Research Statement

Rui Zhang (r.zhang@yale.edu)

The natural language is a fundamental form of information and communication. As my ultimate research goal, I aim to design machine learning approaches to build intelligent natural language processing systems that can understand and communicate in different contexts, domains, text styles, and languages. To be specific, my research focuses on building intelligent systems that (1) first understand the contextual meaning of language grounded in various contexts where the language is used and (2) then generate effective language response in different forms for information request and human-computer communication. To this end, during my Ph.D. study, I decompose this long-term goal into the following research questions:

- How to enable machines to understand language at different levels of granularity including entities, sentences, documents, and conversations [1, 2, 3]?
- How to effectively gather information relevant to the user request from multilingual documents in low-resource languages [4, 5]?
- How can agents help people summarize the information in the email, news, and scientific articles [6, 7, 8]?
- How can we synthesize formal programs (e.g., SQL) from the natural language for multi-turn question answering between humans and machines [9, 10, 11]?

The challenges and solutions are inherently multi-disciplinary, spanning a range of areas including natural language processing, deep learning, information retrieval, database, and human-computer interaction. The sections below present my completed projects towards the answers to each question, followed by future directions in which I would like to make progress.

Deep Neural Modeling of Text Units. Natural language consists of text units at different granularity levels such as entity mentions, sentences, and documents, and conversations. The understanding of textual units depends on each other, and small text units make up larger ones to convey meaning. Therefore, modeling text units and dependency relationships among them is a fundamental question for natural language understanding and its applications. In particular, I am interested in how deep learning models can help us better understand and utilize the dependency relationship among text units in different NLP tasks. To be specific, I have designed end-to-end deep neural networks for (1) coreference resolution in documents [2], (2) sentiment analysis and text classification for sentences and documents [1], and (3) addressee and response selection for multi-turn multi-party conversations [3]. The third one is described in detail as follows.

Real-world conversations often involve more than two speakers. In the Ubuntu Internet Relay Chat channel (IRC), for example, one user can initiate a discussion about an Ubuntu-related technical issue, and many other users can work together to solve the problem. Therefore, a multi-party dialog can have complex speaker interactions: at each turn, users play one of three roles (sender, addressee, observer), and those roles vary across turns. In this project, we study the problem of addressee and response selection in multi-party conversations: given a responding speaker and a dialog context, the task is to select an addressee and a response from a set of candidates for the responding speaker. A task example and our model output is given in Table 1 The task requires modeling multi-party conversations and can be directly used to build retrieval-based dialog systems.

To model the complexity of multi-party dialog to solve this task, we introduce the Speaker Interaction Recurrent Neural Network (SI-RNN). SI-RNN redesigns the dialog encoder by updating speaker embeddings in a role-sensitive way. Speaker embeddings are updated in different GRU-based units depending on their roles (sender, addressee, observer). Furthermore, we note that the addressee and response are mutually dependent and view the task as a joint prediction problem. Therefore, SI-RNN models the conditional probability (of addressee given the response and vice versa) and selects the addressee and response pair by maximizing the joint probability. On a public standard benchmark data set, SI-RNN significantly improves the addressee and response selection accuracy, particularly in complex conversations with many speakers and responses to distant messages many turns in the past.

	Sender	Addressee	Utterance
1	codepython	wafflejock	thanks
1	wafflejock	codepython	yup np
2	wafflejock	theoletom	you can use sudo apt-get install packagename – reinstall, to have apt-get install reinstall some package/metapackage and redo the configuration for the program as well
3	codepython	-	i installed ubuntu on a separate external drive. now when i boot into mac, the external drive does not show up as bootable. the blue light is on. any ideas?
4	Guest54977	-	hello there. wondering to anyone who knows, where an ubuntu backup can be retrieved from.
2	theoletom	wafflejock	it's not a program. it's a desktop environment.
4	Guest54977	-	did some searching on my system and googling, but couldn't find an answer
2	theoletom	-	be a trace of it left yet there still is.
2	theoletom	-	i think i might just need a fresh install of ubuntu. if there isn't a way to revert to default settings
5	releaf	-	what's your opinion on a $500 laptop that will be a dedicated ubuntu machine?
5	releaf	-	are any of the pre-loaded ones good deals?
5	releaf	-	if not, are there any laptops that are known for being oem-heavy or otherwise ubuntu friendly?
3	codepython	-	my usb stick shows up as bootable (efi) when i boot my mac. but not my external hard drive on which i just installed ubuntu. how do i make it bootable from mac hardware?
3	Jordan_U	codepython	did you install ubuntu to this external drive from a different machine?
5	Umeaboy	releaf	what country you from?
5	wafflejock		
Model Prediction		Addressee	Response
Direct-Recent+TF-IDF		theoletom	ubuntu install fresh
Dynamic-RNN		codepython	no prime is the replacement
SI-RNN		⋆ releaf	⋆ there are a few ubuntu dedicated laptop providers like umeaboy is asking depends on where you are

Table 1: An example of addressee and response selection in multi-party dialog. SI-RNN chooses to engage in a new sub-conversation by suggesting a solution to "releaf" about Ubuntu dedicated laptops. ⋆ denotes the ground-truth. Sub-conversations are coded with different numbers for the purpose of analysis (sub-conversation labels are not available during training or testing).

Text Summarization in Different Domains and Styles. Summarization is a central problem aiming to produce fluent and coherent synopses covering salient information in the documents. Recently, neural methods have shown promising results in text summarization using both extractive and abstractive approaches. However, most of the work focuses on the single-document setting in the news domain, relying on large training datasets including Gigaword Corpus, the CNN/Daily Mail dataset, the New York Times dataset, and the Newsroom corpus.

My research has been focused on expanding neural summarization methods to various text domains and styles using graph-based representations and reinforcement learning techniques. To be specific, I have worked on summarization in the following scenarios: (1) Using graph convolutional neural networks to summarize multiple long news articles [6], (2) Incorporating citation information for summarizing scientific articles [7], and (3) Generating subject line for personal email messages [8]. The third project is described below.

Email Body: Hi All, I would be grateful if you could get to me today via email a job description for your current role. I would like to get this to the immigration attorneys so that they can finalise the paperwork in preparation for INS filing once the UBS deal is signed. Kind regards,
Subject 1: Current Job Description Needed *(COMMENT: This is good because it is both informative and succinct.)*
Subject 2: Job Description *(COMMENT: This is okay but not informative enough.)*
Subject 3: Request *(COMMENT: This is bad because it does not contain any specific information about the request.)*

Table 2: An email with three possible subject lines.

The email is a ubiquitous form of online communication. An email message consists of two basic elements: an *email subject line* and an *email body*. The subject line, which is displayed to the recipient in the list of inbox messages, should tell what the email body is about and what the sender wants to convey. Table 2 shows an email body with three possible subject lines.

We propose the task of Subject Line Generation (SLG): automatically producing email subjects given the email body. Compared with news headline generation or news single document summarization, email subjects are generally much shorter, which means a system must have the ability to summarize with a high compression ratio. Therefore, we believe this task can also benefit other highly abstractive summarization such as generating section titles for long documents to improve reading comprehension speed and accuracy.

To introduce the task, we build the first dataset, Annotated Enron Subject Line Corpus (AESLC). Furthermore, in order to properly evaluate the subject, we use a combination of automatic metrics from the text summarization and machine translation fields, in addition to building our own regression-based Email Subject Quality Estimator (ESQE). Third, to generate effective email subjects, we propose a method that combines extractive and abstractive summarization using a two-stage process by Multi-Sentence Selection and Rewriting with Email Subject Quality Estimation Reward. The multi-sentence extractor first selects multiple sentences from the input email body. Extracted sentences capture salient information for writing a subject such as named entities and dates. Then, the multisentence abstractor rewrites multiple selected sentences into a succinct subject line while preserving key information. For training the network, we use a multi-stage training strategy incorporating both supervised cross-entropy training and reinforcement learning (RL) by optimizing the reward provided by the ESQE model. Our automatic and human evaluations demonstrate that our model outperforms competitive baselines and approaches human-level quality.

Low-Resource Cross-lingual Information Retrieval. Cross-lingual relevance ranking, or Cross-Lingual Information Retrieval (CLIR), is the task of ranking foreign documents against a user query. As multilingual documents are more accessible, CLIR is increasingly more important whenever the relevant information is in other languages.

Traditional CLIR systems consist of two components: machine translation and monolingual information retrieval. Based on the translation direction, it can be further categorized into the document translation and the query translation approaches. In both cases, we first solve the translation problem, and the task is transformed to the monolingual setting. However, while conceptually simple, the performance of this modular approach is fundamentally limited by the quality of machine translation.

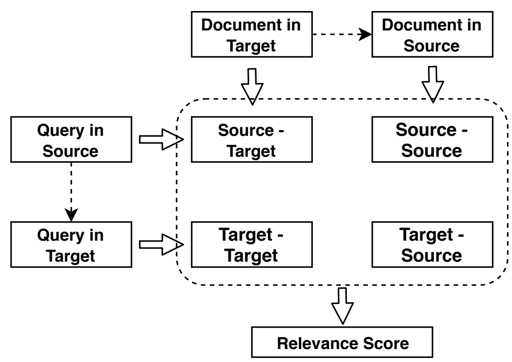

Figure 1: Cross-lingual Relevance Ranking with Bilingual Query and Document Representation.

Recently, many deep neural learning-to-rank models have shown promising results in information retrieval. They learn a scoring function directly from the relevance label of query-document pairs. However, most previous neural IR papers only work with monolingual datasets, primarily for two reasons. First, when queries and documents are in different languages, it is not clear how to measure the similarity of them in distributed representation space. Furthermore, deep neural networks need a large amount of training data to achieve decent performance. The annotation is prohibitively expensive for low-resource language pairs in our cross-lingual case.

In this paper [4], we propose a cross-lingual deep relevance ranking architecture based on a bilingual view of queries and documents. As shown in Figure 1, our model first translates queries and documents and then uses four components to match them in both the source and target language. Each component is implemented as a deep neural network, and the final relevance score combines all components which are jointly trained given the relevance label. We implement this based on state-of-the-art term interaction models because they enable us to make use of cross-lingual embeddings to explicitly encode terms of queries and documents even if they are in different languages. To deal with the small amount of training data, we first perform query likelihood retrieval and include the score as an extra feature in our model. In this way, the model effectively learns to rerank from a small number of relevance labels.

We evaluate our model on the Material CLIR dataset with three language pairs including English to Swahili, English to Tagalog, and English to Somali. Experimental results demonstrate that our model

D_1 : Database about student dormitory containing 5 tables.
C_1 : **Find the first and last names of the students who are living in the dorms that have a TV Lounge as an amenity.**

Q_1 : How many dorms have a TV Lounge?
S_1 : `SELECT COUNT(*) FROM dorm AS T1 JOIN has_amenity AS T2 ON T1.dormid = T2.dormid JOIN dorm_amenity AS T3 ON T2.amenid = T3.amenid WHERE T3.amenity_name = 'TV Lounge'`
Q_2 : What is the total capacity of these dorms?
S_2 : `SELECT SUM(T1.student_capacity) FROM dorm AS T1 JOIN has_amenity AS T2 ON T1.dormid = T2.dormid JOIN dorm_amenity AS T3 ON T2.amenid = T3.amenid WHERE T3.amenity_name = 'TV Lounge'`
Q_3 : How many students are living there?
S_3 : `SELECT COUNT(*) FROM student AS T1 JOIN lives_in AS T2 ON T1.stuid = T2.stuid WHERE T2.dormid IN (SELECT T3.dormid FROM has_amenity AS T3 JOIN dorm_amenity AS T4 ON T3.amenid = T4.amenid WHERE T4.amenity_name = 'TV Lounge')`
Q_4 : Please show their first and last names.
S_4 : `SELECT T1.fname, T1.lname FROM student AS T1 JOIN lives_in AS T2 ON T1.stuid = T2.stuid WHERE T2.dormid IN (SELECT T3.dormid FROM has_amenity AS T3 JOIN dorm_amenity AS T4 ON T3.amenid = T4.amenid WHERE T4.amenity_name = 'TV Lounge')`

D_2 : Database about shipping company containing 13 tables
C_2 : **Find the names of the first 5 customers.**

Q_1 : What is the customer id of the most recent customer?
S_1 : `SELECT customer_id FROM customers ORDER BY date_became_customer DESC LIMIT 1`
Q_2 : What is their name?
S_2 : `SELECT customer_name FROM customers ORDER BY date_became_customer DESC LIMIT 1`
Q_3 : How about for the first 5 customers?
S_3 : `SELECT customer_name FROM customers ORDER BY date_became_customer LIMIT 5`

Figure 2: Two question sequences from SParC. Questions (Q_i) in each sequence query a database (D_m), obtaining information sufficient to complete the interaction goal (C_m). Each question is annotated with a SQL query (S_i). SQL segments from the interaction context are underlined.

D_1 : Database about student dormitories containing 5 tables

Q_1 :	What are the names of all the dorms?	INFORM_SQL
S_1 :	`SELECT dorm_name FROM dorm`	
A_1 :	(Result table with many entries)	
R_1 :	This is the list of the names of all the dorms.	CONFIRM_SQL
Q_2 :	Which of those dorms have a TV lounge?	INFORM_SQL
S_2 :	`SELECT T1.dorm_name FROM dorm AS T1 JOIN has_amenity AS T2 ON T1.dormid = T2.dormid JOIN dorm_amenity AS T3 ON T2.amenid = T3.amenid WHERE T3.amenity_name = 'TV Lounge'`	
A_2 :	(Result table with many entries)	
R_2 :	This shows the names of dorms with TV lounges.	CONFIRM_SQL
Q_3 :	What dorms have no study rooms as amenities?	AMBIGUOUS
R_3 :	Do you mean among those with TV Lounges?	CLARIFY
Q_4 :	Yes.	AFFIRM
S_4 :	`SELECT T1.dorm_name FROM dorm AS T1 JOIN has_amenity AS T2 ON T1.dormid = T2.dormid JOIN dorm_amenity AS T3 ON T2.amenid = T3.amenid WHERE T3.amenity_name = 'TV Lounge' EXCEPT SELECT T1.dorm_name FROM dorm AS T1 JOIN has_amenity AS T2 ON T1.dormid = T2.dormid JOIN dorm_amenity AS T3 ON T2.amenid = T3.amenid WHERE T3.amenity_name = 'Study Room'`	
A_4 :	Fawlty Towers	
R_4 :	Fawlty Towers is the name of the dorm that has a TV lounge but not a study room as an amenity.	CONFIRM_SQL
	...	
Q_8 :	Thanks!	THANK_YOU
R_8 :	You are welcome.	WELCOME

Figure 3: A dialog from CoSQL. Gray boxes separate the user inputs (Q_i) querying the database (D_i) from the SQL queries (S_i), returned answers (A_i), and expert responses (R_i). Users send an input to the expert, who writes the corresponding SQL query (only seen by the expert) if possible and sends an answer and response description back. Dialogue acts are on the right-hand side (e.g., Q_3 is "ambiguous" and R_3 is "clarify").

outperforms other translation-based query likelihood retrieval and monolingual deep relevance ranking approaches. Furthermore, by aligning word embedding spaces for multiple languages, the model can be directly applied under a zero-shot transfer setting when no training data is available for another language pair. We believe the idea of combining bilingual document representations using cross-lingual word embeddings can be generalized to other models as well.

Multi-turn Text-to-SQL Semantic Parsing. Generating SQL queries from user utterances is an important task to help people acquire information from databases. Build such a text-to-SQL semantic parsing system bridges the data and user through an intelligent natural language interface, greatly improving the efficiency of querying relational databases for many users beyond database experts.

In a real-world application, users often access information in a multi-turn interaction with the system by asking a sequence of related questions. Previous studies have shown that by allowing questions to be constructed sequentially, users can explore the data in a more flexible manner, which reduces their cognitive burden and increases their involvement when interacting with the system. The phrasing of such questions depends heavily on the interaction history . The users may explicitly refer to or omit previously mentioned entities and constraints, and may introduce refinements, additions or substitutions to what has already been said. This requires a practical text-to-SQL system to effectively process context information to synthesize the correct SQL logic.

To advance the research progress in this field, we build two multi-turn text-to-SQL data sets:

1. SParC [10] for cross-domain **S**emantic **Par**sing in **C**ontext. It contains 4,298 unique question sequences with 12k+ questions annotated with SQL queries. Examples are in Figure 2.

2. CoSQL [11] for building general-purpose database querying dialogue systems. It consists of 30k+ turns plus 10k+ annotated SQL queries, obtained from a Wizard-of-Oz collection of 3k dialogues. Examples are in Figure 3.

Both of them are built on top of our Spider dataset [12], the largest cross-domain context-independent text-to-SQL dataset available in the field, and thus span 200 complex databases over 138 domains. The large number of domains provide rich contextual phenomena and thematic relations between the questions, which general-purpose natural language interfaces to databases have to address. In addition, it enables us to test the generalization of the trained systems to unseen databases and domains. We are actively maintaining the datasets and leaderboards of our Text-to-SQL Challenge Series including Spider (https://yale-lily.github.io/spider), SParC (https://yale-lily.github.io/sparc), and CoSQL (https://yale-lily.github.io/cosql).

Furthermore, we also propose an editing-based approach for our cross-domain context-dependent text-to-SQL generation task [9]. Based on the observation that adjacent natural language questions are often linguistically dependent and their corresponding SQL queries tend to overlap, we utilize the interaction history by editing the previous predicted query to improve the generation quality. Our editing mechanism views SQL as sequences and reuses generation results at the token level in a simple manner. It is flexible to change individual tokens and robust to error propagation. Furthermore, to deal with complex table structures in different domains, we employ an utterance-table encoder and a table-aware decoder to incorporate the context of the user utterance and the table schema. Experiment results on SParC show that by generating from the previous query, our model delivers an improvement of 7% question match accuracy and 11% interaction match accuracy over the previous state-of-the-art.

Future Research Directions

Even though the past few years have seen tremendous progress in understanding and generating natural language, we have only scratched the surface of what can be achieved in developing intelligent systems. Building upon my past work, I plan to embrace important open challenges by exploring the following new directions.

Grounding Language to Interactive User Actions. Mobile phone applications have been widely used in daily life. To accomplish a job on the mobile phone, the user interacts with an application via a sequence of low-level user interface actions, such as click an element, swipe to check the rest of a list, and type a string. Due to their diversity in content and design, mobile phone applications provide an open-domain learning environment with a rich source of real-world knowledge. We are interested in mapping natural language instructions to sequences of actions in the graphic user interface. In particular, we consider a task setting where the agent only receives a high-level user goal in natural language without step-by-step instructions or human demonstration, and it needs to perform a sequence of user interface actions to accomplish the user goal. This would automate the task completion that currently requires complicated human participation, greatly promoting the efficiency and accessibility for mobile application users. The complex representation of a mobile phone application contains both structured properties (e.g., the internal tree representation of a screen and the spatial relations among elements) and unstructured elements (e.g., text and image icons). Moreover, since we do not provide any human demonstration or step-by-step annotation, the agent needs to reason about the semantics of a high-level user goal and the context of screens, and then learns to perform the correct sequence of actions. Therefore, this presents a challenging natural language grounding and navigation problem in this diverse and open-domain platform.

Multilingual Transfer Learning for Low-resource Languages. Deep neural networks, despite their success to achieve promising results in various natural language processing tasks, still heavily rely on large amounts of in-domain labeled training data. However, only a few languages have large amounts of labeled data, and generalization in low-resource scenarios is still an open challenge. I would like to develop multilingual transfer learning techniques that can leverage annotations in high-resource languages to boost the performance of low-resource languages. I am particularly interested in models that (1) require minimal cross-lingual supervision, (2) leverage knowledge from multiple source languages, and (3) quickly adapt to the target language and task.

Natural Language for Interpretability and Accessibility. While deep learning has become the de facto approach to build intelligent systems, the improvement of performance often comes at a cost of interpretability. Complex neural networks permit easy architectural and operational variations for state-of-the-art accuracy, yet they provide little transparency about their inner decision-making mechanisms. I am interested in how natural language can promote interpretable AI: language is not only the means of communication between humans, but it also offers a media for the intelligent system to explain and rationalize its solutions. To this end, I would like to empower the intelligent systems with abilities to automatically extract or generate human-readable language explanations to justify their predictions or actions. Furthermore, I also plan to use natural language to improve the accessibility of data and AI. I hope natural language can serve as a bridge between humans and machines to help users access data more efficiently, recognize the objects in images and graphic user interfaces for visually impaired people, converse and interact with machines to accomplish daily tasks, etc.

Controllable and Personalized Text Generation. While deep learning models have shown promising results in text generation tasks such as summarization, translation, and dialog response generation, progress remains to be made towards controllable and personalized text generation. In particular, I would like to develop more controllable models such that (1) the generated text stay faithful to the conditioned text input, (2) we can manipulate and transfer the output text attributes (such as formal v.s. informal style, positive v.s. negative sentiment), (3) we should remove social bias and abusive content. Furthermore, I would also like to incorporate personal information such as gender, age, social context, and background knowledge to generate text suitable to individual users.

References

[1] **Rui Zhang**, Honglak Lee, and Dragomir R. Radev. Dependency sensitive convolutional neural networks for modeling sentences and documents. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1512–1521, San Diego, California, June 2016. Association for Computational Linguistics.

[2] **Rui Zhang**, Cícero Nogueira dos Santos, Michihiro Yasunaga, Bing Xiang, and Dragomir Radev. Neural coreference resolution with deep biaffine attention by joint mention detection and mention clustering. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 102–107, Melbourne, Australia, July 2018. Association for Computational Linguistics.

[3] **Rui Zhang**, Honglak Lee, Lazaros Polymenakos, and Dragomir Radev. Addressee and response selection in multi-party conversations with speaker interaction rnns. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[4] **Rui Zhang**, Caitlin Westerfield, Sungrok Shim, Garrett Bingham, Alexander Fabbri, William Hu, Neha Verma, and Dragomir Radev. Improving low-resource cross-lingual document retrieval by reranking with deep bilingual representations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3173–3179, Florence, Italy, July 2019. Association for Computational Linguistics.

[5] Douglas W Oard, Drago Radev, Kathleen McKeown, **Rui Zhang**, et al. Surprise languages: Rapid-response cross-language ir. In *The Ninth International Workshop on Evaluating Information Access*, 2019.

[6] Michihiro Yasunaga, **Rui Zhang**, Kshitijh Meelu, Ayush Pareek, Krishnan Srinivasan, and Dragomir Radev. Graph-based neural multi-document summarization. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pages 452–462, Vancouver, Canada, August 2017. Association for Computational Linguistics.

[7] Michihiro Yasunaga, Jungo Kasai, **Rui Zhang**, Alexander R Fabbri, Irene Li, Dan Friedman, and Dragomir R Radev. Scisummnet: A large annotated corpus and content-impact models for scientific paper summarization with citation networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 7386–7393, 2019.

[8] **Rui Zhang** and Joel Tetreault. This email could save your life: Introducing the task of email subject line generation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 446–456, Florence, Italy, July 2019. Association for Computational Linguistics.

[9] **Rui Zhang**, Tao Yu, He Yang Er, Sungrok Shim, Eric Xue, Xi Victoria Lin, Tianze Shi, Caiming Xiong, Richard Socher, and Dragomir Radev. Editing-based sql query generation for cross-domain context-dependent questions. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing*. Association for Computational Linguistics, 2019.

[10] Tao Yu, **Rui Zhang**, Michihiro Yasunaga, Yi Chern Tan, Xi Victoria Lin, Suyi Li, Heyang Er, Irene Li, Bo Pang, Tao Chen, Emily Ji, Shreya Dixit, David Proctor, Sungrok Shim, Jonathan Kraft, Vincent Zhang, Caiming Xiong, Richard Socher, and Dragomir Radev. SParC: Cross-domain semantic parsing in context. In *Proceedings of ACL*, 2019.

[11] Tao Yu, **Rui Zhang**, He Yang Er, Suyi Li, Eric Xue, Bo Pang, Xi Victoria Lin, Yi Chern Tan, Tianze Shi, Zihan Li, Youxuan Jiang, Michihiro Yasunaga, Sungrok Shim, Tao Chen, Alexander Fabbri, Zifan Li, Luyao Chen, Yuwen Zhang, Shreya Dixit, Vincent Zhang, Caiming Xiong, Richard Socher, Walter Lasecki, and Dragomir Radev. CoSQL: A conversational text-to-sql challenge towards cross-domain natural language interfaces to databases. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing*. Association for Computational Linguistics, 2019.

[12] Tao Yu, **Rui Zhang**, Kai Yang, Michihiro Yasunaga, Dongxu Wang, Zifan Li, James Ma, Irene Li, Qingning Yao, Shanelle Roman, Zilin Zhang, and Dragomir Radev. Spider: A large-scale human-labeled dataset for complex and cross-domain semantic parsing and text-to-SQL task. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, 2018. Association for Computational Linguistics.