

Developing a Uniform Meaning Representation for Natural Language Processing

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

Based on joint work with Jayeol Chun, Kenneth Lai, Sara Moeller, Jiarui Yao, Tim O'Gorman, Andrew Cowell, William Croft, Chu-Ren Huang, Jan Hajič, James Martin, Stephan Oepen, Martha Palmer, James Pustejovsky, Rosa Vallejos

Outline

- ▶ Background
 - ▶ Do we need a new meaning representation?
 - ▶ What makes UMR desirable?
- ▶ Aspects of Uniform Meaning Representation (UMR)
 - ▶ Formal properties of UMR
 - ▶ Cross-lingual applicability - separating the *language-independent* from the *language-specific*
 - ▶ Document-level vs. sentence-level representation
 - ▶ UMR-Writer annotation tool
- ▶ Discussion:
 - ▶ Are symbolic representations of meaning still needed?
 - ▶ Use cases for UMR

Do we need a new meaning representation?

- ▶ Existing meaning representations vary a great deal in their focus and perspective
 - ▶ Formal semantic representations for *logical inference* (e.g. MRS, DRT) focus on the proper representation of:
 - ▶ quantification
 - ▶ negation
 - ▶ tense
 - ▶ modality
 - ▶ Lexical semantic representations (e.g. TR, AMR) focus on the proper representation of:
 - ▶ core predicate-argument structures
 - ▶ word senses
 - ▶ named entities
 - ▶ co-reference

Do we need a new meaning representation?

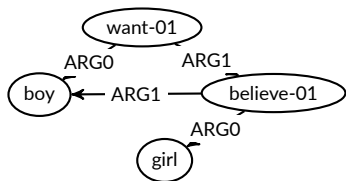
- ▶ Existing meaning representations vary a great deal in the “semantic vocabulary” they use:
 - ▶ One extreme: no classification of named entities at all (MRS)
 - ▶ Other extreme: over 100 types of named entities (AMR)
- ▶ Existing meaning representations are often developed based on English / high-resource languages
 - ▶ Their structures and workflows therefore pose challenges for the annotation of typologically different languages

In comes UMR

- ▶ UMR (Uniform Meaning Representation) is an NSF-funded collaborative project between Brandeis University, University of Colorado, and University of New Mexico
- ▶ Our starting point is AMR, which has a number of attractive properties:
 - ▶ Easy to read
 - ▶ Scalable (does not rely on syntactic structures)
 - ▶ Information that is important to downstream applications (e.g., semantic roles, named entities and coreference)
 - ▶ Well-defined mathematical structure (single-rooted, directed, acyclical graph)
- ▶ UMR augments AMR with meaning components that are missing and adapts it to cross-lingual settings

The starting point: AMR

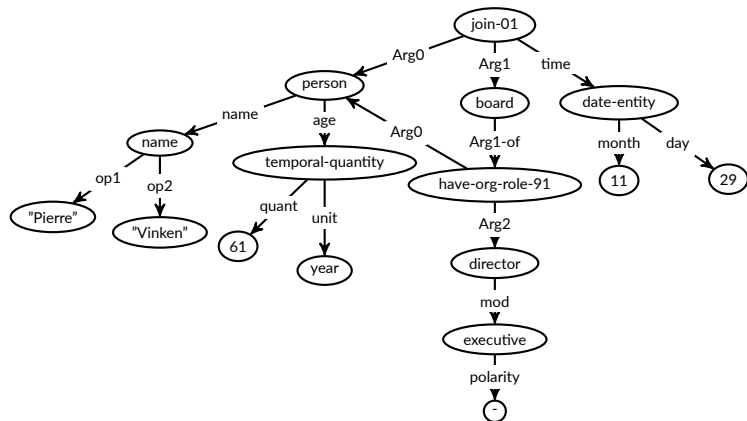
- ▶ Single-rooted, directed, acyclic graph
- ▶ Nodes are concepts (sense-disambiguated predicates, named entity types, plain lemmas)
- ▶ Edges are relations (participant roles, other semantic relations)



```
(w / want-01
:ARG0 (b / boy)
:ARG1 (b2 / believe-01
:ARG0 (g / girl)
:ARG1 b))
```

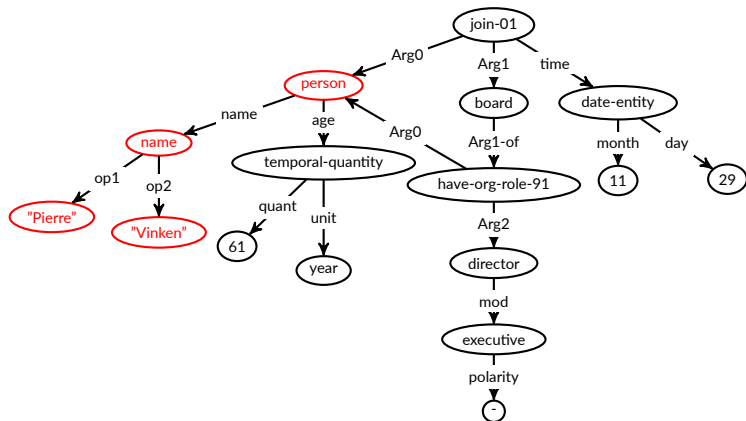
“The boy wants the girl to believe him.”
Banarescu et al. (2013)

The starting point: AMR



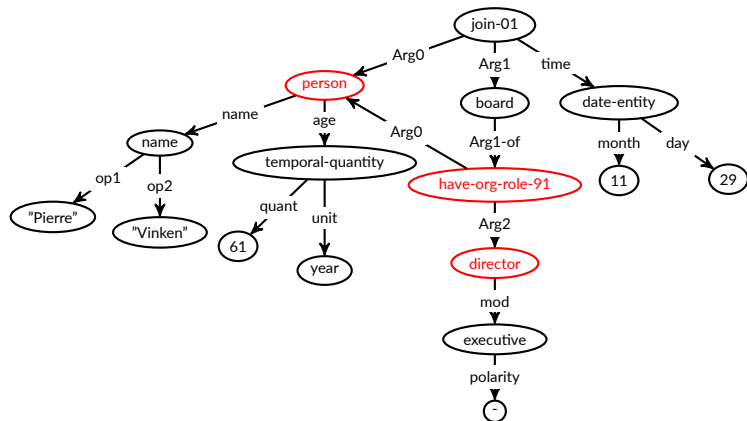
“Pierre Vinken , 61 years old , will join the board as a nonexecutive director Nov. 29 .”

AMR: Named Entities



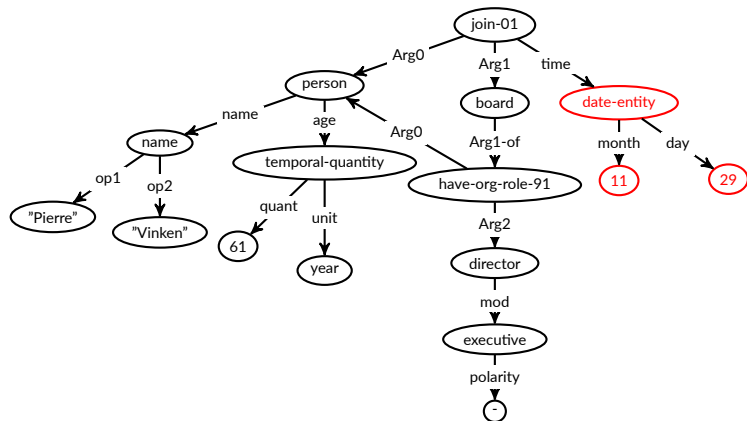
"**Pierre Vinken** , 61 years old , will join the board as a nonexecutive director Nov. 29 ."

AMR: Relations



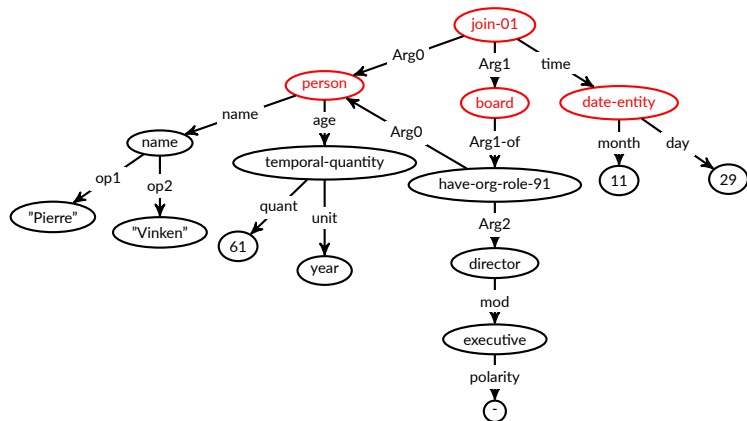
"Pierre Vinken , 61 years old , will join the board as a nonexecutive director Nov. 29 ."

AMR: Date entities



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

AMR: Word sense and semantic roles



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

From AMR to UMR (Van Gysel et al., 2021)

- ▶ To make UMR cross-linguistically applicable, it:
 - ▶ defines a set of language-independent abstract concepts and participant roles,
 - ▶ uses lattices to accommodate linguistic variability,
 - ▶ provides meaning-based guidelines for the identification of events,
 - ▶ designs specifications for complicated mappings between words and UMR concepts,
 - ▶ is organized as a road map so that languages at different stages of documentation and description can use UMR at an appropriate level of detail.

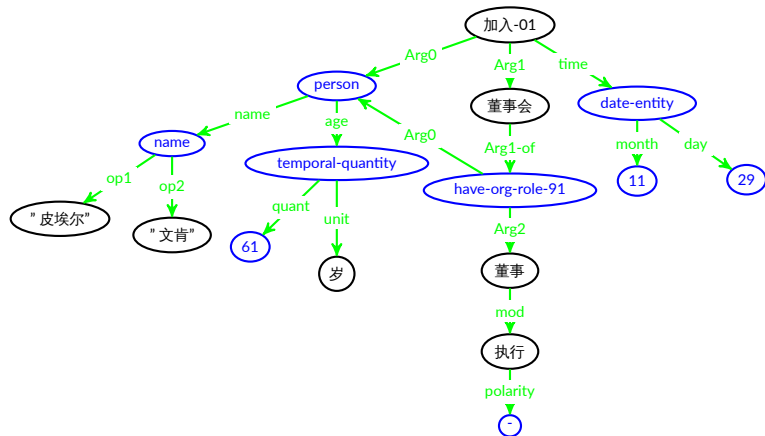
From AMR to UMR (Van Gysel et al., 2021)

- ▶ At the sentence level, UMR adds:
 - ▶ An *Aspect* attribute to eventive concepts
 - ▶ *Person* and *Number* attributes for pronouns and other nominal expressions
 - ▶ A principled set of discourse relations
 - ▶ 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

UMR is a cross-lingual meaning representation

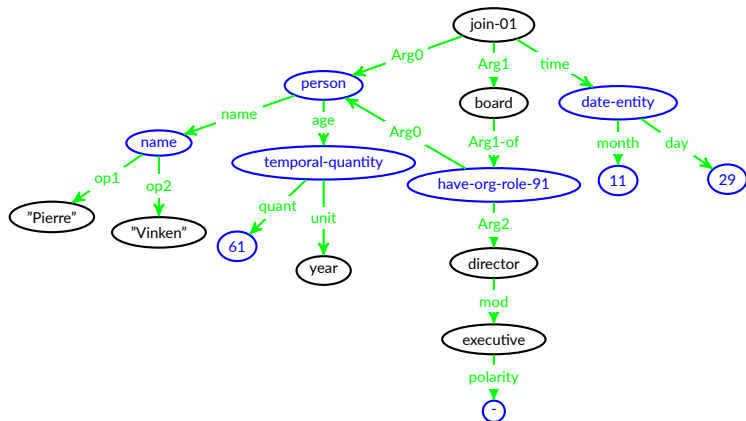
- ▶ Abstract concepts (e.g., *person*, *thing*, *have-org-role-91*) are uniform across languages
 - ▶ Concepts that do not always have explicit lexical support but can be inferred from context
- ▶ UMR defines a set of general participant roles (e.g., *agent*, *theme*, *causer*) and non-participant relations that are uniform across languages
- ▶ But UMR is still not an Interlingua:
 - ▶ Lexical concepts include sense-disambiguated lemmas or simple lemmas and are language-specific (e.g., Mandarin 加入.01 vs. English *join-01* vs. Sanapaná *empahlkay'a*)
 - ▶ Languages can define their own lexicalized participant roles (e.g., :ARGO of 加入.01)
- ▶ In general, grammatical meaning is language-independent while lexical meaning is language-specific

Language-independent vs language-specific aspects



“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



"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).

Where do we find abstract eventive concepts?

Semantic type and information packaging (Croft 2001):

	Reference	Modification	Predication
Entities	UNMARKED NOUNS	relative clauses, PPs on nouns	predicate nominals, complements
States	deadjectival nouns	UNMARKED ADJECTIVES	predicate adjectives, complements
Processes	event nom- inals, com- plements, infinitives, gerunds	participles, rel- ative clauses	UNMARKED VERBS

Where do we find abstract eventive concepts?

- ▶ Sentence-level information packaging is not always predicational:
 - ▶ *I have a book* - “thetic”, “all-new”, “presentational”
 - ▶ *The book belongs to me* - “predicative”, possessee is known information
- ▶ AMR does not distinguish these meanings, UMR does only in typically “non-verbal” contexts:
 - ▶ Possession
 - ▶ Location
 - ▶ Object/Property predication

Where do we find abstract eventive concepts?

- ▶ Languages use different strategies to express these meanings:
 - ▶ Overt copula: English *I have a book*
 - ▶ Juxtaposition: Tiwi *ngawa mantani teraka* 'Our friend has a wallaby, lit. [As for] our friend, wallaby.'
 - ▶ Predicativized possessum: Yukaghir *pulundie jowjen'i* 'The old man has a net, lit. The old man net-has.'
- ▶ UMR assumes annotators are able to recognize the semantics of these constructions and select the appropriate abstract predicate and its participant roles
- ▶ UMR does not require alignment between concepts and words

Sample abstract events

Clause type	Predicate	ARG0	ARG1
thetic possession	have-03	possessor	possessum
predicative possession	belong-01	possessum	possessor
thetic 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	object category
equational	identity-91	theme	equated referent

Example abstract events

an-yetn-eye' ko'o vakka-hak ah-angkok
2/3F-exist-V1.NFUT 1SG:PRO cow-old/broken 1SG-POS
'I have a book.' lit. 'My book exists.'
(h / have-03
 :ARG0 (p / person
 :ref-person 1st
 :ref-number Singular)
 :ARG1 (v/ vakkahak 'book')
 :aspect State)

Named entities

Type	Subtype (AMR NE Type)
	person, family, animal, language, nationality, ethnic-group, regional-group, religious-group, political-movement
Organization	commerical-org (company), political-org (political-party), government-org (government-organization), military-org (military), criminal-org (criminal-organization), academic-org (school, university, research-institute), sports-org (team, league), market-sector
Geographic-entity	ocean, sea, lake, river, gulf, bay, strait, canal, peninsula, mountain, volcano, valley, canyon, island, desert, forest
Celestial-body	moon, planet, star, constellation region local-region, country-region, world-region GPE city, city-district, county, state, province, territory, country
facility	airport, station, port, tunnel, bridge, road, railway-line, canal, building, theater, museum, palace, hotel, worship-place, market, sports-facility, park, zoo, amusement-park
event	incident, natural-disaster, earthquake, war, conference, game, festival
product	vehicle, ship, aircraft, aircraft-type, spaceship, car-make, work-of-art, picture, music, show, broadcast-program
publication	book, newspaper, magazine, journal

Language-independent vs language-specific participant roles

- ▶ Core participant roles are defined in a set of frame files (valency lexicon). 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. 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

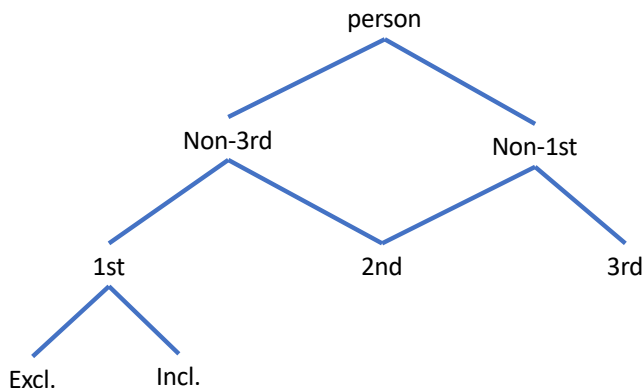
UMR Annotation	Definition
Actor	Animate entity that initiates the action
Undergoer	Entity (animate or inanimate) affected by the action
Theme	Entity (animate or inanimate) moving from one entity to another, spatially or metaphorically
Recipient	Animate entity that gains possession (or at least temporary control) of another entity
Force	Inanimate entity that initiates the action
Causer	Animate entity that acts on another animate entity to initiate the action
Experiencer	Animate entity that cognitively or sensorily experiences a stimulus
Stimulus	Entity (animate or inanimate) that is experienced by an experiencer

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 and relations to accommodate variability across languages.
 - ▶ Languages with 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.
- ▶ Lattices for Aspect, Modal Strength, Person, Number, Discourse Relations, Modification Relations



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 map to one concepts

- ▶ UMR is working on consistent standards for annotating MWEs cross-linguistically

(i / intrigue-01	(x0/敲竹杠-01
:Aspect Performance	:arg0 (x1/他)
:ARG0 (a / aspect	:arg1 (x2/人
:ARG1-of (m / moral-02)	:mod (x3/老)
:poss (m2 / movement-07))	:mod (x4/可怜)
:ARG1 (h / he)	:mod (x5/那)
:mod (a2 / as-well))	:cunit (x6/□)))

The moral aspects of the movement intrigued him as well

对那个可怜的老人，他还敲竹杠。

Concepts can map to words that are discontinuous

(x0/□忙-01

:aspect Performance

:arg0 (x1/地理□)

:beneficiary (x2/我)

:degree (x3/大))

(w / want-01

:ARG0 (p / person)

:ref-person 3rd

:ref-number singular

:ARG1 (g / give-up-07

:ARG0 h

:ARG1 (t / that))

:ARG1-of (c / cause-01

:ARG0 (a / amr-un

地理学帮了我很大的忙。

“Why would he want to give
that up?”

UMR sentence-level additions

- ▶ An *aspect* attribute to eventive concepts
 - ▶ *Aspect* refers to the internal constituency of events, and indicates whether it is on-going or completed
- ▶ *Person* and *number* attributes for pronouns and other nominal expressions
- ▶ Quantification scope between quantified expressions to facilitate translation of UMR to logical expressions

Aspect

- ▶ *Process*: unspecified type of process
- ▶ *State*: unspecified type of state
 - ▶ *reversible state*: acquired state that is not permanent
 - ▶ *irreversible state*: acquired state that is permanent
 - ▶ *inherent state*: state that is not acquired and permanent
 - ▶ *point state*: state that is acquired and reversed at a single point in time
- ▶ *Habitual*: an event that occurs regularly in the past or present.
- ▶ *Activity*: an event has not necessarily ended and may be ongoing at Document Creation Time (DCT).
- ▶ *Endeavor*: used for processes that end without reaching completion (i.e., termination)
- ▶ *Performance*: used for processes that reach a completed result state.

UMR attribute: aspect

He denied any wrongdoing.

(d / deny-01

:Aspect Performance

:ARG0 (p / person)

:ref-person 3rd

:ref-number Singular

:ARG1 (t / thing

:ARG1-of (d2 / do-02

:ARG0 h

:ARG1-of (w / wrong-02))))

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, but pronouns differ from language to language, so in UMR they are decomposed into person and number attributes.

AMR:

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

UMR:

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

“He saw rare birds today.”

UMR attributes: Person and number

- ▶ Person and number attributes are not limited to pronouns

```
(s / see-01
  :ARG0 (p / person
         :name (n / name
                :op1 "Bill"))
  :ARG1 (b / bird
         :mod (r / rare)
         :ref-number Plural))
```

Bill saw rare birds today.

UMR attributes: Person and number

- ▶ Person and number can be applied uniformly across languages

```
(e / entoma-00 'eat'  
  :Actor (p / person  
           :ref-person 3rd  
           :ref-number Plural)  
  :Undergoer (t / thing  
               :ref-number Singular)  
  :aspect Performance  
  :polarity -)  
m-e-hl-t-om-o=hlta  
NEG-2/3M.IRR-DSTR-eat-PST/HAB-SBJ=PHOD  
"They did not eat it."
```

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))

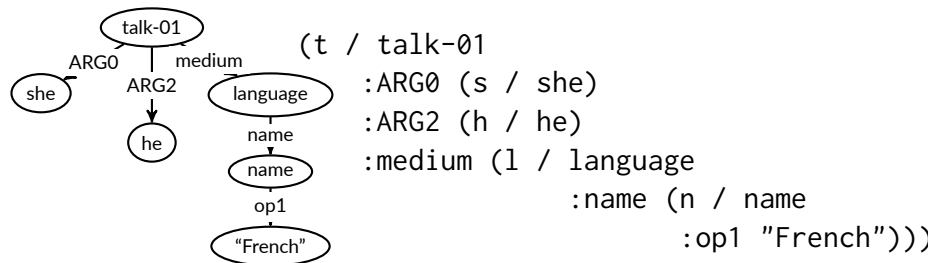
$\exists p(\text{person}(p) \wedge \neg \forall q(\text{question}(q) \rightarrow$

$\exists a(\text{answer-01}(a) \wedge \text{ARG1}(a, q) \wedge \text{ARG0}(a, p))))$

UMR 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 graph as identity or subset relations between named entities or events.

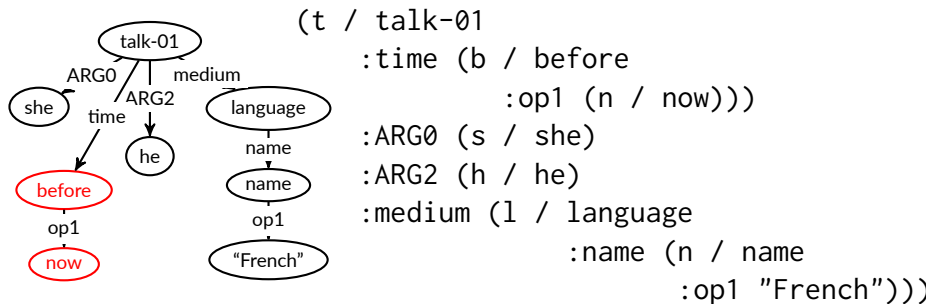
No representation of tense in AMR



- ▶ "She talked to him in French."
- ▶ "She is talking to him in French."
- ▶ "She will talk to him in French."

Adding tense is seemingly straightforward

Adding tense to AMR involves essentially defining a temporal relation between the time of the event and the Document Creation Time (DCT) or speech time (Donatelli et al 2019).



“She talked to him in French.”

Limitations of simply adding tense

- ▶ For some events, its temporal relation to the DCT or speech time is undefined. “John said he would go to the florist shop”.
 - ▶ Is “going to the florist shop” before or after the DCT?
 - ▶ Its temporal relation is more naturally defined with respect to “said”.
- ▶ In quoted speech, the speech time has shifted. “I visited my aunt on the weekend,” Tom said.
 - ▶ The reference time for “visited” has shifted to the time when Tom said this. We only know the “visiting” event happened before the DCT indirectly.
- ▶ Tense is not universally grammaticalized, e.g., Chinese

Limitations of simply adding tense

- ▶ Even in cases when tense, i.e., the temporal relation between an event and the DCT is clear, tense may not give us the most precise temporal location of the event.
 - ▶ John **went** into the florist shop.
 - ▶ He **had promised** Mary some flowers.
 - ▶ He **picked out** three red roses, two white ones and one pale pink
 - ▶ Example from (Webber 1988)
- ▶ All three events happened in the past, i.e., before the DCT, but we also know that the “going” event happened after the “promising” event, but before the “picking out” event.

A structured approach to temporal interpretation

So it seems that we can't properly interpret the temporal relations without a clear notion of *reference time*. So in the UMR framework, we propose to:

- ▶ Explicitly represent the temporal location of an event as a relation between the event and its reference time
- ▶ In addition to the speech time or DCT, possible reference times also include other events, time expressions, or a general past, present, or future reference
- ▶ The events and their reference times in a text will form a dependency graph where the nodes are events and time expressions and the edges are temporal relations

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

Identifying reference times for events

- ▶ Reference time is the DCT
 - ▶ The Pentagon said today that it will re-examine the question.
 - ▶ DCT → re-examine
 - ▶ The Pentagon said today that it would re-examine the question.
 - ▶ said → re-examine

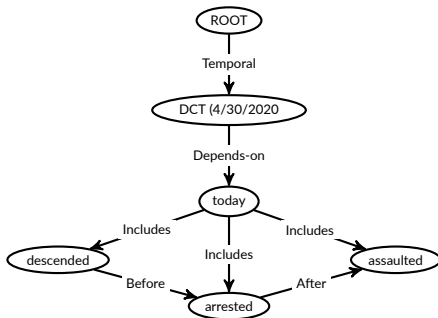
Identifying reference times for events

- ▶ The reference time of an event is another event
 - ▶ John went into the florist shop. He had promised Mary some flowers. He picked out three red roses, two white ones and one pale pink
 - ▶ went → had promised
 - ▶ went → picked out



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."

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 represents 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."
- ▶ The 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

The interaction of modality and polarity

- ▶ When assessing that reality status of a statement, it is often necessary consider the interaction between modality and polarity (Vigus et al 2019):

Label	Value	Example
AFF	full affirmative	The dog barked
PARTAFF	partial affirmative	The dog probably barked
NEUTFF	neutral affirmative	The dog might have barked
NEUTNEG	neutral negative	The dog might not have barked
PARTNEG	partial negative	The dog probably didn't bark
NEG	full negative	The did not bark

Modal Source or Conceiver

- ▶ In order to assess the validity or factuality of a statement, we also need to take into consideration the credibility of the source (Vigus et al 2019).
 - ▶ “Mary might have walked the dog”. It is the author’s belief that it is possible that Mary has walked the dog.
- ▶ The modal source or conceiver can be nested:
 - ▶ “Mary said that Henry told her that John thinks the cat is hungry.”
- ▶ To assess if the cat is really hungry or not, the credibility of each one of sources will affect the outcome.

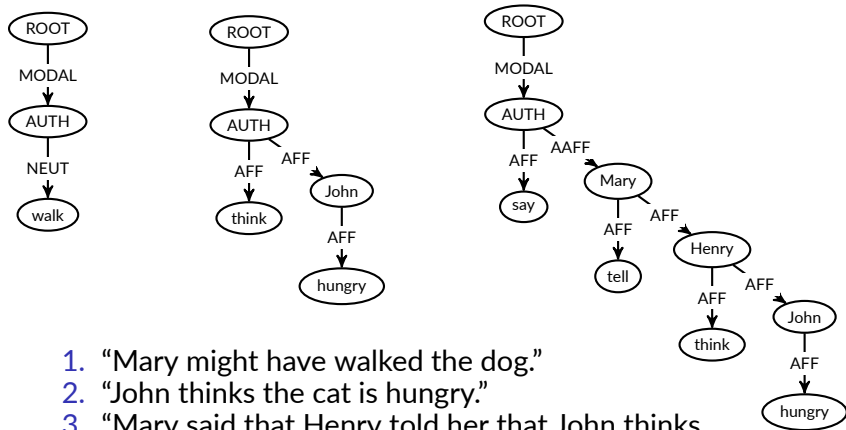
Why is source important to factuality assessment

“WBUR: A man in his 20s from Worcester County tested positive Tuesday for the new, apparently more contagious coronavirus variant, public health officials said. The variant was first detected in the United Kingdom, and experts have warned that it could soon become widespread in the U.S. ”

- ▶ Is the event “testing positive” as credible if it comes from your neighbor?

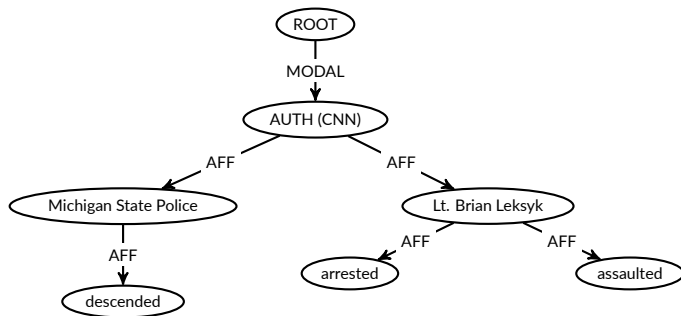
A structured representation of modality and polarity

- ▶ We can use a dependency structure to model the nestedness of modal source Vigus et al. (2019); Yao et al. (2021):



1. "Mary might have walked the dog."
2. "John thinks the cat is hungry."
3. "Mary said that Henry told her that John thinks the cat is hungry."

Modal dependency structure (MDS)



“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

- ▶ Entity coreference
 - ▶ 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 possessive and controlling but he has no right to be as we are not together.
- ▶ Event coreference

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 / contrast-01
  :ARG1 (s4c3 / control-01
    :ARG0 (s4p2 / person
      :ref-person 3rd
      :ref-number Singular)
    :degree (s4v / very)))
  :ARG2 (s4r / right-05 :polarity -
    :ARG1 s4p2
    :ARG1-of (s4c2 / cause-01
      :ARG0 (s4t / together
        :Aspect State
        :domain (s4p3 / person
          :ref-person 1st
          :ref-number Plural))))))
(s / sentence
  :coref ((s4p2 :subset-of s4p3)))
```

Implicit arguments

- ▶ Implicit arguments can be inferred from context and can be annotated for coreference like overt (pronominal) expressions

(s3d / deny-01
:Aspect Performance
:ARG0 (s3p / person
:ref-number Singular
:ref-person 3rd)
:ARG1 (s3t / thing
:ARG1-of s3d2 / do-02
:ARG0 s3p
:ARG1-of
(s3w / wrong-02))))

(x0 / person
:ARG0-of (x1 / 卖-01)
:Aspect Process
:ARG1 (x2/水果))
“卖水果的”

“He denied any wrongdoing.”

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 ((s1t2 :before DCT)

(s1t3 :depends-on DCT))

:modal ((s1t2 :AFF AUTH)))

Sentence-level representation vs document-level representation

Pope is the American businessman who was convicted last week on spying charges and sentenced to 20 years in a Russian prison.

(s2b2 / businessman

:Aspect State

:mod (s2c5 / country

:name (s2n6 / name :op1 "America"))

:domain (s2p / person

:name (s2n5 / name :op1 "Pope"))

:ARG1-of (s2c4 / convict-01

:Aspect Performance

:ARG2 (c / charge-05

:ARG1 s2b2

:ARG2 (s2s2 / spy-01 :ARG0 s2p))

:time (s2w / week :mod (s2l / last)))

:ARG1-of (s2s / sentence-01

:Aspect Performance

:ARG2 (s2p2 / prison

:mod (s2c3 / country

:name (s2n4 / name :op1 "Russia"))

:duration (s2t3 / temporal-quantity

:quant 20

:unit (s2y2 / Year)))

:ARG3 s2s2))

(s2 / sentence

:temporal ((s2c4 :before s1t2)
(s2s :after s2c4))

:modal ((s2c4 :AFF AUTH)
(s2s :AFF AUTH))

Sentence-level representation vs document-level representation

He denied any wrongdoing.

(s3d / deny-01

:Aspect Performance

:ARG0 (s3p / person

:ref-number Singular

:ref-person 3rd)

:ARG1 (s3t / thing

:ARG1-of (s3d2 / do-02

:ARG0 s3p

:ARG1-of

(s3w / wrong-02))))

(s3 / sentence

:temporal ((s3d :before DCT))

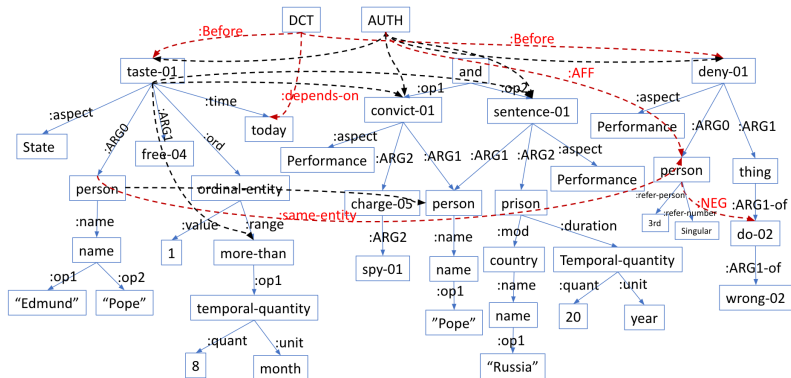
:modal ((s3d :AFF AUTH)

(s3d2 :NEG

(s3p :AFF AUTH)))

:coref ((s3p :same-entity s1p)))

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."

UMR writer: sentence-level interface

UMR Annotation Tool

Home About

Annotate New Post Account Logout

File Name: sample_snts_english.txt File Language: English

Doc ID in database: 18

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

2 He denied any wrongdoing .

Sentence ID: 1

save&go

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

UMR Penman Annotation:

{s1t / taste-01
:ARG0 {s1p / person
:name {s1n / name
:op1 "Edmund"
:op2 "Pope")
:ARG1 {s1f / freedom
:time {s1t2 / today}}

Alignments:

Edmund(): 1-1
Pope(): 2-2
name{s1n}: -1--1

Roles(English)

Attributes

Abstract Concept

Named Entity Types

Add abstract concept

Lexicalized Concept:
↓

Add attribute

export

replace

delete

undo

redo

doc-level-annot

Quick links

UMR Guidelines AMR Guidelines User Guide

Contact

jinzhaob@brandeis.edu

UMR Writer: document-level interface

UMR Annotation Tool

Home

About

Annotate

New Post

Account

Logout

Current User: xue

User ID: 3

File Name: sample_snts_english.txt

File Language: English

Doc ID in database: 18

1. Sentence: Edmund Pope tasted freedom today for the first time in more than eight months .

(s1t / taste-01

:ARG0 (s1p / person

:name (s1n / name

:op1 "Edmund"

:op2 "Pope")

:ARG1 (s1f / freedom)

:time (s1t2 / today))

2. Sentence: He denied any wrongdoing .

(s2d / deny-01

:ARG0 (s2h / he))

Doc-Level Annotation:

(s1 / sentence

:temporal ((s1t :before DCT)

(s1t :overlap s1t2)))

current sentence: 1

save&go

Relations: ▾

Child:

Parent:

submit

export

delete

undo

redo

sent-level-annot

Doc-Level Annotation:

(s2 / sentence

:temporal ((s2d :before s1t)

(s2d :before DCT)))

Quick links

UMR Guidelines

AMR Guidelines

User Guide

Contact

jinzhao@brandeis.edu

Discussion question: Do we still need symbolic meaning representations?

- ▶ End-to-end neural models changed the landscape of NLP, e.g., Neural MT, machine reading
- ▶ Diminished returns for linguistic structures as an intermediate representation in some end-to-end systems, e.g., NMT
- ▶ However, neural models do have their limitations: hard to interpret, hard to anticipate errors, not naturally suited for logical/quantitative reasoning that human routinely perform. Systems based on symbolic meaning representations can provide a viable alternative
- ▶ Hard NLP problems cannot be solved without world knowledge, and a general purpose symbolic meaning representation can be used to distill structured knowledge from natural language text

Adapting to the new environment when developing linguistic resources

- ▶ Do not design meaning representations as an intermediate representation, but rather as an end (or near-end) representation
- ▶ Support applications where neural models do not provide a good solution
 - ▶ Temporal reasoning that answers questions that cannot answered with a machine reading approach (not no comprehension)

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

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 representation
- ▶ UMR aims to achieve semantic stability across syntactic variations and support logical inference
- ▶ UMR is a cross-lingual meaning representation that separates aspects of meaning from those that are language-specific
- ▶ We are testing UMR English, Chinese, Arabic, Arapahoe, Kukama, Sanapana, Navajo

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*.