Frame First, Then Extract: A Frame-Semantic Reasoning Pipeline for Zero-Shot Relation Triplet Extraction

Zehan Li, Fu Zhang*, Wenqing Zhang, Jiawei Li,
Zhou Li, Jingwei Cheng Tianyue Peng
School of Computer Science and Engineering, Northeastern University, China
lizehan1999@163.com, zhangfu@neu.edu.cn

Abstract

Large Language Models (LLMs) have shown impressive capabilities in language understanding and generation, leading to growing interest in zero-shot relation triplet extraction (ZeroRTE), a task that aims to extract triplets for unseen relations without annotated data. However, existing methods typically depend on costly fine-tuning and lack the structured semantic guidance required for accurate and interpretable extraction. To overcome these limitations, we propose FrameRTE, a novel ZeroRTE framework that adopts a "frame first, then extract" paradigm. Rather than extracting triplets directly, FrameRTE first constructs high-quality Relation Semantic Frames (RSFs) through a unified pipeline that integrates frame retrieval, synthesis, and enhancement. These RSFs serve as structured and interpretable knowledge scaffolds that guide frozen LLMs in the extraction process. Building upon these RSFs, we further introduce a human-inspired three-stage reasoning pipeline consisting of semantic frame evocation, frame-guided triplet extraction, and core frame elements validation to achieve semantically constrained extraction. Experiments demonstrate that FrameRTE achieves competitive zero-shot performance on multiple benchmarks. Moreover, the RSFs we construct serve as high-quality semantic resources that can enhance other extraction methods, showcasing the synergy between linguistic knowledge and foundation models.

1 Introduction

Relation Triplet Extraction (RTE) aims to identify `<subject, relation, object>` triplets from unstructured text, which serve as foundational structures for downstream applications (Zhao et al., 2024). Recently, increasing attention has been paid to a more challenging variant, **Zero**-Shot **RTE** (**ZeroRTE**) (Chia et al., 2022), which aims to extract triplets for unseen relations without annotated data. Unlike **Zero**-Shot **R**elation **C**lassification (**ZeroRC**) (Chen and Li, 2021), which classifies relations between given entities, ZeroRTE must also detect the subject and object spans, making it considerably more difficult.

Most existing approaches to ZeroRTE rely on supervision from seen relations (Lv et al., 2023). They typically fine-tune two separate models for entity recognition and relation classification (Gong and Eldardiry, 2024), or train lightweight generative models to directly produce triplets (Kim et al., 2023). With the rise of Large-scale generative Language Models (LLMs), researchers have begun exploring their potential for ZeroRTE (Xu et al., 2024a). These efforts can be grouped into two categories. ***Tuning-based methods*** improve performance by generating pseudo-labeled data (Sun et al., 2024) or fine-tuning LLMs on annotated corpora (Gui et al., 2024), but they are often resource-intensive and less adaptable to rapidly evolving foundation models. In contrast, ***Tuning-free methods*** rely on chain-of-thought prompting (Wei et al., 2024) or in-context learning (Pang et al., 2023) to guide frozen LLMs using a few demonstrations. While more efficient and generalizable, these methods may still mislead LLMs if the demonstrations are inappropriately selected (Guo et al., 2025).

To address these limitations, considering that *determining the relation between entities often relies on contextual scenes and is triggered by lexical cues*, we for the first time introduce **Frame Semantics** (Fillmore, 1976), a theory that models words through structured conceptual scenarios known as **Frames**. We present a frame structure in Figure 1. Each frame includes a name, definition, Core and Non-Core Frame Elements (FEs), Lexical Units (LUs) that evoke the frame, and annotated demonstrations. Core FEs capture essential semantic roles, while Non-Core FEs provide supplementary con-

* Corresponding author.

Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, pages 27352–27365
November 4-9, 2025

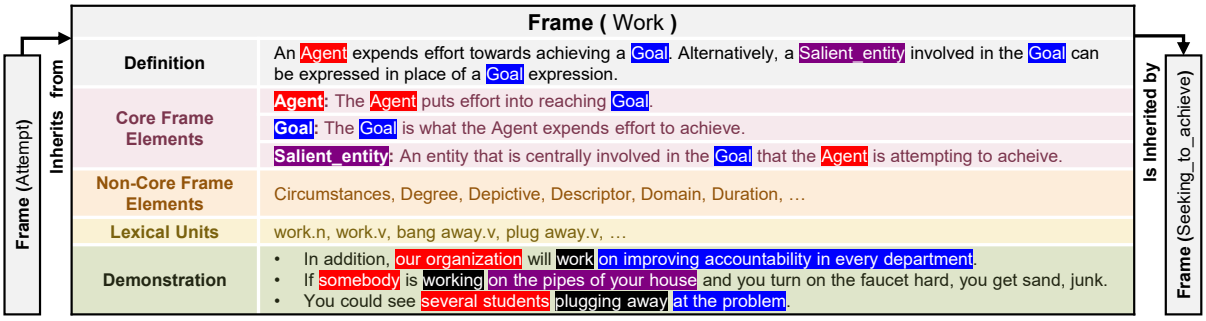

Figure 1: This figure illustrates the structure of the "Work" semantic frame as defined in the public resource FrameNet (Baker et al., 1998). The frame includes a *definition*, *core frame elements* (such as Agent, Goal, and Salient entity), *non-core frame elements*, associated *lexical units*, and *demonstrations*. It inherits from the "Attempt" frame and is inherited by the "Seeking_to_achieve" frame.

text. LUs act as lexical triggers, and frames are further connected through a hierarchical structure.

In this work, we propose **FrameRTE**, a zero-shot Relation Triplet Extraction system that integrates LLMs with Frame semantics to inject interpretable, structured knowledge into frozen models without fine-tuning. To address the limited coverage of open-domain relations in public frame resources, we design a pipeline to construct and enhance **R**elation **S**emantic **F**rames (**RSFs**) by retrieving top candidate frames via embedding similarity and generating structured RSFs through LLM-guided synthesis, aggregation, and demonstration construction. Based on these RSFs, we further propose a three-stage heuristic reasoning pipeline that emulates human-like semantic interpretation. We begin by evoking contextually relevant RSFs using frame definitions and LUs. Next, candidate triplets are generated under the structural constraints of the corresponding RSFs. Finally, we validate the predicted entities against the definitions of the Core FEs to ensure semantic consistency. This *frame first, then extract* paradigm introduces structured semantic guidance and improves interpretability in ZeroRTE. Our contributions are as follows:

- We introduce frame semantics into the RTE task for the first time by defining Relation Semantic Frames (RSFs) based on frame-semantic theory, and propose a method for RSFs construction and enhancement.
- Based on these RSFs, we propose a three-stage triplet extraction pipeline that incorporates heuristic prompts and interpretable semantics, effectively enhancing the extraction accuracy of frozen LLMs without fine-tuning.
- Experiments show that FrameRTE achieves competitive results across multiple benchmarks. Additionally, we demonstrate the transferability of RSFs as high-quality external semantic resources that can further enhance the performance of other methods.

2 Related Work

2.1 LLMs for Relation Extraction

The rise of generative Large Language Models (LLMs) has brought new research perspectives to relation extraction (RE), particularly in low-resource domains (Xu et al., 2024a). Existing methods address zero-shot RE through prompt engineering such as question answering (Zhang et al., 2023) and summarization prompting (Li et al., 2023). Recent studies have also explored LLM-driven data augmentation methods (Zhou et al., 2024), but these are limited to scenarios with given entities, which fall under the task of zero-shot relation classification (ZeroRC) (Chen and Li, 2021).

Zero-shot relation triplet extraction (ZeroRTE) is more challenging than ZeroRC, as it requires identifying the subject, object, and relation type simultaneously (Chia et al., 2022). Previous efforts have primarily relied on large-scale, well-annotated data with seen relation labels as supervision signals to fine-tune small-scale models (Li et al., 2025b; Kim et al., 2023; Li et al., 2024a). Recent work has also fine-tuned LLMs using large-scale external data (Li et al., 2024b; Gui et al., 2024), achieving notable performance (Zhang et al., 2025b). To alleviate the scarcity of labeled data, some methods leverage LLMs as data augmenters (He et al., 2024; Zheng et al., 2024), generating pseudo-instances via multi-round interaction (Xu et al., 2024b) or code annotation (Sainz et al., 2023), and subse-

quently fine-tune LLMs on the augmented data (Xue et al., 2024). However, these approaches often require costly data preparation and training, limiting their scalability to general-purpose LLMs. To address this, recent studies propose tuning-free methods, such as two-stage question answering (Wei et al., 2024) and multi-agent collaboration (Shi et al., 2024). Building on this direction, we introduce frame semantics as external guidance to unlock the potential of frozen LLMs for ZeroRTE.

2.2 Frame Semantics

Frame Semantics (Fillmore, 1976) is a linguistic theory that represents word meaning through conceptual structures called frames. Based on this theory, FrameNet (Baker et al., 1998) was developed as a large lexical resource, defining over 1,200 semantic frames. It has become a cornerstone for many NLP tasks, including semantic role labeling (Zheng et al., 2023; Ai and Tu, 2024), question answering (Shen and Lapata, 2007), and fact verification (Devasier et al., 2024).

In event extraction, frame semantics is especially useful due to the structural similarity between frames and events. Prior work has leveraged this to expand datasets (Liu et al., 2016; Huang et al., 2018), define event schemas (Wang et al., 2020, 2023), and enrich semantic features (Spiliopoulou et al., 2017). However, resources like FrameNet (Baker et al., 1998) still fall short of covering the full lexical diversity of real-world texts. Although some studies have attempted to extend FrameNet annotations (Pancholy et al., 2021; Cui and Swayamdipta, 2024), they primarily focus on frame completion tasks. Zhao et al. (2020) incorporate frame semantics into representations for relation classification, but their method lacks entity extraction and depends on distant supervision.

To address this gap, we introduce frame semantics into relation triplet extraction (RTE) for the first time. Considering FrameNet's rich, structured, and semantically grounded organization of event frames, which goes beyond surface-level syntactic roles to support context-sensitive modeling of predicate–argument structures, we propose a tuning-free framework that models relations as semantic frames and guides LLMs through a three-stage extraction process. This enables fine-grained semantic reasoning without additional training and advances zero-shot RTE with structured linguistic knowledge.

3 FrameRTE pipeline

3.1 Problem Formulation

Zero-shot **R**elation **T**riplet **E**xtraction (**ZeroRTE**) aims to extract a set of triplets $\mathcal{T} = \{(e_s, r, e_o)\}_{i=1}^{|\mathcal{T}_i|}$ from an input sentence $s \in \mathcal{X}$, where the relation type r is drawn from a predefined set $\mathcal{R}$ that lacks manually annotated data. Each triplet consists of a subject entity $e_s \in \mathcal{E}$ and an object entity $e_o \in \mathcal{E}$, with $\mathcal{E}$ denoting the set of potential entities in s, and r describing their interaction.

Prior work on ZeroRTE mainly adopts transfer learning, training on seen relations to generalize to unseen ones. While this satisfies the zero-shot condition via disjoint label sets, it still depends on human-labeled data. In contrast, we explore the feasibility of performing ZeroRTE using general-purpose LLMs without any supervised annotations.

3.2 Overview

We propose the **FrameRTE** system for zero-shot relation triplet extraction, as shown in Figure 2. It leverages **Relation Semantic Frames (RSFs)** to provide interpretable semantic priors for LLMs. The pipeline consists of two main modules:

The *RSFs construction* module (§3.3) retrieves relevant frames from FrameNet to guide LLMs in generating RSFs, which are then enhanced through aggregation and demonstration synthesis. The *triplet extraction* module (§3.4) follows a three-stage process: evoking candidate RSFs, extracting triplets guided by RSF structures, and validating them using Core FEs for semantic consistency.

3.3 RSF Construction and Enhancement

We argue that existing public semantic frame resources (e.g., FrameNet (Baker et al., 1998) and Chinese FrameNet (You and Liu, 2005)) fall short in adequately capturing relational semantics. To assess the coverage of relation types in existing frame resources, we conduct a statistical analysis on three representative RTE datasets. Specifically, we use the embedding model $\mathcal{M}_e$ (`text-embedding-3-small` [1]) to embed relation names and the frame definitions and lexical units of semantic frames. For each relation, the top 10 most similar candidate frames are retrieved based on embedding similarity, and human evaluation is conducted to determine whether these frames accurately express the relational semantics.

[1] https://platform.openai.com/docs/guides/embeddings

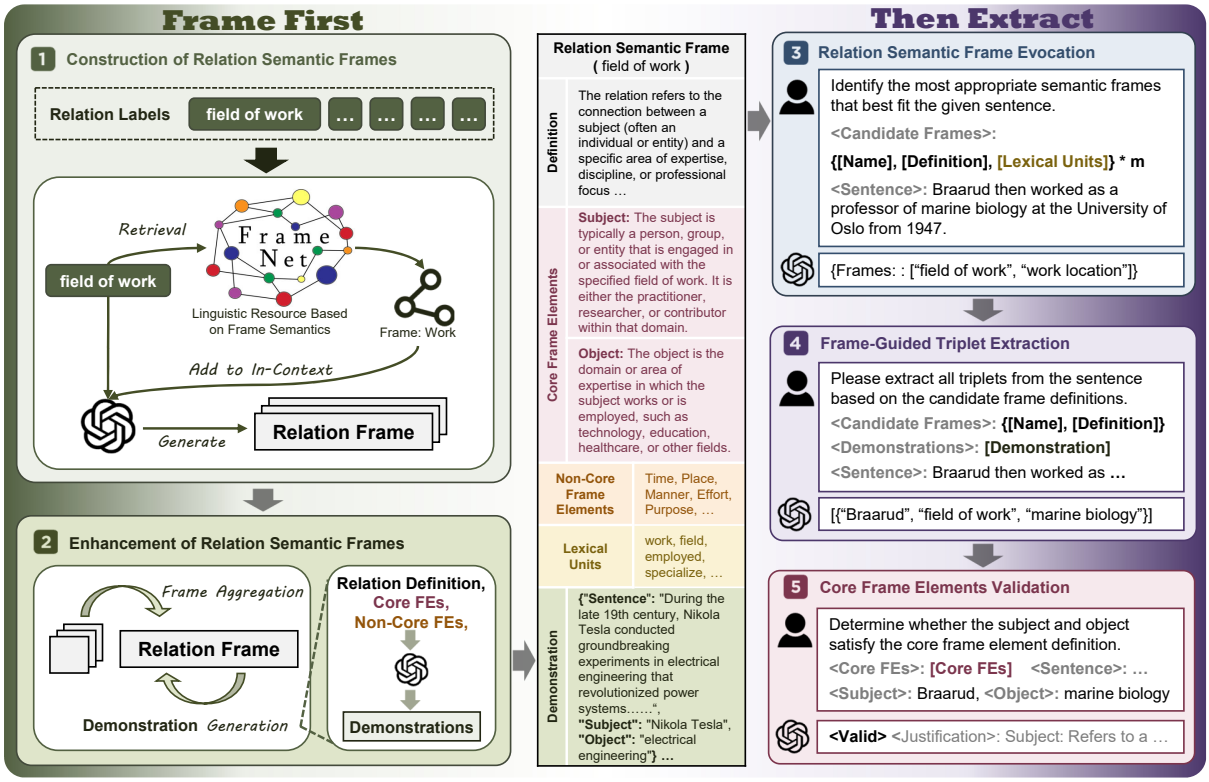

Figure 2: An illustration of the proposed **FrameRTE** pipeline for constructing and utilizing Relation Semantic Frames (RSFs). It consists of: **(1)** retrieving relevant FrameNet frames to guide RSFs generation, **(2)** enhancing RSFs via aggregation and demonstration synthesis, **(3)** evoking sentence-matched candidate RSFs, **(4)** extracting relation triplets guided by RSFs, and **(5)** validating triplets using Core FEs to ensure semantic consistency.

Metric	Wiki-ZSL	FewRel	DuIE
# Relation Types	113	80	48
Frame Coverage	49.56%	58.75%	22.92%
LU Coverage	43.36%	53.75%	-

Table 1: The table reports the number of relation types, Frame Coverage, and Lexical Unit (LU) Coverage across three datasets, which are detailed in Section 4.1.

Table 1 summarizes the coverage results. Existing resources cover only about half of the relation types, and some relations appear as LUs without a complete frame description, limiting their usefulness in relation modeling. To address this gap, we propose the **R**elation **S**emantic **F**rame (**RSF**), along with a two-stage construction method to more systematically model relational semantics.

3.3.1 Structural Components of RSF

We draw inspiration from the frame construction theory (Ruppenhofer et al., 2016) in FrameNet and define the components of RSF based on the characteristics of relational semantics. Figure 2 presents an illustrative example of an RSF, which includes the following components:

The **Definition** provides a concise explanation of the relation, highlighting the conceptual link between the subject and the object. **Core Frame Elements (FEs)** represent the essential semantic roles in the relation. As their meanings depend entirely on the specific relation and cannot be inferred from syntax alone, both the subject and object are designated as Core FEs. **Non-Core FEs**, such as time, place, manner, and purpose, enrich the semantic context and offer finer-grained expressiveness. **Lexical units (LUs)** are words or phrases that evoke the frame and serve as lexical triggers for identifying the relation in text. Finally, the **Demonstration** presents a concrete example sentence annotated with semantic roles, illustrating how the frame is instantiated in natural language.

3.3.2 Construction of RSFs

Given target relations, we first retrieve the top-k semantically relevant frames from FrameNet as prior knowledge for RSFs construction. Specifically, we employ the embedding model $\mathcal{M}_e$ to encode each relation r and all frame names ϕ in FrameNet. The

similarity between r and f is computed as:

$$\text{score}(r, \phi) = \cos(\mathcal{M}_e(r), \mathcal{M}_e(\phi)). \quad (1)$$

We then select the top-k frames with the highest similarity scores. Each retrieved frame is incorporated into a prompt template as contextual information, resulting in k distinct prompts p_i. These prompts are used to guide the LLM $\mathcal{M}_g$ (`GPT-4.1-mini`[2]) to generate initial RSFs that are semantically aligned with the target relation:

$$\mathcal{F}_{\text{initial}} = \{\mathcal{M}g(r, p_i)\}_{i=1}^{k}. \quad (2)$$

A simplified prompt example is shown below, and all complete prompt versions are provided in Appendix A. This process helps mitigate the randomness of LLM generation and promotes diversity among the constructed RSFs, which are further refined in the subsequent enhancement stage.

RSFs Construction Prompt

```
Your task is to generate a JSON-formatted
semantic frame for a given relation label.
Each frame consists of a name, a definition,
frame elements, and lexical units. You
may refer to manually annotated frames
from FrameNet that most closely resemble
the relation label to assist in generating.

<Manually Annotated Reference Frame>: ...
<Relation Label>: ...
<JSON Format>: ...
```

3.3.3 Enhancement of RSFs

To further improve the semantic expressiveness of the constructed RSFs while mitigating generation noise and enhancing diversity, we propose an RSFs enhancement method. Specifically, we first aggregate the k initial RSFs $\mathcal{F}_{\text{initial}}$. For each relation, we consolidate the descriptions of the Definition and Core FEs from the k candidates to produce a concise and semantically accurate version. Formally:

$$x' = \mathcal{M}g\left(\{x_i\}_{i=1}^{k}\right), x_i \in \{\text{Definition, Core FEs}\}. \quad (3)$$

In addition, we merge the Non-Core FEs and LUs across all k RSFs corresponding to the same relation to obtain the enhanced RSF:

$$\mathcal{F}_{\text{enhanced}} = x' \cup \text{Merge(Non-Core FEs)} \cup \text{Merge(LUs)}, \quad (4)$$

where $\text{Merge}(\cdot)$ denotes the union of all elements in each category across the k inputs.

Finally, we leverage $\mathcal{F}_{\text{enhanced}}$ to generate high-quality demonstrations for in-context learning with LLMs. Each prompt includes the Definition and Core FEs to ensure a clear relational pattern and subject-object alignment. Additionally, we explicitly instruct the LLMs to consider incorporating Non-Core FEs as auxiliary elements, thereby introducing greater diversity and semantic complexity into the demonstrations. Although the RSFs are primarily generated by LLMs, we apply minimal human revision to address occasional inconsistencies and ensure the overall quality of the RSFs.

3.4 Three-Stage Zero-Shot Triplet Extraction Driven by Relation Semantic Frames

To leverage Relation Semantic Frames (RSFs) for ZeroRTE, we design a three-stage reasoning framework that mirrors human interpretation: evoking contextual frames, extracting semantic roles, and validating their coherence.

Semantic Frame Evocation. Given an input sentence s, we first identify the most contextually relevant RSFs as the evoked background knowledge, so as to narrow the candidate space. As illustrated in stage (3) of Figure 2, we construct a structured prompt by concatenating each frame's definition and its associated lexical units (LUs). LUs act as semantic triggers in natural language and play a crucial role in evoking specific frames. This process guides the LLMs to generate the most likely activated frames, which serve as candidate relations for downstream triplet extraction.

Frame-Guided Extraction. After evoking the candidate RSFs, we leverage their semantic structure to guide the extraction of relation triplets from input sentence. As illustrated in stage (4) of Figure 2, we construct structured prompts based on each RSF's definition and high-quality demonstrations, and use them to instruct the LLM to generate semantically constrained candidate triplets. These demonstrations are pseudo-examples automatically constructed from the Core and Non-Core FEs, as described in Section 3.3.3, and are beneficial for improving in-context learning performance.

Core Frame Elements Validation. To mitigate the over-generation issue often observed in LLMs (Li et al., 2024d), we introduce a semantic consistency check based on Core Frame Elements (FEs). As illustrated in stage (5) of Figure 2, we incorporate the Core FEs definitions from the corresponding RSF into a validation prompt, guiding the LLM to assess whether the subject and object in a predicted triplet align with the expected semantic roles. A triplet is considered valid only if both the

[2]https://platform.openai.com/docs/models

subject and object conform to the Core FEs definitions, ensuring semantic precision in the final output. If the LLM returns <Invalid>, the triplet is discarded from the final results.

4 Experimental Setup

4.1 Dataset and Evaluation Metrics

We conduct experiments on three relation triplet extraction (RTE) datasets as follows: **Wiki-ZSL** (Chen and Li, 2021) and **FewRel** (Han et al., 2018) are two widely used benchmark datasets for the zero-shot RTE task. **Wiki-ZSL**, consisting of 113 relation types, is constructed via distant supervision from Wikipedia. **FewRel**, containing 80 relation types, is a manually annotated dataset originally designed for few-shot RTE. We follow the data splits and evaluation protocol proposed by Chia et al. (2022). Specifically, we randomly sample $m \in 5, 10, 15$ unseen relations to evaluate generalization under different levels of difficulty. To control evaluation cost, 1,000 test instances are sampled for each m setting. **DuIE** is a large-scale Chinese RTE dataset developed by Baidu, with data sourced from Baidu Baike and Baidu Tieba, covering 48 relation types. Following Shi et al. (2024), we evaluate these datasets using the full relation candidate set. We follow the settings of previous work and report the standard micro F_1 score.

4.2 Implementation Details

To ensure the quality of relation semantic frames (RSFs), we use `GPT-4.1-mini` during the frame construction stage and keep RSFs fixed for all subsequent extraction and verification processes. We adopt FrameNet (Baker et al., 1998) as the frame resource for Wiki-ZSL and FewRel, and use Chinese FrameNet (You and Liu, 2005) for DuIE. For these downstream tasks, we experiment with three LLMs: `GPT-3.5-turbo`, `GPT-4o`, and `Llama-4-Scout`, all accessed via official APIs without any fine-tuning. We set the temperature to 0.7 during frame construction to promote diversity, and fix it to 0 during extraction and verification for reproducibility. We limit the number of initial RSFs k to 5. The implementation code will be made publicly available.

4.3 Baselines

We compare FrameRTE with two representative categories of baselines for the ZeroRTE task.

Fine-tuning methods rely on supervised training over seen relations and aim to transfer learned knowledge to unseen relations. **REPrompt** (Chia et al., 2022) uses `GPT-2` (Radford et al., 2019) to generate synthetic data for unseen relations and then trains a `BART`-based extractor (Lewis et al., 2020) on the generated data. **TAG** (Xu et al., 2024b) extends this approach with a two-agent framework, introducing a cycle of attempting, criticizing, and rectifying to iteratively improve the quality of synthetic samples. **ZETT** (Kim et al., 2023) constructs relation-specific templates manually and directly generates entities for each unseen relation. Building on ZETT, **HCDR** (Li et al., 2025a) introduces a discriminative reranking task to improve ranking accuracy. More recently, Zhang et al. (2025b) demonstrated state-of-the-art performance by fine-tuning `LlamA2-13B-Chat` and `Qwen1.5-14B-Chat` using Low-Rank Adaptation (LoRA) (Hu et al., 2021), which is denoted as **FT**.

Prompting methods leverage general-purpose LLMs to directly generate triplets via carefully designed instructions and demonstrations. **ChatIE** (Wei et al., 2024) reformulates ZeroRTE as a two-stage QA task: first identifying potential relations, then extracting entities accordingly. **AgentRE** (Shi et al., 2024) adopts an agent-based framework incorporating retrieval, memory, and extraction modules; we apply its zero-shot setting in our comparison. To assess the basic capability of LLMs for ZeroRTE, we use a simple prompting baseline, referred to as **Vanilla Prompt**, which directly instructs LLMs to generate triplets. Additionally, we provide 3 demonstrations in the prompt to investigate the impact of **In-Context Learning**.

5 Results and Discussion

5.1 Main results

We evaluate the proposed FrameRTE on three public benchmark datasets, with Table 2 presenting a comprehensive comparison against existing fine-tuning (`FT`) and prompting (`PT`) methods.

Under the same frozen LLM backbone (e.g., `GPT-3.5-turbo`), FrameRTE consistently outperforms all prompting baselines. For instance, compared to the state-of-the-art multi-agent model AgentRE, FrameRTE achieves an average F1 gain of **9.07** across seven settings, highlighting the effectiveness of structured semantic frames in enhancing relational understanding.

In addition, FrameRTE also surpasses fine-tuned small-scale generative models (e.g., ZETT, HCDR) in most settings, even without parameter

Methods	Backbone	Paradigm	*Wiki-ZSL*			*FewRel*			*DuIE*
			m=5	m=10	m=15	m=5	m=10	m=15	*all*
REPrompt (Chia et al., 2022)†	BART & GPT-2	FT	30.01	31.19	28.85	22.34	24.61	20.08	-
TAG (Xu et al., 2024b)†	BART & GPT-2	FT	38.24	31.88	29.18	38.81	32.18	25.59	-
ZETT (Kim et al., 2023)†	T5-base	FT	31.74	24.87	21.21	33.71	31.28	24.39	-
HCDR (Li et al., 2025a)†	T5-base	FT	38.56	29.14	21.97	38.24	34.89	25.43	-
FT (Zhang et al., 2025b)†	Llama2-13B-Chat	FT	27.40	44.63	44.30	51.79	33.67	27.71	-
FT (Zhang et al., 2025b)†	Qwen1.5-14B-Chat	FT	30.09	46.57	44.87	52.88	33.74	27.41	-
Vanilla Prompt	GPT-3.5-turbo	PT	12.01	9.28	7.59	14.07	11.00	12.22	10.80
In-Context Learning	GPT-3.5-turbo	PT	22.60	18.30	13.81	28.43	18.51	16.12	13.18
ChatIE (Wei et al., 2024)	GPT-3.5-turbo	PT	19.72	15.66	13.81	23.78	22.35	20.89	27.82†
AgentRE (Shi et al., 2024)	GPT-3.5-turbo	PT	15.83	14.77	19.40	29.58	25.25	21.46	32.10†
FrameRTE (ours)	GPT-3.5-turbo	PT	36.88	31.72	22.30	34.75	33.91	28.23	34.07
Vanilla Prompt	GPT-4o	PT	20.58	16.23	15.74	34.62	25.20	24.06	14.84
In-Context Learning	GPT-4o	PT	20.55	23.97	17.39	26.96	22.27	26.24	16.08
FrameRTE (ours)	GPT-4o	PT	**39.14**	32.08	23.53	**38.62**	34.11	**29.47**	**36.82**
Vanilla Prompt	Llama-4-Scout	PT	22.36	19.31	16.53	32.47	25.97	20.83	21.69
In-Context Learning	Llama-4-Scout	PT	23.25	27.18	18.03	30.22	31.45	28.10	20.60
FrameRTE (ours)	Llama-4-Scout	PT	37.88	**34.06**	**25.21**	36.42	**35.35**	28.87	35.18

Table 2: Main results on three benchmarks. We report F_1 performance under the multi-triplet setting, where m represents the number of unseen relations. On DuIE, all 48 relation types are treated as unseen. Results marked with † are taken from the original papers. **Bold** indicates the highest score under the setting where LLMs parameters are frozen. FT refers to *Fine-tuning methods*, while PT refers to *Prompting methods*.

updates. However, consistent with previous findings, our prompting-based framework still lags behind fine-tuned LLM methods. Nevertheless, the performance gap has been substantially narrowed. For example, on the FewRel dataset (m=10,15), FrameRTE outperforms fine-tuned LLMs by **1.61** and **1.76** points respectively, without using any labeled data for tuning, demonstrating its advantage in complex low-resource scenarios.

Moreover, FrameRTE shows consistently strong performance across both closed-source GPT series models and the latest open-source backbones (Llama-4-Scout). Notably, stronger LLMs (e.g., GPT-4o vs. GPT-3.5-turbo) yield further improvements, with an average F1 gain of **1.70**, validating the scalability of our approach.

5.2 Ablation Study

To evaluate the effectiveness of each proposed module, we conduct ablation studies to assess the contribution of key components in Table 3.

RSFs Construction. When we do not retrieve similar frames from FrameNet to construct RSFs (**w/o Frame Retrieval**), performance drops slightly, indicating that these manually curated frame resources serve as useful priors for guiding RSF generation. When we skip the aggregation step and directly use the initial frames (**w/o Frame Aggregation**), performance degrades significantly,

Methods	*Pre.*	*Rec.*	F_1
FrameRTE	**30.04**	39.44	**34.11**
RSFs Construction			
-w/o Frame Retrieval	29.40	35.44	32.13
-w/o Frame Aggregation	24.46	29.63	26.80
-w/o Non-Core FEs	28.07	38.52	32.47
Triplet Extraction			
-w/o RSF Evocation	17.57	21.11	19.16
-w/o Demonstrations	23.47	32.78	27.36
-w/o Core FE Validation	29.36	**39.63**	33.73

Table 3: Ablation study of FrameRTE. We report the Precision (*Pre.*), Recall (*Rec.*), and F_1 performance (%) on FewRel under the setting of m=10.

suggesting that multi-round generation and aggregation help mitigate semantic noise caused by the stochastic nature of LLMs. In addition, incorporating Non-Core Frame Elements in the demonstrations (**w/o Non-Core FEs**) further improves performance by enhancing demonstration quality.

Triplet Extraction. Removing the RSF evocation stage (**w/o RSF Evocation**) leads to a substantial performance drop, highlighting the importance of narrowing the candidate relation space during extraction. Excluding demonstrations derived from RSFs (**w/o Demonstrations**) also causes a notable decline in performance, showing that synthesized demonstrations effectively guide the LLM. Skip-

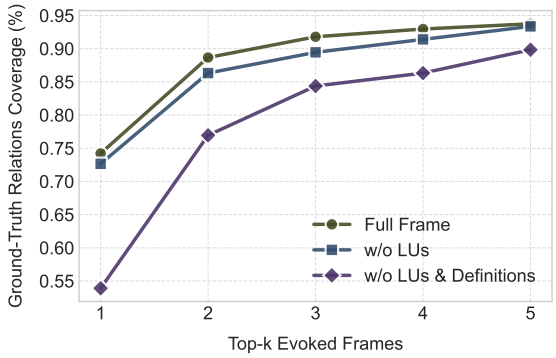

Figure 3: Top-k ground-truth relation coverage under different semantic frame components on the FewRel dataset with $m = 15$.

Method	m=5	m=10	m=15
ZS-BERT	77.90	57.25	36.82
+ RSFs	**82.48** (+4.58)	**72.90** (+15.65)	**59.07** (+22.25)
RE-Matching	92.58	82.93	73.66
+ RSFs	**93.11** (+0.53)	**85.75** (+2.82)	**77.87** (+4.21)
AlignRE	93.09	85.75	77.31
+ RSFs	**94.67** (+1.58)	**87.20** (+1.45)	**80.20** (+2.89)
EMMA	94.67	87.22	80.10
+ RSFs	**95.37** (+0.70)	**90.54** (+3.32)	**83.24** (+3.14)
CE-DA	95.17	88.10	83.31
+ RSFs	**95.60** (+0.43)	**91.66** (+3.56)	**85.20** (+1.89)

Table 4: F_1 scores (%) on FewRel under different m settings (m =5,10,15). **+ RSFs** indicates the incorporation of Relation Semantic Frames. F_1 improvements are highlighted in blue.

ping semantic validation (**w/o Core FE Validation**) leads to a slight decline, reflecting the utility of core elements in filtering inconsistent predictions.

5.3 Effect of Frame Components on Evocation

To investigate the impact of different semantic components within RSFs on candidate frame evocation, we conduct an ablation study by progressively removing key elements . We measure the coverage of ground-truth relations by the top-k evoked frames under each setting.

As shown in Figure 3, the full frame configuration consistently achieves the highest coverage across all values of k, with over 90% of ground-truth relations covered at Top-3. Removing LUs leads to a slight performance drop, indicating that LUs play an important role in activating contextually relevant frames. When both LUs and definitions are removed, the coverage drops significantly, with Top-1 performance falling to nearly 55%, highlighting the crucial role of relation definitions in resolving semantic ambiguity and improving frame selection accuracy.

5.4 ZeroRC with Frame Augmentation

To further assess the transferability of our relation semantic frame (RSF) resource, we adopt the task of Zero-Shot Relation Classification (ZeroRC) as a testbed to investigate whether this resource can enhance the performance of existing methods.

Specifically, we integrate different components of the relation semantic frame into several representative prototype-based ZeroRC methods: **ZS-BERT** (Chen and Li, 2021) formulates ZSRE as a semantic matching task by minimizing the embedding distance between input sentences and relation prototypes, where the prototype is defined by a relation description. We replace the original `Wikidata` descriptions with our frame definitions. **RE-Matching** (Zhao et al., 2023) and **EMMA** (Li et al., 2024c) leverage human-annotated entity descriptions or virtual entity representations derived from descriptions to support classification. We substitute the entity representations with corresponding Core Frame Element definitions. **AlignRE** (Li et al., 2024e) constructs relation prototypes by aggregating relation names, descriptions, and aliases, while **CE-DA** (Zhang et al., 2025a) further incorporates a dynamic aggregation mechanism. For both methods, we augment the prototypes with Lexical Units to enrich their semantic representations.

Table 4 compares baseline methods with and without frame-semantic integration. Results show consistent performance gains across all models, underscoring the superior semantic expressiveness of our RSF over manual descriptions and its effectiveness in improving relation discrimination and generalization in zero-shot settings.

5.5 Case Study

We provide a complete case to analyze the potential difficulties and failure points in the three-stage reasoning pipeline of FrameRTE in Table 5.

First, in the **Frame Evocation** stage, the LLM is required to accurately distinguish between multiple semantically similar frames. The system identifies the relevance of the sentence to professional roles and disciplinary fields based on the lexical trigger "chair" and contextual expressions such as "mathematics" and "astrology." As a result, it effectively evokes the `field of work` frame. However, the

Frame Evocation [Input]	You are an expert in frame semantics ... Candidate Frames: - Frame Name: field of work, Definition: The 'field of work' relation refers to the connection between a subject (often an individual or entity) and a specific area of expertise, discipline, or professional focus. It indicates the subject's involvement in or contribution to that field through research, application, or specialization, typically requiring relevant expertise, skills, or knowledge. Lexical Units: work in, be employed in, career, field. - Frame Name: occupation, Definition: ... Lexical Units: profession, job, position, occupation. - Frame Name: winner, Definition: ... Lexical Units: win, victor, champion, winner. Example: ... Input: <Sentence>: Regiomontanus served as the chair in mathematics while the court astrologer for Corvinus, Marcin Bylica, was the chair in astrology from the university's inception. Hayton, Darin. <Output>:
Frame Evocation [Output]	{"Frames": ["occupation", "field of work"]}
Triplet Extraction [Input]	Please extract all triplets from the sentence based on the candidate frame definitions, and output them in JSON format. Candidate Frames: - Frame Name: field of work, Definition: ... - Frame Name: occupation, Definition: ... Example:... Input: <Sentence>: Regiomontanus served as the chair in mathematics while the court astrologer for Corvinus, Marcin Bylica, was the chair in astrology from the university's inception. Hayton, Darin. <Output>:
Triplet Extraction [Output]	{ "Triplets": [{"Subject": "Regiomontanus", "Object": "chair", "Relation": "occupation"}, {"Subject": "Regiomontanus", "Object": "mathematics", "Relation": "field of work"}, {"Subject": "Marcin Bylica", "Object": "chair", "Relation": "occupation"}, {"Subject": "Marcin Bylica", "Object": "astrology", "Relation": "field of work"}] }
Core FEs Validation [Input]	You are an expert in frame semantics ... Input: <Core Frame Elements>: - Subject: The subject is the individual or entity who holds a specific occupation or professional role, actively performing tasks or duties associated with that role, often as part of their work or employment in a stable or long-term capacity. - Object: The object is the occupation, job, or profession that the subject holds, defining the specific type of work, tasks, or responsibilities they perform, and the field, industry, or domain in which they are engaged or recognized. <Sentence>: ... <Subject>: Marcin Bylica, <Object>: chair, <Relation>: occupation.
Core FEs Validation [Output]	<Invalid> <Justification>: The object "chair" by itself is too generic and does not specify the occupation or professional role clearly. According to the frame definition, the object should denote a specific job or profession (e.g., "astrologer" or "chair in astrology"). In this sentence, "chair" needs to be qualified to align with the expected semantic role, such as "chair in astrology," to represent a valid occupation.

Table 5: A case study of the three-stage reasoning pipeline of FrameRTE.

frame `occupation`, which is not part of the ground truth, is still evoked. This introduces unnecessary burden to downstream extraction, potentially due to inconsistent granularity.

Then, in the **Triplet Extraction** stage, guided by the definitions and core element structures of the evoked frames, the system identifies each person's professional position ("chair") and corresponding field of work ("mathematics," "astrology"). However, the LLM may also over-rely on surface-level lexical cues while neglecting the requirement for semantic completeness. Especially in long or nested sentences, it either overgenerates or omits modifiers, resulting in incorrect entity spans.

Finally, in the **Core FEs Validation** stage, we apply semantic role validation to filter out semantically ambiguous or underspecified triplets. Specifically, the triplets with "chair" as the object are discarded due to lack of specificity, resulting in a final set of semantically coherent triplets. Introducing semantic similarity measures during this stage improves the flexibility and robustness of the validation process.

6 Conclusion

We propose FrameRTE, a zero-shot relation triplet extraction framework that leverages frame semantics to provide structured and interpretable guidance for frozen large language models (LLMs). In contrast to existing approaches that rely on costly fine-tuning, FrameRTE adopts a frame-first-then-extract paradigm: it first constructs high-quality Relation Semantic Frames (RSFs) through a pipeline that combines frame retrieval with high-quality demonstration synthesis. These RSFs are then utilized in a three-stage reasoning process comprising semantic evocation of candidate RSFs, triplet generation under structural constraints, and core frame elements validation. Experiments across multiple benchmarks demonstrate that FrameRTE achieves strong zero-shot performance. Furthermore, RSFs prove to be transferable external semantic resources that can enhance other extraction methods. This work highlights the potential of integrating linguistic knowledge into LLMs and paves the way for applying frame semantics to a broader range of information extraction tasks in the future.

Limitations

Despite the strong zero-shot performance of FrameRTE on multiple benchmark datasets, several limitations remain. First, compared with methods that fine-tune LLMs using annotated seen data, our prompt-based approach still exhibits a noticeable performance gap. This highlights the inherent limitation of relying solely on prompt learning for information extraction, which may be attributed to the built-in biases of LLMs.

Second, although our method incorporates semantic knowledge from relation semantic frames, it has yet to fully exploit the hierarchical semantic relationships between frames in public semantic resources such as FrameNet—e.g., *Inherits from* and *Is Inherited by*. These structural relationships

could provide valuable inductive signals for frame selection and semantic disambiguation.

Third, although our method demonstrates both performance advantages and zero-shot flexibility, it still incurs the overhead of multiple LLM calls. While relation semantic frames can be precomputed and reused, the need to invoke LLMs multiple times for each sentence introduces non-negligible latency and cost.

In future work, we plan to explore methods for integrating frame hierarchies into multi-stage inference processes to further enhance the accuracy and interpretability of ZeroRTE.

Acknowledgments

The authors sincerely thank the reviewers for their valuable comments, which improved the paper. The work is supported by the National Natural Science Foundation of China (62276057).

References

Chaoyi Ai and Kewei Tu. 2024. Frame semantic role labeling using arbitrary-order conditional random fields. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 17638–17646.

Collin F Baker, Charles J Fillmore, and John B Lowe. 1998. The berkeley framenet project. In *COLING 1998 Volume 1: The 17th International Conference on Computational Linguistics*.

Chih-Yao Chen and Cheng-Te Li. 2021. Zs-bert: Towards zero-shot relation extraction with attribute representation learning. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3470–3479.

Yew Ken Chia, Lidong Bing, Soujanya Poria, and Luo Si. 2022. Relationprompt: Leveraging prompts to generate synthetic data for zero-shot relation triplet extraction. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 45–57.

Xinyue Cui and Swabha Swayamdipta. 2024. Annotating FrameNet via structure-conditioned language generation. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 681–692. Association for Computational Linguistics.

Jacob Devasier, Rishabh Mediratta, Phuong Le, David Huang, and Chengkai Li. 2024. Claimlens: Automated, explainable fact-checking on voting claims using frame-semantics. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 311–319.

Charles J Fillmore. 1976. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, 280(1):20–32.

Jiaying Gong and Hoda Eldardiry. 2024. Prompt-based zero-shot relation extraction with semantic knowledge augmentation. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 13143–13156.

Honghao Gui, Lin Yuan, Hongbin Ye, Ningyu Zhang, Mengshu Sun, Lei Liang, and Huajun Chen. 2024. Iepile: Unearthing large-scale schema-based information extraction corpus. *arXiv preprint arXiv:2402.14710*.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Xu Han, Hao Zhu, Pengfei Yu, Ziyun Wang, Yuan Yao, Zhiyuan Liu, and Maosong Sun. 2018. Fewrel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4803–4809.

Li He, Hayilang Zhang, Jie Liu, Kang Sun, and Qing Zhang. 2024. Zero-shot relation triplet extraction via knowledge-driven llm synthetic data generation. In *International Conference on Intelligent Computing*, pages 329–340. Springer.

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*.

Lifu Huang, Heng Ji, Kyunghyun Cho, Ido Dagan, Sebastian Riedel, and Clare Voss. 2018. Zero-shot transfer learning for event extraction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2160–2170.

Bosung Kim, Hayate Iso, Nikita Bhutani, Estevam Hruschka, Ndapandula Nakashole, and Tom Mitchell. 2023. Zero-shot triplet extraction by template infilling. In *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 272–284.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in neural information processing systems*, 33:9459–9474.

Diyou Li, Lijuan Zhang, Juncheng Zhou, Jie Huang, Neal Xiong, Lei Zhang, and Jian Wan. 2024a. An improved two-stage zero-shot relation triplet extraction model with hybrid cross-entropy loss and discriminative reranking. *Expert Systems with Applications*, page 126077.

Diyou Li, Lijuan Zhang, Juncheng Zhou, Jie Huang, Neal Xiong, Lei Zhang, and Jian Wan. 2025a. An improved two-stage zero-shot relation triplet extraction model with hybrid cross-entropy loss and discriminative reranking. *Expert Systems with Applications*, 265:126077.

Guozheng Li, Peng Wang, and Wenjun Ke. 2023. Revisiting large language models as zero-shot relation extractors. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 6877–6892.

Guozheng Li, Peng Wang, Jiajun Liu, Yikai Guo, Ke Ji, Ziyu Shang, and Zijie Xu. 2024b. Meta in-context learning makes large language models better zero and few-shot relation extractors. *arXiv preprint arXiv:2404.17807*.

Shilong Li, Ge Bai, Zhang Zhang, Ying Liu, Chenji Lu, Daichi Guo, Ruifang Liu, and Sun Yong. 2024c. Fusion makes perfection: An efficient multi-grained matching approach for zero-shot relation extraction. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers)*, pages 79–85.

Yinghao Li, Rampi Ramprasad, and Chao Zhang. 2024d. A simple but effective approach to improve structured language model output for information extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 5133–5148.

Zehan Li, Fu Zhang, and Jingwei Cheng. 2024e. Alignre: An encoding and semantic alignment approach for zero-shot relation extraction. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 2957–2966.

Zehan Li, Fu Zhang, Kailun Lyu, Jingwei Cheng, and Tianyue Peng. 2025b. Re-cent: A relation-centric framework for joint zero-shot relation triplet extraction. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 7344–7354.

Shulin Liu, Yubo Chen, Shizhu He, Kang Liu, and Jun Zhao. 2016. Leveraging framenet to improve automatic event detection. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2134–2143.

Bo Lv, Xin Liu, Shaojie Dai, Nayu Liu, Fan Yang, Ping Luo, and Yue Yu. 2023. Dsp: Discriminative soft prompts for zero-shot entity and relation extraction. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 5491–5505.

Ayush Pancholy, Miriam RL Petruck, and Swabha Swayamdipta. 2021. Sister help: Data augmentation for frame-semantic role labeling. In *Proceedings of the Joint 15th Linguistic Annotation Workshop (LAW) and 3rd Designing Meaning Representations (DMR) Workshop*, pages 78–84.

Chaoxu Pang, Yixuan Cao, Qiang Ding, and Ping Luo. 2023. Guideline learning for in-context information extraction. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 15372–15389.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

Josef Ruppenhofer, Michael Ellsworth, Myriam Schwarzer-Petruck, Christopher R Johnson, and Jan Scheffczyk. 2016. Framenet ii: Extended theory and practice. Technical report, International Computer Science Institute.

Oscar Sainz, Iker García-Ferrero, Rodrigo Agerri, Oier Lopez de Lacalle, German Rigau, and Eneko Agirre. 2023. Gollie: Annotation guidelines improve zero-shot information-extraction. *arXiv preprint arXiv:2310.03668*.

Dan Shen and Mirella Lapata. 2007. Using semantic roles to improve question answering. In *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)*, pages 12–21.

Yuchen Shi, Guochao Jiang, Tian Qiu, and Deqing Yang. 2024. Agentre: An agent-based framework for navigating complex information landscapes in relation extraction. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pages 2045–2055.

Evangelia Spiliopoulou, Eduard Hovy, and Teruko Mitamura. 2017. Event detection using frame-semantic parser. In *Proceedings of the Events and Stories in the News Workshop*, pages 15–20.

Qi Sun, Kun Huang, Xiaocui Yang, Rong Tong, Kun Zhang, and Soujanya Poria. 2024. Consistency guided knowledge retrieval and denoising in llms for zero-shot document-level relation triplet extraction. *arXiv preprint arXiv:2401.13598*.

Sijia Wang, Mo Yu, and Lifu Huang. 2023. The art of prompting: Event detection based on type specific prompts. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 1286–1299.

Xiaozhi Wang, Ziqi Wang, Xu Han, Wangyi Jiang, Rong Han, Zhiyuan Liu, Juanzi Li, Peng Li, Yankai Lin, and Jie Zhou. 2020. Maven: A massive general domain event detection dataset. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1652–1671.

Xiang Wei, Xingyu Cui, Ning Cheng, Xiaobin Wang, Xin Zhang, Shen Huang, Pengjun Xie, Jinan Xu, Yufeng Chen, Meishan Zhang, Yong Jiang, and Wenjuan Han. 2024. Chatie: Zero-shot information extraction via chatting with chatgpt. *Preprint*, arXiv:2302.10205.

Derong Xu, Wei Chen, Wenjun Peng, Chao Zhang, Tong Xu, Xiangyu Zhao, Xian Wu, Yefeng Zheng, Yang Wang, and Enhong Chen. 2024a. Large language models for generative information extraction: A survey. *Frontiers of Computer Science*, 18(6):186357.

Ting Xu, Haiqin Yang, Fei Zhao, Zhen Wu, and Xinyu Dai. 2024b. A two-agent game for zero-shot relation triplet extraction. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 7510–7527.

Lilong Xue, Dan Zhang, Yuxiao Dong, and Jie Tang. 2024. Autore: document-level relation extraction with large language models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, pages 211–220.

Liping You and Kaiying Liu. 2005. Building chinese framenet database. In *2005 international conference on natural language processing and knowledge engineering*, pages 301–306. IEEE.

Fu Zhang, He Liu, Zehan Li, and Jingwei Cheng. 2025a. Ce-da: Custom embedding and dynamic aggregation for zero-shot relation extraction. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 9814–9823.

Kai Zhang, Bernal Jiménez Gutiérrez, and Yu Su. 2023. Aligning instruction tasks unlocks large language models as zero-shot relation extractors. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 794–812.

Yifang Zhang, Pengfei Duan, Yiwen Yang, and Shengwu Xiong. 2025b. Enhancing zero-shot relation extraction through staged interaction with large language models. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE.

Hongyan Zhao, Ru Li, Xiaoli Li, and Hongye Tan. 2020. Cfsre: Context-aware based on frame-semantics for distantly supervised relation extraction. *Knowledge-Based Systems*, 210:106480.

Jun Zhao, Wenyu Zhan, Wayne Xin Zhao, Qi Zhang, Tao Gui, Zhongyu Wei, Junzhe Wang, Minlong Peng, and Mingming Sun. 2023. Re-matching: A fine-grained semantic matching method for zero-shot relation extraction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6680–6691.

Xiaoyan Zhao, Yang Deng, Min Yang, Lingzhi Wang, Rui Zhang, Hong Cheng, Wai Lam, Ying Shen, and Ruifeng Xu. 2024. A comprehensive survey on relation extraction: Recent advances and new frontiers. *ACM Computing Surveys*, 56(11):1–39.

Ce Zheng, Yiming Wang, and Baobao Chang. 2023. Query your model with definitions in framenet: an effective method for frame semantic role labeling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 14029–14037.

Yifan Zheng, Wenjun Ke, Qi Liu, Yuting Yang, Ruizhuo Zhao, Dacheng Feng, Jianwei Zhang, and Zhi Fang. 2024. Making llms as fine-grained relation extraction data augmentor. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, pages 6660–6668. International Joint Conferences on Artificial Intelligence Organization.

Sizhe Zhou, Yu Meng, Bowen Jin, and Jiawei Han. 2024. Grasping the essentials: Tailoring large language models for zero-shot relation extraction. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 13462–13486.

A Detailed Prompts

A.1 RSFs Construction

RSFs Construction Prompt

You are a domain expert in frame semantics, specializing in the construction and annotation of semantic frames. Your task is to create a JSON-formatted semantic frame for a given relation label, based on the guidelines below.
Each semantic frame should include the following components:

1. **Frame Name**: This should exactly match the provided relation label and requires no modification.
2. **Definition**: Provide a concise and precise description of the conceptual scenario evoked by the relation.
3. **Frame Elements**: Categorize into:
   - **Core Frame Elements**: Must include at least two required roles — `Subject` and `Object`.
     - The `Subject` typically denotes the initiator or experiencer of the relation.
     - The `Object` typically represents the target, recipient, or affected entity.
     - Each core element must be defined clearly and comprehensively.
   - **Non-Core Frame Elements**: Include optional participants or adjuncts that enrich the frame (e.g., Time, Manner, Location). Provide at least one illustrative example with its definition.
4. **Lexical Units**: List key verbs or nominal expressions that can evoke the frame in natural language. For each lexical unit, provide a short definition that captures its role in expressing the relation.

You may consult manually annotated frames from FrameNet that most closely correspond to the given relation to guide your construction.

Output Format:
```
{ "Frame_Name": "{relation_name}", "Definition": "", "Frame_Elements": { "Core": [{ "Name": "Subject", "Definition": ""}, { "Name": "Object", "Definition": ""} ], "Non-Core": [{ "Name": "", "Definition": ""}] }, "Lexical_Units": [{ "Name": "", "Definition": ""}] }
```

Input:
<Manually Annotated Reference Frame>: `{similar_frame}`
<Relation Label>: `{relation_name}`

A.2 RSFs Enhancement

Definition Aggregation Prompt

You are an expert in frame semantics. Please consolidate the following definitions of the same `{type}` into a single, concise, accurate, and comprehensive definition. Note: If any part of the definitions is unclear or ambiguous, clarify it through reasonable inference or adjusted wording. While avoiding the inclusion of irrelevant information, retain as much critical information from all definitions as possible. Ensure the final definition is logically clear, succinctly expressed, and free of redundancy or unnecessary complexity.

Input: <Definitions>: `{definitions}`

Output Format:
```
{ "{type} Definition": ""}
```

A.3 Semantic Frame Evocation

Frame Evocation Prompt

You are an expert in frame semantics. Your task is to identify the most appropriate semantic frames (i.e., relations) that best fit the given sentence.
You are provided with a list of candidate frames, each accompanied by a definition and relevant lexical units (i.e., triggering words or expressions).
Please output your selected frame names in **descending order of relevance**, using the following JSON format:

Output Format:
{ "Frames": ["<FrameName1>",...]}

Candidate Frames:
- Frame Name: {frame name}
Definition: {definition}
Lexical Units: {lexical units}
...

Example:
<Sentence>: {sentence}
<Output>: { "Frames": ["FrameName"]}

Input:
<Sentence>: {sentence}
<Output>:

A.4 Frame-Guided Extraction

Triplet Extraction Prompt

Please extract all triplets from the sentence based on the candidate frame definitions, and output them in JSON format.

Candidate Frames:
- Frame Name: {frame name}
Definition: {definition}
...

Example:
<Sentence>: {sentence}
<Output>: "Triplets":["Subject": "subject", "Object": "object", "Relation": "relation"]

Input:
<Sentence>: {sentence}
<Output>:

A.5 Core Frame Elements Validation

Core Frame Elements Validation Prompt

You are an expert in frame semantics. Your task is to verify whether a predicted triplet aligns with the definitions of Core Frame Elements. If both match the expected semantic roles, output <Valid>. Otherwise, output <Invalid> and briefly explain why.

Output Format:
<Valid/Invalid>
<Justification>: ...

Input:

<Core Frame Elements>:
- Subject: {definition}
- Object: {definition}

<Sentence>: {sentence}
<Subject>: {subject}
<Object>: {object}
<Relation>: {relation}

<Output>: