

# Fully-Semantic Parsing and Generation: the BabelNet Meaning Representation

Abelardo Carlos Martínez Lorenzo<sup>1,2</sup>, Marco Maru<sup>1</sup>, and Roberto Navigli<sup>1</sup>

<sup>1</sup>Sapienza NLP Group  
Sapienza University of Rome

<sup>2</sup>Babelscape, Rome

first.lastname@uniroma1.it, surname@babelscape.com

## Abstract

A language-independent representation of meaning is one of the most coveted dreams in Natural Language Understanding. With this goal in mind, several formalisms have been proposed as frameworks for meaning representation in Semantic Parsing. And yet, the dependencies these formalisms share with respect to language-specific repositories of knowledge make the objective of closing the gap between high- and low-resourced languages hard to accomplish. In this paper, we present the BabelNet Meaning Representation (BMR), an interlingual formalism that abstracts away from language-specific constraints by taking advantage of the multilingual semantic resources of BabelNet and VerbAtlas. We describe the rationale behind the creation of BMR and put forward BMR 1.0, a dataset labeled entirely according to the new formalism. Moreover, we show how BMR is able to outperform previous formalisms thanks to its fully-semantic framing, which enables top-notch multilingual parsing and generation. We release the code at <https://github.com/SapienzaNLP/bmr>.

## 1 Introduction

Natural Language Understanding (NLU) enables machines to understand human language. A key enabling task in NLU is that of Semantic Parsing, whose longed-for dream is that of developing a formalism that can be used as an interlingual representation of meaning, i.e., one that, independently of the language, can explicitly embed sentence meaning into a machine- and human-readable form (Navigli, 2018). To this end, different formalisms such as Abstract Meaning Representation (Banarescu et al., 2013, AMR), Universal Conceptual Cognitive Annotation (Abend and Rappoport, 2013, UCCA) and Universal Meaning Representation (Van Gysel et al., 2021, UMR), have been proposed over the years.

As of now though, AMR is the most popular

formalism for Semantic Parsing, being widely applied to a variety of areas of NLP, such as Machine Translation (Song et al., 2019), Question Answering (Lim et al., 2020; Bonial et al., 2020b; Kapaniathi et al., 2021), Human-Robot Interaction (Bonial et al., 2020a), Text Summarization (Hardy and Vlachos, 2018; Liao et al., 2018) and Information Extraction (Rao et al., 2017).

The primary precept of AMR is that different sentences carrying the same meaning should have the same graph representation. Nonetheless, a few inherent properties of AMR make it inappropriate for the purpose of providing a language-agnostic representation of meaning. In fact, nodes within AMR graphs are represented by means of either English lemmas or OntoNotes frames (Hovy et al., 2006) which, in turn, are based on PropBank (Kingsbury and Palmer, 2002). The issue with lemmas is that they are merely surface forms devoid of semantics, whereas, with respect to frames, even though analogous repositories exist in other languages such as AnCora for Spanish (Aparicio et al., 2008) or the Chinese PropBank (Xue and Palmer, 2009), they are not mutually interlinked, hence making the cross-lingual application of AMR arduous to achieve (Conia et al., 2021).

Against this background, we follow the ideas put forward by Navigli et al. (2022) and develop the BabelNet Meaning Representation (BMR), a formalism providing the building blocks for a language-agnostic representation of meaning by exploiting the wealth of multilingual knowledge contained in BabelNet (Navigli and Ponzetto, 2010; Navigli et al., 2021)<sup>1</sup> and VerbAtlas (Di Fabio et al., 2019)<sup>2</sup>.

In outline, the main contributions of this paper are as follows: (i) we introduce BMR, a new Semantic Parsing formalism that can be used as an interlingua, (ii) we produce BMR 1.0, i.e., the first

<sup>1</sup><https://babelnet.org/>

<sup>2</sup><https://verbatlas.org/>

lexical-semantic dataset annotated according to the BMR formalism, (iii) we create and release models that can generate BMR graphs from text and text from BMR graphs in English, German, Spanish, and Italian, and (iv) we describe a sound experimental setup to show how, thanks to its fully semantic framing, BMR outdoes previous formalisms in both preserving and encoding textual information, as well as in being used as an interlingua in downstream tasks such as Machine Translation.

## 2 Related Work

Even though the vast majority of formalisms for Semantic Parsing have been designed with English in mind, several approaches have attempted to narrow the gap between English and other languages. For instance, Universal Conceptual Cognitive Annotation (Abend and Rappoport, 2013, UCCA) was proposed as a cross-lingual annotation formalism in which words in a sentence are connected using semantic relations not tied to specific languages. And yet, while UCCA reflects the semantic relations between nodes via a set of coarse-grained roles, it represents concepts by means of simple lemmas, hence preventing an abstraction from language-specific constraints. Parallel Meaning Bank (Abzianidze et al., 2017, PMB), an approach based on the Discourse Representation Theory (Kamp and Reyle, 1993, DRT),<sup>3</sup> also emerged. In PMB, English sentences are parsed with labels that are automatically projected to non-English translations. PMB too, however, cannot be seen as a unified interlingual representation, since it uses English-specific meaning repositories.

As regards Abstract Meaning Representation (Banarescu et al., 2013, AMR), instead, several approaches have tried to adapt it for cross-lingual use. As a case in point, Xue et al. (2014) analyzed the viability of tailoring the AMR formalism to fit other languages by making use of language-specific repositories similar to PropBank (Aparicio et al., 2008; Xue and Palmer, 2009).<sup>4</sup> On a different note, Damonte and Cohen (2018) and Biloshmi et al. (2020) attempted to adopt AMR as an interlingual formalism, despite its English-centric nature, by assuming that the AMR graph of an English sentence is also representative of translations of that sentence in other languages. Once again, these

approaches testify to the limits of AMR as an interlingua, given the drawbacks of dealing with structural divergences among different languages. In recent years, Zhu et al. (2019) have recommended abstracting the AMR formalism away in order to reduce its language-specific complexity by preserving just the predicate roles and relations that constitute the core semantic information of sentences. Conversely, rather than decreasing the complexity of AMR, the Universal Meaning Representation (Van Gysel et al., 2021, UMR) extends it by including new features that render the formalism less tied to a specific language. In particular, UMR enriches the verbal predicates with information about grammatical aspect and scope, while introducing temporal and modal dependencies at the document level. Finally, it enhances AMR to use it as a cross-lingual formalism by employing language-specific repositories and relations. Yet, the focus of UMR is that of providing languages with the necessary resources to parse texts, rather than being an interlingual representation.

In contrast to previous approaches, and thanks to the multilingually-shared word meanings and semantic roles taken from the interlinked repositories of BabelNet (Navigli et al., 2021) and VerbAtlas (Di Fabio et al., 2019), we put forward BMR, a formalism that fully detaches from syntax and thus stands as a lexical-semantic representation that is able to bring different languages together.

## 3 Preliminaries

To accomplish the goal of an interlingual meaning representation, we disconnect our formalism from language-specific constraints of any kind. To this end, we draw on resources that inherently connect word meanings and predicate-argument structures across languages, i.e., BabelNet and VerbAtlas.

**BabelNet** (Navigli et al., 2021) is a multilingual encyclopedic dictionary and semantic knowledge base in which concepts are represented as synsets (sets of synonyms that convey the same meaning), linked via semantic relation edges like hypernymy or meronymy. BabelNet was built by the aggregation of several knowledge resources including WordNet (Fellbaum, 1998), Wikipedia and Wiktionary, resulting in a remarkable ontology of concepts and named entities covering 500 languages. Given its versatility, which makes it suitable for a wide range of tasks across languages, we employ its most recent version 5.0 as a tool to switch the

<sup>3</sup>DRT is a framework that embeds the semantics of an utterance employing a formal logic semantic structure.

<sup>4</sup>For a similar approach, see also the Prague Tectogrammatical Graphs of Hajič et al. (2012, PTG).

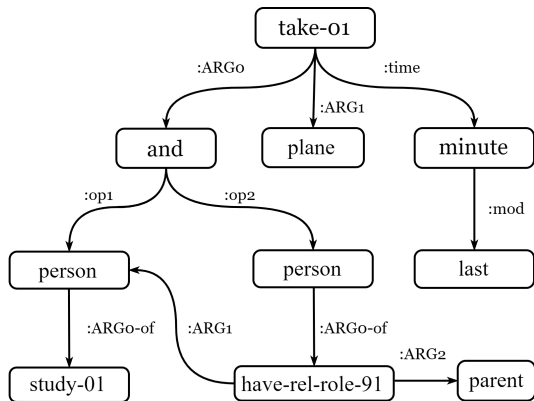


Figure 1: AMR graph for the sentence “*The students and their parents will take the plane at the last minute*”.

focus of Semantic Parsing formalisms from words to multilingual concepts.

**VerbAtlas** (Di Fabio et al., 2019) is a manually-curated lexical-semantic inventory that collapses the BabelNet verbal synsets into around 450 semantically-coherent frames, each defining prototypical argument structures via human-readable relationships (e.g. AGENT, THEME). Thanks to its linkage to BabelNet, VerbAtlas represents the best option for handling predicate-argument relations in BMR in a language-independent manner.

## 4 BabelNet Meaning Representation

Like AMR, BMR embeds the semantics of a sentence in a directed acyclic graph, with nodes and edges connecting them. However, where AMR relies on English lemmas and OntoNotes frames to represent nodes and relations (see Figure 1), BMR disposes of language-specific constraints, and employs multilingual concepts and self-explanatory semantic roles (see Figure 2).<sup>5</sup> In what follows (Sections 4.1 to 4.4), we will describe and detail the features that make BMR stand out with respect to a widely-employed Semantic Parsing formalism such as AMR, as well as their integration into the AMR 3.0 dataset (Knight et al., 2020) to produce the BMR 1.0 dataset.<sup>6</sup>

### 4.1 Self-explanatory Semantic Relations

As briefly mentioned in Section 1, AMR derives its coarse-grained frames and argument structures

<sup>5</sup>Appendix A details how to read BMR graphs.

<sup>6</sup>AMR 3.0 is licensed by LDC at <https://catalog.ldc.upenn.edu/LDC2020T02>. For this reason, we do not make the BMR-annotated dataset (BMR 1.0) publicly available, but rather provide tools to convert the original AMR 3.0 dataset, provided its rightful ownership.

from the English PropBank section of OntoNotes, a repository which is circumscribed to the English language and that features semantic relations that are both predicate-specific and largely unintelligible without a gloss. For example, in Figure 1, the subgraph representation of *students’ parents* is pivoted on the frame `have-rel-role-91`, where the relations `:ARG0`, `:ARG1`, and `:ARG2` identify the first entity, the second entity, and the role of the first entity, respectively. As importantly, even though language-specific repositories similar to PropBank have been used to annotate non-English sentences with structures comparable to those of AMR (Aparicio et al., 2008; Xue and Palmer, 2009), there is not an exact one-to-one mapping between the frames they define, meaning that, e.g., the frame `have-rel-role-91` might not be featured in the other inventories. Therefore, with the aim of overcoming language specificity, we replace PropBank with VerbAtlas as an alternative repository of predicate-argument structure information, which, as explained above, inherently accounts for multilingually-shared semantics.

To build the BMR 1.0 dataset, we exploit the mapping provided by Di Fabio et al. (2019), which links VerbAtlas frames and arguments to PropBank, and use it to replace the original frames and semantic roles in the AMR 3.0 dataset with those of VerbAtlas (e.g., the frame `take-01` corresponds to `MOVE_BY_MEANS_OF` in VerbAtlas, and its `ARG0` to `AGENT`). However, this mapping is incomplete and, as a result, several predicates found within AMR 3.0 can not be transitioned directly. Among these, two kinds of predicates can be identified, (i) predicates that OntoNotes labels as verbal, and (ii) non-verbal predicates and special predicates which AMR uses to define special semantic structures (e.g., `have-rel-role-91`). To deal with these predicates, we asked a linguist<sup>7</sup> to create a mapping between PropBank and VerbAtlas for the missing verbal predicates, and, with respect to the others instead, to map them to BMR adapting previous semantic roles and creating new ones to better accommodate their argument structures.<sup>8</sup>

### 4.2 Node Merging

Multiword expressions and idioms are rendered word by word in AMR, using node composition.

<sup>7</sup>Annotators share effective operational English proficiency and received a wage in line with their country of residence.

<sup>8</sup>See Appendix B for the list of BMR semantic roles, and Appendix C for mapping examples.

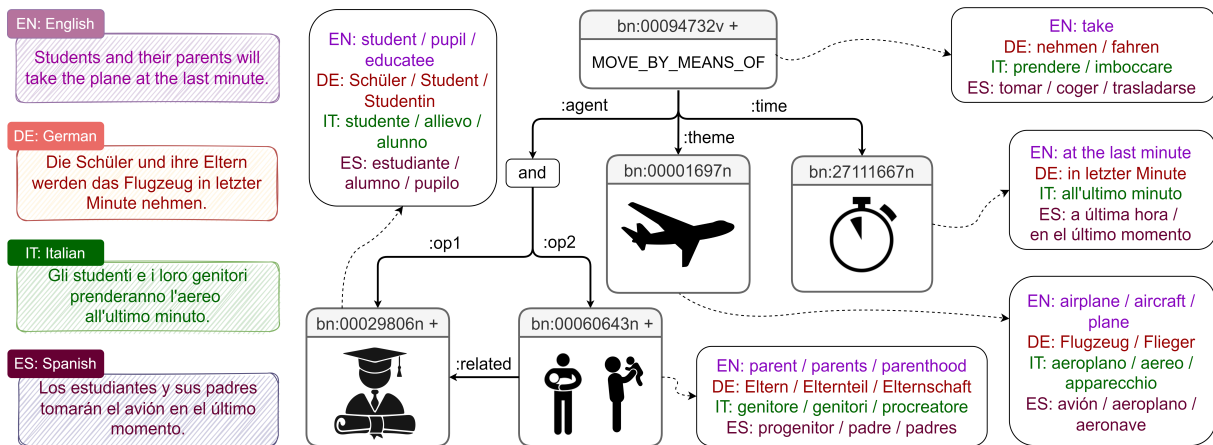


Figure 2: Equivalent sentences in different languages (left) and their shared BMR graph (right).

Nevertheless, such an approach is not feasible for an interlingual representation, since the overall meaning of an expression can not, as a general rule, be compositionally inferred from the meanings of its individual words. Therefore, in BMR we make use of the available BabelNet synsets to identify the meaning of a multiword expression or idiom, and hence we represent it with a single node. As a case in point, the idiom *at the last minute* which, according to Wiktionary, is defined as “very close to a deadline or potential crucial event”, does not entail that something will happen precisely *in* the last minute. This exact expression, that in AMR 3.0 is represented using two nodes (*m* and *l*) as:

$$(m / \textit{minute} :mode (l / \textit{last}))$$

appears in BMR as a single node *m*:

$$(m / \textit{at\_the\_last\_minute} / bn:00114428r)$$

As a result, we are both able to (i) abstract away from language-specific lexicons making use of concepts connected across languages and, concurrently, to (ii) reduce the graph density, hence easing the computational burden for systems.

Another intrinsic limit of AMR as an interlingual representation is that, since the meaning of nodes can only be partially identified using OntoNotes frames, AMR maximizes their usage so as to express as many concepts as possible, even non-verbal ones. The main reason this constitutes an issue is that the OntoNotes frame composition used to define a concept and the concept itself are not semantically equivalent. For example, the concept of *student*, which AMR represents as “a person who studies” by means of the connection between the node of *person* with the OntoNotes frame

*study-01*, is arguably different from the definition of *student* as, quoting the BabelNet synset gloss, “a learner who is enrolled in an educational institution”. Additionally, these language-specific rules are not transferable across languages, and they are not consistent even within AMR itself, as, whenever a verbalization is not viable (AMR does not render *professor* as “a person who professes”), the word is included in the graph as it is.

In the remainder of this Section, we describe the strategies by means of which we remodel AMR 3.0 to obtain BMR 1.0 employing node merging.

**Multiword Expression Identification** To merge nodes, we must first identify the words or multiword expressions that are represented by several nodes in the AMR graph. In this regard, we proceed by lemmatizing the original sentences in AMR 3.0 using the 3.1 version of the SpaCy software library (Honnibal and Johnson, 2015). At this stage, for each sentence, we check for the longest concatenations of lemmas that match a BabelNet synset lexicalization in BabelNet 5.0. Once the expressions have been identified, we use the automatic AMR aligner of Flanigan et al. (2014) to get the alignments between the tokens in the original sentence (and, consequently, the identified words and multiwords) and the graph nodes.

**Manual Validation** The automatic identification of multiwords can be noisy and lead to poor node merging choices which, in turn, can result in wrong sense attributions. For instance, in the sentence “the rest of the world knows the same”, the multiword *rest of the world* is identified, even though its only meaning in BabelNet is that of “a team of players from many countries”, which is clearly

not appropriate in the reported context. To address this issue, we asked our expert linguist to manually inspect all of the automatically detected multiword instances within the AMR 3.0 dataset in order to maintain, modify or delete them.

**Graph Conversion** Finally, using the multiwords and the alignments derived from the previous steps, we navigate the AMR graphs bottom-to-top and collapse together nodes referring to the same word or multiword expression (i.e., first reducing nodes closer to the graph leaves and then moving towards the graph root).

As a result, we move from the original figure of 936,769 nodes of AMR 3.0 to 828,483 in BMR 1.0, reducing the graph density by a notable 11.6%.

### 4.3 Number, Tense and Aspect

Even though AMR is able to encode textual information in its semantic structure, its formalism does not account for the inclusion of word components that are crucial for understanding meaning, and that languages express via the grammatical categories of number, tense and aspect. This, along with the fact that the importance of incorporating such details in Semantic Parsing formalisms has already been stressed in the literature (Donatelli et al., 2018; Bonial et al., 2019), leads us to implement these features to further enhance the representative power of BMR. To this end, we employ SpaCy in order to retrieve the Penn Treebank part-of-speech tags (Marcus et al., 1993), which inherently provide information with respect to number, tense, and aspect, for all the words and multiword expressions aligned with a node in the graphs. In practice, we account for tense by enriching each verbal node with the semantic role `:timing` showing a value of `+` or `-` to indicate events that will take place in the future or that happened in the past, respectively. Similarly, we handle plurality of the nominal nodes by adding the `:quantity` relation followed by a `+` value (see Figure 2). Lastly, we account for aspect by adding the relation `:ongoing` followed by a `+` mark to verbal nodes expressing the imperfective aspect (ongoing or usual actions).

### 4.4 Graph Disambiguation

An interlingual representation of meaning has the basic requirement of being fully linked to an inventory of meanings which can be expressed in multiple languages. For this reason, in order to make nodes in BMR graphs language-independent,

we enhance them with BabelNet synsets information. An example of why this is needed is provided in Figure 1, where the predicate `take-01` employed in AMR is defined in OntoNotes with the very coarse-grained gloss of “take, acquire, come to have, choose, bring with you from somewhere, receiving, internalizing, bringing along, enacting”, and the ambiguous word *plane* is merely represented as a lexical node, which provides no cues for understanding whether it refers to an airplane, a geometric plane, or a carpenter’s plane, inter alia. Moreover, the combination of the two does not clarify whether “take the plane” means “to take a flight” or “to take the carpenter’s plane somewhere”.

Lacking a pointer to a more fine-grained and multilingual word sense inventory also has the disadvantage of preventing the use of the formalism as a means of moving across languages effectively. For example, if the word *parents* is not assigned the proper word sense, it would lead to ambiguous translations in languages such as Spanish, where the corresponding word *padres* can indicate both the meaning of “parents”, but also the meaning of “fathers”. Therefore, the advantages that come from the disambiguation of nodes with BabelNet are twofold: (i) resolving language ambiguity while representing word meaning explicitly, and (ii) interconnecting the same meanings across languages.

Adding the disambiguation information to AMR 3.0 graphs is our last step in order to complete its conversion to BMR 1.0. To this end, we employ a set of different strategies: (a) we exploit the mapping from VerbAtlas frames to BabelNet synsets to assign word senses to nodes based on their lemmas, (b), we use the Wikipedia page information featured in AMR nodes representing named entities to retrieve the corresponding synset BabelNet identifies that page with, and (c), we make use of ESCHER (Barba et al., 2021),<sup>9</sup> a state-of-the-art system for Word Sense Disambiguation, i.e., the task of automatically assigning a meaning to a word in context (Bevilacqua et al., 2021b), to disambiguate the nodes without word senses.

As a result, we succeed in assigning a BabelNet synset to an overall figure of 92% AMR content nodes (i.e., nodes aligned with content words), with 42,549 fully disambiguated graphs out of 59,255.

<sup>9</sup>We employ the code available at <https://github.com/SapienzaNLP/esc>.

## 5 Experimental Setup

To demonstrate the importance of BMR’s semantic framing, its aptness at preserving lexical information, and its effectiveness in acting as an interlingual representation, we devise three experiments to assess its performance in comparison with AMR. Before delving into their details (Section 5.2), as well as describing our models and the evaluation measures we employ (Sections 5.3 and 5.4, respectively), we first provide thorough information regarding the datasets used in our experiments.

### 5.1 Datasets

Aside from the original AMR 3.0 and BMR 1.0 datasets described in Section 4,<sup>10</sup> the following datasets are employed in our experiments, namely: (i) AMR+, which features the set of enhancements applied to the English AMR 3.0, as described from Section 4.1 to 4.3 (excluding node disambiguation), and (ii) BMR\*, i.e., a version of BMR 1.0 that does not include lemma information.

For each dataset, we also create language-specific versions in German (DE), Italian (IT) and Spanish (ES): starting from the English AMR 3.0, we followed Biloshmi et al. (2020) and create training and development sets for these languages by using gold AMR graphs – and their converted AMR+, BMR and BMR\* versions – and pairing them with silver sentences translated with the machine translation models of Tiedemann and Thottingal (2020, OPUS-MT). As test data, we use the 1, 371 parallel sentences of Abstract Meaning Representation 2.0 - Four Translations,<sup>11</sup> that translate into our set of non-English languages their English (EN) counterparts (a subset of AMR 3.0) found in the AMR 2.0 test split.<sup>12</sup>

### 5.2 Tasks

**Graph-to-Text (GtoT)** Our first experiment concerns the Graph-to-Text generation task, i.e., the task of transforming graph meaning representations into their corresponding text, and has the goal of appraising the effectiveness of BMR as a tool for generating texts in different languages. In this context, we also conduct an ablation study on AMR+

to assess the individual impact brought about by each feature described in Section 4.

**Text-to-Graph (TtoG)** Our second experiment deals instead with the Text-to-Graph generation task (Semantic Parsing), i.e., the task of generating a graph according to a given formalism, starting from raw text. The aim of TtoG is to assess the complexity of generating BMR graphs compared to AMR ones.

**Text-to-Graph-to-Text (TGT)** Finally, in the third experiment, we evaluate the suitability of AMR and BMR to be used as interlingual representations by means of the combination of Text-to-Graph and Graph-to-Text parsing going from a source to a target language. In the same context, we also conduct an ablation study on BMR to assess the impact of the disambiguation in the graphs.

### 5.3 Models

All models employed in our experiments are built on top of SPRING (Bevilacqua et al., 2021a), an auto-regressive model for AMR parsing and generation based on the BART (Lewis et al., 2020) pretrained language model. Since the original SPRING works with pairs of sentences and linearized versions of the graphs, we modify its tokenizer to account for BMR nodes, since they contain BabelNet synset IDs too. Furthermore, we add all synsets that appear more than once<sup>13</sup> within BMR 1.0 to the model’s vocabulary and adapt SPRING to the mBART language model (Liu et al., 2020) in order to account for multiple languages in the GtoT and TGT experiments.

Given the datasets described in Section 5.1, we confront models trained on AMR 3.0, BMR 1.0, AMR+ and BMR\* for each language (AMR/BMR/AMR+/BMR\*<sub>EN,DE,IT,ES</sub>). As regards the ablation study of the GtoT experiment, we apply each modification introduced to AMR 3.0 one at a time, and obtain several versions of the dataset, each of which is used to train additional models, namely, AMR 3.0 (i) including self-explanatory relations (AMR<sub>REL</sub>), (ii) including self-explanatory relations and node merging (AMR<sub>NOD</sub>), (iii) featuring the number category (AMR<sub>NUM</sub>), (iv) featuring the tense and aspect categories (AMR<sub>TEN</sub>), and (v) featuring the number, tense and aspect categories together (AMR<sub>NT</sub>).

<sup>10</sup>We use the training/development/test splits of AMR 3.0 for both AMR 3.0 and BMR 1.0 datasets.

<sup>11</sup><https://catalog.ldc.upenn.edu/LDC2020T07>

<sup>12</sup><https://catalog.ldc.upenn.edu/LDC2017T10>

<sup>13</sup>Based on model tuning.

Language	English (EN)				German (DE)				Italian (IT)				Spanish (ES)			
Model	AMR	AMR+	BMR	BMR*	AMR	AMR+	BMR	BMR*	AMR	AMR+	BMR	BMR*	AMR	AMR+	BMR	BMR*
BLEU	44.8	49.8	<b>50.7</b>	45.7	23.2	24.3	<b>24.8</b>	22.2	29.0	31.3	<b>31.4</b>	29.1	34.6	36.8	<b>37.3</b>	35.5
chrF++	73.4	76.0	<b>76.3</b>	72.1	55.8	57.0	<b>57.1</b>	54.7	60.7	62.1	<b>62.2</b>	60.0	64.0	65.2	<b>65.5</b>	63.7
METEOR	42.2	43.9	<b>44.3</b>	42.4	25.4	26.4	<b>26.4</b>	25.3	28.9	30.4	<b>30.5</b>	29.2	32.4	33.5	<b>33.7</b>	32.8
Rouge-L	68.2	71.7	<b>72.8</b>	69.7	49.3	50.7	<b>51.1</b>	49.7	51.9	54.2	<b>54.3</b>	52.4	57.4	60.9	<b>61.0</b>	59.8

Table 1: Results for the GtoT experiment. Row blocks: models (EN, DE, IT, ES), measures. **Bold** is best.

Model	BLE	CH+	MET	R-L
AMR <sub>EN</sub>	44.8	73.4	42.2	68.2
AMR <sub>REL</sub>	44.9	73.8	42.4	68.7
AMR <sub>NOD</sub>	45.5	73.8	42.5	68.9
AMR <sub>NUM</sub>	46.9	74.3	42.9	69.8
AMR <sub>TEN</sub>	47.6	74.9	43.0	70.2
AMR <sub>NT</sub>	49.0	75.4	43.5	71.1
AMR+ <sub>EN</sub>	49.8	76.0	43.9	71.7
BMR <sub>EN</sub>	<b>50.7</b>	<b>76.3</b>	<b>44.3</b>	<b>72.8</b>

Table 2: Results for the ablation study of the GtoT experiment. Left to right: model, BLEU (BLE), chrF++ (CH+), METEOR (MET), ROUGE-L (R-L). **Bold** is best.

Model	SMT.	unlab.	noWSD	conc.	NER	neg.	reent.
AMR <sub>EN</sub>	<b>82.1</b>	85.3	<b>82.6</b>	88.0	<b>89.0</b>	67.0	<b>73.0</b>
AMR+ <sub>EN</sub>	<b>82.1</b>	<b>85.8</b>	82.1	<b>90.0</b>	<b>89.0</b>	<b>75.0</b>	70.0
BMR <sub>EN</sub>	78.6	82.2	78.6	82.0	83.0	63.0	65.0
BMR* <sub>EN</sub>	78.7	82.2	78.6	82.0	83.0	63.0	65.0

Table 3: Results for the TtoG experiment. Left to right: models, Smatch (SMT.), fine-grained scores (Damonte et al., 2017). **Bold** is best.

## 5.4 Evaluation Measures

To evaluate the text generation tasks (i.e., GtoT and TGT), we use five standard Natural Language Generation measures, namely, BLEU (Papineni et al., 2002), chrF++ (Popović, 2017), METEOR (Banerjee and Lavie, 2005), and ROUGE-L (Lin, 2004), tokenizing system predictions with the JAMR script (Flanigan et al., 2014). For the TtoG experiment, instead, as is customary, we employ the Smatch measure (Cai and Knight, 2013).

## 6 Results

### 6.1 GtoT

Results for the GtoT experiment are reported in Table 1. As can be seen, BMR obtains the highest scores for all the measures across the board, testifying to its effectiveness at generating text in multiple languages. Interestingly, when BMR is confronted with AMR+, the benefits of featuring disambiguation information immediately become

evident, with highest scores on each measure.

Results for the ablation study are, instead, shown in Table 2. Even though the impact of self-explanatory relations is not striking in this scenario (AMR<sub>REL</sub> model), the use of node merging already leads to an evident performance boost, particularly for BLEU and ROUGE-L (AMR<sub>NOD</sub>). Not surprisingly, the addition of the grammatical categories of number, tense, and aspect to AMR 3.0 corroborates the thesis of Donatelli et al. (2018) and Bonial et al. (2019), with results for the different measures growing between 1.3 to 4.2 points for AMR<sub>NT</sub> compared to the baseline AMR<sub>EN</sub> model. Moreover, demonstrating the beneficial interaction of all features described in Section 4, the AMR+<sub>EN</sub> model significantly outperforms the baseline model by 1.7 points on METEOR (lowest) and 5.0 points on BLEU (highest), while also outscoring each other model featuring only specific modifications.

### 6.2 TtoG

Results for this experiment are shown in Table 3 and provide evidence for the high degree of complexity that BMR graphs have in comparison to their AMR counterparts. In particular, AMR+<sub>EN</sub> (which, except for the disambiguated nodes, has the same graph structure as BMR<sub>EN</sub>) outperforms BMR<sub>EN</sub> by 3.5 Smatch points, demonstrating that the extra layer represented by the inclusion of disambiguation information makes BMR graphs harder to generate automatically starting from raw text. As a matter of fact, a model attempting to generate BMR graphs needs to provide disambiguation for each node (and not just for the verbal predicates), hence it faces a much more difficult task.

### 6.3 TGT

Finally, in Table 4 we report the scores for the TGT experiment, by means of which we appraise the capability of formalisms to act as bridges to translate sentences, first, performing a Text-to-Graph step, and then a Graph-to-Text one. Despite having shown lower performances in comparison to AMR

Pairs	EN – EN				EN – DE				EN – IT				EN – ES			
	AMR	AMR+	BMR	BMR*	AMR	AMR+	BMR	BMR*	AMR	AMR+	BMR	BMR*	AMR	AMR+	BMR	BMR*
BLEU	45.3	49.3	<b>50.1</b>	45.1	23.0	<b>25.1</b>	24.4	22.8	29.0	30.7	<b>30.9</b>	29.1	34.0	36.5	<b>36.6</b>	35.4
chrF++	73.5	75.2	<b>75.4</b>	71.4	55.6	<b>56.8</b>	56.1	54.1	60.2	<b>61.4</b>	61.2	59.8	63.3	64.6	<b>64.8</b>	63.3
METEOR	42.3	43.5	<b>43.7</b>	41.9	25.4	<b>26.4</b>	26.0	25.1	28.7	29.8	<b>29.9</b>	29.1	32.0	<b>33.3</b>	33.2	32.6
Rouge-L	68.8	71.8	<b>73.0</b>	69.5	49.6	<b>50.8</b>	<b>50.8</b>	49.3	51.9	<b>53.5</b>	53.2	52.6	57.2	61.1	<b>62.2</b>	59.7

Table 4: Results for the TGT experiment. Row blocks: language pairs (EN-EN, EN-DE, EN-IT, EN-ES), models (AMR, AMR+, BMR, BMR\*), generation measures. **Bold** is best.

Pairs	EN – EN		EN – DE		EN – IT		EN – ES	
	AMR+*	BMR	AMR+*	BMR	AMR+*	BMR	AMR+*	BMR
BLEU	47.7	<b>50.1</b>	23.4	<b>24.4</b>	30.2	<b>30.9</b>	35.4	<b>36.6</b>
chrF++	74.2	<b>75.4</b>	55.5	<b>56.1</b>	60.8	<b>61.2</b>	63.7	<b>64.8</b>
METEOR	42.8	<b>43.7</b>	25.6	<b>26.0</b>	29.6	<b>29.9</b>	32.5	<b>33.2</b>
Rouge-L	70.8	<b>73.0</b>	49.8	<b>50.8</b>	53.2	<b>53.2</b>	59.6	<b>62.2</b>

Table 5: Results for the ablation study on the TGT experiment. Row blocks: language pairs (EN-EN, EN-DE, EN-IT, EN-ES), models (AMR+\*, BMR), generation evaluation measures. **Bold** is best.

in the TtoG experiment, the high scores obtained by BMR in this experiment demonstrate that it is better suited as an interlingua. Nevertheless, AMR+ outperforms BMR in a few settings, likely due to the higher complexity entailed by BMR parsing, as explained in Section 6.2. Corroborating this thesis are the results shown in Table 5, where BMR scores are compared against a model (AMR+\*) in which, to perform the Graph-to-Text step, AMR+ uses a BMR parser with the synset information removed, rather than its own parser. The outcome of this ablation study, with BMR now systematically outscoring its competitor, sheds further light upon the effectiveness of synset-driven disambiguation for encoding valuable sentence information.

Returning to the results given in Table 4, even though performances for BMR\* models are the lowest (yet competitive, and sometimes higher than AMR) on the board, it is worth remarking that this setting does not feature the lemma information. In fact, in order to be purely semantic, BMR graphs should solely feature the BabelNet synset information. However, given that state-of-the-art Semantic Parsing and generation models make use of pre-trained language models such as BART and mBART, which are trained with data in human language (hence devoid of synset information), the performance of fully-semantic models drops if lemmas are not taken into account. Additionally, currently available text generation metrics are sub-optimal when employed to assess semantics, since these measures evaluate similarities at the lemma level. Therefore, though a fully-semantic model

could infer the meaning of a BabelNet synset, its performances will be penalized for not generating specific lemmas while outputting perfectly suitable synonyms. In view of this, BMR 1.0 incorporates the lemma information along with the BabelNet synset specifying its meaning (see also Appendix A), demonstrating that lexical-semantic representations improve over purely lexical ones.

## 7 BMR\*: A Case Study

Results for the experiments we conducted depict BMR\* as the model that, on the whole, achieves the lowest scores. With the aim of showing how such results might arise due to inadequate evaluation measures (see Section 6.3), we propose a focused case study in which we qualitatively inspect the differences between graphs and sentences generated by means of the AMR and BMR\* models. Starting from the sentence “*My friend did not tolerate his father’s behaviour*” (Figure 3), it can be seen how the grammatical categories of number and tense for the words *friend* and *tolerate* are correctly preserved by BMR\* only. Additionally, it can be noted how the complex structure that defines *child* in AMR can confuse the model when there is a reentrant node (in this case, the model does not know to whom the father is related). As interestingly, the sentence generated via BMR\* replaces *tolerate* with the synonym *put up with*, which worsens its performance according to exact string matching metrics, but, at the same time, provides an insight of a higher level of abstraction when lemmas are omitted.

## 8 Error Analysis

Although the experiments reported in Section 5 testify to the quality of BMR, following an in-house behavioral analysis inspired by the work of Ribeiro et al. (2020), we identify three main classes of errors that undermine the application of BMR as an interlingua, one concerning the formalism (*repository constraints*), one tied to the data contained in the BMR 1.0 dataset (*disambiguation constraints*),



and one concerning the language-specific lexicons (*language-specific constraints*).

**Repository constraints** BabelNet features a wealth of synsets covering content words in a multilingual setting, but, at the same time, does not provide information regarding parts of speech other than nouns, verbs, adjectives and adverbs. As a result, BMR uses language-specific lemmas to represent conjunctions or ambiguous pronouns such as *anyone*, which can mean either “not a single person” or “everyone”, depending on the use of negative or positive phrasing. On a different note, with roughly 6,500 languages spoken in the world and BabelNet 5.0 featuring a subset of them, the definition of BMR as an interlingua is actually constrained to the number, albeit large, of 500 BabelNet languages.

**Disambiguation constraints** The creation of BMR 1.0 is based upon the Word Sense Disambiguation task carried out via a state-of-the-art system (Barba et al., 2021, ESCHER). And yet, this neural architecture is trained to predict word senses featured in the WordNet 3.0 sense inventory only. By virtue of the fact that, following the node merging strategy (Section 4.2), we can obtain polysemous multiwords found in BabelNet but not in WordNet (as is the case of *run off at the mouth*), we cannot provide disambiguation for such instances. This justifies the fact that 8% of content nodes in BMR are not disambiguated (see also Section 4.4).

**Language-specific constraints** The number of items in a lexicon and the degree of word polysemy vary from language to language (Talmy, 2000). Using BabelNet synsets to represent abstract concepts and connect them multilingually is certainly a desirable feature. However, there are concepts and expressions that exist in a given language only, e.g., owing to their being culturally connoted. For example, the Spanish word *espeto*, which refers to a traditional way of cooking freshly-caught sea fish, has no equivalent in English. Though the concept is featured in BabelNet, it has no lexicalizations in other languages and, as such, it would need to be paraphrased in order to be rendered.

## 9 Conclusion

Current Semantic Parsing formalisms share tight dependencies with semantic repositories which are both language-specific and isolated from word senses in other languages. As a result, they are not fit to be used as interlingual representations of

```

—Original Sentence—
My friend did not tolerate his father's behaviour

—AMR—
(t / tolerate-01 :polarity -
  :ARG0 (p / person
    :ARG0-of (h / have-rel-role-91
      :ARG1 (i / i)
      :ARG2 (f / friend)))
  :ARG1 (b / behave-01
    :ARG0 (p2 / person
      :ARG0-of (h2 / have-rel-role-91
        :ARG1 p
        :ARG2 (f2 / father))))))

—BMR*—
(t / bn:00082138v :polarity - :timing -
  :agent (f / bn:00036538n
    :related (i / i))
  :theme (b / bn:00009656n
    :related (f2 / bn:00009616n
      :related f)))

—AMR Generated Sentence—
My friends do not tolerate the behavior of my father.

—BMR* Generated Sentence—
My friend did not put up with the behaviour of his father.

```

Figure 3: AMR and BMR\* graph representations and generated sentences for the original sentence “My friend did not tolerate his father’s behaviour”. Best seen in color.

meaning. In this paper, we put forward BMR, a new language-independent formalism that abstracts away from language-specific constraints thanks to two multilingual semantic resources, namely, BabelNet and VerbAtlas. To put our formalism into practice, we also created BMR 1.0, the first dataset labeled according to BMR.

Our experiments demonstrate the impact that the fully-semantic framing of our formalism has in comparison to the widely-employed formalism of AMR, as well as showing its ability to be a better tool at encoding textual information, and a much more effective interlingua in a text-to-graph-to-text machine translation task.

As future work, we plan to (i) create a single multilingual model to parse graphs and generate text in any language, (ii) apply BMR crosslingually to other downstream tasks such as text summarization, (iii) evolve the formalism to prevent the inclusion of lexical information of any kind. We make our code and data available to the research community at <https://github.com/SapienzaNLP/bmr>.

## Acknowledgments

The authors gratefully acknowledge the support of the ERC Consolidator Grant MOUSSE No. 726487, the Marie Skłodowska-Curie project *Knowledge Graphs at Scale* (Knowledge Graphs) No. 860801, and the ELEXIS project No. 731015 under the European Union's Horizon 2020 research and innovation programme.



## 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.
- 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.
- Juan Aparicio, Mariona Taulé, and Maria Antònia Martí. 2008. [Ancora-verb: A Lexical Resource for the Semantic Annotation of Corpora](#). In *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2008, 26 May - 1 June 2008, Marrakech, Morocco*. European Language Resources Association.
- 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.
- Satanjeev Banerjee and Alon Lavie. 2005. [METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments](#). In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 65–72, Ann Arbor, Michigan. Association for Computational Linguistics.
- Edoardo Barba, Tommaso Pasini, and Roberto Navigli. 2021. [ESC: Redesigning WSD with Extractive Sense Comprehension](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4661–4672, Online. Association for Computational Linguistics.
- Michele Bevilacqua, Rexhina Blloshmi, and Roberto Navigli. 2021a. [One SPRING to Rule Them Both: Symmetric AMR semantic Parsing and Generation without a Complex Pipeline](#). *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(14):12564–12573.
- Michele Bevilacqua, Tommaso Pasini, Alessandro Raganato, and Roberto Navigli. 2021b. [Recent Trends in Word Sense Disambiguation: A Survey](#). In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 4330–4338, Montreal, Canada. International Joint Conferences on Artificial Intelligence Organization. Survey Track.
- 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, Online. Association for Computational Linguistics.
- Claire Bonial, Lucia Donatelli, Mitchell Abrams, Stephanie M. Lukin, Stephen Tratz, Matthew Marge, Ron Artstein, David Traum, and Clare Voss. 2020a. [Dialogue-AMR: Abstract Meaning Representation for Dialogue](#). In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 684–695, Marseille, France. European Language Resources Association.
- 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.
- Claire Bonial, Stephanie M. Lukin, David Doughty, Steven Hill, and Clare Voss. 2020b. [InfoForager: Leveraging semantic search with AMR for COVID-19 research](#). In *Proceedings of the Second International Workshop on Designing Meaning Representations*, pages 67–77, Barcelona, Spain (online). Association for Computational Linguistics.
- Shu Cai and Kevin Knight. 2013. [Smatch: an Evaluation Metric for Semantic Feature Structures](#). In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 748–752, Sofia, Bulgaria. Association for Computational Linguistics.
- Simone Conia, Andrea Bacciu, and Roberto Navigli. 2021. [Unifying Cross-Lingual Semantic Role Labeling with Heterogeneous Linguistic Resources](#). In

- Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 338–351, Online. Association for Computational Linguistics.
- Marco Damonte and Shay B. Cohen. 2018. **Cross-Lingual Abstract Meaning Representation Parsing**. 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 1146–1155, New Orleans, Louisiana. Association for Computational Linguistics.
- Marco Damonte, Shay B. Cohen, and Giorgio Satta. 2017. **An incremental parser for Abstract Meaning Representation**. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 536–546, Valencia, Spain. 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. *Where have all the adjectives gone?: and other essays in semantics and syntax*, volume 107. De Gruyter Mouton.
- 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.
- Christiae Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*, volume 22. MIT Press, Cambridge, MA, USA.
- Jeffrey Flanigan, Sam Thomson, Jaime Carbonell, Chris Dyer, and Noah A. Smith. 2014. **A Discriminative Graph-Based Parser for the Abstract Meaning Representation**. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1426–1436, Baltimore, Maryland. Association for Computational Linguistics.
- Jan Hajič, Eva Hajičová, Jarmila Panevová, Petr Sgall, Ondřej Bojar, Šilvie Cinková, Eva Fučíková, Marie Mikulová, Petr Pajas, Jan Popelka, Jiří Semecký, Jana Šindlerová, Jan Štěpánek, Josef Toman, Zdeňka Uřešová, 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).
- Hardy Hardy and Andreas Vlachos. 2018. **Guided Neural Language Generation for Abstractive Summarization using Abstract Meaning Representation**. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 768–773, Brussels, Belgium. Association for Computational Linguistics.
- Matthew Honnibal and Mark Johnson. 2015. **An improved non-monotonic transition system for dependency parsing**. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1373–1378, Lisbon, Portugal. Association for Computational Linguistics.
- Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. 2006. **OntoNotes: The 90% Solution**. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, pages 57–60, New York City, USA. Association for Computational Linguistics.
- Hans Kamp and Uwe Reyle. 1993. *From Discourse to Logic: Introduction to Model-theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, volume 42 of *Studies in Linguistics and Philosophy*. Springer, Dordrecht.
- Pavan Kapanipathi, Ibrahim Abdelaziz, Srinivas Ravishankar, Salim Roukos, Alexander Gray, Ramón Fernández Astudillo, Maria Chang, Cristina Cornelio, Saswati Dana, Achille Fokoue, Dinesh Garg, Alfio Gliozzo, Sairam Gurajada, Hima Karanam, Naweed Khan, Dinesh Khandelwal, Young-Suk Lee, Yunyao Li, Francois Luus, Ndivhuwo Makondo, Nandana Mihindukulasooriya, Tahira Naseem, Sumit Neelam, Lucian Popa, Revanth Gangi Reddy, Ryan Riegel, Gaetano Rossiello, Udit Sharma, G P Shrivatsa Bhargav, and Mo Yu. 2021. **Leveraging Abstract Meaning Representation for Knowledge Base Question Answering**. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3884–3894, Online. Association for Computational Linguistics.
- Paul Kingsbury and Martha Palmer. 2002. **From TreeBank to PropBank**. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC'02)*, Las Palmas, Canary Islands - Spain. European Language Resources Association (ELRA).
- Kevin Knight, Bianca Badarau, Laura Baranescu, Claire Bonial, Madalina Bardocz, Kira Griffitt, Ulf Hermjakob, Daniel Marcu, Martha Palmer, Tim O’Gorman, et al. 2020. **Abstract Meaning Representation (AMR) Annotation Release 3.0 (LDC2020T02)**. Philadelphia. Linguistic Data Consortium.

- Valentina Leone, Giovanni Siragusa, Luigi Di Caro, and Roberto Navigli. 2020. [Building Semantic Grams of Human Knowledge](#). In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 2991–3000, Marseille, France. European Language Resources Association.
- Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. [BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.
- Xexin Liao, Logan Lebanoff, and Fei Liu. 2018. [Abstract Meaning Representation for Multi-Document Summarization](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1178–1190, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- Jungwoo Lim, Dongsuk Oh, Yoonna Jang, Kisu Yang, and Heuiseok Lim. 2020. [I Know What You Asked: Graph Path Learning using AMR for Commonsense Reasoning](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2459–2471, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Chin-Yew Lin. 2004. [ROUGE: A Package for Automatic Evaluation of Summaries](#). In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.
- Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. 2020. [Multilingual Denoising Pre-training for Neural Machine Translation](#). *Transactions of the Association for Computational Linguistics*, 8:726–742.
- Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. [Building a Large Annotated Corpus of English: The Penn Treebank](#). *Computational Linguistics*, 19(2):313–330.
- Roberto Navigli. 2018. [Natural language understanding: Instructions for \(present and future\) use](#). In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 5697–5702. International Joint Conferences on Artificial Intelligence Organization.
- Roberto Navigli, Michele Bevilacqua, Simone Conia, Dario Montagnini, and Francesco Cecconi. 2021. [Ten years of BabelNet: A survey](#). In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021*, pages 4559–4567. ijcai.org.
- Roberto Navigli, Rexhina Blloshmi, and Abelardo Carlos Martinez Lorenzo. 2022. [BabelNet Meaning Representation: A Fully Semantic Formalism to Overcome Language Barriers](#). *Proceedings of the AAAI Conference on Artificial Intelligence*, 36.
- Roberto Navigli and Simone Paolo Ponzetto. 2010. [BabelNet: Building a very large multilingual semantic network](#). In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 216–225, Uppsala, Sweden. Association for Computational Linguistics.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. [Bleu: a Method for Automatic Evaluation of Machine Translation](#). In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.
- Maja Popović. 2017. [chrF++: words helping character n-grams](#). In *Proceedings of the Second Conference on Machine Translation*, pages 612–618, Copenhagen, Denmark. Association for Computational Linguistics.
- Sudha Rao, Daniel Marcu, Kevin Knight, and Hal Daumé III. 2017. [Biomedical Event Extraction using Abstract Meaning Representation](#). In *BioNLP 2017*, pages 126–135, Vancouver, Canada. Association for Computational Linguistics.
- Marco Tulio Ribeiro, Tongshuang Wu, Carlos Guestrin, and Sameer Singh. 2020. [Beyond Accuracy: Behavioral Testing of NLP Models with CheckList](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4902–4912, Online. Association for Computational Linguistics.
- Linfeng Song, Daniel Gildea, Yue Zhang, Zhiguo Wang, and Jinsong Su. 2019. [Semantic Neural Machine Translation Using AMR](#). *Transactions of the Association for Computational Linguistics*, 7:19–31.
- Leonard Talmy. 2000. [Toward a cognitive semantics - volume 1: Concept structuring systems](#).
- Jörg Tiedemann and Santhosh Thottingal. 2020. [OPUS-MT – Building open translation services for the World](#). In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, pages 479–480, Lisboa, Portugal. European Association for Machine Translation.
- Jens E. L. Van Gysel, Meagan Vigus, Jayeol Chun, Kenneth Lai, Sarah Moeller, Jiarui Yao, Tim O’Gorman, Andrew Cowell, William Croft, Chu-Ren Huang, Jan Hajič, James H. Martin, Stephan Oepen, Martha Palmer, James Pustejovsky, Rosa Vallejos, and Nianwen Xue. 2021. [Designing a Uniform Meaning Representation for Natural Language Processing](#). *KI - Künstliche Intelligenz*, 35(0):343–360.

- 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).
- Niawen Xue and Martha Palmer. 2009. **Adding semantic roles to the chinese treebank.** *Natural Language Engineering*, 15(1):143–172.
- Huaiyu Zhu, Yunyao Li, and Laura Chiticariu. 2019. **Towards universal semantic representation.** In *Proceedings of the 1st International Workshop on Designing Meaning Representations*, pages 177–181, Florence, Italy.

## A Reading BMR Graphs

The BMR formalism renders graphs in a text-friendly fashion following the AMR custom. Specifically, edges are represented by means of standardized semantic relations names preceded by a colon (e.g. `:agent`, or its inverse relation `:agent_of`), and nodes are identified by triplets:

*(id / lemma / Babel synset id)*

Left to right, the triplet shows: (i) the unique identifier (id) of the node,<sup>14</sup> (ii) the lemma for the word (or multiword expression) in the original sentence, and (iii) the BabelNet synset id taken from BabelNet 5.0 that is assigned to disambiguate the node.<sup>15</sup>

Lastly, node hierarchy in BMR is represented by means of open and closed round brackets, and special characters such as `+` indicate special features of some nodes, such as, e.g., grammatical tense information. Figure 4 shows an example of a BMR graph (bottom) in comparison to the AMR graph (top) for the same sentence, in text-friendly format.

## B Semantic Roles in BMR

Semantic roles in BMR (Table 6) are largely based on the VerbAtlas inventory and have been modified to account for non-verbal entities drawing inspiration from property lists available in the literature (Dixon, 2010; Leone et al., 2020). Similarly to AMR, each relation has its inverse, expressed by appending `_of` to it (e.g., `:purpose` versus `:purpose_of`). Roles in AMR which are not listed in Table 6 are preserved in BMR (e.g., `:age`, `:degree`, `:frequency` or `:manner`).

## C AMR 3.0 to BMR 1.0 Manual Mapping Examples

Non-verbal predicates, as well as special predicates found within AMR 3.0 have been mapped to the BMR formalism according to the set of semantic relations described in Appendix B (see also Section 4.1) by means of an in-house annotation interface. The choice of BabelNet synsets to express the meaning of the original predicates followed a simple set of annotation strategies, sorted

<sup>14</sup>Similarly to AMR, if a node is referred to anew in the same BMR graph, only the id is used to identify it.

<sup>15</sup>Note that, even though the lemma changes according to the source language used to produce the graph, it is the synset information that serves as the interlingual component. In fact, graphs for the same sentence translated in different languages will show different lemmas, but the same BabelNet synsets.

```
--AMR--
(t/ take-01
 :ARG0 (a / and
 :op1 (p / person
 :ARG0-of (s / study-01))
 :op2 (p2 / person
 :ARG0-of (h / have-rel-role-91
 :ARG1 p
 :ARG2 (p3 / parent))))
 :ARG1 (p4 / plane)
 :time (t2/ minute
 :mod (l / last)))

--BMR--
(t / take / bn:00094732v
 :timing +
 :agent (a / and
 :op1 (s / student / bn:00029806n
 :quantity +)
 :op2 (p / parent / bn:00060643n
 :quantity +
 :related s)
 :theme (p2 / plane / bn:00001697n)
 :timing (t2 / at_the_last_minute
 / bn:00114428r))
```

Figure 4: AMR (top) and BMR (bottom) text-friendly form of the graphs for the sentence “*The students and their parents will take the plane at the last minute*”.

by desired priority: using the predicate name as a query to look for available synsets, (i) pick a nominal synset also featured in WordNet 3.0 (e.g., querying with the lemma *liberality* for the predicate `liberal.02`), (ii) pick an adjectival synset featured in WordNet 3.0, (iii) pick a nominal synset not featured in WordNet 3.0, (iv) pick an adjectival synset not featured in WordNet 3.0, (v) pick a synonym to query for available synsets (e.g., querying with the lemma *correct* for predicate `be_it.07`). See Table 7 for a random sample of AMR 3.0 to BMR 1.0 mappings.

BMR	AMR	VerbAtlas	examples
agent	-	AGENT	marry <sub>V</sub> >man <sub>N</sub> ; suggest <sub>V</sub> >mum <sub>N</sub>
appearance	mod	-	complexion <sub>N</sub> >pale <sub>A</sub> ; road <sub>N</sub> >snowy <sub>A</sub>
cause	cause	CAUSE/STIMULUS	issue <sub>N</sub> >society <sub>N</sub> ; kill <sub>V</sub> >evil <sub>N</sub>
co-agent	accompanier	CO_AGENT	(same as agent)
co-patient	accompanier	CO_PATIENT	(same as patient)
co-theme	accompanier	CO_THEME	(same as theme)
composition	consist-of	MATERIAL	cup <sub>N</sub> >metal <sub>N</sub> ; army <sub>N</sub> >idiot <sub>N</sub>
context	location, topic	TOPIC/ATTRIBUTE	boldness <sub>N</sub> >mission <sub>N</sub> ; excellence <sub>N</sub> >sports <sub>N</sub>
cost	cost	ASSET	computer <sub>N</sub> >euro <sub>N</sub> ; tuition <sub>N</sub> >free <sub>A</sub>
experiencer	-	EXPERIENCER	badness <sub>N</sub> >customer <sub>N</sub> ; react <sub>V</sub> >sector <sub>N</sub>
extent	duration, extent	EXTENT	trip <sub>N</sub> >mile <sub>N</sub> ; work <sub>V</sub> >day <sub>N</sub>
identity	domain/meaning/role/example	-	boyfriend <sub>N</sub> >lawyer <sub>N</sub> ; dog <sub>N</sub> >animal <sub>N</sub>
instrument	instrument	INSTRUMENT	book <sub>N</sub> >eye <sub>N</sub> ; pasta <sub>N</sub> >fork <sub>N</sub>
interactor	-	-	book <sub>N</sub> >librarian <sub>N</sub> ; treaty <sub>N</sub> >actor <sub>N</sub>
location	location	LOCATION	hotel <sub>N</sub> >beach <sub>N</sub> ; water <sub>N</sub> >jar <sub>N</sub>
membership	employed-by/have-org-role-91	-	friend <sub>N</sub> >company <sub>N</sub> ; nurse <sub>N</sub> >hospital <sub>N</sub>
part	part/subset/superset	-	book <sub>N</sub> >beginning <sub>N</sub> ; car <sub>N</sub> >wheel <sub>N</sub>
patient	-	PATIENT	dry <sub>V</sub> >skin <sub>N</sub> ; kick <sub>V</sub> >ball <sub>N</sub>
physical_prop	mod	-	hand <sub>N</sub> >cold <sub>A</sub> ; train <sub>N</sub> >fast <sub>A</sub>
property	poss	-	patient <sub>N</sub> >health <sub>N</sub> ; professor <sub>N</sub> >book <sub>N</sub>
purpose	purpose	PURPOSE	treaty <sub>N</sub> >extradition <sub>N</sub> ; book <sub>N</sub> >coloring <sub>N</sub>
quality	mod	-	book <sub>N</sub> >available <sub>A</sub> ; city <sub>N</sub> >beautiful <sub>A</sub>
quantity	quant	VALUE	cow <sub>N</sub> >few <sub>A</sub> ; degree <sub>N</sub> >42
related	have-rel-role-91	-	city <sub>N</sub> >suburb <sub>N</sub> ; father <sub>N</sub> >son <sub>N</sub>
result	-	PRODUCT, RESULT	become <sub>V</sub> >farmer <sub>N</sub> ; revolution <sub>N</sub> >overthrow <sub>V</sub>
source	source	SOURCE	book <sub>N</sub> >author <sub>N</sub> ; trip <sub>N</sub> >San Diego <sub>N</sub>
target	beneficiary/destination/direction	BENEFICIARY/DESTINATION/GOAL/RECIPIENT	brutality <sub>N</sub> >person <sub>N</sub> ; food <sub>N</sub> >dog <sub>N</sub>
theme	-	THEME	read <sub>V</sub> >book <sub>N</sub> ; require <sub>V</sub> >vitamin <sub>N</sub>
timing	time	TIME	marry <sub>V</sub> >then <sub>R</sub> ; struggle <sub>N</sub> >current <sub>A</sub>
url	hyperlink-91	-	website <sub>N</sub> >https://verbatlas.org/

Table 6: Semantic roles in BMR. Left to right: BMR role names (BMR), AMR role(s) equivalent (AMR), VerbAtlas role(s) equivalent (VerbAtlas), role usage example(s). Examples read as follows: father node<sub>PoS</sub>>child node<sub>PoS</sub>.

<p><b>be temporally_at.91 (reification of :time)</b>  ARG1 (entity); ARG2 (time)  AMR_ARG1 :theme_of (bn:00083185v :timing (AMR_ARG2))</p>
<p><b>explicit.03 (clear, detailed)</b>  ARG0 (causer of clarification); ARG1 (thing becoming clearer); ARG2 (explained to)  AMR_ARG1 :quality (bn:00019459n :cause (AMR_ARG0) :experiencer (AMR_ARG2))</p>
<p><b>loose.04 (not tight fitting or compacted)</b>  ARG0 (causer of looseness); ARG1 (non-compact substance, may be abstract); ARG2 (instrument of loosening, if in addition to ARG0)  AMR_ARG1 :quality (bn:00106169a :cause (AMR_ARG0) :instrument (AMR_ARG2))</p>
<p><b>regular.02 (occurring on a consistent schedule; periodic)</b>  ARG1 (thing occurring regularly); ARG2 (specific activity/aspect of ARG1 that occurs regularly, if in addition); ARG3 (measurement of the period)  AMR_ARG1 :part (AMR_ARG2 :timing (bn:00066931n :extent (AMR_ARG3)))</p>
<p><b>sterile.02 (inhospitable to the growth of life)</b>  ARG1 (sterile location/entity); ARG2 (new life)  AMR_ARG1 :physical_prop (bn:00046772n :context (AMR_ARG2))</p>

Table 7: Mapping examples from AMR 3.0 to BMR 1.0. Each row block lists (top to bottom) original OntoNotes predicate names and glosses, original glosses for the predicate arguments, predicate rendering in BMR. AMR\_ARGX is a placeholder that is replaced with the name of the node having the relation AMRX in the AMR 3.0 graph.