MR4AP: Meaning Representation for Application Purposes

Bastien Giordano and **Cédric Lopez**
Emvista
Immeuble Le 610
10 rue Louis Breguet
34830 Jacou, France
{firstname.lastname}@emvista.com

Abstract

Despite the significant progress made in Natural Language Processing (NLP) thanks to deep learning techniques, efforts are still needed to model explicit, factual, and accurate meaning representation formalisms. In this article, we present a comparative table of ten formalisms that have been proposed over the last thirty years, and we describe and put forth our own, Meaning Representation for Application Purposes (MR4AP), developed in an industrial context with a definitive applicative aim.

1 Introduction

An efficient human-machine interface is one of the dreams of Artificial Intelligence (AI). In NLP, despite the dazzling progress of the last few years with the emergence of large language models, it is necessary to resort to formal representations of textual statements, so that the machine can structure its result and reason while providing the explanation.

The last thirty years have witnessed numerous formalism proposals, the most recent of which are Universal Conceptual Cognitive Annotation (UCCA, Abend and Rappoport, 2013), Abstract Meaning Representation (AMR, Banarescu et al., 2013), Uniform Meaning Representation (UMR, Van Gysel et al., 2021) and BabelNet Meaning Representation (BMR, Navigli et al., 2022). The adoption of a meaning representation formalism is not a trivial choice, especially in an industrial context, as is the case of the authors of this paper. In this context, it is required that a formalism be explicit and factual while maximizing the richness and accuracy of the most semantically salient linguistic phenomena (Abzianidze and Bos, 2019).

The contribution of this article is twofold. On the one hand, we facilitate the comparison of ten formalisms *via* a table (section 2). To the best of our knowledge, although it is one of the shortcomings expressed by the community (Abend and Rappoport, 2017), only Koller et al. (2019) and Žabokrtský et al. (2020) have established such a comparison[1][2], but their studies and ours do not overlap much. Moreover, we include the most recent formalisms. It is on this basis that we present our own, Meaning Representation for Application Purposes (MR4AP, section 3), which we are already exploiting in an industrial context with an applicative focus. In section 4, we put forward three examples of our representation choices, while section 5 describes the first version of an annotated corpus following our formalism as well as a first small-scale manual annotation effort, accompanied by the annotation guidelines. Before concluding, we discuss some limitations and prospects for future work (section 6).

2 Meaning Representations comparison

In this section, we compare ten meaning representation formalisms, with which we compare our own (see Table 1). Each of the formalisms occupies a column (from oldest to newest), while the rows represent some of the linguistic features and phenomena that are fully covered (✓), partially covered (#), or not covered at all (empty space). The rows are grouped into five clusters, respectively related to genericity, structure, explicitness, various intra- and inter-sentence relations, and diversity of annotated attributes. For this last characteristic, we symbolize it from the least rich (+) to the richest (+++).

Partial coverage (#) has several meanings. It can mean that a feature is covered, but only in one of the formalism's extensions. This is the

[1] Other works address and compare in a more or less extensive way a number of formalisms (Bonn et al., 2023; Hershcovich et al., 2020; Pavlova et al., 2022; *inter alia*).

[2] Flanigan et al. (2022) have also prepared a tutorial in which they present and compare several meaning representation formalisms, but this tutorial was unknown to the authors at the time of writing and was not yet available.

	DRT	UNL	MRS	PDT	GMB	UCCA	AMR	UDS	UMR	BMR	MR4AP
Multilingual	✓	✓	✓	#	#	✓	#	✓	✓	✓	✓
Invariance	#	#	#		#	✓	✓		✓	✓	✓
Multi-sentence	✓			✓	✓	✓	#		✓	#	✓
P-A structure	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Named rel.		✓	#	#	✓	#	#		#	✓	✓
Sem. typing	✓	✓	#	#	#		✓	✓	✓	✓	✓
Anaph. & coref.	✓			✓	#	#	#		✓	✓	✓
Event coref.				✓		#	#		✓	✓	✓
Temporal rel.	✓	#	#	✓	✓	✓	#	#	✓	#	✓
Discourse rel.	✓	#	#	✓	#	✓	#		#	#	✓
Modal rel.									✓		
Attr. richness	++	+++	++	+++	+	+	+	+	+++	++	++

Table 1: Comparative table of meaning representation formalisms

case, for example, for the multilingual nature of the Prague Dependency Treebank's Tectogrammatical Layer (PDT-TL, whose original version was only for Czech (Mikulová et al., 2006) before covering English as well (Hajič et al., 2012)) and Groningen Meaning Bank (GMB, which was only for English (Basile et al., 2012) before being extended to German, Dutch, and Italian with Parallel Meaning Bank (PMB, Abzianidze et al., 2017)). The need to use an extension also holds true for AMR, a formalism for which various works have aimed at making it multi-sentence (O'Gorman et al., 2018) or enriching the annotated attributes (such as tense and aspect in Donatelli et al., 2018). This is also the case for UCCA with respect to coreference resolution (entities and events), which is dependent on a layer over the foundational one (Prange et al., 2019b)[3].

Partial coverage can also mean that a feature is covered but only in a limited way. This is the case for relations that are numbered rather than named (`:ARG0`, `:ARG1`, etc., in AMR), for labels that are insufficiently fine-grained (UCCA), or when nodes carry the label that is traditionally assigned to the arcs (PDT-TL). This partial coverage mainly concerns the group of intra- and inter-sentence relations: when one of these relations is taken into account by the formalism, however this coverage is only realized at one of the two levels (for instance, UMR's discourse relations), we consider that it is incomplete.

The different formalisms proposed over the years diverge on many points and converge on others. One of the most salient points of divergence is the distance between the meaning representation and the surface syntactic form. The Czech school with the PDT-TL formalism is among those that remain closest to syntax. Several layers of annotation are superimposed, the highest of which being the so-called tectogrammatical layer (`t-layer`), which combines syntax and semantics. Many linguistic phenomena are encoded (grammatical tense, coreference and anaphora, semantic types), some of which are largely discarded by other formalisms (ellipsis, focus/topicalization). However, its obvious proximity to syntax means that complex sentences that are semantically similar, but whose main and subordinate clauses would have been inverted produce drastically different results (Abend and Rappoport, 2017).

The invariance of representations for semantically close segments, regardless of their syntactic configuration (active/passive voice, paraphrasing, cleft sentences), is a consensus feature. AMR, UMR and BMR are among the formalisms that adhere to it. All three belong to the non-anchored semantic graphs (*i.e.*, of type 2 according to the typology of Kuhlmann and Oepen, 2016), that is to say that there is no direct and explicit correspondence between the graph's nodes and the source tokens. AMR uses PropBank (Palmer et al., 2005), a resource whose concepts are represented by frames and whose relations are symbolized by an enumeration of arguments noted `:ARG0`, `:ARG1`, etc. This

[3]Other works have enriched UCCA at different levels: role labeling of core (Shalev et al., 2019) and non-core (Prange et al., 2019a) arguments based on the supersenses of Schneider et al. (2018), refinement of implicit argument types (Cui and Hershcovich, 2020), *inter alia*.

opacity has been preserved by UMR, but not by BMR, whose authors consider that it prevents an explicit understanding of the semantics attached to the relation. A selection of 25 of the 39 thematic roles of VerbNet (Schuler, 2005) were preferred.

Although these two resources (PropBank and VerbNet) are regularly chosen for relation labeling, other formalisms deviate from them. This is the case with UCCA, which exploits a smaller set of relations. UCCA is a multilingual formalism based on Basic Linguistic Theory (Dixon, 2010) that uses acyclic directed graphs (DAGs). Unlike AMR and its followers, UCCA is an anchored multi-layer formalism: for a given text, each token constitutes a leaf of the graph. The textual content is seen as a set of *Scenes* that can describe actions or states. Each *Scene* has a root node linked to the main relation (or main process) of the statement. To represent relations, the UCCA foundational layer has a dedicated set of only twelve labels, rendering the annotation process, according to the authors, quite simple, even for people without linguistic training. However, the semantics attached to a predicate's participants (all represented with the single label `A`) is far from fine-grained. In contrast, Universal Decompositional Semantics (UDS, White et al., 2016) is a formalism that does not use any discrete values to symbolize the relations between predicates and their arguments. Instead, the authors use proto-roles from Dowty (1991), which have numerical values appended to them. Instead of being labeled `Agent`, an argument can have a value related to its attributes `Awareness`, `Volition`, `Instigation`, etc. This representation, described as feature-based and opposed to traditional systems (White et al., 2020), has been extended to different phenomena, namely semantic typing of entities, factuality of events (Rudinger et al., 2018), genericity of entities and events (Govindarajan et al., 2019), and temporal relations between events (Vashishtha et al., 2019).

3 MR4AP's position

In this section, we focus on positioning MR4AP with respect to the other formalisms on the points that seem most salient to us.

Applicative aim. MR4AP is a formalism that has been designed with an industrial and, therefore, applicative aim. Although we base our choices on existing research works, we have made them with the requirement of being factual, meaning that the annotation should not be left to the subjective interpretation of the annotator. There should not be several possible annotations for the annotator to choose from. Therefore, despite the originality of their approach compared to other formalisms, we detach ourselves from UDS's choices of continuous representation, mainly because such representation using probabilities can make the annotation process complex and be difficult to assess accurately. Moreover, we move away from theoretical formalisms such as Discourse Representation Theory (DRT, Kamp et al., 1993) and Minimal Recursion Semantics (MRS, Copestake et al., 2005).

Genericity. MR4AP has been designed with *genericity* as its watchword. This applies both to the multilingual character of the representation and the invariance of the representations despite syntactic idiosyncrasies. Most recent formalisms aim at abstracting away from syntax, and MR4AP joins them on this point. Therefore, we detach ourselves from those that have a strong correlation with syntactic representations, as is the case of PDT-TL and UDS. On the same note, and although they are only notation variants (Oepen et al., 2019), the inverted arguments of AMR and its extensions (`:ARG0-of`) force parsers on the one hand to normalize relations (making graphs *de facto* multi-rooted), and on the other hand modify the graph, furthermore creating more cycles in supposedly acyclic graphs (Kuhlmann and Oepen, 2016). MR4AP being multi-rooted does not allow inverted arguments.

Explicitness. From our point of view, a meaning representation must be as explicit as possible. This explicitness is expressed at several levels. On the one hand, we agree with Di Fabio et al. (2019) on the need to name all relations between nodes: if a relation is not typed with a sufficiently specified label (UCCA), or is not usable without glossing (AMR/UMR), or is not represented by a discrete value (UDS), much of the semantics attached to the relationship is lost. Like BMR, MR4AP, therefore, uses a subset of VerbNet roles, to which some labels are added to specify temporal, spatial, discourse, and coreference relations. Likewise, it seems to us necessary to make entities' types as explicit as possible thanks to a label, mainly to avoid having to gloss their meanings.

Intra- and inter-sentence relations. We consider that a meaning representation would not be complete if it did not include the different relations that exist in a document at the intra- and inter-sentence levels. In this respect, MR4AP is close to UMR, because the latter includes a representation at the document level, although UMR's is parallel to the one at the sentence level while MR4AP's isn't. UMR's parallel document-level structure includes anaphora and coreference relations (between entities and between events), temporal relations between events, and modal relations, a representation unique to UMR and based on the work of Vigus et al. (2019). Thus, all discourse relations are excluded from their document-level representation, despite the carryover of the modal strength corresponding to the `:condition` and `:purpose` relations. MR4AP differs from UMR on several points regarding inter-sentence representation. On the one hand, there is no distinction between the two levels, which are perfectly inseparable. On the other hand, and this follows from this single structure, in addition to coreference and temporal relations, all discourse relations are represented at both levels, simply because they can occur in adjacent sentences (see section 4). Finally, modality is represented by an attribute linked to the predicate that is modified.

Attribute richness. Following Bonial et al. (2019), we believe that a meaning representation of the text must include a certain amount of information conveyed by the morphosyntax. Among this information, we can count grammatical tense, aspect, and number. It is precisely these three elements that are missing in AMR and that motivated BMR's authors to incorporate them in their formalism (Martínez Lorenzo et al., 2022), although in a minimal way. On the contrary, UMR adds a lot of complexity by introducing deep lattices, multiplying the possible labels for each phenomenon. That holds true in particular for aspectual values with twenty-three possible labels against two for BMR (`:ongoing +` and `:ongoing -`) and seven for MR4AP, based on UMR's work (`habitual`, `state`, `process`, `atelic process`, `activity`, `endeavor` and `performance`). This important multiplication of attributes and associated values is also visible for Universal Networking Language (UNL, Uchida et al., 1999) and PDT-TL. We prefer a smaller set of attributes while keeping those necessary for an objective and factual representation of the textual content.

4 MR4AP representation examples

In this section, we apply our formalism to represent three distinct examples. Each of these examples illustrates one or more parts of the formalism that we consider important.

4.1 Document-level representation and main points

The first example will be used to introduce the formalism in its broad outline. It will also allow us to demonstrate that a representation at the document level, taking into account all intra- and inter-sentence relations, is possible. It consists of the following three sentences:

1. *Luke and John are singing songs.*
2. *As a result, Mary cannot sleep.*
3. *She will reprimand them tomorrow morning.*

Predicate-argument structure. In Figure 1, the three main predicates (red squares with solid edges) are `vn:performance-26.7`, `vn:snooze-40.4`, and `vn:judgment-33`[4]. Each of these predicates is linked to its arguments by a thematic role (bold arcs). The conjunction of the proper nouns in (1) gives rise to the reification (red square with dotted edges) of an `:addition` node, whose arguments are the `:Agents` of the predicate `vn:performance-26.7`.

Inter-sentence relations. Inter-sentence relations are resolved at several levels. At the coreference level, the tokens *She* and *them* are linked to their respective antecedents (or to what symbolizes them), namely *Mary* and `:addition`, *via* the `:SameAs` relation. At the discourse level, the causal relation between the predicates `vn:performance-26.7` and `vn:snooze-40.4` is represented by the `:Cause` and `:Consequence` relations. At the temporal level, the predicate `vn:snooze-40.4` is linked by a relation `:TimeMax` to `vn:judgment-33`, *i.e.*, the former is realized before the latter.

Attributes. Each predicate and each entity has its own attributes (dotted arcs). The verbal predicates can have a modal value, an aspectual

[4]Even though those three are VerbNet's classes (hence the *vn:* prefix), MR4AP does not cling to one resource in particular. We consider that the formalism must remain at the conceptual level and that linking a specific resource to it would already be tantamount to instantiating it. This instantiation could be done from any resource, or even from any conjunction of resources, as is the case in Figure 2.

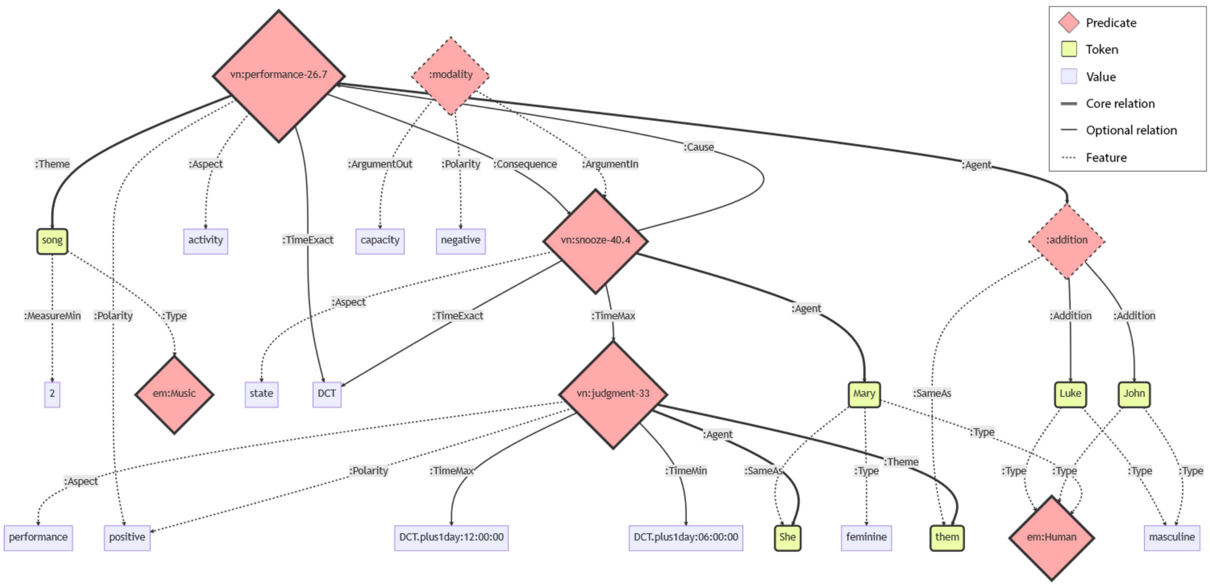

Figure 1: Document-level MR4AP representation example (see subsection 4.1). Note that all the attribute arcs are in fact by default reified, as is the case for the *:modality* node. Solely for readability reasons is it the only attribute node visibly reified in the graph. It is made so in order to take into account the negation's scope, hence the negative polar value for this node. The relations *:Argument*$\{In,Out\}$ are empty relations meant to link a node to its value. They are used for every reified attribute node. To be perfectly clear, triples like *("Luke" :Type "masculine")* are in fact always two triples such as the following: *(:type :ArgumentIn "Luke")* and *(:type :ArgumentOut "masculine")*.

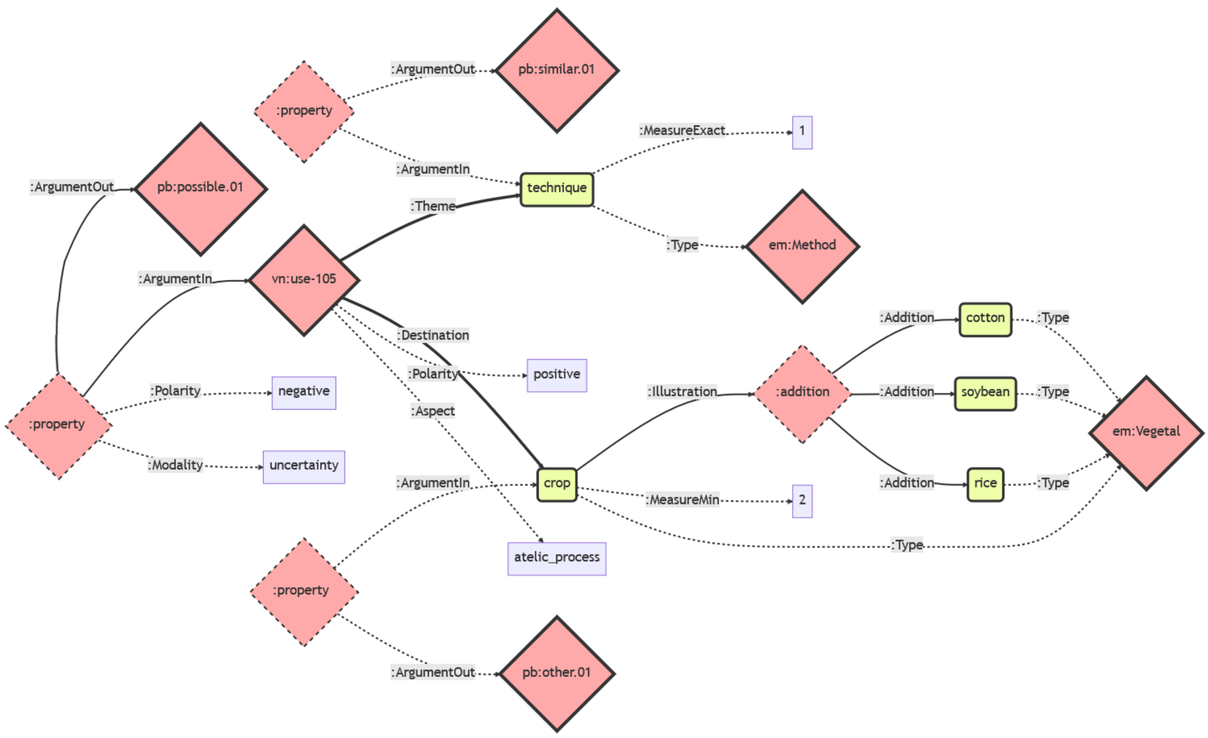

Figure 2: MR4AP representation example of MRP's running example (see subsection 4.2)

value, a polarity value, and temporal attributes (`:Time{Min,Max,Exact,Fuzzy,Duration}`). Nominal entities can have attributes related to their number, semantic type, and gender. Both semantic types (introduced by the `em:` prefix) and gender values are introduced with a `:Type` attribute.

4.2 MRP's running example

This subsection is dedicated to the representation of the running example used during the 2019 and 2020 Meaning Representation Parsing (MRP) shared tasks (Oepen et al., 2019, 2020). This will allow readers to more easily compare the frameworks that took part in those campaigns with our formalism. Here is the sentence:

4. *A similar technique is almost impossible to apply to other crops, such as cotton, soybean, and rice.*

This example was chosen because it presents a number of difficulties, namely a *tough* adjective *(impossible)*, a scopal adverb *(almost)*, and an appositive conjunction of more than two terms *(cotton, soybean, and rice)* illustrating a collection *(crops)* (Oepen et al., 2019).

***Tough* adjective.** *Tough* constructions (TCs) are a syntactic turn in which the logical object of an embedded non-finite verb is the main verb's syntactic subject (Hicks, 2009). In (4), the seemingly missing object of *to apply* is in fact the syntactic subject of *be (almost) impossible* (that is to say *technique)*. This can be paraphrased into two other configurations: either, acting as the subject, an expletive *it (it is almost impossible to apply a similar technique)* or an infinitival clause *(to apply a similar technique is almost impossible)*. To be as factual as possible and leave the annotator no choice, we always represent adjectives, whether attributive or predicative, with a `:property` attribute node linking the object and the adjective using the `:Argument{In,Out}` empty relations. Therefore, the attributive adjectives *other* and *similar*, which respectively trigger the `pb:other.01` and `pb:similar.01` nodes, are linked to *crop* and *technique via* `:property` nodes. As for *impossible*, it is treated in the same way. We consider that *to apply a technique is impossible* is similar to *the impossible application of a technique*. Thus, the two surface forms should produce the same graph. As a result, we link the `pb:possible.01` node to `vn:use-105` *via* a `:property` node. Also note that said node has a negative polar value triggered by its source token's prefix *(im-)*.

Scopal adverb. Regarding the scopal adverb *almost*, it modifies the adjective *impossible* and makes it uncertain. Consequently, the modal value `uncertainty` is added to the `:property` node linked to `pb:possible.01`. It should be remembered that the `:Modality` relation is in fact a reified node by default. Therefore, had the adverb been preceded by a negation particle (*i.e.*, *not almost impossible*), the reified `:modality` node would have had a negative polar value.

Enumeration in apposition. This representation does not differ from the conjunction of proper nouns in (1). An `:addition` node is reified, and each of the terms of the enumeration is linked to this node by an `:Addition` relation. The term referring to the collection of these examples is *crops*, and the `:addition` node is linked to it by the discourse relation `:Illustration` (*such as*).

4.3 Other difficult phenomena

The last example focuses on the representation of three arguably difficult elements: event coreference, interrogative sentences, and multiword expressions (MWEs), especially when they include event nominals. In addition, it helps to demonstrate the formalism's resistance to paraphrasing. The representation is the same for the following pair of paraphrases (see Figure 3):

5. *Who committed the murder of that police officer, and was it for revenge or for love?*
6. *Was this murder perpetrated out of revenge or out of love? And who killed that policeman?*

Event coreference. The event coreference appears in (5) between *(committed the) murder* and *it*, then in (6) between *murder (perpetrated)* and *killed*. In the same way that we represent coreference between entities in (1), we do not merge the nodes corresponding to each mention of the event, but rather join them with the `:SameAs` relation. The light verbs accompanying the event nominals are dropped from the representation. We discuss this further below.

Interrogative sentences. As for the representation of interrogative sentences, and especially that of unknown elements during the utterance of these sentences, we distinguish three types, to which are attached three different labels linked to the `:Type` attribute: polar questions (`question-closed`, whose answers are

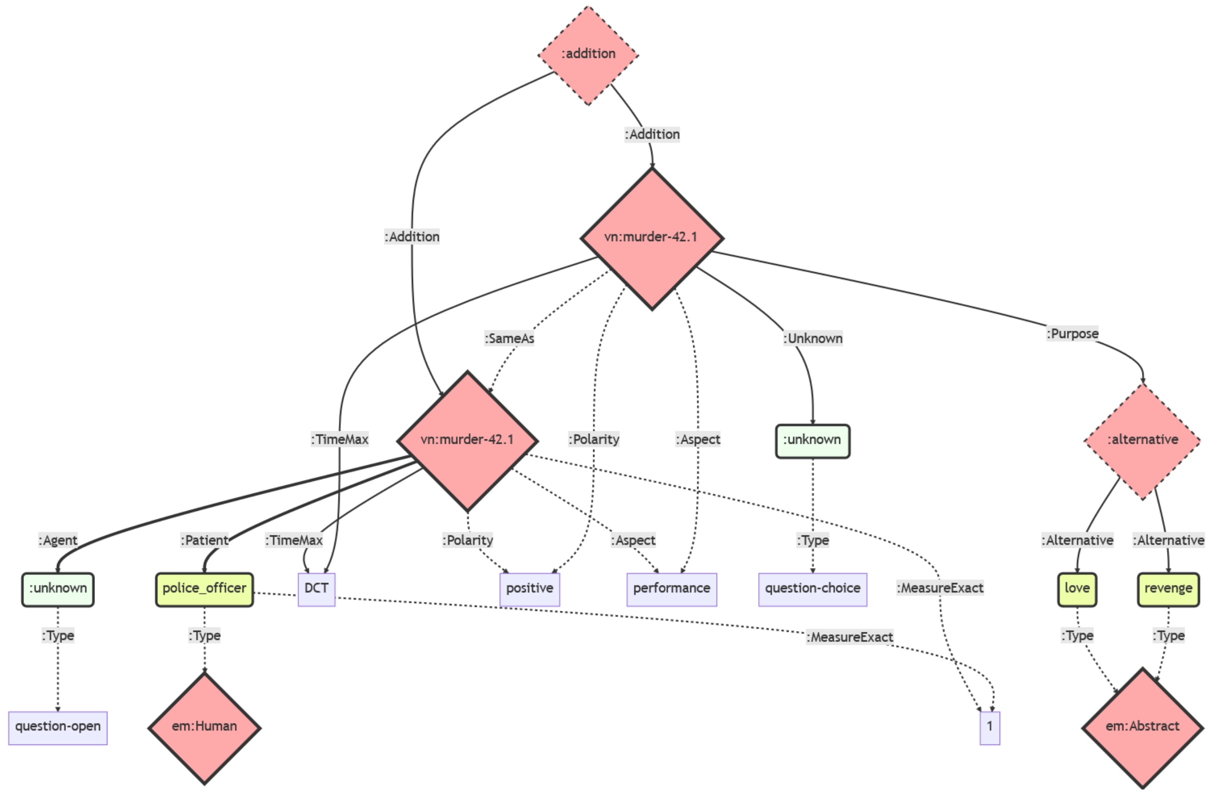

Figure 3: MR4AP representation example for subsection 4.3

either positive, negative, or doubtful), alternative questions (`question-choice`, whose answers are mentioned in the question with an alternative being offered), and variable questions (`question-open`, whose answers are quite open-ended, although governed by the nature of the unknown element). In examples (5-6), two elements are unknown: the perpetrator (*i.e.*, the `:Agent` of `vn:murder-42.1`) and the motive of the latter (*i.e.*, the possible choices linked to the `:alternative` node[5]).

For the first one, it is an `:Agent` relation to an `:unknown` node that is created from `vn:murder-42.1`. This same node is thus typed `question-open`. For the second one, the `:alternative` node, linked to its predicate with a `:Purpose` relation, offers a choice between two known options. To mark the interrogativeness attached to the `vn:murder-42.1` node, we use an `:Unknown` relation to point towards an `:unknown` node, itself pointing towards a `question-choice` value. Moreover, had the coordination of the questions not been made explicit by the conjunction *and* but had remained implicit with a paratactic conjunction, the representations would have remained identical. Our example does not display any polar questions, but had the question been *Was the police officer murdered?*, the `vn:murder-42.1` node would have had an `:Unknown` relation pointing towards an `:unknown` node whose `:Type` value would have been `question-closed`.

Multiword expressions. MWEs are known to be a *pain in the neck* (Sag et al., 2002), both because of their heterogeneity and their pervasiveness. We choose to consider them non-compositionally as most MWEs' meaning can not be broken down according to their constituents (Constant et al., 2017). Considering an MWE as a single semantic entity enables a greater graph similarity from one language to another (Navigli et al., 2022), or even when comparing a set of paraphrases. For instance, in (5), *police officer* (which could have been translated into *officier de police*, *agente di polizia*, or even *ḍābit aš-šurṭa* in French, Italian, and Arabic respectively) becomes in (6) the single word token *policeman* (which could have been translated into *policier*, *poliziotto*, or even *šurṭiyy*).

[5] Such nodes also appear in non-interrogative statements to represent disjunctions *(I reckon John wrote novels* ***or*** *poems)*.

Having a single node for both expressions of the same real world's concept seems mandatory as far as uniformity is concerned. MWEs can be the source of a relation (the prepositional locution *out of* for `:Purpose`), of a non-event entity (the compound *police officer*), or of an event (nominal or verbal, like the light verb construction *to commit murder*). In the latter case, the light verb is simply dropped because it is redundant, but the accompanying event nominal inherits the arguments and the various linguistic features attached to it. For example, *committed* in (5) is a preterite verb, hence the `:TimeMax` relation towards the DCT; it denotes a completed action that led to a result, hence the `performance` aspectual value; and there is no negation particle, hence the `positive` polar value. The *police officer* is the `:Patient` of the murder rather than its commission's, and the unknown subject is its `:Agent`. The `vn:murder-42.1` node actually inherits the arguments and linguistic features of what could have been a `vn:complete-55.2` node with the former as its `:Theme`.

5 Annotated corpora

We first applied MR4AP to French short sentences from the TaPaCo corpus (Scherrer, 2020) to ensure its usability on low-complexity sentences. We then applied it to more complex sentences from Wikipedia articles in five languages to also test its multilingual compatibility. Both datasets are available on the GitHub repository[6].

MR4AP-tapaco. In order to demonstrate the viability of our formalism, we produced an annotated dataset. Version 0.1 of the MR4AP-tapaco corpus[7] is relatively small, but it is bound to grow as contributions are made. So far, 100 short sentences in French from the TaPaCo paraphrase corpus have been automatically annotated using our own tool before being manually checked and validated. Choosing paraphrases is not insignificant as it will allow us to gauge the similarity of the graphs obtained after annotating sets of paraphrases once we have enough data. Some statistics regarding this corpus can be found on the remote repository.

Multilingual compatibility experiment. In order to validate MR4AP's multilingual compatibility in practice, as well as to explore ways to propose a protocol and a manual annotation tool, we started a first small-scale annotation project. First, we wrote guidelines[8] that present MR4AP in an exhaustive and extensive way. In a second step, to avoid content bias, we randomly selected five Wikipedia articles in French and kept the first three sentences of each. After setting up all the necessary parameters in the INCEpTION platform (Klie et al., 2018), our annotation tool of choice that allows the annotation of both explicit and implicit elements, we annotated the five texts. In a third step, we automatically translated them into English, Spanish, Italian and Modern Standard Arabic (MSA), and annotated them[9]. Despite the small scale of the annotation effort and the relatively modest language panel considered, we were able to determine that MR4AP seems to be multilingually compatible. Pursuing the annotation effort is however mandatory.

MR4AP-wikipedia. Having established that the formalism can be used with different languages, the five annotated texts in French, English, Spanish, Italian and MSA constitute the first annotated texts for the MR4AP-wikipedia corpus[10]. The objective is to obtain a fully manually annotated dataset that would serve as a gold standard. This dataset will be regularly enriched with new annotated texts.

Data format. We use JSON files with three fields: `id` (the document identifier), `text` (the document's textual content), and `rdf` (the RDF[11] representation of the text). An RDF data model consists of RDF triples where each RDF triple codifies a statement in the form of subject–predicate–object expressions. RDF triples have no ordering and triples can be linked to other triples according to their common elements (so that a graph is obtained). Using RDF, we make the assumption that regardless of the order in which the sentences are written, the text will systematically produce the same semantic graph (*i.e.*, the same set of triples). Finally, RDF triples applied with OWL[12] can be used as input for a reasoner that in turn could be used to saturate the graph with inferred annotations. Our dataset's RDF graphs can be viewed with an appli-

[6] https://github.com/Emvista/MR4AP/tree/main/corpora

[7] https://github.com/Emvista/MR4AP/tree/main/corpora/MR4AP-tapaco

[8] https://github.com/Emvista/MR4AP/tree/main/guidelines/guidelines.md

[9] The annotation was carried out by the two authors.

[10] https://github.com/Emvista/MR4AP/tree/main/corpora/MR4AP-wikipedia

[11] Resource Description Framework: https://www.w3.org/RDF/

[12] Web Ontology Language: https://www.w3.org/OWL/

cation such as Protégé[13] (Musen, 2015).

6 Limitations and perspectives

Although our formalism is able to address some shortcomings that other formalisms can't, some limitations remain. On the one hand, given all the elements that MR4AP represents, annotation remains a time-consuming and rather complex process. Moreover, this complexity only increases with the length of the texts. The multilingual annotation experiment described above only exacerbated the need for a perhaps more efficient annotation strategy.

On the other hand, while the annotation was carried out on texts in five different languages, the variety was somewhat limited: three of them are Romance languages with few differences; English is a standard; only MSA, due to its important differences with the other four languages, really allows us to conclude that MR4AP seems compatible with multilingualism. Continuing the annotation effort with languages from different families or low-resource languages would enable us to support this assertion.

Moreover, from a cross-formalism perspective and following the recent mapping effort made between AMR and UMR (Bonn et al., 2023), we would like to follow suit and align MR4AP to these two formalisms. This mapping would allow the production of a multi-formalism corpus, which could in turn allow the implementation of comparative experiments on the performance of each formalism from the same source material.

Important questions remain to be tackled: Which tool to use/develop to annotate more efficiently with a formalism such as MR4AP? And how to ensure annotation completeness for a given text? Does the level of anchoring have an impact on the explainability of semantic parsers (*e.g.*, to source graph nodes)? Knowing that graph linearization is an important topic and that edge ordering can have a "big negative effect" on the evaluation measures of some tasks (Bevilacqua et al., 2021), is RDF appropriate and what impact would this have on the performance of state-of-the-art parsers?

7 Conclusion

We have highlighted the divergences and convergences between ten meaning representation formalisms. On this basis, we have put forth and positioned MR4AP, our application-oriented formalism. We have extensively described it both through guidelines and through several examples demonstrating its efficiency in representing meaning at the document level by taking into account discourse, coreference and temporal relations, its potential to represent some of the most complex linguistic phenomena, and its robustness to paraphrasing and multilingualism. We have also briefly presented the first version of the MR4AP-tapaco corpus as well as a first small-scale manual annotation effort to assert the multilingual compatibility of the formalism in practice. We concluded that MR4AP is usable regardless of the text's language, and this annotation effort allowed us to create the MR4AP-wikipedia corpus, which will serve as a gold standard. Note that a hybrid semantic parser, which does not need any training data to annotate textual content, has been developed with this formalism, is already in production, and will be the subject of a future publication.

[13]Protégé: `https://protege.stanford.edu/`

Acknowledgements

We would like to warmly thank the three anonymous reviewers who had us thinking and helped in improving the content of the article thanks to their accurate and thorough feedback.

References

Omri Abend and Ari Rappoport. 2013. Universal Conceptual Cognitive Annotation (UCCA). In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 228–238, Sofia, Bulgaria. Association for Computational Linguistics.

Omri Abend and Ari Rappoport. 2017. The state of the art in semantic representation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 77–89, Vancouver, Canada. Association for Computational Linguistics.

Lasha Abzianidze, Johannes Bjerva, Kilian Evang, Hessel Haagsma, Rik van Noord, Pierre Ludmann, Duc-Duy Nguyen, and Johan Bos. 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*, pages 242–247, Valencia, Spain. Association for Computational Linguistics.

Lasha Abzianidze and Johan Bos. 2019. Thirty musts for meaning banking. In *Proceedings of the First In-*

ternational Workshop on Designing Meaning Representations*, pages 15–27, 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.

Valerio Basile, Johan Bos, Kilian Evang, and Noortje Venhuizen. 2012. Developing a large semantically annotated corpus. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 3196–3200, Istanbul, Turkey. European Language Resources Association (ELRA).

Michele Bevilacqua, Rexhina Blloshmi, and Roberto Navigli. 2021. One spring to rule them both: Symmetric amr semantic parsing and generation without a complex pipeline. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 12564–12573.

Claire Bonial, Lucia Donatelli, Stephanie M. Lukin, Stephen Tratz, Ron Artstein, David Traum, and Clare Voss. 2019. Augmenting Abstract Meaning Representation for human-robot dialogue. In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 199–210, Florence, Italy. Association for Computational Linguistics.

Julia Bonn, Skatje Myers, Jens E. L. Van Gysel, Lukas Denk, Meagan Vigus, Jin Zhao, Andrew Cowell, William Croft, Jan Hajič, James H. Martin, Alexis Palmer, Martha Palmer, James Pustejovsky, Zdenka Urešová, Rosa Vallejos, and Nianwen Xue. 2023. Mapping AMR to UMR: Resources for adapting existing corpora for cross-lingual compatibility. In *Proceedings of the 21st International Workshop on Treebanks and Linguistic Theories (TLT, GURT/SyntaxFest 2023)*, pages 74–95, Washington, D.C. Association for Computational Linguistics.

Mathieu Constant, Gülşen Eryiğit, Johanna Monti, Lonneke van der Plas, Carlos Ramisch, Michael Rosner, and Amalia Todirascu. 2017. Survey: Multiword expression processing: A Survey. *Computational Linguistics*, 43(4):837–892.

Ann Copestake, Dan Flickinger, Carl Pollard, and Ivan A Sag. 2005. Minimal recursion semantics: An introduction. *Research on language and computation*, 3:281–332.

Ruixiang Cui and Daniel Hershcovich. 2020. Refining implicit argument annotation for UCCA. In *Proceedings of the Second International Workshop on Designing Meaning Representations*, pages 41–52, Barcelona Spain (online). Association for Computational Linguistics.

Andrea Di Fabio, Simone Conia, and Roberto Navigli. 2019. VerbAtlas: a novel large-scale verbal semantic resource and its application to semantic role labeling. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 627–637, Hong Kong, China. Association for Computational Linguistics.

Robert MW Dixon. 2010. *Basic linguistic theory volume 2: Grammatical topics*, volume 2. Oxford University Press on Demand.

Lucia Donatelli, Michael Regan, William Croft, and Nathan Schneider. 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)*, pages 96–108, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

David Dowty. 1991. Thematic proto-roles and argument selection. *language*, 67(3):547–619.

Jeffrey Flanigan, Ishan Jindal, Yunyao Li, Tim O'Gorman, Martha Palmer, and Nianwen Xue. 2022. Meaning representations for natural languages: Design, models and applications. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: Tutorial Abstracts*, pages 1–8, Abu Dubai, UAE. Association for Computational Linguistics.

Venkata Govindarajan, Benjamin Van Durme, and Aaron Steven White. 2019. Decomposing generalization: Models of generic, habitual, and episodic statements. *Transactions of the Association for Computational Linguistics*, 7:501–517.

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).

Daniel Hershcovich, Nathan Schneider, Dotan Dvir, Jakob Prange, Miryam de Lhoneux, and Omri Abend. 2020. Comparison by conversion: Reverse-engineering UCCA from syntax and lexical semantics. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2947–2966, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Glyn Hicks. 2009. Tough-constructions and their derivation. *Linguistic Inquiry*, 40(4):535–566.

Hans Kamp, Uwe Reyle, Hans Kamp, and Uwe Reyle. 1993. Tense and aspect. *From Discourse to Logic:*

Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory, pages 483–689.

Jan-Christoph Klie, Michael Bugert, Beto Boullosa, Richard Eckart de Castilho, and Iryna Gurevych. 2018. The INCEpTION platform: Machine-assisted and knowledge-oriented interactive annotation. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9, Santa Fe, New Mexico. Association for Computational Linguistics.

Alexander Koller, Stephan Oepen, and Weiwei Sun. 2019. Graph-based meaning representations: Design and processing. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, pages 6–11, Florence, Italy. Association for Computational Linguistics.

Marco Kuhlmann and Stephan Oepen. 2016. Squibs: Towards a catalogue of linguistic graph Banks. *Computational Linguistics*, 42(4):819–827.

Abelardo Carlos Martínez Lorenzo, Marco Maru, and Roberto Navigli. 2022. Fully-Semantic Parsing and Generation: the BabelNet Meaning Representation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1727–1741, Dublin, Ireland. Association for Computational Linguistics.

Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolářová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, et al. 2006. Annotation on the tectogrammatical level in the prague dependency treebank. *annotation manual. Technical Report*, 30:5–11.

Mark A Musen. 2015. The protégé project: a look back and a look forward. *AI matters*, 1(4):4–12.

Roberto Navigli, Rexhina Blloshmi, and Abelardo Carlos Martinez Lorenzo. 2022. Babelnet meaning representation: A fully semantic formalism to overcome language barriers. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 12274–12279.

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, 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, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

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

Siyana Pavlova, Maxime Amblard, and Bruno Guillaume. 2022. How much of UCCA can be predicted from AMR? In *Proceedings of the 18th Joint ACL - ISO Workshop on Interoperable Semantic Annotation within LREC2022*, pages 110–117, Marseille, France. European Language Resources Association.

Jakob Prange, Nathan Schneider, and Omri Abend. 2019a. Made for each other: Broad-coverage semantic structures meet preposition supersenses. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 174–185, Hong Kong, China. Association for Computational Linguistics.

Jakob Prange, Nathan Schneider, and Omri Abend. 2019b. Semantically constrained multilayer annotation: The case of coreference. In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 164–176, Florence, Italy. Association for Computational Linguistics.

Rachel Rudinger, Aaron Steven White, and Benjamin Van Durme. 2018. Neural models of factuality. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 731–744, New Orleans, Louisiana. Association for Computational Linguistics.

Ivan A Sag, Timothy Baldwin, Francis Bond, Ann Copestake, and Dan Flickinger. 2002. Multiword expressions: A pain in the neck for nlp. In *Computational Linguistics and Intelligent Text Processing: Third International Conference, CICLing 2002 Mexico City, Mexico, February 17–23, 2002 Proceedings 3*, pages 1–15. Springer.

Yves Scherrer. 2020. TaPaCo: A corpus of sentential paraphrases for 73 languages. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 6868–6873, Marseille, France. European Language Resources Association.

Nathan Schneider, Jena D. Hwang, Vivek Srikumar, Jakob Prange, Austin Blodgett, Sarah R. Moeller, Aviram Stern, Adi Bitan, and Omri Abend. 2018. Comprehensive supersense disambiguation of English prepositions and possessives. In *Proceedings* *of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 185–196, Melbourne, Australia. Association for Computational Linguistics.

Karin Kipper Schuler. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. University of Pennsylvania.

Adi Shalev, Jena D. Hwang, Nathan Schneider, Vivek Srikumar, Omri Abend, and Ari Rappoport. 2019. Preparing SNACS for subjects and objects. In *Proceedings of the First International Workshop on Designing Meaning Representations*, pages 141–147, Florence, Italy. Association for Computational Linguistics.

Hiroshi Uchida, Meiying Zhu, and Tarcisio Della Senta. 1999. A gift for a millennium. *IAS/UNU, Tokyo*.

Jens EL Van Gysel, Meagan Vigus, Jayeol Chun, Kenneth Lai, Sarah Moeller, Jiarui Yao, Tim O'Gorman, Andrew Cowell, William Croft, Chu-Ren Huang, et al. 2021. Designing a uniform meaning representation for natural language processing. *KI-Künstliche Intelligenz*, 35(3-4):343–360.

Siddharth Vashishtha, Benjamin Van Durme, and Aaron Steven White. 2019. Fine-grained temporal relation extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2906–2919, Florence, Italy. Association for Computational Linguistics.

Meagan Vigus, Jens E. L. 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, Florence, Italy. Association for Computational Linguistics.

Aaron Steven White, Drew Reisinger, Keisuke Sakaguchi, Tim Vieira, Sheng Zhang, Rachel Rudinger, Kyle Rawlins, and Benjamin Van Durme. 2016. Universal decompositional semantics on Universal Dependencies. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1713–1723, Austin, Texas. Association for Computational Linguistics.

Aaron Steven White, Elias Stengel-Eskin, Siddharth Vashishtha, Venkata Subrahmanyan Govindarajan, Dee Ann Reisinger, Tim Vieira, Keisuke Sakaguchi, Sheng Zhang, Francis Ferraro, Rachel Rudinger, Kyle Rawlins, and Benjamin Van Durme. 2020. The universal decompositional semantics dataset and decomp toolkit. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 5698–5707, Marseille, France. European Language Resources Association.

Zdeněk Žabokrtský, Daniel Zeman, and Magda Ševčíková. 2020. Sentence meaning representations across languages: What can we learn from existing frameworks? *Computational Linguistics*, 46(3):605–665.