Noname manuscript No.
(will be inserted by the editor)

Designing a Uniform Meaning Representation for Natural Language Processing

Jens E. L. Van Gysel · Meagan Vigus · Jayeol Chun · Kenneth Lai · Sarah Moeller · Jiarui Yao · Tim O'Gorman · Andrew Cowell · William Croft · Chu-Ren Huang · Jan Hajič · James H. Martin · Stephan Oepen · Martha Palmer · James Pustejovsky · Rosa Vallejos · Nianwen Xue

Abstract In this paper we present Uniform Meaning Representation (UMR), a meaning representation designed to annotate the semantic content of a text. UMR is primarily based on Abstract Meaning Representation (AMR), an annotation framework initially designed for English, but also draws from other meaning representations. UMR extends AMR to other languages, particularly morphologically complex, low-resource languages. UMR also adds features to AMR that are critical to semantic interpretation and enhances AMR by proposing a companion document-level representation that captures linguistic phenomena such as coreference as well as temporal and modal dependencies that potentially go beyond sentence boundaries.

Keywords Natural Language Processing · meaning representation

This work is supported in part by a grant from the IIS Division of National Science Foundation (Awards No. 1763926, 1764048, 1764091) entitled Building a Uniform Meaning Representation for Natural Language Processing awarded to Nianwen Xue, James Pustejovsky, Martha Palmer and William Croft. All views expressed in this paper are those of the authors and do not necessarily represent the view of the National Science Foundation. This work is supported in part by a grant from the Ministry of Education, Youth and Sports of the Czech Republic (Project No. LM2018101) and in part by the Czech Science Foundation (Award No. GX20-16819X).

Jens E. L. Van Gysel, Meagan Vigus, William Croft, and Rosa Vallejos
University of New Mexico
jelvangysel,mlvigus,wcroft,rvallejos@unm.edu

Jayeol Chun, Kenneth Lai, Jiarui Yao, James Pustejovsky, Nianwen Xue
Brandeis University
jchun,klai12,jryao,jamesp,xuen@brandeis.edu

Sarah Moeller, Andrew Cowell, James H. Martin, Martha Palmer
University of Colorado at Boulder
sarah.moeller,james.cowell,james.martin,
martha.palmer@colorado.edu

Tim O'Gorman
University of Massachusetts at Amherst
timjogorman@gmail.com

Chu-Ren Huang
Hong Kong Polytechnic University
churen.huang@polyu.edu.hk

Jan Hajič
Charles University
hajic@ufal.mff.cuni.cz

Stephan Oepen
University of Oslo
oe@ifi.uio.no

1 Introduction

It is undeniable that neural network-based end-to-end systems have led to fundamental changes in the landscape of NLP. In areas where large training data sets exist, such as machine translation [4] and machine reading [63, 80, 83], end-to-end systems enabled by neural network models have reduced reliance on intermediate semantic (and other) representations. The successful application of such end-to-end systems has caused many to wonder whether it is still necessary to invest time and money in building such linguistically annotated resources. We argue that the emergence of a host of new application scenarios makes the need for deep semantic analysis and representations more urgent than ever. In human-robot interactions, meaning representations are needed as the medium of communication between human users and robots. Intelligent agents in the medical field require intermediate meaning representations in order to provide interpretable background for predictions, judgments and diagnoses. Even in machine translation, while end-to-end systems have made impressive advances, especially in fluency [10, 11, 18], intermediate structures can provide "scaffolding" for the learning process [69], improving the faithfulness of translations [31, 82].

This paper presents the fundamentals of Uniform Meaning Representation (UMR), a practical and cross-linguistically valid meaning representation designed to meet the needs of a wide range of NLP applications. The remainder of the paper is organized as follows. In Section 2, we lay out four desiderata that guide the design of UMR. In Section 3, we present an overview of Abstract Meaning Representation that serves as the starting point of UMR. We present UMR sentence-level extensions to AMR in Section 4, and document-level extensions in Section 5. We discuss how UMR addresses cross-linguistic diversity in linguistic distinctions and in mapping words to UMR concepts in Section 6. In Section 7, we present our strategy for applying UMR to minority languages that face cultural and technological challenges and lack of foundational resources. Although annotated full UMRs are not yet available as we actively develop tools to support UMR annotation, we present experiments on novel aspects of UMR in Section 8 that show they can be annotated reliably. In Section 9, we discuss how UMR is related to existing meaning representations, and in particular, we compare UMR with existing meaning representations in how they address each aspect of our four desiderata. We conclude in 10.

2 Design goals for UMR

The design of UMR is guided by the following four desiderata:

- **Scalability/learnability.** Meaning representations are expected to be automatically reproduced by machine learning systems trained on data annotated with this representation. As such, it is important for the meaning representations to be annotated at scale on large data sets. This means that the meaning representation needs to be intuitive so that it does not put too many constraints on the pool of annotators who are capable of performing the annotation. The meaning representation also needs to be a formal object such as a tree or a graph that is easy to manipulate algorithmically.
- Supporting similarity-based **lexical inference.** Natural languages are known to be both *variable* (the same meaning can be expressed through different morphosyntactic constructions) and *ambiguous* (the same surface string can have different meanings in different contexts). For a meaning representation to support lexical inference, different natural language expressions that have the same meaning should be expressed in the same way. This means that the meaning representation needs to abstract away from the morphosyntactic variations, disambiguate the senses of a word or phrase, and resolve references of referring expressions such as proper nouns and pronouns.
- Supporting **logical inference**. Supporting logical inference has been the primary goal for classical meaning representations, which aim to be easily translatable to logical form – typically first-order logic. Logical systems allow new statements to be inferred from known facts, and linguistic phenomena such as quantification, negation, tense and aspect, and modality have traditionally figured prominently in logic-based meaning representations. First-order logic formalisms have also played a key role in grounded semantic parsing, the goal of which is to parse natural language queries into first-order logic-based meaning representations that can be executed against knowledge bases [35, 46, 81, 49, 16, 17, 64]. It is also important to canonicalize referring expressions etc. so that they can be easily grounded to external knowledge bases.
- **Cross-linguistic plausibility and portability**. We envision a meaning representation that is uniform across languages, so that a wide variety of languages can be annotated in a comparable way. It must thus be able to deal with variability in morphosyntax (e.g. constituent order, degree of inflectional synthesis of the verb), grammaticalization of different ways of dividing up conceptual space [75], and different morphosemantic mappings between concepts and words.

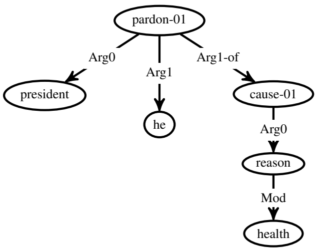
3 Overview of Abstract Meaning Representation

When designing UMR we use Abstract Meaning Representation (AMR) [6], a meaning representation designed for English, as a starting point, and extend it to other languages and enhance its expressiveness. AMR has attracted significant attention in recent years due to its simplicity and its focus on semantic content such as predicate-argument structure, named entities, and word senses that are key to many NLP applications. Its formal properties as a single-rooted, node- and edge-labeled directed graph also make it amenable to machine learning based parsing algorithms [32, 79, 50, 87, 15], adding to its attractiveness. In this sense it has already satisfied the first two of our UMR design goals.

An example AMR is provided in (1). In this example, the AMR of the sentence "The president pardoned him for health reasons" is formally a graph where the nodes represent semantic concepts and edges represent relations. The semantic concepts can be word senses (e.g., *pardon-01*, *cause-01*) or word lemmas when the senses of a word are yet to be defined (e.g., *president*, *he*, *reason*, *health*). The concepts can also be entity types (e.g., *person*, *date-entity*), or quantity types (e.g., *monetary-quantity*, *distance-quantity*). AMR relations include participant roles that are defined for each predicate (e.g., *ARG0*, *ARG1*), as well as general semantic relations (e.g., *MOD*).

(1) The president pardoned him for health reasons.

```
(p3 / pardon-01
    :ARG0 (p / president)
    :ARG1 (h2 / he)
    :ARG1-of (c /cause-01
                :ARG0 (r / reason
                        :MOD (h /health))))
```



UMR extends AMR in three core ways. First, while it has been shown that AMR can be extended to languages like Chinese, Czech, or Korean [47, 48, 19] that have existing foundational resources like valency lexicons or frame files, it is not immediately clear how to extend it to language families with different morphosyntactic properties and especially to low-resource languages. Second, while existing meaning representations such as Minimal Recursion Semantics (MRS) [21, 34] and Discourse Representation Structures (DRS) [45, 13] have been designed to support logical inference, AMR lacks modal, aspectual, and scopal annotation that is crucial to logical inference. Finally, while multi-sentence AMR [58] includes inter-sentential coreference that goes beyond AMR's original sentence-level focus and allow, for example, the concept *he* posited for the pronoun "him" in (1) to be linked to its referent in a preceding sentence, it lacks annotation of temporal and modal relations that can also go beyond sentence boundaries.

In the next four sections, we will present our extensions and refinements that extend AMR in the three ways described in the previous paragraph. In Section 4, we present our extension and refinements to AMR at the sentence level, specifically how UMR annotates aspect (Section 4.1) and scope (Section 4.2). We also show how UMR scope annotation can be used to support conversion to first-order logical expressions. In Section 5, we discuss document-level refinements and extensions to AMR. In particular, we discuss how we add coreference and temporal and modal dependencies to the UMR document-level representation, and how sentence-level and document-level representations are combined. In Section 6, we refine UMR to accommodate cross-linguistic variation in semantic distinctions that are encoded in languages. Specifically, we discuss how to annotate grammatical semantic distinctions in different languages in a comparable way (Section 6.1), and cross-linguistic issues in mapping word tokens in sentences to UMR concepts (Section 6.2). Finally, in Section 7, we present a road map for annotating UMRs for low-resource languages, particularly focusing on annotating participant roles in languages without frame files.

4 Extensions: expanding the semantic range of AMR at the sentence level

Sentence-level representation refers to semantic categories pertaining to single events. These include participants in the event (predicate-argument structure) semantic categories such as valency, aspect, and participant roles; and quantification and scope relations among participants. Predicate-argument structure is well represented in current AMR; in this section we describe the extension of AMR to include annotation of aspect, quantification and scope relations.

4.1 Aspect in UMR

The UMR aspect annotation, building on [27, 28], marks a feature on events that captures their internal temporal and qualitative structure. The annotation values do not correspond to specific verbs or constructions in a language, but characterize the event in context.

The annotation distinguishes five base level aspectual values – ***State***, ***Habitual***, ***Activity***, ***Endeavor***, and ***Performance*** – and a range of more fine-grained and more coarse-grained values organized in a lattice format [75] as described in Section 6.1. The State value corresponds to stative events in [76]; no change occurs during the event. It also includes predicate nominals (*be a doctor*), predicate locations (*be in the forest*), and thetic (presentational) possession (*have a cat*). The Habitual value is annotated on events that occur regularly in the past or present. The Activity value indicates an event has not necessarily ended and may be ongoing at Document Creation Time (DCT). Endeavor is used for processes that end without reaching completion (i.e., termination), whereas Performance is used for processes that reach a completed result state. The Performance value corresponds to achievements and accomplishments [76]. Event nominals are typically hard to annotate for aspect, since they lack the grammatical cues that verbs often show. Therefore, they are all annotated with the coarse-grained value ***Process***.

The aspect annotation is implemented as an aspect feature (e.g., :aspect Peformance) in UMR. For examples, please refer to Figure 1.

4.2 Scope in UMR

One notable shortcoming of AMR is its lack of representation for scoping relations, leading to scope ambiguity even in cases where it is not warranted, a problem when translating AMRs to first-order logic expressions. While other meaning representations, including MRS and DRS, explicitly represent scope, we want to preserve the advantages of AMR, including its relative simplicity and focus on predicate-argument structure. We therefore follow [62] and augment predicates with an optional *scope node*, that specifies the relative scope ordering of each of its arguments. For example, consider the UMR for the sentence in (2):

(2) "Someone didn't answer all the questions"
    (a / answer-01
        :ARG0 (p / person)
        :ARG1 (q / question :quant A :polarity -)
        :pred-of (s / scope :ARG0 p :ARG1 q))

The scope node indicates that "someone" takes wide scope over (not) "all the questions" (i.e., there exists someone who didn't answer all the questions). If the argument order were reversed (:ARG0 q and :ARG1 p), another interpretation arises, where some questions were not answered by anyone. Note *pred-of* is an inverse relation that indicates *answer-01* is a predicate under the scope node.

We adopt a continuation-passing style semantics for scope [7], inspired by the semantics for AMRs in [12]. Briefly, the relative scopes of each argument are determined by the order of evaluation. A scope node, if present, then acts as a restriction on the possible orderings. For example, a continuized representation for the above AMR, with ⟦someone⟧ evaluated before ⟦not all the questions⟧, is given below in (3a), with the corresponding first-order logic expression in (3b):

(3) a. $\lambda k.[\![\text{someone}]\!](\lambda n.[\![\text{not all the questions}]\!]$
$(\lambda o.[\![\text{answered}]\!]$
$(\lambda m.\text{ARG1}(m,o) \wedge \text{ARG0}(m,n) \wedge k(m))))$
b. $\exists p(\text{person}(p) \wedge \neg\forall q(\text{question}(q) \rightarrow \exists a(\text{answer-01}(a)$
$\wedge\, \text{ARG1}(a,q) \wedge \text{ARG0}(a,p))))$

5 Extensions: Expanding AMR to document-level representation

The sentence-level representation presented in Section 4 in and of itself is insufficient to properly interpret the semantic content of a text, as some semantic relations go beyond sentence boundaries. Such semantic relations include entity and event coreference, temporal relations between events, and modal dependencies. We represent semantic relations that go beyond sentence boundaries in a *document-level representation* that complements the sentence-level representation. It is important to note that the document-level semantic relations *can but do not necessarily* go beyond sentence boundaries. For instance, coreference can occur across sentence boundaries, or within the same sentence. The same is true for temporal relations and modal dependencies.

5.1 Coreference

As we can see from our AMR example in (1), resolving anaphoric expressions such as pronouns is essential to the interpretation of semantic content of a text. The UMR entity coreference annotation, like AMR, includes both entity and event coreference, and extends it to inter-sentential coreference. The entity annotation includes identity relations where an anaphoric expression refers to the same entity as another expression (***same-entity***) in a document as in (4), and ***subset*** relations where the referent of an anaphoric expression is a subset of the referents for another expression, as in (5).

(4) a. **Edmund Pope** tasted freedom today for the first time in more than eight months.
b. **He** denied any wrongdoing.

(5) **He** is very possesive and controlling but **he** has no right to be as **we** are not together.

The UMR event coreference annotation includes event identity where there are multiple mentions of the same event (***same-event***) as in (6), as well as cases where one event mention is a ***subset*** of another event mention as in (7). The decision to annotate event coreference is partially motivated by the need to make inferences on the temporal relations between events, which we discuss in the next section.

(6) a. El-Shater and Malek's property was **confiscated** and is believed to be worth millions of dollars.
b. Abdel-Maksoud stated the **confiscation** will affect the Brotherhood's financial bases

(7) a. Demirtas was 1 of 10 people **arrested** in May 2008 during a crackdown led by France on people suspected of helping fund the Islamic Movement of Uzbekistan (IMU).
b. 8 **arrests** took place in a suburb of the eastern French city of Mulhouse and in the central Rhone region.

Clearly, there are other types of coreference such as bridging that UMR does not currently consider, in order to keep UMR simple and practical. We demonstrate how coreference is annotated in UMR in Section 5.4.

5.2 Temporal dependencies

UMR adopts the TimeML view that the temporal relations in a text can be interpreted in terms of relations between a time expression and an event, between two events, and between two time expressions [61]. This is a broader temporal annotation approach than annotating just tense, which is the temporal relation between the time when an event occurs and the document creation time (DCT). There are two reasons for this. One is that for many languages (e.g., Chinese), tense is not overtly grammaticalized, and as a result, the relation is not intuitive to speakers of those languages [84]. Another reason is that tense annotation alone is insufficient for interpreting the temporal relations in a text. Two events that are both in the past may have a temporal precedence that cannot be captured with tense alone. UMR further adopts the idea that the temporal relations in a document are hierarchically organized in a temporal dependency structure [88], a view that is compatible with graph representation of the rest of the UMR annotation. For the most part, the event-time relations are annotated as part of the predicate-argument structure annotation at the sentence-level. In example (9) in Section 5.4, for example, based on the predicate-argument structure annotation at the sentence level, we know that Edmund tasted freedom "today". To properly interpret the temporal content of this sentence, however, we also need to properly interpret when "today" is. This is done at the UMR document-level annotation, which focuses on event-event and time-time relations. The UMR temporal annotation proceeds as follows. For each relative time expression (e.g., "today", "yesterday"), we identify another time expression it *depends on* to resolve it to an absolute time. For example "today" in (9) depends on the DCT of the sentence in order to be resolved. For each event, we identify another *reference event* with respect to which the temporal location of this event can be mostly specifically defined. The determination of the most specific event can be determined based on grammatical or contextual clues. For example, from the context, we can determine that the *convict* event in (10) happened before the *tasted* event in (9). The implementation of UMR document-level annotation is illustrated in Figure 1.

5.3 Modal dependencies

The UMR modal dependency captures the epistemic strength and polarity of events, as related to conceivers (or, sources) [77]. The epistemic strength and polarity relations are largely based on FactBank [67]: ***full affirmative (:AFF)***, ***partial affirmative (:PARTAFF)***, ***neutral affirmative (:NEUTAFF)***, ***neutral negative (:NEUTNEG)***, ***partial negative (:PARTNEG)*** and ***negative (:NEG)***. Events and conceivers (sources) make up the nodes in the dependency structure and epistemic strength and polarity characterize the edges. The dependency structure parallels the annotation for temporal relations [88]. The dependency structure permits the representation of nested modal values, necessary to annotate certain linguistic constructions, as in (8), where the author of the text has only a partial affirmative commitment to the senator's belief, which in turn represents a neutral affirmative commitment to the bill's passing tomorrow.

(8) The senator probably thinks that the bill could pass tomorrow.

```
:modal ((pass :NEUTAFF senator)
     (senator :PARTAFF AUTH))
```

The dependency structure uses the same modal strength values for the links between two conceivers, two events, or a conceiver and an event. Scope relations between modality and negation are represented in the dependency structure.

When annotating modal dependencies, annotators do not need to construct the entire dependency structure themselves, especially at Stage 0 of the road map (see Section 7, simplifying the annotation process). Annotators are expected to give events a MODSTR ("modal strength") value, but conceivers remain unspecified. Events under the scope of a modal predicate and events under the scope of a reporting predicate receive a special annotation: MODAL and QUOT, respectively. Events with a MODAL annotation do not receive a MODSTR annotation, as this can be automatically inferred from the main verb (e.g. *want* conveys a NEUTRAL modal strength onto its complement). Events with a QUOT annotation do receive a regular MODSTR value in addition. The MODSTR, MODAL and QUOT annotations, together with the argument structure and the lexical semantics of modal verbs, can be used to automatically create the dependency structure.

5.4 Integrating document-level and sentence-level annotation

We provide an integrated example for a short text in Figure 1 to illustrate how UMR is implemented. For each sentence, the sentence-level representation is on the left and the document-level representation is on the right[1]. The document-level representation makes reference to sentence-level concepts, using an ID that combines the sentence ID and the concept ID. For instance, "s1t2" refers the concept "t2" in the sentence "s1". The document-level representation consists of a coreference, temporal and modal relations that link entity and event concepts in the current sentence to other concepts, potentially in a previous sentence.

[1] As can be seen from this example, the document-level representation is a list of triples in the form of <dependent relation parent>, and deviates from the Penn notation used for the sentence-level representation.

(9) 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)
                  (s1m :before s1t2)
                  (s1t3 :depends-on DCT))
      :modal ((s1t2 :AFF AUTH)))
```

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

```
(s2i / identity-91
      :Aspect State
      :ARG0 (s2p / person
            :name (s2n5 / name :op1 "Pope"))
      :ARG1 (s2b2 / businessman
            :mod (s2c5 / country
                  :name (s2n6 / name :op1 "America"))
            :ARG1-of (s2c4 / convict-01
                  :Aspect Performance
                  :ARG2 (s2c / charge-05
                        :ARG1 s2b2
                        :ARG2 (s2s2 / spy-01 :ARG0 s2b2))
                  :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)))
```

(11) He **denied** any wrongdoing.

```
(s3d / deny-01
      :Aspect Performance
      :ARG0 (s3p / person
                  :ref 3s)
      :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 (s3h :AFF AUTH)))
      :coref ((s3p :same-entity s1p)))
```

Fig. 1: An integrated UMR example

6 Refinements: Typological adaptation of AMR

One goal of UMR is application to as many languages as possible. Therefore, the annotation values must accommodate cross-linguistic diversity in linguistic distinctions in semantic space; and the annotation process must accommodate cross-linguistic diversity in the semantic distinctions encoded in a language, and in the mapping between words and concepts.

6.1 Adapting UMR to accommodate cross-linguistic diversity in linguistic distinctions

The cross-lingual annotation is guided by two main principles. First, the default categories reflect semantic distinctions overtly expressed in as large as possible a proportion of the world's languages, so that annotators for the majority of languages need not infer meanings that are not expressed through overt forms, except for easily-defined distinctions such as past vs. present. Second, the scheme is flexible enough to accommodate typological diversity in grammaticalized semantic distinctions, and yet preserve cross-linguistic comparability of annotations. Annotation values are organized in typologically-motivated lattices, as proposed in [75]. A paradigmatic lattice organizes potentially overlapping categories of greater or lesser generality. For instance, annotators are encouraged to use the default (bolded) level of modal annotation in Figure 2. Annotators for languages with different grammaticalized distinctions may use labels from the higher and lower levels of the paradigmatic lattice for access to more coarse-grained and more fine-grained categories, respectively. The levels are connected, keeping annotations within and across languages comparable. In addition to the lattice for modal categories presented in Figure 2 lattices for number, certain spatial relations, aspect and time reference can be found in [75]. The categories represented in the lattice are based on existing typological work in [22, 24] and [14],

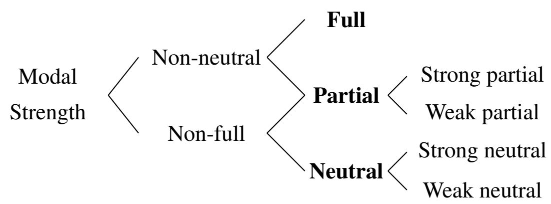


Fig. 2: Paradigmatic lattice for modal strength annotation

6.2 Mapping between words and concepts across languages

Taking a cross-linguistic perspective on concept-word mappings raises issues that have not typically been at the forefront in computational linguistics. There is, for one, no accepted cross-linguistic definition for what constitutes a "word" across languages ([26], but see [89]). Even in closely related languages like English and German, the same concept may be considered one word (German *Schifffahrten*) or two words (English *boat rides*). Despite the typological differences, our annotations do not depend on language-internal wordhood tests, as they are based in semantic rather than formal criteria.

Criteria for concepthood are equally challenging, since languages differ in how they allocate senses to words. But how many concepts are combined into a single word? For example, English *cut* could be broken down into concepts of causation, instrumentality (a bladelike implement) and change of state; and the Arapaho example in (12) divides it into morphemes that way.

(12) nih-teb-e'ei-s-o'

nih- teb -e'ei
PAST- break/remove.stick.like -head
-s -o'
-by.blade.CAUS -1S/3S

'I cut his head off with a knife'

```
(n / teb-e'ei-s
    :actor (p / person :ref 1s)
    :undergoer (p2 / person :ref 3s)
    :theme (t /thing :ref obviative)
    :Aspect Performance)
```

In practice, descriptive linguists and typologists use as a common denominator a set of concepts that are expressed by single morphemes in at least some languages. Typologically, the mapping between such concepts and words is highly variable. This fact has not been fully appreciated because high-resource languages are skewed towards languages with relatively little morphology. Therefore, prior work in computational linguistics has been focused on *multiword expressions (MWEs)* [66], where a single concept is expressed with multiple words. UMR builds on such previous research, allowing individual languages to propose criteria on how to map MWEs to UMR concepts for that language.

Many low-resource languages are morphologically complex, using a single word to express concepts for which English needs multiple words. Such "multiconcept words" pose a different issue for semantic annotation. In such multiconcept words, concepts may be expressed by clearly distinct morphemes (*teb* 'break / remove.stick.like' and *-e'ei* 'head' in (12)), a portmanteau morpheme (*-o'* 1S/3S in (12)), or a single, unanalyzable morpheme (English *kill* encodes both the concepts *die* and CAUSE).

As a matter of principle, *UMR does not require the decomposition of morphologically complex words into morphemes* that map to UMR concepts. Instead, such words can as a whole map to multiple concepts. A number of considerations lead to this design decision. Portmanteau morphemes generally cannot be split into separate parts for each concept. Even when morphemes are separable, annotators with less linguistic training or field linguists in the early stages of analysis may not be aware of where morpheme boundaries lie. Finally, concept-word mismatches (multi-word concepts and multi-concept words) threaten the consistency of annotations across languages. Broadly, we adopt four different solutions to these issues. Each solution depends on the semantic categories involved and their behavior across languages.

First, for categories in which concepts are clearly distinct from each other semantically, we will ask annotators to identify multiple concepts in one word. For now, we apply this solution to **argument indexation** (often called **pronominal affixation**) and certain types of **noun incorporation**.

Many languages use verbal affixes to index arguments, which then may not be expressed elsewhere in the clause. Such indexed arguments are treated in the same way as free pronouns (i.e., identified as arguments of the verb). For example, the affix *-o'* in (12) indicates that the verbal predicate takes a 1st person *actor* and a 3rd person *undergoer*. This portmanteau morpheme cannot be further decomposed into an actor morpheme and an undergoer morpheme. The event is annotated as having two arguments, since these are clearly identifiable.

The Arapaho example in (12) also has implications for the annotation of pronominal references cross-linguistically. AMR currently treats person and number lexically, inserting English lexical pronoun forms as concepts in the AMR graph, as "he" in (1). However, it is cross-linguistically useful for the semantic representation to be independent of whether a participant is referred to with pronouns and with morphological agreement marking [23, 40]. To make this consistent, we encode reduced reference with a named entity type (such as "person") and add additional referential information such as person or number through the use of additional ":ref" attributes. Person and number lattices are being defined according to typological study [25], and also use ":ref" to encode definiteness, obviation, and language-specific referential information such as grammatical gender.

There are different types of noun incorporation [54, 55], where one word expresses both an event and a participant. In less grammaticalized constructions, as in (12) above, the incorporated noun (*-e'ei* 'head') functions as an argument of the verb, meaning that no overt NP can fill this semantic role. While UMR treats the whole word *teb-e'ei-s* as the predicate, it does posit a separate theme concept for 'head'. For more grammaticalized types of noun incorporation where the incorporated noun does not replace one of the verbal arguments, such as Arapaho instrumental suffixes like *-s* 'by blade' in (12), we do not posit a separate concept for the instrument.

Second, verb forms that differ in valency, e.g., **causatives**, **passives** and **applicatives**, are not decomposed into multiple concepts. Instead, their semantics can be inferred from the participant role annotations associated with the verb. For example, the difference between an intransitive verb and its causative (compare intransitive *nih-teb-e'ei-t* [PAST-break/remove.stick.like-head-3S] 'his head broke off' to the causative in (12)), are reflected by annotating only an Undergoer participant for the former, but both an Actor (the causer) and an Undergoer for the latter.

Third, some categories expressed by either separate words or verbal affixes will not be treated as separate concepts from the verb they modify. For example, **aspect**, which may be expressed by verbal morphology or auxiliaries, will simply be annotated with the aspect feature, such as the *Performance* value annotated for the verb in (12); see Section 4.1. The same is true of modal and tense constructions, which will inform the temporal and modal dependency annotations (see Sections 5.2 and 5.3, respectively). Certain semi-modals, such as *want*, and associated motion constructions will be identified as independent predicates when there is evidence that they are interpreted as independent events (e.g. when an associated motion construction can take locative or directional NPs as its arguments, or when a desiderative construction can be modalized with a modal scoping only over the desire). When such evidence is lacking, they will not be identified as independent events.

Example (13) illustrates associated motion in Sanapaná: the morphemes expressing motion are suffixed to the verb root 'see'. In the English translation in (14), arrival and motion are expressed as separate predicates. In (14), *arrive* can be modified by a locative phrase, as in *They arrived at the village...*. The Sanapaná associated motion construction in (13) can also occur with an NP expressing a location, but such NPs are likely better analyzed as circumstantial locatives expressing the location of the seeing-event than arguments of the arriving-event. The associated motion verb form in (13) is therefore represented as denoting a single complex event.

(13) netamen
afterwards
apk-el-vet-**angv**-ay-**akm**-e'
2/3M-DISTR-see-**LOC**-PST/HAB-**APPRX**-V1.NFUT
hlema nenhlet, ang-kelvana.
one person 2/3F-woman
'Afterwards, they arrived and saw a person, a woman.'

```
(v / engvetangvayam 'arrive (there) and see'
 :Aspect State
 :experiencer (p / person :ref 3pl)
 :stimulus (n / nenhlet 'person'
       :mod (a / angkelvana 'woman')
       :quant 1))
```

(14) Afterwards, they arrived and saw a person, a woman.

```
(a2 / and
 :op1 (a / arrive-01
     :Aspect Performance
     :ARG1 (p / person :ref 3pl))
 :op2 (s / see-01
     :Aspect State
     :ARG0 p
     :ARG1 (p2 / person
         :mod (w / woman)
         :quant 1))
```

Fourth, certain categories are expressed across languages by either derivational verbal morphology or by separate words. This includes types of "nonverbal clauses", such as locatives, nominals, property predication, and possession. Languages differ in how the strategies they use to express these meanings package concepts into words. There are three common strategies, two of which are problematic for the predicate-argument structure of existing meaning representations such as AMR. The cross-linguistic distribution of these strategies is based on our own research, and re-interpretation of the data in [70, 71].

In English examples, such as *John has a book* or *John is a doctor*, a predicate can easily be identified (*have* and *be*, respectively), and so can the NPs that function as its arguments. However, in the Kukama object predication in (15), the predicate does not map to a specific word: object predication is expressed through juxtaposition of two NPs, with the predicational meaning implicit, but inherent in the construction. In the Kukama thetic (presentational) possession construction in (16), the possessum and the possession relation are combined in a single word which functions as a predicate: something typically thought of as an "argument" is predicativized.

From a perspective of cross-linguistic portability, it is important that these different strategies are annotated in comparable ways, but from the perspective of ease of annotation, one may want to make allowances for annotators of individual languages.

(15) *ajan kunumi tsumi*
this young.man shaman
'This young man is a shaman.'

```
(h / have-role-91
   :ARG0 (k / kunumi 'young man')
   :ARG1 (t / tsumi 'shaman')
   :Aspect State
   :modstr Aff)
```

(16) *Mijiri-tin ɨara-yara*
Miguel-CER canoe-owner
'Miguel does have a canoe.' lit. 'Miguel is a canoe-owner.'

```
(e / ɨara-yara 'has canoe'
   :ARG0 (m / Mijiri 'Miguel')
   :ARG1 (i / ɨara 'canoe')
   :Aspect State
   :modstr Aff)
```

Different solutions are proposed for these two cases: in constructions with predicativized arguments, such as (16), a non-verbal clause function and an argument are identified and linked to the same word (since this word contains both the predication meaning and the argument-like meaning). When there is no overt predicate-word, such as in (15), we assume that annotators will be able to recognize the type of non-verbal clause function, and use an abstract predicate from Table 1. The resulting annotations have a comparable structure as seen above.

7 Strategies for applying UMR to all languages, including low-resource languages

7.1 A "roadmap" approach to achieve UMR annotation for low-resource languages

To allow annotation of indigenous minority languages, certain social and practical concerns must be dealt with too. Firstly, field linguists and native speaker communities must be convinced that semantic annotation can contribute to descriptive analysis, language documentation, and revitalization. It may be argued that semantic annotation can be a useful way of engaging with the meanings conventionalized in a language, and may bring up semantic questions that may not have otherwise come up. Annotated corpora can also aid in language documentation, by enriching (learners') dictionaries with frame files created during argument structure annotation.

Secondly, for many languages, annotators both highly proficient in the language and familiar with linguistic theory and computational tools may be scarce. Additionally, linguistic attitudes, such as cultural taboos against representing or sharing a language in written form, may further complicate annotation.

Thirdly, there is considerable cross-linguistic diversity in the availability of computational resources. AMR annotation of predicate-argument structure currently relies on the availability of a verbal lexicon with frame files. However, many languages do not have a standard dictionary or comprehensive grammatical description available, or even grammatical analysis or orthography.

UMR is being structured as a "road map" that (i) allows for flexibility depending on the material, technological, and linguistic resources available, and (ii) allows field linguists to base annotation on existing data collected in widely used software such as Flex and Tool-

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

Table 1: Nonverbal clause predicates

box, and to create resources such as PropBank-style frame files [59] from those data in a relatively short time frame. This will lower the threshold for linguists and communities to start doing semantic annotation.

We envision this road map as having two extreme points. Stage 0 will be a starting point for annotation of languages with few resources and little description. UMR annotation in this stage would be based on whole words (more precisely, stems whose inflections have been analyzed), without relying on lexical resources. Stage 1 will be a fully specified end point for annotation of languages with significant corpora, description, and other resources, taking advantage of morphological analysis and lexical resources comparable to PropBank's frame files. These should not be seen as discrete annotation stages, but instead languages can gradually move from Stage 0 to Stage 1 as resources are developed, and annotations at later stages will be compatible with those at the earlier stages. This road map proposal is treated in more depth in [78].

7.2 Lexical vs Non-lexical roles

UMR aims for a representation in which concepts in the graph are labeled with language-specific word senses, and where core roles of a predicate are defined using predicate-specific terms. However, we view this as a destination rather than a starting point, as many languages lack lexicons. To resolve this, we define frameworks for both non-lexicalized and lexicalized annotation of predicates and semantic roles, and propose that projects should expand lexical coverage during annotation.

For the non-lexicalized UMR predicates, the predicate is simply annotated with a lemmatized form and the arguments of that predicate are annotated using a general inventory of core participant roles given in Table 2, which extends AMR's non-core roles based on cross-linguistic argument realization patterns in ValPaL [41]. The first row of the following Arapaho example illustrates such a non-lexicalized annotation.

(17) he'ihnooko'wuuteen

he'ih-	nooko'wuutee	-ni
NARRPAST	white.streak.in/on.ground	OBV

The ground had a white mark/streak in it.

nooko'wuutee-01:
ground marked white
ARG1: ground

```
(n / nooko'wuutee
    :theme (x / thing
        :ref obviative)
    :Aspect State)

(n / nooko'wuutee-01
    :arg1 (x / thing
        :ref obviative)
    :Aspect State)
```

Transitioning from such non-lexicalized annotations to lexicalized semantic roles requires the construction of predicate-specific role definitions, as illustrated in the lower row of this example. This shift from a non-lexicalized representation to a lexicalized one is necessary for both establishing consistent argument annotations, and for slowly developing the language-specific conventions regarding multi-word expressions and the decomposition of multiconcept words. We suggest that these lexical entries should be mapped to non-lexicalized roles (i.e. the general inventory of participant roles), to enable automatic or semi-automatic conversion of existing annotations to the lexicalized form. The end result will be a purely lexicalized annotation.

While the same road map approach might be adopted for entity typing and word senses, we expect that word senses will be defined for individual languages in UMR annotation. While the general inventory of entity types currently used in AMR can be reframed into a more cross-linguistically robust form, we leave that to future work.

8 Pilot annotation experiments

As of now, we have not been able yet to conduct significant amounts of annotation with the full UMR scheme in any language. We have, however, conducted a number of annotation experiments aimed specifically at testing the robustness of the proposed UMR annotation

Central roles	Actor, Undergoer, Theme, Recipient, Force, Causer, Experiencer, Stimulus
Peripheral roles	Instrument, Companion, Material/Source, Place, Start, Goal, Affectee
Roles for entities and events	Cause, Manner, Reason, Purpose, Temporal, Extent

Table 2: UMR non-lexical roles

schemes for newly added semantic domains (temporality, modality, and aspect), and at evaluating the efficiency of lattice-style annotation schemes in maintaining cross-linguistic comparability. These experiments were all conducted in English (annotation pilots of Arapaho, Kukama, Navajo, and Sanapaná are under way). As long as these new semantic domains can be reliably annotated, we are confident that UMR as a whole can be reliably annotated as it builds on AMR which has been successfully annotated in large-scale annotation projects [6], and the typological adaptations are based on large-scale cross-linguistic studies of the categories and constructions described above.

Experiments on temporal dependency annotation have been carried out in a crowd-sourcing setting with untrained annotators on Amazon Mechanical Turk [86]. This is arguably a more challenging annotation setting as it is impractical to ensure that the annotators have proper linguistic background and require that they read detailed guidelines. In our temporal dependency annotation experiments, for each relative time expression (e.g., *today*), the crowd-worker is asked to identify its *reference time* so that it can be resolved to an absolute time that can be properly interpreted. For each event, the crowd-worker is asked to identify either a *reference event* or a *reference time* and then determine the temporal relations (e.g., *Before*, *After*) between them. We measure the agreement between an expert annotator and the aggregated majority opinion of crowd-workers. The agreement is measured separately for events and time expressions, and we use two metrics, unlabeled and labeled F1 score. Unlabeled F1 score measures whether the annotators agree on the reference time or reference event, while labeled F1 is a more stringent metric that also measures whether the same temporal relation is identified. Assuming that the time expressions and events are properly identified, which is a reasonable assumption in UMR annotation as the sentence-level annotation is already in place, the unlabeled and labeled agreement is 0.85 and 0.77 for time expressions and 0.75 and 0.83 for events. A more detailed breakdown shows that annotators can more reliably determine temporal relations but find it more challenging to agree on the same reference time or event. Since UMR annotation is designed for "traditional" annotation approaches that assume detailed annotation guidelines and trained annotators, we expect annotation agreement on temporal dependency can only improve in such a setting.

The most extensive annotation experiment was conducted for the modal dependency annotation scheme. As reported in [77], six English texts were annotated by two independent expert annotators using this scheme, amounting to 377 events expressed in 108 sentences. In the first pass of annotation, the identification of events to be given a modal value, inter-annotator agreement scores were very high (precision: 0.94, recall: 0.93, F-score: 0.93). For the second annotation pass, setting up the modal superstructure (the relations between conceivers), precision was high (0.91) but recall was much lower (0.77). The F-score for this pass was 0.83. For the third pass, the attribution of modal strength values to each event, inter-annotator agreement was once again high (precision, recall, and F-score all 0.88). Even though there were significant and interesting genre-based differences in inter-annotator agreement - agreement was consistently much higher for events from newswire text than for events from messages on discussion forums - these results show that the annotation scheme can be successfully implemented.

To test the robustness of the aspect annotation, secondly, five English texts were annotated by the same two expert annotators as in the modal annotation experiment described above. A total of 238 events were given one of six aspect labels: **State**, **Habitual**, **Activity**, **Endeavor**, **Performance**, or **N/A** (the latter was used for event nominals and future events, and changed to **Process** later). Inter-annotator agreement for this task was slightly lower than for the modal annotation, but still much higher than chance: Cohen's K = 0.84; Siegel & Castellan's adjusted K to account for bias = 0.84; 2 * observed proportion of agreement - 1 = 0.80. Agreement was again considerably higher in news text than in narrative text - for one news text there was perfect agreement. Once again, these agreement results are encouraging. Both for the modal and the aspectual annotation, these results await confirmation from annotation in other languages.

Lastly, a short cross-linguistic annotation experiment was conducted to gauge whether the organization of semantic categories in a lattice leads to higher cross-linguistic inter-annotator agreement [75]. Thirty-six English sentences expressing spatial relations and their Czech, Dutch, and Korean translations were annotated, each by one native speaker of the language in question. Annotators used a lattice with **support**, **attachment**, **containment** as the default level values. On the level above, **non-containment** grouped together **support** and **attachment**, while **non-support** grouped together **attachment** and **containment**. More fine-grained values were **adhesion** (intermediate between support and attachment, e.g. a band-aid on an arm), and **attached containment** (intermediate between attachment and containment, e.g. an apple on a tree branch, where it is also enveloped by leaves).

A fairly large proportion of sentences were annotated with an identical value across languages: for the lowest-scoring language pair (Czech and English), Cohen's K was 0.64 for exact agreement, while for the highest-scoring language pair (Czech and Dutch), Cohen's K was 0.86. However, when looking at sentences with *compatible* annotations, rather than *identical* ones (e.g. an event annotated as **attachment** in one language but **non-containment** in another), scores went up considerably: the lowest-scoring language pair was now Czech and Korean (Cohen's K = 0.79), while the highest scoring language pair was now Dutch and Korean (Cohen's K = 0.94). Therefore, the proposed lattice architecture seems fairly successful at abstracting away from language-specific differences in category boundaries. These results as well await further confirmation from cross-linguistic annotation of different semantic domains.

9 Related work

Though necessarily somewhat superficially, we discuss the design choices of UMR in relation to five existing meaning representations: Abstract Meaning Representation (AMR) [6], Discourse Representation Structures (DRS) [44, 13], the tectogrammatical layer of the Prague Dependency TreeBank (PDT) [37, 38], Minimal Recursion Semantics (MRS) [21], and Universal Conceptual Cognitive Annotation (UCCA) [1]. These five meaning representations are selected as a basis for comparison because: i) they have all been deployed in large-scale annotation projects; ii) they all provide some degree of abstraction from surface text spans, and iii) they provide a complete meaning representation for at least the entire sentence if not the whole text. These requirements preclude from consideration partial meaning representations such as semantic role labeling frameworks like FrameNet [5] and Propbank [59] where the focus is on the argument structure of verbal and nominal predicates. Of course, the meaning representations we have chosen for comparative discussion are not intended to be a complete list of semantic representations that have been proposed over the years.

Abstract Meaning Representation (AMR) [6], firstly, represents sentence meanings as single-rooted, directed, and acyclic graphs in which the nodes are concepts and the edges are relations. AMR concepts include word sense-disambiguated verbal and nominal predicates, entities (e.g. "person") and relations (e.g.,"have-org-role-91"), or simple lemmas. AMR relations include Propbank style semantic roles (e.g., "Arg0"), general semantic relations (":degree"), discourse relations (e.g., ":condition"), etc.

Discourse Representation Theory (DRT) [44] proposes a discourse-level meaning representation for an entire text that can be easily translated into logical form. The Groningen Meaning Bank (GMB) [13] is a large data set annotated with DRS that makes use of word senses from the WordNet, semantics roles from VerbNet, and rhetorical relations from SDRT [3] and puts the theoretical foundation of DRT into practice. The GMB is produced by associating Combinatory Categorial Grammar (CCG) parses [72] with semantic forms. Semantic forms of the entire sentence can be constructed from primitive semantic forms associated with lexical entries in the CCG lexicon. As the CCG tree of a sentence is constructed so is its semantic representation. Parallel Meaning Bank (PMB) [2] is a more recent effort to extends GMB annotation to multiple languages to create a parallel corpus annotated with DRS.

The tectogrammatical layer of the Prague Dependency TreeBank (PDT) [37, 38] covers many of the same semantic distinctions covered by AMR such as the argument structure (semantic roles, called "functors" in PDT), word senses, coreference, and intra- and inter-sentential discourse relations. Additionally it also annotates tense, modalities, and a host of other "semantic" node attributes, bridging and textual coreference as well as topic/focus (information structure) which are not part of AMR annotation. PDT uses a multi-layered annotation framework where the tectogrammatical layer is explicitly linked by individual node references to the other (lower) layers of linguistic analysis.

Minimal Recursion Semantics (MRS) [21] is a sentence-level meaning representation that also focuses on representing predicate-argument structure, sense distinctions where they are grammaticalized, logical semantic phenomena such as quantification and operator-like scopal predicates, and tense, aspect, modality etc. as determined by morpho-syntax. MRS emphasizes semantic compositionality [20, 9], and full representations are typically derived in conjunction with grammar-based parsers, e.g. the English Resource Grammar (ERG) [33].

Universal Conceptual Cognitive Annotation (UCCA) [1] has a foundational layer that focuses on the predicate-argument structure. The UCCA foundational layer views text as a collection of *scenes*, and each scene contains a main relation (a state or process) that is the anchor of the scene, as well as participants of the relation. As it currently stands, UCCA does not annotate word senses, named entities, relations as AMR does, nor does it annotate tense, aspect, modality, and quantification scope like MRS. However, it has a multi-layered design like PDT that allows extensions, and there is ongoing research to add coreference annotation to UCCA [60].

We compare UMR with these existing meaning representations against the four desiderata we have outlined in Section 2, as well as whether they support discourse-level semantic processing.

9.1 Scalability

Scalability is a key consideration in the design of UMR, since it needs to be applicable in large-scale annotation settings. We argue that representations that can be annotated independently of other layers of linguistic analysis – particularly syntactic annotation, as this is a highly complex task in itself – have an advantage in this regard. Such independence from syntactic layers of annotation not only improves the scalability of an annotation system within one language, but also improves cross-linguistic portability by making semantic annotation possible for languages which may not have syntactically annotated corpora, nor the resources to create them. Ultimately, we aim to train accurate semantic parsers which is only possible given sufficient amounts of training data.

AMR relaxes the strict correspondence between the meaning representation and morphosyntax: the concepts and relations in AMR need not be linked to constituents in a morphosyntactic structure, or even sub-segments of the surface linguistic signal. This contributes to scalability by eliminating the time needed to build morphosyntactic structures preceding semantic annotation. It allows great freedom in handling syntax-semantics mismatches, such as "contentless" function words which can be left out of the meaning representation (e.g. infinitival *to* in English), constructs in the meaning representation which do not correspond to any words in the text but can be inferred from the context (e.g. a concept "person" can be inferred from the surface phrase "the young"), and complicated correspondences between the meaning representation and the surface syntactic structure (e.g. discontinuous constructions such as English "as $\cdots$ as $\cdots$", which is collapsed into a single AMR relation *:compared-to*).

Like AMR, UCCA is also a stand-alone meaning representation that does not have to be linked to a syntactic representation. In addition, like AMR it also allows the annotation of *implicit arguments*, arguments that do are not lexicalized. In this sense, it has the same advantage as AMR in terms of scalability. In fact, it is shown that shows that UCCA does not require linguistically trained annotators [1], adding to its attractiveness. However, UCCA currently lacks many of the semantic elements that other meaning representation has, and these include named entities and relations between them, logical constructs such as tense, aspect, modality, polarity, and quantification scope, so it remains to be seen if the simplificity (thus scalability) of the UCCA foundational layers can be maintained if it is extended to account for these additional semantic elements.

The three other semantic representation frameworks, except in part for the PDT which keeps the layers separated but interlinked, are more tightly tied to syntactic annotation and harder to annotate independently. Minimal Recursion Semantics (MRS) [21] and Discourse Representation Theory [45] are not intended to be created directly through manual annotation, and they are typically produced compositionally via the mediation of syntactic structures. Examples of large-scale initiatives in producing these meaning representations include the Lingo Redwoods Initiative [57] and the Groningen Meaning Bank (GMB) [8], respectively. The goal of the GMB is to generate Discourse Representation Structures (DRS), mediated with syntactic structures produced by Combinatory Categorial Grammar [72] parsers. The Lingo Redwoods Initiative adopts a similar approach in generating MRS representations via Head-Driven Phrase Structure [65] parses.

UMR inherits the approach of AMR by allowing meaning representation to be annotated independently of syntactic structures. This makes it possible to annotate the meaning of morphologically complex low-resource languages without having to tackle their morphosyntactic complexities and so it is easier for semantic annotation to get off the ground more quickly, and thus more scalable. However, see [9] for a different perspective on this trade-off, where it is argued that basing semantic annotation off syntactic annotation makes it more likely to be consistent. We believe that with detailed guidelines that specify when a new concept can be inferred and when a discontinuous pattern can be mapped to a single concept/relation, annotation consistency can be achieved.

9.2 Supporting both lexical and logical inference

Minimal Recursion Semantics has the clear goal of supporting *logical* inference in reasoning-based AI systems: it is easily translatable into predicate logic. Much of their focus is on proper representation of semantic components such as quantification, negation, tense, and modality. Less emphasis is placed on the representation of lexical semantic information such as semantic roles and word senses, or entities and relations, which are prioritized in AMR.

On the other end of the spectrum are meaning representations aimed at supporting similarity-based *lexical* inferences. These are crucial to modern practical natural language understanding applications, and include distinguishing different senses of words (e.g. *John ran a race* versus *John runs a company*), and identifying which entities play which semantic roles in events (e.g. *John* is the Actor (Arg0) of both examples above). Both AMR and the Tectogrammatical (TG) annotation in the PDT family of treebanks [52, 38, 39] focus on facilitating such inferences. Similarly to AMR, every verb in the TG annotation is sense-disambiguated and linked to a lexical entry in a valency lexicon of the language being annotated [36, 73]. These lexical entries are further linked to semantic classes of verbal synonyms [74] and consequently thus also to entries in other lexical resources, such as FrameNet [5], VerbNet [68], Propbank [59] and WordNet [53]. Several TG-annotated corpora are available, including the parallel Prague Czech-English Dependency Treebank PTB/WSJ corpus [37, 38, 39].

UCCA does not currently annotate word senses, but it achieves "semantic stability" by annotating scenes that consists of participants in a relation. As a result, it is indifferent to variations as a result of paraphrasing. In this sense, it supports lexial inference. As it currently does not annotate elements that are important to logical inference, in this aspect it is more in line with AMR and the tectogrammatical layer of PDT.

Discourse Representation Theory sits somewhere in the middle in the spectrum. It has a clear focus on supporting logical inference. At the same time, it also draws from WordNet [53] and VerbNet [68] in that it uses the word sense information in the former and the semantic role labels in the latter.

UMR adds the annotation of quantifier scope and negation to AMR to support logical inference. [12] has shown that with disambiguated quantifier scope, AMR can be translated into first-order logic. Apart from scope, UMR adds layers with temporal, aspectual, and modal information to AMR, as described in Section 5. Such additional layers on top of AMR expand its capability for logical inference.

9.3 Achieving cross-linguistic uniformity

For a meaning representation to be uniform across languages, it must provide mechanisms for abstracting away from language-specific morphosyntax without losing semantic information. An advantage of AMR is that it already abstracts away from constituent order which varies widely across languages. This has been shown to make AMR more portable across languages, for example to Chinese and Czech [85, 47, 48]. UCCA, by intentionally attempting to annotate semantic relations that are indifferent to syntactic relations, have also proved to be readily portable to other languages [43]. Similarly, DRS, MRS, and PDT have all been annotated in multilingual contexts [2, 51, 39], and have all demonstrated some level of cross-lingual validity.

To our knowledge, however, UMR is the first to explicitly take low-resource languages into account and is general enough to eventually encompass the range of variation exhibited by the world's roughly 7000 languages. Unlike Interlingua [56, 29, 30, 42], UMR does not aim to build a common cross-linguistic vocabulary for all concepts. Instead, UMR aims to **factor out what is common for all languages** - roughly, it proposes the use of language-specific lexical databases and Propbank-style frame files, and cross-linguistically portable annotation layers for "grammatical" semantics such as aspect and modality. UMR thus uses a combination of language-specific (concrete, lexical) concepts and a shared inventory of (abstract, grammatical) concepts, features, and relations for all languages. This shared inventory is expected to have a manageable vocabulary size of a few hundred items. As it is based on cross-linguistic typological research from the last six decades, UMR can represent robust cross-linguistic commonalities in a uniform manner while allowing for cross-linguistic variation.

9.4 Document-level representation

We have shown in Section 5 that it is impossible to properly interpret the meaning of a text without representing document-level language phenomena. Existing meaning representations are either sentence-level representations (MRS), or have mostly focused on coreference and limited discourse relations across sentences (the tectogrammatical layer of PDT). For example, the Groningen Meaning Bank, based on DRS, includes coreference annotation and rhetorical relations based on Segmented Discourse Representation Theory (SDRT) [3]. The Multi-Sentence AMRs [58] focus on various forms of coreference annotation. UCCA annotation can go over sentence boundaries and can include several paragraphs, and there is also research to add coreference annotation to UCCA. UMR advances the state of the art by providing a robust document-level representation that represents coreference, temporal and modal relations.

10 Conclusion and future work

In this paper we presented the design of a Uniform Meaning Representation (UMR) that draws on existing meaning representations while substantially extending them. To make UMR cross-linguistically plausible, we proposed a staged strategy for annotating morphologically complex low-resource languages with UMR. We also proposed additional features to UMR, and a document-level representation that captures semantic relations that potentially go beyond sentence-boundaries.

Designing a meaning representation that captures the essential semantic content of a text is crucial to making progress in natural language understanding. While each aspect of the UMR annotation has been performed individually to some extent in prior work, packing all of them in one simple (and thus practical and scalable) framework is still an incredible challenge. We believe UMR has taken a significant step in that direction and hope to produce UMR annotations for many languages. We are currently developing UMR annotation guidelines and tools, and as they become available, plan to annotate UMR on data from multiple languages.

References

1. Abend O, Rappoport A (2013) UCCA. A semantics-based grammatical annotation scheme. In: Proceedings of the 10th International Conference on Computational Semantics, Potsdam, Germany, p 1 – 12
2. Abzianidze L, Bjerva J, Evang K, Haagsma H, van Noord R, Ludmann P, Nguyen DD, Bos J (2017) The Parallel Meaning Bank: Towards a multilingual corpus of translations annotated with compositional meaning representations. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, Valencia, Spain, pp 242–247
3. Asher N, Asher NM, Lascarides A (2003) Logics of conversation. Cambridge University Press
4. Bahdanau D, Cho K, Bengio Y (2015) Neural machine translation by jointly learning to align and translate. In: Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)
5. Baker CF, Fillmore CJ, Lowe JB (1998) The berkeley framenet project. In: Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1, Association for Computational Linguistics, pp 86–90
6. Banarescu L, Bonial C, Cai S, Georgescu M, Griffitt K, Hermjakob U, Knight K, Koehn P, Palmer M, Schneider N (2013) Abstract meaning representation for sembanking. In: Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse, pp 178–186
7. Barker C (2002) Continuations and the nature of quantification. Natural language semantics 10(3):211–242
8. Basile V, Bos J, Evang K, Venhuizen N (2012) Developing a large semantically annotated corpus. In: LREC, vol 12, pp 3196–3200
9. Bender EM, Flickinger D, Oepen S, Packard W, Copestake A (2015) Layers of interpretation: On grammar and compositionality. In: Proceedings of the 11th International Conference on Computational Semantics, London, UK, p 239 – 249
10. Bentivogli L, Bisazza A, Cettolo M, Federico M (2016) Neural versus phrase-based machine translation quality: a case study. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, Texas, pp 257–267
11. Bojar O, Chatterjee R, Federmann C, Graham Y, Haddow B, Huck M, Jimeno Yepes A, Koehn P, Logacheva V, Monz C, Negri M, Neveol A, Neves M, Popel M, Post M, Rubino R, Scarton C, Specia L, Turchi M, Verspoor K, Zampieri M (2016) Findings of the 2016 conference on machine translation. In: Proceedings of the First Conference on Machine Translation, Berlin, Germany, pp 131–198
12. Bos J (2016) Expressive power of Abstract Meaning Representations. Computational Linguistics 42(3):527–535
13. Bos J, Basile V, Evang K, Venhuizen N, Bjerva J (2017) The Groningen Meaning Bank. In: Ide N, Pustejovsky J (eds) Handbook of Linguistic Annotation, vol 2, Springer, pp 463–496
14. Boye K (2012) Epistemic meaning: A crosslinguistic and functional-cognitive study, Empirical Approaches to Language Typology, vol 43. De Gruyter Mouton, Berlin
15. Cai D, Lam W (2020) AMR parsing via graph-sequence iterative inference. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, pp 1290–1301, DOI 10.18653/v1/2020.acl-main.119
16. Cai Q, Yates A (2013) Large-scale semantic parsing via schema matching and lexicon extension. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), vol 1, pp 423–433
17. Cai Q, Yates A (2013) Semantic parsing freebase: Towards open-domain semantic parsing. In: Second Joint Conference on Lexical and Computational Semantics (* SEM), Volume 1: Proceedings of the Main Conference and the Shared Task: Semantic Textual Similarity, vol 1, pp 328–338
18. Castilho S, Moorkens J, Gaspari F, Sennrich R, Sosoni V, Georgakopoulou Y, Lohar P, Way A, Barone AVM, Gialama M (2017) A Comparative Quality Evaluation of PBSMT and NMT using Professional Translators. In: In Proceedings of Machine Translation Summit XVI, Nagoya, Japan
19. Choe H, Han J, Park H, Kim H (2019) Copula and

case-stacking annotations for korean amr. In: Proceedings of the First International Workshop on Designing Meaning Representations, pp 128–135
20. Copestake A, Lascarides A, Flickinger D (2001) An algebra for semantic construction in constraint-based grammars. In: Proceedings of the 39th Meeting of the Association for Computational Linguistics, Toulouse, France, p 140 – 147
21. Copestake A, Flickinger D, Pollard C, Sag IA (2005) Minimal Recursion Semantics: An introduction. Research on Language and Computation 3(2-3):281–332
22. Croft W (2012) Verbs, Aspect and Causal Structure. Oxford University Press
23. Croft W (2013) Agreement as anaphora, anaphora as coreference. Languages across boundaries: Studies in memory of Anna Siewierska 95:117
24. Croft W, Pešková P, Regan M (2017) Integrating decompositional event structures into storylines. In: Proceedings of the Events and Stories in the News Workshop, Association for Computational Linguistics, Vancouver, Canada, pp 98–109, DOI 10.18653/v1/W17-2713
25. Cysouw M (2003) The paradigmatic structure of person marking. Oxford University Press
26. Dixon RM, Aikhenvald AY, et al. (2002) Word: A cross-linguistic typology, chap Word: a typological framework, pp 1–41
27. Donatelli L, Regan M, Croft W, Schneider N (2018) Annotation of tense and aspect semantics for sentential AMR. In: Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Constructions (LAW-MWE-CxG-2018), pp 96–108
28. Donatelli L, Schneider N, Croft W, Regan M (2019) Tense and aspect semantics for sentential AMR. In: Proceedings of the Society for Computation in Linguistics, vol 2, pp 346–348
29. Dorr BJ (1993) Machine translation: a view from the Lexicon. MIT press
30. Dorr BJ (1994) Machine translation divergences: A formal description and proposed solution. Computational Linguistics 20(4):597–633
31. Eriguchi A, Tsuruoka Y, Cho K (2017) Learning to Parse and Translate Improves Neural Machine Translation. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Vancouver, Canada, pp 72–78
32. Flanigan J, Thomson S, Carbonell JG, Dyer C, Smith NA (2014) A discriminative graph-based parser for the abstract meaning representation. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp 1426–1436
33. Flickinger D (2000) On building a more efficient grammar by exploiting types. Natural Language Engineering 6 (1):15 – 28
34. Flickinger D, Bender EM, Oepen S (2014) Towards an Encyclopedia of Compositional Semantics: Documenting the Interface of the English Resource Grammar. In: LREC, pp 875–881
35. Ge R, Mooney RJ (2005) A statistical semantic parser that integrates syntax and semantics. In: Proceedings of the ninth conference on computational natural language learning, pp 9–16
36. Hajič J, Panevová J, Urešová Z, Bémová A, Kolářová V, Pajas P (2003) PDT-VALLEX: Creating a large-coverage valency lexicon for treebank annotation. In: Nivre J, Hinrichs E (eds) Proceedings of The Second Workshop on Treebanks and Linguistic Theories, Vaxjo University Press, Vaxjo, Sweden, vol 9, pp 57–68
37. Hajič J, Panevová J, Hajičová E, Sgall P, Pajas P, Štěpánek J, Havelka J, Mikulová M, Žabokrtský Z, Ševčíková-Razímová M, Urešová Z (2006) Prague Dependency Treebank 2.0 (PDT 2.0). URL `http://hdl.handle.net/11858/00-097C-0000-0001-B098-5`
38. Hajič J, Bejček E, Bémová A, Buráňová E, Hajičová E, Havelka J, Homola P, Kárník J, Kettnerová V, Klyueva N, Kolářová V, Kučová L, Lopatková M, Mikulová M, Mírovský J, Nedoluzhko A, Pajas P, Panevová J, Poláková L, Rysová M, Sgall P, Spoustová J, Straňák P, Synková P, Ševčíková M, Štěpánek J, Urešová Z, Vidová Hladká B, Zeman D, Zikánová Š, Žabokrtský Z (2018) Prague dependency treebank 3.5. URL `http://hdl.handle.net/11234/1-2621`
39. Hajič J, Hajičová E, Panevová J, Sgall P, Bojar O, Cinková S, Fučíková E, Mikulová M, Pajas P, Popelka J, Semecký J, Šindlerová J, Štěpánek J, Toman J, Urešová Z, Žabokrtský Z (2012) Announcing Prague Czech-English Dependency Treebank 2.0. In: Proceedings of the eighth international conference on Language Resources and Evaluation, Istanbul, pp 3153–3160
40. Haspelmath M (2013) Argument indexing: A conceptual framework for the syntactic status of bound person forms. Languages across boundaries: Studies in memory of Anna Siewierska 197:226
41. Haspelmath M, Hartmann I (2015) Comparing verbal valency across languages. Valency classes in the worlds languages 1:41–72
42. Helmreich S, Farwell D, Dorr B, Habash N, Levin L, Mitamura T, Reeder F, Miller K, Hovy E, Rambow O, et al. (2004) Interlingual annotation of multilingual text corpora. In: Proc. of the HLT-EACL Workshop on Frontiers in Corpus Annotation

43. Hershcovich D, Aizenbud Z, Choshen L, Sulem E, Rappoport A, Abend O (2019) SemEval-2019 task 1: Cross-lingual semantic parsing with UCCA. In: Proceedings of the 13th International Workshop on Semantic Evaluation, Minneapolis, Minnesota, USA, pp 1–10, DOI 10.18653/v1/S19-2001
44. Kamp H, Reyle U (1993) From discourse to logic: Introduction to model theoretic semantics of natural language, formal logic and discourse representation theory. Kluwer, Dordrecht
45. Kamp H, Reyle U (2013) From discourse to logic: Introduction to model-theoretic semantics of natural language, formal logic and discourse representation theory, vol 42. Springer Science & Business Media
46. Kate RJ, Mooney RJ (2007) Learning language semantics from ambiguous supervision. In: AAAI, vol 7, pp 895–900
47. Li B, Wen Y, QU W, Bu L, Xue N (2016) Annotating The Little Prince with Chinese AMRs. In: Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016 (LAW-X 2016), Association for Computational Linguistics, pp 7–15, DOI 10.18653/v1/W16-1702
48. Li B, Wen Y, Song L, Qu W, Xue N (2019) Building a Chinese AMR Bank with Concept and Relation Alignments. LiLT (Linguistic Issues in Language Technology) 18
49. Liang P, Jordan MI, Klein D (2009) Learning semantic correspondences with less supervision. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1, pp 91–99
50. Lyu C, Titov I (2018) AMR parsing as graph prediction with latent alignment. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia, pp 397–407, DOI 10.18653/v1/P18-1037
51. Marimon M (2010) The spanish resource grammar. In: Chair) NCC, Choukri K, Maegaard B, Mariani J, Odijk J, Piperidis S, Rosner M, Tapias D (eds) Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), European Language Resources Association (ELRA), Valletta, Malta
52. Mikulová M, Bémová A, Hajič J, Hajičová E, Havelka J, Kolářová V, Kučová L, Lopatková M, Pajas P, Panevová J, Razímová M, Sgall P, Štěpánek J, Urešová Z, Veselá K, Žabokrtský Z (2006) Annotation on the tectogrammatical level in the Prague Dependency Treebank. Annotation manual. Tech. Rep. 30, Institute of Formal and Applied Linguistics, Charles Univ., Prague, Czech Rep.
53. Miller GA (1998) WordNet: An electronic lexical database. MIT press
54. Mithun M (1984) The evolution of noun incorporation. Language 60:847–94
55. Mithun M (2015) Morphological complexity and language contact in languages indigenous to north america. Linguistic Discovery 13(2):37–59
56. Nyberg EH, Mitamura T (1992) The KANT system: Fast, accurate, high-quality translation in practical domains. In: Proceedings of the 14th conference on Computational linguistics-Volume 3, pp 1069–1073
57. Oepen S, Flickinger D, Toutanova K, Manning CD (2004) Lingo redwoods. Research on Language and Computation 2(4):575–596
58. O'Gorman, Tim and Regan, Michael and Griffitt, Kira and Hermjakob, Ulf and Knight, Kevin and Palmer, Martha (2018) AMR beyond the sentence: the Multi-sentence AMR corpus. In: Proceedings of the 27th International Conference on Computational Linguistics, pp 3693–3702
59. Palmer M, Gildea D, Kingsbury P (2005) The Proposition Bank: An annotated corpus of semantic roles. Computational linguistics 31(1):71–106
60. Prange J, Schneider N, Abend O (2019) Semantically constrained multilayer annotation: The case of coreference. In: Proceedings of the First International Workshop on Designing Meaning Representations, Florence, Italy, pp 164–176, DOI 10.18653/v1/W19-3319
61. Pustejovsky J, Castano JM, Ingria R, Sauri R, Gaizauskas RJ, Setzer A, Katz G, Radev DR (2003) TimeML: Robust specification of event and temporal expressions in text. New directions in question answering 3:28–34
62. Pustejovsky J, Xue N, Lai K (2019) Modeling quantification and scope in abstract meaning representations. In: Proceedings of the First International Workshop on Designing Meaning Representations, pp 28–33
63. Rajpurkar P, Zhang J, Lopyrev K, Liang P (2016) SQuAD: 100,000+ questions for machine comprehension of text. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, Texas, pp 2383–2392, DOI 10.18653/v1/D16-1264
64. Reddy S, Täckström O, Petrov S, Steedman M, Lapata M (2017) Universal semantic parsing. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, pp 89–101, DOI 10.18653/v1/D17-1009
65. Sag IA, Wasow T, Bender EM, Sag IA (1999) Syntactic theory: A formal introduction, vol 92. Center for the Study of Language and Information Stan-

ford, CA
66. Sag IA, Baldwin T, Bond F, Copestake A, Flickinger D (2002) Multiword expressions: A pain in the neck for NLP. In: International conference on intelligent text processing and computational linguistics, Springer, pp 1–15
67. Saurí R, Pustejovsky J (2009) FactBank: a corpus annotated with event factuality. Language Resources and Evaluation 43:227–268
68. Schuler KK (2005) Verbnet: A broad-coverage, comprehensive verb lexicon. PhD thesis, University of Pennsylvania
69. Smith N (2017) Squashing computational linguistics, invited talk at the 55th acl. `https://homes.cs.washington.edu/~nasmith/slides/acl-8-1-17.pdf`
70. Stassen L (1997) Intransitive Predication. Oxford University Press, Oxford, UK
71. Stassen L (2009) Predicative Possession. Oxford University Press, Oxford, UK
72. Steedman M (2000) The syntactic process, vol 24. MIT press Cambridge, MA
73. Urešová Z, Fučíková E, Šindlerová J (2016) CzEngVallex: a bilingual Czech-English valency lexicon. The Prague Bulletin of Mathematical Linguistics 105:17–50
74. Ureov Z, Fukov E, Hajiov E, Haji J (2020) Synsemclass linked lexicon: Mapping synonymy between languages. In: Proceedings of the 2020 Globalex Workshop on Linked Lexicography (LREC 2020), European Language Resources Association, Marseille, France, pp 10–19
75. Van Gysel JEL, Vigus M, Kalm P, Lee Sk, Regan M, Croft W (2019) Cross-linguistic semantic annotation: Reconciling the language-specific and the universal. In: Proceedings of the First International Workshop on Designing Meaning Representations, Florence, Italy, pp 1–14, DOI 10.18653/v1/W19-3301
76. Vendler Z (1967) Linguistics in philosophy, Cornell University Press, Ithaca, chap Verbs and times, pp 97–121
77. Vigus M, Van Gysel JE, Croft W (2019) A dependency structure annotation for modality. In: Proceedings of the First International Workshop on Designing Meaning Representations, pp 182–198
78. Vigus M, Van Gysel JEL, O'Gorman T, Cowell A, Vallejos R, Croft W (2020) Cross-lingual annotation: a road map for low- and no-resource languages. In: Proceedings of the Second International Workshop on Designing Meaning Representations, Barcelona Spain (online), pp 30–40, URL `https://www.aclweb.org/anthology/2020.dmr-1.4`
79. Wang C, Xue N, Pradhan S (2015) A transition-based algorithm for AMR parsing. In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp 366–375
80. Wang Z, Mi H, Hamza W, Florian R (2016) Multi-perspective context matching for machine comprehension. arXiv preprint arXiv:161204211
81. Wong YW, Mooney RJ (2006) Learning for semantic parsing with statistical machine translation. In: Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, pp 439–446
82. Wu S, Zhang D, Yang N, Li M, Zhou M (2017) Sequence-to-dependency neural machine translation. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vancouver, Canada, pp 698–707
83. Xiong C, Zhong V, Socher R (2016) Dynamic coattention networks for question answering. arXiv preprint arXiv:161101604
84. Xue N, Zhong H, Chen KY (2008) Annotating 'tense' in a tense-less language. In: LREC, Marrakech, Morocco
85. Xue N, Bojar O, Hajič J, Palmer M, Urešová Z, Zhang X (2014) Not an interlingua, but close: Comparison of English AMRs to Chinese and Czech. In: Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC)
86. Yao J, Qiu H, Min B, Xue N (2020) Annotating temporal dependency graphs via crowdsourcing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp 5368–5380
87. Zhang S, Ma X, Duh K, Van Durme B (2019) AMR parsing as sequence-to-graph transduction. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, pp 80–94, DOI 10.18653/v1/P19-1009
88. Zhang Y, Xue N (2018) Structured interpretation of temporal relations. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), European Language Resources Association (ELRA), Miyazaki, Japan
89. Zingler T (2020) Wordhood issues: typology and grammaticalization. PhD thesis, University of New Mexico