Hybrid LLM-Symbolic NLU for Qualitative Process Frame Generation

Xin Lian and Kenneth D. Forbus
Qualitative Reasoning Group, Northwestern University
x.lian@u.northwestern.edu, forbus@northwestern.edu

Abstract

Natural-language texts used in qualitative reasoning tasks are often syntactically complex and contain covert qualitative semantics, making it difficult for symbolic natural language understanding systems to generate useful formal representations consistently. We present a hybrid approach that combines LLMs with the Companion Natural Language Understanding (CNLU) system for generating qualitative process frame representations from natural-language causal statements. In our approach, LLMs simplify causal statements into constrained simple sentences and causal sentence pairs, while CNLU generates logical forms. We improve CNLU's ability to construct QP frames in cases where complete qualitative information is distributed across multiple lexical items. We evaluate the system on grounding facts from the QuaRTz dataset, and present pilot results demonstrating promising performance.

1 Introduction

Symbolic natural language understanding (NLU) systems produce structured, inspectable, and correctable representations, but often struggle with syntactic complexity and covert semantics from unrestricted language. These limitations become especially apparent in qualitative reasoning tasks, where natural language (NL) statements frequently imply changes, comparisons, and causal relationships between quantities without expressing them explicitly. The *Companion Natural Language Understanding* (CNLU) system [Tomai and Forbus, 2009] has been successfully demonstrated its usefulness in qualitative reasoning domains such as QuaRel [Tafjord *et al.*, 2019a; Crouse, 2021; Forbus and Demel, 2022; Forbus, 2025], but it does not consistently derive useful qualitative representations from general NL text. Long sentences, multiple interacting quantities, and implicit qualitative semantics often induce parsing failures or incomplete formal representations. For example, CNLU can readily derive a quantity representation from "The particles have energy," but comparative expressions such as "The particles have more energy" or "The particles have higher energy levels" require additional representations of comparison events relative to a previous state. More complex examples such as "The smoother the surface, the less friction against an object, so the faster it will move," introduces multiple entities, quantities, and causal relationships that can overwhelm CNLU. Large language models (LLMs), despite limitations in robust semantic reasoning [Asher *et al.*, 2023], are highly effective at linguistic transformation tasks such as paraphrasing and simplification. Recent work has explored incorporating LLMs into cognitive architectures and NLU systems (e.g., Nakos and Forbus [2023], Zhao and Forbus [2025], Lian and Forbus [2025]).

This paper explores a hybrid approach that tightly integrates LLMs with symbolic NLU systems. LLMs transform grounding facts into constrained simple sentences and causal sentence pairs, while CNLU is extended to derive more robust qualitative process (QP) frame representations [Forbus & Hinrichs, 2020]. The extensions address two classes of qualitative expressions: (1) cases in which the quantity type is explicitly expressed by some token, while other qualitative constituents of the QP frame are distributed across multiple tokens, including comparatives ("less friction"), change derivatives ("force decrease"), and quantity-neutral adjectives expressing qualitative values ("a great amount of blood"). (2) cases in which no quantity type is explicit by introducing a quantity *wrapper* for a non-quantity noun (e.g. interpreting "more atoms" in terms of an implicit count quantity). We evaluate the system on grounding facts from the QuaRTz dataset and present pilot results demonstrating promising performance in generating qualitative representations from NL.

2 Background

The Companion Natural Language Understanding (CNLU) system [Tomai and Forbus, 2009] is a knowledge-rich symbolic NLU system designed to interpret NL text and derive logical forms suitable for reasoning with the *Companion* cognitive architecture [Forbus and Hinrichs, 2017]. It integrates parsing, semantic disambiguation, interpretation, and abductive reasoning using the large knowledge base *NextKB*[1].

CNLU first tokenizes the input using its lexicon, and each token may introduce one or more *semantic translations* (*semtrans*) that map tokens to concepts in the ontology. For example, the noun *energy* may introduce semantics associated with the quantity type under the FrameNet[2] frame `FN_Dynamism`:

```
(fnSemtrans (TheList) Energy-TheWord Noun
  (and (isa :NOUN EnergyQuantity))
  (frame FN_Dynamism)
  (bindingTemplate (Thelist ……))
  (groupPatterns (TheSet ……)))
```

CNLU then parses the sentence using a highly modified version of the TRAINS parser [Allen, 1995]. Parsing proceeds bottom-up and may generate multiple candidate parse trees that combine semantics from the tokens to capture meaning. Since lexical items often have several possible meanings, disambiguation is required to determine which parse structure and semantic choices best fit the context. CNLU now supports several disambiguation strategies including heuristic rules, BERT-based frame scoring [Ribeiro and Forbus, 2021], and LLM-assisted disambiguation [Zhao and Forbus, 2025]. Once the semantic choices have been selected, interpretation rules derive logical forms suitable for downstream reasoning. For qualitative reasoning tasks, additional interpretation rules operate over the parse and semantic information to construct formal representations involving quantities, comparisons, and causal dependencies.

2.1 QP Theory and QP Frames

Qualitative Process (QP) *Theory* [Forbus, 1984] provides a formal theory for qualitative reasoning, and we focus here on quantities and their properties. *Qualitative proportionalities* (*qprops*) represent monotonic dependencies between quantities without requiring precise numerical values: a positive qprop indicates that increasing one quantity tends to increase another, whereas a negative qprop indicates an inverse relationship. Kuhene [2004] illustrated a close correspondence between QP Theory and NL semantics—NL contains syntactic patterns and semantic relations that can be used to construct representations of physical phenomena, showing a natural alignment between the organization of physical knowledge in QP theory and the structure of frame semantics. Quantities may be expressed either through quantity nouns and morphologically related words, such as *length*, *long*, and *lengthen*, or evoked by words that are not morphologically derived from a quantity noun, such as *hot*, which implies temperature. For the latter case, quantities may be conveyed by either *quantity-specific* or *quantity-neutral* adjectives and adverbs. For example, *fast* in "move fast" is quantity-specific because it evokes a speed or velocity quantity across contexts, while *high* in "high temperature" is quantity-neutral, since it does not independently identify a quantity type and instead specifies the relative magnitude of the contextually supplied quantity, temperature. NL also provides information about other constituents of qualitative processes through recurring syntactic and semantic patterns. For instance, ordinal relations may be expressed through quantity-specific comparisons, such as *hotter*, or through constructions involving quantity-neutral comparison terms, such as *more* in "more heat." Changes may be expressed directly by verbs or nouns such as *increase*, or by constructions that specify the direction or path such as "transfer from… to…" Interpreting these descriptions requires identifying the relevant quantities along with their qualitative constituents, including their qualitative values, ordering relations, and the processes or influences that cause them to change.

The resulting QP Theory constructs can be encoded in a frame-based format compatible with frame-semantic resources such as FrameNet. Subsequent work extended this frame notion to a broader range of qualitative models and narrative functions for extracting them from language [McFate *et al.* 2014; Forbus and Hinrichs 2020]. We collectively refer to these structures as *QP frames*, which include but are not limited to *quantity frames*, *ordinal frames*, and *influence frames*. A quantity frame represents a quantity associated with other qualitative information, including its entity, qualitative values, or the discourse variables from which the information is derived, if applicable. For example, "The particles have energy" produces a frame for the particles' energy:

```
(QuantityFrame1
  (isa QuantityFrame1 QuantityFrame)
  (quantityEntity QuantityFrame1 particle1)
  (quantityType QuantityFrame1 EnergyQuantity)
  (quantityTypeDV QuantityFrame1 energy1)
  (quantityEntityDV QuantityFrame1 particle1))
```

Qualitative information may also be introduced through adjectives or adverbs directly. For example, "the water is hot" introduces temperature with the qualitative value `Hot`:

```
(QuantityFrame2
  (isa QuantityFrame2 QuantityFrame)
  (quantityEntity QuantityFrame2 water1)
  (quantityEntityDV QuantityFrame2 water1)
  (quantityType QuantityFrame2 Temperature)
  (quantityValue QuantityFrame2 Hot))
```

An ordinal frame represents an ordering or comparison between quantities or between states of the same quantity. For example, "The book is heavier than the cup" produces quantity frames for the masses of the book and cup, together with an ordinal frame relating them:

```
(QuantityFrame3 ……
  (quantityEntity QuantityFrame3 book1)
  (quantityType QuantityFrame3 Mass) ……)
(QuantityFrame4
  (quantityEntity QuantityFrame4 cup1)
  (quantityType QuantityFrame4 Mass) ……)
(OrdinalFrame1 ……
  (ordinalReln OrdinalFrame1 greaterThan)
```

[1] https://www.qrg.northwestern.edu/nextkb/index.html

[2] https://framenet.icsi.berkeley.edu/

```
    (quantity1 OrdinalFrame1 QuantityFrame3)
    (quantity2 OrdinalFrame2 QuantityFrame4))
```

An influence frame represents a causal dependency between quantities. For example, "The density of the stone depends on its mass" produces an influence frame in which mass constrains density:

```
(InfluenceFrame1 ……
   (constrained InfluenceFrame1 QuantityFrame6)
   (constrainer InfluenceFrame1 QuantityFrame5)
   (influenceSign InfluenceFrame1 1)
   (influenceType InfluenceFrame1 IndirectInfluences))
```

Here the positive influence sign (1) indicates that the constrained quantity changes in the same direction as the constrainer. The influence type distinguishes direct from indirect influences; this paper focuses on indirect influences, or qprops.

2.2 *QuaRTz*

QuaRTz [Tafjord *et al.*, 2019b] is the first open-domain dataset for reasoning about comparative analysis problems. It includes 3,864 crowdsourced questions, each annotated with the properties being compared and a corresponding *grounding fact*, which is an NL statement conveying the causal qualitative knowledge required to answer the question. For example, the question:

> *When Katie was in the hospital with high blood pressure, she understood that the volume of her blood was*
> *(a) high (b) low*

is paired with the grounding fact:

> *As blood volume in the body increases, blood pressure increases.*

Answering the question requires identifying the positive causal relationship between blood volume and blood pressure. In QP terms, blood volume acts as the constrainer quantity, while blood pressure acts as the constrained quantity. Importantly, grounding facts often contain multiple quantities, entities, and causal relationships within a single sentence. Generating formal qualitative representations from these texts requires high level of both linguistic interpretation and qualitative semantic analysis.

2.3 LLM-Assisted NLU

Texts used in qualitative reasoning tasks are typically written in unrestricted everyday language like the ones in QuaRTz, while symbolic NLU systems like CNLU generally operate best on more constrained linguistic input. Recent work suggests that LLMs can help bridge this gap.

One promising direction is to use LLMs as hybrids with symbolic NLU systems. Because LLMs are highly effective at text transformation tasks such as paraphrasing and simplification, they can restructure NL input into forms that are easier for symbolic NLUs, or ideally, curated based on the NLUs' capability to process. Nakos and Forbus [2023], for example, proposed simplifying text before symbolic interpretation to improve CNLU's coverage. Similarly, Kirk et al. [2022, 2024] explored prompting LLMs to generate text that could be interpreted by symbolic cognitive agents.

Specifically, in work that uses LLMs to generate qualitative causal structures, Friedman et al. [2021] employ a domain-specific transformer architecture to predict the elements of a knowledge graph from an NL passage under a predefined schema with fixed sets of allowable elements; the graph nodes consist of token sequences extracted from the text. Similarly, Barres et al. [2025] use LLMs to construct QP causal networks from documents retrieved in response to NL queries. However, neither approach constructs or exploits ontology-grounded entities and relations.

Another direction is to use LLMs to improve specific stages of symbolic NLU systems. Zhao and Forbus [2025] incorporated LLMs into CNLU's semantic disambiguation process, while Lian and Forbus [2025] explored using LLMs to expand lexicons and semtranses.

In this paper, we focus on a hybrid approach. LLMs are used primarily as constrained linguistic preprocessors that simplify grounding facts and generate causal sentence pairs, while CNLU performs symbolic parsing and formal representation generation. Rather than replacing symbolic reasoning, the LLM component is intended to produce language that preserves the original semantics while being easier for interpretation and QP frame construction.

3 Task Description

Our goal is to generate QP frames, including influence-frame representations (qprops), from QuaRTz grounding facts. Consider the surface-object example: "The smoother the surface, the less friction against an object, so the faster it will move." The original sentence contains multiple quantities (surface smoothness, friction, event rate or speed), entities (surface and object), and causal relationships expressed within a single syntactically complex statement. Because such complexity can make parsing and interpretation unstable for symbolic NLU systems, our hybrid system first uses an LLM to transform the sentence into constrained simple sentences, each describing a single quantity or quantity change:

> *The surface is smoother.*
> *There is less friction against an object on the surface.*
> *The object moves faster.*

The LLM also identifies causal pairs among these sentences:

> *("The surface is smoother."*
> *"There is less friction against an object on the surface.")*
> *("There is less friction against an object on the surface."*
> *"The object moves faster.")*

CNLU then generates quantity and, when applicable, ordinal frames for each simplified sentence. For example, "The surface is smoother" produces quantity frames for the surface's smoothness and an ordinal frame representing a comparison to a previous state:

```
(QuantityFrame7 ……
   (quantityType QuantityFrame7 SurfaceSmoothness)
   (quantityEntity QuantityFrame7 surface1))
```

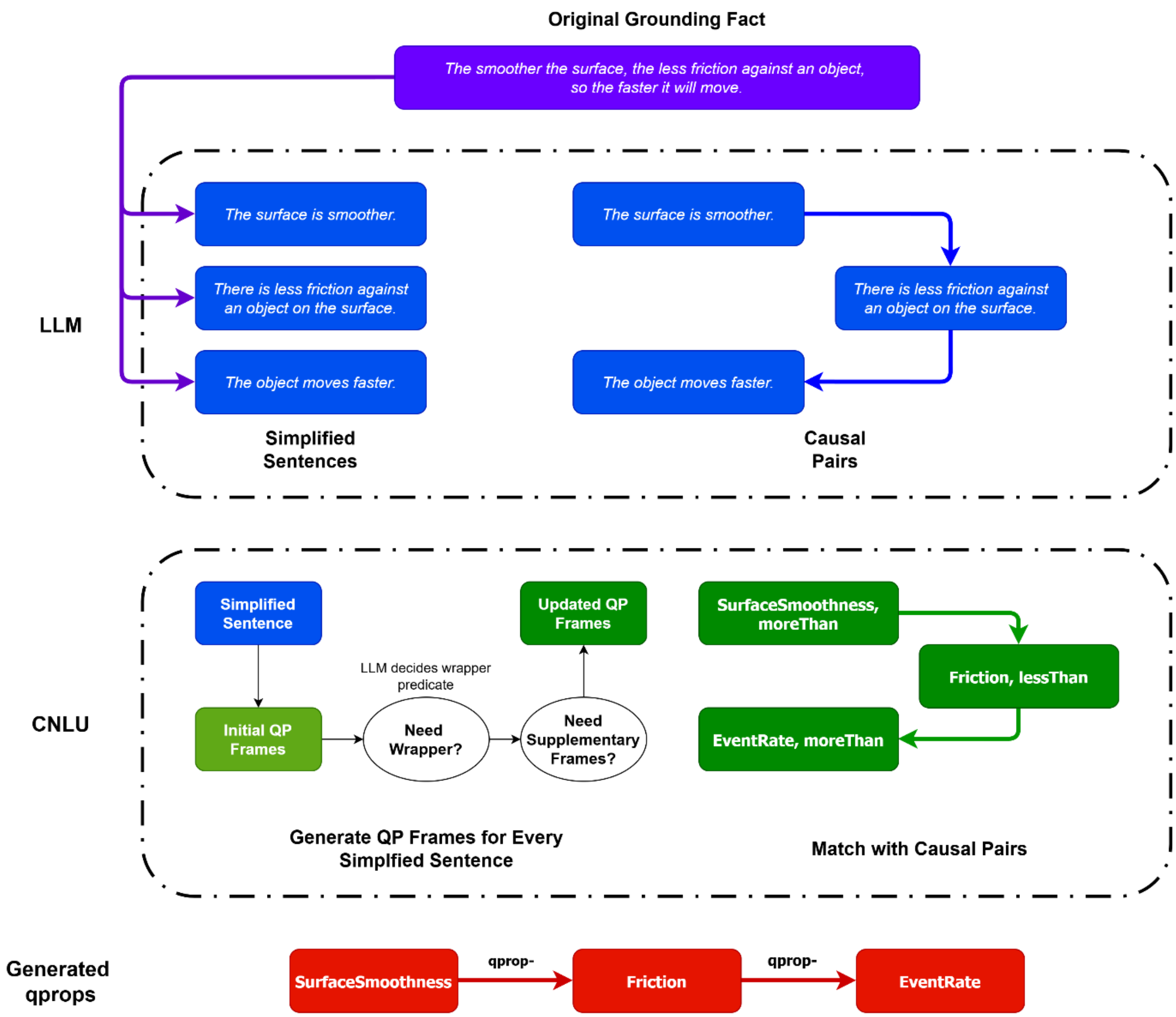

Figure 1: Procedure for deriving qprops from a QuaRTz grounding fact using the hybrid LLM-CNLU system. The LLM generates constrained simplified sentences and their causal pairs. For each sentence, CNLU generates initial quantity and ordinal frames, introduces quantity wrappers when comparative or change expressions modify nouns not recognized as quantity types, and derives supplementary frames when qualitative information is distributed across multiple tokens. After aligning and updating the frames, the system uses the causal pairs to derive qprops.

```
(QuantityFrame8 ……
  (quantityType QuantityFrame8 SurfaceSmoothness)
  (quantityEntity QuantityFrame8 (GapFn :NOUN)))
(OrdinalFrame2 ……
  (quantity1 OrdinalFrame2 QuantityFrame7)
  (quantity2 OrdinalFrame2 QuantityFrame8)
  (ordinalReln OrdinalFrame2 greaterThan))
```

The system combines these QP frames with the causal pairs to derive influence frames. For instance, the first pair produces a negative qprop from surface smoothness to friction:

```
(InfluenceFrame2 ……
  (constrained InfluenceFrame2 QuantityFrame9)
  (constrainer InfluenceFrame2 QuantityFrame7)
  (influenceSign InfluenceFrame2 -1)
  (influenceType InfluenceFrame2 IndirectInfluences))
```

Here `QuantityFrame9` represents the friction between the object and the surface. The resulting influence frames support downstream qualitative reasoning tasks, including answering QuaRTz questions.

4 Methodology

We use a hybrid LLM-symbolic pipeline consisting of three stages: (1) the LLM simplifies grounding facts into constrained simple sentences and corresponding casual sentence pairs; (2) CNLU generates quantity and ordinal frames from the simplified sentences; (3) CNLU pairs the generated frames based on causal sentence pairs to derive influence frames. An overview of the procedure is shown in Figure 1.

4.1 Sentence Simplification

The LLM-generated sentences must satisfy several constraints. First, they should be grammatical and structurally simple enough for CNLU to parse reliably. Ideally, each simplified sentence describes a single quantity, its value or change, and one primary entity when applicable. For example, back to the surface-object example, a less desirable simplification would be "The surface is smoother and has less friction against on," because it combines multiple quantities. These should instead be separated into "The surface is smoother" and "There is less friction against an object on the surface." Quantities associated with different entities or disjunctions in general should also be expressed separately. Consider "The larger a telescope's mirror or lens, the better it is at seeing narrowly separated objects as individual objects and the sharper the images look." A poor simplification would be "The mirror or lens of the telescope is larger," because the mirror and lens may participate different qualitative relationships even though they share the same quantity type here. Better would be: "The mirror of the telescope is larger" and "The lens of the telescope is larger."

Additionally, the LLM should preserve much of the original lexical content as possible, so CNLU can make use of more complete set of semantics. For example, "It moves faster" omits entity information available, better would be "The object moves faster." Such pronouns can complicate discourse-variable coordination and frame alignment.

Finally, the LLM should avoid introducing quantity terms or semantics absent from the original text. For example, the surface-object example could be reformulated as "The surface has less smoothness" or "The object has more speed," which would simplify QP-frame generation but introduce the quantity-denoting terms smoothness and speed that do not appear explicitly in the original text. We therefore use the LLM primarily for linguistic restructuring rather than semantic reinterpretation—preserving the original lexical content helps keep the simplified sentences semantically faithful to the grounding fact.

Overall, this stage produces sentences that are grammatically simple, semantically faithful to the original text, and structurally compatible with QP frame generation.

4.2 Causal Sentence Pairing

In addition to generating simplified sentences, the system identifies causal relationships among the events implied by each grounding fact. These relationships are required to derive influence frames (qprops) from the corresponding quantity and ordinal frames. The LLM pairs simplified sentences from the same grounding fact, with each ordered pair representing a constrainer event followed by a constrained event. For the surface-object example, it generates the causal pairs shown in Section 3. The prompt instructs the LLM to use only the simplified sentences, preserve the casual direction expressed in the original text, and avoid introducing intermediate events or additional semantics.

An alternative would be to generate explicit causal reformulations such as "The friction against the object depends negatively on the smoothness of the surface" and "The speed of the object depends negatively on the friction against the object" for the surface-object example. However, these formulations tend to be longer and syntactically more complex than the simplified sentences, making them less reliable inputs for CNLU and more likely to cause parsing instability where additional tokens tend to induce drastic changes in parsing, e.g. chart overflows. They also assign the LLM greater responsibility for formal semantic interpretation. Using ordered sentence pairs instead maintains a clearer separation between NL preprocessing and symbolic reasoning.

4.3 QP Frames Generation

For each LLM-generated simplified sentence, CNLU generates quantity and ordinal frames that are subsequently used to derive influence frames. Unless the sentence specifies otherwise, comparative expressions receive a forward temporal interpretation: the current state of a quantity is compared with its previous state. To improve QP frame generation, we extend CNLU to handle three broad classes of qualitative expressions: expressions (1) whose quantity and qualitative information are supplied directly by lexical semantics, (2) whose qualitative information is distributed across multiple tokens, and (3) implying quantities not directly introduced by any lexical item. We discuss each in turn.

Explicit quantity semantics. The simplest cases are those in which a lexical item directly introduces both a quantity type and its qualitative information. In the surface-object example, *smoother* and *faster* introduce `SurfaceSmoothness` and `EventRate` (or `Speed`), respectively. Similarly, *energy* in "The particles have energy" introduces `EnergyQuantity`, while *fast* in "The object moves fast" introduces `EventRate` together with a relatively high qualitative value. CNLU's existing mechanisms handle these cases because the relevant quantity semantics are localized within individual lexical items.

Supplementary frame generation. When qualitative meaning is distributed across multiple tokens, CNLU first generates quantity frames using its existing mechanisms, and then derives supplementary quantity or ordinal frames from the sentence's syntactic and semantic information. We consider four cases. (1) Comparatives modifying quantity nouns: In "There is less friction against the object on the surface," *friction* introduces the quantity type Friction, while the quantity-neutral comparative token *less* supplies the comparison. CNLU identifies the modified quantity and generates an ordinal frame with a `lessThan` relation to its previous state. We focus here on *more* and *less*. (2) Comparatives modifying quantity-specific adjectives or adverbs. In "The object moves more rapidly," *rapidly* introduces high `EventRate` while *more* supplies the comparison. CNLU generates a supplementary ordinal frame with a `greaterThan` relation; *less* rapidly instead yields `lessThan`. (3) Terms directly expressing quantity change. Words such as *increase* and *decrease* indicate directional change without independently supplying the quantity type. In “The gravitational force between them decreases,” *gravitational force* introduces `GravitationalForceQuantity` and *decreases* induces an ordinal frame relating its current and previous states with `lessThan`. We also consider related terms such as *rise*, *fall*, *gain*, and *lose*. (4) Quantity-neutral adjectives modifying quantity nouns. Adjectives such as *high*, *low*, *great*, and *small*, including their comparative forms, supply qualitative information but do not independently identify a quantity type. In “The heart expels a greater amount of blood with each stroke,” *amount* introduces the quantity type and *greater* induces an ordinal frame with a `greaterThan` relation.

Quantity wrappers. Some sentences imply quantity changes even though no lexical item directly introduces a quantity type. For instance, “The object has more atoms” compares the number of atoms associated with the object, although *atoms* denotes an entity type rather than a quantity type. When no appropriate quantity frame is initially generated, CNLU checks for comparative or change constructions involving nouns that can be interpreted quantitatively. It then wraps the ontologized concept with `CountableQuantityFn` for countable entities or `MassNounFn` for continuous quantities. Thus, *atoms* is represented as `(CountableQuantityFn Atom)`. CNLU then generates quantity and ordinal frames for the wrapped representation. The choice between `CountableQuantityFn` and `MassNounFn` is currently guided by the LLM using the noun and its linguistic context. This mechanism extends QP frame generation to everyday expressions in which quantities are implied rather than explicitly named.

Note that these classes extend Kuehne’s [2004] account by reorganizing and expanding the relevant linguistic constructions according to the information that CNLU must recover to generate QP frames.

4.4 qprops Derivation

After generating quantity and ordinal frames for all simplified sentences, the system aligns discourse variables by assigning shared variables to references that denote the same entities or quantities. It then combines the aligned QP frames with the corresponding causal sentence pairs to derive influence frames. For example, the pair

(“The surface is smoother.”
“There is less friction against the object on the surface.”)

yields an influence frame representing a negative relationship between `SurfaceSmoothness` and `Friction`, as shown in Section 3. The resulting influence frames serve as the final qualitative representations for downstream reasoning.

4.5 Evaluation

We evaluate four stages of qprop generation: simplified-sentence generation, causal-pair construction, quantity and ordinal frame generation, and qprop generation.

Simplified-sentence score. We first evaluate the LLM’s ability to generate simplified sentences. Each generated sentence receives a maximum score of 1 point, and it receives the full point when it (1) correctly describes a single event involving either a change in one quantity or a temporary qualitative value of one entity and (2) preserves all qualitative constituents needed to construct the corresponding QP frame when that information is present in the grounding fact. Partially correct sentences receive 0.5 points; incorrect sentences receive 0. For example, consider:

Leaves can have the following adaptations to reduce transpiration rate: small or narrow leaves to reduce the surface area over which water vapor is lost.

“Leaves are small” and “Leaves are narrow” each receive 1 point because each describes a single qualitative property of a single entity, while “Leaves are small or narrow” receives a half point for combining two alternative qualitative properties in one sentence.

Causal-pair score. We next evaluate the generated causal pairs, each consisting of a cause and an effect sentence. Each pair receives up to 3 points: 1 for the correct constrainer event, 1 for the correct constrained event, and 1 the correct causal direction. Partial credit is awarded when a missing causal relationship can be recovered indirectly from other causal pairs generated from the same grounding fact by inference. For the leaf example, the following pairs each receive full credit:

(“Leaves are small.”
“The surface area for water vapor loss is reduced.”)
(“Leaves are narrow.”
“The surface area for water vapor loss is reduced.”)
(“The surface area for water vapor loss is reduced.”
“The transpiration rate of leaves is lower.”)

Quantity and ordinal frame score. The extended CNLU generates quantity or ordinal frames for each simplified sentence. Each frame receives up to 5 points: 2 for the correct quantity type, 2 for the correct ordinal relation or qualitative value, and 1 for the correct associated entity. For “Leaves are narrow,” the following quantity frame receives full credit since it correctly identifies the quantity type, qualitative value, and relevant entity:

```
(QuantityFrame9
  (quantityEntity QuantityFrame9 leaf1)
  (quantityType QuantityFrame9 Distance)
  (isa QuantityFrame9 QuantityFrame)
  (quantityEntityDV QuantityFrame9 leaf1)
  (quantityValue QuantityFrame9 Narrow))
```

By contrast, the following frame for “The surface area for water vapor loss is reduced” receives 2 points—it identifies the correct quantity type `Area` but associates it with an inappropriate entity, here the water vapor rather than the surface, and does not represent the decrease:

```
(QuantityFrame10
  (quantityEntity QuantityFrame10 water-vapor1)
  (quantityType QuantityFrame10 Area)
```

```
(isa QuantityFrame10 QuantityFrame)
(quantityEntityDV QuantityFrame10 water-vapor1)
(quantityTypeDV QuantityFrame10 area1))
```

A complete representation would additionally encode either a supplementary ordinal relation frame with `lessThan` or a negative change derivative in this quantity frame (`-1`).

qprop score, precision, and recall. qprops are then generated from the QP frames associated with the simplified sentences and from the cause-effect positions of those sentences in the causal pairs. Each generated qprop receives up to 6 points: 2 for the correct constrainer quantity, 2 for the correct constrained quantity, and 2 for the correct influence sign. Given the causal pair

("leaves are narrow"
"the surface area for water vapor loss is reduced")

the following influence frame receives full credit because it links the appropriate constrainer and constrained quantity frames and assigns the correct positive influence sign:

```
(InfluenceFrame3
  (constrained InfluenceFrame3 QuantityFrame10)
  (constrainer InfluenceFrame3 QuantityFrame9)
  (influenceSign InfluenceFrame3 1)
  (influenceType InfluenceFrame3 IndirectInfluences)
  (isa InfluenceFrame3 InfluenceFrame))
```

We report three metrics for the generated qprops: (1) the average score vs. maximum score per qprop, (2) partial-credit precision, defined as the total credit awards divided by the maximum possible credit for all generated qprops, and (3) partial-credit recall, defined as the total credit awarded divided by the maximum possible credit for all gold qprops.

5 Pilot Results and Discussions

We used *Phi-4* [Abdin et al., 2024] as the LLM component because it produces relatively stable outputs while remaining relatively efficient based on our observations. A single rater evaluated all 81 distinct grounding facts in the QuaRTz test set. Table 1 reports performance at the four stages of the pipeline. For simplified sentence generation, most errors occurred when the LLM failed to separate multiple quantity-related events into independent simplified sentences each of which is supposed to describe one single quantity of one single entity. For quantity/ordinal frames or qprops generation, the results include cases where CNLU parsing failed because of timeouts or because lexical items were missing from the lexicon or lacked appropriate semtranses.

Many qprop-generation failures occurred because either the constrainer or the constrained sentence in a causal pair failed to produce usable QP frames, preventing the system from deriving the corresponding qprop. Consequently, although the grounding facts contained 129 gold qprops in total, the system only generated 64 qprops, some of which did not fully match their corresponding gold qprops. We expect the performance to improve as additional lexicon expansion, semantic translation expansion, and improved disambiguation

Evaluation Stage	**Avg Score/Max Score**
Simplified sentence	0.96/1.00
Causal pair	2.72/3.00
Quantity/ordinal frame	3.30/5.00
qprop	2.56/6.00
qprop-level Metric	**Value (%)**
Precision	85.29
Recall	42.31

Table 1: Evaluation results for the four stages of qprop generation from QuaRTz grounding facts. The upper panel reports the mean score relative to the maximum score at each stage. The lower panel reports partial-credit precision and recall for the generated qprops.

mechanisms are incorporated into the system [Lian and Forbus, 2025; Zhao and Forbus, 2025].

The evaluation also revealed several broader limitations. First, not all causal events correspond directly to quantity changes—some grounding facts contain causal events that trigger quantity changes without themselves expressing qualitative information about quantities explicitly. For example, given "the carbon taxes encourage people to use less fossil fuel, which reduces carbon dioxide emissions," one corresponding causal pair is

("Carbon taxes are in place." "People use less fossil fuel.")

The constrainer sentence describes the presence of a policy rather than an explicit quantity change, although it contrasts with a state in which no carbon tax is applied. The current system handles such events inconsistently. Second, some LLM-generated sentences still combine multiple events or remain syntactically complex. For example, "More of the solid is exposed and able to react," should ideally be decomposed into "More of the solid is exposed" and "More of the solid is able to react." These cases suggest the need for stronger prompting or output constraints when using a certain LLM. Third, supplementary-frame generation still covers a limited range of quantity-change expressions. Although the system handles terms such as *increase*, *decrease*, *rise*, and *fall*, it remains unreliable for semantically context-dependent verbs such as *stimulate*, *improve*, and *reduce*, which may imply qualitative changes or causal relationships in QuaRTz. Fourth, some simplified sentences remained too syntactically complex for reliable parsing. Finally, some grounding facts involve highly implicit quantity semantics that remain difficult for the current system to capture. For example, "the rays of the Sun are more focused" implicitly suggests that the exposed area of the rays becomes smaller, but the relationship is not expressed explicitly through lexical quantity semantics. These highly covert semantic relationships remain challenging for the current system and likely require more advanced interpretation mechanisms and commonsense knowledge.

Overall, the pilot results suggest that combining LLM-based linguistic preprocessing with extended symbolic QP frame generation is a promising direction for qualitative reasoning from natural language. The observed limitations highlight the importance of improving semantic coverage, supplementary-frame generation, and symbolic robustness for more complex qualitative language.

6 Summary and Future Work

In this paper, we presented a hybrid LLM-symbolic NLU system for generating QP frame representations from natural-language grounding facts. Our approach combines the linguistic transformation capabilities of LLMs with the structured representation and reasoning capabilities of CNLU. LLMs simplify grounding facts into constrained simplified sentences and generate causal sentence pairs through prompt engineering, while CNLU generates quantity, ordinal, and influence frames from the resulting text. We additionally extended CNLU to support several classes of qualitative linguistic constructions that frequently occur in everyday causal language. In particular, the system now handles cases where (1) complete qualitative information is distributed across multiple lexical items, (2) comparative constructions involving *more/less* and quantity-neutral qualitative adjectives, (3) derivative constructions involving *increase/decrease*, and (4) quantity-wrapper representations for nouns that do not explicitly denote quantity types, which are extended from scenarios of introducing qualitative information about quantities from NL from Kuehne [2004] and more specific to CNLU.

Pilot evaluation on grounding facts from the QuaRTz dataset suggests that this hybrid architecture is a promising direction for generating formal qualitative representations from unrestricted natural language. At the same time, the observed limitations highlight the continuing challenges of robust symbolic interpretation for covert qualitative semantics and syntactically complex language.

Several directions remain for future work. First, we plan to improve CNLU's qualitative interpretation capabilities further by extending supplementary-frame generation to additional classes of derivative and causal expressions. We also plan to improve the handling of highly covert quantity semantics that are only implied indirectly by linguistic context. Second, we intend to incorporate more advanced lexicon expansion, semantic translation expansion, and semantic disambiguation mechanisms [Lian and Forbus, 2025; Zhao and Forbus, 2025]. These additions should improve both semantic coverage and parsing robustness when processing unrestricted NL input. Third, we plan to investigate how analogical retrieval and analogical generalization can support qualitative reasoning and learning by reading at larger scales. Since Companions already supports analogical reasoning mechanisms, future systems may be able to retrieve structurally similar qualitative situations and use them to guide representation generation and causal inference. Finally, we are interested in more developed automatic reading strategies in which the system dynamically decides how to process and diagnose linguistic input. Rather than following a fixed pipeline, future versions of the system may learn when to simplify text further, expand lexical semantics, invoke supplementary-frame generation, or apply additional reasoning procedures depending on the characteristics of the input sentence, and these might be achieved via a solve-based approach with reading plans learning.

Overall, we believe that combining constrained LLM-based linguistic preprocessing with extensible symbolic NLU systems provides a promising path toward more robust qualitative reasoning from natural language.

Ethical Statement

There are no ethical issues.

Acknowledgments

This research was supported by the US Office of Naval Research.

References

[Abdin *et al.*, 2024] Abdin, M., Aneja, J., Behl, H., Bubeck, S., Eldan, R., Gunasekar, S., ... & Zhang, Y. Phi-4 technical report. *arXiv preprint arXiv:2412.08905*, 2024.

[Allen, 1995] Allen, J. *Natural Language Understanding (2nd ed)*. Benjamin-Cummings Publishing Co., Inc., 1995.

[Asher *et al.*, 2023] Asher, N., Bhar, S., Chaturvedi, A., Hunter, J., & Paul, S. Limits for learning with language models. In *Proceedings of the 12th Joint Conference on Lexical and Computational Semantics (* SEM 2023)*, pp. 236-248, 2023.

[Barres *et al.*, 2025] Barres, V., McFate, C. J., Kalyanpur, A., Saravanakumar, K. K., Moon, L., Seifu, N., & Bautista-Castillo, A., From Generating Answers to Building Explanations: Integrating Multi-Round RAG and Causal Modeling for Scientific QA. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 3: Industry Track)*, pp. 515-522. 2025.

[Crouse, 2021] Crouse, M. *Question-Answering with Structural Analogy*. PhD dissertation, Northwestern University, 2021.

[Forbus, 1984] Forbus, K. D. Qualitative process theory. *Artificial intelligence*, 24, 85–168, 1984.

[Forbus, 2026] Forbus, K. D. Exploring hybrid model formulation strategies for qualitative reasoning. In *Proceedings of the 38th International Workshop on Qualitative Reasoning*, 2026.

[Forbus and Demel, 2022] Forbus, K. D., & Demel, W. Integrating qr quantity representations with the semantic web: a progress report. In *Proceedings of the 35th International Workshop on Qualitative Reasoning*, 2022.

[Forbus and Hinrichs, 2017] Forbus, K. D., & Hinrich, T. Analogy and relational representations in the companion cognitive architecture. *AI Magazine*, 38.4 (2017): 34-42. 2017.

[Forbus and Hinrichs, 2020] Forbus, K. D., & Hinrich, T. Unifying instance-level and type-level QP frames for natural language understanding. In *Proceedings of the 33rd International Workshop on Qualitative Reasoning*. 2020.

[Friedman *et al.*, 2021] Friedman, S. E., Magnusson, I. H., & Schmer-Galunder, S. M. Extracting Qualitative Causal Structure with Transformer-Based NLP. In *Proceedings of the 34th International Workshop on Qualitative Reasoning*. 2021.

[Kirk *et al.*, 2022] Kirk, J. R., Wray, R. E., Lindes, P., & Laird, J. E. Improving language model prompting in support of semi-autonomous task learning. *arXiv preprint arXiv:2209.07636*. 2022.

[Kirk *et al.*, 2024] Kirk, J. R., Wray, R. E., Lindes, P., & Laird, J. E. Improving knowledge extraction from llms for task learning through agent analysis. In *Proceedings of the AAAI conference on artificial intelligence*, vol. 38, no. 16, pp. 18390-18398. 2024.

[Kuehne, 2004] Kuehne, S. E. *Understanding Natural Language Descriptions of Physical Phenomena*. PhD dissertation, Northwestern University, 2004.

[Lian and Forbus, 2025] Lian, X., & Forbus, K. D. LLM-Augmented Symbolic NLU System for More Reliable Continuous Causal Statement Interpretation. *Twelfth Annual Conference on Advances in Cognitive Systems*. 2025.

[McFate *et al.,* 2014] McFate, C., Forbus, K. D., & Hinrichs, T. Using narrative function to extract qualitative information from natural language texts. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 28, no. 1. 2014.

[Nakos and Forbus, 2023] Nakos, C., & Forbus, K. D. Using large language models in the companion cognitive architecture: A case study and future prospects. *Proceedings of the AAAI Symposium Series*. Vol. 2. No. 1. 2023.

[Ribeiro and Forbus, 2021] Ribeiro, D. N., & Forbus, K. D. Combining analogy with language models for knowledge extraction. In *3rd Conference on automated knowledge base construction*. 2021.

[Tafjord *et al.*, 2019a] Tafjord, O., Clark, P., Gardner, M., Yih, W. T., & Sabharwal, A. Quarel: A dataset and models for answering questions about qualitative relationships. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 7063-7071. 2019.

[Tafjord *et al.*, 2019b] Tafjord, O., Gardner, M., Lin, K., & Clark, P. QuaRTz: An open-domain dataset of qualitative relationship questions. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 5941-5946. 2019.

[Zhao and Forbus, 2025] Zhao, K., & Forbus, K. D. Integrating Symbolic Natural Language Understanding and Language Models for Word Sense Disambiguation. *Twelfth Annual Conference on Advances in Cognitive Systems*. 2025.