Large Language Models as a Normalizer for Transliteration and Dialectal Translation

Md Mahfuz Ibn Alam[1], **Antonios Anastasopoulos**[1,2]
[1]Department of Computer Science, George Mason University, USA
[2]Archimedes/Athena RC, Greece
{malam21,antonis}@gmu.edu

Abstract

NLP models trained on standardized language data often struggle with non-standard variations. We assess various Large Language Models (LLMs) for transliteration and dialectal normalization. Tuning open-source LLMs with as little as 10,000 parallel examples using LoRA can achieve results comparable to or better than closed-source LLMs. We perform dialectal normalization experiments for twelve South Asian languages and dialectal translation experiments for six language continua worldwide. The dialectal normalization task can also be a preliminary step for the downstream dialectal translation task. Among the six languages used in dialectal translation, our approach enables Italian and Swiss German to surpass the baseline model by 21.5 and 25.8 BLEU points, respectively.[1]

1 Introduction

Language variation encompasses how language manifests across different regions, social groups, and individual speakers. One prominent form of this variation is dialects, distinct forms of a language spoken by particular groups, often defined by geographical or social boundaries. Dialects include vocabulary, pronunciation, grammar, and usage variations, reflecting the rich tapestry of human experience and cultural identity. Additionally, we encounter phenomena such as transliteration in language use, which involves converting text from one script to another while preserving its phonetic characteristics. Transliteration, relying on mapping the pronunciation of words (their sounds) from one language into the orthography of another, is common practice in contexts where languages with different writing systems interact (Ahmadi and Anastasopoulos, 2023).

Translating language varieties presents a unique and complex challenge for linguists and translators. Dialects, with their distinct vocabularies, pronunciations, and grammatical structures, reflect their speakers' cultural and regional identities. Capturing these nuances in translation requires a deep understanding of both the source and target languages and the cultural contexts from which they arise. In the case of transliteration, unlike a few languages where the transliterated script serves as a standard means of input (as seen in systems like Pinyin for Chinese), most languages lack universally established transliteration systems. When individuals use scripts other than the formal script of the language to write, they do not always adhere to a specific standard (Ryskina et al., 2020). Instead, they typically employ the informal script to offer a rough phonetic transcription of the intended word. This transcription can vary significantly from person to person due to various factors, including regional or dialectal variations in pronunciation, different transcription conventions, or individual idiosyncrasies.

In the evolution of language and speech technology (LST) for a given language, varieties and dialects that have more data are initially prioritized. This results in a disparity in technology usage among speakers of different dialects of the same language. For example, despite the extensive work done in English, only a few studies focus on dialects or varieties such as African-American Vernacular English compared to Mainstream American English (Blodgett et al., 2018). Historically, Roman and related scripts have enjoyed widespread support across various platforms and devices for digital content creation. Although native language keyboards in numerous languages are available, most users still prefer using the Roman keyboard due to its comfort and familiarity.

In this work, we try to address both of these shortcomings. We build models that can translate

[1]https://github.com/mahfuzibnalam/LLM-Normalizer-Dialectal-Transaltion

Proceedings of the Twelfth Workshop on NLP for Similar Languages, Varieties and Dialects, pages 39–67
January 19, 2025.

dialectal varieties through a normalization step. We also build models that will be greatly valued by users and involve the automatic transliteration normalization of Romanized input into the native orthography. In summary, our contributions are:

1. We demonstrate using LLMs for two NLP tasks: transliteration and dialectal normalization.
2. We show that with a small amount of data, one can easily adapt (through finetuning with low-rank adaptors) an open-source LLM to achieve higher performance in both tasks.
3. We demonstrate that incorporating a dialectal normalization step before translation enhances performance for downstream dialectal translation tasks.

2 Task Definitions and Datasets

2.1 Transliteration Normalization

The process of transliteration involves representing a word, phrase, or text in a different script or writing system in an intentional manner. Transliterations aim to show how the original word sounds in a different script so people who use that script can get an idea of how to say the word. For example, instead of writing the Bengali sentence "আমি তোমাকে ভালোবাসি" in Bengali script, we can transliterate it using the Roman script, resulting in "Ami tomake valobashi."

The transliteration normalization task is essentially the reverse of transliteration. In this task, given a sentence transliterated into an informal writing system, our goal is to convert it back to the original writing system of that language.

Dakshina Dataset For the transliteration normalization task, we use the Dakshina dataset (Roark et al., 2020) as the primary resource for testing and training. This dataset includes three data sources focused on transliteration: Native Script Wikipedia, Romanization Lexicon, and Romanized Wikipedia. The Romanized Wikipedia is most relevant to our work, providing romanizations of complete Wikipedia sentences. The dataset supports twelve South Asian languages: Bengali, Gujarati, Hindi, Kannada, Malayalam, Marathi, Punjabi, Sindhi, Sinhala, Tamil, Telugu, and Urdu. For each language, native speakers romanized 10,000 sentences. The instruction for the annotators was to transcribe the given sentences as they would naturally write them in the Latin script. For our experiments, we randomly divided the 10,000 sentences into training and testing sets using an 80-20 split.

Aksharantar Dataset We also use the Aksharantar dataset (Madhani et al., 2022) to conduct an ablation study for the transliteration normalization task. Aksharantar is the largest publicly available transliteration dataset for Indian languages, created by mining from monolingual and parallel corpora and human annotators' contributions. It contains 26 million transliteration word pairs for 21 Indic languages, making it 21 times larger than existing datasets. However, we do not use this dataset for training and testing because it only includes word-level transliteration pairs, whereas our work focuses on sentence-level transliteration.

2.2 Dialectal Normalization

A dialect is a specific form of a language unique to a particular region or social group. Dialectal normalization involves converting a dialectal variation of a sentence into its standard form within that language. For instance, the Alassio dialect sentence corresponding to the English sentence `"They stole the painting"` is `"I han rubbau u quaddru"`. In contrast, the standard Italian variant is `"Hanno rubato il quadro"`.

CODET We use the CODET dataset (Alam et al., 2024) for the dialectal translation task. CODET is a contrastive dialectal benchmark encompassing 891 different varieties from 12 different languages. In this work, we consider six languages that have a good amount of dialect coverage: Arabic (25 vernaculars), Bengali (5 varieties), Basque (39 varieties), Italian (439 varieties), Kurdish (4 varieties), and Swiss German (368 varieties). Even though the dataset covers a vast range of dialects, the number of sentences for each language is small and can only be used as a testing set. Only five dialects of Arabic have more than 10,000 sentences, and precisely, these are the ones for which we can create a training set.

3 Methods

3.1 Zero-shot Prompting

In NLP, zero-shot learning for a model involves categorizing objects or concepts without having seen examples of those categories or concepts during training. This promising technique enhances the utility of LLMs across various tasks. Zero-shot prompting means that the prompt used to interact with the model does not include examples or demonstrations. The zero-shot prompt directly instructs the model to perform a task without providing any additional examples to guide it.

3.2 LoRA-tuning

A significant paradigm in natural language processing involves large-scale pre-training on general domain data followed by further adaptation to specific tasks or domains. One adaptation method is full fine-tuning, which retrains all model parameters. However, this approach becomes less feasible with the rise of large billion-parameter models, as deploying independent instances of fine-tuned models with billions of parameters is prohibitively expensive.

Hu et al. (2021) introduced Low-Rank Adaptation (LoRA), which addresses this issue by freezing the pre-trained model weights and injecting trainable rank decomposition matrices into each layer of the Transformer architecture. This method significantly decreases the number of parameters that need to be trained for downstream tasks. Their research demonstrates that LoRA, when compared to fine-tuned GPT-3 175B with Adam, can reduce the number of trainable parameters by 10,000 times and the GPU memory requirement by three times. Additionally, LoRA performs on par with or better than traditional fine-tuning in model quality.

3.3 Evaluation Metrics

BLEU Bilingual Evaluation Understudy (Papineni et al., 2002) is a metric for comparing a candidate translation to one or more reference translations. It is quick and inexpensive to calculate, language-independent, and highly correlated with human evaluation.

SPBLEU This is a modified version of BLEU where both the candidate and reference texts are tokenized using a single language-agnostic and publicly available fixed SentencePiece subword model (Kudo and Richardson, 2018). Unlike BLEU, which operates on words determined by whitespace, SPBLEU calculates BLEU scores over sub-words.

WER Word Error Rate (WER) is calculated by dividing the number of errors by the total number of words. Errors include substitutions, insertions, and deletions in a sequence of recognized words. Substitutions happen when a word is replaced, insertions occur when an extra word is added, and deletions occur when a word is omitted from the transcript.

Hyper Parameters	
Sub-word Tokens	7500, 15000, 30000, 60000, 90000
Learning Rate	0.01, 0.001, 0.0001
Dropout	0.2, 0.36, 0.5
Encoder-Decoder Layers	4, 6, 8

Table 1: Hyper-parameter search space for tuning the Scratch model.

SPWER Similar to SPBLEU, SPWER is a modified version of WER where the calculation is performed over sub-words rather than words. A SentencePiece model is used to generate the sub-words.

4 Experimental Setup

4.1 Transliteration Normalization

Baseline We use the IndicXlit model (Madhani et al., 2022) as our baseline model. IndicXlit is a transformer-based multilingual transliteration normalization model with approximately 11 million parameters. It supports transliteration conversions between Roman and native scripts for 21 Indic languages. Madhani et al. (2022) use the Aksharantar dataset to train the model, the largest publicly available parallel corpus, containing 26 million word pairs across 20 Indic languages.

Scratch The Scratch model employs a sequence-to-sequence Transformer architecture (Vaswani et al., 2017). It takes transliterated text in Roman script as input to the encoder and produces text in the original script as output from the decoder. The model is trained similarly to Machine Translation, utilizing sub-word tokens during training. The encoder and decoder have separate vocabularies, with the source vocabulary consisting of English and the target vocabulary combining all twelve languages' scripts. To inform the model which script to translate from Roman, we prepend a language-specific token (e.g., $< bn >$) to the source sentence.

In our experiments, we set the model dimension to 256, attention heads to 4, and hidden dimension to 1024. We employ the Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.98$, and $\epsilon = 10^{-6}$. Training lasts for 50 epochs with a batch size of 128 and utilizes the GLEU activation function. We perform extensive hyperparameter tuning to optimize model performance. Table 1 illustrates the hyper-parameters used. Through experimentation, we determine that setting the sub-word tokens to 7500, learning rate to 0.001, dropout to 0.2, and using six layers for both encoder and decoder yields the best average performance across all languages.

LoRA-Tuning We rely on the implementation provided by Li et al. (2023) to perform LoRA-tuning on our open-sourced LLM models. We conduct LoRA-tuning on ten models, with five models having 7B parameters and the remaining five with 13B parameters. This allows us to investigate any potential performance discrepancies due to model size. These models are BactrianX 7B and 13B (Li et al., 2023), Bloomz 7B and MT0 13B (Muennighoff et al., 2022), Gemma 7B (Team et al., 2024), Mistral Instruct 7B (Jiang et al., 2023), Tower Instruct 7B (Alves et al., 2024), ALMA 13B (Xu et al., 2024), Aya 13B (Üstün et al., 2024), Llama2 Chat 13B (Touvron et al., 2023). Among these models, Aya 13B and MT0 13B are encoder-decoder models, while the rest are causal language models (decoder-only).

For LoRA-tuning, we incorporate training data from all twelve languages in a multilingual fashion. We train the model for two epochs with a 3×10^{-4} learning rate. LoRA's rank, alpha, and dropout are configured to 64, 16, and 0.05, respectively. Furthermore, we convert the loaded model into a mixed-8bit quantized model. Prompt used during LoRA-tuning and to perform inference:

`Transliteration Normalization:`
1: Given a phonetic transcription of a Bengali sentence into Roman script. Translate it to Bengali script. Show just the translation. Roman: Trimatrik gathane dimatrik pristho katake ched bole.
2: Given a phonetic transcription of a Hindi sentence into Roman script. Translate it to Devanagari script. Show just the translation. Roman: 1947 men Dara Singh Singapore aa gaye.

4.2 Dialectal Normalization

LoRA-Tuning We employ the same implementation and settings as described in subsection 4.1. However, in this scenario, only data from five Arabic dialects was sufficient for LoRA-tuning. Thus, we train the model multilingually using the combined data from these five dialects. Prompt examples:

`Dialectal Normalization:`
1: Given an Italian sentence from Alassio. Translate it to standard Italian. Show just the translation. Alassio: Quelle garçune i fumman tante sigarette.
2: Given a German sentence from Aarau. Translate it to standard German. Show just the translation. Aarau: Oh, sie ist nicht da, sie ist einkaufen gegangen.

4.3 Dialectal Translation

In this downstream task, our objective is to demonstrate the benefit of incorporating a normalization step before translation instead of directly translating the dialectal variation. We utilize the NLLB-200 3.3B model (NLLB Team et al., 2022) for translation. Following the approach outlined in (Alam et al., 2024), our baseline model does not incorporate the normalization step before translation. This baseline model is referred to as "Without Normalization" in our study.

4.4 Evaluation Metrics

For evaluation, we utilize four metrics. The BLEU score is calculated using the SacreBLEU library (Post, 2018). We compute the WER score using the JiWER Python package[2]. To calculate SPBLEU and SPWER, we tokenize the texts using the SentencePiece model from FLORES-200[3]. This model trains a single SentencePiece (SPM) model for all 200 languages, ensuring representation across a broad spectrum of languages. It employs a vocabulary size of 256,000 to adequately cover both low- and high-resource languages, with careful down-sampling and up-sampling to balance representation.

5 Results

5.1 Transliteration Normalization

Zero-Shot Table 3 showcases our zero-shot prompting analysis outcomes across ten publicly available LLMs and one proprietary LLM. This experiment was conducted exclusively in Bengali to gauge the performance of open-source LLMs against both the Baseline and Scratch models. As anticipated, the open-source LLMs yield subpar results, with BLEU scores consistently below nine across all instances. Particularly noteworthy is the superior performance of the GPT4 model within this framework, surpassing the Baseline model by

[2] https://pypi.org/project/jiwer/
[3] https://github.com/facebookresearch/flores/blob/main/flores200/README.md

	BN	GU	HI	KN	ML	MR	PU	SD	SI	TA	TE	UR	Average
Baseline	53.8	53.6	63.5	69.5	47.7	62.1	50.0	35.4	37.4	54.6	65.9	30.0	52.0
Scratch	54.7	69.7	65.2	57.8	44.4	57.7	59.1	**62.0**	**51.3**	51.0	51.8	65.1	57.5
BactrianX 7B	39.5	22.7	49.8	19.4	29.3	49.3	23.4	45.3	21.6	37.2	18.9	53.5	34.2
Bloomz 7B	42.0	47.6	58.6	31.0	26.3	45.1	44.6	45.1	19.6	29.2	31.6	55.4	39.7
Gemma 7B	62.8	**72.5**	72.0	63.0	52.9	62.5	**62.2**	60.4	51.1	57.9	58.2	70.7	**62.2**
Llama 7B	41.1	22.6	50.7	19.9	30.1	49.2	24.6	46.8	22.6	37.4	19.2	53.6	34.8
Mistral 7B	54.0	34.7	57.3	44.8	20.8	58.8	27.4	54.6	31.3	46.0	38.8	61.7	44.2
Tower 7B	48.0	26.4	54.9	23.4	38.1	56.2	28.4	53.2	27.7	43.3	23.3	59.7	40.2
ALMA 13B	46.7	26.3	54.0	23.1	37.0	55.9	27.8	50.7	26.1	41.0	22.4	58.3	39.1
Aya 13B	52.3	62.0	67.8	46.7	39.5	56.0	57.0	51.2	33.6	40.9	42.4	67.4	51.4
BactrianX 13B	45.9	25.5	53.4	22.5	36.9	53.9	26.9	50.1	25.8	41.0	22.3	57.4	38.5
Llama 13B	44.9	24.8	52.0	21.2	31.7	52.6	26.0	48.5	23.9	29.8	20.1	55.4	35.9
Llama 13B	46.0	25.4	51.5	21.9	35.2	54.0	26.8	49.9	25.2	40.4	22.3	57.9	38.0
MT0 13B	52.7	60.9	68.3	46.4	38.9	55.7	57.0	50.7	34.3	38.9	43.8	67.5	51.3
GPT4 Turbo	**67.0**	70.7	**77.6**	67.2	**53.6**	**70.7**	59.6	27.8	42.0	**60.0**	**68.3**	**77.3**	61.8

Table 2: LoRA-tuned performance of the open-sourced LLMs in BLEU ↑ metric. The performances of the open-sourced LLMs improved greatly compared to their zero-shot performance. Gemma 7B and GPT4 models outperform the Baseline model. Gemma 7B is the best-performing model.

	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓
Baseline	67.8	53.8	21.47	24.41
Scratch	66.2	54.7	22.08	23.94
BactrianX 7B	11.3	3.5	83.37	88.94
Bloomz 7B	1.4	0.3	153.50	166.54
Gemma 7B	17.6	7.0	77.00	77.38
Mistral 7B	7.4	2.5	128.40	130.31
Tower 7B	16.9	5.9	81.21	78.49
ALMA 13B	13.7	5.5	96.18	99.51
Aya 13B	18.3	8.3	83.16	94.31
BactrianX 13B	16.5	5.9	83.18	82.96
Llama2 13B	21.1	8.8	73.49	74.45
MT0 13B	6.5	2.1	114.16	121.17
GPT4 Turbo	**77.7**	**67.0**	**14.37**	**17.41**

Table 3: Zero-shot performance of the LLMs in Bengali transliteration normalization task. All open-sourced LLMs perform poorly. GPT4 is the only LLM to outperform the Baseline model.

14.2 BLEU points. However, owing to the proprietary nature of GPT4, it remains uncertain whether the model was exposed to the test set during training. In subsequent phases, we aim to explore strategies to improve the performance of both the Baseline and GPT4 models utilizing open-source alternatives.

LoRA-Tuning Tables 2, 7, 8, 9 show the results of the open-sourced LLM models after LoRA-tuning (Hu et al., 2021) using the training data for four evaluation metric. For space constraint, the results with the SPBLEU, WER, and SPWER metrics are in the Appendix A. In the case of BLEU, Table 2 we can see that the Gemma 7B model outperforms the Baseline model. It even outperforms the GPT4 model on average for all twelve languages. Individually, we see the Gemma 7B model perform better for languages like Gujarati, Punjabi, and Sindhi, probably because the GPT4 has not seen much data in those languages. Results are consistent across all metrics.

Ablation Study The data in Table 2 indicates that the average BLEU score is higher for the Scratch model than the Baseline model. This raises an intriguing question: Why is this happening? One plausible explanation could be attributed to the phenomenon of "word leakage" between the training and testing data of the Scratch model, both originating from the same source. By its nature, transliteration lacks a predefined structure, leaving the form of writing entirely to the author's discretion. Given that both the training and test sets stem from the same dataset, there exists a likelihood that certain transliterated words remain consistent across both sets.

Consequently, it is plausible that the Scratch and LoRA-tuned models may become accustomed to normalizing specific variations and struggle to generalize to alternative transliterated forms of the same word. To illustrate, consider the Bengali word সঙ্গীত, which can be transliterated in various ways; two commonly used forms are "songit" and "sangeet". Our hypothesis regarding the Scratch model posits that if the model encounters a particular variation during training and subsequently en-

	Original Dakshina					Modified Dakshina				
	Scratch	**Leakage**	**Baseline**	**Leakage**	**Gemma 7B**	**Scratch**	**Leakage**	**Baseline**	**Leakage**	**Gemma 7B**
Bengali	66.2	47.5	67.8	26.9	72.3	54.6	33.8	61.8	27.4	66.7
Gujrati	77.3	48.1	67.5	26.5	78.6	65.8	37.3	64.7	26.6	69.6
Hindi	66.8	51.5	67.4	20.6	73.3	56.0	43.9	61.5	21.3	70.1
Kannada	73.9	38.8	82.0	34.0	75.5	69.9	31.2	79.3	33.7	72.4
Malayalam	68.1	30.9	73.1	29.5	73.2	65.7	25.5	72.0	29.5	71.3
Marathi	66.9	44.4	74.3	26.6	70.9	61.3	35.3	71.0	26.8	67.4
Punjabi	65.9	49.8	60.2	18.7	68.1	59.5	41.3	56.0	19.4	63.1
Sindhi	66.8	54.4	45.7	-	64.4	58.9	48.4	43.7	-	57.0
Sinhala	68.4	50.1	60.2	-	66.5	65.4	43.2	60.1	-	64.1
Tamil	68.3	33.5	72.2	31.4	72.3	64.2	26.3	69.3	31.4	68.8
Telegu	68.3	37.5	80.1	33.1	72.3	63.2	28.0	77.8	34.0	69.3
Urdu	66.0	59.3	38.6	18.4	70.9	55.3	43.8	36.1	20.0	64.3
Average	68.6	45.5	65.8	26.6	71.5	61.7	36.5	62.8	27.0	67.0

Table 4: Ablation study for the high-performance of the Scratch model on the Dakshina test-set in SPBLEU ↑ metric. When the leakage decreases, the performance of the Scratch model also decreases drastically. Whereas the Gemma 7B model still outperforms the Baseline model.

counters the same variation in the test set, it would yield a higher score. Conversely, the score would likely be lower if, during inference, we encounter a different variation.

We introduce a novel metric termed "Leakage" to quantify the percentage of words from the test set present in the training set. As depicted in Table 4, on the left side, the Scratch model exhibits an average leakage of 45.48% for the Original Dakshina test set. In contrast, the Baseline model demonstrates an average leakage of 26.57%. We utilize the Aksharantar training data to ascertain the baseline model's leakage. To validate our hypothesis, we construct a new dataset derived from the Original test set, the Modified Dakshina test set. Leveraging the same Aksharantar training data, which lists several variations of each word, we replace any word appearing in the Dakshina test set with an alternative variation found in the Aksharantar dataset. For instance, if "songit" appears for the Bengali word সঙ্গীত in the test set, we substitute it with "sangeet" based on the Aksharantar dataset. In Table 4, on the right side, for the Modified Dakshina test set, we observe that the average leakage for the Scratch model decreases by 9%. However, the leakage for the Baseline model remains unchanged.

Now, let us examine the scores of three models for these two test sets. Notably, the SPBLEU score decreases by 9 points for the Scratch model, confirming our hypothesis that the model tended to replicate specific variations rather than generalize to different ones. Consequently, the Scratch model fails to surpass the baseline model's performance on this new test set. While a similar trend is evident for the Baseline and The Gemma 7B models, the disparity is less substantial than observed with the Scratch model. Furthermore, the Gemma 7B model consistently outperforms the Baseline model, underscoring the robust generalization ability of these open-source LLM models across various transliterated variations.

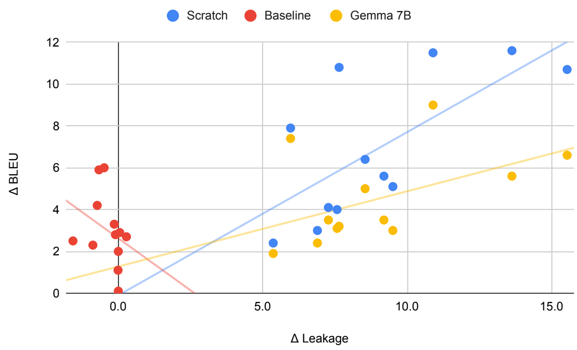

Figure 1: Correlation between Δ Leakage and Δ BLEU of the three models (Scratch, Baseline, and Gemma 7B).

Figure 1 shows the correlation between leakage and the models' performance (We calculate Δ Leakage and BLEU by subtracting scores from the Original Dakshina to the Modified Dakshina). Our hypothesis again gets verified by the trendline of the models. The Scratch model correlates higher with leakage than the Gemma 7B model. The Gemma 7B model has a higher generalizing ability for different variations than the Scratch model.

Arabic Variety		Zero Shot				LoRA Tuned		
		Gemma 7B	Aya 13B	MT0 13B	GPT4 Turbo	Gemma 7B	Aya 13B	MT0 13B
Cairo	SPBLEU↑	5.8	8.8	9.3	21.0	24.6	24.6	**25.0**
	BLEU↑	3.2	4.2	5.6	14.2	16.7	22.6	**23.4**
Tunis	SPBLEU↑	3.0	5.3	6.2	14.3	**21.6**	19.1	19.1
	BLEU↑	1.7	2.1	3.2	8.7	14.6	**17.9**	17.8
Rabat	SPBLEU↑	3.2	6.6	7.8	17.4	**23.4**	20.9	20.8
	BLEU↑	2.0	2.9	4.7	11.9	16.0	**19.5**	19.3
Beirut	SPBLEU↑	4.0	6.7	7.3	18.0	**24.0**	22.3	22.8
	BLEU↑	2.0	2.5	3.7	11.6	16.3	20.8	**21.5**
Doha	SPBLEU↑	7.8	9.5	10.3	19.6	**25.2**	24.3	24.7
	BLEU↑	3.4	4.4	5.9	13.1	17.0	22.7	**22.9**
Average	SPBLEU↑	4.8	7.5	8.2	18.1	**23.8**	22.2	22.5
	BLEU↑	2.5	3.2	4.6	11.9	16.1	20.7	**21.0**

Table 5: Zero-shot and LoRA-tuned performance of the open-sourced LLMs in Arabic normalization task. The LoRA-tuned models outperform the base models like before. In this task, the open-sourced models even outperform the GPT4 model.

5.2 Dialectal Normalization

Zero-shot and LoRA-tuned Among the six languages involved in the Dialectal normalization task, only five Arabic dialects possess sufficient data to enable LoRA-tuning of an open-source LLM. In light of this, for experiments within this setup, we solely consider three open-source LLMs, a decision informed by the outcomes of the previous task. Table 5 illustrates the results for these three open-source models. Analogous to the transliteration normalization task, the performance of the open-source models in zero-shot prompting scenarios proves subpar compared to GPT4. However, the LoRA-tuned variants perform superior to the GPT4 model across the five dialects.

Conversely, the remaining five languages need more training data to facilitate the LoRA-tuning of an open-source model. Consequently, to utilize normalization as a precursor to the downstream dialectal translation task, we will employ the best-performing zero-shot model, GPT4.

5.3 Dialectal Translation

Table 6 conveys the results of the downstream task for all six languages. We average the scores of the overall dialects of the language. As mentioned, we performed the normalization step using the LoRA-tuned MT0 model for Arabic. We did the normalization step for the other languages using the GPT4 model. The BLEU score, on average, for all six languages goes up by 9.56 points when we complete the normalization step beforehand. Apart from Kurdish, the BLEU score goes up for all five languages. The jump in quality for Italian and Swiss German is enormous, 21.55 and 25.79 BLEU points, respectively. We believe this is because of the vast amount of data available on the internet for these two languages, as GPT4 is likely being trained on data from all these varieties. For space constraint we show the performance of individual dialects of six languages in Tables 11, 12, 13, 14, 15, 16 of Appendix A.

Language	Without Normalizing (BLEU ↑)	With Normalizing (BLEU ↑)
Arabic*	37.90	**42.93**
Bengali	17.04	**20.06**
Basque	13.51	**16.24**
Italian	21.90	**43.45**
Swiss German	47.77	**73.56**
Kurdish	**9.35**	8.60
Average	24.58	**34.14**

Table 6: performance of the translation task with or without the normalization step. We had the data for Arabic to do LoRA-tuning on an open-sourced LLM for that language. For the other languages, we did the normalization using the GPT4 model in a zero-shot manner. The normalization step helps outperform the previous baseline (without normalization) model for all the languages except Kurdish.

6 Related Work

6.1 Dialectal

Most of the previous work on developing machine translation (MT) technologies for dialects and varieties has focused on Arabic (Zbib et al., 2012;

Harrat et al., 2019), Swiss German (Garner et al., 2014; Honnet et al., 2017), Kurdish (Ahmadi et al., 2022), Portuguese (Fancellu et al., 2014), and French (Garcia and Firat, 2022). One of the main challenges in this field is identifying potential translation sources and creating corpora and datasets for translating these dialects and varieties (Zampieri et al., 2020). Considering this, Alam et al. (2023) attempted to quantify dialectal translation disparities across as many languages as possible. Their study shows that general machine translation systems struggle to comprehend and accurately translate dialectal varieties. Building on their work, we propose a prior step of dialectal normalization before performing translation.

6.2 Transliteration

Several transliteration systems were recently proposed during the Named Entities Workshop evaluation campaigns in 2018 (Chen et al., 2018). These campaigns comprise transliterating tasks from English to other languages with various writing systems. The transliteration models typically mentioned in the literature include a combination of neural and non-neural models. Kundu et al. (2018); Le and Sadat (2018) used deep attention-based RNN encoder-decoder models and Merhav and Ash (2018); Roark et al. (2020); Moran and Lignos (2020) used neural transformer-based models. Kunchukuttan et al. (2021) use multilingual training to train their transliteration system. They recommend using single-script models to train separate models for two different language families. To our knowledge, we are the first ones to use LLMs for transliteration.

6.3 Using Large Language Models for Translation

Using LLMs for multilingual machine translation is garnering increasing attention. Lin et al. (2022) evaluate GPT-3 and XGLM-7.5B across 182 translation directions. Similarly, Bawden and Yvon (2023) assess BLOOM in 30 directions. Evaluations of ChatGPT by Bang et al. (2023); Jiao et al. (2023); Hendy et al. (2023) cover 6 to 18 directions. Zhu et al. (2023) comprehensively evaluates multilingual translation performance for popular LLMs in 102 languages and 606 directions, comparing them with state-of-the-art translation engines like NLLB and Google Translate. This extensive benchmark highlights the challenges in optimizing this emerging translation paradigm.

Significant efforts have focused on designing exemplar selection strategies to improve in-context learning (ICL) for machine translation. Agrawal et al. (2023); Zhang et al. (2023); Moslem et al. (2023) contribute to this area, with Zhang et al. (2023) finding that random selection can be a simple yet effective strategy. Wei et al. (2022) demonstrate that few-shot exemplars enhance translation performance. Moreover, Vilar et al. (2023) note that selecting ICL examples from a high-quality pool, such as a development set, is more beneficial, and (Zhang et al., 2023) analyze the importance of exemplar quality in translation outcomes. In this work, we do not use large language models (LLMs) to translate sentences directly. Instead, we employ LLMs as a preliminary step for normalization, which then facilitates further downstream translation tasks.

7 Conclusion

In this work, we show that it is possible to use the closed-sourced LLM for the new tasks: transliteration normalization and dialectal normalization, even if we do not have data for training. We also show that if we have a small quantity of data for training (ten thousand), we can LoRA-tune open-sourced LLMs to be on par or even better in performance than the closed-source ones. These open-sourced models are significantly smaller and cheaper to run than closed-source ones. Finally, one can use the dialectal normalization step as a prior step for the dialectal translation task.

Regarding the transliteration, we only use the Romanized Wikipedia data from the Dakshina dataset. We do not use other data sources like native script Wikipedia or the Romanization lexicon. The Aksharantar dataset also contains 26 million Romanization lexicon pairs for 21 Indic languages. In this work, we focused on sentence-level transliteration. In the future, we plan on using these vast data sources for model training.

Limitations

One limitation of our approach to dialectal normalization is the usage of a closed-sourced model like GPT4, which can be very expensive. As mentioned earlier, one way around this is to use open-sourced models for fine-tuning. However, this can not be done for dialects as very few training datasets exist. For our dialectal experiments, we spent around a thousand dollars.

Acknowledgements

The authors are thankful to the anonymous reviewers for their feedback. This work has been supported by the National Science Foundation under awards IIS-2125466 and IIS-2327143, and by the Microsoft Accelerate Foundation Models Research (AFMR) grant program. Computational resources for experiments were provided by the Office of Research Computing at George Mason University (URL: `https://orc.gmu.edu`) and funded in part by grants from the National Science Foundation (Awards Number 1625039 and 2018631). AA is also partially supported by project MIS 5154714 of the National Recovery and Resilience Plan Greece 2.0 funded by the European Union under the NextGenerationEU Program.

References

Sweta Agrawal, Chunting Zhou, Mike Lewis, Luke Zettlemoyer, and Marjan Ghazvininejad. 2023. In-context examples selection for machine translation. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 8857–8873, Toronto, Canada. Association for Computational Linguistics.

Sina Ahmadi and Antonios Anastasopoulos. 2023. Script normalization for unconventional writing of under-resourced languages in bilingual communities. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 14466–14487, Toronto, Canada. Association for Computational Linguistics.

Sina Ahmadi, Hossein Hassani, and Daban Q Jaff. 2022. Leveraging Multilingual News Websites for Building a Kurdish Parallel Corpus. *Transactions on Asian and Low-Resource Language Information Processing*, 21(5):1–11.

Md Mahfuz Ibn Alam, Sina Ahmadi, and Antonios Anastasopoulos. 2023. CODET: A Benchmark for Contrastive Dialectal Evaluation of Machine Translation. *arXiv preprint arXiv:2305.17267*.

Md Mahfuz Ibn Alam, Sina Ahmadi, and Antonios Anastasopoulos. 2024. CODET: A benchmark for contrastive dialectal evaluation of machine translation. In *Findings of the Association for Computational Linguistics: EACL 2024*, pages 1790–1859, St. Julian's, Malta. Association for Computational Linguistics.

Duarte M. Alves, José Pombal, Nuno M. Guerreiro, Pedro H. Martins, João Alves, Amin Farajian, Ben Peters, Ricardo Rei, Patrick Fernandes, Sweta Agrawal, Pierre Colombo, José G. C. de Souza, and André F. T. Martins. 2024. Tower: An open multilingual large language model for translation-related tasks. *Preprint*, arXiv:2402.17733.

Yejin Bang, Samuel Cahyawijaya, Nayeon Lee, Wenliang Dai, Dan Su, Bryan Wilie, Holy Lovenia, Ziwei Ji, Tiezheng Yu, Willy Chung, Quyet V. Do, Yan Xu, and Pascale Fung. 2023. A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity. 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 675–718, Nusa Dua, Bali. Association for Computational Linguistics.

Rachel Bawden and François Yvon. 2023. Investigating the translation performance of a large multilingual language model: the case of BLOOM. In *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*, pages 157–170, Tampere, Finland. European Association for Machine Translation.

Su Lin Blodgett, Johnny Wei, and Brendan O'Connor. 2018. Twitter Universal Dependency parsing for African-American and mainstream American English. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1415–1425, Melbourne, Australia. Association for Computational Linguistics.

Nancy Chen, Rafael E. Banchs, Min Zhang, Xiangyu Duan, and Haizhou Li. 2018. Report of NEWS 2018 named entity transliteration shared task. In *Proceedings of the Seventh Named Entities Workshop*, pages 55–73, Melbourne, Australia. Association for Computational Linguistics.

Federico Fancellu, Andy Way, and Morgan O'Brien. 2014. Standard language variety conversion for content localisation via SMT. In *Proceedings of the 17th Annual conference of the European Association for Machine Translation*, pages 143–149.

Xavier Garcia and Orhan Firat. 2022. Using natural language prompts for machine translation. *arXiv preprint arXiv:2202.11822*.

Philip N. Garner, David Imseng, and Thomas Meyer. 2014. Automatic speech recognition and translation of a Swiss German dialect: Walliserdeutsch. In *Proc. Interspeech 2014*, pages 2118–2122.

Salima Harrat, Karima Meftouh, and Kamel Smaili. 2019. Machine translation for Arabic dialects (survey). *Information Processing & Management*, 56(2):262–273.

Amr Hendy, Mohamed Abdelrehim, Amr Sharaf, Vikas Raunak, Mohamed Gabr, Hitokazu Matsushita, Young Jin Kim, Mohamed Afify, and Hany Hassan Awadalla. 2023. How good are gpt models at machine translation? a comprehensive evaluation. *Preprint*, arXiv:2302.09210.

Pierre-Edouard Honnet, Andrei Popescu-Belis, Claudiu Musat, and Michael Baeriswyl. 2017. Machine translation of low-resource spoken dialects: Strategies for normalizing Swiss German. *arXiv preprint arXiv:1710.11035*.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *Preprint*, arXiv:2106.09685.

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. *Preprint*, arXiv:2310.06825.

Wenxiang Jiao, Wenxuan Wang, Jen tse Huang, Xing Wang, Shuming Shi, and Zhaopeng Tu. 2023. Is chatgpt a good translator? yes with gpt-4 as the engine. *Preprint*, arXiv:2301.08745.

Taku Kudo and John Richardson. 2018. SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 66–71, Brussels, Belgium. Association for Computational Linguistics.

Anoop Kunchukuttan, Siddharth Jain, and Rahul Kejriwal. 2021. A large-scale evaluation of neural machine transliteration for Indic languages. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 3469–3475, Online. Association for Computational Linguistics.

Soumyadeep Kundu, Sayantan Paul, and Santanu Pal. 2018. A deep learning based approach to transliteration. In *Proceedings of the Seventh Named Entities Workshop*, pages 79–83, Melbourne, Australia. Association for Computational Linguistics.

Ngoc Tan Le and Fatiha Sadat. 2018. Low-resource machine transliteration using recurrent neural networks of Asian languages. In *Proceedings of the Seventh Named Entities Workshop*, pages 95–100, Melbourne, Australia. Association for Computational Linguistics.

Haonan Li, Fajri Koto, Minghao Wu, Alham Fikri Aji, and Timothy Baldwin. 2023. Bactrian-x : A multilingual replicable instruction-following model with low-rank adaptation. *Preprint*, arXiv:2305.15011.

Xi Victoria Lin, Todor Mihaylov, Mikel Artetxe, Tianlu Wang, Shuohui Chen, Daniel Simig, Myle Ott, Naman Goyal, Shruti Bhosale, Jingfei Du, Ramakanth Pasunuru, Sam Shleifer, Punit Singh Koura, Vishrav Chaudhary, Brian O'Horo, Jeff Wang, Luke Zettlemoyer, Zornitsa Kozareva, Mona Diab, Veselin Stoyanov, and Xian Li. 2022. Few-shot learning with multilingual generative language models. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 9019–9052, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Yash Madhani, Sushane Parthan, Priyanka A. Bedekar, Ruchi Khapra, Vivek Seshadri, Anoop Kunchukuttan, Pratyush Kumar, and Mitesh M. Khapra. 2022. Aksharantar: Towards building open transliteration tools for the next billion users. *ArXiv*, abs/2205.03018.

Yuval Merhav and Stephen Ash. 2018. Design challenges in named entity transliteration. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 630–640, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Molly Moran and Constantine Lignos. 2020. Effective architectures for low resource multilingual named entity transliteration. In *Proceedings of the 3rd Workshop on Technologies for MT of Low Resource Languages*, pages 79–86, Suzhou, China. Association for Computational Linguistics.

Yasmin Moslem, Rejwanul Haque, John D. Kelleher, and Andy Way. 2023. Adaptive machine translation with large language models. In *Proceedings of the 24th Annual Conference of the European Association for Machine Translation*, pages 227–237, Tampere, Finland. European Association for Machine Translation.

Niklas Muennighoff, Thomas Wang, Lintang Sutawika, Adam Roberts, Stella Biderman, Teven Le Scao, M Saiful Bari, Sheng Shen, Zheng-Xin Yong, Hailey Schoelkopf, et al. 2022. Crosslingual generalization through multitask finetuning. *arXiv preprint arXiv:2211.01786*.

Team 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, Loic 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, Sergey Edunov, Angela Fan, Cynthia Gao, Vedanuj Goswami, Francisco Guzmán, Philipp Koehn, Alexandre Mourachko, Christophe Ropers, Safiyyah Saleem, Holger Schwenk, and Jeff Wang. 2022. No language left behind: Scaling human-centered machine translation. *Preprint*, arXiv:2207.04672.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Matt Post. 2018. A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Brussels, Belgium. Association for Computational Linguistics.

Brian Roark, Lawrence Wolf-Sonkin, Christo Kirov, Sabrina J. Mielke, Cibu Johny, Isin Demirsahin, and Keith Hall. 2020. Processing South Asian languages written in the Latin script: the Dakshina dataset. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 2413–2423, Marseille, France. European Language Resources Association.

Maria Ryskina, Ella Rabinovich, Taylor Berg-Kirkpatrick, David Mortensen, and Yulia Tsvetkov. 2020. Where new words are born: Distributional semantic analysis of neologisms and their semantic neighborhoods. In *Proceedings of the Society for Computation in Linguistics 2020*, pages 367–376, New York, New York. Association for Computational Linguistics.

Gemma Team, Thomas Mesnard, Cassidy Hardin, Robert Dadashi, Surya Bhupatiraju, Shreya Pathak, Laurent Sifre, Morgane Rivière, Mihir Sanjay Kale, Juliette Love, Pouya Tafti, Léonard Hussenot, Pier Giuseppe Sessa, Aakanksha Chowdhery, Adam Roberts, Aditya Barua, Alex Botev, Alex Castro-Ros, Ambrose Slone, Amélie Héliou, Andrea Tacchetti, Anna Bulanova, Antonia Paterson, Beth Tsai, Bobak Shahriari, Charline Le Lan, Christopher A. Choquette-Choo, Clément Crepy, Daniel Cer, Daphne Ippolito, David Reid, Elena Buchatskaya, Eric Ni, Eric Noland, Geng Yan, George Tucker, George-Christian Muraru, Grigory Rozhdestvenskiy, Henryk Michalewski, Ian Tenney, Ivan Grishchenko, Jacob Austin, James Keeling, Jane Labanowski, Jean-Baptiste Lespiau, Jeff Stanway, Jenny Brennan, Jeremy Chen, Johan Ferret, Justin Chiu, Justin Mao-Jones, Katherine Lee, Kathy Yu, Katie Millican, Lars Lowe Sjoesund, Lisa Lee, Lucas Dixon, Machel Reid, Maciej Mikuła, Mateo Wirth, Michael Sharman, Nikolai Chinaev, Nithum Thain, Olivier Bachem, Oscar Chang, Oscar Wahltinez, Paige Bailey, Paul Michel, Petko Yotov, Rahma Chaabouni, Ramona Comanescu, Reena Jana, Rohan Anil, Ross McIlroy, Ruibo Liu, Ryan Mullins, Samuel L Smith, Sebastian Borgeaud, Sertan Girgin, Sholto Douglas, Shree Pandya, Siamak Shakeri, Soham De, Ted Klimenko, Tom Hennigan, Vlad Feinberg, Wojciech Stokowiec, Yu hui Chen, Zafarali Ahmed, Zhitao Gong, Tris Warkentin, Ludovic Peran, Minh Giang, Clément Farabet, Oriol Vinyals, Jeff Dean, Koray Kavukcuoglu, Demis Hassabis, Zoubin Ghahramani, Douglas Eck, Joelle Barral, Fernando Pereira, Eli Collins, Armand Joulin, Noah Fiedel, Evan Senter, Alek Andreev, and Kathleen Kenealy. 2024. Gemma: Open models based on gemini research and technology. *Preprint*, arXiv:2403.08295.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. 2023. Llama 2: Open foundation and fine-tuned chat models. *Preprint*, arXiv:2307.09288.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

David Vilar, Markus Freitag, Colin Cherry, Jiaming Luo, Viresh Ratnakar, and George Foster. 2023. Prompting palm for translation: Assessing strategies and performance. *Preprint*, arXiv:2211.09102.

Jason Wei, Maarten Bosma, Vincent Y. Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M. Dai, and Quoc V. Le. 2022. Finetuned language models are zero-shot learners. *Preprint*, arXiv:2109.01652.

Haoran Xu, Amr Sharaf, Yunmo Chen, Weiting Tan, Lingfeng Shen, Benjamin Van Durme, Kenton Murray, and Young Jin Kim. 2024. Contrastive preference optimization: Pushing the boundaries of llm performance in machine translation. *Preprint*, arXiv:2401.08417.

Marcos Zampieri, Preslav Nakov, and Yves Scherrer. 2020. Natural language processing for similar languages, varieties, and dialects: A survey. *Natural Language Engineering*, 26(6):595–612.

Rabih Zbib, Erika Malchiodi, Jacob Devlin, David Stallard, Spyros Matsoukas, Richard Schwartz, John Makhoul, Omar Zaidan, and Chris Callison-Burch. 2012. Machine translation of Arabic dialects. In *Proceedings of the 2012 conference of the north american chapter of the association for computational linguistics: Human language technologies*, pages 49–59.

Biao Zhang, Barry Haddow, and Alexandra Birch. 2023. Prompting large language model for machine translation: A case study. *Preprint*, arXiv:2301.07069.

Wenhao Zhu, Hongyi Liu, Qingxiu Dong, Jingjing Xu, Shujian Huang, Lingpeng Kong, Jiajun Chen, and Lei Li. 2023. Multilingual machine translation with large language models: Empirical results and analysis. *Preprint*, arXiv:2304.04675.

Ahmet Üstün, Viraat Aryabumi, Zheng-Xin Yong, Wei-Yin Ko, Daniel D'souza, Gbemileke Onilude, Neel Bhandari, Shivalika Singh, Hui-Lee Ooi, Amr Kayid, Freddie Vargus, Phil Blunsom, Shayne Longpre, Niklas Muennighoff, Marzieh Fadaee, Julia Kreutzer, and Sara Hooker. 2024. Aya model: An instruction finetuned open-access multilingual language model. *arXiv preprint arXiv:2402.07827*.

A All Results

	BN	GU	HI	KN	ML	MR	PU	SD	SI	TA	TE	UR	Average
Baseline	67.8	67.5	67.4	82.0	73.1	74.3	60.2	45.7	60.2	72.2	80.1	38.6	65.8
Scratch	66.2	77.3	66.8	73.9	68.1	66.9	65.9	66.8	68.4	68.3	68.3	66.0	68.6
Bactrian 7B	50.7	32.6	52.1	31.5	50.6	58.9	29.0	50.1	35.8	52.6	32.3	55.0	44.3
Bloomz 7B	51.5	56.1	59.2	46.0	44.9	53.3	49.1	49.7	34.7	43.1	47.0	54.8	49.1
Gemma 7B	72.3	78.6	73.3	75.5	73.2	70.9	68.1	64.4	66.5	72.3	72.3	70.9	**71.5**
Llama 7B	52.0	32.7	52.9	32.0	51.4	59.1	30.3	51.2	36.5	53.2	32.8	55.4	45.0
Mistral 7B	63.9	43.9	59.5	57.4	38.1	67.7	32.8	58.3	45.3	60.5	52.8	63.1	53.6
Tower 7B	58.3	35.6	57.0	34.7	58.7	65.5	34.0	56.9	41.4	58.1	35.8	61.3	49.8
ALMA 13B	57.0	35.3	56.3	34.3	57.8	65.3	33.3	54.7	40.1	56.1	35.1	60.0	48.8
Aya 13B	63.1	70.1	68.2	62.9	59.7	64.1	62.3	55.5	57.1	55.3	59.0	67.0	62.0
Bactrian 13B	56.3	35.2	55.7	34.5	57.1	63.2	32.6	54.2	39.6	56.3	35.6	58.8	48.3
Llama 13B	55.4	34.5	54.2	32.7	50.8	61.6	31.1	52.3	36.8	40.7	32.5	56.6	44.9
Llama2 13B	56.3	34.7	53.9	33.5	56.4	63.8	32.2	54.1	39.2	55.7	34.8	59.5	47.8
MT0 13B	63.1	68.9	68.8	63.4	59.6	63.9	62.7	54.9	58.6	54.2	60.1	67.2	62.1
GPT4 Turbo	77.7	78.5	79.8	79.7	75.1	78.1	69.8	34.1	62.4	74.6	81.1	78.7	**72.5**

Table 7: LoRA-tuned performance of the open-sourced LLMs in SPBLEU ↑ metric. The performance of the open-sourced LLMs improved a lot compared to their zero-shot performance. Gemma 7B and GPT4 models outperform the Baseline model. GPT4 is the best-performing model.

	BN	GU	HI	KN	ML	MR	PU	SD	SI	TA	TE	UR	Average
Baseline	24.4	25.6	18.5	17.3	29.4	20.2	26.7	36.4	36.5	26.1	19.2	41.7	26.8
Scratch	23.9	16.3	17.9	25.5	32.9	23.0	21.4	21.6	28.3	28.5	30.1	20.8	24.2
Bactrian 7B	40.2	64.8	35.5	65.1	48.9	31.9	62.7	38.4	61.1	43.9	65.2	33.5	49.3
Bloomz 7B	36.3	31.5	22.9	46.1	52.0	33.9	31.9	34.4	61.1	51.1	47.7	28.6	39.8
Gemma 7B	20.7	16.8	15.4	22.5	27.5	21.1	22.5	24.7	30.9	25.1	26.2	18.6	**22.7**
Llama 7B	39.4	65.0	35.1	64.6	48.2	32.1	62.1	37.8	60.3	43.3	64.9	33.4	48.8
Mistral 7B	29.8	54.3	30.7	40.4	61.7	25.7	61.3	33.2	52.7	37.1	45.6	28.6	41.8
Tower 7B	34.8	62.2	32.7	61.7	41.6	27.4	59.4	34.2	56.4	39.2	61.4	30.2	45.1
ALMA 13B	36.1	62.5	33.7	62.2	42.7	27.9	60.2	36.2	57.9	41.3	62.2	31.3	46.2
Aya 13B	27.0	21.5	17.8	34.4	39.4	25.8	24.8	29.9	48.8	38.7	37.8	20.7	30.6
Bactrian 13B	36.3	62.9	33.5	62.6	42.8	28.9	60.5	35.9	57.8	40.6	62.3	31.4	46.3
Llama 13B	37.2	63.6	35.4	64.4	49.0	30.6	62.3	38.3	60.8	57.1	65.5	34.0	49.9
Llama2 13B	36.6	63.2	35.9	63.0	44.0	28.8	60.9	36.5	58.4	41.4	62.5	31.6	46.9
MT0 13B	26.0	22.2	17.4	34.0	39.4	25.8	24.7	30.4	47.6	39.5	36.9	20.6	30.4
GPT4 Turbo	17.4	15.9	11.2	19.5	28.0	16.4	21.6	46.5	35.0	24.3	19.4	13.5	**22.4**

Table 8: LoRA-tuned performance of the open-sourced LLMs in WER ↓ metric. The performance of all the open-sourced LLMs improved a lot compared to their zero-shot performance. Gemma 7B and GPT4 models outperform the Baseline model. GPT4 is the best-performing model.

	BN	GU	HI	KN	ML	MR	PU	SD	SI	TA	TE	UR	Average
Baseline	21.5	21.7	21.3	12.1	17.3	16.6	28.2	38.2	25.7	18.3	13.0	42.8	23.1
Scratch	22.1	14.8	21.9	17.4	20.3	21.1	22.3	23.1	20.2	20.1	20.9	24.2	**20.7**
Bactrian 7B	38.5	60.7	38.1	61.0	37.4	29.7	61.9	39.8	54.0	36.9	58.9	35.5	46.0
Bloomz 7B	37.8	32.7	29.2	41.9	44.6	35.0	37.1	38.1	53.4	46.3	40.8	34.9	39.3
Gemma 7B	19.2	15.6	18.2	17.7	17.8	20.3	22.5	26.7	23.5	18.6	19.5	21.2	**20.1**
Llama 7B	37.4	60.8	37.5	60.6	36.5	29.4	60.9	39.1	53.5	36.4	58.7	35.2	45.5
Mistral 7B	27.3	50.7	32.1	36.0	53.6	23.2	60.0	34.0	46.2	30.6	39.3	29.4	38.5
Tower 7B	32.6	58.7	34.3	58.7	31.2	24.9	58.2	35.3	50.0	32.6	56.7	31.1	42.0
ALMA 13B	33.9	59.1	35.2	59.2	32.1	25.4	59.0	37.2	51.4	34.6	57.5	32.1	43.1
Aya 13B	26.2	20.4	21.5	26.3	28.8	24.9	26.4	33.1	30.6	32.7	29.2	24.1	27.0
Bactrian 13B	34.0	58.9	35.4	58.8	32.5	26.4	59.1	37.1	51.3	34.1	56.5	32.8	43.1
Llama 13B	35.0	59.5	37.1	60.8	39.4	28.3	61.2	39.4	54.8	51.9	60.2	35.5	46.9
Llama2 13B	34.5	59.6	37.6	59.9	33.1	26.1	59.8	37.4	51.9	34.7	57.5	32.7	43.7
MT0 13B	25.5	21.0	21.0	25.5	28.2	24.6	26.0	33.5	28.7	33.1	28.3	23.7	26.6
GPT4 Turbo	14.4	13.7	12.7	13.7	16.0	14.0	20.0	54.2	24.8	16.6	12.7	14.9	**19.0**

Table 9: LoRA-tuned performance of the open-sourced LLMs in SPWER ↓ metric. The performance of all the open-sourced LLMs improved a lot compared to their zero-shot performance. Gemma 7B and GPT4 models outperform the Baseline model. GPT4 is the best-performing model.

		Zero Shot				LORA Tuned		
		Gemma 7B	Aya 13B	MT0 13B	GPT4 Turbo	Gemma 7B	Aya 13B	MT0 13B
Cairo	**SPWER↓**	115.96	88.12	90.01	70.7	67.69	62.66	62.09
	WER↓	101.76	90.41	93.55	76.6	64.25	65.70	65.60
Tunis	**SPWER↓**	134.16	103.61	100.20	79.96	72.20	69.01	69.31
	WER↓	110.55	97.19	96.72	82.4	66.57	69.93	70.41
Rabat	**SPWER↓**	145.09	99.27	91.91	75.37	69.10	67.85	67.98
	WER↓	112.96	95.12	94.15	79.15	65.27	69.29	69.73
Beirut	**SPWER↓**	131.98	89.80	88.88	73.99	69.16	65.15	64.21
	WER↓	113.20	92.98	92.56	79.14	64.00	67.13	66.47
Doha	**SPWER↓**	105.21	83.18	82.71	70.02	65.59	61.95	61.56
	WER↓	97.96	89.66	89.25	76.87	62.33	65.09	64.79
Average	**SPWER↓**	126.48	92.80	90.74	74.01	68.75	65.32	**65.03**
	WER↓	107.29	93.07	93.24	78.83	**64.49**	67.43	67.40

Table 10: Zero-shot and Lora-tuned performance of the open-sourced LLMs in Arabic normalization task. The Lora-tuned models outperform the base models same as before. In this task the open-sourced models even outperform the GPT4 model.

Vernacular	Without Normalizing				With Normalizing			
	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓
Cairo*	45.1	43	43.07	49.38	47.1	45.8	39.7	46.11
Tunis*	28.7	27.2	60.23	66.7	36.3	35.5	48.4	55.62
Rabat*	35.5	33.7	53.62	59.4	39.7	38.4	45.65	53.13
Beirut*	36.8	34.5	50.88	57.71	42.9	41.6	42.32	49.08
Doha*	38.1	36.7	47.92	53.68	45.8	44.7	39.35	45.54
Aleppo	38.4	36.2	50.56	56.84	46.2	45.8	40.18	46.17
Aswan	41.8	39.5	45.91	52.67	46.7	35.5	39.56	46.18
Benghazi	37.5	35.2	50.31	56.53	45.9	38.4	39.99	46.3
Fes	43.5	42	45.47	50.29	47.2	41.6	39.33	45.69
Muscat	45.6	44.2	41.13	46.63	49.1	44.7	37.39	43.43
Sanaa	41.7	39.6	44.9	51.25	45.9	45.1	39.93	46.92
Mosul	43.5	41.8	43.8	49.29	43.2	45.4	42.59	49.11
Salt	44.9	43	42.68	49.19	47.4	44.6	38.38	44.52
Tripoli	34.2	32.2	53.7	59.81	42.4	45.9	43.43	50
Alexandria	47.3	45.1	40.11	45.96	**50.7**	47.7	36.11	41.78
Baghdad	42.1	40.2	45.58	51.28	44.6	44	41.21	47..34
Jeddah	38.4	36.7	47.85	54.12	44.5	42.1	39.89	46.18
Algiers	29.6	28.3	59.77	66.38	38.2	46.1	47.52	54.79
Basra	40.4	38.8	46	51.44	42.1	41.2	42.97	49.59
Damascus	40.8	39	47.58	53.6	46.8	**49.5**	38.85	45.09
Jerusalem	39.5	37.6	46.53	53.5	45.9	43.4	38.97	45.23
Sfax	24	22.6	64.97	71.55	31.9	43.3	53.36	61.54
Amman	42.8	40.8	44.75	51.25	47.3	36.9	38.5	44.95
Khartoum	44	42	44.13	48.93	48	40.7	38	44.06
Riyadh	49.2	47.7	37.06	42.35	**50.7**	45.3	36.42	42.1
Average	39.74	37.90	47.94	53.99	**44.66**	**42.93**	**41.12**	**47.63**

Table 11: Performance of the translation task with or without the normalization step in Arabic. *: for these vernaculars we had the data to do LoRA-tuning on an open-sourced LLM for those vernaculars. For the other languages, we used the LoRA-tuned model thus can be said we are normalized in a zero-shot setup. The normalization step helps outperform the previous baseline(without normalization) model in all the vernacular except Mosul.

Dialect	Without Normalizing				With Normalizing			
	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓
Barisal	11.1	9.1	92.17	97.27	16.5	14.1	74.99	83.37
Dhakaiya	18	15.5	77.81	86.56	22.3	20.1	67.3	75.05
Jessore	23.8	21.6	67.64	73.92	24.5	**22.5**	64.91	72.79
Khulna	22	19.4	71.39	78.78	23.4	21.2	65.35	72.99
Kushtia	22.5	19.6	69.98	76.71	**25.3**	22.4	**62.34**	**69.66**
Average	19.48	17.04	75.80	82.65	**22.4**	**20.06**	**66.98**	**74.77**

Table 12: Performance of the translation task with or without the normalization step in Bengali. The normalization step helps outperform the previous baseline(without normalization) model in all the dialects.

Dialect	Without Normalizing				With Normalizing			
	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓
Hewlêr	10.1	8.4	84.59	91.33	9	7.7	89.47	96.34
Mehabad	11.3	10.5	86.54	89.78	9.6	8.7	**83.11**	89.66
Silêmanî	**12.7**	**11.6**	84.44	**88.64**	10.7	9.6	87.05	93.2
Sine	8.6	6.9	93.14	96.09	10.1	8.4	85.83	92.79
Average	**10.67**	**9.35**	87.18	**91.46**	9.85	8.6	**86.37**	93.0

Table 13: Performance of the translation task with or without the normalization step in Kurdish. The normalization step helps outperform the previous baseline(without normalization) in just one dialect (Sine).

Dialect	Without Normalizing				With Normalizing			
	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓
Ahetze	15.07	15.80	82.41	79.41	17.37	18.46	78.96	76.77
Bidarrai	12.85	14.30	85.71	79.94	15.12	16.30	82.95	79.40
Iholdi	11.09	11.71	94.92	88.70	13.19	13.65	84.36	80.55
Mitikile	9.53	10.46	94.87	87.40	15.72	16.73	82.28	79.70
Uharte-Garazi	13.00	13.84	84.11	79.17	16.97	18.39	81.02	77.02
Aloze	7.13	6.94	107.14	78.57	11.04	11.04	**71.43**	78.57
Bidarte	13.95	15.30	84.37	79.80	17.69	18.52	78.94	75.33
Isturitze	8.36	9.37	95.96	84.89	13.21	14.38	87.87	81.19
Mugerre	14.58	15.71	84.23	78.63	17.23	18.46	80.69	77.25
Urdinarbe	3.69	3.75	114.29	97.97	7.31	7.35	102.43	90.69
Amenduze-Unaso	16.09	17.63	80.69	76.17	18.61	19.40	76.06	74.79
Donibane-Lohizune	12.39	13.11	89.96	86.70	18.37	20.13	77.93	75.10
Itsasu	15.16	15.68	83.91	79.40	5.60	6.01	105.41	100.31
Muskildi	4.71	4.71	124.18	102.96	8.22	8.41	100.34	89.93
Urepele	13.57	14.01	85.80	82.10	16.04	16.95	83.15	80.09
Arbona	15.82	17.12	79.99	75.76	17.85	18.83	77.37	74.28
Ezpeize-Undureine	7.56	8.35	102.19	95.50	12.77	13.92	89.11	85.94
Jatsu	10.69	11.75	94.14	87.01	13.78	14.67	86.55	82.55
Pagola	5.45	5.84	100.19	92.75	9.02	8.58	88.57	87.19
Urruna	19.76	21.42	73.88	70.09	**22.15**	**23.67**	71.79	**70.01**
Azkaine	17.42	18.66	79.21	74.79	18.82	19.83	75.10	72.64
Gabadi	11.99	12.97	86.72	80.48	19.33	20.82	78.30	73.97
Jutsi	16.32	18.01	80.05	75.94	18.22	19.92	76.58	73.64
Ziburu	15.19	16.75	80.89	75.51	16.96	18.04	77.86	74.75
Baigorri	13.52	14.41	85.39	80.63	17.19	18.58	79.09	76.10
Garruze	17.01	18.52	79.67	74.48	17.33	19.25	78.25	73.71
Larraine	5.87	5.82	102.73	93.36	10.06	9.72	91.88	86.18
Sara	16.32	17.19	82.82	78.55	20.71	21.67	72.84	70.33
Barkoxe	7.27	7.10	99.29	92.93	11.59	11.93	86.10	82.63
Hazparne	11.81	13.10	90.48	78.93	12.35	12.98	90.30	79.12
Larzabale-Arroze	14.93	15.90	81.72	77.63	17.52	18.17	80.05	75.79
Senpere	16.61	17.44	79.09	75.56	7.44	8.38	103.99	99.15
Behorlegi	16.63	17.25	79.09	76.56	18.36	19.31	75.55	74.17
Heleta	14.19	15.69	81.47	77.17	18.24	18.85	78.76	76.79
Luhuso	15.68	16.91	79.47	75.63	18.00	19.79	80.44	75.48
Beskoitze	16.38	17.52	80.17	75.64	20.38	21.75	77.00	73.86
Hendaia	15.39	16.62	82.20	78.45	19.53	20.75	79.23	75.62
Maule-Lextarre	5.77	6.49	118.66	106.23	11.59	12.48	90.73	88.01
Suhuskune	13.00	13.84	84.11	79.17	16.58	17.47	82.11	79.17
Average	12.61	13.51	89.13	82.32	**15.32**	**16.24**	**83.11**	**79.43**

Table 14: Performance of the translation task with or without the normalization step in Basque. The normalization step helps outperform the previous baseline(without normalization) model in all the dialects except Senpere, and Itsasu.

Dialect	Without Normalizing				With Normalizing			
	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓
Agugliaro	35.93	33.42	56.72	66.67	60.23	55.03	25.37	31.11
Alassio	21.73	24.11	83.91	75.15	47.96	45.66	47.15	50.00
Alba	18.96	17.31	79.51	81.93	49.40	48.00	43.85	47.26
Albosaggia	12.37	11.30	86.84	90.64	28.82	27.10	65.15	70.18
Aldeno1	33.66	32.25	60.24	64.72	51.52	49.95	39.36	44.27
Aldeno2	32.66	31.22	59.17	64.81	55.13	53.47	37.81	42.59
Aldeno3	35.74	34.02	56.86	62.09	55.35	53.69	37.92	42.60
Altare	8.26	7.94	94.40	97.69	27.84	26.67	64.89	70.44
Altavilla_Vicentina	33.59	31.15	57.66	61.40	60.34	58.33	30.86	34.63
Alte_Ceccato	38.22	36.31	55.87	59.13	64.80	63.30	29.27	32.34
Amblar	20.56	19.40	74.53	78.14	46.85	45.63	47.37	51.20
Andreis	18.31	16.11	78.44	85.18	45.02	43.58	48.72	52.25
Aquilano	42.73	42.73	20.00	25.00	100.00	100.00	0.00	0.00
Aquileia	16.20	14.24	80.29	85.25	38.45	35.88	53.88	59.19
Arcola	16.71	15.81	81.44	86.04	38.70	36.95	52.77	57.86
Arenzano	11.82	10.72	90.65	94.78	28.98	27.02	66.28	71.15
Ariano_Irpino	24.76	23.22	67.71	74.78	56.97	54.32	34.10	40.06
Arsiero	40.87	39.50	49.64	52.98	63.15	61.93	27.91	31.29
Arzeno	20.01	18.62	79.15	85.08	44.00	42.24	47.44	52.78
Bagnoli_Irpino	17.61	14.65	83.29	87.76	47.97	45.08	41.86	48.44
Bagnolo_S._Vito	14.83	15.13	84.77	89.14	42.74	41.94	49.68	53.05
Bagnoregio	39.59	36.93	51.84	58.80	56.84	52.72	35.59	41.39
Barcis	20.04	19.04	76.09	79.94	50.45	48.20	42.68	47.01
Bari	13.48	10.33	80.88	86.69	24.81	20.27	68.27	77.13
Bergantino	13.15	11.93	86.65	92.47	30.43	28.49	63.56	70.26
Biancavilla	37.92	36.58	51.41	56.52	69.20	67.26	24.47	28.31
Bitti	9.77	8.98	94.79	102.38	34.17	32.04	56.65	64.41
Bologna1	2.39	3.02	96.77	95.65	20.09	19.40	158.06	160.87
Bondeno	17.86	16.91	78.77	82.05	44.40	43.21	49.34	53.66
Borghetto_di_Vara	23.31	20.75	70.35	74.76	45.37	43.12	46.30	50.84
Borgo_San_Martino	10.74	10.20	89.21	97.22	44.63	43.72	48.66	55.50
Borgofranco_dIvrea	11.66	9.88	83.71	86.46	35.25	34.42	56.58	59.51
Borgomanero	11.98	12.36	92.13	88.85	33.33	32.05	61.75	66.48
Borgonato1	13.68	11.93	84.36	88.77	27.40	25.67	71.84	76.65
Borgonato2	16.72	14.54	82.79	87.43	33.35	31.99	60.56	64.37
Borgonato3	17.17	14.81	79.55	83.23	37.87	35.80	57.09	61.53
Borgonato4	14.37	12.45	82.01	85.93	33.88	33.38	60.34	63.77
Borgonato5	16.42	14.24	79.78	83.83	30.98	29.40	65.59	69.16
Borgonato6	16.33	14.33	90.61	98.80	31.35	30.21	62.79	65.72
Borgonato7	15.00	12.59	84.58	87.57	26.68	24.66	70.73	76.65
Borgoricco_1	37.59	36.29	54.53	56.59	60.42	59.61	31.84	32.49
Bormio	13.03	12.82	85.76	93.17	44.42	42.66	46.26	50.49
Bovolone	36.74	35.05	54.19	58.98	56.86	54.55	35.42	38.02
Briana	37.09	35.76	54.86	56.59	56.29	54.56	37.77	41.17
Brione	18.62	17.13	81.76	83.24	41.89	41.23	50.14	55.77
Cairo_Montenotte	16.69	16.86	79.23	83.83	35.23	33.16	56.47	62.28
Calalzo_di_Cadore	26.54	23.94	65.94	72.66	47.46	44.38	41.53	48.92
Calcinate	10.24	8.83	83.13	88.47	22.97	21.73	71.96	76.50
Caldogno	38.66	36.61	54.30	58.38	58.90	56.72	35.42	39.22
Calitri	13.94	11.41	81.38	86.74	34.03	31.38	54.89	62.61
Calizzano	14.26	13.65	85.92	90.95	38.44	36.68	53.49	57.75
Calliano	23.66	22.46	74.53	80.09	42.91	41.37	51.40	56.59
Camisano_Vicentino	33.74	32.34	55.20	59.28	63.55	61.50	26.93	30.99
Campagnola	33.49	32.46	60.11	62.50	63.18	61.83	29.10	31.48
Campi_Salentina	28.64	26.29	66.13	72.43	43.63	41.24	53.15	59.19
Campobasso	18.47	15.67	77.09	81.08	30.73	29.43	71.16	76.99
Capurso	10.66	8.45	86.10	94.74	28.66	25.87	62.23	72.37
Carcare	16.21	14.82	91.11	98.50	36.06	33.99	57.47	62.19
Cardito	15.56	15.24	81.93	88.79	43.12	41.70	53.01	57.72
Cardito1	16.32	13.03	80.67	87.78	42.57	39.42	51.39	59.92
Cardito2	15.32	13.88	82.20	89.15	44.31	41.74	50.07	56.25
Cardito3	18.24	16.20	78.63	86.68	44.43	41.73	49.32	56.47
Cardito4	17.86	16.70	83.78	91.80	47.07	43.87	51.35	58.20
Carife	9.39	8.46	96.74	101.81	39.74	37.66	50.67	56.73
Carmignano_di_Brenta	23.33	22.56	82.28	89.21	46.64	45.33	48.95	55.36
Carmignano_di_Brenta1	35.31	33.44	56.76	58.98	65.20	63.23	28.49	32.04
Carosino	20.08	17.96	76.43	83.15	31.70	29.58	65.40	71.72

Dialect	Without Normalizing				With Normalizing			
	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓
Carpi	18.24	17.12	81.11	85.73	48.87	47.05	43.78	49.18
Carrara	7.74	7.50	95.26	101.85	47.34	45.83	44.60	49.21
Casalmaggiore	11.13	11.78	101.64	97.34	31.32	31.00	60.38	64.95
Casarza_Ligure	19.56	18.31	77.52	82.91	39.36	36.76	53.68	58.77
Castellano	38.80	40.37	55.69	56.36	52.35	51.62	38.84	42.95
Castiglione_Messer_Marino	7.92	6.43	95.50	98.87	14.40	13.19	94.82	101.13
Castrignano_del_Capo	21.76	20.53	71.81	77.02	43.27	40.95	46.44	53.24
Catania1	25.63	24.29	66.22	73.78	51.13	49.37	39.79	45.68
Catania2	20.88	17.53	71.64	81.00	42.93	39.95	47.09	54.14
Catania3	14.04	12.02	83.56	92.87	27.80	24.79	63.67	73.41
Catania4	17.65	16.58	78.09	83.43	39.53	36.17	52.61	58.36
Cazet	16.04	14.62	81.88	88.22	33.51	31.51	61.88	66.33
Cencenighe_Agordino	17.67	17.30	80.43	81.18	37.52	35.80	56.36	60.51
Ceneda	32.79	29.88	58.61	66.21	54.27	51.82	38.32	43.87
Cesarolo1	30.01	29.25	62.58	68.40	47.85	47.25	48.23	50.72
Cesarolo2	20.22	18.66	78.96	82.77	41.23	38.30	49.70	56.18
Cesena2	14.45	12.55	82.28	87.28	38.53	36.67	56.40	61.13
Cesesa1	5.48	5.79	98.09	99.86	19.73	17.93	74.35	80.74
Cesiomaggiore	37.07	35.15	54.94	60.42	62.87	61.08	28.70	31.60
Chiavari1	23.21	21.16	74.33	78.67	56.15	54.18	37.06	40.63
Chiavari2	21.25	19.41	80.16	82.86	46.24	43.00	45.99	50.71
Chies_dAlpago	33.42	31.34	57.73	64.20	60.24	58.24	31.75	36.08
Chioggia	41.23	38.89	48.13	54.92	64.04	62.25	28.76	31.55
Cicagna	15.56	13.37	81.95	85.50	35.38	34.16	58.17	63.15
Cimolais	18.89	18.51	78.10	80.99	42.97	42.24	53.52	55.09
Cirvoi	27.34	25.94	65.14	72.07	54.00	52.30	36.85	42.01
Cividale	18.18	17.75	78.42	82.53	38.48	36.19	53.77	59.31
Civita_di_Bagnoregio_1	35.49	33.15	58.53	64.07	41.91	39.54	55.49	63.67
Claut	14.82	12.37	81.11	85.33	40.43	36.93	49.32	54.61
Colle_Val_dElsa	49.77	50.79	49.08	49.54	44.73	45.84	67.23	66.51
Collina	11.06	10.88	91.25	95.28	34.72	32.36	58.06	64.81
Colognola_ai_Colli	23.55	22.53	72.63	77.25	42.47	40.46	49.16	53.14
Comano	16.84	16.69	79.34	82.58	38.56	37.30	52.70	56.60
Copertino	17.72	15.74	81.83	86.11	31.98	30.65	68.55	75.64
Cordenons	18.38	17.22	82.25	87.21	44.87	43.21	46.33	50.39
Corigliano_dOtranto	30.42	28.92	58.97	66.00	51.31	49.74	40.00	45.35
Corleone	32.60	30.60	56.33	62.18	57.03	56.07	34.10	39.05
Correzzola	43.07	41.78	48.94	51.18	66.37	65.21	26.12	28.66
Corvara	10.53	8.81	87.55	95.19	27.35	25.24	68.56	77.66
Cosenza	23.66	23.33	70.40	76.94	49.68	47.34	41.00	47.24
Crotone	14.60	14.32	80.90	85.01	47.56	45.94	41.65	47.62
Cutrofiano	21.32	20.42	77.64	81.80	21.66	19.80	87.15	92.83
Due_Carrare	38.54	37.46	54.53	57.34	63.74	62.56	29.05	31.89
Due_Carrare2	37.37	36.33	54.30	56.89	60.39	59.39	32.85	35.63
Due_Carrare3	34.58	32.77	56.98	60.48	58.73	56.42	33.18	35.18
Facca	36.99	36.55	57.54	60.78	64.57	63.78	28.49	30.99
Faggiano	16.58	14.81	79.79	85.85	30.41	28.64	61.31	68.75
Falzè_di_Piave	34.38	32.53	59.11	60.93	61.93	59.32	30.84	34.43
Farra_di_Soligo	34.51	32.76	59.48	64.77	52.98	50.66	40.01	45.66
Favale_di_Malvaro	18.79	16.35	76.58	80.78	39.00	36.46	52.84	58.22
Ferrara1	15.93	14.59	74.08	80.71	42.47	40.94	47.63	52.75
Ferrara2	8.98	8.89	101.79	105.06	33.81	32.44	61.32	66.99
Finale_Ligure	14.94	14.88	90.87	88.79	39.16	38.30	51.89	55.74
Firenze	64.54	65.27	31.38	29.52	73.62	72.36	22.06	25.12
Forlì	13.23	13.37	86.30	89.64	37.62	36.50	56.57	61.57
Francavilla_Fontana	19.77	16.76	79.52	85.85	44.69	43.54	50.74	56.25
Frontale_di_Sondalo	20.02	18.31	76.68	81.36	38.18	36.19	56.94	63.43
Galliera_Veneta	36.14	34.59	58.10	60.93	64.13	62.14	28.38	31.14
Galliera_Veneta1	34.98	34.56	59.55	61.68	62.39	60.45	32.07	36.23
Gallipoli1	15.58	13.98	77.07	85.37	41.27	38.78	49.10	57.20
Gazzo	29.45	27.22	63.58	66.32	54.72	52.68	38.55	43.41
Gazzolo	32.65	30.11	61.56	64.37	55.17	52.26	38.66	42.51
Ghizzole_di_Montegaldella	36.78	32.97	54.19	59.88	63.21	59.96	28.72	32.49
Giazza	2.88	3.89	91.67	113.33	2.94	3.47	83.33	106.67
Gorizia	24.08	22.35	68.19	73.64	43.70	41.86	50.17	54.67
Gragnano	11.45	9.27	85.71	92.16	34.01	32.21	68.44	74.44
Granarola	15.82	14.27	77.85	82.54	43.36	41.95	44.83	49.40

Dialect	Without Normalizing				With Normalizing			
	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓
Grosio	18.60	17.70	74.10	80.93	48.10	47.52	45.80	49.53
Grottaglie	16.84	12.89	80.59	90.63	30.32	27.38	68.27	75.00
Iglesias	9.65	7.61	95.85	101.84	34.49	31.62	58.37	67.65
Illasi	23.21	21.32	72.38	78.57	48.53	46.42	46.62	51.87
Iseo1	19.53	17.86	74.19	78.14	36.86	35.34	55.64	60.78
Iseo2	16.98	14.71	79.33	83.83	36.20	34.81	56.98	62.13
Iseo3	13.51	11.39	85.47	90.12	29.62	27.96	65.03	69.31
Iseo4	15.85	14.08	78.32	82.19	34.69	33.98	58.99	63.02
Iseo5	17.64	16.74	89.39	91.77	37.70	35.67	56.09	59.43
Iseo6	17.48	15.45	78.77	80.84	38.12	37.38	57.65	61.98
Iseo7	18.07	15.50	78.10	82.78	33.09	32.56	61.45	65.12
Iseo8	14.95	13.47	88.38	91.32	33.22	33.24	61.23	63.02
Jesolo	34.11	32.01	57.73	61.84	57.34	55.57	33.24	37.81
La_Spezia	19.40	19.53	79.86	81.11	43.52	41.98	48.69	53.35
Laino_Castello	24.18	21.33	66.84	73.41	48.72	46.47	40.69	48.09
Lamon	23.60	22.33	68.94	75.92	51.98	49.64	39.53	46.72
Lanciano	19.22	16.55	74.03	81.99	30.65	29.41	76.17	83.46
Laste_di_Rocca_Pietore	15.21	14.64	83.15	85.80	39.01	37.84	53.38	58.59
Lecce	16.36	13.88	78.24	84.42	32.79	29.00	64.05	71.06
Lecce2	23.05	21.42	73.20	79.81	48.33	46.30	45.71	51.96
Lecco	20.94	19.70	74.30	76.54	38.81	37.21	53.03	58.85
Lesina	15.48	13.77	77.74	85.37	41.42	38.42	48.28	56.82
Lion	32.11	29.18	59.44	63.17	60.47	58.99	32.07	33.68
Liscia	3.57	2.47	106.92	114.00	15.14	12.09	76.88	84.79
Livigno1	11.59	10.04	86.82	91.28	32.27	29.89	64.19	68.81
Livigno2	9.77	8.63	93.68	98.72	22.16	20.16	71.51	78.87
Lizzano	7.35	5.67	90.91	85.71	7.16	8.91	118.18	114.29
Locorotondo	7.30	5.48	92.48	100.66	23.24	20.77	67.86	74.63
Locri	23.50	21.60	66.26	73.38	41.08	39.40	47.06	53.74
Lonato	18.02	16.69	76.05	79.95	41.49	39.55	52.09	56.63
Longare	35.51	34.33	58.79	61.31	58.41	56.82	35.51	38.58
Lubriano	20.15	18.46	74.93	84.50	33.40	30.51	57.18	68.60
Lucanico	18.20	18.12	75.20	80.99	42.09	40.38	51.40	54.19
Lucinico	14.18	11.97	86.00	90.79	35.72	31.77	55.33	64.04
Lughignano	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Lupia_di_Sandrigo	38.97	37.33	51.96	55.24	65.61	64.12	26.15	28.74
Luserna	2.38	1.89	158.33	146.67	0.00	0.00	100.00	100.00
Luzzara1	12.85	11.19	84.02	91.77	41.28	39.13	53.63	58.68
Macerata	24.69	23.05	72.00	76.66	48.24	44.83	41.60	48.63
Maglie	29.22	25.70	64.76	71.72	46.36	44.50	48.03	54.12
Malonno	12.46	11.33	90.15	94.83	28.09	26.46	66.03	71.37
Mantova	17.07	16.40	78.46	77.56	34.70	33.21	55.47	59.64
Marchigiano	33.60	28.72	57.49	65.21	50.92	45.65	39.34	46.12
Marcianise	35.22	33.66	53.62	59.80	56.56	53.64	34.89	40.27
Marostica	37.08	35.34	55.53	60.13	63.18	61.37	28.87	32.84
Marostica2	35.82	35.10	59.55	63.92	60.05	58.40	33.30	36.53
Martina_Franca	4.40	2.64	98.53	102.57	15.72	13.23	97.86	105.51
Martinsicuro	11.24	9.57	91.21	98.68	32.76	29.34	65.80	73.35
Maserà_di_Padova	35.00	33.82	57.88	60.63	63.42	61.37	28.60	31.59
Mason_Vicentino	35.15	33.20	57.11	61.79	66.29	65.07	25.98	28.21
Massafra	10.00	7.68	95.05	100.55	23.25	21.52	88.62	94.67
Mazara_del_Vallo	20.90	18.68	74.97	79.23	46.55	45.57	49.53	55.51
Melfi	16.60	13.98	75.85	83.48	41.84	38.37	50.32	56.18
Mellame_d'Arsiè	25.01	24.45	66.84	69.96	58.77	55.96	31.94	36.71
Messina1	28.41	26.24	59.79	67.80	54.53	52.23	35.48	41.43
Messina2	25.02	23.45	66.91	74.28	53.50	51.28	37.29	43.37
Messina3	24.61	22.63	67.50	74.64	51.85	49.61	39.68	45.89
Mestre	37.21	38.40	55.80	57.59	50.62	50.13	42.30	46.65
Milano1	19.55	18.10	73.26	76.31	40.76	38.13	46.52	52.48
Milano2	17.62	16.30	78.53	82.18	38.88	36.85	53.72	59.47
Milano3	26.71	25.64	66.03	70.36	53.77	53.29	40.67	44.31
Milano4	17.64	16.76	75.73	80.24	43.07	41.29	50.08	55.08
Milano5	17.01	17.05	77.04	83.30	37.61	35.09	50.42	58.90
Mirano	40.96	38.30	49.69	57.13	61.66	60.24	30.16	34.21
Moimacco	21.74	21.05	72.12	76.01	40.83	38.53	49.77	55.39
Molfetta1	7.80	6.72	101.39	103.34	31.80	29.61	58.91	65.82

Dialect	Without Normalizing				With Normalizing			
	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓
Molfetta2	8.39	6.92	92.07	98.34	30.19	27.49	60.43	68.23
Molfetta3	9.21	8.79	93.88	96.33	37.63	35.85	53.62	60.16
Molfetta4	9.61	8.27	92.45	96.76	35.40	32.87	54.89	61.24
Molfetta5	9.00	8.47	88.56	93.66	36.43	34.45	55.37	62.46
Molfetta6	8.87	7.81	92.34	96.18	33.19	30.86	57.61	65.49
Molfetta7	9.69	8.00	92.71	96.40	33.47	30.07	56.76	65.13
Monasterace1	24.08	22.72	65.69	73.70	46.59	44.20	42.82	50.79
Monasterace2	19.27	18.08	72.61	80.19	39.10	37.45	51.54	57.93
Moncalieri	14.15	13.89	83.37	84.40	40.14	38.88	50.22	53.35
Mondovì	12.60	12.42	87.16	85.71	31.54	30.13	59.71	64.82
Monno	11.21	11.09	93.77	94.28	27.09	25.55	68.81	75.46
Monselice	32.27	29.69	58.32	63.17	55.92	53.33	35.42	38.62
Montecalvo_Irpino	17.88	16.58	75.96	82.42	46.39	43.66	46.60	52.38
Montecchio_Precalcino	31.76	28.21	58.44	65.42	61.73	59.62	31.06	35.93
Monteiasi	21.38	18.73	77.91	84.19	35.83	34.41	54.89	60.48
Monteiasi_2	17.60	14.63	80.59	88.24	34.69	32.25	58.90	64.52
Montella	17.38	14.60	81.86	90.32	38.62	34.70	51.17	59.03
Montereale_Valcellina	24.46	23.36	67.76	71.71	47.17	45.94	43.21	47.66
Monteroni	17.57	16.47	80.29	85.37	39.69	36.38	55.74	62.93
Monterotondo	55.69	53.13	40.88	47.93	58.46	55.72	34.16	40.63
Montesover	37.50	37.56	56.25	57.73	55.58	55.55	36.72	38.85
Morolo	34.21	32.22	57.43	63.12	42.18	39.27	46.32	54.44
Motta_di_Livenza	39.14	38.60	53.18	55.82	62.59	60.82	28.84	32.26
Mussomeli	20.65	19.68	80.86	83.46	42.41	40.78	54.08	58.82
Napoli	12.42	10.10	84.45	90.42	38.38	36.31	53.50	61.11
Nardò	18.80	17.20	80.68	86.27	35.10	32.00	61.35	69.20
Nimis	21.76	21.29	71.63	78.23	47.17	44.47	43.17	49.88
Noale	33.88	31.80	57.77	59.43	63.29	60.29	29.27	32.34
Nones_	18.85	17.68	73.22	80.17	43.51	40.38	44.85	54.21
Novi_Ligure	8.65	5.73	100.35	101.40	21.10	18.66	93.75	98.60
Oneglia	22.07	21.34	73.42	77.69	49.67	47.78	43.61	47.62
Ortelle	29.42	28.06	61.12	67.94	49.91	47.98	40.64	46.69
Ortisei	7.72	8.39	92.04	95.95	7.49	6.37	123.88	129.73
Orvietano	34.45	31.90	56.06	62.42	51.33	48.12	44.91	49.67
Osimo	34.15	35.56	61.08	63.35	61.58	59.74	33.86	37.56
Ossi	14.64	13.83	82.83	89.86	39.57	37.66	51.58	59.06
Paciano	45.17	44.14	43.62	46.97	65.86	64.07	27.66	32.06
Padola	9.09	8.33	119.24	117.30	29.82	28.17	67.84	73.36
Padova1	29.92	30.49	71.65	75.10	46.19	45.27	49.33	55.68
Padova100	0.00	0.00	100.00	100.00	0.00	0.00	100.00	100.00
Padova3	36.43	34.78	55.26	59.94	65.95	65.04	25.29	28.95
Padova4	33.09	31.56	61.79	64.37	61.76	61.58	31.28	33.98
Padova5	32.88	32.59	59.58	61.51	54.08	52.24	37.60	42.11
Padova6	40.19	38.68	53.25	56.05	61.32	60.44	31.36	33.92
Padova7	38.05	35.60	53.68	57.55	63.46	61.15	26.97	30.54
Padova8	36.94	34.71	55.64	59.88	62.82	60.21	28.27	32.04
Padova9	38.52	36.51	54.08	57.04	63.02	60.82	29.50	32.04
Palazzolo_dello_Stella_	13.43	13.41	81.92	83.31	34.45	32.36	56.81	61.83
Palermo10	21.34	19.60	78.91	81.87	50.75	49.87	43.40	47.66
Palermo2	14.12	13.28	75.63	83.59	45.92	42.05	45.29	53.89
Palermo3	20.78	19.46	75.81	80.15	51.35	49.01	43.65	48.42
Palermo4	18.37	17.50	85.81	88.13	43.79	42.32	54.73	59.74
Palermo5	20.24	18.31	78.78	82.62	49.45	48.37	42.86	47.85
Palermo6	15.44	13.97	83.47	91.75	46.83	43.94	42.77	50.52
Palermo7	16.70	15.30	76.55	84.54	46.33	43.68	45.20	52.31
Palermo8	18.26	16.18	75.25	82.83	49.18	45.85	40.51	48.64
Palermo9	13.14	10.99	77.95	85.18	39.34	35.23	50.61	58.68
Palmanova	42.89	43.23	47.99	51.71	57.74	56.85	34.04	37.07
Palù_del_Fersina	2.42	2.69	95.83	113.33	4.68	3.30	95.83	100.00
Papasidero	20.33	17.27	75.45	83.81	40.61	36.39	52.38	60.32
Peaio	20.75	18.92	71.72	78.95	44.11	42.15	45.45	53.23
Pennapiedimonte	6.68	6.08	95.29	100.28	27.08	24.69	62.97	69.73
Pescara1	15.64	13.75	79.65	86.21	35.52	32.85	62.25	71.32
Pianella1	13.29	12.13	89.02	97.43	36.02	34.01	60.78	66.36
Pianella10	9.54	9.11	115.24	106.37	27.02	24.65	68.98	75.66
Pianella2	11.18	12.53	108.82	100.20	35.25	32.23	63.87	72.60
Pianella3	11.42	10.81	104.86	99.07	33.92	31.70	65.54	71.24

Dialect	Without Normalizing				With Normalizing			
	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓
Pianella4	7.98	7.08	101.81	111.35	27.61	25.87	71.55	75.96
Pianella5	13.24	11.45	95.14	105.75	26.21	24.13	75.54	83.49
Pianella6	9.86	9.52	108.51	102.23	32.41	31.35	64.46	69.20
Pianella7	10.45	9.23	98.78	108.99	30.06	26.58	70.48	79.59
Pianella8	7.51	4.58	106.42	113.81	21.00	18.44	91.49	98.57
Pianella9	8.02	6.38	101.81	109.93	26.59	23.81	72.56	78.41
Pianiga	35.93	34.36	56.42	58.83	60.64	59.54	32.07	34.73
Pianiga1	39.56	39.18	53.63	55.69	62.56	61.64	30.50	32.19
Pianiga2	34.61	33.08	56.20	58.38	57.27	55.78	34.64	37.43
Pianiga3	34.12	33.05	58.32	59.88	58.96	56.53	32.96	36.98
Piove_di_Sacco	38.66	38.72	54.56	55.81	63.30	61.09	29.36	33.48
Piove_di_Sacco2	37.70	37.15	52.92	55.12	59.17	57.58	32.36	35.39
Piove_di_Sacco3	37.02	35.48	55.64	58.08	62.22	60.67	29.05	33.08
Poirino	12.63	11.10	87.73	89.76	35.67	33.69	55.67	62.30
Pontevigodarzere_1	36.39	34.24	54.19	57.34	64.43	63.01	27.82	29.94
Pontevigodarzere_2	35.73	32.69	56.74	62.89	57.81	56.04	32.48	35.05
Pontevigodarzere_3	41.49	39.86	50.61	52.84	63.85	63.32	28.83	30.24
Pontinvrea	14.76	13.90	82.51	87.24	39.85	38.02	52.46	57.00
Posada	12.81	11.33	95.85	100.07	37.12	35.15	54.21	61.63
Pozza_di_Fassa	11.78	11.17	86.98	93.68	34.58	32.41	60.57	64.74
Pozzale_di_cadore	24.01	22.00	69.83	75.86	44.22	42.20	48.11	54.67
Pramaggiore	35.50	34.17	55.51	58.94	56.51	54.50	34.83	39.08
Prà_del_Torno	10.56	9.03	86.44	90.32	26.62	23.96	61.69	64.98
Puos_dAlpago	31.10	29.26	60.69	66.98	56.95	54.63	34.64	39.41
Qualso	16.75	15.96	81.26	86.03	5.72	4.95	101.81	108.26
Quinto_Vicentino	31.69	29.06	62.57	66.77	59.91	57.33	33.85	39.07
Ragusa	10.01	9.86	97.88	99.76	38.34	35.61	56.71	66.34
Ramats	5.78	5.60	100.94	105.54	17.44	16.60	78.69	87.23
Redondesco	13.70	12.41	84.47	90.68	34.51	33.65	61.26	68.15
Reisoni	24.37	22.35	68.94	73.57	48.35	46.67	43.01	47.09
Remanzacco	14.62	13.86	82.31	85.88	33.78	31.37	57.80	63.55
Revò	19.08	17.50	80.11	81.89	38.49	36.20	53.52	59.43
Rimini	11.21	11.54	85.94	86.46	27.48	26.63	67.19	70.04
Riomaggiore	17.50	16.80	82.56	85.35	36.45	35.44	55.47	59.37
Riva_presso_Chieri	15.55	13.95	81.08	83.86	38.18	36.87	55.80	60.44
Rivai_di_Arsiè	25.50	23.44	64.22	69.49	57.88	54.95	31.24	36.59
Rivarossa_Canavese	15.51	15.30	81.03	81.94	38.80	37.57	53.24	57.87
Rocca_Pietore	14.51	13.02	85.56	88.65	33.06	31.70	57.95	63.74
Rodoretto	8.72	8.66	93.82	93.57	31.58	31.42	62.29	63.94
Roma	100.00	100.00	0.00	0.00	37.00	69.14	57.14	40.00
Romanesco	39.07	37.75	52.05	60.24	55.62	52.57	39.96	47.48
Romano_DEzzelino	40.93	38.83	49.90	54.21	68.86	66.93	24.26	28.40
Ronzone	13.63	11.69	87.15	91.47	26.75	24.45	67.60	73.80
Ronzone_2	24.33	23.63	73.30	75.45	47.45	46.06	43.91	48.50
Rovereto	41.33	42.19	52.79	53.63	58.24	56.85	32.70	36.25
Rovigo	41.07	39.46	49.68	52.64	66.40	64.82	26.68	30.45
Rovolon	37.92	36.83	52.92	55.81	59.91	58.04	31.00	33.53
Salerno	7.65	5.88	109.01	118.16	33.32	32.86	62.72	69.73
Salzano	38.60	38.18	54.30	58.17	57.75	56.58	35.19	38.90
San_Cesario_di_Lecce	30.41	27.77	59.23	67.01	51.83	48.69	37.96	45.28
San_Leonardo	13.31	11.44	82.66	89.71	27.07	24.15	66.11	74.88
San_Marco_in_Lamis	24.07	22.75	68.01	73.46	51.35	48.56	38.42	44.12
San_Marco_in_Lamis2	14.48	14.23	83.57	85.96	36.34	35.50	52.83	57.38
San_Martino_di_Lupari	29.83	29.44	62.35	64.22	62.17	61.97	29.16	31.29
San_Martino_di_Lupari1	33.97	32.21	59.66	62.28	61.25	58.16	33.41	37.13
San_Martino_di_Lupari2	31.95	30.18	60.67	63.47	63.84	61.38	29.94	32.63
San_Martino_di_Lupari_4	31.33	30.49	62.68	64.82	57.12	55.82	34.30	36.98
San_Martino_di_Lupari_5	36.77	34.90	56.31	58.98	62.88	62.22	30.61	31.59
San_Martino_di_Lupari_6	37.00	35.49	55.75	58.23	66.63	64.69	26.26	29.34
San_Martino_di_Lupari_7	36.40	35.43	56.09	60.18	59.78	58.24	33.52	36.08
San_Martino_in_Pensilis	9.39	9.71	89.47	93.13	24.70	22.57	71.62	79.38
San_Michele_al_Tagliamento1	15.09	15.29	86.72	85.88	39.21	37.68	53.93	58.72
San_Michele_al_Tagliamento2	21.88	20.67	71.60	76.94	44.66	41.91	47.31	53.18
San_Pietro_in_Gu	37.23	35.31	54.09	59.30	68.21	66.91	24.74	27.94
San_Pietro_in_Gu1	38.27	37.24	55.08	58.23	66.21	64.30	27.60	31.74
San_Pietro_in_Gu2	32.41	30.96	58.77	62.13	56.69	55.29	34.08	36.83
San_Valentino	7.39	5.84	91.92	96.31	21.54	18.21	82.59	94.33

Dialect	Without Normalizing				With Normalizing			
	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓
San_Valentino_in_Abruzzo	8.22	7.66	88.57	90.56	19.33	16.70	73.06	81.11
San_martino_di_lupari_3	35.48	35.23	58.88	61.08	65.99	64.91	27.04	29.34
Santa_Croce_Bigolina	31.43	29.67	62.12	66.47	63.31	61.76	29.83	32.49
Santa_Maria_di_Sala	37.10	35.70	54.75	57.49	59.27	57.53	31.51	34.73
Santa_Maria_di_Sala_1	28.05	23.97	64.00	69.20	52.32	47.67	38.37	44.16
Santa_Maria_di_Sala_2	37.96	36.46	55.64	57.63	65.24	63.66	27.15	29.49
Santa_Maria_di_Sala_3	25.92	23.24	64.87	69.47	47.56	44.77	44.00	48.42
Santa_Maria_di_Sala_4	48.95	48.52	42.79	44.11	69.71	68.97	23.09	24.54
Santa_Maria_di_Sala_5	35.65	33.80	55.92	59.83	66.82	65.82	26.32	28.40
Savona	22.31	20.14	74.47	82.25	50.33	47.54	42.33	47.91
Scampitella	9.96	8.63	94.52	100.14	33.63	31.81	59.79	66.43
Schenone	7.09	6.19	108.59	111.47	31.47	31.17	63.09	68.46
Schio	34.64	33.12	56.87	60.33	60.87	58.94	31.28	35.48
Sciacca	20.44	19.69	68.62	75.15	49.42	47.32	41.06	46.12
Scorzé	39.57	39.54	49.78	53.21	52.99	52.32	35.83	39.81
Selva_di_Val_Gardena	9.80	7.62	103.37	109.41	20.37	18.72	79.43	83.77
Selvazzano_Dentro	36.24	34.60	56.65	59.73	61.30	58.85	29.27	32.78
Semogo	16.25	15.49	84.62	90.93	35.44	34.41	55.69	62.27
Sinagra	22.64	21.21	75.31	78.19	38.56	35.95	62.79	66.67
Solesino	34.05	32.74	59.78	62.28	67.88	66.28	24.58	27.40
Soleto	25.53	24.07	72.44	75.79	40.46	38.40	55.83	60.53
Squinzano	16.04	13.47	88.55	92.16	39.44	36.83	56.71	62.50
Standard	100.00	100.00	0.00	0.00	75.64	74.10	22.12	25.23
Sutrio	14.00	16.41	87.88	87.50	43.46	39.96	57.58	62.50
Tabarchino	8.15	7.28	91.55	102.73	28.51	26.01	68.91	78.36
Taggia	29.14	27.58	61.76	68.97	60.31	58.63	30.77	35.51
Taglio_di_Po1	22.97	22.27	70.24	76.63	39.07	37.90	52.70	58.50
Taglio_di_Po2	26.67	25.77	67.71	72.59	43.10	40.76	47.38	53.87
Tai_di_Cadore	31.85	29.15	59.95	67.51	58.27	55.84	31.34	36.10
Taranto	7.76	5.75	95.94	102.42	28.95	27.48	74.38	78.69
Teglio_Veneto	20.31	18.65	79.03	86.57	47.85	45.58	42.40	46.64
Teolo	28.41	25.66	70.39	79.79	55.33	52.70	36.20	39.82
Termoli	16.37	14.17	72.07	77.59	37.13	36.71	52.31	56.64
Terranegra	34.05	31.87	57.65	61.53	64.55	62.24	28.04	30.54
Terravecchia	14.39	12.57	80.56	89.33	33.37	29.67	58.55	68.65
Tezze_sul_Brenta	35.43	34.56	57.15	60.02	55.82	54.16	36.12	39.44
Tignes_di_Pieve_dAlpago	32.85	31.38	60.85	65.61	57.97	55.76	35.76	41.40
Tollegno	13.40	12.34	86.60	93.78	37.46	35.14	54.03	62.32
Torino	15.07	16.25	86.57	82.35	43.02	41.42	48.11	53.22
Torino1	12.55	13.45	90.55	86.31	36.77	34.46	55.47	61.73
Torino2	14.18	13.61	87.64	88.62	42.23	40.03	49.69	54.86
Torino3	15.63	16.64	87.05	81.49	43.74	41.36	47.63	52.11
Torino4	16.20	15.93	81.35	82.56	40.75	39.14	50.84	55.68
Torino5	18.25	17.48	86.55	89.86	50.21	48.22	40.96	46.06
Torino6	16.57	15.53	88.87	93.48	39.05	36.03	47.42	55.16
Torre_del_Greco	11.51	10.42	87.77	91.44	32.37	30.77	61.31	67.76
Torre_del_Greco1	16.92	15.97	80.74	84.99	37.39	34.62	63.96	69.49
Trecate	7.98	8.00	99.00	95.62	20.45	18.90	73.88	77.56
Treia	38.06	38.04	59.44	62.57	66.53	64.86	28.49	32.34
Trepuzzi	21.22	19.31	78.37	83.97	43.41	41.05	49.27	56.55
Trevico	15.02	13.89	80.37	85.59	37.68	35.24	53.99	61.46
Treviso	37.75	37.90	54.46	57.05	58.28	57.10	33.48	36.66
Tricase	22.54	21.41	67.86	71.22	44.42	43.02	46.00	50.51
Trieste1	34.29	33.95	66.23	67.95	55.12	53.82	38.59	43.40
Trieste2	40.32	37.88	49.37	55.70	60.56	58.62	33.12	36.94
Triggiano	9.80	9.75	89.42	95.74	43.33	41.21	49.48	55.50
Trissino	43.95	43.86	47.68	50.18	59.86	58.71	31.27	34.95
Troina1	22.42	20.85	69.47	76.22	48.13	45.53	43.14	49.78
Troina10	29.36	27.30	61.76	67.87	54.70	52.50	35.64	41.07
Troina2	25.91	24.54	65.21	73.13	52.92	49.35	37.55	45.32
Troina3	21.32	19.11	69.52	78.10	43.20	40.01	47.82	54.97
Troina4	26.46	24.43	66.12	72.77	54.11	52.06	38.40	43.52
Troina5	27.23	25.96	63.72	69.73	51.52	48.52	36.01	42.19
Troina6	20.45	18.46	71.15	79.97	43.17	40.37	46.61	54.74
Troina7	26.50	24.74	63.88	71.04	54.52	51.98	35.90	43.23
Troina8	29.59	28.39	62.07	68.37	58.22	55.03	33.40	39.84
Troina9	25.86	24.75	64.80	70.95	55.50	53.65	34.88	40.82

Dialect	Without Normalizing				With Normalizing			
	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓
Udine	10.68	14.43	114.29	120.00	100.00	100.00	0.00	0.00
Valdagno	33.94	32.69	56.89	63.17	59.45	57.54	31.50	35.76
Valfurva1	13.99	13.26	82.58	88.80	40.82	39.83	51.14	57.41
Valfurva2	16.21	15.36	79.59	84.47	41.77	39.63	48.06	54.98
Vallecrosia	18.99	18.08	79.51	83.15	43.83	41.51	47.66	52.15
Valmorbia	34.79	32.04	58.06	64.44	61.01	57.43	30.07	35.81
Vaprio_dAdda	13.93	12.27	85.43	90.15	35.42	33.51	59.71	66.12
Venezia1	39.23	38.71	53.07	54.49	61.90	59.94	30.28	34.73
Venezia_6	38.16	36.02	52.59	57.02	67.87	64.89	27.36	31.62
Veneziano	40.61	38.97	51.11	53.32	67.49	65.98	27.84	30.42
Venosa	9.13	6.75	89.84	96.54	25.57	22.54	67.71	75.65
Verona	34.53	33.75	59.04	62.26	57.91	56.55	34.71	38.71
Veternigo	33.92	32.32	56.86	60.31	66.78	64.67	26.14	29.72
Vicenza	36.81	34.22	55.29	59.62	66.49	64.87	29.93	32.52
Vicenza2	35.46	33.18	57.39	61.89	63.28	60.29	31.63	35.31
Vidor	37.65	37.95	55.42	57.87	56.71	55.51	36.08	39.51
Vidor2	35.64	35.71	57.25	61.89	60.41	57.52	33.46	38.46
Villa_di_Chiavenna	10.72	10.95	94.34	101.12	29.84	28.37	66.24	71.30
Villa_di_Tirano	16.39	15.61	82.58	87.73	40.66	39.60	50.28	56.84
Villacidro	7.78	5.09	99.10	102.73	33.72	31.49	64.62	71.22
Villafranca_Padovana	31.99	31.15	61.57	62.06	62.05	59.92	30.72	33.57
Villaverla	30.60	29.56	62.09	66.78	59.17	58.30	32.68	34.79
Villorba	34.67	33.02	53.77	61.15	60.65	59.25	30.09	34.68
Vione	13.86	14.42	84.71	82.49	28.78	26.95	60.27	65.53
Vitigliano	17.23	12.93	82.33	89.26	32.72	28.96	59.26	67.34
Vodo_Di_Cadore	0.00	0.00	133.33	150.00	0.00	0.00	133.33	200.00
Vodo_di_Cadore	15.96	14.35	84.36	88.73	45.92	42.32	51.65	59.72
Zero_Branco	36.28	36.17	55.82	57.34	65.79	64.41	29.02	30.24
Zianigo	38.13	36.64	52.94	55.42	66.44	65.36	28.37	29.90
Zianigo2	37.88	35.96	52.55	54.20	64.59	62.10	28.89	32.87
Zianigo3	40.16	37.52	49.02	53.50	67.42	64.60	26.27	29.55
Zianigo4	37.84	35.43	53.07	55.59	71.12	69.50	24.18	27.27
Zianigo5	33.52	32.09	57.25	59.27	60.66	57.96	30.46	34.79
Zianigo6	32.21	30.64	61.31	64.34	56.61	54.05	34.25	38.11
padova2	37.06	36.35	55.80	58.28	49.67	47.58	41.63	45.83
Average	23.21	21.90	73.32	77.68	**45.31**	**43.45**	**48.27**	**53.46**

Table 15: Performance of the translation task with or without the normalization step in Italian. The normalization step helps outperform the previous baseline(without normalization) model in all the dialects.

Dialect	Without Normalizing				With Normalizing			
	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓
Aarau,AG	51.08	47.46	35.88	43.13	78.10	76.37	15.20	18.13
Aarberg,BE	51.27	47.93	35.46	42.40	77.43	75.54	15.33	17.91
Aarburg,AG	51.08	48.02	34.89	41.67	74.35	72.62	18.81	22.60
Adelboden,BE	45.40	42.92	41.66	48.04	77.16	75.41	16.86	20.28
Aedermannsdorf,SO	49.27	46.19	37.19	44.21	77.07	75.29	16.44	19.57
Aesch,BL	49.79	46.35	36.84	44.21	75.74	73.65	17.67	21.46
Aeschi,SO	49.43	46.35	37.46	44.41	73.06	70.58	19.80	24.25
Agarn,VS	47.62	45.70	40.93	46.25	71.78	69.48	20.02	23.25
Alpnach,OW	49.68	47.90	37.54	42.91	73.09	71.46	19.83	22.73
Alpthal,SZ	48.71	46.17	38.37	44.83	73.75	72.45	18.81	22.15
Altdorf,UR	50.08	47.54	36.56	43.18	73.78	71.42	19.25	22.94
Altstätten,SG	50.80	47.37	37.06	44.09	75.34	73.45	17.17	20.20
Amden,SG	53.99	50.05	33.18	40.18	76.00	75.14	16.35	19.12
Amriswil,TG	52.00	48.03	35.42	41.95	75.51	73.93	17.38	20.71
Andelfingen,ZH	54.43	50.68	32.83	39.48	75.77	73.60	17.12	20.65
Andermatt,UR	49.28	47.46	38.55	44.34	75.11	73.33	18.06	21.59
Andwil,SG	52.19	49.21	35.85	42.73	79.19	77.48	15.01	17.70
Appenzell,AI	53.42	49.29	33.58	41.05	71.96	69.66	19.91	23.79
Arosa,GR	50.60	47.20	35.96	43.22	74.89	72.47	17.71	21.61
Ausserberg,VS	50.24	47.68	37.60	44.35	74.94	72.69	18.35	22.11
Avers,GR	50.82	47.79	35.86	43.74	73.13	71.31	19.49	23.58
Baldingen,AG	53.24	50.78	35.51	41.01	74.93	72.45	17.84	22.01
Basadingen-Schlattingen,TG	53.35	49.89	34.23	41.57	77.15	75.21	16.33	19.69
Basel,BS	53.06	49.29	33.95	41.09	76.57	74.59	16.56	19.77
Bassersdorf,ZH	52.91	49.32	34.43	41.53	76.89	75.14	16.10	19.16
Bauma,ZH	51.93	48.73	35.66	42.65	75.94	73.65	16.85	20.68
Belp,BE	52.68	49.89	34.87	41.18	75.24	73.50	17.95	20.78
Benken,SG	55.41	52.45	32.64	38.34	82.87	80.90	11.57	13.66
Bern,BE	50.35	47.44	36.13	43.12	78.30	76.68	15.50	18.83
Berneck,SG	49.71	46.68	37.57	44.52	76.20	74.42	17.67	20.84
Betten,VS	47.50	45.54	40.81	46.89	77.46	75.22	16.95	20.40
Bettingen,BS	53.35	50.17	34.06	40.68	78.16	76.75	16.04	19.04
Bettlach,SO	48.77	46.21	38.16	45.15	75.00	73.11	16.86	20.19
Bibern,SH	51.39	48.12	35.76	42.93	77.89	75.38	15.59	19.07
Bibern,SO	75.98	59.46	16.67	33.33	54.11	35.36	33.33	33.33
Binn,VS	50.60	48.39	37.33	43.39	76.95	74.89	16.13	19.51
Birmenstorf,AG	52.77	49.23	34.51	41.04	77.58	75.33	15.62	19.32
Birwinken,TG	53.36	49.32	33.64	40.31	74.13	71.73	18.44	22.46
Blatten,VS	48.09	45.82	40.14	46.78	78.08	76.06	16.13	18.71
Bleienbach,BE	48.67	45.50	36.77	43.43	79.09	77.17	14.42	17.53
Boltigen,BE	46.85	44.34	39.59	47.10	75.70	73.34	18.07	21.64
Boniswil,AG	49.73	46.93	37.25	44.04	76.93	75.46	16.94	20.48
Boswil,AG	49.75	47.01	37.35	44.13	77.37	75.63	17.02	20.28
Bottighofen,TG	52.81	48.62	34.14	41.37	76.61	74.75	16.70	19.84
Bremgarten,AG	52.56	49.84	34.20	40.82	75.47	73.92	17.10	19.81
Brienz,BE	49.08	46.49	37.88	43.93	75.50	73.31	17.27	21.03
Brig-Glis,VS	48.14	46.33	40.05	45.92	78.77	76.99	14.80	17.60
Brugg,AG	52.10	48.24	34.50	41.09	74.37	71.77	18.01	22.03
Brunegg,AG	50.22	46.83	36.31	42.80	76.09	74.25	17.09	20.46
Brunnadern,SG	52.53	48.75	35.03	41.49	75.01	73.40	17.98	21.26
Buchberg,SH	52.49	49.51	34.60	41.92	75.00	72.72	17.89	21.21
Buckten,BL	48.20	46.13	39.12	45.35	73.21	70.98	19.61	23.59
Buochs,NW	48.82	46.75	38.48	44.66	76.53	74.95	16.97	19.89
Bäretswil,ZH	52.22	49.20	35.42	41.72	74.66	72.89	18.73	21.95
Bühler,AR	51.42	47.62	36.49	43.31	74.35	72.89	18.39	21.52
Bülach,ZH	53.34	49.41	33.30	40.60	77.87	76.11	15.61	18.80
Bürchen,VS	49.68	46.71	38.32	45.36	79.10	76.75	14.86	18.04
Chur,GR	52.60	48.71	34.85	42.42	79.65	77.54	15.15	18.05
Churwalden,GR	52.81	49.95	35.10	41.55	76.77	74.84	17.08	20.00
Dagmersellen,LU	49.65	46.90	37.32	43.95	78.78	76.81	16.02	19.21
Davos,GR	50.54	47.75	37.09	43.39	72.88	71.36	20.75	23.83
Degersheim,SG	52.73	48.82	35.42	41.41	72.63	70.60	19.83	23.04
Densbüren,AG	50.22	47.26	37.50	43.63	76.11	74.09	16.94	20.81
Diemtigen,BE	48.05	45.88	39.74	45.32	77.02	75.46	16.94	19.62
Diepoldsau,SG	52.11	48.78	35.35	42.15	75.56	73.80	17.29	20.21
Düdingen,FR	50.41	47.39	36.47	42.38	75.10	73.02	18.04	22.16
Ebnat-Kappel,SG	51.42	47.77	35.70	42.89	77.48	75.43	15.92	19.62

Dialect	Without Normalizing				With Normalizing			
	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓
Egg,ZH	52.35	49.32	35.73	41.56	72.44	69.69	19.55	23.55
Eglisau,ZH	54.02	51.05	33.99	39.79	78.47	76.77	15.46	18.35
Einsiedeln,SZ	50.35	47.75	37.30	43.89	73.10	70.72	19.46	23.22
Elfingen,AG	51.80	48.77	35.32	42.08	77.11	74.80	15.71	19.05
Elgg,ZH	52.29	48.74	34.62	41.98	78.57	76.37	15.70	18.58
Elm,GL	52.81	50.20	35.92	41.95	73.73	71.33	18.86	22.47
Engelberg,OW	49.32	46.96	37.84	43.26	73.28	70.96	19.06	23.77
Engi,GL	51.32	48.82	36.45	42.86	74.99	72.68	17.95	21.49
Entlebuch,LU	50.84	48.14	36.73	43.04	78.59	77.63	15.58	18.50
Erlach,BE	49.92	46.53	36.69	44.11	77.04	75.68	16.67	19.76
Ermatingen,TG	51.88	48.02	35.23	42.97	77.28	75.79	16.34	19.92
Erschwil,SO	49.61	47.11	37.40	44.09	74.66	73.28	18.31	22.25
Eschenbach,LU	51.54	48.33	35.64	41.49	77.12	75.29	16.43	20.21
Eschenbach,SG	3.63	3.75	333.33	225.00	5.26	3.39	216.67	225.00
Escholzmatt,LU	49.99	47.45	37.11	43.43	77.40	75.99	16.37	20.19
Ettingen,BL	52.41	48.33	34.89	42.88	74.92	72.61	17.45	21.10
Ferden,VS	48.66	47.09	41.02	46.75	76.39	74.08	17.68	21.15
Fiesch,VS	47.61	46.26	41.10	46.22	77.59	76.25	15.76	18.78
Fischingen,TG	54.15	50.41	33.30	39.66	76.07	74.18	16.70	20.29
Flaach,ZH	52.01	48.63	35.27	42.29	76.93	74.90	16.78	20.61
Flawil,SG	51.87	48.18	35.65	41.78	74.01	71.33	18.48	22.06
Flums,SG	52.59	49.74	34.19	40.82	75.27	73.80	17.73	20.15
Fläsch,GR	52.11	48.46	34.85	42.29	75.86	74.08	17.24	20.18
Flühli,LU	48.07	46.08	40.25	45.77	76.28	74.87	17.14	20.24
Frauenfeld,TG	51.18	48.38	36.40	42.46	76.36	74.76	17.24	20.69
Frauenkappelen,BE	50.60	47.27	35.73	43.43	76.54	74.22	16.53	19.85
Fribourg,FR	49.39	46.68	37.44	44.36	74.25	72.08	18.67	22.44
Frick,AG	51.72	48.20	35.46	42.14	73.92	71.06	18.58	22.82
Frutigen,BE	47.79	45.90	39.71	44.60	74.45	72.43	19.21	21.86
Fällanden,ZH	51.57	48.32	35.71	42.11	75.03	72.95	18.41	21.82
Gadmen,BE	50.04	46.97	36.83	43.28	79.94	77.59	13.91	17.54
Gais,AR	52.84	48.77	34.05	41.30	76.15	73.88	17.07	20.52
Gelterkinden,BL	49.88	46.90	36.92	43.85	76.94	74.71	15.98	19.76
Giffers,FR	48.93	46.16	38.50	44.76	76.65	74.99	17.13	20.16
Giswil,OW	49.76	47.43	37.38	43.60	74.71	73.11	18.79	21.80
Glarus,GL	53.97	51.84	33.99	39.60	74.71	73.18	19.06	22.04
Gossau,ZH	50.98	48.62	37.18	43.14	74.07	71.83	19.04	22.52
Grabs,SG	52.10	48.19	34.82	42.48	74.62	72.48	18.60	21.77
Grafenried,BE	49.48	46.35	37.47	44.71	76.20	73.70	16.96	21.02
Grindelwald,BE	50.70	48.38	36.55	43.58	75.47	73.73	17.63	20.71
Grosswangen,LU	48.99	46.20	37.59	44.18	74.79	72.65	18.09	21.16
Gsteig,BE	47.22	44.60	39.57	45.50	77.72	75.60	15.86	18.90
Guggisberg,BE	46.95	43.73	39.67	46.50	71.51	69.49	21.24	25.74
Gurmels,FR	52.52	49.87	35.53	41.58	74.06	71.17	19.69	24.11
Gurtnellen,UR	51.20	48.92	37.41	43.63	72.21	70.06	19.60	24.01
Guttannen,BE	48.12	45.30	38.45	45.24	74.37	72.62	18.32	22.06
Guttet-Feschel,VS	49.62	47.70	38.80	43.70	76.88	74.34	15.87	18.89
Gächlingen,SH	50.09	47.36	37.07	44.37	73.18	71.37	20.15	23.66
Göschenen,UR	51.56	49.00	35.93	42.36	77.76	75.85	15.69	19.43
Habkern,BE	46.60	43.99	40.10	46.85	75.98	73.94	17.00	20.62
Hallau,SH	51.84	47.93	35.04	42.91	75.27	73.30	18.55	21.72
Hedingen,ZH	52.85	50.16	34.80	40.86	76.04	74.22	17.17	19.97
Heiden,AR	52.28	48.18	34.64	41.40	74.51	72.71	18.11	21.60
Heitenried,FR	47.71	45.58	39.72	45.42	72.73	70.59	19.81	23.90
Herisau,AR	52.22	48.13	34.14	41.10	75.61	73.90	17.21	20.62
Homburg,TG	53.04	48.46	33.83	40.85	74.16	72.33	18.10	21.60
Horw,LU	50.70	47.55	36.24	43.20	75.16	73.23	18.12	21.53
Huttwil,BE	49.30	46.11	36.86	44.04	77.48	75.74	16.45	19.92
Hägglingen,AG	49.81	46.60	36.40	43.33	78.58	76.81	15.23	18.27
Hölstein,BL	49.54	46.44	36.82	43.79	76.17	73.91	17.17	20.36
Hünenberg,ZG	50.90	48.24	36.42	43.07	75.09	73.63	17.78	21.20
Hütten,ZH	52.47	49.38	35.09	41.47	76.27	74.50	16.36	19.85
Hüttwilen,TG	54.20	50.27	32.86	39.74	76.04	73.78	16.43	20.00
Illnau-Effretikon,ZH	51.21	48.02	36.80	43.32	76.47	75.01	17.27	20.15
Inden,VS	50.42	47.39	37.06	43.51	76.61	74.27	15.80	19.80
Ingenbohl,SZ	51.13	48.94	36.01	42.27	73.84	71.45	18.47	22.81
Innerthal,SZ	49.78	47.67	38.71	44.13	76.08	73.85	17.99	21.72

Dialect	Without Normalizing				With Normalizing			
	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓
Innertkirchen,BE	47.47	44.14	38.12	45.25	74.33	72.19	18.97	23.07
Ins,BE	49.28	46.51	37.39	44.28	77.26	75.86	16.36	19.48
Interlaken,BE	48.74	45.32	37.44	44.55	77.75	75.74	16.71	20.08
Iseltwald,BE	48.50	45.78	38.39	45.68	77.12	75.30	16.74	20.00
Isenthal,UR	53.47	50.90	34.33	40.34	70.44	68.58	21.14	25.16
Ittigen,BE	49.77	46.60	36.58	43.82	78.47	76.59	15.64	18.41
Jaun,FR	46.37	43.78	40.68	47.94	74.17	71.63	19.68	24.44
Jenins,GR	51.76	48.47	36.18	42.78	76.80	74.89	16.42	19.39
Kaiserstuhl,AG	53.66	49.56	33.27	39.97	74.23	72.71	18.39	21.07
Kaisten,AG	53.74	50.30	33.73	40.45	74.27	71.85	18.40	22.24
Kandersteg,BE	48.33	45.60	38.17	45.38	77.75	75.58	16.94	20.52
Kerns,OW	49.84	47.59	37.08	43.63	73.69	71.56	19.26	22.94
Kesswil,TG	52.08	48.46	34.60	41.53	75.49	73.48	17.30	20.63
Kirchberg,SG	53.97	50.02	33.24	40.45	76.36	74.10	16.57	20.29
Kirchleerau,AG	50.87	47.81	36.12	43.25	77.87	75.80	15.61	18.79
Kleinlützel,SO	48.61	45.22	37.88	45.48	77.94	75.94	15.46	18.35
Klosters-Serneus,GR	53.19	50.84	34.31	40.43	74.13	71.74	18.62	22.45
Konolfingen,BE	49.42	46.92	37.40	44.80	79.91	77.85	14.69	17.87
Kradolf-Schönenberg,TG	52.59	48.47	34.20	41.18	72.55	70.11	19.32	23.28
Krauchthal,BE	50.18	47.20	36.61	43.78	76.63	74.55	16.56	20.24
Krinau,SG	51.55	47.81	35.58	42.06	76.60	74.98	16.73	19.63
Küblis,GR	52.57	49.44	34.74	41.23	72.63	70.04	20.31	23.96
Küsnacht,ZH	53.65	50.53	33.33	40.45	75.19	73.30	17.39	20.97
Lachen,SZ	54.91	51.52	32.68	38.87	75.70	74.01	17.04	20.59
Langenbruck,BL	50.56	48.09	36.34	42.48	78.09	76.22	16.73	19.86
Langenthal,BE	49.55	46.02	36.33	43.26	73.82	71.40	19.61	24.12
Langwies,GR	52.12	48.50	35.11	42.13	74.20	72.44	19.39	22.85
Laufen,BL	49.08	46.25	37.97	45.06	74.57	71.58	18.84	22.60
Laupen,BE	48.78	45.65	37.68	45.44	77.44	75.85	15.92	19.30
Lauterbrunnen,BE	48.38	46.31	38.77	45.00	75.69	73.71	17.63	20.49
Leibstadt,AG	53.08	49.15	34.47	41.23	75.96	74.02	17.37	21.18
Leissigen,BE	46.79	43.46	39.01	47.19	75.77	73.80	17.77	22.22
Lenk,BE	48.30	45.85	38.91	45.32	76.91	74.58	16.00	19.24
Lenzburg,AG	50.80	47.65	37.17	43.56	73.98	71.74	19.22	23.74
Liesberg,BL	48.96	46.18	37.69	45.03	77.52	75.11	16.11	19.74
Liestal,BL	49.51	46.95	37.30	44.41	77.82	75.74	16.59	20.59
Ligerz,BE	49.60	46.45	38.32	46.79	83.84	82.44	11.43	13.55
Linthal,GL	51.57	49.32	36.61	42.42	76.41	74.75	16.85	19.95
Luchsingen,GL	54.02	51.10	33.01	39.31	77.78	76.30	15.61	18.55
Lungern,OW	49.04	46.02	37.66	44.77	75.62	73.95	17.96	21.58
Lupfig,AG	50.72	47.32	35.46	42.21	76.57	74.95	17.09	20.36
Luzern,LU	50.53	47.42	36.25	43.04	78.37	76.56	15.84	19.25
Lützelflüh,BE	47.57	44.46	38.58	45.90	74.97	72.35	18.64	23.28
Magden,AG	49.45	46.28	37.04	44.22	76.42	74.22	16.57	20.41
textbfMaisprach,BL	49.34	47.00	37.94	44.34	77.94	76.42	16.68	19.97
Malans,GR	52.56	49.13	34.36	42.74	77.52	75.11	16.12	19.35
Malters,LU	48.60	45.51	38.39	44.46	72.15	69.57	20.95	25.73
Mammern,TG	53.87	50.37	33.88	40.30	75.34	73.18	17.75	21.28
Marbach,LU	51.03	48.60	37.05	42.63	76.55	74.27	17.05	20.32
Marthalen,ZH	53.57	49.63	34.22	41.11	76.79	74.75	16.59	20.69
Maur,ZH	52.90	50.55	35.42	41.56	73.67	71.47	19.11	22.42
Meikirch,BE	48.66	45.45	37.34	44.59	74.02	71.36	18.91	22.83
Meilen,ZH	50.25	46.98	37.24	44.44	74.86	72.75	18.84	22.72
Meiringen,BE	48.31	45.23	38.19	44.80	74.46	72.51	18.22	21.95
Melchnau,BE	49.44	45.99	36.09	42.91	78.82	77.55	15.56	18.22
Mels,SG	51.37	48.31	36.33	42.75	72.06	70.15	19.58	22.48
Menzingen,ZG	52.57	49.63	34.38	40.29	76.44	74.21	16.86	20.08
Merenschwand,AG	49.20	46.68	38.39	44.39	74.73	72.28	18.33	21.93
Merishausen,SH	52.14	49.47	35.25	42.08	75.70	73.28	17.25	21.11
Metzerlen,SO	53.41	50.65	34.78	40.93	76.88	74.87	17.04	20.47
Mollis,GL	52.78	50.35	35.17	40.86	73.19	71.46	19.01	22.33
Mosnang,SG	51.58	48.18	36.27	42.43	75.25	72.88	17.99	21.80
Muhen,AG	48.97	45.83	36.98	44.15	79.00	77.21	15.39	18.30
Muotathal,SZ	48.64	46.25	37.88	44.50	72.00	70.09	20.21	23.95
Murgenthal,AG	50.92	47.71	35.86	43.20	76.65	74.27	17.03	20.85
Murten,FR	48.81	45.91	37.33	44.59	75.18	73.33	17.70	21.35
Mutten,GR	54.23	51.70	33.56	39.76	77.47	75.62	15.90	19.00

Dialect	Without Normalizing				With Normalizing			
	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓
Muttenz,BL	52.98	49.66	34.21	40.63	78.88	76.73	15.14	18.36
Möhlin,AG	50.40	47.72	37.51	43.84	74.21	71.43	18.67	23.12
Mörel,VS	49.70	46.72	37.79	44.32	74.95	72.81	17.52	21.36
Mörschwil,SG	52.20	48.27	35.32	41.69	73.80	71.87	18.73	21.82
Mümliswil-Ramiswil,SO	49.39	45.84	37.17	44.16	73.60	71.98	19.32	22.69
Münchenbuchsee,BE	50.43	47.37	35.74	42.55	76.44	74.63	16.86	20.40
Neftenbach,ZH	53.21	49.76	34.01	40.87	78.58	76.72	15.06	18.25
Neuenegg,BE	49.76	46.42	36.02	43.49	80.49	78.33	14.46	17.32
Neuenkirch,LU	50.62	47.16	35.92	41.97	76.43	74.63	17.09	20.24
Niederbipp,BE	49.15	46.14	37.42	44.53	75.88	73.48	17.99	21.47
Niederrohrdorf,AG	52.09	48.43	35.09	42.30	76.02	74.71	16.41	19.19
Niederweningen,ZH	51.84	49.33	36.16	42.05	74.93	72.95	17.68	21.52
Nunningen,SO	48.66	46.08	38.70	44.99	73.34	71.31	19.16	23.59
Näfels,GL	53.95	51.09	34.32	40.08	73.13	70.88	18.73	22.15
Oberhof,AG	47.99	44.62	38.56	45.76	72.63	70.11	19.79	24.51
Oberiberg,SZ	48.89	46.83	38.60	44.84	75.60	73.73	17.88	21.16
Oberriet,SG	51.66	48.07	35.17	42.26	73.42	71.51	19.27	22.63
Obersaxen,GR	51.38	47.89	35.87	43.52	77.43	74.59	15.94	19.50
Oberwald,VS	46.84	45.06	41.30	46.86	71.23	68.68	21.22	25.90
Oberwichtrach,BE	49.46	46.01	36.41	43.84	77.09	75.10	17.11	20.13
Oberägeri,ZG	49.90	47.65	37.68	43.39	73.00	71.15	19.64	23.81
Obstalden,GL	51.42	48.43	36.02	42.91	79.01	77.01	14.93	17.59
Pfaffnau,LU	51.63	48.42	35.10	41.50	77.12	75.45	16.49	19.28
Pfäfers,SG	52.07	49.43	36.11	42.77	77.01	75.40	15.63	18.65
Pfäffikon,ZH	54.18	50.39	33.55	39.82	74.76	72.69	17.94	21.66
Pieterlen,BE	49.23	46.25	37.09	44.46	76.12	74.16	16.88	20.40
Plaffeien,FR	47.32	44.32	39.23	45.92	70.75	68.54	21.06	25.27
Pratteln,BL	48.95	45.45	37.42	45.17	73.73	71.81	19.54	23.04
Quarten,SG	53.60	50.60	34.86	40.99	76.31	74.30	17.01	19.97
Rafz,ZH	51.94	49.24	36.05	42.24	76.83	74.74	16.49	19.42
Ramsen,SH	52.11	49.13	35.12	42.22	76.38	74.42	17.37	20.52
Randa,VS	49.35	47.20	38.67	45.01	79.19	77.18	15.07	18.11
Rapperswil,BE	52.94	49.93	34.51	40.96	78.87	77.09	15.63	18.60
Reckingen,VS	49.60	48.28	39.17	44.49	76.91	75.29	15.96	18.72
Regensberg,ZH	53.82	50.69	34.06	40.81	75.88	73.97	17.49	20.91
Reutigen,BE	50.11	46.74	36.87	43.71	74.24	71.68	19.12	23.00
Rheineck,SG	53.21	50.00	34.55	40.41	75.10	73.20	17.46	20.33
Rickenbach,SO	47.98	45.09	39.31	45.81	74.65	72.32	18.49	22.58
Rifferswil,ZH	53.11	49.05	33.65	40.21	74.51	72.36	18.81	22.19
Risch,ZG	51.15	49.30	37.12	42.88	77.42	75.49	16.95	20.32
Roggenburg,BL	50.19	47.29	36.53	44.00	76.95	75.07	16.44	19.72
Roggwil,TG	51.72	47.81	35.24	42.47	78.34	76.36	15.04	18.15
Romanshorn,TG	53.99	50.15	33.09	39.51	75.82	73.63	17.29	20.59
Rorbas,ZH	53.17	50.40	34.62	40.18	76.46	74.68	17.31	20.66
Rubigen,BE	49.50	45.81	36.95	44.87	78.22	75.96	16.48	20.64
Ruswil,LU	52.29	49.43	35.05	42.22	77.73	76.19	16.87	19.74
Römerswil,LU	49.47	46.63	37.44	44.39	76.83	74.59	16.91	20.59
Rüeggisberg,BE	52.22	49.39	34.68	40.77	76.64	75.10	16.71	19.79
Rümlang,ZH	53.25	49.98	34.39	40.88	79.13	77.58	14.88	17.74
Rüte,AI	51.14	47.70	35.96	42.31	74.39	72.26	18.03	21.68
Saanen,BE	46.01	43.37	40.81	47.67	79.39	77.98	14.50	17.31
Safien,GR	51.46	48.47	37.07	44.61	78.63	76.33	15.52	19.14
Salgesch,VS	49.32	47.48	38.76	44.21	76.67	74.37	15.76	19.43
Sarnen,OW	49.04	46.63	38.49	44.72	75.31	73.43	18.02	21.37
Schaffhausen,SH	54.64	51.03	33.24	40.03	77.25	75.78	15.38	18.23
Schangnau,BE	50.94	48.26	35.50	41.84	77.19	75.34	16.17	19.34
Schiers,GR	51.70	48.57	35.47	41.71	74.71	73.14	19.24	22.33
Schlatt-Haslen,AI	50.59	46.47	36.63	43.15	73.35	71.01	19.36	22.82
Schleitheim,SH	53.00	49.61	34.72	41.93	75.01	73.88	17.93	20.50
Schnottwil,SO	50.07	47.05	36.72	43.94	80.20	78.52	14.33	17.71
Schwanden,GL	53.29	50.37	34.13	40.68	74.95	73.13	17.75	21.42
Schwyz,SZ	50.79	48.08	35.89	41.69	70.91	68.64	20.84	25.06
Schänis,SG	53.65	50.13	33.62	40.79	77.61	75.56	14.94	18.15
Schönenbuch,BL	49.18	46.62	38.12	45.01	77.79	75.94	16.92	19.84
Schüpfheim,LU	49.72	46.71	36.95	44.28	74.79	72.14	18.47	22.34
Seftigen,BE	50.33	47.29	36.12	43.27	77.66	75.80	16.72	20.25

Dialect	Without Normalizing				With Normalizing			
	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓
Sempach,LU	50.38	47.51	36.54	43.80	76.88	75.58	17.00	19.66
Sennwald,SG	51.43	48.11	36.28	42.95	74.94	73.24	17.68	21.22
Sevelen,SG	51.34	47.73	35.21	43.35	76.53	74.71	16.41	19.44
Siglistorf,AG	54.08	51.04	33.96	40.23	76.92	74.85	16.14	19.66
Signau,BE	50.68	47.66	35.79	42.50	80.33	78.83	13.96	16.81
Silenen,UR	51.75	48.92	35.73	41.66	78.80	77.25	15.78	18.66
Simplon,VS	51.72	49.26	36.96	43.04	78.45	76.27	14.66	17.69
Solothurn,SO	51.37	48.86	35.47	41.33	73.06	71.61	19.65	22.80
Spiez,BE	48.74	46.07	39.35	46.48	76.41	75.05	17.49	20.70
St.Antönien,GR	51.72	48.70	35.63	42.65	75.05	73.07	18.34	22.05
St.Gallen,SG	52.35	48.71	33.93	41.54	75.83	74.34	17.71	20.70
St.Niklaus,VS	46.72	44.95	41.37	47.03	73.69	71.24	19.26	23.25
St.Stephan,BE	48.03	45.60	39.48	45.42	74.42	72.74	18.44	21.16
Stadel,ZH	54.74	51.75	33.39	38.90	78.48	77.14	15.04	17.73
Stallikon,ZH	50.79	47.79	36.90	43.70	76.74	75.09	16.83	19.60
Stans,NW	50.74	48.22	36.16	42.75	74.20	72.44	18.55	22.67
Steffisburg,BE	49.21	46.28	37.11	44.28	75.92	74.24	17.94	22.21
Steg,VS	50.21	47.60	37.89	44.44	78.20	76.49	14.97	17.39
Stein,AG	53.70	50.01	33.61	40.36	72.89	70.17	18.89	23.52
Stein,SG	75.98	59.46	16.67	33.33	54.11	35.36	33.33	33.33
Sternenberg,ZH	51.21	47.78	35.59	42.86	76.06	73.77	17.11	21.05
Stüsslingen,SO	50.31	46.73	35.87	42.70	75.26	73.52	18.13	21.02
Sumiswald,BE	48.45	45.13	37.04	44.23	76.34	74.83	16.96	19.67
Sursee,LU	49.88	46.95	37.25	43.69	77.47	74.66	15.78	19.38
Tafers,FR	46.87	43.99	39.90	46.38	73.70	71.71	19.51	23.74
Tamins,GR	53.34	50.51	34.17	41.83	76.63	74.41	17.63	21.11
Teufenthal,AG	49.87	46.27	36.29	43.47	74.39	72.36	18.47	22.12
Thalwil,ZH	55.16	51.65	32.50	38.77	77.24	75.43	16.06	19.26
Thun,BE	48.64	45.51	37.55	45.09	77.76	75.85	15.87	18.83
Thundorf,TG	53.81	50.17	32.93	39.84	76.69	75.10	16.23	19.01
Thusis,GR	52.90	50.06	34.70	41.89	77.86	75.72	15.72	18.81
Triengen,LU	49.09	45.48	37.50	44.60	76.25	74.01	17.82	21.85
Trimmis,GR	52.15	49.04	34.99	42.28	77.10	74.97	16.60	19.71
Trogen,AR	51.82	47.51	34.67	42.36	73.94	71.89	19.20	22.41
Trub,BE	49.65	46.52	36.35	42.84	74.79	72.83	18.42	22.09
Tuggen,SZ	52.72	49.37	34.33	41.09	76.97	74.64	16.49	19.60
Turbenthal,ZH	53.34	50.44	35.14	41.04	77.47	75.95	15.93	18.42
Täuffelen,BE	47.94	44.82	38.27	46.19	78.86	77.15	14.92	18.37
Tüscherz-Alfermée,BE	50.89	47.56	35.87	43.06	78.52	76.75	14.93	18.19
Ueberstorf,FR	51.30	48.25	35.86	41.92	77.71	76.02	15.63	18.69
Unterschächen,UR	46.26	43.88	40.96	46.98	75.76	74.56	17.53	20.72
Unterstammheim,ZH	51.61	48.52	36.26	43.42	74.70	72.77	17.94	21.84
Untervaz,GR	52.49	49.39	35.17	42.46	76.12	73.40	17.54	21.98
Urdorf,ZH	53.93	50.36	33.24	39.61	75.54	73.82	17.33	21.05
Urnäsch,AR	50.43	46.40	36.96	44.39	73.70	71.96	18.76	21.80
Ursenbach,BE	48.82	45.98	36.97	44.31	77.64	75.43	16.40	19.71
Uster,ZH	53.74	50.91	34.35	39.95	74.22	72.83	18.65	21.64
Utzenstorf,BE	49.12	45.76	37.32	44.49	77.83	75.81	15.86	19.12
Vals,GR	50.36	47.32	36.71	43.64	73.41	70.90	19.82	24.68
Villigen,AG	51.58	47.85	35.40	41.94	76.91	74.48	16.08	20.20
Visp,VS	48.83	46.94	39.04	44.72	76.38	74.45	17.23	19.69
Visperterminen,VS	47.76	45.96	40.32	46.01	75.07	72.52	18.02	21.54
Wahlern,BE	49.44	45.85	36.56	43.49	73.64	70.96	19.44	24.03
Walchwil,ZG	51.13	48.53	35.96	42.66	75.51	73.66	17.31	20.73
Wald,ZH	53.70	49.85	33.55	40.10	75.08	72.98	17.34	20.70
Waldstatt,AR	51.26	47.32	35.07	42.29	76.25	74.67	17.58	21.01
Walenstadt,SG	51.12	48.42	36.72	43.05	76.85	74.83	16.42	19.56
Wartau,SG	50.55	47.73	37.38	43.91	76.27	74.56	16.71	19.70
Wattwil,SG	51.72	48.00	35.53	42.06	76.20	74.15	16.88	20.13
Wegenstetten,AG	51.43	48.42	35.75	42.75	74.26	71.83	17.70	21.31
Weggis,LU	49.25	47.10	38.09	43.77	76.34	74.46	17.82	20.84
Weinfelden,TG	52.70	48.50	34.30	41.25	76.09	74.70	16.87	20.23
Welschenrohr,SO	47.81	44.35	38.34	45.73	77.41	76.16	15.64	18.47
Wengi,BE	48.76	45.30	37.49	44.94	76.51	74.20	17.12	20.78
Wiesen,GR	54.81	51.74	32.64	39.19	75.83	73.46	17.31	20.84
Wil,SG	53.23	48.53	33.33	41.01	73.54	71.06	19.18	23.03
Wilchingen,SH	51.08	47.21	35.89	43.79	76.92	74.80	16.97	20.36

Dialect	Without Normalizing				With Normalizing			
	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓	SPBLEU ↑	BLEU ↑	SPWER ↓	WER ↓
Wildhaus,SG	51.41	47.79	36.68	43.28	74.72	72.69	18.53	22.24
Winterthur,ZH	52.18	48.17	34.59	41.90	79.25	78.02	14.66	17.54
Wolfenschiessen,NW	50.33	48.56	37.64	43.37	74.40	71.93	18.16	21.88
Wolhusen,LU	50.43	48.38	37.75	43.46	77.22	75.36	17.09	20.16
Wollerau,SZ	51.66	48.68	35.92	42.57	76.83	75.69	16.31	18.81
Worb,BE	49.12	45.73	36.80	44.23	74.77	72.27	17.98	21.79
Wynigen,BE	48.64	45.86	38.06	44.63	76.45	74.69	16.85	20.31
Wädenswil,ZH	54.59	51.57	33.83	40.15	78.84	77.56	15.11	17.76
Wängi,TG	53.83	50.26	33.05	39.32	72.36	70.64	21.03	25.00
Würenlos,AG	53.78	50.19	33.90	40.77	78.12	76.38	15.43	18.34
Zell,LU	50.53	47.29	36.22	42.46	77.99	76.37	16.52	19.92
Zermatt,VS	49.41	48.10	39.46	45.06	71.04	69.22	21.53	25.41
Ziefen,BL	49.03	45.76	37.58	44.76	75.32	73.18	18.09	22.12
Zihlschlacht-Sitterdorf,TG	53.39	49.19	33.36	40.13	77.43	75.24	16.36	19.61
Zofingen,AG	51.06	47.32	35.57	41.98	76.47	74.57	16.79	20.05
Zug,ZG	51.45	49.06	36.12	41.62	74.88	73.21	18.64	22.30
Zunzgen,BL	49.98	46.42	36.46	43.70	78.05	76.09	15.60	19.17
Zweisimmen,BE	48.59	45.73	38.97	45.48	76.02	73.90	16.54	19.62
Zürich,ZH	52.56	48.98	34.83	41.65	75.52	73.39	17.73	21.65
Average	50.88	47.77	37.06	43.46	**75.63**	**73.56**	**17.98**	**21.39**

Table 16: Performance of the translation task with or without the normalization step in Swiss German. The normalization step helps outperform the previous baseline (without normalization) in all the dialects.