Soft Well-Formed Semantic Parsing with Score-Based Selection

Jiangming Liu
Yunnan University
Yunnan Key Laboratory of Intelligent Systems and Computing, Yunnan University
jiangmingliu@ynu.edu.cn

Abstract

Semantic parsing is the task of translating natural language into a structured, formal semantic representation that can be interpreted by machines. These semantic representations are organized with complex structures. While various models have been developed for semantic parsing, there has been limited focus on generating semantic representations with well-formed structures. In this study, we introduce a score-based method to select well-formed outputs from candidates generated by beam search algorithms. Our experiments focus on parsing texts into discourse representation structures, which are innovative semantic representations designed to capture the meaning of texts with arbitrary lengths across languages. Our experimental results demonstrate that models utilizing the proposed method can reduce the number of ill-formed outputs and improve F1 scores in English. Furthermore, our final model achieves significant improvements in German, Italian and Dutch zero-shot DRS parsing by effectively preventing ill-formed outputs.

Keywords: semantic parsing, discourse representation structure, well-formed structure, score-based selection

1. Introduction

Semantic parsing is the task of translating natural language into a formal semantic representation that can be comprehended and interpreted by machines using logical structures. These semantic representations can take the form of executable programming code that can be directly run by operating systems (Zettlemoyer and Collins, 2005). Also, they can manifest as symbolic representations of the meaning within textual content, allowing for logical inference and language reasoning (Bos, 2008; Banarescu et al., 2013).

Whether the goal is to execute programming code or make logical inferences, it is crucial that the semantic representations produced by the parser are well-formed and adhere to the syntactic and semantic rules of the respective formal systems. However, the neural network-based approaches (Dong and Lapata, 2016; Alvarez-Melis and Jaakkola, 2017; Cheng et al., 2017) in semantic parsing, such as AMR parsing (Damonte et al., 2017; Konstas et al., 2017; Lee et al., 2022; Bai et al., 2022) and DRS parsing (Liu et al., 2018; van Noord et al., 2018; Liu et al., 2021; Wang et al., 2023), has predominantly focused on enhancing the parser's performance in terms of accuracy, often at the expense of ensuring that the generated outputs are well-formed.

Discourse Representation Structure (DRS) is an innovative semantic representation based on Discourse Representation Theory (DRT; Kamp and Reyle 1993), designed to capture a wide range of linguistic phenomena, including discourse relations, the interpretation of pronouns, and temporal expressions, in texts of varying lengths and across multiple languages. Due to its intricate structures, the issue of producing ill-formed semantic representations poses a more significant challenge in DRS parsing compared to other semantic representations. However, recent DRS parsers focusing on clauses (van Noord et al., 2018, 2019; Liu et al., 2019b; van Noord et al., 2020; Wang et al., 2021; Liu et al., 2021), trees (van Noord et al., 2018, 2019; Liu et al., 2019b; van Noord et al., 2020; Wang et al., 2021; Liu et al., 2021), and graphs (Poelman et al., 2022; Wang et al., 2023) have made limited efforts in ensuring the generated DRSs with well-formed strcutures.

In order to alleviate the problem of generating ill-formed semantic representation, we propose a score-based selection method with the help of beam search algorithms. The beam search algorithm is applied to generate a batch of DRS candidates together with scores. We rank the candidates according to the scores and select the well-formed DRSs as the final outputs.

The experiments are conducted on the Parallel Meaning Bank (PMB; Abzianidze et al. 2017), a standard benchmark for DRS parsing. We specifically transform DRSs into their graph format in the experiments, focusing on DRS graph parsing. The results of the experiments demonstrate that the final models using the score-based selection method, exhibit an improved capability to generate well-formed DRSs compared to the baseline models. Additionally, the models achieve a substantial increase in F1 scores on the test data. The key contributions of this work can be summarized as follows:

LREC-COLING 2024, pages 15037–15043

- Our primary focus centers on the task of well-formed DRS parsing. To achieve this, we propose score-based method with the help of the beam search algorithm, a commonly used technique in sequence generation tasks, to generate a greater number of well-formed DRS representations.
- The models equipped with the score-based selection method yield significant improvements in F1 scores for DRS graph parsing, primarily through the mitigation of ill-formed DRSs.
- The proposed method enhances the robustness of DRS parsers, leading to a substantial reduction in ill-formed DRSs in multilingual zero-shot DRS parsing.

We release the detailed experimental settings and code that are available at https://github.com/LeonCrashCode/DRS-Cross-Lingual-Training.

2. Related Works

2.1. Semantic Parsing

Semantic parsing is a well-established task with diverse applications (Woods, 1973). Semantic representations can be broadly categorized into task-specific and general-purpose formalisms. Task-specific semantic parsers primarily depend on lexical semantics, hand-crafted templates, and features specific to the task (Dong and Lapata, 2016; Alvarez-Melis and Jaakkola, 2017; Cheng et al., 2017; Dong and Lapata, 2018; Cai et al., 2021; Scholak et al., 2021; Qi et al., 2022; Li et al., 2023; Ni et al., 2023; Nguyen et al., 2023). In contrast, general-purpose semantic parsers are based on linguistic features and are designed to handle a wide range of linguistic phenomena, such as semantic roles, presuppositions, and anaphora (Damonte et al., 2017; Konstas et al., 2017; Lyu and Titov, 2018; Liu et al., 2018; van Noord et al., 2018; Liu et al., 2019a; van Noord et al., 2019, 2020; Cai and Lam, 2020; Poelman et al., 2022; Lee et al., 2022; Bai et al., 2022; Drozdov et al., 2022; Wang et al., 2023; Vasylenko et al., 2023; Martínez Lorenzo et al., 2023).

2.2. Well-Formed Discourse Representation Structure Parsing

The early work on DRS parsing introduced Boxer (Bos, 2008), an open-domain semantic parser that generates DRS representations in box form by leveraging the syntactic analysis provided by a robust CCG parser (Curran et al., 2007). Due to the constraints imposed by CCG derivations, this parser tends to produce well-formed DRS, although its accuracy is relatively low. With the advent of neural networks, recent DRS parsing models predominantly rely on sequence-to-sequence architectures. For instance, Liu et al. (2018) employed a finite-state machine to govern sequential generation, aiming to prevent the generation of invalid tokens during inference.

Algorithm 1 Score-Based Selection

Input: $\{(Y^j, s^j)\}^n$
Output: $Y^* = \text{NULL}$
$[Y^0, Y^1, \ldots, Y^n]$ = $\textsc{Sort}(\{(Y^j, s^j)\}^n)$
for $j \leftarrow 1$ to n **do**
  if $\textsc{IsWellFormed}(Y^j)$ **then**
    $Y^* \leftarrow Y^j$
    Break
  end if
end for
if Y^* is NULL **then**
  $Y^* \leftarrow Y^0$
end if

3. Method

In this section, we first present the base model employed for building a DRS parser. Then, we implement the beam search algorithm to generate output candidates and introduce the score-based method, which is used to choose the well-formed candidate as the final output.

3.1. Models

Following the previous work (Wang et al., 2023), DRS graph parsing is modelled as sequence-to-sequence generation task. We adopt pre-trained language models as our baseline models. The input text is a sequence of words, $X = [x_0, x_1, \ldots, x_n]$, where n is the length of the input words. The trained model generate a sequence of DRS symbols, $Y = [y_0, y_1, \ldots, y_m]$, where m is the length of the outputted sequence.

3.2. Candidate Generation

The models are trained using input text sequences paired with the corresponding gold DRS symbols. After training, we adopt the beam search algorithm to generate a set of output candidates. In the ith step of this process, a set of candidate tokens T_i is predicted:

$$T_i = \textsc{BeamSearch}(T_{i-1}, X), \quad (1)$$

where $|T_i| = |T_{i-1}| = n$, n is the beam size. At the end of the inference process, we obtain a set of sequences Y^n, where $Y^j = [t_0^j, t_1^j, \ldots, t_m^j]$ and $t_i^j \in T_i$ is a candidate token. Each sequence Y^j

is associated with a score, denoted as s^j, which is computed by averaging the scores of all the tokens within the sequence:

$$s^j = \frac{1}{m}\sum_{i=0}^{m} s_i^j, \quad (2)$$

where s_i^j is the score assigned to token t_i^j, provided by the beam search algorithm.

3.3. Score-Based Selection

Algorithm 1 show the score-based selection approach. This algorithm operates on a set of candidate sequences, each associated with corresponding scores, $\{(Y^j, s^j)\}^n$. First, we sort the set of candidate sequences by their scores, ordering them from highest to lowest using the function SORT. Then, we iterate through the candidates and assess whether they are well-formed or not using the function ISWELLFORMED to obtain the first well-formed sequence chosen as the final output.[1] If all candidates are ill-formed, we choose the one with the highest score by default.

4. Experiments

We conduct experiments to investigate the effectiveness of the score-based method along with the beam search.

4.1. Experimental Settings

Data. We conducted our experiments on the standard benchmark: the Parallel Meaning Bank (PMB; Abzianidze et al. 2017) in versions 4.0.0. These benchmarks include English, German, Italian, and Dutch data annotated with DRS graphs. The data is categorized into gold and silver subsets. The gold data is annotated by experts, while the silver data is automatically annotated and corrected by humans. Following the previous work, we divided the data into train, development, and test sets.

4.2. Results

Models. The baseline model is built on the mT0 pre-trained language model (Muennighoff et al., 2023) with the model card `mT0-large`.[2] A summary of the models used in our experiments is shown below:

- **mt0-g** The model is fine-tuned with the gold train data and utilizes a greedy inference strategy for parsing.
- **mt0-sg** The model is initially fine-tuned on a combination of silver data and gold train data. Following this, the model is further fine-tuned on the gold train data and utilizes a greedy inference strategy for parsing.
- **w/ bs** Given the well-trained model, the beam search algorithm is utilized to produce a set of candidate outputs, and the candidate with the highest score is selected as the final output.
- **w/ sbs** Given the well-trained model, the beam search algorithm is employed to generate a set of candidate outputs, and the score-based selection method is used to select well-formed outputs.

		dev		test	
		F1 ↑	IF ↓	F1 ↑	IF ↓
mt0-g		95.16	1.28	95.33	1.04
w/ bs	4	95.24	1.34	95.47	0.95
	8	95.24	1.34	95.47	0.95
	16	95.24	1.34	95.47	0.95
	32	95.24	1.34	95.47	0.95
w/ sbs	4	95.76	0.29	96.09	**0.00**
	8	95.76	0.29	96.09	**0.00**
	16	95.85	0.19	96.09	**0.00**
	32	95.85	0.19	96.09	**0.00**
mt0-sg		95.81	1.03	95.93	0.57
w/ bs	4	95.86	0.95	95.94	0.48
	8	95.86	0.95	95.94	0.48
	16	95.86	0.95	95.94	0.48
	32	95.86	0.95	95.94	0.48
w/ sbs	4	96.19	0.19	96.34	0.09
	8	96.19	0.19	96.20	0.09
	16	**96.22**	**0.10**	**96.40**	**0.00**
	32	**96.22**	**0.10**	96.26	**0.00**

Table 1: Results on development and test data for English DRS graph parsing with various beam sizes. The IF (%) column is the percentage of ill-formed outputs. The best scores are highlighted in bold.

Training. The models are fine-tuned with an initial learning rate of 0.001 in 100 epochs. The optimizer used is AdamW (Loshchilov and Hutter, 2019), along with a linear learning rate scheduler. The batch size is 8.

Evaluation Metrics. We follow the previous work (Wang et al., 2023) to use the F1 score calculated by SMATCH (Cai and Knight, 2013) to evaluate

[1] The function ISWELLFORMED is realized according to the ud-boxer at `https://github.com/WPoelman/ud-boxer/blob/master/ud_boxer/sbn.py`

[2] `https://huggingface.co/bigscience/mt0-large`

		de				it				nl			
		dev		test		dev		test		dev		test	
		F1 ↑	IF ↓	F1 ↑	IF ↓	F1 ↑	IF ↓	F1 ↑	IF ↓	F1 ↑	IF ↓	F1 ↑	IF ↓
mt0-g		85.84	4.83	86.27	3.66	84.92	6.30	86.55	3.90	86.35	2.52	85.88	3.46
w/ bs	8	86.90	3.94	86.88	3.11	85.90	5.00	87.32	3.90	87.09	2.06	86.43	3.05
	16	86.92	3.94	86.89	3.11	85.95	5.00	87.30	3.69	87.10	2.06	86.44	3.05
	32	86.92	3.94	86.90	3.11	85.93	5.00	87.28	3.69	87.12	2.06	86.44	3.05
w/ sbs	8	88.38	0.72	88.28	0.37	87.92	0.56	88.68	0.65	87.73	0.23	87.69	0.61
	16	88.49	0.54	88.28	0.37	**88.03**	0.37	88.79	0.43	87.74	0.23	87.78	0.41
	32	**88.73**	**0.18**	**88.39**	**0.00**	88.00	0.37	**88.86**	0.22	**87.90**	**0.00**	**87.90**	**0.00**
mt0-sg		82.70	1.43	82.94	1.83	76.68	4.07	78.24	3.04	83.86	0.69	82.54	2.65
w/ bs	8	82.63	1.79	83.52	1.10	77.35	2.41	78.56	3.25	83.59	1.14	82.90	2.04
	16	82.64	2.15	83.48	1.10	77.32	2.41	78.56	3.25	83.55	1.14	82.69	2.24
	32	82.55	1.97	83.46	1.10	77.32	2.41	78.56	3.25	83.55	1.14	82.92	2.04
w/ sbs	8	83.34	0.72	83.93	**0.00**	78.61	0.37	79.87	**0.00**	84.07	**0.00**	83.52	0.41
	16	83.55	0.72	83.89	**0.00**	78.63	0.19	79.90	**0.00**	84.03	**0.00**	83.61	0.20
	32	83.31	0.72	83.52	0.18	78.68	**0.00**	79.90	**0.00**	84.04	**0.00**	83.70	0.20

Table 2: Results on development and test data for DRS graph parsing for German (de), Italian (it) and Dutch (nl) in zero-shot settings. The IF (%) column is the percentage of ill-formed outputs. The best scores are highlighted in bold.

	F1 ↑	IF ↓
UD-Box (Poelman et al., 2022)	81.8	**0.0**
N-Boxer (Poelman et al., 2022)	92.5	2.3
MLM (Wang et al., 2023)	94.7	0.3
ours	**96.4**	**0.0**

Table 3: Results on test data for English DRS graph parsing, comparing to the state-of-the-art systems. The IF (%) is the percentage of ill-formed outputs.

DRS parsing . Furthermore, we compute the percentage of ill-formed outputs to measure the quality of the generated DRS graphs.

Engish DRS Parsing. Table 1 shows the development and test results for English DRS parsing. With more training data, mt0-g outperforms mt0-sg by 0.65% and 0.60% F1 scores in development and test data, respectively. Additionally, mt0-sg generates more well-formed DRSs. The models exhibit marginal improvement in F1 scores with the beam search algorithm, and they also reduce the generation of ill-formed DRSs. By using the proposed score-based selection method, the models achieve the best results and significantly reduce the number of ill-formed DRSs. In the test data, all the DRSs generated by mt0-sg with score-based selection are well-formed.

Multilingual Zero-shot DRS Parsing. Table 2 shows the development and test results of multilingual DRS parsing in zero-shot settings, where we use a model trained on English data without any additional language-specific training data to parse texts in other languages. Interestingly, with more training data (silver), mt0-sg underperforms mt0-g. One reason for this is that zero-shot parsing sensitive to the quality of the supervision, and zero-shot parsers suffers from the potentially negative impact of low-quality data (silver). The models using beam search algorithms (w/bs) exhibit marginal improvements in F1 score but do not succeed in reducing the number of ill-formed DRSs, and, as the batch size increases, the number of generated ill-formed DRSs does not decrease. On the other hand, the models equipped with the score-based selection method (w/ sbs) significantly prevent the generation of ill-formed DRSs.

Comparison with Previous Works. According to the results in Table 1, we select mt0-sg using score-based selection with a beam size of 16 as our final model. As shownin Table 3, our final model attains state-of-the-art results, achieving an F1 score of 96.40% and generating well-formed DRSs (0.0% ill-formed DRSs).

5. Conclusions

We introduced a score-based selection method for semantic parsing, with the goal of generating well-formed semantic structures. The experiments conducted on DRS parsing show that the models equipped with this method not only achieve significant improvements in F1 scores but also generate more well-formed structures, both in monolingual supervised settings and multilingual zero-shot settings. Moreover, our final model attains state-of-the-art results in the standard benchmarks of DRS parsing.

6. acknowledgments

This work was supported by the Natural Science Foundation of Yunnan Province of China (No. 202301CF070086) and the Open Project Program of Yunnan Key Laboratory of Intelligent Systems and Computing (No. ISC23Z01, 202205AG070003).

7. Bibliographical References

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.

David Alvarez-Melis and Tommi S. Jaakkola. 2017. Tree-structured decoding with doubly-recurrent neural networks. In *Proceedings of the 5th International Conference on Learning Representation (ICLR)*, Toulon, France.

Xuefeng Bai, Yulong Chen, and Yue Zhang. 2022. Graph pre-training for AMR parsing and generation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6001–6015, Dublin, Ireland. Association for Computational Linguistics.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract Meaning Representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria. Association for Computational Linguistics.

Johan Bos. 2008. Wide-coverage semantic analysis with Boxer. In *Semantics in Text Processing. STEP 2008 Conference Proceedings*, pages 277–286. College Publications.

Deng Cai and Wai Lam. 2020. AMR parsing via graph-sequence iterative inference. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1290–1301, Online. Association for Computational Linguistics.

Ruichu Cai, Jinjie Yuan, Boyan Xu, and Zhifeng Hao. 2021. Sadga: Structure-aware dual graph aggregation network for text-to-sql. *Advances in Neural Information Processing Systems*, 34:7664–7676.

Shu Cai and Kevin Knight. 2013. Smatch: an evaluation metric for semantic feature structures. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 748–752, Sofia, Bulgaria. Association for Computational Linguistics.

Jianpeng Cheng, Siva Reddy, Vijay Saraswat, and Mirella Lapata. 2017. Learning structured natural language representations for semantic parsing. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, Canada*, pages 44–55.

James Curran, Stephen Clark, and Johan Bos. 2007. Linguistically motivated large-scale nlp with C&C and Boxer. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 33–36, Prague, Czech Republic.

Marco Damonte, Shay B. Cohen, and Giorgio Satta. 2017. An incremental parser for Abstract Meaning Representation. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 536–546, Valencia, Spain. Association for Computational Linguistics.

Li Dong and Mirella Lapata. 2016. Language to logical form with neural attention. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany.*, pages 33–43.

Li Dong and Mirella Lapata. 2018. Coarse-to-fine decoding for neural semantic parsing. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 731–742, Melbourne, Australia. Association for Computational Linguistics.

Andrew Drozdov, Jiawei Zhou, Radu Florian, Andrew McCallum, Tahira Naseem, Yoon Kim, and Ramón Astudillo. 2022. Inducing and using alignments for transition-based AMR parsing. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1086–1098, Seattle, United States. Association for Computational Linguistics.

Hans Kamp and Uwe Reyle. 1993. *From Discourse to Logic; An Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Kluwer, Dordrecht.

Ioannis Konstas, Srinivasan Iyer, Mark Yatskar, Yejin Choi, and Luke Zettlemoyer. 2017. Neural AMR: Sequence-to-sequence models for parsing and generation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 146–157, Vancouver, Canada. Association for Computational Linguistics.

Young-Suk Lee, Ramón Astudillo, Hoang Thanh Lam, Tahira Naseem, Radu Florian, and Salim Roukos. 2022. Maximum Bayes Smatch ensemble distillation for AMR parsing. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5379–5392, Seattle, United States. Association for Computational Linguistics.

Haoyang Li, Jing Zhang, Cuiping Li, and Hong Chen. 2023. Resdsql: Decoupling schema linking and skeleton parsing for text-to-sql. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 13067–13075.

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.

Jiangming Liu, Shay B. Cohen, and Mirella Lapata. 2019a. Discourse representation parsing for sentences and documents. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 6248–6262, Florence, Italy. Association for Computational Linguistics.

Jiangming Liu, Shay B. Cohen, and Mirella Lapata. 2019b. Discourse representation structure parsing with recurrent neural networks and the transformer model. In *Proceedings of the IWCS Shared Task on Semantic Parsing*, Gothenburg, Sweden. Association for Computational Linguistics.

Jiangming Liu, Shay B. Cohen, Mirella Lapata, and Johan Bos. 2021. Universal discourse representation structure parsing. *Computational Linguistics*, 47(2):445–476.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. In *International Conference on Learning Representations*.

Chunchuan Lyu and Ivan Titov. 2018. AMR parsing as graph prediction with latent alignment. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 397–407, Melbourne, Australia. Association for Computational Linguistics.

Abelardo Carlos Martínez Lorenzo, Pere Lluís Huguet Cabot, and Roberto Navigli. 2023. AMRs assemble! learning to ensemble with autoregressive models for AMR parsing. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1595–1605, Toronto, Canada. Association for Computational Linguistics.

Niklas Muennighoff, Thomas Wang, Lintang Sutawika, Adam Roberts, Stella Biderman, Teven Le Scao, M Saiful Bari, Sheng Shen, Zheng Xin Yong, Hailey Schoelkopf, Xiangru Tang, Dragomir Radev, Alham Fikri Aji, Khalid Almubarak, Samuel Albanie, Zaid Alyafeai, Albert Webson, Edward Raff, and Colin Raffel. 2023. Crosslingual generalization through multitask finetuning. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15991–16111, Toronto, Canada. Association for Computational Linguistics.

Phuong Minh Nguyen, Tung Le, Huy Tien Nguyen, Vu Tran, and Minh Le Nguyen. 2023. Phrasetransformer: an incorporation of local context information into sequence-to-sequence semantic parsing. *Applied Intelligence*, 53(12):15889–15908.

Ansong Ni, Srini Iyer, Dragomir Radev, Veselin Stoyanov, Wen-tau Yih, Sida Wang, and Xi Victoria Lin. 2023. Lever: Learning to verify language-to-code generation with execution. In *International Conference on Machine Learning*, pages 26106–26128. PMLR.

Wessel Poelman, Rik van Noord, and Johan Bos. 2022. Transparent semantic parsing with Universal Dependencies using graph transformations.

In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4186–4192, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Jiexing Qi, Jingyao Tang, Ziwei He, Xiangpeng Wan, Yu Cheng, Chenghu Zhou, Xinbing Wang, Quanshi Zhang, and Zhouhan Lin. 2022. RASAT: Integrating relational structures into pretrained Seq2Seq model for text-to-SQL. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 3215–3229, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Torsten Scholak, Nathan Schucher, and Dzmitry Bahdanau. 2021. PICARD: Parsing incrementally for constrained auto-regressive decoding from language models. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9895–9901, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Rik van Noord, Lasha Abzianidze, Antonio Toral, and Johan Bos. 2018. Exploring neural methods for parsing discourse representation structures. *Transactions of the Association for Computational Linguistics*, 6:619–633.

Rik van Noord, Antonio Toral, and Johan Bos. 2019. Linguistic information in neural semantic parsing with multiple encoders. In *Proceedings of the 13th International Conference on Computational Semantics - Short Papers*, pages 24–31, Gothenburg, Sweden. Association for Computational Linguistics.

Rik van Noord, Antonio Toral, and Johan Bos. 2020. Character-level representations improve DRS-based semantic parsing even in the age of BERT. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4587–4603, Online. Association for Computational Linguistics.

Pavlo Vasylenko, Pere Lluís Huguet Cabot, Abelardo Carlos Martínez Lorenzo, and Roberto Navigli. 2023. Incorporating graph information in transformer-based AMR parsing. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 1995–2011, Toronto, Canada. Association for Computational Linguistics.

Chunliu Wang, Huiyuan Lai, Malvina Nissim, and Johan Bos. 2023. Pre-trained language-meaning models for multilingual parsing and generation. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 5586–5600, Toronto, Canada. Association for Computational Linguistics.

Chunliu Wang, Rik van Noord, Arianna Bisazza, and Johan Bos. 2021. Input representations for parsing discourse representation structures: Comparing English with Chinese. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 767–775, Online. Association for Computational Linguistics.

William A. Woods. 1973. Progress in natural language understanding: an application to lunar geology. In *Proceedings of the National Computer Conference and Exposition, New York, USA*, pages 441–450. ACM.

Luke S. Zettlemoyer and Michael Collins. 2005. Learning to map sentences to logical form: Structured classification with probabilistic categorial grammars. In *PProceedings of the 21st Conference in Uncertainty in Artificial Intelligence*, pages 658–666, Edinburgh, Scotland, UK.