

# Planning in the Dark: LLM-Symbolic Planning

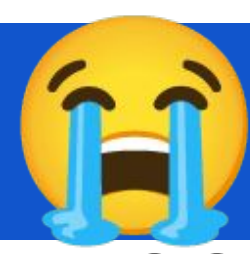
## Pipeline

# Without Experts

Sukai Huang,  
Nir Lipovetzky,  
and Trevor Cohn



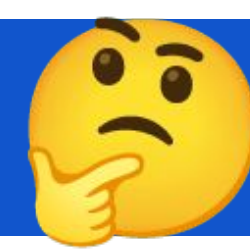
## 1. Limitations in Existing Pipeline



**Fragile Pipeline:** LLM-generated PDDL fail >99.9% of the time—requires *expert!*

**Expert Bottleneck & Bias:** Heavy expert refinement (about. *59 iterations*) + single-perspective bias

## 2. Solvable Schemas: A Simple Fix!



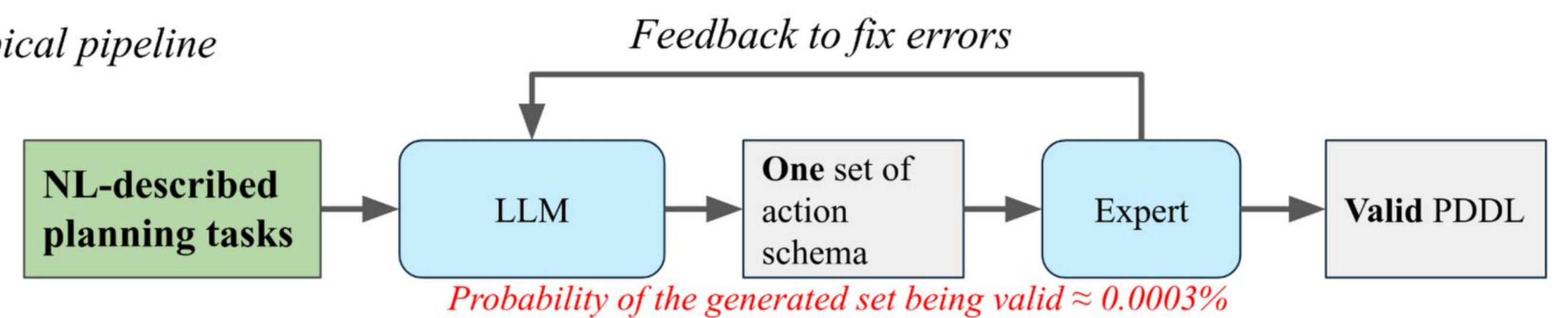
**Multiple LLMs + Inter Schema Set**

**combination:** the probability of *not* finding a solvable set becomes  $(1 - p^M)^{N^M} \rightarrow 0$  where N is #LLMs, M is #actions, p is the prob. of valid action schema (single LLM)

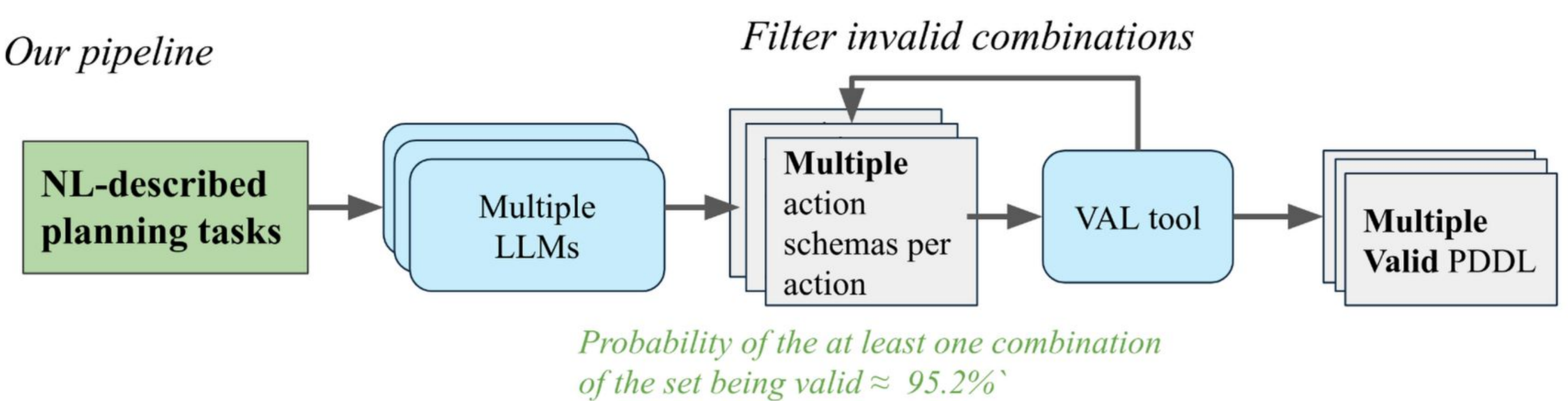
**Adv:** Solvable Schema Without Experts!

**Disadv:** brutal force, semantic misalign

Typical pipeline



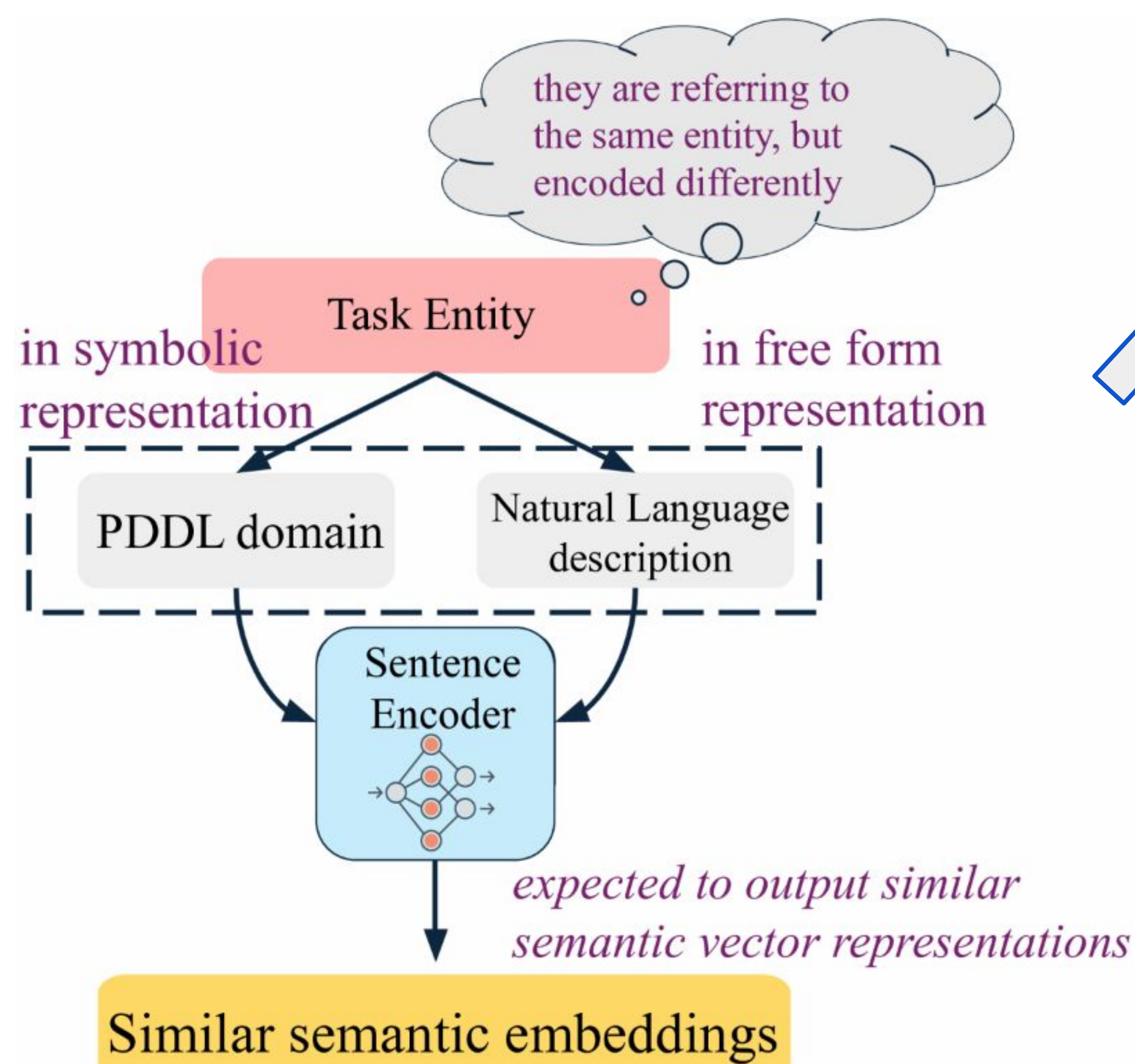
Our pipeline



## 3. Weaver (1952)'s assumption



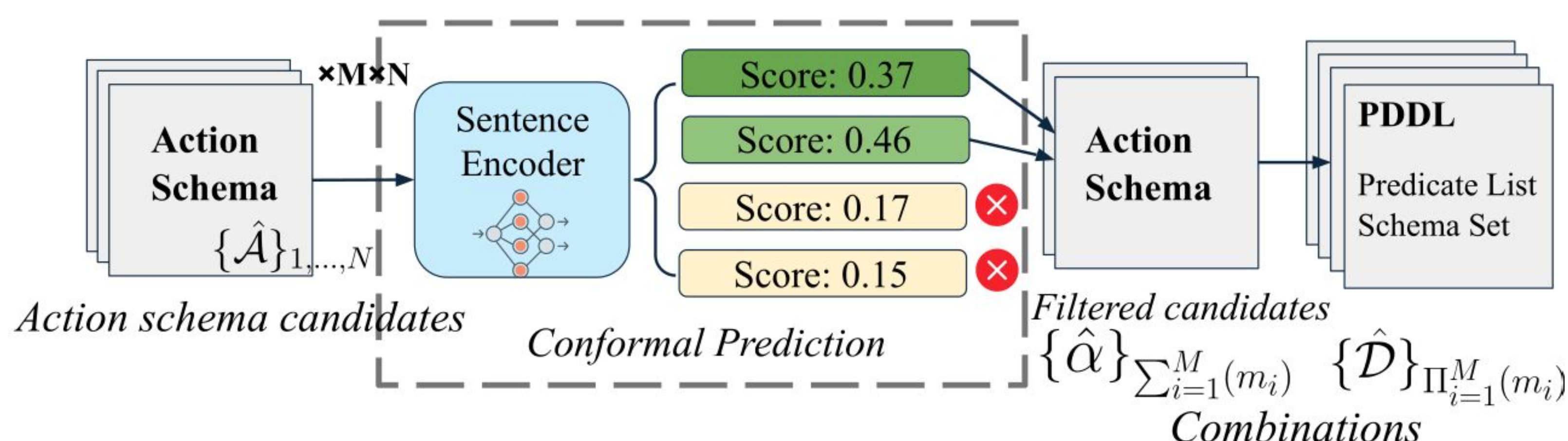
Like translation, there is a “common base of meaning” between natural language task and symbolic schemas.



Assumption in the filtering mechanism

## 4. Filtering and Ranking Inspired by Weaver (1952)

Semantic Coherence Filtering



Semantic score: **Schema Filter** and *even* Plan ranking!

## 5. Fine-tuning the Sentence Encoder is Convenient!

**Contrastive training** with hard negatives synthesized via precondition & effect manipulation

Manipulation Type	Description	Example
Swap	Exchanges a predicate between preconditions and effects	Precondition: (at ?x ?y) Effect: (not (at ?x ?z)) → Precondition: (not (at ?x ?z)) Effect: (at ?x ?y)
Negation	Negates a predicate in either preconditions or effects	Precondition: (clear ?x) → Precondition: (not (clear ?x))
Removal	Removes a predicate from either preconditions or effects	Precondition: (and (on ?x ?y) (clear ?x)) → Precondition: (on ?x ?y)
Addition	Adds mutually exclusive (mutex) predicates to preconditions or effects (Helmert 2009)	Effect: (on-table ?x) → Effect: (and (on-table ?x) (holding ?x))

## 6. Contributions & find out more



1. Address natural language *ambiguity* by having *diverse interpretation* of the action schema
2. Semantic validation, filtering and ranking *without experts*
3. In fact, the proposed pipeline also allows *lightweight* expert intervention to further enhance accuracy too! If you are curious, find our paper to see the details!

