Abstract

We introduce an open-domain neural semantic parser which generates formal meaning representations in the style of Discourse Representation Theory (DRT; Kamp and Reyle 1993). We propose a method which transforms Discourse Representation Structures (DRSs) to trees and develop a structure-aware model which decomposes the decoding process into three stages: basic DRS structure prediction, condition prediction (i.e., predicates and relations), and referent prediction (i.e., variables). Experimental results on the Groningen Meaning Bank (GMB) show that our model outperforms competitive baselines by a wide margin.

1 Introduction

Semantic parsing is the task of mapping natural language to machine interpretable meaning representations. A variety of meaning representations have been adopted over the years ranging from functional query language (FunQL; Kate et al. 2005) to dependency-based compositional semantics (λ-DCS; Liang et al. 2011), lambda calculus (Zettlemoyer and Collins, 2005), abstract meaning representations (Banarescu et al., 2013), and minimal recursion semantics (Copestake et al., 2005).

Existing semantic parsers are for the most part data-driven using annotated examples consisting of utterances and their meaning representations (Zelle and Mooney, 1996; Wong and Mooney, 2006; Zettlemoyer and Collins, 2005). The successful application of encoder-decoder models (Sutskever et al., 2014; Bahdanau et al., 2015) to a variety of NLP tasks has provided strong impetus to treat semantic parsing as a sequence transduction problem where an utterance is mapped to a target meaning representation in string format (Dong and Lapata, 2016; Jia and Liang, 2016; Kočiský et al., 2016). The fact that meaning representations do not naturally conform to a linear ordering has also prompted efforts to develop recurrent neural network architectures tailored to tree or graph-structured decoding (Dong and Lapata, 2016; Cheng et al., 2017; Yin and Neubig, 2017; Alvarez-Melis and Jaakkola, 2017; Rabinovich et al., 2017; Buys and Blunsom, 2017).

Most previous work focuses on building semantic parsers for question answering tasks, such as querying a database to retrieve an answer (Zelle and Mooney, 1996; Cheng et al., 2017), or conversing with a flight booking system (Dahl et al., 1994). As a result, parsers trained on query-based datasets work on restricted domains (e.g., restaurants, meetings; Wang et al. 2015), with limited vocabularies, exhibiting limited compositionality, and a small range of syntactic and semantic constructions. In this work, we focus on open-domain semantic parsing and develop a general-purpose system which generates formal meaning representations in the style of Discourse Representation Theory (DRT; Kamp and Reyle 1993).

DRT is a popular theory of meaning representation designed to account for a variety of linguistic phenomena, including the interpretation of pronouns and temporal expressions within and across sentences. Advantageously, it supports meaning representations for entire texts rather than isolated sentences which in turn can be translated into first-order logic. The Groningen Meaning Bank (GMB; Bos et al. 2017) provides a large collection of English texts annotated with Discourse Representation Structures (see Figure 1 for an example). GMB integrates various levels of semantic annotation (e.g., anaphora, named entities, thematic roles, rhetorical relations) into a unified formalism providing expressive meaning representations for open-domain texts.

We treat DRT parsing as a structure prediction problem. We develop a method to transform DRSs to tree-based representations which can be further linearized to bracketed string format. We examine a series of encoder-decoder models (Bahdanau et al., 2015) differing in the way tree-
The statement says each of the dead men wore magazine vests and carried two hand grenades.

structured logical forms are generated and show that a structure-aware decoder is paramount to open-domain semantic parsing. Our proposed model decomposes the decoding process into three stages. The first stage predicts the structure of the meaning representation omitting details such as predicates or variable names. The second stage fills in missing predicates and relations (e.g., thing, Agent) conditioning on the natural language input and the previously predicted structure. Finally, the third stage predicts variable names based on the input and the information generated so far.

Decomposing decoding into these three steps reduces the complexity of generating logical forms since the model does not have to predict deeply nested structures, their variables, and predicates all at once. Moreover, the model is able to take advantage of the GMB annotations more efficiently, e.g., examples with similar structures can be effectively used in the first stage despite being very different in their lexical make-up. Finally, a piecemeal mode of generation yields more accurate predictions; since the output of every decoding step serves as input to the next one, the model is able to refine its predictions taking progressively more global context into account. Experimental results on the GMB show that our three-stage decoder outperforms a vanilla encoder-decoder model and a related variant which takes shallow subtree structures that have a model-theoretic interpretation and can be translated into first-order logic (Kamp and Reyle, 1993). Basic DRSs consist of discourse referents (e.g., x,y) representing entities in the discourse and discourse conditions (e.g., \( \text{man}(x) \), \( \text{magazine}(y) \)) representing information about discourse referents. Following conventions in the DRT literature, we visualize DRSs in a box-like format (see Figure 1).

GMB adopts a variant of DRT that uses a neo-Davidsonian analysis of events (Kipper et al., 2008), i.e., events are first-order entities characterized by one-place predicate symbols (e.g., say(e1) in Figure 1). In addition, it follows Projective Discourse Representation Theory (PDRT; Venhuizen et al. 2013) an extension of DRT specifically developed to account for the interpretation of presuppositions and related projection phenomena.

Our contributions in this work are three-fold: an open-domain semantic parser which yields discourse representation structures; a novel end-to-end neural model equipped with a structured decoder which decomposes the parsing process into three stages; a DRS-to-tree conversion method which transforms DRSs to tree-based representations allowing for the application of structured decoders as well as sequential modeling. We release our code\(^1\) and tree formatted version of the GMB in the hope of driving further research in open-domain semantic parsing.

### 2 Discourse Representation Theory

In this section we provide a brief overview of the representational semantic formalism used in the GMB. We refer the reader to Bos et al. (2017) and Kamp and Reyle (1993) for more details.

Discourse Representation Theory (DRT; Kamp and Reyle 1993) is a general framework for representing the meaning of sentences and discourse which can handle multiple linguistic phenomena including anaphora, presuppositions, and temporal expressions. The basic meaning-carrying units in DRT are Discourse Representation Structures (DRSs), which are recursive formal meaning structures that have a model-theoretic interpretation and can be translated into first-order logic (Kamp and Reyle, 1993). Basic DRSs consist of discourse referents (e.g., x,y) representing entities in the discourse and discourse conditions (e.g., \( \text{man}(x) \), \( \text{magazine}(y) \)) representing information about discourse referents. Following conventions in the DRT literature, we visualize DRSs in a box-like format (see Figure 1).

\[^1\]https://github.com/EdinburghMLP/EncDecDRSparSing
Basic DRSs consist of a set of referents (\(\text{ref}\)) (denoting individuals or discourse referents) and (\(\exp\)) (i.e., truth values):
\[
\begin{align*}
\exp & := \text{ref}, \quad \exp := \text{drs} | \text{drs}, \\
\text{drs} & := \text{pvar} | \text{num}.
\end{align*}
\]
(1)

More formally, DRSs are expressions of type \(\langle \exp \rangle\) (e.g., factive verbs) with propositional content (e.g., conventional implicatures). In PDRT, each basic DRS introduces a label, which can be bound by a pointer indicating the interpretation site of semantic content. To account for the rhetorical structure of texts, GMB adopts Segmented Discourse Representation Theory (SDRT; Asher and Lascarides 2003). In SDRT, discourse segments are linked with rhetorical relations reflecting different characteristics of textual coherence, such as temporal order and communicative intentions (see continuations \(k_1, k_2\) in Figure 1).

As mentioned earlier, DRSs are expressions of type \(\langle \exp \rangle\) (denoting individuals or discourse referents) and (\(\exp\)) (i.e., truth values):
\[
\begin{align*}
\exp & := \langle \exp \rangle | \langle \exp \rangle | \langle \exp \rangle, \\
\text{drs} & := \text{pvar} | \text{num}.
\end{align*}
\]
(2)

Discourse referents \(\text{ref}\) are in turn classified into six categories, namely common referents \(x_n\), event referents \(e_n\), state referents \(s_n\), segment referents \(k_n\), proposition referents \(\pi_n\), and time referents \(t_n\). \(\langle \text{drs} \rangle\) and \(\langle \text{drs} \rangle\) denote basic and segmented DRSs, respectively:
\[
\langle \text{drs} \rangle := \frac{k_1 : \langle \exp \rangle, k_2 : \langle \exp \rangle}{\text{coo}(k_1, k_2)}, \quad \frac{k_1 : \langle \exp \rangle, k_2 : \langle \exp \rangle}{\text{sub}(k_1, k_2)}.
\]
(3)

Basic DRSs consist of a set of referents \(\langle \text{ref} \rangle\) and conditions \(\langle \text{condition} \rangle\), whereas segmented DRSs are 

\[
\langle \text{condition} \rangle := \langle \text{basic} \rangle | \langle \text{complex} \rangle.
\]
(4)

Basic conditions express properties of discourse referents or relations between them:
\[
\langle \text{basic} \rangle := \langle \text{sym}_1 \rangle (\langle \exp \rangle) | \langle \text{sym}_2 \rangle (\langle \exp \rangle, \langle \exp \rangle) | \langle \exp \rangle = \langle \exp \rangle | \langle \exp \rangle = \langle \text{num} \rangle | \text{timex}(\langle \exp \rangle, \langle \text{sym}_0 \rangle) | \text{named}(\langle \exp \rangle, \langle \text{sym}_0 \rangle, \text{class}).
\]
(5)

where \(\text{sym}_n\) denotes \(n\)-place predicates, \(\langle \text{num} \rangle\) denotes cardinal numbers, \(\text{timex}\) expresses temporal information (e.g., \(\text{timex}(x_7,2005)\) denotes the year 2005), and \(\text{class}\) refers to named entity classes (e.g., location).

Complex conditions are unary or binary. Unary conditions have one DRS as argument and represent negation (\(\Box\)) and modal operators expressing necessity (\(\Box\)) and possibility (\(\Diamond\)). Condition

<table>
<thead>
<tr>
<th>sections</th>
<th># doc</th>
<th># sent</th>
<th># token</th>
<th>avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>00-99</td>
<td>10,000</td>
<td>62,010</td>
<td>1,354,149</td>
<td>21.84</td>
</tr>
<tr>
<td>20-99</td>
<td>7,970</td>
<td>49,411</td>
<td>1,078,953</td>
<td>21.83</td>
</tr>
<tr>
<td>10-19</td>
<td>1,038</td>
<td>6,483</td>
<td>142,344</td>
<td>21.95</td>
</tr>
<tr>
<td>00-09</td>
<td>992</td>
<td>6,116</td>
<td>132,852</td>
<td>21.72</td>
</tr>
</tbody>
</table>

Table 1: Statistics on the GMB (avg denotes the average number of tokens per sentence).

Structural Representations (DRSs) in GMB were obtained from Boorer (Bos, 2018, 2015), and then refined using expert linguists and crowdsourcing methods. Boxer constructs DRSs based on a pipeline of tools involving POS-tagging, named entity recognition, and parsing. Specifically, it relies on the syntactic analysis of the C&C parser (Clark and Curran, 2007), a general-purpose parser using the framework of Combinatory Categorial Grammar (CCG; Steedman 2001). DRSs are obtained from CCG parses, with semantic composition being guided by the CCG syntactic derivation.

Documents in the GMB were collected from a variety of sources including Voice of America (a newspaper published by the US Federal Government), the Open American National Corpus, Aesop’s fables, humorous stories and jokes, and country descriptions from the CIA World Factbook. The dataset consists of 10,000 documents each annotated with a DRS. Various statistics on the GMB are shown in Table 1. Bos et al. (2017) recommend sections 20–99 for training, 10–19 for tuning, and 00–09 for testing.

3 The Groningen Meaning Bank Corpus

Corpus Creation DRSs in GMB were obtained from Boxer (Bos, 2008, 2015), and then refined using expert linguists and crowdsourcing methods. Boxer constructs DRSs based on a pipeline of tools involving POS-tagging, named entity recognition, and parsing. Specifically, it relies on the syntactic analysis of the C&C parser (Clark and Curran, 2007), a general-purpose parser using the framework of Combinatory Categorial Grammar (CCG; Steedman 2001). DRSs are obtained from CCG parses, with semantic composition being guided by the CCG syntactic derivation.

Documents in the GMB were collected from a variety of sources including Voice of America (a newspaper published by the US Federal Government), the Open American National Corpus, Aesop’s fables, humorous stories and jokes, and country descriptions from the CIA World Factbook. The dataset consists of 10,000 documents each annotated with a DRS. Various statistics on the GMB are shown in Table 1. Bos et al. (2017) recommend sections 20–99 for training, 10–19 for tuning, and 00–09 for testing.

DRS-to-Tree Conversion As mentioned earlier, DRSs in the GMB are displayed in a box-like format which is intuitive and easy to read but not particularly amenable to structure modeling. In this section we discuss how DRSs were post-processed and simplified into a tree-based format, which served as input to our models.

The GMB provides DRS annotations per-document. Our initial efforts have focused on sentence-level DRS parsing which is undoubtedly
a necessary first step for more global semantic representations. It is relatively straightforward to obtain sentence-level DRSs from document-level annotations since referents and conditions are indexed to tokens. We match each sentence in a document with the DRS whose content bears the same indices as the tokens occurring in the sentence. This matching process yields 52,268 sentences for training (sections 20–99), 5,172 sentences for development (sections 10–19), (development), and 5,440 sentences for testing (sections 00–09).

In order to simplify the representation, we omit referents in the top part of the DRS (e.g., \( x_1 \), \( e_1 \) and \( \pi_1 \) in Figure 1) but preserve them in conditions without any information loss. Also we ignore pointers to DRSs since this information is implicitly captured through the typing and co-indexing of referents. Definition (1) is simplified to:

\[
\text{(drs)} := \text{DRS}((\text{condition})^*),
\]

where DRS() denotes a basic DRS. We also modify discourse referents to SDRSs (e.g., \( k_1 \), \( k_2 \) in Figure 1) which we regard as elements bearing scope over expressions \( \langle \exp \rangle \) and add a 2-place predicate \( \langle \text{sym}_2 \rangle \) to describe the discourse relation between them. So, definition (3) becomes:

\[
\text{(sdrs)} := \text{SDRS}((\langle \text{ref} \rangle)(\langle \exp \rangle))^* \]

\[
(\langle \text{sym}_2 \rangle((\langle \exp \rangle),((\langle \text{ref} \rangle))))^*,
\]

where SDRS() denotes a segmented DRS, and \( \langle \text{ref} \rangle \) are segment referents.

We treat cardinal numbers \( \langle \text{num} \rangle \) and \( \langle \text{sym}_0 \rangle \) in relation \textit{timex} as constants. We introduce the binary predicate “card” to represent cardinality (e.g., \( |x_8| = 2 \) is card(\( x_8, \text{NUM} \)). We also simplify \( \langle \exp_e \rangle = \langle \exp_n \rangle \) to \( \text{eq}(\langle \exp_n \rangle, \langle \exp_n \rangle) \) using the binary relation “eq” (e.g., \( x_1 = x_2 \) becomes \( \text{eq}(x_1, x_2) \)). Moreover, we ignore \textit{class} in \textit{named} and transform \textit{named}(\( \langle \exp_n \rangle, \langle \text{sym}_0 \rangle, \text{class} \) into \( \langle \text{sym}_1 \rangle((\langle \exp_n \rangle)) \) (e.g., named(\( x_2, \text{mongolia,geo} \) becomes mongolia(\( x_2 \))). Consequently, basic conditions (see definition (5)) are simplified to:

\[
\text{(basic)} := \langle \text{sym}_1 \rangle((\langle \exp_n \rangle))\langle \text{sym}_2 \rangle((\langle \exp_n \rangle, \langle \exp_n \rangle))
\]

Analogously, we treat \textit{unary} and \textit{binary} conditions as scoped functions, and definition (6) becomes:

\[
\text{(unary)} := \neg \mid \Box \mid \Diamond \mid \langle \text{ref} \rangle((\langle \exp_n \rangle))
\]

\[
\text{(binary)} := \rightarrow \mid \lor \mid ?(\langle \exp_n \rangle, \langle \exp_n \rangle).
\]

Following the transformations described above, the DRS in Figure 1 is converted into the tree in Figure 2, which can be subsequently linearized into a PTB-style bracketed sequence. It is important to note that the conversion does not diminish the complexity of DRSs. The average tree width in the training set is 10.39 and tree depth is 4.64.

4 Semantic Parsing Models

We present below three encoder-decoder models which are increasingly aware of the structure of the DRT meaning representations. The models take as input a natural language sentence \( X \) represented as \( w_1, w_2, \ldots, w_n \), and generate a sequence \( Y = (y_1, y_2, \ldots, y_m) \), which is a linearized tree (see Figure 2 bottom), where \( n \) is the length of the sentence, and \( m \) the length of the generated DRS sequence. We aim to estimate \( p(Y|X) \), the conditional probability of the semantic parse tree \( Y \) given natural language input \( X \):

\[
p(Y|X) = \prod_j p(y_j|Y_i^{j-1}, X^n)
\]

4.1 Encoder

An encoder is used to represent the natural language input \( X \) into vector representations. Each token in a sentence is represented by a vector \( x_k \) which is the concatenation of randomly initialized embeddings \( e_w \), \( w \) is the pre-trained word embeddings \( \bar{e}_w \), and lemma embeddings \( e_l \): \( x_k = \text{tanh}([e_w; \bar{e}_w; e_l] * W_1 + b_1) \), where \( W_1 \in \mathbb{R}^{d_w} \) and \( D \) is a shorthand for \( (d_w + d_p + d_l) \times d_{\text{input}} \) (subscripts \( w, p, \) and \( l \) denote the dimensions of word embeddings, pre-trained embeddings, and lemma embeddings, respectively); \( b_1 \in \mathbb{R}^{d_{\text{input}}} \) and the symbol ; denotes concatenation. Embeddings \( e_w \),
and $e_t$ are randomly initialized and tuned during training, while $\tilde{e}_t$ are fixed.

We use a bidirectional recurrent neural network with long short-term memory units (bi-LSTM; Hochreiter and Schmidhuber 1997) to encode natural-language sentences:

$$[h_{e1} : h_{en}] = \text{bi-LSTM}(x_1 : x_n),$$

where $h_{ei}$ denotes the hidden representation of the encoder, and $x_i$ refers to the input representation of the $ith$ token in the sentence. Table 2 summarizes the notation used throughout this paper.

### 4.2 Sequence Decoder

We employ a sequential decoder (Bahdanau et al., 2015) as our baseline model with the architecture shown in Figure 3(a). Our decoder is a (forward) LSTM, which is conditionally initialized with the hidden state of the encoder, i.e., we set $h_{idb} = h_{en}$ and $c_{idb} = c_{en}$, where $c$ is a memory cell:

$$h_{dj} = \text{LSTM}(e_{y_{j-1}}),$$

where $h_{dj}$ denotes the hidden representation of $y_j$, $e_{y_j}$ are randomly initialized embeddings tuned during training, and $y_0$ denotes the start of sequence.

The decoder uses the contextual representation of the encoder together with the embedding of the previously predicted token to output the next token from the vocabulary $V$:

$$s_j = [h_{ctj}; e_{y_{j-1}}] * W_2 + b_2,$$

where $W_2 \in \mathbb{R}^{(d_{enc}+d_y) \times |V|}$, $b_2 \in \mathbb{R}^{|V|}$, $d_{enc}$ and $d_y$ are the dimensions of the encoder hidden unit and output representation, respectively, and $h_{ctj}$ is obtained using an attention mechanism:

$$h_{ctj} = \sum_{i=1}^{n} \beta_{ji} h_{ei},$$

where the weight $\beta_{ji}$ is computed by:

$$\beta_{ji} = \frac{e^{f(h_{dj}, h_{ei})}}{\sum_k e^{f(h_{dj}, h_{ek})}},$$

and $f$ is the dot-product function. We obtain the probability distribution over the output tokens as:

$$p_j = p(y_j|Y_1^{j-1}, X_1^n) = \text{SOFTMAX}(s_j)$$

### 4.3 Shallow Structure Decoder

The baseline decoder treats all conditions in a DRS uniformly and has no means of distinguishing between conditions corresponding to tokens in a sentence (e.g., the predicate say($e_1$) refers to the verb said) and semantic relations (e.g., Cause($e_1, x_1$)). Our second decoder attempts to take this into account by distinguishing conditions which are local and correspond to words in a sentence from items which are more global and express semantic content (see Figure 3(b)). Specifically, we model sentence specific conditions using a copying mechanism, and all other conditions $G$ which do not correspond to sentential tokens (e.g., thematic roles, rhetorical relations) with an insertion mechanism.

Each token in a sentence is assigned a copying score $o_{ji}$:

$$o_{ji} = h_{dj}^T W_3 h_{ei},$$

where subscript $ji$ denotes the $ith$ token at $jth$ time step, and $W_3 \in \mathbb{R}^{d_{enc} \times d_{enc}}$. All other conditions $G$ are assigned an insertion score:

$$s_j = [h_{ctj}; e_{y_{j-1}}] * W_4 + b_4,$$

where $W_4 \in \mathbb{R}^{(d_{enc}+d_y) \times |G|}$, $b_4 \in \mathbb{R}^{|G|}$, and $h_{ctj}$ are the same with the baseline decoder. We obtain the probability distribution over output tokens as:

$$p_j = p(y_j|Y_1^{j-1}, X_1^n) = \text{SOFTMAX}([o_j; s_j])$$

### Table 2: Notation used throughout this paper.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X; Y$</td>
<td>sequence of words; outputs</td>
</tr>
<tr>
<td>$w_i; y_i$</td>
<td>the $ith$ word; output</td>
</tr>
<tr>
<td>$X_i^j; Y_i^j$</td>
<td>word; output sequence from position $i$ to $j$</td>
</tr>
<tr>
<td>$e_w; e_y$</td>
<td>random embedding of word $w_i$; of output $y_i$</td>
</tr>
<tr>
<td>$\tilde{e}_w; \tilde{e}_y$</td>
<td>fixed pretrained embedding of word $w_i$</td>
</tr>
<tr>
<td>$e_p$</td>
<td>random embedding for lemma $l_i$</td>
</tr>
<tr>
<td>$d_{w}$</td>
<td>dimension of random word embedding</td>
</tr>
<tr>
<td>$d_p$</td>
<td>dimension of pretrained word embedding</td>
</tr>
<tr>
<td>$d_{input}$</td>
<td>input dimension of encoder</td>
</tr>
<tr>
<td>$d_{enc}; d_{dec}$</td>
<td>hidden dimension of encoder; decoder matrix of model parameters</td>
</tr>
<tr>
<td>$x_i$</td>
<td>representation of $ith$ token</td>
</tr>
<tr>
<td>$h_{ei}$</td>
<td>hidden representation of $ith$ token</td>
</tr>
<tr>
<td>$e_{ci}$</td>
<td>memory cell of $ith$ token in encoder</td>
</tr>
<tr>
<td>$h_{ci}$</td>
<td>hidden representation of $ith$ token in decoder</td>
</tr>
<tr>
<td>$s_j$</td>
<td>score vector of $jth$ output in decoder</td>
</tr>
<tr>
<td>$h_{ctj}$</td>
<td>context representation of $jth$ output</td>
</tr>
<tr>
<td>$\beta_j$</td>
<td>alignment from $jth$ output to $ith$ token</td>
</tr>
<tr>
<td>$o_j$</td>
<td>copy score of $jth$ output from $ith$ token</td>
</tr>
<tr>
<td>$-^*$</td>
<td>indicates tree structure (e.g., $\tilde{Y}, \tilde{y}_j, \tilde{s}_j$)</td>
</tr>
<tr>
<td>$-^*$</td>
<td>indicates DRS conditions (e.g., $\tilde{Y}, \tilde{y}_j, \tilde{s}_j$)</td>
</tr>
<tr>
<td>$-^*$</td>
<td>indicates referents (e.g., $Y, \bar{y}_j, s_j$)</td>
</tr>
</tbody>
</table>
4.4 Deep Structure Decoder

As explained previously, our structure prediction problem is rather challenging: the length of a bracketed DRS is nearly five times longer than its corresponding sentence. As shown in Figure 1, a bracketed DRS, \( y_1, y_2, \ldots, y_n \) consists of three parts: internal structure \( \bar{Y} = \hat{y}_1, \hat{y}_2, \ldots, \hat{y}_r \) (e.g., DRS( \( \pi_1( \text{SDRS}(k_1(\text{DRS}(s_1(\text{DRS}()) \text{DRS}()))) \)) \)) \( k_2(\text{DRS}() \text{DRS}() \text{DRS}()) \)), conditions \( \bar{Y} = \hat{y}_1, \hat{y}_2, \ldots, \hat{y}_r \) (e.g., statement, say, Topic), and referents \( \bar{Y} = 1, y_2, \ldots, y_v \) (e.g., \( x_1, e_1, \pi_1 \)), where \( t + r + 2 + v = n \).

Our third decoder (see Figure 3(c)) first predicts the structural make-up of the DRS, then the conditions, and finally their referents in an end-to-end framework. The probability distribution of structured output \( Y \) given natural language input \( X \) is rewritten as:

\[
p(Y|X) = p(\bar{Y}, \hat{Y}, \bar{Y} | \bar{Y}) = \prod_j p(\hat{y}_j | \hat{y}_j^{-1}, X) \times \prod_j p(\bar{y}_j | \bar{y}_j^{-1}, \hat{y}_j, X) \times \prod_j p(\hat{y}_j | \hat{y}_j^{-1}, \bar{y}_j, \hat{y}_j, X) \tag{11}
\]

where \( \hat{y}_j^{-1}, \bar{y}_j^{-1}, \) and \( \hat{y}_j^{-1} \) denote the tree structure, conditions, and referents predicted so far.

\( \hat{y}_j^r \) denotes the structure predicted before conditions \( \bar{y}_j; \hat{y}_j^r \) and \( \bar{Y}_j^r \) are the structures and conditions predicted before referents \( \hat{y}_j \). We next discuss how each decoder is modeled.

Structure Prediction To model basic DRS structure we apply the shallow decoder discussed in Section 4.3 and also shown in Figure 3(c.1). Tokens in such structures correspond to parent nodes in a tree; in other words, they are all inserted from \( G \), and subsequently predicted tokens are only scored with the insert score, i.e., \( \hat{s}_i = s_i \). The hidden units of the decoder are:

\[
\hat{h}_{d_j} = \text{LSTM}(e_{\bar{y}_{j-1}}),
\]

And the probabilistic distribution over structure denoting tokens is:

\[
p(y_j | y_j^{-1}, X) = \text{SOFTMAX}(\hat{s}_j)
\]

Condition Prediction DRS conditions are generated by taking previously predicted structures into account, e.g., when “DRS(” or “SDRS(“ are predicted, their conditions will be generated next. By mapping \( j \) to \( (k, m_k) \), the sequence of conditions can be rewritten as \( \bar{y}_1, \ldots, \bar{y}_j, \ldots, \bar{y}_r = \bar{y}_1(1), \bar{y}_2(1), \ldots, \bar{y}_m(k), \ldots, \) where \( \bar{y}_m(k) \) is \( m_k \)th
condition of structure token $\hat{y}_k$. The corresponding hidden units $\hat{h}_{d_k}$ act as conditional input to the decoder. Structure denoting tokens (e.g., “DRS(“) or “SDRS(“) are fed into the decoder one by one to generate the corresponding conditions as:

$$e_{y(k)} = \hat{h}_{d_k} \ast W_5 + b_5,$$

where $W_5 \in \mathbb{R}^{d_{dec} \times d_i}$ and $b_5 \in \mathbb{R}^{d_i}$. The hidden unit of the conditions decoder is computed as:

$$\hat{h}_{d_j} = \hat{h}_{d(k,m_k)} = \text{LSTM}(e_{\hat{y}(k,m_k-1)}),$$

Given hidden unit $\hat{h}_{d_j}$, we obtain the copy score $\hat{\sigma}_j$ and insert score $\hat{s}_j$. The probabilistic distribution over conditions is:

$$p(\hat{y}_j|\hat{Y}_i^{j-1}, \hat{Y}_i^{j'}, X) = \text{SOFTMAX}(\hat{\sigma}_j; \hat{s}_j).$$

## Referent Prediction
Referents are generated based on the structure and conditions of the DRS. Each condition has at least one referent. Similar to condition prediction, the sequence of referents can be rewritten as $\hat{y}_1, \ldots, \hat{y}_j, \ldots, \hat{y}_k = \hat{y}(1.1) , \hat{y}(1.2), \ldots, \hat{y}(k,m_k)\ldots$. The hidden units of the conditions decoder are fed into the referent decoder $e_{\hat{y}(k,0)} = \hat{h}_{d_k} \ast W_6 + b_6$, where $W_6 \in \mathbb{R}^{d_{dec} \times d_i}$, $b_6 \in \mathbb{R}^{d_i}$. The hidden unit of the referent decoder is computed as:

$$\hat{h}_{d_j} = \hat{h}_{d(k,m_k)} = \text{LSTM}(e_{\hat{y}(k,m_k-1)}),$$

All referents are inserted from $\mathcal{G}$, given hidden unit $\hat{h}_{d_j}$ (we only obtain the insert score $\hat{s}_j$). The probabilistic distribution over predicates is:

$$p(\hat{y}_j|\hat{Y}_i^{j-1}, \hat{Y}_i^{j'}, \hat{Y}_i^{j''}, X) = \text{SOFTMAX} (\hat{s}_j).$$

Note that a single LSTM is adopted for structure, condition and referent prediction. The mathematical symbols are summarized in Table 2.

## 4.5 Training
The models are trained to minimize a cross-entropy loss objective with $\ell_2$ regularization:

$$L(\theta) = - \sum_j \log p_j + \frac{\lambda}{2} ||\theta||^2,$$

where $\theta$ is the set of parameters, and $\lambda$ is a regularization hyper-parameter ($\lambda = 10^{-6}$). We used stochastic gradient descent with Adam (Kingma and Ba, 2014) to adjust the learning rate.

## 5 Experimental Setup
### Settings
Our experiments were carried out on the GMB following the tree conversion process discussed in Section 3. We adopted the training, development, and testing partitions recommended in Bos et al. (2017). We compared the three models introduced in Section 4, namely the baseline sequence decoder, the shallow structured decoder and the deep structure decoder. We used the same empirical hyper-parameters for all three models. The dimensions of word and lemma embeddings were 64 and 32, respectively. The dimensions of hidden vectors were 256 for the encoder and 128 for the decoder. The encoder used two hidden layers, whereas the decoder only one. The dropout rate was 0.1. Pre-trained word embeddings (100 dimensions) were generated with Word2Vec trained on the AFP portion of the English Gigaword corpus.\(^3\)

### Evaluation
Due to the complex nature of our structured prediction task, we cannot expect model output to exactly match the gold standard. For instance, the numbering of the referents may be different, but nevertheless valid, or the order of the children of a tree node (e.g., “DRS(india(x_1) say(e_1))” and “DRS(say(e_1) india(x_1))” are the same). We thus use $F_1$ instead of exact match accuracy. Specifically, we report D-match\(^4\) a metric designed to evaluate scoped meaning representations and released as part of the distribution of the Parallel Meaning Bank corpus (Abzianidze et al., 2017). D-match is based on Smatch\(^5\), a metric used to evaluate AMR graphs (Cai and Knight, 2013); it calculates $F_1$ on discourse representation graphs (DRGs), i.e., triples of nodes, arcs, and their referents, applying multiple restarts to obtain a good referent (node) mapping between graphs.

We converted DRSs (predicted and goldstandard) into DRGs following the top-down procedure described in Algorithm 1.\(^6\) $\text{ISCONDITION}$ returns $true$ if the child is a condition (e.g., $\text{india(x_1)}$), where three arcs are created, one is connected to a parent node and the other two are connected to $\text{arg}_1$ and $\text{arg}_2$, respectively (lines 7–12). $\text{ISQUANTIFIER}$ returns $true$ if the child is a quantifier (e.g., $\pi_1$, $\neg$ and $\exists$) and three arcs are created; one is connected to the parent node, one to the referent that is created if and only...
Algorithm 1 DRS to DRG Conversion

**Input:** $T$, tree-like DRS

**Output:** $G$, a set of edges

1: $n_b \leftarrow 0; n_c \leftarrow 0; G \leftarrow \emptyset$
2: stack $\leftarrow [[]; R \leftarrow \emptyset$
3: procedure TRAVELDRS(parent)
4: stack.append($b_n$); $n_b \leftarrow n_b + 1$
5: $node_p \leftarrow stack.top$
6: for child in parent do
7: if $IS\text{CONDITION}(child)$ then
8: $G \leftarrow G \cup \{node_p \xrightarrow{\text{child} \_\text{rel}} c_n\}$
9: $G \leftarrow G \cup \{c_n \xrightarrow{\text{arg} 1} \text{child}.\text{arg} 1\}$
10: $G \leftarrow G \cup \{c_n \xrightarrow{\text{arg} 2} \text{child}.\text{arg} 2\}$
11: $n_c \leftarrow n_c + 1$
12: ADDREFERENT($node_p, child$)
13: else if $IS\text{QUANTIFIER}(child)$ then
14: $G \leftarrow G \cup \{node_p \xrightarrow{\text{child} \_\text{class}} c_n\}$
15: $G \leftarrow G \cup \{c_n \xrightarrow{\text{arg} 1} \text{child}.\text{arg} 1\}$
16: $G \leftarrow G \cup \{c_n \xrightarrow{\text{arg} 1} b_{n+1}\}$
17: $n_c \leftarrow n_c + 1$
18: if $IS\text{PROPSEG}(child)$ then
19: ADDREFERENT($node_p, child$)
20: end if
21: TRAVELDRS(child.nextDRS)
22: end if
23: stack.pop() procedure
24: procedure ADDREFERENT($node_p, child$)
25: if child.arg 1 not in $R$ then
26: $G \leftarrow G \cup \{node_p \xrightarrow{\text{ref} \_\text{rel}} \text{child}.\text{arg} 1\}$
27: $R \leftarrow R \cup \text{child}$.arg 1
28: end if
29: if child.arg 2 not in $R$ then
30: $G \leftarrow G \cup \{node_p \xrightarrow{\text{ref} \_\text{rel}} \text{child}.\text{arg} 2\}$
31: $R \leftarrow R \cup \text{child}.arg 2$
32: end if
33: procedure
34: TRAVELDRS($T$)
35: return $G$

if the child is a proposition or segment (e.g., $p_1$ and $k_1$), and one is connected to the next DRS or SDRS nodes (lines 13–20). The algorithm will recursively travel all DRS or SDRS nodes (line 21). Furthermore, arcs are introduced to connect DRS or SDRS nodes to the referents that first appear in a condition (lines 26–35).

When comparing two DRGs, we calculate the $F_1$ over their arcs. For example consider the two DRGs (a) and (b) shown in Figure 4. Let $\{b_0 : b_0, x_1 : x_2, x_2 : x_3, c_0 : c_0, c_1 : c_2, c_2 : c_3\}$ denote the node alignment between them. The number of matching arcs is eight, the number of arcs in the gold DRG is nine, and the number of arcs in the predicted DRG is 12. So recall is 8/9, precision is 8/12, and $F_1$ is 76.19.

Figure 4: (a) is the gold DRS and (b) is the predicted DRS (condition names are not shown).

6 Results

Table 3 compares our three models on the development set. As can be seen, the shallow structured decoder performs better than the baseline decoder, and the proposed deep structure decoder outperforms both of them. Ablation experiments show that without pre-trained word embeddings or word lemma embeddings, the model generally performs worse. Compared to lemma embeddings, pre-trained word embeddings contribute more.

Table 4 shows our results on the test set. To assess the degree to which the various decoders contribute to DRS parsing, we report results when predicting the full DRS structure (second block), when ignoring referents (third block), and when ignoring both referents and conditions (fourth block). Overall, we observe that the shallow structure model improves precision across the baseline with a slight loss in recall, while the deep structure model performs best by a large margin. When referents are not taken into account (compare the second and third blocks in Table 4), performance improves across the board. When conditions are additionally omitted, we observe further performance gains. This is hardly surprising, since errors propagate from one stage to the next when predicting full DRS structures. Further analysis revealed that the parser performs slightly better on (copy) conditions which correspond to natural language tokens compared to (insert) conditions (e.g., Topic, Agent) which are generated from global semantic content (83.22 vs 80.63 $F_1$). The parser is also better on sentences which do not represent SDRSs (79.12 vs 68.36 $F_1$) which is expected given that they usually correspond to more elaborate structures. We also found that rhetorical relations (linking segments) are predicted fairly accurately, especially if they are frequently attested (e.g., Continuation, Parallel), while the parser has difficulty with relations denoting contrast.
Figure 5 shows F$_1$ performance for the three parsers on sentences of different length. We observe a similar trend for all models: as sentence length increases, model performance decreases. The baseline and shallow models do not perform well on short sentences which despite containing fewer words, can still represent complex meaning which is challenging to capture sequentially. On the other hand, the performance of the deep model is relatively stable. LSTMs in this case function relatively well, as they are faced with the easier task of predicting meaning in different stages (starting with a tree skeleton which is progressively refined). We provide examples of model output in the supplementary material.

7 Related Work

Tree-structured Decoding A few recent approaches develop structured decoders which make use of the syntax of meaning representations. Dong and Lapata (2016) and Alvarez-Melis and Jaakkola (2017) generate trees in a top-down fashion, while in other work (Xiao et al., 2016; Krishnamurthy et al., 2017) the decoder generates from a grammar that guarantees that predicted logical forms are well-typed. In a similar vein, Yin and Neubig (2017) generate abstract syntax trees (ASTs) based on the application of production rules defined by the grammar. Rabinovich et al. (2017) introduce a modular decoder whose various components are dynamically composed according to the generated tree structure. In comparison, our model does not use grammar information explicitly. We first decode the structure of the DRS, and then fill in details pertaining to its semantic content. Our model is not strictly speaking top-down, we generate partial trees sequentially, and then expand non-terminal nodes, ensuring that when we generate the children of a node, we have already obtained the structure of the entire tree.

Wide-coverage Semantic Parsing Our model is trained on the GMB (Bos et al., 2017), a richly annotated resource in the style of DRT which provides a unique opportunity for bootstrapping wide-coverage semantic parsers. Boxer (Bos, 2008) was a precursor to the GMB, the first semantic parser of this kind, which deterministically maps CCG derivations onto formal meaning representations. Le and Zuidema (2012) were the first to train a semantic parser on an early release of the GMB (2,000 documents; Basile et al. 2012), however, they abandon lambda calculus in favor of a graph based representation. The latter is closely related to AMR, a general-purpose meaning representation language for broad-coverage text. In AMR the meaning of a sentence is represented as a rooted, directed, edge-labeled and leaf-labeled graph. AMRs do not resemble classical meaning representations and do not have a model-theoretic interpretation. However, see Bos (2016) and Artzi et al. (2015) for translations to first-order logic.

8 Conclusions

We introduced a new end-to-end model for open-domain semantic parsing. Experimental results on the GMB show that our decoder is able to recover discourse representation structures to a good degree (77.54 F$_1$), albeit with some simplifications. In the future, we plan to model document-level representations which are more in line with DRT and the GMB annotations.

Acknowledgments We thank the anonymous reviewers for their feedback and Johan Bos for answering several questions relating to the GMB. We gratefully acknowledge the support of the European Research Council (Lapata, Liu; award number 681760) and the EU H2020 project SUMMA (Cohen, Liu; grant agreement 688139).
References


Hans Kamp and Uwe Reyle. 1993. From discourse to logic: an introduction to modelfheoretic semantics of natural language, formal logic and DRT.


