DeDisCo at the DISRPT 2025 Shared Task: A System for Discourse Relation Classification

Zhuoxuan Ju, Jingni Wu, Abhishek Purushothama, Amir Zeldes
Corpling Lab
Georgetown University

GEORGETOWN UNIVERSITY

C@GU

Code Demo

Introduction

The DISRPT 2025 Discourse Relation Classification task aims to classify both explicit and implicit discourse relations, encompassing a wide variety of data from 16 different languages and 6 distinct annotation frameworks.

We tested two architectures – one based on an encoder model and one based on a decoder.

Our final submission, DeDisCo (Decoder-based Discourse Cognoscente), is a decoder-only system, fine-tuned with instruction prompts for multilingual discourse relation classification.

Approach

Features:

- **Language Corpus Framework (LCF):** Incorporate metadata into the input.
- **DisCoDisCo Features (Gessler et al., 2021):** Selectively incorporate a subset of potent, hand-crafted discourse features from the DisCoDisCo system to improve classification accuracy.
- **Direction:** Encode the explicit relation direction (e.g., 1 > 2), provided directly by a dedicated column in the dataset, as a key input feature.
- **Context:** Expand the model's input to include the surrounding text of the two segments, providing it with essential context information.

Data Augmentation:

Targeted data augmentation was used to enhance performance for six languages with limited training data and low accuracy: Czech (ces), Dutch (nld), French (fra), Basque (eus), German (deu), and Persian (fas), covering a total of seven datasets.

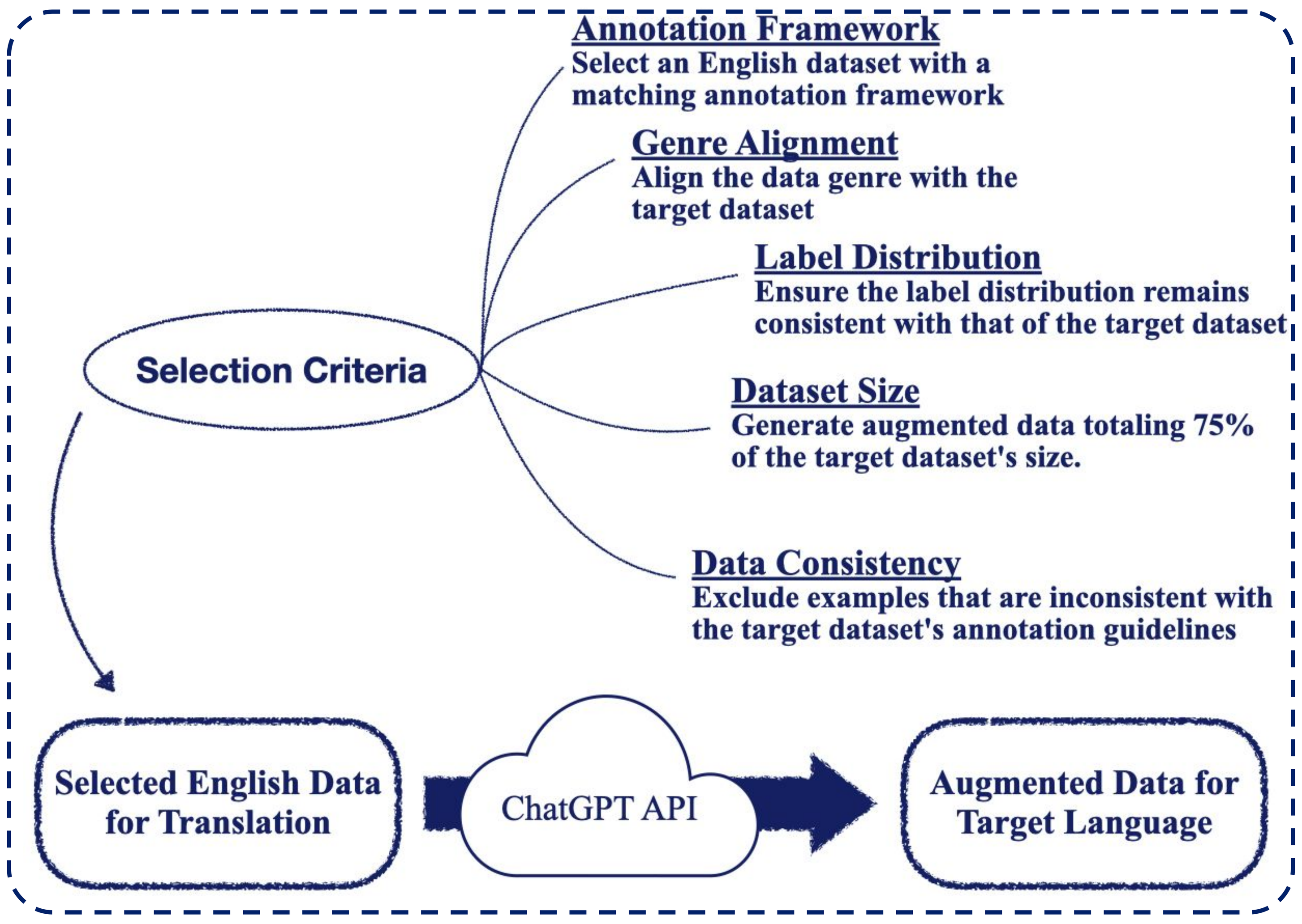

Figure 1: Illustration of Data Augmentation Strategy.

Decoder (DeDisCo)

We selected Qwen3-4B model, instructing it to select the correct label from a predefined set provided in the prompt.

Pruning: We pruned the Qwen3-4B model via layer removal to meet the task's parameter limit.

Supervised Fine-Tuning: We performed full-parameter supervised fine-tuning on the model, using the task-specific dataset reformatted into instruction prompts.

Prompt Design

```
## Role and Goal:
You are an expert in discourse analysis,
tasked with identifying the discourse
relation between two sentence units based
on the provided label. Your goal is to
accurately determine the relationship
between these two units.

## Guidelines:
1. You will receive Unit1 and Unit2. Unit1
appears before Unit2 in the original text.
2. You will also be informed about the
language of these units.
3. You will also be informed of the corpus
from which the data is drawn, which may
help guide your analysis.
...

## Labels:
contrast, condition, mode…

## Language:
eng

## Corpus:
gum
...
```

Figure 2: Illustration of the Prompt used in Qwen3-4B experiments,see paper for full prompt.

mT5 Encoder

We selected the mT5-XL variant, comprising 3.7 billion parameters. We used only the encoder and added a classification head.

Feature Injection via Input Tokens:

- We prepended LCF features as special input tokens, so that the model can incorporate this metadata directly in its tokenized input sequence.
- We encoded DisCoDisCo feature set, which capture properties such as whether the units are full sentences, whether the relation is discontinuous, and whether the two units share the same speaker, as explicit key-value tokens.
- We implemented pseudo-directional features from DisCoDisCo. These directional cues are lightweight but informative, and help disambiguate argument structure across instances.

The resulting input sequence is organized as follows: metadata (LCF features), followed by categorical DisCoDisCo features, and finally the target argument span:

```
LANG_eng FW_erst CORP_gum [SEP]
IS_SENTENCE_1 DISCONTINUOUS_0
SAME_SPEAKER_1 GENRE_academic [SEP] }
Aesthetic Appreciation and Spanish
Art: > Arg2: In this study we used
eye-tracking in the first stage
```

Results

Corpus	Decoder (DeDisCo)	Encoder
<u>ces.rst.crdt</u>	52.70	51.35
<u>deu.pdtb.pcc</u>	67.01	56.19
<u>deu.rst.pcc</u>	67.03	49.82
eng.dep.covdtb	68.21	**73.05**
eng.dep.scidtb	83.66	79.58
eng.erst.gentle	67.08	61.29
eng.erst.gum	73.45	62.98
eng.pdtb.gentle	67.94	61.07
eng.pdtb.gum	71.39	65.20
eng.pdtb.pdtb	83.77	77.32
eng.pdtb.tedm	71.79	61.54
eng.rst.oll	62.73	51.66
eng.rst.rstdt	73.27	62.60
eng.rst.sts	58.54	49.39
eng.rst.umuc	67.36	59.09
eng.sdrt.msdc	90.00	84.11
eng.sdrt.stac	75.80	65.96
<u>eus.rst.ert</u>	54.64	**55.67**
<u>fas.rst.prstc</u>	60.47	57.77
<u>fra.sdrt.annodis</u>	60.39	52.82
ita.pdtb.luna	70.13	66.13
<u>nld.rst.nldt</u>	68.62	53.85
pcm.pdtb.disconaija	59.39	41.40
pol.iso.pdc	74.02	55.05
por.pdtb.crpc	79.17	75.64
por.pdtb.tedm	68.41	64.84
por.rst.cstn	70.22	69.85
rus.rst.rrt	74.85	68.95
spa.rst.rststb	69.72	64.55
spa.rst.sctb	83.02	76.73
tha.pdtb.tdtb	96.73	**96.80**
tur.pdtb.tdb	64.13	**65.08**
tur.pdtb.tedm	59.23	54.55
zho.dep.scidtb	80.00	68.37
zho.pdtb.cdtb	88.65	86.54
zho.pdtb.ted	75.49	66.24
zho.rst.gcdt	75.55	65.37
zho.rst.sctb	74.21	66.67
Macro Average	71.28	64.34
Micro Average	76.13	69.74

Table 1: Accuracy on 38 test corpora (bold: encoder > decoder; underline: data augmentation applied).

Error Analysis

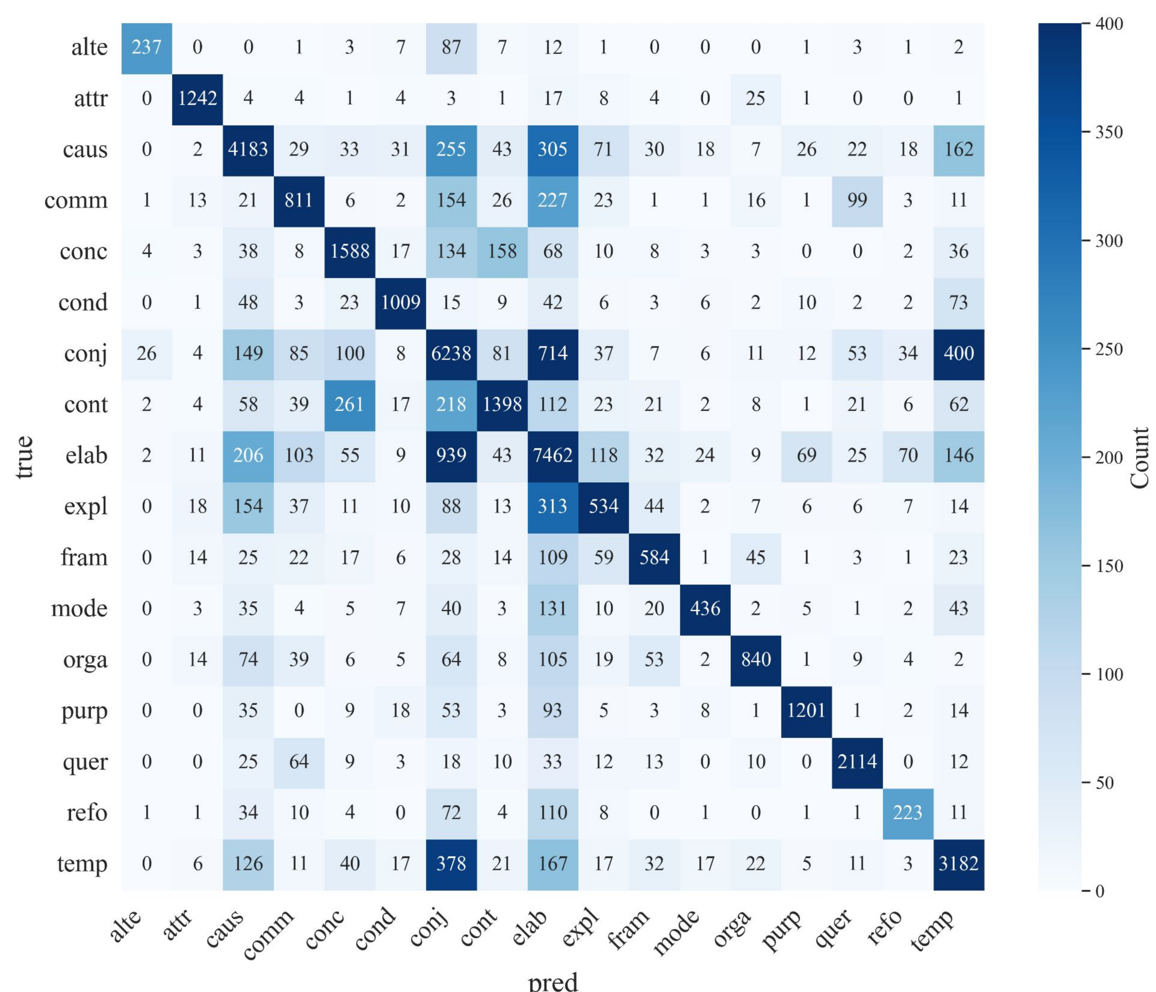

Figure 3: Confusion matrix over the entire dataset.

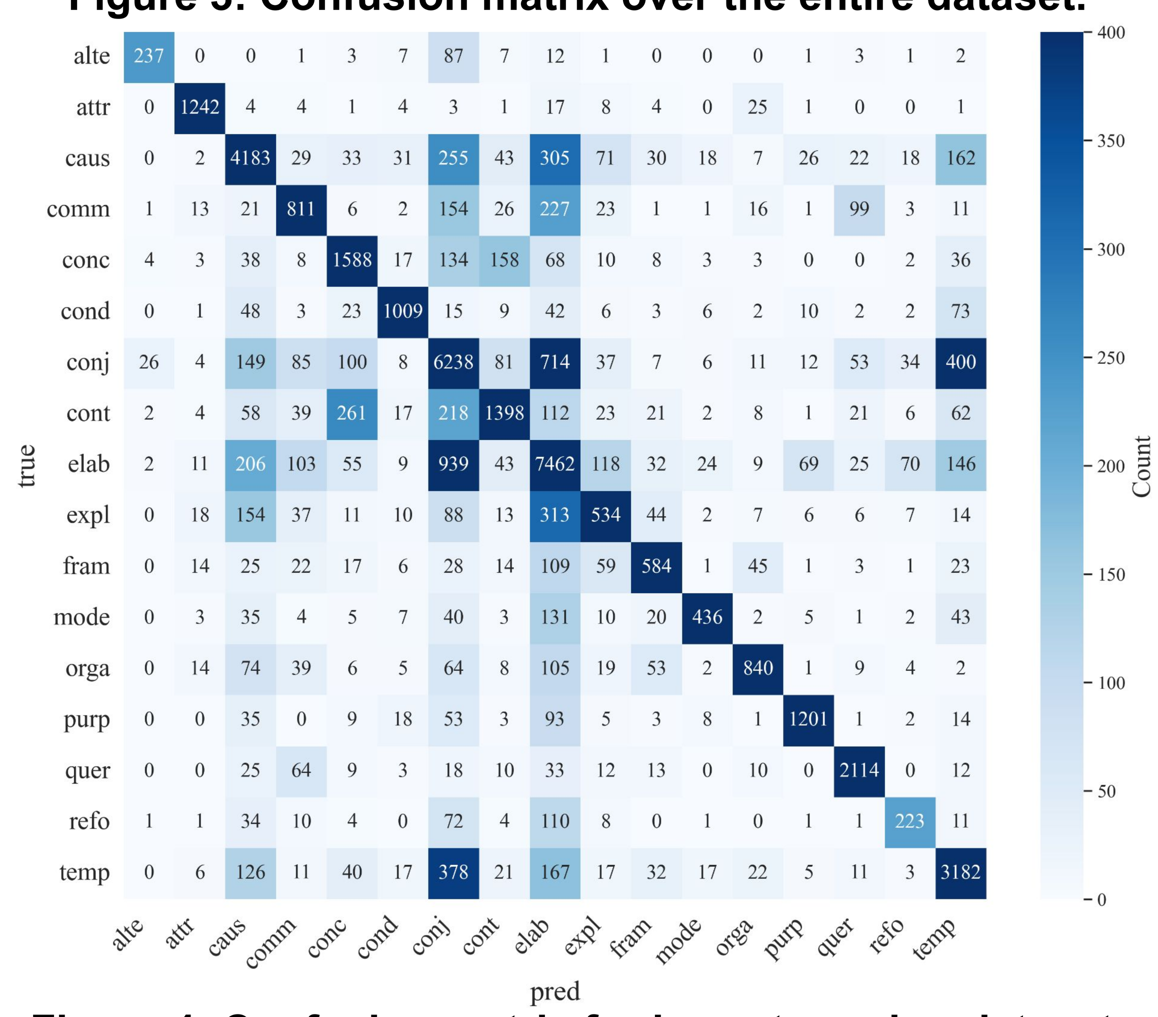

Figure 4: Confusion matrix for lowest-scoring dataset, ces.rst.crdt.