

A Pilot Study of Text-to-SQL Semantic Parsing for Vietnamese

Anh Tuan Nguyen^{1,*}, Mai Hoang Dao² and Dat Quoc Nguyen²

¹NVIDIA, USA; ²VinAI Research, Vietnam

tuananh@nvidia.com, {v.maidh3, v.datnq9}@vinai.io

Abstract

Semantic parsing is an important NLP task. However, Vietnamese is a low-resource language in this research area. In this paper, we present the *first* public large-scale Text-to-SQL semantic parsing dataset for Vietnamese. We extend and evaluate two strong semantic parsing baselines EditSQL (Zhang et al., 2019) and IRNet (Guo et al., 2019) on our dataset. We compare the two baselines with key configurations and find that: automatic Vietnamese word segmentation improves the parsing results of both baselines; the normalized pointwise mutual information (NPMI) score (Bouma, 2009) is useful for schema linking; latent syntactic features extracted from a neural dependency parser for Vietnamese also improve the results; and the monolingual language model PhoBERT for Vietnamese (Nguyen and Nguyen, 2020) helps produce higher performances than the recent best multilingual language model XLM-R (Conneau et al., 2020).

1 Introduction

Semantic parsing is the task of converting natural language sentences into meaning representations such as logical forms or standard SQL database queries (Mooney, 2007), which serves as an important component in many NLP systems such as Question answering and Task-oriented dialogue (Androustopoulos et al., 1995; Moldovan et al., 2003; Guo et al., 2018). The significant availability of the world’s knowledge stored in relational databases leads to the creation of large-scale Text-to-SQL datasets, such as WikiSQL (Zhong et al., 2017) and Spider (Yu et al., 2018), which help boost the development of various state-of-the-art sequence-to-sequence (seq2seq) semantic parsers (Bogin et al., 2019; Zhang et al., 2019; Guo

et al., 2019). Compared to WikiSQL, the Spider dataset presents challenges not only in handling complex questions but also in generalizing to unseen databases during evaluation.

Most SQL semantic parsing benchmarks, such as WikiSQL and Spider, are exclusively for English. Thus the development of semantic parsers has largely been limited to the English language. As SQL is a database interface and universal semantic representation, it is worth investigating the Text-to-SQL semantic parsing task for languages other than English. Especially, the difference in linguistic characteristics could add difficulties in applying seq2seq semantic parsing models to the non-English languages (Min et al., 2019). For example, about 85% of word types in Vietnamese are composed of at least two syllables (Thang et al., 2008). Unlike English, in addition to marking word boundaries, white space is also used to separate syllables that constitute words in Vietnamese written texts. For example, an 8-syllable written text “Có bao nhiêu quốc gia ở châu Âu” (How many countries in Europe) forms 5 words “Có_bao_nhiêu_How_many_quốc_gia_country_ở_in_châu_Âu_Europe”. Thus it is interesting to study the influence of word segmentation in Vietnamese on its SQL parsing, i.e. syllable level vs. word level.

In terms of Vietnamese semantic parsing, previous approaches construct rule templates to convert single database-driven questions into meaning representations (Nguyen and Le, 2008; Nguyen et al., 2009, 2012; Tung et al., 2015; Nguyen et al., 2017). Recently, Vuong et al. (2019) formulate the Text-to-SQL semantic parsing task for Vietnamese as a sequence labeling-based slot filling problem, and then solve it by using a conventional CRF model with handcrafted features, due to the simple structure of the input questions they deal with. Note that seq2seq-based semantic parsers have not yet been explored in any previous work w.r.t. Vietnamese.

*Work done during internship at VinAI Research.

Semantic parsing datasets for Vietnamese include a corpus of 5460 sentences for assigning semantic roles (Phuong et al., 2017) and a small Text-to-SQL dataset of 1258 simple structured questions over 3 databases (Vuong et al., 2019). However, these two datasets are not publicly available for research community.

In this paper, we introduce the *first* public large-scale Text-to-SQL dataset for the Vietnamese semantic parsing task. In particular, we create this dataset by manually translating the Spider dataset into Vietnamese. We empirically evaluate strong seq2seq baseline parsers EditSQL (Zhang et al., 2019) and IRNet (Guo et al., 2019) on our dataset.

Extending the baselines, we extensively investigate key configurations and find that: (1) Our human-translated dataset is far more reliable than a dataset consisting of machine-translated questions, and the overall result obtained for Vietnamese is comparable to that for English. (2) Automatic Vietnamese word segmentation improves the performances of the baselines. (3) The NPMI score (Bouma, 2009) is useful for linking a cell value mentioned in a question to a column in the database schema. (4) Latent syntactic features, which are dumped from a neural dependency parser pre-trained for Vietnamese (Nguyen and Verspoor, 2018), also help improve the performances. (5) Highest improvements are accounted for the use of pre-trained language models, where PhoBERT (Nguyen and Nguyen, 2020) helps produce higher results than XLM-R (Conneau et al., 2020).

We hope that our dataset can serve as a starting point for future Vietnamese semantic parsing research and applications. We publicly release our dataset at: <https://github.com/VinAIRResearch/ViText2SQL>.

2 Our Dataset

We manually translate all English questions and the database schema (i.e. table and column names as well as values in SQL queries) in Spider into Vietnamese. Note that the original Spider dataset consists of 10181 questions with their corresponding 5693 SQL queries over 200 databases. However, only 9691 questions and their corresponding 5263 SQL queries over 166 databases, which are used for training and development, are publicly available. Thus we could only translate those available ones.

The translation work is performed by 1 NLP researcher and 2 computer science students (IELTS

	#Qu.	#SQL	#DB	#T/D	#Easy	#Med.	#Hard	#ExH
all	9691	5263	166	5.3	2233	3439	2095	1924
train	6831	3493	99	5.4	1559	2255	1502	1515
dev	954	589	25	4.2	249	405	191	109
test	1906	1193	42	5.7	425	779	402	300

Table 1: Statistics of our human-translated dataset. “#Qu.”, “#SQL” and “#DB” denote the numbers of questions, SQL queries and databases, respectively. “#T/D” abbreviates the average number of tables per database. “#Easy”, “#Med.”, “#Hard” and “#ExH” denote the numbers of questions categorized by their SQL queries’ hardness levels of “easy”, “medium”, “hard” and “extra hard”, respectively (as defined by Yu et al.).

7.0+). Every question and SQL query pair from the same database is first translated by one student and then cross-checked and corrected by the second student; and finally the NLP researcher verifies the original and corrected versions and makes further revisions if needed. Note that in case we have literal translation for a question, we stick to the style of the original English question as much as possible. Otherwise, for complex questions, we will rephrase them based on the semantic meaning of the corresponding SQL queries to obtain the most natural language questions in Vietnamese.

Following Yu et al. (2018) and Min et al. (2019), we split our dataset into training, development and test sets such that no database overlaps between them, as detailed in Table 1. Examples of question and SQL query pairs from our dataset are presented in Table 2. Note that translated question and SQL query pairs in our dataset are written at the syllable level. To obtain a word-level version of the dataset, we apply RDRSegmenter (Nguyen et al., 2018) from VnCoreNLP (Vu et al., 2018) to perform automatic Vietnamese word segmentation.

Original (Easy question–involving one table in one database): What is the number of cars with more than 4 cylinders? SELECT count(*) FROM CARS_DATA WHERE Cylinders > 4
Translated: Cho biết số lượng những chiếc xe có nhiều hơn 4 xi lanh. SELECT count(*) FROM [dữ liệu xe] WHERE [số lượng xi lanh] > 4
Original (Hard question–with a nested SQL query): Which countries in europe have at least 3 car manufacturers? SELECT T1.CountryName FROM COUNTRIES AS T1 JOIN CONTINENTS AS T2 ON T1.Continent = T2.ContId JOIN CAR_MAKERS AS T3 ON T1.CountryId = T3.Country WHERE T2.Continent = "europe" GROUP BY T1.CountryName HAVING count(*) >= 3
Translated: Những quốc gia nào ở châu Âu có ít nhất 3 nhà sản xuất xe hơi? SELECT T1.[tên quốc gia] FROM [quốc gia] AS T1 JOIN [lục địa] AS T2 ON T1.[lục địa] = T2.[id lục địa] JOIN [nhà sản xuất xe hơi] AS T3 ON T1.[id quốc gia] = T3.[quốc gia] WHERE T2.[lục địa] = "châu Âu" GROUP BY T1.[tên quốc gia] HAVING count(*) >= 3

Table 2: Syllable-level examples. Word segmentation outputs are not shown for simplification.

3 Baseline Models and Extensions

3.1 Baselines

Recent state-of-the-art results on the Spider dataset are reported for RYANSQL (Choi et al., 2020) and RAT-SQL (Wang et al., 2020), which are based on the seq2seq encoder-decoder architectures. However, their implementations are not published at the time of our empirical investigation.¹ Thus we select seq2seq based models EditSQL (Zhang et al., 2019) and IRNet (Guo et al., 2019) with publicly available implementations as our baselines, which produce near state-of-the-art scores on Spider. We briefly describe the baselines EditSQL and IRNet as follows:

- EditSQL is developed for a context-dependent Text-to-SQL parsing task, consisting of: (1) a BiLSTM-based question-table encoder to explicitly encode the question and table schema, (2) a BiLSTM-based interaction encoder with attention to incorporate the recent question history, and (3) a LSTM-based table-aware decoder with attention, taking into account the outputs of both encoders to generate a SQL query.
- IRNet first performs an n-gram matching-based schema linking to identify the columns and the tables mentioned in a question. Then it takes the question, a database schema and the schema linking results as input to synthesize a tree-structured SemQL query—an intermediate representation bridging the input question and a target SQL query. This synthesizing process is performed by using a BiLSTM-based question encoder and an attention-based schema encoder together with a grammar-based LSTM decoder (Yin and Neubig, 2017). Finally, IRNet deterministically uses the synthesized SemQL query to infer the SQL query with domain knowledge.

See Zhang et al. (2019) and Guo et al. (2019) for more details of EditSQL and IRNet, respectively.

3.2 Our Extensions

NPMI for schema linking: IRNet essentially relies on the large-scale knowledge graph ConceptNet (Speer et al., 2017) to link a cell value mentioned in a question to a column in the database schema, based on two ConceptNet categories ‘is a type of’

¹The implementations are still not yet publicly available on 03/06/2020—the EMNLP 2020’s submission deadline.

and ‘related terms’. However, these two ConceptNet categories are not available for Vietnamese. Thus we propose a novel use of the NPMI collocation score (Bouma, 2009) for the schema linking in IRNet, which ranks the NPMI scores between the cell values and column names to match a cell value to its column.

Latent syntactic features: Previous works have shown that syntactic features help improve semantic parsing (Monroe and Wang, 2014; Jie and Lu, 2018). Unlike these works that use handcrafted syntactic features extracted from dependency parse trees, and inspired by Zhang et al. (2017)’s relation extraction work, we investigate whether latent syntactic features, extracted from the BiLSTM-based dependency parser jPTDP (Nguyen and Verspoor, 2018) pre-trained for Vietnamese, would help improve Vietnamese Text-to-SQL parsing. In particular, our approach is that we dump latent feature representations from jPTDP’s BiLSTM encoder given our word-level inputs, and directly use them as part of input embeddings of EditSQL and IRNet.

Pre-trained language models: Zhang et al. (2019) and Guo et al. (2019) make use of BERT (Devlin et al., 2019) to improve their model performances. Thus we also extend EditSQL and IRNet with the use of pre-trained language models XLM-R-base (Conneau et al., 2020) and PhoBERT-base (Nguyen and Nguyen, 2020) for the syllable- and word-level settings, respectively. XLM-R is the recent best multi-lingual model, based on RoBERTa (Liu et al., 2019), pre-trained on a 2.5TB multilingual corpus which contains 137GB of syllable-level Vietnamese texts. PhoBERT is a monolingual variant of RoBERTa for Vietnamese, pre-trained on a 20GB of word-level Vietnamese texts.

4 Experiments

4.1 Experimental Setup

We conduct experiments to study a quantitative comparison between our human-translated dataset and a machine-translated dataset,² the influence of Vietnamese word segmentation (i.e. syllable level and word level), and the usefulness of the latent syntactic features, the pre-trained language models and the NPMI-based approach for schema linking.

For both baselines EditSQL and IRNet which require input pre-trained embeddings for syllables

²We employ a well-known machine translation engine to translate the English questions into Vietnamese.

Approach	Easy	Medium	Hard	ExH	SELECT	WHERE	ORDER BY	GROUP BY	KEYWORDS
EditSQL _{DeP}	65.7	46.1	37.6	16.8	75.1	44.6	65.6	63.2	73.5
EditSQL _{XLM-R}	75.1	56.2	45.3	22.4	82.7	60.3	70.7	67.2	79.8
EditSQL _{PhoBERT}	75.6	58.0	47.4	22.7	83.3	61.8	72.5	67.9	80.6
IRNet _{DeP}	71.8	51.5	47.4	18.5	79.3	48.7	71.8	63.4	74.3
IRNet _{XLM-R}	76.2	57.8	46.8	23.5	83.5	59.1	74.4	68.3	80.5
IRNet _{PhoBERT}	76.8	57.5	47.2	24.8	84.5	59.3	76.6	68.2	80.3

Table 4: Exact matching accuracy categorized by 4 different hardness levels, and F_1 scores of different SQL components on the test set. “ExH” abbreviates Extra Hard.

	Approach	dev	test	Approach	dev	test
Vi-Syllable	EditSQL [MT]	21.5	16.8	IRNet [MT]	25.4	20.3
	EditSQL	28.6	24.1	IRNet	43.3	38.2
	EditSQL _{XLM-R}	55.2	51.3	IRNet _{XLM-R}	58.6	52.8
	EditSQL [MT]	22.8	17.4	IRNet [MT]	27.4	21.6
Vi-Word	EditSQL	33.7	30.2	IRNet	49.7	43.6
	EditSQL _{DeP}	45.3	42.2	IRNet _{DeP}	52.2	47.1
	EditSQL _{PhoBERT}	56.7	52.6	IRNet _{PhoBERT}	60.2	53.2
En	EditSQL _{RoBERTa}	58.3	53.6	IRNet _{RoBERTa}	63.8	55.3

Table 3: Exact matching accuracies of EditSQL and IRNet. “Vi-Syllable” and “Vi-Word” denote the results w.r.t. the syllable level and the word level, respectively. [MT] denotes accuracy results with the machine-translated questions. The subscript “DeP” refers to the use of the latent syntactic features. Other subscripts denote the use of the pre-trained language models. “En” denotes our results on the English Spider dataset but under our training/development/test split w.r.t. the total 9691 public available questions.

and words, we pre-train a set of 300-dimensional syllable embeddings and another set of 300-dimensional word embeddings using the Word2Vec skip gram model (Mikolov et al., 2013) on syllable- and word-level corpora of 20GB Vietnamese texts (Nguyen and Nguyen, 2020). In addition, we also use these 20GB syllable- and word-level Vietnamese corpora as our external datasets to compute the NPMI score (with a window size of 20) for schema linking in IRNet.

Our hyperparameters for EditSQL and IRNet are taken from Zhang et al. (2019) and Guo et al. (2019), respectively. The pre-trained syllable and word embeddings are fixed, while the pre-trained language models XLM-R and PhoBERT are finetuned during training.

Following Yu et al. (2018), we use two commonly used metrics for evaluation. The first one is the exact matching accuracy, which reports the percentage of input questions that have exactly the same SQL output as its gold reference. The second one is the component matching F_1 , which reports F_1 scores for SELECT, WHERE, ORDER

BY, GROUP BY and all other keywords.

We run for 10 training epochs and evaluate the exact matching accuracy after each epoch on the development set, and then select the best model checkpoint to report the final result on the test set.

4.2 Main Results

Table 3 shows the overall exact matching results of EditSQL and IRNet on the development and test sets. Clearly, IRNet does better than EditSQL, which is consistent with results obtained on the original English Spider dataset.

We find that our human-translated dataset is far more reliable than a dataset consisting of machine-translated questions. In particular, at the word level, compared to the machine-translated dataset, our dataset obtains about 30.2-17.4 \approx 13% and 43.6-21.6 = 22% absolute improvements in accuracies of EditSQL and IRNet, respectively (i.e. 75%–100% relative improvements). In addition, the word-based Text-to-SQL parsing obtains about 5+% absolute higher accuracies than the syllable-based Text-to-SQL parsing (EditSQL: 24.1% \rightarrow 30.2% ; IRNet: 38.2% \rightarrow 43.6%), i.e. automatic Vietnamese word segmentation improves the accuracy results.

Furthermore, latent syntactic features dumped from the pre-trained dependency parser jPTDP for Vietnamese help improve the performances of the baselines (EditSQL: 30.2% \rightarrow 42.2%; IRNet: 43.6% \rightarrow 47.1%). Also, biggest improvements are accounted for the use of pre-trained language models. In particular, PhoBERT helps produce higher results than XLM-R (EditSQL: 52.6% vs. 51.3%; IRNet: 53.2% vs. 52.8%).

We also retrain EditSQL and IRNet on the English Spider dataset with the use of the strong pre-trained language model RoBERTa instead of BERT, but under our dataset split. We find that the overall results for Vietnamese are smaller but comparable to the English results. Therefore, Text-to-SQL semantic parsing for Vietnamese might not be significantly more challenging than that for English.

Table 4 shows the exact matching accuracies of EditSQL and IRNet w.r.t. different hardness levels of SQL queries and the F_1 scores w.r.t. different SQL components on the test set. Clearly, in most cases, the pre-trained language models PhoBERT and XLM-R help produce substantially higher results than the latent syntactic features, especially for the WHERE component.

NPMI-based schema linking: We also investigate the contribution of our NPMI-based extension approach for schema linking in applying IRNet for Vietnamese. Without using NPMI for schema linking,³ we observe 6+% absolute decrease in the exact matching accuracies of IRNet on both development and test sets, thus showing the usefulness of our NPMI-based approach for schema linking.

4.3 Error Analysis

To understand the source of errors, we perform an error analysis on the development set which consists of 954 questions. Using IRNet_{PhoBERT} which produces the best result, we identify several causes of errors from 382/954 failed examples.

For 121/382 cases (32%), IRNet_{PhoBERT} makes incorrect predictions on the column names which are not mentioned or only partially mentioned in the questions. For example, given the question “Hiển thị tên và năm phát hành của những bài hát thuộc về ca sĩ trẻ tuổi nhất” (Show the name and the release year of the song by the youngest singer),⁴ the model produces an incorrect column name prediction of “tên” (name) instead of the correct one “tên bài hát” (song name). Errors related to column name predictions can either be missing the entire column names or inserting random column names into the WHERE component of the predicted SQL queries.

About 12% of failed examples (47/382) in fact have an equivalent implementation of their intent with a different SQL syntax. For example, the model produces a ‘failed’ SQL output “SELECT MAX [sức chứa] FROM [sân vận động]” which is equivalent to the gold SQL query of “SELECT [sức chứa] FROM [sân vận động] ORDER BY [sức chứa] DESC LIMIT 1”, i.e. the SQL output would be valid if we measure an execution accuracy.

About 22% of failed examples (84/382) are caused by nested and complex SQL queries which mostly belong to the Extra Hard category. With

³Without schema linking, IRNet assigns a ‘NONE’ type for column names.

⁴Word segmentation is not shown for simplification.

18% of failed examples (70/382), incorrectly predicting operators is another common type of errors. For example, given the phrases “già nhất” (oldest) and “trẻ nhất” (youngest) in the question, the model fails to predict the correct operators `max` and `min`, respectively. The remaining 60/382 cases (16%) are accounted for an incorrect prediction of table names in a FROM clause.

5 Conclusion

In this paper, we have presented the first public large-scale dataset for Vietnamese Text-to-SQL semantic parsing. We also extensively experiment with key research configurations using two strong baseline models on our dataset and find that: the input representations, the NPMI-based approach for schema linking, the latent syntactic features and the pre-trained language models all have the influence on this Vietnamese-specific task. We hope that our dataset can serve as the starting point for further research and applications in Vietnamese question answering and dialogue systems.

References

- I. Androustopoulos, G.D. Ritchie, and P. Thanisch. 1995. [Natural language interfaces to databases – an introduction](#). *Natural Language Engineering*, 1(1):29–81.
- Ben Bogin, Matt Gardner, and Jonathan Berant. 2019. Global Reasoning over Database Structures for Text-to-SQL Parsing. In *Proceedings of EMNLP-IJCNLP*, pages 3659–3664.
- G. Bouma. 2009. Normalized (pointwise) mutual information in collocation extraction. In *Proceedings of the Biennial GSCL Conference*, pages 31–40.
- DongHyun Choi, Myeong Cheol Shin, EungGyun Kim, and Dong Ryeol Shin. 2020. RYANSQL: Recursively Applying Sketch-based Slot Fillings for Complex Text-to-SQL in Cross-Domain Databases. *arXiv preprint*, arXiv:2004.03125.
- Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised Cross-lingual Representation Learning at Scale. In *Proceedings of ACL*, pages 8440–8451.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, pages 4171–4186.

- Daya Guo, Duyu Tang, Nan Duan, Ming Zhou, and Jian Yin. 2018. Dialog-to-Action: Conversational Question Answering Over a Large-Scale Knowledge Base. In *NIPS*, pages 2942–2951.
- Jiaqi Guo, Zecheng Zhan, Yan Gao, Yan Xiao, Jian-Guang Lou, Ting Liu, and Dongmei Zhang. 2019. Towards Complex Text-to-SQL in Cross-Domain Database with Intermediate Representation. In *Proceedings of ACL*, pages 4524–4535.
- Zhanming Jie and Wei Lu. 2018. Dependency-based Hybrid Trees for Semantic Parsing. In *Proceedings of EMNLP*, pages 2431–2441.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv preprint*, arXiv:1907.11692.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Advances in Neural Information Processing Systems 26*, pages 3111–3119.
- Qingkai Min, Yuefeng Shi, and Yue Zhang. 2019. A Pilot Study for Chinese SQL Semantic Parsing. In *Proceedings of EMNLP-IJCNLP*, pages 3652–3658.
- Dan Moldovan, Christine Clark, Sanda Harabagiu, and Steve Maorano. 2003. COGEX: A Logic Prover for Question Answering. In *Proceedings of HLT-NAACL*, pages 166–172.
- Will Monroe and Yushi Wang. 2014. [Dependency Parsing Features for Semantic Parsing](#).
- Raymond J. Mooney. 2007. Learning for semantic parsing. In *Proceedings of CICLing*, pages 311–324.
- Anh Kim Nguyen and Huong Thanh Le. 2008. Natural Language Interface Construction Using Semantic Grammars. In *Proceedings of PRICAI*, pages 728–739.
- Dai Quoc Nguyen, Dat Quoc Nguyen, and Son Bao Pham. 2009. A Vietnamese Question Answering System. In *Proceedings of KSE*, pages 26–32.
- Dai Quoc Nguyen, Dat Quoc Nguyen, and Son Bao Pham. 2012. A Semantic Approach for Question Analysis. In *Proceedings of IEA/AIE*, pages 156–165.
- Dat Quoc Nguyen and Anh Tuan Nguyen. 2020. PhoBERT: Pre-trained language models for Vietnamese. *arXiv preprint*, arXiv:2003.00744.
- Dat Quoc Nguyen, Dai Quoc Nguyen, and Son Bao Pham. 2017. Ripple Down Rules for Question Answering. *Semantic Web*, 8(4):511–532.
- Dat Quoc Nguyen, Dai Quoc Nguyen, Thanh Vu, Mark Dras, and Mark Johnson. 2018. A Fast and Accurate Vietnamese Word Segmenter. In *Proceedings of LREC*, pages 2582–2587.
- Dat Quoc Nguyen and Karin Verspoor. 2018. An improved neural network model for joint POS tagging and dependency parsing. In *Proceedings of the CoNLL 2018 Shared Task*, pages 81–91.
- Le Hong Phuong, Pham Hoang, Pham Khoai, Nguyen Huyen, Nguyen Luong, and Nguyen Hiep. 2017. Vietnamese Semantic Role Labelling. *VNU Journal of Science: Computer Science and Communication Engineering*, 33(2).
- Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Proceedings of AAAI*, pages 4444–4451.
- Dinh Quang Thang, Le Hong Phuong, Nguyen Thi Minh Huyen, Nguyen Cam Tu, Mathias Rossignol, and Vu Xuan Luong. 2008. Word segmentation of Vietnamese texts: a comparison of approaches. In *Proceedings of LREC*, pages 1933–1936.
- Vu Xuan Tung, Le Minh Nguyen, and Duc Tam Hoang. 2015. Semantic Parsing for Vietnamese Question Answering System. In *Proceedings of KSE*, pages 332–335.
- Thanh Vu, Dat Quoc Nguyen, Dai Quoc Nguyen, Mark Dras, and Mark Johnson. 2018. VnCoreNLP: A Vietnamese Natural Language Processing Toolkit. In *Proceedings of NAACL: Demonstrations*, pages 56–60.
- Thi-Hai-Yen Vuong, Thi-Thu-Trang Nguyen, Nhu-Thuat Tran, Le-Minh Nguyen, and Xuan-Hieu Phan. 2019. Learning to Transform Vietnamese Natural Language Queries into SQL Commands. In *Proceedings of KSE*.
- Bailin Wang, Richard Shin, Xiaodong Liu, Oleksandr Polozov, and Matthew Richardson. 2020. RAT-SQL: Relation-Aware Schema Encoding and Linking for Text-to-SQL Parsers. In *Proceedings of ACL*, pages 7567–7578.
- Pengcheng Yin and Graham Neubig. 2017. A Syntactic Neural Model for General-Purpose Code Generation. In *Proceedings of ACL*, pages 440–450.
- Tao Yu, Rui Zhang, Kai Yang, Michihiro Yasunaga, Dongxu Wang, Zifan Li, James Ma, Irene Li, Qingning Yao, Shanelle Roman, Zilin Zhang, and Dragomir R. Radev. 2018. Spider: A Large-Scale Human-Labeled Dataset for Complex and Cross-Domain Semantic Parsing and Text-to-SQL Task. In *Proceedings of EMNLP*, pages 3911–3921.
- Meishan Zhang, Yue Zhang, and Guohong Fu. 2017. End-to-End Neural Relation Extraction with Global Optimization. In *Proceedings of EMNLP*, pages 1730–1740.

Rui Zhang, Tao Yu, Heyang Er, Sungrok Shim, Eric Xue, Xi Victoria Lin, Tianze Shi, Caiming Xiong, Richard Socher, and Dragomir R. Radev. 2019. Editing-Based SQL Query Generation for Cross-Domain Context-Dependent Questions. In *Proceedings of EMNLP-IJCNLP*, pages 5337–5348.

Victor Zhong, Caiming Xiong, and Richard Socher. 2017. Seq2SQL: Generating Structured Queries from Natural Language using Reinforcement Learning. *arXiv preprint*, arXiv:1709.00103.