Mapping AMR to UMR: Resources for Adapting Existing Corpora for Cross-Lingual Compatibility

Julia Bonn[1], **Skatje Myers**[1], **Jens E. L. Van Gysel**[2], **Lukas Denk**[2], **Meagan Vigus**[2], **Jin Zhao**[3], **Andrew Cowell**[1], **William Croft**[2], **Jan Hajič**[4], **James H. Martin**[1], **Alexis Palmer**[1], **Martha Palmer**[1], **James Pustejovsky**[3], **Zdenka Urešová**[4], **Rosa Vallejos**[2], **Nianwen Xue**[3]

[1]University of Colorado, Boulder, [2]University of New Mexico
[3]Brandeis University, [4]Charles University, Prague
Corresponding author: julia.bonn@colorado.edu

Abstract

This paper presents detailed mappings between the structures used in Abstract Meaning Representation (AMR) and those used in Uniform Meaning Representation (UMR). These structures include general semantic roles, rolesets, and concepts that are largely shared between AMR and UMR, but with crucial differences. While UMR annotation of new low-resource languages is ongoing, AMR-annotated corpora already exist for many languages, and these AMR corpora are ripe for conversion to UMR format. Rather than focusing on semantic coverage that is new to UMR (which will likely need to be dealt with manually), this paper serves as a resource (with illustrated mappings) for users looking to understand the fine-grained adjustments that have been made to the representation techniques for semantic categories present in both AMR and UMR.

1 Introduction

Even with the overwhelming improvement in performance in Natural Language Processing (NLP) brought about by recent transformer architectures (Vaswani et al., 2017; Radford et al., 2018; Liu et al., 2021), there is an enduring interest in symbolic meaning representations in the community. In the decade since English Abstract Meaning Representations (AMRs) were launched, they have become increasingly popular and feature in several successful NLP applications (Bonial et al., 2020; Zhang and Ji, 2021; Fu et al., 2021). The 60K sentence dataset of English annotations available via LDC[1] has contributed significantly to this popularity. There is also growing interest in projecting English AMRs onto other languages to explore their suitability as a cross-lingual representation (Damonte and Cohen, 2017; Blloshmi et al., 2020; Uhrig et al., 2021; Oral et al., 2022; Cabezudo et al., 2022; Damonte and Cohen, 2022), with mixed success (Wein et al., 2022b; Wein and Schneider, 2022). In parallel with this exploration, a serious study has been made of the more English-centric aspects of AMRs, with the goal of moving AMRs in the direction of Uniform Meaning Representations (UMRs) as an annotation framework that can more readily be applied to all languages. As discussed in Van Gysel et al. (2021), UMR was developed with cross-linguistic scope in mind, paying careful attention to linguistic typology and typologically diverse languages. For researchers familiar with AMR, and especially for those who have already been annotating datasets in other languages with AMR-like annotations, it is important to understand exactly the ways in which AMR and UMR are similar and the crucial ways in which they differ, as discussed here. This should be of interest to anyone familiar with AMR who wants to adapt it to another language, or who already has an AMR dataset that can be retrofit to be more compatible with UMR.

In our previous UMR publications, we introduced the schema by explaining how it carves up conceptual space into categories that are applicable to typologically diverse languages (Vigus et al., 2020, 2019; Van Gysel et al., 2019). Special consideration has been given to the needs of field linguists who may be approaching semantic annotation for the first time as well as to the semantic coverage that is new for UMR, such as modal (Vigus et al., 2019), aspectual (Vigus et al., 2020) and scopal relations (Pustejovsky et al., 2019). In this paper we turn to more direct mappings from AMR-elements to UMR-elements.

We start by reviewing how AMR carves up conceptual space into graph elements (section 2), and then show how these elements overlay those of the UMR schema, focusing on retention, alteration and removal. Section 3 focuses on role-role mappings, section 4 on abstract roleset mappings, and section 5 on abstract concept mappings. We limit our discussion to sentence-level graphs. In the

[1]AMR 3.0: https://doi.org/10.35111/44cy-bp51

Proceedings of the 21st International Workshop on Treebanks and Linguistic Theories (TLT, GURT/SyntaxFest 2023), pages 74 - 95
March 9-12, 2023

appendix, we illustrate these mappings with easily-digestible graphics.

As the developers of UMR, nothing would delight us more than a rush to convert existing AMR datasets to UMR. Indeed, AMR corpora already exist in a variety of languages: English[2], Chinese[3], Czech (Xue et al., 2014), Spanish (Migueles-Abraira et al., 2018; Wein et al., 2022a), Turkish (Oral et al., 2022; Azin and Eryiğit, 2019), Vietnamese (Linh and Nguyen, 2019), Portuguese (Anchiêta and Pardo, 2018; Sobrevilla Cabezudo and Pardo, 2019), Korean (Choe et al., 2019), Persian (Takhshid et al., 2022), and more. Some parallel corpora also exist (Damonte and Cohen, 2022; Li et al., 2017), which will be especially useful if converted to UMR. Much conversion of AMR corpora to UMR corpora should be able to be accomplished deterministically, with unique issues arising for each language.

The body of this paper pertains most directly to LDC's 3.0 English AMR release, which consists of 60,000 sentences, 7800 of which are annotated with Multisentence AMR (O'Gorman et al., 2018). English AMR has also been extended to special domain projects outside of the LDC release. [4] Sections 6.1 and 6.2 focus on existing Czech and Chinese data sets, considering issues that are expected to arise when converting existing AMR corpora to UMR format.

2 From AMR to UMR

UMR begins with sentence level graphs largely inherited from AMR. This paper describes the changes to AMR graph structures that will produce a first-pass UMR graph. In this section, we break AMR structures down into types, with subsequent sections describing how to map each type onto its UMR counterparts.

UMR improves on AMR in two major ways: first by adjusting the AMR schema to make it more cross-linguistically applicable, and second, by adding new semantic coverage to the schema in the form of sentence-level graph elements for tense, aspect, modality, and scope, as well as document-level dependency structures for temporal relations, modality, and coreference. See Vigus et al. (2019), Vigus et al. (2020), and Pustejovsky et al. (2019) for in depth instruction on these additions. The UMR-Writer (Zhao et al., 2021) also creates alignments between tokens in a surface level representation and elements in its graph.

AMR fails many non-English languages in two basic ways. First, the categories AMR uses to divide semantic space sometimes fail to cover necessary distinctions in a given language or fail to align with the language's category boundaries. Second, the style of morphosyntactic expression used by a language may be incompatible with conversion from surface forms to graph structures. AMR is based on predicate argument structures and assumes that a predicate and its arguments are distinct tokens that can each be plugged into its own node in the graph. This creates problems for polysynthetic and agglutinating languages in which an event, its modifiers, and its participants may all be morphologically bound together. Applying English-based AMR practices either disrupts surface lexical units or produces single node graphs so semantically specific that they are unlikely to come up twice in a corpus. Figure 3 (Appendix A.2) demonstrates the second point with a side-by-side comparison of AMR and UMR graphs for an Arapaho sentence and its English translation.

Language-specific Rolesets: AMR represents the semantics of a given sentence as a rooted, directed, acyclic graph consisting of nested predicate argument structures (Banarescu et al., 2013). Most of the predicate argument structures come in the form of lexical rolesets, taken from a valency lexicon associated with the language to be annotated. AMR was originally created with English in mind, and English AMR uses the English PropBank Roleset Lexicon (Palmer et al., 2005; Pradhan et al., 2022; Bonial et al., 2014; O'Gorman et al., 2018a,b). Chinese AMR relies on the Chinese PropBank, and Czech AMR uses the Czech PDT-Vallex valency lexicon (Hajič et al., 2003; Urešová et al., 2021).

In general, lexical rolesets are created for eventualities (events and states) and are ambiguous for the parts of speech used to express them. Rolesets consist of a predicate ID, for sense disambiguation, and a set of numbered roles with lexicalized descriptions that are considered semantically primary to the sense.

General Roles: In order to give structure to

[2]ISI's AMR main page: https://amr.isi.edu/

[3]Chinese AMR main page with release: https://www.cs.brandeis.edu/ clp/camr/camr.html

[4]These range in size from just over 1000 with DialAMR (Bonial et al., 2020) to 8000 with NIH THYME (Wright-Bettner et al., 2020) to over 10,000 for Minecraft SpatialAMR (Bonn et al., 2020), to mention only those we know well.

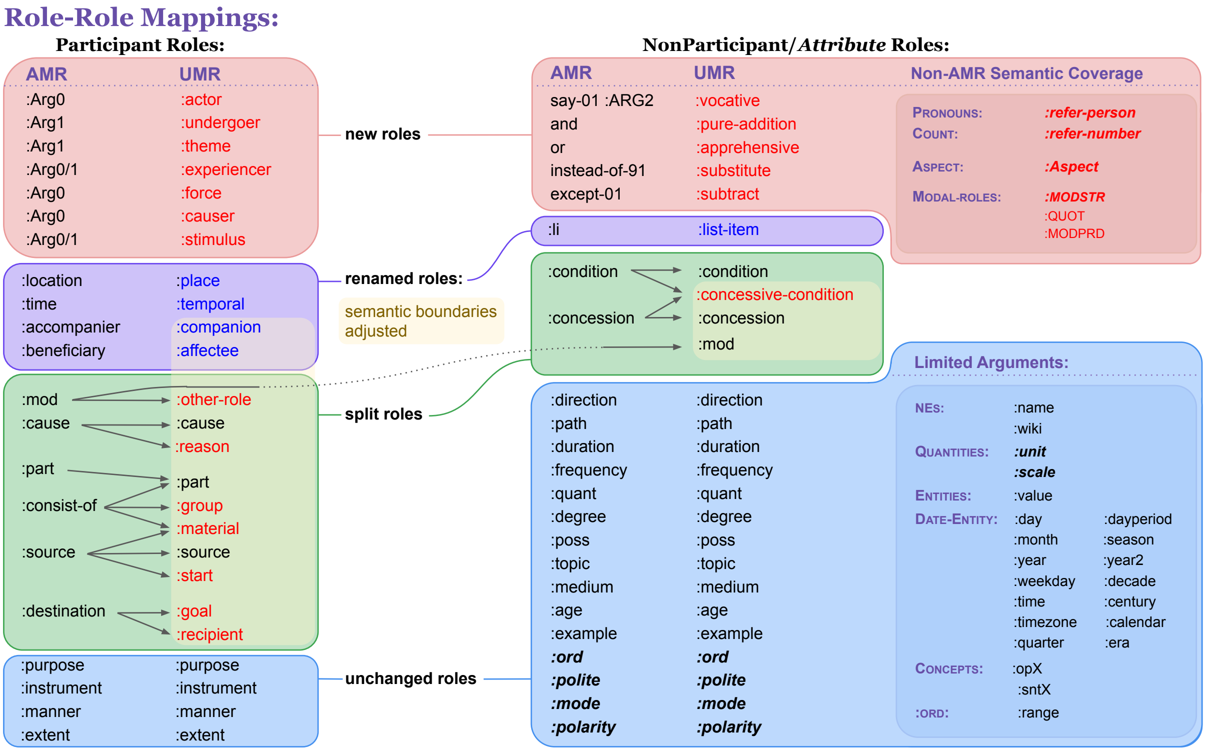


Figure 1: All AMR and UMR sentence-level roles, mapped. *Red text = new roles, blue text = role name changed, yellow highlight = semantic boundary shift.*

relations not covered by language-specific rolesets, AMR starts with general semantic roles (e.g., `:location`). These allow modification of entities and eventualities in a graph. Each role has a corresponding inverse role in the form `:<role>-of` (e.g., `:location-of`).

Abstract Rolesets: Sometimes, relations not covered by language-specific lexical rolesets need to be represented in a graph with a predicate (which has a variable), rather than a role (which does not). AMR uses Abstract Rolesets for this, broken down into several categories. First, most general roles have a corresponding reifying roleset (e.g., `:location`, `be-located-at-91`, with ARG1-theme and ARG2-location). Other Abstract Rolesets cover predication of implicitly understood relations such as entity-entity role relationships (`have-rel-role-91`) or inclusion (`include-91`).[5] In certain cases, AMR uses an existing language-specific roleset from English PropBank as an abstract roleset (e.g., `last-01` as a reification of `:duration`; `contrast-01` for contrast between clauses). There is no single comprehensive list of Abstract Rolesets readily available to AMR annotators either in the AMR editor[6] or the AMR guidelines.[7]

Abstract Concepts: AMR uses a limited set of Abstract Concepts in the graphs. These do not come with numbered arguments, but they may project one or more general roles as arguments (see the Limited Arguments in figure 1). Abstract Concepts include discourse relations such as `and`, `amr-unknown` (used for questions), quantity types (e.g., `temporal-quantity`, `volume-quantity`), entity types (e.g., `date-entity`), and concepts from the Named Entity Hierarchy. The NE concepts can be used to characterize implicit participants when needed.

3 Mapping Roles

In order to better serve a typologically diverse range of languages, UMR uses an updated set of general semantic roles. Many are taken directly from AMR, although some have been renamed, semantically expanded or contracted, or split. Others have been replaced with non-role strategies in UMR, and a small handful have been discontinued and not replaced. Changes are motivated by the cross-linguistic argument realization patterns in the ValPaL database (Vigus et al., 2020; Hartmann

[5] AMR frequently uses a `-91` suffix for these rolesets.

[6] https://amr.isi.edu/editor.html

[7] https://github.com/amrisi/amr-guidelines/blob/master/amr.md

et al., 2013). Each role in UMR still has an inverse and a roleset that reifies it. This section presents the ways in which roles have been adjusted to be more semantically comprehensive for UMR, as illustrated in Figure 1.

A major feature of UMR is the ability to use only general semantic roles for languages without roleset lexicons (see Stage 0 annotation in Van Gysel et al. (2021)). Due to this change, rather than having a single list of roles as in AMR, UMR groups its roles into Participant roles (primary arguments of eventualities), NonParticipant roles (other modifiers), and Attributes (roles that take only a fixed set of values). Most adjustments from AMR to UMR target Participant roles.

New roles: the concepts/roles in the top (pink) boxes in Figure 1 occur in AMR, but not as general roles. Of these, new participant roles (left) appear exclusively as numbered arguments (typically ARG0 or ARG1) of language-specific lexical rolesets, and new NonParticipant roles (right) as abstract concepts or abstract rolesets. The new NonParticipant roles are all discourse relation roles (see section 4) with the exception of **`:vocative`**, which AMR deals with by inferring an English-specific **`say-01`** roleset. A general role for vocatives is preferable for non-English corpora.

Renamed roles: the roles in the next boxes down in Figure 1 (purple) have been renamed so as to better describe the semantic categories they cover. **`:companion`** and **`:affectee`** both have shifted semantic boundaries compared to their AMR counterparts. **`:companion`** now applies to animate participants only ('I traveled *with Mary*') and is no longer to be used with entity modification (as in 'a pizza *with pineapple*'). Instances of the latter sort should be handled now with the appropriate topological (**`:part`**) or discourse relations (**`and`**) depending on the situation. **`:affectee`** now includes Maleficiaries as well as Beneficiaries, filling a sometimes uncomfortable gap in AMR. The other roles (**`:place`**, **`:temporal`**, and **`:list-item`**) are more cross-linguistically descriptive than their previous forms. **`:place`** and **`:temporal`** are better labels for participants that are abstract or metaphorically extended. Note that **`:time`** is still used for clock times under **`date-entity`**. **`:list-item`** replaces **`:li`**, which was considered too opaque a label.

Split roles: roles that are split into finer-grained categories to better support cross-linguistic semantic role diversity are next in Figure 1 (green). In most cases, the original role name has been retained but applied to a narrower category. The exceptions are AMR's **`:consist-of`** and **`:destination`**. **`:consist-of`** is dispreferred as a label because of its '**`-of`**' ending, which is ambiguous with inverse roles. **`:destination`** has been split across **`:goal`** and **`:recipient`** because of how **`:destination`**'s split contrasts with **`:source`**'s split. **`:source`** was split into three roles: **`:source`** (the Start-Point in motion events that involves separation of a Part from a Whole), **`:start`** (the Start-Point for other motion events), and **`:material`** (the Material in a creation event). **`:destination`** on the other hand was split along animacy lines, with **`:recipient`** being used for animate End-Points of sent-motion events, and **`:goal`** being used, in effect, for participants that contrast with any of the three **`:source`**-based roles. In other words, **`:goal`** captures End-points in motion events regardless of whether the entity in motion becomes part of the End-point, and also Products in creation events. **`:goal`** is considered to be more appropriately general than **`:destination`** across these types. **`:consist-of`** is also split across three narrower roles–**`:part`**, **`:group`**, and **`:material`**. These are largely topological distinctions.

In AMR, **`:mod`** is used as a catch-all role for modifiers with no clear home in one of the other roles. In UMR, we split this duty between **`:mod`** and **`:other-role`**. **`:other-role`** should be used as a catch-all Participant role, whereas **`:mod`** should be used as a catch-all NonParticipant role. In other words, **`:mod`** should still be used for demonstratives and the like. We note too that UMR is much more flexible about leaving surface concepts unannotated if they do not fit clearly into the schema, somewhat reducing the need for a catch-all role. (Exactly which concepts should be left unannotated for this reason is a question that needs to be resolved on a language-by-language basis, depending on how conceptual space is morphosyntactically distributed. Consider the option to omit 'just' and 'back-and-forth' as their own :mod nodes in the English UMR in Appendix A.2.) **`:other-role`** is more likely to be used by low-resource languages undergoing Stage 0 annotation.

Removed roles: UMR discontinues use of a number of roles (Appendix A.1). Some of these were simply shortcuts, while others do not fit into the current schema as roles. The most important removed role is `:domain`, which AMR used for identity relations in place of a copular roleset. `:domain` was used in places where property, object, or identity relations were expressed clausally. It was considered an inverse of `:mod`, so `:mod` has been given a new inverse, `:mod-of`. In UMR, these clausal expressions are now handled using the rolesets of nonprototypical predication (formerly ‘nonverbal clauses’), discussed in section 4.

`:subset`, `:superset`, and `:subevent` have also been removed, although `:subset` is still used for entity coreference at the document level. `:subset` and `:superset` relations can still be represented in sentence-level graphs with `include-91` from AMR. Subevent relations can still be represented at the document level in UMR as part of the temporal dependency annotation (see UMR guidelines [8]). As for the shortcuts: these were not fully ‘roles’, they just triggered automatic creation of a related abstract roleset in the graph. The shortcuts are no longer used, but the rolesets they pointed to still are, in their UMR forms. The exception is `:cost`, which should now be treated with `:other-role`.

Inverse roles appear in the form :<role>/:<role>-of (as seen with `:mod`/`:Mod-of` above), with the inverse capitalized. As noted with `:consist-of`, UMR does not use base roles ending in `-of`, thus eliminating ambiguity.

4 Mapping Abstract Rolesets

4.1 Role-Reification Rolesets: Sometimes during annotation, relations typically handled with general roles need to be represented as a predicate in a graph. This need may arise for three different reasons. First, the relation might be expressed clausally (*’I’m in Chicago’*). Second, the relation may be split across sentences, as (*’What did you make?’ ’Blankets.’*). This occurs especially frequently in casual dialogue corpora. Third, sometimes the relation itself needs a variable, either for coreference or modification.

AMR provides reified rolesets for most of its general roles, although there are a few holes, as with `:direction` and `:path`. Most of AMR’s reifications are constructed fairly consistently. Those that were created for AMR had `-91` suffixes, but sometimes, English-specific rolesets were used. This turns out to be an uncomfortable set-up for annotation of non-English languages, since English-sourced roleset names may not be obviously identifiable as reifications (e.g., `last-01` for `:duration`, `concern-02` for `:topic`). Also problematically, the English-sourced rolesets don’t always have exactly the same semantic coverage as the roles they reify. For example, `age-01` is aspectually-incompatible with `:age`, as it pertains to an aging event rather than a property. `age-01` also includes an `:ARG0` for the causer/agent of the aging event, which is not applicable to the `:age` role.

Because of these issues, the entire set of reified rolesets has been overhauled to bring them into alignment and make them more cross-linguistically appropriate. This involves new conventions for naming the rolesets and new conventions for naming and structuring their arguments. We believe these changes are a great improvement that will benefit all annotators moving forward.

Naming conventions: Reification rolesets are now named consistently, as follows: 1) each has a `-91` suffix. If multiple rolesets exist for the same concept, numbering continues with `-92` and so on; 2) each starts with a `have-` prefix, with the exception of `be-polite-91`, which we keep as a stylistic choice; 3) the content between ’`have-`’ and ’`-91`’ is the name of the role being reified. Appendix A.3 shows reification rolesets for all UMR roles.

Argument structuring conventions: Reification rolesets are also structured and numbered consistently now, as follows: 1) each roleset starts with `:ARG1`, as `:ARG0` is reserved for agentive/causal arguments; 2) `:ARG1` is used for the event or entity that would serve as the head of the unreified role in a graph, and `:ARG2` is for the value that would be annotated under the unreified role. Other arguments may be possible but are less conventionalized. See how the old reification for `:accompanier` was reconfigured as `have-companion-91` below:

```
accompany-01
    :ARG0 accompanier
    :ARG1 accompanied
    :ARG2 start point
    :ARG3 end point
have-companion-91
    :ARG1 event
    :ARG2 accompanier
```

Argument structure changes particularly affect mappings between prior English-sourced reification rolesets and their new UMR counterparts, since many of the English predicates (`cause-01`, `concern-02`) originated as rolesets for agentive verbs that started with `:ARG0`.

[8] UMR website: https://umr4nlp.github.io/web/

4.2 Rolesets for Nonprototypical Predication. Following Croft (2022) and Heine (1997), UMR has a set of **-91** rolesets for representing nonprototypical predication (previously referred to as nonverbal clause rolesets). Also following Croft, UMR has abandoned the 'nonverbal clause' terminology because annotators found it to be unclear. Although the term sounds like it excludes verbal expressions, it is in fact used to describe a set of semantic categories that can be expressed in many different ways cross-linguistically, including verbally.

Rolesets of nonprototypical predication cover five semantic categories first (possession, location, property predication, object predication, and identity relationships), and describe syntactic realization second and less strongly. While it is true that these semantic categories can be expressed using many different syntactic strategies across languages (attributive expressions, predication through juxtaposition, etc.), the rolesets for location, possession and property predication are to be used for *clausal* expressions, rather than *phrasal* expressions, which are covered in UMR with **:place**, **:poss** and **:mod** roles. In fact, the rolesets for nonprototypical predication of location, possession and property serve *as* the reifications of **:place**, **:poss**, and **:mod**. Clausal expressions of location and possession are divided discourse-pragmatically depending on which argument is presented as new information. As for object and identity predication, due to the complexity and structure of the relationships involved, their **-91** rolesets are used for both phrasal and clausal expressions. Appendix A.4 shows the nonprototypical predication rolesets.

In a recent change, all of the above rolesets now follow the argument numbering conventions used for reification rolesets. This means that **have-rel-role** and **have-org-role** have had their argument numbers shifted so that they start with **:ARG1** instead of **:ARG0**. Because of this change, and in the interest of avoiding confusion between versions, the renumbered UMR versions have been given a **-92** suffix.

However, **have-91** and **belong-91** started out as the English have-03 and belong-01. Their arguments have also been shifted to avoid using **:ARG0**. A sticky issue arises when we consider which instances of 'have' and 'belong' in English annotation to retrofit. The rolesets of nonprototypical predication are intended to be used any time the semantic relationship is expressed clausally, and this includes verbal expressions. This suggests that all instances previously annotated as **have-03** in AMR should be converted to **have-91**, and the **have-03** roleset should be retired. But how do we determine when 'the semantic relationship' has been expressed when it comes to verbal rolesets?

For example, **belong-01** has been used for verbal expressions of membership and possession. Should the roleset be retained for instances of membership-belonging, while instances of possessor-focused-possession are represented with **belong-91**? What about verbs expressing location? Some postural verbs can be used in bleached form for generic expressions of location (*'the radio tower lies 10 miles south of town'*), which also seem to fit the criteria for using **have-location-91** or **exist-91**. We doubt annotators would mark postural instance (*'the radio tower lay on its side'*) with a **-91** roleset, but bleached usages need more consideration.

Another interesting issue involves clauses of property predication (**have-mod-91**), often expressed adjectivally in English. English PropBank includes many rolesets for adjectives. Originally, these were limited to predicating adjectives only, but it became apparent that adjectival rolesets were useful in cases where multiple arguments were involved or where multiple senses existed. Over time, single-place adjective rolesets were added even outside of these constraints. AMR did not adopt all of these adjectival rolesets, although they did adopt some. The convention has become to use any roleset that shows up in the editor any time it fits semantically, which means plenty of modifying adjectives have been annotated with adjective rolesets in English AMR. To put it plainly, English AMR has not been consistent on the adjective front. Cross-linguistically, languages express property predication in an even wider variety of ways, including with lexical verbs (Croft, 2022; Heine, 1997). While some linguists may not wish to use language-specific property rolesets for property predication, for languages like Arapaho in which properties are frequently expressed as verbs, it would seem very strange not to. Ultimately, we propose that individual languages make this decision for themselves.

4.3 Other Abstract -91 Rolesets. AMR has a number of rolesets that are used for other abstract relationships in the graphs. Many of these are **-91** rolesets (**publication-91**, **correlate-91**, etc.) but some are also English-specific rolesets (e.g., `mean-01`, `resemble-01`, `contrast-01`) Appendix A.5 presents a full list of these rolesets and their UMR resolutions. UMR's versions all use a `-91` sufix; argument numbering has been retained.

4.4 Discourse Relations. UMR provides a lattice of categories to annotate relations between clauses in complex sentences. The upper levels represent discourse relations as abstract concepts (which can take a variable number of interchangeable `:opX` arguments, e.g. junction) or abstract rolesets (whose arguments are fixed in number and ordered, e.g. contrast). The bottom level concepts can be expressed roles or their inverses or reifications. Appendix A.6 shows the lattice with AMR concepts overlaying UMR concepts.

The categories are based on typological work by Croft (2022), Malchukov (2004) and Thompson & Longacre (1985). Some of the more coarse-grained categories on the lattice are inherited directly from AMR, as are some of the most fine-grained categories. Most new additions can be found in the intermediate levels. Specifically, UMR maintains the use of coarse-grained AMR `(o/ or)` and `(a/ and)` with their numbered `:opX` arguments, which did the heavy lifting for conjunction and disjunction, and adds finer-grained `(o / or-incl)` and `(o /or-excl)`. As for contrast between clauses, AMR's English-sourced `contrast-01` roleset has been swapped out for a version with a `-91` suffix and no `ARG0`. Given that many languages do not have clearly distinct morphosyntactic strategies for expressing conjunction and contrast, various high- and mid-level values in the UMR lattice combine concepts formerly annotated with `and` and those formerly annotated with `contrast-01`. This is the case for `and-unexpected`, `and-contrast`, and `and-but`. Here, manual reannotation of existing corpora may be necessary if the more fine-grained options are to be used.

On the lowest level of the lattice, UMR uses a number of fine-grained roles to annotate subtypes of conjunction, disjunction, and contrast. Some, like `:concession` and `:manner`, are inherited from AMR and simply placed in the discourse lattice under the appropriate higher-level value (`unexpected-co-occurrence-91` and `consecutive`, respectively). Others are newly introduced (e.g. `:subtraction`, `:pure-addition`, `:apprehensive`). All have inverses and related reified rolesets created as discussed above.

As with other domains where UMR organizes category values in lattices, annotators from individual languages will decide which values are appropriate for (i.e. explicitly expressed in) their language and use only those – the lattice helps make their annotations compatible to those in other languages.

5 Mapping Abstract Concepts

As mentioned in section 2, AMR includes a number of abstract concepts such as `X-entity` and `X-quantity` types, named entity types, mathematical concepts, and other annotation-support concepts like `AMR-unintelligible`. By-and-large, these are unchanged in UMR. Of course those referencing '`AMR-`' have been updated to reference '`UMR-`' (e.g. `UMR-unknown`, etc.). All `X-entities` and `X-quantities` are still in use. See Appendices A.7 and A.8.

When a named entity is identified, UMR assigns a semantic type to it, following the practice of AMR. These types are organized hierarchically, and annotators aim to use the most specific type possible. UMR makes a number of adaptations to the AMR hierarchy in order to make the named entity types more cross-linguistically applicable and practical. Following the general spirit of making UMR ontologies hierarchical so that annotators of each language can select the level of granularity they are comfortable with, UMR arranges the named entity types hierarchically in a lattice. UMR also adds categories that are needed when annotating indigenous languages, as their societies are often organized in such a way that the AMR types did not fit well. For instance, entity types like "clan" are not available in AMR but are added in UMR. The UMR named entity type heierarchy is "backward compatible" in that it is a superset of the AMR entity types, and existing AMR named entity types can be automatically mapped to UMR. See Appendix A.9.

6 Czech and Chinese Corpora

Adapting UMR annotation to other languages need not necessarily mean starting from scratch. In this section, we describe preliminary work planning the use of existing resources for Czech and Chinese to build UMR-annotated corpora. For both languages, there are existing AMR corpora (of varying size), as well as semantically-annotated corpora, for example in the styles used for the shared tasks on Meaning Representation Parsing (Oepen et al., 2019, 2020). Either type of resource can provide a starting point for developing a UMR corpus with less effort.

6.1 Czech Corpora

In addition to Czech's corpus of 100 AMR-annotated sentences (Xue et al., 2014), Czech tectogrammatical (TR) annotation has undergone a preliminary comparison with the current UMR style and guidelines (Oepen et al., 2019, 2020). Xue et al. 2014 describe the relation between Czech TR annotation and AMR, and here we will briefly describe the possibilities for using TR for pre-annotation of Czech to UMR.

The tectogrammatical annotation has been developed for the Prague Dependency Treebank (Hajič et al., 2020) as well as for some of its sibling treebanks, in particular for the Prague Czech English Dependency Treebank (Hajič et al., 2012) which is a parallel treebank based on the WSJ portion of the Penn Treebank. Tectogrammatical annotation focuses on the syntactic-semantic properties of language; while it largely keeps the dependency structure used at the surface-syntactic level, it adds a number of semantic properties relevant to a possible conversion to UMR:

- argument (valency) structure and predicate senses similar to PropBank (Hajič et al., 2003), though the approach to argument labeling is different. Technically, the conversion to the equivalent frame files should be feasible (to allow the UMR annotation tool to be used for Czech structural annotation or conversion);
- elided arguments as separate nodes, with the possibility of linking them by coreference or other relations (see also below);
- removal of function words, replacing them with largely semantic relations similar in number and nature to UMR roles;
- semantic attributes on each node (depending on its type), such as tense (preceding/concurrent), aspect (regardless of lexical vs. syntactic expression), number, modality, etc.;
- coreference relations, both grammatical (wh-clauses, attribute clauses, etc.) and textual (pronominal);
- discourse relations that go beyond sentence boundaries, and paratactic relations within sentences, which can serve as the basis for logical predicates;
- information structure annotation for determining scope in the focus part of sentences; and
- multiword expression annotation for both named entities and terminology.

These features should simplify and automate a large part of the conversion to UMRs. As described in more detail in Xue et al. (2014), about half of the sentences in a parallel Czech-English corpus have the same structure (between the TR annotation and the AMR structure, which is identical or easily convertible into UMR), needing to convert the labels only or do simple structural changes by deterministic rules (such as mapping multiword expressions into a single lexeme if the ontology used requires it). Other algorithmic changes involve TR attributes for modality (some of which will be converted to a structure headed by predicates such as `possible-01`, or to the `:modstr` attribute), TR rhematizer nodes for negation (to be converted to the `:polarity -` attribute), and others.

The other half of the TR structures will have to be checked and possibly corrected by hand; still, most of the annotation contained in the TR-labelled and structured trees will be valid and could remain intact.

6.2 Chinese Corpora

Existing Chinese AMR data sets (Li et al., 2016, 2019) are drawn from the Chinese version of the Little Prince (1562 sentences) as well as the Chinese TreeBank (Xue et al., 2005, 5000 sentences), with semantic roles that are defined in the frame files for the Chinese Propbank (Xue and Palmer, 2009). They generally adopt the AMR annotation style, with adaptations to handle Chinese-specific constructions. The adaptations that are needed to map Chinese AMRs to UMRs are thus very similar to those discussed in previous sections.

Chinese AMR does extend English in a number of ways, and they include the following:

- Chinese AMR adds a few Chinese-specific roles, including :cunit, :tense, and :aspect. :cunit indicates a relation between a noun and a measure word (e.g., 本, a measure word or classifier for books). :tense indicates a relation between a verb and an adverb (e.g., 将 "will"), and :aspect indicates a relation between a verb and an aspect marker (e.g., 着, 了, 过, 正在). The Chinese tense and aspect annotation are thus very superficial. However, when mapping to UMRs, they can be used to help determine the UMR aspect attribute and temporal relations.

- Chinese discourse relations in Chinese AMR are based on discourse relations defined in the Chinese Discourse TreeBank (Zhou and Xue, 2012, 2015). In addition to *and* and *or*, they also include *causation*, *condition*, *contrast*, *temporal*, *concession*, *progression*, *purpose*, *expansion*, and *multi-sentence*. They are annotated as Chinese AMR concepts and can be mapped to UMR discourse relations.

A significant departure from English AMR is that Chinese AMR data sets include concept-to-word alignments, as well as relation-to-word alignments. Since UMR also captures alignment between UMR concepts and word tokens from the source language, the alignment in Chinese AMR data sets will help map Chinese AMRs to UMRs. However, Chinese AMR's alignments are different from UMR's in some ways, so some changes to the alignments are needed to convert between them during AMR-to-UMR conversion. In particular, the relation-to-word alignments will need to be stripped off.

7 Conclusion

As more languages are annotated with UMR, we continue to identify ways the schema can be further refined to support cross-lingual expressivity. However, the process with new languages can be slow. Converting existing AMR corpora to UMR is an efficient way to grow the overall UMR corpus. Also, users with expertise in how AMR categories map to real language conceptual space can help identify the more nuanced areas in which UMR can improve on AMR. We expect the mappings outlined in this paper to be important support for users who wish to be a part of this process.

Acknowledgements

This work is supported in part by a grant from the CNS Division of National Science Foundation (Awards no: 2213804, 2213805) entitled "Building a Broad Infrastructure for Uniform Meaning Representations". It has also been supported by the Czech Science Foundation project "LUSyD", Award No. GX20-16819X, and the by LINDAT/CLARIAH-CZ Large Research Infrastructure, supported by the Czech Ministry of Education, Youth and Sports, Awards No. LM2023062 and LM2018101. All views expressed in this paper are those of the authors and do not necessarily represent the view of the National Science Foundation.

References

Rafael Anchiêta and Thiago Pardo. 2018. Towards AMR-BR: A SemBank for Brazilian Portuguese language. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Zahra Azin and Gülşen Eryiğit. 2019. Towards Turkish Abstract Meaning Representation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 43–47, Florence, Italy. Association for Computational Linguistics.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract Meaning Representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria. Association for Computational Linguistics.

Rexhina Blloshmi, Rocco Tripodi, and Roberto Navigli. 2020. XL-AMR: Enabling cross-lingual AMR parsing with transfer learning techniques. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2487–2500.

Claire Bonial, Julia Bonn, Kathryn Conger, Jena D. Hwang, and Martha Palmer. 2014. PropBank: Semantics of new predicate types. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3013–3019, Reykjavik, Iceland. European Language Resources Association (ELRA).

Claire Bonial, Lucia Donatelli, Mitchell Abrams, Stephanie Lukin, Stephen Tratz, Matthew Marge, Ron Artstein, David Traum, and Clare Voss. 2020. Dialogue-AMR: abstract meaning representation for dialogue. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 684–695.

Julia Bonn, Martha Palmer, Jon Cai, and Kristin Wright-Bettner. 2020. Spatial AMR: Expanded spatial annotation in the context of a grounded Minecraft corpus. In *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020),*.

Marco Antonio Sobrevilla Cabezudo, Rafael Torres Anchiêta, and Thiago Alexandre Salgueiro Pardo. 2022. Comparison of Cross-lingual Strategies for AMR-to-Brazilian Portuguese Generation.

Hyonsu Choe, Jiyoon Han, Hyejin Park, and Hansaem Kim. 2019. Copula and case-stacking annotations for Korean AMR. In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 128–135, Florence, Italy. Association for Computational Linguistics.

William Croft. 2022. *Morphosyntax: constructions of the world's languages*. Cambridge University Press.

Marco Damonte and Shay Cohen. 2022. Abstract Meaning Representation 2.0-Four Translations.

Marco Damonte and Shay B Cohen. 2017. Cross-lingual abstract meaning representation parsing. *arXiv preprint arXiv:1704.04539*.

Qiankun Fu, Linfeng Song, Wenyu Du, and Yue Zhang. 2021. End-to-end AMR coreference resolution. 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 4204–4214.

Jan Hajič, Eduard Bejček, Jaroslava Hlavacova, Marie Mikulová, Milan Straka, Jan Štěpánek, and Barbora Štěpánková. 2020. Prague dependency treebank - consolidated 1.0. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 5208–5218, Marseille, France. European Language Resources Association.

Jan Hajič, Eva Hajičová, Jarmila Panevová, Petr Sgall, Ondřej Bojar, Silvie Cinková, Eva Fučíková, Marie Mikulová, Petr Pajas, Jan Popelka, Jiří Semecký, Jana Šindlerová, Jan Štěpánek, Josef Toman, Zdeňka Urešová, and Zdeněk Žabokrtský. 2012. Announcing Prague Czech-English Dependency Treebank 2.0. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 3153–3160, Istanbul, Turkey. European Language Resources Association (ELRA).

Jan Hajič, Jarmila Panevová, Zdeňka Urešová, Alevtina Bémová, Veronika Kolářová, and Petr Pajas. 2003. PDT-VALLEX: Creating a Large-coverage Valency Lexicon for Treebank Annotation. In *Proceedings of The Second Workshop on Treebanks and Linguistic Theories*, pages 57–68, Vaxjo, Sweden. Vaxjo University Press.

Iren Hartmann, Martin Haspelmath, and Bradley Taylor, editors. 2013. *The Valency Patterns Leipzig online database*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Bernd Heine. 1997. *Cognitive foundations of grammar*. Oxford University Press.

Bin Li, Yuan Wen, Lijun Bu, et al. 2017. A comparative analysis of the AMR graphs from English and Chinese corpus of The Little Prince. *Journal of Chinese Information Processing*, 31(1):50–57.

Bin Li, Yuan Wen, Weiguang Qu, Lijun Bu, and Nianwen Xue. 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)*, pages 7–15.

Bin Li, Yuan Wen, Li Song, Weiguang Qu, and Nianwen Xue. 2019. Building a Chinese AMR bank with concept and relation alignments. *Linguistic Issues in Language Technology*, 18.

Ha Linh and Huyen Nguyen. 2019. A case study on meaning representation for Vietnamese. In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 148–153, Florence, Italy. Association for Computational Linguistics.

Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2021. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *arXiv preprint arXiv:2107.13586*.

Andrej L Malchukov. 2004. Towards a semantic typology of adversative and contrast marking. *Journal of semantics*, 21(2):177–198.

Noelia Migueles-Abraira, Rodrigo Agerri, and Arantza Diaz de Ilarraza. 2018. Annotating abstract meaning representations for Spanish. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.

Stephan Oepen, Omri Abend, Lasha Abzianidze, Johan Bos, Jan Hajic, Daniel Hershcovich, Bin Li, Tim O'Gorman, Nianwen Xue, and Daniel Zeman. 2020. MRP 2020: The second shared task on cross-framework and cross-lingual meaning representation parsing. In *Proceedings of the CoNLL 2020 Shared Task: Cross-Framework Meaning Representation Parsing*, pages 1–22, Online. Association for Computational Linguistics.

Stephan Oepen, Omri Abend, Jan Hajic, Daniel Hershcovich, Marco Kuhlmann, Tim O'Gorman, Nianwen Xue, Jayeol Chun, Milan Straka, and Zdenka Uresova. 2019. MRP 2019: Cross-framework meaning representation parsing. In *Proceedings of the Shared Task on Cross-Framework Meaning Representation Parsing at the 2019 Conference on Natural Language Learning*, pages 1–27, Hong Kong. Association for Computational Linguistics.

Tim O'Gorman, Sameer Pradhan, Martha Palmer, Julia Bonn, Katie Conger, and James Gung. 2018a. The new Propbank: Aligning Propbank with AMR through POS unification. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Tim O'Gorman, Michael Regan, Kira Griffitt, Ulf Hermjakob, Kevin Knight, and Martha Palmer. 2018b.

AMR beyond the sentence: the multi-sentence AMR corpus. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3693–3702, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Elif Oral, Ali Acar, and Gülşen Eryiğit. 2022. Abstract meaning representation of Turkish. *Natural Language Engineering*, pages 1–30.

Tim O'Gorman, Michael Regan, Kira Griffitt, Ulf Hermjakob, Kevin Knight, and Martha Palmer. 2018. AMR beyond the sentence: the multi-sentence AMR corpus. In *Proceedings of the 27th international conference on computational linguistics*, pages 3693–3702.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–106.

Sameer Pradhan, Julia Bonn, Skatje Myers, Kathryn Conger, Tim O'gorman, James Gung, Kristin Wright-bettner, and Martha Palmer. 2022. PropBank comes of Age—Larger, smarter, and more diverse. In *Proceedings of the 11th Joint Conference on Lexical and Computational Semantics*, pages 278–288, Seattle, Washington. Association for Computational Linguistics.

James Pustejovsky, Ken Lai, and Nianwen Xue. 2019. Modeling quantification and scope in Abstract Meaning Representations. In *Proceedings of the first international workshop on designing meaning representations*, pages 28–33.

Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. 2018. Improving language understanding by generative pre-training.

Marco Antonio Sobrevilla Cabezudo and Thiago Pardo. 2019. Towards a general Abstract Meaning Representation corpus for Brazilian Portuguese. In *Proceedings of the 13th Linguistic Annotation Workshop*, pages 236–244, Florence, Italy. Association for Computational Linguistics.

Reza Takhshid, Razieh Shojaei, Zahra Azin, and Mohammad Bahrani. 2022. Persian Abstract Meaning Representation. *arXiv preprint arXiv:2205.07712*.

Sandra A Thompson, Robert E Longacre, and Shin Ja J Hwang. 1985. Adverbial clauses. *Language typology and syntactic description*, 2:171–234.

Sarah Uhrig, Yoalli Rezepka Garcia, Juri Opitz, and Anette Frank. 2021. Translate, then parse! A strong baseline for cross-lingual AMR parsing. *arXiv preprint arXiv:2106.04565*.

Zdeňka Urešová, Alevtina Bémová, Eva Fučíková, Jan Hajič, Veronika Kolářová, Marie Mikulová, Petr Pajas, Jarmila Panevová, and Jan Štěpánek. 2021. PDT-vallex: Czech valency lexicon linked to treebanks 4.0 (PDT-vallex 4.0). LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Jens EL 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 Martin, Stephan Oepen, Martha Palmer, James Pustejovsky, Rosa Vallejos, and Nianwen Xue. 2021. Designing a uniform meaning representation for natural language processing. *KI-Künstliche Intelligenz*, 35(3):343–360.

Jens EL Van Gysel, Meagan Vigus, Pavlina Kalm, Sookkyung Lee, Michael Regan, and William Croft. 2019. Cross-linguistic semantic annotation: Reconciling the language-specific and the universal. In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 1–14.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Meagan Vigus, Jens EL Van Gysel, and William Croft. 2019. A dependency structure annotation for modality. In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 182–198.

Meagan Vigus, Jens EL Van Gysel, Tim O'Gorman, Andrew Cowell, Rosa Vallejos, and William Croft. 2020. Cross-lingual annotation: a road map for low- and no-resource languages. In *Proceedings of the Second International Workshop on Designing Meaning Representations*, pages 30–40.

Shira Wein, Lucia Donatelli, Ethan Ricker, Calvin Engstrom, Alex Nelson, and Nathan Schneider. 2022a. Spanish Abstract Meaning Representation: Annotation of a General Corpus. *arXiv preprint arXiv:2204.07663*.

Shira Wein, Wai Ching Leung, Yifu Mu, and Nathan Schneider. 2022b. Effect of source language on AMR structure. In *Proceedings of the 16th Linguistic Annotation Workshop (LAW-XVI) within LREC2022*, pages 97–102, Marseille, France. European Language Resources Association.

Shira Wein and Nathan Schneider. 2022. Accounting for language effect in the evaluation of cross-lingual AMR parsers. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 3824–3834, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Kristin Wright-Bettner, Chen Lin, Timothy Miller, Steven Bethard, Dmitriy Dligach, Martha Palmer, James H. Martin, and Guergana Savova. 2020. Defining and learning refined temporal relations in the clinical narrative. In *Proceedings of the 11th International Workshop on Health Text Mining and Information Analysis*, pages 104–114, Online. Association for Computational Linguistics.

Nianwen Xue, Ondřej Bojar, Jan Hajič, Martha Palmer, Zdeňka Urešová, and Xiuhong Zhang. 2014. Not an interlingua, but close: Comparison of English AMRs to Chinese and Czech. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 1765–1772, Reykjavik, Iceland. European Language Resources Association (ELRA).

Nianwen Xue and Martha Palmer. 2009. Adding semantic roles to the Chinese Treebank. *Natural Language Engineering*, 15(1):143–172.

Nianwen Xue, Fei Xia, Fu-Dong Chiou, and Marta Palmer. 2005. The Penn Chinese TreeBank: Phrase structure annotation of a large corpus. *Natural language engineering*, 11(2):207–238.

Zixuan Zhang and Heng Ji. 2021. Abstract meaning representation guided graph encoding and decoding for joint information extraction. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 39–49.

Jin Zhao, Nianwen Xue, Jens Van Gysel, and Jinho D Choi. 2021. UMR-Writer: A Web Application for Annotating Uniform Meaning Representations. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 160–167.

Yuping Zhou and Nianwen Xue. 2012. PDTB-style discourse annotation of Chinese text. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 69–77.

Yuping Zhou and Nianwen Xue. 2015. The Chinese Discourse TreeBank: A Chinese corpus annotated with discourse relations. *Language Resources and Evaluation*, 49(2):397–431.

A Appendices

A.1 Removed Roles

	AMR	UMR
ROLES:	:domain	NonPrototypical predication rolesets
	:subevent	Still available in document-level temporal dependency
	:prep-x	-
	:conj-as-if	-
	:subevent	-
SHORTCUTS:	:cost	:other-role
	:employed-by	have-org-role-92
	:meaning	mean-91
	:role	have-role-91, have-rel-role-92
	:subset	include-91
	:superset	include-91
	:instead-of	instead-of-91

Figure 2: Roles and shortcuts used in AMR but not used in UMR.

A.2 Graph Differences, Arapaho-English, AMR-UMR

ARAPAHO SENTENCE:

Text:	Beni'beebee3sohowuuneti3i' .
Morphological breakdown:	beni'- bee- bee3sohowuuneti -3i'
English glosses:	IC.just- REDUP- do sign language to each other -3PL
Parts of speech:	prefix- prefix- vai.RECIP -infl
English Translation:	*'[They didn't speak.] They were just doing sign language back and forth."*

	Arapaho:	English translation:
AMR:	(b / beni'beebee3sohowuuneti3i'-00)	(s / sign-00 :actor (t / they) :recipient t :mod (j / just) :manner (b / back-and-forth))
UMR:	(b / beebee3sohowuuneti-00 :actor (p / person ***/-3i'/* :refer-person 3rd :refer-number Plural) :recipient p :ARG1-of (c / contrast-91) ***/beni'-/* :Aspect Activity ***/bee-/* :modstr FullAff)	(s / sign-00 :actor (p / person ***they* :refer-person 3rd :refer-number Plural) :recipient p ***back and forth* :ARG1-of (c / contrast-91) ***just* :Aspect Activity ***back and forth* :modstr FullAff)

Figure 3: Comparison of AMR vs UMR graphs for a sentence from Arapaho (a polysynthetic language) and its English translation (non-polysynthetic). Note the disparity between the capturable semantics for Arapaho vs English in AMR. Conversely, UMR's schema allows semantically parallel sentences to appear in structurally-similar graphs. Alignments between tokens and graph elements ensure that tokens not appearing directly in the graph (e.g., 'they') may still be identified with their semantic representations in the graph.

A.3 Reification Roleset Mappings

AMR	+ reification	UMR	+ reification	
:Arg0	-	:actor	have-actor-91	new roles
:Arg1	-	:undergoer	have-undergoer-91	
:Arg1	-	:theme	have-theme-91	
:Arg0/1	-	:experiencer	have-experiencer-91	
:Arg0	-	:force	have-force-91	
:Arg0	-	:causer	have-causer-91	
:Arg0/1	-	:stimulus	have-stimulus-91	
say-01 :ARG2		:vocative	have-vocative-91	
and		:pure-addition	have-pure-addition-91	
or		:apprehensive	have-apprehensive-91	
instead-of-91		:substitute	instead-of-91	
except-91		:subtract	have-subtraction-91	
:location	be-located-at-91	:place	have-location-91	renamed roles
:time	be-temporally-at-91	:temporal	have-temporal-91	
:accompanier	**accompany-01**	:companion	have-companion-91	
:beneficiary	**benefit-01**	:affectee	have-affectee-91	
:li	have-li-91	:list-item	have-list-item-91	
:mod	have-mod-91	:mod	have-mod-91	split roles
		:other-role	have-other-role-91	
:cause	**cause-01**	:cause	have-cause-91	
		:reason	have-reason-91	
:part	have-part-91	:part	have-part-91	
:consist-of	**consist-01**	:group	have-group-91	
		:material	have-material-91	
:source	be-from-91	:source	have-source-91	
		:start	have-start-91	
:destination	be-destined-for-91	:goal	have-goal-91	
		:recipient	have-recipient-91	
:direction	-	:direction	have-direction-91	unchanged roles
:path	-	:path	have-path-91	
:duration	**last-01**	:duration	have-duration-91	
:frequency	have-frequency-91	:frequency	have-frequency-91	
:quant	have-quant-91	:quant	have-quant-91	
:degree	have-degree-91	:degree	have-degree-91	
	have-degree-91		have-degree-92	
:poss	**have-03, own-01**	:poss	have-91	
:topic	**concern-02**	:topic	have-topic-91	
:medium	-	:medium	have-medium-91	
:age	**age-01**	:age	have-age-91	
:example	**exemplify-01**	:example	have-example-91	
:ord	have-ord-91	:ord	have-ord-91	
:range	-	:range		
:polite	be-polite-91	:polite	be-polite-91	
:mode	have-mode-91	:mode	have-mode-91	
:polarity	have-polarity-91	:polarity	have-polarity-91	
:name	have-name-91	:name	have-name-91	
:wiki	-	:wiki	-	
:unit	-	:unit	have-unit-91	
:scale	-	:scale	-	
:value	have-value-91	:value	have-value-91	

Figure 4: Reification rolesets. *new roleset*, *renamed roleset*, *English-sourced roleset*.

A.4 Nonprototypical Predication Mappings

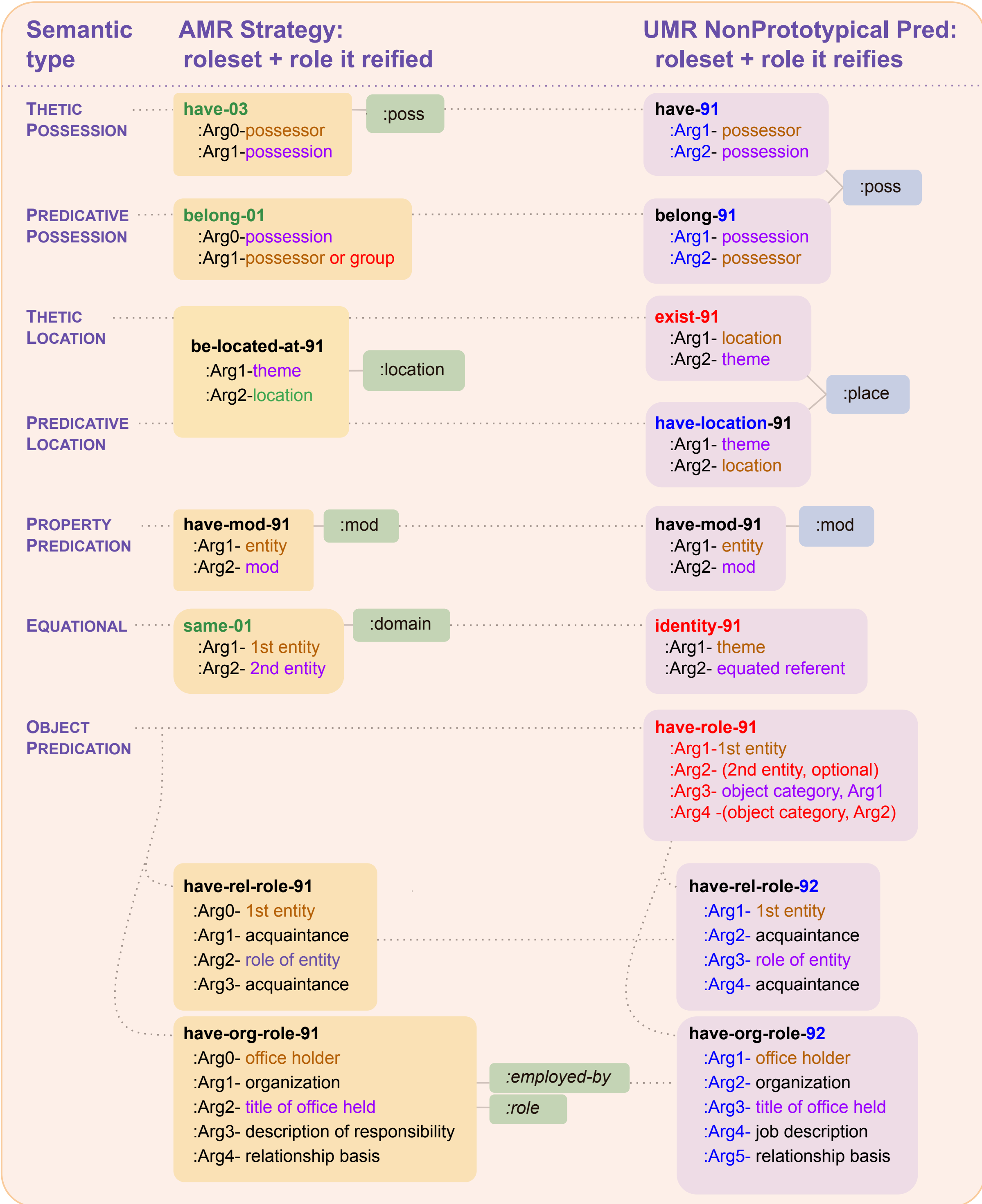


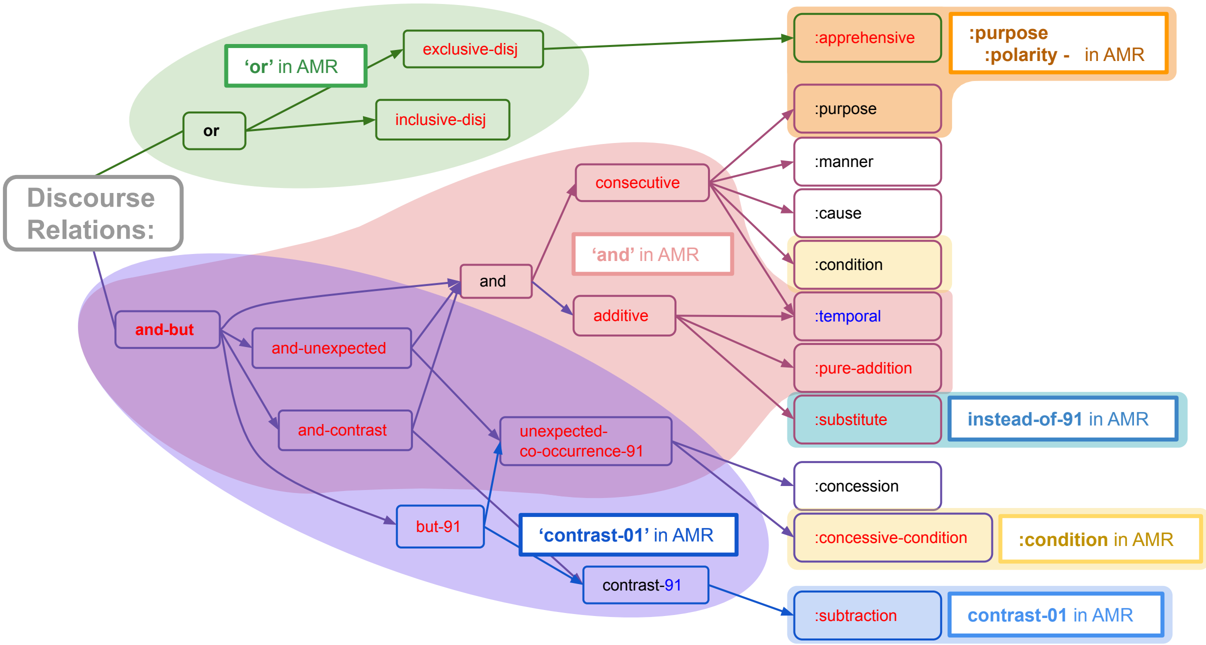
Figure 5: Nonprototypical Predication: *new element*, *renamed element*, *English-sourced roleset*, *topicalized argument*, *focused argument*

A.5 Other Abstract Roleset Mappings

AMR	UMR
-91 Rolesets:	
byline-91	byline-91
correlate-91	correlate-91
course-91	course-91
distribution-range-91	distribution-range-91
have-degree-of-resemblance-91	- (use: have-degree-91 + resemble-91)
hyperlink-91	hyperlink-91
include-91	include-91
instead-of-91	instead-of-91
publication-91	publication-91
rate-entity-91	rate-entity-91
regardless-91	- (use: :concessive-conditional)
request-confirmation-91	- (investigate further)
request-response-91	- (investigate further)
score-on-scale-91	score-on-scale-91
statistical-test-91	statistical-test-91
street-address-91	street-address-91
English-Sourced Rolesets:	
cite-01	cite-91
cost-01	- (use: :other-role)
counter-01 (for 'anti')	- (investigate further)
infer-01	- (use: infer-91 or :reason, depending on context)
mean-01	mean-91
oppose-01 (for 'anti')	- (investigate further)
protest-01 (for 'anti')	- (investigate further)
resemble-01	resemble-91
Modal Rolesets:	**(See Vigus et al. (2019) for full modal dependency annotation guidelines)**
obligate-01	:modal PrtAff
possible-01	:modal NeutAff
recommend-01	:modal PrtAff
permit-01	:modal NeutAff
wish-01	:modal NeutAff

Figure 6: Other Abstract Rolesets. Arguments unchanged where rolesets have been retained. *new roleset*, *renamed roleset*, *English-sourced roleset*, *commentary*.

A.6 Discourse Relation Mappings



Upper Levels: Abstract Concepts/Rolesets

Meaning	Concept/Roleset	Roles
Disjunctive	or	:op1 :op2 :opX
Exclusive Disjunctive	exclusive-disj	:op1 :op2 :opX
Inclusive Disjunctive	inclusive-dis	:op1 :op2 :opX
And	and	:op1 :op2 :opX
Consecutive	consecutive	:op1 :op2 :opX
Additive	additive	:op1 :op2 :opX
And + But	and-but	:op1 :op2 opX
And + Unexpected	and-unexpected	:op1 :op2 :opX
And + Contrast	and-contrast	:op1 :op2 :opX
Unexpected Co-Occurrence	unexpected-co-occurrence-91	:ARG1 :ARG2
But	but-91	:ARG1 :ARG2
Pure Contrast	contrast-91	:ARG1 :ARG2

Lower Levels: Roles & Reification Rolesets

Meaning	Relation	Reification	Roles
Apprehensive	:apprehensive	have-apprehensive–91	:ARG1 :ARG2
Purpose	:purpose	have-purpose-91	:ARG1 :ARG2
Means	:manner	have-manner-91	:ARG1 :ARG2
Cause	:cause	have-cause-91	:ARG1 :ARG2
Conditional	:condition	have-condition-91	:ARG1 :ARG2
Posterior Anterior Simultaneous	:temporal	have-temporal-91	:ARG1 :ARG2
Pure-Addition	:pure-addition	have-pure-addition-91	:ARG1 :ARG2
Substitution	:substitute	instead-of-91	:ARG1 :ARG2
Concessive	:concession	have-concession-91	:ARG1 :ARG2
Concessive-Conditional	:concessive-condition	have-concessive-condition-91	:ARG1 :ARG2
Subtraction	:subtraction	have-subtraction-91	:ARG1 :ARG2

Figure 7: Discourse Relations. Above: lattice, with concepts and rolesets in the upper levels and roles in the lower level. AMR concept mappings overlap the lattice. Below: Argument structures. *new relation*, *renamed relation*.

A.7 Abstract Concepts

AMR	+ Roles	UMR	+ Roles
AMR-/UMR- & Polarity-Related:			
amr-unknown	-	umr-unknown	
amr-choice	:op1 :op2 :opX	umr-choice	:op1 :op2 :opX
amr-empty	-	umr-empty	
amr-unintelligible	-	umr-unintelligible	
truth-value	:Polarity-of	truth-value	:Polarity-of
Entity Types:			
date-entity	(list unchanged, see Figure 1)	date-entity	(list unchanged, see Figure 1)
email-address-entity	:value	email-address-entity	:value
ordinal-entity	:value :range	ordinal-entity	:value :range :range-start
percentage-entity	:value	percentage-entity	:value
phone-number-entity	:value	phone-number-entity	:value
score-entity	:op1 :op2 :opX	score-entity	:op1 :op2 :opX
string-entity	:value	string-entity	:value
url-entity	:value	url-entity	:value
Interval Types:			
value-interval	:op1 :op2	value-interval	:op1 :op2
date-interval	:op1 :op2	date-interval	:op1 :op2
slash	:op1 :op2	slash	:op1 :op2
Other			
name	:op1 :op2 :opX	name	:op1 :op2 :opX
emoticon	:value	emoticon	:value
relative-position	:op1 :direction :quant	relative-position	:op1 :direction :quant
Count & Math			
more-than	:op1	more-than	:op1
less-than	:op1	less-than	:op1
at-most	:op1	at-most	:op1
at-least	:op1	at-least	:op1
sum-of	:op1	sum-of	:op1
product-of	:op1	product-of	:op1
difference-of	:op1	difference-of	:op1
quotient-of	:op1	quotient-of	:op1
power-of	:op1	power-of	:op1
root-of	:op1	root-of	:op1
logarithm-of	:op1	logarithm-of	:op1
ratio-of	:op1	ratio-of	:op1
Generic Concepts for Participant/NonParticipant Roles:			
thing		thing	:refer-number
person		person	:refer-person :refer-number
dummy (Chinese AMR)		dummy	
location		place	:refer-number
manner		manner	:refer-number
quantity		quantity	:Quant-of
event		event	:refer-number
Removed:			
either	:op1 :op2	- (use or/inclusive-disj/exclusive-disj :op1 :op2)	
neither	:op1 :op2	- (use or/inclusive-disj/exclusive-disj :op1 :op2 :polarity -)	
multiple	:op1	- (see mensural constructions, UMR-guidelines)	

Figure 8: Abstract Concepts, not including X-quantities or Named Entities. These are largely unchanged from AMR. *new concept*, *renamed concept*, *commentary*.

A.8 Quantity Types

Quantity-type	Arguments + Suggested Values	
monetary-quantity	:unit	dollar, euro, pound, yen, yuan
distance-quantity	:unit	meter, kilometer, inch, foot, yard, mile, light-year, kilo-base-pair
area-quantity	:unit	square-meter, square-kilometer, square-foot, acre, hectare, square-mile
volume-quantity	:unit	liter, cubic-meter, fluid-ounce, pint, gallon, cubic-mile
temporal-quantity	:unit	second, minute, hour, day, week, month, year, decade, century
frequency-quantity	:unit	hertz
speed-quantity	:unit	meter-per-second, mile-per-hour
acceleration-quantity	:unit	meter-per-second-squared
mass-quantity	:unit	kilogram, ounce, pound, ton, atomic-mass-unit, kilodalton
force-quantity	:unit	newton
pressure-quantity	:unit	pascal, bar, psi, atmosphere, torr
energy-quantity	:unit	joule, calorie, kilowatt-hour, btu, electron-volt
power-quantity	:unit	watt, horsepower
charge-quantity	:unit	coulomb
potential-quantity	:unit	volt
resistance-quantity	:unit	ohm
inductance-quantity	:unit	henry
magnetic-field-quantity	:unit	tesla, gauss
magnetic-flux-quantity	:unit	maxwell, weber
radiation-quantity	:unit	becquerel, curie, sievert, rem, gray, rad
fuel-consumption-quantity	:unit	liter-per-100-kilometer, mile-per-gallon
numerical-quantity	:unit	point, mole
information-quantity	:unit	bit, byte, kilobyte, megabyte, terabyte, petabyte, exabyte, zettabyte, yottabyte, nibble
concentration-quantity	:unit	molar (1M = 1 molar = 1 mole/liter), micromolar (μM), kilogram-per-cubic-meter, parts-per-million
catalytic-activity-quantity	:unit	katal (kat), microkatal, nanokatal, enzyme-unit (U)
acidity-quantity	:scale	ph
seismic-quantity	:scale	richter
temperature-quantity	:unit	degree
	:scale	celsius, kelvin, farenheit
angle-quantity	:unit	degree, radian

Figure 9: X-Quantity Types. Unchanged from AMR. More values are possible in UMR than just those listed as suggested.

A.9 UMR Named Entity Types

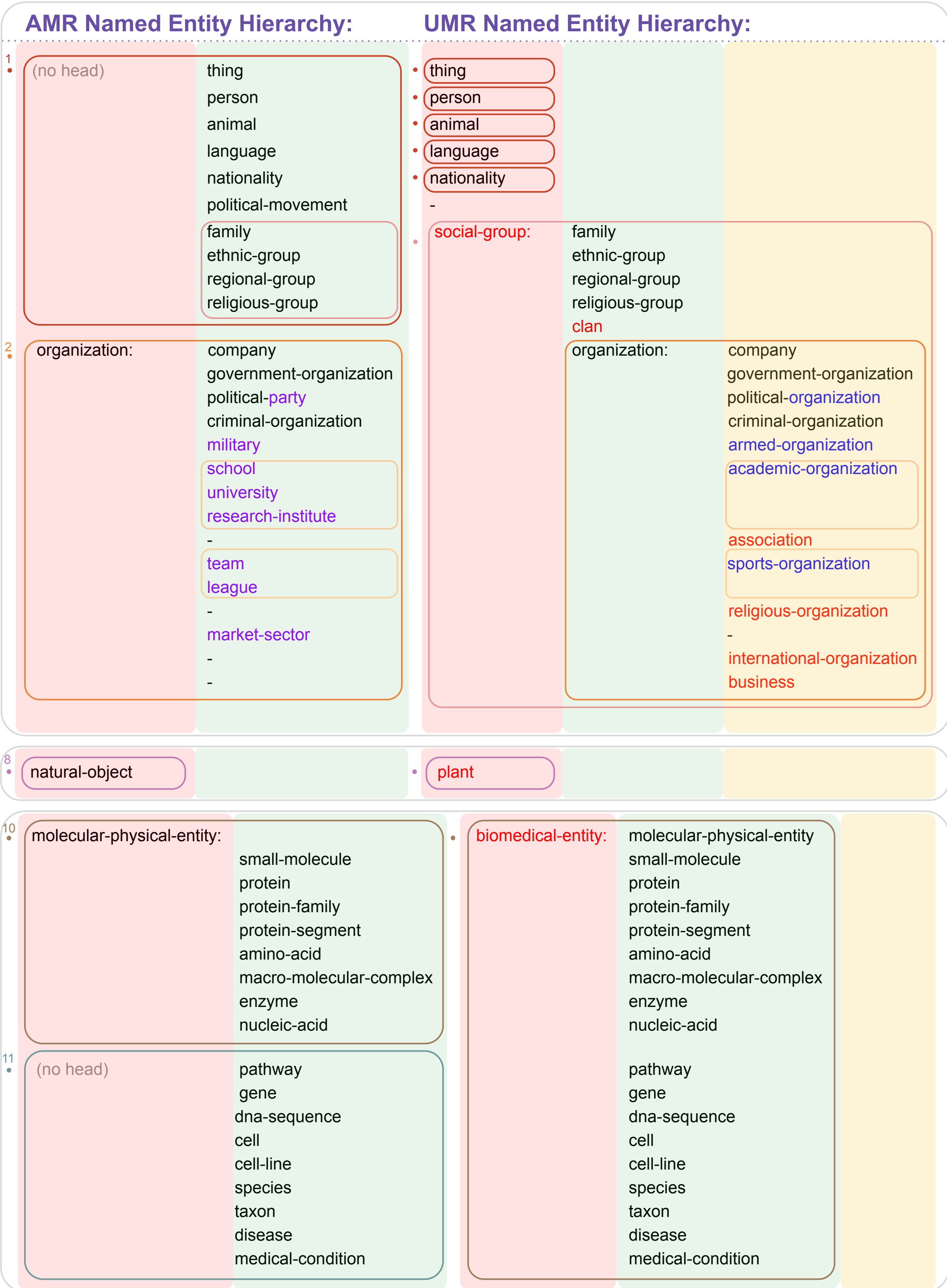


Figure 10: Named Entity Hierarchy Mapping. The far left bullet points on the AMR side are the eleven top-level groupings given in the AMR editor, numbered according to the order in which they appear there. Colored columns distinguish category levels for each hierarchy. *New type*, *renamed type*, *old name*.

	AMR Named Entity Hierarchy:		UMR Named Entity Hierarchy:	
3	location:	ocean	geographic-entity:	ocean
		sea		sea
		lake		lake
		river		river
		gulf		gulf
		bay		bay
		strait		strait
		canal		- (duplicate)
		peninsula		peninsula
		mountain		mountain
		volcano		volcano
		valley		valley
		canyon		canyon
		island		island
		desert		desert
		forest		forest
		continent		continent
		moon		moon
		planet		planet
		star		star
		constellation		constellation
		local-region	region:	local-region
		country-region		country-region
		world-region		world-region
		city	geo-political-region:	city
		city-district		city-district
		county		county
		state		state
		province		province
		territory		territory
		country		country
4	facility:	airport	facility:	airport
		station		station
		port		port
		tunnel		tunnel
		bridge		bridge
		road		road
		railway-line		railway-line
		canal		canal
		building		building
		theater		theater
		museum		museum
		palace		palace
		hotel		hotel
		worship-place		worship-place
		sports-facility		sports-facility
		market		market
		park		park
		zoo		zoo
		amusement-park		amusement-park

Figure 11: Named Entity Hierarchy Mapping, 3rd, 4th AMR bullet points. *New type*, *renamed type*, *old name*, *commentary*.

Figure 12: Named Entity Hierarchy Mapping, 5th, 6th, 7th, and 9th AMR bullet points. *New type*, *renamed type*, *old name*, *commentary*.