

DRS at MRP 2020: Dressing up Discourse Representation Structures as Graphs

Lasha Abzianidze*
UiL OTS
Utrecht University
l.abzianidze@uu.nl

Johan Bos
CLCG
University of Groningen
johan.bos@rug.nl

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

Abstract

Discourse Representation Theory (DRT) is a formal account for representing the meaning of natural language discourse. Meaning in DRT is modeled via a Discourse Representation Structure (DRS), a meaning representation with a model-theoretic interpretation, which is usually depicted as nested boxes. In contrast, a directed labeled graph is a common data structure used to encode semantics of natural language texts. The paper describes the procedure of dressing up DRSs as directed labeled graphs to include DRT as a new framework in the 2020 shared task on Cross-Framework and Cross-Lingual Meaning Representation Parsing. Since one of the goals of the shared task is to encourage unified models for several semantic graph frameworks, the conversion procedure was biased towards making the DRT graph framework somewhat similar to other graph-based meaning representation frameworks.

1 Introduction

Graphs are a common data structure for representing meaning of natural language sentences or texts. Several shared tasks on semantic parsing have been organized, and the target meaning representations of the shared tasks were predominantly encoded as directed labeled graphs:¹ Semantic Dependency Graphs (Oepen et al., 2014, 2015), Abstract Meaning Representation (May, 2016; May and Priyadarshi, 2017), and Universal Conceptual Cognitive Annotation (Hershcovich et al., 2019). Some of these graphs are presented in Figure 1. Recently, Oepen et al. (2019) packaged several meaning representation graphs in a uniform graph

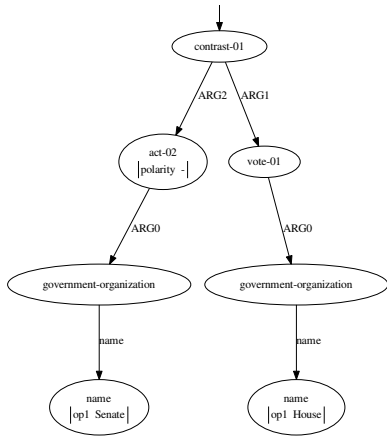
^{*}Part of the work was done while the author was at the University of Groningen.

¹Throughout the paper, we mean a directed labeled graph when simply talking about graphs, unless stated otherwise.

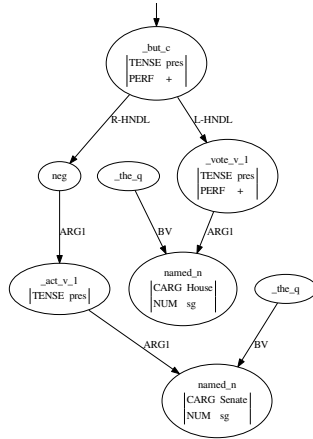
abstraction and serialization for cross-framework meaning representation parsing.

Parallel to these developments our point of departure is Discourse Representation Theory (DRT, Kamp and Reyle, 1993), a well-studied framework for studying formal semantics beyond sentences. Its meaning representation structures, Discourse Representation Structure (DRS), are directly translatable into formal logic. A sample DRS, in its traditional box format, is illustrated in Figure 2. We will discuss the DRS in more details in Section 2.

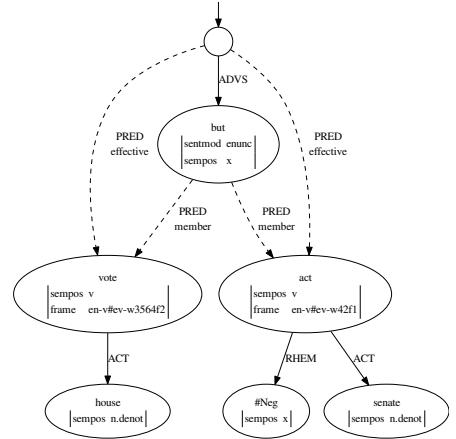
Obviously, DRSs are meaning representation structures, but they are different from the already mentioned graph-based meaning representations in two aspects. First, **DRSs are not inherently graphs**. A DRS is more like a formula of predicate logic which is further organized in sub-formulas and governed with additional operations that account for co-reference and presupposition. That’s why DRSs are usually not considered as graph-based meaning representations. For example, DRT was not among the frameworks of the shared task on cross-framework meaning representation parsing (MRP 2019, Oepen et al., 2019) since the meaning representations at the shared task were all uniformly formatted as graphs. Žabokrtský et al. (2020) excluded DRSs when surveying sentence meaning representations as they “limit [themselves] to meaning representations whose backbone structure can be described as a graph over words (possibly with added non-lexical nodes) [...]”. The second main contrast between DRSs and several of the graph-based meaning representations is that **DRSs are very different from syntactic structures**. DRSs have roots in formal semantics, and they are geared to account for negation, quantification, and semantic scope rather than for syntactic



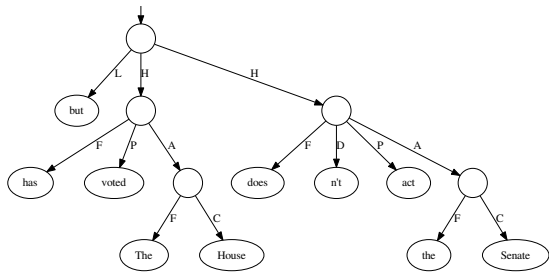
(a) AMR: Abstract Meaning Representation (Banarescu et al., 2013)



(b) EDS: Elementary Dependency Structures (Oepen and Lønning, 2006)



(c) PTG: Prague Tectogrammatical Graphs (Sgall et al., 1986; Hajič et al., 2012; Zeman and Hajič, 2020)



(d) UCCA: Universal Conceptual Cognitive Annotation (Abend and Rappoport, 2013)

b1 REF x1	0:3	b3 REF x2	24:27
b1 PRESUPPOSITION b2	0:3	b3 PRESUPPOSITION b4	24:27
b1 Name x1 "house"	4:9	b3 Name x2 "Senate"	28:34
b1 house "n.05" x1	4:9	b3 senate "n.01" x2	28:34
b2 REF e1	14:19	b4 REF t2	35:39
b2 REF t1	10:13	b4 EQU t2 "now"	35:39
b2 TPR t1 "now"	10:13	b4 time "n.08" t2	35:39
b2 Time e1 t1	10:13	b5 Time e2 t2	35:39
b2 time "n.08" t1	10:13	b4 NEGATION b5	39:42
b2 Agent e1 x1	14:19	b5 REF e2	43:46
b2 vote "v.01" e1	14:19	b5 Agent e2 x2	43:46
b2 CONTRAST b4	20:23	b5 act "v.01" e2	43:46

(e) DRS: Discourse Representation Structure in a clausal form (Kamp and Reyle, 1993; Bos et al., 2017; Abzianidze et al., 2017)

Figure 1: The meaning representation graphs (a-d) of the MRP 2020 frameworks for the sentence *The House has voted but the Senate doesn't act*. (e) is the DRS of Figure 2 in the clausal form, a suitable format for semantic parsing. The goal is to convert (e) into a graph somewhat similar to (a-d).

structures.²

Given that graphs are mainstream when it comes to representing meaning and semantically parsing wide-coverage natural language texts, it is important that DRSs are also convertible into graphs, and we refer to these structures as Discourse Representation Graphs (DRGs). This will make DRSs accessible for researchers that primarily focus on graph-based meaning representations and parsing: (a) already existing graph-based semantic parsing models can be re-used or tested on DRGs; and (b) the specific structure of DRGs, reflecting formal semantics of the meaning, will pose new challenges for graph representation learning.

In a nutshell, to embrace DRSs in the second edition of the shared task on cross-framework (and cross-lingual) meaning representation pars-

²For instance, this fact is another reason for excluding DRSs from the survey by Žabokrtský et al. (2020): “we do not include primarily logical representations which are too distant from sentence structures; this leaves out some prominent frameworks such as the Groningen Meaning Bank [...]”.

ing (MRP 2020; Oepen et al., 2020), we investigate the conversion of DRSs from clausal form (the form adapted to semantic parsing, see Figure 1e) into graphs. While doing so, our goal is to (i) make DRGs structurally as close as possible to the graphs of other frameworks in MRP 2020 (see Figure 1), and (ii) keeping redundant information in DRGs to a minimum to prevent graphs of extensive size and to avoid inflation of the evaluation score. Our efforts contribute to unified parsing models and evaluation tools across the frameworks. Hopefully, it will also save the time of participants by preventing them from developing a completely new parsing model for DRGs.

The rest of the paper is organized as follows. First, Section 2 briefly describes the building blocks of DRSs, and then Section 3 outlines already existing approaches of converting DRSs into graphs. In addition to the existing ones, Section 4 introduces several candidate graph-based encodings of DRSs. In Section 5, we compare several

Class	Type	symbol	SDRS signature	Examples
Entity	t	C	constant	now, house, senate
		r	discourse referent	$x_1, x_2, e_1, e_2, t_1, t_2$
	B	box label	b1, b2, b3, b4, b5	
Predicate	B	S	semantic role	Agent, Name, Time
		M	comparison relation	<, =
	C	concept	house.n.05, act.v.01	
Discourse	R	discourse relation	CONTRAST	
connective	O	DRS operator	NEGATION, PRESUPPOSITION	

Table 1: Classification of the DRS signature. Each element of the signature has a type symbol (in a bold font). t is for terms, which might be a constant or a discourse referent, while B stands for binary relations, which are semantic roles and comparison relations.

DRG formats on the computational feasibility of finding maximum common edge subgraph (MCES) because the computational feasibility is crucial for evaluating the meaning representation graphs against the gold standard. In the end, based on the findings of the MCES experiment and our desire for similarity with other graph-based frameworks, we select the specific DRG format that is included in MRP 2020.

2 Discourse Representation Structures

DRT is a framework that dates back to the early 1980s (Kamp, 1981; Heim, 1982). Since then, the framework has gone through several extensions and modifications to account for certain semantic or pragmatic phenomena. Throughout the paper we use DRSs that are derived from the Parallel Meaning Bank (PMB, Abzianidze et al., 2017). One such DRS is presented in Figure 2. The DRS signature is given in Table 1.

The PMB incorporates several extensions to DRSs. On a micro level, the extensions aim to make DRSs language-neutral by disambiguating non-logical symbols with WordNet (Miller, 1995) synsets and VerbNet (Bonial et al., 2011) roles, where the VerbNet roles are used in combination with neo-Davidsonian event semantics (Parsons, 1990). On a macro level, presuppositions are modeled and explicitly represented following Van der Sandt (1992) and Projective DRT (Venhuizen et al., 2013) while discourse is analyzed following Segmented DRT (Asher and Lascarides, 2003) and flattened by treating discourse relations and DRS operators in a unified way. Due to these extensions,

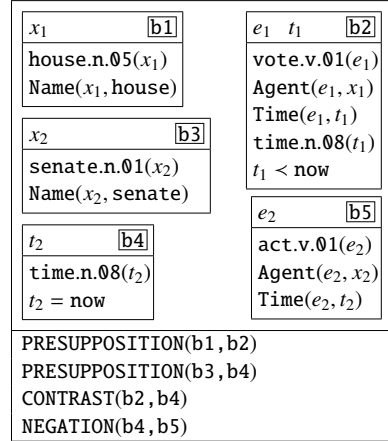


Figure 2: A flat visualization of a box-formatted DRS for the sentence *The House has voted but the Senate doesn't act.*

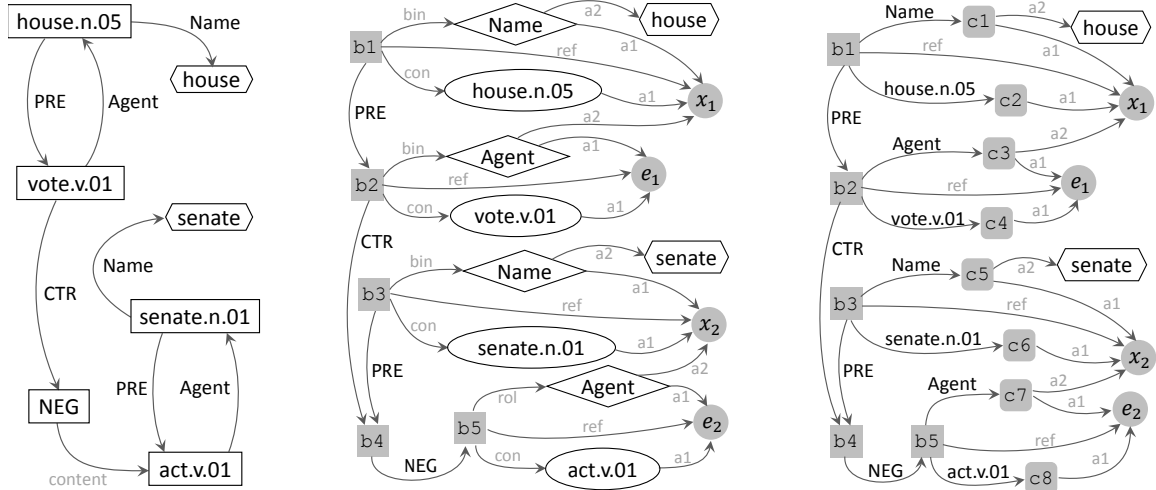
all boxes are labeled with identifiers.

Let's decipher what the DRS in Figure 2 is expressing. It consists of two parts: a set of boxes and a set of *discourse connectives* applied to box labels (i.e., identifiers). Boxes can be seen as sub-formulas whose separation is relevant for fine-grained semantics. Each box includes a (possibly empty) set of *discourse referents* stacked on a (possibly empty) set of *conditions*. The example sentence contains two clauses, corresponding to boxes b2 and b4, that are related with each other via the CONTRAST discourse relation. Both b2 and b4 presuppose the existence of entities x_1 (for *the House*) and x_2 (for *the Senate*), which are further characterized with concepts (using WordNet synsets) and the naming semantic role. The presuppositions are put in separate boxes labeled with b1 and b3. The presupposition relations are explicitly stated with the binary PRESUPPOSITION DRS operator. Since we use a flat visualization of DRSs, b5, which is negated and nested in b4 (expressed by NEGATION(b4, b5)), is depicted outside b4. In addition to modeling verb argument structure via neo-Davidsonian event semantics and semantic roles, the DRS also contains information about tense.³

3 Related Work

There have been several approaches to represent DRSs as graphs. These representations are put side-by-side in Figure 3.

³Note that t_2 is in b4 because it has to be out of the scope of negation: there is a time t_2 , and it is not the case that at t_2 *the Senate acts*.



(a) The augmented graph of Power (1999) corresponding to the simplified sample DRS. The graph is a felicitous extension of Power’s original proposal over DRSs with presuppositions and discourse relations.

(b) The BB* encoding largely follows Basile and Bos (2013) and incorporates several additional simplifications. The encoding is node-centric. B and C are encoded as labeled nodes while R , O and argument positions (A) as labeled edges. Only B and r are unlabeled nodes.

(c) L18 is the edge-centric encoding by Liu et al. (2018). B and C are represented as unlabeled nodes with B - and C -labeled incoming edges. R , O and argument positions (A) are encoded as labeled edges. Unlabeled nodes are introduced not only by B and r but also by B and C .

Figure 3: Contrasting the existing graph representations of DRSs. The graphs encode the sample DRS from Figure 2. For brevity, the tense information is omitted from the DRS. Unlabeled nodes have a gray background. The shapes of nodes are not part of the graphs but simply help with reading to distinguish the types of symbols.

The work by Power (1999) doesn’t aim to convert DRSs into graphs as such, but it proposes to augment object-oriented knowledge representation (OOKR) graphs with additional scope information to establish correspondence with DRSs. Although the correspondence is incomplete, e.g., some OOKR graphs might have no corresponding DRS. The augmentation of Power (1999) doesn’t cover DRSs with discourse relations, presuppositions (e.g., b1 to b2 in Figure 2) or with an embedded box that contains base and complex conditions (like b4 in Figure 2). Nevertheless, for demonstration purposes, we still present Power (1999)’s augmented graph for a felicitous, simplified DRS of Figure 2.

Basile and Bos (2013) proposed converting DRSs into graphs, calling them Discourse Representation Graphs (DRGs). Their goal was to facilitate word-level alignment between surface forms and the corresponding DRSs to generate texts from DRSs. The graph encoding, with several simplifications, is exemplified in Figure 3b.⁴ The simplifications decrease the number of nodes and out-

⁴Originally Basile and Bos (2013) use more labels for edges that expresses type-specific information of nodes. For example, they use different edge labels to distinguish the first argument position of B from the only argument position of C while in the paper we use the same label for both. Basile and Bos (2013) also encodes O as a reified node that introduces

of-signature labels in the graph. The encoding can be seen as node-centric since the most frequent signature symbols, namely the symbols of type B and C , are modeled as labeled nodes. Argument positions (A) of binary predicates are distinguished via edge labels. We call this DRG format BB*.

To evaluate the output of their DRS parser, Liu et al. (2018) converted DRSs into graphs, demonstrated in Figure 3c. This graph encoding, in contrast to BB*, is edge-centric as the symbols of type B and C are used as edge labels. Moreover, compared to BB*, the encoding contains more unlabeled nodes since B and C are also modeled with reified nodes. We call Liu et al. (2018)’s encoding L18.

Interestingly, in contrast to the proposed graph encodings of DRS, van Noord et al. (2018a) refused to convert DRSs into graphs and instead used so-called *clausal form* of DRSs (see Figure 1e). The clauses in clausal form are triples, e.g., $\langle b4, \text{NEGATION}, b5 \rangle$, or quadruples, e.g., $\langle b2, \text{Agent}, e1, x1 \rangle$, where the quadruples are hyper-edges and fall out of the scope of standard graph encodings. The official evaluation of the shared task on DRS parsing (Abzianidze et al.,

two edges $b4 \xrightarrow{\text{unary}} \text{NEG} \xrightarrow{\text{scope}} b5$. Instead, we simply model O with a single edge $b4 \xrightarrow{\text{NEG}} b5$.

2019) was also based on clausal form of DRSs.

4 More Graph-based Encodings of DRS

As illustrated in the previous section, there is no agreement on how DRSs should be converted into graphs (or whether they should be converted at all). The range of graph encodings in Figure 3 presents anything but an exhaustive list. Some encoding can even be further refined and compressed without affecting the readability or expressiveness. For instance, as explained in footnote 4, BB^* represents a refined version of DRGs proposed by Basile and Bos (2013). L18 can also be further compressed by discarding reified concept nodes and their outgoing a1 edges, e.g., replacing $b5 \xrightarrow{\text{act.v.}\theta 1} c8 \xrightarrow{a1} e_2$ with $b5 \xrightarrow{\text{act.v.}\theta 1} e_2$. We will use L18* to refer to the DRGs refined in such a way.

In general, the choices in which DRG formats might differ are several. Here we will discuss some of them, namely (see also Table 2):

- (A) Expressing Argument positions of B via forking and labeled edges ($\textcircled{B} \xrightarrow{1} 2$, like BB^*) or solely via graph configuration ($1 \rightarrow \textcircled{B} \rightarrow 2$, without labeled edges), e.g., encoding $\text{Agent}(e_1, x_1)$ as $e_1 \rightarrow \text{Agent} \rightarrow x_1$;
- (B) Representing Binary predicates as labeled nodes (\textcircled{B} , like BB^*) or unlabeled nodes with B -labeled edges ($\xrightarrow{B} \circ$, like L18);
- (C) Encoding Concepts as labeled nodes (\textcircled{C} , like BB^*), unlabeled nodes with incoming C -labeled edges ($\xrightarrow{C} \circ$, like L18), labeled edges ($\xrightarrow{C} \bullet$, like L18*), or as a label on an r node ($c \textcircled{\bullet}$, which will be discussed further);
- (I) Expressing box membership explicitly (Exp) or implicitly (Imp). Whether a node (corresponding to B , C , or r) is *In* a particular B , can be depicted via an explicit connecting edge or implicitly via graph configuration.

Here we would like to elaborate more on (I). The box membership in DRT directly accounts for a semantic scope. Like discourse referents, conditions are also members of boxes. So, we also need to express the box membership of condition predicates in the graphs. All the encodings in Figure 3 explicitly express box membership. For instance, $\text{Agent}(e_1, x_1)$ belonging to b2 is expressed via connecting b2 to the Agent node (see Figure 3b) or via the outgoing Agent edge from b2 to c3. Explicating all box memberships via labeled edges

DRG encoding	Args	B	C	In-box
BB^* +typed edges	$\textcircled{B} \xrightarrow{1} 2$	\textcircled{B}	\textcircled{C}	Exp
$A_a^< B^\circ C^\circ$	$\textcircled{B} \xrightarrow{1} 2$	\textcircled{B}	\textcircled{C}	Exp
$A_a^< C^\downarrow B^\downarrow$ (L18)	$\textcircled{B} \xrightarrow{1} 2$	$\xrightarrow{B} \circ$	$\xrightarrow{C} \circ$	Exp
$A_a^< B^\downarrow C^\downarrow$ (L18*)	$\textcircled{B} \xrightarrow{1} 2$	$\xrightarrow{B} \circ$	$\xrightarrow{C} \bullet$	Exp
$A_a^< B^\circ C^\downarrow$	$\textcircled{B} \xrightarrow{1} 2$	\textcircled{B}	$\xrightarrow{C} \bullet$	Exp
$A^\downarrow B^\circ C^\downarrow$	$1 \rightarrow \textcircled{B} \rightarrow 2$	\textcircled{B}	$\xrightarrow{C} \bullet$	Exp
$A_a^< B^\downarrow C$	$\textcircled{B} \xrightarrow{1} 2$	$\xrightarrow{B} \circ$	$c \textcircled{\bullet}$	Exp
$A_a^< B^\circ C$	$\textcircled{B} \xrightarrow{1} 2$	\textcircled{B}	$c \textcircled{\bullet}$	Exp
$A^\downarrow B^\circ C$	$1 \rightarrow \textcircled{B} \rightarrow 2$	\textcircled{B}	$c \textcircled{\bullet}$	Exp
$A^\downarrow B^\circ C I$	$1 \rightarrow \textcircled{B} \rightarrow 2$	\textcircled{B}	$c \textcircled{\bullet}$	Im-a1
$A^\downarrow_a B^\circ C I$	$1 \rightarrow \textcircled{B} \rightarrow 2$	\textcircled{B}	$c \textcircled{\bullet}$	Im-a1

Table 2: Several combinations of the choices in DRG design. The choices concern representation of argument positions, B symbols, C symbols, and in-box relations. The names of encodings visually follow the combinations of the choices.

increases the graphs in size. To prevent this, one can make box membership of certain predicates or their arguments implicit but at the same time easily and unambiguously recoverable from the graphs. For example, if we assume that directionality of arrows carries the in-box inheritance and consider the case when argument positions are configurationally encoded ($1 \rightarrow \textcircled{B} \rightarrow 2$), then there is no need to explicate the in-box relation for Name in $x_1 \rightarrow \text{Name} \rightarrow \text{house}$ whenever the Name condition and x_1 are in the same box.⁵ We dub such an implication of box membership of B from the first argument as ‘Im-a1’.

Table 2 lists several DRG formats based on combinations of how argument positions, binary predicates, concepts, and in-box relations are represented in a graph. While modeling the argument position, $1 \rightarrow \textcircled{B} \rightarrow 2$ is preferred over $\textcircled{B} \xrightarrow{1} 2$ from a theoretical point of view because a1 and a2 labels are not part of the DRS signature; They are ad-hoc ingredients only helping with distinguishing argument positions. When it comes to modeling concepts, as we already discussed, $\xrightarrow{C} \bullet$ leads to more economic graphs than $\xrightarrow{C} \circ$.

In the PMB annotation, for almost any discourse referent, there exists the most specific concept among the concepts applied to it. For example, a discourse referent might have only two con-

⁵Remember that a discourse referent is considered to be *in a box* if it is introduced in the top row of the box.

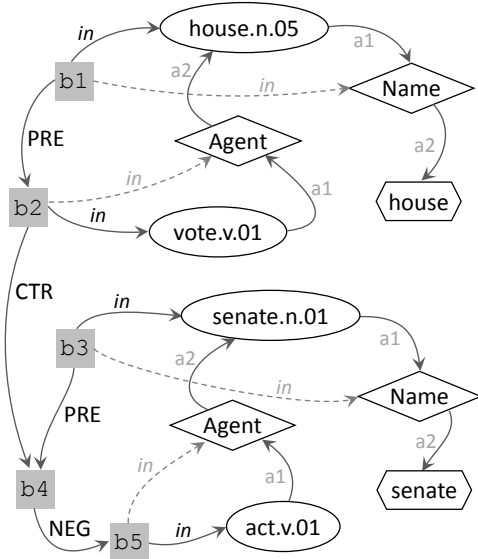


Figure 4: $A \uparrow_a B^\circ C$ and $A \uparrow B^\circ C I$ encodings. The r nodes are labeled with C oncepts and connected to the boxes via in -edges. Dashed in -edges for B inary predicates and the aN edge labels are recoverable. $A \uparrow B^\circ C I$ is obtained by ignoring dashed edges and gray edge labels. Unlabeled nodes are colored in grey where their labels merely serve to match graph components across the different visualizations.

cepts, `male.n.02` and `person.n.01`, applied to it, but among these concepts there exists the most specific concept, namely `male.n.02`, as `male.n.02` is a hyponym of `person.n.01` according to WordNet. The $c \odot$ choice exploits this annotation property of concepts in the PMB and labels the node of a discourse referent with the corresponding most specific concept. This type of encoding of C is shown in Figure 4.

Figure 4 also depicts $A \uparrow B^\circ C I$ DRG encoding with implicit box membership of B . Though all the box membership edges of B are made implicit in the encoding example, this is not the case in general. For example, attributive and predicative adjectives usually introduce $\langle b1, \text{Attribute}, x1, s1 \rangle$ clause, where $x1$ is the attributed entity which is not necessarily introduced in the same $b1$ box as the attributing state $s1$. Another example is a construction with a locative preposition and a definite noun phrase, e.g., *hid a parcel under the bed*, whose DRS contains the following fragment:

```

b2 REF e1          b2 Location e1 x3
b2 hide "v.01" e1  b2 SZP x2 x3
b2 REF x1          b3 REF x2
b2 parcel "n.01" x1 b3 bed "n.01" x2
b2 Patient e1 x1

```

where the binary relation SZP (spatial above) is in

DRS parsers	DRS	BB^*	$A_d^{\circ} B^\circ C^b$	$A_d^{\circ} B^\circ C^i$	$A_d^{\circ} B^\circ C^g$	$A \uparrow B^\circ C I$
		DRGs				
chLSTM \downarrow	64.6	79.6	74.3	77.7	77.9	78.2
Boxer \downarrow	78.2	89.5	86.8	87.5	87.6	87.7
chLSTM \uparrow	84.3	92.3	88.4	90.9	90.9	91.1
Boxer \uparrow	87.2	94.2	92.3	92.9	92.9	93.0

Table 3: Macro F-scores of the models when their output is treated as DRS or DRG. F-score for DRS is computed with Counter while for DRG with mtool.

a different box than its first argument.

As we have shown, there are at least a dozen ways to dress up DRSs as graphs. Some of the DRG formats are verbose, some can employ default rules to ignore certain redundancies, some require out-of-signature symbols, and some prefer labeled edges over labeled nodes. There isn't enough space to illustrate the graphs listed in Table 2, but each of the mentioned encoding choices is demonstrated by at least one of the graphs from Figure 3 and Figure 4.

5 Matching & Evaluating DRGs

In graph-based semantic parsing, system outputs are conventionally evaluated against the gold standard graphs by finding the maximum common edge subgraph (MCES) for each pair of produced and gold graphs, and then calculating macro-average F-score (Oepen et al., 2019). In general, the MCES problem is NP-complete, and finding the maximum subgraph *shared* between two relatively large graphs is sometimes computationally infeasible. In this section, we experiment on how computationally expensive is the MCES problem for each DRG design.

5.1 Data & Tools

We run the experiments on the output of existing DRS parsers. Four distinct parsing models are selected to achieve diversity in the system output graphs. Two of the parsers are end-to-end character-based LSTM models from van Noord et al. (2018b): one is their best model (chLSTM \uparrow) while another one is trained on fewer data on purpose to have mediocre performance (chLSTM \downarrow). Another two parsers are based on the semantic

DRS parser	\times	$\mathcal{G}\%$	BB^*	$A_a^<B^oC^o$	$A_a^<C^bB^d$	$A_a^<B^bC^i$	$A_a^<B^oC^i$	$A_a^<B^oC^i$	$A_a^<B^oC^i$	$A_a^<B^b\mathcal{G}$	$A_a^<B^o\mathcal{G}$	$A_a^<B^o\mathcal{G}$	$A_a^<B^o\mathcal{G}$	$A_a^<B^o\mathcal{G}$	$A_a^<B^o\mathcal{G}$
chLSTM [↓]	13	11	27.2	30.4	28.0	12.5	15.0	18.1	15.0	9.3	10.3	12.8	10.3	8.1	7.0
Boxer [↓]	7	9	9.2	10.3	9.7	3.6	4.8	6.7	4.8	2.0	2.2	3.5	2.2	1.0	0.1
chLSTM [↑]	4	3	8.3	9.8	13.1	4.3	5.2	6.6	5.2	3.0	3.2	4.4	3.2	2.2	1.6
Boxer [↑]	7	3	4.4	5.1	4.8	1.3	2.7	3.3	2.7	1.0	1.0	1.8	1.0	0.9	0.3

Table 4: The percentage of approximate (i.e., non-exact) matches w.r.t. the total non-null DRGs. Lower numbers are better as more graph matches corresponding to MCES are found. The total number of DRGs is 885. While converting DRGs into DRGs, \times -number of DRGs become null due to ill-formed DRGs and are excluded during calculating the percentages. Encoding with \mathcal{G} additionally renders $\mathcal{G}\%$ -number of DRGs untranslatable.

parser Boxer (Bos, 2008), which is used in the PMB to *pack* all annotations layers into DRS boxes. Boxer[↓] is Boxer based on the NLP tools of the PMB pipeline⁶, on the other hand, Boxer[↑] is Boxer employing annotation layers output by MaChAMP (van der Goot et al., 2020). As the names suggest, Boxer[↑] is a better model than Boxer[↓]. The output DRGs are obtained by parsing the development set (885 documents) of the PMB v3.0.0.⁷ Evaluation of the models based on the DRGs of the dev set is given in Table 3. DRGs are scored with Counter (van Noord et al., 2018a), the clause matching tool for DRGs in clausal form.⁸

For MCES-based matching of DRGs, we use mtool⁹, the Swiss Army Knife for Graph-Based Meaning Representation. Based on the graph configurations, mtool schedules potential node-to-node mappings between two graphs. This information is used to initialize *promising* node-to-node mappings that might lead to finding the MCES early. mtool is the official scorer in both the MRP 2019 and MRP 2020 shared tasks.

All types of graph encodings employed in the experiments are obtained with the DRS2Graph tool.¹⁰ This new converter from clause-based DRGs to labeled directed graphs is one of the contributions of the paper.

5.2 Results & Analysis

The results of finding MCES between the system generated and converted DRGs and reference

DRGs are provided in Table 4. The reference DRGs were obtained by converting the gold standard DRS of the PMB 3.0.0 development set. We run experiments with 13 DRG formats. All 885 DRGs were converted in each DRG format without problems. In principle, the encodings with the \mathcal{G} choice are lossy, however, they were successfully applied to the gold DRGs. Several parser-produced DRGs were not converted according to the \mathcal{G} choice since the parsers assert the inconsistent concepts for discourse referents. For example, Boxer[↑] produced a DRS with *measure.n.02* and *book.n.01* applied to the same discourse referent. Since these senses are not in hyponymy/hypernymy relation, the DRS didn’t meet the requirement from \mathcal{G} and was one of the three DRGs of Boxer[↑] that couldn’t be dressed up as \mathcal{G} -based graphs.¹¹

Table 4 shows the computational (in)feasibility of the MCES problem across the combinations of parsing models and graph encodings (using the mtool implementation with default limits on its search space). Given that models are sorted according to their performance in ascending order from top to bottom, the table shows that for relatively distinct graphs it can be difficult to guarantee the MCES solution.¹² But things are not so straightforward as chLSTM[↑] outperforms Boxer[↓] but finding MCES for Boxer[↓] is easier for 10 encodings out of 13. This can be explained by the fact that gold DRGs are obtained from Boxer[↓] while taking into account added human annotations. Given this, it is

¹¹From 6620 gold DRGs of the PMB 3.0.0 training part, only 16 (0.24%) DRGs didn’t satisfy the constraint of \mathcal{G} .

¹²When exhausting its search space limits, mtool falls back to an anytime strategy, returning the best overall match found up to that point. This match will often correspond to the MCES, but there is no correctness guarantee in this mode.

⁶<https://pmb.let.rug.nl/software.php>

⁷<https://pmb.let.rug.nl/data.php>

⁸https://github.com/RikVN/DRS_parsing

⁹<https://github.com/cfmrp/mtool>

¹⁰<https://github.com/kovvalsky/DRS2Graph>

expected that gold and Boxer[↓]'s DRSs have in common substantial chunks of boxes, and this sharing is transferred on the DRGs too.

Interestingly, the encodings BB^* (Basile and Bos, 2013) and $A_a^{\leftarrow}B^{\downarrow}C^{\downarrow}$ (Liu et al., 2018) are one of the most inefficient encodings across all the models. For instance, non-exact (i.e., approximate) MCES was found for 237 DRG pairs out of 872 for chLSTM[↓] and BB^* encoding. For other encodings the ratio of approximate matches halves.

Among the encodings with the C^{\downarrow} choice, $A_a^{\leftarrow}B^{\downarrow}C^{\downarrow}$ appears to provide most computationally *friendly* graphs. Every encoding with C^{\downarrow} becomes even better when C^{\downarrow} is replaced with \mathcal{G} . This is because \mathcal{G} brings at least a 16% reduction in the number of edges and increases the number of labeled nodes. The latter apparently helps `mtool` to get better initializations for node mappings.

$A_a^{\leftarrow}B^{\downarrow}\mathcal{G}$ is the best among \mathcal{G} -featured encodings with explicit box membership. It doesn't improve further when changing its encoding choices, including switching to A^{\downarrow} . The results show that A_a^{\leftarrow} is consistently better than A^{\downarrow} . Even when they are combined, A^{\downarrow} adds no value to A_a^{\leftarrow} . However, the advantage of A^{\downarrow} over A_a^{\leftarrow} is that it configurationally distinguishes argument positions and there is no need for out-of-signature labels. Moreover, A^{\downarrow} invites the intuitive inheritance property about in-box relation (see (I) discussed in Section 4). When incorporating the implicit in-box relation with A^{\downarrow} , the combination $A^{\downarrow}B^{\circ}\mathcal{G}I$ yields a substantial decrease in the number of approximate matches. This is explained by the fact that the number of edges decreases by at least 23%. Adding the out-of-signature edge labels for marking argument positions further improves the encoding.

Differences between F-scores calculated over DRS (with Counter) and DRGs (with `mtool`) are significant (see Table 3). The gap between low- and high-performing model is greater than 10% and 5%, respectively. The DRS-based score is more strict than the DRG-based one because DRSs are evaluated in the clausal form, where some DRSs conditions (e.g., built with B) are modeled via quadruples, i.e., hyper-edges. In DRGs, the hyper-edges are represented by multiple triples ($\langle \text{nodeID}, \text{edgeLabel}, \text{nodeID} \rangle$ or $\langle \text{nodeID}, \text{label}, \text{labelValue} \rangle$), and this additionally rewards the models when they get parts of hyper-edges correctly.

6 Conclusion

There have been several approaches that encoded DRSs as graphs (surveyed in Section 3), but their objectives were to transform DRSs in a suitable format for particular applications rather than exploring and comparing different types of DRG encodings. This paper fills this gap. We have systematically characterized a dozen of DRG encodings and contrasted them with each other, and compared them to the DRS clausal form from an evaluation perspective.

We opt for the $A^{\downarrow}B^{\circ}\mathcal{G}I$ DRG encoding (see Figure 4) to represent DRSs at the MRP 2020 shared task. Despite the encoding being lossy, it represents an excellent trade-off due to the advantages it brings: (a) the encoding has at least 23% fewer edges than other encodings, which makes the DRGs more compact and easier to read; (b) given that scope information inflates DRSs, learning relatively compact DRGs seems a good starting point for the shared task; (c) only less than 0.25% DRSs are lost when applying the encoding; (d) it doesn't employ the out-of-signature labels `a1` and `a2`; (e) for the DRGs obtained from the average-performing DRS parsers, the evaluation tool can find exact maximal matches for at least 98.4% of DRG pairs.

When abstracting from the reification of the roles as nodes, the chosen DRG encoding and the graphs of other frameworks in MRP 2020 have abstractly parallel graph topologies for linguistically parallel predicate-argument structures.

Acknowledgments

We thank Rik van Noord for providing us with outputs of the DRS parsers. We acknowledge access to the Peregrine HPC cluster provided by the CIT of the University of Groningen, and to the NLPL infrastructure provided by Sigma2 in Norway. The first two authors were supported by the NWO-VICI grant (288-89-003). The first author was additionally supported by the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement No. 742204).

References

- Omri Abend and Ari Rappoport. 2013. [UCCA: A semantics-based grammatical annotation scheme](#). In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013) – Long Papers*, pages 1–12, Potsdam, Germany. Association for Computational Linguistics.
- Lasha Abzianidze, Johannes Bjerva, Kilian Evang, Hessel Haagsma, Rik van Noord, Pierre Ludmann, Duc-Duy Nguyen, and Johan Bos. 2017. [The Parallel Meaning Bank: Towards a multilingual corpus of translations annotated with compositional meaning representations](#). In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 242–247, Valencia, Spain. Association for Computational Linguistics.
- Lasha Abzianidze, Rik van Noord, Hessel Haagsma, and Johan Bos. 2019. [The first shared task on discourse representation structure parsing](#). In *Proceedings of the IWCS Shared Task on Semantic Parsing*, Gothenburg, Sweden. Association for Computational Linguistics.
- N. Asher and A. Lascarides. 2003. *Logics of conversation*. Studies in natural language processing. Cambridge University Press.
- Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. [Abstract Meaning Representation for sembanking](#). In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria. Association for Computational Linguistics.
- Valerio Basile and Johan Bos. 2013. [Aligning formal meaning representations with surface strings for wide-coverage text generation](#). In *Proceedings of the 14th European Workshop on Natural Language Generation*, pages 1–9, Sofia, Bulgaria. Association for Computational Linguistics.
- Claire Bonial, William J. Corvey, Martha Palmer, Volha Petukhova, and Harry Bunt. 2011. A hierarchical unification of LIRICS and VerbNet semantic roles. In *Proceedings of the 5th IEEE International Conference on Semantic Computing (ICSC 2011)*, pages 483–489.
- Johan Bos. 2008. [Wide-Coverage Semantic Analysis with Boxer](#). In Johan Bos and Rodolfo Delmonte, editors, *Semantics in Text Processing. STEP 2008 Conference Proceedings*, volume 1 of *Research in Computational Semantics*, pages 277–286. College Publications.
- Johan Bos, Valerio Basile, Kilian Evang, Noortje Venhuizen, and Johannes Bjerva. 2017. [The Groningen Meaning Bank](#). In Nancy Ide and James Pustejovsky, editors, *Handbook of Linguistic Annotation*. Springer Netherlands.
- Rob van der Goot, Ahmet Üstün, Alan Ramponi, and Barbara Plank. 2020. [Massive choice, ample tasks \(machamp\): A toolkit for multi-task learning in nlp](#).
- Jan Hajič, Eva Hajičová, Jarmila Panevová, Petr Sgall, Ondřej Bojar, Silvie Cinková, Eva Fučíková, Marie Mikulová, Petr Pajas, Jan Popelka, Jiří Semecký, Jana Šindlerová, Jan Štěpánek, Josef Toman, Zdeňka Urešová, and Zdeněk Žabokrtský. 2012. [Announcing Prague Czech-English Dependency Treebank 2.0](#). In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)*, pages 3153–3160, Istanbul, Turkey. European Language Resources Association (ELRA).
- Irene Heim. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. thesis, University of Massachusetts, Amherst.
- Daniel Hershcovich, Zohar Aizenbud, Leshem Choshen, Elior Sulem, Ari Rappoport, and Omri Abend. 2019. [SemEval-2019 task 1: Cross-lingual semantic parsing with UCCA](#). In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 1–10, Minneapolis, Minnesota, USA. Association for Computational Linguistics.
- Hans Kamp. 1981. A theory of truth and semantic representation. In J. A. G. Groenendijk, T. M. V. Janssen, and M. B. J. Stokhof, editors, *Formal Methods in the Study of Language*, volume 1, pages 277–322. Mathematisch Centrum, Amsterdam.
- Hans Kamp and Uwe Reyle. 1993. *From Discourse to Logic; An Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and DRT*. Kluwer, Dordrecht.
- Jiangming Liu, Shay B. Cohen, and Mirella Lapata. 2018. [Discourse representation structure parsing](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 429–439, Melbourne, Australia. Association for Computational Linguistics.
- Jonathan May. 2016. [SemEval-2016 task 8: Meaning representation parsing](#). In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 1063–1073, San Diego, California. Association for Computational Linguistics.
- Jonathan May and Jay Priyadarshi. 2017. [SemEval-2017 task 9: Abstract Meaning Representation parsing and generation](#). In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 536–545, Vancouver, Canada. Association for Computational Linguistics.
- George A Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.
- Rik van Noord, Lasha Abzianidze, Hessel Haagsma, and Johan Bos. 2018a. [Evaluating scoped meaning representations](#). In *Proceedings of the Eleventh*

- International Conference on Language Resources and Evaluation (LREC-2018)*, Miyazaki, Japan. European Languages Resources Association (ELRA).
- Rik van Noord, Lasha Abzianidze, Antonio Toral, and Johan Bos. 2018b. [Exploring neural methods for parsing discourse representation structures](#). *Transactions of the Association for Computational Linguistics*, 6:619–633.
- Stephan Oepen, Omri Abend, Lasha Abzianidze, Johan Bos, Jan Hajič, Daniel Hershcovich, Bin Li, Tim O’Gorman, Nianwen Xue, and Daniel Zeman. 2020. [MRP 2020: The Second Shared Task on Cross-framework and Cross-Lingual Meaning Representation Parsing](#). In *Proceedings of the CoNLL 2020 Shared Task: Cross-Framework Meaning Representation Parsing*, pages 1–22, Online.
- Stephan Oepen, Omri Abend, Jan Hajič, Daniel Hershcovich, Marco Kuhlmann, Tim O’Gorman, Nianwen Xue, Jayeol Chun, Milan Straka, and Zdeňka Urešová. 2019. [MRP 2019: Cross-framework Meaning Representation Parsing](#). In *Proceedings of the Shared Task on Cross-Framework Meaning Representation Parsing at the 2019 Conference on Computational Natural Language Learning*, pages 1–27, Hong Kong, China.
- 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 *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 915–926, Denver, Colorado. Association for Computational Linguistics.
- 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 *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 63–72, Dublin, Ireland. Association for Computational Linguistics.
- Stephan Oepen and Jan Tore Lønning. 2006. [Discriminant-based MRS banking](#). In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC’06)*, Genoa, Italy. European Language Resources Association (ELRA).
- Terence Parsons. 1990. *Events in the Semantics of English: A Study in Subatomic Semantics*. MIT Press.
- Richard Power. 1999. [Controlling logical scope in text generation](#). In *Proceedings of the 7th European Workshop on Natural Language Generation (EWNLG’99)*, pages 1–9, Toulouse.
- Rob A. Van der Sandt. 1992. [Presupposition projection as anaphora resolution](#). *Journal of Semantics*, 9(4):333–377.
- Petr Sgall, Eva Hajičová, and Jarmila Panevová. 1986. *The Meaning of the Sentence and Its Semantic and Pragmatic Aspects*. Academia/Reidel Publishing Company, Prague, Czech Republic/Dordrecht, Netherlands.
- Noortje J. Venhuizen, Johan Bos, and Harm Brouwer. 2013. [Parsimonious semantic representations with projection pointers](#). In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013) – Long Papers*, pages 252–263, Potsdam, Germany. Association for Computational Linguistics.
- Daniel Zeman and Jan Hajič. 2020. [FGD at MRP 2020: Prague Tectogrammatical Graphs](#). In *Proceedings of the CoNLL 2020 Shared Task: Cross-Framework Meaning Representation Parsing*, pages 33–39, Online.
- Zdeněk Žabokrtský, Daniel Zeman, and Magda Ševčíková. 2020. [Sentence meaning representations across languages: What can we learn from existing frameworks?](#) *Computational Linguistics*, 0(0):605–665.