Language Technologies Institute

# Thesis Defense

## Wednesday, August 4, 2021| 11:00 am

## Learning Structured Neural Semantic Parsers

**Pengcheng Yin**

**COMMITTEE:**

**Graham Neubig**
**(Chair)**

**Tom Mitchell**

**Yonatan Bisk**

**Luke Zettlemoyer**
**(U. Wash/FB AI Research)**

**Abstract**

Semantic parsing, the task of translating user-issued natural language (NL) utterances (e.g., flights from Pittsburgh to New York) into formal meaning representations (MRs, e.g., an SQL database query or a Python program), has become an important direction in developing natural language interfaces to computational systems. Recent years have witnessed the burgeoning of applying neural network-based semantic parsers in various tasks and domains. However, meaning representations typically exhibit strong syntactic structure, and are defined following domain-specific structured knowledge schemas (e.g., a database schema or Python API specification), which are not easily captured by standard neural sequence transduction models. Neural semantic parsers are also data-hungry, requiring non-trivial manual annotation effort by domain experts. These issues limit the scope of applications supported by a neural semantic parser, impeding the progress of applying the system to broader scenarios, especially those with diverse and complex structures of meaning representations.

In this thesis, we explore developing neural semantic parsing models that could better capture the structures in various types of logical formalisms and knowledge schemas, while providing approaches to mitigate the cost of labeled data acquisition. The dissertation consists of three parts. The first part introduces a general-purpose parsing model with built-in syntactic knowledge of the grammatical structures of meaning representations. Next, in the second part, we investigate approaches to encode structured information in domain knowledge schemas (e.g., database tables) useful to understand user-issued utterances. Specifically, we focus on grounding elements in the schema (e.g., columns like departure_city in database tables, or functions like GetFlight(from=GetCityByName(·)) in API specifications) to their corresponding NL constituents (e.g., from Pittsburgh) in utterances. Finally, in the third part, we aim to improve the data efficiency of semantic parsers via semi-supervised learning, while developing machine-assisted approaches to accelerate training data acquisition.

**https://www.dropbox.com/sh/33dwoag3zsxvhtf/AAABr_pjcRdDMA8q5UZtr6F-a?dl=0**