Effective Self-Mining of In-Context Examples for Unsupervised Machine Translation with LLMs

Abdellah El Mekki [ξ] **Muhammad Abdul-Mageed** [λ,ξ,γ]
[λ]The University of British Columbia [ξ]MBZUAI [γ]Invertible AI
abdellah.elmekki@mbzuai.ac.ae, muhammad.mageed@ubc.ca

Abstract

Large Language Models (LLMs) have demonstrated impressive performance on a wide range of natural language processing (NLP) tasks, notably through in-context learning (ICL). In ICL, an LLM is provided with examples that represent a given task such that it learns to generate answers for test inputs. However, access to these in-context examples is not guaranteed especially for low-resource or massively multilingual tasks. In this work, we propose an unsupervised approach to mine in-context examples for machine translation (MT), enabling unsupervised MT (UMT) across different languages. Our approach begins with word-level mining to acquire word translations that are then used to perform sentence-level mining. As the quality of mined parallel pairs may not be optimal due to noise or mistakes, we introduce a filtering criterion to select the optimal in-context examples from a pool of unsupervised parallel sentences. We evaluate our approach using two multilingual LLMs on 288 directions from the FLORES-200 dataset (Team et al., 2022) and analyze the impact of various linguistic features on performance. Our findings demonstrate the effectiveness of our unsupervised approach in mining in-context examples for MT, leading to better or comparable translation performance as translation with regular in-context samples (extracted from human-annotated data), while also outperforming the other state-of-the-art UMT methods by an average of 7 BLEU points. Our code is available at: https://github.com/UBC-NLP/sm-umt.

1 Introduction

Large language models (LLMs) have significantly advanced various natural language processing (NLP) tasks (Brown et al., 2020; Kaplan et al., 2020; Ouyang et al., 2022; Touvron et al., 2023), with generative pretrained Transformer (GPT) models, which follow a decoder-only architecture (Radford and Narasimhan, 2018; Radford et al., 2019), demonstrating particularly admirable outcomes. Performance of LLMs is especially effective when using few-shot learning through *in-context learning* (ICL, Brown et al. 2020). In ICL, the LLM is provided with task-specific input-output examples, allowing it to learn how to make predictions for a test input based on these examples without further fine-tuning. This approach has shown promising results in tasks such as question answering (Li et al., 2023a), common sense reasoning (Geva et al., 2021; Wu et al., 2023), and text classification (Milios et al., 2023; Khondaker et al., 2023; Chandra et al., 2024). However, ICL is sensitive to several constraints. For example, it is highly dependent on the selected in-context samples. In addition, small changes in the prompt can lead to high variance in acquired outputs (Lu et al., 2022; Chang and Jia, 2023). To address these issues, researchers have proposed several methods to choose the best examples that can lead to optimal performance (Rubin et al., 2022; Li et al., 2023b; Luo et al., 2023; Agrawal et al., 2023). For instance, Agrawal et al. (2023) propose selecting examples that are closer to the test input in terms of BM25 score. Another challenge is that performance of ICL can be affected if the task is expressed in a language not well represented in the LLM pretraining data (Huang et al., 2023). This is especially relevant to work attempting to leverage multilingual LLMs such as XGLM (Lin et al., 2022), Bloom (Scao et al., 2022), Gemma (Mesnard et al., 2024), and Llama-3 (AI@Meta, 2024).

While previous efforts have demonstrated the effectiveness of in-context learning through careful example selection, this approach assumes that relevant in-context examples are readily available. However, in tasks such as machine translation (MT), especially for low-resource languages, this assumption often does not hold, leading to challenges in leveraging LLMs effectively. In this paper, we address this limitation by focusing on

Findings of the Association for Computational Linguistics: NAACL 2025, pages 4229–4256
April 29 - May 4, 2025

unsupervised machine translation (UMT) in scenarios where in-context examples are not available. Recent studies have tackled UMT by leveraging monolingual data available in the source and target languages (Lample et al., 2018a; Artetxe et al., 2019) using techniques such as alignment of word embeddings (Lample et al., 2018a) and back-translation (Sennrich et al., 2016; Lu and Zhang, 2024). In contrast, our approach assumes access to only a small portion of unlabeled data in each language (less than 1,000 sentences) and a multilingual LLM. To tackle the UMT problem, we divide it into two stages: (i) **a word-level translation stage**, where we use the LLM to generate high-quality word translations that we exploit to create synthetic word-by-word translated parallel data; and (ii) **a sentence-level translation stage**, where we leverage the synthetic parallel data to create better in-context examples that we can use to translate the test input. Additionally, we study several methods for in-context example selection from the unsupervised mined sentence pairs.

We experiment with two multilingual LLMs as our base testbeds, namely, Llama-3 (8B) (AI@Meta, 2024) and Bloom (7B) (Scao et al., 2022). We conduct extensive experiments on the FLORES-200 (Team et al., 2022) evaluation sets, considering three major differences between languages: language *resource level*, language *family*, and language *script*. To summarize, we offer the following contributions:

- We propose a two-stage unsupervised method to extract in-context examples for any language pair. Initially, our approach involves word-level translation, which is subsequently expanded to sentence-level translation.

- We propose an input-specific in-context example selection method for unsupervised mined sentence pairs, which combines sentence embedding similarity scores and BM25 (Trotman et al., 2014) to identify the most relevant and informative examples for each test input sentence.

- We perform extensive evaluations on 288 language directions covering various *resource levels* (High, Medium, Low, and Very Low), *scripts* (e.g., Latin, Arabic, Hangul), and *language families* (e.g., Romance, Afro-Asiatic, Turkic) of the source and target languages.

- We compare our method to robust baselines such as existing state-of-the-art UMT methods and regular (and BM25) k-shot ICL MT. Our unsupervised method not only outperforms previous UMT approaches but also achieves results comparable to, or surpassing, those of the regular ICL method in most of the experiments we conduct.

- We conduct an extensive analysis to explore the factors influencing UMT performance, focusing on resource availability, language family, and script. Additionally, we examine various linguistic features, including language distances, uncovering notable correlations between these features and UMT performance.

2 Related work

Unsupervised Word Translation. Bilingual lexicon induction (BLI) is a crucial task for UMT (Artetxe et al., 2018c; Marchisio et al., 2020). It is the task of inducing word translations across languages from monolingual corpora. BLI approaches can be categorized as either supervised or unsupervised. Supervised methods rely on a seed dictionary, monolingual word embeddings, and an alignment algorithm to learn the alignment between the source and target languages (Artetxe et al., 2016, 2018a). In contrast, unsupervised methods rely solely on monolingual word embeddings and an unsupervised alignment algorithm (Lample et al., 2018a; Artetxe et al., 2018b). Recent work has started exploring the use of multilingual LLMs in BLI by prompting or fine-tuning them to translate words from one language to another (Li et al., 2023c; El Mekki et al., 2023; Li et al., 2024), leading to results better than relying on static word embeddings alone. More recently, Li et al. (2024) proposed a fully unsupervised method that leverages an in-context seed dictionary to translate words using multilingual LLMs.

Unsupervised Machine Translation (UMT). A key technique that has gained prominence in recent UMT research is back-translation, first introduced by Sennrich et al. (2016). This approach involves translating monolingual data in the target language to create a synthetic parallel corpus for training, which is especially beneficial for language pairs with limited parallel examples. Modern UMT approaches have made remarkable progress by incorporating representation learning (Lample et al., 2018a) and pre-trained generative models (Artetxe et al., 2018c; Song et al., 2019; He et al., 2022; Lu and Zhang, 2024) into the existing framework of BLI and back-translation. These advancements allow UMT models to learn cross-lingual representations and generate translations without the need for parallel supervision. Recently, a study by Zhang et al. (2023) has introduced a novel perspective on UMT by showcasing the potential of LLMs in creating pseudo-parallel examples through prompting leading to better UMT performance.

In-Context Examples Selection. Several methods have been proposed to select optimal in-context examples for language tasks, including MT (Liu et al., 2022; Rubin et al., 2022; Li et al., 2023b; Luo et al., 2023), often relying on BM25 or fine-tuned dense retrievers to find examples similar to the test input. In MT, Agrawal et al. (2023) and Sia and Duh (2023) found that selecting examples with greater n-gram overlap or from the same domain improves translation performance, with longer examples generally yielding better results. However, Zhang et al. (2023) showed that MT performance does not always correlate with the selected in-context examples, noting high variance even with the same examples, and that 1-shot learning can underperform zero-shot approaches. To address example selection challenges, Kumar et al. (2023) proposed a scoring model evaluating sentence pairs based on semantic similarity, translation quality, and perplexity to select top-K in-context examples, but it requires annotated data and tools trained on parallel data. Nguyen et al. (2024) introduced Linguistically-Diverse Prompting (LDP), leveraging LLMs and synthetic exemplars from high-resource languages. For X→English translation, it uses diverse language pairs; for other directions, a two-step back-translation process. They found that using diverse languages for exemplars is more effective than relying on a single related language and helps models identify the correct target language. In contrast, we focus on unsupervised mining of parallel pairs, proposing a method to filter and select optimal examples from generated synthetic data to enhance UMT performance using LLMs.

3 Method

UMT Task Description. We consider two distinct languages, denoted as L_s (source) and L_t (target), each with access to its corresponding most frequent words vocabulary V_s and V_t, respectively. Additionally, we assume access to a set of unlabeled sentences D_U in L_t. Access to a multilingual LLM is also assumed. Our UMT method uses these resources to develop an MT system that translates sentences from L_s to L_t without needing human-labeled parallel data. Furthermore, we assume access to a test set D_T, containing parallel sentences between L_s and L_t, which we use for evaluation. Importantly, *no parallel data are available for the learning phase*. Meanwhile, we also assume access to an unsupervised similarity function $sim(x, y)$ that computes the cosine similarity between sentences x and y, each from a different language. In this work, we focus on UMT using ICL of a multilingual LLM. Our task involves the unsupervised extraction (mining) of k examples, which will be utilized as in-context examples to perform MT between L_s and L_t, as getting these examples is still a challenge for under-represented languages. We now describe our method.

Methodology. Our primary objective is to mine an effective set of k parallel sentences in an unsupervised manner. These k examples can then be used to translate input sentences from the source language L_s in the test set D_T. Our approach consists of two main phases: (i) a *word-by-word* translation phase, and (ii) a *sentence-level* translation phase. It begins with performing word-level translation that we can leverage to produce synthetically annotated parallel data, which we use to initialize our sentence-level MT system.

3.1 Word-by-Word Translation Phase

A typical word translation task requires access to a seed dictionary containing word pairs between L_s and L_t. Then, one can use a word embedding mapping algorithm (Artetxe et al., 2016, 2018a) or fine-tune a language model (El Mekki et al., 2023; Li et al., 2023c) to learn the word translation task. However, in the absence of such a dictionary, we rely on LLMs to achieve this. In this section, we describe the process we follow to build our word translation model in an unsupervised manner.

3.1.1 Word Pair Mining

We adopt an approach inspired by Li et al. (2024) to perform unsupervised word translation using LLMs. Given a multilingual LLM, we can mine word pairs from it using zero-shot prompting. To achieve this, we take each word from the source vocabulary V_s and prompt the LLM to translate it from L_s to L_t using the following prompt template:[1]

"The L_s word "w_s" in L_t is:",

where L_s, L_t, and w_s are placeholders for the *source* language, *target* language, and the *query* word, respectively.

Following a random sampling decoding (we discuss the use of other decoding strategies in Figure 7, Appendix B.2), the LLM can generate n sequences (generation stops upon reaching the space token)[2] ranked by their sequence scores. Then, we filter out the predicted translations that are not included in the vocabulary V_t. This step generates a pool of word pairs $P_{s\to t}$, containing a maximum of $n \times |V_s|$ pairs (as some pairs are filtered out). The next step is to reverse the process, going from the target language back to the source language (using greedy sampling decoding), focusing on target words in $P_{s\to t}$. This results in another pool of word pairs, $P_{t\to s}$, from L_t to L_s.

Finally, we identify and retain the best pairs from both $P_{s\to t}$ and $P_{t\to s}$ that accurately represent the word translations between the source language L_s and the target language L_t. Unlike Li et al. (2024), this selection is performed in two steps:

1. Keep pairs $(w_s, \hat{w}_t)$ present in both $P_{s\to t}$ and $P_{t\to s}$, where w_s is the back-translation of $\hat{w}_t$.
2. Rank pairs by cosine similarity scores using $sim(w_s, \hat{w}_t)$. Select top k_{wp} pairs. This sorts pairs by semantic similarity, ensuring a higher quality of word pairs.

3.1.2 k-Shot Word Pair Mining

We replicate the steps from the previous mining process (i.e., in Section 3.1.1), but with two key modifications. First, we substitute zero-shot prompting with k-shot ICL. Second, we employ the top k_{wp} word pairs selected in Section 3.1.1 as examples for the ICL process, essentially bootstrapping the learning from our initial results. This step produces a more refined version of k_{wp} word pairs, leveraging the knowledge gained from the initial selection to improve our word-by-word sentence translation accuracy and reliability in subsequent steps.

[1]We experimented with several prompt templates and found that the presented one yields optimal results. The templates are inspired by Zhang et al. (2023) and Li et al. (2024).

[2]We follow this method as it aligns with established static word embedding alignment methods for BLI, which typically focus on single-word alignments.

3.1.3 Weakly-Annotated Synthetic Parallel Data.

We use the k_{wp} mined word pair as in-context examples to translate words in each sentence of D_T, resulting in a set of sentences D_{w2w} that have been translated word-by-word from L_s to L_t. Although these translations do not maintain accurate word order or grammatical structure, we hypothesize they would preserve the semantic meaning of the original sentences.

3.2 Sentence-Level Translation Phase

In the following, we use D_{w2w} to generate more accurate translations. The process is as follows:

3.2.1 L_t to L_s Back-Translation

Given that the sentences in D_{w2w} are translated on a word-by-word basis, the quality of these translations may be suboptimal compared to the ground truth, with potential issues such as missing words, incorrect word order, or simple inaccuracies in the translation. However, the overall meaning is observably largely preserved. To address these issues, we follow a back-translation step: we reverse the translation direction by using the word-by-word target translations as the source and the original source sentences as the target. We select k instances from D_{w2w}, along with their corresponding original sentences from D_T. These selected sentence pairs will serve as ICL examples in the next step to translate sentences in D_U from L_t to L_s using the LLM, thereby achieving more natural translations than word-by-word ones.

3.2.2 L_s to L_t ICL Example Mining

Following the back-translation step, we acquire a set of sentence pairs that both have natural language and that correct the word-by-word translation word order and grammar issues. Our next step involves using these pairs for ICL to translate sentences in the test set D_T. For that, we propose **TopK+BM25**, a method to select the optimal k examples from these generated pairs. Specifically, for each test example in D_T, we identify the most relevant sentence pairs based on two criteria: (i) selecting pairs where the cosine similarity score (computed using $sim(.,.)$) between the source and target sentences exceeds a specified threshold τ, and (ii) from these selected pairs, choosing the k pairs that yield the highest BM25 (Trotman et al., 2014) similarity score for each test example.

4 Experimental Setup

Models. We perform our experiments using two different models, namely Bloom (7B) (Scao et al., 2022) and Llama-3 (8B) (AI@Meta, 2024).We use their base model versions for all the experiments. For the similarity function $sim(x, y)$, we employ an XLM-R (Conneau et al., 2020) version of Sentence-BERT (Reimers and Gurevych, 2019) to compute embeddings for input texts. We then use cosine similarity to measure the similarity score between embeddings of two sentences.

Datasets. We conduct our evaluation on the FLORES-200 dataset (Team et al., 2022), which covers 200 languages.[3] Each language has a test set of 1,012 examples (we use it as our test set D_T) and a validation set of 997 examples (we use it as our unlabeled sentences D_U). We filter out these languages to end up with 64 languages (listed in Table 7 in Appendix A.1) on which we evaluate our approach. For filtering, we keep languages for which we have access to a list of vocabularies (V_s and V_t from Section 3). From our 64 languages, we create 288 directions by forming different language pair combinations. Our combinations cover four main scenarios: (i) *English-centric* experiments, with English as source or as target; (ii) *language-family-centric* experiments (12 language families); (iii) *script-centric* experiments (14 language scripts), and (iv) *resource-level-centric* experiments (from very low-resource to high-resource languages). For vocabularies per language, V_s and V_t, we utilize FastText embedding models (Grave et al., 2018). These models contain up to 200,000 frequency-sorted words per language; we select the top 10,000 for each language.

Baselines and Comparisons. We compare our method to a wide range of baselines involving ICL, as follows (comparisons to UMT SOTA in 5.2):

- ***Zero*-Shot**: The LLM generates translations without any examples.
- **k-Shot (Regular ICL)**: The LLM uses k random example translations from the validation set as prompts.
- **k-Shot (Regular BM25 ICL)**: The LLM uses k example translations from the validation set as prompts. These examples are selected specifically for each input using BM25.
- **Unsupervised Word-by-Word (UW2W)**: Translates individual words using the word translation model from Section 3.1.3.
- **Unsupervised k-Shot (Random)**: Uses k randomly sampled mined sentence pairs (Section 3.2.2) as prompts.
- **Unsupervised k-Shot (TopK)**: Employs k mined sentence pairs with the highest cosine similarity scores (Section 3.2.2) as prompts.
- **Unsupervised k-Shot ICL (TopK+BM25)**: Selects the top k mined sentence pairs sorted by cosine similarity and BM25 (Section 3.2.2) as prompts.

For simplicity and reproducibility of the random selection, we select the first k examples and consider them as randomly sampled.

Evaluation Metrics. We use two metrics that reflect performance across various levels, following Team et al. (2022): (i) spBLEU (Goyal et al., 2022), a variant of BLEU (Papineni et al., 2002), and (ii) chrF++ (Popović, 2017). spBLEU accounts for the morphological richness of languages by tokenizing sentences into subwords before computing the BLEU score, while chrF++ evaluates character n-gram matches and has been shown to correlate with human annotations, particularly for low-resource languages (Popović, 2017).

Implementation Details. We use vLLM (Kwon et al., 2023) for LLMs inference in our experiments. For the word translation phase, we set $k_{wp} = 10$ as the number of k_{wp} word pairs to mine for ICL word translation; and for the iteration from source to target language, we generate ten word translations for each source word. For the sentence translation phase, we set $k = 8$ as the number of k in-context sentence pair examples.[4] Concerning the threshold τ, we set it as 0.90.[5] If this criterion is not met, we select the top 20 pairs and apply BM25 selection from them. Our experiments are conducted on a server with two A100 (40GB) GPUs. To make our inference more efficient, we used prefix caching as most of the prompt instructions are redundant.

[3] https://github.com/openlanguagedata/flores

[4] Based on the experiments conducted with various values of k (presented in Figure 5 in B.2), $k = 8$ was the optimal.

[5] Experiments were conducted with τ values ranging from 0.0 to 0.90 (presented in Table 10 in Appendix B.2). Values between 0.8 and 0.9 yielded optimal spBLEU scores.

5 Results and Analysis

5.1 Results

Table 1 shows a subset of our Llama-3 experimental results (full results in Table 8, Appendix B.1). It focuses on language pairs with English as the source or target, covering four resource levels: high, medium, low, and very low. Three non-English languages per level are randomly selected. Our unsupervised TopK+BM25 method achieves an average spBLEU of 55.76 on this subset, outperforming the k-shot regular ICL (55.07 spBLEU) and on par with the k-shot regular BM25 ICL (56.93 spBLEU), without relying on human-annotated data.

Additionally, the results indicate that our *UW2W* translation method (40.82 spBLEU) outperforms *zero*-shot MT (37.79 spBLEU). Furthermore, our *TopK+BM25* selection algorithm further enhances the results of these *UW2W* translations compared to the *Random ICL* and *TopK ICL*. For instance, in the English to Italian (*eng_Latn* $\rightarrow$ *ita_Latn*) translation, our *TopK+BM25* achieves a 68.34 spBLEU. This approach outperforms the UW2W method, *Random ICL*, *TopK ICL*, and the k-shot regular BM25 method, which achieve 54.77, 66.91, 66.57, and 64.70 spBLEU scores, respectively.

The full results for Llama-3 in Table 8 in Appendix B.1 show that our approach achieves an average spBLEU of 41.75 and a chrF++ of 28.30. In comparison, the regular BM25-Based k-shot method achieves 42.19 spBLEU and 28.89 chrF++. For Bloom, the results are provided in Table 9 in Appendix B.1. Consistent with the trend observed in Llama-3, Bloom achieves an average spBLEU score of 25.34 and a chrF++ score of 15.29. These results surpass those of the regular k-shot, which achieves an average spBLEU of 23.17 and a chrF++ of 14.95.

5.2 Analysis

How does our TopK+BM25 compare to existing UMT approaches? Using the WMT-14 and WMT-16 benchmarks covering English-French and English-German, respectively, we directly compare our TopK+BM25 (with LLaMA-3) to state-of-the-art UMT methods—including Statistical Machine Translation (SMT), Neural Machine Translation (NMT), and hybrid SMT+NMT methods—as well as strong UMT baselines built on top of XLM (see Table 2). On average, our method achieves a BLEU score of 40.13, outperforming all the previous UMT baselines including the XLM-based ones by an average of 7 BLEU score points. The best baseline is XLM with Online Self-Training followed by self-correction (Lu and Zhang, 2024), which achieves an average BLEU score of 33.68. Specifically, the same trend is observed for all pairs that we evaluate. This shows the strength of our method compared to leading UMT models.

How does our TopK+BM25 compare to out-of-domain (OOD) k-shot? To simulate a more realistic use case of our method. We evaluate the performance achieved using our TopK+BM25 method when applied on unlabeled in-domain texts and compare it to regular k-shot when the shots are coming from out-of-domain (OOD) pairs. The parallel texts used in this experiment come from the following datasets: Flores-200, OPUS-100 (Zhang et al., 2020), WMT-16, and WebCrawl African (Vegi et al., 2022). Table 3 presents the results of this experiment. Results show that our method consistently outperforms the baseline shots in the majority of performed experiments. For example, when the test data is from Flores Dataset and OOD is from WebCrawl African, the performance when choosing in-domain shots for Yoruba to English is 39.29 spBLEU, while OOD shots result in 34.90, widely underperforming our unsupervised method which achieves 40.54 spBLEU. The same scenario applies to the opposite direction, English to Yoruba, where the in-domain shots achieve 26.23 spBLEU, outperforming our unsupervised method which achieves 22.63 spBLEU. However, our method still largely outperforms the OOD shots method, which achieves 14.10 spBLEU.

Does the language resource level affect the UMT performance? Figure 1 shows the mean spBLEU scores from the TopK+BM25 experiments by language resource level. The results demonstrate that the performance of our unsupervised approach is influenced by both the source and target language resource levels, with generally higher scores when the *target* language has high resource levels. Notably, the average spBLEU is highest when both languages are high-resource, reaching a score of 70.39. Conversely, performance decreases as the resource level of the target language diminishes. Unsurprisingly, the lowest scores are observed when both source and target are very low-resource (25.57 spBLEU). This trend highlights the impact of resource level on MT performance.

Resource Level	Language	Direction	Baselines *Zero-Shot*	Baselines *k*-Shot	Baselines BM25	Ours (Unsupervised) UW2W	Ours (Unsupervised) Random	Ours (Unsupervised) TopK	Ours (Unsupervised) TopK+BM25
High	ita_Latn	→	56.64	64.02	64.70	54.77	66.91	66.57	**68.34**
		←	55.48	71.55	73.10	49.79	73.63	70.51	**73.91**
	rus_Cyrl	→	41.95	56.81	60.76	45.98	58.54	59.64	**63.47**
		←	55.15	68.93	71.53	34.09	72.09	71.81	**74.38**
	kor_Hang	→	22.32	39.77	40.47	28.94	37.68	40.94	**41.23**
		←	37.08	60.61	63.73	49.51	61.94	41.19	**67.74**
Medium	arb_Arab	→	13.28	33.11	38.40	**48.23**	32.01	30.67	36.85
		←	43.19	66.92	71.84	48.23	66.34	72.53	**74.35**
	nld_Latn	→	55.75	65.10	63.08	57.46	66.98	67.30	**67.86**
		←	63.24	70.75	69.83	56.95	71.94	**73.91**	73.36
	cat_Latn	→	60.83	69.05	70.84	57.92	69.90	69.75	**72.95**
		←	53.65	76.82	78.96	55.52	77.62	77.40	**80.28**
Low	zsm_Latn	→	53.66	68.85	70.88	57.41	69.82	70.55	**71.20**
		←	54.11	73.96	75.83	57.25	70.17	71.11	**77.07**
	amh_Ethi	→	4.72	14.48	**17.62**	7.23	12.43	4.92	5.50
		←	18.94	34.52	**42.68**	20.99	33.13	12.32	21.48
	som_Latn	→	9.02	26.31	**32.10**	29.64	22.35	23.03	23.02
		←	22.30	35.44	**40.37**	33.25	31.07	26.12	34.23
Very Low	bel_Cyrl	→	17.19	36.40	37.76	29.40	43.50	42.17	**45.39**
		←	46.66	60.52	62.47	38.16	**65.18**	57.05	65.08
	asm_Beng	→	8.16	34.79	**37.29**	5.24	28.88	5.42	5.54
		←	26.14	54.26	**59.31**	6.22	45.63	32.28	50.92
	oci_Latn	→	26.96	58.37	**62.22**	53.25	59.61	61.83	61.91
		←	60.48	80.42	**82.58**	54.38	81.43	63.83	82.13
Average			37.79	55.07	**56.93**	40.82	54.95	50.54	55.76

Table 1: spBLEU scores for a subset of language pairs involving English as either the source (→) or target (←) language, using the Flores-200 dataset across various resource levels. The results compare our UMT approaches to Zero-Shot and regular k-shot baselines using Llama-3. The highest scores for each language pair are highlighted in bold. The "Average" row displays the mean score for all included pairs. (Full results in Table 8 in Appendix B.1).

Type	Method	WMT-14 En→Fr	WMT-14 Fr→En	WMT-16 En→De	WMT-16 De→En	Avg
NMT	Lample et al. (2018b)	25.10	24.20	17.20	21.00	21.88
SMT	Lample et al. (2018b)	28.10	27.20	17.90	22.90	24.03
	Artetxe et al. (2019)	27.80	27.90	19.40	24.80	24.98
SMT+NMT	Lample et al. (2018b)	27.60	27.70	20.20	25.20	25.18
	Artetxe et al. (2019)	33.60	33.20	26.40	33.80	31.75
XLM	He et al. (2022)	36.40	34.30	28.30	34.50	33.38
	Lu and Zhang (2024)	36.90	34.70	28.30	34.80	33.68
Ours	TopK+BM25	**42.47**	**39.78**	**35.04**	**43.26**	**40.13**

Table 2: Comparison of our UMT method (using Beam Search Decoding) against existing UMT methods relying on NMT and SMT. The baseline results presented in the table are BLEU scores reported in their paper. Our BLEU scores were computed using the `mteval-v13a.pl` script.

Pair	Test Data	OOD Data	ID shots	OOD shots	Ours
asm → eng	Flores	OPUS	**54.26**	52.82	50.92
bel → eng			60.52	63.57	**65.08**
bul → eng			73.23	71.29	**77.52**
ben → eng			**62.15**	61.13	57.13
cat → eng			76.82	75.99	**80.28**
deu → eng			77.23	73.12	**79.14**
ell → eng			72.13	68.59	**73.86**
eng → gle			38.00	30.21	**38.90**
yor → eng	Flores	WebCrawl African	39.29	34.90	**40.54**
eng → yor			**26.23**	14.10	22.63
eng → deu	Flores	WMT-16	64.99	60.34	**67.82**
eng → deu	WMT-16	Flores	64.37	62.31	**68.03**

Table 3: spBLEU scores of our UMT method compared to in-domain (ID) and out-of-domain (OOD) regular k-shot.

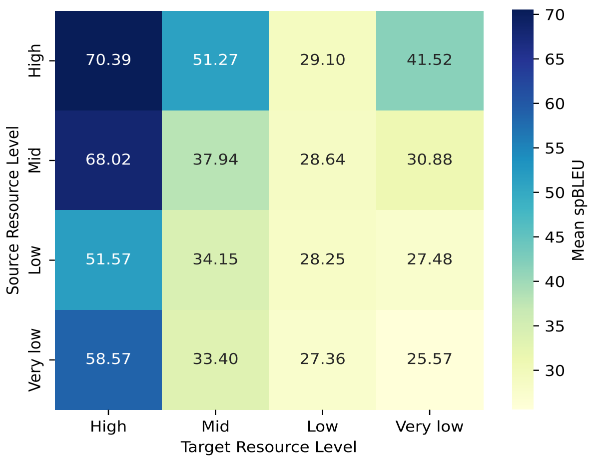

Figure 1: Mean spBLEU TopK+BM25 scores using Llama-3, showing MT performance across resource levels. (Bloom results in Appendix B.2, Figure 3).

Does the language family and script affect ICL performance in UMT? Tables 4 and 5 present the average spBLEU scores for each language family and script when used as the target language in our approach (TopK+BM25), respectively. Results from the language family experiments indicate high variance in performance among different language families. For instance, the performance is optimal for Romance and Germanic languages. These groups comprise most of the Indo-European languages, which are generally well-resourced. Meanwhile, performance declines when the target languages belong to the Nilotic+Other AC or Indo-Aryan language families. These families primarily include African and South Asian languages, many of which are considered low to very-low resourced. Similarly, performance per script results also show high variance, where languages written in Latin script have a better average spBLEU and under-represented scripts such as Ge'ez and Bengali have lower scores.

Afro-Asiatic 25.90	**Austronesian** 49.50	**Balto-Slavic** 50.30
Bantu 24.60	**Germanic** 63.00	**Indo-Aryan** 23.10
Nilotic+Other AC 17.30	**Romance** 60.60	**Turkic** 31.90
Uralic 45.60		

Table 4: Mean spBLEU TopK+BM25 scores using Llama-3, showing MT performance across language families (as targets).

Arabic 24.00	**Armenian** 30.60	**Bengali** 21.00	**Cyrillic** 34.90
Devanagari 27.70	**Georgian** 23.60	**Ge'ez** 5.30	**Greek** 33.20
Han, Hiragana, Katakana 37.20	**Hangul** 24.90	**Hebrew** 51.10	**Latin** 51.20
Perso-Arabic 23.80	**Thai** 43.20		

Table 5: Mean spBLEU TopK+BM25 scores using Llama-3, showing MT performance across language scripts (as targets).

Which linguistic features impact UMT performance? We analyze the influence of several linguistic features on our UMT performance. These features are categorized as follows:

LLM-Dependent: (i) UW2W spBLEU: The performance of the UW2W translation, reflecting the importance of the word-level translation phase in our method. **(ii) Subword Overlap:** After tokenization, the rate of subwords shared between the source and target languages' parallel data. **(iii)Target Language Compression Ratio (TLCR):** Reflects the compression ratio of the target language and addresses the over-segmentation issue in morphologically rich languages.

LLM-Independent: (i) Word Overlap: The rate of words shared between the source and target languages' parallel data. **(ii) Linguistic Distances:** Four distances from Lin et al. (2019), queried from the URIEL Typological Database (Littell et al., 2017): ***Geographic Distance:*** The orthodromic distance between languages on Earth's surface, divided by the antipodal distance, ***Genetic Distance:*** The genealogical distance between languages, ***Syntactic Distance:*** The cosine distance between feature vectors derived from the syntactic structures of the languages, and ***Phonological Distance:*** The cosine distance between phonological feature vectors of the languages.

Table 6 presents Pearson correlation coefficients between these features and UMT scores using the TopK+BM25 approach. Results are grouped by resource level (RL): high and medium (HM), and low and very low (LVL). Since both Bloom and LLaMa-3 models show similar trends, we focus on LLaMa-3. UW2W spBLEU exhibits the strongest positive correlation with UMT performance (0.71 for HM, 0.78 for LVL), underscoring the importance of the initial word-level translation phase. Subword overlap shows a moderate positive correlation (0.13 for HM, 0.39 for LVL), indicating that languages sharing more subwords achieve better UMT performance. Word overlap also demonstrates a positive correlation (0.37 for HM, 0.57 for LVL), suggesting that shared vocabulary improves UMT tasks. The target language compression ratio (TL CR) has a negative correlation (-0.37 for HM, -0.37 for LVL), implying better UMT performance for languages with less complex morphology or better tokenizer vocabulary coverage.

Among linguistic distances, genetic distance shows the strongest correlation, especially for low-resource languages (-0.42 for LVL), while its impact is less pronounced for high and medium-resource languages (-0.21 for HM). Syntactic distance has a moderate negative correlation for high-resource languages (-0.36 for HM) but is not significant for low-resource languages (-0.13 for

LVL). Other distances—geographic, and phonological—exhibit weak or insignificant correlations with UMT.

	Feature	RL	Bloom	LLaMa-3
LLM-Dependent	UW2W spBLEU	HM	0.85 (0.0)	0.71 (0.0)
		LVL	0.89 (0.0)	0.78 (0.0)
	Subword Overlap	HM	0.18 (0.03)	0.13 (0.09)
		LVL	0.54 (0.0)	0.39 (0.0)
	TL CR	HM	-0.31 (0.0)	-0.37 (0.0)
		LVL	-0.62 (0.0)	-0.37 (0.0)
LLM-Independent	Word Overlap	HM	0.6 (0.0)	0.37 (0.0)
		LVL	0.81 (0.0)	0.57 (0.0)
	Geographic	HM	-0.19 (0.02)	-0.17 (0.02)
		LVL	-0.23 (0.11)	-0.2 (0.12)
	Genetic	HM	-0.34 (0.0)	-0.21 (0.0)
		LVL	-0.4 (0.0)	-0.42 (0.0)
	Syntactic	HM	-0.18 (0.04)	-0.36 (0.0)
		LVL	-0.18 (0.21)	-0.13 (0.31)
	Phonological	HM	0.02 (0.77)	-0.16 (0.03)
		LVL	0.0 (0.99)	-0.08 (0.51)

Table 6: Pearson correlation (p-value) values of our UMT model spBLEU score using Bloom and LLaMa-3 and different linguistic features.

6 Conclusion

In this work, we proposed an unsupervised approach to mine in-context examples for machine translation, enabling effective translation across a wide range of languages, including different language families, scripts, and resource-level settings. Our method combines word-level and sentence-level translation, along with a filtering criterion to select optimal examples using multilingual sentence embedding similarity and BM25. Evaluations using Llama-3 and Bloom on 288 directions from Flores-200 demonstrated that our approach achieved better or comparable performance to regular MT ICL, highlighting its potential to overcome the limitations of access to parallel examples in various language pairs. Our analysis shows the impact of several factors such as resource level, script, language family, and different linguistic distances on the performance of our approach. This work showcases the effectiveness of mining in-context examples for improved translation performance especially for under-represented languages, opening up new possibilities for UMT.

Limitations

While our unsupervised approach demonstrates competitive performance compared to the regular baselines, it is important to highlight the following limitations:

- **Dependency on LLM Language Coverage:** The results presented in this paper indicate that the performance of both regular and unsupervised k-shot ICL for MT is predominantly correlated with the resource level of the target language. This correlation poses significant challenges in developing reliable MT systems for low-resource languages. Primarily, the lack of extensive pre-training corpora for these languages results in poorer generation performance compared to that for high-resource languages. Our main focus in this work was to achieve MT performance comparable to the k-shot performance without the need for human-annotated shots.

- **Dependency on Multilingual Embedding Representation for Filtering.** Although our TopK+BM25 algorithm demonstrates effective performance in selecting semantically similar pairs using the similarity function $sim(x, y)$, it depends on XLM-R embeddings. XLM-R does not support all languages, resulting in noisy and inaccurate representations for several languages, particularly those that are low-resource.

- **Time Complexity for Unsupervised Example Mining:** Our unsupervised method for example mining employs a two-step process—first identifying in-context word pairs and then finding in-context sentence pairs—which requires more computational effort than the regular method. However, our method is computationally efficient compared to other UMT methods (Artetxe et al., 2019; Lu and Zhang, 2024). We also use efficient techniques such as prefix-caching and random sampling (instead of beam search) on two A100 (40GB) GPUs to optimize performance and minimize processing time.

- **Challenges in Optimizing Prompt Templates and In-Context Example Ordering:** The effectiveness of ICL is influenced by the design of prompt templates and the structure and ordering of in-context examples. Although we follow the best practices established in previous work (Zhang et al., 2023), identifying the optimal combination for each language pair and translation direction remains a significant challenge.

Ethics Statement

This research aims to enhance language technology by addressing lexical disparities among languages, groups, and cultures. We focus on utilizing LLMs through the unsupervised mining of in-context sentence pairs, which facilitates sentence-level translation across languages.

Our study includes 64 languages, representing 288 language pairs from 12 language families and 14 scripts, across various resource levels. The goal is to extend unsupervised machine translation to these languages using zero-shot techniques. Ultimately, this work seeks to increase access to technology for diverse populations.

The dataset employed in our research, Flores-200, is publicly available and, in our assessment, poses no risks. However, for any real-world application, we recommend conducting thorough evaluations and analyses before deployment.

Acknowledgments

We acknowledge support from Canada Research Chairs (CRC), the Natural Sciences and Engineering Research Council of Canada (NSERC; RGPIN-2018-04267, RGPIN-2020-05408), the Social Sciences and Humanities Research Council of Canada (SSHRC; 895-2020-1004; 895-2021-1008), Canadian Foundation for Innovation (CFI; 37771), Digital Research Alliance of Canada,[6] and UBC Advanced Research Computing-Sockeye.[7]

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.

AI@Meta. 2024. Llama 3 model card.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2016. Learning principled bilingual mappings of word embeddings while preserving monolingual invariance. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2289–2294, Austin, Texas. Association for Computational Linguistics.

[6]https://alliancecan.ca

[7]https://arc.ubc.ca/ubc-arc-sockeye

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2018a. Generalizing and improving bilingual word embedding mappings with a multi-step framework of linear transformations. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, AAAI'18/IAAI'18/EAAI'18. AAAI Press.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2018b. A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 789–798, Melbourne, Australia. Association for Computational Linguistics.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2018c. Unsupervised statistical machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3632–3642, Brussels, Belgium. Association for Computational Linguistics.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2019. An effective approach to unsupervised machine translation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 194–203, Florence, Italy. Association for Computational Linguistics.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA. Curran Associates Inc.

Manish Chandra, Debasis Ganguly, Yiwen Li, and Iadh Ounis. 2024. 'one size doesn't fit all': Learning how many examples to use for in-context learning for improved text classification. *Preprint*, arXiv:2403.06402.

Ting-Yun Chang and Robin Jia. 2023. Data curation alone can stabilize in-context learning. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8123–8144, Toronto, Canada. Association for Computational Linguistics.

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

Abdellah El Mekki, Muhammad Abdul-Mageed, ElMoatez Billah Nagoudi, Ismail Berrada, and Ahmed Khoumsi. 2023. ProMap: Effective bilingual lexicon induction via language model prompting. 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 577–597, Nusa Dua, Bali. Association for Computational Linguistics.

Mor Geva, Daniel Khashabi, Elad Segal, Tushar Khot, Dan Roth, and Jonathan Berant. 2021. Did Aristotle Use a Laptop? A Question Answering Benchmark with Implicit Reasoning Strategies. *Transactions of the Association for Computational Linguistics*, 9:346–361.

Naman Goyal, Cynthia Gao, Vishrav Chaudhary, Peng-Jen Chen, Guillaume Wenzek, Da Ju, Sanjana Krishnan, Marc'Aurelio Ranzato, Francisco Guzmán, and Angela Fan. 2022. The Flores-101 evaluation benchmark for low-resource and multilingual machine translation. *Transactions of the Association for Computational Linguistics*, 10:522–538.

Edouard Grave, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, and Tomas Mikolov. 2018. Learning word vectors for 157 languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Zhiwei He, Xing Wang, Rui Wang, Shuming Shi, and Zhaopeng Tu. 2022. Bridging the data gap between training and inference for unsupervised neural machine translation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6611–6623, Dublin, Ireland. Association for Computational Linguistics.

Haoyang Huang, Tianyi Tang, Dongdong Zhang, Xin Zhao, Ting Song, Yan Xia, and Furu Wei. 2023. Not all languages are created equal in LLMs: Improving multilingual capability by cross-lingual-thought prompting. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 12365–12394, Singapore. Association for Computational Linguistics.

Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. 2020. Scaling laws for neural language models. *CoRR*, abs/2001.08361.

Md Tawkat Islam Khondaker, Abdul Waheed, Muhammad Abdul-Mageed, et al. 2023. Gptaraeval: A comprehensive evaluation of chatgpt on arabic nlp. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 220–247.

Aswanth Kumar, Ratish Puduppully, Raj Dabre, and Anoop Kunchukuttan. 2023. CTQScorer: Combining multiple features for in-context example selection for machine translation. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 7736–7752, Singapore. Association for Computational Linguistics.

Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph Gonzalez, Hao Zhang, and Ion Stoica. 2023. Efficient memory management for large language model serving with pagedattention. In *Proceedings of the 29th Symposium on Operating Systems Principles*, SOSP '23, page 611–626, New York, NY, USA. Association for Computing Machinery.

Guillaume Lample, Alexis Conneau, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. 2018a. Word translation without parallel data. In *International Conference on Learning Representations*.

Guillaume Lample, Myle Ott, Alexis Conneau, Ludovic Denoyer, and Marc'Aurelio Ranzato. 2018b. Phrase-based & neural unsupervised machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 5039–5049, Brussels, Belgium. Association for Computational Linguistics.

Tianle Li, Xueguang Ma, Alex Zhuang, Yu Gu, Yu Su, and Wenhu Chen. 2023a. Few-shot in-context learning on knowledge base question answering. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6966–6980, Toronto, Canada. Association for Computational Linguistics.

Xiaonan Li, Kai Lv, Hang Yan, Tianyang Lin, Wei Zhu, Yuan Ni, Guotong Xie, Xiaoling Wang, and Xipeng Qiu. 2023b. Unified demonstration retriever for in-context learning. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4644–4668, Toronto, Canada. Association for Computational Linguistics.

Yaoyiran Li, Anna Korhonen, and Ivan Vulić. 2023c. On bilingual lexicon induction with large language models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 9577–9599, Singapore. Association for Computational Linguistics.

Yaoyiran Li, Anna Korhonen, and Ivan Vulić. 2024. Self-augmented in-context learning for unsupervised word translation. *Preprint*, arXiv:2402.10024.

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.

Yu-Hsiang Lin, Chian-Yu Chen, Jean Lee, Zirui Li, Yuyan Zhang, Mengzhou Xia, Shruti Rijhwani, Junxian He, Zhisong Zhang, Xuezhe Ma, Antonios Anastasopoulos, Patrick Littell, and Graham Neubig. 2019. Choosing transfer languages for cross-lingual learning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3125–3135, Florence, Italy. Association for Computational Linguistics.

Patrick Littell, David R. Mortensen, Ke Lin, Katherine Kairis, Carlisle Turner, and Lori Levin. 2017. URIEL and lang2vec: Representing languages as typological, geographical, and phylogenetic vectors. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 8–14, Valencia, Spain. Association for Computational Linguistics.

Jiachang Liu, Dinghan Shen, Yizhe Zhang, Bill Dolan, Lawrence Carin, and Weizhu Chen. 2022. What makes good in-context examples for GPT-3? In *Proceedings of Deep Learning Inside Out (DeeLIO 2022): The 3rd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures*, pages 100–114, Dublin, Ireland and Online. Association for Computational Linguistics.

Jinliang Lu and Jiajun Zhang. 2024. Improving unsupervised neural machine translation via training data self-correction. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 8942–8954, Torino, Italia. ELRA and ICCL.

Yao Lu, Max Bartolo, Alastair Moore, Sebastian Riedel, and Pontus Stenetorp. 2022. Fantastically ordered prompts and where to find them: Overcoming few-shot prompt order sensitivity. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8086–8098, Dublin, Ireland. Association for Computational Linguistics.

Man Luo, Xin Xu, Zhuyun Dai, Panupong Pasupat, Mehran Kazemi, Chitta Baral, Vaiva Imbrasaite, and Vincent Y Zhao. 2023. Dr.icl: Demonstration-retrieved in-context learning. *Preprint*, arXiv:2305.14128.

Kelly Marchisio, Kevin Duh, and Philipp Koehn. 2020. When does unsupervised machine translation work? In *Proceedings of the Fifth Conference on Machine Translation*, pages 571–583, Online. Association for Computational Linguistics.

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, et al. 2024. Gemma: Open models based on gemini research and technology. *Preprint*, arXiv:2403.08295.

Aristides Milios, Siva Reddy, and Dzmitry Bahdanau. 2023. In-context learning for text classification with many labels. In *Proceedings of the 1st GenBench Workshop on (Benchmarking) Generalisation in NLP*, pages 173–184, Singapore. Association for Computational Linguistics.

Xuan-Phi Nguyen, Mahani Aljunied, Shafiq Joty, and Lidong Bing. 2024. Democratizing LLMs for low-resource languages by leveraging their English dominant abilities with linguistically-diverse prompts. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3501–3516, Bangkok, Thailand. Association for Computational Linguistics.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. *Preprint*, arXiv:2203.02155.

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.

Maja Popović. 2017. chrF++: words helping character n-grams. In *Proceedings of the Second Conference on Machine Translation*, pages 612–618, Copenhagen, Denmark. Association for Computational Linguistics.

Alec Radford and Karthik Narasimhan. 2018. Improving language understanding by generative pre-training.

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing* and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3982–3992, Hong Kong, China. Association for Computational Linguistics.

Ohad Rubin, Jonathan Herzig, and Jonathan Berant. 2022. Learning to retrieve prompts for in-context learning. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2655–2671, Seattle, United States. Association for Computational Linguistics.

Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, Roman Castagné, Alexandra Sasha Luccioni, François Yvon, Matthias Gallé, et al. 2022. Bloom: A 176b-parameter open-access multilingual language model. *arXiv preprint arXiv:2211.05100*.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Improving neural machine translation models with monolingual data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 86–96, Berlin, Germany. Association for Computational Linguistics.

Suzanna Sia and Kevin Duh. 2023. In-context learning as maintaining coherency: A study of on-the-fly machine translation using large language models. In *Proceedings of Machine Translation Summit XIX, Vol. 1: Research Track*, pages 173–185, Macau SAR, China. Asia-Pacific Association for Machine Translation.

Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. 2019. Mass: Masked sequence to sequence pre-training for language generation. In *International Conference on Machine Learning*, pages 5926–5936.

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.

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, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *Preprint*, arXiv:2307.09288.

Andrew Trotman, Antti Puurula, and Blake Burgess. 2014. Improvements to bm25 and language models examined. In *Proceedings of the 19th Australasian Document Computing Symposium*, ADCS '14, page 58–65, New York, NY, USA. Association for Computing Machinery.

Pavanpankaj Vegi, Sivabhavani J, Biswajit Paul, Abhinav Mishra, Prashant Banjare, Prasanna K R, and Chitra Viswanathan. 2022. WebCrawl African : A multilingual parallel corpora for African languages. In *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pages 1076–1089, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Zhiyong Wu, Yaoxiang Wang, Jiacheng Ye, and Lingpeng Kong. 2023. Self-adaptive in-context learning: An information compression perspective for in-context example selection and ordering. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1423–1436, Toronto, Canada. Association for Computational Linguistics.

Biao Zhang, Barry Haddow, and Alexandra Birch. 2023. Prompting large language model for machine translation: a case study. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org.

Biao Zhang, Philip Williams, Ivan Titov, and Rico Sennrich. 2020. Improving massively multilingual neural machine translation and zero-shot translation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1628–1639, Online. Association for Computational Linguistics.

A Experimental Setup

A.1 Datasets

B Results and Analysis

B.1 Results

Lang ID	ISO 639-3	Language	Family	Subgrouping	Script	Resource Level
fra_Latn	fra	French	Indo-European	Romance	Latin	High
deu_Latn	deu	German	Indo-European	Germanic	Latin	High
ita_Latn	ita	Italian	Indo-European	Romance	Latin	High
por_Latn	por	Portuguese (Brazil)	Indo-European	Romance	Latin	High
rus_Cyrl	rus	Russian	Indo-European	Balto-Slavic	Cyrillic	High
spa_Latn	spa	Spanish (Latin America)	Indo-European	Romance	Latin	High
eng_Latn	eng	English (Latin America)	Indo-European	Germanic	Latin	High
arb_Arab	ara	Arabic	Afro-Asiatic	Afro-Asiatic	Arabic	Mid
ben_Beng	ben	Bengali	Indo-European	Indo-Aryan	Bengali	Mid
bul_Cyrl	bul	Bulgarian	Indo-European	Balto-Slavic	Cyrillic	Mid
cat_Latn	cat	Catalan	Indo-European	Romance	Latin	Mid
ces_Latn	ces	Czech	Indo-European	Balto-Slavic	Latin	Mid
dan_Latn	dan	Danish	Indo-European	Germanic	Latin	Mid
nld_Latn	nld	Dutch	Indo-European	Germanic	Latin	Mid
fin_Latn	fin	Finnish	Uralic	Uralic	Latin	Mid
glg_Latn	glg	Galician	Indo-European	Romance	Latin	Mid
kat_Geor	kat	Georgian	Kartvelian	Other	Georgian	Mid
ell_Grek	ell	Greek	Indo-European	Other IE	Greek	Mid
heb_Hebr	heb	Hebrew	Afro-Asiatic	Afro-Asiatic	Hebrew	Mid
hin_Deva	hin	Hindi	Indo-European	Indo-Aryan	Devanagari	Mid
hun_Latn	hun	Hungarian	Uralic	Uralic	Latin	Mid
ind_Latn	ind	Indonesian	Austronesian	Austronesian	Latin	Mid
jpn_Jpan	jpn	Japanese	Japonic	Other	Han, Hiragana, Katakana	Mid
jav_Latn	jav	Javanese	Austronesian	Austronesian	Latin	Mid
kor_Hang	kor	Korean	Koreanic	Other	Hangul	Mid
lvs_Latn	lav	Latvian	Indo-European	Balto-Slavic	Latin	Mid
lit_Latn	lit	Lithuanian	Indo-European	Balto-Slavic	Latin	Mid
ltz_Latn	ltz	Luxembourgish	Indo-European	Germanic	Latin	Mid
mkd_Cyrl	mkd	Macedonian	Indo-European	Balto-Slavic	Cyrillic	Mid
mlt_Latn	mlt	Maltese	Afro-Asiatic	Afro-Asiatic	Latin	Mid
pes_Arab	fas	Persian	Indo-European	Indo-Aryan	Perso-Arabic	Mid
pol_Latn	pol	Polish	Indo-European	Balto-Slavic	Latin	Mid
srp_Cyrl	srp	Serbian	Indo-European	Balto-Slavic	Cyrillic	Mid
slk_Latn	slk	Slovak	Indo-European	Balto-Slavic	Latin	Mid
tha_Thai	tha	Thai	Kra-Dai	Sino-Tibetan+Kra-Dai	Thai	Mid
tur_Latn	tur	Turkish	Turkic	Turkic	Latin	Mid
ukr_Cyrl	ukr	Ukrainian	Indo-European	Balto-Slavic	Cyrillic	Mid
vie_Latn	vie	Vietnamese	Austro-Asiatic	Austro-Asiatic	Latin	Mid
afr_Latn	afr	Afrikaans	Indo-European	Germanic	Latin	Low
amh_Ethi	amh	Amharic	Afro-Asiatic	Afro-Asiatic	Ge'ez	Low
hye_Armn	hye	Armenian	Indo-European	Other IE	Armenian	Low
ast_Latn	ast	Asturian	Indo-European	Romance	Latin	Low
bos_Latn	bos	Bosnian	Indo-European	Balto-Slavic	Latin	Low
ceb_Latn	ceb	Cebuano	Austronesian	Austronesian	Latin	Low
gle_Latn	gle	Irish	Indo-European	Other IE	Latin	Low
kaz_Cyrl	kaz	Kazakh	Turkic	Turkic	Cyrillic	Low
kir_Cyrl	kir	Kyrgyz	Turkic	Turkic	Cyrillic	Low
zsm_Latn	msa	Malay	Austronesian	Austronesian	Latin	Low
mar_Deva	mar	Marathi	Indo-European	Indo-Aryan	Devanagari	Low
khk_Cyrl	mon	Mongolian	Mongolic	Other	Cyrillic	Low
pbt_Arab	pus	Pashto	Indo-European	Indo-Aryan	Perso-Arabic	Low
som_Latn	som	Somali	Afro-Asiatic	Afro-Asiatic	Latin	Low
ckb_Arab	ckb	Sorani Kurdish	Indo-European	Indo-Aryan	Arabic	Low
tgk_Cyrl	tgk	Tajik	Indo-European	Indo-Aryan	Cyrillic	Low
urd_Arab	urd	Urdu	Indo-European	Indo-Aryan	Perso-Arabic	Low
cym_Latn	cym	Welsh	Indo-European	Other IE	Latin	Low
yor_Latn	yor	Yoruba	Atlantic-Congo	Nilotic+Other AC	Latin	Low
asm_Beng	asm	Assamese	Indo-European	Indo-Aryan	Bengali	Very low
bel_Cyrl	bel	Belarusian	Indo-European	Balto-Slavic	Cyrillic	Very low
hrv_Latn	hrv	Croatian	Indo-European	Balto-Slavic	Latin	Very low
nso_Latn	nso	Northern Sotho	Atlantic-Congo	Bantu	Latin	Very low
oci_Latn	oci	Occitan	Indo-European	Romance	Latin	Very low
snd_Arab	snd	Sindhi	Indo-European	Indo-Aryan	Perso-Arabic	Very low
uzn_Latn	uzb	Uzbek	Turkic	Turkic	Latin	Very low

Table 7: The list of languages included in our experiments features several attributes for each language: its corresponding Lang ID from Flores-200, ISO 639-3 code, language name, language family, subgrouping, script, and resource level. This metadata was extracted from Goyal et al. (2022).

Table 8: Detailed evaluation of translation scores for 280 language directions from the Flores-200 dataset using the Llama-3 model. Results include Zero-Shot and regular k-shot baselines, along with our unsupervised method employing various selection approaches. Scores for each language pair are presented as chrF++/spBLEU, with the highest scores highlighted in bold. The 'average' row indicates the mean score across all language pairs.

	Baselines (Zero-shot + Regular ICL)			**Ours (Unsupervised ICL)**			
Pair	***Zero*-Shot**	**k-Shot**	**BM25**	**UW2W**	**Random**	**TopK**	**TopK+BM25**
pol_Latn → afr_Latn	30.15 / 44.47	33.58 / 50.83	33.58 / 50.53	28.94 / 48.57	33.65 / 50.88	**39.25 / 59.09**	38.45 / 58.57
afr_Latn → pol_Latn	26.23 / 36.93	34.51 / 50.95	34.39 / 51.26	25.67 / 45.3	**36.25 / 53.39**	34.06 / 51.35	35.98 / 53.18
amh_Ethi → eng_Latn	11.95 / 18.94	21.86 / 34.52	**26.09 / 42.68**	11.53 / 20.99	19.85 / 33.13	5.85 / 12.32	12.58 / 21.48
eng_Latn → amh_Ethi	1.52 / 4.72	6.33 / 14.48	**7.78 / 17.62**	1.51 / 7.23	5.51 / 12.43	1.05 / 4.92	1.29 / 5.5
amh_Ethi → heb_Hebr	6.03 / 4.83	**12.79** / 20.57	2.44 / 0.01	9.86 / **22.25**	6.9 / 9.34	0.62 / 3.84	0.9 / 4.43
heb_Hebr → amh_Ethi	2.19 / 3.73	**5.21 / 11.71**	2.73 / 2.75	1.04 / 5.69	3.03 / 5.53	1.74 / 6.52	1.65 / 6.39
mlt_Latn → amh_Ethi	1.92 / 3.63	5.76 / 14.38	**7.26 / 17.7**	0.79 / 3.78	1.74 / 6.17	0.89 / 3.98	0.86 / 3.88
amh_Ethi → mlt_Latn	9.11 / 11.8	**13.52** / 20.36	9.88 / 10.94	12.34 / **23.84**	10.55 / 14.53	0.92 / 3.39	12.94 / 21.07
tha_Thai → amh_Ethi	1.82 / 1.4	5.68 / 12.55	**7.34 / 17.38**	0.89 / 3.06	3.12 / 8.15	2.8 / 6.32	2.5 / 7.56
amh_Ethi → tha_Thai	3.26 / 2.17	**10.31** / 15.04	9.04 / 10.91	9.91 / **17.27**	7.28 / 11.01	1.19 / 2.39	5.51 / 8.91
arb_Arab → ckb_Arab	4.9 / 5.89	11.72 / 18.55	**15.18 / 26.78**	11.06 / 23.79	8.72 / 11.0	6.92 / 12.82	7.83 / 14.81
ckb_Arab → arb_Arab	7.27 / 9.68	12.88 / 21.79	16.99 / 29.84	**19.05 / 38.47**	11.99 / 19.39	14.59 / 23.93	13.95 / 23.15
arb_Arab → eng_Latn	27.39 / 43.19	51.67 / 66.92	56.13 / 71.84	28.24 / 48.23	51.84 / 66.34	56.51 / 72.53	**58.2 / 74.35**
eng_Latn → arb_Arab	8.74 / 13.28	20.98 / 33.11	24.23 / 38.4	**25.73 / 48.23**	20.39 / 32.01	19.68 / 30.67	23.35 / 36.85
heb_Hebr → arb_Arab	11.33 / 15.09	**20.12** / 33.09	19.07 / 32.42	20.09 / **38.25**	17.67 / 28.05	9.21 / 16.38	19.88 / 33.47
arb_Arab → heb_Hebr	17.98 / 24.86	26.66 / 42.63	30.5 / 49.09	14.63 / 32.35	27.08 / 43.16	28.52 / 45.72	**31.05 / 50.14**
asm_Beng → ben_Beng	16.93 / 35.91	23.44 / 40.95	1.97 / 0.0	17.67 / 38.61	**23.8 / 44.93**	17.91 / 37.96	22.18 / 43.43
ben_Beng → asm_Beng	17.09 / 35.75	20.26 / 38.35	1.15 / 0.0	17.89 / 38.63	21.23 / 42.15	20.01 / 40.99	**22.65 / 44.24**
asm_Beng → eng_Latn	15.67 / 26.14	39.25 / 54.26	**42.09 / 59.31**	1.58 / 6.22	31.13 / 45.63	20.31 / 32.28	35.24 / 50.92
eng_Latn → asm_Beng	4.94 / 8.16	18.48 / 34.79	**20.21 / 37.29**	1.08 / 5.24	15.77 / 28.88	1.14 / 5.42	1.19 / 5.54
hin_Deva → asm_Beng	8.75 / 13.93	19.24 / 35.96	20.47 / 38.25	1.55 / 7.89	**22.55 / 41.95**	22.5 / 41.22	21.59 / 40.66
asm_Beng → hin_Deva	14.9 / 20.85	22.45 / 35.76	26.29 / 42.68	1.23 / 6.52	**26.57 / 43.2**	10.33 / 18.4	8.33 / 16.94
asm_Beng → hun_Latn	15.75 / 18.34	26.73 / 40.51	**26.96 / 40.97**	1.42 / 5.71	23.17 / 33.21	12.03 / 19.54	23.79 / 36.53
hun_Latn → asm_Beng	3.7 / 4.68	16.01 / 31.45	**16.3 / 31.57**	0.95 / 4.92	15.37 / 29.81	1.29 / 5.63	3.57 / 10.56
asm_Beng → snd_Arab	1.41 / 2.17	8.44 / 15.47	**10.08 / 19.94**	0.76 / 4.7	6.43 / 9.28	9.77 / 18.28	9.13 / 16.53
snd_Arab → asm_Beng	4.32 / 4.78	13.77 / 27.99	**15.84 / 31.4**	2.81 / 9.62	10.52 / 19.93	14.88 / 29.81	13.95 / 28.32
asm_Beng → som_Latn	6.47 / 5.46	12.22 / 16.01	**15.13** / 25.05	1.44 / 5.62	13.35 / 21.21	1.41 / 5.24	14.66 / **25.83**
som_Latn → asm_Beng	3.59 / 4.83	11.98 / 23.53	**13.15 / 25.92**	0.96 / 5.0	9.69 / 17.69	1.04 / 4.73	1.21 / 5.08
asm_Beng → uzn_Latn	8.08 / 7.4	20.41 / 31.54	**21.82 / 34.25**	1.32 / 5.34	18.72 / 27.83	3.01 / 7.82	16.59 / 26.74
uzn_Latn → asm_Beng	4.39 / 5.58	15.58 / 30.77	**16.38 / 32.17**	0.87 / 4.61	13.49 / 26.73	0.9 / 4.39	1.01 / 4.67
eng_Latn → ast_Latn	37.72 / 52.87	47.03 / 65.58	46.28 / 64.1	34.61 / 56.42	48.22 / 66.57	48.83 / 67.26	**49.28 / 67.46**
ast_Latn → eng_Latn	37.72 / 54.96	59.54 / 73.55	60.01 / 74.26	35.4 / 56.28	59.82 / 74.23	44.78 / 61.63	**61.72 / 75.88**
ast_Latn → tgk_Cyrl	4.12 / 4.92	14.54 / 21.11	**17.98 / 27.55**	11.01 / 20.41	13.87 / 20.78	2.15 / 6.37	3.09 / 8.14
tgk_Cyrl → ast_Latn	24.71 / 40.51	27.0 / 42.99	**29.79 / 47.1**	20.88 / 36.32	19.43 / 30.13	4.64 / 10.5	26.39 / 42.42
eng_Latn → bel_Cyrl	11.6 / 17.19	23.03 / 36.4	23.43 / 37.76	15.22 / 29.4	27.46 / 43.5	26.92 / 42.17	**28.41 / 45.39**
bel_Cyrl → eng_Latn	30.87 / 46.66	42.78 / 60.52	44.25 / 62.47	23.17 / 38.16	46.12 / **65.18**	42.32 / 57.05	**47.51** / 65.08
hrv_Latn → bel_Cyrl	17.57 / 24.08	23.09 / 36.43	22.6 / 36.48	21.47 / 38.97	28.86 / 46.11	28.59 / 45.27	**29.05 / 46.64**
bel_Cyrl → hrv_Latn	25.65 / 40.55	30.57 / 48.63	28.03 / 43.68	23.57 / 38.29	34.85 / 53.94	34.6 / 53.95	**35.77 / 54.38**
bel_Cyrl → nso_Latn	10.91 / 18.2	14.75 / 24.37	**18.64 / 34.04**	11.2 / 18.35	13.78 / 21.84	9.02 / 17.96	15.2 / 27.61
nso_Latn → bel_Cyrl	7.16 / 7.83	13.64 / 21.83	13.98 / 22.11	10.98 / 22.64	14.42 / 23.95	**15.08 / 25.86**	14.51 / 24.24
bel_Cyrl → tgk_Cyrl	9.77 / 16.55	15.56 / 25.39	**17.32 / 29.15**	13.9 / 24.98	16.4 / 28.42	13.24 / 24.8	13.86 / 25.75
tgk_Cyrl → bel_Cyrl	13.26 / 20.85	18.52 / 30.38	**20.42 / 34.33**	17.39 / 34.04	17.5 / 28.96	15.97 / 29.37	17.28 / 32.26
eng_Latn → ben_Beng	15.37 / 23.23	29.12 / 47.47	**30.01 / 49.23**	1.24 / 5.62	23.03 / 37.8	1.36 / 5.9	1.55 / 6.35
ben_Beng → eng_Latn	19.01 / 32.31	46.79 / 62.15	**49.23 / 65.64**	2.04 / 7.85	12.69 / 22.25	24.25 / 36.11	42.0 / 57.13
bos_Latn → ukr_Cyrl	29.75 / 38.88	37.07 / 53.93	37.21 / 53.79	27.89 / 47.74	41.2 / 58.77	**43.11 / 61.15**	41.95 / 59.87
ukr_Cyrl → bos_Latn	28.74 / 41.97	38.68 / 56.32	37.25 / 54.72	28.49 / 48.23	42.04 / 60.87	42.52 / 62.02	**43.22 / 62.59**
eng_Latn → bul_Cyrl	27.84 / 37.67	41.95 / 57.94	44.98 / 61.4	29.98 / 50.96	44.73 / 61.44	47.07 / 64.1	**47.44 / 64.31**
bul_Cyrl → eng_Latn	34.92 / 51.87	58.02 / 73.23	60.2 / 75.31	23.02 / 39.69	58.97 / 74.53	52.8 / 67.71	**61.81 / 77.52**
ckb_Arab → cat_Latn	25.37 / 38.13	27.76 / 42.91	**32.74 / 50.74**	19.33 / 34.87	22.02 / 32.91	29.04 / 45.45	30.31 / 46.6
cat_Latn → ckb_Arab	3.85 / 4.96	11.1 / 17.65	**13.95 / 24.28**	0.97 / 2.78	11.51 / 17.77	10.08 / 15.18	12.55 / 21.86
dan_Latn → cat_Latn	39.88 / 54.74	41.68 / 58.21	43.34 / 60.1	32.79 / 53.32	45.15 / 62.26	46.95 / 64.51	**47.3 / 65.02**
cat_Latn → dan_Latn	36.57 / 49.7	41.94 / 58.54	43.3 / 60.3	31.35 / 50.84	43.53 / 60.37	46.19 / 63.12	**46.32 / 64.05**
eng_Latn → cat_Latn	47.03 / 60.83	52.86 / 69.05	54.77 / 70.84	37.11 / 57.92	54.11 / 69.9	54.48 / 69.75	**56.84 / 72.95**
cat_Latn → eng_Latn	38.65 / 53.65	62.85 / 76.82	64.96 / 78.96	38.39 / 55.52	63.2 / 77.62	63.36 / 77.4	**66.25 / 80.28**
cat_Latn → fin_Latn	25.21 / 32.76	30.97 / 44.57	33.63 / 49.01	26.21 / 40.52	29.68 / 42.76	**35.82 / 51.19**	35.16 / 50.57
fin_Latn → cat_Latn	33.61 / 47.38	36.34 / 53.32	38.32 / 55.78	28.81 / 46.67	35.67 / 52.31	**40.59 / 59.09**	40.54 / **59.09**
eng_Latn → ceb_Latn	17.99 / 22.74	32.86 / 49.03	**35.97 / 52.85**	29.23 / 47.88	31.69 / 47.74	34.48 / 51.5	34.06 / 51.47
ceb_Latn → eng_Latn	25.82 / 39.62	51.12 / 66.76	**52.93 / 68.65**	27.45 / 43.2	47.07 / 61.68	29.85 / 42.19	43.33 / 60.15

	Baselines (Zero-shot + Regular ICL)			Ours (Unsupervised ICL)			
Pair	***Zero*-Shot**	***k*-Shot**	**BM25**	**UW2W**	**Random**	**TopK**	**TopK+BM25**
hye_Armn → ceb_Latn	8.33 / 10.76	**23.35 / 36.13**	4.17 / 0.01	11.79 / 20.71	15.37 / 22.98	1.72 / 5.5	12.95 / 23.31
ceb_Latn → hye_Armn	8.39 / 6.31	22.62 / 35.53	**24.28 / 37.38**	2.52 / 9.08	18.94 / 29.74	1.73 / 6.51	8.8 / 17.54
jav_Latn → ceb_Latn	15.91 / 23.47	26.03 / 41.25	**28.98 / 45.93**	16.27 / 28.14	22.18 / 35.67	19.84 / 34.62	21.94 / 37.09
ceb_Latn → jav_Latn	14.38 / 22.96	22.56 / 36.62	**26.81 / 43.06**	21.85 / 37.47	16.5 / 22.99	21.73 / 35.82	22.62 / 37.61
ceb_Latn → kat_Geor	9.8 / 9.51	20.31 / 30.62	**21.61 / 31.92**	4.88 / 12.12	19.91 / 29.91	17.63 / 26.08	20.15 / 29.95
kat_Geor → ceb_Latn	9.47 / 12.18	19.11 / 26.11	0.66 / 0.0	12.9 / 22.27	**19.44 / 31.13**	1.75 / 5.0	7.82 / 13.89
ceb_Latn → zsm_Latn	28.41 / 42.21	37.03 / 53.49	**39.67 / 56.8**	27.95 / 45.45	30.02 / 44.13	34.38 / 50.81	36.69 / 53.67
zsm_Latn → ceb_Latn	16.97 / 23.57	28.09 / 44.36	**31.07 / 48.67**	16.23 / 27.19	26.66 / 42.07	19.83 / 32.9	20.57 / 34.14
kor_Hang → ces_Latn	17.65 / 21.81	25.89 / 39.46	26.02 / 40.01	22.03 / 38.8	21.51 / 30.37	15.2 / 22.25	**27.74 / 42.33**
ces_Latn → kor_Hang	9.09 / 19.33	15.53 / 33.18	15.44 / 32.45	10.52 / 27.47	13.2 / 27.78	**16.9 / 35.56**	16.52 / 35.18
eng_Latn → ckb_Arab	2.74 / 5.04	10.31 / 16.69	**14.03 / 24.66**	9.43 / 23.28	9.48 / 13.91	10.93 / 17.6	11.67 / 20.23
ckb_Arab → eng_Latn	16.83 / 28.4	37.51 / 53.31	**42.51 / 60.19**	22.44 / 41.04	35.76 / 51.38	29.78 / 44.39	40.98 / 58.62
ckb_Arab → nso_Latn	10.92 / 18.28	16.31 / 28.8	**19.11 / 35.49**	11.97 / 21.39	11.16 / 15.07	7.13 / 15.24	12.65 / 24.28
nso_Latn → ckb_Arab	2.53 / 4.45	12.66 / 22.63	**13.62 / 24.17**	7.28 / 17.09	12.58 / 22.24	9.99 / 18.04	10.9 / 17.84
fra_Latn → cym_Latn	18.99 / 26.71	26.73 / 40.72	**29.33 / 44.83**	15.1 / 24.62	23.34 / 35.22	20.84 / 35.12	23.34 / 37.86
cym_Latn → fra_Latn	39.53 / 51.62	47.38 / 63.72	**48.1** / 64.84	25.09 / 44.12	44.82 / 60.5	48.01 / **64.99**	45.32 / 61.83
fra_Latn → dan_Latn	38.63 / 52.54	45.71 / 62.85	43.09 / 60.21	32.49 / 49.26	45.8 / 63.17	44.86 / 61.78	**46.45 / 64.11**
dan_Latn → fra_Latn	45.65 / 58.51	51.66 / 67.39	50.42 / 66.33	33.15 / 52.75	49.1 / 64.69	**54.77 / 70.89**	53.71 / 70.1
dan_Latn → hun_Latn	26.77 / 35.83	35.3 / 51.64	34.85 / 50.5	28.19 / 47.74	34.01 / 49.78	36.76 / 53.05	**36.88 / 53.65**
hun_Latn → dan_Latn	29.22 / 39.5	36.68 / 52.83	38.03 / 54.8	30.29 / 51.67	38.15 / 55.39	**40.93 / 58.44**	39.79 / 57.0
dan_Latn → kaz_Cyrl	8.43 / 11.42	19.49 / 30.18	**20.82 / 32.5**	12.18 / 21.39	16.62 / 25.85	15.74 / 25.24	18.66 / 28.81
kaz_Cyrl → dan_Latn	19.39 / 25.24	32.19 / 48.07	**34.19 / 51.4**	23.41 / 40.26	25.0 / 37.47	2.95 / 8.04	31.47 / 45.46
deu_Latn → eng_Latn	46.06 / 62.29	62.75 / 77.23	62.34 / 77.04	32.02 / 48.79	61.99 / 76.8	58.83 / 73.72	**64.3 / 79.14**
eng_Latn → deu_Latn	39.55 / 51.19	52.3 / 67.39	52.59 / 67.98	31.37 / 51.21	53.6 / 68.94	53.28 / 68.52	**54.22 / 69.83**
ita_Latn → deu_Latn	41.9 / 58.51	42.81 / 60.38	40.82 / 57.7	28.46 / 46.73	44.22 / 62.33	**45.48** / 63.66	45.35 / **64.11**
deu_Latn → ita_Latn	41.02 / 55.78	42.93 / 60.0	42.11 / 58.97	31.95 / 53.55	43.8 / 61.16	45.72 / 63.3	**46.1 / 63.83**
ell_Grek → eng_Latn	23.26 / 37.64	55.73 / 72.13	55.92 / 72.34	34.0 / 47.13	57.06 / 73.74	54.3 / 68.72	**58.02 / 73.86**
eng_Latn → ell_Grek	12.7 / 14.56	27.81 / 39.79	29.24 / 42.19	25.66 / **45.56**	29.14 / 41.69	28.06 / 40.89	**30.06** / 43.13
ell_Grek → heb_Hebr	16.98 / 23.72	25.67 / 42.96	26.83 / 44.44	16.94 / 35.13	25.79 / 41.39	**28.09 / 44.98**	23.46 / 40.44
heb_Hebr → ell_Grek	8.7 / 9.27	19.92 / 29.38	**22.06** / 33.01	18.63 / **35.12**	20.53 / 30.39	3.12 / 6.28	21.27 / 32.28
ell_Grek → ita_Latn	38.98 / 55.35	39.28 / 56.52	40.5 / 57.97	29.99 / 50.52	43.21 / 61.42	38.61 / 55.84	**44.24 / 63.09**
ita_Latn → ell_Grek	12.27 / 17.0	18.97 / 30.25	18.73 / 29.18	**24.41 / 44.68**	21.42 / 33.87	11.9 / 21.78	22.89 / 36.43
jpn_Jpan → ell_Grek	4.21 / 2.36	15.64 / 22.48	**16.89 / 25.16**	0.33 / 0.65	14.17 / 20.03	2.66 / 3.27	10.94 / 15.49
ell_Grek → jpn_Jpan	6.04 / 13.16	14.85 / 31.92	**15.68 / 32.49**	7.92 / 10.9	12.29 / 23.74	13.03 / 27.32	15.14 / 31.31
kor_Hang → ell_Grek	4.73 / 3.12	14.33 / 19.41	15.58 / 22.14	**19.15 / 33.56**	10.72 / 13.52	9.08 / 12.24	14.21 / 20.81
ell_Grek → kor_Hang	6.54 / 13.18	14.12 / 30.52	**14.21 / 30.91**	9.61 / 22.49	11.22 / 22.07	5.69 / 15.04	13.87 / 29.34
fin_Latn → eng_Latn	42.19 / 58.24	52.75 / 69.18	53.14 / 69.64	32.95 / 53.25	53.24 / 69.8	34.24 / 48.15	**54.3 / 71.08**
eng_Latn → fin_Latn	27.63 / 34.87	37.07 / 52.21	37.83 / 52.62	29.81 / 48.57	38.47 / 54.71	38.83 / 54.77	**40.86 / 57.23**
fra_Latn → eng_Latn	43.2 / 58.6	63.39 / 77.47	63.76 / 78.08	39.68 / 52.35	64.85 / 79.19	64.69 / 78.0	**65.72 / 79.54**
eng_Latn → fra_Latn	51.68 / 62.52	59.91 / 73.55	59.93 / 73.71	37.35 / 57.06	61.34 / 74.8	62.23 / 75.69	**62.99 / 76.64**
eng_Latn → gle_Latn	14.27 / 19.14	24.32 / 38.0	**26.09 / 40.52**	22.18 / 38.52	22.7 / 34.98	16.17 / 27.36	24.58 / 38.9
gle_Latn → eng_Latn	27.17 / 40.82	48.46 / 65.35	**51.2 / 67.96**	25.17 / 43.09	44.67 / 60.71	20.45 / 31.68	37.37 / 52.88
eng_Latn → heb_Hebr	18.45 / 26.97	36.2 / 54.7	39.19 / 58.54	17.36 / 37.22	33.71 / 51.38	37.34 / 56.35	**40.14 / 59.77**
heb_Hebr → eng_Latn	21.76 / 37.09	56.66 / 71.34	59.08 / **74.25**	30.19 / 51.05	55.33 / 69.83	31.92 / 43.09	**59.13** / 74.14
hin_Deva → eng_Latn	23.2 / 37.95	52.73 / 67.88	54.45 / 70.0	3.13 / 8.78	49.58 / 64.59	19.77 / 31.08	**55.91 / 71.47**
eng_Latn → hin_Deva	16.8 / 23.99	29.95 / 45.36	31.58 / 47.9	1.57 / 6.37	29.46 / 44.96	2.63 / 8.75	**32.23 / 49.17**
eng_Latn → hrv_Latn	31.08 / 44.01	39.77 / 57.57	41.74 / 59.46	30.47 / 51.29	43.02 / 61.38	42.4 / 60.22	**44.05 / 62.61**
hrv_Latn → eng_Latn	42.32 / 58.8	55.64 / 71.1	56.96 / 72.74	34.33 / 54.9	57.31 / 73.07	40.28 / 56.13	**59.15 / 75.09**
eng_Latn → hun_Latn	26.77 / 34.49	37.72 / 53.52	38.7 / 54.32	30.68 / 50.86	38.55 / 54.11	**40.35 / 57.13**	40.05 / 56.35
hun_Latn → eng_Latn	41.41 / 57.45	52.47 / 68.14	54.27 / 70.42	19.38 / 32.65	51.51 / 67.24	30.88 / 44.72	**55.35 / 71.68**
eng_Latn → hye_Armn	8.97 / 12.98	28.5 / 41.06	**31.19 / 44.55**	1.12 / 4.91	28.04 / 40.47	1.32 / 5.36	25.98 / 37.2
hye_Armn → eng_Latn	25.45 / 39.37	50.0 / 65.54	21.99 / 11.2	26.14 / 42.84	48.85 / 64.97	26.83 / 38.75	**50.65 / 66.51**
ind_Latn → eng_Latn	51.47 / 66.27	60.25 / 74.59	61.19 / 75.73	34.56 / 57.51	61.28 / 75.4	29.19 / 42.97	**62.85 / 77.29**
eng_Latn → ind_Latn	49.14 / 61.83	55.98 / 70.83	58.07 / 72.74	37.9 / 58.35	57.93 / 72.51	56.83 / 71.94	**59.53 / 73.84**
ita_Latn → eng_Latn	40.58 / 55.48	55.0 / 71.55	56.42 / 73.1	36.3 / 49.79	56.39 / 73.63	56.05 / 70.51	**58.79 / 73.91**
eng_Latn → ita_Latn	43.25 / 56.64	47.79 / 64.02	48.28 / 64.7	34.5 / 54.77	50.43 / 66.91	50.25 / 66.57	**51.58 / 68.34**
eng_Latn → jav_Latn	16.14 / 23.16	29.45 / 43.84	**32.83 / 48.79**	29.02 / 48.62	29.07 / 43.72	29.95 / 46.39	32.08 / 48.01
jav_Latn → eng_Latn	25.11 / 39.1	48.47 / 64.24	**51.08 / 66.74**	27.22 / 48.06	45.79 / 61.66	21.09 / 35.26	49.22 / 65.48
eng_Latn → jpn_Jpan	11.24 / 20.64	20.91 / 41.63	20.82 / 41.49	1.26 / 1.59	18.63 / 38.66	1.88 / 1.99	**22.03 / 43.12**
jpn_Jpan → eng_Latn	24.41 / 39.17	44.43 / 61.08	**46.73 / 63.53**	0.56 / 1.4	29.78 / 39.29	31.36 / 41.16	37.9 / 50.74
kat_Geor → eng_Latn	20.0 / 33.1	41.05 / 56.53	2.86 / 0.0	16.02 / 25.68	**41.69 / 58.75**	24.54 / 36.65	38.43 / 53.9
eng_Latn → kat_Geor	8.58 / 10.94	22.33 / 31.58	**25.53 / 36.41**	5.61 / 14.13	23.61 / 33.83	23.0 / 32.78	25.14 / 35.81
eng_Latn → kaz_Cyrl	4.85 / 7.76	19.68 / 29.05	**23.52 / 35.62**	12.78 / 23.19	18.84 / 28.38	2.38 / 6.56	20.76 / 31.22
kaz_Cyrl → eng_Latn	20.7 / 34.21	41.95 / 57.33	**45.17 / 61.58**	22.08 / 38.75	39.6 / 55.0	17.51 / 28.81	44.45 / 59.75
khk_Cyrl → eng_Latn	14.31 / 23.12	32.27 / 47.71	**34.9 / 51.44**	17.92 / 31.63	24.97 / 37.71	12.09 / 22.5	30.69 / 45.07
eng_Latn → khk_Cyrl	3.09 / 4.61	13.89 / 22.91	**16.2 / 26.58**	7.77 / 15.09	12.2 / 19.17	1.85 / 5.6	9.78 / 16.74

	Baselines (Zero-shot + Regular ICL)			Ours (Unsupervised ICL)			
Pair	***Zero*-Shot**	**k-Shot**	**BM25**	**UW2W**	**Random**	**TopK**	**TopK+BM25**
eng_Latn → kir_Cyrl	5.19 / 7.67	16.72 / 26.27	**19.01 / 29.51**	13.5 / 24.93	14.95 / 22.84	3.09 / 7.79	5.24 / 11.45
kir_Cyrl → eng_Latn	16.73 / 27.39	35.68 / 51.44	**37.56 / 54.16**	20.68 / 36.8	34.27 / 50.78	15.07 / 26.15	36.68 / 52.47
eng_Latn → kor_Hang	10.82 / 22.32	19.8 / 39.77	20.56 / 40.47	11.74 / 28.94	18.9 / 37.68	20.14 / 40.94	**20.68 / 41.23**
kor_Hang → eng_Latn	23.34 / 37.08	44.59 / 60.61	47.36 / 63.73	30.11 / 49.51	45.52 / 61.94	30.48 / 41.19	**50.35 / 67.74**
eng_Latn → lit_Latn	21.23 / 28.48	30.65 / 45.92	32.66 / 48.24	28.51 / 45.5	32.65 / 48.69	32.78 / 48.09	**34.52 / 51.41**
lit_Latn → eng_Latn	39.2 / 56.03	50.36 / 66.35	50.99 / 67.15	31.03 / 50.98	48.58 / 64.09	21.08 / 33.64	**52.67 / 69.25**
lvs_Latn → eng_Latn	39.64 / 57.08	52.22 / 68.81	52.93 / 69.79	31.52 / 52.02	49.99 / 65.99	27.55 / 42.03	**54.38 / 71.5**
eng_Latn → lvs_Latn	19.96 / 27.19	30.49 / 45.37	32.98 / 48.68	28.21 / 49.05	31.76 / 46.91	32.33 / 47.45	**33.43 / 49.66**
nld_Latn → eng_Latn	46.53 / 63.24	54.07 / 70.75	53.2 / 69.83	39.12 / 56.95	54.92 / 71.94	**56.72 / 73.91**	56.23 / 73.36
eng_Latn → nld_Latn	39.99 / 55.75	47.41 / 65.1	45.79 / 63.08	36.6 / 57.46	48.89 / 66.98	49.41 / 67.3	**49.49 / 67.86**
eng_Latn → nso_Latn	13.37 / 19.08	19.79 / 34.11	**24.13 / 41.58**	15.58 / 24.73	16.91 / 27.58	17.16 / 26.88	17.42 / 27.46
nso_Latn → eng_Latn	13.49 / 20.47	23.87 / 37.15	**26.13 / 39.39**	17.67 / 27.03	22.75 / 35.32	18.58 / 27.38	19.52 / 30.67
eng_Latn → oci_Latn	22.64 / 26.96	41.23 / 58.37	**45.44 / 62.22**	34.77 / 53.25	41.99 / 59.61	44.22 / 61.83	44.5 / 61.91
oci_Latn → eng_Latn	46.13 / 60.48	67.58 / 80.42	70.16 / **82.58**	38.72 / 54.38	68.45 / 81.43	51.61 / 63.83	**70.46** / 82.13
eng_Latn → pbt_Arab	2.21 / 3.89	8.98 / 15.99	11.62 / 23.06	**15.66 / 33.01**	8.23 / 14.29	9.96 / 18.13	10.21 / 19.81
pbt_Arab → eng_Latn	16.64 / 27.15	38.3 / 53.71	**42.74 / 59.66**	16.79 / 27.01	31.53 / 43.39	18.92 / 30.93	34.25 / 49.72
pol_Latn → eng_Latn	42.69 / 60.5	51.07 / 67.86	51.58 / 68.55	31.49 / 51.92	52.44 / 69.93	22.07 / 33.18	**54.79 / 72.54**
eng_Latn → pol_Latn	32.2 / 44.19	39.26 / 56.66	38.35 / 55.65	27.43 / 47.12	40.47 / 58.14	40.33 / 58.33	**41.65 / 59.88**
eng_Latn → por_Latn	51.49 / 64.35	61.82 / 75.86	62.32 / 76.43	40.67 / 62.63	64.42 / 77.98	64.43 / 78.19	**64.91 / 78.73**
por_Latn → eng_Latn	64.14 / 77.85	66.7 / 79.98	67.47 / 80.67	41.42 / 61.4	67.89 / 80.9	68.47 / 81.64	**69.16 / 82.17**
rus_Cyrl → eng_Latn	38.87 / 55.15	53.02 / 68.93	55.43 / 71.53	20.58 / 34.09	55.35 / 72.09	55.81 / 71.81	**58.18 / 74.38**
eng_Latn → rus_Cyrl	33.18 / 41.95	41.77 / 56.81	44.81 / 60.76	26.46 / 45.98	43.15 / 58.54	44.52 / 59.64	**47.3 / 63.47**
eng_Latn → slk_Latn	24.47 / 33.49	35.12 / 50.21	36.19 / 51.81	28.0 / 48.55	37.91 / 54.47	38.05 / 53.9	**39.04 / 55.58**
slk_Latn → eng_Latn	45.43 / 62.28	56.72 / 72.38	57.2 / 72.65	35.08 / 55.43	58.59 / 74.69	25.83 / 39.43	**59.62 / 75.89**
eng_Latn → snd_Arab	2.13 / 4.84	9.82 / 17.57	**14.22** / 26.31	11.3 / **27.48**	9.91 / 18.41	10.46 / 19.25	10.92 / 20.62
snd_Arab → eng_Latn	15.32 / 25.31	40.91 / 57.46	**44.94 / 61.91**	23.57 / 41.07	36.5 / 50.95	14.01 / 24.53	41.95 / 59.49
eng_Latn → som_Latn	8.87 / 9.02	16.35 / 26.31	**19.31 / 32.1**	16.69 / 29.64	14.97 / 22.35	14.16 / 23.03	14.17 / 23.02
som_Latn → eng_Latn	14.75 / 22.3	23.34 / 35.44	**26.5 / 40.37**	18.91 / 33.25	20.65 / 31.07	16.4 / 26.12	19.89 / 34.23
eng_Latn → tgk_Cyrl	3.23 / 6.15	14.73 / 21.67	**19.57 / 29.7**	12.05 / 22.13	15.12 / 23.07	2.43 / 6.72	2.75 / 7.35
tgk_Cyrl → eng_Latn	17.71 / 30.77	40.38 / 56.78	**44.78 / 62.8**	16.05 / 25.3	38.82 / 55.39	34.49 / 51.54	29.38 / 45.01
eng_Latn → tha_Thai	8.98 / 10.04	24.42 / 38.59	**28.28 / 43.7**	14.92 / 25.9	25.9 / 40.21	23.55 / 35.09	28.05 / 43.19
tha_Thai → eng_Latn	21.25 / 35.46	47.71 / 63.6	49.36 / 66.01	1.72 / 4.62	48.15 / 65.47	36.5 / 50.96	**51.19 / 69.1**
tur_Latn → eng_Latn	38.14 / 53.09	50.87 / 66.6	53.88 / 69.78	33.02 / 54.3	47.01 / 62.64	42.96 / 58.85	**54.61 / 70.78**
eng_Latn → tur_Latn	24.34 / 31.36	34.09 / 46.78	36.45 / 50.22	29.98 / 49.24	33.66 / 46.39	35.13 / 49.2	**36.46 / 50.71**
eng_Latn → vie_Latn	39.13 / 53.53	46.82 / 65.04	47.3 / 65.96	26.06 / 48.13	47.06 / 65.58	**50.08 / 68.78**	48.59 / 67.14
vie_Latn → eng_Latn	35.35 / 52.2	52.11 / 67.73	54.51 / 70.54	24.67 / 37.92	53.96 / 69.68	49.16 / 65.83	**55.74 / 72.21**
eng_Latn → yor_Latn	8.22 / 11.02	10.52 / 18.91	**13.38 / 26.23**	11.64 / 20.46	11.99 / 23.06	12.96 / 22.53	12.99 / 22.63
yor_Latn → eng_Latn	13.34 / 20.18	21.91 / 33.52	**25.23** / 39.29	16.06 / 29.26	16.03 / 28.69	14.88 / 27.16	24.44 / **40.54**
eng_Latn → zsm_Latn	42.35 / 53.66	53.34 / 68.85	55.41 / 70.88	38.22 / 57.41	54.52 / 69.82	54.96 / 70.55	**55.44 / 71.2**
zsm_Latn → eng_Latn	38.07 / 54.11	60.05 / 73.96	62.0 / 75.83	34.25 / 57.25	57.49 / 70.17	56.27 / 71.11	**62.67 / 77.07**
fin_Latn → hun_Latn	24.62 / 32.99	31.91 / 47.49	32.52 / 48.18	26.99 / 44.59	30.87 / 46.54	34.44 / 50.99	**35.11 / 51.82**
hun_Latn → fin_Latn	23.65 / 30.89	29.46 / 43.33	30.88 / 45.97	28.44 / 47.13	30.77 / 45.6	**33.19 / 49.04**	32.88 / 48.79
fin_Latn → kir_Cyrl	6.56 / 8.6	16.15 / 26.69	17.15 / 28.27	12.1 / 21.0	15.63 / 26.18	16.41 / 26.43	**17.59 / 29.16**
kir_Cyrl → fin_Latn	15.29 / 18.19	24.81 / 36.86	**25.61** / 38.28	22.38 / **38.82**	21.42 / 31.0	5.83 / 11.65	17.93 / 28.61
snd_Arab → fin_Latn	13.7 / 15.37	24.26 / 36.68	**26.73 / 40.92**	20.12 / 34.34	19.7 / 28.51	12.27 / 19.93	21.06 / 32.14
fin_Latn → snd_Arab	2.44 / 3.69	9.05 / 16.53	**11.79 / 22.81**	2.3 / 7.89	9.88 / 18.56	10.0 / 19.53	10.43 / 20.86
tur_Latn → fin_Latn	19.61 / 23.8	27.17 / 39.87	27.67 / 40.8	27.33 / **46.42**	22.09 / 30.71	27.57 / 40.01	**27.97** / 41.24
fin_Latn → tur_Latn	18.94 / 26.18	27.75 / 41.1	26.67 / 39.22	27.03 / **46.68**	23.58 / 33.22	27.35 / 40.79	**28.29** / 42.11
vie_Latn → fin_Latn	19.37 / 23.21	28.08 / 40.78	29.92 / 43.86	22.99 / 32.59	27.16 / 40.18	11.84 / 20.83	**31.85 / 46.91**
fin_Latn → vie_Latn	27.38 / 39.23	36.66 / 54.54	36.22 / 55.9	26.62 / 48.13	34.79 / 53.48	**38.33 / 57.56**	36.44 / 55.87
glg_Latn → hun_Latn	25.18 / 34.58	31.83 / 46.17	32.91 / 48.31	27.82 / 47.89	32.27 / 47.3	35.93 / 52.16	**35.97 / 52.44**
hun_Latn → glg_Latn	30.14 / 42.84	34.82 / 51.13	37.14 / 54.22	29.34 / 49.54	36.76 / 53.8	**40.63 / 58.83**	38.57 / 56.61
glg_Latn → oci_Latn	29.1 / 39.01	35.06 / 52.64	37.54 / 55.18	36.79 / 57.23	38.99 / 58.04	41.74 / 61.2	**42.04 / 61.42**
oci_Latn → glg_Latn	43.37 / 60.65	44.07 / 62.44	44.92 / 63.03	38.97 / 58.79	48.48 / 67.46	47.43 / 67.11	**49.11 / 68.04**
heb_Hebr → rus_Cyrl	31.05 / 41.77	37.25 / 52.84	39.51 / 56.26	24.4 / 42.31	38.78 / 54.89	39.76 / 55.73	**40.37 / 56.63**
rus_Cyrl → heb_Hebr	20.39 / 29.77	28.58 / 46.0	30.81 / 49.96	16.25 / 35.36	29.18 / 46.84	31.69 / 51.14	**33.55 / 54.25**
hin_Deva → hye_Armn	13.21 / 12.04	24.74 / 37.07	**25.66 / 38.52**	1.9 / 7.93	22.18 / 32.64	16.66 / 24.89	24.49 / 36.08
hye_Armn → hin_Deva	8.35 / 10.66	21.33 / 34.43	2.17 / 0.0	3.63 / 12.9	21.08 / 33.38	21.97 / 35.93	**23.37 / 38.23**
hin_Deva → kat_Geor	13.05 / 12.37	19.78 / 28.73	**21.01** / 29.11	2.24 / 7.05	19.65 / **30.36**	10.36 / 17.33	16.63 / 25.39
kat_Geor → hin_Deva	9.95 / 11.78	18.87 / 30.76	0.33 / 0.0	6.92 / 15.28	**22.46 / 37.52**	1.89 / 6.34	13.72 / 24.82
mar_Deva → hin_Deva	21.88 / 36.93	29.17 / 46.42	**30.2 / 48.39**	19.18 / 36.42	19.45 / 37.01	19.45 / 37.01	20.07 / 38.03
hin_Deva → mar_Deva	15.14 / 27.02	14.15 / 22.28	17.6 / 29.01	18.87 / 36.91	18.96 / 37.02	18.98 / 37.05	**19.06 / 37.16**
hin_Deva → mlt_Latn	10.23 / 12.49	18.07 / 28.46	**21.74 / 35.94**	2.55 / 7.8	12.67 / 17.79	10.91 / 16.08	15.3 / 24.47
mlt_Latn → hin_Deva	10.54 / 14.78	22.67 / 36.75	**23.54 / 38.44**	1.51 / 5.5	11.29 / 18.58	2.1 / 6.85	2.73 / 8.52
hin_Deva → nso_Latn	9.87 / 15.62	16.08 / 28.0	**20.12 / 37.06**	2.52 / 7.38	12.7 / 24.94	3.79 / 9.79	16.05 / 31.1
nso_Latn → hin_Deva	6.0 / 8.37	12.86 / 23.93	**14.36 / 26.59**	5.97 / 16.12	11.2 / 20.32	11.82 / 23.87	12.6 / 22.34

	Baselines (Zero-shot + Regular ICL)			Ours (Unsupervised ICL)			
Pair	***Zero*-Shot**	**k-Shot**	**BM25**	**UW2W**	**Random**	**TopK**	**TopK+BM25**
hin_Deva → slk_Latn	15.91 / 21.19	24.42 / 37.72	25.34 / **39.91**	3.47 / 10.37	19.56 / 29.0	16.99 / 26.07	**25.53** / 39.1
slk_Latn → hin_Deva	10.18 / 13.68	20.86 / 34.67	**22.42 / 37.12**	1.39 / 5.76	19.41 / 31.42	1.48 / 6.02	22.05 / 36.71
hrv_Latn → ita_Latn	37.85 / 51.31	40.49 / 56.71	40.42 / 57.01	29.4 / 48.42	42.62 / 59.92	**44.32** / 61.48	44.05 / **61.7**
ita_Latn → hrv_Latn	30.57 / 45.57	34.87 / 53.28	33.2 / 50.71	28.01 / 43.73	37.04 / 56.52	37.67 / 57.16	**38.96 / 58.97**
hrv_Latn → pol_Latn	31.66 / 44.18	35.29 / 51.88	35.09 / 52.23	29.49 / 49.46	38.3 / 56.04	39.28 / 57.66	**39.61 / 57.91**
pol_Latn → hrv_Latn	28.68 / 42.18	33.39 / 51.74	32.79 / 50.23	30.0 / 50.05	37.4 / 57.15	37.93 / 57.55	**38.54 / 58.87**
kat_Geor → hun_Latn	22.06 / 31.68	26.32 / 38.83	0.62 / 0.0	22.78 / 39.42	26.88 / 40.09	3.61 / 8.35	**28.51 / 43.45**
hun_Latn → kat_Geor	10.05 / 10.86	21.37 / 31.34	22.93 / 33.14	2.23 / 5.82	19.4 / 28.19	2.34 / 6.21	**23.56 / 34.06**
tur_Latn → hun_Latn	23.2 / 30.45	31.08 / 45.74	30.26 / 44.61	27.46 / 45.85	29.57 / 43.38	**33.55 / 49.69**	32.43 / 47.78
hun_Latn → tur_Latn	20.95 / 28.16	30.16 / 43.46	28.29 / 41.03	28.08 / **47.24**	26.01 / 38.84	**30.2** / 44.09	29.97 / 44.05
hun_Latn → uzn_Latn	11.1 / 13.43	23.24 / 35.81	**23.8** / 36.57	23.39 / **40.0**	21.51 / 33.57	21.62 / 33.82	23.26 / 35.57
uzn_Latn → hun_Latn	21.12 / 28.99	27.35 / 40.94	**28.95 / 43.34**	23.44 / 39.78	26.37 / 40.6	14.18 / 23.37	21.88 / 35.86
hun_Latn → vie_Latn	29.14 / 41.39	35.38 / 52.34	36.49 / 54.78	25.93 / 48.58	35.05 / 52.42	**39.0 / 58.66**	37.03 / 55.89
vie_Latn → hun_Latn	19.85 / 24.58	29.48 / 43.65	29.58 / 43.94	21.5 / 33.92	29.39 / 44.0	28.48 / 41.59	**30.79 / 45.59**
hye_Armn → mar_Deva	2.15 / 2.72	8.23 / 14.24	0.67 / 0.0	1.37 / 6.21	7.81 / 10.92	9.01 / 14.92	**9.56 / 15.17**
mar_Deva → hye_Armn	10.17 / 7.41	21.19 / **33.23**	**21.71** / 31.82	0.99 / 4.82	17.22 / 26.14	12.92 / 19.7	21.1 / 31.72
ind_Latn → zsm_Latn	48.33 / 65.67	50.45 / 66.73	50.09 / 66.88	21.39 / 35.28	48.18 / 66.55	47.99 / 66.35	**50.99 / 68.52**
zsm_Latn → ind_Latn	50.9 / 68.8	52.22 / 68.37	52.59 / 68.58	46.97 / 63.93	**54.53 / 71.23**	51.82 / 68.8	54.44 / 70.96
ita_Latn → nso_Latn	12.83 / 20.76	17.63 / 31.62	**21.24 / 38.2**	15.02 / 25.07	16.28 / 28.25	15.93 / 29.28	16.29 / 30.48
nso_Latn → ita_Latn	14.03 / 18.24	21.05 / 33.47	**21.58** / 33.59	18.81 / **35.63**	18.79 / 28.26	15.71 / 29.45	20.89 / 34.38
jav_Latn → uzn_Latn	11.68 / 14.94	**21.38 / 33.35**	21.04 / 31.91	6.24 / 12.21	17.36 / 26.36	15.58 / 25.8	15.59 / 25.8
uzn_Latn → jav_Latn	11.89 / 17.65	19.51 / 30.01	**22.48 / 36.2**	20.88 / 36.16	17.95 / 27.72	16.64 / 26.36	20.04 / 33.6
jav_Latn → zsm_Latn	35.81 / 52.86	42.8 / 59.54	**44.53 / 62.0**	17.92 / 30.29	43.13 / 60.04	36.73 / 54.54	41.8 / 59.76
zsm_Latn → jav_Latn	17.9 / 26.29	30.77 / 47.42	34.1 / 51.7	31.73 / 52.11	33.75 / 52.56	34.07 / 52.92	**35.13 / 54.43**
kat_Geor → khk_Cyrl	3.2 / 4.09	**8.86** / 12.78	0.04 / 0.0	8.7 / **17.19**	4.39 / 8.98	1.16 / 4.16	2.96 / 7.08
khk_Cyrl → kat_Geor	7.7 / 11.86	**18.26 / 27.19**	9.11 / 3.71	3.86 / 10.02	10.64 / 16.24	2.28 / 6.17	14.32 / 23.55
kor_Hang → kat_Geor	9.85 / 8.0	18.78 / 27.23	**21.08 / 31.12**	1.6 / 3.95	17.29 / 24.48	12.22 / 16.89	13.44 / 18.2
kat_Geor → kor_Hang	5.92 / 12.63	**11.67 / 25.6**	0.9 / 0.0	5.85 / 14.02	9.75 / 19.49	1.48 / 4.92	3.78 / 10.29
ltz_Latn → kat_Geor	8.05 / 5.41	19.94 / 29.17	**23.02 / 33.35**	2.22 / 5.9	20.83 / 31.05	3.1 / 7.87	3.43 / 8.5
kat_Geor → ltz_Latn	10.88 / 13.86	**17.36** / 24.39	0.59 / 0.0	14.76 / 25.83	15.31 / 22.31	3.31 / 7.33	17.24 / **28.49**
mar_Deva → kat_Geor	9.87 / 7.56	18.17 / 26.98	16.49 / 19.38	1.78 / 5.01	9.61 / 14.29	10.14 / 16.73	**19.09 / 28.26**
kat_Geor → mar_Deva	2.72 / 2.51	7.27 / 9.96	0.11 / 0.0	1.86 / 5.51	7.31 / 10.78	9.13 / 13.99	**9.69 / 15.45**
nso_Latn → kat_Geor	7.89 / 9.9	15.93 / 24.32	**16.71** / 24.83	8.22 / 16.71	16.22 / **25.34**	12.99 / 21.97	12.25 / 20.93
kat_Geor → nso_Latn	8.06 / 13.29	12.49 / 18.96	0.26 / 0.0	11.72 / 19.54	**14.44 / 27.31**	2.07 / 5.89	10.87 / 20.28
pes_Arab → kat_Geor	12.83 / 12.89	18.95 / 27.37	**22.16 / 32.4**	1.48 / 4.68	16.79 / 23.52	4.43 / 9.61	8.66 / 14.88
kat_Geor → pes_Arab	7.73 / 8.39	14.55 / 24.69	1.09 / 0.0	**16.71 / 35.46**	13.59 / 21.06	4.07 / 9.54	2.44 / 7.51
kat_Geor → yor_Latn	4.96 / 5.48	7.51 / 10.64	0.1 / 0.0	8.77 / **18.59**	**8.81** / 16.38	3.33 / 7.19	7.64 / 13.99
yor_Latn → kat_Geor	5.81 / 5.84	**15.23 / 22.95**	12.69 / 15.42	6.67 / 14.46	14.26 / 21.97	12.29 / 20.05	13.3 / 19.75
kaz_Cyrl → nso_Latn	10.57 / 17.26	16.15 / 27.88	**19.45 / 35.9**	11.92 / 17.78	11.79 / 17.07	8.52 / 15.08	12.54 / 21.54
nso_Latn → kaz_Cyrl	6.78 / 7.3	13.83 / 21.12	**14.87 / 22.91**	11.2 / 17.53	12.45 / 20.73	4.9 / 10.38	11.82 / 17.75
tur_Latn → kaz_Cyrl	10.28 / 15.0	22.18 / **35.52**	**22.41** / 35.27	13.37 / 24.27	17.03 / 25.09	21.32 / 33.64	20.77 / 32.24
kaz_Cyrl → tur_Latn	21.81 / 30.34	30.54 / 44.56	30.27 / 44.79	25.15 / 43.3	27.14 / 38.53	**31.95 / 47.8**	31.58 / 46.4
kaz_Cyrl → uzn_Latn	10.57 / 17.63	30.51 / 45.17	30.26 / 45.52	20.69 / 37.16	28.86 / 43.29	31.12 / 47.67	**33.75 / 50.54**
uzn_Latn → kaz_Cyrl	12.06 / 19.29	23.15 / 36.35	25.68 / 39.98	19.47 / 36.09	**26.19 / 42.13**	24.04 / 38.38	25.53 / 40.53
khk_Cyrl → srp_Cyrl	9.28 / 16.2	18.71 / 32.0	**19.18** / 32.27	17.1 / **32.56**	15.07 / 23.65	11.69 / 19.91	12.94 / 22.83
srp_Cyrl → khk_Cyrl	5.9 / 6.46	14.6 / 25.58	**16.68 / 29.7**	5.59 / 11.38	10.89 / 16.12	11.37 / 19.82	11.8 / 20.87
khk_Cyrl → ukr_Cyrl	9.84 / 15.03	19.56 / 30.91	**20.98 / 33.39**	15.88 / 29.32	16.19 / 23.39	12.92 / 22.76	15.01 / 24.48
ukr_Cyrl → khk_Cyrl	6.76 / 8.89	14.81 / 24.97	**16.34 / 28.0**	11.88 / 20.19	14.3 / 24.62	12.07 / 21.25	12.26 / 21.72
kir_Cyrl → nso_Latn	9.73 / 15.26	16.04 / 28.41	**18.84 / 35.37**	12.03 / 18.63	15.0 / 26.58	2.67 / 7.11	8.98 / 15.94
nso_Latn → kir_Cyrl	7.03 / 8.29	13.39 / 21.24	**13.8** / 21.0	11.81 / 19.93	12.09 / 20.75	11.07 / 20.13	12.15 / **22.02**
kir_Cyrl → tgk_Cyrl	11.28 / 17.36	14.56 / 23.03	**16.32** / 26.76	14.93 / 27.22	14.7 / 25.75	14.72 / 26.31	15.93 / **28.56**
tgk_Cyrl → kir_Cyrl	13.18 / 22.73	16.33 / 27.45	**17.89 / 30.68**	17.53 / 30.42	15.03 / 24.83	16.18 / 28.24	15.9 / 27.87
tur_Latn → kir_Cyrl	8.87 / 12.81	18.77 / 31.72	**19.37** / 31.57	12.7 / 22.5	18.27 / 30.48	18.95 / 31.86	19.19 / **32.09**
kir_Cyrl → tur_Latn	20.24 / 30.34	26.81 / 40.18	26.58 / 40.03	22.15 / 39.27	25.82 / 38.63	11.48 / 20.85	**27.46 / 42.39**
kor_Hang → nso_Latn	8.66 / 11.53	14.59 / 23.9	**18.7 / 33.61**	11.81 / 16.53	11.94 / 17.65	5.41 / 8.99	12.32 / 21.43
nso_Latn → kor_Hang	3.58 / 7.99	7.73 / 17.42	**7.96** / 17.17	5.67 / 12.15	7.25 / 16.03	5.32 / 13.44	7.29 / **17.52**
yor_Latn → kor_Hang	2.93 / 6.76	6.75 / 14.48	**7.18 / 16.67**	5.1 / 11.82	5.76 / 11.82	6.17 / 14.73	6.89 / 15.56
kor_Hang → yor_Latn	5.49 / 6.35	8.13 / 12.63	**10.36 / 20.01**	9.34 / 14.47	6.84 / 9.9	3.86 / 6.66	6.01 / 9.41
rus_Cyrl → ltz_Latn	17.02 / 23.19	22.3 / 33.62	24.5 / 37.62	21.43 / 37.9	22.79 / 34.99	**24.72 / 38.52**	24.49 / 38.45
ltz_Latn → rus_Cyrl	20.94 / 22.4	33.51 / 48.14	**37.84 / 53.76**	20.46 / 35.94	29.62 / 42.41	32.15 / 46.29	35.86 / 51.83
spa_Latn → lvs_Latn	16.5 / 21.6	23.52 / 36.03	26.52 / 41.54	25.1 / 40.76	27.59 / 43.21	26.43 / 40.98	**29.17 / 45.68**
lvs_Latn → spa_Latn	37.96 / 54.81	38.49 / 55.78	38.75 / 56.77	26.07 / 43.83	**40.56** / 58.83	39.9 / 58.36	40.54 / **59.62**
rus_Cyrl → mkd_Cyrl	31.84 / 48.08	34.62 / 52.88	36.84 / 55.19	31.15 / 52.85	40.63 / 61.18	41.0 / 61.57	**42.24 / 63.03**
mkd_Cyrl → rus_Cyrl	32.6 / 47.58	39.17 / 54.95	40.53 / 57.13	27.26 / 47.8	42.64 / 59.38	43.91 / 61.68	**44.22 / 61.88**
snd_Arab → mkd_Cyrl	14.76 / 21.99	22.83 / 36.03	**26.71 / 43.49**	21.85 / 40.87	14.48 / 19.69	19.61 / 30.86	24.23 / 39.79
mkd_Cyrl → snd_Arab	2.76 / 3.13	8.75 / 15.68	**12.43 / 24.38**	1.22 / 4.7	8.78 / 14.58	9.72 / 18.48	10.66 / 21.36

	Baselines (Zero-shot + Regular ICL)			Ours (Unsupervised ICL)			
Pair	***Zero*-Shot**	**k-Shot**	**BM25**	**UW2W**	**Random**	**TopK**	**TopK+BM25**
som_Latn → mlt_Latn	10.81 / 14.48	13.77 / 20.44	**16.66** / 26.68	16.25 / **29.29**	12.34 / 17.7	13.93 / 24.09	14.12 / 21.68
mlt_Latn → som_Latn	8.76 / 8.67	15.0 / 23.77	**18.02 / 30.89**	13.1 / 22.55	14.62 / 25.66	13.73 / 24.07	13.88 / 24.35
mlt_Latn → uzn_Latn	11.22 / 13.02	20.36 / 32.02	**22.1 / 35.01**	17.25 / 30.95	17.4 / 26.31	17.02 / 27.46	17.65 / 25.79
uzn_Latn → mlt_Latn	9.76 / 11.49	15.55 / 23.65	**19.13 / 31.48**	17.97 / 30.98	14.01 / 19.89	14.9 / 24.33	15.56 / 26.07
nso_Latn → tur_Latn	12.06 / 16.74	15.2 / 21.23	**15.75** / 22.59	15.15 / **24.51**	13.76 / 19.54	14.37 / 22.28	14.4 / 23.08
tur_Latn → nso_Latn	10.56 / 14.78	17.12 / 29.89	**20.57 / 37.28**	14.03 / 21.82	16.21 / 28.84	13.77 / 22.41	14.05 / 23.31
yor_Latn → nso_Latn	11.23 / 18.22	16.66 / 28.23	**18.68 / 33.2**	11.23 / 21.68	15.5 / 28.9	11.15 / 21.5	14.9 / 27.56
nso_Latn → yor_Latn	8.18 / 13.42	11.15 / 21.92	11.78 / **23.27**	11.15 / 19.74	11.91 / 21.2	12.03 / 21.08	**12.05** / 21.27
spa_Latn → oci_Latn	24.67 / 33.37	33.29 / 51.5	33.61 / 50.55	34.98 / 56.89	37.95 / 59.17	38.99 / 60.05	**39.46 / 60.73**
oci_Latn → spa_Latn	44.94 / 62.11	46.31 / 63.71	46.1 / 63.42	36.8 / 58.54	49.06 / 67.34	49.11 / 67.14	**49.24 / 67.63**
pbt_Arab → pes_Arab	10.81 / 20.29	17.07 / 29.12	**20.63** / 35.95	17.65 / **39.2**	13.87 / 25.18	13.95 / 30.98	19.34 / 37.3
pes_Arab → pbt_Arab	6.22 / 9.48	10.61 / 18.79	12.29 / 24.43	**14.59 / 33.18**	10.36 / 18.74	14.18 / 30.65	13.98 / 29.93
por_Latn → pbt_Arab	3.39 / 4.38	9.16 / 16.31	**11.8 / 23.81**	1.94 / 6.51	7.76 / 13.19	2.55 / 7.88	5.03 / 13.47
pbt_Arab → por_Latn	21.42 / 26.94	30.61 / 45.46	**33.5 / 50.31**	23.95 / 42.49	25.2 / 35.77	26.89 / 41.71	31.83 / 48.19
snd_Arab → pbt_Arab	5.83 / 10.15	10.73 / 21.21	11.96 / 25.16	**13.41 / 29.81**	7.89 / 12.23	10.95 / 24.82	11.67 / 25.96
pbt_Arab → snd_Arab	6.7 / 13.09	9.81 / 18.75	11.92 / 24.16	**14.23 / 31.5**	8.73 / 14.52	13.99 / 29.92	12.45 / 25.72
pol_Latn → urd_Arab	3.07 / 3.19	13.1 / 23.77	**14.01 / 26.23**	0.68 / 5.03	9.38 / 15.78	0.49 / 3.88	10.22 / 20.22
urd_Arab → pol_Latn	15.81 / 20.1	26.56 / 41.16	**28.21 / 44.24**	20.53 / 34.96	21.68 / 32.17	7.15 / 14.57	20.96 / 33.84
rus_Cyrl → por_Latn	44.94 / 61.51	44.55 / 61.79	45.43 / 63.1	29.69 / 49.86	48.11 / 66.34	**50.42 / 68.49**	49.94 / 68.21
por_Latn → rus_Cyrl	35.74 / 47.07	36.86 / 51.8	39.88 / 56.17	25.94 / 45.33	42.57 / 59.1	43.12 / 59.53	**44.08 / 61.1**
rus_Cyrl → spa_Latn	41.69 / 59.11	40.73 / 57.96	39.89 / 57.4	26.79 / 46.2	43.03 / 61.38	44.16 / 62.94	**44.48 / 63.32**
spa_Latn → rus_Cyrl	29.26 / 39.38	34.97 / 50.71	35.68 / 52.52	24.52 / 40.63	39.17 / 57.14	40.46 / 58.89	**40.89 / 59.5**
slk_Latn → zsm_Latn	33.01 / 47.09	37.57 / 53.99	39.93 / 56.84	30.66 / 51.97	38.8 / 55.58	40.89 / 58.58	**41.61 / 59.9**
zsm_Latn → slk_Latn	19.22 / 25.65	28.29 / 43.59	27.63 / 42.06	24.1 / 42.77	28.04 / 42.62	29.85 / **46.13**	**30.1** / 45.44
snd_Arab → tur_Latn	14.56 / 18.55	24.25 / 36.58	**25.05 / 38.15**	18.71 / 35.44	15.89 / 21.89	12.69 / 20.16	17.68 / 26.8
tur_Latn → snd_Arab	3.24 / 4.44	10.95 / 20.8	**12.59 / 23.56**	1.43 / 5.39	9.04 / 15.72	11.48 / 23.15	11.31 / 22.01
snd_Arab → urd_Arab	16.87 / 28.31	23.6 / 41.12	23.55 / 41.69	20.04 / 42.33	26.1 / 45.1	11.73 / 29.16	**27.79 / 49.29**
urd_Arab → snd_Arab	10.55 / 19.73	16.0 / 29.24	18.23 / 34.58	19.69 / 40.97	20.29 / 40.32	**23.81 / 45.43**	23.47 / 44.43
uzn_Latn → snd_Arab	2.57 / 4.16	9.77 / 18.26	**12.17 / 23.42**	1.74 / 5.93	6.77 / 9.71	1.4 / 4.62	1.89 / 5.98
snd_Arab → uzn_Latn	7.24 / 6.05	21.31 / 33.23	**22.96 / 36.16**	16.14 / 30.59	17.42 / 26.3	6.98 / 12.89	12.94 / 20.56
yor_Latn → spa_Latn	15.75 / 22.42	21.32 / 33.91	**22.19 / 35.83**	13.69 / 26.57	18.97 / 29.24	12.94 / 23.71	15.49 / 27.22
spa_Latn → yor_Latn	6.18 / 7.72	9.78 / 18.59	**12.05 / 24.06**	10.72 / 18.08	11.06 / 22.74	10.97 / 19.22	10.99 / 19.22
Average	19.07 / 26.65	28.53 / 41.69	**28.89 / 42.19**	18.39 / 32.03	27.51 / 39.98	23.12 / 34.88	28.3 / 41.75

Table 9: Detailed evaluation of translation scores for 222 language directions from the Flores-200 dataset using the Bloom model. Results include Zero-Shot and regular k-shot baselines, along with our unsupervised method employing various selection approaches. Scores for each language pair are presented as chrF++/spBLEU, with the highest scores highlighted in bold. The 'average' row indicates the mean score across all language pairs.

	Baselines (Zero-shot + Regular ICL)		**Ours (Unsupervised ICL)**			
Pair	***Zero*-Shot**	**k-Shot**	**UW2W**	**Random**	**TopK**	**TopK+BM25**
afr_Latn → pol_Latn	10.25 / 12.93	11.42 / 15.29	**17.79 / 29.78**	11.54 / 18.14	14.71 / 24.81	14.17 / 23.61
pol_Latn → afr_Latn	11.67 / 16.45	12.49 / 18.24	**16.67 / 29.78**	13.87 / 23.35	15.02 / 25.52	14.70 / 25.46
eng_Latn → amh_Ethi	0.21 / 1.06	2.83 / 7.35	0.87 / 4.43	2.11 / 5.34	1.03 / 4.86	1.06 / 4.89
amh_Ethi → eng_Latn	0.77 / 0.00	12.14 / 18.55	10.55 / 17.84	11.40 / 17.58	11.72 / 19.91	12.61 / 22.11
mlt_Latn → amh_Ethi	0.38 / 2.55	3.43 / 9.38	0.79 / 3.80	1.16 / 4.53	0.85 / 3.86	0.85 / 3.85
amh_Ethi → mlt_Latn	4.17 / 2.84	8.88 / 12.18	7.67 / 11.78	**10.06** / 15.76	1.54 / 4.70	4.80 / 9.27
tha_Thai → amh_Ethi	0.19 / 1.55	0.28 / 1.19	**0.83 / 2.76**	0.18 / 0.46	0.17 / 0.83	0.26 / 1.46
amh_Ethi → tha_Thai	2.22 / 0.57	4.15 / 7.19	6.59 / 8.71	5.08 / 8.10	2.29 / 3.96	6.52 / 10.71
ckb_Arab → arb_Arab	4.28 / 4.74	10.19 / 17.14	**12.75 / 26.67**	10.65 / 18.70	7.62 / 14.68	9.55 / 18.67
arb_Arab → ckb_Arab	1.49 / 1.58	**8.71** / 15.00	7.30 / 14.17	8.60 / **15.26**	6.98 / 12.98	7.00 / 13.05
eng_Latn → arb_Arab	2.21 / 2.54	26.99 / 42.68	24.69 / 45.91	25.65 / 40.15	**31.81 / 49.96**	30.22 / 47.56
arb_Arab → eng_Latn	18.75 / 9.10	41.18 / 53.60	28.86 / 46.12	40.92 / 54.41	26.59 / 38.37	**46.12 / 60.94**
heb_Hebr → arb_Arab	3.12 / 4.78	11.38 / 19.96	**12.82 / 26.62**	10.81 / 19.02	12.51 / 25.43	12.76 / 25.94
arb_Arab → heb_Hebr	2.10 / 5.35	8.38 / 15.59	**11.17 / 25.14**	8.31 / 15.16	9.18 / 19.57	8.48 / 16.86
ben_Beng → asm_Beng	9.15 / 17.59	9.71 / 17.41	17.47 / 37.60	18.19 / 37.57	18.08 / 38.68	**18.70 / 39.39**
asm_Beng → ben_Beng	4.76 / 11.12	16.63 / 29.61	17.81 / 38.70	**26.82 / 47.41**	20.78 / 41.85	22.26 / 43.27
asm_Beng → eng_Latn	21.50 / 25.77	29.72 / 44.22	2.03 / 7.64	20.91 / 33.78	9.19 / 18.68	15.05 / 27.15
eng_Latn → asm_Beng	0.40 / 0.36	4.23 / 8.60	**6.05 / 16.65**	4.22 / 7.91	3.91 / 8.04	3.97 / 8.18
asm_Beng → hun_Latn	7.21 / 7.44	**10.90 / 16.25**	1.58 / 6.11	4.47 / 9.01	1.90 / 6.00	3.91 / 8.60
hun_Latn → asm_Beng	1.25 / 3.17	**3.81 / 8.33**	0.96 / 4.89	3.58 / 7.79	3.00 / 7.84	2.75 / 7.01
som_Latn → asm_Beng	0.45 / 1.32	**4.29** / 8.24	0.94 / 4.47	4.16 / **8.25**	2.96 / 7.75	2.88 / 7.24
asm_Beng → som_Latn	5.57 / 5.04	**11.36 / 18.36**	1.38 / 5.41	1.73 / 5.84	1.73 / 5.70	2.06 / 6.26
eng_Latn → ast_Latn	13.26 / 11.23	31.94 / 49.99	31.19 / 52.32	31.61 / 49.73	31.81 / 49.85	**34.65 / 52.81**
ast_Latn → eng_Latn	9.32 / 0.20	49.71 / 65.04	33.05 / 53.84	45.78 / 58.45	**53.04 / 69.12**	52.77 / 68.70
tgk_Cyrl → ast_Latn	12.09 / 18.53	12.68 / 19.43	13.71 / 21.97	12.89 / 20.46	14.31 / 26.52	14.49 / **26.91**
ast_Latn → tgk_Cyrl	1.37 / 2.13	7.93 / 13.48	**10.08 / 18.76**	6.82 / 11.72	6.41 / 11.77	7.54 / 13.01
bel_Cyrl → eng_Latn	2.73 / 0.01	15.87 / 26.74	16.05 / 26.95	15.76 / 27.43	12.93 / 24.36	**18.11** / 33.02
eng_Latn → bel_Cyrl	1.72 / 0.98	6.73 / 10.25	**11.13 / 20.55**	5.75 / 8.02	7.06 / 11.89	6.98 / 12.20
bel_Cyrl → hrv_Latn	10.50 / 15.92	11.42 / 17.24	**16.29 / 27.24**	12.55 / 21.48	4.67 / 9.55	12.50 / 22.70
hrv_Latn → bel_Cyrl	3.59 / 4.86	4.67 / 7.12	**11.04 / 20.15**	4.92 / 8.39	6.34 / 11.67	6.84 / 11.49
bel_Cyrl → nso_Latn	7.46 / 9.13	11.59 / 18.96	11.15 / 18.94	12.53 / 23.44	**15.67 / 31.66**	13.84 / 26.97
nso_Latn → bel_Cyrl	3.07 / 2.12	5.91 / 8.32	**9.78 / 19.62**	5.56 / 9.72	4.48 / 7.44	5.80 / 9.69
tgk_Cyrl → bel_Cyrl	9.30 / 15.35	7.44 / 12.10	**12.39 / 22.49**	11.68 / 21.29	11.82 / 21.52	11.79 / 21.58
bel_Cyrl → tgk_Cyrl	4.15 / 7.03	7.92 / 12.96	4.86 / 10.72	11.81 / 21.57	**11.94** / 21.53	11.90 / **21.92**
eng_Latn → ben_Beng	3.03 / 1.83	19.67 / 33.80	19.38 / **38.77**	20.45 / 36.14	19.19 / 32.73	**20.69** / 36.13
ben_Beng → eng_Latn	25.13 / 26.46	36.51 / 51.88	2.05 / 7.91	19.57 / 31.66	11.12 / 20.99	15.67 / 27.03
eng_Latn → bul_Cyrl	2.69 / 2.08	12.51 / 22.40	**17.75 / 33.67**	12.74 / 22.76	12.76 / 22.31	12.88 / 22.07
bul_Cyrl → eng_Latn	8.15 / 1.45	26.43 / 41.54	22.24 / 39.62	23.11 / 35.47	14.61 / 26.50	**29.89 / 48.71**
cat_Latn → ckb_Arab	0.58 / 0.64	**8.82 / 14.88**	0.95 / 2.68	7.57 / 11.63	4.90 / 11.44	8.50 / 14.40
ckb_Arab → cat_Latn	7.98 / 6.77	11.79 / 16.61	13.32 / **25.66**	11.62 / 16.32	11.49 / 16.27	12.55 / 19.21
cat_Latn → dan_Latn	13.89 / 20.85	19.96 / 33.73	**23.08 / 38.27**	20.27 / 33.79	21.80 / 37.99	21.11 / 36.50
dan_Latn → cat_Latn	22.47 / 33.43	26.64 / 40.45	23.92 / 43.54	26.26 / 40.62	24.85 / 39.72	**28.20 / 44.72**
fin_Latn → cat_Latn	14.29 / 21.58	14.90 / 22.47	15.88 / 26.64	14.88 / 23.09	16.31 / 27.39	15.78 / 25.84
cat_Latn → fin_Latn	8.99 / 9.33	12.11 / 16.47	**16.53 / 25.07**	12.00 / 16.34	14.62 / 23.55	13.85 / 21.68
ceb_Latn → eng_Latn	16.06 / 17.56	24.34 / 36.64	22.18 / 33.91	25.26 / 41.05	21.24 / 32.09	**25.56 / 42.61**
eng_Latn → ceb_Latn	8.19 / 5.75	**18.97** / 30.27	16.27 / 28.05	18.31 / 29.39	18.82 / 30.39	18.85 / **30.44**
zsm_Latn → ceb_Latn	13.12 / 17.32	15.46 / 24.03	17.18 / 30.41	18.09 / 29.90	**19.80 / 32.85**	19.29 / 31.89
ceb_Latn → zsm_Latn	13.38 / 18.99	19.68 / 31.28	18.77 / 31.21	16.43 / 25.53	**20.59** / 33.21	20.45 / **33.66**
ckb_Arab → eng_Latn	1.35 / 0.00	11.84 / 16.59	13.79 / 23.54	11.81 / 17.47	15.02 / **26.17**	13.07 / 20.21
eng_Latn → ckb_Arab	0.47 / 0.32	**9.08** / 15.03	6.34 / 11.75	8.44 / 13.54	8.99 / 14.76	8.49 / 14.22
ckb_Arab → nso_Latn	7.34 / 8.57	10.80 / 17.13	**11.88 / 22.96**	10.53 / 18.34	8.97 / 14.62	9.82 / 14.96
nso_Latn → ckb_Arab	0.82 / 1.29	**9.23 / 17.08**	5.16 / 7.79	8.61 / 16.25	7.40 / 13.01	8.22 / 16.06
cym_Latn → fra_Latn	12.26 / 16.58	16.08 / 26.52	16.32 / 29.55	14.47 / 23.23	14.04 / 24.20	15.13 / 26.39
fra_Latn → cym_Latn	10.83 / 15.09	11.24 / 16.86	**15.14** / 24.95	11.71 / 18.59	14.54 / 25.27	14.31 / **25.62**
dan_Latn → fra_Latn	20.68 / 30.21	31.05 / 47.48	24.44 / 44.24	27.35 / 42.16	22.10 / 35.96	**31.27 / 49.16**
fra_Latn → dan_Latn	14.25 / 21.87	20.65 / 35.74	**24.23 / 38.41**	20.58 / 34.93	20.35 / 34.57	21.88 / 37.87
hun_Latn → dan_Latn	11.98 / 18.12	11.80 / 16.92	**17.54 / 31.42**	12.92 / 20.52	15.68 / 26.20	15.58 / 26.45
dan_Latn → hun_Latn	8.37 / 9.58	11.46 / 16.63	**16.77 / 29.99**	12.47 / 19.73	15.22 / 25.77	15.07 / 25.40

	Baselines (Zero-shot + Regular ICL)		Ours (Unsupervised ICL)			
Pair	***Zero*-Shot**	**k-Shot**	**UW2W**	**Random**	**TopK**	**TopK+BM25**
ita_Latn → deu_Latn	15.24 / 22.99	17.07 / 25.86	**21.51 / 38.10**	16.29 / 24.16	20.23 / 32.49	20.13 / 32.07
deu_Latn → ita_Latn	16.27 / 22.62	20.65 / 32.47	**26.37 / 46.54**	20.59 / 32.44	22.99 / 36.93	23.18 / 37.21
ell_Grek → eng_Latn	2.97 / 0.01	19.52 / 33.18	20.60 / 33.50	18.84 / 32.95	19.97 / 35.96	**21.76 / 38.53**
eng_Latn → ell_Grek	2.09 / 1.88	8.94 / 16.02	**14.67 / 28.58**	8.76 / 15.33	9.05 / 17.09	9.20 / 17.27
heb_Hebr → ell_Grek	1.51 / 1.57	**6.67 / 8.22**	2.70 / 6.59	2.99 / 4.58	1.47 / 2.57	1.71 / 3.34
ell_Grek → heb_Hebr	1.85 / 4.85	7.66 / 13.02	**12.34 / 22.66**	7.73 / 13.69	7.18 / 11.46	6.90 / 12.91
kor_Hang → ell_Grek	0.88 / 1.00	4.44 / 6.08	**9.59 / 15.85**	3.68 / 4.62	3.07 / 4.98	3.26 / 5.35
ell_Grek → kor_Hang	1.69 / 4.41	2.76 / 6.13	2.37 / 4.34	1.37 / 1.98	2.08 / 4.37	2.17 / 4.70
eng_Latn → fin_Latn	5.97 / 2.89	12.30 / 16.66	**18.33 / 31.30**	11.15 / 14.18	14.98 / 23.16	14.24 / 21.60
fin_Latn → eng_Latn	6.93 / 2.66	15.86 / 23.57	17.37 / 30.21	15.15 / 23.38	16.17 / 23.91	16.30 / 25.36
fra_Latn → eng_Latn	31.30 / 26.17	55.63 / 69.99	39.01 / 53.19	54.34 / 67.65	56.57 / 69.54	**59.79 / 74.83**
eng_Latn → fra_Latn	27.53 / 26.44	53.29 / 67.54	37.09 / 56.79	52.41 / 66.20	52.76 / 64.96	**56.42 / 70.34**
heb_Hebr → eng_Latn	4.09 / 0.24	14.97 / 23.90	**16.31 / 30.33**	13.30 / 20.64	13.73 / 25.12	14.51 / 25.59
eng_Latn → heb_Hebr	2.00 / 3.10	8.03 / 15.47	**13.84 / 30.66**	8.23 / 15.46	9.26 / 18.88	8.71 / 18.23
hin_Deva → eng_Latn	27.36 / 27.16	41.53 / 56.82	4.35 / 12.97	29.16 / 42.76	13.57 / 24.23	**44.77 / 61.48**
eng_Latn → hin_Deva	5.16 / 4.69	27.00 / 45.09	24.14 / 45.83	26.53 / 44.89	28.18 / 46.58	**29.00 / 47.82**
hun_Latn → eng_Latn	11.14 / 11.46	15.93 / 23.93	**19.92 / 34.36**	16.02 / 24.57	15.54 / 25.39	17.65 / 30.85
eng_Latn → hun_Latn	7.56 / 6.55	11.92 / 17.07	**18.93 / 33.94**	11.76 / 16.34	13.11 / 19.87	12.82 / 19.40
eng_Latn → hye_Armn	0.45 / 0.84	4.98 / 7.80	1.10 / 4.90	4.84 / 7.72	4.21 / 8.86	3.74 / 7.77
hye_Armn → eng_Latn	1.02 / 0.00	12.63 / 19.84	12.37 / 17.76	12.12 / 19.01	13.34 / 23.14	14.73 / **27.20**
ita_Latn → eng_Latn	25.56 / 22.97	47.32 / 65.26	34.39 / 50.75	45.28 / 62.67	46.38 / 63.46	**50.96 / 67.49**
eng_Latn → ita_Latn	13.27 / 12.09	25.19 / 38.01	31.02 / **50.52**	28.63 / 43.41	27.51 / 39.29	**31.05** / 46.69
eng_Latn → jav_Latn	9.14 / 6.70	16.15 / 25.20	**24.13 / 42.16**	15.39 / 24.59	17.76 / 31.18	17.05 / 29.23
jav_Latn → eng_Latn	14.43 / 11.42	27.98 / 41.73	24.92 / 42.51	24.55 / 36.87	26.07 / 42.11	**29.08 / 45.69**
jpn_Jpan → eng_Latn	7.62 / 0.65	**29.33 / 45.05**	0.56 / 1.40	6.74 / 9.50	3.54 / 5.33	4.84 / 7.19
eng_Latn → jpn_Jpan	1.24 / 2.25	6.71 / **16.24**	4.69 / 7.47	5.97 / 13.50	**7.53** / 16.02	6.70 / 15.81
eng_Latn → kat_Geor	0.72 / 0.32	9.89 / 18.60	2.09 / 5.49	9.24 / 17.56	7.86 / 14.73	8.70 / 16.29
kat_Geor → eng_Latn	0.59 / 0.00	12.13 / 18.00	14.42 / 22.21	11.67 / 17.06	5.56 / 12.66	12.34 / 21.31
eng_Latn → kaz_Cyrl	1.23 / 0.61	7.68 / 9.97	**11.31 / 20.29**	7.03 / 8.81	8.07 / 11.97	8.01 / 12.17
kaz_Cyrl → eng_Latn	2.55 / 0.01	13.07 / 18.85	13.95 / 23.31	13.15 / 20.15	3.18 / 8.64	15.17 / 26.50
khk_Cyrl → eng_Latn	3.01 / 0.06	12.57 / 18.72	13.70 / 22.45	12.76 / 20.03	5.08 / 11.85	13.07 / 21.36
eng_Latn → khk_Cyrl	1.20 / 0.32	8.18 / 11.52	9.99 / 18.13	7.36 / 10.12	7.56 / 10.11	8.38 / 12.67
kor_Hang → eng_Latn	4.78 / 0.23	17.42 / 27.79	19.85 / **35.86**	17.01 / 27.28	12.65 / 24.10	**20.43** / 35.63
eng_Latn → kor_Hang	1.41 / 2.29	3.53 / 9.41	**6.32 / 16.35**	3.05 / 7.88	3.60 / 9.22	3.67 / 9.78
lit_Latn → eng_Latn	5.69 / 0.88	16.64 / 25.46	**19.86 / 34.96**	16.43 / 26.62	15.62 / 25.82	16.78 / 28.63
eng_Latn → lit_Latn	6.67 / 4.68	11.89 / 16.79	**18.84 / 33.30**	11.97 / 16.92	15.49 / 25.50	14.44 / 23.13
eng_Latn → lvs_Latn	5.42 / 2.85	10.83 / 14.96	**18.34 / 33.49**	11.58 / 16.77	13.65 / 22.72	12.62 / 19.25
lvs_Latn → eng_Latn	7.01 / 2.58	17.03 / 26.95	**19.37 / 34.01**	16.23 / 25.85	14.93 / 25.07	15.64 / 27.36
eng_Latn → nld_Latn	11.58 / 11.33	23.58 / 38.53	**28.77 / 48.71**	23.90 / 39.37	24.00 / 38.76	25.46 / 41.23
nld_Latn → eng_Latn	21.29 / 20.81	36.01 / 53.94	31.69 / 49.97	36.79 / 55.41	36.78 / 55.63	**38.90 / 58.22**
eng_Latn → nso_Latn	4.18 / 0.49	12.73 / 19.66	15.56 / 24.68	15.21 / 24.79	**17.14** / 26.85	**17.14 / 26.88**
nso_Latn → eng_Latn	9.99 / 8.79	16.20 / 23.25	17.83 / 25.61	18.02 / 30.60	**18.45** / 27.27	18.14 / 28.16
eng_Latn → pbt_Arab	1.28 / 0.72	7.19 / 12.91	**8.24 / 17.32**	6.13 / 9.56	7.94 / 15.84	7.27 / 14.19
pbt_Arab → eng_Latn	3.66 / 0.09	15.04 / 22.83	18.17 / **30.25**	14.40 / 22.54	14.94 / 22.75	17.65 / 30.13
eng_Latn → pol_Latn	7.18 / 5.05	15.40 / 24.06	**20.58 / 36.37**	15.78 / 25.48	15.19 / 23.24	16.87 / 27.53
pol_Latn → eng_Latn	9.47 / 3.38	**26.04** / 41.08	23.21 / 39.20	24.93 / 39.66	17.41 / 28.06	25.32 / **42.86**
por_Latn → eng_Latn	45.73 / 51.75	59.01 / 72.33	39.79 / 60.33	60.00 / 73.46	62.60 / 76.96	**63.79 / 78.00**
eng_Latn → por_Latn	31.28 / 35.24	52.95 / 66.53	39.07 / 61.24	53.82 / 67.40	54.92 / 68.56	**57.81 / 72.72**
eng_Latn → slk_Latn	7.18 / 5.48	13.66 / 20.96	**20.34 / 37.06**	13.34 / 20.36	13.98 / 21.42	14.71 / 24.55
slk_Latn → eng_Latn	12.20 / 9.25	22.14 / 33.28	22.55 / **40.67**	21.78 / 34.33	17.35 / 29.26	20.98 / 35.67
eng_Latn → snd_Arab	0.83 / 0.52	6.83 / 13.57	**9.78 / 22.67**	6.57 / 12.34	6.98 / 14.19	7.45 / 15.11
snd_Arab → eng_Latn	3.16 / 0.03	14.06 / 20.93	16.40 / 28.64	14.03 / 21.71	14.97 / 22.72	15.29 / 24.57
eng_Latn → som_Latn	5.24 / 2.23	11.36 / 17.64	13.11 / 21.54	13.85 / 22.46	**14.21 / 23.14**	14.20 / 23.13
som_Latn → eng_Latn	5.63 / 1.36	13.84 / 19.13	14.98 / 23.14	14.47 / 24.29	15.28 / 23.75	15.31 / 24.86
eng_Latn → tgk_Cyrl	0.98 / 0.64	7.93 / 11.98	**11.11** / 20.01	7.89 / 11.91	8.51 / 13.94	7.96 / 12.61
tgk_Cyrl → eng_Latn	2.40 / 0.01	12.76 / 18.75	14.31 / 20.53	12.43 / 18.67	4.53 / 10.53	15.29 / 27.67
tha_Thai → eng_Latn	1.55 / 0.00	13.31 / 20.57	1.73 / 4.73	12.82 / 20.30	7.73 / 14.43	10.72 / 19.19
eng_Latn → tha_Thai	1.14 / 0.84	6.63 / 10.78	2.15 / 4.20	5.31 / 6.83	3.26 / 5.17	3.39 / 5.38
tur_Latn → eng_Latn	5.17 / 0.36	17.44 / 27.02	**21.26 / 37.42**	17.10 / 27.20	15.94 / 26.50	17.44 / 29.81
eng_Latn → tur_Latn	7.19 / 5.60	13.18 / 20.14	**18.92 / 34.36**	12.62 / 18.69	13.32 / 19.86	14.31 / 22.44
eng_Latn → vie_Latn	15.62 / 15.24	40.48 / 58.23	29.72 / 54.14	42.08 / 60.99	42.82 / 60.83	**44.17 / 63.28**
vie_Latn → eng_Latn	32.53 / 36.70	45.11 / 60.09	25.21 / 37.73	41.92 / 56.17	44.60 / 58.67	**47.39 / 63.41**
eng_Latn → zsm_Latn	13.00 / 9.17	38.60 / 54.15	33.42 / 54.85	38.56 / 53.62	37.74 / 54.05	**40.24 / 56.52**
zsm_Latn → eng_Latn	20.98 / 11.39	46.98 / 60.53	32.40 / 52.00	45.86 / 59.78	27.47 / 42.02	**52.74 / 68.60**
fin_Latn → hun_Latn	9.47 / 12.59	11.02 / 15.84	14.17 / **23.50**	13.10 / 20.76	**14.20** / 22.95	14.15 / 22.92
hun_Latn → fin_Latn	9.67 / 11.51	10.06 / 11.47	**15.84 / 26.73**	13.76 / 22.13	14.37 / 23.51	14.33 / 23.39

Pair	Baselines (Zero-shot + Regular ICL)		Ours (Unsupervised ICL)			
Pair	***Zero*-Shot**	**k-Shot**	**UW2W**	**Random**	**TopK**	**TopK+BM25**
tur_Latn → fin_Latn	9.71 / 11.68	10.82 / 14.10	**15.37 / 25.66**	13.19 / 20.99	14.24 / 23.47	14.19 / 23.44
fin_Latn → tur_Latn	10.79 / 15.62	10.61 / 15.04	13.91 / 23.57	14.07 / 23.48	**14.48** / 23.48	**14.48 / 23.65**
fin_Latn → vie_Latn	9.26 / 16.95	13.59 / 27.39	12.59 / 28.00	13.25 / 28.83	12.79 / 27.74	**13.61 / 28.93**
vie_Latn → fin_Latn	9.26 / 10.72	11.74 / 16.03	**14.53 / 20.09**	11.49 / 15.63	10.67 / 18.24	12.00 / 19.06
glg_Latn → hun_Latn	8.84 / 10.85	11.05 / 15.89	**15.10 / 26.28**	12.74 / 20.23	14.47 / 24.35	14.28 / 24.23
hun_Latn → glg_Latn	14.33 / 23.12	14.17 / 20.87	**17.49 / 31.29**	14.17 / 23.39	13.96 / 23.73	14.16 / 24.21
oci_Latn → glg_Latn	29.62 / 45.17	35.05 / 54.75	32.86 / 52.36	38.09 / 59.31	37.16 / 58.58	**38.50 / 60.00**
glg_Latn → oci_Latn	24.65 / 40.18	28.95 / 47.20	26.36 / 46.47	30.90 / 50.70	30.81 / 50.80	**32.68 / 52.91**
rus_Cyrl → heb_Hebr	1.36 / 3.66	6.02 / 10.15	2.19 / 5.57	5.34 / 9.55	**6.74 / 13.52**	5.80 / 11.69
heb_Hebr → rus_Cyrl	2.58 / 2.79	4.40 / 6.50	**12.02 / 20.56**	3.10 / 4.25	3.29 / 5.40	3.86 / 5.81
hin_Deva → hye_Armn	0.77 / 1.77	4.34 / 7.36	0.98 / 4.87	3.09 / 8.67	1.18 / 5.03	1.34 / 5.34
hye_Armn → hin_Deva	4.76 / 5.57	**12.51 / 28.54**	2.36 / 9.61	10.98 / 24.03	10.90 / 25.05	11.46 / 25.37
kat_Geor → hin_Deva	4.11 / 4.98	11.10 / **23.39**	9.33 / 19.42	10.22 / 21.40	10.39 / 22.07	10.65 / 22.62
hin_Deva → kat_Geor	1.07 / 1.42	9.42 / 16.46	1.61 / 5.16	9.48 / 16.83	2.68 / 6.73	3.75 / 8.53
mar_Deva → hin_Deva	18.31 / 29.68	**32.25 / 51.61**	19.13 / 36.27	19.37 / 35.37	19.84 / 38.12	20.29 / 38.50
hin_Deva → mar_Deva	14.76 / 27.51	15.11 / 26.21	18.24 / 35.64	20.04 / 38.07	18.93 / 36.94	**22.12 / 38.18**
hin_Deva → mlt_Latn	6.61 / 6.55	**10.92 / 17.31**	2.68 / 7.23	7.70 / 13.08	7.30 / 13.13	8.84 / 14.63
mlt_Latn → hin_Deva	6.47 / 12.16	**13.81** / 28.60	3.08 / 8.72	13.76 / **29.59**	5.75 / 15.07	7.27 / 17.87
nso_Latn → hin_Deva	3.79 / 6.30	**12.61 / 26.10**	10.05 / 22.47	12.35 / 25.90	7.26 / 18.28	5.43 / 14.77
hin_Deva → nso_Latn	5.98 / 4.97	**13.77 / 24.44**	2.64 / 7.65	4.25 / 10.76	5.33 / 12.66	5.48 / 12.54
hrv_Latn → ita_Latn	12.82 / 16.10	16.09 / 24.75	**20.52 / 36.13**	14.70 / 24.02	15.43 / 24.66	16.22 / 25.83
ita_Latn → hrv_Latn	12.35 / 18.69	14.60 / 23.68	**20.65 / 33.50**	13.85 / 21.68	17.61 / 31.44	16.97 / 29.04
pol_Latn → hrv_Latn	13.62 / 21.98	13.75 / 22.54	**19.70 / 34.35**	17.23 / 31.27	18.40 / 31.49	17.93 / 29.22
hrv_Latn → pol_Latn	11.92 / 16.91	13.12 / 19.65	**19.55 / 34.86**	16.84 / 28.93	17.43 / 30.85	17.69 / 30.87
hun_Latn → kat_Geor	1.45 / 2.98	10.54 / 19.47	2.27 / 5.70	7.25 / 12.53	2.33 / 6.19	2.32 / 6.05
kat_Geor → hun_Latn	6.30 / 5.32	10.14 / 14.10	**12.31 / 20.70**	5.10 / 8.94	2.37 / 6.60	4.77 / 9.66
hun_Latn → tur_Latn	10.69 / 15.73	10.05 / 13.39	**15.20 / 26.17**	13.78 / 22.87	14.20 / 22.97	14.11 / 23.11
tur_Latn → hun_Latn	8.28 / 9.13	11.01 / 15.99	**14.55 / 24.55**	13.63 / 22.01	14.01 / 22.92	13.81 / 22.41
vie_Latn → hun_Latn	8.71 / 10.71	11.61 / 16.84	**16.54 / 27.62**	11.68 / 17.35	12.52 / 18.88	11.92 / 17.87
hun_Latn → vie_Latn	10.22 / 19.77	13.83 / 28.33	14.20 / **32.70**	12.88 / 26.92	**14.21** / 30.70	13.77 / 29.56
mar_Deva → hye_Armn	0.65 / 1.62	5.50 / 9.47	0.82 / 4.20	1.82 / 5.35	1.00 / 4.13	1.00 / 4.13
hye_Armn → mar_Deva	3.30 / 4.11	6.96 / 12.08	0.87 / 4.71	6.44 / 11.07	6.35 / 11.15	5.49 / 9.69
ind_Latn → zsm_Latn	39.58 / 53.54	45.95 / 62.76	33.10 / 53.10	48.54 / 66.02	48.35 / 66.48	**49.14 / 66.71**
zsm_Latn → ind_Latn	44.04 / 57.49	49.06 / 64.65	42.87 / 57.99	50.26 / 68.01	49.53 / 67.24	**50.48 / 68.16**
nso_Latn → ita_Latn	9.22 / 8.86	13.63 / 20.54	**18.31 / 32.70**	13.20 / 19.70	15.22 / 28.29	15.17 / 28.19
ita_Latn → nso_Latn	6.66 / 4.96	11.90 / 18.52	14.82 / 25.37	15.41 / 28.91	**15.80** / 29.40	15.73 / **29.43**
zsm_Latn → jav_Latn	23.43 / 38.12	19.21 / 29.88	26.02 / 42.13	23.90 / 39.87	27.76 / 45.62	**28.22 / 46.11**
jav_Latn → zsm_Latn	19.00 / 27.28	26.83 / 41.80	19.85 / 34.44	27.22 / 44.93	25.91 / 43.84	**27.83 / 46.30**
kat_Geor → khk_Cyrl	2.06 / 2.13	10.08 / 19.01	7.90 / 14.87	7.24 / 12.85	1.69 / 5.12	1.70 / 5.21
khk_Cyrl → kat_Geor	1.28 / 2.69	**10.20 / 18.77**	1.80 / 5.07	9.21 / 17.21	1.31 / 3.11	1.75 / 4.99
kor_Hang → kat_Geor	0.94 / 1.10	7.72 / 12.53	1.78 / 4.58	5.33 / 8.05	1.67 / 3.18	1.63 / 3.34
kat_Geor → kor_Hang	1.58 / 3.00	3.14 / 7.37	3.18 / **9.78**	2.17 / 5.69	2.27 / 4.82	2.74 / 6.96
ltz_Latn → kat_Geor	1.86 / 3.40	10.54 / 19.03	2.23 / 6.01	9.59 / 18.24	2.33 / 6.32	2.35 / 6.35
kat_Geor → ltz_Latn	5.89 / 3.20	8.88 / 11.61	**11.64 / 20.03**	9.72 / 13.70	3.44 / 8.18	9.06 / 16.23
mar_Deva → kat_Geor	1.04 / 1.34	9.83 / 18.33	1.65 / 4.56	7.48 / 13.86	2.65 / 6.19	3.11 / 6.88
kat_Geor → mar_Deva	2.49 / 2.16	7.19 / 11.49	**8.86 / 16.89**	6.32 / 9.60	6.62 / 11.10	7.01 / 10.82
kat_Geor → nso_Latn	4.45 / 2.63	10.66 / 17.80	11.26 / 19.54	**13.54 / 26.45**	2.50 / 7.41	10.24 / 19.49
nso_Latn → kat_Geor	0.92 / 1.01	**11.11 / 21.58**	1.99 / 5.66	9.53 / 18.90	2.10 / 5.85	8.45 / 16.74
pes_Arab → kat_Geor	1.25 / 2.59	10.30 / 19.23	1.34 / 4.27	**10.38 / 20.20**	1.50 / 4.10	1.78 / 4.29
kat_Geor → pes_Arab	2.47 / 2.88	7.39 / 13.60	4.64 / 11.47	4.65 / 5.72	5.13 / 12.60	1.67 / 5.40
yor_Latn → kat_Geor	0.89 / 0.99	**11.06 / 21.78**	1.80 / 5.07	5.71 / 11.34	1.97 / 5.20	2.14 / 5.53
kat_Geor → yor_Latn	3.70 / 3.41	5.17 / 7.45	**8.10 / 13.96**	4.59 / 7.44	1.69 / 5.36	1.78 / 5.60
nso_Latn → kaz_Cyrl	3.01 / 2.86	7.42 / 10.52	**10.26 / 16.86**	6.89 / 11.44	6.63 / 10.40	6.52 / 11.69
kaz_Cyrl → nso_Latn	8.12 / 9.95	12.28 / 21.00	11.89 / 18.44	**12.36** / 22.01	8.39 / 16.83	11.89 / **23.02**
tur_Latn → kaz_Cyrl	3.33 / 4.51	7.22 / 11.64	**11.14 / 21.58**	6.75 / 11.38	6.22 / 11.73	7.06 / 12.83
kaz_Cyrl → tur_Latn	9.03 / 12.68	10.85 / 15.51	**14.86 / 27.28**	11.65 / 18.08	10.45 / 18.48	12.88 / 22.20
uzn_Latn → kaz_Cyrl	2.28 / 4.58	9.26 / 16.33	**11.24 / 20.97**	8.34 / 15.94	4.61 / 10.09	8.82 / 17.03
kaz_Cyrl → uzn_Latn	3.44 / 6.96	**14.60 / 25.79**	14.05 / 25.59	14.15 / 23.98	6.52 / 13.26	13.66 / 23.59
nso_Latn → kir_Cyrl	2.50 / 2.91	6.71 / 9.96	**11.32 / 18.84**	3.83 / 4.91	4.97 / 9.27	5.60 / 9.45
kir_Cyrl → nso_Latn	8.00 / 9.91	11.78 / 20.38	12.12 / 19.55	14.26 / 27.06	7.61 / 16.00	**14.84 / 29.22**
tgk_Cyrl → kir_Cyrl	12.10 / 22.65	9.83 / 17.34	14.51 / **26.34**	14.55 / 25.74	**15.06** / 26.06	14.92 / 26.12
kir_Cyrl → tgk_Cyrl	7.09 / 11.72	9.49 / 15.05	4.34 / 9.24	13.71 / 24.31	**14.51 / 25.88**	14.40 / 25.59
tur_Latn → kir_Cyrl	1.99 / 3.77	5.57 / 9.91	**11.37 / 21.27**	5.41 / 9.23	2.98 / 7.03	4.54 / 8.78
kir_Cyrl → tur_Latn	8.30 / 11.18	10.63 / 15.06	**13.62 / 24.68**	11.77 / 19.03	6.75 / 13.64	10.40 / 18.50
kor_Hang → nso_Latn	6.25 / 6.08	13.55 / 24.84	11.74 / 20.84	**13.56 / 26.60**	5.80 / 8.86	11.87 / 22.17
nso_Latn → kor_Hang	1.60 / 3.28	2.86 / 6.68	**4.42 / 9.94**	2.92 / 6.94	2.76 / 7.39	3.11 / 8.14

	Baselines (Zero-shot + Regular ICL)		Ours (Unsupervised ICL)			
Pair	***Zero*-Shot**	**k-Shot**	**UW2W**	**Random**	**TopK**	**TopK+BM25**
yor_Latn → kor_Hang	1.66 / 3.25	2.84 / 6.23	**3.78 / 8.66**	2.08 / 5.45	2.72 / 5.82	2.62 / 5.64
kor_Hang → yor_Latn	5.54 / 7.05	6.72 / 11.90	**8.34 / 14.10**	5.70 / 10.35	2.17 / 4.84	4.12 / 8.04
ltz_Latn → rus_Cyrl	6.55 / 8.83	10.76 / 15.36	**14.06 / 26.39**	11.53 / 18.03	9.90 / 16.13	9.48 / 16.11
rus_Cyrl → ltz_Latn	11.20 / 15.12	11.59 / 18.09	**14.01 / 23.10**	12.18 / 19.63	2.61 / 7.25	10.59 / 17.87
rus_Cyrl → mkd_Cyrl	12.66 / 21.46	13.86 / 23.88	20.81 / 38.84	20.29 / 37.15	**22.59 / 41.82**	22.49 / 41.59
mkd_Cyrl → rus_Cyrl	10.26 / 17.87	12.02 / 19.60	**22.74 / 41.64**	18.94 / 34.62	20.25 / 36.28	20.52 / 36.73
mlt_Latn → som_Latn	8.49 / 11.91	9.15 / 12.26	13.32 / 23.30	13.54 / 23.73	**13.73 / 24.07**	13.69 / 24.02
som_Latn → mlt_Latn	7.95 / 8.93	9.42 / 12.43	**13.97** / 23.84	13.38 / 23.15	13.91 / 24.14	13.90 / **24.28**
tur_Latn → nso_Latn	7.44 / 6.92	11.38 / 18.45	**13.75 / 23.51**	13.37 / 22.58	13.65 / 22.10	13.65 / 22.30
nso_Latn → tur_Latn	7.57 / 7.98	9.96 / 14.26	**14.58 / 22.78**	12.22 / 20.11	14.30 / 22.26	14.18 / 22.36
yor_Latn → nso_Latn	7.29 / 7.86	10.27 / 17.08	**12.52 / 24.12**	10.86 / 20.66	10.71 / 20.59	10.80 / 20.65
nso_Latn → yor_Latn	6.30 / 8.44	6.04 / 9.47	11.14 / 18.11	9.37 / 17.57	**12.04 / 21.13**	11.92 / 20.89
spa_Latn → oci_Latn	20.18 / 30.36	27.81 / 46.36	27.59 / 48.73	28.92 / 48.40	**31.46 / 52.54**	31.41 / 52.09
oci_Latn → spa_Latn	17.36 / 17.38	43.76 / 60.88	35.26 / 56.39	44.41 / 62.35	44.41 / 63.08	**45.55 / 64.41**
pes_Arab → pbt_Arab	4.33 / 10.13	7.98 / 14.25	13.85 / **30.66**	13.92 / 30.18	**14.18** / 30.65	14.15 / 30.63
pbt_Arab → pes_Arab	5.77 / 12.65	6.97 / 13.50	**14.29 / 32.98**	12.32 / 28.14	13.85 / 30.57	13.63 / 30.81
por_Latn → pbt_Arab	1.55 / 1.26	6.25 / 10.49	1.93 / 6.63	6.48 / 10.98	4.27 / 11.55	**6.96 / 16.43**
pbt_Arab → por_Latn	5.08 / 0.92	14.31 / 22.97	18.22 / **32.17**	12.91 / 19.69	12.44 / 17.31	14.88 / 23.77
pbt_Arab → snd_Arab	6.66 / 14.90	7.02 / 13.22	**11.45 / 25.82**	9.89 / 22.02	11.03 / 24.77	10.95 / 24.81
snd_Arab → pbt_Arab	3.64 / 7.88	6.74 / 12.05	**11.63 / 25.41**	10.49 / 24.34	10.66 / 24.58	10.58 / 24.27
pol_Latn → urd_Arab	3.01 / 7.87	**12.89 / 27.40**	8.44 / 21.23	12.73 / 27.05	10.22 / 23.49	11.24 / 25.56
urd_Arab → pol_Latn	9.36 / 11.88	**13.78 / 22.74**	12.12 / 20.91	13.63 / 22.70	11.28 / 17.70	10.77 / 16.51
spa_Latn → rus_Cyrl	4.00 / 6.14	11.21 / 18.28	**20.37 / 36.06**	11.67 / 19.56	10.48 / 18.15	11.19 / 19.09
rus_Cyrl → spa_Latn	10.39 / 3.28	28.07 / 43.94	22.82 / 41.15	23.38 / 35.09	29.45 / 45.96	**30.39 / 47.55**
tur_Latn → snd_Arab	0.87 / 2.05	6.28 / 13.74	2.03 / 7.12	5.31 / 11.35	3.47 / 8.31	6.16 / 13.11
snd_Arab → tur_Latn	8.62 / 11.67	10.42 / 14.65	**12.54 / 23.03**	9.98 / 14.14	11.59 / 18.47	10.92 / 17.28
snd_Arab → urd_Arab	3.46 / 9.04	11.73 / 22.21	**13.24 / 31.44**	11.72 / 26.55	11.78 / 28.55	11.92 / 25.73
urd_Arab → snd_Arab	7.43 / 18.12	6.29 / 12.93	12.28 / **29.51**	6.44 / 13.24	**12.34** / 29.20	12.31 / 29.13
uzn_Latn → snd_Arab	1.27 / 2.99	**5.97 / 12.20**	1.32 / 4.42	5.18 / 10.82	1.40 / 4.61	1.71 / 5.48
snd_Arab → uzn_Latn	5.82 / 6.87	**10.95 / 16.59**	6.53 / 12.73	10.32 / 15.64	1.59 / 4.80	2.44 / 6.15
yor_Latn → spa_Latn	11.01 / 9.43	15.22 / 23.17	15.02 / 29.19	17.49 / 27.79	13.63 / 24.83	16.82 / 28.43
spa_Latn → yor_Latn	4.58 / 3.44	6.15 / 11.05	10.72 / 18.72	7.95 / 15.07	**10.98** / 18.81	10.73 / **19.30**
Average	8.04 / 9.50	14.95 / 23.17	14.21 / 25.33	14.72 / 23.73	13.70 / 23.03	**15.29 / 25.34**

B.2 Analysis

Does the multilingual LLM testbed affect translation quality? Our experiments were conducted using Llama-3 and Bloom. It is important to note that translation performance is highly correlated with the choice of the LLM. Overall, Llama-3 outperforms Bloom in terms of translation spBLEU score. As illustrated in Figure 2, Llama-3 consistently outperforms Bloom across various language pairs, notably in *tur_Latn* $\rightarrow$ *eng_Latn* and *ben_Beng* $\rightarrow$ *eng_Latn*. However, there are exceptions where Bloom slightly surpasses Llama-3, such as in the *kaz_Latn* $\rightarrow$ *nso_Latn* pair.

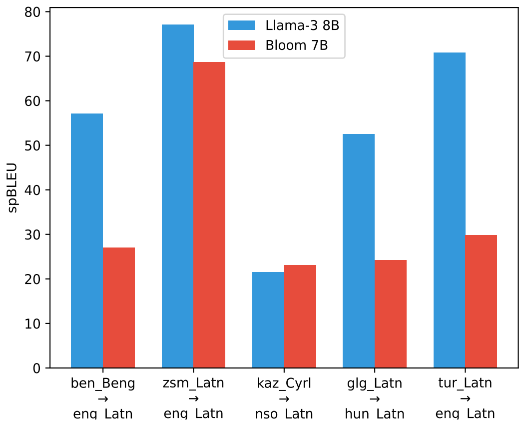

Figure 2: Comparison of the spBLEU score of our unsupervised TopK+BM25 using Llama-3 and Bloom for a randomly selected subset of language pairs.

How does iterative k-shot re-mining improve the quality of UMT? Figure 4 illustrates the spBLEU scores of five randomly selected language pairs from our experiments, where an additional iteration of the sentence-level translation phase described in Section 3.2 was performed. The results of this random subset indicate that, overall, the performance of UMT declines after another iteration of our algorithm. This can be observed notably in pairs such as *kor_Hang* $\rightarrow$ *nso_Latn* and *ltz_Latn* $\rightarrow$ *kat_Geor*, where the spBLEU performance using our UMT approach decreased considerably in the second iteration. Based on these findings, we conclude that applying our approach for a single iteration generally yields the best results.

Does the number of unsupervised ICL examples affect the UMT performance? Figure 5 illustrates the influence of the number of chosen ICL examples on the spBLEU score when employing

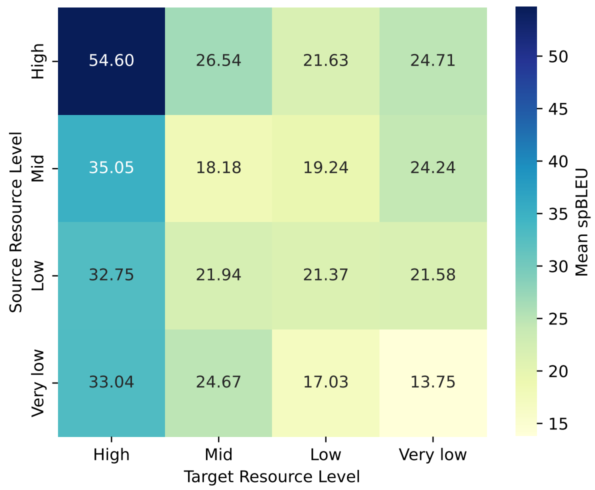

Figure 3: Average spBLEU scores from TopK+BM25 experiments across different resource levels for different language pairs using Bloom. Each cell represents the mean spBLEU score for translations from a source resource level to a target resource level.

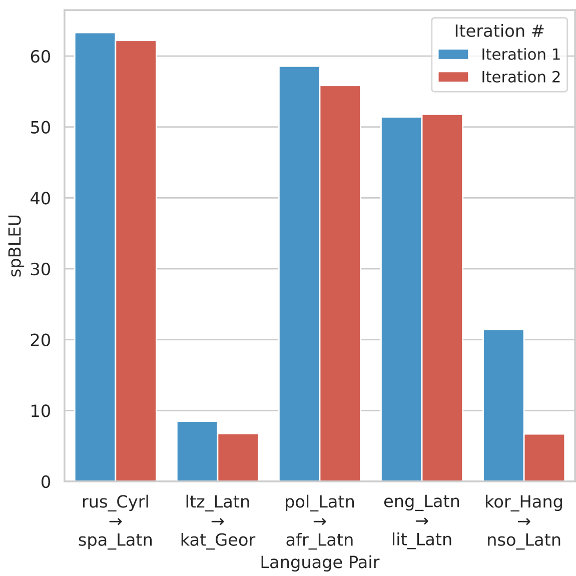

Figure 4: The impact of multiple iterations of our approach using TopK+BM25 on the spBLEU score for a subset of language pairs using Llama-3.

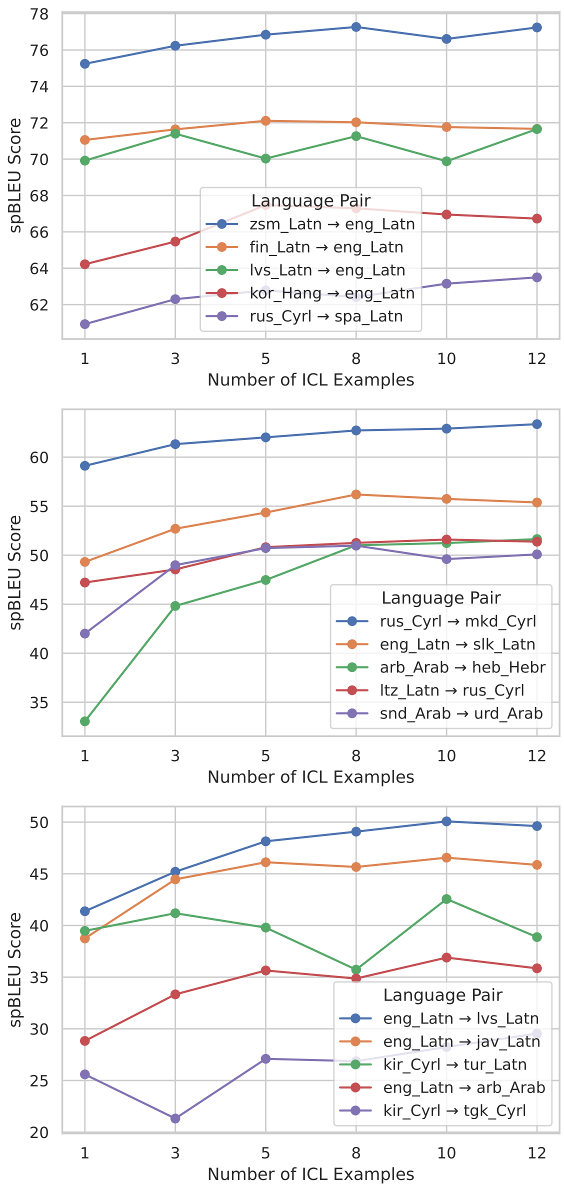

Figure 5: The impact of the quantity of ICL examples on the spBLEU score for UMT employing our TopK+BM25 approach with Llama-3.

the TopK+BM25 selection approach. The figure indicates that selecting a greater number of ICL examples generally enhances performance. This effect is particularly significant between 1 and 8 examples, where the spBLEU score shows significant increases for the majority of pairs. However, the increments in spBLEU score become less substantial between 8 and 12 examples.

Why does the translation performance in TopK ICL decrease compared to Random ICL? Table 1 illustrates a decline in performance when selecting the most confident pairs (*TopK ICL)* compared to *Random ICL*. Our analysis suggests that this decrease occurs because the similarity function prioritizes semantic confidence while disregarding linguistic details, consequently pairing examples from identical languages at times. In contrast, the BM25 selection method integrates both semantic relevance and surface-level linguistic comparisons. This approach results in pairs that not only represent the source and target languages accurately but also maintain semantic similarity.

Does applying BM25 filtering to the most similar sentence pairs enhance the quality of the resulting in-context examples? In our TopK+BM25 results, we applied our BM25 filtering criteria exclusively to sentence pairs that exceed a specific similarity threshold, τ. Figure 6 compares the results between applying BM25 selection on the entire pool of mined sentences (full pool) and applying it only to pairs filtered by the similarity threshold (TopK). The results indicate that using BM25 on the most similar parallel pairs generally yields higher spBLEU scores. This improvement is particularly noticeable in pairs such as *kor_Hand* $\rightarrow$ *eng_Latn* and *asm_Beng* $\rightarrow$ *som_Latn*. Table 12 in Appendix B.2 presents results for 19 language pairs, demonstrating that the TopK+BM25 selection strategy outperforms the application of BM25 on the full pool, with average spBLEU scores of 48.80 and 48.30, respectively.

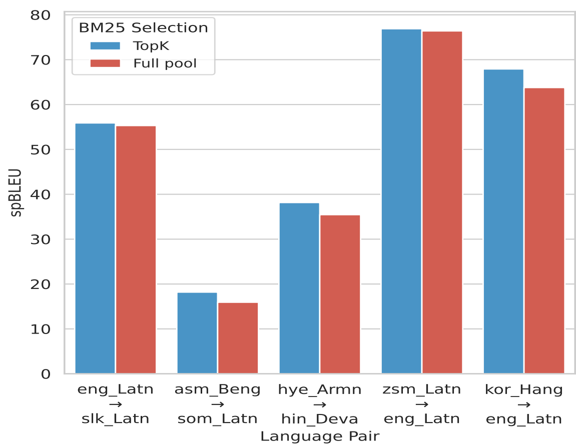

Figure 6: The impact of selecting the unsupervised k-shot from the most confident pairs (TopK) compared to selecting from all pairs (Full pool) on the spBLEU score for a subset of language pairs using Llama-3. (Table 12 in Appendix B.2 presents results for additional pairs.)

What is the impact of the selection threshold τ on the TopK+BM25 performance? Our proposed TopK+BM25 in-context selection for UMT

Lang. Pair	Threshold τ				
	0.0	**0.3**	**0.5**	**0.7**	**0.9**
ita_Latn	67.89	67.93	68.03	68.13	68.34
cat_Latn	71.92	72.14	72.76	72.81	73.95
zsm_Latn	70.67	70.98	71.01	71.50	71.20
bel_Cyrl	44.37	44.65	44.81	45.58	45.39

Table 10: Impact of Threshold τ on our TopK+BM25 UMT spBLEU Scores using Llama-3. These experiments were conducting by having the English language as source and the cited language in the Lang. Pair column as target.

depends on the threshold τ to filter only the most confident semantically similar pairs. We conduct experiments varying the similarity threshold τ to analyze its impact on in-context example quality and translation performance. Our findings across multiple language pairs are presented in Table 10. From this subset, we conclude that the selection of the threshold τ has an impact on the TopK+BM25 spBLEU score, with higher τ values generally improving spBLEU scores, indicating better translation quality.

Does the decoding strategy affect performance? We performed the experiments in Table 1, but using Beam Search instead of Random Sampling. Results are presented in Figure 7 and compare both decoding strategies. Experiments performed using Beam Search are indeed slower than those using Random Sampling; however, the scores resulting from both models are comparable. This led us to choose random sampling in all our experiments.

Examples of translations made using our approach compared to the ground truth references and the regular k-shot baseline. Table 11 presents a sample of translations generated using our TopK+BM25 ICL approach compared to the regular k-shot and our unsupervised word-by-word approach. The examples indicate that our TopK+BM25 ICL approach can generate translations of comparable or superior quality to those produced by the regular ICL. Additionally, the examples demonstrate that while the word-by-word translation approach preserves the overall meaning, it often results in grammatical errors and incorrect word order. We discussed this issue in Section 3.2.1, where we explained the necessity for a back-translation step to mitigate these problems and produce more natural language translations from these word-by-word translations.

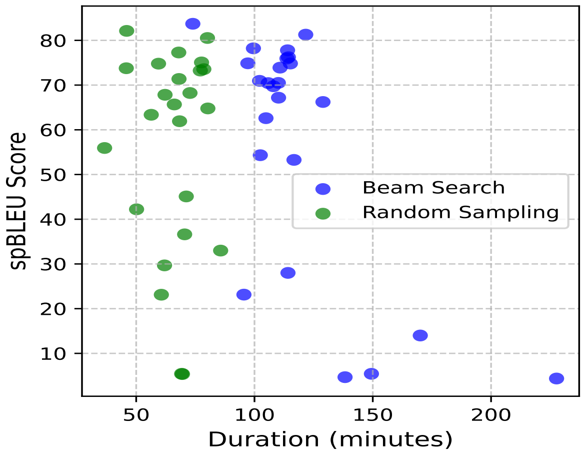

Figure 7: spBLEU scores and durations (in minutes) for the MT performance achieved using TopK+BM25 with Llama-3.

ast_Latn → eng_Latn	**Source**	La corteya tien alredor de 70 km d'espesor nel llau más cercanu y de 100 km nel más llonxanu.
	Reference	The crust is about 70 km thick on the near side and 100 km thick on the far side.
	Supervised k-shot	The crown has a thickness of about 70 km on the closest side and 100 km on the farthest side.
	Unsupervised W2W	the cut ten around of 70 km d'espesor in cry more close and of 100 km in more llonxanu.
	TopK+BM25	The crust has about 70 km of thickness on the near side and 100 km on the far side.
kir_Cyrl → eng_Latn	**Source**	Agencija Xinhua izvijestila je da su vladini istražitelji u srijedu pronašli dva snimača podataka o letu iz „crne kutije".
	Reference	Xinhua reported that government investigators recovered two 'black box' flight recorders on Wednesday.
	Supervised k-shot	Chinese state investigators found two flight recorders in a "black box" on Wednesday, according to the Xinhua news agency.
	Unsupervised W2W	agency xinhua reported yes yes you governmental researchers in wednesday find two cameras data o flight from „crne kutije".
	TopK+BM25	Agency Xinhua reported that government investigators had found two flight recorders from "the black box" on Wednesday.
cat_Latn → eng_Latn	**Source**	Vladajuća stranka, Narodna organizacija jugozapadne Afrike (South West Africa People's Organisation, SWAPO), također je zadržala većinu na parlamentarnim izborima.
	Reference	The ruling party, South West Africa People's Organisation (SWAPO), also retained a majority in the parliamentary elections.
	Supervised k-shot	The ruling party, the South West Africa People's Organisation (SWAPO), also kept its parliamentary majority.
	Unsupervised W2W	ruling stranka, national organization southwestern africa (South west africa People's Organisation, SWAPO), also yes kept most on parliamentary izborima.
	TopK+BM25	The ruling party, the South West Africa People's Organisation (SWAPO), also retained a majority in parliamentary elections.
hrv_Latn → eng_Latn	**Source**	Velika područja dalje na sjeveru vrlo su rijetko naseljena, a nekima gotovo vlada nenaseljena divljina.
	Reference	Large areas further north are quite sparsely populated and some is nearly uninhabited wilderness.
	Supervised k-shot	Large areas further north are sparsely inhabited, and some remain wild.
	Unsupervised W2W	great areas further on north very you rarely naseljena, a some almost government uninhabited divljina.
	TopK+BM25	Large areas further north are extremely sparsely populated and some are virtually wilderness.

Table 11: Examples of translations made by our proposed unsupervised Word-by-word (W2W) and TopK+BM25 ICL methods compared with regular k-shot ICL using Llama-3.

Pair	TopK Pairs	Full pool
eng_Latn → slk_Latn	**55.90**	55.30
kor_Hang → eng_Latn	**67.90**	63.80
eng_Latn → jav_Latn	46.50	**49.20**
cat_Latn → ckb_Arab	19.70	**23.60**
eng_Latn → lvs_Latn	49.90	**50.60**
fin_Latn → eng_Latn	**71.30**	69.70
rus_Cyrl → spa_Latn	**62.80**	62.00
rus_Cyrl → mkd_Cyrl	63.10	**63.40**
arb_Arab → heb_Hebr	**51.20**	47.10
lvs_Latn → eng_Latn	**70.80**	68.70
kir_Cyrl → tgk_Cyrl	**26.90**	23.80
kat_Geor → hun_Latn	31.70	**40.40**
asm_Beng → som_Latn	**18.20**	15.90
snd_Arab → urd_Arab	50.30	**50.40**
ltz_Latn → rus_Cyrl	**52.10**	48.40
hun_Latn → kat_Geor	32.30	**34.70**
hye_Armn → hin_Deva	**38.10**	35.40
kir_Cyrl → tur_Latn	**40.70**	37.90
zsm_Latn → eng_Latn	**76.90**	76.40
Average	**48.80**	48.30

Table 12: spBLEU scores evaluating the performance of our unsupervised sentence-level translation model using LLaMa-3. The scores compare two methods: selecting the top-k parallel sentences with the highest similarity scores between source and target languages using BM25 (TopK), and applying BM25 to the entire pool of parallel sentences (Full Pool).