

**Motivation:** In this paper, the authors discuss the issue of inference (i.e. learning and modeling textual entailment patterns) in semantic parsing corpora. While a lot of work in semantic parsing has been focused on learning formal meaning representations, they argue that inference has received considerably less attention, despite being one of fundamental aspects of semantic theory.

In particular, they show specific cases where entailment patterns (i.e. judgements about whether one sentence follows from another) mismatch the target semantic representations being learned, which shows that such representations do not adequately capture the full meaning of the sentences. Two of their examples are shown below:

1. purple 5 quickly kicks the ball (rep= `kick(purple5)`)
2. purple 5 kicks (rep=`kick(purple5)`)

Notice that the first sentence has more information than the second, since it contains an adverbial modifier “quickly”. As such, we can say that example 2 follows from (or is *entailed by*) example 1. Since, however, the modifier is not explicitly expressed in the target representation (right), one cannot capture these entailment using the representations alone.

Their goal is to learn better representations that capture the correct semantics and entailments, and they propose a new method for doing this.

**Method:** Their method involves using judgements about entailments between pairs of sentences as additional supervision for training a semantic parsing. They add such pairs to the ordinary semantic parsing corpus and, using a logical calculus, reason about the meaning of the symbols in the logic. The idea is that proofs in this logical calculus reveal facts not captured in the target representations, which allows the system to learn missing information.

Modeling is done using a single probabilistic generative model, in this case a PCFG. Rules of the PCFG are described in the form of templates throughout the paper. To train this model, they use a version of the expectation maximization algorithm (EM).

**Evaluation:** They describe a new RTE style experiment (i.e. they attempt to see if, given such pairs, the system can determine the correct entailments). They compare their models to several baselines (a discriminative, max-entropy classifier, and other such methods) on this task and achieve an accuracy of nearly 73%.

**Problems:** It’s not always clear how the rule templates for the PCFG work. For example.... (specific examples).

In general, this method works nice for their small domain, but it’s not clear how this approach could scale to larger problems. For example, ....