

# Beyond Supervised Clarification: Input Rewriting with LLMs for Dialogue Discourse Parsing

Yiming Liu<sup>1</sup>, Ziyue Zhang<sup>1</sup>, Zhichao Xu<sup>2</sup>, Xin Yu<sup>2</sup>,  
Yingheng Tang<sup>3</sup>, Tianyu Jiang<sup>4</sup>, Jie Cao<sup>1</sup>

<sup>1</sup>University of Oklahoma, <sup>2</sup>University of Utah

<sup>3</sup>Lawrence Berkeley National Laboratory, <sup>4</sup>University of Cincinnati

{ymliu, ziyue.zhang-1, jie.cao}@ou.edu, {zhichao.xu, xin.yu}@utah.edu

ytang4@lbl.gov, tianyu.jiang@uc.edu

## Abstract

Rewriting inputs to improve frozen downstream models has become a common strategy in modern NLP pipelines. Prior work on incremental dialogue discourse parsing (DDP) shows that supervised clarification models can rewrite fragmentary or underspecified utterances—such as resolving ellipsis or references—to improve parsing accuracy. In this work, we revisit this idea under realistic deployment conditions, where no clarification supervision is available and the clarifier must rely on zero-shot prompting or feedback from a frozen parser. Across three Segmented Discourse Representation Theory (SDRT) datasets and multiple parsers, we find that last-utterance clarification is far less reliable than suggested by supervised settings. Parser-agnostic rewriting often introduces more regressions than repairs, as edits that enable fixes also disrupt discourse cues relied upon by the parser. A best-of-8 rewriting analysis further reveals a practical ceiling: a large fraction of errors are not repairable through input rewriting alone. A parser-aware clarifier trained with GRPO reduces regressions by up to 37% by learning conservative abstention, yet still fails to produce selectivity-aware clarifications that consistently improve parsing. Together, these findings recast clarification as a selective intervention problem. We identify **rewritability prediction**—deciding whether an utterance is repairable before intervention—as the key missing capability for input-side optimization of frozen discourse parsers, and a critical direction for improving agentic pipelines more broadly.<sup>1</sup>

## 1 Introduction

Recent work has shown that rewriting the input to a frozen system can improve NLP pipelines without retraining, including in machine translation (Ki and Carpuat, 2025), retrieval-augmented

generation (Wu et al., 2022), and conversational search (Zhu et al., 2025). Fan et al. (2025) apply this idea to dialogue discourse parsing (DDP) through *last-utterance* clarification, training an LLM clarifier on supervised clarification data and then optimizing it with preference-based reinforcement learning (RL) to rewrite the final utterance before parsing. However, such supervision is costly to obtain and difficult to reuse across parsers. We therefore ask whether last-utterance clarification can improve DDP without access to supervised clarification examples.

Unlike retrieval or translation, SDRT discourse parsing requires semantic inference over structured dialogue context (Lascarides and Asher, 2007). An incremental parser such as Llamipa (Thompson et al., 2024a) must determine how the current utterance connects to prior discourse. In a sliding-window setting, the last utterance becomes a natural intervention point—it is the only part of the current input that can still be rewritten before prediction. This makes last-utterance clarification particularly interesting, but also challenging: while clarification may reduce ambiguity, it can also alter surface cues (e.g., fragmentary syntax, pronouns, connectives) that the parser relies on.

Following Fan et al. (2025), we study this setting under incremental SDRT parsing, where no future context is available and past dialogue history and decisions cannot be revised. As illustrated in Figure 1, the ambiguous final utterance (e.g.,  $u_{17}$ : “finally,”) is interpreted given the preceding context ( $u_{11}$ – $u_{16}$ ), and a clarifier must decide whether to intervene before passing it to a frozen parser. To go beyond supervised clarification, we consider two settings that reflect common agentic patterns: (1) prompt-based interaction with a frozen tool or system (*parser-agnostic*, §3.1), and (2) adaptation through downstream feedback-driven learning to optimize performance (*parser-aware*, §3.2).

We first instantiate the parser-agnostic setting

<sup>1</sup>Data and code are available at <https://github.com/ounlp/Clarification-for-DDP>.

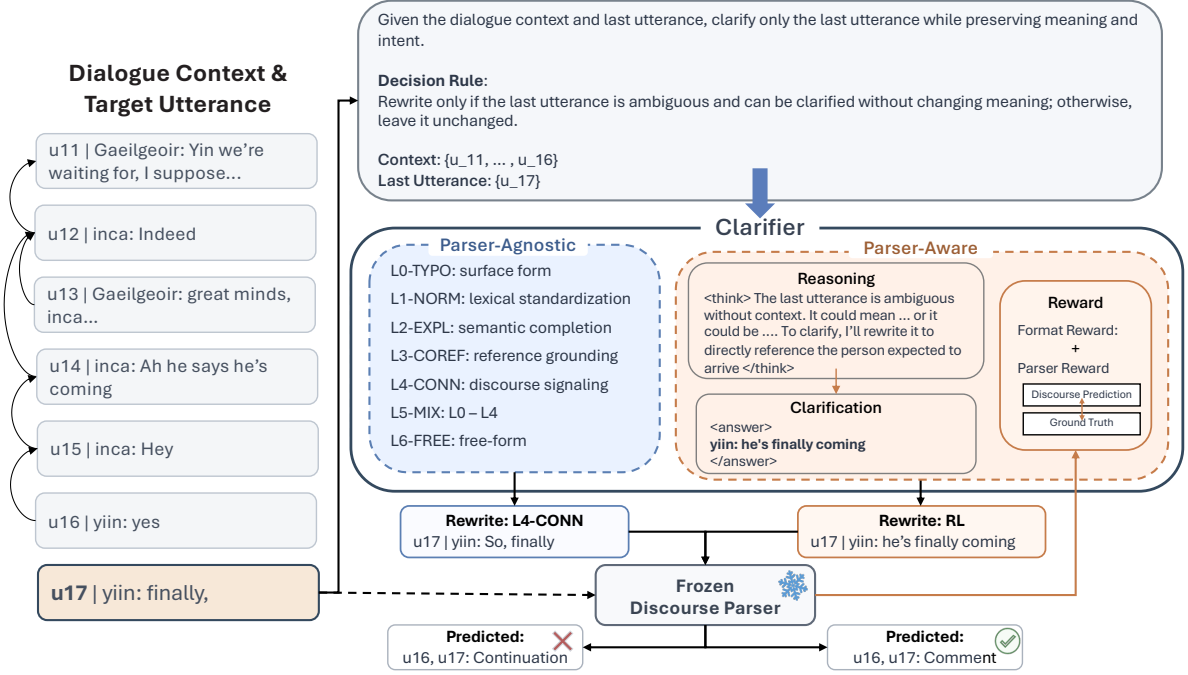


Figure 1: **Overview of our last-utterance clarification framework for incremental dialogue discourse parsing.** Given a dialogue context, the clarifier rewrites only the final utterance, replaces it in the parser’s context window, and passes the modified context to a frozen discourse parser. Parser-agnostic clarification uses predefined rewrite strategies, whereas parser-aware clarification optimizes the clarifier with reinforcement learning from parser-based reward computed by comparing the predicted discourse structure on the rewritten context against the ground truth.

by evaluating seven clarification strategies, ranging from low-level typo correction to free-form semantic rewriting, without any access to parser feedback. In contrast, the parser-aware setting asks whether a clarifier can *learn when and how to intervene* purely from downstream signals. To this end, we train a clarifier with GRPO (Guo et al., 2025), enabling it to acquire rewrite-or-copy behavior directly from frozen-parser feedback. Experiments are conducted on three SDRT datasets (STAC, Molweni, MSDC) and four parsers spanning both generative and discriminative families.

Our results reveal a consistent pattern. Parser-agnostic strategies often introduce more regressions than repairs, as edits that enable fixes can also disrupt cues the parser relies on. A parser-aware RL clarifier reduces regressions by up to 37% by learning when to abstain, but still fails to produce clarifications that reliably improve parsing. A best-of-8 rewriting analysis further shows that about 80% of errors are not repairable by any rewrite, pointing to **rewritability prediction — deciding when to intervene — as a key direction forward**. Together, these findings suggest that last-utterance rewriting does not transfer cleanly to dialogue discourse parsing in an incremental setting,

highlighting a limitation of common agentic designs for black-box components. More broadly, they indicate that upstream modules — whether rewriting, planning, or tool orchestration — are most effective not as generic prompt-driven interfaces, but when they are selective, context-aware, and aligned with downstream signals.

## 2 Related Work

**Dialogue Discourse Parsing.** DDP has been studied with both discriminative and generative parsers under multiple discourse formalisms, including Segmented Discourse Representation Theory (SDRT, Lascarides and Asher, 2007; Asher et al., 2016), Rhetorical Structure Theory (RST, Mann and Thompson, 1988), the Penn Discourse Treebank framework (PDTB, Prasad et al., 2008), and Dependency Dialogue Acts (DDA, Cai et al., 2023, 2025). Early work focused on discriminative structure prediction for multi-party dialogue (Afan-tenos et al., 2015; Perret et al., 2016; Shi and Huang, 2018; Liu and Chen, 2021; Chi and Rudnicky, 2022), while more recent work has shown that LLM-based generative parsers can achieve strong performance (Li et al., 2024; Thompson et al., 2024a) on STAC (Asher et al., 2016), Mol-

weni (Li et al., 2020), and MSDC (Thompson et al., 2024b). Prior work has identified several input-side phenomena that are plausibly relevant to relation prediction, including unresolved references (Choi et al., 2021), fragmentary utterances (Li et al., 2023), and explicit versus implicit discourse marking (Liu et al., 2024). Our work treats these phenomena as potential targets of upstream clarification and tests whether resolving them actually helps a frozen parser.

**Input Rewriting for Frozen Systems.** A growing body of work improves frozen downstream systems by rewriting their inputs at inference time, including in machine translation (Sun et al., 2024; Ki and Carpuat, 2025), conversational retrieval and search (Wu et al., 2022; Zhu et al., 2025; Cao et al., 2025; Xu et al., 2026), retrieval-augmented generation (Ma et al., 2023; Ye et al., 2025) and recommendation system (Lin et al., 2025). This line of work is also related to broader meaning-preserving reformulation methods, such as decontextualization and incomplete utterance rewriting, which transform an input into a more explicit form without changing its underlying meaning (Choi et al., 2021; Li et al., 2023). Our setting differs from these tasks in that the downstream objective is not directly tied to local surface matching: in discourse parsing, rewriting a single utterance affects the parser only indirectly through its interaction with the broader dialogue context.

**Clarification for Discourse Parsing.** Most related to our work, Fan et al. (2025) train an LLM clarifier for DDP using supervised clarification data followed by preference-based reinforcement learning (RLHF; Ouyang et al., 2022). We study a different setting: the clarifier has access only to frozen parser feedback, with no supervised clarification data. This lets us test clarification as input-side intervention, rather than as supervised adaptation to parser-specific errors. Aktas and Roth (2025) study connective insertion for underspecified discourse relations in instructional texts under PDTB, reporting gains from making implicit connectives explicit. Their work also frames clarification as a discourse-oriented rewriting operation, but in a different task setting. We therefore position our work as a study of the limits of clarification as input-side optimization for a frozen discourse parser: whether it remains effective without supervised clarification data, and whether frozen-parser feedback alone is sufficient to support useful intervention.

## 3 Methods and Experiments

### 3.1 Parser-Agnostic Clarification Strategies

Prior work has identified three input-side phenomena that specifically challenge discourse relation labeling (Choi et al., 2021; Ma et al., 2023; Sun et al., 2023; Ki and Carpuat, 2025). These observations motivate targeting these phenomena through upstream rewriting. However, whether resolving them actually helps a frozen parser, or whether the parser has learned to handle them through its training distribution, is precisely the empirical question we study. We therefore define clarification strategies at increasing levels of intervention, from conservative surface edits that are unlikely to disturb any parser cues, to deeper semantic edits that directly target the discourse-relevant phenomena above. This ordering lets us jointly ask: at what depth of intervention do repairs emerge, and at what depth do regressions begin to dominate?

**Surface form (L0-TYPO)** corrects only obvious typos and surface errors, which may resolve entity mentions or cue words that affect attachment and relation labeling. Prior work on lexical normalization of non-canonical forms has been shown to improve downstream tagging performance, suggesting that restoring corrupted cue words can benefit structure prediction (van der Goot and Çetinoğlu, 2021).

**Lexical Standardization (L1-NORM)** expands abbreviations/shorthand/slang into standard forms (e.g., “idk”→“I don’t know”), without altering wording or adding information. This reduces surface noise by restoring lexical cues that guide DDP, consistent with evidence that task-agnostic normalization improves robustness under noisy inputs (Bitton et al., 2022).

**Semantic Completion (L2-EXPL)** targets ellipsis and fragmentary turns by minimally restoring omitted but context-entailed material, making the last-utterance more self-contained. Li et al. (2023) show that rewriting underspecified dialogue utterances supports the benefit of meaning-preserving completion for dialogue understanding.

**Reference Grounding (L3-COREF)** makes coreference explicit by replacing pronouns/deictics with contextually unambiguous antecedents to improve entity continuity and reduce referential ambiguity (Choi et al., 2021). This aligns with evidence from conversational QA/retrieval that making context-dependent turns more explicit can improve downstream performance (Wu et al., 2022).

**Discourse Signaling (L4-CONN)** adds a semanti-

cally compatible cue (e.g., “because,” “but,” “so,” “by the way”) to signal the intended discourse relation explicitly. Such connectives often constrain relation interpretation, and their removal may trigger label shift (Liu et al., 2024). Therefore, we treat connective insertion as a natural parser-agnostic clarification strategy, following Aktas and Roth (2025).

In addition to the five base levels (L0 to L4), we include a mixed-rule strategy (L5-Mix) and a minimally constrained free-form clarification strategy (L6-FREE), yielding 7 parser-agnostic clarifiers. All strategies share the same prompt skeleton and differ only in their level-specific rules. L5-Mix exposes all five base-level edit operations (L0 to L4) within a single prompt and permits the clarifier to apply any subset whose per-rule safety conditions are satisfied. As shown in Table 8, for each level, we select the prompt that ensures clarification while maximizing meaning preservation metric BERTScore (Zhang et al., 2019) on STAC validation set under that level’s corresponding edit constraints<sup>2</sup>.

### 3.2 RL-based Parser-Aware Clarification

We train a rewriting policy  $\pi_\phi$  with reinforcement learning to improve a frozen discourse parser. At each dialogue step  $t$ , the policy observes the current dialogue prefix  $s_t = (u_1, \dots, u_t)$ , where  $u_t$  is the final utterance to be clarified, and generates either the original utterance or a rewritten version  $\tilde{u}_t$ . We then replace  $u_t$  with  $\tilde{u}_t$  in the dialogue prefix and run the frozen parser on both the original and rewritten contexts. The parser’s predictions are compared against the gold discourse structure, and the resulting change in parsing quality is converted into a scalar reward for policy learning.

We optimize  $\pi_\phi$  with GRPO (Guo et al., 2025), using a clipped policy-gradient objective with group-relative advantages and KL regularization to prevent degenerate rewrites. We use a verifiable reward defined entirely by the frozen parser’s behavior on the rewritten utterance. The reward has two components: a format reward and a parsing reward. Following Guo et al. (2025) and Jiang et al. (2025), who demonstrate that enforcing structured output via format rewards stabilizes GRPO training and prevents degenerate generations, we assign  $r_{fmt} = 1$  when the output contains exactly

<sup>2</sup>Our prompting experiments suggest that stronger constraints are necessary to elicit repairs. Please refer to Appendix B.

| Dataset | Train | Dev   | Test  |
|---------|-------|-------|-------|
| STAC    | 9,507 | 1,084 | 1,045 |
| Molweni | 9,215 | 1,026 | 1,107 |
| MSDC    | 1,732 | 1,016 | 989   |

Table 1: Dataset statistics for SDRT dialogue discourse parsing datasets: STAC, Molweni, and MSDC. Counts are the number of last-utterance parsing instances (one per dialogue turn) in each split. We follow the standard splits provided with each dataset.

one valid `<think>` . . . `<answer>` block in order,  $r_{fmt} = -2$  otherwise. For parsing rewards, we assign +2 for correcting errors, -2 for introducing errors, +1 for partial fixes, and 0 otherwise. This shaping encourages well-formed clarifications that measurably improve discourse parsing. Detailed RL implementation using verl (Sheng et al., 2024) is provided in §Appendix C.

### 3.3 Experiment Setup

**Datasets.** We evaluate on three SDRT-based dialogue discourse parsing datasets: STAC, Molweni, and MSDC, using their standard train/dev/test splits (Table 1). STAC contains multi-party game dialogues, Molweni includes Ubuntu technical chats, and MSDC consists of collaborative Minecraft dialogues. Importantly, we do not use any clarification or rewriting supervision—our methods operate without clarification data.

**Models.** On each dataset, we train two LLM-based generative parsers by supervised LoRA fine-tuning (Hu et al., 2021) from Qwen3-8B and Qwen3-14B (Yang et al., 2025), outperforming the original Llama3-based Llamipa models (as Table 2, additional training details are provided in §D.1). To test whether our findings generalize beyond generative parsers, we also evaluate SDDP (Chi and Rudnicky, 2022), a discriminative graph-based discourse parser that uses a BERT-based encoder to score candidate parent-child links and relation types via biaffine attention (Chen and Manning, 2014), then selects a globally consistent set of links. We adapt SDDP to our incremental setting by parsing each dialogue prefix and evaluating only the relations predicted for the current utterance.

We study seven parser-agnostic (§3.1) and one parser-aware (§3.2) clarifiers based on Qwen3 and Qwen2.5-7B-Instruct respectively.<sup>3</sup> We train the

<sup>3</sup>We use Qwen2.5-7B-Instruct rather than the Qwen3 for the parser-aware clarifier due to compatibility constraints with our RL training framework (verl).



| Method          | STAC         |            | Molweni      |              | MSDC         |              |
|-----------------|--------------|------------|--------------|--------------|--------------|--------------|
|                 | F1           | Fix/Reg    | F1           | Fix/Reg      | F1           | Fix/Reg      |
| Llamipa         | 0.577        | -          | -            | -            | 0.795        | -            |
| Qwen3-8B        | 0.598        | -          | 0.571        | -            | 0.800        | -            |
| <b>L0-TYPO</b>  | <b>0.596</b> | <b>2/3</b> | <u>0.565</u> | <u>28/51</u> | <b>0.800</b> | <b>11/11</b> |
| <b>L1-NORM</b>  | 0.594        | 5/9        | 0.564        | 31/59        | <u>0.797</u> | <u>12/24</u> |
| <b>L2-EXPL</b>  | 0.582        | 7/23       | <b>0.568</b> | <b>33/47</b> | 0.795        | 24/48        |
| <b>L3-COREF</b> | 0.594        | 5/8        | 0.556        | 38/96        | 0.794        | 14/40        |
| <b>L4-CONN</b>  | 0.540        | 21/88      | 0.531        | 74/234       | 0.771        | 61/212       |
| <b>L5-MIX</b>   | 0.589        | 9/20       | 0.555        | 50/114       | 0.793        | 25/56        |
| <b>L6-FREE</b>  | 0.577        | 17/38      | 0.565        | 104/128      | 0.789        | 43/105       |
| Qwen3-14B       | 0.591        | -          | 0.554        | -            | 0.805        | -            |
| <b>L0-TYPO</b>  | <b>0.589</b> | <b>2/4</b> | <u>0.552</u> | <u>17/27</u> | 0.803        | 2/10         |
| <b>L1-NORM</b>  | 0.585        | 6/12       | <b>0.552</b> | <b>34/42</b> | 0.804        | 9/12         |
| <b>L2-EXPL</b>  | 0.576        | 10/26      | 0.551        | 36/51        | <b>0.799</b> | <b>21/47</b> |
| <b>L3-COREF</b> | <u>0.587</u> | <u>3/7</u> | 0.538        | 30/94        | 0.799        | 12/45        |
| <b>L4-CONN</b>  | 0.532        | 22/89      | 0.512        | 73/239       | 0.779        | 60/192       |
| <b>L5-MIX</b>   | 0.584        | 9/17       | 0.549        | 49/74        | <u>0.799</u> | <u>13/44</u> |
| <b>L6-FREE</b>  | 0.567        | 16/44      | 0.549        | 90/111       | 0.797        | 40/84        |

Table 2: Parser-agnostic clarification introduces more regressions than repairs, with discourse connective insertion (**L4-CONN**) producing the largest degradations.

parser-aware clarifier against the Qwen3-8B parser and report its results in comparison with the **L6-FREE** baseline on the same parser. Because the clarifier and parser do not share parameters and interact only through rewritten text, this model mismatch does not affect the validity of our setup.

**Evaluation.** We use micro-F1 for link attachment and full structure (link+relation) to measure model performance across all datasets, parsers, and clarifiers. We also report the number of fixes (wrong→correct; “fix”) and regressions (correct→wrong; “reg”) after clarifying the last utterance (Table 2).

## 4 Main Results and Analysis

Our results show a clear pattern: last-utterance clarification often harms more than it helps, and even with RL, the challenge is not how to rewrite, but when to intervene. Parser-agnostic strategies (§4.1) introduce more regressions than repairs, while parser-aware RL (§4.2) learns dataset-level intervention regimes without achieving reliable per-instance selectivity. Finally, we provide further insights via best-effort baseline analysis and ablation studies (§5).

### 4.1 Parser-Agnostic Rewriting Induces Regressions

Table 2 summarizes the impact of seven parser-agnostic clarification strategies applied to the last utterance under two frozen parsers (Qwen3-8B/14B) across three datasets. A consistent pattern emerges: *parser-agnostic clarification introduces more regressions than repairs, indicating*

| Metric                       | STAC  | Molweni | MSDC  |
|------------------------------|-------|---------|-------|
| <i>LLM-rewrite (L6-FREE)</i> |       |         |       |
| Link                         | 0.752 | -       | -     |
| Link+Rel                     | 0.577 | 0.565   | 0.789 |
| Repairs                      | 17    | 104     | 43    |
| Regressions                  | 38    | 128     | 105   |
| Edit%                        | 0.499 | 0.882   | 0.453 |
| <i>Parser-aware RL</i>       |       |         |       |
| Link                         | 0.754 | 0.846   | 0.872 |
| Link+Rel                     | 0.586 | 0.551   | 0.784 |
| Repairs                      | 7     | 110     | 51    |
| Regressions                  | 24    | 113     | 125   |
| Edit%                        | 0.214 | 0.988   | 0.338 |

Table 3: Results of parser-aware clarification compared with the L6-FREE baseline across three datasets using the Qwen3-8B parser. Edit% is the fraction of examples where the last utterance is rewritten. The learned policy behaves differently across datasets: it becomes more conservative on STAC, rewrites nearly all utterances on Molweni, and remains harmful on MSDC.

*that improved surface clarity does not reliably benefit frozen discourse parsing.* For Qwen3-8B parser, all strategies either leave performance nearly unchanged (e.g., STAC **L0-TYPO**: 0.598 → 0.596) or reduce Link+Rel F1, with the largest drops consistently observed for discourse connective insertion (**L4-CONN**). Qwen3-14B shows the same trend, including the same failure mode for **L4-CONN**, suggesting that the limitation is not specific to smaller models.

A second pattern reveals a clear safety–utility trade-off across intervention depth. Conservative strategies (**L0-TYPO** and **L1-NORM**) are relatively safe but offer limited gains (e.g., on STAC with Qwen3-8B: 2 vs. 3 and 5 vs. 9 for repairs vs. regressions). In contrast, more semantic interventions create greater repair opportunities but incur substantially more regressions: **L4-CONN** yields 21 vs. 88, and **L6-FREE** 17 vs. 38. This imbalance persists across models and datasets, with heavier interventions showing greater instability. *In short, the same edits that enable fixes also disproportionately increase the risk of harming correct predictions.*

### 4.2 Parser-Aware RL Learns Intervention Regimes, Not Selective Policies

The key question for parser-aware clarification is whether frozen-parser feedback can train a policy that intervenes selectively—rewriting when helpful and abstaining when harmful. Table 3 and Figure 2 show that RL does learn coherent, convergent behaviors within each dataset. However, rather than

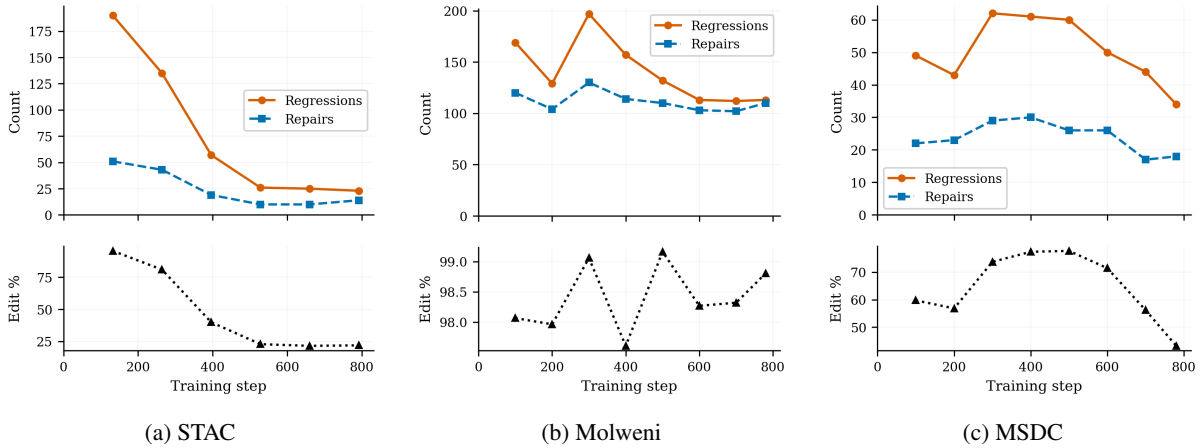


Figure 2: Validation-set learning dynamics of the parser-aware clarifier on STAC, Molweni, and MSDC. Across datasets, GRPO training produces different intervention behaviors rather than a single stable selective policy: conservative abstention on STAC, near-universal rewriting on Molweni, and unstable behavior on MSDC.

collapsing to a single universal strategy, the learned policies diverge in systematic ways: conservative abstention on STAC, near-universal rewriting on Molweni, and a more unstable mixed behavior on MSDC. This variation suggests that the learning signal shapes how often to intervene, but not yet when to intervene at the instance level.

On STAC, the learned policy improves Link+Rel F1 over the **L6-FREE** baseline while substantially reducing regressions (38  $\rightarrow$  24, -37%). Figure 2 shows a corresponding drop in Edit% from 49.9% to 21.4%, indicating that the policy learns to abstain. Notably, this gain comes not from better rewrites — repairs decrease (17 to 7) — but from avoiding harmful ones. RL therefore discovers a conservative regime where doing less is better.

This abstention dynamic does not generalize to the other two datasets. On Molweni, the learned policy rewrites nearly every utterance, reaching a final Edit% of 98.8%. As shown in Table 3, regressions fall only modestly (128  $\rightarrow$  113) and repairs increase only slightly (104  $\rightarrow$  110), while Link+Rel F1 drops relative to the **L6-FREE** baseline. The policy therefore learns to intervene almost universally rather than selectively. MSDC exhibits a different behavior. Figure 2 shows that Edit% eventually declines during training, but regressions remain high and the final policy is still harmful: compared with **L6-FREE**, regressions increase (105  $\rightarrow$  125) while repairs rise only modestly (43  $\rightarrow$  51), resulting in a lower Link+Rel F1. Unlike STAC, reducing the intervention rate is not sufficient here to avoid harmful rewrites.

Taken together, although the learned policies

vary across datasets, they converge to structured intervention regimes rather than random behavior, yet still fall short of true per-instance selectivity.

## 5 Ablation Studies and Discussion

The main results suggest that the central challenge is not whether clarification can help, but when and why it does. To better understand this, we conduct further analyses to characterize the repairability ceiling of last-utterance rewriting (§5.1), examine whether the observed instability depends on parser family (§5.2), and synthesize key implications for future work (§5.3).

### 5.1 How Repairable Are Parser Errors?

To estimate the headroom of clarification on DDP, we construct a best-effort baseline by generating eight independent rewrites on the validation set with the base Qwen3-8B model (our skeleton model for RL-post-training). For each incorrect prediction with the original last utterance, we use the same **L6-FREE** strategy to sample 8 free-form rewrites of the last utterance, rerun the frozen parser, and mark the example as repairable if any rewrite yields the correct prediction. This estimates the upper bound of recoverable errors under repeated rewriting, rather than the performance of a practical clarifier.

Table 4 shows that the repairability ceiling is low across all three datasets. Most parser errors remain unrepaired even under candidate selection: 80.0% errors on STAC, 72.5% on Molweni, and 71.8% on MSDC are not repaired by any of the eight sampled rewrites. At the same time, repairability appears as a spectrum rather than a binary property. A non-

| Category              | Succ. (of 8) | STAC  |      | Molweni |      | MSDC  |      |
|-----------------------|--------------|-------|------|---------|------|-------|------|
|                       |              | # Ex. | %    | # Ex.   | %    | # Ex. | %    |
| Not repairable        | 0            | 475   | 80.0 | 1173    | 72.5 | 265   | 71.8 |
| Rarely repairable     | 1–2          | 66    | 11.1 | 182     | 11.2 | 35    | 9.5  |
| Moderately repairable | 3–5          | 39    | 6.6  | 147     | 9.1  | 52    | 14.1 |
| Highly repairable     | 6–8          | 14    | 2.4  | 116     | 7.2  | 17    | 4.6  |

Table 4: Distribution of originally incorrect examples on validation set by estimated repairability under a best-of-8 free-form rewriting baseline. Repairability is defined by how many of eight sampled rewrites, generated from the same free-form prompt, yield a correct prediction.

| Dataset | Not rep. | Rarely | Moderately  | Highly      |
|---------|----------|--------|-------------|-------------|
| STAC    | 0.9      | 4.4    | <b>12.5</b> | 11.1        |
| Molweni | 0.8      | 7.7    | 19.7        | <b>31.9</b> |
| MSDC    | 1.5      | 5.7    | 9.6         | <b>23.5</b> |

Table 5: Repair rate (%) of the parser-aware RL clarifier on baseline-wrong validation examples, grouped by estimated repairability from Table 4. For each bucket, the repair rate is the fraction of examples for which the final RL policy changes an originally incorrect prediction into a correct one.

trivial minority of cases are moderately repairable, accounting for 6.6% of errors on STAC, 9.1% on Molweni, and 14.1% on MSDC, while only a much smaller fraction are highly repairable. These results indicate that repeated free-form rewriting of the final utterance offers genuine headroom for a limited subset of cases, but leaves most errors untouched.

To examine whether the parser-aware RL policy aligns with the repairable subset – and potentially goes beyond it – we group baseline-wrong validation examples by their estimated repairability (Table 4) and compute RL repair rates within each bucket. Table 5 shows a clear trend: the policy rarely repairs examples deemed not repairable by the candidate rewrites (<2% across datasets), while repair rates increase substantially with estimated difficulty. On Molweni and MSDC, this rise is nearly monotonic, reaching 31.9% and 23.5% in the highly repairable bucket; STAC shows a similar pattern, with minor variance due to small sample size. Notably, even within the “not-repairable” bucket, the small fraction of successful repairs (<2%) highlights the exploratory power of RL: the policy can occasionally discover effective rewrites beyond the finite candidate rewrites. **Overall, candidate rewriting repairability is a strong predictor of RL success, while RL retains a limited but meaningful ability to search beyond it.** However, this alignment reflects the quality of rewrites, not the decision to rewrite: on STAC, RL’s rewrite

| Method          | Link  | Link + Rel   | Fix/Reg      | Edit% |
|-----------------|-------|--------------|--------------|-------|
| STAC            |       |              |              |       |
| SDDP            | 0.729 | 0.557        | –            | –     |
| <b>L0-TYPO</b>  | 0.727 | <b>0.552</b> | <b>0/4</b>   | 0.098 |
| <b>L1-NORM</b>  | 0.727 | <b>0.551</b> | <b>6/10</b>  | 0.207 |
| <b>L2-EXPL</b>  | 0.726 | <b>0.539</b> | <b>9/27</b>  | 0.216 |
| <b>L3-COREF</b> | 0.725 | 0.549        | 3/13         | 0.083 |
| <b>L4-CONN</b>  | 0.706 | 0.497        | 19/81        | 0.462 |
| <b>L5-MIX</b>   | 0.721 | 0.547        | 10/22        | 0.282 |
| <b>L6-FREE</b>  | 0.717 | 0.525        | 17/47        | 0.499 |
| Molweni         |       |              |              |       |
| SDDP            | 0.798 | 0.543        | –            | –     |
| <b>L0-TYPO</b>  | 0.795 | <b>0.537</b> | <b>16/40</b> | 0.425 |
| <b>L1-NORM</b>  | 0.797 | <b>0.536</b> | <b>18/46</b> | 0.448 |
| <b>L2-EXPL</b>  | 0.799 | <b>0.535</b> | <b>22/54</b> | 0.244 |
| <b>L3-COREF</b> | 0.796 | 0.529        | 22/75        | 0.297 |
| <b>L4-CONN</b>  | 0.797 | 0.522        | 76/158       | 0.684 |
| <b>L5-MIX</b>   | 0.793 | 0.530        | 45/96        | 0.691 |
| <b>L6-FREE</b>  | 0.794 | 0.533        | 74/111       | 0.882 |
| MSDC            |       |              |              |       |
| SDDP            | 0.779 | 0.696        | –            | –     |
| <b>L0-TYPO</b>  | 0.778 | <b>0.696</b> | <b>11/18</b> | 0.070 |
| <b>L1-NORM</b>  | 0.778 | <b>0.696</b> | <b>13/20</b> | 0.123 |
| <b>L2-EXPL</b>  | 0.776 | 0.695        | 8/24         | 0.147 |
| <b>L3-COREF</b> | 0.776 | 0.694        | 19/37        | 0.123 |
| <b>L4-CONN</b>  | 0.768 | 0.688        | 20/79        | 0.458 |
| <b>L5-MIX</b>   | 0.775 | 0.694        | 12/32        | 0.196 |
| <b>L6-FREE</b>  | 0.775 | 0.693        | 19/43        | 0.453 |

Table 6: Parser-agnostic clarification results with the SDDP parser on STAC, Molweni, and MSDC. The same pattern holds across all three corpora: conservative edits are near-neutral, while discourse connective insertion (**L4-CONN**) produces the largest degradation.

rate stays near the 23.0% base rate across all four buckets, confirming global abstention rather than per-instance selectivity.

Tables 4 and 5 do more than explain why overall gains are limited—they reveal a structured landscape of opportunities. While a large portion of cases appears inherently unrepairable under local rewriting, a meaningful subset remains consistently repairable, and RL shows a clear ability to align with — even occasionally go beyond — this structure through exploration. This perspective reframes the problem: rather than treating clarification as uniformly applicable, it suggests a more selective paradigm. A natural next step is rewritability prediction (identifying when intervention is likely to help), so that future systems can focus effort where improvement is possible, rather than attempting to rewrite indiscriminately.

## 5.2 Do Different Parser Families Matter?

We evaluate SDDP on STAC, Molweni and MSDC (Table 6) and compare it with the LLM-

based parsers (Table 2), and we observe that the same clarification strategies yield qualitatively similar patterns across parser families. Conservative surface-level edits (**L0-TYPO**, **L1-NORM**) are largely neutral across both LLM-based and SDDP parsers, producing small changes in Link+Rel with relatively few repairs or regressions. In contrast, discourse connective insertion (**L4-CONN**) produces the largest degradations on both families.

Notably, structured decoding does not eliminate clarification-induced regressions. Despite SDDP’s explicit global structure modeling, it exhibits the same overall pattern as the LLM-based parsers: conservative edits are mostly neutral, while semantically heavier rewrites reduce Link+Rel accuracy. Strategies with higher intervention rates, especially **L4-CONN**, **L5-MIX**, and **L6-FREE**, are again associated with larger performance drops. These results show that the instability of parser-agnostic clarification is not specific to LLM-based parsers, but appears across parser families. On MSDC, SDDP exhibits the same qualitative pattern: conservative edits (**L0-TYPO**, **L1-NORM**) remain near-neutral, while **L4-CONN** causes the largest degradation (20/79 fixes vs. regressions), consistent with the LLM parsers on the same dataset.

One caveat: since choosing an appropriate connective is itself a hard generation task (Liu et al., 2024), **L4-CONN**’s regressions may partly reflect poor marker choice rather than distribution shift. We read distribution shift as the main driver, since the easy-to-execute conservative strategies (**L0-TYPO**, **L1-NORM**) still edit non-trivially yet stay near-neutral, and leave fully disentangling the two to future work.

### 5.3 What Works and What Remains Challenging?

First, clarification is effective on a well-defined subset of cases. The repairability analysis (§5.1) shows that a non-trivial portion of errors are consistently recoverable through last-utterance rewriting, establishing genuine headroom for improvement. The parser-aware RL results further demonstrate that **learning-based approaches can align with this structure**: RL increases repair rates on more repairable instances and even exhibits exploratory capability, occasionally discovering successful rewrites beyond the finite rewriting candidates. In addition, across parser families, conservative interventions (**L0-TYPO**, **L1-NORM**) provide stable behavior with minimal regressions, indicat-

ing that **safe intervention regimes are achievable**.

Meanwhile, clarification is inherently selective. A large portion of errors might lie outside the reach of local rewriting, and more aggressive interventions — while enabling additional repairs — introduce more regression risk. This trade-off appears consistently across both LLM-based and structured parsers, suggesting it reflects a general property of discourse parsing rather than a model-specific limitation. In the parser-aware setting, RL can influence global intervention behavior (e.g., learning to abstain), but does not yet yield a stable, fine-grained per-instance decision policy across datasets. Together, these results indicate that the main challenge is not only generating better rewrites, but **reliably identifying when rewriting is beneficial**.

These findings point to several promising directions. First, the structured repairability landscape motivates **rewritability prediction** as a key next step: estimating whether an instance is locally repairable before applying any intervention. Second, the presence of unrecoverable cases suggests moving beyond single-utterance rewriting **toward broader, context-level interventions** that can address dependencies outside the local window. More broadly, our results refine a common agentic-AI pattern: when intervention is limited to local, prompt-based rewriting over a fixed interface, gains are constrained by both repairability and the difficulty of selective application. This suggests that effective pipelines require **selective intervention, broader control over context, and better alignment with downstream signals**. *This extends beyond rewriting: upstream components — planning, routing, tool orchestration — are unlikely to be reliably effective as generic, prompt-driven front-ends; their success depends on being selective, context-aware, and tightly coupled with downstream structure and feedback.*

### 5.4 Qualitative Analysis

**SFT parser sensitivity to surface form.** Fixing a typo or expanding an abbreviation can flip the parser’s discourse prediction, even when the meaning is unchanged to human (see Table 11 in the appendix). On the fix side, normalizing *yeah* to *yes* changes the parser’s prediction from Acknowledgement to the correct Question-answer\_pair; expanding the abbreviation *go nz* to *go New Zealand* shifts a mispredicted Continuation into the correct Result. On the regression side, expanding *lol* to *laughing out loud* breaks a correct Comment prediction



| Original               | RL rewrite                                       | Candidate rewrites                                  | Gold                   |
|------------------------|--------------------------------------------------|-----------------------------------------------------|------------------------|
| A really               | oh, you mean 7 again?                            | {“really”}                                          | CLARIFICATION_QUESTION |
| B Sorry, sheep         | Sorry, I meant sheep                             | {“But, Sorry, sheep”, “Sorry, ore”, “Sorry, sheep”} | CORRECTION             |
| C I have no sheep :) ) | I don’t have any sheep but I’m offering wheat :) | {“But I have no sheep :)”, “I have no sheep :)”}    | ELABORATION            |
| D tough times..        | I’m out of resources.                            | {“So tough times..”}                                | EXPLANATION            |

Table 7: RL-generated rewrites that lie outside the LLM rewrite candidate pool yet improve parser output. The “Candidate rewrites” column lists only representative examples from the rewrites produced by the template strategies (L2-EXPL–L6-FREE); additional candidates may exist but are omitted for space. The RL rewrite is absent from the complete pool for every utterance.

to Elaboration; capitalizing *peurto rico* to *Puerto Rico*—a change no human would register—causes a correct Contrast to be mispredicted as Comment. These cases show the SFT parser has not fully abstracted away lexical surface form, which is part of what motivates the rewriting pipeline and the shift to RL training on parser reward.

**When not to rewrite, and when rewriting is not enough.** The RL agent also learns when to abstain. Bare fragments like “*what for?*” and non-verbal tokens like “^^” are already parsed correctly in context; the agent leaves them unchanged, avoiding the noise that would come from always appending a paraphrase. At the same time, some errors lie beyond the reach of any surface rewrite. When “*if you roll a 7*” is completed to “*if you roll a 7, you can move him*”—a factually correct answer—both the original and the rewrite still attach to the wrong antecedent question. The mistake is in discourse antecedent selection across two consecutive questions in a multi-party context, a structural ambiguity that rephrasing the answer cannot resolve.

**RL discovering rewrites beyond the LLM rewriting candidates.** Table 7 gives four cases where the RL rewrite is absent from the LLM rewrite candidate pool (§5.1, §5.3). Rule strategies produce only surface variants within their rules; RL instead, guided by parser reward, escapes these constraints in two distinct ways.

When the candidate rewrites cannot effectively recover the context, RL recovers the correct discourse function from scratch: recognizing that “*really*” functions as a Clarification\_question and rewriting explicitly with a question mark accordingly (“*oh, you mean 7 again?*”). In row B, RL instead finds the minimal precise repair—inserting “*I meant*”—to mark the self-correction explicitly as a Correction (“*Sorry, I meant sheep*”). In row C the candidates can only prepend “*But*”; RL reads the preceding turn (“*I’m definitely giving wheat*”) and incorporates it explicitly, recovering an Elabo-

ration relation the candidate rewrites cannot reach. And in row D it can only prepend a connective (leaving the utterance idiomatic), yet RL translates the colloquial expression into an explicit proposition (“*tough times..*” → “*I’m out of resources.*”), recovering an Explanation relation the candidate rewriting strategies cannot reach. Together, these cases suggest RL acquires a sense of discourse function that template strategies, by design, cannot express.

## 6 Conclusion and Future Work

We examine a simple question: can making dialogue utterances clearer reliably improve a frozen discourse parser? Our results suggest a nuanced answer: clarification helps, but only under the right conditions. Two structural factors shape this outcome. First, repairability defines a natural boundary: many parser errors cannot be resolved through local rewriting of the final utterance, even under a best-of-8 candidate pool, while a smaller subset remains consistently repairable. Second, effectiveness depends on selectivity. Within this subset, naive clarification often causes more harm than benefit, whereas a parser-aware RL policy can reduce regressions by learning when to abstain. Together, these findings suggest that the key challenge is not just generating better rewrites, but deciding when to intervene. This reframes clarification as a selective intervention problem. A key next step is **rewritability prediction**: estimating whether an instance is locally repairable before rewriting. More broadly, the prevalence of unrecoverable cases motivates moving beyond single-utterance rewriting toward context-level interventions, while the limits of black-box interaction suggest tighter integration between clarifier and parser.

## Limitations

Our study examines last-utterance clarification within an incremental discourse parsing setting. While incremental discourse dependency parsing

(DDP) is commonly adopted to model the streaming and real-time nature of dialogue (Naim et al., 2025), this setting represents only one possible modeling choice and may limit the generality of our findings. We note several limitations below.

First, our evaluation considers two representative parsers from the generative and discriminative families. Although these parsers reflect common design choices, other parsing paradigms—such as graph-based (Afantenos et al., 2015; Perret et al., 2016) and transition-based approaches (Wang et al., 2017)—were not examined and may exhibit different sensitivities to clarification.

Second, our experiments are restricted to a single discourse formalism. It remains an open question whether similar patterns would be observed under alternative frameworks, such as RST (Mann and Thompson, 1988) or PDTB (Prasad et al., 2008).

Third, due to computational budget constraints and the scale of experimentation, we rely on a single type of open-weight LLM (e.g., Qwen3 in 8B and 14B). Results may differ with other model families or more recent frontier LLMs.

Finally, we investigate seven representative clarification strategies with manually selected conservative prompts optimized for meaning preservation, which capture common classes of linguistic edits but do not aim to exhaustively enumerate the space of possible clarification operations. We also do not directly compare against supervised clarification (Fan et al., 2025) or evaluate a learned rewrite-or-not classifier; both are natural baselines that our repair/regression framework enables as future work.

## Acknowledgments

The authors wish to thank the anonymous reviewers and members of the OU NLP group for their valuable feedback. This research was supported by NSF ACCESS Allocation Request CIS250873, a compute grant from Modal.com, and a GPU gift from NVIDIA Corporation.

## References

Stergos Afantenos, Eric Kow, Nicholas Asher, and J  r  my Perret. 2015. [Discourse parsing for multi-party chat dialogues](#). In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 928–937, Lisbon, Portugal. Association for Computational Linguistics.

Berfin Aktas and Michael Roth. 2025. [Clarifying under-](#)

[specified discourse relations in instructional texts](#). In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 12237–12256, Vienna, Austria. Association for Computational Linguistics.

Nicholas Asher, Julie Hunter, Mathieu Morey, Benamara Farah, and Stergos Afantenos. 2016. [Discourse structure and dialogue acts in multiparty dialogue: the STAC corpus](#). In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, pages 2721–2727, Portoro , Slovenia. European Language Resources Association (ELRA).

Joanna Bitton, Maya Pavlova, and Ivan Evtimov. 2022. [Adversarial text normalization](#). In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Industry Track*, pages 268–279, Hybrid: Seattle, Washington + Online. Association for Computational Linguistics.

Jon Z. Cai, Brendan King, Peyton Cameron, Susan Windisch Brown, Miriam Eckert, Dananjay Srinivas, George Arthur Baker, V Kate Everson, Martha Palmer, James Martin, and Jeffrey Flanigan. 2025. [In search of the lost arch in dialogue: A dependency dialogue acts corpus for multi-party dialogues](#). In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 20135–20149, Vienna, Austria. Association for Computational Linguistics.

Jon Z Cai, Brendan King, Margaret Perkoff, Shiran Dudy, Jie Cao, Marie Grace, Natalia Wojarnik, Ananya Ganesh, James H Martin, Martha Palmer, and 1 others. 2023. [Dependency dialogue acts—annotation scheme and case study](#). *arXiv preprint arXiv:2302.12944*.

Zhiyu Cao, Peifeng Li, and Qiaoming Zhu. 2025. [ICR: Iterative clarification and rewriting for conversational search](#). In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 9810–9824, Suzhou, China. Association for Computational Linguistics.

Danqi Chen and Christopher D Manning. 2014. [A fast and accurate dependency parser using neural networks](#). In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 740–750.

Ta-Chung Chi and Alexander Rudnicky. 2022. [Structured dialogue discourse parsing](#). In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 325–335, Edinburgh, UK. Association for Computational Linguistics.

Eunsol Choi, Jennimaria Palomaki, Matthew Lamm, Tom Kwiatkowski, Dipanjan Das, and Michael Collins. 2021. [Decontextualization: Making sentences stand-alone](#). *Transactions of the Association for Computational Linguistics*, 9:447–461.

- Yaxin Fan, Peifeng Li, and Qiaoming Zhu. 2025. [Improving dialogue discourse parsing through discourse-aware utterance clarification](#). In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 18800–18816, Vienna, Austria. Association for Computational Linguistics.
- Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shitong Ma, Peiyi Wang, Xiao Bi, and 1 others. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.
- Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.
- Pengcheng Jiang, Jiacheng Lin, Lang Cao, Runchu Tian, SeongKu Kang, Zifeng Wang, Jimeng Sun, and Jiawei Han. 2025. [Deepretrieval: Hacking real search engines and retrievers with large language models via reinforcement learning](#). In *Second Conference on Language Modeling*.
- Dayeon Ki and Marine Carpuat. 2025. [Automatic input rewriting improves translation with large language models](#). In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 10829–10856, Albuquerque, New Mexico. Association for Computational Linguistics.
- Alex Lascarides and Nicholas Asher. 2007. Segmented discourse representation theory: Dynamic semantics with discourse structure. In *Computing meaning*, pages 87–124. Springer.
- Chuyuan Li, Yuwei Yin, and Giuseppe Carenini. 2024. [Dialogue discourse parsing as generation: A sequence-to-sequence LLM-based approach](#). In *Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 1–14, Kyoto, Japan. Association for Computational Linguistics.
- Jiaqi Li, Ming Liu, Min-Yen Kan, Zihao Zheng, Zekun Wang, Wenqiang Lei, Ting Liu, and Bing Qin. 2020. [Molweni: A challenge multiparty dialogues-based machine reading comprehension dataset with discourse structure](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2642–2652, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Zitong Li, Jiawei Li, Haifeng Tang, Kenny Zhu, and Ruolan Yang. 2023. [Incomplete utterance rewriting by a two-phase locate-and-fill regime](#). In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 2731–2745, Toronto, Canada. Association for Computational Linguistics.
- Jiacheng Lin, Tian Wang, and Kun Qian. 2025. Rec-r1: Bridging generative large language models and user-centric recommendation systems via reinforcement learning. *arXiv preprint arXiv:2503.24289*.
- Wei Liu, Stephen Wan, and Michael Strube. 2024. [What causes the failure of explicit to implicit discourse relation recognition?](#) In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 2738–2753, Mexico City, Mexico. Association for Computational Linguistics.
- Zhengyuan Liu and Nancy Chen. 2021. Improving multi-party dialogue discourse parsing via domain integration. In *Proceedings of the 2nd Workshop on Computational Approaches to Discourse*, pages 122–127.
- Xinbei Ma, Yeyun Gong, Pengcheng He, Hai Zhao, and Nan Duan. 2023. [Query rewriting in retrieval-augmented large language models](#). In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 5303–5315, Singapore. Association for Computational Linguistics.
- William C Mann and Sandra A Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text-interdisciplinary Journal for the Study of Discourse*, 8(3):243–281.
- Jannatun Naim, Jie Cao, Fareen Tasneem, Jennifer Jacobs, Brent Milne, James Martin, and Tamara Sumner. 2025. [Towards Actionable Pedagogical Feedback: A Multi-Perspective Analysis of Mathematics Teaching and Tutoring Dialogue](#). *arXiv preprint. ArXiv:2505.07161 [cs]*.
- Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, and 1 others. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744.
- Jérémy Perret, Stergos Afantenos, Nicholas Asher, and Mathieu Morey. 2016. Integer linear programming for discourse parsing. In *Proceedings of the 2016 conference of the north american chapter of the association for computational linguistics: Human language technologies*, pages 99–109.
- Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltasakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. [The Penn Discourse TreeBank 2.0](#). In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC’08)*, Marrakech, Morocco. European Language Resources Association (ELRA).
- Guangming Sheng, Chi Zhang, Zilingfeng Ye, Xibin Wu, Wang Zhang, Ru Zhang, Yanghua Peng, Haibin Lin, and Chuan Wu. 2024. Hybridflow: A flexible and efficient rlhf framework. *arXiv preprint arXiv:2409.19256*.



- Zhouxing Shi and Minlie Huang. 2018. [A deep sequential model for discourse parsing on multi-party dialogues](#). *Preprint*, arXiv:1812.00176.
- Zengkui Sun, Yijin Liu, Fandong Meng, Jinan Xu, Yufeng Chen, and Jie Zhou. 2024. [LCS: A language converter strategy for zero-shot neural machine translation](#). In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 9201–9214, Bangkok, Thailand. Association for Computational Linguistics.
- Zhongkai Sun, Yingxue Zhou, Jie Hao, Xing Fan, Yanbin Lu, Chengyuan Ma, Wei Shen, and Chenlei Guo. 2023. [Improving contextual query rewrite for conversational AI agents through user-preference feedback learning](#). In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track*, pages 432–439, Singapore. Association for Computational Linguistics.
- Kate Thompson, Akshay Chaturvedi, Julie Hunter, and Nicholas Asher. 2024a. [Llamipa: An incremental discourse parser](#). In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 6418–6430, Miami, Florida, USA. Association for Computational Linguistics.
- Kate Thompson, Julie Hunter, and Nicholas Asher. 2024b. [Discourse structure for the Minecraft corpus](#). In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 4957–4967, Torino, Italia. ELRA and ICCL.
- Rob van der Goot and Özlem Çetinoğlu. 2021. [Lexical normalization for code-switched data and its effect on POS tagging](#). In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2352–2365, Online. Association for Computational Linguistics.
- Yizhong Wang, Sujian Li, and Houfeng Wang. 2017. [A two-stage parsing method for text-level discourse analysis](#). In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 184–188, Vancouver, Canada. Association for Computational Linguistics.
- Zequ Wu, Yi Luan, Hannah Rashkin, David Reitter, Hannaneh Hajishirzi, Mari Ostendorf, and Garav Singh Tomar. 2022. [CONQRR: Conversational query rewriting for retrieval with reinforcement learning](#). In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 10000–10014, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
- Zhichao Xu, Shengyao Zhuang, Xueguang Ma, Bingsen Chen, Yijun Tian, Fengran Mo, Tao Li, Jie Cao, and Vivek Srikumar. 2026. [Rethinking on-policy optimization for query augmentation](#). *Transactions on Machine Learning Research*.
- An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, and 41 others. 2025. [Qwen3 technical report](#). *Preprint*, arXiv:2505.09388.
- Xiaopeng Ye, Chen Xu, Chaoliang Zhang, Zhaocheng Du, Jun Xu, Gang Wang, and Zhenhua Dong. 2025. [Q-PRM: Adaptive query rewriting for retrieval-augmented generation via step-level process supervision](#). In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 15113–15128, Suzhou, China. Association for Computational Linguistics.
- Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. BERTscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.
- Changtai Zhu, Siyin Wang, Ruijun Feng, Kai Song, and Xipeng Qiu. 2025. [ConvSearch-r1: Enhancing query reformulation for conversational search with reasoning via reinforcement learning](#). In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 26547–26564, Suzhou, China. Association for Computational Linguistics.

## A Parser-Agnostic Clarification and Dataset Statistics

**Meaning Preservation Analysis.** We report BERTScore between the original and rewritten utterances as a post-hoc diagnostic to verify that the clarifications are broadly meaning-preserving. These scores are computed on the test set, but are not used for prompt selection, model tuning, or evaluation, and do not affect the reported parsing results. Their role is purely to confirm that the rewriting strategies do not **substantially** alter the underlying intent of the utterance. Since no model or prompt decisions are based on these scores, reporting them on the test set does not introduce evaluation leakage.



| Dataset | Method   | Edit% (n_changed) | BERTScore-f1 |
|---------|----------|-------------------|--------------|
| stac    | L0-TYPO  | 9.76 (102)        | 0.9468       |
|         | L1-NORM  | 20.67 (216)       | 0.9369       |
|         | L2-EXPL  | 21.63 (226)       | 0.8907       |
|         | L3-COREF | 8.33 (87)         | 0.9180       |
|         | L4-CONN  | 46.22 (483)       | 0.9227       |
|         | L5-MIX   | 28.23 (295)       | 0.9367       |
| molweni | L6-FREE  | 49.86 (521)       | 0.9285       |
|         | L0-TYPO  | 42.52 (1671)      | 0.9721       |
|         | L1-NORM  | 44.81 (1757)      | 0.9653       |
|         | L2-EXPL  | 22.43 (906)       | 0.9511       |
|         | L3-COREF | 29.75 (1159)      | 0.9485       |
|         | L4-CONN  | 68.42 (2773)      | 0.9620       |
| msdc    | L5-MIX   | 69.06 (2711)      | 0.9609       |
|         | L6-FREE  | 88.19 (3466)      | 0.9451       |
|         | L0-TYPO  | 6.96 (342)        | 0.9517       |
|         | L1-NORM  | 12.31 (605)       | 0.9445       |
|         | L2-EXPL  | 14.63 (718)       | 0.9379       |
|         | L3-COREF | 12.39 (607)       | 0.9307       |
|         | L4-CONN  | 44.83 (2231)      | 0.9418       |
|         | L5-MIX   | 19.58 (963)       | 0.9364       |
|         | L6-FREE  | 45.34 (2229)      | 0.9331       |

Table 8: Average BERTScore over 3 clarification runs across three datasets via different clarification strategies.

## B Selected Prompts for Parser-Agnostic Clarification

The following boxes show the selected prompts for each clarification strategies we studied in our paper.

| L0-TYPO Prompt                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>You are an expert dialogue rewriting assistant. Given the dialogue context and the last utterance, you will clarify the last utterance, but only when it is necessary and safe. Your task is to:</p> <ol style="list-style-type: none"> <li>Identify if the last utterance has the following form-only issue: <ul style="list-style-type: none"> <li>A high-confidence typo/misspelling with a single obvious correction (no guessing).</li> </ul> </li> <li>Then choose one of the following actions: <ul style="list-style-type: none"> <li>If there is an above issue and the fix is safe, output the minimal allowed rewrite to fix only those issues.</li> <li>Otherwise, output the last utterance exactly unchanged.</li> </ul> </li> </ol> <p>Allowed edits:</p> <ul style="list-style-type: none"> <li>Fix high-confidence typos with one obvious correction.</li> </ul> <p>Important:</p> <ul style="list-style-type: none"> <li>If you are not fully confident the edit is meaning-preserving, output the last utterance unchanged.</li> <li>Do NOT make any edits beyond the allowed edits above.</li> <li>Do NOT add new facts or information not explicitly in the context.</li> <li>Do NOT change meaning and intent.</li> <li>Do NOT change stance/tone/sentiment. Minimal casing/punctuation changes are allowed only as required by the allowed edits.</li> </ul> <p>Show your reasoning only in &lt;think&gt; &lt;/think&gt; tags. Final output of the last utterance must be inside &lt;answer&gt; &lt;/answer&gt; as plain text (no speaker tag).</p> <p>Perform clarification for the following dialogue context and the last utterance:</p> <p>Context:<br/>{context}</p> <p>Last Utterance:<br/>{utterance}</p> |

| L1-NORM Prompt                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>You are an expert dialogue rewriting assistant. Given the dialogue context and the last utterance, you will clarify the last utterance, but only when it is necessary and safe. Your task is to:</p> <ol style="list-style-type: none"> <li>Identify if the last utterance has the following normalization issue: <ul style="list-style-type: none"> <li>It contains informal chat shorthand/abbreviation/slang that has ONE clear, widely accepted expansion in this context (no plausible alternative meanings).</li> </ul> </li> <li>Then choose one of the following actions: <ul style="list-style-type: none"> <li>If there is an above issue and the fix is safe, output the minimal allowed rewrite to fix only those issues.</li> <li>Otherwise, output the last utterance exactly unchanged.</li> </ul> </li> </ol> <p>Allowed edits:</p> <ul style="list-style-type: none"> <li>Expand the shorthand token(s) to their single canonical expansion.</li> <li>Make the minimum number of expansions needed.</li> </ul> <p>Important:</p> <ul style="list-style-type: none"> <li>If you are not fully confident the edit is meaning-preserving, output the last utterance unchanged.</li> <li>Do NOT make any edits beyond the allowed edits above.</li> <li>Do NOT add new facts or information not explicitly in the context.</li> <li>Do NOT change meaning and intent.</li> <li>Do NOT change stance/tone/sentiment. Minimal casing/punctuation changes are allowed only as required by the allowed edits.</li> </ul> <p>Show your reasoning only in &lt;think&gt; &lt;/think&gt; tags. Final output of the last utterance must be inside &lt;answer&gt; &lt;/answer&gt; as plain text (no speaker tag).</p> <p>Perform clarification for the following dialogue context and the last utterance:</p> <p>Context:<br/>{context}</p> <p>Last Utterance:<br/>{utterance}</p> |

| L2-EXPL Prompt                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>You are an expert dialogue rewriting assistant. Given the dialogue context and the last utterance, you will clarify the last utterance, but only when it is necessary and safe. Your task is to:</p> <ol style="list-style-type: none"> <li>Identify if the last utterance has the following explicitness issue caused by ellipsis/fragmentation: <ul style="list-style-type: none"> <li>The last utterance is NOT a complete, stand-alone clause on its own (e.g., a fragment, short answer, or bare reply), AND</li> <li>The missing material (predicate / object / complement) is uniquely and explicitly recoverable from the dialogue context, AND</li> <li>Adding the missing words would NOT introduce new facts, new intentions, or new discourse relations beyond what is already entailed by the context.</li> </ul> </li> <li>Then choose one of the following actions: <ul style="list-style-type: none"> <li>If there is an above issue and the fix is safe, output the minimal allowed rewrite to fix only those issues.</li> <li>Otherwise, output the last utterance exactly unchanged.</li> </ul> </li> </ol> <p>Allowed edits:</p> <ul style="list-style-type: none"> <li>Complete short answers/fragments by copying ONLY explicitly stated information from the context. <ul style="list-style-type: none"> <li>YES/NO answers: If the previous turn asks a yes/no question or proposes an action, you may expand "yes/no/okay/sure" into a full answer that repeats the proposition from context.</li> <li>WH-answers (time/place/object): If the previous turn asks "when/where/which/what", you may expand a fragment (e.g., "tomorrow", "at 5", "the red one") into a complete clause that answers that question.</li> <li>Justification-fragments: If the previous turn states the proposition being justified, you may attach it (e.g., "Because I was busy" -&gt; "I can't make it because I was busy") ONLY if that proposition is explicitly stated in context.</li> </ul> </li> <li>If the completion needs a subject, use only "I" (speaker of the last utterance) or "you" (addressing the other speaker) WHEN it is obvious from the previous utterances.</li> </ul> <p>Important:</p> <ul style="list-style-type: none"> <li>If you are not fully confident the edit is meaning-preserving, output the last utterance unchanged.</li> <li>Do NOT make any edits beyond the allowed edits above.</li> <li>Do NOT add new facts or information not explicitly in the context.</li> <li>Do NOT change meaning and intent.</li> <li>Do NOT change stance/tone/sentiment. Minimal casing/punctuation changes are allowed only as required by the allowed edits.</li> </ul> <p>Show your reasoning only in &lt;think&gt; &lt;/think&gt; tags. Final output of the last utterance must be inside &lt;answer&gt; &lt;/answer&gt; as plain text (no speaker tag).</p> |

Perform clarification for the following dialogue context and the last utterance:

Context:  
{context}

Last Utterance:  
{utterance}

### L3-COREF Prompt

You are an expert dialogue rewriting assistant. Given the dialogue context and the last utterance, you will clarify the last utterance, but only when it is necessary and safe. Your task is to:

- Identify if the last utterance has the following coreference issue:
  - The last utterance contains a pronoun or deictic (e.g., it/this/that/these/those/he/she/they/him/her/them), AND
  - The dialogue context contains exactly one unambiguous antecedent for it that is explicitly mentioned.
- Then choose one of the following actions:
  - If there is an above issue and the fix is safe, output the minimal allowed rewrite to fix only those issues.
  - Otherwise, output the last utterance exactly unchanged.

Allowed edits:

- Replace the pronoun/deictic with its antecedent ONLY when the antecedent is explicit and unambiguous.
- Make the minimum number of replacements needed.

Important:

- If you are not fully confident the edit is meaning-preserving, output the last utterance unchanged.
- Do NOT make any edits beyond the allowed edits above.
- Do NOT add new facts or information not explicitly in the context.
- Do NOT change meaning and intent.
- Do NOT change stance/tone/sentiment. Minimal casing/punctuation changes are allowed only as required by the allowed edits.

Show your reasoning only in <think> </think> tags. Final output of the last utterance must be inside <answer> </answer> as plain text (no speaker tag).

Perform clarification for the following dialogue context and the last utterance:

Context:  
{context}

Last Utterance:  
{utterance}

### L4-CONN Prompt

You are an expert dialogue rewriting assistant. Given the dialogue context and the last utterance, you will clarify the last utterance, but only when it is necessary and safe. Your task is to:

- Identify if the last utterance has the following discourse-marker issue:
  - The last utterance's intended discourse move is clear from the context, but the relation to the immediately preceding utterance is left implicit (e.g., topic shift, continuation/addition, contrast, consequence/next-step), AND
  - Adding a very short discourse marker would make this relation explicit WITHOUT changing the propositional content, intent, or tone, AND
  - There is ONE single obvious discourse marker that fits (no plausible alternatives).
- Then choose one of the following actions:
  - If there is an above issue and the fix is safe, output the minimal allowed rewrite to fix only those issues.
  - Otherwise, output the last utterance exactly unchanged.

Allowed edits:

- Add AT MOST ONE short discourse marker phrase, typically at the beginning of the last utterance, you may choose from the following set:
  - Topic shift/return: "By the way,", "Anyway,", "Back to that,"
  - Addition/continuation: "Also,", "And,", "In addition,"
  - Contrast/correction: "But,", "However,", "Instead,"
  - Result/next step: "So,", "Then,", "In that case,", "As a result,"
  - Example/clarification: "For example,", "For instance,", "In other words,"
- Do not add any other words besides the marker (and required comma/spacing).
- The marker's casing must match the style of the last utterance.

Important:

- If you are not fully confident the edit is meaning-preserving, output the last utterance unchanged.
- Do NOT make any edits beyond the allowed edits above.

- Do NOT add new facts or information not explicitly in the context.
- Do NOT change meaning and intent.
- Do NOT change stance/tone/sentiment. Minimal casing/punctuation changes are allowed only as required by the allowed edits.

Show your reasoning only in <think> </think> tags. Final output of the last utterance must be inside <answer> </answer> as plain text (no speaker tag).

Perform clarification for the following dialogue context and the last utterance:

Context:  
{context}

Last Utterance:  
{utterance}

### L5-MIX Prompt

You are an expert dialogue rewriting assistant. Given the dialogue context and the last utterance, you will clarify the last utterance, but only when it is necessary and safe. Your task is to:

- Identify if the last utterance has ANY of the following issues:
  - Form-only: a high-confidence typo/misspelling with one obvious correction.
  - Normalization: informal shorthand/abbreviation/slang with ONE clear expansion in this context.
  - Explicitness: ellipsis/fragment/underspecified answer where the missing words are uniquely recoverable from context and adding them is meaning-preserving.
  - Coreference: a pronoun/deictic with exactly one explicit, unambiguous antecedent in context.
  - Discourse marker: the connection to the immediately preceding utterance is clear but implicit, and exactly one obvious short discourse marker would make it explicit without changing meaning.
- Then choose one of the following actions:
  - If there is an above issue and the fix is safe, output the minimal allowed rewrite to fix only those issues.
  - Otherwise, output the last utterance exactly unchanged.

Allowed edits:

- Fix high-confidence typos (one obvious correction).
- Expand shorthand tokens to their single canonical expansion.
- Complete ellipsis/fragments ONLY when missing material is explicitly recoverable from context (no guessing).
- Replace pronouns/deictics with their explicit unambiguous antecedent (minimum replacements).
- Add AT MOST ONE discourse marker phrase at the beginning, you may choose from:
  - Topic shift/return: "By the way,", "Anyway,", "Back to that,"
  - Addition/continuation: "Also,", "And,", "In addition,"
  - Contrast/correction: "But,", "However,", "Instead,"
  - Result/next step: "So,", "Then,", "In that case,", "As a result,"
  - Example/clarification: "For example,", "For instance,", "In other words,"
- The marker's casing must match the style of the last utterance.

Important:

- If you are not fully confident the edit is meaning-preserving, output the last utterance unchanged.
- Do NOT make any edits beyond the allowed edits above.
- Do NOT add new facts or information not explicitly in the context.
- Do NOT change meaning and intent.
- Do NOT change stance/tone/sentiment. Minimal casing/punctuation changes are allowed only as required by the allowed edits.

Show your reasoning only in <think> </think> tags. Final output of the last utterance must be inside <answer> </answer> as plain text (no speaker tag).

Perform clarification for the following dialogue context and the last utterance:

Context:  
{context}

Last Utterance:  
{utterance}

### L6-FREE Prompt

You are an expert dialogue rewriting assistant. Given the dialogue context and the last utterance, you will clarify ONLY the last utterance to be clearer and easier to understand, while preserving the original meaning and intent. Your task is to:

- Identify if the last utterance has ambiguity or implicitness that could cause misunderstanding (e.g., typo, abbreviations, slang, vague references, incomplete phrasing, etc).
- Then choose one of the following actions:

```

- If yes and you can safely preserve meaning, rewrite the last
  utterance to be clearer.
- Otherwise, output the last utterance exactly unchanged.

Important:
- Do NOT add new facts or information not explicitly in the context.
- Do NOT change meaning and intent.
- Do NOT change stance/tone/sentiment.
- The rewrite must sound like a natural next turn after the given
  context and not disrupting the dialogue flow.
- Make only the minimal changes needed for clarity.

Show your reasoning only in <think></think> tags. Final output must be
inside <answer></answer> as plain text (no speaker tag).

Perform clarification for the following dialogue context and the last
utterance:

Context:
{context}

Last Utterance:
{utterance}

```

## C RL-based Parser-Aware Clarification

We now describe how we train the rewriting policy  $\pi_\phi$  with reinforcement learning to improve the frozen parser’s output. At a high level, each training example is treated as a one-step episode: the policy rewrites the current utterance, the parser returns a structured prediction, and a scalar reward is computed from the change in discourse quality.

**RL training with GRPO.** For each dialogue  $x = (u_1, \dots, u_T)$  and position  $t$ , we define a state  $s_t$  as the dialogue prefix and current utterance,  $s_t = (u_1, \dots, u_t)$ . The policy  $\pi_\phi$  takes  $s_t$  as input and generates a textual output that contains both reasoning and the clarified utterance. We form a clarified prefix by replacing  $u_t$  with  $\tilde{u}_t$  and run the frozen parser on both the original and clarified prefixes. The environment then returns a scalar reward  $r$  computed from these two outputs and the gold discourse annotation.

We optimize  $\pi_\phi$  with GRPO: for each state we sample a group of candidate clarifications, query the frozen parser on each, and compute each candidate’s advantage by normalizing its reward within the sampled group rather than learning a value network. To keep the policy close to the base LLM and preserve fluent, semantically plausible clarifications, we include a KL penalty that discourages large deviations from the initial policy distribution. Throughout training, the parser remains frozen and is only used to produce discrete discourse graphs for reward computation.

**Reward Design.** To facilitate effective RL training, we design verifiable reward signals tailored to the discourse parsing task. We decompose the reward into two parts: *format* rewards and *parsing* rewards.

We encourage models to follow a strict response format of the form `<think>...</think><answer>...</answer>`. The output must contain exactly one `<think>` block and one `<answer>` block in this order, with no stray text outside the tags. If the format is correct, we assign a format reward  $r_{\text{fmt}} = +1$ ; otherwise we assign a penalty  $r_{\text{fmt}} = -2$  and *do not* compute any parsing reward for this episode (i.e., we set  $r_{\text{parse}} = 0$ ). The total reward is

$$r = r_{\text{fmt}} + r_{\text{parse}}. \quad (1)$$

When the format is correct, we compute a parsing reward based on how the clarification changes the frozen parser’s prediction for the current step. Let  $c_{\text{orig}}$  indicate whether the parser’s prediction on the original utterance is fully correct (both link and relation match the gold), and  $c_{\text{rew}}$  indicate whether the prediction on the clarified utterance is fully correct. We also define  $p_{\text{rew}}$  to indicate a *partial match* on the clarified utterance: the parser attaches the link to the correct target but predicts an incorrect relation label. The parsing reward is then

$$r_{\text{parse}} = \begin{cases} +2 & \text{if } \neg c_{\text{orig}} \wedge c_{\text{rew}} \\ -2 & \text{if } c_{\text{orig}} \wedge \neg c_{\text{rew}} \\ +1 & \text{if } \neg c_{\text{orig}} \wedge \neg c_{\text{rew}} \wedge p_{\text{rew}} \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

Thus, the clarifier receives a large positive reward when it fixes an originally incorrect prediction, a large negative reward when it breaks an originally correct one, and a smaller positive reward when it at least corrects the attachment while still predicting the wrong relation. Neutral or ambiguous changes receive zero parsing reward. Together with the format term, this encourages the policy to produce well-structured outputs and clarifications that genuinely help the frozen parser on its structured discourse objective.

## D Experiment Setup

### D.1 LLM-based SFT Parser

Table 9 describes hyperparameters we used when finetuning Qwen3-8B (<https://huggingface.co/Qwen/Qwen3-8B>).

### D.2 SDDP Parser

We use the official code ([https://github.com/chijames/structured\\_dialogue\\_discourse\\_parsing](https://github.com/chijames/structured_dialogue_discourse_parsing)) released by the SDDP authors and train the model following their default hyperparameter settings.

| Hyperparameter | Value |
|----------------|-------|
| batch size     | 16    |
| optimizer      | AdamW |
| learning rate  | 5e-5  |
| weight decay   | 0.01  |
| lora dropout   | 0.05  |
| lora rank      | 64    |
| lora alpha     | 128   |

Table 9: Hyperparameters of finetuning Qwen3-8B on Stac, Molweni, MSDC

### D.3 GRPO-based Parser-aware Clarification

The follow table describes hyperparameters we used for RL training. We use the Instruct-tuned version of Qwen2.5-7B for all experiments.

| Hyperparameter       | Value |
|----------------------|-------|
| epoch                | 8     |
| max prompt length    | 1024  |
| max response length  | 768   |
| batch size           | 96    |
| rollout              | 8     |
| learning rate        | 1e-6  |
| sampling temperature | 0.7   |
| lora rank            | 32    |
| lora alpha           | 32    |

Table 10: Hyperparameters of GRPO training.

**Training Details.** All RL models are trained using the verl (<https://github.com/volcengine/verl>) framework and are conducted on 4 NVIDIA RTX 6000 96GB GPUs.

## E More Analysis on Results

Our results point to a central limitation of input-side intervention for frozen dialogue discourse parsers. The main bottleneck is not simply generating better rewrites, but aligning the scope of intervention with the source of parser error and identifying when intervention is warranted. Across parser-agnostic strategies, clarification rarely yields net gains because the same edits that create headroom for repairs also introduce regressions, often overwhelming the benefit of the repaired cases. This pattern holds across datasets and across both generative and discriminative parsers, suggesting that the instability is not tied to one particular model family, but to the mismatch between local surface editing and discourse decisions that depend on broader

structured context.

**Why is the selectivity problem so hard?** The regression patterns in Section 3.1 reveal that parser behavior is not a smooth function of surface form. Even semantically similar rewrites of the same utterance can produce opposite parser outcomes: one may correct an attachment error, while another creates a new one. These outcomes are not reliably predictable from surface form alone. Such unpredictability distinguishes discourse parsing from tasks like retrieval or translation where improvements in input quality often yield more gradual gains in output quality. The RL clarifier’s behavior in Section 4.2 reflects this directly: the policy learns *when not to act* but struggles to learn *how to act well*.

Negative signal from regressions is frequent and consistent, teaching the policy to be conservative. But positive signal from repairs is sparse and noisy — not because the policy generates poor rewrites, but because most errors fall in the unrepairable 80%, meaning that even high-quality rewrites receive zero reward. The policy cannot distinguish a good rewrite that fails due to structural unrepairability from a bad rewrite that fails on its own terms. This asymmetry explains why GRPO training converges toward selective abstention rather than toward better clarification. Recent work has increasingly sophisticated mechanisms for densifying rewards and improving credit assignment in exactly this kind of outcome-reward-through-a-frozen-model setup, including gradient attribution from the frozen judge, implicit process rewards derived from outcome labels, and entropy-weighted per-token credit assignment within GRPO. Our analysis suggests none of these would help in our setting because they do not address the scope mismatch described in Section 4.2, where 80% of errors originate outside the intervention window and are therefore unreachable regardless of how finely credit is attributed. The ceiling is structural rather than a consequence of coarse training signal.

**Why does surface clarification fail for discourse parsing?** Our results provide empirical support for a theoretically grounded explanation. In SDRT, discourse relations are determined through semantic inference over structured context, not through surface pattern matching. A parser trained on naturalistic dialogue learns to exploit subtle interactions between context and the current utterance, including implicit signals such as the absence of a connec-



| Eff.  | Strat. | Original                    | Rewrite                     | Gold  | Before | After |
|-------|--------|-----------------------------|-----------------------------|-------|--------|-------|
| fix   | L1     | <i>yeah</i>                 | <i>yes</i>                  | QAP   | ACK    | QAP   |
| fix   | L1     | <i>go nz</i>                | <i>go New Zealand</i>       | RES   | CONT   | RES   |
| regr. | L1     | <i>lol</i>                  | <i>laughing out loud</i>    | CMT   | CMT    | ELAB  |
| regr. | L0     | <i>not like peurto rico</i> | <i>not like Puerto Rico</i> | CONTR | CONTR  | CMT   |

Table 11: Surface-form changes that flip the Qwen3-8B parser output despite carrying no meaning difference for a human reader (STAC test). Before/After = parser prediction on original/rewritten utterance. QAP = Question-answer\_pair; ACK = Acknowledgement; CONT = Continuation; RES = Result; CMT = Comment; ELAB = Elaboration; CONTR = Contrast.

tive, a pronoun left unresolved, or a fragmentary form that signals continuation. When a clarifier makes these signals explicit, it does not simply add information; it replaces one surface realization with another that falls outside the parser’s learned distribution. This explains why connective insertion (L4-CONN) is consistently the most harmful strategy: it directly targets the discourse-level signal, but the parser was never trained on utterances with explicit connectives in positions where the original data left them implicit.

#### What should the community pursue instead?

Our findings point to three directions. First, the 80% unreachable ceiling suggests that multi-utterance or context-level rewriting may be necessary to address errors that originate outside the last utterance, such as long-range dependencies and context-side ambiguity. Second, the selectivity bottleneck motivates rewritability prediction as a standalone task: a lightweight classifier that estimates whether a given utterance is likely to benefit from rewriting before any rewrite is attempted. Such a classifier could be trained on the repair/regression labels generated by our experimental framework, and it could condition on parser uncertainty signals (e.g., entropy over candidate relations) to improve triggering precision. Third, our results suggest that tighter integration between clarifier and parser may be needed for consistent gains. Rather than treating the parser as a fully black-box system, approaches that expose partial parser state (attention distributions, candidate rankings) to the clarifier could enable more targeted interventions. This would move beyond the strict frozen-parser constraint we study here, but our findings suggest that this constraint is precisely what limits the approach.

More broadly, our results carry a cautionary message for agentic AI pipelines that compose upstream rewriters with frozen downstream tools. While this modular pattern works well for tasks

where surface form directly determines output quality (retrieval, translation), it is less effective for tasks governed by semantic inference, where the downstream tool’s decisions depend on implicit distributional signals that rewriting may disrupt. Evaluating such pipelines requires the kind of repair-vs-regression accounting we employ here, as aggregate metrics can mask harmful failure modes.