

Yale SPaC: Cross-Domain Semantic Parsing in Context

salesforce

Tao Yu¹, Rui Zhang¹, Michihiro Yasunaga¹, Yi Chern Tan¹, Xi Victoria Lin²,
Suyi Li¹, Heyang Er¹, Irene Li¹, Bo Pang¹, Tao Chen¹, Emily Ji¹, Shreya Dixit¹,
David Proctor¹, Sungrok Shim¹, Jonathan Kraft¹, Vincent Zhang¹,
Caiming Xiong², Richard Socher², Dragomir Radev¹
¹Yale University ²Salesforce Research

<https://yale-lily.github.io/sparc>
<https://yale-lily.github.io/spider>

Overview

Users tend to ask a sequence of thematically related questions to learn about a particular topic or to achieve a complex goal. To simulate these data querying scenarios, we present **SPaC**, a **large-scale, multi-turn, and cross-domain** semantic parsing dataset.

Given a complex interaction goal in **Spider**, a CS college student comes up with sequences of inter-related questions (and write SQL answers) to obtain the information demanded by the interaction goal (**not just decomposition**).

Interaction Examples

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?

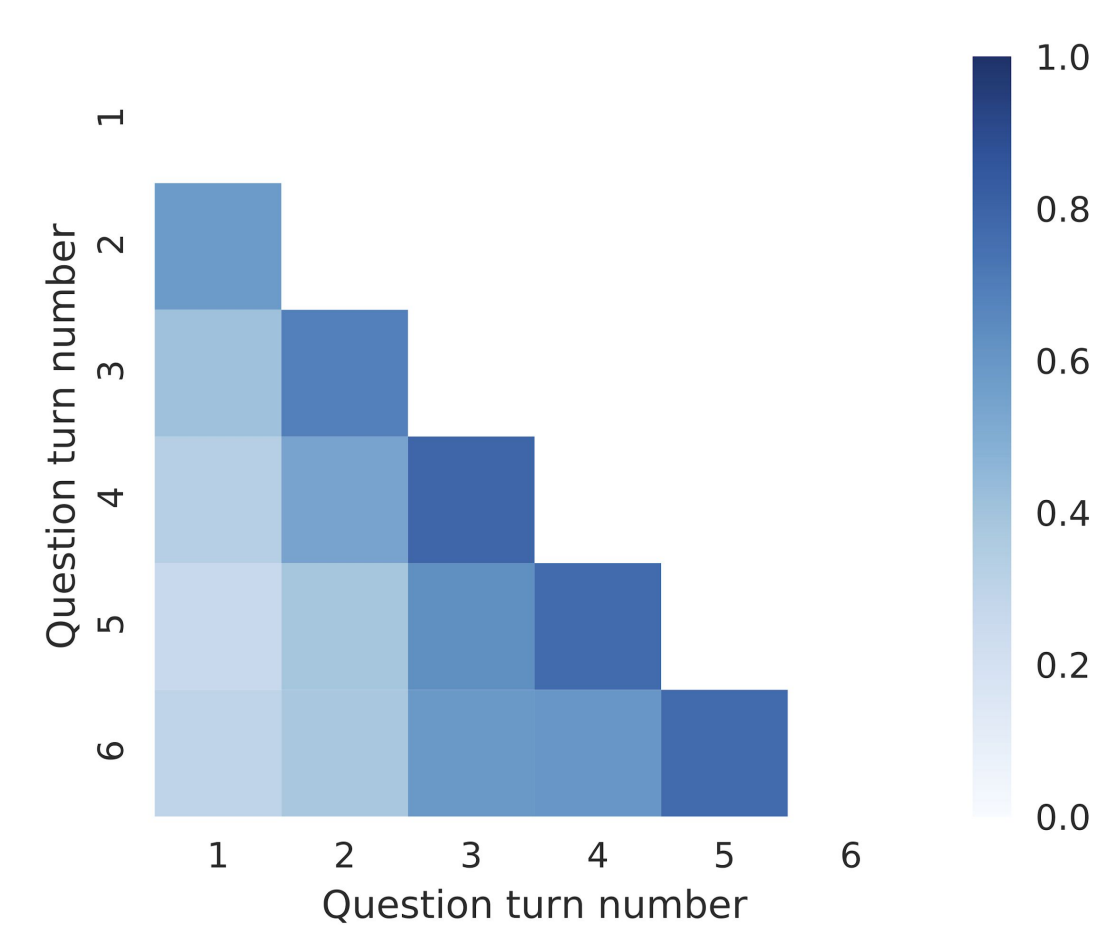
S_3 : `SELECT customer_name FROM customers ORDER BY date_became_customer LIMIT 5`

Data Statistics

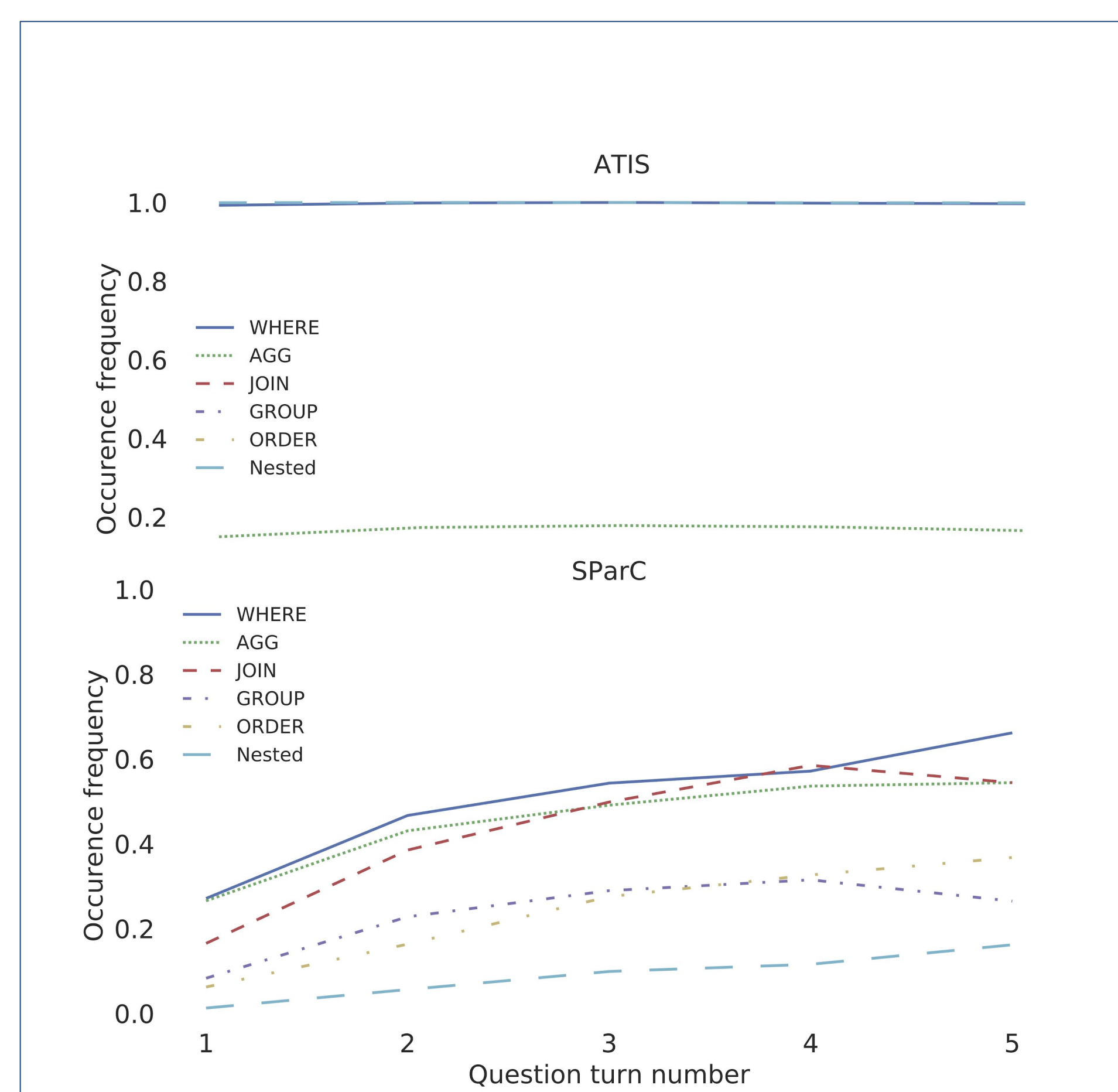
	SPaC	ATIS
Sequence #	4298	1658
Question #	12,726	11,653
Database #	200	1
Table #	1020	27
Avg. Q len	8.1	10.2
Vocab #	3794	1582
Avg. turn #	3.0	7.0

Data Analysis

- demonstrates complex contextual dependencies
- has greater semantic diversity
- requires generalization on new domains



(a) The heatmap shows the percentage of SQL token overlap between questions in different turns



(b) Percentage of question sequences that contain a particular SQL keyword at a given turn. ATIS (top) v.s. SPaC (bottom)

	Train	Dev	Test
# Q sequences	3034	422	842
# Q-SQL pairs	9025	1203	2498
# Databases	140	20	40

(c) **Cross Domain**: SPaC contains questions over 200 databases in 138 different domains. Each database appears in only one of train, development and test sets.

(d) **Thematic relations between questions in a database QA system defined by Bertomeu et al. (2006). The entities (bold), properties (italics) and constraints (underlined) are highlighted in each question.**

Thematic relation	Description	Example	Percentage
Refinement	The current question asks for the same type of entity as a previous question with a different constraint.	Prev_Q: Which major has <u>the fewest students</u> ? Cur_Q: What is <u>the most popular one</u> ?	33.8%
Theme-entity (topic exploration)	The current question asks for other properties about the same entity as a previous question.	Prev_Q: What is <i>the capacity</i> of Anonymous Donor Hall ? Cur_Q: List <i>all of the amenities</i> which it has.	48.4%
Theme-property (participant shift)	The current question asks for the same property about another entity.	Prev_Q: Tell me the <i>rating</i> of the episode named "Double Down" . Cur_Q: How about for "Keepers" ?	9.7%
Answer refinement/theme (answer exploration)	The current question asks for a subset of the entities given in a previous answer or asks about a specific entity introduced in a previous answer.	Prev_Q: Please list all the different department names . Cur_Q: What is the <i>average salary</i> of all instructors in the Statistics department ?	8.1%

Experiments

Evaluation Metrics: Measure whether the predicted query as a whole matches the gold query using set comparison for each SQL clause.

Baselines

- CD-Seq2Seq**: a cross domain Seq2Seq based text-to-SQL model extended with the turn-level history encoder proposed in (Suhr et al., 2018)
- SyntaxSQL-con**: a cross-domain syntax tree based text-to-SQL model extended by encoding the previous question.

Performance Analysis

Model	Question Match		Interaction Match	
	Dev	Test	Dev	Test
CD-Seq2Seq	17.1	18.3	6.7	6.4
SyntaxSQL-con	18.5	20.2	4.3	5.2
SyntaxSQL-sta	15.2	16.9	0.7	1.1

(1) Overall performance over all questions and all interactions

Turn #	CD-Seq2Seq	SyntaxSQL-con
1 (422)	31.4	38.6
2 (422)	12.1	11.6
3 (270)	7.8	3.7
≥ 4 (89)	2.2	1.1

(2) Performance stratified by question turns (dev)

Goal Difficulty	CD-Seq2Seq	SyntaxSQL-con
Easy (483)	35.1	38.9
Medium (441)	7.0	7.3
Hard (145)	2.8	1.4
Extra hard (134)	0.8	0.7

(3) Performance stratified by question difficulty (dev)

Thematic relation	CD-Seq2Seq	SyntaxSQL-con
Refinement	8.4	6.5
Theme-entity	13.5	10.2
Theme-property	9.0	7.8
answer refine./them.	12.3	20.4

4) Performance stratified by thematic relations (dev)

Need to better modelling

- SQL semantic in questions
- information flow between questions