Can dependency parses facilitate generalization in language models? A case study of cross-lingual relation extraction

Ritam Dutt, Shounak Sural, Carolyn Penstein Rosé
Carnegie Mellon University
{rdutt,ssural,cprose}@andrew.cmu.edu

Abstract

In this work, we propose DEPGEN, a framework for evaluating the generalization capabilities of language models on the task of relation extraction, with dependency parses as scaffolds. We use a GNN-based framework that takes dependency parses as input and learns embeddings of entities which are augmented to a baseline multilingual encoder. We also investigate the role of dependency parses when they are included as part of the prompt to LLMs in a zero-shot learning setup. We observe that including off-the-shelf dependency parses can aid relation extraction, with the best performing model having a mild relative improvement of 0.91% and 1.5% in the in-domain and zero-shot setting respectively across two datasets. For the in-context learning setup, we observe an average improvement of 1.67%, with significant gains for low-performing LLMs. We also carry out extensive statistical analysis to investigate how different factors such as the choice of the dependency parser or the nature of the prompt impact performance. We make our code and results publicly available for the research community at https://github.com/ShoRit/multilingual-re.git

1 Introduction

Information packaging in language does not happen arbitrarily (Croft, 2022). The "internal structure" of a text message, which determines how the message is constructed or parsed, is grounded in predefined linguistic rules in the form of syntax and semantics. Linguistic structures such as dependency graphs (Zeman et al., 2019; Chomsky, 2002) or semantic parses (Banarescu et al., 2013; Reddy et al., 2017) have been pivotal in the history of NLP research both for their intrinsic merit i.e. developing frameworks that can construct or interpret such structures automatically (Chen et al., 2024; Gu et al., 2024), and their external value as augmentations to aid language understanding tasks (Ding et al., 2024; Şahin, 2022).

Information extraction or IE is one such field which had relied heavily on linguistic information ever since its inception; some notable examples include few-shot named entity recognition or NER (Chen et al., 2023; Xie et al., 2024), relation extraction (Li et al., 2023; Zhou et al., 2024), open-domain question answering, (Zhang et al., 2023b, 2024) amongst others. However, recent years have witnessed a decline in the adoption of linguistic frameworks in favor of large scale pre-trained language models (Devlin et al., 2018; Liu et al., 2019; Conneau et al., 2020a; Sainz et al., 2024) which are shown to encode syntactic and semantic information within their parameters (Starace et al., 2023; Liu et al., 2024) and have also demonstrated significant improvements on IE (Sainz et al., 2024; Efeoglu and Paschke, 2024).

Moreover, as we usher into an era of large language models, the question which looms over our head like the proverbial sword of Damocles "Are dependency parses helpful for information extraction?" We are motivated to answer this question based on the past work of Sachan et al. (2021) which showed the utility of adding syntactic information for different information extraction tasks in English. However, the observed benefits hold true only when the **gold parses** are available, with no improvements over the baseline in presence of off-the-shelf parses. In this study we expand upon this idea and investigate whether off-the-shelf dependency parses can assist language models in multilingual information extraction for both indomain and zero-shot transfer settings.

We specifically deal with the task of multilingual relation extraction, wherein we identify the nature of relationship between two annotated entities in a document. We show in Figure 1 how we can connect the entities wood and fences by traversing the dependency graph that connects these two entities, highlighting the potential utility of linguistic frameworks for this task. We explore the role

Proceedings of the 4th International Workshop on Knowledge-Augmented Methods for Natural Language Processing (KnowledgeNLP'25), pages 338–358
May 3, 2025 ©2025 Association for Computational Linguistics

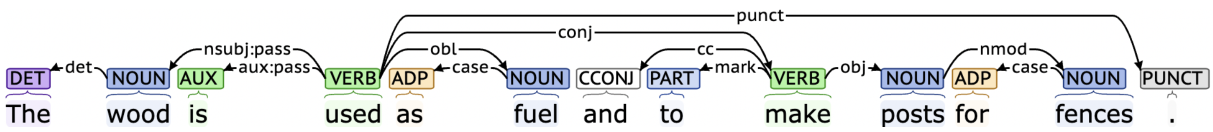

Figure 1: Example depicting the supplemental information provided by the *dependency tree*. The entities of interest are wood and fences, having the relationship material_used. The path *wood* ← *used* → *make* → *posts* → *fences* elicits this relationship.

of dependency parses for cross-lingual relation extraction in both a fine-tuned supervised setting and a prompting/ in-context learning setup.

We propose a framework, DEPGEN, built on top of a pretrained multi-lingual language model that uses dependency parse information to perform relation extraction for both in-domain and zero-shot cross-lingual transfer settings. Through a comprehensive set of 2440 experiments spanning 10 languages over 2 datasets, we observe that incorporating dependency information brings about modest improvements for in-domain and cross-lingual fine-tuning setups by 0.9% and 1.5% respectively.

We also carry out extensive statistical analysis to identify which factors significantly impact performance. Our observations highlight that performance improvements is mostly predicated by the choice of the target language, and the choice of the pre-trained language model rather than the choice of the dependency parser for all cases. However, for the in-context learning setup, we demonstrate that the performance is determined by the choice of the prompting strategy, with our proposed approach boasting the highest gains, i.e. an absolute improvement of 1.67 F1 score over the baseline.

2 Related Work

2.1 Generalization in Information Extraction

Recent years bear witness to countless research endeavors to facilitate generalizability and transfer across domains for several information extraction (IE) tasks. Such works include zero-shot relation extraction (Wang et al., 2022b; Jun et al., 2022; Li et al., 2023), zero-shot or few-shot NER (Zeng et al., 2022; Das et al., 2022; Xie et al., 2024) , zero-shot KBQA (Gu et al., 2021; Dutt et al., 2023), cross-lingual KBQA (Zhang et al., 2023a), and open domain QA (Min et al., 2020; Zhang et al., 2023b), amongst others (Fritzler et al., 2019; Zhou et al., 2019). This interest is in part due to the advent of large scale pre-trained language models such as Devlin et al. (2018); Liu et al. (2019); Conneau et al. (2020a); Sainz et al. (2024) which have shown significant improvements on IE. Recent works on domain adaptation and transfer learning have advocated different pre-training objective functions to ensure the model is well adapted to the particular domain. Other multi-lingual/ cross-lingual transfer works employ different data augmentation techniques such as translation into the target data to aid transfer. In this work, we investigate approaches to perform multi-lingual information extraction in a zero-shot setting without any additional data in the target language.

2.2 Relation Extraction

The goal of relation extraction or relation classification is to detect and classify the relation between specified entities in a text according to some predefined schema. Current research in RE has mostly been carried out in a few-shot or a zero-shot setting to address the dearth of training data (Liu et al., 2022; Li et al., 2023) and the "long-tail" problem of skewness in relation classes (Ye and Ling, 2019b; Liang et al., 2023). Salient work in that direction includes (i) designing RE-specific pretraining objectives for learning better representations (Baldini Soares et al., 2019; Wang et al., 2022a), (ii) incorporating meta-information such as relation descriptions (Yang et al., 2020; Chen and Li, 2021), a global relation graph, (Qu et al., 2020), or entity types (Peng et al., 2020), and (iii) leveraging additional information in the form of dependency parses (Yu et al., 2022), translated texts for multilingual RE (Nag et al., 2021), or distantly supervised instances (Zhao et al., 2021; Ye and Ling, 2019a). T-5 based models have shown to perform well in relation extraction settings with few-shot finetuning (Diaz-Garcia and Lopez, 2024).

Recently, LLMs have shown promise in zero-shot relation extraction. Challenging cases such as overlapping relations and none-of-the-above (nota) relations have been handled effectively by LLMs in zero-shot settings (Li et al., 2023). LLMs have also outperformed smaller models for RE with larger, document-level context sizes in models such as

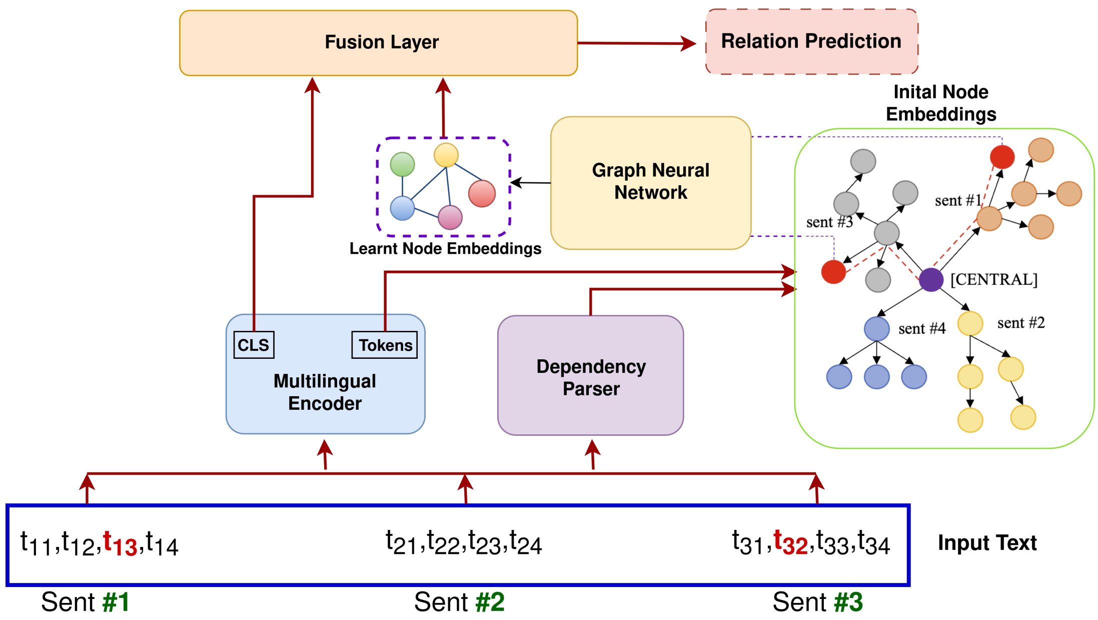

Figure 2: An overview of our proposed framework DEPGEN. The architecture takes as input a document, which comprises a sequence of sentences, with the entities highlighted in red. This document passes through a multilingual encoder to obtain the token embeddings, and a dependency parser that generates dependency parses for each sentence. The individual sentences in the dependency parser are connected using a central [CENTRAL] node to obtain a connected graph. The nodes are initialized using the embeddings obtained from the multilingual encoder and updated using a Graph Neural Network. The final representations of the entities obtained from the GNN are fused with the entity embeddings and concatenated with the [CLS] token of the document to predict the relation.

AutoRE (Xue et al., 2024). All of these techniques seek to alleviate the need for expensive human-annotated training data. In this work, we question whether incorporating linguistic structures in the form of dependency parsing as an explicit addition to the input in existing models can aid learning robust representations that can be transferred to other languages.

3 Methodology

We investigate the role of dependency parses for zero-shot cross-lingual relation extraction in two setups, namely (i) a fine-tuned setup where a model is first trained on a given source language and then evaluated on a target language, and (ii) an in-context-learning setup where we prompt an LLM to predict the relation between two specified entities in a zero-shot setting to test the innate capabilities of the LLM for RE.

3.1 Fine-Tuning Setting

We present a detailed description of our proposed framework, DEPGEN here. Our framework leverages the internal structure of a document text to aid relation classification. We define internal structure as the linguistic information encoded within the document based on syntactic rules in the form of dependency parses. This section describes the individual components that constitute our framework DEPGEN, namely the multilingual encoder, dependency parser, graph neural network, and the fusion layer. We dive deep into the methodology for representing the textual content, and elaborate on the approach employed for incorporating dependency parses for a given input sentence. Finally, we end the section with how the different modes of information are fused, and the classification setup. A pictorial representation of our framework can be seen in Figure 2 Our architecture involves the following components.

3.1.1 Multilingual Encoder

We experiment with mBERT (Devlin et al., 2018) and XLMR (Conneau et al., 2020c) as our multilingual text encoder to obtain representations of the input sentence(s). Past work has shown the efficacy of such contextual multilingual encoders in capturing long-range semantic dependency in text (Litschko et al., 2021). Similar to these works, we consider the final encoder layer representation of [CLS] token as the text representation. The sentence(s) are fed as input to the MULTILINGUAL

ENCODER (Figure 2) and the [CLS] token representation from the final layer is fed into the FUSION LAYER. The individual token representations from the final layer are used to initialize the node embeddings in the dependency graph of the INTERNAL STRUCTURE module, which we describe below.

3.1.2 Internal Structure

We incorporate the internal structure information by learning the syntactic dependency information between the tokens in the input sentence. We first pass the input tokens through a DEPENDENCY PARSER to obtain the dependency tree for each sentence. We then construct a dependency graph from the constituent dependency trees, which is then fed as input to a Graph Neural Network (GNN) (Scarselli et al., 2008). The various components of this module are as follows.

Dependency Parser To generate the dependency tree, we use off-the-shelf multilingual dependency parsing modules, i.e. Stanza (Qi et al., 2020a) and Trankit (Nguyen et al., 2021). The resulting dependency tree represents the syntactic dependency relations between the words in a sentence; the dependencies follow the Universal Dependencies formalism (Nivre et al., 2016; Zeman et al., 2019), resulting in 76 types of dependencies across the different languages for our experiments.

Dependency Graph Since the dependency tree is defined for a sentence, the output from DEPENDENCY PARSER will be in the form of a forest of disconnected dependency trees; for example 4 trees for 4 sentences in Figure 2. We add a pseudo node [CENTRAL] and add a new type of dependency relation [SENT] between the [CENTRAL] and all the [ROOT] nodes of the sentences. The proposed design has two benefits - (1) The [CENTRAL] node allows for information exchange between the sentences, which otherwise would probably lead to different clusters of representations (represented by colors in Figure 2) for nodes in different sentences, (2) The distance between the two entities is reduced (dotted red line in Figure 2) when the entities are present across two different sentences, resulting in an efficient information flow between them.

Graph Neural Network We represent each word as a node in the dependency graph and the dependency relations as the edges between the nodes. Each node in the graph is initialized with the representations obtained from the final layer of the MULTILINGUAL ENCODER. We aggregate the subtoken representations via max-pooling and obtain the final representation of a word. This initialization helps incorporate the semantic relationship between the nodes and facilitates end-to-end joint training of the MULTILINGUAL ENCODER and the INTERNAL STRUCTURE modules. The relation embeddings for the all the relation types are initialized at random and learnt jointly along with the node embeddings. The representations of the two entities from the multi-layer GNN are then fed to the FUSION LAYER along with the sentence representation for relation prediction.

3.1.3 Relation Prediction

We concatenate the representations obtained from the MULTILINGUAL ENCODER and the INTERNAL STRUCTURE modules in the FUSION LAYER and perform a multi-class classification for predicting the relation. During training, we compute the standard Cross Entropy loss, and back-propagate it jointly through all the components of the network.

3.2 In-context Learning Setting

In addition to the DEPGEN framework that encapsulates the fine-tuned setting, we also explore the role of dependency parses when provided as additional inputs to LLMs in a zero-shot prompting setup. We experiment with three different types of prompt formats that encodes the dependency information which we describe below.

Tuple Format: In the tuple-based prompt format, we simply provide the dependency parse as a list of tuples or dictionary keys. Each tuple comprises three elements, i.e. a node in the dependency graph or a word, the corresponding head node of that word, and the relation that connects the head node to the word. For example, the phrase “Porsche Panamera”, would have the following information in the form of a tuple.

```
{
    word: Porsche
    head: Panamera,
    rel: compound
}
```

Text Format: Instead of providing the dependency parse information in the form of tuples, we verbalize the dependency relations between the words in the sentence in natural language format. In the above example of

"Porsche Panamera", we re-write the tuple information as `"Porsche is Compound noun modifier of Panamera"`. We do this for all the tuples in the dependency graph.

Filtered Text Format: As opposed to verbalizing all the tuples in the dependency graph, we filter out only the tuples that connect the two entities in the sentence via the dependency relations. Not only does this reduce the number of input tokens to the LLM, it also helps filter out redundant information.

As a control, we also prompt the models with only the text, without any dependency information, which serves as a baseline. The details of the prompts are in the Appendix.

4 Experimental Setup

4.1 Dataset

We conduct our experiments on relation extraction on two datasets i.e. IndoRE and REDFM.

IndoRE (Nag et al., 2021) The IndoRE dataset covers a diverse and rich set of entity and relation annotated sentences in three low resource Indian languages — Bengali (bn), Hindi (hi) and Telugu (te). To study protocols for transferring RE capability across languages, it also has labeled English (en) RE instances as an example of a resource-high language. The dataset consists of 32,610 sentences combining all four languages from Wikidata where each language contains 51 unique relations. Out of these languages, we exclude Bengali from our experiments because the dependency parsers' inability to parse the language.

REDFM (Huguet Cabot et al., 2023) We use this dataset consisting of examples from 7 languages. These languages include English (en), Arabic (ar), Spanish (es), German (de), Italian (it), French (fr), and Chinese (zh), which are hand-annotated. There are a total of about 15,400 examples in the dataset with a total of 32 types of relations. We use the languages en, es, de, it, and fr for training (i.e. source languages), and all the 7 languages for testing in a zero-shot setting (i.e. target languages). We exclude Arabic and Chinese as source language due to the unavailability of a training split in the REDFM dataset. We use the train/validation/test splits as in the original paper.

4.2 Fine-tuned Experimental Setup

We experiment with the following settings:

1. **Baseline:** We experiment with mBERT (Devlin et al., 2019) and XLMR (Conneau et al., 2020b) as our choices to encode the document text and the entity spans. We concatenate the pooled representation of the entities and the [CLS] embedding and use it for relation classification.

2. **Dependency Parses:** We experiment with Trankit (Nguyen et al., 2021) and Stanza (Qi et al., 2020b) as the choice of the dependency parser across all languages for both datasets.

3. **Graph Neural Network:** We experiment with RGCN (Schlichtkrull et al., 2018) and RGAT (Busbridge et al., 2019) as the backbone GNN architecture to encode the dependency information between words in the document. We use a GNN with 2 hidden layers for all our experiments.

4.3 In-context Learning Experimental Setup

We employ three different instruction-tuned LLMs for our in-context learning experiments, i.e. LLaMA (Meta-Llama-3-8B-Instruct) (Grattafiori et al., 2024), Mistral (Mistral-7B-Instruct-v0.3) (Jiang et al., 2023) and Qwen (Qwen2-7B-Instruct) (Yang et al., 2024). We use instruction-tuned LLMs since we wish to employ these LLMs in a zero-shot setup for relation extraction without fine-tuning or additional training. Similar to the fine-tuned experimental setup, the dependency parse information are obtained from two sources, i.e. Stanza and Trankit.

4.4 Experiment Counts

For in-domain, we have a total of 8 languages (5 for RedFM, 3 for IndoRE) for 2 given choices of encoder, parser and GNN. Each experiment is repeated for 5 seeds resulting in a total of 320 experiments, that include dependency information and an additional 80 experiments (over 8 languages, 2 encoders, and 5 seeds) as the baseline. For the zero-shot cross-lingual transfer setting, for RedFM, we have 6 possible target languages for each of the 5 source languages, while for IndoRE, we have 2 possible targets for each of the 3 source languages. Thus for each 36 possible cross-lingual pairs, we evaluate 50 different combination of encoder, parser, GNN, and seeds, resulting in another 1800 experiments. Finally, in the in-context learning setup for LLMs, we experiment with 3 LLMs for 10 languages over 4 kinds of prompts (including

		RedFM					IndoRE		
					mBERT				
DEP	GNN	en	es	fr	it	de	en	hi	te
-	-	84.3±0.7	80.0±0.6	78.6±0.3	76.3±0.8	78.7±0.3	94.3±0.6	89.6±0.4	84.9±0.4
stanza	rgcn	**85.7±0.8**	80.5±1.0	79.7±1.0	78.2±0.5	80.0±0.9	94.4±0.2	**90.9±0.3**	**86.1±0.9**
stanza	rgat	85.2±1.4	**82.2±0.6**	79.9±0.4	77.9±1.2	**80.5±0.6**	**94.9±0.3**	89.5±1.4	85.9±1.1
trankit	rgcn	84.3±0.4	81.8±0.8	**80.7±0.8**	**78.9±0.7**	79.7±0.9	94.0±0.2	89.7±0.1	85.9±1.9
trankit	rgat	85.5±1.3	80.9±0.3	80.2±0.2	77.3±0.8	78.9±0.7	94.1±0.5	88.9±0.5	84.6±0.8
					XLMR				
-	-	**84.0±1.1**	77.2±2.0	76.2±1.0	74.8±1.2	75.2±0.6	92.1±0.8	88.7±0.9	**86.3±1.1**
stanza	rgcn	83.7±0.6	76.8±0.8	**76.7±0.9**	73.3±0.7	75.7±1.5	91.8±0.8	89.6±1.1	85.6±0.7
stanza	rgat	**84.0±0.8**	**77.5±1.4**	74.4±0.9	**75.6±1.2**	**76.2±1.1**	92.2±0.4	**89.9±0.9**	85.7±0.6
trankit	rgcn	83.8±0.5	76.4±1.1	74.7±1.0	72.6±2.3	73.9±2.6	91.9±0.9	**89.9±0.8**	85.2±0.5
trankit	rgat	82.6±0.8	77.3±0.2	75.0±0.3	74.0±1.7	75.9±0.1	**92.6±0.7**	89.2±1.0	85.9±1.6

Table 1: In-domain RE performance of mBERT and XLMR on RedFM and IndoRE, with dependency information (i.e. choice of the parser or DEP, and the choice of the GNN used to encode the information, i.e. GNN). Results are averaged across the top 3 seeds, with the highest values in each column bolded.

		RedFM							IndoRE		
						mBERT					
DEP	GNN	en	es	fr	it	de	ar	zh	en	hi	te
-	-	77.5±1.1	81.0±1.1	78.8±1.1	76.7±1.1	75.6±1.1	72.5±1.1	70.0±1.1	57.5±1.8	57.6±2.7	42.4±2.4
stanza	rgcn	78.2±0.8	81.0±0.8	**79.5±0.8**	**76.8±0.8**	77.1±0.8	72.6±0.8	70.0±0.8	57.0±1.0	57.1±0.8	44.6±1.2
stanza	rgat	78.0±1.0	81.1±1.0	78.8±1.0	76.5±1.0	77.2±1.0	**73.2±1.0**	**70.4±1.0**	56.4±1.2	57.7±1.2	**45.2±1.4**
trankit	rgcn	**78.7±0.8**	**81.3±0.8**	79.3±0.8	75.4±0.8	77.8±0.8	72.8±0.8	70.0±0.8	**57.9±0.8**	**59.1±0.6**	44.9±1.6
trankit	rgat	77.9±0.8	80.6±0.8	79.1±0.8	76.3±0.8	**77.9±0.8**	73.1±0.8	**70.4±0.8**	57.1±1.4	57.9±1.8	45.1±1.7
						XLMR					
-	-	72.7±1.4	74.2±1.4	72.2±1.4	66.8±1.4	70.7±1.4	61.8±1.4	63.1±1.4	50.0±2.2	55.1±1.5	45.9±1.6
stanza	rgcn	**73.4±1.4**	74.5±1.4	73.2±1.4	67.7±1.4	70.3±1.4	61.2±1.4	**63.9±1.4**	49.3±1.8	55.4±1.4	46.1±1.7
stanza	rgat	73.3±1.5	74.3±1.5	**73.4±1.5**	**67.9±1.5**	68.4±1.5	61.1±1.5	63.2±1.5	50.0±1.6	53.8±2.8	46.3±2.0
trankit	rgcn	73.1±1.3	74.7±1.3	73.1±1.3	66.8±1.3	69.5±1.3	**62.7±1.3**	63.8±1.3	50.7±0.7	**56.3±1.1**	45.5±2.9
trankit	rgat	73.1±1.1	**75.7±1.1**	**73.4±1.1**	65.9±1.1	**70.9±1.1**	62.1±1.1	63.6±1.1	**50.8±1.4**	56.0±2.2	**46.9±2.6**

Table 2: Zero-shot Cross-lingual RE performance on RedFM and IndoRE with mBERT and XLMR as the multilingual encoders with different combinations of dependency information. For a given target language, we average the performance across the different source languages. The highest values in each column are highlighted in bold. Detailed individual cross-lingual performance metrics are given in the Appendix.

the baseline), and 2 kinds of parsers (Stanza and Dependency), resulting in a suite of 240 prompting experiments. Our final experimental suite thus comprises 2440 experiments.

5 Results and Insights

In this section, we pose the following research questions (RQs) and attempt to answer the same.

RQ1. Impact of dependency parses on RE for indomain and cross-lingual transfer ?

We report the in-domain and cross-lingual relation extraction performance with mBERT and XLMR as the multilingual encoders, stanza and trankit as the choice of the off-shelf-parsers, and RGCN and RGAT being the backbone GNN for both the IndoRE and RedFM datasets, in Tables 1 and 2 respectively.

At the outset, we observe that across both datasets, adding dependency information generally improves performance over the baseline in the in-domain setting; we see higher gains when we have mBERT as the MLM as opposed to XLMR. We also observe that the gains are higher for the REDFM dataset than IndoRE, possibly due to the poorer quality of dependency parses on low-resource languages like Hindi and Telugu, as opposed to standard high-resource cases like English, Spanish, and Italian. In fact, for all languages other than English, we see a consistent improvement in F1-score of approximately 2.0% and 1.0% with

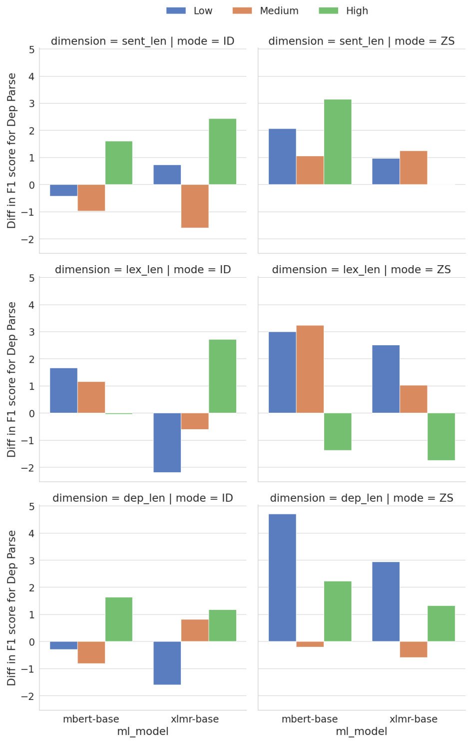

Figure 3: Performance of DEPGEN for in-domain and zero-shot cross-lingual transfer settings on the IndoRE dataset analyzed across variations in sentence, lexical and dependency length

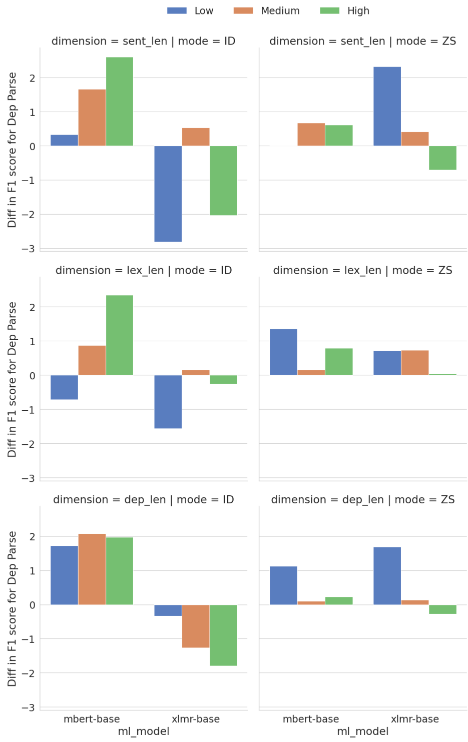

Figure 4: Performance of DEPGEN for in-domain and zero-shot cross-lingual transfer settings on the RedFM dataset analyzed across variations in sentence, lexical and dependency length

the mBERT model on the REDFM and IndoRE dataset respectively, for the best combination of dependency parser and GNN.

In the zero-shot cross-lingual transfer scenario from Table 2 we observe trends that are markedly different from the in-domain setting. Each entry in this Table is computed by averaging the macro-F1 score over the other source languages, apart from itself, for the top 3 seeds. We notice only slight improvements in RE performance for mBERT but higher gains for XLMR. We hypothesize that since XLMR has a worse performance than mBERT, it benefits more from the dependency information in the zero-shot setting. In a similar vein, we observe much higher gains for Hindi and Telugu (around 2.6% and 6.6% relative performance improvements respectively) in the zero-shot setting for mBERT. The markedly lower scores in IndoRE in the zero-shot transfer setup as compared to REDFM can be attributed to the higher number of relations in the dataset (32 for IndoRE vs 51 for REDFM).

RQ2. Which scenarios benefit the most with additional information in the fine-tuned setup?

In the fine-tuned setup, we analyze which scenarios or inputs benefit the most from including dependency information. We thus group the test instances according to three different dimensions, i.e. (1) input sentence length (2) lexical distance between two entities in the sentence and (3) dependency path length. Figures 3 and 4 show the effect of these components for the in-domain and zero-shot cross-lingual transfer settings for the IndoRE and RedFM datasets respectively. The blue, orange and green plots reflect the bottom quartile, inter-quartile range and the top quartile respectively for each of these three dimensions.

Sentence Length: We quantify sentence length based on the total number of tokens in the document. For both zero-shot and in-domain settings across the two datasets, adding linguistic information in the form of dependency graphs improves relation extraction for longer sentences. We posit that including dependency information helps to capture long range dependencies across words and thus the observed gains for longer sentences.

Lexical Distance: We quantify the lexical distance as the number of tokens between the two entities. Here, we observe that dependency information is more helpful for cases where the dis-

Model	Parser	Prompting	RedFM							IndoRE			Average
			ar	de	en	es	fr	it	zh	en	hi	te	
Llama	None	-	25.6	25.7	27.0	27.0	16.7	**36.7**	**37.1**	47.6	39.0	21.9	30.4
	Stanza	Tuple	24.3	19.3	23.9	17.0	18.4	19.3	29.6	30.2	28.3	10.9	22.1
	Stanza	Text	25.1	24.5	22.6	23.1	23.5	24.0	30.6	44.4	37.4	22.9	27.8
	Stanza	Filtered Text	**33.5**	35.0	32.3	**31.6**	**30.5**	34.0	36.1	**48.0**	**44.5**	**29.8**	**35.5** (↑**5.1%**)
	Trankit	Tuple	30.3	17.1	**37.3**	17.2	18.3	22.1	32.4	27.5	30.6	10.5	24.3
	Trankit	Text	23.4	25.4	22.7	22.6	23.8	25.6	30.5	44.8	38.4	24.0	28.1
	Trankit	Filtered Text	33.1	**35.2**	35.6	31.4	28.7	30.3	35.3	46.2	42.8	29.5	34.8
Mistral	None	-	36.7	38.2	39.0	35.8	36.0	38.3	35.6	**51.3**	**38.5**	10.6	36.0
	Stanza	Tuple	27.2	35.9	30.9	31.9	28.1	35.1	30.9	48.4	30.6	9.8	30.9
	Stanza	Text	29.2	32.0	34.4	32.6	30.4	33.4	33.2	47.5	37.1	8.7	31.9
	Stanza	Filtered Text	39.1	39.5	**40.9**	**37.1**	36.6	**40.2**	36.7	50.8	**38.5**	10.3	**37.0** (↑**1.0%**)
	Trankit	Tuple	27.4	35.3	32.5	31.5	26.9	30.6	31.3	48.0	30.5	10.8	30.5
	Trankit	Text	27.9	32.0	34.7	30.7	31.0	32.7	34.1	46.8	36.4	**11.2**	31.7
	Trankit	Filtered Text	**39.3**	**39.7**	39.3	36.3	**36.9**	37.8	**38.1**	50.9	38.3	**11.2**	36.8
Qwen	None	-	**44.3**	**39.6**	**40.3**	38.0	**36.8**	**43.0**	**40.8**	42.7	39.2	29.1	**39.4**
	Stanza	Tuple	35.4	32.0	34.6	31.8	31.9	37.8	31.4	38.3	38.2	26.1	33.8
	Stanza	Text	33.8	34.8	36.0	33.3	33.3	33.3	29.9	39.5	41.1	**30.6**	34.6
	Stanza	Filtered Text	42.1	32.8	39.8	37.3	33.6	38.4	40.4	**44.7**	45.4	28.6	38.3(↓1.1%)
	Trankit	Tuple	34.3	30.7	35.2	34.1	28.1	35.1	33.6	39.6	37.4	21.8	33.0
	Trankit	Text	35.4	35.2	34.2	33.1	34.0	33.4	30.2	40.5	40.5	27.3	34.4
	Trankit	Filtered Text	39.9	36.0	35.4	**39.4**	34.7	38.6	34.5	44.0	**45.9**	26.6	37.5

Table 3: Effect of dependency parses and prompting techniques for LLM-based relation extraction for the REDFM and IndoRE datasets. Performance reported in terms of F1-Score. Best performing methods are shown in bold.

tance between the entities is not high, i.e. Low and Medium categories.

Dependency Path Length: We quantify the dependency path length as the number of dependency relations that separate the two entities in the dependency graph. We see prominent gains for both short and long range dependency paths, especially for the ZS case for IndoRE. However, similar to lexical distance, the gains are more prominent when the dependency path between the entities is small. Since our chosen GNN has only two layers, we hypothesize that it is unable to capture signals across long dependency paths effectively.

RQ3. Can dependency parses help improve relation extraction performance for LLMs?

Table 3 summarizes the performance of three LLMs - LLaMA (Grattafiori et al., 2024), Mistral (Jiang et al., 2023) and Qwen (Yang et al., 2024) for zero-shot relation extraction on the IndoRE and RedFM datasets. To account for the skew in distribution of relations, we employ the macro-F1 score as the primary evaluation metric. We observe that for the LLama-3 and Mistral models, incorporating dependency parses improves performance across several cases. The gains are most prominent when the dependency information is presented in the form of natural language text; we see consistent improvements for the Text Prompt Format over the Tuple Prompt Format, where the information is presented as a list of tuples. We see that the filtered prompt that removes information not pertaining to the two entities, improves performance further.

The improvement can be as significant as 1% to 5% in some cases in terms of absolute F1-score for Mistral and LLama-3 respectively. For the Qwen model, dependency parses do not afford much benefits. Thus the choice of the LLM and the description of the prompt, play a significant role in zero-shot relation extraction performance. It should be noted, however, that the zero-shot performance for the in-context learning setup is significantly worse than the zero-shot cross-lingual performance in the fine-tuned setup. With LLMs, we see an average absolute improvement of 1.67% across all models and languages with the Filtered Text Prompt.

RQ4. Which factors influence generalization?

We now inspect the factors that characterize performance improvements over the baseline for the two datasets in the fine-tuned learning and in-context learning setup. We perform a multivariate ANOVA analysis with the relative performance difference (expressed as a percentage over the baseline), from including the dependency parses, as the dependent variable.

The independent variables chosen are the choice of the multilingual encoder, (mBERT or XLMR), dependency parser (Stanza or Trankit), GNN employed (RGCN or RGAT), and the source and target language [1]. We also consider the pair-wise interaction effects of each of these variables, and note the F-statistic and their corresponding p-value for the indomain (Tables 9 and 11) and zero-shot cross-lingual (Tables 10 and 12) respectively.

For the indomain setting in IndoRE, we observe that the relative performance change hinges most on the choice of the dependency parser followed by source language. Although the choice of the encoder and the GNN do not have any significant effect on relative performance, their pair-wise interactions is indeed significant. The story is remarkably different for REDFM where only the choice of the encoder has any significant effect on RE.

In the zero-shot cross-lingual setting for IndoRE, we see significant effects arising from the choice of the target language and the pairwise interaction between the choice of the source/target language with that of the encoder. A similar story also holds for REDFM, wherein we notice the only significant interactions are between the choice of the source/-target language and the encoder, and also between the choice of the source/target language pairs themselves. Simply put in the zero-shot setting the role of the dependency information is insignificant for both datasets.

We carry out a similar statistical analysis for the zero-shot ICL setup, with the relative performance change over the baseline as the dependent variable, and the choice of the LLM (i.e. LLama-3, Qwen, and Mistral), the prompt (i.e. Tuple Format, Text Format, and Filtered Text Format), the language (7 for RedFM and 3 for IndoRE), and the choice of the dependency parser (i.e. Trankit and Stanza) as the independent variables. We also consider the pair-wise interaction effects of each of these variables, and note the F-statistic and their corresponding p-value for the IndoRE and REDFM dataset respectively in Tables 13 and 14 respectively. We observe, over both datasets, significant effects arising from the choice of the LLM, and the choice of the prompt, as well as the pairwise interaction between the choice of the prompt and LLM, and the choice of the source language and LLM. Once again, we see that the choice of the dependency parser, i.e. the Stanza or Trankit, does not play a significant role.

[1]For the indomain setting we consider only the target language

6 Conclusion and Future Work

In this paper we propose a multi-component framework for multi-lingual relation extraction. Our finetuned framework DepGen, combines the signals from the input sentence with dependency parses that are encoded through a GNN. Through extensive evaluations, we have reported the implications of our work for both in-domain and zero-shot transfer settings across multiple languages. We observe that including off-the-shelf dependency parses can aid relation extraction, with the best performing model having a mild relative improvement of 0.91% and 1.5% in the in-domain and zero-shot setting respectively across two datasets. We also develop an in-context learning prompting approach that incorporates dependency information to bring about an average improvement of 1.67%, with significant gains for low-performing LLMs.

In this work, we investigate how augmenting dependency parses in language models can facilitate information extraction tasks in low-resource settings. Specifically, our contributions being independent of the language setting makes our model portable to other languages in a zero-shot transfer setup. Future avenues will explore the role of these linguistic frameworks for other information extraction or reasoning tasks, and the impact of different kinds of linguistic frameworks such as AMRs or UMRs.

Acknowledgments

We thank the anonymous reveiwers for their constructive feedback. This project was funded in part by NSF grant 1949110.

References

Livio Baldini Soares, Nicholas FitzGerald, Jeffrey Ling, and Tom Kwiatkowski. 2019. Matching the blanks: Distributional similarity for relation learning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2895–2905, Florence, Italy. Association for Computational Linguistics.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin

Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract meaning representation for sembanking. In *Proceedings of the 7th linguistic annotation workshop and interoperability with discourse*, pages 178–186.

Dan Busbridge, Dane Sherburn, Pietro Cavallo, and Nils Y. Hammerla. 2019. Relational graph attention networks.

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, Online. Association for Computational Linguistics.

Jiawei Chen, Yaojie Lu, Hongyu Lin, Jie Lou, Wei Jia, Dai Dai, Hua Wu, Boxi Cao, Xianpei Han, and Le Sun. 2023. Learning in-context learning for named entity recognition. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13661–13675, Toronto, Canada. Association for Computational Linguistics.

Junjie Chen, Xiangheng He, and Yusuke Miyao. 2024. Language model based unsupervised dependency parsing with conditional mutual information and grammatical constraints. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 6355–6366, Mexico City, Mexico. Association for Computational Linguistics.

Noam Chomsky. 2002. *Syntactic structures*. Mouton de Gruyter.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020a. Unsupervised cross-lingual representation learning at scale. In *ACL*.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020b. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020c. Unsupervised cross-lingual representation learning at scale.

William Croft. 2022. Morphosyntax: constructions of the world's languages.

Sarkar Snigdha Sarathi Das, Arzoo Katiyar, Rebecca Passonneau, and Rui Zhang. 2022. CONTaiNER: Few-shot named entity recognition via contrastive learning. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6338–6353, Dublin, Ireland. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Jose A. Diaz-Garcia and Julio Amador Diaz Lopez. 2024. A survey on cutting-edge relation extraction techniques based on language models.

Qiang Ding, Lvzhou Luo, Yixuan Cao, and Ping Luo. 2024. Attention with dependency parsing augmentation for fine-grained attribution.

Ritam Dutt, Sopan Khosla, Vinayshekhar Bannihatti Kumar, and Rashmi Gangadharaiah. 2023. GrailQA++: A challenging zero-shot benchmark for knowledge base question answering. 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 897–909, Nusa Dua, Bali. Association for Computational Linguistics.

Sefika Efeoglu and Adrian Paschke. 2024. Relation extraction with fine-tuned large language models in retrieval augmented generation frameworks.

Alexander Fritzler, Varvara Logacheva, and Maksim Kretov. 2019. Few-shot classification in named entity recognition task. In *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, pages 993–1000.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, Arun Rao, Aston Zhang, Aurelien Rodriguez, Austen Gregerson, Ava Spataru, Baptiste Roziere, Bethany Biron, Binh Tang, Bobbie Chern, Charlotte Caucheteux, Chaya Nayak, Chloe Bi, Chris Marra, Chris McConnell, Christian Keller, Christophe Touret, Chunyang Wu, Corinne Wong, Cristian Canton Ferrer, Cyrus Nikolaidis, Damien Allonsius, Daniel Song, Danielle Pintz, Danny Livshits,

Danny Wyatt, David Esiobu, Dhruv Choudhary, Dhruv Mahajan, Diego Garcia-Olano, Diego Perino, Dieuwke Hupkes, Egor Lakomkin, Ehab AlBadawy, Elina Lobanova, Emily Dinan, Eric Michael Smith, Filip Radenovic, Francisco Guzmán, Frank Zhang, Gabriel Synnaeve, Gabrielle Lee, Georgia Lewis Anderson, Govind Thattai, Graeme Nail, Gregoire Mialon, Guan Pang, Guillem Cucurell, Hailey Nguyen, Hannah Korevaar, Hu Xu, Hugo Touvron, Iliyan Zarov, Imanol Arrieta Ibarra, Isabel Kloumann, Ishan Misra, Ivan Evtimov, Jack Zhang, Jade Copet, Jaewon Lee, Jan Geffert, Jana Vranes, Jason Park, Jay Mahadeokar, Jeet Shah, Jelmer van der Linde, Jennifer Billock, Jenny Hong, Jenya Lee, Jeremy Fu, Jianfeng Chi, Jianyu Huang, Jiawen Liu, Jie Wang, Jiecao Yu, Joanna Bitton, Joe Spisak, Jongsoo Park, Joseph Rocca, Joshua Johnstun, Joshua Saxe, Junteng Jia, Kalyan Vasuden Alwala, Karthik Prasad, Kartikeya Upasani, Kate Plawiak, Ke Li, Kenneth Heafield, Kevin Stone, Khalid El-Arini, Krithika Iyer, Kshitiz Malik, Kuenley Chiu, Kunal Bhalla, Kushal Lakhotia, Lauren Rantala-Yeary, Laurens van der Maaten, Lawrence Chen, Liang Tan, Liz Jenkins, Louis Martin, Lovish Madaan, Lubo Malo, Lukas Blecher, Lukas Landzaat, Luke de Oliveira, Madeline Muzzi, Mahesh Pasupuleti, Mannat Singh, Manohar Paluri, Marcin Kardas, Maria Tsimpoukelli, Mathew Oldham, Mathieu Rita, Maya Pavlova, Melanie Kambadur, Mike Lewis, Min Si, Mitesh Kumar Singh, Mona Hassan, Naman Goyal, Narjes Torabi, Nikolay Bashlykov, Nikolay Bogoychev, Niladri Chatterji, Ning Zhang, Olivier Duchenne, Onur Çelebi, Patrick Alrassy, Pengchuan Zhang, Pengwei Li, Petar Vasic, Peter Weng, Prajjwal Bhargava, Pratik Dubal, Praveen Krishnan, Punit Singh Koura, Puxin Xu, Qing He, Qingxiao Dong, Ragavan Srinivasan, Raj Ganapathy, Ramon Calderer, Ricardo Silveira Cabral, Robert Stojnic, Roberta Raileanu, Rohan Maheswari, Rohit Girdhar, Rohit Patel, Romain Sauvestre, Ronnie Polidoro, Roshan Sumbaly, Ross Taylor, Ruan Silva, Rui Hou, Rui Wang, Saghar Hosseini, Sahana Chennabasappa, Sanjay Singh, Sean Bell, Seohyun Sonia Kim, Sergey Edunov, Shaoliang Nie, Sharan Narang, Sharath Raparthy, Sheng Shen, Shengye Wan, Shruti Bhosale, Shun Zhang, Simon Vandenhende, Soumya Batra, Spencer Whitman, Sten Sootla, Stephane Collot, Suchin Gururangan, Sydney Borodinsky, Tamar Herman, Tara Fowler, Tarek Sheasha, Thomas Georgiou, Thomas Scialom, Tobias Speckbacher, Todor Mihaylov, Tong Xiao, Ujjwal Karn, Vedanuj Goswami, Vibhor Gupta, Vignesh Ramanathan, Viktor Kerkez, Vincent Gonguet, Virginie Do, Vish Vogeti, Vítor Albiero, Vladan Petrovic, Weiwei Chu, Wenhan Xiong, Wenyin Fu, Whitney Meers, Xavier Martinet, Xiaodong Wang, Xiaofang Wang, Xiaoqing Ellen Tan, Xide Xia, Xinfeng Xie, Xuchao Jia, Xuewei Wang, Yaelle Goldschlag, Yashesh Gaur, Yasmine Babaei, Yi Wen, Yiwen Song, Yuchen Zhang, Yue Li, Yuning Mao, Zacharie Delpierre Coudert, Zheng Yan, Zhengxing Chen, Zoe Papakipos, Aaditya Singh, Aayushi Srivastava, Abha Jain, Adam Kelsey, Adam Shajnfeld, Adithya Gangidi, Adolfo Victoria, Ahuva Goldstand, Ajay Menon, Ajay Sharma, Alex Boesenberg, Alexei Baevski, Allie Feinstein, Amanda Kallet, Amit Sangani, Amos Teo, Anam Yunus, Andrei Lupu, Andres Alvarado, Andrew Caples, Andrew Gu, Andrew Ho, Andrew Poulton, Andrew Ryan, Ankit Ramchandani, Annie Dong, Annie Franco, Anuj Goyal, Aparajita Saraf, Arkabandhu Chowdhury, Ashley Gabriel, Ashwin Bharambe, Assaf Eisenman, Azadeh Yazdan, Beau James, Ben Maurer, Benjamin Leonhardi, Bernie Huang, Beth Loyd, Beto De Paola, Bhargavi Paranjape, Bing Liu, Bo Wu, Boyu Ni, Braden Hancock, Bram Wasti, Brandon Spence, Brani Stojkovic, Brian Gamido, Britt Montalvo, Carl Parker, Carly Burton, Catalina Mejia, Ce Liu, Changhan Wang, Changkyu Kim, Chao Zhou, Chester Hu, Ching-Hsiang Chu, Chris Cai, Chris Tindal, Christoph Feichtenhofer, Cynthia Gao, Damon Civin, Dana Beaty, Daniel Kreymer, Daniel Li, David Adkins, David Xu, Davide Testuggine, Delia David, Devi Parikh, Diana Liskovich, Didem Foss, Dingkang Wang, Duc Le, Dustin Holland, Edward Dowling, Eissa Jamil, Elaine Montgomery, Eleonora Presani, Emily Hahn, Emily Wood, Eric-Tuan Le, Erik Brinkman, Esteban Arcaute, Evan Dunbar, Evan Smothers, Fei Sun, Felix Kreuk, Feng Tian, Filippos Kokkinos, Firat Ozgenel, Francesco Caggioni, Frank Kanayet, Frank Seide, Gabriela Medina Florez, Gabriella Schwarz, Gada Badeer, Georgia Swee, Gil Halpern, Grant Herman, Grigory Sizov, Guangyi, Zhang, Guna Lakshminarayanan, Hakan Inan, Hamid Shojanazeri, Han Zou, Hannah Wang, Hanwen Zha, Haroun Habeeb, Harrison Rudolph, Helen Suk, Henry Aspegren, Hunter Goldman, Hongyuan Zhan, Ibrahim Damlaj, Igor Molybog, Igor Tufanov, Ilias Leontiadis, Irina-Elena Veliche, Itai Gat, Jake Weissman, James Geboski, James Kohli, Janice Lam, Japhet Asher, Jean-Baptiste Gaya, Jeff Marcus, Jeff Tang, Jennifer Chan, Jenny Zhen, Jeremy Reizenstein, Jeremy Teboul, Jessica Zhong, Jian Jin, Jingyi Yang, Joe Cummings, Jon Carvill, Jon Shepard, Jonathan McPhie, Jonathan Torres, Josh Ginsburg, Junjie Wang, Kai Wu, Kam Hou U, Karan Saxena, Kartikay Khandelwal, Katayoun Zand, Kathy Matosich, Kaushik Veeraraghavan, Kelly Michelena, Keqian Li, Kiran Jagadeesh, Kun Huang, Kunal Chawla, Kyle Huang, Lailin Chen, Lakshya Garg, Lavender A, Leandro Silva, Lee Bell, Lei Zhang, Liangpeng Guo, Licheng Yu, Liron Moshkovich, Luca Wehrstedt, Madian Khabsa, Manav Avalani, Manish Bhatt, Martynas Mankus, Matan Hasson, Matthew Lennie, Matthias Reso, Maxim Groshev, Maxim Naumov, Maya Lathi, Meghan Keneally, Miao Liu, Michael L. Seltzer, Michal Valko, Michelle Restrepo, Mihir Patel, Mik Vyatskov, Mikayel Samvelyan, Mike Clark, Mike Macey, Mike Wang, Miquel Jubert Hermoso, Mo Metanat, Mohammad Rastegari, Munish Bansal, Nandhini Santhanam, Natascha Parks, Natasha White, Navyata Bawa, Nayan Singhal, Nick Egebo, Nicolas Usunier, Nikhil Mehta, Nikolay Pavlovich Laptev, Ning Dong, Norman Cheng, Oleg Chernoguz, Olivia Hart, Omkar Salpekar, Ozlem Kalinli, Parkin Kent, Parth Parekh, Paul Saab, Pavan Balaji, Pedro Rittner, Philip Bontrager, Pierre Roux, Piotr

Dollar, Polina Zvyagina, Prashant Ratanchandani, Pritish Yuvraj, Qian Liang, Rachad Alao, Rachel Rodriguez, Rafi Ayub, Raghotham Murthy, Raghu Nayani, Rahul Mitra, Rangaprabhu Parthasarathy, Raymond Li, Rebekkah Hogan, Robin Battey, Rocky Wang, Russ Howes, Ruty Rinott, Sachin Mehta, Sachin Siby, Sai Jayesh Bondu, Samyak Datta, Sara Chugh, Sara Hunt, Sargun Dhillon, Sasha Sidorov, Satadru Pan, Saurabh Mahajan, Saurabh Verma, Seiji Yamamoto, Sharadh Ramaswamy, Shaun Lindsay, Shaun Lindsay, Sheng Feng, Shenghao Lin, Shengxin Cindy Zha, Shishir Patil, Shiva Shankar, Shuqiang Zhang, Shuqiang Zhang, Sinong Wang, Sneha Agarwal, Soji Sajuyigbe, Soumith Chintala, Stephanie Max, Stephen Chen, Steve Kehoe, Steve Satterfield, Sudarshan Govindaprasad, Sumit Gupta, Summer Deng, Sungmin Cho, Sunny Virk, Suraj Subramanian, Sy Choudhury, Sydney Goldman, Tal Remez, Tamar Glaser, Tamara Best, Thilo Koehler, Thomas Robinson, Tianhe Li, Tianjun Zhang, Tim Matthews, Timothy Chou, Tzook Shaked, Varun Vontimitta, Victoria Ajayi, Victoria Montanez, Vijai Mohan, Vinay Satish Kumar, Vishal Mangla, Vlad Ionescu, Vlad Poenaru, Vlad Tiberiu Mihailescu, Vladimir Ivanov, Wei Li, Wenchen Wang, Wenwen Jiang, Wes Bouaziz, Will Constable, Xiaocheng Tang, Xiaojian Wu, Xiaolan Wang, Xilun Wu, Xinbo Gao, Yaniv Kleinman, Yanjun Chen, Ye Hu, Ye Jia, Ye Qi, Yenda Li, Yilin Zhang, Ying Zhang, Yossi Adi, Youngjin Nam, Yu, Wang, Yu Zhao, Yuchen Hao, Yundi Qian, Yunlu Li, Yuzi He, Zach Rait, Zachary DeVito, Zef Rosnbrick, Zhaoduo Wen, Zhenyu Yang, Zhiwei Zhao, and Zhiyu Ma. 2024. The llama 3 herd of models.

Yanggan Gu, Yang Hou, Zhefeng Wang, Xinyu Duan, and Zhenghua Li. 2024. High-order joint constituency and dependency parsing. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 8144–8154, Torino, Italia. ELRA and ICCL.

Yu Gu, Sue Kase, Michelle Vanni, Brian Sadler, Percy Liang, Xifeng Yan, and Yu Su. 2021. Beyond iid: three levels of generalization for question answering on knowledge bases. In *Proceedings of the Web Conference 2021*, pages 3477–3488.

Pere-Lluís Huguet Cabot, Simone Tedeschi, Axel-Cyrille Ngonga Ngomo, and Roberto Navigli. 2023. RED$^{\rm fm}$: a filtered and multilingual relation extraction dataset. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4326–4343, Toronto, Canada. Association for Computational Linguistics.

Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7b.

Zhao Jun, Hu Yuan, Xu Nuo, Gui Tao, Zhang Qi, Chen Yunwen, and Gao Xiang. 2022. An exploration of prompt-based zero-shot relation extraction method. In *Proceedings of the 21st Chinese National Conference on Computational Linguistics*, pages 786–797, Nanchang, China. Chinese Information Processing Society of China.

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, Singapore. Association for Computational Linguistics.

Tianming Liang, Yang Liu, Xiaoyan Liu, Hao Zhang, Gaurav Sharma, and Maozu Guo. 2023. Distantly-supervised long-tailed relation extraction using constraint graphs. *IEEE Transactions on Knowledge and Data Engineering*, 35(7):6852–6865.

Robert Litschko, Ivan Vulić, Simone Paolo Ponzetto, and Goran Glavaš. 2021. On cross-lingual retrieval with multilingual text encoders.

Yang Liu, Jinpeng Hu, Xiang Wan, and Tsung-Hui Chang. 2022. A simple yet effective relation information guided approach for few-shot relation extraction. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 757–763, Dublin, Ireland. Association for Computational Linguistics.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Zhu Liu, Cunliang Kong, Ying Liu, and Maosong Sun. 2024. Fantastic semantics and where to find them: Investigating which layers of generative LLMs reflect lexical semantics. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 14551–14558, Bangkok, Thailand. Association for Computational Linguistics.

Sewon Min, Julian Michael, Hannaneh Hajishirzi, and Luke Zettlemoyer. 2020. AmbigQA: Answering ambiguous open-domain questions. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5783–5797, Online. Association for Computational Linguistics.

Arijit Nag, Bidisha Samanta, Animesh Mukherjee, Niloy Ganguly, and Soumen Chakrabarti. 2021. A data bootstrapping recipe for low-resource multilingual relation classification. In *Proceedings of the 25th Conference on Computational Natural Language Learning*, pages 575–587, Online. Association for Computational Linguistics.

Minh Van Nguyen, Viet Dac Lai, Amir Pouran Ben Veyseh, and Thien Huu Nguyen. 2021. Trankit: A light-weight transformer-based toolkit for multilingual natural language processing. In *Proceedings of the 16th*

Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations, pages 80–90, Online. Association for Computational Linguistics.

Joakim Nivre, Marie-Catherine De Marneffe, Filip Ginter, Yoav Goldberg, Jan Hajic, Christopher D Manning, Ryan McDonald, Slav Petrov, Sampo Pyysalo, Natalia Silveira, et al. 2016. Universal dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 1659–1666.

Hao Peng, Tianyu Gao, Xu Han, Yankai Lin, Peng Li, Zhiyuan Liu, Maosong Sun, and Jie Zhou. 2020. Learning from Context or Names? An Empirical Study on Neural Relation Extraction. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3661–3672, Online. Association for Computational Linguistics.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020a. Stanza: A Python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020b. Stanza: A python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 101–108, Online. Association for Computational Linguistics.

Meng Qu, Tianyu Gao, Louis-Pascal Xhonneux, and Jian Tang. 2020. Few-shot relation extraction via Bayesian meta-learning on relation graphs. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 7867–7876. PMLR.

Siva Reddy, Oscar Täckström, Slav Petrov, Mark Steedman, and Mirella Lapata. 2017. Universal semantic parsing. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 89–101.

Devendra Sachan, Yuhao Zhang, Peng Qi, and William L. Hamilton. 2021. Do syntax trees help pre-trained transformers extract information? In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2647–2661, Online. Association for Computational Linguistics.

Gözde Gül Şahin. 2022. To augment or not to augment? a comparative study on text augmentation techniques for low-resource NLP. *Computational Linguistics*, 48(1):5–42.

Oscar Sainz, Iker García-Ferrero, Rodrigo Agerri, Oier Lopez de Lacalle, German Rigau, and Eneko Agirre. 2024. GoLLIE: Annotation guidelines improve zero-shot information-extraction. In *The Twelfth International Conference on Learning Representations*.

Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. 2008. The graph neural network model. *IEEE transactions on neural networks*, 20(1):61–80.

Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *European semantic web conference*, pages 593–607. Springer.

Giulio Starace, Konstantinos Papakostas, Rochelle Choenni, Apostolos Panagiotopoulos, Matteo Rosati, Alina Leidinger, and Ekaterina Shutova. 2023. Probing LLMs for joint encoding of linguistic categories. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 7158–7179, Singapore. Association for Computational Linguistics.

Chenguang Wang, Xiao Liu, Zui Chen, Haoyun Hong, Jie Tang, and Dawn Song. 2022a. DeepStruct: Pretraining of language models for structure prediction. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 803–823, Dublin, Ireland. Association for Computational Linguistics.

Shusen Wang, Bosen Zhang, Yajing Xu, Yanan Wu, and Bo Xiao. 2022b. RCL: Relation contrastive learning for zero-shot relation extraction. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 2456–2468, Seattle, United States. Association for Computational Linguistics.

Tingyu Xie, Qi Li, Yan Zhang, Zuozhu Liu, and Hongwei Wang. 2024. Self-improving for zero-shot named entity recognition with large language models. 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 583–593, Mexico City, Mexico. Association for Computational Linguistics.

Lilong Xue, Dan Zhang, Yuxiao Dong, and Jie Tang. 2024. Autore: Document-level relation extraction with large language models.

An Yang, Baosong Yang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Zhou, Chengpeng Li, Chengyuan Li, Dayiheng Liu, Fei Huang, Guanting Dong, Haoran Wei, Huan Lin, Jialong Tang, Jialin Wang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Ma, Jianxin Yang, Jin Xu, Jingren Zhou, Jinze Bai, Jinzheng He, Junyang Lin, Kai Dang, Keming Lu, Keqin Chen, Kexin Yang, Mei Li, Mingfeng Xue, Na Ni, Pei Zhang, Peng Wang, Ru Peng, Rui Men, Ruize Gao, Runji Lin, Shijie Wang, Shuai Bai, Sinan Tan, Tianhang Zhu, Tianhao Li, Tianyu Liu, Wenbin Ge, Xiaodong Deng, Xiaohuan Zhou, Xingzhang Ren, Xinyu Zhang, Xipin Wei, Xuancheng Ren, Xuejing Liu, Yang Fan, Yang Yao, Yichang Zhang, Yu Wan, Yunfei Chu, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, Zhifang Guo, and Zhihao Fan. 2024. Qwen2 technical report.

Kaijia Yang, Nantao Zheng, Xinyu Dai, Liang He, Shujian Huang, and Jiajun Chen. 2020. Enhance prototypical network with text descriptions for few-shot relation classification. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 2273–2276.

Zhi-Xiu Ye and Zhen-Hua Ling. 2019a. Distant supervision relation extraction with intra-bag and inter-bag attentions. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2810–2819, Minneapolis, Minnesota. Association for Computational Linguistics.

Zhi-Xiu Ye and Zhen-Hua Ling. 2019b. Multi-level matching and aggregation network for few-shot relation classification. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2872–2881, Florence, Italy. Association for Computational Linguistics.

Tianshu Yu, Min Yang, and Xiaoyan Zhao. 2022. Dependency-aware prototype learning for few-shot relation classification. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 2339–2345, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Daniel Zeman, Joakim Nivre, Mitchell Abrams, Noëmi Aepli, Željko Agić, Lars Ahrenberg, Gabrielė Aleksandravičiūtė, Lene Antonsen, Katya Aplonova, Maria Jesus Aranzabe, Gashaw Arutie, Masayuki Asahara, Luma Ateyah, Mohammed Attia, Aitziber Atutxa, Liesbeth Augustinus, Elena Badmaeva, Miguel Ballesteros, Esha Banerjee, Sebastian Bank, Verginica Barbu Mititelu, Victoria Basmov, Colin Batchelor, John Bauer, Sandra Bellato, Kepa Bengoetxea, Yevgeni Berzak, Irshad Ahmad Bhat, Riyaz Ahmad Bhat, Erica Biagetti, Eckhard Bick, Agnė Bielinskienė, Rogier Blokland, Victoria Bobicev, Loïc Boizou, Emanuel Borges Völker, Carl Börstell, Cristina Bosco, Gosse Bouma, Sam Bowman, Adriane Boyd, Kristina Brokaitė, Aljoscha Burchardt, Marie Candito, Bernard Caron, Gauthier Caron, Tatiana Cavalcanti, Gülşen Cebiroğlu Eryiğit, Flavio Massimiliano Cecchini, Giuseppe G. A. Celano, Slavomír Čéplö, Savas Cetin, Fabricio Chalub, Jinho Choi, Yongseok Cho, Jayeol Chun, Alessandra T. Cignarella, Silvie Cinková, Aurélie Collomb, Çağrı Çöltekin, Miriam Connor, Marine Courtin, Elizabeth Davidson, Marie-Catherine de Marneffe, Valeria de Paiva, Elvis de Souza, Arantza Diaz de Ilarraza, Carly Dickerson, Bamba Dione, Peter Dirix, Kaja Dobrovoljc, Timothy Dozat, Kira Droganova, Puneet Dwivedi, Hanne Eckhoff, Marhaba Eli, Ali Elkahky, Binyam Ephrem, Olga Erina, Tomaž Erjavec, Aline Etienne, Wograine Evelyn, Richárd Farkas, Hector Fernandez Alcalde, Jennifer Foster, Cláudia Freitas, Kazunori Fujita, Katarína Gajdošová, Daniel Galbraith, Marcos Garcia, Moa Gärdenfors, Sebastian Garza, Kim Gerdes, Filip Ginter, Iakes Goenaga, Koldo Gojenola, Memduh Gökırmak, Yoav Goldberg, Xavier Gómez Guinovart, Berta González Saavedra, Bernadeta Griciūtė, Matias Grioni, Normunds Grūzītis, Bruno Guillaume, Céline Guillot-Barbance, Nizar Habash, Jan Hajič, Jan Hajič jr., Mika Hämäläinen, Linh Hà Mỹ, Na-Rae Han, Kim Harris, Dag Haug, Johannes Heinecke, Felix Hennig, Barbora Hladká, Jaroslava Hlaváčová, Florinel Hociung, Petter Hohle, Jena Hwang, Takumi Ikeda, Radu Ion, Elena Irimia, Ọlájídé Ishola, Tomáš Jelínek, Anders Johannsen, Fredrik Jørgensen, Markus Juutinen, Hüner Kaşıkara, Andre Kaasen, Nadezhda Kabaeva, Sylvain Kahane, Hiroshi Kanayama, Jenna Kanerva, Boris Katz, Tolga Kayadelen, Jessica Kenney, Václava Kettnerová, Jesse Kirchner, Elena Klementieva, Arne Köhn, Kamil Kopacewicz, Natalia Kotsyba, Jolanta Kovalevskaitė, Simon Krek, Sookyoung Kwak, Veronika Laippala, Lorenzo Lambertino, Lucia Lam, Tatiana Lando, Septina Dian Larasati, Alexei Lavrentiev, John Lee, Phng Lê Hồng, Alessandro Lenci, Saran Lertpradit, Herman Leung, Cheuk Ying Li, Josie Li, Keying Li, KyungTae Lim, Maria Liovina, Yuan Li, Nikola Ljubešić, Olga Loginova, Olga Lyashevskaya, Teresa Lynn, Vivien Macketanz, Aibek Makazhanov, Michael Mandl, Christopher Manning, Ruli Manurung, Cătălina Mărănduc, David Mareček, Katrin Marheinecke, Héctor Martínez Alonso, André Martins, Jan Mašek, Yuji Matsumoto, Ryan McDonald, Sarah McGuinness, Gustavo Mendonça, Niko Miekka, Margarita Misirpashayeva, Anna Missilä, Cătălin Mititelu, Maria Mitrofan, Yusuke Miyao, Simonetta Montemagni, Amir More, Laura Moreno Romero, Keiko Sophie Mori, Tomohiko Morioka, Shinsuke Mori, Shigeki Moro, Bjartur Mortensen, Bohdan Moskalevskyi, Kadri Muischnek, Robert Munro, Yugo Murawaki, Kaili Müürisep, Pinkey Nainwani, Juan Ignacio Navarro Horñiacek, Anna Nedoluzhko, Gunta Nešpore-Bērzkalne, Lng Nguyễn Thị, Huyền Nguyễn Thị Minh, Yoshihiro Nikaido, Vitaly Nikolaev, Rattima Nitisaroj, Hanna Nurmi, Stina Ojala, Atul Kr. Ojha, Adédayọ Olúòkun, Mai Omura, Petya Osenova, Robert Östling, Lilja Øvrelid, Niko Partanen, Elena Pascual, Marco Passarotti, Agnieszka Patejuk, Guilherme Paulino-Passos, Angelika Peljak-Łapińska, Siyao Peng, Cenel-Augusto Perez, Guy Perrier, Daria Petrova, Slav Petrov, Jason Phelan, Jussi Piitulainen, Tommi A Pirinen, Emily Pitler, Barbara Plank, Thierry Poibeau, Larisa Ponomareva, Martin Popel, Lauma Pretkalniņa, Sophie Prévost, Prokopis Prokopidis, Adam Przepiórkowski, Tiina Puolakainen, Sampo Pyysalo, Peng Qi, Andriela Rääbis, Alexandre Rademaker, Loganathan Ramasamy, Taraka Rama, Carlos Ramisch, Vinit Ravishankar, Livy Real, Siva Reddy, Georg Rehm, Ivan Riabov, Michael Rießler, Erika Rimkutė, Larissa Rinaldi, Laura Rituma, Luisa Rocha, Mykhailo Romanenko, Rudolf Rosa, Davide Rovati, Valentin Roșca, Olga Rudina, Jack Rueter, Shoval Sadde, Benoît Sagot, Shadi Saleh, Alessio Salomoni, Tanja Samardžić, Stephanie Samson, Manuela Sanguinetti, Dage Särg, Baiba Saulīte, Yanin Sawanakunanon, Nathan Schneider, Sebastian Schuster, Djamé Seddah, Wolfgang Seeker, Mojgan Seraji, Mo Shen, Atsuko Shimada, Hiroyuki Shirasu, Muh Shohibussirri, Dmitry Sichinava, Aline Silveira, Natalia Silveira, Maria Simi, Radu Simionescu, Katalin Simkó, Mária Šimková, Kiril Simov, Aaron Smith, Isabela Soares-Bastos, Carolyn Spadine, Antonio Stella, Milan Straka, Jana Strnadová, Alane Suhr, Umut Sulubacak, Shingo Suzuki, Zsolt Szántó, Dima Taji, Yuta Takahashi, Fabio Tamburini, Takaaki Tanaka, Isabelle Tellier, Guillaume Thomas, Liisi Torga, Trond Trosterud, Anna Trukhina, Reut Tsarfaty, Francis Tyers, Sumire Uematsu, Zdeňka Urešová, Larraitz Uria, Hans Uszkoreit, Andrius Utka, Sowmya Vajjala, Daniel van Niekerk, Gertjan van Noord, Viktor Varga, Eric Villemonte de la Clergerie, Veronika Vincze, Lars Wallin, Abigail Walsh, Jing Xian Wang, Jonathan North Washington, Maximilan Wendt, Seyi Williams, Mats Wirén, Christian Wittern, Tsegay Woldemariam, Tak-sum Wong, Alina Wróblewska, Mary Yako, Naoki Yamazaki, Chunxiao Yan, Koichi Yasuoka, Marat M. Yavrumyan, Zhuoran Yu, Zdeněk Žabokrtský, Amir Zeldes, Manying Zhang, and Hanzhi Zhu. 2019. Universal dependencies 2.5. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

Jiali Zeng, Yufan Jiang, Yongjing Yin, Xu Wang, Binghuai Lin, and Yunbo Cao. 2022. DualNER: A dual-teaching framework for zero-shot cross-lingual named entity recognition. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 1837–1843, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Chen Zhang, Yuxuan Lai, Yansong Feng, Xingyu Shen, Haowei Du, and Dongyan Zhao. 2023a. Cross-lingual question answering over knowledge base as reading comprehension. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 2439–2452, Dubrovnik, Croatia. Association for Computational Linguistics.

Heidi Zhang, Sina Semnani, Farhad Ghassemi, Jialiang Xu, Shicheng Liu, and Monica Lam. 2024. SPAGHETTI: Open-domain question answering from heterogeneous data sources with retrieval and semantic parsing. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 1663–1678, Bangkok, Thailand. Association for Computational Linguistics.

Qin Zhang, Shangsi Chen, Dongkuan Xu, Qingqing Cao, Xiaojun Chen, Trevor Cohn, and Meng Fang. 2023b. A survey for efficient open domain question answering. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 14447–14465, Toronto, Canada. Association for Computational Linguistics.

Xinyu Zhao, Shih-Ting Lin, and Greg Durrett. 2021. Effective distant supervision for temporal relation extraction. In *Proceedings of the Second Workshop on Domain Adaptation for NLP*, pages 195–203, Kyiv, Ukraine. Association for Computational Linguistics.

Ben Zhou, Daniel Khashabi, Chen-Tse Tsai, and Dan Roth. 2019. Zero-shot open entity typing as type-compatible grounding. *arXiv preprint arXiv:1907.03228*.

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, Miami, Florida, USA. Association for Computational Linguistics.

A Appendices

In this section, we provide extra figures and tables to further corroborate our experiments in this work. Additionally, we also present a statistical significance analysis of our results using the Anova method.

Dataset	Source	Encoder	Parser	Sent Length		Lex Length		Dep Length		# Docs	# Rels
				Mean	Median	Mean	Median	Mean	Median		
Indore	en	mBERT	stanza	31.23	29.0	13.92	11.0	5.43	5.0	8486	51
	hi	mBERT	stanza	66.76	56.0	27.29	21.0	5.70	5.0	6963	51
	te	mBERT	stanza	151.45	140.0	44.74	31.0	6.05	6.0	8154	51
	en	mBERT	trankit	31.23	29.0	13.92	11.0	5.42	5.0	8486	51
	hi	mBERT	trankit	66.76	56.0	27.29	21.0	5.85	6.0	6963	51
	te	mBERT	trankit	151.45	140.0	44.74	31.0	5.69	6.0	8154	51
	en	XLMR	stanza	34.40	32.0	15.95	13.0	5.43	5.0	8486	51
	hi	XLMR	stanza	56.25	48.0	22.85	17.0	5.70	5.0	6963	51
	te	XLMR	stanza	125.60	113.0	37.43	26.0	6.05	6.0	8154	51
	en	XLMR	trankit	34.40	32.0	15.95	13.0	5.42	5.0	8486	51
	hi	XLMR	trankit	56.25	48.0	22.85	17.0	5.85	6.0	6963	51
	te	XLMR	trankit	125.60	113.0	37.43	26.0	5.69	6.0	8154	51
RedFM	en	mBERT	stanza	117.53	107.0	27.96	17.0	6.40	6.0	10899	32
	es	mBERT	stanza	103.77	91.0	25.97	17.0	6.24	6.0	6538	32
	fr	mBERT	stanza	92.18	78.0	23.19	15.0	5.86	5.0	7383	32
	it	mBERT	stanza	79.31	65.0	20.56	14.0	5.80	5.0	6812	32
	de	mBERT	stanza	88.79	79.0	22.97	15.0	5.28	5.0	7497	32
	ar	mBERT	stanza	107.15	90.0	29.68	22.0	6.08	6.0	1846	32
	zh	mBERT	stanza	118.18	101.0	35.25	22.0	6.69	6.0	1384	32
	en	mBERT	trankit	117.53	107.0	27.96	17.0	6.37	6.0	10899	32
	es	mBERT	trankit	103.77	91.0	25.97	17.0	6.16	6.0	6538	32
	fr	mBERT	trankit	92.18	78.0	23.19	15.0	5.68	5.0	7383	32
	it	mBERT	trankit	79.31	65.0	20.56	14.0	5.64	5.0	6812	32
	de	mBERT	trankit	88.79	79.0	22.97	15.0	5.16	5.0	7497	32
	ar	mBERT	trankit	107.15	90.0	29.68	22.0	6.18	6.0	1846	32
	zh	mBERT	trankit	118.18	101.0	35.25	22.0	6.75	6.0	1384	32
	en	XLMR	stanza	130.33	119.0	31.52	19.0	6.40	6.0	10899	32
	es	XLMR	stanza	112.22	100.0	28.86	19.0	6.24	6.0	6538	32
	fr	XLMR	stanza	103.20	86.0	26.77	17.0	5.86	5.0	7383	32
	it	XLMR	stanza	85.14	71.0	22.72	16.0	5.80	5.0	6812	32
	de	XLMR	stanza	96.23	87.0	25.53	17.0	5.28	5.0	7497	32
	ar	XLMR	stanza	95.81	81.0	26.64	19.0	6.08	6.0	1846	32
	zh	XLMR	stanza	96.10	81.0	28.35	18.0	6.69	6.0	1384	32
	en	XLMR	trankit	130.33	119.0	31.52	19.0	6.37	6.0	10899	32
	es	XLMR	trankit	112.22	100.0	28.86	19.0	6.16	6.0	6538	32
	fr	XLMR	trankit	103.20	86.0	26.77	17.0	5.68	5.0	7383	32
	it	XLMR	trankit	85.14	71.0	22.72	16.0	5.64	5.0	6812	32
	de	XLMR	trankit	96.23	87.0	25.53	17.0	5.16	5.0	7497	32
	ar	XLMR	trankit	95.81	81.0	26.64	19.0	6.18	6.0	1846	32
	zh	XLMR	trankit	96.10	81.0	28.35	18.0	6.75	6.0	1384	32

Table 4: Combined Statistics for Indore and RedFM Datasets

Without Any Dependency Information:
Given the sentence: "The Porsche Panamera is a mid/full-sized luxury vehicle (E-segment in Europe) manufactured by the <e2>German</e2> automobile manufacturer <e1>Porsche</e1>. It is front-engined and has a rear-wheel-drive layout, with all-wheel drive versions also available.", which one of the following relations between the two entities <e1> and <e2> is being discussed? Choose one from this list of 32 options:\n0: country\n1: place of birth \n2: spouse\n3: country of citizenship\n4: instance of\n5: capital\n6: child\n7: shares border with\n8: author\n9: director\n10: occupation\n11: founded by\n12: league\n13: owned by\n14: genre\n15: named after\n16: follows\n17: headquarters location\n18: cast member\n19: manufacturer\n20: located in or next to body of water\n21: location\n22: part of\n23: mouth of the watercourse\n24: member of\n25: sport \n26: characters\n27: participant\n28: notable work\n29: replaces\n30: sibling\n31: inception\n\n. The answer is :
Tuple Format Prompt:
Given the sentence: "The Porsche Panamera is a mid/full-sized luxury vehicle (E-segment in Europe) manufactured by the <e2>German</e2> automobile manufacturer <e1>Porsche</e1>. It is front-engined and has a rear-wheel-drive layout, with all-wheel drive versions also available.", which one of the following relations between the two entities <e1> and <e2> is being discussed? We also provide the dependency parse in the form of head, rel, and word: {"head": "Panamera", "rel": "det", "word": "The"}, {"head": "Panamera", "rel": "compound", "word": "Porsche"}, {"head": "vehicle", "rel": "nsubj", "word": "Panamera"}, {"head": "vehicle", "rel": "cop", "word": "is"}, {"head": "vehicle", "rel": "det", "word": "a"}, {"head": "sized", "rel": "compound", "word": "mid"}, {"head": "sized", "rel": "punct", "word": "/"}, {"head": "sized", "rel": "amod", "word": "full"}, {"head": "sized", "rel": "punct", "word": "-"}, {"head": "vehicle", "rel": "amod", "word": "sized"}, {"head": "vehicle", "rel": "compound", "word": "luxury"}, {"head": "ROOT", "rel": "root", "word": "vehicle"}, {"head": "segment", "rel": "punct", "word": "("}, {"head": "segment", "rel": "compound", "word": "E"}, {"head": "segment", "rel": "punct", "word": "-"}, {"head": "vehicle", "rel": "appos", "word": "segment"}, {"head": "Europe", "rel": "case", "word": "in"}, {"head": "segment", "rel": "nmod", "word": "Europe"}, {"head": "segment", "rel": "punct", "word": ")"}, {"head": "vehicle", "rel": "acl", "word": "manufactured"}, {"head": "manufacturer", "rel": "case", "word": "by"}, {"head": "manufacturer", "rel": "det", "word": "the"}, {"head": "manufacturer", "rel": "amod", "word": "German"}, {"head": "manufacturer", "rel": "compound", "word": "automobile"}, {"head": "manufactured", "rel": "obl", "word": "manufacturer"}, {"head": "manufacturer", "rel": "appos", "word": "Porsche"}, {"head": "vehicle", "rel": "punct", "word": "."}, {"head": "engined", "rel": "nsubj", "word": "It"}, {"head": "engined", "rel": "cop", "word": "is"}, {"head": "engined", "rel": "obl:npmod", "word": "front"}, {"head": "engined", "rel": "punct", "word": "-"}, {"head": "ROOT", "rel": "root", "word": "engined"}, {"head": "has", "rel": "cc", "word": "and"}, {"head": "engined", "rel": "conj", "word": "has"}, {"head": "layout", "rel": "det", "word": "a"}, {"head": "drive", "rel": "amod", "word": "rear"}, {"head": "drive", "rel": "punct", "word": "-"}, {"head": "drive", "rel": "compound", "word": "wheel"}, {"head": "drive", "rel": "punct", "word": "-"}, {"head": "layout", "rel": "amod", "word": "drive"}, {"head": "has", "rel": "obj", "word": "layout"}, {"head": "layout", "rel": "punct", "word": ","}, {"head": "available", "rel": "mark", "word": "with"}, {"head": "drive", "rel": "det", "word": "all"}, {"head": "drive", "rel": "punct", "word": "-"}, {"head": "drive", "rel": "compound", "word": "wheel"}, {"head": "versions", "rel": "compound", "word": "drive"}, {"head": "available", "rel": "nsubj", "word": "versions"}, {"head": "available", "rel": "advmod", "word": "also"}, {"head": "layout", "rel": "acl", "word": "available"}, {"head": "engined", "rel": "punct", "word": "."}. Choose one from this list of 32 options:\n0: country\n1: place of birth \n2: spouse\n3: country of citizenship\n4: instance of\n5: capital\n6: child\n7: shares border with\n8: author\n9: director\n10: occupation\n11: founded by\n12: league\n13: owned by\n14: genre\n15: named after\n16: follows\n17: headquarters location\n18: cast member\n19: manufacturer\n20: located in or next to body of water\n21: location\n22: part of\n23: mouth of the watercourse\n24: member of\n25: sport \n26: characters\n27: participant\n28: notable work\n29: replaces\n30: sibling\n31: inception\n\n. The answer is :

Table 5: Prompt without dependency information and the tuple format prompt are used for relation extraction on the English subset of the RedFM dataset with Trankit as the dependency parser.

Text Prompt:
Given the sentence: T̈he Porsche Panamera is a mid/full-sized luxury vehicle (E-segment in Europe) manufactured by the <e2>German</e2> automobile manufacturer <e1>Porsche</e1>. It is front-engined and has a rear-wheel-drive layout, with all-wheel drive versions also available.̈, which one of the following relations between the two entities <e1> and <e2> is being discussed?\We also provide the dependency parses as follows: T̈he is Determiner of Panamera, Porsche is Compound noun modifier of Panamera, Panamera is Nominal subject of vehicle, is is Copula of vehicle, a is Determiner of vehicle, mid/ is Adverbial modifier of sized, full is Adjectival modifier of sized, - is Punctuation of sized, sized is Adjectival modifier of vehicle, luxury is Compound noun modifier of vehicle, vehicle is the root word, (is Punctuation of E, E is Appositional modifier of vehicle, - is Punctuation of segment, segment is Unspecified dependency of E, in is Case marker of Europe, Europe is Nominal modifier of segment,) is Punctuation of segment, manufactured is Clausal modifier of noun of vehicle, by is Case marker of Porsche, the is Determiner of Porsche, German is Adjectival modifier of Porsche, automobile is Compound noun modifier of manufacturer, manufacturer is Compound noun modifier of Porsche, Porsche is Oblique nominal of manufactured, . is Punctuation of vehicle, It is Nominal subject of engined, is is Copula of engined, front is Adjectival modifier of engined, - is Punctuation of front, engined is the root word, and is Coordinating conjunction of has, has is Conjunction of engined, a is Determiner of layout, rear is Compound noun modifier of drive, - is Punctuation of wheel, wheel is Compound noun modifier of drive, - is Punctuation of drive, drive is Compound noun modifier of layout, layout is Object of has, , is Punctuation of available, with is Marker of available, all is Determiner of wheel, - is Punctuation of all, wheel is Compound noun modifier of drive, drive is Compound noun modifier of versions, versions is Nominal subject of available, also is Adverbial modifier of available, available is Adverbial clause modifier of has, . is Punctuation of engined, \Choose one from this list of 32 options:\n0: country\n1: place of birth \n2: spouse\n3: country of citizenship\n4: instance of\n5: capital\n6: child\n7: shares border with\n8: author\n9: director\n10: occupation\n11: founded by\n12: league\n13: owned by\n14: genre\n15: named after\n16: follows\n17: headquarters location\n18: cast member\n19: manufacturer\n20: located in or next to body of water\n21: location\n22: part of\n23: mouth of the watercourse\n24: member of\n25: sport \n26: characters\n27: participant\n28: notable work\n29: replaces\n30: sibling\n31: inception\n\nThe answer is : "
Filtered Text Prompt:
Given the sentence: T̈he Porsche Panamera is a mid/full-sized luxury vehicle (E-segment in Europe) manufactured by the <e2>German</e2> automobile manufacturer <e1>Porsche</e1>. It is front-engined and has a rear-wheel-drive layout, with all-wheel drive versions also available.̈, which one of the following relations between the two entities <e1> and <e2> is being discussed?\n We also provide the dependency parses as follows: P̈orsche is Adjectival modifier of German, \n Choose one from this list of 32 options:\n0: country\n1: place of birth \n2: spouse\n3: country of citizenship\n4: instance of\n5: capital\n6: child\n7: shares border with\n8: author\n9: director\n10: occupation\n11: founded by\n12: league\n13: owned by\n14: genre\n15: named after\n16: follows\n17: headquarters location\n18: cast member\n19: manufacturer\n20: located in or next to body of water\n21: location\n22: part of\n23: mouth of the watercourse\n24: member of\n25: sport \n26: characters\n27: participant\n28: notable work\n29: replaces\n30: sibling\n31: inception\n\n. The answer is :

Table 6: Text prompt and Filtered Text prompts used for relation extraction on the English subset of the RedFM dataset with Trankit as the dependency parser.

Table 7: Zero-shot cross-lingual performance for Relation Extraction on the RedFM dataset using different combinations of multi-lingual encoder and dependency parse information and GNN. Highest values in each column are in bold. The rows and columns correspond to the source and target language respectively.

					mBERT				
Src	DEP	GNN	en	es	fr	it	de	ar	zh
en	-	-	-	80.4±0.2	80.7±0.4	**77.3±1.3**	78.8±0.9	72.7±0.8	70.4±0.6
en	stanza	rgcn	-	79.6±0.8	**80.9±1.4**	76.2±1.0	**80.2±0.5**	74.4±0.9	72.0±0.8
en	stanza	rgat	-	80.3±0.4	80.3±0.2	74.8±1.2	79.5±0.3	74.1±0.9	**72.3±0.4**
en	trankit	rgcn	-	80.1±0.4	80.8±0.5	73.8±0.2	79.3±0.7	73.8±1.8	69.5±0.6
en	trankit	rgat	-	**80.8±0.3**	80.7±0.2	74.4±1.8	79.0±0.7	**74.5±0.7**	70.1±0.6
es	-	-	77.6±0.1	-	77.2±0.8	76.4±0.6	75.9±0.7	70.9±1.6	70.8±1.1
es	stanza	rgcn	78.0±0.4	-	**82.6±0.8**	77.6±1.4	76.9±1.3	73.2±0.5	69.2±0.6
es	stanza	rgat	79.1±0.2	-	78.4±0.5	77.4±1.3	76.2±0.7	**73.5±0.9**	69.7±1.5
es	trankit	rgcn	79.3±0.9	-	80.6±1.6	76.3±0.6	77.1±1.0	73.3±0.3	**71.5±1.6**
es	trankit	rgat	**80.0±1.1**	-	78.7±0.5	**78.3±1.0**	**77.7±0.8**	72.6±1.2	71.2±2.6
fr	-	-	76.6±2.9	80.4±1.3	-	76.9±2.0	74.8±1.6	70.2±1.1	66.4±2.6
fr	stanza	rgcn	76.6±0.3	82.1±1.0	-	77.7±0.7	76.6±0.2	70.4±0.8	66.8±0.9
fr	stanza	rgat	80.0±0.7	82.1±0.9	-	77.0±1.0	77.5±1.5	71.5±1.0	67.5±1.2
fr	trankit	rgcn	78.6±0.3	**83.3±1.6**	-	**78.7±1.1**	**78.8±2.5**	**72.4±0.5**	**69.7±0.7**
fr	trankit	rgat	**80.1±0.8**	79.7±2.1	-	76.6±1.5	77.4±0.1	70.9±0.8	68.4±0.5
it	-	-	75.4±0.4	83.1±0.5	77.7±1.1	-	72.9±1.1	**73.0±2.0**	70.8±1.0
it	stanza	rgcn	**79.0±0.6**	83.0±0.7	77.2±1.0	-	74.7±1.4	70.8±0.3	70.0±0.7
it	stanza	rgat	76.7±0.9	**83.8±0.7**	77.5±0.5	-	75.7±1.5	72.2±1.6	70.5±0.4
it	trankit	rgcn	77.1±1.4	82.3±0.3	77.2±0.6	-	76.0±1.2	71.0±1.0	69.2±1.9
it	trankit	rgat	77.1±0.1	82.5±0.4	**77.8±0.5**	-	**76.3±0.1**	71.7±1.0	**71.5±0.9**
de	-	-	80.4±1.0	80.0±0.4	**78.3±0.1**	76.1±1.5	-	**75.8±1.9**	71.6±1.2
de	stanza	rgcn	80.0±0.2	80.4±0.7	76.7±0.3	75.8±0.8	-	74.2±0.8	70.0±1.9
de	stanza	rgat	79.2±0.4	**81.3±1.1**	78.1±1.4	76.6±2.7	-	74.6±0.5	**71.7±0.6**
de	trankit	rgcn	79.7±0.3	80.6±1.4	77.9±0.3	75.1±0.4	-	73.3±1.0	70.1±0.1
de	trankit	rgat	**80.7±0.7**	79.2±0.1	77.8±0.6	**77.4±0.5**	-	73.7±0.0	70.6±0.8
					XLMR				
en	-	-	-	73.1±1.8	**72.8±2.8**	64.2±3.7	**75.6±1.7**	61.7±1.8	64.4±1.0
en	stanza	rgcn	-	74.4±1.3	**72.7±0.5**	**67.4±1.3**	74.6±0.7	**63.2±1.5**	65.1±0.9
en	stanza	rgat	-	73.1±0.7	**72.7±1.4**	66.5±3.5	71.1±1.0	59.6±2.7	62.2±0.4
en	trankit	rgcn	-	74.4±1.5	72.0±1.8	65.4±2.2	71.5±1.6	62.6±1.8	64.6±1.3
en	trankit	rgat	-	**74.9±0.7**	70.3±0.1	62.4±1.6	73.9±0.4	61.5±1.7	**66.5±1.7**
es	-	-	73.3±0.4	-	74.3±0.4	**70.1±1.4**	**70.6±0.7**	63.2±3.1	**65.9±1.9**
es	stanza	rgcn	73.4±2.2	-	75.1±0.3	68.3±2.5	67.3±0.6	61.9±1.2	62.4±1.4
es	stanza	rgat	72.7±1.9	-	75.2±1.0	69.3±1.6	67.3±0.3	60.5±1.4	62.8±1.8
es	trankit	rgcn	**73.8±1.0**	-	75.9±1.5	69.8±1.8	70.0±2.5	**64.3±2.1**	65.6±2.7
es	trankit	rgat	71.4±1.2	-	**76.2±1.2**	68.0±1.5	68.7±2.0	60.0±0.9	62.5±2.3
fr	-	-	71.1±0.9	75.0±0.6	-	68.9±0.6	68.5±1.3	**61.5±1.2**	59.4±2.6
fr	stanza	rgcn	**74.3±1.7**	74.1±1.1	-	**69.7±0.6**	**72.2±1.3**	58.7±0.6	**62.9±2.7**
fr	stanza	rgat	70.1±1.5	73.9±1.3	-	67.0±1.5	66.2±1.0	59.0±0.9	60.3±1.6
fr	trankit	rgcn	70.0±0.2	74.4±0.5	-	68.4±0.7	66.4±0.7	58.9±2.2	59.5±1.8
fr	trankit	rgat	71.8±1.3	**76.0±0.7**	-	68.2±0.8	70.6±1.0	**61.5±1.2**	59.9±1.3
it	-	-	71.2±1.1	76.1±1.6	72.2±0.9	-	68.2±1.7	60.8±0.5	62.0±1.7
it	stanza	rgcn	73.3±2.0	76.1±0.8	74.3±1.3	-	67.2±2.1	61.8±0.3	63.1±0.3
it	stanza	rgat	**74.9±1.0**	76.0±0.2	74.2±1.3	-	68.9±0.2	62.2±0.1	**64.7±1.5**
it	trankit	rgcn	73.3±1.2	77.0±0.7	74.8±1.6	-	70.0±1.7	**64.5±1.0**	**64.7±1.0**
it	trankit	rgat	72.6±1.9	**78.7±0.5**	**76.6±0.2**	-	**70.2±1.0**	63.6±3.4	64.6±1.5
de	-	-	75.0±1.5	72.4±0.9	69.3±1.3	64.1±0.3	-	60.8±0.7	64.0±1.2
de	stanza	rgcn	72.6±1.5	73.4±2.1	70.8±1.9	65.2±0.5	-	60.6±0.8	**66.0±1.9**
de	stanza	rgat	**76.1±1.5**	**73.5±0.2**	**71.5±1.3**	**69.0±2.8**	-	**64.0±1.6**	65.8±1.7
de	trankit	rgcn	74.1±1.0	72.8±0.8	69.6±1.8	63.6±2.3	-	63.4±1.0	64.5±1.9
de	trankit	rgat	75.0±0.5	73.2±1.6	70.3±1.3	64.9±1.0	-	63.7±0.5	64.4±3.5

Table 8: Zero-shot cross-lingual performance for Relation Extraction on the IndoRE dataset using different combinations of multi-lingual encoder and dependency parse information and GNN. Highest values in each column are in bold. The rows and columns correspond to the source and target language respectively.

Src	DEP	GNN	en	hi	te
mBERT					
en	-	-	-	60.7±0.6	35.3±0.8
en	stanza	rgcn	-	60.1±0.4	38.3±1.2
en	stanza	rgat	-	58.7±0.3	**40.6**±**2.2**
en	trankit	rgcn	-	**62.5**±**0.8**	38.0±1.4
en	trankit	rgat	-	61.8±1.0	37.8±1.8
hi	-	-	**69.7**±**1.9**	-	49.5±2.3
hi	stanza	rgcn	68.6±0.6	-	49.4±0.8
hi	stanza	rgat	67.8±2.3	-	49.7±0.6
hi	trankit	rgcn	68.1±0.8	-	49.6±2.2
hi	trankit	rgat	68.0±1.6	-	**53.9**±**0.9**
te	-	-	45.3±1.7	54.4±2.6	-
te	stanza	rgcn	45.6±1.4	54.0±1.3	-
te	stanza	rgat	44.8±0.3	**56.6**±**0.3**	-
te	trankit	rgcn	**47.7**±**0.8**	54.2±0.1	-
te	trankit	rgat	46.1±1.2	54.2±2.5	-
XLMR					
en	-	-	-	57.4±2.3	37.2±2.5
en	stanza	rgcn	-	55.3±1.2	37.0±1.6
en	stanza	rgat	-	55.5±2.3	37.8±1.9
en	trankit	rgcn	-	58.8±0.5	36.4±3.8
en	trankit	rgat	-	**61.0**±**2.5**	**39.0**±**4.0**
hi	-	-	59.1±1.8	-	53.7±1.0
hi	stanza	rgcn	57.4±1.3	-	**54.7**±**1.2**
hi	stanza	rgat	**61.0**±**2.5**	-	**54.8**±**2.1**
hi	trankit	rgcn	59.5±0.8	-	54.3±1.8
hi	trankit	rgat	57.3±2.4	-	**54.8**±**2.3**
te	-	-	40.9±2.6	52.8±0.7	-
te	stanza	rgcn	41.2±2.2	**55.5**±**0.9**	-
te	stanza	rgat	39.0±0.7	52.0±3.2	-
te	trankit	rgcn	**41.8**±**0.6**	53.7±0.6	-
te	trankit	rgat	41.4±0.3	53.7±1.8	-

Source	sum_sq	df	F	P(>F)
C(src)	1.844	2.000	6.265	**0.020**
C(GNN)	0.185	1.000	1.258	0.291
C(DEP)	1.226	1.000	8.330	**0.018**
C(ENC)	0.308	1.000	2.094	0.182
C(src):C(DEP)	0.165	2.000	0.56	0.590
C(src):C(ENC)	7.124	2.000	24.20	**0.000**
C(src):C(GNN)	1.335	2.000	4.534	**0.043**
C(DEP):C(GNN)	0.055	1.000	0.371	0.557
C(ENC):C(GNN)	1.045	1.000	7.098	**0.026**
C(DEP):C(ENC)	1.005	1.000	6.827	**0.028**
Residual	1.325	9.000	NaN	NaN

Table 9: Indore In-Domain ANOVA Results

Source	sum_sq	df	F	P(>F)
C(src)	48.606	2.000	2.449	0.108
C(GNN)	4.009	1.000	0.404	0.531
C(DEP)	23.301	1.000	2.348	0.139
C(ENC)	20.426	1.000	2.058	0.164
C(tgt)	199.051	2.000	10.030	**0.001**
C(tgt):C(DEP)	13.604	2.000	0.686	0.513
C(tgt):C(ENC)	85.332	2.000	4.300	**0.025**
C(tgt):C(GNN)	19.710	2.000	0.993	0.385
C(tgt):C(src)	12.388	4.000	0.312	0.735
C(src):C(DEP)	6.487	2.000	0.327	0.724
C(src):C(ENC)	73.878	2.000	3.723	**0.039**
C(src):C(GNN)	7.459	2.000	0.376	0.691
C(DEP):C(GNN)	0.845	1.000	0.085	0.773
C(ENC):C(GNN)	0.923	1.000	0.093	0.763
C(DEP):C(ENC)	1.561	1.000	0.157	0.695
Residual	238.143	24.000	NaN	NaN

Table 10: Indore Cross-Domain ANOVA Results

Source	sum_sq	df	F	P(>F)
C(src)	1.862	4.000	0.408	0.800
C(GNN)	0.719	1.000	0.630	0.438
C(DEP)	3.613	1.000	3.167	0.093
C(ENC)	51.586	1.000	45.228	**0.000**
C(src):C(DEP)	2.027	4.000	0.444	0.775
C(src):C(ENC)	9.053	4.000	1.984	0.143
C(src):C(GNN)	3.373	4.000	0.739	0.578
C(DEP):C(GNN)	0.221	1.000	0.194	0.665
C(ENC):C(GNN)	1.773	1.000	1.555	0.229
C(DEP):C(ENC)	1.601	1.000	1.403	0.252
Residual	19.390	17.000	NaN	NaN

Table 11: RedFM In-domain ANOVA Results

Source	sum_sq	df	F	P(>F)
C(src)	14.700	4.000	0.988	0.322
C(GNN)	0.109	1.000	0.029	0.864
C(DEP)	1.111	1.000	0.299	0.585
C(ENC)	4.923	1.000	1.323	0.252
C(tgt)	10.040	6.000	0.450	0.718
C(tgt):C(DEP)	25.753	6.000	1.154	0.334
C(tgt):C(ENC)	106.197	6.000	4.757	**0.000**
C(tgt):C(GNN)	1.642	6.000	0.074	0.998
C(tgt):C(src)	314.185	24.000	3.518	**0.000**
C(src):C(DEP)	23.724	4.000	1.594	0.178
C(src):C(ENC)	323.737	4.000	21.752	**0.000**
C(src):C(GNN)	49.322	4.000	3.314	**0.012**
C(DEP):C(GNN)	0.615	1.000	0.165	0.685
C(ENC):C(GNN)	2.771	1.000	0.745	0.389
C(DEP):C(ENC)	0.389	1.000	0.105	0.747
Residual	647.408	174.000	NaN	NaN

Table 12: RedFM Cross-Domain ANOVA Results

Source	sum_sq	df	F	P(>F)
C(src)	58.4	2	0.657	5.26E-01
C(DEP)	2.2	1	0.048	8.28E-01
C(LLM)	1260.3	2	14.17	**6.18E-05**
C(PRM)	3042.5	2	34.22	**3.94E-08**
C(src):C(DEP)	16.7	2	0.187	8.30E-01
C(src):C(LLM)	543.7	4	3.058	**3.36E-02**
C(src):C(PRM)	426.9	4	2.401	7.46E-02
C(DEP):C(LLM)	62.3	2	0.708	5.05E-01
C(DEP):C(PRM)	48.0	2	0.54	5.87E-01
C(LLM):C(PRM)	2205.3	4	12.40	**7.47E-06**
Residual	1200.1	27	NaN	NaN

Table 13: Indore Zero-shot ICL ANOVA Results

Source	sum_sq	df	F	P(>F)
C(src)	6123.02	6	13.34	**2.91E-10**
C(DEP)	5.09	1	0.07	7.97E-01
C(LLM)	4945.81	2	32.32	**6.97E-11**
C(PRM)	12473.39	2	81.51	**1.23E-19**
C(src):C(DEP)	178.97	6	0.39	8.83E-01
C(src):C(LLM)	13819.12	12	15.05	**1.46E-15**
C(src):C(PRM)	1727.37	12	1.88	5.01E-02
C(DEP):C(LLM)	131.03	2	0.86	4.29E-01
C(DEP):C(PRM)	101.88	2	0.67	5.17E-01
C(LLM):C(PRM)	3130.31	4	10.23	**1.12E-06**
Residual	5815.44	76	NaN	NaN

Table 14: RedFM Zero-shot ICL ANOVA Results