(1) Collect from existing benchmarks

Spider	FOLIO		SCAN
WebQSP	Code		CLEVR

$\mathcal{D}_{s_1}$ (88.3%)
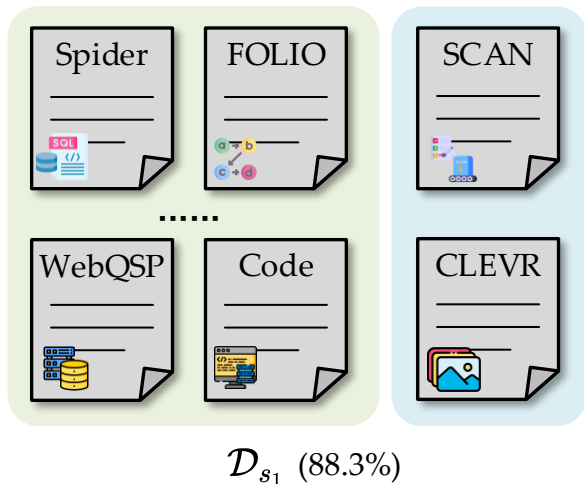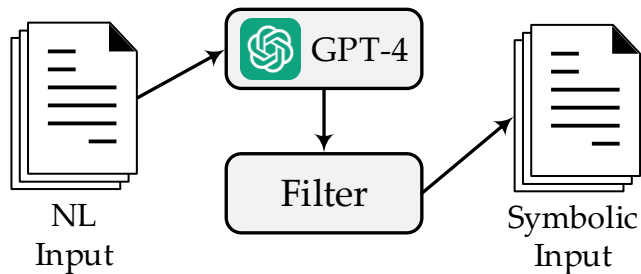
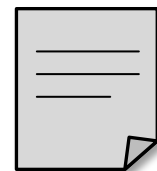## (3) Diversify by *Symbol-evol* strategy

Origin (one example)

Transform the natural language sentence into action sequence from the candidate actions. I_TURN_RIGHT means turning right. I_JUMP means jump...

Symbol-Evol

Transform the natural language sentence into action sequence from the candidate actions. shy2sW means turning right. PMx1Ua means jump...

$\mathcal{D}_{s_3}$ (5.9%)

(2) Prompt GPT-4

NL Input → GPT-4 → Filter → Symbolic Input → New Pairs

$\mathcal{D}_{s_2}$ (5.8%)