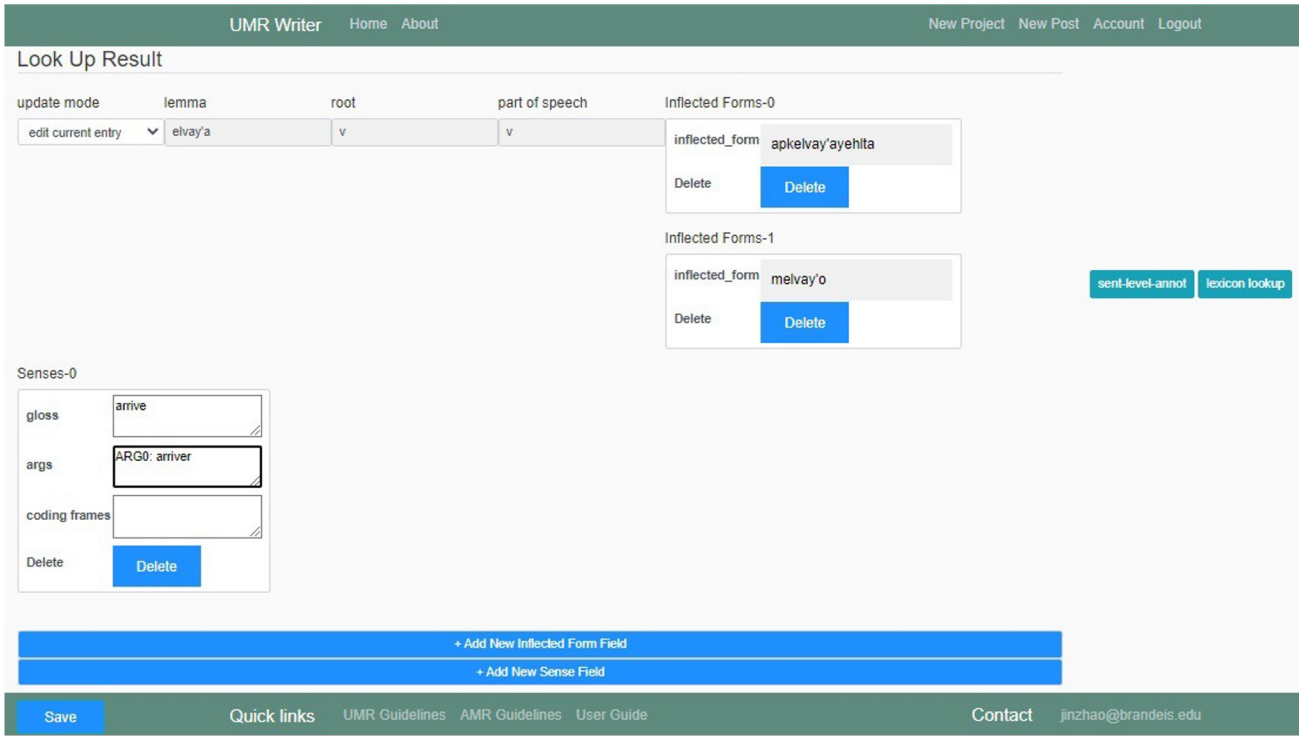
UMR writer: Project management



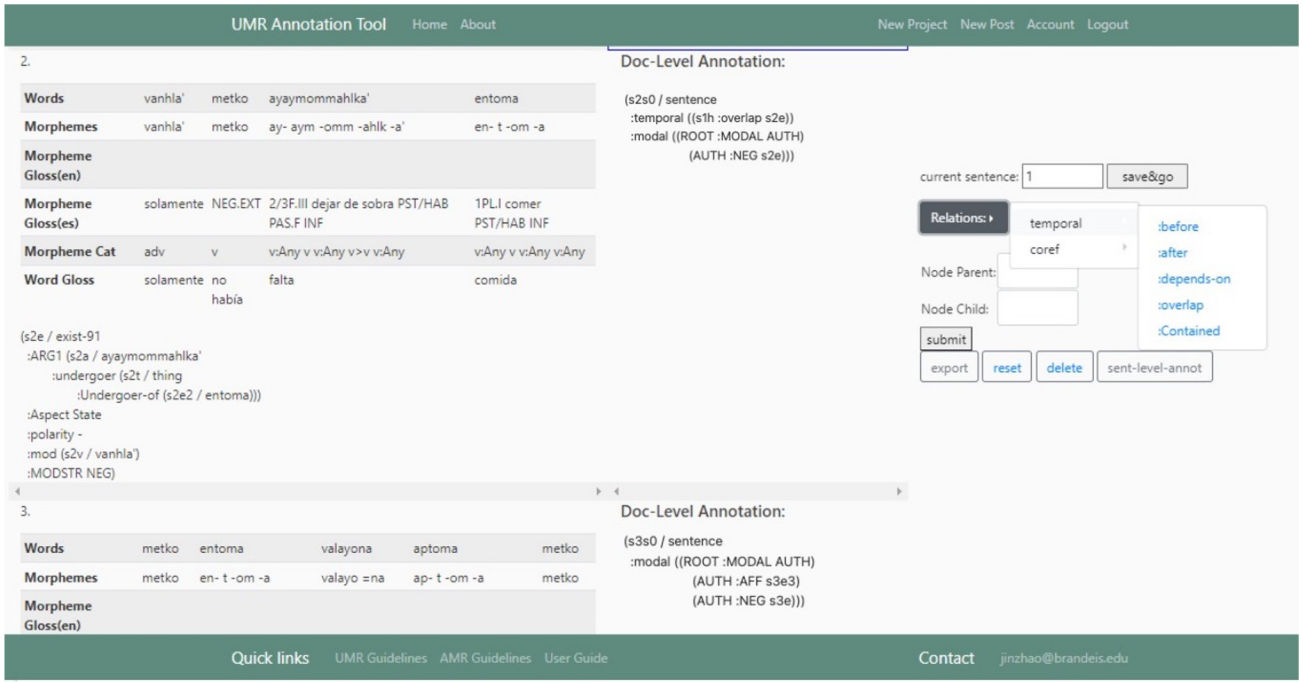
UMR writer: Sentence-level interface



UMR writer: Lexicon interface



UMR Writer: Document-level interface



UMR summary

- UMR is a rooted directed node-labeled and edge-labeled document-level graph.
- UMR is a document-level meaning representation that builds on sentence-level meaning representations
- UMR aims to achieve semantic stability across syntactic variations and support logical inference
- UMR is a cross-lingual meaning representation that separates language-general aspects of meaning from those that are language-specific
- We are doing UMR English, Chinese, Arabic, Arapaho, Kukama, Sanapana, Navajo, Quechua

Use cases of UMR

- Temporal reasoning
  - UMR can be used to extract temporal dependencies, which can then be used to perform temporal reasoning
- Knowledge extraction
  - UMR annotates aspect, and this can be used to extract habitual events or state, which are typical knowledge forms
- Factuality determination
  - UMR annotates modal dependencies, and this can be used to verify the factuality of events or claims
- As intermediate representation for dialogue systems where control is more needed.
  - UMR annotates entities and coreferences, which helps tracking dialogue states

Planned UMR activities

- The DMR international workshops. The 5th DMR workshop planned in 2024, in conjunction with one of the major NLP/CL conferencess
- UMR summer schools, tentatively in 2024 and 2025.
- UMR shared tasks once we have sufficient amount of UMR-annotated data as well as evaluation metrics and baseline parsing models

UMR 1.0

released via

https://umr4nlp.github.io/web/

References

Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., and Schneider, N. (2013). Abstract meaning representation for sembanking. In *Proceedings of the 7th linguistic annotation workshop and interoperability with discourse*, pages 178–186.

Hartmann, I., Haspelmath, M., and Taylor, B., editors (2013). *The Valency Patterns Leipzig online database*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Van Gysel, J. E. L., Vigus, M., Chun, J., Lai, K., Moeller, S., Yao, J., O'Gorman, T. J., Cowell, A., Croft, W. B., Huang, C. R., Hajic, J., Martin, J. H., Oepen, S., Palmer, M., Pustejovsky, J., Vallejos, R., and Xue, N. (2021). Designing a uniform meaning representation for natural language processing. *Künstliche Intelligenz*, pages 1–18.

Vigus, M., Van Gysel, J. E., and Croft, W. (2019). A dependency structure annotation for modality. In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 182–198.

Yao, J., Qiu, H., Min, B., and Xue, N. (2020). Annotating temporal dependency graphs via crowdsourcing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5368–5380.

Yao, J., Qiu, H., Zhao, J., Min, B., and Xue, N. (2021). Factuality assessment as modal dependency parsing. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1540–1550.

Zhang, Y. and Xue, N. (2018). Structured interpretation of temporal relations. In *Proceedings of LREC 2018*.

Acknowledgements

We would like to acknowledge the support of National Science Foundation:

- NSF IIS (2018): “Building a Uniform Meaning Representation for Natural Language Processing” awarded to Brandeis (Xue, Pustejovsky), Colorado (M. Palmer, Martin, and Cowell) and UNM (Croft).

- NSF CCRI (2022): ``Building a Broad Infrastructure for Uniform Meaning Representations'', awarded to Brandeis (Xue, Pustejovsky) and Colorado (A. Palmer, M. Palmer, cowell, Martin), with Croft as consultant

All views expressed in this paper are those of the authors and do not necessarily represent the view of the National Science Foundation.