A Survey of Meaning Representations – From Theory to Practical Utility

Zacchary Sadeddine[1], **Juri Opitz**[2], **Fabian M. Suchanek**[1]
[1] Télécom Paris, Institut Polytechnique de Paris, France
[2] University of Zurich, Switzerland
{zacchary.sadeddine, fabian.suchanek}@telecom-paris.fr
opitz.sci@gmail.com

Abstract

Symbolic meaning representations of natural language text have been studied since at least the 1960s. With the availability of large annotated corpora, and more powerful machine learning tools, the field has recently seen several new developments. In this survey, we study today's most prominent Meaning Representation Frameworks. We shed light on their theoretical properties, as well as on their practical research environment, i.e., on datasets, parsers, applications, and future challenges.

1 Introduction

Being able to represent the semantic structure of a text has been an important research goal since the early days of NLP. Early works developed natural language interfaces for specific databases. They transformed raw text into an executable language, using formalisms such as SQL, first-order logic or lambda-calculus (Mooney, 1996; Wong and Mooney, 2006; Mooney, 2007). Another avenue of research, which is the focus of this work, has developed general-purpose, non-executable Meaning Representations (MRs), inspired by formal grammars. These often take the form of human-readable graphs. Figure 1 shows an example.

Such MRs are used to improve the accuracy of NLP systems in tasks such as summarization or machine translation (Gao and Vogel, 2011; Liu et al., 2015; Mohamed and Oussalah, 2019; Liao et al., 2018; Song et al., 2019; Ribeiro et al., 2022). In the age of large language models (LLMs), they also get leveraged for their interpretability, e.g., to enhance semantic search (Bonial et al., 2020; Cai et al., 2022; Opitz and Frank, 2022b) or natural language inference (Opitz et al., 2023b). They are also used to generate paraphrases (Cai et al., 2021), augment training data (Shou et al., 2022), or to do style-transfer (Jangra et al., 2022; Shi et al., 2023).

In this survey, we provide a structured overview of current Meaning Representation Frameworks. Several other surveys have discussed MRs before us. However, they are either focused on linguistic theory (Abend and Rappoport, 2017; Žabokrtský et al., 2020; Pavlova et al., 2023) and thus tend to neglect applications, parsers, and resources, or they focus on the practical application only (Verrev, 2023). Our survey aims to strike a middle ground: It presents both the different formalisms and their applications, resources, and parsers. This balance allows us to describe a bigger picture of the field and outline open challenges. Our survey thus aims to be a handy reference for anyone who wishes to choose, understand, build, or use a Meaning Representation.

In Section 2 we introduce the main concepts and properties of MRs. Section 3 tackles Shallow MRs, and Section 4 Deep MRs. Finally, Section 5 discusses open challenges in the domain.

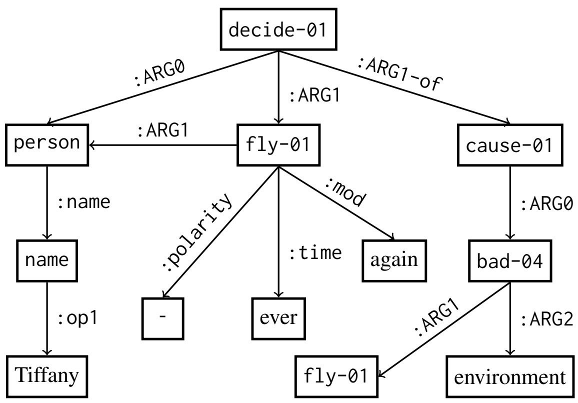

Figure 1: AMR graph for the sentence "Tiffany decided that she would never fly again, because it is bad for the environment".

2 Meaning Representations

Given a text in natural language, MR parsing is the task of producing a symbolic representation of its meaning, as it is understood by a language speaker (Abend and Rappoport, 2017). Different Meaning Representation Frameworks (MRFs)

Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), pages 2877–2892
June 16-21, 2024 ©2024 Association for Computational Linguistics

MRF	Subevents	Shape	Compositional	Node type (Flavor)	Edge type
SR	-	Tree	-	Text spans (1)	Numbered
RST	-	Tree	✓	Text spans (1)	Theory-oriented
UDS	-	Tree	✓	Text spans (1)	Numbered & theory-oriented
SD	✓	Graph	-	Augmented Text Spans (0)	Numbered
EDS	✓	Graph	-	Augmented Text Spans (1)	Numbered
UCCA	✓	Tree	✓	Text spans (1)	Theory-oriented
AMR	✓	Graph	-	Synsets (Propbank) (2)	Predicate-dependent
DRS	✓	Graph	✓	Synsets (WordNet) (2)	Predicate-independent

Table 1: Properties of the Meaning Representation Frameworks that we survey. The middle line separates shallow and deep formalisms.

have been developed. In this survey, we are interested in graph-like MRFs with their corpora and parsers. We focus on **Semantic Roles (SR)**, **Rhetorical Structure Theory (RST)**, **Universal Decompositional Semantics (UDS)**, **Semantic Dependencies (SD)**, **Elementary Discourse Structures (EDS)**, **Universal Conceptual Cognitive Annotation (UCCA)**, **Abstract Meaning Representation (AMR)**, and **Discourse Representation Structure (DRS)**. We will focus on English sentences, and discuss the aspects of multi-sentence and multi-lingual support where relevant.

MRFs are often inspired by a Neo-Davidsonian semantics, and see *events* as the main elements of sentences. The *predicate* of an event defines the type of the event, and is most often a verb (`decide-01` or `fly-01` in Figure 1). The *arguments* of the event are the entities that participate in the event ("Tiffany" in the example), or the circumstances of the event, such as its place or manner (a negative polarity "-", in our example). The *semantic role* of an argument specifies the role that the participant plays in the event. In our example, the semantic role of *Tiffany* in the `decide-01` event is the subject/agent (`ARG0` in AMR jargon).

Based on this, the semantic information in a sentence can be decomposed into three levels: The level in the middle describes events. The sub-event level decomposes the arguments of events into smaller components – up to words and possibly even sub-words. In our example, "bad for the environment" is modeled by the link from `bad-4` to `environment` with the semantic role `ARG2`. The supra-event level links events to other events. This is done using *discourse relations*. In our example, the `cause-01` node connects `decide-01` and `bad-04`, meaning that the decision was taken *because* flying is bad for the environment. Discourse relations can even link events across sentences.

Different MRFs vary these general ideas along several axes, which we show in Table 1. First, not all MRFs can represent sub-events (Column 2 in Table 1). We call a MRF *deep* if it can represent sub-events, and *shallow* otherwise. Second, MRFs construct either trees (where each node has at most one parent) or full-fledged graphs (Column 3). Our example in Figure 1 is not a tree: `person` has two different parents, because it plays two different roles. Third, some MRFs are *compositional* (Column 4), which here means that nodes can denote sub-graphs/trees. Our example in Figure 1 is not compositional: every node contains the same level of information. However, we can imagine creating a node that represents the fact that the `fly-01` event is negated. This would then be a compositional node. The difference between compositional and non-compositional MRFs is thus similar to the difference between dependency and constituency trees in syntactic parsing.

MRFs can further be distinguished by how abstract their node labels are (Column 5): Nodes can be labeled with a span from the text, but they can also be augmented with extra information such as a POS tag. Some representations go as far as using abstractions such as synsets from predefined vocabularies, to help reduce (or even eliminate) lexical ambiguity, and make events invariant to surface form. In the case of compositional MRFs, this property applies to leaf nodes. The node type can be completed with the Flavor hierarchy proposed by Oepen et al. (2019). This hierarchy differentiates MRFs based on anchoring, i.e. on the explicit correspondence between nodes and the input sentence. Flavor 0 means that each node injectively corresponds to one word; Flavor 1 relaxes the anchoring constraints, allowing a node to correspond to a whole span, and the same span to correspond to several nodes; and Flavor 2 marks the absence of explicit links between the nodes and the text.

Finally, the MRFs differ in their edge type (Column 6): Some MRFs use roles that depend on a specific linguistic theory, like `elaboration` (discourse theory) or `scene` (cognitive science). These schemes can describe only a limited array of relations, and do not distinguish the agents and patients of events. Other representations are more specific and use numbered semantic roles (`A0`, `A1`, ...). In these schemes, `A0` and `A1` usually correspond to Dowty's Proto-Agent and Proto-Patient (Dowty, 1991), respectively. These proto-roles are defined by their features: Typical agent features are *awareness, movement*, and *volition*, while typical patient features are *change of state*, *being stationary*, etc. The other semantic roles (`A2`, `A3`, ...) usually do not have such a predefined meaning. Again other MRFs are more specific, and use predicate-independent semantic roles that distinguish finer roles such as `Agent` and `Patient`. Finally, some MRFs make the meaning of the role dependent on the predicate: in Figure 1, `ARG0` means "pilot, agentive entity capable of flight" for `fly-01`, while it means "decider" for `decide-01`. These MRFs thus describe their arguments very specifically.

3 Shallow Meaning Representation Frameworks

3.1 Semantic Roles

A prominent Shallow Meaning Representation Framework is the Semantic Roles framework (SR, Gildea and Jurafsky, 2000). Given an input sentence and a predicate, its purpose is to determine the arguments of the predicate and their semantic roles – a task known as Semantic Role Labeling (SRL). SRL focuses on event-level relations, which means that its predicates are verbs. There are (at least) three different implementations of semantic roles. The most popular one is *PropBank SRL*, where semantic roles are split into core and non-core roles according to PropBank (Palmer et al., 2005). The non-core roles are also called modifiers, and they always have the same meaning: `ARGM-CAU` indicates cause, `ARGM-LOC` indicates location, etc. The meaning of core roles ($\texttt{ARG}_{2...n}$) depends on the predicate. However, `ARG0` and `ARG1` usually correspond to Dowty's Proto-Agent and Proto-Patient (Dowty, 1991). Other paradigms exist: *FrameNet SRL* generalizes descriptions across similar verbs (e.g., say, speak) as well as similar nouns and other words (e.g., speech), based on FrameNet (Baker et al., 1998). *Semantic Proto-Role Labeling* (SPRL) aims at directly approximating Dowty's Proto-Roles with features such as movement, awareness, etc (Dowty, 1991).

Figure 2 shows a merger of three parsings for our example (in PropBank-SRL style), for the predicates "decided", "fly", and "is". Since an SRL graph consists of only one predicate node and its arguments, the graph is a dependency tree, with text spans as nodes. SRL is a rather light annotation, and it is used to enhance LLMs (Zhang et al., 2020b), e.g., for downstream tasks such as Fact Checking (Zhong et al., 2020), Question Answering (Pillai et al., 2018), and Summarization (Mohamed and Oussalah, 2019; Zhang and Bansal, 2021).

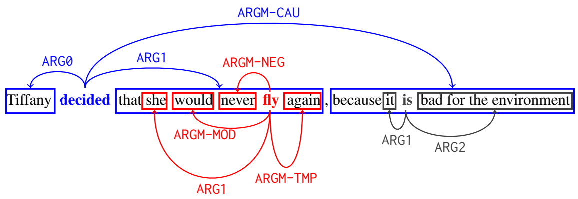

Figure 2: Semantic Role Labeling of our example sentence in span-graph style. Each color corresponds to a predicate (bold) and its arguments (solid)

Resources. PropBank-SRL has been the focus of several shared tasks, which produced datasets that are used to this day. CoNLL 2005 (Carreras and Màrquez, 2004, 2005) introduced span-based SRL, while CoNLL 2008 (Surdeanu et al., 2008) and 2009 (Hajič et al., 2009) introduced dependency-based SRL (which labels only the syntactic heads of the arguments). These works are based on the expert annotations of the WSJ section of the Penn Treebank (in English) from Propbank. The resulting training sets consist of 40,000 sentences each. Other datasets provide FrameNet SRL (Burchardt and Pennacchiotti, 2008; Das and Smith, 2011; Hartmann et al., 2017) and SPRL annotations (Reisinger et al., 2015; White et al., 2016).

Parsing. Regardless of the flavor of SRL, many approaches for parsing (or labeling) are heavily reliant on syntactic features (Pradhan et al., 2005; Punyakanok et al., 2008; Li et al., 2018; Fei et al., 2021). The progress in Neural Networks has allowed systems to become more syntax-agnostic (Zhou and Xu, 2015; He et al., 2017; Tan et al., 2018; Rudinger et al., 2018; Arora et al., 2022; Spaulding et al., 2023), so much that recent approaches extract not just the arguments, but also the predicates themselves (Cai et al., 2018; He et al., 2018; Zhang et al., 2022). This is particularly appealing, as it allows to directly transform a text into a MR.

3.2 Rhetorical Structure Theory

Rhetorical Structure Theory (RST, Mann and Thompson, 1988) takes interest in discourse relations. It sees the text as a sequence of *Elementary Discourse Units (EDUs)*, which roughly correspond to events, and seeks to identify the relations between these units, such as `Condition`, `Contrast`, `Cause`, `Result`, or `Elaboration`. RST models a text as a tree, in which discourse relations are recursively applied to connect discourse units. Leaf nodes are EDUs (text spans), while inner nodes are unlabeled compositional nodes. Figure 3 shows the RST MR of our example sentence. The EDUs coincide with the spans delimited by predicates and arguments in the SRL graph. Each discourse relation links a *satellite* (supporting information) to a *nucleus* (central information). In our example, the nucleus of the `Reason` relation is the fact that Tiffany decided to never fly again, and the satellite is the reason for that decision. The repertoire of discourse relations depends on the dataset.

Discourse relations can cross sentence boundaries, which means that one rhetorical structure can represent a multi-sentence document. RST has been used for Summarization (Xu et al., 2020) and Question Answering (Ouyang et al., 2021), and even for Argument Mining (Peldszus and Stede, 2013; Mitrović et al., 2017; Chakrabarty et al., 2019).

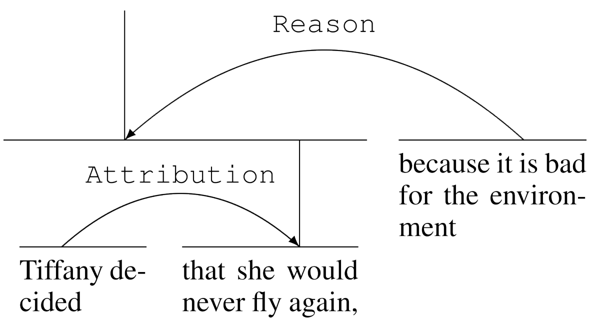

Figure 3: RST-DT style annotation for our example.

Resources. The main dataset for RST is RST-DT (Carlson et al., 2001), which defines 78 discourse relations, divided into 16 classes. The dataset contains 385 documents from the Wall Street Journal corpus, which have been annotated with around 20,000 EDUs by expert linguists on the basis of an extensive annotation manual.

Parsing is usually performed in two steps: EDU Segmentation and Tree Building. Wang et al. (2018) achieves a near-perfect performance on segmentation using a Bi-LSTM-CRF based model. First approaches for Tree Building (Soricut and Marcu, 2003; Hernault et al., 2010) used handcrafted features. Ji and Eisenstein (2014) introduced the first RST-DT neural parser, followed by bottom-up parsers (Li et al., 2016; Braud et al., 2017; Wang et al., 2017; Yu et al., 2018), and more recently top-down ones (Lin et al., 2019; Zhang et al., 2020a; Kobayashi et al., 2020). Though they have different approaches, Nguyen et al. (2021) and Koto et al. (2021) are the current best-performing systems for this task.

3.3 Universal Decompositional Semantics

Universal Decompositional Semantics (UDS) is a multi-layer semantic annotation scheme, which means that it allows annotating the same sentence on different dimensions. These dimensions include factuality and time for predicates, and genericity and word sense for arguments. UDS builds a semantic compositional tree, where the leaf nodes are the words of the sentence (or special tokens) and inner nodes represent larger semantic units. The graph structure is based on PredPatt (White et al., 2016), a pattern-based framework for predicate-argument extraction that operates on (syntactic) Universal Dependencies (UD, de Marneffe et al., 2021). It focuses on event-level relations, which means that the extracted structure is close to that of merged SRL graphs. UDS uses Dowty's Proto Roles decomposition, which, as mentioned above, describe features of event participants and how they are affected by the event (*movement*, *volition*, *change of state*, and so on).

Resources. The UDS dataset can be accessed through the Decomp Toolkit (White et al., 2020). The original annotations include proto-roles (Reisinger et al., 2015), word sense, and factuality. Around 10,000 arguments were annotated with proto-roles, using answers of Mechanical Turk workers to simple questions about the arguments. The framework was enriched with annotations on time (Vashishtha et al., 2019), and generalizing statements (Govindarajan et al., 2019), and also some discourse relations (Gantt et al., 2022).

Parsing. UDS Parsing is a fairly unexplored task. Zhang et al. (2018) performs cross-lingual UDS parsing with a pipeline approach performing graph transduction, coreference resolution and semantic proto-role labeling. Stengel-Eskin et al. (2020) proposes an end-to-end parser with an encoder-decoder structure, while Stengel-Eskin et al. (2021) parses UD and UDS jointly.

4 Deep Meaning Representation Frameworks

Deep Meaning Representation Frameworks go further than shallow ones by representing relations at all levels of the text, in particular at the sub-event level. They aim to model the meaning of the text exhaustively, representing as many phenomena as possible (negations, comparisons, modifiers, time, cause, etc.).

4.1 Semantic Dependencies

Semantic Dependencies (SD) is a family of MR frameworks that are used in the SemEval 2014 & 2015 challenges (Oepen et al., 2014, 2015). Their aim is to go further than syntactic dependency parsing, and to represent the semantic structure of a sentence – a process called Semantic Dependency Parsing (SDP). Four main frameworks have been proposed, derived from independent annotation schemes with different formalisms: DM (DELPH-IN MRS-Derived Bi-Lexical Dependencies, Flickinger et al., 2012), PAS (Enju Predicate-Argument Structures, Miyao, 2006), PSD (Prague Semantic Dependencies, Hajič et al., 2012), and CCD (Combinatory Categorial Grammar Dependencies, Hockenmaier and Steedman, 2007).

All frameworks see the semantic structure as a dependency (non-compositional) graph with Flavor 0 (every node corresponds to exactly one word in the sentence). In contrast to syntactic dependency trees, the modeling of semantic dependencies requires a graph, as nodes can have several incoming edges (a phenomenon called *re-entrancy*) if a word is the argument of several predicates, as well as none if they are semantically vacuous. A node is a word that can be augmented with its lemma, POS-tag or framework-specific identifier. The exact vocabulary of semantic roles, as well as the way the graph models different phenomena, varies across frameworks. Most of them use unspecific semantic roles (`ARG1`, `ARG2`, `ARG3`, ...). Nevertheless, similar to SRL, `ARG1` and `ARG2` usually correspond to Dowty's Proto-Agent and Proto-Patient.

Still, SD has the advantage to be easily understandable by human readers. Figure 4 shows DM annotations for our example sentence. The event decomposition goes all the way to the token level: the adjective phrase "bad for the environment" is seen as an object of interest, with "for" being a predicate, with the arguments "bad" and "environment".

Resources. Oepen et al. (2016) proposes a corpus with annotations for all four frameworks, with close to 37,000 English sentences from the WSJ corpus, which were obtained through expert annotation. The dataset also provides a corpus of annotations in ohter languages: Chinese for PAS, and Czech for PSD. Other corpora are formalism-specific: DeepBank for DM (Flickinger et al., 2012), the Enju Treebank for PAS (Miyao, 2006), the Prague Czech-English Dependency Treebank for PAS (Hajič et al., 2012), CCGBank for CCD (Hockenmaier and Steedman, 2007).

Parsing. Most parsing approaches for SDP are inspired by syntactic dependency parsing (Dozat and Manning, 2018; Fernández-González and Gómez-Rodríguez, 2020). The best results across the different SDP variants are achieved by a multi-task system (Wang et al., 2021b).

Variations. English Resource Grammar (ERG), of which DM is a reduction, produces MRs in the Minimal Recursion Semantics (Copestake et al., 2005). These structures are particularly expressive and can model scope, but they are also complex to read and exploit. Elementary Discourse Structures (EDS, Oepen and Lønning, 2006) try to reduce this complexity by making the graph non-compositional. The main difference between EDS and DM is that EDS are Flavor 1 graphs, which has more abstract node labels: in addition to POS tags and identifiers, nodes can be labeled with properties, such as time or number.

4.2 Universal Conceptual Cognitive Annotation

The Universal Conceptual Cognitive Annotation (UCCA, Abend and Rappoport, 2013) is a semantic annotation scheme aiming to be "universal", i.e., it aims to be resistant to syntactic variation within and across languages. An UCCA Representation takes the form of a compositional tree whose leaf nodes are the words of the sentence, and intermediate

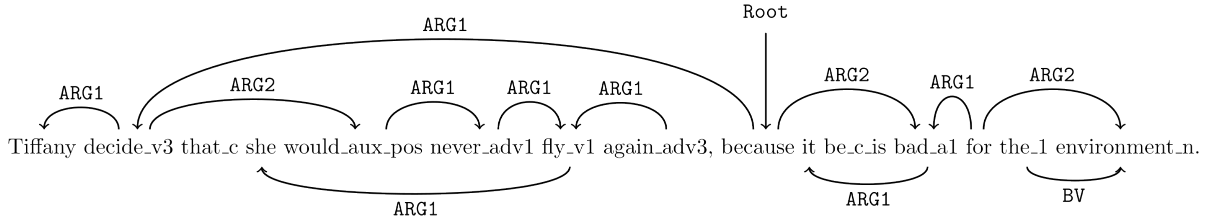

Figure 4: A Semantic Dependency Parse in DM-style for our running example.

nodes, called `units`, are unlabeled. UCCA identifies 3 levels of semantic information. On the central level, scene units correspond to events. They are linked to a predicate, to its core arguments by a generic label `participant`, as well as to non-core arguments using several other labels (see Figure 5). On the lower level, *sub-scene units* help specify the participants of a scene. Finally, *superparallel units* can link two scenes with generic `parallel scene` edges, and possibly a cue word indicating the type of discourse relation with a `linker` edge. At any level, *functional units* can represent phenomena such as prepositions, articles, or expletive pronouns. UCCA can annotate several sentences in a single graph.

There are very few semantic roles in UCCA, which makes the annotation task more accessible to non-experts and portable to other languages. Semantic roles have a generic interpretability, but it can be hard to exploit them directly: for instance, the `participant` role doesn't make a difference between what would be labeled as `ARG0` (Agent) and `ARG1` (Patient) in other frameworks. UCCA is multi-layered, which makes it possible to add extensions to the representation, for instance to annotate co-reference links, more specific semantic roles, or more abstract node types. UCCA is cross-lingual, and as such found applications in Machine Translation (Slobodkin et al., 2022; Birch et al., 2016), but also in Text Simplification (Sulem et al., 2018a,b).

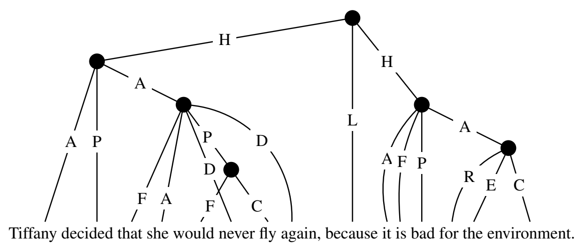

Figure 5: UCCA graph for our example. H: Parallel Scene, L: Linker, P: Process, A: Participant, D: Adverbial, F: Function C: Center, E: Elaborator, R: Relator.

Resources. UCCA comes with a large expert-annotated multilingual corpus (Abend and Rappoport, 2013). Its English version annotates a total of 1350 passages (more than 200,000 tokens). This includes not only elements from Wikipedia, the Web- and Penn Treebanks, but also from the literature (e.g. *The Little Prince*).

Parsing. The first proposed parser for UCCA (Hershcovich et al., 2017) was transition-based. Other methods exploit constituency parsers (Jiang et al., 2019; Bölücü and Can, 2021). Nowadays, the best parsers are sequence-to-sequence models (Ozaki et al., 2020; Samuel and Straka, 2020).

4.3 Abstract Meaning Representation

Abstract Meaning Representation (AMR, Banarescu et al., 2013) aims at further abstracting away from syntax, even mapping named entities to Wikipedia. AMR has no explicit alignments between nodes and the text. The representation itself takes the form of a rooted, acyclic, directed dependency graph, where each node (aka *variable*) is labeled with a *concept*, and represents an instance of this concept.[1] The root of an AMR is used for modeling the focus, or main event, of a text. Figure 1 shows the AMR graph for our running example.

AMR has an abstract node type: a node can be labeled with PropBank frames (for events and entities), English word lemmas, or special frames (e.g. for dates, modalities, negations, comparisons, or family relationships). Semantic roles are either PropBank roles, which have accessible predicate-specific interpretation, or manually-crafted ones (e.g. `:name`, `:location`, `:cause`, `:concession`, `:month`, `:poss`, `degree`...).

Many AMR roles can be reified and used as concepts, thus allowing the focus to be on the relation itself. AMR also makes use of re-entrancy:

[1] There are cases where a variable is the same as the concept (e.g. for negation modelled with an '-'). We generalize over such special cases.

in our example, Tiffany appears only once as a node, and is linked to both `decide-01` and `fly-01`. AMR also represents explicit quantities and temporal relations. This makes AMR graphs nearly unambiguous.

Of all MRFs, AMR has probably garnered the most attention in recent years. It has been used in tasks such as Machine Translation (Song et al., 2019), Question Answering (Kapanipathi et al., 2021; Lim et al., 2020; Xu et al., 2021), Toxic Content Detection (Elbasani and Kim, 2022), Semantic Search and Natural Language Inference (Opitz and Frank, 2022b; Opitz et al., 2023b), and Social Reasoning (Chanin and Hunter, 2023).

Resources. The most important AMR corpus is the AMR Annotation Release (Banarescu et al., 2013). It was constructed fully manually by trained annotators, and contains about 60,000 English AMR graphs in its latest (3.0) version, including multi-sentence graphs (O'Gorman et al., 2018). AMR graphs are often linearized in the 'Penman' form (Kasper, 1989), which is easy to read, and allows processing with neural models in a sequence-to-sequence manner (the Penman form uses a depth-first traversal and can, in principle, linearize any directed rooted graph).

Parsing. Many AMR parsers have been proposed through the years, graph-based (Flanigan et al., 2014; Werling et al., 2015; Cai and Lam, 2020), transition-based (Wang et al., 2015; Vilares and Gómez-Rodríguez, 2018; Lee et al., 2020), or seq-2-seq (Barzdins and Gosko, 2016; Peng et al., 2018; Bevilacqua et al., 2021), possibly leveraging adapters to better incorporate graph topology (Vasylenko et al., 2023). Most systems of the 2020s leverage large pre-trained language models and achieve strong performance on AMR 3.0.

Extensions. AMR has been extended to model tense and aspect (Donatelli et al., 2018), as well as scope (Pustejovsky et al., 2019), and larger documents (Naseem et al., 2022). The BabelNet Meaning Representation (Navigli et al., 2022) aims at making it multilingual by using BabelNet synsets for concepts (Navigli et al., 2021) and semantic roles from VerbAtlas (Di Fabio et al., 2019). Perhaps even more ambitiously, the Universal Meaning Representation (UMR, Van Gysel et al., 2021) aims at compensating all main shortcomings of AMR, adding aspect and scope, integrating document-level annotations with coreference, temporal and modal relations between sentences, and making the representation language-agnostic.

4.4 Discourse Representation Structure

Discourse Representation Structure (DRS) is the fruit of Discourse Representation Theory (DRT, Kamp, 1981; Kamp and Reyle, 1993) and provides a meaning representation that fully integrates with first order logic. We focus here on the characteristics of the DRS format used in the Parallel Meaning Bank (PMB, Abzianidze et al., 2017), based on Segmented Discourse Representation Theory (Asher and Lascarides, 2003), which augments DRT with discourse relations.

A Discourse Representation Structure (DRS) is a recursive structure of nested boxes. Figure 6 shows the representation of our example sentence.

While AMR links concepts to PropBank, concepts in DRS are linked to *Wordnet* synsets (Miller, 1995). Wordnet has a very wide coverage of English, which means that most words can be mapped to such Wordnet synsets. Semantic roles are taken from VerbNet (Kipper et al., 2000), augmented by hand-crafted roles (e.g. `Quantity`, `Name`, `Owner`, `Time`). These roles are generic, dispensing of predicate-specific interpretation.

Usually, a simple box represents a situation (an *event* in our terminology, similar to an EDU). Discourse relations are represented similarly to semantic roles, but with boxes as arguments. This means that DRS is compositional, and naturally equipped for multi-sentence representation. Modal logic operators can also be applied to boxes (negation, possibility, and necessity), which allows for a precise scoping of these operators: in the example, "she will never fly again" is represented as the negation of the box expressing that Tiffany flies at some point in the future.

Even though there may be no ideal way to transform a DRS into a graph (Abzianidze et al., 2020), we can see concepts as nodes, and semantic roles as labels of the edges between these nodes. Boxes would be another type of nodes, with discourse relations linking them. The most recent development of DRS, the Sequence Notation (Bos, 2023), proposes a similar graph equivalent. With this view, DRS are compositional graphs, where high-level nodes represent scope.

Resources. DRS annotations are hard to produce even for experts, which makes constructing large corpora difficult. The Groningen Meaning Bank (GMB, Basile et al., 2012) was the first DRS corpus, followed by the Parallel Meaning Bank (PMB, Abzianidze et al., 2017). These banks were built using an semi-automatic pipeline based on the rule-based parser DRS parser 'Boxer' (Bos, 2008) and a CCG parser. The latest release contains almost 10,000 human-corrected English 'gold' documents. In addition, the PMB tries to make DRS language-neutral by associating English documents with translations to one or several languages.

Parsing. Several DRS parsers are available, exploiting transition-based parsing (Evang, 2019), DAG Grammars (Fancellu et al., 2019) or POS-tags and dependency graphs (van Noord, 2019). Modern parsers use LLMs (van Noord et al., 2018, 2020) and generally outperform older ones.

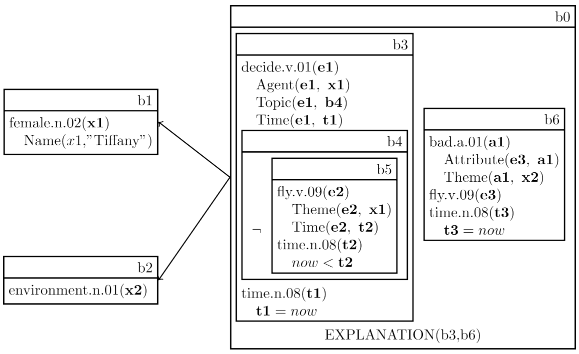

Figure 6: DRS for our running example.

5 Current Research Trends

Synthesizing insights from our overview of MRFs, we see research trends in three main areas for Deep MRFs: MRF design, MR parsing, and MR applications.

5.1 Trends in MRF design

MRs seem to lend themselves to multi-linguality, since they represent semantic concepts such as *agent*, *patient*, *instrument*, and *cause* that appear to be universal. However, these concepts, and more generally their structure, are still based on English semantics. Only UCCA, built on Basic Linguistic Theory (Dixon, 2009), is natively fully language independent. To make them more language-agnostic, some MRFs are being equipped with parallel corpora, node labels, and even more neutral structure (Abzianidze et al., 2020; Navigli et al., 2022; Giordano and Lopez, 2023; Van Gysel et al., 2021).

Multi-sentence representation is also a topic of research. Compositional MRFs are naturally well-equipped for this, to the point that RST and DRS are already able to represent multi-sentence document in one MR. Several AMR extensions (Naseem et al., 2022; Van Gysel et al., 2021) also work in this direction.

Another trend is to make MRFs more expressive. This happens along three axes: One axis extends existing MRFs (as illustrated by AMR extensions for tense or scope modeling, see above); another axis uses multi-layer annotation schemes (as exemplified by UCCA or UDS); and yet another one employs more complex structures (as DRS does).

However, there is a trade-off between expressivity and simplicity of a MRF. The simpler an MRF is, the easier it is to generate training data sets. In fact, several works aim at reducing the annotation load: some aim at crowd-sourcing MRs (e.g., by re-formulating annotation tasks into simple questions White et al., 2016); others improve annotation tools (e.g., with a CodePilot machine-in-the-loop Cai et al., 2023); and again others create new, simpler MRFs, based on AMR (Feng et al., 2023) or DRS (Bos, 2023).

It is interesting to note that most works on AMR focus on increasing expressivity, while works on DRS tend to focus on improving simplicity. This might indicate hat the right balance lies somewhere between the two.

5.2 Trends in MR Parsing

Parsing. For humans, producing an MR is an arduous task, particularly for abstract frameworks: a trained annotator needs about 10 minutes to annotate a sentence in AMR (Banarescu et al., 2013). Therefore, much research has been dedicated to building automatic parsing systems, with many ideas shared between frameworks in Deep Meaning Representations Parsing, as highlighted by the SemEval shared tasks (Oepen et al., 2019, 2020). Traditional approaches to Neural MR Parsing usually fall into two main categories: graph-prediction, which try to identify nodes and the best edge assignment, and transition-based parsers, which build the graph iteratively with a restrained set of actions and a stack-buffer structure. Graph-prediction seems particularly suited for SDP, as the nodes are the input tokens bearing strong relations to syntactic dependency parsing (Almeida and Martins, 2015; Dozat and Manning, 2018). However, when integrated into a pipeline that performs concepts identification and afterwards relations prediction, it straightforwardly extends to AMR (Flanigan et al., 2014; Werling et al., 2015; Lyu and Titov, 2018) or EDS (Cao et al., 2021; Chen et al., 2019). Transition-based parsers seem suited to predict abstract structures, and were used for building UCCA parsers (Hershcovich et al., 2017; Jiang et al., 2019; Bölücü and Can, 2021), DRS parsers (Evang, 2019; Fancellu et al., 2019) and AMR parsers (Wang et al., 2015; Vilares and Gómez-Rodríguez, 2018). Most parsers now use sequence-to-sequence architectures (Ozaki et al., 2020; Samuel and Straka, 2020; van Noord et al., 2018, 2020; Bevilacqua et al., 2021; Zhou et al., 2021). These models take text as input, and output the linearized graph. However, there is a wide variety of learning strategies: graph pre-training (Bai et al., 2022; Wang et al., 2023), instruction fine-tuning (Lee et al., 2023), graph information distillation (Vasylenko et al., 2023), or even prompting (Ettinger et al., 2023). Other approaches mix deep learning with classical ideas, using the representations of language models in transition-based parsing (Astudillo et al., 2020; Zhou et al., 2021), graph-prediction parsing (Lyu and Titov, 2018), or ensembling (Hoang et al., 2021; Lorenzo et al., 2023).

Evaluation. The evaluation of graph-based MRs is 'classically' addressed through metrics such as SMATCH that measures the structural similarity of the output graph to a reference graph (Allen et al., 2008; Cai and Knight, 2013; Opitz, 2023). Three main issues have been observed in SMATCH: First, the procedure it is inefficient, because computing graph isomorphism is NP complete. Different heuristics have been developed to remedy this problem, based on graph traversals (Song and Gildea, 2019; Liu et al., 2020) or SMATCH distillation (Opitz et al., 2023a). The second issue is that SMATCH evaluations consider only the graph structure, and fails to see, e.g., that a node `cat` is similar to a node `kitten` or a sub-graph `cat :mod young`. Different neural networks and graph algorithms are developed to remedy this issue (Opitz et al., 2020, 2021; Shou and Lin, 2023). Finally, SMATCH struggles to discriminate between strong parsers (Opitz and Frank, 2022a). Finer semantic graph measures are thus being developed, using neural networks or graph algorithms (Opitz et al., 2020, 2021; Shou and Lin, 2023; Kachwala et al., 2024). As an alternative approach to metric evaluation, MRF-specific 'challenge sets' are proposed for AMR (Groschwitz et al., 2023) and DRS (Wang et al., 2021a), to test parsers across a suite of tasks, e.g., difficult 'Winograd' pronouns (Levesque et al., 2012), or tense.

5.3 Trends in MR Application

We may wonder what is the place of MRFs in a domain dominated by always better-performing large language models (LLMs). However, different from LLMs, MRs make all facets of the meaning of a text explicit, which can provide **accuracy**, **control**, **robustness**, and **explainability** to any NLP pipeline. And indeed, these assets have been leveraged in several ways, also in combination with LLMs.

A 'classic' strategy is to use the MR as *supporting information*, which is exploited in a neural architecture. For this, MRs can be fed into sequence encoders as linearized strings (Ouyang et al., 2021; Xu et al., 2020), or into graph neural networks that exploit structure (Song et al., 2019; Xu et al., 2021; Lim et al., 2020; Ribeiro et al., 2022). Other works use discourse-level information to perform scene-aware attention (Slobodkin et al., 2022), or combine sentence and MR embeddings to refine representations (Cai et al., 2022).

Another strategy is to *exploit the graphs directly in symbolic or neuro-symbolic pipelines*, so as to control the results or explain them. Some works leverage MRs for improved paraphrasing (Cai et al., 2021; Huang et al., 2023) and style transfer (Jangra et al., 2022), neutralizing 'translationese' in translation references (Wein and Schneider, 2024), or link prediction (Yang et al., 2023). Other approaches apply MR-to-text generation after manipulating or splitting MR subgraphs, e.g., for data augmentation (Shou et al., 2022) or text simplification (Sulem et al., 2018b). Graph metrics are used to assist textual inference between pairs of sentences (Bonial et al., 2020; Opitz et al., 2023b). Other works use MRs for symbolic reasoning (Kapanipathi et al., 2021; Chanin and Hunter, 2023).

Yet another strategy is to *indirectly exploit* MRs: Opitz and Frank (2022b) partition text embeddings into interpretable linguistic features by binding distances between embedding parts to distances between MR-subgraphs that elicit, e.g., polarity, or semantic roles. A technical advantage of this is that a parser is not required at inference.

MRFs are thus being combined fruitfully with LLMs, contributing interpretability, useful intermediate representations, and a bridge towards formal logic.

Acknowledgement

This work was partially funded by the NoRDF project (ANR-20-CHIA-0012-01). We thank Johan Bos and Nathan Schneider for correcting minor inaccuracies for this updated version of the paper.

6 Limitations

Our survey is limited to graph-like meaning representations. While these are indeed the most popular meaning representations these days, there are others that could be discussed in this survey. Raymond Mooney's ground-breaking works (Mooney, 1996; Wong and Mooney, 2006; Mooney, 2007), e.g., or L. Zettlemoyer's work on CCG parsing (Kwiatkowski et al., 2011; Wang et al., 2014; Dasigi et al., 2019), aim at building MRs from a corpus for a target application. The compactness of this survey also prevents us from going more into detail for the parsing techniques. While we do discuss current methods and future trends, parsing itself could merit a survey. The majority of applications presented in Section 5 stem from AMR. This is simply because AMR is the most popular MRF, with very well-performing parsers. However, this does not mean that the other MRFs are less useful. They each have their unique properties that predispose them to different applications.

References

Omri Abend and Ari Rappoport. 2013. Universal Conceptual Cognitive Annotation (UCCA). In *ACL*.

Omri Abend and Ari Rappoport. 2017. The state of the art in semantic representation. In *ACL*.

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 *EACL*.

Lasha Abzianidze, Johan Bos, and Stephan Oepen. 2020. DRS at MRP 2020: Dressing up discourse representation structures as graphs. In *CoNLL*.

James Allen, Mary Swift, and William de Beaumont. 2008. Deep semantic analysis of text. In *STEP 2008*.

Mariana S. C. Almeida and André F. T. Martins. 2015. Lisbon: Evaluating TurboSemanticParser on multiple languages and out-of-domain data. In *SemEval*.

Aashish Arora, Harshitha Malireddi, Daniel Bauer, Asad Sayeed, and Yuval Marton. 2022. Multi-task learning for joint semantic role and proto-role labeling. *arXiv:2210.07270*.

Nicolas Asher and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press.

Ramón Astudillo, Miguel Ballesteros, Tahira Naseem, Austin Blodgett, and Radu Florian. 2020. Transition-based parsing with stack-transformers. In *EMNLP Findings*.

Xuefeng Bai, Yulong Chen, and Yue Zhang. 2022. Graph pre-training for AMR parsing and generation. In *ACL*.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *COLING*.

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 *LAW*.

Guntis Barzdins and Didzis Gosko. 2016. RIGA at SemEval-2016 task 8: Impact of Smatch extensions and character-level neural translation on AMR parsing accuracy. In *SemEval*.

Valerio Basile, Johan Bos, Kilian Evang, and Noortje Venhuizen. 2012. Developing a large semantically annotated corpus. In *LREC*.

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 *AAAI*.

Alexandra Birch, Omri Abend, Ondřej Bojar, and Barry Haddow. 2016. HUME: Human UCCA-based evaluation of machine translation. In *EMNLP*.

Necva Bölücü and Burcu Can. 2021. Self-attentive constituency parsing for ucca-based semantic parsing. *arXiv:2110.00621*.

Claire Bonial, Stephanie M. Lukin, David Doughty, Steven Hill, and Clare Voss. 2020. InfoForager: Leveraging semantic search with AMR for COVID-19 research. In *DMR*.

Johan Bos. 2008. Wide-coverage semantic analysis with Boxer. In *STEP*.

Johan Bos. 2023. The sequence notation: Catching complex meanings in simple graphs. In *IWCS*.

Chloé Braud, Maximin Coavoux, and Anders Søgaard. 2017. Cross-lingual RST discourse parsing. In *EACL*.

Aljoscha Burchardt and Marco Pennacchiotti. 2008. FATE: a FrameNet-annotated corpus for textual entailment. In *LREC*.

Deng Cai and Wai Lam. 2020. AMR parsing via graph-sequence iterative inference. In *ACL*.

Deng Cai, Xin Li, Jackie Chun-Sing Ho, Lidong Bing, and Wai Lam. 2022. Retrofitting multilingual sentence embeddings with Abstract Meaning Representation. In *EMNLP*.

Jiaxun Cai, Shexia He, Zuchao Li, and Hai Zhao. 2018. A full end-to-end semantic role labeler, syntactic-agnostic over syntactic-aware? In *COLING*.

Jon Cai, Shafiuddin Rehan Ahmed, Julia Bonn, Kristin Wright-Bettner, Martha Palmer, and James H. Martin. 2023. CAMRA: Copilot for AMR annotation. In *EMNLP*.

Shu Cai and Kevin Knight. 2013. Smatch: An evaluation metric for semantic feature structures. In *ACL*.

Yitao Cai, Yue Cao, and Xiaojun Wan. 2021. Revisiting pivot-based paraphrase generation: Language is not the only optional pivot. In *EMNLP*.

Junjie Cao, Zi Lin, Weiwei Sun, and Xiaojun Wan. 2021. Comparing knowledge-intensive and data-intensive models for English resource semantic parsing. *Computational Linguistics*, 47(1).

Lynn Carlson, Daniel Marcu, and Mary Ellen Okurovsky. 2001. Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. In *SIGDIAL*.

Xavier Carreras and Lluís Màrquez. 2004. Introduction to the CoNLL-2004 shared task: Semantic role labeling. In *CoNLL*.

Xavier Carreras and Lluís Màrquez. 2005. Introduction to the CoNLL-2005 shared task: Semantic role labeling. In *CoNLL*.

Tuhin Chakrabarty, Christopher Hidey, Smaranda Muresan, Kathy McKeown, and Alyssa Hwang. 2019. AMPERSAND: Argument mining for PERSuAsive oNline discussions. In *EMNLP-IJCNLP*.

David Chanin and Anthony Hunter. 2023. Neuro-symbolic commonsense social reasoning. *arXiv:2303.08264*.

Yufei Chen, Yajie Ye, and Weiwei Sun. 2019. Peking at MRP 2019: Factorization- and composition-based parsing for elementary dependency structures. In *CoNLL*.

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

Dipanjan Das and Noah A. Smith. 2011. Semi-supervised frame-semantic parsing for unknown predicates. In *ACL*.

Pradeep Dasigi, Matt Gardner, Shikhar Murty, Luke Zettlemoyer, and Eduard Hovy. 2019. Iterative search for weakly supervised semantic parsing. In *NAACL*.

Marie-Catherine de Marneffe, Christopher D. Manning, Joakim Nivre, and Daniel Zeman. 2021. Universal Dependencies. *Computational Linguistics*, 47(2).

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 *EMNLP-IJCNLP*.

Robert MW Dixon. 2009. *Basic linguistic theory volume 1: Methodology*, volume 1. OUP Oxford.

Lucia Donatelli, Michael Regan, William Croft, and Nathan Schneider. 2018. Annotation of tense and aspect semantics for sentential AMR. In *LAW-MWE-CxG*.

David Dowty. 1991. Thematic proto-roles and argument selection. *Language*, 67(3).

Timothy Dozat and Christopher D. Manning. 2018. Simpler but more accurate semantic dependency parsing. In *ACL*.

Ermal Elbasani and Jeong-Dong Kim. 2022. AMR-CNN: Abstract Meaning Representation with convolution neural network for toxic content detection. *Journal of Web Engineering*, 21(03).

Allyson Ettinger, Jena Hwang, Valentina Pyatkin, Chandra Bhagavatula, and Yejin Choi. 2023. "You are an expert linguistic annotator": Limits of LLMs as analyzers of Abstract Meaning Representation. In *EMNLP Findings*.

Kilian Evang. 2019. Transition-based DRS parsing using stack-LSTMs. In *IWCS*.

Federico Fancellu, Sorcha Gilroy, Adam Lopez, and Mirella Lapata. 2019. Semantic graph parsing with recurrent neural network DAG grammars. In *EMNLP-IJCNLP*.

Hao Fei, Shengqiong Wu, Yafeng Ren, Fei Li, and Donghong Ji. 2021. Better combine them together! integrating syntactic constituency and dependency representations for semantic role labeling. In *ACL-IJCNLP Findings*.

Lydia Feng, Gregor Williamson, Han He, and Jinho D Choi. 2023. Widely interpretable semantic representation: Frameless meaning representation for broader applicability. *arXiv:2309.06460*.

Daniel Fernández-González and Carlos Gómez-Rodríguez. 2020. Transition-based semantic dependency parsing with pointer networks. In *ACL*.

Jeffrey Flanigan, Sam Thomson, Jaime Carbonell, Chris Dyer, and Noah A. Smith. 2014. A discriminative graph-based parser for the Abstract Meaning Representation. In *COLING*.

Dan Flickinger, Yi Zhang, and Valia Kordoni. 2012. Deepbank. a dynamically annotated treebank of the wall street journal. In *TLT*.

William Gantt, Lelia Glass, and Aaron Steven White. 2022. Decomposing and recomposing event structure. *Transactions of the ACL*, 10.

Qin Gao and Stephan Vogel. 2011. Corpus expansion for statistical machine translation with semantic role label substitution rules. In *ACL*.

Daniel Gildea and Daniel Jurafsky. 2000. Automatic labeling of semantic roles. In *ACL*.

Bastien Giordano and Cédric Lopez. 2023. MR4AP: Meaning representation for application purposes. In *DMR*.

Venkata Govindarajan, Benjamin Van Durme, and Aaron Steven White. 2019. Decomposing generalization: Models of generic, habitual, and episodic statements. *Transactions of the ACL*, 7.

Jonas Groschwitz, Shay Cohen, Lucia Donatelli, and Meaghan Fowlie. 2023. AMR parsing is far from solved: GrAPES, the granular AMR parsing evaluation suite. In *EMNLP*.

Jan Hajič, Massimiliano Ciaramita, Richard Johansson, Daisuke Kawahara, Maria Antònia Martí, Lluís Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Štěpánek, Pavel Straňák, Mihai Surdeanu, Nianwen Xue, and Yi Zhang. 2009. The CoNLL-2009 shared task: Syntactic and semantic dependencies in multiple languages. In *CoNLL*.

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 *LREC*.

Silvana Hartmann, Ilia Kuznetsov, Teresa Martin, and Iryna Gurevych. 2017. Out-of-domain FrameNet semantic role labeling. In *EACL*.

Luheng He, Kenton Lee, Omer Levy, and Luke Zettlemoyer. 2018. Jointly predicting predicates and arguments in neural semantic role labeling. In *ACL*.

Luheng He, Kenton Lee, Mike Lewis, and Luke Zettlemoyer. 2017. Deep semantic role labeling: What works and what's next. In *ACL*.

Hugo Hernault, Helmut Prendinger, David duVerle, and Mitsuru Ishizuka. 2010. HILDA: A Discourse Parser Using Support Vector Machine Classification. *Dialogue & Discourse*, 1(3).

Daniel Hershcovich, Omri Abend, and Ari Rappoport. 2017. A transition-based directed acyclic graph parser for UCCA. In *ACL*.

Thanh Lam Hoang, Gabriele Picco, Yufang Hou, Young-Suk Lee, Lam Nguyen, Dzung Phan, Vanessa Lopez, and Ramon Fernandez Astudillo. 2021. Ensembling graph predictions for AMR parsing. In *NeurIPS*.

Julia Hockenmaier and Mark Steedman. 2007. CCGbank: A corpus of CCG derivations and dependency structures extracted from the Penn Treebank. *Computational Linguistics*, 33(3).

Kuan-Hao Huang, Varun Iyer, I-Hung Hsu, Anoop Kumar, Kai-Wei Chang, and Aram Galstyan. 2023. ParaAMR: A large-scale syntactically diverse paraphrase dataset by AMR back-translation. In *ACL*.

Anubhav Jangra, Preksha Nema, and Aravindan Raghuveer. 2022. T-STAR: Truthful style transfer using AMR graph as intermediate representation. In *EMNLP*.

Yangfeng Ji and Jacob Eisenstein. 2014. Representation learning for text-level discourse parsing. In *ACL*.

Wei Jiang, Zhenghua Li, Yu Zhang, and Min Zhang. 2019. HLT@SUDA at SemEval-2019 task 1: UCCA graph parsing as constituent tree parsing. In *SemEval*.

Zoher Kachwala, Jisun An, Haewoon Kwak, and Filippo Menczer. 2024. Rematch: Robust and efficient matching of local knowledge graphs for improved structural and semantic similarity. In *NAACL*.

Hans Kamp. 1981. A theory of truth and semantic representation. In *Formal Semantics - the Essential Readings*. Blackwell.

Hans Kamp and Uwe Reyle. 1993. *From Discourse to Logic*. Springer Dordrecht.

Pavan Kapanipathi, Ibrahim Abdelaziz, Srinivas Ravishankar, Salim Roukos, Alexander Gray, Ramón Fernandez 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 *ACL-IJCNLP Findings*.

Robert T. Kasper. 1989. A flexible interface for linking applications to Penman's sentence generator. In *HLT*.

Karin Kipper, Hoa Trang Dang, and Martha Palmer. 2000. Class-based construction of a verb lexicon. In *AAAI*.

Naoki Kobayashi, Tsutomu Hirao, Hidetaka Kamigaito, Manabu Okumura, and Masaaki Nagata. 2020. Top-down RST parsing utilizing granularity levels in documents. In *AAAI*.

Fajri Koto, Jey Han Lau, and Timothy Baldwin. 2021. Top-down discourse parsing via sequence labelling. In *EACL*.

Tom Kwiatkowski, Luke Zettlemoyer, Sharon Goldwater, and Mark Steedman. 2011. Lexical generalization in CCG grammar induction for semantic parsing. In *EMNLP*.

Young-Suk Lee, Ramón Fernandez Astudillo, Radu Florian, Tahira Naseem, and Salim Roukos. 2023. Amr parsing with instruction fine-tuned pre-trained language models. *arXiv:2304.12272*.

Young-Suk Lee, Ramón Fernandez Astudillo, Tahira Naseem, Revanth Gangi Reddy, Radu Florian, and Salim Roukos. 2020. Pushing the limits of AMR parsing with self-learning. In *EMNLP Findings*.

Hector Levesque, Ernest Davis, and Leora Morgenstern. 2012. The winograd schema challenge. In *KR*.

Qi Li, Tianshi Li, and Baobao Chang. 2016. Discourse parsing with attention-based hierarchical neural networks. In *EMNLP*.

Zuchao Li, Shexia He, Jiaxun Cai, Zhuosheng Zhang, Hai Zhao, Gongshen Liu, Linlin Li, and Luo Si. 2018. A unified syntax-aware framework for semantic role labeling. In *EMNLP*.

Kexin Liao, Logan Lebanoff, and Fei Liu. 2018. Abstract Meaning Representation for multi-document summarization. In *COLING*.

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 *COLING*.

Xiang Lin, Shafiq Joty, Prathyusha Jwalapuram, and M Saiful Bari. 2019. A unified linear-time framework for sentence-level discourse parsing. In *ACL*.

Fei Liu, Jeffrey Flanigan, Sam Thomson, Norman Sadeh, and Noah A. Smith. 2015. Toward abstractive summarization using semantic representations. In *NAACL*.

Jiangming Liu, Shay B. Cohen, and Mirella Lapata. 2020. Dscorer: A fast evaluation metric for discourse representation structure parsing. In *ACL*.

ACM Lorenzo, Pere Lluís Huguet Cabot, and Roberto Navigli. 2023. AMRs assemble! learning to ensemble with autoregressive models for AMR parsing. In *ACL*.

Chunchuan Lyu and Ivan Titov. 2018. AMR parsing as graph prediction with latent alignment. In *ACL*.

William C. Mann and Sandra A. Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text - Interdisciplinary Journal for the Study of Discourse*, 8(3).

George A. Miller. 1995. WordNet: A Lexical Database for English. *Communications of the ACM*, 38(11).

Jelena Mitrović, Cliff O'Reilly, Miljana Mladenović, and Siegfried Handschuh. 2017. Ontological representations of rhetorical figures for argument mining. *Argument & Computation*, 8(3).

Yusuke Miyao. 2006. *From linguistic theory to syntactic analysis. Corpus-oriented grammar development and feature forest model*. Ph.D. thesis, University of Tokyo, Tokyo, Japan.

Muhidin Mohamed and Mourad Oussalah. 2019. Srl-esa-textsum: A text summarization approach based on semantic role labeling and explicit semantic analysis. *Information Processing & Management*, 56(4).

Raymond J. Mooney. 1996. Inductive logic programming for natural language processing. In *ILP*. Springer.

Raymond J. Mooney. 2007. Learning for semantic parsing. In *CICLing*.

Tahira Naseem, Austin Blodgett, Sadhana Kumaravel, Tim O'Gorman, Young-Suk Lee, Jeffrey Flanigan, Ramón Astudillo, Radu Florian, Salim Roukos, and Nathan Schneider. 2022. DocAMR: Multi-sentence AMR representation and evaluation. In *NAACL*.

Roberto Navigli, Michele Bevilacqua, Simone Conia, Dario Montagnini, and Francesco Cecconi. 2021. Ten years of BabelNet: A survey. In *IJCAI*.

Roberto Navigli, Rexhina Blloshmi, and Abelardo Carlos Martínez Lorenzo. 2022. BabelNet Meaning Representation: A fully semantic formalism to overcome language barriers. In *AAAI*.

Thanh-Tung Nguyen, Xuan-Phi Nguyen, Shafiq Joty, and Xiaoli Li. 2021. RST parsing from scratch. In *NAACL*.

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 *CoNLL*.

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 *CoNLL*.

Stephan Oepen, Marco Kuhlmann, Yusuke Miyao, Daniel Zeman, Silvie Cinková, Dan Flickinger, Jan Hajič, Angelina Ivanova, and Zdeňka Urešová. 2016. Towards comparability of linguistic graph Banks for semantic parsing. In *LREC*.

Stephan Oepen, Marco Kuhlmann, Yusuke Miyao, Daniel Zeman, Silvie Cinková, Dan Flickinger, Jan Hajič, and Zdeňka Urešová. 2015. SemEval 2015 task 18: Broad-coverage semantic dependency parsing. In *SemEval*.

Stephan Oepen, Marco Kuhlmann, Yusuke Miyao, Daniel Zeman, Dan Flickinger, Jan Hajič, Angelina Ivanova, and Yi Zhang. 2014. SemEval 2014 task 8: Broad-coverage semantic dependency parsing. In *SemEval*.

Stephan Oepen and Jan Tore Lønning. 2006. Discriminant-based MRS banking. In *LREC*.

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 *COLING*.

Juri Opitz. 2023. SMATCH++: Standardized and extended evaluation of semantic graphs. In *EACL Findings*.

Juri Opitz, Angel Daza, and Anette Frank. 2021. Weisfeiler-leman in the bamboo: Novel AMR graph metrics and a benchmark for AMR graph similarity. *Transactions of the ACL*, 9.

Juri Opitz and Anette Frank. 2022a. Better Smatch = better parser? AMR evaluation is not so simple anymore. In *Eval4NLP*.

Juri Opitz and Anette Frank. 2022b. SBERT studies meaning representations: Decomposing sentence embeddings into explainable semantic features. In *AACL-IJCNLP*.

Juri Opitz, Philipp Meier, and Anette Frank. 2023a. SMARAGD: Learning SMatch for accurate and rapid approximate graph distance. In *IWCS*.

Juri Opitz, Letitia Parcalabescu, and Anette Frank. 2020. AMR similarity metrics from principles. *Transactions of the ACL*, 8.

Juri Opitz, Shira Wein, Julius Steen, Anette Frank, and Nathan Schneider. 2023b. AMR4NLI: Interpretable and robust NLI measures from semantic graphs. In *IWCS*.

Siru Ouyang, Zhuosheng Zhang, and Hai Zhao. 2021. Dialogue graph modeling for conversational machine reading. In *ACL-IJCNLP Findings*.

Hiroaki Ozaki, Gaku Morio, Yuta Koreeda, Terufumi Morishita, and Toshinori Miyoshi. 2020. Hitachi at MRP 2020: Text-to-graph-notation transducer. In *CoNLL*.

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

Siyana Pavlova, Maxime Amblard, and Bruno Guillaume. 2023. Structural and global features for comparing semantic representation formalisms. In *DMR*.

Andreas Peldszus and Manfred Stede. 2013. From argument diagrams to argumentation mining in texts: A survey. *International Journal of Cognitive Informatics and Natural Intelligence*, 7(1).

Xiaochang Peng, Linfeng Song, Daniel Gildea, and Giorgio Satta. 2018. Sequence-to-sequence models for cache transition systems. In *ACL*.

Lekshmi R Pillai, Veena G., and Deepa Gupta. 2018. A combined approach using semantic role labelling and word sense disambiguation for question generation and answer extraction. In *ICAECC*.

Sameer Pradhan, Wayne Ward, Kadri Hacioglu, James Martin, and Daniel Jurafsky. 2005. Semantic role labeling using different syntactic views. In *ACL*.

Vasin Punyakanok, Dan Roth, and Wen-tau Yih. 2008. The importance of syntactic parsing and inference in semantic role labeling. *Computational Linguistics*, 34(2).

James Pustejovsky, Ken Lai, and Nianwen Xue. 2019. Modeling quantification and scope in Abstract Meaning Representations. In *DMR*.

Drew Reisinger, Rachel Rudinger, Francis Ferraro, Craig Harman, Kyle Rawlins, and Benjamin Van Durme. 2015. Semantic proto-roles. *Transactions of the ACL*, 3.

Leonardo F. R. Ribeiro, Mengwen Liu, Iryna Gurevych, Markus Dreyer, and Mohit Bansal. 2022. FactGraph: Evaluating factuality in summarization with semantic graph representations. In *NAACL*.

Rachel Rudinger, Adam Teichert, Ryan Culkin, Sheng Zhang, and Benjamin Van Durme. 2018. Neural-Davidsonian semantic proto-role labeling. In *EMNLP*.

David Samuel and Milan Straka. 2020. ÚFAL at MRP 2020: Permutation-invariant semantic parsing in PERIN. In *CoNLL*.

Kaize Shi, Xueyao Sun, Li He, Dingxian Wang, Qing Li, and Guandong Xu. 2023. AMR-TST: Abstract Meaning Representation-based text style transfer. In *ACL*.

Ziyi Shou, Yuxin Jiang, and Fangzhen Lin. 2022. AMR-DA: Data augmentation by Abstract Meaning Representation. In *ACL*.

Ziyi Shou and Fangzhen Lin. 2023. Evaluate AMR graph similarity via self-supervised learning. In *ACL*.

Aviv Slobodkin, Leshem Choshen, and Omri Abend. 2022. Semantics-aware attention improves neural machine translation. In **SEM*.

Linfeng Song and Daniel Gildea. 2019. SemBleu: A robust metric for AMR parsing evaluation. In *ACL*.

Linfeng Song, Daniel Gildea, Yue Zhang, Zhiguo Wang, and Jinsong Su. 2019. Semantic Neural Machine Translation Using AMR. *Transactions of the ACL*, 7.

Radu Soricut and Daniel Marcu. 2003. Sentence level discourse parsing using syntactic and lexical information. In *NAACL*.

Elizabeth Spaulding, Gary Kazantsev, and Mark Dredze. 2023. Joint end-to-end semantic proto-role labeling. In *ACL*.

Elias Stengel-Eskin, Kenton Murray, Sheng Zhang, Aaron Steven White, and Benjamin Van Durme. 2021. Joint universal syntactic and semantic parsing. *Transactions of the ACL*, 9.

Elias Stengel-Eskin, Aaron Steven White, Sheng Zhang, and Benjamin Van Durme. 2020. Universal decompositional semantic parsing. In *ACL*.

Elior Sulem, Omri Abend, and Ari Rappoport. 2018a. Semantic structural evaluation for text simplification. In *NAACL*.

Elior Sulem, Omri Abend, and Ari Rappoport. 2018b. Simple and effective text simplification using semantic and neural methods. In *ACL*.

Mihai Surdeanu, Richard Johansson, Adam Meyers, Lluís Màrquez, and Joakim Nivre. 2008. The CoNLL 2008 shared task on joint parsing of syntactic and semantic dependencies. In *CoNLL*.

Zhixing Tan, Mingxuan Wang, Jun Xie, Yidong Chen, and Xiaodong Shi. 2018. Deep semantic role labeling with self-attention. In *AAAI*.

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. *Künstliche Intelligenz*, 35(3).

Rik van Noord. 2019. Neural boxer at the IWCS shared task on DRS parsing. In *IWCS*.

Rik van Noord, Lasha Abzianidze, Antonio Toral, and Johan Bos. 2018. Exploring neural methods for parsing discourse representation structures. *Transactions of the ACL*, 6.

Rik van Noord, Antonio Toral, and Johan Bos. 2020. Character-level representations improve DRS-based semantic parsing even in the age of BERT. In *EMNLP*.

Siddharth Vashishtha, Benjamin Van Durme, and Aaron Steven White. 2019. Fine-grained temporal relation extraction. In *ACL*.

Pavlo Vasylenko, Pere Lluís Huguet Cabot, Abelardo Carlos Martínez Lorenzo, and Roberto Navigli. 2023. Incorporating graph information in transformer-based AMR parsing. In *ACL Findings*.

Martin Verrev. 2023. Evaluation of semantic parsing frameworks for automated knowledge base construction. In *ISDA*.

David Vilares and Carlos Gómez-Rodríguez. 2018. A transition-based algorithm for unrestricted AMR parsing. In *NAACL*.

Adrienne Wang, Tom Kwiatkowski, and Luke Zettlemoyer. 2014. Morpho-syntactic lexical generalization for CCG semantic parsing. In *EMNLP*.

Chuan Wang, Nianwen Xue, and Sameer Pradhan. 2015. A transition-based algorithm for AMR parsing. In *NAACL*.

Chunliu Wang, Huiyuan Lai, Malvina Nissim, and Johan Bos. 2023. Pre-trained language-meaning models for multilingual parsing and generation. In *ACL Findings*.

Chunliu Wang, Rik van Noord, Arianna Bisazza, and Johan Bos. 2021a. Evaluating text generation from discourse representation structures. In *GEM*.

Xinyu Wang, Yong Jiang, Nguyen Bach, Tao Wang, Zhongqiang Huang, Fei Huang, and Kewei Tu. 2021b. Automated concatenation of embeddings for structured prediction. In *ACL-IJCNLP*.

Yizhong Wang, Sujian Li, and Houfeng Wang. 2017. A two-stage parsing method for text-level discourse analysis. In *ACL*.

Yizhong Wang, Sujian Li, and Jingfeng Yang. 2018. Toward fast and accurate neural discourse segmentation. In *EMNLP*.

Shira Wein and Nathan Schneider. 2024. Lost in translationese? reducing translation effect using Abstract Meaning Representation. In *EACL*.

Keenon Werling, Gabor Angeli, and Christopher D. Manning. 2015. Robust subgraph generation improves Abstract Meaning Representation parsing. In *ACL-IJCNLP*.

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 *EMNLP*.

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 *LREC*.

Yuk Wah Wong and Raymond Mooney. 2006. Learning for semantic parsing with statistical machine translation. In *NAACL*.

Jiacheng Xu, Zhe Gan, Yu Cheng, and Jingjing Liu. 2020. Discourse-aware neural extractive text summarization. In *ACL*.

Weiwen Xu, Huihui Zhang, Deng Cai, and Wai Lam. 2021. Dynamic semantic graph construction and reasoning for explainable multi-hop science question answering. In *ACL-IJCNLP Findings*.

Yuqing Yang, Qipeng Guo, Xiangkun Hu, Yue Zhang, Xipeng Qiu, and Zheng Zhang. 2023. An AMR-based link prediction approach for document-level event argument extraction. In *ACL*.

Nan Yu, Meishan Zhang, and Guohong Fu. 2018. Transition-based neural RST parsing with implicit syntax features. In *COLING*.

Longyin Zhang, Yuqing Xing, Fang Kong, Peifeng Li, and Guodong Zhou. 2020a. A top-down neural architecture towards text-level parsing of discourse rhetorical structure. In *ACL*.

Sheng Zhang, Xutai Ma, Rachel Rudinger, Kevin Duh, and Benjamin Van Durme. 2018. Cross-lingual decompositional semantic parsing. In *EMNLP*.

Shiyue Zhang and Mohit Bansal. 2021. Finding a balanced degree of automation for summary evaluation. In *EMNLP*.

Yu Zhang, Qingrong Xia, Shilin Zhou, Yong Jiang, Guohong Fu, and Min Zhang. 2022. Semantic role labeling as dependency parsing: Exploring latent tree structures inside arguments. In *COLING*.

Zhuosheng Zhang, Yuwei Wu, Hai Zhao, Zuchao Li, Shuailiang Zhang, Xi Zhou, and Xiang Zhou. 2020b. Semantics-aware bert for language understanding. In *AAAI*.

Wanjun Zhong, Jingjing Xu, Duyu Tang, Zenan Xu, Nan Duan, Ming Zhou, Jiahai Wang, and Jian Yin. 2020. Reasoning over semantic-level graph for fact checking. In *ACL*.

Jiawei Zhou, Tahira Naseem, Ramón Fernandez Astudillo, and Radu Florian. 2021. AMR parsing with action-pointer transformer. In *NAACL*.

Jie Zhou and Wei Xu. 2015. End-to-end learning of semantic role labeling using recurrent neural networks. In *ACL-IJCNLP*.

Zdeněk Žabokrtský, Daniel Zeman, and Magda Ševčíková. 2020. Sentence Meaning Representations Across Languages: What Can We Learn from Existing Frameworks? *Computational Linguistics*, 46.