Adapting Biomedical Relation Extraction to Turkish and Azerbaijani

Team Member:
Matin Huseynzade 280201086

Advisor: Dr. Emrah Inan

A Thesis Report Submitted to the
Faculty of Engineering in Partial Fulfilment of the
Requirements for the Degree of

BACHELOR OF SCIENCE

(Academic Project)

Department of Computer Engineering
İzmir Institute of Technology
İzmir, Turkey

June, 2025

Abstract

This thesis addresses the challenge of adapting biomedical relation extraction systems to low-resource languages by developing a translation-aware pipeline tailored for Turkish and Azerbaijani. Building on an English-based dataset, we use AI agents equipped with prompt-based instructions to automate the translation process while preserving named entity annotations.

A core aspect of our method involves guiding language models with few-shot examples, significantly improving translation quality—particularly for complex, tagged biomedical text. The translated outputs are manually evaluated using a 5-point grading scale to assess syntactic and semantic fidelity.

We then fine-tune a relation extraction model on both Turkish and Azerbaijani translations, comparing the performance under zero-shot and few-shot configurations. Experimental results confirm that few-shot prompting leads to substantial gains in accuracy and generalization.

This work demonstrates that with careful design of prompts and minimal in-language examples, reliable biomedical NLP systems can be developed for underrepresented languages.

Keywords: Biomedical NLP, Machine Translation, Turkish, Azerbaijani, Named Entity Recognition, Few-Shot Learning, In-Context Learning

Contents

1. INTRODUCTION

The recent success of large language models has created new opportunities for low-resource language processing, yet biomedical natural language processing (BioNLP) remains heavily underrepresented in many languages such as Turkish and Azerbaijani. These languages often lack annotated corpora, translation infrastructure, and general-purpose models fine-tuned on domain-specific data. This scarcity poses a major barrier for relation extraction tasks that rely on syntactic consistency and accurate preservation of named entities such as chemicals and diseases.

This project addresses this challenge by building a translation-aware biomedical relation extraction pipeline tailored for Turkish and Azerbaijani. Rather than manually annotating text from scratch, we start from an English biomedical dataset annotated with chemical-disease relations and apply automatic translation strategies to adapt it to the target languages.

Our methodology combines the use of advanced AI agents for translation with prompt-based control mechanisms to ensure correct handling of biomedical entity tags. A particular emphasis is placed on evaluating the effectiveness of few-shot learning—providing the model with just a handful of examples during inference—to improve translation fidelity and downstream task accuracy. This approach is compared against zero-shot learning to quantify the performance gains attributable to in-context learning.

The final system is capable of:

- Translating biomedical data from English into Turkish and Azerbaijani with entity-preserving consistency.
- Scoring and visualizing translation quality using a manual evaluation interface.
- Fine-tuning and testing biomedical relation extraction models under few-shot and zero-shot settings.

This thesis demonstrates how combining prompt design, AI-assisted translation, and minimal supervision can yield scalable NLP systems for underrepresented languages, offering a generalizable methodology for biomedical domain adaptation.

2. RELATED WORK

Recent advancements in machine translation have focused on expanding support for low-resource languages. The *No Language Left Behind* (NLLB) project introduced a universal translation model capable of handling over 200 languages, emphasizing improvements in translation quality for underrepresented languages [2]. Complementing this, the CCMatrix initiative mined billions of high-quality parallel sentences across 38 languages, providing a substantial resource for training multilingual models [1].

In the realm of biomedical natural language processing, the BioCreative V Chemical-Disease Relation (CDR) corpus has been pivotal. It offers a comprehensive dataset of annotated chemical and disease entities, facilitating research in relation extraction tasks [5].

Furthermore, the concept of in-context learning has gained traction, with studies demonstrating that providing models with multiple examples (many-shot learning) significantly enhances performance across various tasks, including low-resource machine translation [4].

3. RESEARCH METHODOLOGY

This chapter outlines the methodology used to translate, annotate, and evaluate biomedical text for relation extraction in Turkish and Azerbaijani. Our approach is designed for low-resource languages and leverages AI agents for translation, prompt-based control for tag alignment, and contrastive evaluation using few-shot learning.

3.1 Automated Translation with AI Agents

Given the scarcity of biomedical corpora in Turkish and Azerbaijani, we developed a pipeline to translate large English biomedical datasets into these target languages. Instead of relying on manual translation alone, we employed advanced AI agents—language models guided via structured prompts—to automate the translation process.

These agents received not only the English sentences but also few translation examples from our existing dataset to learn stylistic and semantic preferences. This strategy ensured greater consistency and fluency in translated output, particularly around complex biomedical terminology.

3.2 Prompt-Based Tag Preservation

Named entities such as chemical and disease names must be preserved with high fidelity during translation. To address this, we created a tag translation prompt system that explicitly instructed the AI agents how to handle entity spans. Prompts included:

- Highlighted entity regions in the input sentence.
- Sample translations where tags were successfully aligned.
- Instructions to retain entity boundaries while allowing grammatical restructuring.

This ensured high-quality entity alignment, which is critical for maintaining label consistency in downstream relation extraction tasks.

3.3 Translation Quality Evaluation

The translated sentences were manually graded using a 5-point rubric:

- **0**: Semantically incorrect translation.
- **1**: Major syntactic errors.
- **2**: Minor syntactic issues.
- **3**: Correct translation.
- **4**: Not applicable or irrelevant case.

The resulting quality scores were recorded in a CSV file for Turkish and Azerbaijani translations. These grades were later used to interpret the correlation between translation fidelity and downstream model performance.

3.4 Few-Shot vs. Zero-Shot Learning

To train relation extraction models on the translated data, we tested both zero-shot and few-shot configurations. In the zero-shot setting, the model received no in-language examples and was prompted solely with English context. In the few-shot setting, we provided several Turkish or Azerbaijani examples as in-context demonstrations.

Results showed that few-shot learning significantly outperformed zero-shot setups, especially in Azerbaijani. This supports the claim that even limited annotated examples in the target language can drastically improve generalization for low-resource biomedical NLP.

3.5 Pipeline Summary

Our full methodology consists of:

1. Automatic translation of English biomedical data using AI agents with prompt guidance.
2. Tag-aware alignment strategies to preserve named entity boundaries.
3. Manual quality grading to assess translation accuracy.
4. Evaluation of downstream performance using both few-shot and zero-shot learning.

The complete system pipeline is illustrated in Figure 4.3, while the logic behind tag-preserving prompts is visualized in Figure 4.4.

4. EXPERIMENTAL STUDY

4.1 Dataset Preparation

The original English dataset used in this study was derived from biomedical relation extraction tasks focusing on chemical-disease interactions. These instances were first translated into Turkish and later into Azerbaijani. The translations aimed to preserve the biomedical semantics while adapting to the syntax and structure of the target languages.

Figure 4.1 illustrates example translations from the dataset along with their manually assigned quality grades.

Example Translation and Evaluation:

Original: Three patients received high doses of chlormethiazole for alcohol withdrawal symptoms, and one took a suicidal overdose of nitrazepam.

Turkish: Üç hasta alkol yoksunluk semptomları için yüksek dozda klormetiazol aldı ve biri intihar amaçlı nitrazepam aşırı dozu aldı.
Grade: 2

Azerbaijani: Üç xəstə spirtin tərkibindən imtina simptomları üçün yüksək dozada xlorometiazol aldı, bir xəstə isə nitrazepamın intihar məqsədli dozada artıq qəbul etdi.
Grade: 1

Figure 4.1: Example original sentences and their Turkish and Azerbaijani translations with quality scores.

Translations were manually scored using a 5-point rubric:

- 0 = Semantically wrong
- 1 = Major syntactic errors
- 2 = Minor syntactic errors
- 3 = Correct translation
- 4 = Not applicable

4.2 Evaluation Methodology

To evaluate the translation quality, a Python-based tool was implemented to load parallel translations, allow manual evaluation, and output the scores in CSV format. The resulting dataset includes original English text, its Azerbaijani and Turkish translations, and their respective quality grades.

In total, **50** samples were evaluated.

4.3 Translation Quality Distribution

Figure 4.2 shows the distribution of translation grades for both Azerbaijani and Turkish. As seen in the chart, Turkish translations performed more consistently at the highest quality level (grade 3), while Azerbaijani translations had a more diverse spread including minor and major errors.

Figure 4.2: Distribution of manual translation quality grades for Azerbaijani and Turkish.

4.4 Workflow and Tagging System

The translation process was accompanied by tag alignment mechanisms to ensure entity-level consistency in downstream NLP tasks. The overall workflow of the system—from English sentence parsing to model evaluation—is illustrated in Figure 4.3.

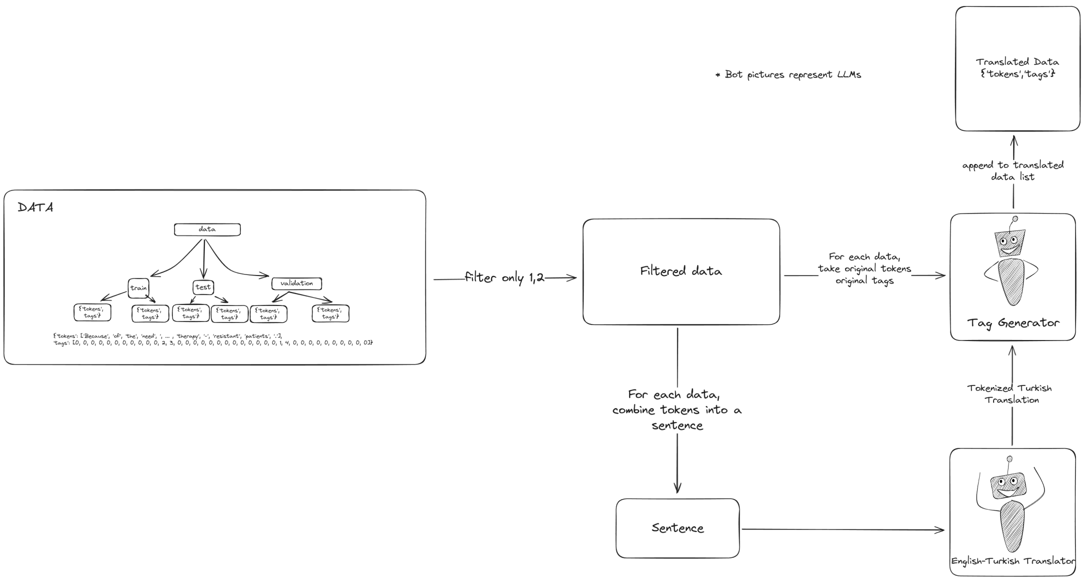

Figure 4.3: System workflow diagram illustrating the translation, tagging, and evaluation pipeline.

4.5 Tag Translator Design

Figure 4.4 demonstrates the design logic behind the tag translator. The system ensures that biomedical named entities are preserved accurately through translation, especially when using few-shot prompting strategies.

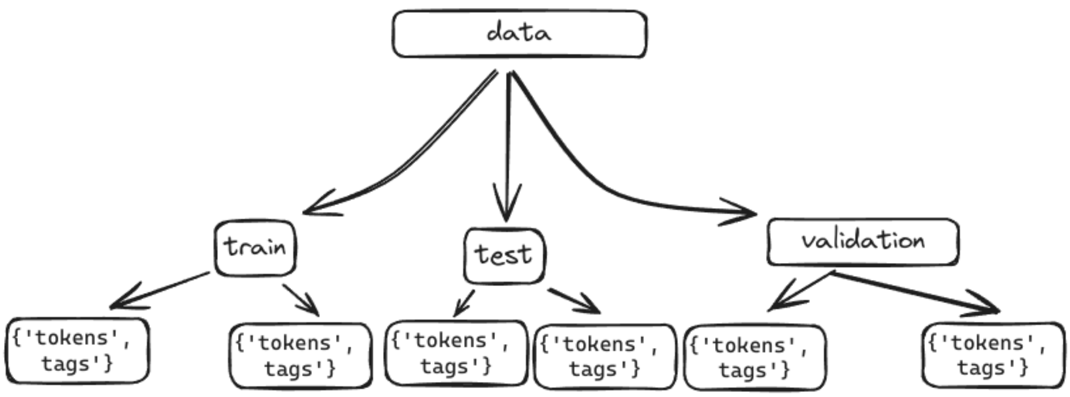

Figure 4.4: Tag translation logic flow for aligning biomedical entity tags between original and translated data.

5. CONCLUSIONS AND FUTURE WORK

This study presents a translation-aware adaptation of biomedical relation extraction systems for low-resource languages, specifically Turkish and Azerbaijani. By utilizing AI agents, structured prompts, and few-shot learning strategies, we demonstrated that high-quality, tag-preserving translations can be automatically generated and effectively leveraged for downstream NLP tasks.

Our experiments indicate that few-shot prompting significantly improves translation quality and downstream performance over zero-shot approaches. Even a small number of in-context examples within the prompt enable AI models to generalize structural patterns of tagged biomedical data, particularly for languages lacking large annotated corpora. This confirms that minimal supervised intervention can yield strong improvements in both linguistic fidelity and extraction accuracy.

Future Work

There are several promising directions for extending this work:

- **Many-Shot Prompting:** Expanding prompt context with additional translated samples or diverse examples from external biomedical corpora may enhance model reliability and semantic coverage, especially in handling complex syntactic structures or rare entities.
- **Cross-Language Transfer Learning:** Incorporating multilingual embeddings or fine-tuning a single model on both Turkish and Azerbaijani may reveal shared representational structures and further improve generalization.
- **Human-in-the-Loop Corrections:** Introducing minimal human edits on AI-translated output could create feedback loops for adaptive fine-tuning, improving translation robustness across unseen datasets.
- **Extending to Other Domains:** This approach could be adapted to domains beyond biomedical NLP, such as legal or environmental corpora, where structured and tagged data are similarly scarce in low-resource languages.

In summary, this work lays a foundation for combining translation automation with prompt-based learning to bridge the gap in biomedical NLP for underrepresented languages. Through continued refinement of prompts, data sources, and learning strategies, the quality and applicability of these systems can be substantially extended.

REFERENCES

[1] Holger Schwenk, Guillaume Wenzek, Sergey Edunov, Edouard Grave, and Armand Joulin, *CCMatrix: Mining Billions of High-Quality Parallel Sentences on the Web*, arXiv preprint arXiv:1911.04944, 2019.

[2] NLLB Team, Marta R. Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, Anna Sun, Skyler Wang, Guillaume Wenzek, Al Youngblood, Bapi Akula, Loïc Barrault, Gabriel Mejia Gonzalez, Prangthip Hansanti, John Hoffman, Semarley Jarrett, Kaushik Ram Sadagopan, Dirk Rowe, Shannon Spruit, Chau Tran, Pierre Andrews, Necip Fazil Ayan, Shruti Bhosale, Angela Fan, and Francisco Guzmán, *No Language Left Behind: Scaling Human-Centered Machine Translation*, arXiv preprint arXiv:2207.04672, 2022.

[3] Myle Ott, Sergey Edunov, Alexei Baevski, Angela Fan, Sam Gross, Nathan Ng, David Grangier, and Michael Auli, *fairseq: A Fast, Extensible Toolkit for Sequence Modeling*, arXiv preprint arXiv:1904.01038, 2019.

[4] Rishabh Agarwal, Avi Singh, Lei M. Zhang, Bernd Bohnet, Stephanie C.Y. Chan, Ankesh Anand, Zaheer Abbas, Azade Nova, John D. Co-Reyes, Eric Chu, Feryal Behbahani, Aleksandra Faust, and Hugo Larochelle, *Many-Shot In-Context Learning*, arXiv preprint arXiv:2404.11018, 2024.

[5] Jiao Li, Yueping Sun, Robert J. Johnson, Daniela Sciaky, Chih-Hsuan Wei, Robert Leaman, Allan Peter Davis, Carolyn J. Mattingly, Thomas C. Wiegers, and Zhiyong Lu, *BioCreative V CDR Task Corpus: A Resource for Chemical Disease Relation Extraction*, Database, 2016.

[6] Duygu Altinok, *A Diverse Set of Freely Available Linguistic Resources for Turkish*, Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 13739–13750, 2023.

[7] Turkish NLP Suite, *Vitamins and Supplements NER Dataset*, GitHub repository, 2023. Available at: `https://github.com/turkish-nlp-suite/Vitamins-Supplements-NER-dataset`

[8] Hugging Face, *Token Classification with Hugging Face Transformers*, 2023. Available at: `https://huggingface.co/docs/transformers/tasks/token_classification`

[9] Dina Demner-Fushman, Sophia Ananiadou, Paul Thompson, and Brian Ondov (Eds.), *Proceedings of the First Workshop on Patient-Oriented Language Processing (CL4Health) @ LREC-COLING 2024*, ELRA and ICCL, Torino, Italia, 2024.

A. APPENDIX

A.1

Codebase and Script Descriptions

The project includes several Python scripts designed to support translation evaluation, dataset formatting, and visualization. All scripts are located in the project directory and can be run using Python 3.8+.

A.1.1

translation_eval.py This script provides the core functionality for evaluating translation quality. It loads a CSV file containing the original English sentence along with its Turkish and Azerbaijani translations, and allows for manual assignment of a quality grade.

- **Input:** CSV file with columns: 'original_tokens', 'az_tokens', 'tr_tokens', 'az_grade', 'tr_grade'
- **Output:** Updated CSV with completed manual grades
- **Usage:**

  ```
  python translation_eval.py
  ```

- Bar plots comparing grade distributions across languages
- Rendering of translation samples for review

A.1.2

tag_translator_system_prompt.txt This file contains the system prompt instructions used for tag-preserving translation. It can be passed to AI agents like GPT models to guide the translation process. Prompts instruct the model to retain biomedical entity tags.

A.1.3

translated_ds_az.json / translated_ds_tr.json These are the translated datasets in JSON format, used as input for model training or evaluation. Each entry includes the original sentence, its translation, and tagged entity spans.